



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

**Centro Nacional de Investigación
y Desarrollo Tecnológico**

Tesis de Doctorado

**Detección unimodal de estados emocionales en
Ambientes Virtuales de Aprendizaje Inmersivo
mediante técnicas de Inteligencia Artificial aplicadas
al análisis del comportamiento y datos espaciales**

presentada por

MC. Jorge Enrique Velázquez Cano

como requisito para la obtención del grado de
Doctor en Ciencias de la Computación

Director de tesis

Dr. Juan Gabriel González Serna

Codirectora de tesis

Dra. Leonor Rivera Rivera

Cuernavaca, Morelos, México. Septiembre de 2026.

 Centro Nacional de Investigación y Desarrollo Tecnológico	ACEPTACIÓN DE IMPRESIÓN DEL DOCUMENTO DE TESIS DOCTORAL		Código: CENIDET-AC-006-D20
			Revisión: 0
	Referencia a la Norma ISO 9001:2008 7.1, 7.2.1, 7.5.1, 7.6, 8.1, 8.2.4		Página 1 de 1

Cuernavaca, Mor., a 12 de agosto de 2026

DR. CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
P R E S E N T E

AT'n: DRA. ANDREA MAGADÁN SALAZAR
PRESIDENTA DEL CLAUSTRO DOCTORAL DEL
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Los abajo firmantes, miembros del Comité Tutorial del estudiante **M.C. JORGE ENRIQUE VELÁZQUEZ CANO** manifiestan que después de haber revisado el documento de tesis titulado **"Detección unimodal de estados emocionales en Ambientes Virtuales de Aprendizaje Inmersivo mediante técnicas de Inteligencia Artificial aplicadas al análisis del comportamiento y datos espaciales"**, realizado bajo la dirección del **Dr. Juan Gabriel González Serna** y codirigido por la **Dra. Leonor Rivera Rivera**, el trabajo se **ACEPTA** para proceder a su impresión.

A T E N T A M E N T E

Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"


DR. JUAN GABRIEL GONZÁLEZ SERNA
TECNM/CENIDET


DRA. LEONOR RIVERA RIVERA
INSTITUTO NACIONAL DE SALUD PÚBLICA (INSP)


DR. NIMROD GONZÁLEZ FRANCO
TECNM/CENIDET


DRA. BLANCA DINA VALENZUELA ROBLES
TECNM/CENIDET


DR. MÁXIMO LÓPEZ SÁNCHEZ
TECNM/CENIDET


DR. JOSÉ ALEJANDRO REYES ORTIZ
UNIVERSIDAD AUTÓNOMA METROPOLITANA
(UAM) UNIDAD AZCAPOTZALCO

c.c.p: Verónica Sotelo Boyas /Jefa del Departamento de Servicios Escolares
c.c.p: Née Alejandro Castro Sánchez / Jefe del Departamento de Ciencias Computacionales
c.c.p: Expediente



Centro Nacional de Investigación y Desarrollo Tecnológico
Subdirección Académica

Cuernavaca Mor, 21/agosto/2026

Oficio No. SAC/266/2026

Asunto: Autorización de impresión de tesis

JORGE ENRIQUE VELÁZQUEZ CANO
CANDIDATO AL GRADO DE DOCTOR EN CIENCIAS
DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Detección unimodal de estados emocionales en Ambientes Virtuales de Aprendizaje Inmersivo mediante técnicas de Inteligencia Artificial aplicadas al análisis del comportamiento y datos espaciales"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

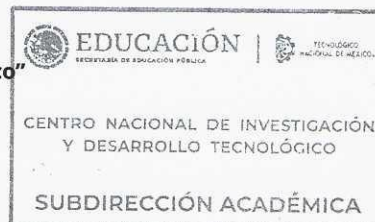
Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al servicio de México"


CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2026
año de
Margarita
Maza



Dedicatoria

A Dios. De Él son los caminos y el tiempo. De Él son las habilidades. De Él proviene todo.

A mi esposa Lluvia Arely.

A mi madre Doricela y mi hermano Carlos.

Agradecimientos

Al Tecnológico Nacional de México campus Centro Nacional de investigación y Desarrollo Tecnológico (CENIDET) por aceptarme para seguir formando parte de su cuerpo estudiantil y darme la oportunidad de realizar el programa de Doctorado en Ciencias en esta tan reconocida institución.

A la Secretaria de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por haberme proporcionado el apoyo que hizo posible la presente tesis, y que en los temas que la subyacen se vislumbran caminos hacia un país mejor, poniendo la tecnología al servicio de la sociedad.

A mi director de tesis el Dr. Juan Gabriel González Serna por todo el tiempo dedicado a guiarme y señalarme el camino; por la confianza que depositó en mí no solo para desarrollar este trabajo sino reflejada también en otras actividades de difusión y colaboración.

A la Dra. Leonor Rivera Rivera y la Dra. Marina Séris Martínez junto con su equipo del Instituto Nacional de Salud Pública (INSP). Su apoyo fue fundamental para realización de las pruebas con voluntarios. Sus contribuciones han hecho del proyecto una autentica herramienta que busca construir una mejor realidad para las niñas y los niños.

A mi comité revisor. Sus comentarios y recomendaciones esculpieron no solo el proyecto, sino a mi mismo como persona e investigador. Estableciendo en conjunto un referente digno del investigador que aspiro ser y alcanzar.

A mi familia. A mi esposa, quien me escuchó en mis momentos de frustración y duda, pero también de éxito y logro.

A Dios. Sin él nada es posible.

Tabla de contenido

Lista de figuras	VII
Lista de tablas	VIII
Resumen	IX
Abstract	X
1 Introducción	1
1.1 Antecedentes	3
1.1.1 Entorno Virtual de Aprendizaje para evaluar niveles de inmersión, interacción y estados emocionales durante un proceso de intervención educativa de prevención del Abuso Sexual infantil en niños y niñas de edad escolar (J.-E. Velázquez-Cano, 2021)	3
1.1.2 Tutor Virtual Inteligente para Ambiente Virtual de Aprendizaje Inmersivo mediante Tecnología de Cognición Aumentada (Cisneros-Brito, 2018) . . .	3
1.1.3 Ambiente virtual de aprendizaje inmersivo para realizar intervenciones educativas para la prevención del Abuso Sexual Infantil (Herrera, s.f.) . . .	3
2 Marco Conceptual	5
2.1 Realidad Virtual (RV)	5
2.2 AVAI.	7
2.3 Intelligent Tutoring System (ITS).	7
2.4 Emoción.	8
2.5 Affective Tutoring System (ATS).	8
3 Estado del arte	10

4	Planteamiento del problema	22
5	Propuesta de solución	25
5.1	Objetivos	25
5.1.1	General	25
5.1.2	Específicos	25
5.2	Justificación	26
5.3	Hipótesis	27
5.4	Alcance y limitaciones	27
5.4.1	Alcances	27
5.4.2	Limitaciones	28
6	Metodología	29
6.1	Enfoque y tipo de investigación	29
6.2	Definición de la población muestra	29
6.2.1	Criterios de inclusión	30
6.2.2	Criterios de exclusión	30
6.3	Materiales e instrumentos	30
6.3.1	Equipo de Realidad Virtual Inmersiva	30
6.3.2	Plataforma tecnológica	31
6.3.3	Instrumentos	31
6.4	Procedimiento	34
6.5	Consideraciones éticas	35
7	Diseño y desarrollo del modelo unimodal de detección emocional.	37
7.1	Arquitectura general del modelo	37
7.2	Análisis de los datos	38
7.3	Cálculo de características	39

