

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

El estudio universitario en modalidad virtual se ha consolidado como una práctica habitual en distintos niveles de formación superior. Esta forma de aprendizaje, sostenida por plataformas tecnológicas y conexión remota, ha permitido dar continuidad a los procesos educativos en circunstancias diversas. Sin embargo, también ha evidenciado con claridad las múltiples condiciones que influyen en el desempeño académico. Para Liu et al. [11] opina que estudiar desde casa implica, para muchos estudiantes, reorganizar su vida cotidiana, responder a responsabilidades familiares, adaptarse a entornos reducidos o poco silenciosos, y mantener la atención en situaciones que no siempre son óptimas. En este contexto, asistir a clases, participar, cumplir con tareas o rendir evaluaciones, se vuelve una tarea que requiere más que conocimientos: exige persistencia, autorregulación y una capacidad constante de sostenerse.

A pesar de cumplir con lo académico, gran parte de lo que el estudiante experimenta durante su jornada virtual no queda registrado. Fatiga visual, tensión corporal, desconexión emocional o saturación cognitiva son condiciones frecuentes que inciden en el proceso de aprendizaje y que, sin embargo, suelen pasar desapercibidas. La tecnología educativa ha comenzado a responder a esta situación con enfoques más sensibles. El uso de inteligencia artificial orientado al análisis atencional permite identificar de forma no invasiva ciertos indicadores visuales que reflejan el estado del estudiante. Movimientos del cuerpo, expresiones faciales o cambios en la dirección de la mirada son señales que, al ser observadas con respeto y procesadas adecuadamente, ofrecen información útil para acompañar mejor el proceso formativo. Investigaciones como las de Liu et al. [11] y Abate et al. [1] han validado estos métodos, subrayando su valor como apoyo pedagógico.

La comprensión de que el aprendizaje también involucra lo físico y lo emocional ha ampliado las posibilidades de uso responsable de estas tecnologías. Estudios como los de Ikram [9] y Márquez-Carpintero [14] han evidenciado que

el rendimiento académico está vinculado a factores como el bienestar emocional, la fatiga acumulada y la postura corporal. Herramientas que reconocen estas dimensiones permiten intervenir sin juzgar, identificar con más claridad cuándo una persona está teniendo dificultades, y ofrecer una respuesta más ajustada a sus necesidades. En América Latina, trabajos como los de Piedrahíta-Carvajal et al. [17] y Farsani y Villa-Ochoa [6] han mostrado cómo el análisis postural y expresivo puede integrarse con fines pedagógicos, brindando retroalimentación oportuna en tiempo real y facilitando un acompañamiento más cercano, incluso en escenarios remotos.

La aplicación propuesta en este proyecto se ha desarrollado a partir de estas premisas. Su función principal es registrar, mediante la cámara del computador, información visual relacionada con el comportamiento atencional del estudiante durante actividades virtuales. Esta información incluye la postura, la dirección de la mirada y ciertas expresiones faciales que se analizan a través de un modelo no invasivo de inteligencia artificial. Ha sido pensada para implementarse en la Universidad de las Fuerzas Armadas ESPE, y requiere solo de una conexión estable, condiciones adecuadas de iluminación y un dispositivo con cámara funcional. El objetivo no es reemplazar al docente ni reducir la interacción educativa a un sistema automático. Al contrario, se trata de ofrecer una herramienta que permita observar con mayor precisión lo que ocurre en la experiencia cotidiana de aprendizaje, sin añadir presión ni interferir en el ritmo del estudiante.

En un entorno donde muchas personas estudian en silencio, desde sus hogares, sin contacto directo con sus docentes o compañeros, este tipo de herramientas representa un aporte concreto. En la realidad actual, donde la presencia del estudiante se reduce muchas veces a una imagen en pantalla, contar con un sistema que pueda detectar, de forma respetuosa y efectiva, momentos de desconexión o señales de agotamiento, contribuye a una educación más cercana, más clara y más justa.

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 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, como se muestra en la Figura 1 se aplicó la metodología PRISMA para la selección final de artículos [16], esta metodología dió como resultado 23 artículos, los cuales cumplían con todos los requisitos de la investigación y fueron seleccionados para conformar el grupo de estudios relacionados.

