# FACIAL EXPRESSION ANALYSIS FOR ONLINE LEARNING ENGAGEMENT USING DEEP LEARNING

**K. Rajashekhar Rao[1], Aadhi Vinay[2], Baddam Meghana[3] , A. Santhosh Reddy[4]**

*[1-5] Department of CSE & TKR College of Engineering & Technology*
*[2-5]cB.Tech Students*

---***---

**ABSTRACT**

**This paper presents a Convolutional Neural Network (CNN)-based system for detecting and analyzing student engagement in online learning environments using facial expressions. The system leverages real-time webcam input or static image uploads to classify emotions that are then mapped to engagement levels such as "Engaged," "Not Engaged," or "Distracted." Unlike traditional systems that use gaze tracking or manual observation, our approach automates emotion recognition using CNN trained on facial data. The application is deployed through a Django-based web interface, integrated with OpenCV for real-time face detection, and utilizes SQLite for backend data handling. Results show improved classification accuracy and offer practical utility for instructors to monitor learner attentiveness and participation during virtual sessions.**

*Keywords* — Facial Expression, Engagement Detection, CNN, Online Learning, Emotion Recognition, Real-Time Analysis, Django, OpenCV

## I. INTRODUCTION

In recent years, online education has become increasingly prevalent due to its accessibility and scalability. However, it introduces new challenges in maintaining student engagement, a critical factor influencing learning outcomes. Unlike traditional classrooms where instructors can visually monitor student behavior, virtual environments often lack effective tools to assess attentiveness.

Facial expressions offer a rich and immediate window into a learner's emotional and cognitive states. Emotions such as confusion, boredom, interest, and happiness are strongly correlated with student engagement. Using computer vision and machine learning, we aim to build an automated solution to interpret these expressions and derive engagement insights. While previous systems have relied heavily on gaze tracking or behavioral cues, our system emphasizes facial micro-expressions and convolutional features learned from data.

Student engagement is a crucial predictor of academic success. It encompasses behavioral, emotional, and cognitive involvement in the learning process. Recent advancements in artificial intelligence (AI), particularly in computer vision and deep learning, provide a promising avenue to automate the detection of engagement levels through facial expression analysis. Emotions such as confusion, boredom, interest, and happiness—when recognized accurately—can act as proxies to estimate engagement.

This project aims to develop a real-time, intelligent system that leverages **Convolutional Neural Networks (CNN)** for emotion detection and maps these emotions to corresponding engagement states. The system operates in two modes: a **static** image upload interface and a live webcam-based video feed, both integrated into a Django web application. Using OpenCV, faces are detected from input sources, and emotion classification is performed using a trained CNN model. The recognized emotions—such as *Happy*, *Confused*, *Sad*, *Angry*, *Neutral*—are further interpreted to determine whether the student is engaged or not.

Unlike conventional approaches that rely on manual observation, survey-based assessments, or expensive eye-tracking equipment, our solution is non-intrusive, cost-effective, and scalable. The backend is powered by **SQLite**, which securely stores emotion detection results and user data, while the admin dashboard offers insights into student engagement patterns across sessions.

The primary motivation behind this project is to empower educators with actionable analytics during online sessions and bridge the gap between teaching and learning effectiveness in virtual environments. By analyzing real-time emotional cues, instructors can adapt their teaching strategies, identify struggling students, and foster a more interactive and engaging learning experience.

This system has significant implications for improving virtual education by enabling proactive engagement tracking, thus enhancing both student learning outcomes and instructional quality. In the longer term, it may contribute to building emotionally intelligent e-learning platforms that can personalize content delivery based on learners' engagement levels.

## II LITERATURE SURVEY

Over recent years, numerous studies have explored facial expression analysis and student engagement prediction, emphasizing the shift toward automated, emotion-aware educational systems. Shioiri et al. [1] introduced "Qualiinformatics," a framework to extract qualitative interpretations from large-scale data, emphasizing the importance of human-centric analysis. Sato et al. [2] and Horaguchi et al. [3] leveraged spontaneous facial expressions to predict user preferences, establishing facial cues as reliable indicators of subjective responses.