7.3.1	Posición	39
7.3.2	Rotación Real	39
7.3.3	Rotación	40
7.3.4	Velocidad	40
7.3.5	Velocidad angular	40
7.3.6	Aceleración	40
7.3.7	Aceleración angular	41
7.3.8	Curvatura de la trayectoria	41
7.3.9	Vector HR	42
7.3.10	Vector HL	42
7.3.11	Vector RL	42
7.3.12	Joint Relative Triangle Area (JRTA)	43
7.3.13	Joint Relative Angle (JRA)	43
7.3.14	Joint relative cosine dissimilarity (JRCD)	43
7.3.15	Foco de atención	44
7.3.16	Simetría distal simple	45
7.3.17	Simetría vertical	45
7.3.18	Volumen	45
7.3.19	Delta de la simetría distal	46
7.3.20	Variables con prefijo A	46
7.3.21	Variables categorizadas	46
7.3.22	Estructura del vector de características	47
7.4	Selección de características	47
7.5	Algoritmos de regresión	48
7.6	Entrenamiento y validación del modelo	49
8	Resultados	50

8.1	Desempeño de los modelos	50
8.1.1	Resultado del modelo Random Forest	50
8.1.2	Resultado del modelo Redes Neuronales	53
8.1.3	Resultado del modelo XGBoost	55
8.2	Análisis exploratorio	55
8.2.1	Análisis por edad	55
8.2.2	Análisis por género	56
8.2.3	Análisis por escena	57
8.2.4	Análisis del REN por escena	58
8.2.5	Resumen de resultados	60
9	Discusión	61
10	Conclusiones y trabajos futuros	64
10.1	Limitaciones	65
10.2	Aportaciones	66
10.2.1	AVAI ASI extendido	66
10.2.2	Modelo unimodal de detección emocional en Realidad Virtual Inmersiva	66
10.2.3	Corpus	66
10.2.4	Metodología	67
10.2.5	Publicaciones	67
10.3	Trabajos futuros	68
11	Cierre	69
	Referencias	71

Lista de figuras

2.1	La inmersión como resultado del realismo, involucramiento e interacción.	5
2.2	Diferencias en la presentación de un ambiente virtual entre dispositivos inmersivos y no inmersivos.	6
2.3	Arquitectura general de un ITS (Diseño propio)	7
2.4	Arquitectura general de un ATS (Diseño propio)	8
6.1	Interfaz del cuestionario demográfico en RV	32
6.2	Interfaz del Emoji-SAM RV	33
7.1	Arquitectura general para la detección de estados emocionales.	38
7.2	Mecanismo de detección del foco de atención del usuario	44
8.1	Matrices de confusión de los resultados categorizados de la dimensión valencia . .	52
8.2	Matrices de confusión de los resultados categorizados de la dimensión activación	53
8.3	Distribución porcentual por edad	56
8.4	Distribución de los valores de valencia activación con respecto a la edad	56
8.5	Distribución porcentual por género	57
8.6	Distribución de los valores de valencia activación con respecto a género	57
8.7	Distribución de los valores de valencia activación con respecto a las escenas . . .	58
8.8	Boxplot de las respuestas del REN por escena	59
8.9	Distribución de los valores de valencia activación con respecto al REN	59

Lista de tablas

3.1	Estado del arte en materia de prevención del Abuso Sexual Infantil usando TIC. .	11
3.2	Estado del arte en materia de prevención del Abuso Sexual Infantil usando RV como elemento completo.	13
3.3	Estado del arte respecto a tutores afectivos.	15
3.4	Estado del arte respecto a tutores afectivos y RVI.	17
7.1	Valores registrados de acuerdo al nivel de atención detectado.	44
7.2	Categorización de valencia y activación.	47
7.3	Características seleccionadas para los vectores de medias y medianas.	48
8.1	Mejores hiperparámetros.	50
8.2	Resultados de los modelos para el 20 % de los datos de prueba.	51
8.3	Resultados de los modelos para el 10 % de los datos previos al SMOTE.	51
8.4	Metricas de las matrices de confusión de la dimensión valencia	52
8.5	Metricas de las matrices de confusión de la dimensión activación	53
8.6	Resultados de los modelos MLP para la dimensión valencia usando el vector de medias	54
8.7	Resultados de los modelos MLP para la dimensión valencia usando el vector de medianas	54
8.8	Hiperparametros y resultados de las pruebas con XGBoost	55

Resumen

La presente tesis documenta la investigación realizada, los instrumentos desarrollados, análisis de los datos y resultados de la detección de estados emocionales en ambientes virtuales usando técnicas de inteligencia artificial con información espacial de manos y cabeza.

Particularmente, este trabajo da seguimiento al uso del ambiente virtual de aprendizaje inmersivo (AVAI) “pinta tu raya ASI” (AVAI ASI) y su uso como la plataforma donde se implementan los distintos mecanismos de detección e instrumentos de auto reporte emocional. La investigación considera esta perspectiva como parte fundamental del proyecto, abordándolo desde el estado del arte como un proyecto multidisciplinario. No obstante, se enfocan los procesos y resultados al área de las ciencias de la computación.

Se describen los mecanismos de detección desarrollados y los instrumentos de medición utilizados, destacando su diseño y desarrollo como parte indispensable para el modelo propuesto, así como hallazgos relacionados al diseño de mejores interfaces para sistemas en realidad virtual (RV). Se analizan estudios exploratorios adicionales que pueden abrir nuevas oportunidades de investigación y para los cuales se diseñaron e implementaron instrumentos nuevos.

Posteriormente se presentan las intervenciones realizadas en campo junto con el protocolo seguido. Se registró el comportamiento de niños de 8 a 10 años de edad durante el uso del AVAI ASI para generar un *dataset* con un *ground truth* compuesto de las respuestas de los instrumentos de autoevaluación.

Finalmente se describe el tratamiento de los datos obtenidos y los resultados de los diferentes modelos de regresión utilizados. Se encontró el uso de bosques aleatorios como el modelo con mejores resultados, se hallaron diferencias importantes con respecto a la medida central utilizada que apuntan a establecer balances entre dimensionalidad y precisión requerida. En cuanto al grupo etario hubo hallazgos relevantes en relación a la percepción emocional con la edad del participante. Las conclusiones presentan las consideraciones bajo las cuales el modelo propuesto funciona y cómo pueden estos resultados adaptarse a otros proyectos de naturaleza similar.

Abstract

The current thesis documents the research, the developed instruments, tests, data analysis, and results on emotional state detection in virtual environments using artificial intelligence techniques with spatial information from the head and hands.

Particularly, this work follows the use of the immersive virtual learning environment “Pinta tu raya ASI” as a platform where different detection mechanisms and emotional self-report instruments are implemented. This research takes this perspective as a fundamental part of the project, approaching it from the state of art as a multidisciplinary project. However computer science processes and results are the main focus.

Subsequently, the field experiments are presented along with the protocol followed. The behavior of children aged 8 to 10 years during the use of the virtual environment was recorded to generate a data set with its corresponding ground truth which is conformed of the self-assessment instruments’ evaluations.

Finally, the treatment of data and the results of the different regression models’ are described. Random Forest were found to be the best model approach. Important differences associated to the central tendency measure used were found that point towards stablishing a balance between dimensionality and required precision. Regarding the age group, there were relevant findings related to the emotional perception with the participants’ age. The conclusions present the considerations under which the proposed model functions and how these results can be adapted to other projects of a similar nature.

1

Introducción

Las emociones son respuestas sistemáticas coordinadas de un individuo ante distintos eventos y situaciones. Por esta razón, el contexto en el que se presentan resulta ineludible y ha sido ampliamente aprovechado en la literatura para inducir y detectar estados afectivos, replicando las situaciones que las generan de manera real o virtual. Sin embargo, cuando se abordan experiencias educativas sensibles como los cursos de prevención del abuso sexual infantil (ASI), el estado emocional del estudiante cobra una relevancia crítica. Aunque la ludificación ha demostrado ser efectiva (Zainuddin et al., 2020) para mantener el involucramiento mediante sensaciones de logro y avance (Plass et al., 2015), rara vez se considera el impacto emocional que puede generar el contenido, especialmente emociones como ansiedad, miedo o vergüenza (Allen et al., 2020) (Wurtele, 2009 como fue citado en (Tutty, 2014)) y en general el estado emocional del estudiante, un elemento crucial en el motor de aprendizaje (D'Mello & Graesser, 2012) que tiene efecto también sobre la motivación y el involucramiento (Vilhunen et al., 2022).

