SMARTCLASS: A FRAMEWORK FOR STUDENT MONITORING AND ENGAGEMENT EVALUATION USING COMPUTER VISION AND BIO PRINT METHOD

Tien Do[1, 2], Xuan Le[3], Phong Nguyen A[3], Phong Nguyen*[3]
[1]University of Information Technology, Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam.
[2]Vietnam National University Ho Chi Minh City, Vietnam, Ho Chi Minh City, Vietnam.
[3]Ho Chi Minh University of Technology, Vietnam, Ho Chi Minh City, Vietnam.
*Corresponding Author: Phong Nguyen (Phone: +84 38 6101703; Email: ntphong2702@gmail.com).

ABSTRACT

Evaluating classroom effectiveness through children's emotions is crucial for optimizing educational outcomes. This study introduces SmartClass, a multi-modal framework crafted for real-time assessment of student engagement and emotional responses within classrooms. Designed for children aged 5–12, SmartClass employs YOLOv11-Face for face detection, YOLOv11-Classify for emotion recognition, and ArcFace for identity verification, achieving an overall accuracy of 92.6%. The framework is underpinned by a curated dataset of 50,000 real-life classroom images, ensuring robust performance across diverse learning environments. To bolster security, SmartClass incorporates biometric authentication methods, including fingerprint scanning and NFC card identification—enabling precise student verification upon school entry. A multicamera setup ensures comprehensive emotion analysis, effectively mitigating challenges such as occlusions and misclassifications. By offering real-time analytics and actionable insights aligned with a positive emotion policy, SmartClass equips educators with an intelligent tool to refine teaching strategies, foster student engagement, and drive adaptive learning experiences. This work contributes to bridging critical gaps in educational technology, presenting a scalable and data-driven solution for modern classrooms.

Keywords: Deep Learning, Smart System, Biometric Authentication, Emotion Recognition, Classroom Analytics

1. Introduction

High-quality education is fundamental to children's development, particularly during middle childhood (ages 5–12), a critical period for emotional, social, and cognitive growth (Allen et al., 2013), (Pope et al., 2012). Emotional regulation and processing during this stage significantly impact academic performance, classroom behavior, and peer interactions, shaping long-term educational and personal outcomes (Loevaas et al., 2018). While a positive emotional climate fosters student engagement and achievement, challenges in emotion regulation can hinder learning, underscoring the importance of integrating emotional well-being into educational frameworks (Frenzel et al., 2021), (Braet et al., 2014).

Despite growing recognition of the role of emotions in classroom dynamics, existing research lacks comprehensive datasets and systems that analyze the interplay between emotional and academic factors in real-world educational settings (Veraksa et al., 2020). Many studies have examined classroom quality and emotional regulation separately, failing to provide an integrated perspective (Jones et al., 2014). Furthermore, current datasets often suffer from limited representation of middle childhood or employ oversimplified emotional metrics, reducing their applicability to educational contexts (Brown et al., 2024).

For example, datasets such as AffectNet and FER-2013 offer large-scale annotated facial emotion images but primarily focus on adult populations or restricted emotion categories, limiting their relevance to children's classroom environments (Mollahosseini et al., 2019), (Zahara et al., 2020). Similarly, studies like DEAP and GoEmotions provide valuable physiological and textual emotion recognition insights but lack contextual data from school settings (Koelstra et al., 2012), (De Silva et al., 2022). This gap is particularly critical for children aged 5–12, where a targeted dataset is necessary to capture the nuanced relationship between emotions and learning experiences (Verpaalen et al., 2019).

To address these limitations, this paper introduces SmartClass, a novel multimodal framework designed for evaluating classroom effectiveness, emotional dynamics, and identity recognition in children aged 5–12. The key contributions of this work include:

- Development of the SmartClass dataset, comprising 50,000 images, to analyze emotional states and classroom interactions in children.
- Integration of YOLOv11 and ArcFace for face detection, emotion classification, and identity verification, achieving 92.6% accuracy.
- Implementation of biometric authentication methods, such as fingerprint scanning and NFC cards, to enhance student identification and security.
- Deployment of a multicamera system to improve emotion recognition accuracy by mitigating occlusions and misclassifications.
- Establishment of a classroom effectiveness policy framework based on emotional dynamics and student engagement analytics.

