Open Access
Full Length Article

ORIGINAL ARTICLE

Smart Class Analytics for Assessing Group Understanding via Facial Expressions

Haider Ali[1],Muhammad Adnan Kaim Khani[2],Abdul Salam Shah[3],Asif Ali Laghari[4],Muhammad Minam[5]

[1]Department of Computer Science, ILMA University, Karachi, Pakistan
[2]Department of Computer Science, ILMA University, Karachi, Pakistan
[3]School of Computer Science, Faculty of Innovation and Technology, Taylor's University, Subang Jaya, Selangor, Malaysia
[4]Software Collage Shenyang Normal University Liaoning, China
[5]Department of Computer Science, ILMA University, Karachi, Pakistan

ABSTRACT

Abstract: Understanding the comprehension level of students in a communicating classroom is critical. As technology becomes increasingly integral to our lives, advanced facial emotion recognition becomes a powerful tool for assessing comprehension levels in a real-time scenario. We created a system to measure cognizance levels and analyze facial emotions to ascertain comprehension. In this research, we presented a new system for assessing the comprehension level of a group setting through facial expressions in a real-time. Our techniques, supported by advanced DL techniques like OpenCV and Deep Face provide us with different backend options that are very effective for facial emotion recognition and analysis such as MTCNN and Retina face.

Initially, we used OpenCV for real-time capturing video by computer webcam and capture the frames at regular intervals. The frames are stored in a pre-defined directory. After that, we defined a function that iterates over each frame for detecting emotions with facial landmarks and also shows the dominant emotion with each face in the frame. A pre-trained MTCNN model (one of the state-of-the-art models provided by Deep Face) has been utilized for detecting the face present in each frame and facial landmark localization. MTCNN is a lightweight model that detects faces with high accuracy and speed.

Finally, The Deep Face analyze () function has been utilized for emotion analysis in the wild with the MTCNN backend which provides insights into the dominant emotion related to each face. Comprehension levels were measured as per emotional states detected, with positive affect ['happy', 'neutral' (when confidence is >70), and 'surprised' (when confidence is <= 60%)] implying higher comprehension, negative affect ['sad', 'disgust', 'fear', and 'surprised when the confidence level is above 60%'] hinting at lower comprehension, and when ['happy', 'neutral'] emotions having confidence <=70% hinting to moderate affect. The measured comprehension level is then disseminated into high, moderate, and low.

The finding exposed the complex interrelation among facial expressions, emotions, and comprehension levels in intra-group interactions. This research supports the development of understanding human behavior and interaction dynamics, with connection to domains such as psychology, human-computer interaction, and education.

Keywords: *Emotion Recognition, Image Analysis, Computer Vision, Image Processing, Understanding Assessment*

Author`s Contribution
[1,2] Data analysis, interpretation, and manuscript writing, Active participation in data collection
[34,,5] Conception, synthesis, planning of research. Interpretation and discussion

Address of Correspondence
Haider Ali
Email: haiderali51214318@gmail.com

Article info.
Received: June,15, 6, 2025
Accepted: November 13, 2025
Published: December20, 2025

Cite this article: *Ali H, khani K A M, Shah S A, Laghari A A.Minam A. Smart Class Analytics for Assessing Group Understanding via Facial ExpressionsJ. inf. commun. technol. robot. appl.2025; 14(1):1-08*

Funding Source: Nil
Conflict of Interest: Nil

INTRODUCTION

A systematic investigation of emotion recognition was first introduced in the 1990s [1]. In the realm of educational psychology, emotions play a significant role because they are closely linked with how we think and learn. Understanding students' emotions in the classroom is inviable for educators as it enables them to tailor their teaching methods and enhance individual student's development [2]. By analyzing emotions teachers can adopt their instructional strategies to better engage students.

Facial emotional expression is super useful in a lot of areas such as it helps in psychology, and law, making animations more real, and even in security and counterterrorism. The process of recognizing facial emotions involves the main steps: image pre-processing [3], identifying the face [4], then extracting features [5], and ultimately categorizing the expressions displayed.

