# Emotion-Aware Adaptive Learning: Enhancing Engagement and Performance in STEM Education Using AI-Powered Emotion Analysis

Nisrine El Ayat[1*], Mohammed Boutalline[1], Adil Tannouche[2], and Hamid Ouanan[3]

[1*]Engineering Laboratory for Intelligent Technologies and Transformation (ELITT-Lab), Higher School of Technology, University Abdelmalek Essaadi, Tetouan, Morocco
n.elayat@uae.ac.ma
boutalline@gmail.com

[2]Laboratory of Engineering and Applied Technology, Higher School of Technology of Beni Mellal , Sultan Moulay Slimane University, Morocco
a.tannouche@usms.ma

[3]Information Processing and Decision Support Laboratory, National School of Applied Sciences of Beni Mellal, Sultan Moulay Slimane University, Beni Mellal, Morocco
ham.ouanan@gmail.com

**Abstract**. Adaptive learning systems that monitor emotions have attracted interest due to their capabilities to customize educational experiences. The process of selecting optimal emotion recognition models faces challenges due to discrepancies in accuracy levels along with computational constraints and requirements for real-time learning capabilities. The researchers conduct a comparative evaluation between existing AI emotion recognition approaches for adaptive learning which includes deep learning models and traditional machine learning methods alongside multimodal fusion techniques.

This research examines various methodologies starting with CNNs used to detect facial expressions followed by LSTM and transformer methods for physiological and textual analysis together with hybrid approaches that merge multiple modalities. An evaluation of each technique relies on accuracy, computational efficiency, scalability, and real-time capability within e-learning systems. The research also investigates both ethical standards and privacy challenges that emerge from using AI emotion analysis in educational contexts.

By providing a benchmark analysis of AI-driven emotion recognition models, this paper aims to guide future research and development in adaptive learning systems. Our finding identified key strengths and limitations of different approaches that can lead to effective integration of emotion analysis within personalized learning modules across STEM education.

**Keywords**: Adaptive learning, emotion recognition, AI in education, deep learning, affective computing, intelligent tutoring systems.



M. Khaldi (ed.), *Proceedings of the E-Learning and Smart Engineering Systems (ELSES 2025)*, Atlantis Highlights in Social Sciences, Education and Humanities 46,
https://doi.org/10.2991/978-94-6239-634-0_21

# 1 Introduction

AI-enabled education has revolutionized the personalization and adaptation of learning experiences, especially for STEM teaching. Adaptive learning environments design content and pace based on learner preferences to enhance engagement and maximize learning outcomes. Conventional systems usually rely on cognitive or behavioral indices, such as performance metrics or response times, without taking into account the learners' emotional states, which greatly affect their motivation, attention, and retention [1].

That void is filled by emotion-aware adaptive learning systems that employ affective computing techniques for detecting and reacting to students' emotional cues in real time. The development of AI, especially from the disciplines of deep learning and multimodal data fusion, has made the recognition of emotions increasingly capable, scalable, and diversified across different modalities such as facial expressions and physiological signals (e.g., EEG, ECG), and textual cues [2, 3]. This scenario paints a favorable option for making adaptive learning systems more enriched by adapting pedagogical strategies, feedback, and content delivery dynamically based on learners' emotional states.

Choosing the right emotion recognition models remains a very serious challenge, despite all these advances. Existing approaches being considered in the past range widely with respect to accuracy, computational cost, modality of data accepted, and degrees of appropriateness to real-time pedagogical settings [2]. Also, they lack an extensive comparison of these choices in the context of adaptive learning for STEM education, where cognitive load and motivation are attributed to affective states [1, 4]

This paper looks at the AI-powered emotion recognition techniques and their applications to adaptive learning systems via a structured conceptual analysis. It proposes a taxonomy of models, sorting them according to input modalities and algorithmic approaches, and discussing their pros and cons from the educationally relevant criteria, along with talking about important ethical and design implications that result from this. The research and implementation of emotion-aware models in support of increased engagement and performance in STEM environments is thereby hoped to be aided by this framework.