Through this research, SmartClass bridges the gap between emotional and academic evaluations, offering a scalable, data-driven solution for modern classrooms. By providing real-time insights, this framework empowers educators to refine teaching strategies, enhance student engagement, and foster an emotionally supportive learning environment.

2. Related Works

In recent years, the demand for datasets aimed at assessing classroom effectiveness, identity recognition, and emotional analysis especially for children aged 5–12 has surged. This rise is largely driven by advancements in deep neural networks, which excel at processing multimodal data from real-world educational environments (Mollahosseini et al., 2019), (Zahara et al., 2020). A key focus has been on "in-the-wild" datasets, which capture spontaneous classroom interactions, providing a foundation for bridging the gap between theoretical research and practical applications in education.

Emotion recognition datasets have been instrumental in advancing affective computing. The OMG Emotion Behavior dataset, for instance, captures spontaneous emotion expressions in real-world contexts, aligning with the need for authenticity in classroom-based emotion analysis. Meanwhile, physiological datasets such as DEAP (Koelstra et al., 2012) and SEED (Zheng & Lu, 2015) leverage EEG signals to analyze emotional states, showcasing the significance of multimodal approaches. While valuable for studying neurophysiological responses, these datasets offer limited applicability in classroom environments, where non-intrusive data collection methods are preferred.

For text-based emotion analysis, datasets like GoEmotions (comprising 58k annotated Reddit comments) provide granular emotional categorization, demonstrating the scalability of linguistic-based affect detection (Demszky et al., 2020). In contrast, FER-2013 (Zahara et al., 2020) and AffectNet (Mollahosseini et al., 2019) focus on facial emotion recognition through deep learning techniques, paving the way for real-time classroom emotion detection. AffectNet, in particular, is one of the largest datasets for valence-arousal computation, making it a key resource for facial expression analysis in natural settings.

Datasets specifically designed for child emotion recognition remain scarce. The Child Affective Facial Expression (CAFE) Set (LoBue & Thrasher, 2015) and KDEF-PT (Garrido & Prada, 2017) provide age-specific facial emotion data, addressing some of the unique challenges associated with recognizing emotions in younger populations. However, these datasets are often static and lack the dynamic, interaction-based data necessary for understanding emotional fluctuations in learning environments.

The application of emotion recognition in education has gained traction in recent years. Some studies have explored how affective computing can improve e-learning experiences by analyzing students' engagement and focus levels (Krithika L.B & Lakshmi Priya GG, 2016). Others have integrated AI driven monitoring systems to track student behavior and provide feedback, enabling adaptive teaching methods in real-time (Parambil et al., 2022). Additionally, deep learning models have been employed to assess student engagement and classroom interactions, offering insights into how emotional states influence learning outcomes (Li & Yao, 2023).

Several studies emphasize the importance of emotions in shaping classroom dynamics. Research such as Emotions Matter underscores how sentiment analysis can enhance student experiences by tailoring instruction to emotional responses (Anwar et al., 2023). Similarly, approaches leveraging facial expression analysis have been used to adapt teaching strategies, demonstrating the practical applications of emotion recognition in classroom settings (Ramos et al., 2020), (Chu et al., 2018).

Despite these advances, a significant gap remains in the development of datasets and systems specifically designed for middle childhood (ages 5–12) a stage where emotional dynamics play a pivotal role in cognitive and social development. Many existing datasets prioritize adult populations or focus on non-educational settings, limiting their applicability to young learners in real-world classroom environments. Addressing this gap requires the creation of context-aware, multimodal datasets that accurately capture children's emotional states and interactions in natural learning conditions.

3. Proposed Framework

3.1. Proposed Dataset

The dataset was gathered in a classroom environment utilizing two distinct seating configurations to guarantee varied data capture. The Straightlook Layout arranged desks in a U-shaped pattern, fostering interaction while providing an unobstructed camera view. In contrast, the Forward-Facing Layout positioned all tables toward the front, ensuring a uniform perspective for facial expression analysis.