Facial emotion recognition is essential for effective human communication and interaction. Detecting prominent concentration levels of students depends on ayes contact and mouth regions to successfully recognize emotions, sadness and fear belong to eye contact while disgust and happiness belong to mouth expression [2,6]. Understanding how people understand and distinguish facial expressions is important in several fields such as Artificial Intelligence, Human-computer interaction [7], mental health assessment, and Psychology [8]. In the present age, humans can not only communicate with each other but also with machines. As humans interact with machines the question arises whether machines can understand and interpret human emotions. Automatic facial recognition emerges as a tool for bridging this gap.

The classroom environment has a profound impact on students' emotions. Emotions are crucial in education just like in everything else we do. Emotions have an immense influence on learning how well students perform academically [9]. Positive emotions like interest and enthusiasm can significantly boost student's motivation and engagement levels. On the other hand, negative emotions like anxiety or frustration make it difficult for students to learn and can slow down their academic evolution [10].

Students can express their emotions through many ways such as facial expression, body language, and interaction with a peer or their teachers. By recognizing and understanding these emotions/expressions, educationalists can get a valuable understanding of students' emotional states and address their needs more effectively. A student ERS is a technical tool intended to detect and analyze students' emotions in an educational setting [11]. The ERS student coalition uses sensors such as cameras, microphones, and psychological sensors to collect data on the facial expressions, voice data, and body language of students. These are then examined by AI algorithms or ML models to determine what emotional state students are in the system [9]. An ERS can serve in the area of teaching and learning by providing valued student perceptions of their emotional experiences and needs.

LITERATURE REVIEW

Problem Statement:

In education sitting truthfully evaluating group comprehension level remains a persistent challenge due to the restrictions of traditional education methods. Existing methods to assess group understanding depend upon subjective judgment or limited feedback mechanisms. This can lead to a lack of accuracy and inconsistency in calculating understanding levels. Moreover, conventional assessment methods might be missed out on excellently apprehended subtle expressions and nonverbal signals which play a crucial role in group understanding. Due to this reason, we need for precise and thorough system to evaluate comprehension levels within group settings. As a result, there are often inaccuracies and gaps that lead us to how we can measure comprehension level which can hinder the efficiency of teaching technique and eventually affect student learning outcomes. Consequently, there is a clear need for a state-of-the-art assessment system that leverages facial expression analysis offering a more nuanced, dependable, and automated evaluation of group comprehension levels.

Problem Solution:

The suggested approach entails the development of a facial expression-based understanding assessment system for evaluating group comprehension levels built on facial emotions in real time. The system will utilize progressive technology such as facial recognition, and machine learning techniques that aim to provide instructors with added precision and effective means of assessing comprehension within a group environment.

OBJECTIVES:

Objective 1: Development of a facial emotions recognition system utilizing facial landmarking techniques, feature extraction, feature classification, and prediction to analyze facial emotions.

Objective 2: Investigate the potential connection between Detected FEs and traditional indicators of comprehension.

Objective 3: Extract important features to classify the facial emotions to evaluate the comprehension assessment in a group setting environment.

Research Questions:

Research Question 1: What are the key emotions that affect the efficiency of FER in real-time the understanding measurement system or comprehension assessment system in instructive group setting?

Research Question 2: When should educators apply a comprehension assessment system to enrich educational methods?

Research Question 3: Why do educators use facial emotion analysis techniques to improve academic Performance?

Related Work:

The assessment of students' comprehension conventionally depends on direct feedback mechanisms like verbal responses, quizzes, and exams. However, these approaches frequently slip the instant emotional and reasoning state of students during learning. Current expansions in facial expression analysis deliver auspicious substitutes for real-time assessment of student engagement and understanding. The FACS, formulated by Ekman and Friesen [12] offers a detailed structure for the classification of physical expression related to several emotions. Numerous studies have relied on this system as a cornerstone that aimed to decode facial expressions into recognizable emotional states.