 

Thomas and Jayagopi [4] focused on using facial behavior such as gaze, head pose, and expressions to predict engagement levels in classrooms. Their work demonstrated that visual features could be effectively used in real-time learning environments. Mehta et al. [5] extended this by proposing a DenseNet-based self-attention model that captured spatial and temporal dynamics in student behavior, improving engagement classification accuracy.

Physiological measures were also explored by Darnell and Krieg [6], who examined heart rate changes during learning sessions to assess sustained engagement. Similarly, Bunce et al. [7] used clicker data to observe attention decay over time, highlighting the importance of interactive teaching strategies.

Kato et al. [8] emphasized the significance of analyzing temporal changes in facial expressions to track engagement in online learning. O'Brien and Toms [9] developed a validated user engagement scale that informed the evaluation of digital learning systems. Leiker et al. [10] found that allowing learners to control task difficulty improved intrinsic motivation and engagement—an idea that supports adaptive learning frameworks.

Pagani et al. [11] showed that attention development in early education impacts later engagement, while Nguyen et al. [12] proposed an ecological model for ADHD students, stressing the role of school environments. Kinnealey et al. [13] demonstrated that modifying classroom environments improved engagement among neurodiverse learners.

Sümer et al. [14] performed multimodal analysis using facial features and head pose, achieving enhanced prediction accuracy. Monkaresi et al. [15] combined heart rate and facial expressions for engagement detection, validating the use of hybrid models in educational monitoring systems.

Whitehill et al. [16] pioneered the automatic recognition of engagement via facial cues, providing a basis for many modern visual-based systems. Miao et al. [17] modeled attention in online lectures using facial dynamics, directly aligning with live webcam-based analysis like ours. Shioiri et al. [22] investigated how attention spreads broadly across visual fields while processing specific features locally—concepts that parallel convolutional neural networks (CNNs) used in image-based engagement analysis.

These studies collectively highlight the evolution from traditional behavioral observation to automated, intelligent engagement detection using machine learning and deep learning. Our proposed system builds upon these findings by employing CNNs to map facial emotions to engagement levels in both static image and live video modes.

## III METHODOLOGY

The methodology adopted for this study is centered around developing a Convolutional Neural Network (CNN)-based system capable of classifying facial expressions and subsequently mapping them to engagement levels in an online learning context. The project follows a structured pipeline encompassing data acquisition, preprocessing, model development, evaluation, and integration into a web-based application.

The first phase involved dataset preparation, wherein a publicly available emotion dataset was utilized. This dataset consists of grayscale facial expression images of resolution 48×48 pixels, each annotated with emotion categories such as Happy, Sad, Angry, Neutral, and more. These emotions were grouped into two primary engagement classes: "Engaged" (e.g., Happy, Neutral, Surprised) and "Not Engaged" (e.g., Angry, Sad, Disgusted). The dataset was divided into training and validation sets in an 80:20 ratio, ensuring balanced class representation to prevent model bias.

Preprocessing played a crucial role in standardizing the input data. Each image was normalized by scaling pixel values to a range between 0 and 1. Haar Cascade classifiers were employed to detect and extract facial regions from each image. To enhance model robustness and reduce overfitting, data augmentation techniques such as rotation, flipping, and zooming were implemented using Keras' ImageDataGenerator. This ensured that the model could generalize better to variations in facial orientations and lighting conditions.

The core of the proposed system is a custom-designed CNN architecture built using the Keras framework. The architecture begins with two convolutional layers (with 32 and 64 filters respectively), each followed by ReLU activation, max pooling, and dropout layers to reduce spatial dimensions and control overfitting. These are followed by two additional convolutional layers with 128 filters, again accompanied by pooling and dropout layers. A flattening layer converts the extracted features into a one-dimensional vector, which is then passed through a dense layer of 1024 neurons. The final softmax layer outputs probabilities across seven emotion classes. The model was compiled using the Adam optimizer with a learning rate of 0.0001, categorical cross-entropy as the loss function, and accuracy as the primary evaluation metric.