# 2 Background and Related Work

Artificial intelligence technology enables adaptive learning methods that respond to individual emotions in STEM education to boost student participation and academic achievement [5, 6].The connection between student emotions and attention as well as motivation and cognitive functions creates significant potential for using affective feedback in intelligent systems [7, 8].

### 2.1 Emotion Recognition in Adaptive Learning

Researchers have been investigating how emotion detection systems can be integrated into adaptive learning systems through recent research. Bala et al. developed an adaptive electronic learning system which uses facial emotion detection for on-the-fly content modifications that resulted in better student involvement and content satisfaction [5]. Salloum et al. examined the possibilities of artificial intelligence in recognizing emotions from facial expressions to optimize learning effectiveness through emotional response systems [6].

### 2.2 AI Frameworks for Emotion Detection

Educational settings require robust AI frameworks to achieve effective detection of emotions. A new AI framework called Emotion AWARE enables adaptable and explainable detection of multi-granular emotional expressions to handle the complex digital emotion expressions in learning environments [8]. The researchers behind the study conducted by Vistorte et al. performed a systematic literature review to evaluate artificial intelligence applications in educational emotion assessment while examining present obstacles and their practical implementations [7].

### 2.3 Multimodal Emotion Recognition

AI progress brought about the development of multimodal emotion recognition systems, which combine facial expressions with speech and physiological responses and behavior logs. Through deep learning models that include convolutional and recurrent neural networks, these systems analyze emotional data in real-time with high dimensionality [6]. The integration of different input sources enables these systems to detect complex emotional states, which leads to better educational content adjustment.

### 2.4 Ethical Considerations

Emotion-aware systems create multiple educational advantages but generate important ethical issues when they are implemented in learning institutions. The implementation of AI in educational settings requires addressing four core problems, which are data privacy violations, informed consent issues, emotional manipulation risks and algorithmic bias challenges [7, 8]. Organizations must develop systems, which provide transparency and understandability to users while allowing regulatory control to achieve stable operational deployment.

## 3 Taxonomy of Emotion Recognition Models

The classification of emotional data interpretation systems for adaptive learning occurs through the alignment of three main factors which include input modality, algorithmic approach and fusion level. The analysis framework establishes an organized method for studying how learning systems collect and analyze emotional information to improve individualized learning and student involvement [3, 9, 10].

As presented above in Figure 1, this taxonomy has been visually summarized to illustrate the relationships between data types, model architectures, and some of the fusion strategies frequently used in emotion-aware educational systems.

### 3.1 Classification by Input Modality

Emotion recognition systems base their operations on various learner data types:

- **Facial Expression:** Computer vision systems detect emotions through muscle movement analysis and facial landmark evaluation. Real-time facial expression detection benefits from the superior accuracy of CNN models that identify minor facial expression details [9].
- **Speech and Audio:** The detection of emotional content in speech depends on analyzing pitch changes along with tone variations and tempo fluctuations. The combination of BiLSTM and CNN models demonstrates strong performance in vocal emotion recognition tasks [11].
- **Textual Data:** Natural language processing models including BERT analyze learner text input from chats, essays and discussion forums to detect both sentiment and emotional tone [12].
- **Physiological Signals:** Bio-signals like EEG, ECG, and EDA provide detailed information about emotional states. Time-series data analysis benefits from LSTM models as their main tool of choice [13].

### 3.2 Classification by Algorithmic Approach

Models fall into groups based on their input characteristics and system objectives:

- **Traditional Machine Learning**: The performance of Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN) algorithms remains steady for basic features, but they show limitations in recognizing intricate emotional patterns [14].
- **Deep Learning:** Through the development of CNNs for visual data and LSTMs for sequences and transformers for text input, recognition performance dramatically improved across multiple domains [15, 16].
- **Hybrid Architectures:** Multimodal systems use the integration of CNN with LSTM and attention-based models to process various input types and complex time-dependent relationships [3].

### 3.3 Classification by Fusion Strategy

Combining different emotion channels becomes essential when developing multimodal systems:

- **Early Fusion:** Systematically integrates raw features from multiple modalities through the learning process. This method maintains simplicity yet struggles to keep specific modal characteristics intact [17].