A total of 1 hour and 15 minutes of classroom recordings were processed through three primary stages: video segmentation, face extraction, and data preprocessing. The recordings were divided into 0.2-second frames, and YOLOv11-Face was employed to extract and normalize faces to 160×160 pixels for standardized emotion and identity recognition. The dataset, hosted at SmartClass, was annotated with seven emotions—angry, disgusted, fear, happy, sad, surprise, and neutral using a two step labeling process that combined automatic classification via DeepFace with manual verification to ensure accuracy.

The final dataset comprises 50,000 images, augmented through techniques such as rotations, cropping, brightness/contrast adjustments, and the application of Gaussian noise, ensuring both robustness and real world adaptability. This dataset addresses a critical gap in classroom specific emotion recognition, thereby supporting advancements in AI-driven educational analysis. Additionally, it includes 100 biometric prints (fingerprints/NFC card data) from 100 different individuals, which enhances identity recognition and security applications. A multicamera setup was also integrated, offering comprehensive coverage and further improving the accuracy of facial expression and identity recognition. In summary, the final dataset fills an essential gap in classroom specific emotion recognition, paving the way for further advancements in AI-driven educational analysis.

3.2. SmartClass Framework

The SmartClass framework seamlessly integrates face detection, emotion classification, identity recognition, and classroom effectiveness evaluation into a unified system designed to analyze and enhance emotional dynamics in educational settings. Utilizing a multicamera setup, SmartClass ensures precise analysis by mitigating occlusions, motion variations, and other challenging classroom conditions. The system workflow is illustrated in Figure 1.

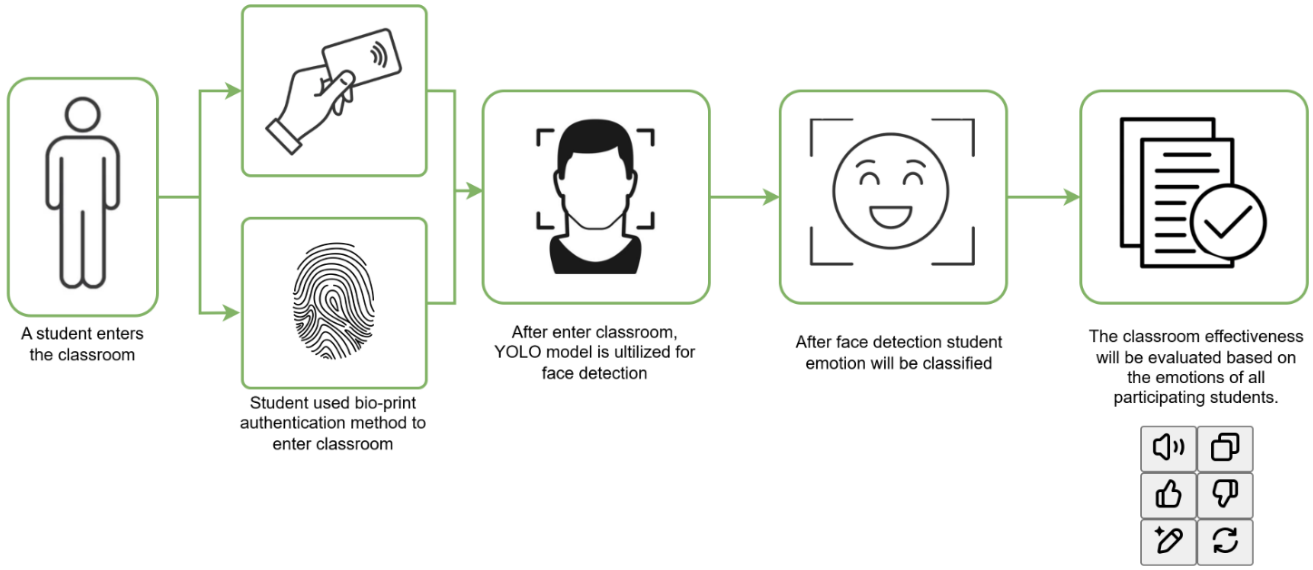
A student enters the classroom
Student used bio-print authentication method to enter classroom
After enter classroom, YOLO model is ultilized for face detection
After face detection student emotion will be classified
The classroom effectiveness will be evaluated based on the emotions of all participating students.