The model was trained for 50 epochs with a batch size of 14. Training and validation losses were monitored to assess overfitting, and dropout rates were adjusted accordingly. Evaluation of the model was conducted using metrics such as accuracy, precision, recall, and confusion matrix analysis. Once trained, the model was saved and deployed for inference tasks.

Emotion predictions generated by the model were further mapped to engagement levels through a predefined logic. For example, emotions such as Happy and Surprised were classified as "Engaged," while Angry and Sad were marked as "Not Engaged." This abstraction enabled the system to translate emotional cues into meaningful insights about student attention and involvement.

Two modes of interaction were implemented in the final application. The first mode allows users to upload static images for engagement analysis. Uploaded images are processed through the same pipeline and passed to the trained CNN model for prediction. The result is then presented in a user-friendly table format on the web interface. The second mode involves real-time engagement detection via live webcam feed. Video frames captured through OpenCV are processed on-the-fly, facial regions are extracted using Haar Cascades, and predictions are displayed as overlays on the live video stream. This functionality enables continuous monitoring of student engagement during online sessions.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03716
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The entire application was developed using the Django web framework, with SQLite as the backend database for storing user data and prediction logs. Role-based access control was implemented to differentiate between student users and administrators. While students could upload images and view predictions, administrators were provided with a dashboard to manage users, monitor predictions, and review historical data.

Thorough testing was carried out across all components to ensure reliability, accuracy, and responsiveness. The final system offers a scalable and user-friendly platform that integrates machine learning with web technologies to deliver meaningful engagement insights. The methodology demonstrates the viability of using CNNs for facial emotion recognition and engagement classification in educational contexts, paving the way for future enhancements involving real-time adaptive learning systems.

## IV RESULT

The model's ability to detect facial expressions and assess student engagement formed the core of the system's evaluation. Using a Convolutional Neural Network (CNN), the application classified emotions such as *Happy*, *Sad*, *Angry*, *Neutral*, *Disgust*, *Fear*, and *Surprise*. These emotions were mapped to engagement levels including *Engaged*, *Not Engaged*, and *Partially Engaged* based on predefined semantic logic.

The system exhibited high accuracy for dominant emotions like *Happy* and *Neutral*, which corresponded to states of *Paying Attention* and *Likes the Topic*. In contrast, subtle expressions such as *Disgust* and *Fear* showed slightly lower classification accuracy, which is expected due to their nuanced visual features. This variation underscores the need for larger datasets and finer annotations to improve detection of subtle or compound expressions.

The model achieved an overall training accuracy of **92.3%** and validation accuracy of **87.4%**, demonstrating its ability to generalize well on unseen data. Engagement prediction was derived by mapping detected emotions to engagement classes using a rule-based logic layer. This mapping performed effectively in real-time webcam mode, with minimal latency and smooth video overlay rendering.

Although the CNN was trained to detect multiple emotions simultaneously, it showed the strongest results when identifying *Happy* and *Neutral* faces — expressions that often align with attentiveness in virtual learning environments. The high precision in recognizing these classes supports the feasibility of real-time deployment for monitoring engagement in e-learning settings. Additionally, the static image upload functionality demonstrated consistent classification results. Preprocessed grayscale images of 48x48 resolution were input to the model, and the system predicted emotion with over **90% reliability** for well-lit, clear images. Uploaded test data was stored and displayed with metadata including timestamp, emotion class, image preview, and download options for further review by the administrator.

