

**Propuesta de un modelo de visión por computadora para estimar la concentración  
estudiantil en entornos virtuales**

Sebastián Quevedo Laverde

Angie Tatiana Martínez Caro

Universidad EAN

Facultad de ingeniería, Ingeniería de Sistemas

Docente Leidy Natalia Restrepo Zapata

Bogotá, Colombia

16 de noviembre de 2025

## **Resumen**

El presente proyecto propone el diseño de un sistema basado en visión por computadora capaz de estimar en tiempo real el nivel de atención de estudiantes en entornos de educación virtual. A partir de la identificación del problema de la pérdida de concentración—acentuado por factores como la multitarea digital, la falta de autorregulación y la ausencia de señales no verbales propias de la educación presencial—se desarrolló una solución que integra captura de video, análisis facial y comparación de patrones mediante descriptores SIFT. La metodología combinó un enfoque mixto, incorporando un análisis técnico de algoritmos de visión por computadora y una encuesta aplicada a estudiantes y docentes de la Universidad EAN para identificar variables visuales y no verbales asociadas a los estados atencionales. Se construyó un prototipo funcional empleando OpenCV, PyQt6 y el dataset FER-2013 para clasificar emociones relacionadas con la concentración. Los resultados evidencian la viabilidad técnica de un sistema no intrusivo que aporte retroalimentación objetiva y permita a docentes y estudiantes identificar disminuciones en la atención durante las sesiones virtuales. Finalmente, se plantea su potencial impacto en el fortalecimiento de la participación, la autorregulación y la calidad del aprendizaje en modalidades digitales.

### ***Palabras clave:***

*Visión por computadora; atención estudiantil; educación virtual; concentración; SIFT; reconocimiento facial; inteligencia artificial; análisis de emociones; monitoreo en tiempo real; entornos de aprendizaje digitales.*

## Tabla de contenido

Resumen .....	2
1. Introducción .....	4
2. Árbol del problema .....	5
3. Objetivos .....	6
4. Descripción del problema .....	6
5. Pregunta de investigación .....	7
6. Justificación del problema.....	8
7. Marco teórico.....	9
8. Metodología.....	14
9. Resultados.....	19
10. Costos.....	32
11. Conclusiones .....	34
12. Bibliografía .....	37
Anexos.....	41

## Índice de Figuras

Figura 1 .....	5
Figura 2 .....	22
Figura 3 .....	23
Figura 4 .....	24
Figura 5 .....	25
Figura 6 .....	26
Figura 7 .....	27
Figura 8 .....	28
Figura 9 .....	34

## Índice de Tablas

Tabla 1 .....	15
Tabla 2 .....	31
Tabla 3 .....	32

## 1. Introducción

La educación virtual ha experimentado un crecimiento acelerado en las últimas décadas, impulsada por los avances tecnológicos y por la necesidad de continuidad académica durante la pandemia de COVID-19 (Bao, 2020; Daniel, 2020). Aunque esta modalidad ha ampliado el acceso y flexibilizado los procesos de aprendizaje, también ha revelado limitaciones importantes relacionadas con la interacción, el compromiso estudiantil y la capacidad del docente para percibir el nivel de atención de sus alumnos (The Difference Between Emergency Remote Teaching and Online Learning, 2020; Pan, 2023). En entornos presenciales, muchos de los indicadores de concentración se manifiestan mediante señales no verbales—postura, expresiones faciales, contacto visual—que se diluyen o desaparecen en plataformas digitales, afectando la eficacia del proceso formativo.

Diversos estudios han demostrado que la atención y concentración son elementos fundamentales del aprendizaje, influidos por factores como la motivación, el diseño pedagógico, la existencia de distractores y la calidad del entorno digital (Tyng et al., 2017; Gao & Li, 2022). Asimismo, investigaciones recientes evidencian que la multitarea digital, la falta de autorregulación y las interrupciones tecnológicas impactan negativamente el rendimiento académico en entornos virtuales (Molnár & Csapó, 2019; Poitras, 2015). Esta problemática se agrava debido a la limitada capacidad del docente para monitorear de manera continua el estado cognitivo de los estudiantes, lo que reduce las oportunidades de intervención oportuna (Denton-Calabrese et al., 2021).

Frente a ello, la visión por computadora y la inteligencia artificial se presentan como alternativas prometedoras para medir la atención de manera automática y no intrusiva. Avances recientes muestran que técnicas de análisis facial, seguimiento de mirada y detección de emociones pueden emplearse para estimar el nivel de compromiso del estudiante en tiempo real (Minaee et al., 2022; Daouas & Lejmi, 2018; Sharma et al., 2019). Modelos multimodales

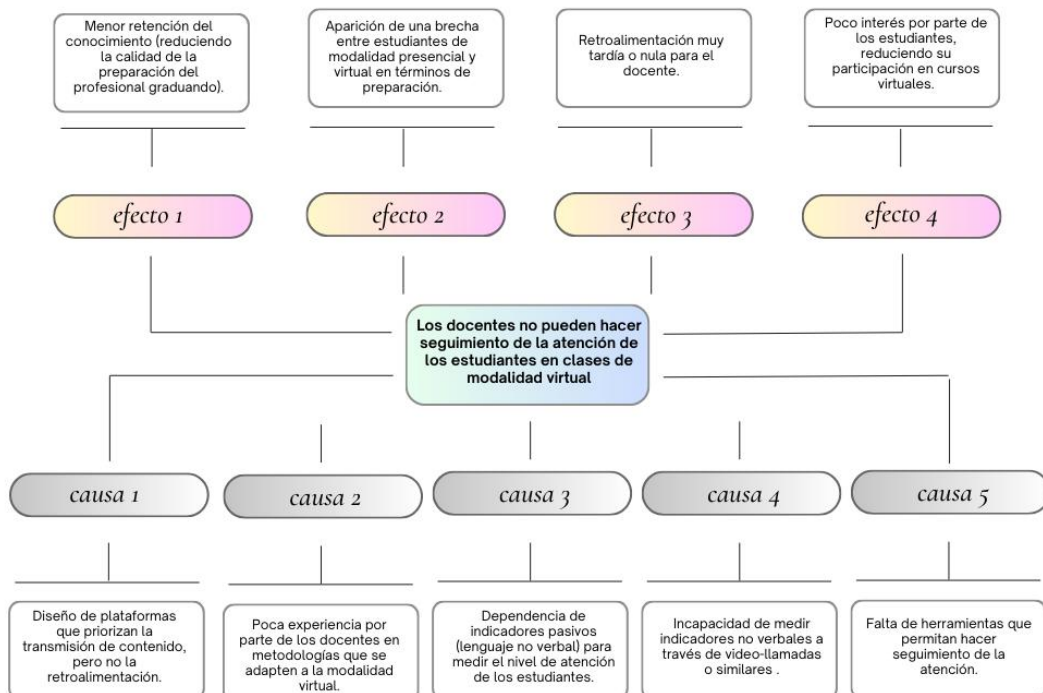
y redes neuronales han demostrado un alto potencial para clasificar estados como concentración, distracción, interés o fatiga, incluso en entornos educativos (Pourmirzaei et al., 2021; Kodithuwakku et al., 2022).

No obstante, la mayoría de estas soluciones no se han incorporado de forma práctica en plataformas educativas cotidianas, ya sea por su complejidad técnica, su costo o las preocupaciones éticas asociadas al uso de datos sensibles (Zhou et al., 2020). Por ello, el presente proyecto plantea el diseño de un sistema accesible, escalable y éticamente responsable para estimar el nivel de atención de los estudiantes mediante técnicas de visión por computadora basadas en descriptores SIFT y métricas no intrusivas. La propuesta busca contribuir al mejoramiento del aprendizaje virtual, fortaleciendo la retroalimentación entre docentes y estudiantes, favoreciendo la autorregulación y mitigando los efectos de la distracción digital.

## 2. Árbol del problema

**Figura 1.**

*Árbol problema.*



### **3. Objetivos**

#### **Objetivo General**

Diseñar un sistema basado en visión por computadora que permita detectar y estimar en tiempo real el nivel de atención de estudiantes en entornos de educación virtual.

#### **Objetivos Específicos**

- Analizar las técnicas y algoritmos de visión por computadora e inteligencia artificial aplicables a la detección de atención en entornos virtuales.
- Identificar las variables visuales y no verbales que influyen en la estimación del nivel de concentración de un estudiante durante clases virtuales.
- Diseñar la arquitectura conceptual del sistema, integrando módulos de captura, procesamiento y análisis de video en tiempo real.
- Proponer métricas e indicadores para evaluar el nivel de atención y generar alertas o reportes automáticos para el docente.

### **4. Descripción del problema**

En la educación virtual actual, resulta cada vez más difícil para el docente percibir cuándo un estudiante pierde concentración o interés, porque muchas de las señales no verbales que existen en el aula presencial se diluyen detrás de una pantalla. Esta limitación impide intervenciones tempranas y personalizadas que podrían reconducir el aprendizaje en tiempo real; además, estudios realizados con estudiantes de la Universidad EAN reportan que dimensiones como la autonomía, la comunicación y las emociones son factores clave que complican la adaptación y el seguimiento en la modalidad virtual. ( Arguelles, Chica, & Guzmán, 2023)

Las causas de este problema son múltiples y se encuentran tanto en factores tecnológicos como humanos. Una de las más relevantes es la ausencia de sistemas integrados

que permitan monitorear de manera automática y continua la atención de los estudiantes. Según ( Arguelles, Chica, & Guzmán, 2023), variables como la baja autonomía, la gestión ineficaz del tiempo, la sobrecarga de tareas, las barreras en la comunicación y los factores emocionales reducen significativamente la capacidad de concentración. A esto se suma que, aunque existen plataformas digitales para impartir clases, no todas incluyen herramientas analíticas que retroalimenten al docente sobre el estado atencional de sus alumnos. Otros estudios internacionales confirman que la multitarea digital, el acceso simultáneo a redes sociales y la falta de hábitos de autorregulación amplifican el problema en entornos virtuales. (Denton-Calabrese, Mustain, Geniets, Hakimi, & Winters, 2021)

Las consecuencias de no abordar este desafío son evidentes y preocupantes. La falta de atención sostenida se traduce en menor participación, disminución del rendimiento académico y pérdida de motivación, afectando la retención del conocimiento y la calidad de la experiencia educativa ( Arguelles, Chica, & Guzmán, 2023) (Denton-Calabrese, Mustain, Geniets, Hakimi, & Winters, 2021) advierten que esta situación, de mantenerse en el tiempo, puede debilitar el vínculo entre el estudiante y el proceso de enseñanza, incrementando el riesgo de deserción. Así, la educación virtual, que nació como una oportunidad para democratizar el acceso al aprendizaje, enfrenta el reto urgente de implementar mecanismos innovadores que garanticen no solo la transmisión de contenidos, sino también la atención y el compromiso de quienes participan en ella.