In the domain of emotion identification, the utilization of DL and TL has been proven particularly effective. Moreover, Pantic and Patras [13] illustrated the adaptation of SVMs for real-time FER within video feeds. Castellano et al. [14] illustrated the importance of body signals and dynamics of gestures in conveying emotional states, emphasizing the complex interconnection between physical movements and internal emotional states.

The research by Ko, and Byoung Chun [15] offers a succinct introduction to DL algorithms to identify and analyze facial expressions in real-time. The CNNs and RNNs are frequently utilized in these systems for the analysis of facial features and forecast emotional states. Molla Hosseini, Chan, and Mahoor [16] illustrated that a pre-existing model could be fine-tuned to limited datasets resulting in notable efficiency of emotion detection. This technique is particularly valuable in education, wherever get-together real-time emotional data from students can be problematic because of limited sample sizes. Ding et al. [17] tackle challenges in real-world emotion acknowledgment, particularly with restricted datasets. Their research highlights the potential of harnessing DNN and multimodal strategies to improve assessment systems. By enhancing emotion recognition in varied educational settings, our system for assessing understanding based on facial expressions for evaluating comprehension levels.

Li, Deng, and Du highlight here [18] the difficulties associated with identifying facial expressions in uncontrolled environments and suggest remedies to enhance the system's resilience and accuracy of the system. Salah et al. [19] explored the complexities of in-the-wild emotion recognition, including signal variation and noise dispersal. They reviewed the recent multimodal methodologies and proposed summarized techniques for visual and audio algorithmic methods.

Through the integration of audio-video attributes with regression-based classifiers which highlighted progressive results in their study on FER. A video-based facial emotion recognition neural network functioning in three dimensions, presented in [20] showed that 3D-CNN is very good for guessing facial emotions expressed over a sequence of frames. To enhance the version of 3D-CNN architecture by hyper-parameters search showed a robust impact on results. Facial emotion recognition has garnered a lot of consideration over the previous years. Rit Lawpanom et al. [21] proposed a HoE-CNN for FER in online learning. their study tackles imbalanced datasets and multi-category classification and enhances classification accuracy to 75.51% on the FER2013 dataset which showcases its capability for real-time submissions. Riffi, Jamal, et al. [22] employ VGG19 for feature extraction and LSTM for apprehending spatio-temporal information in conjunction with HOG-HOF descriptors for capturing dynamic features.

Recent research, such as by Nezami, Omid, et al. [23] has explicitly focused on the education domain, using DL algorithms to differentiate student engagement via facial expressions. Their finding emphasizes the feasibility of utilizing facial expression analysis to observe and improve student engagement for assessing classroom comprehension in real-time scenarios.

RESEARCH METHODOLOGY

We map out the methodology utilized to build a facial emotions-based assessment system that sought to examine the comprehension level of a group setting. This methodology reveals an orderly technique for capturing and analyzing video frames for facial emotion recognition and figuring comprehension levels based on emotions (see Fig 1). This methodology outlines the steps how to proceed for calculating comprehension based on the emotional state of a wild environment where a group of people are interacting. Deepface is an open-source lightweight library for face recognition, emotion detection, facial attribute analysis, age detection, and race detection framework for Python [24].

Positive and Negative Emotions:

Positive emotions are associated with understanding and engagement such as: -

Happy: Display the candidate's understanding of the material well.

Neutral: Indicate the attentive and calm behavior of students.

Surprise: If up to <= % it leads to understanding Negative emotions can hinder comprehension such as Sad: Indicates the lack of understanding Surprise: It can be positive sometimes and can be negative sometimes depending upon the situation. if >60% it leads to low understanding and when <=60% it leads to high understanding. Fear Indicate the hesitation Disgust: Indicate the solid undesirable feelingsAngry: Indicate the frustration and misunderstanding.

Fig. 1. Overview of facial emotions analysis, from video capture to emotion classification and understanding level calculation.