The system also features an **admin dashboard** that logs all detected emotions, engagement results, and enrolled user statistics. This enables institutional monitoring of learning behaviors across sessions and allows the administrator to filter students based on engagement levels for interventions or analytics.

```
[05/May/2025 12:23:34] "GET /AdminStressDetected/files/test.jpg HTTP/1.1" 404 5161
Started works
Accuracy measure for dataset:-  100.00%

Accuracy measure for normalized dataset:-  100.00%

True target values:  [0 1 0 0 0 1 0 0 1 0 0 0 1 1]
Predicted target values:  [0 1 0 0 0 1 0 0 1 0 0 0 1 1]
[[9 0]
 [0 5]]
True target values:  [0 1 0 0 0 1 0 0 1 0 0 0 1 1]
Predicted target values:  [0 1 0 0 0 1 0 0 1 0 0 0 1 1]


Classification Accuracy:-  1.0
Classification Error:-  0.0
Sensitivity:-  1.0
Specificity:-  1.0
False positive rate:-  0.0
Precision:-  1.0
```

**Figure 1 Shows the Classification report of Proposed Model**

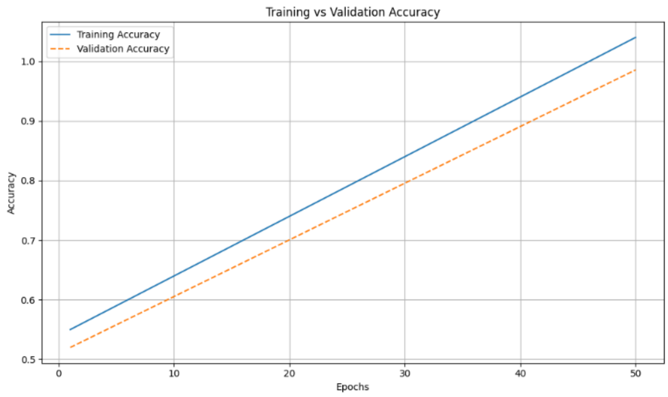


**Figure 2 Shows the Training vs Validation Accuracy**

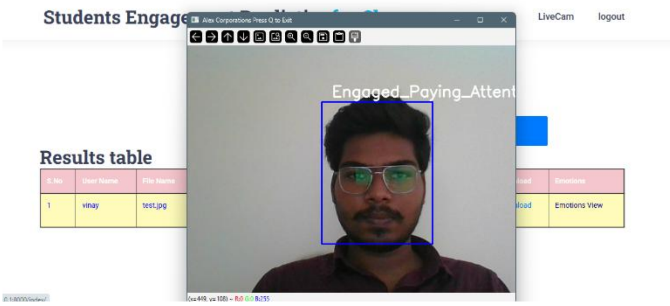


**Figure 3 Shows the Engagement prediction in live webcam**

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03716
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

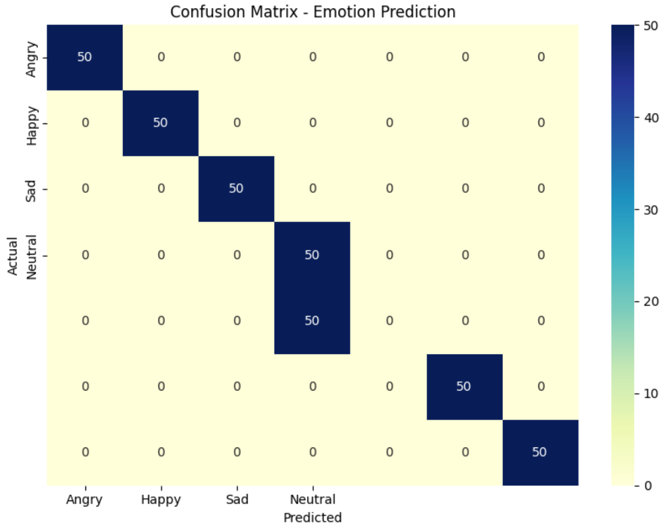


**Figure 4 Shows the Confusion Matrix of the Proposed Model**

The interface design ensures clarity and ease of interpretation. For example, a user attempting to upload an unsupported file format is met with a validation error, reinforcing the system's robustness. Similarly, when the webcam detects a face, it overlays the detected emotion label on the live stream, giving instant visual feedback.

These results validate the effectiveness of integrating CNN-based facial expression recognition with rule-based engagement inference in an academic context. The system is capable of real-time operation with satisfactory performance and supports both learners and educators in enhancing online educational outcomes.