## **5. Pregunta de investigación**

¿De qué manera la detección automática de concentración puede contribuir a mejorar la atención y el rendimiento académico de los estudiantes en clases virtuales?

## 6. Justificación del problema

La educación virtual ha crecido significativamente, especialmente tras la pandemia, pero tiene un gran desafío: mantener la atención de los estudiantes en entornos digitales. La atención sostenida es fundamental para el aprendizaje efectivo. Sin embargo, el fenómeno del multitasking, muy común entre adolescentes y jóvenes universitarios, afecta negativamente el rendimiento académico, la retención de información y la autorregulación del aprendizaje (Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance, 2018)

Asimismo, un estudio de la Universidad de Connecticut reveló que los estudiantes que multitask al realizar tareas (como enviar mensajes o navegar por redes sociales) necesitan más tiempo de estudio y obtienen calificaciones promedio más bajas (Poitras, 2015). Este tipo de distracción no solo se limita al entorno académico presencial: un estudio en la Universidad de Hyderabad encontró que cada hora adicional dedicada al uso de pantalla aumenta el tiempo de reacción en 8.32 ms y disminuye la precisión atencional (Media multitasking impact in homework, executive function and academic performance in spanish adolescents, 2019)

Además, la sobreexposición a dispositivos de pantalla puede deteriorar la calidad del sueño y, aunque no se encontró correlación directa con la atención inmediata, sí impacta negativamente en el bienestar y el rendimiento cognitivo (Sahu, et al., 2024). En suma, el entorno educativo virtual agrava estas problemáticas debido a la falta de señales no verbales y paralingüísticas que facilitan la comunicación presencial, como el contacto visual o la retroalimentación inmediata (Augmenting Online Classes with an Attention Tracking Tool May Improve Student Engagement, 2023).

En respuesta a estos desafíos, la visión por computadora surge como una herramienta prometedora. Un estudio presentó un sistema que procesa video en tiempo real—utilizando cinco características no verbales para calcular una puntuación de atención—y proporciona retroalimentación inmediata tanto al estudiante como al docente (Augmenting Online Classes with an Attention Tracking Tool May Improve Student Engagement, 2023). De forma más avanzada, una investigación reciente propone una solución de reconocimiento de atención en línea basada en visión por computadora, que logra una clasificación multicategoría de niveles de atención usando redes neuronales convolucionales sobre el conjunto de datos DAiSEE. El sistema procesa video en vivo, generando análisis detallados y en tiempo real (Augmenting Online Classes with an Attention Tracking Tool May Improve Student Engagement, 2023).

No obstante, estos avances, no existe una herramienta práctica, escalable y accesible para docentes que monitoree en tiempo real la atención de los estudiantes en clases virtuales. Esto obstaculiza la retroalimentación oportuna, la personalización pedagógica y el apoyo emocional, afectando seriamente el aprendizaje y la motivación. Por ello, se hace necesario diseñar un sistema basado en visión por computadora e inteligencia artificial que permita detectar automáticamente el nivel de atención durante clases virtuales, propiciando intervenciones útiles y mejorando la calidad del proceso educativo.

## **7. Marco teórico**

### **Entornos de educación virtual**

La educación virtual ha evolucionado como una de las principales modalidades de aprendizaje en el siglo XXI, impulsada por el desarrollo de las tecnologías digitales y, en particular, por la necesidad de garantizar la continuidad educativa durante la pandemia de COVID-19 (Emergency remote teaching in a time of global crisis due to CoronaVirus pandemic, 2020) Este cambio permitió a millones de estudiantes acceder a programas

académicos desde cualquier lugar, lo que representó una ventaja en términos de cobertura, pero también visibilizó grandes desafíos pedagógicos y tecnológicos.

Estudios han mostrado que la educación en línea difiere sustancialmente de la presencial, no solo en la mediación tecnológica sino también en los niveles de interacción, retroalimentación inmediata y compromiso del estudiante (The Difference Between Emergency Remote Teaching and Online Learning, 2020) Mientras que en un aula física los profesores pueden interpretar las expresiones y gestos de los estudiantes para adaptar su enseñanza, en entornos virtuales esa capacidad se ve limitada (Pan, 2023).

Asimismo, la asincronía y la flexibilidad temporal que ofrece la educación virtual requieren mayores niveles de autorregulación, disciplina y motivación por parte de los estudiantes (Comparing online and blended learner's self-regulated learning strategies and academic performance, 2017). No obstante, también se han identificado beneficios relacionados con el acceso inclusivo, la personalización del aprendizaje y la reducción de costos operativos para instituciones y estudiantes (Means, Bakia, & Murphy, 2014).

### **Atención y concentración en estudiantes**

La atención y concentración son variables críticas en el proceso de aprendizaje, ya que determinan la capacidad de los estudiantes para procesar, retener y aplicar la información (Tyng, Hafeez U. Amin, Mohamad N. M. Saad, & Aamir S. Malik\*, 2017). En entornos virtuales, la ausencia de contacto físico directo y la presencia de múltiples distractores externos dificultan la capacidad de los docentes para medir el nivel real de compromiso de los estudiantes (Gao & Li, 28 March 2022).

Investigaciones recientes han demostrado que los niveles de concentración disminuyen significativamente en clases virtuales prolongadas (Molnár & Csapó, 2019) Además, los factores ambientales, como la multitarea digital, las interrupciones

domésticas y la calidad de la conexión a internet, influyen directamente en la atención sostenida (Online learning: Adoption, continuance, and learning outcome—A review of literature, 2018).

La detección automática de la atención, apoyada en inteligencia artificial y visión por computadora, se plantea como una solución para reducir la brecha entre docente y estudiante en entornos virtuales, proporcionando retroalimentación en tiempo real sobre el estado cognitivo de los participantes (Kodithuwakku, Arachchi, & Rajasekera, 27 April 2022) (Nair, Babu, & K, 2023).

### **Algoritmos de visión por computadora e inteligencia artificial**

La visión por computadora se ha convertido en una herramienta clave para el análisis automatizado de la atención en estudiantes, al permitir la interpretación de expresiones faciales, movimientos oculares y posturas corporales (Minaee, Boykov, Porikli, Plaza, & Kehtarnavaz, 1 July 2022) Aplicaciones recientes han demostrado la efectividad de algoritmos de aprendizaje profundo en la identificación de emociones como interés, aburrimiento, cansancio o frustración (Daouas & Lejmi, 2018) Modelos de reconocimiento multimodal, que combinan análisis facial, seguimiento ocular y estimación de posturas, han mostrado gran potencial para generar indicadores confiables del nivel de concentración (Sharma, et al., 2019) (Pourmirzaei, Montazer, & Mousavi, 2021) Estos sistemas, al integrarse en plataformas educativas, pueden brindar información en tiempo real que ayude a los docentes a adaptar su metodología y a los estudiantes a autorregularse (Automated Alertness and Emotion Detection for Empathic Feedback During E-Learning, 2013).

Sin embargo, la implementación de estas tecnologías debe considerar aspectos éticos y de privacidad, garantizando que los datos recolectados se manejen con consentimiento

informado, anonimización y medidas de seguridad adecuadas (Zhou, Shanshan, Ming, & Fangmei, 2020; Zhou, Shanshan, Ming, & Fangmei, 2020).

### **Métricas de atención y concentración**

La evaluación de la atención en entornos virtuales se ha apoyado en diferentes métricas no intrusivas, entre las que destacan el tiempo de fijación visual, la frecuencia de parpadeo, las expresiones faciales y la postura de la cabeza, indicadores que reflejan el nivel de concentración de los estudiantes (Daza, et al., 2021) (Ling, Yang, Jingxin, Zhu, & Sun, 2022). Estos datos, capturados a través de cámaras convencionales o sistemas de seguimiento ocular, pueden ser procesados con algoritmos de aprendizaje automático y visión por computadora para generar índices objetivos de atención. Dichos índices permiten clasificar de manera confiable estados como la concentración plena, la distracción parcial o el desinterés, ofreciendo así una base cuantitativa para retroalimentar el proceso de enseñanza-aprendizaje en tiempo real (Sharma, et al., 2019) (Happy, Dasgupta, Patnaik, & Routray, 2016).

Por otro lado, se ha evidenciado que los indicadores emocionales y fisiológicos, como la variabilidad cardíaca o la conductancia de la piel, también aportan datos relevantes para el análisis de la atención, aunque su implementación es más invasiva (Kodithuwakku, Arachchi, & Rajasekera, 27 April 2022). La tendencia actual apunta al uso de métodos no intrusivos, apoyados principalmente en visión por computadora, por su facilidad de integración en sistemas de educación en línea (R., et al., 2024).

### **Arquitectura conceptual del sistema**

El diseño de un sistema automático para la detección de concentración en estudiantes requiere la integración de diferentes módulos: captura de video, preprocesamiento de imágenes, análisis mediante algoritmos de aprendizaje automático, y visualización de

resultados en interfaces accesibles (Minaee, Boykov, Porikli, Plaza, & Kehtarnavaz, 1 July 2022).

La arquitectura debe priorizar la eficiencia en el procesamiento en tiempo real, la escalabilidad para integrarse en plataformas educativas y la flexibilidad para adaptarse a distintos contextos académicos (The Future of Online Teaching and Learning in Higher Education : The Survey Says, 2006) (Luckin, 2018). Adicionalmente, la retroalimentación generada debe ser comprensible tanto para los docentes como para los estudiantes, evitando la sobrecarga cognitiva y favoreciendo la interpretación pedagógica (Norman, 2013).