Fig 1. SmartClass Framework Workflow

The process commences with face detection and tracking. YOLOv11-Face detects faces from multiple camera feeds, and overlapping detections are consolidated using a centroid-based algorithm to maintain a unique representation. Continuous tracking is achieved via the ByteTrack algorithm, ensuring robust performance even in scenarios involving occlusion or rapid movement.

Subsequently, emotion classification is performed using a model trained on the SmartClass dataset, which categorizes emotions into seven distinct states, when the classifier's confidence in a particular emotion reaches or exceeds (≥ 70%), that emotion is considered the primary state for the subject (Zeng et al., 2009), (Gunes & Schuller, 2013). In other words, if one emotion accounts for over 70% of the classification output, it is deemed the dominant emotion, while predictions falling below this threshold are treated as ambiguous and thus discarded. This threshold-based strategy helps to ensure that only highly reliable predictions contribute to the final emotional assessment. Moreover, integrating data from multiple cameras further strengthens the reliability of the classification by reducing the impact of potential misclassifications from any single viewpoint.

For identity recognition, ArcFace ensures the accurate association of emotional states with individual identities throughout the session. Additionally, the dataset includes 100 biometric prints (fingerprint/NFC card data) from specific individuals, providing an extra layer of security when students enter the classroom.

The multicamera system comprises four strategically placed cameras positioned at the front, back, and sides of the classroom to offer a 360-degree view. This configuration significantly enhances facial recognition accuracy by capturing multiple perspectives, reducing occlusion, and filtering out low-confidence detections. Figure 2 illustrates this setup.

Lastly, the framework evaluates classroom effectiveness using a positive emotion policy. Sessions are considered effective if at least 70% of students display happy, surprise, or neutral emotions over a given period. This evaluation method aligns with research suggesting that a predominantly positive emotional climate is indicative of effective classroom engagement. For example, in (Calvo & D'Mello, 2010) discuss how real-time affect detection can be leveraged to inform pedagogical strategies by monitoring students' emotional states. In parallel, Fredrickson's broaden-and-build theory posits that positive emotions—such as happiness and surprise—expand cognitive and social resources, ultimately enhancing learning and classroom performance (Fredrickson, 2001). Although the specific threshold of 70% is a design parameter, it is conceptually supported by these findings, implying that when a substantial majority of students exhibit positive or neutral emotions, the session is likely fostering an effective learning environment. Furthermore, the system generates real-time analytics, including visual graphs that track positive emotions over time, enabling educators to optimize teaching strategies and enhance student engagement.

4. Model Training and Framework Performance

The SmartClass framework utilizes YOLOv11 for emotion classification, having been trained on the SmartClass dataset. It achieved an accuracy of 90.0% on the original dataset and 92.8% on the augmented dataset, outperforming both the Detection Transformer (DET) and ResNet50, as detailed in Table 1. The confusion matrix of all models are also illustrated in Figure 3 bellow, indicate their proficiency in emotion classification process.

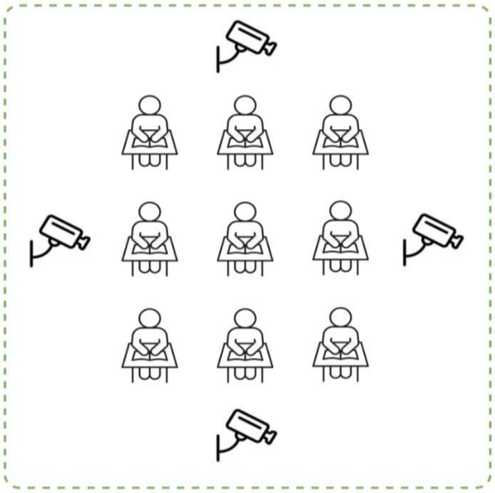

Fig 2. Classroom setup for the multicamera method of the SmartClass framework. Four cameras are strategically placed around the classroom environment to capture students' emotions comprehensively.