En el ámbito específico de la prevención de ASI, padres y docentes expresan preocupación constante por posibles efectos adversos, aun cuando estudios en la literatura muestran que estos son escasos o transitorios (Hébert, Lavoie, Vitaro, McDuff, Tremblay, 2008; Nibert, Cooper, Ford, 1989; Taal Edelaar, 1995; Tutty, 1992, 1997; Wurtele Miller-Perrin, 1987, como fueron citados en (Tutty, 2014)) (Walsh et al., 2015 como fue citado en (Kenny et al., 2020)), y que incluso entre sobrevivientes los beneficios superan con creces los riesgos cuando el programa está bien diseñado (Gubbels, van der Put et al., 2021; Jones et al., 2020). A pesar de la evidencia favorable, los programas tradicionales de prevención siguen dependiendo casi exclusivamente de facilitadores humanos que, se asume, manejarán las respuestas emocionales en tiempo real. Esto deja un vacío tecnológico importante: el uso de tutores afectivos es prácticamente inexistente en este dominio. Asimismo, aunque las Tecnologías de la Información y la Comunicación (TIC) ya forman parte de algunos programas (Jones et al., 2020; Malamsha et al., 2021), la Realidad Virtual Inmersiva (RVI) aparece solo marginalmente como herramienta pedagógica y sin capacidades afectivas integradas. En el campo más amplio de la computación afectiva e interacciones humano-computadora, a pesar del crecimiento exponencial de dispositivos y técnicas de captura, la interacción sigue siendo predominantemente rígida, poco natural y mecanizada. La detección automática de emociones ha avanzado principalmente mediante reconocimiento facial, análisis de voz,

y con sensores fisiológicos, configurando una tendencia multimodal que suele mejorar la precisión, pero incrementa la complejidad, costo e invasividad. Por otro lado, el análisis de postura, gestos y comportamiento espacial de cabeza y manos permanece sub-explorado y suele depender de dispositivos externos (sensores Kinect, cámaras Web) o datos de emociones actuadas estereotípicas que introducen sesgos culturales y pérdida de autenticidad emocional.

Este trabajo se inserta precisamente en esa intersección abordando el desarrollo un tutor afectivo en Realidad Virtual Inmersiva desde tres capas:

- Capa de aplicación: enfocada en la prevención de ASI con impacto social, atendiendo tanto los vacíos pedagógicos y emocionales de los programas actuales como la percepción cualitativa de niños y facilitadores.
- Capa de método: usa la RVI como alternativa escalable y segura a los métodos presenciales o basados en medios 2D, aprovechando su capacidad demostrada para generar sentido de presencia y respuestas emocionales auténticas.
- Capa de técnica (núcleo computacional de esta tesis): se compone de un módulo unimodal de detección emocional basado exclusivamente en características espacio-temporales del movimiento natural de cabeza y manos capturadas por los sensores de los dispositivos de RVI, evitando sensores adicionales y emociones actuadas o etiquetadas a priori.

El principal aporte científico-técnico radica en demostrar que es posible alcanzar niveles de precisión competitivos utilizando únicamente información de comportamiento espacial no invasiva y en contexto real de aprendizaje de temas de naturaleza sensible, reduciendo drásticamente los requisitos hardware (e.g. sensores fisiológicos) y eliminando la dependencia de expresiones faciales (que pueden estar parcial o totalmente ocultas por el visor) o actuación forzada o estereotípica. Además, al derivar tanto las características como las etiquetas emocionales directamente del participante (mediante auto reportes post-escena) se minimiza el sesgo cultural de observadores externos y se respeta la subjetividad individual de la experiencia emocional.

Para validar la propuesta, se diseñó y ejecutó un protocolo con 65 niñas y niños de 8 a 10 años de edad, voluntarios de una escuela primaria del estado de Morelos, México, en un estudio no experimental en el que se recolectó el conjunto de datos a ser utilizados para el entrenamiento de los modelos de inteligencia artificial (IA) y además conformar el *ground truth*.

1.1. Antecedentes

1.1.1. Entorno Virtual de Aprendizaje para evaluar niveles de inmersión, interacción y estados emocionales durante un proceso de intervención educativa de prevención del Abuso Sexual infantil en niños y niñas de edad escolar (J.-E. Velázquez-Cano, 2021)

Este trabajo aborda el desarrollo de la aplicación “Pinta tu raya ASI” en dos versiones: No Inmersiva (i.e. de escritorio) e Inmersiva. Se describe un modelo de diseño y desarrollo de mejores entornos virtuales, particularmente aquellos que pretendan ser usados con fines de estudio. Posteriormente se presentan las pruebas realizadas junto con los protocolos para cada modalidad y los instrumentos de medición utilizados (Escala de Afecto Positivo y Negativo, Cuestionario de Depresión Infantil y Adolescente, Self-Assesment Manikin versión emoticono, Inventario de Motivación Intrínseca e Inventario de Sensación de Presencia). Se buscaron cambios en la percepción emocional auto reportada, su experiencia durante la intervención, su sensación de presencia y la emoción inducida por el entorno virtual. Se encontraron diferencias y cambios positivos significativos en el estado emocional asociados a la versión inmersiva de la aplicación, y también al uso general de la misma.

1.1.2. Tutor Virtual Inteligente para Ambiente Virtual de Aprendizaje Inmersivo mediante Tecnología de Cognición Aumentada (Cisneros-Brito, 2018)

En esta tesis de maestría se diseñó e implementó un entorno virtual de aprendizaje inmersivo para la educación en línea, en este AVAI se monitorea el estado cognitivo del estudiante. Se incluyó un tutor que monitorea el estado cognitivo del estudiante con la finalidad de identificar su estado cognitivo. Si el estado no es el adecuado, el tutor presenta ejercicios para retomar la atención del estudiante. Se diseñó un sistema embebido para capturar las señales electroencefalográficas por medio de la diadema de Emotiv Epoc. Este sistema es el encargado de obtener el estado cognitivo del estudiante y enviarlo al entorno virtual a través de un servicio web.

1.1.3. Ambiente virtual de aprendizaje inmersivo para realizar intervenciones educativas para la prevención del Abuso Sexual Infantil (Herrera, s.f.)

En este trabajo es una tesis de maestría, se diseñó, implementó y evaluó un prototipo para aplicar intervención educativa preventiva del abuso sexual infantil (ASI) en colaboración con

el Instituto Nacional de Salud Pública (INSP), utilizando un visor de Realidad Virtual Inmersiva Autónomo con el objetivo capacitar a niños de entre 6 y 12 años a identificar situaciones de posible abuso sexual.

2

Marco Conceptual

2.1. Realidad Virtual (RV)

Es la representación sintética y abstracta de un ambiente que pretende ser una recreación de algo real o bien una creación ficticia. Posee distintos medios y dispositivos que permiten que el usuario interactúe con el entorno y que a su vez representan un nivel de inmersión de acuerdo con la cantidad de canales sensoriales que interviene (Capece et al., 2019; Chirico et al., 2018; Mann et al., 2018). La Figura 2.1 muestra como la inmersión es el aprovechamiento de las prestaciones tecnológicas de los dispositivos para generar entornos realistas (nivel de fidelidad respecto a la realidad física) e interactivos (cómo las acciones del usuario modifican el entorno virtual) que involucren más al usuario en la tarea a realizar.