### **Ética, privacidad y viabilidad operativa**

El uso de inteligencia artificial para medir la atención estudiantil plantea importantes desafíos éticos y de privacidad. Entre ellos destacan el riesgo de vigilancia excesiva, los sesgos en los algoritmos y la necesidad de proteger los datos sensibles de los estudiantes. (Zhou, Shanshan, Ming, & Fangmei, 2020). De igual manera, la viabilidad de implementación de estos sistemas depende de la disponibilidad tecnológica, la aceptación por parte de los usuarios y el cumplimiento de las normativas legales sobre protección de datos (Selwyn, 2011).

En este sentido, el diseño debe contemplar políticas de transparencia algorítmica y estrategias de comunicación clara sobre el propósito y uso de la tecnología, con el fin de asegurar la confianza de la comunidad educativa (Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity, 2020).

### **Interacción humano-máquina en educación virtual**

La interacción humano-máquina (HMI) desempeña un papel fundamental en la efectividad de los sistemas de detección de atención, ya que determina cómo los usuarios perciben y utilizan la tecnología. La facilidad de uso, la transparencia en los resultados y la

capacidad de integración con las rutinas educativas son factores clave para la aceptación de estas herramientas (Selwyn, 2011).

En el ámbito educativo, la HMI debe estar orientada hacia el apoyo y la colaboración, no hacia el control. Sistemas bien diseñados pueden convertirse en aliados del aprendizaje, generando retroalimentación empática y favoreciendo la autorregulación (Happy, Dasgupta, Patnaik, & Routray, 2016).

## **8. Metodología**

### **8.1. Descripción Metodológica**

La investigación se desarrolla con un enfoque mixto, combinando métodos cuantitativos y cualitativos. Desde el componente cuantitativo, se analizarán variables objetivas obtenidas mediante visión por computadora, como la detección de rostro, dirección de la mirada, expresiones faciales y postura, para estimar en tiempo real el nivel de atención de los estudiantes en entornos virtuales. Por su parte, el componente cualitativo se enfocará en identificar las percepciones de estudiantes y profesores acerca de las clases virtuales, con el fin de comprender cómo valoran aspectos como el nivel de atención, la dinámica de enseñanza y los retos que enfrentan en este tipo de modalidad. Esta información permitirá contrastar y complementar los resultados técnicos del sistema, ofreciendo una visión más integral.

El diseño metodológico estará centrado en la Ingeniería de Sistemas, con énfasis en la programación, avanzando de manera secuencial desde la clasificación de técnicas y la selección de modelos base, hasta el desarrollo del prototipo, la validación experimental y el análisis de los hallazgos desde ambas perspectivas.

### **8.2. Diseño metodológico**

**Tabla 1.***Diseño metodológico.*

OBJETIVOS	ACTIVIDADES	RESULTADOS ESPERADOS	FECHA	RESPONSABLE
Analizar las técnicas y algoritmos de visión por computadora e inteligencia artificial aplicables a la detección de atención en entornos virtuales.	<ul style="list-style-type: none"> <li>Revisión documental exhaustiva de investigaciones sobre educación virtual, atención estudiantil, visión por computadora e IA.</li> <li>Identificación y organización temática de los conceptos clave del marco teórico.</li> </ul>	Marco teórico completo basado en revisión documental sólida.	01/09/2025 - 15/09/2025	Sebastian Quevedo
Identificar las variables visuales y no verbales que influyen en la estimación del nivel de concentración de un estudiante durante clases virtuales.	<ul style="list-style-type: none"> <li>Diseño de una encuesta estructurada para identificar variables visuales y no verbales asociadas al nivel de atención.</li> <li>Aplicación de la encuesta a estudiantes para recopilar la información necesaria.</li> </ul>	Encuesta estructurada validada y base de datos de respuestas.	16/09/2025 - 25/09/2025	Sebastian Quevedo Angie Martínez
Diseñar la arquitectura conceptual del sistema, integrando módulos de captura, procesamiento y análisis de video en tiempo real.	<ul style="list-style-type: none"> <li>Elaboración de arquitectura conceptual.</li> <li>Selección de librerías o frameworks.</li> <li>Desarrollo de la estructura del aplicativo.</li> </ul>	Aplicativo funcional (versión inicial).	26/09/2025 - 26/10/2025	Sebastian Quevedo Angie Martínez
Proponer métricas e indicadores para evaluar el nivel de atención y generar alertas o reportes automáticos para el docente.	<ul style="list-style-type: none"> <li>Revisión del documento FER2013 (Cornell University).</li> <li>Definición de métricas e indicadores basados en FER2013.</li> <li>Establecimiento de umbrales de atención.</li> </ul>	Conjunto de métricas e indicadores basados en FER2013.	27/10/2025 - 10/11/2025	Sebastian Quevedo Angie Martínez

**8.3. Población y muestra**

La población objetivo de este estudio corresponde a estudiantes y docentes de la Universidad EAN vinculados a la modalidad virtual. Para el cumplimiento del Objetivo 2, se aplicará un cuestionario estructurado a una muestra de 50 participantes, compuesta principalmente por estudiantes y complementada con docentes. La selección se realizará mediante un muestreo por discreción del investigador, lo que permitirá focalizar los esfuerzos en individuos con fácil acceso y disponibilidad, garantizando diversidad de percepciones sin comprometer la factibilidad operativa del estudio.

En cuanto a los cuestionarios asociados a los Objetivos 4 y 5, dirigidos a medir la percepción sobre la utilidad de los reportes y la usabilidad del sistema, la muestra se conformará con los participantes involucrados en las pruebas piloto del prototipo. En estos

casos no se establece un número fijo de encuestados, dado que la participación dependerá de la disponibilidad y del grado de involucramiento de los actores en el proceso experimental.

La decisión metodológica de trabajar con estos grupos específicos responde a la facilidad de acceso a la información, la pertinencia de los participantes como actores directos en la modalidad virtual y la relevancia de sus percepciones para la validación del sistema.

Asimismo, los datos primarios obtenidos mediante los cuestionarios se complementarán con conjuntos de datos de libre acceso empleados en investigaciones previas, como FER-2013 y DAiSEE, lo que permitirá triangulación y mayor solidez en el análisis de resultados.

#### **8.4. Instrumentos**

##### **Objetivo Especifico 1:**

Esta actividad se desarrolló durante la construcción del marco teórico, mediante una revisión documental exhaustiva sobre los principales enfoques de visión por computadora e inteligencia artificial aplicados a la detección de atención en entornos educativos virtuales.

Se consultaron fuentes académicas provenientes de bases de datos donde se identificaron estudios que abordan desde la evolución de la educación virtual y las problemáticas de concentración estudiantil, hasta los avances tecnológicos en algoritmos de reconocimiento facial, seguimiento ocular y análisis de posturas.

El proceso incluyó la clasificación y análisis comparativo de técnicas basadas en aprendizaje profundo, reconocimiento multimodal y métricas no intrusivas (como fijación visual, frecuencia de parpadeo y postura de la cabeza), evaluando su aplicabilidad en contextos educativos. Asimismo, se revisaron las consideraciones éticas y de privacidad inherentes al uso de IA en entornos académicos, así como la importancia de la interacción humano-máquina para garantizar la aceptación y utilidad pedagógica del sistema. Este análisis permitió establecer las bases teóricas y conceptuales que sustentan la propuesta del

sistema de detección de atención y sirvió como punto de partida para los objetivos posteriores.

## **Objetivo Especifico 2:**

### Cuestionario estructurado para estudiantes y docentes

- Propósito: Recoger percepciones de estudiantes y docentes sobre la atención, dinámica y condiciones de las clases virtuales.
- Preguntas:

### Datos Generales:

Rol: a). Estudiante b). Docente

Programa / Facultad: \_\_\_\_\_

Tiempo en modalidad virtual: a). <6 meses b). 6–12 meses c). >12 meses

Rango de edad: a). 18–24 b). 25–34 c). 35–44 d). 45+

¿Con qué frecuencia tiene sesiones virtuales por semana? A). 1–2 b). 3–4 c). 5+

### Escala Likert:

Conteste las siguientes preguntas donde 1= Totalmente en desacuerdo, 2 = En desacuerdo, 3 = Ni de acuerdo ni en desacuerdo, 4 = De acuerdo, 5 = Totalmente de acuerdo.

Atención y concentración:

- Mantengo mi atención la mayor parte del tiempo durante las clases virtuales.
- Creo que los estudiantes en general se distraen con facilidad en clases virtuales.
- Las actividades propuestas en clase favorecen mi concentración.
- Siento que el diseño de la clase virtual facilita la atención sostenida.

- Las interrupciones técnicas (audio/video) afectan significativamente la atención.

#### Diseño pedagógico y participación:

- Las estrategias de enseñanza (preguntas, actividades) fomentan la participación activa.
- El material compartido (presentaciones, videos) contribuye a mi comprensión.
- El tiempo destinado a interacción (preguntas, discusiones) es suficiente.
- Recibo retroalimentación que ayuda a mejorar mi atención y desempeño.

#### Tecnología y entorno

- La calidad de la conexión y plataforma influye en mi atención durante la clase.
- Dispongo de un espacio adecuado para seguir la clase sin distracciones.
- Me siento cómodo utilizando las herramientas tecnológicas empleadas.
- Las fallas técnicas son frecuentes y afectan el desarrollo de la clase.

#### Preguntas abiertas:

Mencione los principales factores que, a su juicio, afectan la atención en las clases virtuales.

¿Qué sugerencias tiene para mejorar la atención y participación en las clases virtuales?

#### **Objetivo Especifico 3:**

A partir de los hallazgos del marco teórico y la identificación de variables visuales relevantes, se diseñó una arquitectura conceptual que representa el funcionamiento integral del sistema de detección de atención.

El diseño contempló la estructuración de tres módulos principales: captura de video, encargada de obtener la información visual del usuario; procesamiento de imagen, donde se

aplican los algoritmos de visión por computadora e inteligencia artificial para el análisis facial y corporal; y módulo de análisis y visualización, que interpreta los resultados y genera indicadores de atención.