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 [4] 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 [19] 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, este fue incorporado en el 70% de los artículos seleccionados finales. Se observó que en la mayoría de los artículos se prioriza el uso de expresiones faciales como indicador clave de atención en contextos educativos, debido a su accesibilidad visual y la interacción humano-computadora, uno de los artículos que aplicaba el reconocimiento de expresiones faciales fue el de Luo [7], quién desarrolló un sistema basado en redes neuronales convolucionales con atención dirigida (DA-CNNs) para estimar la intensidad de la sonrisa caracterizándola como emoción positiva. De manera similar, Kumar [8] exploró el uso de rostros animados en contenidos educativos, demostrando que las expresiones faciales percibidas pueden influir significativamente en la atención y curiosidad de los estudiantes. No obstante, algunos estudios optan por otros enfoques como señales fisiológicas, considerando que las expresiones faciales pueden ser ambiguas o poco confiables en ciertos contextos, por ejemplo, el sistema HELS propuesto por Fordson [12], que se centró principalmente en señales fisiológicas como la electroencefalografía (EEG) y la respuesta galvánica de la piel (GSR), omitiendo deliberadamente el reconocimiento facial. El autor

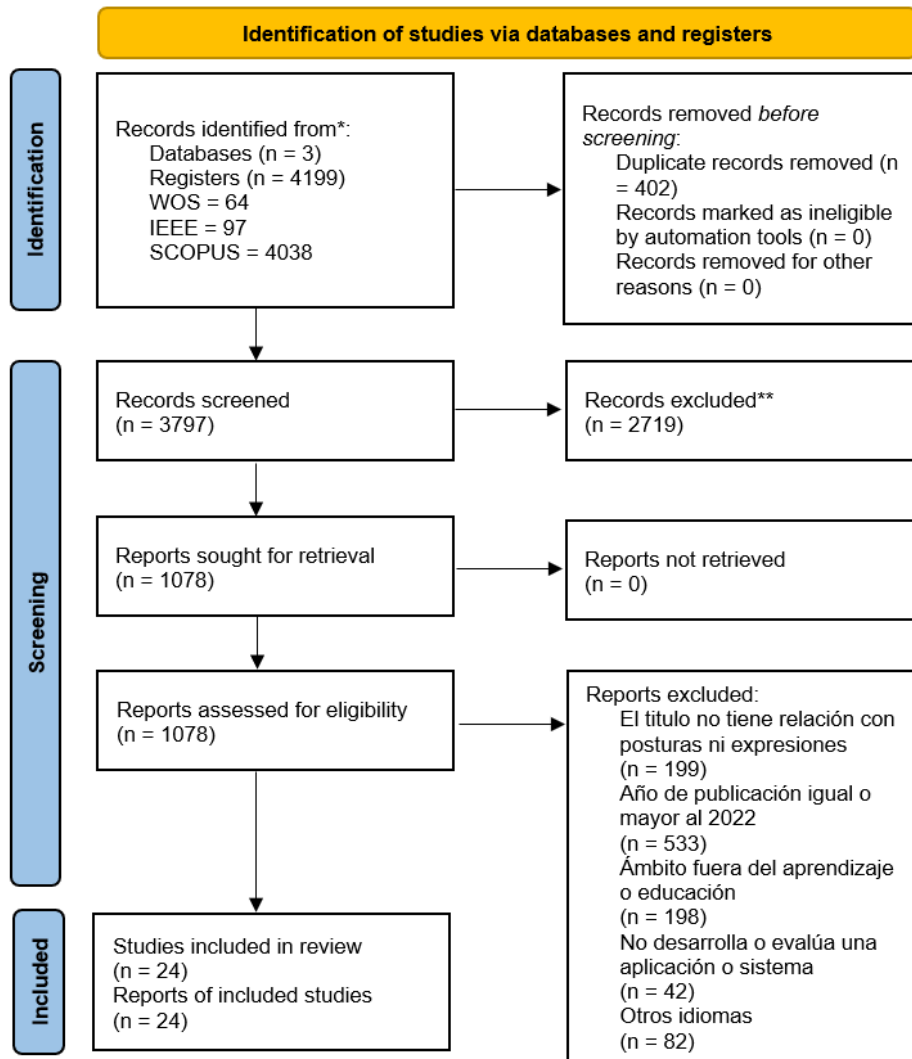


Fig. 1. Diagrama del método PRISMA. Elaboración propia basada en [16]

argumentó que las expresiones faciales pueden resultar ambiguas o poco fiables en ciertos contextos educativos.