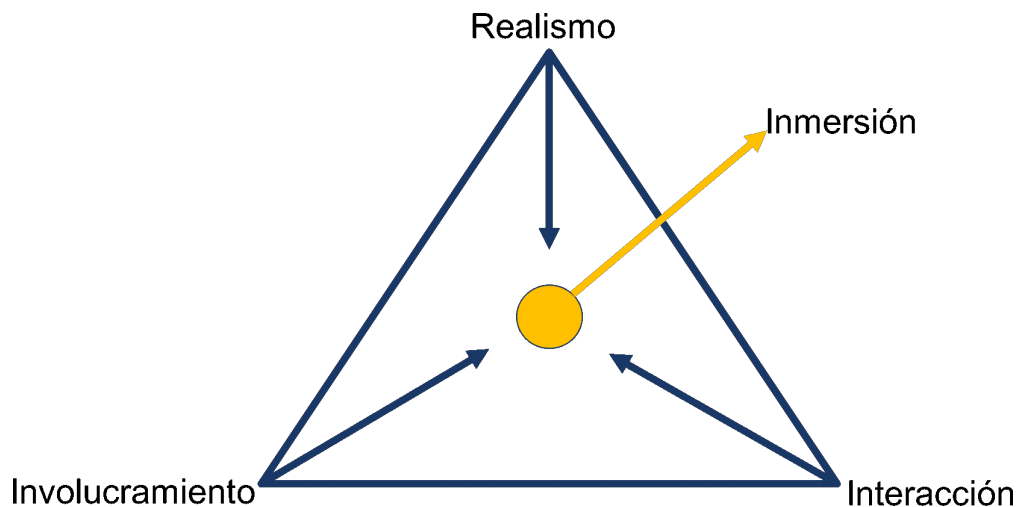


Figura 2.1

La inmersión como resultado del realismo, involucramiento e interacción. Diversas capacidades tecnológicas y su aprovechamiento modifican el nivel de cada eje.

La RV se clasifica en dos tipos (Makransky & Lilleholt, 2018):

- **No inmersiva.** No tienen la capacidad para excluir o intervenir señales sensoriales provenientes del mundo real. Se conoce también como realidad virtual de escritorio (*desktop VR*) aunque se puede extender el concepto a dispositivos como tabletas o teléfonos celulares con capacidades para manejar este tipo de aplicaciones (figura 2.2).
- **Inmersiva.** Puede excluir señales sensoriales provenientes del mundo real al tiempo que ofrece alta fidelidad en los ambientes representados. Hace uso de dispositivos tecnológicos más avanzados y dedicados a la intervención de algún sentido en particular, principalmente el *Head Mounted Display* (HMD) que tiene la función de intervenir las señales visuales y reemplazarlas con las provenientes de la representación sintética. La inmersión y el nivel de interacción posible con el entorno virtual genera en el usuario el llamado sentido de presencia o la sensación de “*estar ahí*” (Cummings & Bailenson, 2016).

La figura 2.2 muestra cómo el usuario percibe la representación sintética (i.e. ambiente virtual) en cada una de las variantes anteriores. Mientras que en las inmersivas el entorno parece ocurrir “alrededor” del usuario haciéndolo parte del mismo, en las no inmersivas el dispositivo mediador es, de manera clara y evidente, un elemento separado del usuario y parte del mundo real.

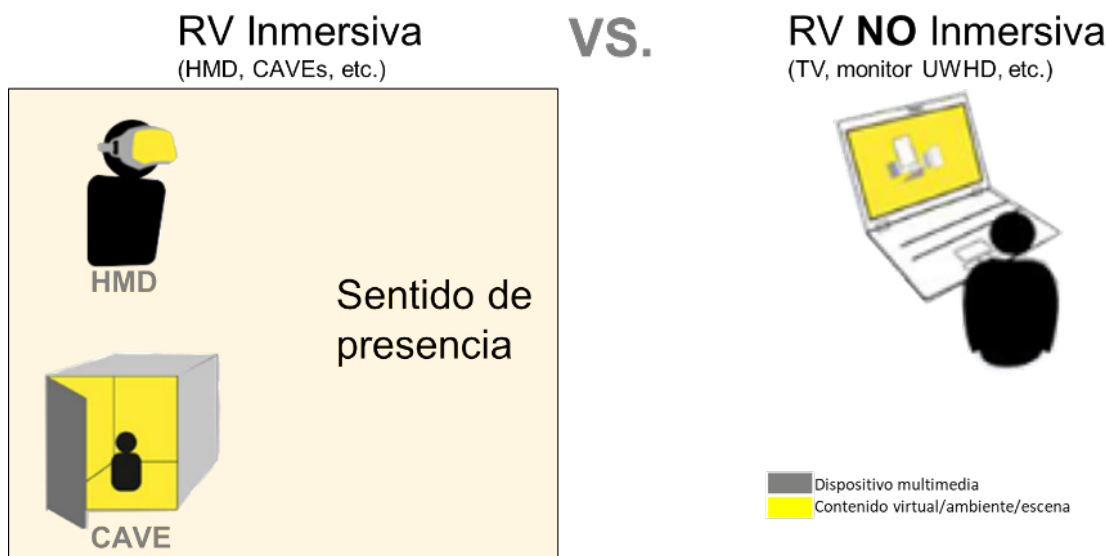


Figura 2.2

Diferencias en la presentación de un ambiente virtual entre dispositivos inmersivos y no inmersivos. El término CAVE proviene de “cueva” ya que el usuario literalmente entra a un espacio físico adaptado para proyectar el entorno virtual a su alrededor. En un HMD el mismo efecto se logra mediante el uso de sensores que hacen corresponder la proyección con los movimientos del usuario.

2.2. AVAI.

Mikropoulos y Natsis (2011), definen un Ambiente Virtual de Aprendizaje (AVA) como un Ambiente Virtual que, basado en cierto modelo pedagógico, incorpora uno o más objetivos didácticos y provee a los usuarios con experiencias que no podrían tener en el mundo físico o en su defecto, experiencias cuya realización implica riesgos físicos u otros impedimentos (e.g. económicos). Un AVA proporciona un aprendizaje específico. Un AVA se considera AVAI (i.e. inmersivo) cuando incorpora tecnologías de Realidad Virtual Inmersiva (RVI).

2.3. Intelligent Tutoring System (ITS).

Es un sistema de aprendizaje asistido por computadora que intenta replicar a un tutor humano mediante instrucciones y lecciones personalizadas formuladas de acuerdo a las características del estudiante. Transmite conocimiento con técnicas de enseñanza uno a uno con el objetivo de mejorar el proceso de aprendizaje (Hasan et al., 2020).

La aplicación principal de un ITS es personalizar la guía, adaptar a los materiales de aprendizaje, analizar los estilos de aprendizaje de los estudiantes aplicando múltiples técnicas que permitan un proceso de enseñanza suave.

Su arquitectura está compuesta por los módulos principales Domino, Estudiante y Pedagógico y un cuarto módulo de interfaz de usuario (Figura 2.3). El módulo de Dominio contiene el conocimiento del tema que se pretende enseñar; el módulo de Estudiante, contiene el perfil del estudiante, preferencias y un diagnóstico sobre su progreso; el módulo pedagógico, determina las estrategias de enseñanza. Distintos autores pueden extender la arquitectura especificando elementos como bases de datos, módulos específicos con las técnicas y mecanismos de selección, perfilado y evaluación o considerando la interfaz de usuario también como un módulo principal.