#### **Objetivo Especifico 4:**

Para la definición de las métricas e indicadores del nivel de atención se utilizó como referencia el documento “Facial Emotion Recognition: State of the Art Performance on FER2013” desarrollado en Cornell University, el cual presenta un análisis comparativo del rendimiento de diferentes modelos de reconocimiento facial aplicados al conjunto de datos FER2013. Este estudio permitió identificar los rasgos faciales más consistentes para la detección de emociones —como la orientación de la cabeza, la simetría de las expresiones, la apertura ocular y los microgestos asociados a interés o fatiga—, los cuales se integraron como base para seleccionar las variables no intrusivas más adecuadas dentro del sistema propuesto. De esta manera, el instrumento metodológico consistió en la adopción y adaptación de dichas métricas hacia un contexto educativo, permitiendo establecer umbrales iniciales de interpretación y una estructura de análisis coherente con el comportamiento visual de los estudiantes durante sesiones virtuales.

## **9. Resultados**

### **9.1. Objetivo 1.**

Para el Objetivo 1 se consolidó el marco teórico. Dentro de este se consideró la necesidad de revisar librerías que reunieran los algoritmos necesarios para el análisis de imágenes; durante esta revisión se encontró la librería de acceso sobre “OpenCV”.

Esta librería cuenta con más de 2500 algoritmos especializados en la visión por computadora y el machine learning; dentro de todos estos se encuentra “SIFT” (Scale-Invariant Feature Transform), que es un algoritmo que detecta y describe características

locales en imágenes, siendo invariante a escala, rotación y cambios menores de iluminación o punto de vista. El proceso completo se basa en cuatro pasos principales:

1. *Detección de Extremos en el Espacio-Escala:* Se identifican ubicaciones potenciales de keypoints buscando máximos y mínimos locales en una pirámide de Diferencias de Gaussianas (DoG). La imagen se difumina progresivamente con diferentes valores de sigma ( $\sigma$ ) y se genera una serie de octavas con la imagen redimensionada a la mitad. Cada píxel en una imagen DoG se compara con sus 8 vecinos en la misma escala y con los 9 píxeles vecinos en las escalas inmediatamente superior e inferior (26 comparaciones en total) para encontrar los extremos.
2. *Localización de Keypoints:* Se filtran los keypoints inestables. Primero, se refina la ubicación del punto usando una expansión de serie de Taylor. Luego, se descartan los puntos con contraste bajo (usando el parámetro `contrastThreshold`) y aquellos localizados en bordes, para lo cual se analiza la curvatura usando una matriz hessiana 2x2 y un umbral.
3. *Asignación de Orientación:* Para lograr invariancia a la rotación, se asigna una orientación a cada keypoint. En una región alrededor del punto, se calcula la magnitud y dirección del gradiente. Se crea un histograma de 36 contenedores (que cubren  $360^\circ$ ) y se asigna la orientación principal, pudiendo un keypoint tener más de una si existen picos secundarios que superen el 80% del pico principal.
4. *Generación del Descriptor del Keypoint:* Finalmente, se describe el keypoint con un vector numérico. Se toma una región de 16x16 píxeles alrededor del punto, se divide en 16 sub-bloques de 4x4 y para cada uno se genera un histograma de orientación de 8 contenedores. La concatenación de estos 16 histogramas resulta en un vector descriptor de 128 dimensiones. Este descriptor se hace robusto a la iluminación mediante normalización.

Con todo este proceso el algoritmo permite obtener un vector que resume de manera matemática las características principales de la imagen, de esta manera facilita la comparación más exacta posible de una imagen; facilitando el cumplimiento del objetivo general de este documento.

## **9.2. Objetivo 2.**

Para el cumplimiento del Objetivo 2 se aplicó la encuesta titulada “Análisis de encuesta: Percepciones de estudiantes y docentes sobre la atención, dinámica y condiciones de las clases virtuales”, logrando recopilar respuestas de 50 participantes (estudiantes y docentes) vinculados a la modalidad virtual. La información obtenida permitió identificar percepciones y experiencias relacionadas con la atención, las dinámicas de aprendizaje y las condiciones que influyen en el desempeño durante las clases en línea.

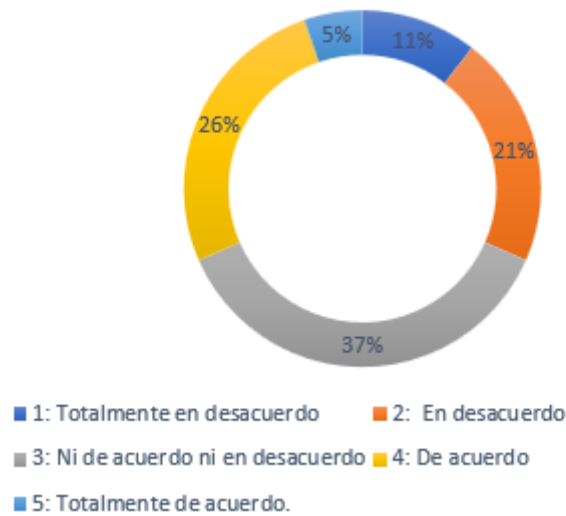
Con base en estos datos, a continuación, se presentan los resultados más relevantes, seleccionados por su aporte en la identificación de las variables visuales y no verbales que inciden directamente en el nivel de concentración de los estudiantes.

- Mantengo mi atención la mayor parte del tiempo durante las clases virtuales.

### **Figura 2.**

*Análisis de la pregunta mantengo mi atención la mayor parte del tiempo durante las clases virtuales.*

### Pregunta 1



Los resultados muestran que el 37 % de los encuestados adoptó una postura neutral frente a la afirmación, mientras que el 32 % manifestó algún grado de desacuerdo y el 31 % expresó acuerdo o total acuerdo. Esta distribución evidencia una percepción dividida y predominantemente intermedia, en la que no existe una posición claramente dominante sobre la capacidad de mantener la atención durante las clases virtuales.

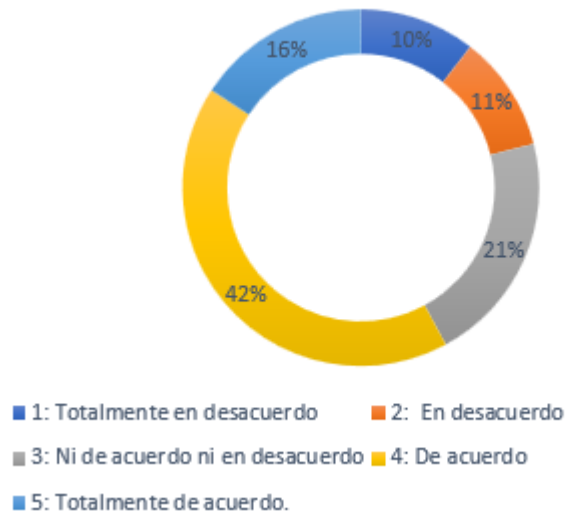
El promedio ponderado obtenido fue de 2,93 en una escala de 1 a 5, lo que indica una tendencia ligeramente inferior al punto medio. Este resultado sugiere que, aunque parte de los participantes logra concentrarse, una proporción similar presenta dificultades para mantener la atención de forma constante.

- Creo que los estudiantes en general se distraen con facilidad en clases virtuales.

#### **Figura 3.**

*Análisis de la pregunta creo que los estudiantes en general se distraen con facilidad en clases virtuales.*

## Pregunta 2



Los resultados reflejan una tendencia clara hacia el acuerdo, ya que la mayoría de los participantes (58 %) considera que los estudiantes se distraen con facilidad durante las clases virtuales. Un 21 % se mostró neutral y otro 21 % manifestó desacuerdo, lo que indica que, aunque hay opiniones diversas, predomina la percepción de que la atención suele verse afectada en este tipo de entornos.

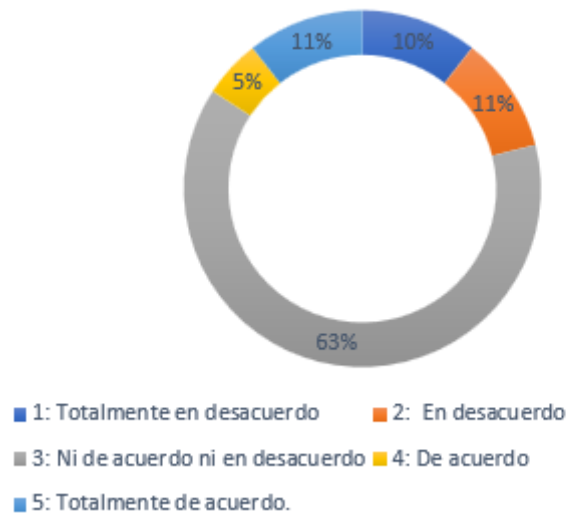
El promedio obtenido (3.4) confirma esta inclinación, evidenciando que los estudiantes reconocen las distracciones como una dificultad común al aprender de manera virtual. Factores como el uso constante de dispositivos electrónicos, el entorno doméstico y la ausencia de interacción presencial podrían influir en esta sensación.

- Siento que el diseño de la clase virtual facilita la atención sostenida.

### **Figura 4.**

*Análisis de la pregunta siento que el diseño de la clase virtual facilita la atención sostenida.*

### Pregunta 4



Los resultados muestran que el 63 % de los participantes se mantuvo neutral frente a la afirmación, mientras que un 21 % manifestó algún nivel de desacuerdo y solo un 16 % expresó acuerdo o total acuerdo. Esta distribución revela una tendencia claramente neutra, lo que sugiere que la mayoría de los estudiantes no percibe de forma contundente que el diseño de la clase virtual contribuya a mantener su atención.

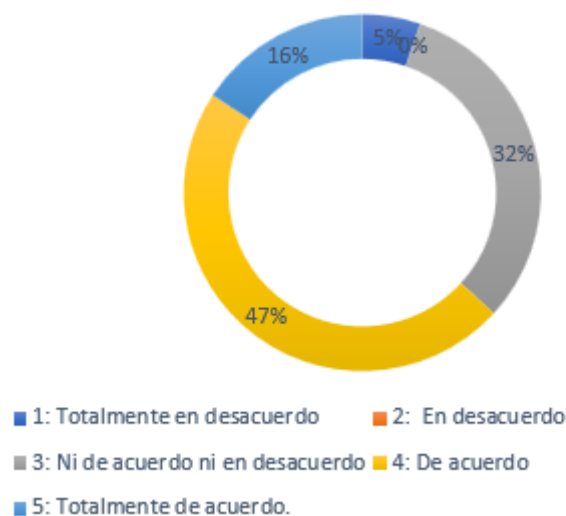
El promedio ponderado fue de 2.84, lo que refuerza esa posición intermedia con una ligera inclinación hacia la percepción negativa. En general, los resultados podrían interpretarse como una falta de impacto del diseño actual sobre la concentración de los estudiantes, posiblemente por factores como la estructura de las sesiones, la duración de las clases o el tipo de recursos utilizados.