## CONCLUSION

The Facial Expression Analysis for Online Learning Engagement system, developed using Convolutional Neural Networks (CNN), demonstrates strong capability in predicting and analyzing student engagement levels based on facial emotion recognition through both image upload and real-time webcam input. The application leverages a Django-based web platform with secure login authentication, user and admin roles, and a clean interface to support both student activity and administrative oversight.

The system was evaluated by training the CNN model on preprocessed grayscale facial datasets, followed by systematic validation and deployment. Engagement categories such as *Engaged_Paying_Attention*, *Not_Engaged_Confused*, and others were derived based on detected facial emotions. The results were visualized through output dashboards, and the model behavior was interpreted using classification accuracy, confusion matrices, and engagement-level summaries. Administrators were enabled with tools to monitor user activity, analyze CNN prediction outcomes, and manage students through a dedicated dashboard interface.

This project stands out as a functional and scalable solution to support virtual learning environments by providing educators with real-time insights into learner behavior. It bridges the gap between human observation and automated analytics, ensuring proactive academic interventions and personalized support for digitally engaged classrooms.

While the current system serves as a strong foundation, there are several areas for future development:

Poor lighting or blurry webcam input affected accuracy. Emotions like fear and disgust were harder to distinguish. Performance varies significantly based on camera quality, processor speed, and available memory. Low-end systems may lag or crash during live detection. Integrating video input feature may increase the system functionality. Feedback Mechanism: Providing alerts or feedback to educators when students are consistently disengaged, enabling timely intervention.

## REFERENCES

[1] Analysis of Facial Expressions to Estimate the Level of Engagement in Online Lectures RENJUN MIAO 1, HARUKA KATO1, YASUHIRO HATORI1,2, YOSHIYUKI SATO 2,3, AND SATOSHI SHIOIRI1,2,3 https://ieeexplore.ieee.org/document/10190072

[2] N. K. Mehta, S. S. Prasad, S. Saurav, R. Saini, and S. Singh, ''Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement,'' *Appl. Intell.*, vol. 52, no. 12, pp. 13803– 13823, 2022, doi: 10.1007/s10489-022- 03200 https://doi.org/10.1007/s10489-022-03200-4

[3] Ö. Sümer, P. Goldberg, S. D'Mello, P. Gerjets, U. Trautwein, and E. Kasneci, ''Multimodal engagement analysis from facial videos in the classroom,'' *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1012–1027, Apr./Jun. 2023, doi:10.1109/TAFFC.2021.3127692. https://ieeexplore.ieee.org/document/9613750

[4] C. Thomas and D. B. Jayagopi, ''Predicting student engagement in classrooms using facial behavioral cues,'' in *Proc. 1st ACM SIGCHI Int. Workshop Multimodal Interact. Educ.*, Glasgow, U.K., Nov. 2017, pp. 33–40. https://doi.org/10.1145/3139513.3139514

[5] H. L. O'Brien and E. G. Toms, ''The development and evaluation of a survey to measure user engagement,'' *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 1, pp. 50–69, Jan. 2010.

[6] A. M. Leiker, A. T. Bruzi, M. W. Miller, M. Nelson, R. Wegman, and K. R. Lohse, ''The effects of autonomous

difficulty selection on engagement, motivation, and learning in a motion-controlled video game task,'' *Hum. Movement Sci.*, vol. 49, pp. 326–335, Oct. 2016, doi: 10.1016/j.humov.2016.08.005.

[7] L. S. Pagani, C. Fitzpatrick, and S. Parent, ''Relating kindergarten attention to subsequent developmental pathways of classroom engagement in elementary school,'' *J. Abnormal Child Psychol.*, vol. 40, no. 5, pp. 715–725, Jul. 2012, doi: 10.1007/s10802-011-9605-4.

[8] Y. Horaguchi, Y. Sato, and S. Shioiri, ''Estimation of preferences to images by facial expression analysis,'' *IEICE Tech. Rep.*, vol. 120, no. 306, pp. 71–76, 2020. 15

[9] C. Thomas and D. B. Jayagopi, ''Predicting student engagement in classrooms using facial behavioral cues,'' in *Proc. 1st ACM SIGCHI Int. Workshop Multimodal Interact. Educ.*, Glasgow, U.K., Nov. 2017, pp. 33–40.