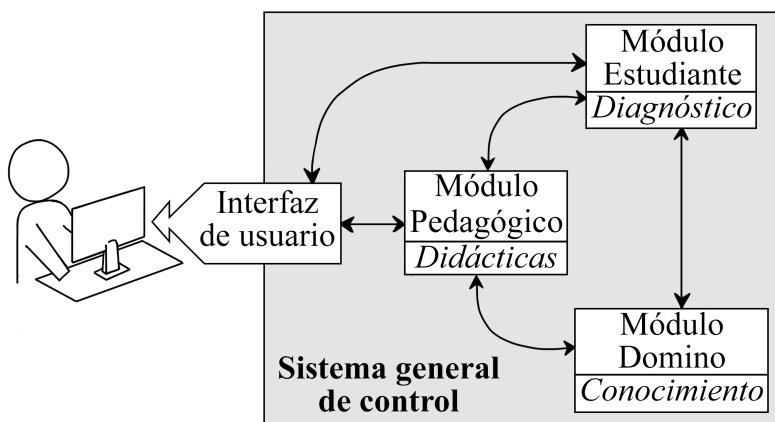


Figura 2.3
Arquitectura general de un ITS (Diseño propio)

2.4. Emoción.

En la teoría cognitiva se definen como un episodio de cambios sincronizados interrelacionados en los estados de todos o la mayoría de los cinco subsistemas orgánicos: cognitivo, neurofisiológico, motivacional, expresión motora, sentimiento subjetivo. Estos cambios se dan en respuesta a la evaluación de un estímulo externo o interno tan relevante como para concernir al organismo (Scherer, 2005). Las emociones son respuestas del organismo a una excitación fisiológica e involucran un conjunto de experiencias, actitudes y creencias sobre el mundo que influyen en la percepción de eventos y situaciones (Universidad de Alicante, 2007). Madera-Carrillo et al. (2015) lo sintetiza como un cambio fisiológico, una reacción orgánica producto de respuestas sistemáticas que reflejan la activación central de diversos sistemas del cuerpo en preparación a una reacción o acción física.

2.5. Affective Tutoring System (ATS).

Un ATS es un sistema basado en la computación afectiva propuesta por Picard (2000). Similar a un ITS, un ATS intenta replicar a un tutor humano en sus capacidades de adaptar el contenido didáctico, pero en este caso al estado emocional del aprendiz. Esto lo logra reaccionando en las situaciones donde el estado emocional pudiera afectar el proceso de aprendizaje y, por ende, en sus resultados (Petrovica et al., 2017). La figura 2.4 muestra cómo la arquitectura de un ATS es esencialmente similar a la de un ITS con algunas adiciones.

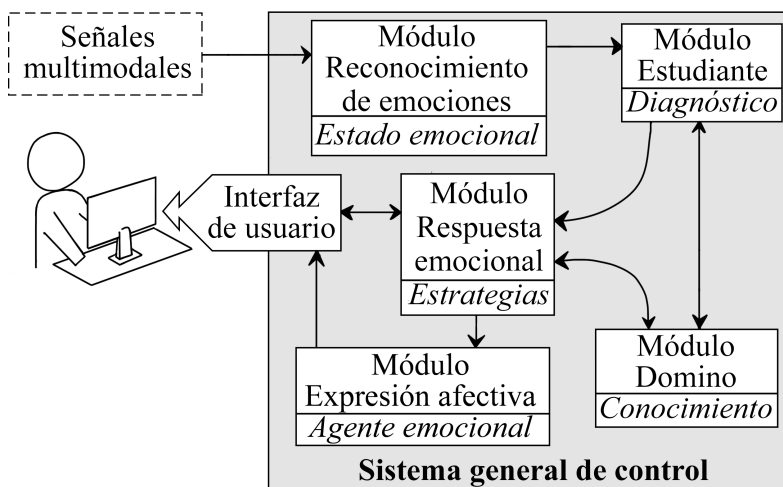


Figura 2.4
Arquitectura general de un ATS (Diseño propio)

El módulo de reconocimiento de emociones detecta diferentes características físicas y/o fisiológicas que permiten identificar, mediante un clasificador automático, el estado emocional

del estudiante; el módulo de respuesta emocional se encarga de evaluar la situación de la tutoría basándose en el estado emocional del estudiante y de responder adaptando el proceso de enseñanza mediante la aplicación de estrategias de regulación emocional; opcionalmente puede incorporarse un módulo de expresión afectiva que extiende la interfaz con un tutor virtual con sus propias emociones, que expresa en respuesta a las acciones y emociones del estudiante. Azevedo et al. (2017) orientan este agente a técnicas de aprendizaje autorregulado, donde se da a conocer al usuario cuál es su estado emocional, los factores que lo podrían estar ocasionando y estrategias de regulación que puede aplicar para regresar a un estado emocional favorable. Aunque Petrovica et al. (2017) manejan el ATS como una extensión de un ITS dada su similitud, en realidad, son sistemas independientes que pueden complementarse en diferentes grados.

3

Estado del arte

El estado del arte se aborda desde una perspectiva multidisciplinaria que recorre los niveles de aplicación, método y técnica descritos en el Capítulo 1 y se obtuvo mediante una búsqueda exhaustiva y sistemática de la literatura. Se emplearon diversas cadenas de búsqueda temáticas en Google Scholar y ACM Digital Library, complementadas con exploraciones en ResearchGate, Springer, IEEE Xplore, Taylor & Francis Online y PubMed. Adicionalmente, se aplicó la estrategia de la bola de nieve, siguiendo referencias cruzadas de los artículos identificados para descubrir trabajos relevantes no capturados en las búsquedas iniciales, que resultó particularmente eficaz para identificar enfoques cada vez más específicos en cuanto al tipo de señales (temporales, posturales) y su integración en entornos de RVI. El proceso de selección incluyó un primer filtro por título, luego por resumen (abstract), posteriormente por conclusiones y, finalmente, por lectura completa del texto. Se eliminaron los resultados duplicados entre diferentes búsquedas. Se priorizaron publicaciones en revistas de cuartiles 1 y 2 según Scimago, con un rango principal de 2018 a 2025; artículos de mayor antigüedad se incluyeron solo cuando su relevancia era considerada fundamental. La búsqueda inicial exploró el nivel de aplicación mediante términos como *Qualitative analysis*, *Virtual Reality*, *Child Sexual Abuse Prevention*. Se revisaron los primeros 150 resultados y de estos, se seleccionaron trabajos clave siguiendo los filtros descritos y solicitando acceso a aquellos no disponibles inicialmente. Otras búsquedas se centraron en la detección de emociones mediante gestos, movimientos y posturas en contextos de realidad virtual, utilizando cadenas como “*detecting emotions in virtual reality*” (primeros 120 resultados) y “*specific posture detection in temporal signals -visual +Virtual Reality*” (primeros 150 resultados). Finalmente, se llevó a cabo una búsqueda dedicada a instrumentos de registro emocional, con énfasis en el *Self-Assessment Manikin* (SAM), sus variantes y/o alternativas adaptadas al grupo etario infantil, priorizando propuestas que faciliten la autoevaluación subjetiva en contextos educativos o experimentales sensibles. Los trabajos identificados abordan perspectivas diversas: desde el análisis cualitativo de programas de prevención de abuso sexual infantil mediante realidad virtual, pasando por los efectos de la RVI en el compromiso y el desempeño del usuario, hasta la importancia de la regulación emocional y métodos de clasificación emocional basados en señales multimodales o unimodales. De particular relevancia para este proyecto es el enfoque propuesto por Lin et al.

(2023), que considera las condiciones operativas y restricciones específicas del uso de sistemas de RVI para la identificación de estados emocionales. En conjunto, los estudios sustentan la necesidad de enfoques no invasivos, unimodales y contextualizados en RVI, alineados con los objetivos de la presente investigación.