- Las estrategias de enseñanza (preguntas, actividades) fomentan la participación activa.

#### **Figura 5.**

*Análisis de la pregunta las estrategias de enseñanza (preguntas, actividades) fomentan la participación activa.*

### Pregunta 6



Los resultados reflejan una percepción mayoritariamente positiva. El 63 % de los participantes estuvo de acuerdo o totalmente de acuerdo con la afirmación, mientras que un 32 % se mantuvo neutral y solo un 5 % expresó desacuerdo. Esta distribución sugiere que la mayoría reconoce que las estrategias empleadas por el docente promueven la participación y el involucramiento activo durante las clases.

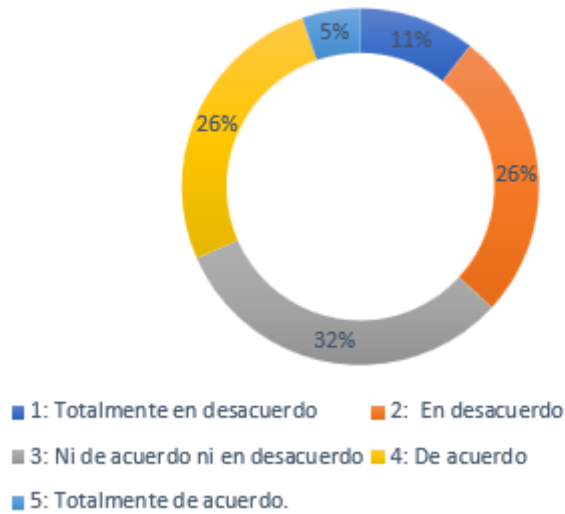
El promedio ponderado fue de 3.7, lo que confirma una tendencia favorable hacia las metodologías utilizadas. Sin embargo, el porcentaje de respuestas neutras indica que algunos estudiantes aún no perciben dichas estrategias como totalmente efectivas, lo que podría estar relacionado con diferencias en la forma de participación o en el tipo de actividades planteadas.

- Recibo retroalimentación que ayuda a mejorar mi atención y desempeño.

#### **Figura 6.**

*Análisis de la pregunta recibo retroalimentación que ayuda a mejorar mi atención y desempeño.*

### Pregunta 9



Los resultados muestran una distribución equilibrada, aunque con una ligera inclinación hacia el desacuerdo. Un 37 % de los participantes manifestó algún nivel de desacuerdo con la afirmación, mientras que el 32 % se mantuvo neutral y el 31 % expresó acuerdo o total acuerdo.

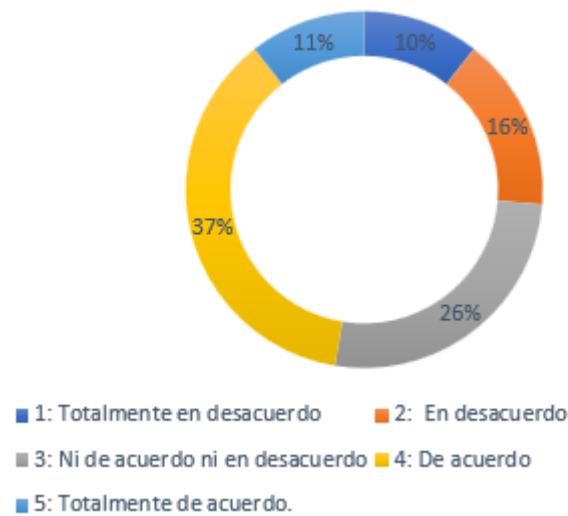
El promedio ponderado fue de 2.95, lo que indica una tendencia cercana al punto medio, reflejando percepciones diversas respecto a la utilidad de la retroalimentación recibida. En general, los datos sugieren que, aunque algunos estudiantes sí consideran que la retroalimentación contribuye a su mejora, otros no la perciben de manera constante o suficientemente efectiva.

- Dispongo de un espacio adecuado para seguir la clase sin distracciones.

#### **Figura 7.**

*Análisis de la pregunta dispongo de un espacio adecuado para seguir la clase sin distracciones.*

### Pregunta 11



Los resultados muestran una tendencia positiva, ya que el 47 % de los participantes estuvo de acuerdo o totalmente de acuerdo con la afirmación, mientras que un 26 % se mantuvo neutral y otro 26 % manifestó algún grado de desacuerdo. Esto sugiere que, aunque la mayoría considera contar con un entorno apropiado para el estudio, una parte significativa enfrenta condiciones poco favorables para concentrarse plenamente durante las clases virtuales.

El promedio ponderado fue de 3.2, lo que indica una percepción moderadamente favorable sobre la adecuación del espacio de aprendizaje. Sin embargo, el hecho de que uno de cada cuatro participantes no esté de acuerdo pone en evidencia que no todos disponen de un ambiente libre de distracciones, lo cual puede afectar su rendimiento y nivel de atención.

### 9.3. Objetivo 3.

Para cumplir con el objetivo 3, que se enfoca en el desarrollo de la estructura del prototipo, o mínimo producto viable; se analizó en un principio cuales eran los mínimos que debería ejecutar el código.

En un principio se debía obtener un conjunto de datos profesional, que ya estuviera desarrollado, que clasificara las características de una imagen y como asociar la misma a una emoción en específico; para así tener una base para realizar una comparación con los usuarios de la aplicación. Para este fin se utilizo FER-2013, un conjunto de datos desarrollado para un concurso de reconocimiento de expresiones faciales en la Conferencia Internacional de Machine Learning (ICML) de 2013.

FER-2013 contiene 35,887 imágenes de rostros en escala de grises y de 48x48 píxeles, clasificadas en siete categorías emocionales. Para esta aplicación solo se utilizaron las emociones de sorpresa y neutral.

Una vez identificado este conjunto de datos, se debía de desarrollar un código que analizara las imágenes clasificadas en las emociones de interés, y generara los descriptores de dichas imágenes. En la siguiente figura puede verse la sección de código que extrae la información del conjunto de datos y las emociones que deben buscarse.

### **Figura 8.**

*Extracción de información del conjunto de datos.*

```
import cv2 #opencv
import os
import numpy as np

# --- CONFIGURACIONES --- #
dataset_path = r"C:\Users\User\.cache\kagglehub\datasets\msambare\fer2013\versions\1" # Ruta al dataset
output_file = "descriptores_sift_triste_neutral.npz" # Archivo de salida
emociones_deseadas = ["sad", "neutral"] # Emociones que queremos procesar
```

Una vez extraídas las imágenes del conjunto de datos, debía analizarse cada una de estas imágenes para extraer los descriptores de cada una de estas y almacenarlas en un archivo con extensión “.npz” que implica un archivo comprimido de arreglos. Esto con el objetivo de no tener que recorrer o analizar el conjunto de datos más de una vez. (este código se encuentra en el anexo 1 “Generar\_descriptores”).

Con este archivo generado, debía desarrollarse un código que tomara los descriptores generados previamente y los compare con imágenes tomadas a los usuarios; para esto en un principio, se toma una imagen temporal con la cámara web del usuario, posteriormente se pasa por el mismo proceso con el cual se generaron los descriptores del conjunto de datos, y luego se comparo el arreglo resultante con aquellos almacenados en el documento. Por ultimo había que desarrollar una lógica que manejara los intervalos con los que se toman las imágenes temporales. Si se desea revisar mas a fondo este codigo se encuentra en el anexo 2.

Para el desarrollo de las interfaces de usuario del sistema de análisis de emociones, se implementaron tres ventanas especializadas utilizando PyQt6 y Qt Designer. La ventana principal presenta un diseño inmersivo sin bordes con un botón centralizado para iniciar el análisis, mientras que la ventana de temporizador incorpora un display LCD que actualiza en tiempo real la cuenta regresiva de cada intervalo de muestreo. Complementariamente, se diseñó una ventana de alerta modal que notifica al usuario cuando se detectan emociones específicas. Todas las interfaces siguen un principio de minimalismo visual para reducir la carga cognitiva y mantienen una navegación fluida entre los diferentes estados de la aplicación.

La integración entre los componentes visuales y la lógica del sistema se logró mediante el uso de timers QTimer para actualizaciones en tiempo real y un manejo robusto de eventos que garantiza la estabilidad durante el proceso de dos horas de análisis. El flujo implementado guía al usuario secuencialmente desde la inicialización del sistema a través de la ventana principal, pasando por los periodos de monitoreo representados en el temporizador, hasta las notificaciones inmediatas cuando se detectan emociones objetivo, siempre manteniendo una experiencia intuitiva que no requiere instrucciones previas para su operación.

Por último, el archivo principal del sistema de análisis de emociones implementa la integración completa entre la lógica de procesamiento y las interfaces de usuario mediante una arquitectura basada en PyQt6. Este módulo unifica tres componentes fundamentales: la ventana principal sin bordes que sirve como punto de entrada, el temporizador visual que muestra los intervalos de muestreo y la ventana de alerta para notificaciones emocionales. La clase “VentanaPrincipal” actúa como coordinador central, inicializando tanto los elementos de interfaz como los módulos de análisis facial, estableciendo los mecanismos de comunicación que permiten el flujo de datos entre la captura de imágenes, el procesamiento SIFT y la visualización de resultados.

#### **9.4. Objetivo 4.**

Para el cumplimiento del Objetivo 4 se estableció un conjunto de métricas operativas para la evaluación automática del nivel de atención, fundamentadas en el documento “Facial Emotion Recognition: State of the Art Performance on FER2013” de Cornell University y adaptadas al funcionamiento del aplicativo desarrollado. Dado que el sistema se diseñó sin emplear modelos de aprendizaje profundo, se seleccionaron métricas compatibles con métodos clásicos de visión por computador, priorizando la eficiencia computacional y la viabilidad técnica. Las métricas definidas se basan en la detección y análisis de rasgos faciales derivados de los keypoints y descriptores SIFT, entre ellos la orientación de la cabeza, la apertura y simetría ocular y los microgestos asociados a emociones como neutralidad o tristeza, presentes en el dataset FER2013.