System Overview:

The system is implemented by setting up a high-resolution camera to capture the video in real-time by using OpenCV and then processing the captured frames by employing the Deep Face analyze () [25] function to identify the most dominant emotion [26] displayed by each face. The emotions are classified as positive emotions such as ['happy', 'neutral', 'moderate-surprise'] and negative emotions such as ['sad', 'angry', 'disgust', 'high-surprise']. Then emotions are aggregated to calculate the comprehension level of students as high-understanding, moderate-understanding, and low-understanding.

Data Collection:

In this study, data collection is a vital component, intended to observe the real-time facial emotions of students during a lecture. The procedure was initiated by capturing video footage of the classroom environment using a webcam synthesized with a computer system. At the outset of our system, we record a frame every 5 seconds within a twenty-minute duration. The seized frames are then preserved in the predetermined directory which was named with a standardized approach to streamline the retrieval and processing. This organized approach to data collection allows us to gather top-quality visual data primarily for precise emotional analysis and comprehension assessment.

Model Used:

We utilized a pre-trained MTCNN model for face detection in real-time. This model is provided by the DeepFace library. DeepFace also provides RetinaFace backend which is the state-of-the-art model for computer vision tasks such as emotions, face, gender, age, and race detection. RetinaFace is more accurate for detecting faces in the wild [27] but it is slow. MTCNN is relatively faster than Retinaface but it faces low accuracy as compared to the RetinaFace backend. OpenCV, Dlib, and SSD are faster but not as accurate as RetinaFace and MTCNN.

Emotion Detection:

Retina face and MTCNN backend are state-of-the-art models for emotion analysis for static images but we are interested in real-time scenarios to analyze the emotions of each detected face. We utilize the DeepFace framework for this purpose. After installing the necessary packages and libraries we can apply these analyzing techniques to get useful information from an image. The Analyze () method contains a robust facial attributes analysis landscape with different parameters such as img-path, actions, and detector-backend. Each backend has a different speed and accuracy [28].

In the process of aggregation of emotions, Deep Face provides us emotions of each detected face in the input image and also gives us the most dominant facial emotion. For each face, we count or extract the dominant facial emotion. Then based on the most dominant emotion we count how many faces show particular emotions such as sad, happy, neutral, etc. Analysis of the dominant emotion enables us to aggregate and gain insights into the general emotional distribution within the group.

RESULTS AND DISCUSSION

The results are derived from analysing facial expressions or emotions captured in real-time using Haarcascades or MTCNN (depending upon choice) face detectors. The main goal is to measure the comprehension level of individuals constructed on their facial emotions. This analysis categorizes emotions that contribute to different levels of understanding High, Moderate, and Low. The rationale for this classification is based on particular criteria that determine the prevailing emotion observed on each face, along with its confidence score. We analysed seven basic emotions (happy, neutral, surprise, sad, angry, fear, disgust). Happiness is strongly linked to a positive reaction to details. When people perceive happiness, they are typically more responsive and likely to grasp new information, making it a sign of strong comprehension. Neutral and happy emotion indicates a standard, composed, and focused state. When it is observed with high confidence (e.g., > 70%), it implies focused engagement and likely strong understanding. Lower confidence in neutral emotion (e.g., ≤ 70%) may indicate moderate understanding due to reduced emotional engagement. Surprise happens as a reaction to an unforeseen event.

A mild surprise (e.g., confidence ≤ 60%) might suggest a rapid adaptation to new information, implying strong understanding. Higher surprise (e.g., confidence > 60%) could suggest confusion or struggle to understand, resulting in reduced understanding. Sadness is linked to emotions of disappointment or hindrance, often delaying mental processing and suggesting reduced understanding. Disgust provokes a strong repulsion, both mentally and physically, resulting in low understanding because of its adverse effect on concentration. Fear, associated with the perception of danger, provokes worry and tension, interrupting cognitive functioning and resulting in low understanding. Anger directs concentration towards the source of displeasure instead of the subject matter, impairing brain activity and leading to low understanding. The emotions are characterized as positive and negative to recognize overall group-level emotions [29].

Setting benchmarks for each emotion consists of determining thresholds explained in Table 1 that classify comprehension levels depending on the type of emotions and confidence score. These thresholds may vary contingent on the dataset or the explicit task, but researchers typically involve fine-tuning these standards to maximize classification accuracy across dissimilar emotions [30].

TABLE I: SUMMARY of Experimental Benchmark Thresholds

Emotions	High	Moderate	Low
Happy Neutral	Confidence >70%	Confidence ≤ 70%	Not Applicable

Surprise	Confidence ≤ 60%	Not Applicable	Confidence > 60%
Sad	Not Applicable	Not Applicable	Detected at any level
Fear			
Angry			
Disgust			

Processing of Frames:

We perform a live demo of our system in a classroom to calculate students' comprehension levels in real-time. A cascade classifier is used for face detector in real-time provided by Opens. The frames are captured in regular intervals and processed using MTCNN provided by the Deep face library. MTCNN provides a robust analysis of facial expressions. In each frame, the emotion of each individual is analysed and the most dominant emotion is captured. Each face's dominant emotions contribute to calculating comprehension levels such as high, moderate, and low. We captured several frames in the specified duration. For example, in our case, we set the duration for 20 minutes and captured the frame after every 5 seconds. Some frames with analysed results after processing are shown as follows (see Fig 2).

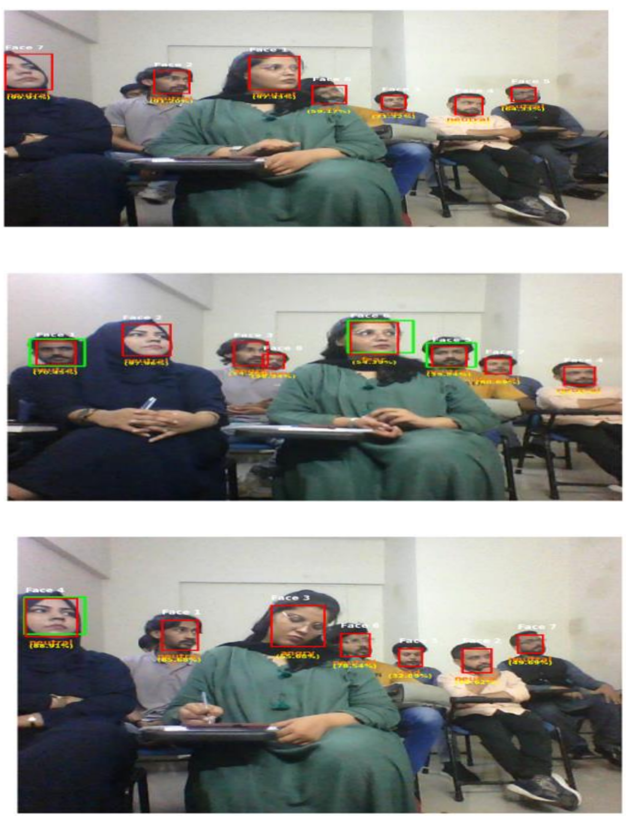


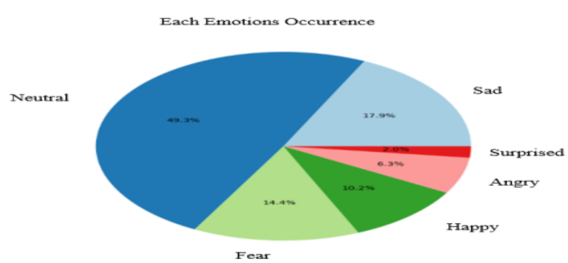
Fig 2. Represents the real-time pictures of testing. (a) shows the understanding level as high, moderate, and low as 3,2, and 3 respectively, (b) shows the understanding level as high, moderate, and low as 6,0, and 1 respectively, (c) shows the understanding level as high, moderate, and low as 3,2, and 2 respectively.