Table 1. Emotion Accuracy Classification Results.

Model	YOLOv11	DETR	ResNet50
Original Dataset	0.910	0.741	0.700
Augmented Dataset	0.932	0.681	0.683

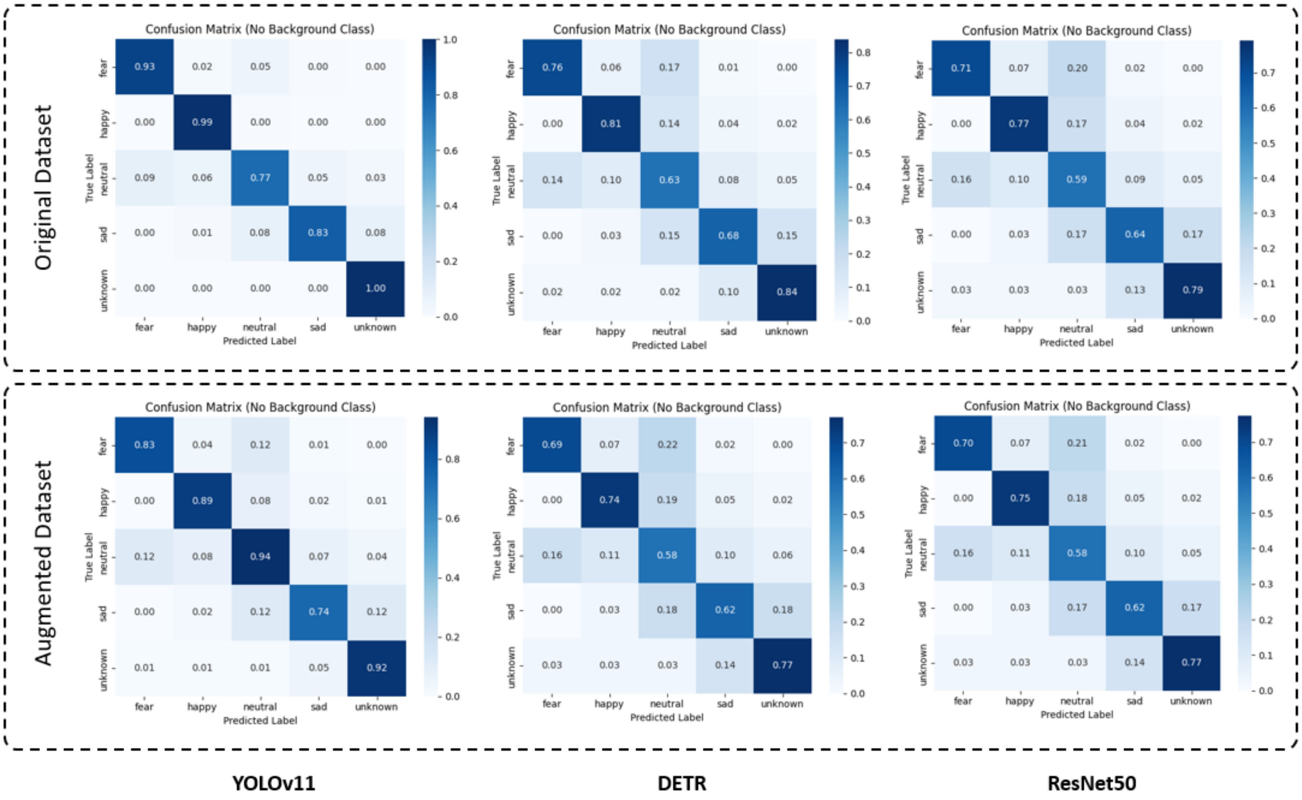

Fig 3. Confusion Matrices of YOLOv11, DETR, and ResNet50 on Original and Augmented Datasets.