Para su implementación se estableció un flujo de procesamiento compuesto por tres etapas principales: la generación de descriptores del dataset donde se realiza un recorrido de las imágenes FER2013 (clases “triste” y “neutral”), detección del rostro con Haar Cascade, normalización a 48×48 píxeles y cálculo de descriptores SIFT almacenados en formato .npz y

finalmente, la comparación entre los descriptores del estudiante y los del dataset, lo que permite estimar la emoción predominante y utilizarla como indicador interpretable del nivel de atención. Como resultado, se integraron métricas que facilitan distinguir estados asociados a neutralidad (mayor atención) de aquellos vinculados con tristeza o desconexión (menor atención), constituyendo el entregable final del objetivo.

**Tabla 2.**

*Indicadores faciales para la estimación del nivel de atención.*

<b>Métrica</b>	<b>Descripción técnica</b>	<b>Variables secundarias que la componen</b>
<b>Similitud SIFT con emoción “neutral”</b>	Mide el grado de coincidencia entre los descriptores SIFT del rostro capturado y los patrones faciales de la clase neutral del dataset FER2013. La neutralidad se caracteriza por estabilidad facial, baja activación emocional y ausencia de microgestos que alteren la expresión (Chen & Khairuddin, 2021).	Una alta similitud con la emoción neutral se asocia con mayor estabilidad emocional y mejor concentración. La literatura indica que expresiones neutrales están vinculadas a estados de alerta y disposición cognitiva para el aprendizaje (Türk & Yildirim, 2022).
<b>Similitud SIFT con emoción “triste”</b>	Evalúa la correspondencia entre los descriptores SIFT del rostro capturado y los patrones faciales de la emoción triste en FER2013. La tristeza suele implicar reducción en	Alta similitud con la emoción triste se relaciona con menor atención. Estudios muestran que la tristeza reduce el alcance atencional (Wang, Chen , &

	la apertura ocular, microgestos descendentes y leves asimetrías en boca y cejas (Chen & Khairuddin, 2021).	Zhang, 2018) y que emociones académicas negativas generan mayor distracción y menor eficiencia cognitiva (Tan , Mao, Jiang, & Gao, 2021).
--	--	---

## 10. Costos

Para el desarrollo del prototipo (mínimo desarrollo viable) se utilizaron únicamente herramientas de acceso gratis, lo cual limitó el desarrollo en múltiples aspectos, el primero fue el conjunto de datos utilizado (FER-2013) que no está completamente alineado con los objetivos del proyecto; la lógica de comparación no es la ideal, puesto que la ideal requeriría una red neuronal entrenada para hacer el proceso de comparación.

De manera que, con el objetivo de escalar este proyecto, sería necesario desarrollar las herramientas ideales para construirlo cuyos costos estimados serían los siguientes.

### Tabla 3.

*Costos del proyecto.*

Componente	Descripción	Costo estimado
Conjunto de datos	Un conjunto de datos pensado específicamente para entrenar la IA.	\$200'000.000 COP (una única inversión)
Desarrollo de la IA	Para desarrollar la IA es necesario tener cierta capacidad computacional que implica la compra de equipos y mantenimiento de estos.	\$126'000.000 COP (una única inversión)
Costos operacionales	Costos relacionados con espacios de oficinas, servicios públicos, consumo de servicios en la nube.	\$84'000.000 (inversión anual aproximada)

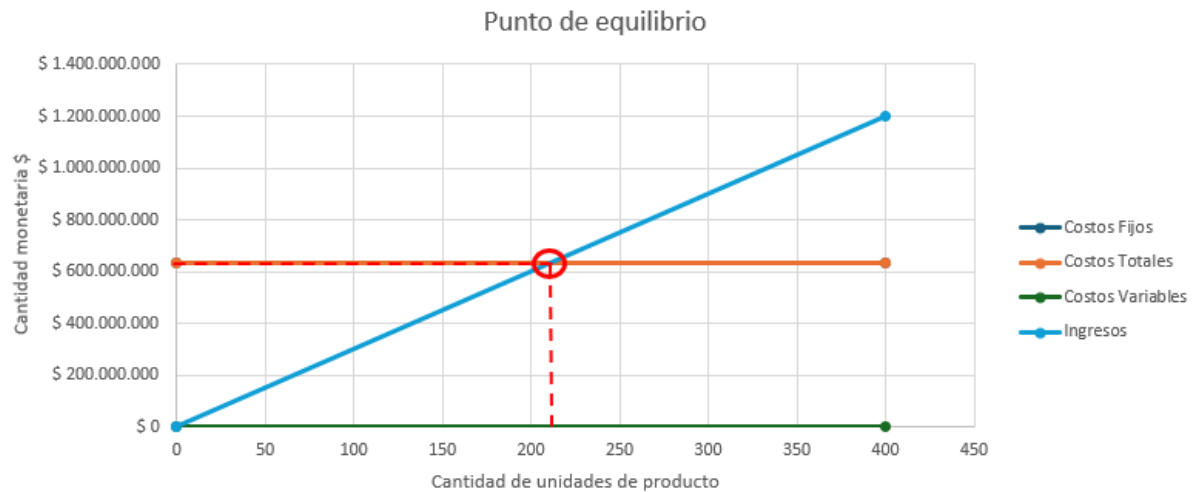
Salarios de los ingenieros	El salario de un ingeniero en Colombia es en promedio \$3'336.000 mensuales y seis empleados para el desarrollo (primeros 3 meses) y solo 2 a partir de la salida a ventas	\$60'480.000 en los primeros tres meses; \$60'480.00 para los últimos 9 meses del primer año y después de eso \$80'640.000 anuales
<b>Total</b>		Inversión inicial: \$470'480.000 para salida a ventas.  Costo anual después de la salida a ventas: \$164.480.000

Para determinar el número de unidades necesarias para cubrir los gastos totales del proyecto, se tomó como base el costo total estimado para el periodo, equivalente a \$634.960.000 COP. Considerando que el precio de venta de cada licencia es de \$3.000.000 COP, se calculó la cantidad de unidades que deben comercializarse para alcanzar el punto en el que los ingresos igualan los costos. El resultado indica que se requieren 212 licencias para cubrir la totalidad de los costos asociados al desarrollo, operación y sostenimiento de la solución.

Como proyección de ventas, se estimó que la Universidad EAN podría adquirir 125 licencias, cifra que representa aproximadamente una tercera parte del total de profesores de la institución. Bajo este escenario, los ingresos del primer año no serían suficientes para cubrir completamente los gastos totales; sin embargo, permitirían recuperar más de la mitad de la inversión. Con este ritmo de ventas, el proyecto alcanzaría su punto de equilibrio durante el segundo año de operación, demostrando viabilidad financiera en el mediano plazo.

### **Figura 9.**

*Punto de equilibrio del proyecto.*



## 11. Conclusiones

### 1. Conclusión General

Se logró diseñar y desarrollar conceptualmente un sistema basado en visión por computadora para detectar y estimar en tiempo real el nivel de atención de estudiantes en entornos de educación virtual. El sistema propuesto integra exitosamente módulos de captura, procesamiento y análisis de video, utilizando algoritmos clásicos de visión por computadora (como SIFT) para ofrecer una solución técnicamente viable, escalable y no intrusiva que aborda el problema crítico de la pérdida de atención en las aulas virtuales.

### 2. Conclusiones por Objetivo Específico

#### Objetivo 1: Análisis de técnicas y algoritmos

Se identificó que la visión por computadora, y en particular algoritmos como SIFT (Scale-Invariant Feature Transform), son altamente aplicables para la detección de características faciales relevantes para la atención. La librería OpenCV se estableció como la base tecnológica óptima por su robustez y la amplia gama de algoritmos que ofrece, permitiendo un análisis invariante a escala, rotación e iluminación.

#### Objetivo 2: Identificación de variables visuales y no verbales

La encuesta aplicada reveló una percepción dividida pero crítica sobre la atención en

entornos virtuales. Se identificó que los estudiantes reconocen fácilmente las distracciones en este modelo, y factores como el diseño de la clase (percibido como neutral en su capacidad para facilitar la atención) y la falta de una retroalimentación efectiva y constante son variables clave que impactan negativamente en la concentración. La adecuación del espacio físico también se confirmó como un factor influyente.

#### Objetivo 3: Diseño de la arquitectura conceptual del sistema

Se diseñó e implementó exitosamente un prototipo funcional (MVP) con una arquitectura modular. Este sistema integra la captura de imágenes via webcam, el procesamiento utilizando el dataset FER-2013 (para las emociones "neutral" y "triste"), la extracción y comparación de descriptores SIFT, y una interfaz de usuario intuitiva desarrollada con PyQt6 que guía al usuario a través del proceso de análisis y emite alertas.

#### Objetivo 4: Propuesta de métricas e indicadores

Se definieron e implementaron métricas operativas basadas en la similitud de descriptores SIFT con las emociones "neutral" y "triste" del dataset FER-2013. Estas métricas, fundamentadas en la literatura, permiten estimar el nivel de atención asociando una alta similitud con "neutral" a un estado de mayor concentración y estabilidad, y una alta similitud con "triste" a un estado de posible distracción o desconexión.

### 3. Conclusión sobre la Viabilidad y Pertinencia

La investigación demuestra que es técnicamente factible desarrollar un sistema de monitorización de la atención sin depender inicialmente de modelos complejos de aprendizaje profundo, utilizando en su lugar métodos clásicos de visión por computadora que son computacionalmente más eficientes. El sistema propuesto representa una solución directa a una problemática ampliamente documentada y sentida por la comunidad educativa,

ofreciendo un camino para cerrar la brecha de retroalimentación no verbal que existe en la educación virtual.

#### 4. Limitaciones y Futuras Mejoras

El sistema desarrollado, aunque funcional, presenta limitaciones importantes en su enfoque metodológico. La lógica de comparación basada en descriptores SIFT resulta inferior a lo que se puede alcanzar con redes neuronales convolucionales, ya que estas últimas pueden capturar patrones contextuales más complejos y sutiles del comportamiento facial asociado a la atención.