- **Late Fusion:** The system deals with every modality separately by unifying their results. This approach proves valuable in contexts where modalities operate at different times [18].

- **Intermediate (Deep) Fusion:** Hidden layers inside deep networks enable the integration of various modalities. The system delivers improved performance while requiring additional complexity to operate [10].

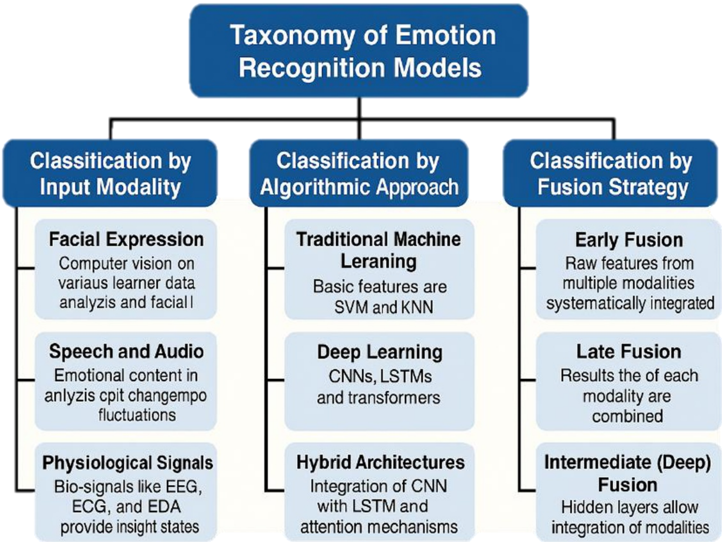


***Figure 1:*** *Taxonomy Recap: Emotion Recognition Models in Adaptive Learning*

# 4 Comparative Analysis Framework

Emotion recognition models in adaptive learning systems can be evaluated across multiple dimensions to determine their suitability for educational applications. This section presents a comparative analysis based on four key frameworks: **Technical Performance, Educational Utility, Multimodal Integration Capability, and Ethical and Practical Feasibility** [12, 19–21].

## 4.1 Evaluation Frameworks

- **Technical Performance:** Analyzes the model's precision and operational speed alongside its ability to function in real-time situations [19, 20].
- **Educational Utility:** The method assesses how well the model helps students learn better while improving their participation and educational context flexibility [12, 21].

- **Multimodal Integration Capability:** The evaluation measures how well the model can handle processing different types of data, such as facial expressions and speech, along with physiological signals [19, 22].
- **Ethical and Practical Feasibility:** The evaluation process involves analyzing privacy risks along with intrusive elements as well as deployment costs and real-world implementation simplicity [21, 23].

## 4.2 Comparative Analysis Table

**Table 1.** Summary of dataset features and metrics.

| Model | Modality | Technical Performance | Educational Utility | Multimodal Integration | Ethical & Practical Feasibility | References |
|---|---|---|---|---|---|---|
| **CNN (Convolutional Neural Network)** | Facial Expressions | High | Medium | Low | Medium | [22, 24] |
| **LSTM (Long Short-Term Memory)** | EEG/Physiological Signals | Medium | High | Medium | Low | [20, 25] |
| **Transformer (e.g., BERT)** | Textual Data | High | High | Medium | High | [12, 21] |
| **CNN-LSTM Hybrid** | Facial + EEG | High | High | High | Medium | [19, 26] |
| **Multimodal Transformer** | Facial + Audio + Text | High | High | High | Medium | [19, 22] |

## 4.3 Observations

The real-time facial emotion recognition capability of CNNs exists but they cannot effectively detect complex emotional states through visual data processing alone [9, 24].

LSTM models demonstrate superior performance in analyzing sequential physiological data to provide internal emotional state insights but their implementation requires specific sensor technologies [20, 25]

The BERT model among other Transformers delivers exceptional results in emotional text analysis especially in situations with abundant textual data [12, 21].

The integration of facial and physiological signals through CNN-LSTM Hybrid models results in superior emotion recognition accuracy by combining spatial and temporal data processing techniques [19].

The integration of multiple data modalities through Multimodal Transformers ensures complete emotion analysis but requires extensive computational power and privacy protection measures [22, 27].

# 5 Ethical and Practical Challenges

The combination of emotion recognition technologies with adaptive learning systems generates both technical capabilities and considerable ethical and practical obstacles. The implementation of these systems needs proper handling of technical and ethical matters because they operate in educational settings where learners' rights to privacy and autonomy should receive top priority.

## 5.1 Privacy and Consent

Adaptive learning systems need emotional data, including facial images and voice patterns, and physiological signals (e.g., EEG, heart rate) to function. The collection and processing of these data types generate substantial privacy challenges. Learners face the danger of surveillance and personal information misuse without receiving proper consent and adequate data governance frameworks that protect their rights [28, 29].

School administrators face major difficulties when trying to receive proper authorization from students as well as informing parents about these systems [30].

## 5.2 Bias and Fairness

Artificial intelligence emotion recognition models develop biases that emerge from gender, race, and cultural disparities within their training datasets. The combination of Western facial expression norms with non-Western faces causes misclassification in datasets that contain an excessive representation of Western facial expressions [31]. The unequal distribution of educational support and feedback becomes intensified when certain groups experience marginalization through these disparities [32].

## 5.3 Intrusiveness and Psychological Impact

Learners might feel uncomfortable and develop performance-related anxiety when they encounter physiological sensors such as EEG headsets and webcams. The benefits of adaptive learning are nullified when students face over-monitoring, which leads to autonomy reduction and stress while suppressing their intrinsic motivation [33].

## 5.4 Cost and Infrastructure

System deployment for multimodal emotion recognition systems demands expensive hardware such as wearable sensors and high-resolution cameras, and powerful computational backends. Educational institutions operating under budget constraints find it challenging to implement these systems in their infrastructure, especially in areas with limited resources [34].

### 5.5 Explainability and Transparency

All educators and students need to comprehend the methods used to adjust learning materials according to emotional states. The inner workings of many deep learning models, such as transformers and multimodal fusion systems, remain mostly unknown to users. When systems lack explainability, students and teachers lose trust in the model, and teachers become unable to react to model errors [35].

## 6 Discussion and Conclusions

Various educational systems can experience improved responsiveness and adaptability through the application of emotion recognition technologies. Real-time identification of emotional states enables these systems to adapt their content delivery process while delivering instant feedback along with motivational support in demanding STEM subjects. These systems require pedagogically sound design alongside meaningful learning outcome focus rather than relying on brief emotional triggers as their design principle [36].

The implementation of facial expressions together with speech and physiological data in multimodal systems shows better performance results than systems, which deploy only single modalities. The implementation of these systems presents various technical issues alongside hardware requirements and deployment scalability obstacles for schools with limited resources [3]. Any educational environment that implements emotional analysis tools must assess the value of integrating these assessments with respect to their costs.

The systems present severe ethical challenges because they handle sensitive emotional and biometric data. The implementation of emotional learning technologies demands the strongest possible privacy protection together with explicit user agreements and transparent data handling methods to prevent monitoring practices while maintaining learner independence [37]. Educational institutions need to establish AI guidelines for ethical design practices along with clearly specified boundaries regarding emotion detection technology in educational environments.

Another challenge involved in this case is the biases inherent in training datasets. In general, any emotion recognition procedures, especially those based on facial expression or speech; replicate the constraints of culture and demographics inherent in the basis itself. Hence, the chance of such systems incorrectly labelling emotions from underrepresented backgrounds increases, which further impairs the fairness of the adaptive response [38]. Such a problem must therefore be addressed with both diverse datasets and fairness-aware training algorithms.

Finally, for emotion-aware systems to see wide and responsible adoption, explainability and human-AI collaboration stand out as two critical approaches in future work. Teachers need to understand how emotion-based recommendations are derived while granting learners control over the use of their emotional data. Integration of these technologies into the learning process in support of not in place of human judgment will require interpretable models and transparent adaptation mechanisms [39].