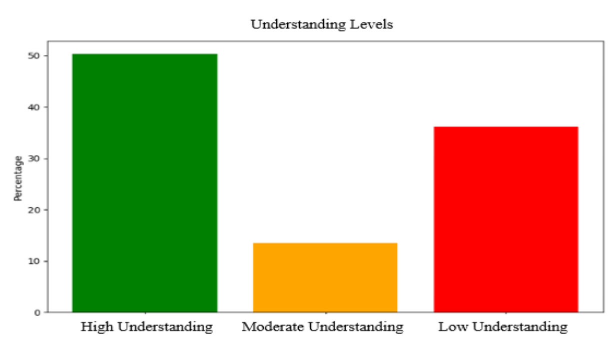
Analysis of Results:

The important part of this research is to calculate the understanding level based on emotions through the frames captured by the webcam at a certain period and then analyze the facial expressions accompanied by these frames. By leveraging the DeepFace library for facial expression analysis the results were used to gauge comprehension which depends on the classification of emotions into positive and negative groups. Initially, the results were taken every 5 seconds over 20 minutes, we can extend the scope of testing as desired. The frames were systematically organized in sequence for easy access and processing. The dominant emotions and their confidence levels are detected through emotional assessment which is subsequently associated with levels of comprehension such as high, moderate, and low.

Figures and Graphs:

The following visualization offers graphical insights into the emotional dynamics during interactions, delineating the distribution of comprehension and understanding of the participants. (see Fig 3).





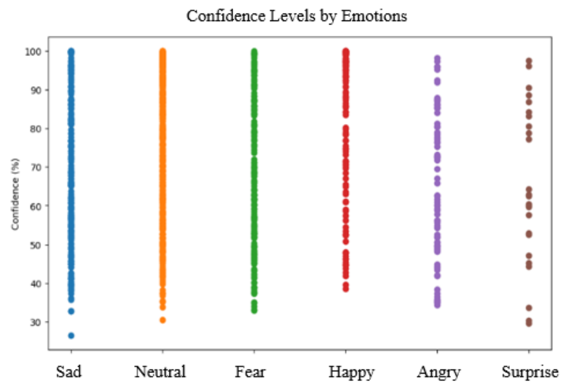


Fig 3: Shows the results of comprehension. (a) represents the occurrence of each emotion (b) shows the understanding level based on facial emotions at certain times as high, moderate, and low (c) shows the confidence level of each emotion.

Limitations:

It does not detect the emotion correctly when the face in the image is incomplete or half face appears in the photos. It leads to incorrect emotion detection which affects the results. The generalizability of results may be limited to a particular educational environment or population, which requires additional validation across different contexts. The accomplishment of the projected system may vary due to some influences such as image quality, light conditions, and diversity of facial expressions/emotions within the group. Resource limitations including funding, limited time, and technology instruments such as backend models with different speeds and accuracy, can restrict the scope of group comprehension assessment by facial emotions. While the study offered appreciated insights, it also acknowledged several zones that require special treatment and enhancements. Furthermore, we use facial emotion recognition to develop a comprehension assessment system. It is not based on psychological signals for capturing emotions. Its mean value equals 0.328 m, and the centered error equals 0.176 m. considering that the reference track radius is 100 m; we can state that the accuracy of the control system is good. On the other hand, Fig. 5d shows that the maximum value of the pitch angle equals 15o (or 0.26 rad) during a very short time, meanwhile during the whole time of tracking the pitch and roll angles do not exceed 10o (or 0.17 rad).

CONCLUSION

In conclusion, this research provides a thorough assessment of facial emotion recognition (FER) technologies and their changing significance in fields like computer vision and human-computer interaction. Through an examination of past innovations and contemporary techniques, we highlighted the profound impact of ML algorithms, especially DL frameworks like Convolutional Neural Networks (CNNs), in improving the precision and performance of emotion detection systems. The increasing incorporation of real-time applications, including adaptive learning systems and behavioural assessment, further demonstrates the significance of FER in both academic research and practical domain implementation. Although obstacles such as data privacy concerns, real-time processing limitations, and model bias persist, ongoing innovation in the field is poised to overcome these barriers and unlock new possibilities for emotion-driven technologies.