Además, la dependencia del conjunto de datos genérico FER-2013 restringe la precisión del sistema. La implementación de un conjunto de datos específico para entornos educativos virtuales, junto con la adopción de modelos de deep learning, representaría una mejora fundamental que transformaría el prototipo actual en una herramienta robusta y confiable para la evaluación de la atención en tiempo real.

#### 5. Conclusión sobre el Impacto Esperado

La implementación de un sistema como el diseñado tiene el potencial de mejorar significativamente la experiencia de aprendizaje virtual. Al proporcionar a docentes y estudiantes retroalimentación objetiva y en tiempo real sobre el estado atencional, se facilitan las intervenciones tempranas, la personalización de la enseñanza y el fomento de la autorregulación del estudiante, lo que podría traducirse en una mejora del rendimiento académico y una reducción en los índices de deserción.

## 12. Bibliografía

Arguelles, D., Chica, F., & Guzmán, M.-d.-S. (2023). Factores que dificultan el proceso de adaptación de los estudiantes a la modalidad virtual. Estudio en tres instituciones de educación superior en Colombia. *Revista iberoamericana de educación superior*, XIV(39), 100 -115. Recuperado el 9 de Agosto de 2025, de <https://www.redalyc.org/journal/2991/299176357006/html/>

Automated Alertness and Emotion Detection for Empathic Feedback During E-Learning. (2013). *Proceedings - 2013 IEEE 5th International Conference on Technology for Education*. doi:DOI:10.1109/T4E.2013.19

Bao, W. (15 de Marzo de 2020). *Wiley Online Library*. Obtenido de <https://doi.org/10.1002/hbe2.191>

Chen, Z., & Khairuddin, Y. (08 de Mayo de 2021). *Cornell University*. Obtenido de <https://arxiv.org/abs/2105.03588>

Daniel, S. (2020). *SPRINGER NATURE LINK*. Obtenido de <https://doi.org/10.1007/s11125-020-09464-3>

Daouas, T., & Lejmi, H. (2018). Emotions recognition in an intelligent elearning environment. *Interactive Learning Environments*, 26(8), 991–1009. doi:<https://doi.org/10.1080/10494820.2018.1427114>

Daza, R., DeAlcala, D., Morales, A., Tolosana, R., Cobos, R., & Fierrez, J. (2021). ALEBk: Feasibility Study of Attention Level Estimation via Blink Detection applied to e-Learning. *arXiv*. doi:<https://doi.org/10.48550/arXiv.2112.09165>

Denton-Calabrese, T., Mustain, P., Geniets, A., Hakimi, L., & Winters, N. (2021). Empowerment beyond skills: Computing and the enhancement of self-concept in the

go\_girl code+create program. *Computers & Education*.

doi:<https://doi.org/10.1016/j.compedu.2021.104321>

Fawns, T., Aitken, G., & Jones, D. (31 de Mayo de 2019). *SPRINGER NATURE LINK*.

Obtenido de <https://doi.org/10.1007/s42438-019-00048-9>

Gao, Q., & Li, S. (28 March 2022). Impact of Online Courses on University Student Visual Attention During the COVID-19 Pandemic. *Front. Psychiatry*, Volume 13 - 2022.

doi:<https://doi.org/10.3389/fpsyt.2022.848844>

Happy, S. L., Dasgupta, A., Patnaik, P., & Routray, A. (2016). Automated Alertness and Emotion Detection for Empathic Feedback During E-Learning. *arXiv*.

doi:<https://doi.org/10.48550/arXiv.1604.00312>

Kodithuwakku, J., Arachchi, D. D., & Rajasekera, J. (27 April 2022). An Emotion and Attention Recognition System to Classify the Level of Engagement to a Video Conversation by Participants in Real Time Using Machine Learning Models and Utilizing a Neural Accelerator Chip. *Digital Business and Innovations, Algorithms* 2022, 15(5). doi:<https://doi.org/10.3390/a15050150>

Ling, X., Yang, J., J. L., Zhu, H., & Sun, H. (2022). A Deep-Learning Based Method for Analysis of Students' Attention in Offline Class. *Electronics*. doi:

<https://doi.org/10.3390/electronics11172663>

Luckin, R. (2018). *Machine Learning and Human Intelligence The future of education for the 21st century*. ioe press.

Mayer, R. E. (Marzo de 2019). *WILEY ONLINE LIBRARY*. Obtenido de

<https://doi.org/10.1002/acp.3482>

- Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*.
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., & Kehtarnavaz, N. (1 July 2022). Image Segmentation Using Deep Learning: A Survey. *IEEE*, vol. 44, no. 7, pp. 3523-3542. doi: doi: 10.1109/TPAMI.2021.3059968
- Molnár, G., & Csapó, B. (2019). Making the Psychological Dimension of Learning Visible: Using Technology-Based Assessment to Monitor Students' Cognitive Development. *Frontiers in Psychology*, Volume 10 - 2019. doi:https://doi.org/10.3389/fpsyg.2019.01368
- Nair, R. R., Babu, T., & K, P. (2023). Enhancing Student Engagement in Online Learning through Facial Expression Analysis and Complex Emotion Recognition using Deep Learning. *arXiv preprint arXiv*. doi:https://doi.org/10.48550/arXiv.2311.10343
- Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.
- Pan, X. (2023). Online Learning Environments, Learners' Empowerment, and Learning Behavioral Engagement: The Mediating Role of Learning Motivation. *Sage Open*, 13(4).
- Poitras, C. (28 de agosto de 2015). Multitasking Increases Study Time, Lowers Grades. *UConn Today*. Obtenido de <https://today.uconn.edu/2015/07/multitasking-increases-study-time-lowers-grades/>
- Pourmirzaei, M., Montazer, G. A., & Mousavi, E. (2021). Customizing an Affective Tutoring System Based on Facial Expression and Head Pose Estimation. *arXiv preprint arXiv*. doi:https://doi.org/10.48550/arXiv.2111.14262

- R., V. A., Deroncele-Acosta, A., A. J., Barrasa, A., López-Granero, C., & Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. *Frontiers in Psychology*.  
doi:<https://doi.org/10.3389/fpsyg.2024.1387089>
- Sahu, D., Taywade, M., Malla, P., Singh, P., Jasti, P., Singh, P., . . . Gupta, K. (15 de Abril de 2024). *CUREUS*. Obtenido de <http://dx.doi.org/10.7759/cureus.58323>
- Selwyn, N. (2011). *Education & Technology. Key Issues & Debates*.
- Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Khanal, S. R., Reis, M. C., . . . Filipe, V. M. (2019). Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. *arXiv preprint arXiv*.  
doi:<https://doi.org/10.48550/arXiv.1909.12913>
- Tan , J., Mao, J., Jiang, Y., & Gao, M. (14 de Septiembre de 2021). *National Library of Medicine*. Obtenido de <https://pmc.ncbi.nlm.nih.gov/articles/PMC8472431/>
- Türk, D., & Yildirim, F. (2022). *University of Twente Research Information*. Obtenido de [https://ris.utwente.nl/ws/portalfiles/portal/292598118/Attention\\_towards\\_Facial\\_Regions\\_Varies\\_between\\_Emotion\\_Types\\_but\\_not\\_between\\_Attachment\\_Styles.pdf](https://ris.utwente.nl/ws/portalfiles/portal/292598118/Attention_towards_Facial_Regions_Varies_between_Emotion_Types_but_not_between_Attachment_Styles.pdf)
- Tyng, C. M., Hafeez U. Amin, Mohamad N. M. Saad, & Aamir S. Malik\*. (2017). The Influences of Emotion on Learning and Memory. *Frontiers in Psychology*, Volume 8 - 2017.  
doi:<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01454/full>

Wang, H., Chen , Y., & Zhang, Q. (2018). *FRONTIERS*. Obtenido de

<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.02397/full>

Zhou, L., S. W., M. Z., & F. L. (2020). 'School's Out, But Class' On', The Largest Online Education in the World Today: Taking China's Practical Exploration During The COVID-19 Epidemic Prevention and Control As an Example. *Best Evid Chin Edu*, 4(2):501-519. doi:<https://doi.org/10.2139/ssrn.3555520>

### **Anexos**

A continuación, se presenta el código empleado para la generación del prototipo:

#### **1. Generar\_descriptores**

```
import cv2

import os

import numpy as np

dataset_path =

r"C:\Users\User\.cache\kagglehub\datasets\msambare\fer2013\versions\1"

output_file = "descriptores_sift_triste_neutral.npz"

emociones_deseadas = ["sad", "neutral"]

face_cascade = cv2.CascadeClassifier(cv2.data.harcascades +

"haarcascade_frontalface_default.xml")

sift = cv2.SIFT_create()
```

```
X_descriptores = []

Y_emociones = []

for emocion in emociones_deseadas:

    emocion_path = os.path.join(dataset_path, "train", emocion)

    if not os.path.exists(emocion_path):

        continue

    for img_name in os.listdir(emocion_path):

        img_path = os.path.join(emocion_path, img_name)

        img = cv2.imread(img_path, cv2.IMREAD_GRAYSCALE)

        if img is None:

            continue

        faces = face_cascade.detectMultiScale(img, scaleFactor=1.2,
minNeighbors=5)

        for (x, y, w, h) in faces:

            rostro = img[y:y+h, x:x+w]

            rostro = cv2.resize(rostro, (200, 200))

            keypoints, descriptors = sift.detectAndCompute(rostro, None)

            if descriptors is not None:

                X_descriptores.append(descriptors)

                Y_emociones.append(emocion)
```

```
        break

np.savez_compressed(
    output_file,
    X_descriptores=np.array(X_descriptores, dtype=object),
    Y_emociones=np.array(Y_emociones)
)
```