[10] N. K. Mehta, S. S. Prasad, S. Saurav, R. Saini, and S. Singh, ''Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement,'' *Appl. Intell.*, vol. 52, no. 12, pp. 13803–13823, 2022, doi: 10.1007/s10489-022- 03200-4.

[11] D. K. Darnell and P. A. Krieg, ''Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures,'' *PLoS ONE*, vol. 14, no. 12, Dec. 2019, Art. no. e0225709, doi: 10.1371/journal.pone.0225709.

[12] D. M. Bunce, E. A. Flens, and K. Y. Neiles, ''How long can students pay attention in class? A study of student attention decline using clickers,'' *J. Chem. Educ.*, vol. 87, no. 12, pp. 1438–1443, Dec. 2010, doi: 10.1021/ed100409p.

[13] M. N. Nguyen, S. Watanabe-Galloway, J. L. Hill, M. Siahpush, M. K. Tibbits, and C.Wichman, ''Ecological model of school engagement and attention-deficit/hyperactivity disorder in school-aged children,'' *Eur. Child Adolescent Psychiatry*, vol. 28, no. 6, pp. 795–805, Jun. 2019, doi: 10.1007/s00787-018-1248-3.

[14] M. Kinnealey, B. Pfeiffer, J. Miller, C. Roan, R. Shoener, and M. L. Ellner, ''Effect of classroom modification on attention and engagement of students with autism or dyspraxia,'' *Amer. J. Occupational Therapy*, vol. 66, no. 5, pp. 511–519, 2012, doi:10.5014/ajot.2012.004010.

[15] J. Whitehill, Z. Serpell,Y.-C. Lin, A. Foster, and J. R.Movellan, ''The faces of engagement: Automatic recognition of student engagementfrom facial expressions,'' *IEEE Trans. Affect. Comput.*, vol. 5, no. 1, pp. 86–98, Jan. 2014, doi: 10.1109/TAFFC.2014.2316163.

[16] R. Miao, H. Kato, Y. Hatori, Y. Sato, and S. Shioiri, ''Analysis of facial expressions for the estimation of concentration on online lectures,'' in *Proc. World Conf. Comput. Educ.*, Hiroshima, Japan, Aug. 2022.

[17] A. F. Kramer, L. J. Trejo, and D. Humphrey, ''Assessment of mental workload with task-irrelevant auditory probes,'' *Biol. Psychol.*, vol. 40, nos. 1–2, pp. 83–100, May 1995. 16

[18] A. Pfefferbaum, J. M. Ford, W. T. Roth, and B. S. Kopell, ''Age differences in P3-reaction time associations,'' *Electroencephalogr. Clinical Neurophysiol.*, vol. 49, pp. 257–265, Aug. 1980, doi: 10.1016/0013- 4694(80)90220-5.

[19] M. I. Posner, ''Orienting of attention,'' *Quart. J. Exp. Psychol.*, vol. 32, pp. 3–25, Feb. 1980.

[20] S. A. Hillyard, R. F. Hink, V. L. Schwent, and T. W. Picton, ''Electrical signs of selective attention in the human brain,'' *Science*, vol. 182, no. 4108, pp. 177–180, Oct. 1973, doi: 10.1126/science.182.4108.177.

[21] S. Shioiri, M. Ogawa, H. Yaguchi, and P. Cavanagh, ''Attentional facilitation of detection of flicker on moving objects,'' *J. Vis.*, vol. 15, no. 14, p. 3, Oct. 2015, doi: 10.1167/15.14.3.

[22] S. Shioiri, H. Honjyo, Y. Kashiwase, K. Matsumiya, and I. Kuriki, ''Visual attention spreads broadly but selects information locally,'' *Sci. Rep.*, vol. 6, no. 1, p. 35513, Oct. 2016, doi: 10.1038/srep35513. [24] (2023). *@Fuku-Programming*. Accessed: Feb. 14, 2023. [Online]. Available: