

Desarrollo de una aplicación para evaluar la atención de los estudiantes universitarios a partir de posturas corporales y expresiones faciales

Christian Vasconez¹  and Kenneth Cortez¹ 

¹ Departamento de Ciencias de la Computación, Universidad de las Fuerzas Armadas
ESPE, Sangolquí, Ecuador
{kacortez,cgvasconez2}@espe.edu.ec

Abstract. Keywords: - - - - -

1 Introducción

En los últimos años, las tecnologías de inteligencia artificial (IA) se han ido incorporando gradualmente al campo de la educación, transformando aspectos como el seguimiento del rendimiento académico y la personalización del aprendizaje, con el objetivo de aumentar las capacidades cognitivas de los estudiantes; Vieri y Petrea [20] han afirmado que estas herramientas generan entornos adaptativos, capaces de ofrecer retroalimentación automatizada y monitoreo constante del progreso del estudiante.

Liu et al. [12] han explorado la aplicación de técnicas de visión por computadora, documentando cómo el análisis de las expresiones faciales, la postura corporal y la dirección de la mirada pueden ser puntos clave con los que se puede inferir el nivel de atención durante, por ejemplo, clases virtuales.

Una de las primeras contribuciones en esta área fue el desarrollo de un modelo de evaluación de la atención basados en el análisis postural el cual fue propuesto por Abate et al. [1], quienes de acuerdo con los resultados de su prototipo, han demostrado que se pueden correlacionar ciertos gestos corporales con el estado de concentración o distracción durante el aprendizaje virtual sincrónico.

Este aporte ha logrado sentar las bases para investigaciones posteriores como las de Ikram [10] y Márquez-Carpintero [15] quienes han logrado ampliar estos modelos combinando datos posturales con el reconocimiento facial de las emociones para mejorar la precisión de la predicción.

Dentro de latinoamérica de igual manera se han logrado el desarrollo de modelos con el mismo objetivo, y entre ellos se encuentra el de Piedrahíta-Carvajal et al. [17] en donde se diseñó una herramienta web que integra análisis de emociones con seguimiento visual de la atención durante las clases virtuales, facilitando la supervisión del profesor en tiempo real. Hubo una investigación en donde incluso se ha podido implementar lo que Farsani y Villa-Ochoa [7] han nombrado como "gafas espía" las cuales son capaces de capturar la interacción conductual de los estudiantes, lo que puede dar indicios a sus rasgos de personalidad así como la expresividad del profesor.

Estas investigaciones son las bases académicas y tecnológicas que sustentan la propuesta de una aplicación web que pueda monitorear automáticamente la atención de los estudiantes universitarios mediante el análisis de sus posturas y expresiones faciales.

En el contexto actual académico universitario, el uso de las tecnologías para llevar acabo el desarrollo de actividades académicas se ha vuelto de uso frecuente, esto conlleva a un aumento del tiempo de los estudiantes al estar frente a un computador. Este suceso ha sido asociado con el aumento de distracciones y perdida de atención de los estudiantes durante las actividades académicas realizadas virtualmente [19]. Además, factores como el mantener una mala postura corporal en un largo período de tiempo estando en sedestación, afecta directamente en la atención y el aprendizaje, debido al aumento de tensión muscular y la incomodidad física que provoca este [4]. En la actualidad, existen pocas aplicaciones que se integren en entornos virtuales de aprendizaje que permitan evaluar al mismo tiempo la atención, el estado emocional, la postura corporal de los estudiantes y que ofrezcan una solución.

Este proyecto tiene como objetivo desarrollar una aplicación que evalúe la atención de los estudiantes universitarios mediante el análisis de posturas corporales y expresiones faciales para identificar su nivel de atención y generar retroalimentación de utilidad para estudiantes y docentes, lo cual delimita su alcance a realizar dicha evaluación a partir de las imágenes captadas por la cámara del computador durante actividades virtuales de estudio. es decir, una evaluación superficial del nivel de atención. La investigación se centrará en el ámbito universitario de la Universidad de las Fuerzas Armadas ESPE. Solo se considerarán indicadores visuales como la alineación del cuerpo, la dirección de la mirada y expresiones faciales. Entre las limitaciones se identifican la necesidad de una conexión estable a internet, condiciones adecuadas de iluminación y dispositivos con cámara funcional. La aplicación tiene como fin complementar el proceso de enseñanza-aprendizaje en el desarrollo de actividades virtuales únicamente con una evaluación superficial del nivel de atención.

2 Trabajos relacionados

Para conformar el grupo de trabajos relacionados, primero se realizó una cadena de búsqueda a partir de los términos más recurrentes de artículos académicos con títulos similares al tema general del proyecto. Dichos artículos se obtuvieron a partir de explorar en la base de datos Google Scholar para, consecuentemente, analizar y recopilar términos recurrentes. De esta manera, se definió un conjunto de palabras clave, mismas que se las agrupó por sinonimia o relación entre palabras, obteniendo la siguiente cadena de búsqueda:

(attention OR engagement) AND (facial expression OR emotion recognition) AND (posture OR body language) AND (machine learning OR artificial intelligence)

La cadena de búsqueda se aplicó en tres bases de datos académicas: ACM, IEEE y Scopus. El proceso de búsqueda en cada base de datos arrojó un total

de 4,199 resultados, siendo Web of Science (n = 64), IEEE Xplore (n = 97) y SCOPUS (n = 4,038). A partir de estos resultados se extrajeron los metadatos, dejando como principales: título, autores, DOI, palabras clave, libro/congreso y año de publicación. Consecuentemente, se aplicó la metodología PRISMA para la selección final de artículos, esta metodología dio como resultado 23 artículos, los cuales cumplían con todos los requisitos de la investigación y fueron seleccionados para conformar el grupo de estudios relacionados.

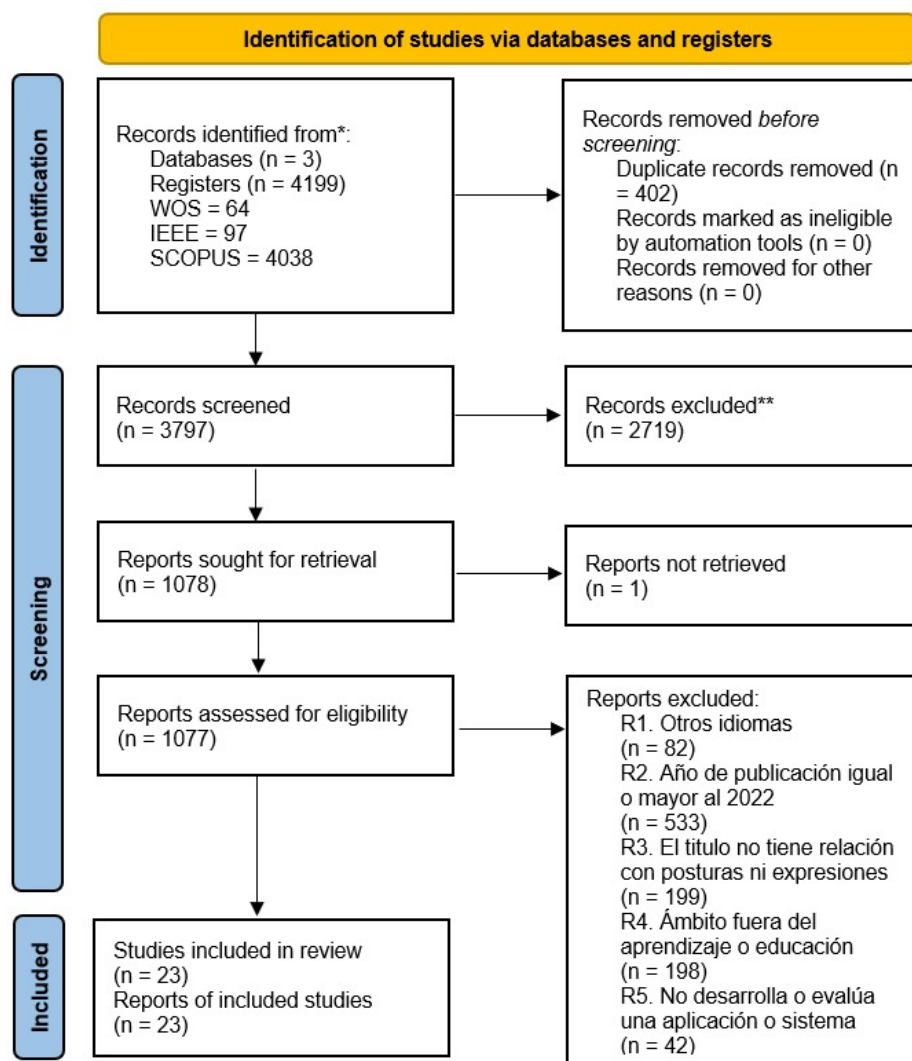


Fig. 1. Diagrama del método PRISMA. Elaboración propia

En base a los artículos relacionados, se logró identificar información significativa que sirve de apoyo y guía para el desarrollo del presente proyecto: desarrollar una aplicación para evaluar la atención de estudiantes universitarios mediante posturas corporales y expresiones faciales. A partir de la tabla 1, se logra destacar la implementación de técnicas de Deep Learning (DL), se observó que el 100% de los autores utilizan estas técnicas, de tal manera que las metodologías basadas en Deep Learning (DL) constituyen el estándar actual para abordar tareas de reconocimiento de emociones y atención por medio de datos visuales como imágenes faciales o registros posturales. Por ejemplo, el sistema AATiENDE propuesto por Escalona [5] utiliza una combinación de CNNs para el análisis de la postura corporal, análisis de expresiones faciales y dirección de la mirada, obteniendo una precisión superior al 95% en entornos controlados. De similar manera, Daza [21] con su modelo de red neuronal DeepFace-Atención, implementa módulos de análisis facial para estimar el nivel de atención, confirmando que las redes neuronales profundas son eficaces en tareas de identificar y diferenciar los estados mentales (cognitivos) y emocionales (afectivos) de una persona. Sin embargo, Yazdani [2] analiza la desalineación emocional entre imágenes y textos, emplearon modelos tradicionales de procesamiento semántico y regresión lineal, lo que indica que, si bien el Deep Learning (DL) es predominante en este tipo de proyecto, algunos enfoques alternativos son igual de válidos dependiendo el enfoque que se le quiera dar.

Respecto al reconocimiento de expresiones faciales, se incorporó en el 70% de los artículos finales. Luo [8] desarrolló un sistema para estimar la intensidad de la sonrisa caracterizándola como emoción positiva, usando DA-CNNs. De similar forma, Kumar [9] exploró más el campo de los rostros animados en contenidos educativos, con el fin de demostrar que las expresiones faciales percibidas son de impacto significativo en la atención y curiosidad de los estudiantes. Por el contrario, Fordson [13], en su sistema HELS, un sistema de reconocimiento emocional confiable, está mayormente basado en señales fisiológicas como Electroencefalografía (EEG) y Respuesta Galvánica de la Piel (GSR), él no consideró incluir las expresiones faciales, mencionando que estas pueden ser ambiguas o poco fiables en ciertos contextos.

En cuanto al análisis de postura corporal, se incorporó en solo el 26% de los artículos finales, es decir, no fue una variable de interés para la mayoría de autores. Huang y Zhou [11] desarrollaron un sistema de detección postural en aulas donde se clasifica en tiempo real si un estudiante está de pie, sentado, entre otras. No obstante, se determina que el resto de autores priorizan expresiones faciales como única fuente de análisis, lo cual, si se enfoca específicamente a identificar el nivel de atención, este limitaría la precisión del mismo.

Una observación de interés es que solo el 52% de los autores abordan la atención como variable de interés principal y de análisis. Por un lado, tenemos a Escalona [5] y Mahmood [14], quienes se enfocaron netamente en la clasificación de la atención de los estudiantes, como atentos o distraídos por medio del reconocimiento facial. Por otra parte, Chaudhari [6] se enfoca en el reconocimiento

emocional, pero no considera la atención como constructor específico, limitando su aplicabilidad al monitoreo educativo.

En relación con el procesamiento en tiempo real, se tiene que el 48% de los autores implementan la toma de datos en tiempo real. Daza [21], y Huang y Zhou [11] demuestran que es viable lograr el análisis visual en vivo con tasas aceptables de Frames Per Second (FPS) mayores a 8, dado que, mientras mayor sea los Frames Per Second (FPS), más fluido y preciso será el seguimiento de la toma de datos en tiempo real. Sin embargo, Chaudhari [3], Zhang [16], entre otros, se limitan al análisis offline sobre datasets previamente grabados.

Por último, se observa que la visión por computadora fue utilizada en un 61% por los autores, siendo una herramienta clave para obtener información visual sin la necesidad de implementar sensores adicionales como lo hizo Chaudhari [6], donde incorpora la visión por computadora para el entrenamiento de CNNs con imágenes de escenas completas. Aun así, Fordson [13], y Mutawa y Hassouneh [18] optan por implementar visión por computadora con señales fisiológicas como Electroencefalografía (EEG) y Respuesta Galvánica de la Piel (GSR), con el fin de mejorar ser más precisos.

Con base en el análisis comparativo de los 23 artículos relacionados, es posible concluir que la pertinencia y viabilidad del presente proyecto, al identificar que el uso de Deep Learning y visión por computadora es una práctica ampliamente adoptada en el reconocimiento de emociones y atención en entornos educativos. Si bien el reconocimiento de expresiones faciales es el enfoque más común, el análisis de postura corporal ha sido poco explorado, lo que representa una oportunidad de innovación. Además, aunque solo la mitad de los estudios abordan explícitamente la atención como variable principal, aquellos que lo hacen demuestran su relevancia en la mejora del aprendizaje. Finalmente, el desarrollo de sistemas en tiempo real y sin sensores invasivos refuerza la aplicabilidad del enfoque propuesto, el cual busca integrar de manera eficiente el análisis facial y postural para evaluar la atención de estudiantes universitarios durante el proceso educativo.

Autores	Expresiones faciales	Posturas corporales	Enfocado en atención	Aplicado en educación	Tiempo real	Visión por computador	Estudiantes universitarios	Sistema funcional	Deep Learning
Escalona et al. (2023)	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Xue & Liu (2024)	Sí	Sí	Sí	Sí	Sí	Sí	No	Sí	Sí
Amish et al. (2024)	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No
Gahlan & Sethia (2025)	No	No	No	Sí	No	No	No	No	Sí
Halkiopoulos et al. (2025)	Sí	No	No	No	No	No	No	No	Sí
Luo et al. (2025)	Sí	No	Sí	Sí	No	Sí	Sí	No	Sí
Kumar (2024)	Sí	No	Sí	Sí	No	Sí	Sí	No	Sí
Mahmood et al. (2024)	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Faria et al. (2024)	No	No	No	No	No	No	No	No	Sí
Banos et al. (2024)	No	No	No	No	No	No	No	No	Sí
Mutawa & Hassouneh (2024)	Sí	No	No	No	Sí	No	No	No	Sí
Samal & Hashmi (2024)	No	No	No	No	No	No	No	No	Sí
Zhang et al. (2024)	Sí	No	No	No	No	No	No	No	Sí
Zhang et al. (ERA, 2024)	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Daza et al. (2024)	Sí	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Huang & Zhou (2023)	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Darejeh et al. (2024)	No	No	No	Sí	No	No	Sí	Sí	Sí
Yazdani et al. (2024)	Sí	No	No	No	No	No	No	No	Sí
Karthika et al. (2023)	No	No	No	Sí	No	No	Sí	Sí	Sí
Zhang et al. (Uni2Mul, 2023)	Sí	No	No	No	No	No	No	No	Sí
Chaudhari et al. (SNCS, 2023)	Sí	No	No	No	No	No	No	No	Sí
Fordson et al. (2023)	No	No	No	No	No	No	No	No	Sí
Chaudhari et al. (CERDL, 2023)	Sí	Sí	No	Sí	No	Sí	Sí	No	Sí

Table 1. Comparación de artículos según características principales

References

1. Abate, A.F., Barsocchi, P., Cimino, M.G.: Attention monitoring for synchronous distance learning. *Future Generation Computer Systems* **127**, 667–678 (2021). <https://doi.org/10.1016/j.future.2021.09.022>
2. Abraham, A., Issac, T., Terrance, J.S.: Investigation of student engagement monitoring system using machine learning. *IEEE ICICI* (2024). <https://doi.org/10.1109/ICICI62254.2024.00017>
3. Banos, O., Comas-González, Z., Medina, J., Polo-Rodríguez, A., et al.: Sensing technologies and ml for emotion recognition in autism. *International Journal of Medical Informatics* (2024). <https://doi.org/10.1016/j.ijmedinf.2024.105469>
4. Chacón-Borrego, F.J., Vidal, J., Noda, M.: Diseño y validación de instrumentos para valorar la postura de sedestación y motricidad fina en alumnado de primaria. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte* **18**(69), 1–16 (2018), <https://www.redalyc.org/articulo.oa?id=237068652010>
5. Escalona, F., Salazar, F., Torres, Y., et al.: Aatiende: Automatic attention evaluation on a non-invasive device. *IWANN* (2023). https://doi.org/10.1007/978-3-031-43078-7_13
6. Faria, D.R., Weinberg, A.I., Ayrosa, P.P.: Multimodal affective communication analysis. *Applied Sciences* (2024). <https://doi.org/10.3390/app14156631>
7. Farsani, D., Villa-Ochoa, J.A.: Análisis de la atención visual de estudiantes a través de gafas espía. *Uniciencia* **36**(1) (2022). <https://doi.org/10.15359/ru.36-1.34>
8. Gahlan, N., Sethia, D.: Federated learning in emotion recognition systems based on physiological signals for privacy preservation: a review. *Multimedia Tools and Applications* (2025). <https://doi.org/10.1007/s11042-024-19467-3>
9. Halkiopoulou, C., Gkintoni, E., Aroutzidis, A., Antonopoulou, H.: Advances in neuroimaging and deep learning for emotion detection. *Diagnostics* (2025). <https://doi.org/10.3390/diagnostics15040456>
10. Ikram, S., Ahmad, H., Mahmood, N., Faisal, C.M.N., Abbas, Q., Qureshi, I., Hussain, A.: Recognition of student engagement state in a classroom environment using deep and efficient transfer learning. *Applied Sciences* **13**(15), 8637 (2023). <https://doi.org/10.3390/app13158637>
11. Kumar, J.A.: Facial animacy in anthropomorphised designs. *Computers Education* (2024). <https://doi.org/10.1016/j.compedu.2024.105150>
12. Liu, Q., Jiang, X., Jiang, R.: Classroom behavior recognition using computer vision: A systematic review. *Sensors* **25**(2), 373 (2025). <https://doi.org/10.3390/s25020373>
13. Luo, Z., Jin, X., Luo, Y., Zhou, Q., Luo, X.: Analysis of students positive emotion and smile intensity using da-cnn. *IEEE/CAA Journal of Automatica Sinica* (2025). <https://doi.org/10.1109/JAS.2024.125016>
14. Mahmood, N., et al.: Measuring student engagement through behavioral and emotional features using deep learning. *Algorithms* (2024). <https://doi.org/10.3390/a17100458>
15. Marquez-Carpintero, L., Pina-Navarro, M., Suescun-Ferrándiz, S., Escalona, F., Gómez-Donoso, F., Roig-Vila, R., Cazorla, M.: Artificial intelligence-based system for detecting attention levels in students. *Journal of Visualized Experiments* (202), e65931 (2023). <https://doi.org/10.3791/65931>
16. Mutawa, A., Hassouneh, A.: Multimodal real-time emotion recognition using facial expressions and eeg. *Biomedical Signal Processing and Control* (2024). <https://doi.org/10.1016/j.bspc.2023.105942>

17. Piedrahíta-Carvajal, A., Rodríguez-Marín, P.A., Terraza-Arciniegas, D.F., Amaya-Gómez, M., Duque-Muñoz, L., Martínez-Vargas, J.D.: Aplicación web para el análisis de emociones y atención de estudiantes. *Tecnológicas* **24**(51), e1821 (2021). <https://doi.org/10.22430/22565337.1821>
18. Samal, P., Hashmi, M.F.: Eeg-based bci emotion recognition: A review. *Artificial Intelligence Review* (2024). <https://doi.org/10.1007/s10462-023-10690-2>
19. Universidad Internacional de La Rioja: El impacto de las nuevas tecnologías en la concentración de los alumnos. *UNIR Revista* (2024), <https://www.unir.net/educacion/revista/impacto-nuevas-tecnologias-concentracion-alumnos/>
20. Vieriu, A.M., Petrea, G.: The impact of artificial intelligence (ai) on students academic development. *Education Sciences* **15**(3), 343 (2025). <https://doi.org/10.3390/educsci15030343>
21. Xue, W., Liu, J.: Design of intelligent teaching evaluation system based on emotion recognition algorithm. *IEEE ICENIT* (2024). <https://doi.org/10.1109/ICENIT61951.2024.00031>