Identity recognition is performed using ArcFace, which achieved an accuracy of 99.8% with Labeled Face In the Wild dataset (Huang et al., 2008), surpassing other models as shown in Table 2. Additionally, the system incorporates biometric prints from 100 individuals, attaining an accuracy of 98.6% and thereby enhancing identification reliability.

The system processes frames at 7.2 FPS with an average confidence score of 0.93, ensuring real-time operation. A video demonstration is available at this link.

Table 2. Identity and Biometric Recognition Accuracy.

Model	ArcFace	VGGFace	FaceNet	Biometric Recognition (Bio-Print)
Accuracy	0.998	0.9871	0.996	0.986

5. Conclusion and Discussion

5.1 Conclusion

This study successfully developed SmartClass, a scalable framework for analyzing classroom dynamics and recognizing emotions in children aged 5–12. By leveraging the SmartClass dataset and state of the art models YOLOv11-Face for face detection, YOLOv11-Classify for emotion recognition, and ArcFace for identity verification the system achieved an overall accuracy of 91.8%. The incorporation of a multicamera setup further enhanced reliability by mitigating occlusions and improving classification accuracy.

Beyond real-time analytics, SmartClass offers policy evaluation based on positive emotions, assisting educators in refining their teaching strategies. By addressing missing data and evaluating both verbal and non-verbal dynamics, the framework significantly advances educational technology.

5.2 Discussion

SmartClass demonstrates the transformative potential of AI in education. The high accuracy of YOLOv11 underscores the necessity for robust models in real-world classroom environments, while ArcFace's biometric tracking ensures dependable identity recognition even in dynamic settings.

However, challenges remain. Misclassification of similar emotional expressions such as disgust versus surprise—indicates a need for more advanced feature extraction and enriched training data. Moreover, although multicamera integration significantly improves accuracy, its adaptability to diverse cultural and environmental contexts requires further exploration.

Future work should focus on expanding the SmartClass dataset by incorporating regional, racial, and linguistic variations to enhance cross-cultural applicability. Additionally, integrating further modalities, such as audio analysis, could enrich classroom insights and support real-time feedback mechanisms.

REFERENCES

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of Effective Teacher–Student Interactions in Secondary School Classrooms: Predicting Student Achievement With the Classroom Assessment Scoring System—Secondary. *School Psychology Review*, *42*(1), 76–98. https://doi.org/10.1080/02796015.2013.12087492

Anwar, A., Rehman, I. U., Nasralla, M. M., Khattak, S. B. A., & Khilji, N. (2023). Emotions Matter: A Systematic Review and Meta-Analysis of the Detection and Classification of Students' Emotions in STEM during Online Learning. *Education Sciences*, *13*(9), Article 9. https://doi.org/10.3390/educsci13090914

Braet, C., Theuwis, L., Van Durme, K., Vandewalle, J., Vandevivere, E., Wante, L., Moens, E., Verbeken, S., & Goossens, L. (2014). Emotion Regulation in Children with Emotional Problems. *Cognitive Therapy and Research*, *38*(5), 493–504. https://doi.org/10.1007/s10608-014-9616-x

Brown, F. L., Lee, C., Servili, C., Willhoite, A., Ommeren, M. V., Hijazi, Z., Kieselbach, B., & Skeen, S. (2024). Psychological interventions for children with emotional and behavioral difficulties aged 5–12 years: An evidence review. *Cambridge Prisms: Global Mental Health*, *11*, e75. https://doi.org/10.1017/gmh.2024.57

Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37. IEEE Transactions on Affective Computing. https://doi.org/10.1109/T-AFFC.2010.1

Chu, H.-C., Tsai, W. W.-J., Liao, M.-J., & Chen, Y.-M. (2018). Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Computing*, *22*(9), 2973–2999. https://doi.org/10.1007/s00500-017-2549-z

De Silva, T. R. S., Dayananda, K. Y., Galagama Arachchi, R. C., Amerasekara, M. K. S. B., Silva, S., & Gamage, N. (2022). Solution to Measure Employee Productivity with Employee Emotion Detection. *2022 4th International Conference on Advancements in Computing (ICAC)*, 210–215. https://doi.org/10.1109/ICAC57685.2022.10025132

Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., & Ravi, S. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp.