Discussion of Findings:

The results of this research highlight the relationship between certain facial expressions and the level of comprehension recorded. Positive emotions including happiness, and neutral with confidence levels > 70% are linked to high comprehension, and when their confidence level is ≤ 70% they lead to moderate comprehension. In the same way, if the surprised emotion confidence level is ≤ 60% it results in high comprehension and if its confidence is > 60% it results in low understanding. In contrast, negative emotions such as sad, fear, and disgust indicate lower comprehension. This approach can be extended to numerous applications.

Future Recommendations:

Future studies could explore the integration of various biometric data, such as consideration of psychological signals that can also be used for advanced emotional assessment based on Arousal (like heart rate skin conductance, and muscle tension) and Valence (like positive and negative emotions assessed by facial expression and body language). Transfer learning can also be important in emotion recognition [31]. Moreover, upgrading the model to accommodate a broader spectrum of environmental conditions and integrating cutting-edge emotion analysis methods such as DL algorithms and multimodal data processing can improve the accuracy and adaptability of the system in real-world scenarios.

Acknowledgment:

We would like to express our gratitude to the reviewers for their precise and succinct recommendations that improved the presentation of the results obtained.

Conflict of Interest:

The authors declare no conflict of interest.

REFERENCE

[1] Picard, R. W. (2000). Affective computing. MIT Press.

[2] Krithika, L. B., & GG, L. P. (2016). Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. Procedia Computer Science, 85, 767-776.

[3] Xu, M. (2024, February). Image Preprocessing Method Applied on Facial Emotion Recognition Problem. In 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023) (pp. 297-304). Atlantis Press.

[4] Jia, S., & Tian, Y. (2024). Face Detection Based on Improved Multi-task Cascaded Convolutional Neural Networks. IAENG International Journal of Computer Science, 51(2).

[5] Zheng, Y. (2023, October). Research on image feature extraction based on graph convolutional network. In Sixth International Conference on Computer Information Science and Application Technology (CISAT 2023) (Vol. 12800, pp. 595-601). SPIE.

[6] Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. PloS one, 12(5), e0177239.

[7] Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. PloS one, 12(5), e0177239.

[8] Cai, Y., Li, X., & Li, J. (2023). Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. Sensors, 23(5), 2455.

[9] Sun, B., Wei, Q., He, J., Yu, L., & Zhu, X. (2016, September). BNU-LSVED: a multimodal spontaneous expression database in an educational environment. In Optics and Photonics for Information Processing X (Vol. 9970, pp. 256-262). SPIE.

[10] Meena, G., Mohbey, K. K., Indian, A., Khan, M. Z., & Kumar, S. (2024). Identifying emotions from facial expressions using a deep convolutional neural network-based approach. Multimedia Tools and Applications, 83(6), 15711-15732.

[11] Chen, Z., Liang, M., Xue, Z., & Yu, W. (2023). STRAN: Student expression recognition based on spatio-temporal residual attention network in classroom teaching videos. Applied Intelligence, 53(21), 25310-25329.

[12] Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiology, 27(4), 363-384.

[13] Michel, P., & El Kaliouby, R. (2003, November). Real-time facial expression recognition in video using support vector machines. In Proceedings of the 5th International Conference on Multimodal Interfaces (pp. 258-264).

[14] Castellano, G., Villalba, S. D., & Camurri, A. (2007, September). Recognizing human emotions from body movement and gesture dynamics. In International conference on affective computing and intelligent interaction (pp. 71-82). Berlin, Heidelberg: Springer Berlin Heidelberg.

[15] Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. sensors, 18(2), 401.

[16] Ng, H. W., Nguyen, V. D., Vonikakis, V., & Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on International Conference on multimodal interaction (pp. 443-449).

[17] Ding, W., Xu, M., Huang, D., Lin, W., Dong, M., Yu, X., & Li, H. (2016, October). Audio and face video emotion recognition in the wild using deep neural networks and small datasets. In Proceedings of the 18th ACM international conference on multimodal interaction (pp. 506-513).