En cuanto al análisis de postura corporal, esta variable fue considerada únicamente en el 26% de los artículos seleccionados finales, indicando que la mayoría de los estudios no la incluyeron como parte de sus metodologías para evaluar la atención estudiantil. Un ejemplo fue el trabajo de Huang y Zhou [10], quienes desarrollaron un sistema de detección postural en aulas donde se clasifica en tiempo real si un estudiante está de pie, sentado, entre otras. Sin embargo, se observó que la mayoría de los autores optaron por centrar sus investigaciones en las expresiones faciales como única fuente de análisis.

Una observación de interés, es que solo el 52% de los artículos seleccionados finales abordan la atención como variable de interés principal y de análisis. Por un lado, tenemos a Escalona [4] y Mahmood [13], 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 [5] 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 artículos seleccionados finales, implementan la toma de datos en tiempo real. Daza [19], y Huang y Zhou [10] 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 [15], entre otros, se limitan al análisis offline sobre datasets previamente grabados.

Por último, se observó que la visión por computadora fue utilizada en un 61% en los artículos seleccionados finales, siendo una herramienta clave para obtener información visual sin la necesidad de implementar sensores adicionales como lo hizo Chaudhari [5], donde incorpora la visión por computadora para el entrenamiento de CNNs con imágenes de escenas completas. Aun así, Fordson [12], 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.

En base al análisis comparativo de los 23 artículos revisados, se concluye que el uso de técnicas de Deep Learning y visión por computadora constituye una práctica ampliamente adoptada para el reconocimiento de emociones y la evaluación de la atención en entornos educativos. Esto evidencia la necesidad de utilizar el presente proyecto hacia dichas tecnologías. Si bien el reconocimiento de expresiones faciales es el enfoque predominante en la literatura consultada, el análisis de la postura corporal ha sido abordado en solo el 26% de los estudios, lo que representa una oportunidad concreta para la innovación tecnológica que este proyecto busca aprovechar. Aunque, la mitad de los artículos analizados considera la atención como variable principal, aquellos que lo hacen resaltan su impacto positivo en la mejora del aprendizaje. Por otra parte, se descarta la implementación de métodos como la electroencefalografía (EEG) y la respuesta

galvánica de la piel (GSR), debido a que requieren equipos costosos, además, que son procedimientos invasivos para los estudiantes, siendo un causante mas de perdida de atención. En este contexto, el desarrollo de sistemas no invasivos y en tiempo real refuerza la aplicabilidad del enfoque propuesto, el cual integra 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. 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
5. Faria, D.R., Weinberg, A.I., Ayrosa, P.P.: Multimodal affective communication analysis. *Applied Sciences* (2024). <https://doi.org/10.3390/app14156631>
6. 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>
7. 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>
8. Halkiopoulos, C., Gkintoni, E., Aroutzidis, A., Antonopoulou, H.: Advances in neuroimaging and deep learning for emotion detection. *Diagnostics* (2025). <https://doi.org/10.3390/diagnostics15040456>
9. 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>
10. Kumar, J.A.: Facial animacy in anthropomorphised designs. *Computers Education* (2024). <https://doi.org/10.1016/j.compedu.2024.105150>
11. 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>
12. 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>
13. Mahmood, N., et al.: Measuring student engagement through behavioral and emotional features using deep learning. *Algorithms* (2024). <https://doi.org/10.3390/a17100458>
14. 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>
15. 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>
16. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Greenhalgh, T., Grimshaw, J.M.H., Hróbjartsson, A., Higgins, M.A.J., Jüni, P., Kahn, J.M., Moher, D.: The prisma 2020 statement: an updated guideline for

- reporting systematic reviews. *BMJ* **372**, n71 (2021). <https://doi.org/10.1136/bmj.n71>, <https://doi.org/10.1136/bmj.n71>, pRISMA 2020 Statement
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. 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>