4040–4054). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.372

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *The American Psychologist*, *56*(3), 218–226. https://doi.org/10.1037//0003-066x.56.3.218

Frenzel, A. C., Daniels, L., & Burić, I. (2021). Teacher emotions in the classroom and their implications for students. *Educational Psychologist*, *56*(4), 250–264. https://doi.org/10.1080/00461520.2021.1985501

Garrido, M. V., & Prada, M. (2017). KDEF-PT: Valence, Emotional Intensity, Familiarity and Attractiveness Ratings of Angry, Neutral, and Happy Faces. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.02181

Gunes, H., & Schuller, B. (2013). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, *31*(2), 120–136. https://doi.org/10.1016/j.imavis.2012.06.016

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). *Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments*.

Jones, S. M., Bailey, R., & Jacob, R. (2014). Social-emotional learning is essential to classroom management. *Phi Delta Kappan*, *96*(2), 19–24. https://doi.org/10.1177/0031721714553405

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31. IEEE Transactions on Affective Computing. https://doi.org/10.1109/T-AFFC.2011.15

Krithika L.B & Lakshmi Priya GG. (2016). Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric. *Procedia Computer Science*, *85*, 767–776. https://doi.org/10.1016/j.procs.2016.05.264

Li, L., & Yao, D. (2023). Emotion Recognition in Complex Classroom Scenes Based on Improved Convolutional Block Attention Module Algorithm. *IEEE Access*, *11*, 143050–143059. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3340510

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01532

Loevaas, M. E. S., Sund, A. M., Patras, J., Martinsen, K., Hjemdal, O., Neumer, S.-P., Holen, S., & Reinfjell, T. (2018). Emotion regulation and its relation to symptoms of anxiety and depression in children aged 8–12 years: Does parental gender play a differentiating role? *BMC Psychology*, *6*(1), 42. https://doi.org/10.1186/s40359-018-0255-y

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, *10*(1), 18–31. https://doi.org/10.1109/TAFFC.2017.2740923

Parambil, M. M. A., Ali, L., Alnajjar, F., & Gochoo, M. (2022). Smart Classroom: A Deep Learning Approach towards Attention Assessment through Class Behavior Detection. *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, 1–6. https://doi.org/10.1109/ASET53988.2022.9735018

Pope, D. J., Butler, H., & Qualter, P. (2012). Emotional Understanding and Color-Emotion Associations in Children Aged 7-8 Years. *Child Development Research*, *2012*(1), 975670. https://doi.org/10.1155/2012/975670

Ramos, A. L. A., Dadiz, B. G., & Santos, A. B. G. (2020). Classifying Emotion based on Facial Expression Analysis using Gabor Filter: A Basis for Adaptive Effective Teaching Strategy. In R. Alfred, Y. Lim, H. Haviluddin, & C. K. On (Eds.), *Computational Science and Technology* (pp. 469–479). Springer. https://doi.org/10.1007/978-981-15-0058-9_45

Veraksa, A. N., Gavrilova, M. N., & Pons, F. (2020). The impact of classroom quality on young children’s emotion understanding. *European Early Childhood Education Research Journal*, *28*(5), 690–700. https://doi.org/10.1080/1350293X.2020.1817240

Verpaalen, I. A. M., Bijsterbosch, G., Mobach, L., Bijlstra, G., Rinck, M., & Klein, A. M. (2019). Validating the Radboud faces database from a child’s perspective. *Cognition and Emotion*, *33*(8), 1531–1547. https://doi.org/10.1080/02699931.2019.1577220

Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I., & Bahri Musa, S. (2020). The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–9. https://doi.org/10.1109/ICIC50835.2020.9288560

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.2008.52

Zheng, W.-L., & Lu, B.-L. (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. IEEE Transactions on Autonomous Mental Development. https://doi.org/10.1109/TAMD.2015.2431497