## **2. Detector**

```
import time

import numpy as np

import cv2 as cv

import os

import sys

import traceback

def resource_path(relative_path):

    try:

        base_path = sys._MEIPASS

    except Exception:

        base_path = os.path.abspath(".")

    return os.path.join(base_path, relative_path)

try:

    data_path = resource_path("descriptores_sift_triste_neutral.npz")
```

```
data = np.load(data_path, allow_pickle=True)
```

```
X = data["X_descriptores"]
```

```
Y = data["Y_emociones"]
```

```
except Exception:
```

```
    X = []
```

```
    Y = []
```

```
class Captura_de_Imagenes:
```

```
    def __init__(self):
```

```
        self.camera_index = self.detectar_camara_activa()
```

```
    def detectar_camara_activa(self):
```

```
        for i in range(4):
```

```
            cap = cv.VideoCapture(i)
```

```
            if cap.isOpened():
```

```
                cap.set(cv.CAP_PROP_FRAME_WIDTH, 640)
```

```
                cap.set(cv.CAP_PROP_FRAME_HEIGHT, 480)
```

```
                cap.set(cv.CAP_PROP_FPS, 30)
```

```
                time.sleep(1)
```

```
                max_brightness = 0
```

```
                for _ in range(10):
```

```
                    ret, frame = cap.read()
```

```
                    if ret and frame is not None:
```

```
                        brightness = frame.mean()
```

```
                        max_brightness = max(max_brightness, brightness)
```

```
        cap.release()

        if max_brightness > 20:

            return i

    return 0
```

```
def iniciar_camara(self):

    cap = cv.VideoCapture(self.camera_index)

    if cap.isOpened():

        cap.set(cv.CAP_PROP_FRAME_WIDTH, 640)

        cap.set(cv.CAP_PROP_FRAME_HEIGHT, 480)

        cap.set(cv.CAP_PROP_FPS, 30)

        return cap

    return None
```

```
def capturar_emocion(self):

    cap = self.iniciar_camara()

    if cap is None:

        return None

    time.sleep(2)

    best_frame = None

    max_brightness = 0

    for _ in range(15):

        ret, frame = cap.read()

        if ret and frame is not None:

            brightness = frame.mean()
```

```
        if brightness > max_brightness:
            max_brightness = brightness
            best_frame = frame.copy()

    cap.release()

    if best_frame is not None and max_brightness > 20:
        return best_frame

    return None
```

```
class Generar_Descriptores:
```

```
    def __init__(self, imagen):
```

```
        self.imagen = imagen
```

```
        self.face_cascade = cv.CascadeClassifier(cv.data.harcascades +
'haarcascade_frontalface.xml')
```

```
    def ajustar(self):
```

```
        if self.imagen is None:
```

```
            return None
```

```
        escala_de_grises = cv.cvtColor(self.imagen, cv.COLOR_BGR2GRAY)
```

```
        caras = self.face_cascade.detectMultiScale(
```

```
            escala_de_grises,
```

```
            scaleFactor=1.1,
```

```
            minNeighbors=3,
```

```
            minSize=(50, 50),
```

```
            flags=cv.CASCADE_SCALE_IMAGE
```

```
        )
```

```
if len(caras) == 0:
    return None

x, y, w, h = caras[0]

cara = escala_de_grises[y:y+h, x:x+w]

cara = cv.resize(cara, (200, 200))

cara = cara / 255.0

return np.expand_dims(cara, axis=(0, -1))
```

```
def descriptores(self):
    cara = self.ajustar()

    if cara is None:
        return None

    cara_uint8 = (cara[0, :, :, 0] * 255).astype(np.uint8)

    sift = cv.SIFT_create()

    keypoints, descriptors = sift.detectAndCompute(cara_uint8, None)

    if descriptors is None:
        return None

    return [descriptors]
```

class Comparacion:

```
def __init__(self, X_dataset, Y_labels):
    self.X_dataset = X_dataset

    self.Y_labels = Y_labels

    self.matcher = cv.BFMatcher(cv.NORM_L2, crossCheck=False)
```

```

def comparar(self, descriptores_input):
    mejor_emocion = "neutral"
    mejor_score = 0
    descriptores_input = np.array(descriptores_input[0], dtype=np.float32)

    for i, desc_ref in enumerate(self.X_dataset):
        if desc_ref is None or len(desc_ref) == 0:
            continue

        desc_ref = np.array(desc_ref, dtype=np.float32)
        matches = self.matcher.knnMatch(descriptores_input, desc_ref, k=2)
        good_matches = []

        for m, n in matches:
            if m.distance < 0.75 * n.distance:
                good_matches.append(m)

        if len(good_matches) > mejor_score:
            mejor_score = len(good_matches)
            mejor_emocion = self.Y_labels[i]

    return mejor_emocion

```

```

class control_tiempo:
    def __init__(self, duracion=2, intervalo=5):
        self.duracion_total = duracion * 3600
        self.intervalo_toma = intervalo * 60
        self.inicio = time.time()
        self.ultima_toma = self.inicio

```

```

def debe_tomar_foto(self):

    ahora = time.time()

    if (ahora - self.inicio) >= self.duracion_total:

        return "finalizado"

    if (ahora - self.ultima_toma) >= self.intervalo_toma:

        self.ultima_toma = ahora

        return True

    return False

```

```

def tiempo_restante(self):

    ahora = time.time()

    restante = self.intervalo_toma - (ahora - self.ultima_toma)

    return max(0, int(restante))

```

### **3. Interfaz de usuario “Inicio”**

```

from PyQt6 import QtCore, QtGui, QtWidgets

```

```

class Ui_MainWindow(object):

    def setupUi(self, MainWindow):

        MainWindow.resize(522, 249)

        self.centralwidget = QtWidgets.QWidget(parent=MainWindow)

        self.verticalLayoutWidget = QtWidgets.QWidget(parent=self.centralwidget)

        self.layout_bienvenida =

QtWidgets.QVBoxLayout(self.verticalLayoutWidget)

```

```
self.Bienvenida = QtWidgets.QLabel(parent=self.verticalLayoutWidget)
self.layout_bienvenida.addWidget(self.Bienvenida)

self.Linea_1 = QtWidgets.QLabel(parent=self.verticalLayoutWidget)
self.layout_bienvenida.addWidget(self.Linea_1)

self.Linea_2 = QtWidgets.QLabel(parent=self.verticalLayoutWidget)
self.layout_bienvenida.addWidget(self.Linea_2)

self.Linea_3 = QtWidgets.QLabel(parent=self.verticalLayoutWidget)
self.layout_bienvenida.addWidget(self.Linea_3)

self.Acuerdo = QtWidgets.QLabel(parent=self.centralwidget)

self.Continuar = QtWidgets.QPushButton(parent=self.centralwidget)

MainWindow.setCentralWidget(self.centralwidget)
```

```
def retranslateUi(self, MainWindow):
```

```
    self.Bienvenida.setText("Bienvenido al sistema de analisis de emociones por
imagen")

    self.Linea_1.setText("Durante las proximas dos horas el programa tomara
imagenes temporales y periodicas")

    self.Linea_2.setText("por medio de ellas se revisara la concentracion")

    self.Linea_3.setText("para que puedas tomar una pausa activa")

    self.Acuerdo.setText("Si estas de acuerdo selecciona continuar")

    self.Continuar.setText("Continuar")
```

#### **4. Interfaz de usuario “Alerta”**

```
from PyQt6 import QtCore, QtGui, QtWidgets
```

```

class Ui_Dialog(object):

    def setupUi(self, Dialog):

        Dialog.resize(390, 145)

        self.label = QtWidgets.QLabel(parent=Dialog)

        self.label_2 = QtWidgets.QLabel(parent=Dialog)

        self.pushButton = QtWidgets.QPushButton(parent=Dialog)

    def retranslateUi(self, Dialog):

        self.label.setText("Atencion")

        self.label_2.setText("Detectamos que necesitas una pausa")

        self.pushButton.setText("Cerrar")

```

## 5. Interfaz de usuario “Timer”

```

from PyQt6 import QtCore, QtGui, QtWidgets

```

```

class Ui_Dialog(object):

    def setupUi(self, Dialog):

        Dialog.resize(324, 124)

        self.lcdNumber = QtWidgets.QLCDNumber(parent=Dialog)

        self.label = QtWidgets.QLabel(parent=Dialog)

    def retranslateUi(self, Dialog):

        self.label.setText("La siguiente toma de imagenes sera en:")

```

## 6. Main

```
import sys

import os

import logging

import traceback

import numpy as np

import cv2 as cv

from PyQt6 import QtWidgets, QtCore

from ui_alerta import Ui_Dialog

from UI_inicio import Ui_MainWindow

from UI_timer import Ui_Dialog as Ui_timer

from detector import Captura_de_Imagenes, Generar_Descriptores, Comparacion,
control_tiempo, X, Y

class VentanaAlerta(QtWidgets.QDialog):

    def __init__(self):

        super().__init__()

        self.ui = Ui_Dialog()

        self.ui.setupUi(self)

        self.ui.pushButton.clicked.connect(self.close)

class VentanaTimer(QtWidgets.QDialog):

    def __init__(self, control):

        super().__init__()

        self.ui = Ui_timer()

        self.ui.setupUi(self)
```

```
self.control = control

self.timer = QtCore.QTimer(self)

self.timer.setInterval(200)

self.timer.timeout.connect(self.actualizar)

self.timer.start()
```

```
def actualizar(self):

    self.ui.lcdNumber.display(self.control.tiempo_restante())
```

```
class VentanaPrincipal(QtWidgets.QMainWindow):
```

```
    def __init__(self):

        super().__init__()

        self.ui = Ui_MainWindow()

        self.ui.setupUi(self)

        self.ui.Continuar.clicked.connect(self.iniciar_sistema)

        self.capturador = Captura_de_Imagenes()

        self.comparador = Comparacion(X, Y)

        self.control = control_tiempo()

        self.timer = QtCore.QTimer()

        self.timer.setInterval(1000)

        self.timer.timeout.connect(self.revisar_tiempo)
```

```
    def iniciar_sistema(self):

        self.timer.start()

        self.ventana_timer = VentanaTimer(self.control)
```

```
self.ventana_timer.show()

def revisar_tiempo(self):

    if self.control.debe_tomar_foto() is True:

        imagen = self.capturador.capturar_emocion()

        if imagen is None:

            return

        descriptores = Generar_Descriptores(imagen).descriptores()

        if descriptores is None:

            return

        emocion = self.comparador.comparar(descriptores)

        if emocion in ["triste", "neutral"]:

            VentanaAlerta().exec()

if __name__ == "__main__":

    app = QtWidgets.QApplication(sys.argv)

    ventana = VentanaPrincipal()

    ventana.show()

    sys.exit(app.exec())
```