Los programas de prevención del abuso sexual infantil (ASI) suelen integrarse en iniciativas más amplias de autocuidado y autoprotección (Gubbels, van der Put et al., 2021). Los estudios cualitativos que analizan estos cursos como un todo revelan que la percepción del contenido didáctico varía significativamente según el grupo etario, lo que lleva a recomendar la segmentación por edad y el ajuste personalizado del material (Allen et al., 2020; Tutty, 2014), especialmente en niños de más de 10 años donde además se recomienda separar por género. En este contexto, las tecnologías de la información y la comunicación (TIC) aparecen como un recurso subutilizado: cuando se incorporan, rara vez se especifica la herramienta empleada y predomina el uso de materiales no interactivos y tradicionales, tales como vídeos, cuadernos de actividades, representaciones teatrales, libros y panfletos. El empleo de TIC se concentra principalmente en la fase de diseño, involucrando a los padres en el proceso (Malamsha et al., 2021) o generando recomendaciones y lineamientos para futuros proyectos (Gubbels, Assink et al., 2021; M. Lu, Barlow, Meinck & Neelakantan, 2023; M. Lu, Barlow, Meinck, Walsh & Wu, 2023). La percepción general de estos esfuerzos es positiva, reconociéndose un impacto favorable en los niños no solo en términos de conocimiento preventivo, sino también en el cultivo de emociones positivas (Gubbels, van der Put et al., 2021; Jones et al., 2020; Kenny et al., 2020; Scholes et al., 2014). Estos beneficios tienden a ser mayores cuando los participantes completan la totalidad del programa. La tabla 3.1 muestra el enfoque de los trabajos y el uso de las tecnologías de la información en programas de prevención ASI.

Tabla 3.1

Estado del arte en materia de prevención del Abuso Sexual Infantil usando TIC.

Autores		Enfoque	ASI	RV/TIC	Resultados
Gubbels, Assink et al. (2021)	P		Sí	No	Lista de componentes importantes que debe contener un curso de prevención de ASI. Se resalta la importancia de recibir esta educación en temprana edad.
Gubbels, van der Put et al. (2021)	P		Sí	No	Hay un incremento de conocimiento en niños que reciben estos cursos. La participación activa en juegos y roles es más efectiva. Abordar las habilidades socio-emocionales y usar juegos y acertijos genera mejores resultados. Ningún estudio aborda si la prevención es lograda.

Continúa en la siguiente página

Tabla 3.1 – Continuación de la página anterior

Autores	Enfoque	ASI	RV/TIC	Resultados
Malamsha et al. (2021)	P	Sí	TIC	Ajusta palabras y considera la opinión de los padres sobre los temas. El 81 % de los padres se dijeron satisfechos con la app y hubo incrementos en su interés y confianza sobre los temas.
Corno y Bouchard (2015)	T	No	RV	Una RV para el tratamiento de PTSD de mujeres abusadas sexualmente y una propuesta temática de un curso de prevención. La RV tiene efectos prometedores en terapias de exposición/simulación.
Jones et al. (2020)	P	Sí	TIC	El curso de prevención incrementó el conocimiento de los niños sobre el abuso. Completar el juego demostró ser más importante que las lecciones. Se requiere evaluar habilidades más que conocimiento.
Scholes et al. (2014)	P	Sí	TIC	Los atributos de los juegos serios pueden superar muchos de los desafíos para proporcionar una prevención efectiva. Sentimientos de positividad ayudan a la autoestima y confianza del aprendiz.
M. Lu, Barlow, Meinck y Neelakantan (2023)	P	Sí	NE	Recomendaciones y lineamientos para el diseño de programas de prevención basado en la revisión de estudios existentes y sus fallas.
M. Lu, Barlow, Meinck, Walsh y Wu (2023)	P	Sí	NE	Intervenciones con más de 3 sesiones tienen mejores efectos de los cursos y parecen ser mayores en niños de 8 años o más. El material principal son libros de trabajo.
Allen et al. (2020)	P	Sí	NE	Diferentes grupos etarios perciben el tema diferente. Estas herramientas permiten la detección de casos de abuso. El miedo y la ansiedad, aunque aparecen, no repercuten en los participantes de forma importante.

Continúa en la siguiente página

Tabla 3.1 – Continuación de la página anterior

Autores	Enfoque	ASI	RV/TIC	Resultados
Kenny et al. (2020)	P	Sí	NE	El 62 % recuerda haber participado en un curso de prevención y reportaron la mitad de incidentes de ASI que aquellos que no participaron. No se reportaron efectos adversos en la satisfacción sexual.
Tutty (2014)	P	Sí	NE	Involucrar a los profesores. Separar por género cuando se imparte a niños de 10 a 12 años. Se documentaron sentimientos incómodos en algunos participantes. Se reconoce que el manejo de los recursos escolares es un problema.

Nota. P: Prevención; T: Tratamiento; NE: No especificado; RV: Realidad Virtual; TIC: Se menciona el uso de alguna tecnología de la información

Profundizando en el uso de tecnologías inmersivas, existen algunas aproximaciones a la RVI en temas relacionados con el ASI, la regulación emocional y tutores inteligentes. No obstante, su aplicación principal es terapéutica y se orienta desde la perspectiva del especialista (Endendijk et al., 2021; Hassan et al., 2022). Solo en casos aislados, como Van Rooij et al. (2016), se observa un enfoque centrado en las respuestas emocionales de los niños, aunque orientado al manejo de ansiedad y no a la prevención del ASI. A pesar de ello, la RVI ha demostrado beneficios en facilitar la comunicación de experiencias traumáticas y la enseñanza de tópicos delicados (tabla 3.2).

Tabla 3.2

Estado del arte en materia de prevención del Abuso Sexual Infantil usando RV como elemento completo.

Autores	Enfoque	ASI	RV/TIC	Resultados
Endendijk et al. (2021)	T	Sí	Sí	Los juegos serios en el tratamiento de niños con posible abuso sexual facilitan la comunicación en la narrativa de los hechos traumáticos y en la enseñanza de tópicos importantes.
Hassan et al. (2022)	T	Sí	Sí	Se reportaron mejoras en la usabilidad del sistema tras familiarizarse con él. Tiene potencial didáctico-educativo, pero deben mejorarse la interactividad y el realismo.

Continúa en la siguiente página

Tabla 3.2 – Continuación de la página anterior

Autores	Enfoque	ASI	RV/TIC	Resultados
Van Rooij et al. (2016)	T	No	Sí	Redujo los niveles de ansiedad autoreportados por los participantes. La mecánica de interacción no fue fácil de comprender.
<i>Nota.</i> T: Tratamiento; RV: Realidad Virtual				

Hasan et al. (2020) proponen una taxonomía que posiciona a los ATS como una evolución de los ITS, al incorporar la capacidad de detectar y responder al estado emocional del estudiante para optimizar el proceso de aprendizaje donde el componente afectivo permite una adaptación más fina. Los ATS entre sí se diferencian principalmente en el modelo emocional referencial, donde destacan los de Pekrun (2006) y Russell (1980), las emociones objetivo consideradas y los mecanismos de regulación o respuesta implementados, que varían considerablemente entre trabajos, por ejemplo, en (Happy et al., 2015) se enfocaron en 10 emociones específicas, mientras que en T S y Guddeti (2020) se enfocaron solo en las involucradas en el proceso de aprendizaje. En la literatura, emociones negativas como el aburrimiento, la confusión y el enojo se identifican consistentemente como obstáculos para el aprendizaje, al interferir con la concentración y la motivación (Coto et al., 2022; Yadegaridehkordi et al., 2019). Por el contrario, emociones positivas como la sorpresa, la alegría y la concentración actúan como facilitadoras, promoviendo la motivación y el involucramiento. No obstante, emociones negativas también pueden preceder acciones favorables o correctivas en ciertos contextos, lo que sugiere una complejidad mayor en su rol (Taub et al., 2021). Es notable que muchos enfoques no distinguen estrictamente entre estados cognitivos y emocionales, agrupándolos bajo la categoría general de “emociones” y clasificándolos solo como deseables o no deseables. Esto es posible observar en trabajos como el Holzwarth et al. (2021) donde reportan usar los movimientos de la cabeza para inferir un estado afectivo durante el uso de una aplicación de RVI pero sus instrumentos se enfocan en usabilidad, presencia y carga cognitiva; o en Cebeci et al. (2019) donde se busca evaluar niveles de felicidad, tristeza, confort y ansiedad, durante escenas diseñadas para inducir calma, estrés, miedo y excitación. También se han propuesto técnicas de diseño basadas en evidencia de respuestas emocionales y la creación de patrones afectivos donde elementos del sistema induzcan respuestas emocionales específicas (Argasiński & Węgrzyn, 2019). Aunque estas ideas se observan parcialmente en implementaciones prácticas por ejemplo, en sistemas que envían mensajes emocionales adaptados al estado cognitivo del estudiante (Jiménez et al., 2018), la evidencia de su aplicación efectiva sigue siendo limitada, particularmente aquellas propuestas por Argasinski.