## References

1. Pekrun R (2006) The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. Educ Psychol Rev 18:315–341. https://doi.org/10.1007/s10648-006-9029-9

2. Kosti R, Alvarez JM, Recasens A, Lapedriza A (2017) Emotion Recognition in Context. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Honolulu, HI, pp 1960–1968

3. Soleymani M, Garcia D, Jou B, Schuller B, Chang S-F, Pantic M (2017) A survey of multimodal sentiment analysis. Image and Vision Computing 65:3–14. https://doi.org/10.1016/j.imavis.2017.08.003

4. D'mello S, Graesser A (2012) AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans Interact Intell Syst 2:1–39. https://doi.org/10.1145/2395123.2395128

5. Bala MM, Akkineni H, Sirivella SA, Ambati S, Potharaju Venkata Sai KV (2023) Implementation of an adaptive E-learning platform with facial emotion recognition. Microsyst Technol 29:609–619. https://doi.org/10.1007/s00542-023-05420-1

6. Salloum SA, Alomari KM, Alfaisal AM, Aljanada RA, Basiouni A (2025) Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. Smart Learn Environ 12:21. https://doi.org/10.1186/s40561-025-00374-5

7. Vistorte AOR, Deroncele-Acosta A, Ayala JLM, Barrasa A, López-Granero C, Martí-González M (2024) Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. Front Psychol 15:1387089. https://doi.org/10.3389/fpsyg.2024.1387089

8. Gamage G, De Silva D, Mills N, Alahakoon D, Manic M (2024) Emotion AWARE: an artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis. J Big Data 11:93. https://doi.org/10.1186/s40537-024-00953-2

9. Zhang Z, Luo P, Loy CC, Tang X (2018) From Facial Expression Recognition to Interpersonal Relation Prediction. Int J Comput Vis 126:550–569. https://doi.org/10.1007/s11263-017-1055-1

10. Baltrusaitis T, Ahuja C, Morency L-P (2019) Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans Pattern Anal Mach Intell 41:423–443. https://doi.org/10.1109/TPAMI.2018.2798607

11. Munshi A (2022) Improved MQTT Secure Transmission Flags in Smart Homes. Sensors 22:2174. https://doi.org/10.3390/s22062174

12. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S (2020) GoEmotions: A Dataset of Fine-Grained Emotions. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics 4040–4054. https://doi.org/10.18653/v1/2020.acl-main.372

13. Alhagry S, Aly A, A. R (2017) Emotion Recognition based on EEG using LSTM Recurrent Neural Network. ijacsa 8:. https://doi.org/10.14569/IJACSA.2017.081046

14. Bhuiyan A, Nath B (2006) Anti-personnel Mine Detection and Classification Using GPR Image. In: 18th International Conference on Pattern Recognition (ICPR'06). IEEE, Hong Kong, China, pp 1082–1085

15. Li Y, Zeng J, Shan S, Chen X (2019) Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism. IEEE Trans on Image Process 28:2439–2450. https://doi.org/10.1109/TIP.2018.2886767

16. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention Is All You Need

17. Liu W, Qiu J-L, Zheng W-L, Lu B-L (2019) Multimodal Emotion Recognition Using Deep Canonical Correlation Analysis

18. Nojavanasghari B, Gopinath D, Koushik J, Baltrušaitis T, Morency L-P (2016) Deep multimodal fusion for persuasiveness prediction. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, Tokyo Japan, pp 284–288

19. Guo Z, Yang M, Lin L, Li J, Zhang S, He Q, Gao J, Meng H, Chen X, Tao Y, Yang C (2024) E-MFNN: an emotion-multimodal fusion neural network framework for emotion recognition. PeerJ Computer Science 10:e1977. https://doi.org/10.7717/peerj-cs.1977

20. Algarni M, Saeed F, Al-Hadhrami T, Ghabban F, Al-Sarem M (2022) Deep Learning-Based Approach for Emotion Recognition Using Electroencephalography (EEG) Signals Using Bi-Directional Long Short-Term Memory (Bi-LSTM). Sensors 22:2976. https://doi.org/10.3390/s22082976