Las señales empleadas para la detección emocional son diversas, pero predominan las fisiológicas (respuesta galvánica de la piel, electroencefalograma, electrocardiograma, electromiograma, frecuencia cardíaca, variación en la frecuencia cardíaca), que requieren la colocación de sensores en el usuario, generando incomodidad, intimidación y posible interferencia en el proceso educativo. El reconocimiento facial mediante visión artificial (Behera et al., 2020; Cebeci et al., 2019) es otra aproximación común, aunque suele provocar en los usuarios la percepción de ser constantemente observados. En conjunto, uno de los principales desafíos radica en el uso de

mecanismos invasivos y en las implicaciones para la privacidad de los datos, especialmente con señales como el reconocimiento facial o datos fisiológicos (Alqahtani et al., 2021; Hasan et al., 2020; Yadegaridehkordi et al., 2019). A pesar de la multiplicidad de enfoques en ATS y tutores afectivos, ninguno de los trabajos identificados (tabla 3.3) implementa estos sistemas en entornos de RVI, ni los aplica con fines de capacitación infantil en prevención del ASI. Esta ausencia representa una brecha significativa, ya que la RVI podría ofrecer un contexto más ecológico y seguro para la detección no invasiva de emociones, alineándose con las necesidades de la propuesta unimodal basada en movimientos de cabeza y manos.

Tabla 3.3

Estado del arte respecto a tutores afectivos.

Autores	ASI	Señal	Modelo emocional	Resultados
Hasan et al. (2020)	No	Varios	Pekrun, Rusell y otros	Taxonomía de ITS y ATS, sus mecanismos de detección de emociones y técnicas de clasificación. También se abordan los estilos de enseñanza y recursos de regulación de contenido.
Argasiński y Węgrzyn (2019)	No	GSR, FC, VFC	NE	Técnica de diseño y creación de patrones afectivos, indicando los elementos involucrados, valores esperados y su mecanismo de detección.
Taub et al. (2021)	No	RF, CIE	NE	La emoción de sorpresa se relaciona inversamente con el nivel de competencia auto percibido. La frustración antecede a tomar mejores apuntes.
Yadegaridehkordi et al. (2019)	No	GSR, FC, VCF, ECG, EEG, EMG	Pekrun, Picard, Rusell y otros	Enfoque centrado en detectar y clasificar emociones pero no en la implementación práctica de estas acciones. Se buscan principalmente las emociones de aburrimiento, enojo y ansiedad.
Behera et al. (2020)	No	Visual	NE	Incremento de los movimientos de la cabeza y ojos conforme avanza la dificultad. La mayor parte del tiempo la emoción detectada fue neutral.
Jiménez et al. (2018)	No	NE	NE	Mensajes de elogio y apoyo mejoraron la motivación en mujeres y estudiantes de ingeniería con bajo desempeño en la prueba.
Alqahtani et al. (2021)	No	EEG, ECG, EMG	Russell	Se logró predecir la dificultad percibida con una precisión de 75 % en lo individual y 66.8 % en lo grupal.

Continúa en la siguiente página

Tabla 3.3 – Continuación de la página anterior

Autores	ASI	Señal	Modelo emocional	Resultados
Coto et al. (2022)	No	Varios	Discreto	Los estados cognitivos y estados emocionales son manejados de forma indistinta. Hay interés en incluir elementos de gamificación en el proceso de inducción y detección emocional.
<i>Nota.</i> GSR: Galvanic Skin Response; NE: No especificado; EEG – Electroencefalograma; ECG – Electrocardiograma; EMG – Electrocardiograma ; RF: Reconocimiento Facial; FC: Frecuencia Cardíaca; VFC: Variabilidad de la Frecuencia Cardíaca.; Comportamiento en Interfaz de Entrada				

Los tutores inteligentes en entornos de Realidad Virtual Inmersiva (RVI) han sido estudiados principalmente en relación con los efectos cognitivos y afectivos de la inmersión. Devon y von Mühlenen Adrian (2018) reportan mejoras en la percepción de emociones y en los niveles básicos de aprendizaje, atribuidas a la mayor inmersión y el sentido de presencia que ofrece la RVI. Marcos Pablos et al. (2022), por su parte, encontró que los tutores afectivos pueden incrementar la aceptación del sistema al generar un sentimiento de empatía en el usuario. En Lin et al. (2023), se evalúa el nivel de atención de los alumnos ante contenidos inmersivos mediante técnicas de reconocimiento facial combinadas con el seguimiento del foco de atención, estableciendo un vínculo directo entre lo observado y la reacción emocional expresada en el rostro. Esta aproximación resulta de particular interés para el presente proyecto, ya que el foco de atención puede determinarse de manera no invasiva a través de los movimientos y la orientación de la cabeza, sin requerir cámaras faciales adicionales.

Sin embargo, a pesar del potencial de tecnologías como la RVI y la realidad aumentada para redefinir los métodos de enseñanza Schweiger et al. (2022), aún se requiere mayor investigación sobre los procesos de detección emocional y sus causas asociadas a los contenidos específicos del curso (Azevedo et al., 2017). Esta necesidad también había sido planteada por Argasiński y Węgrzyn (2019), que enfatizaron la importancia de diseñar patrones afectivos contextualizados, aunque sin evidencia empírica de su implementación. En Yu et al. (2022) se aborda incluso la creación de un nuevo conjunto de datos extraídos de un electroencefalograma bajo la condición de inducción emocional en RVI, al considerar las diferencias dinámicas de ésta en comparación con estímulos emocionales en 2D (imágenes y videos).

La detección emocional en entornos de RVI se ha aplicado predominantemente al tratamiento de traumas y trastornos psicológicos (Reichenberger et al., 2020; Van Rooij et al., 2016), aprovechando la capacidad de la RVI para evocar en un ambiente sintético respuestas emocionales equivalentes a las del mundo real (Lipp et al., 2021). Asimismo, se ha investigado el proceso de inducción emocional mediante el diseño de escenarios para emociones específicas (Cebeci et al., 2019; Hofmann et al., 2021), utilizando principalmente señales fisiológicas como el electroencefalograma y el electrocardiograma para su detección. Debido a este enfoque terapéutico, los estudios suelen centrarse en emociones específicas y, en su mayoría, negativas, tales como la ansiedad, el estrés o el miedo. Un hallazgo relevante es la alta variabilidad interindividual en la expresión de una misma emoción ante un estímulo idéntico: Reichenberger

et al. (2020) asociaron estas diferencias a la gravedad de la ansiedad social padecida entre los participantes y que se reflejó en patrones de movimiento específicos, como hipervigilancia y evasión visual. De manera convergente, los trabajos de Cebeci et al. (2019) y Holzwarth et al. (2021) indican que las emociones también se manifiestan a través de la orientación y los movimientos de la cabeza. Estos resultados proporcionan un indicio prometedor sobre el potencial de señales no invasivas derivadas exclusivamente de la cabeza para la detección emocional en RVI, sin recurrir a sensores fisiológicos adicionales.

Tabla 3.4

Estado del arte respecto a tutores afectivos y RVI.