21. Zhou E, Khatri K, Zhao Y, Krishnamachari B (2025) AffectEval: A Modular and Customizable Framework for Affective Computing

22. Cai L, Hu Y, Dong J, Zhou S (2019) Audio-Textual Emotion Recognition Based on Improved Neural Networks. Mathematical Problems in Engineering 2019:2593036. https://doi.org/10.1155/2019/2593036

23. Vistorte AOR, Deroncele-Acosta A, Ayala JLM, Barrasa A, López-Granero C, Martí-González M (2024) Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. Front Psychol 15:1387089. https://doi.org/10.3389/fpsyg.2024.1387089

24. Sharma M, Mittal A, Rewliya A, Mittal H (2024) Comparative Analysis of Deep Learning Models for Facial Emotion Recognition. In: 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT). IEEE, Gwalior, India, pp 1–6

25. Han Z, Chang H, Zhou X, Wang J, Wang L, Shao Y (2022) E2ENNet: An end-to-end neural network for emotional brain-computer interface. Front Comput Neurosci 16:942979. https://doi.org/10.3389/fncom.2022.942979

26. Wang Y, Guo X, Zhang Y, Ren Y, Huang W, Liu Z, Feng Y, Dai X, Zhang W, Che H (2023) An Adaptive Framework of Multimodal Emotion Recognition Based on Collaborative Discriminative Learning. In: 2023 15th International Conference on Advanced Computational Intelligence (ICACI). IEEE, Seoul, Korea, Republic of, pp 1–8

27. Baltrusaitis T, Ahuja C, Morency L-P (2019) Multimodal Machine Learning: A Survey and Taxonomy. IEEE Trans Pattern Anal Mach Intell 41:423–443. https://doi.org/10.1109/TPAMI.2018.2798607

28. Akar E (2024) Emotional Artificial Intelligence: Introducing the Concept of 'Emotional Privacy.' In: Prifti K, Demir E, Krämer J, Heine K, Stamhuis E (eds) Digital Governance. T.M.C. Asser Press, The Hague, pp 65–87

29. Koops B-J, Leenes R (2014) Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. International Review of Law, Computers & Technology 28:159–171. https://doi.org/10.1080/13600869.2013.801589

30. Anthonysamy P, Greenwood P, Rashid A (2014) A Method for Analysing Traceability between Privacy Policies and Privacy Controls of Online Social Networks. In: Preneel B, Ikonomou D (eds) Privacy Technologies and Policy. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 187–202

31. Rhue L (2018) Racial Influence on Automated Perceptions of Emotions. SSRN Journal. https://doi.org/10.2139/ssrn.3281765

32. Katirai A (2024) Ethical considerations in emotion recognition technologies: a review of the literature. AI Ethics 4:927–948. https://doi.org/10.1007/s43681-023-00307-3

33. D'Mello S, Graesser A (2012) Dynamics of affective states during complex learning. Learning and Instruction 22:145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

34. Pei G, Li H, Lu Y, Wang Y, Hua S, Li T (2024) Affective Computing: Recent Advances, Challenges, and Future Trends. Intell Comput 3:0076. https://doi.org/10.34133/icomputing.0076

35. Lipton ZC (2018) The mythos of model interpretability. Commun ACM 61:36–43. https://doi.org/10.1145/3233231

36. Bosch N, D'Mello S, Baker R, Ocumpaugh J, Shute V, Ventura M, Wang L, Zhao W (2015) Automatic Detection of Learning-Centered Affective States in the Wild. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. ACM, Atlanta Georgia USA, pp 379–388

37. Katirai A (2024) Ethical considerations in emotion recognition technologies: a review of the literature. AI Ethics 4:927–948. https://doi.org/10.1007/s43681-023-00307-3

38. McDuff D, Gontarek S, Picard R (2014) Remote measurement of cognitive stress via heart rate variability. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Chicago, IL, pp 2957–2960

39. Rachha A, Seyam M (2023) Explainable AI In Education : Current Trends, Challenges, And Opportunities. In: SoutheastCon 2023. IEEE, Orlando, FL, USA, pp 232–239