Autores	ASI	Señal	RVI	Resultados
Devon y von Mühlenen Adrian (2018)	No	NA	Sí	Mejoras en la percepción de emociones y en los niveles básicos de aprendizaje.
Holzwarth et al. (2021)	No	CIE RVI	Sí	El movimiento de la cabeza se correlaciona con la presencia, usabilidad, demanda físico-mental y desempeño percibido.
Lin et al. (2023)	No	RF, FA	Sí	Se puede evaluar efectivamente el nivel de atención de los alumnos.
Azevedo et al. (2017)	No	Varios	No	Es necesaria más investigación en los procesos multimodales en la detección de emociones y sus causales asociados a los contenidos de un curso.
Schweiger et al. (2022)	No	NA	Sí	La RVI y RA pueden facilitar el trabajo del profesor. Hay pocos estudios enfocados a la población infantil.
Cebeci et al. (2019)	No	FC, Ocular	Sí	La RVI evoca las mismas emociones que el mismo contexto en la realidad. Hay cambios en la FC y cambios emocionales asociados a los gustos de las personas.
Marcos Pablos et al. (2022)	No	RF, voz	No	Los tutores afectivos pueden incrementar la aceptación de un sistema o del tutor mismo al despertar y aprovechar el sentimiento de empatía.
Lin et al. (2023)	No	RF, Punto Focal	Sí	El nivel de atención puede ser efectivamente evaluado etiquetando áreas de interés de acuerdo a su importancia.
Van Bennekom et al. (2021)	No	VFC	Sí	La frecuencia cardíaca varió significativamente incluso antes de la exposición a la RVI en pacientes con trastorno más severo. Abre la posibilidad al diagnóstico.

Continúa en la siguiente página

Tabla 3.4 – Continuación de la página anterior

Autores	ASI	Señal	RVI	Resultados
Reichenberger et al. (2020)	No	FA, CIE	Sí	Se pudo identificar niveles mayores de miedo en los participantes con ansiedad social alta, además de comportamientos específicos notables.
Hofmann et al. (2021)	No	EEG	Sí	Se obtuvo una precisión del 60.76 % al 60.83 % en la predicción de la excitación de forma unimodal.
<i>Nota.</i> NA: No aplica, o ninguno; NE: No específica; CIE: Comportamiento en Interfaz de entrada; RVI: Realidad Virtual Inmersiva; FC: Frecuencia Cardíaca; RF: Reconocimiento facial; FA: Foco de atención; VFC: Variabilidad de la Frecuencia Cardíaca; EEG: Electroencefalograma				

Los métodos computacionales para la clasificación de estados emocionales se agrupan típicamente en cuatro áreas principales:

1. Reconocimiento facial.
2. Uso de sensores fisiológicos y procesamiento de señales asociadas.
3. Análisis de la voz.
4. Análisis corporal y del comportamiento.

Wang et al. (2022), profundiza en esta taxonomía partiendo de métodos unimodales o multimodales. En este sentido, el análisis corporal se posiciona como un método unimodal de reconocimiento visual, representado claramente por trabajos como Baloch et al. (2023). Este aspecto apoya lo novedoso de la aproximación propuesta en este proyecto. El análisis del comportamiento corporal y postural ha sido ampliamente estudiado como indicador tanto de identidad individual como de estados emocionales. Cada persona exhibe patrones de movimiento únicos bajo diferentes situaciones sociales (Yaremych & Persky, 2019). Siguiendo esta premisa, Pfeuffer et al. (2019) analizaron los movimientos de la cabeza en entornos de RVI para identificar al usuario durante tareas específicas, alcanzando una exactitud del 63 %. Desde una perspectiva emocional, el comportamiento postural y los movimientos corporales constituyen indicadores clave de la intensidad y, en algunos casos, de la especificidad de las emociones (Wallbott, 1998). Wallbott enfatizó que la mayor parte de la información no verbal proviene de la parte superior del cuerpo (incluyendo rostro, cabeza, torso, manos y gestos), mientras que Mehrabian (2017) propuso el principio 7-38-55, según el cual en la comunicación de actitudes y emociones, el 55 % proviene de los movimientos corporales y expresiones faciales, el 38 % del tono y ritmo vocal, y solo el 7 % del contenido verbal. Este concepto ha sido respaldado por estudios posteriores, como el de Sapiński et al. (2019), que analizaron el reconocimiento de emociones a partir del cuerpo completo contra solo la parte superior, encontrando una pérdida mínima de exactitud (alrededor del 4 %) al excluir la información de la parte inferior del cuerpo.

La identificación de estados emocionales mediante postura y movimiento corporal está fuertemente influenciada por el contexto en el que ocurren y el momento de evaluación (i.e., durante la intervención, después de la intervención, de forma continua o solo en una ocasión). En una misma actividad no siempre se manifiestan todas las emociones posibles, y una misma emoción puede expresarse de manera distinta según la tarea o situación.

Generalmente, el ground truth para la clasificación se genera mediante observadores externos (Gerdmann et al., 2025), o a través de actores que representan versiones estereotípicas de las emociones solicitadas (Ahmed et al., 2015; Piana et al., 2016; Sapiński et al., 2019). Distintos factores sociales, culturales e incluso anatómicos influyen en la manera en que hombres y mujeres se acoplan a diferentes posturas y acciones (e.g., sentarse con las piernas cruzadas). Esto deja en segundo plano las sutilezas del comportamiento real (e.g., secuencias de movimientos, amplitud, dirección, frecuencia), que suelen caracterizarse solo cualitativamente (e.g., “manos temblorosas”) y que constituyen elementos fundamentales del comportamiento natural. Estas limitaciones generan una precisión variable en la identificación humana: emociones como alegría y tristeza se reconocen correctamente entre el 81 % y el 87 % (Piana et al., 2014), mientras que sorpresa, miedo y disgusto tienden a confundirse. Esto contrasta con la intensidad subjetiva experimentada, donde miedo, enojo y tristeza suelen ser altas (Magdin et al., 2021). En general, la capacidad humana para identificar estados emocionales ronda el 63 %.

En el contexto específico de la RVI, se ha aprovechado su capacidad para recrear situaciones realistas y evaluar movimientos de la cabeza como indicadores de sesgos en la interacción social, como la ansiedad interracial (Yaremych & Persky, 2019) o estados cognitivos como el estrés (Brambilla et al., 2022). También en (Mancuso et al., 2024), si bien hicieron uso de señales fisiológicas, se encontró relación entre los movimientos de la cabeza y los niveles de valencia, activación y dominio auto reportados. Estos estudios destacan el potencial de señales no invasivas derivadas de la cabeza y del foco de atención. Finalmente, es notable en la literatura que la granularidad, especificidad o número de emociones objetivo impacta directamente en los resultados: cuanto más específica sea la detección, menor será la exactitud obtenida. Esto suele llevar a reducir las categorías en las evaluaciones (Noroozi et al., 2021), llegando incluso a clasificaciones binarias (positivas vs. negativas) para mejorar el rendimiento.

Dada esta variabilidad en el número y tipo de emociones consideradas, las técnicas de identificación de emociones son tan diversas como la cantidad de estados emocionales evaluados en la literatura. Brambilla et al. (2022) emplearon Hidden Markov Models combinados con medidas de tendencia central para modelar secuencias temporales de movimientos de cabeza en RVI. Las redes neuronales convolucionales (CNN), recurrentes (RNN) y Long Short-Term Memory (LSTM) destacan por su capacidad para procesar características espaciales y temporales en posturas y trayectorias (Alqahtani et al., 2021; Behera et al., 2020; Lin et al., 2023; Sapiński et al., 2019; T S & Guddeti, 2020). Ahmed et al. (2015) aplicaron un kernel basado en Dynamic Time Warping (DTW) una técnica de alineamiento secuencial no lineal, para procesar secuencias de longitud variable de las características de interés para el estudio del andar. Pfeuffer et al. (2019) utilizaron un clasificador de bosques aleatorios (Random Forest) para la identificación a partir de movimientos de cabeza, mientras que Marcos Pablos et al. (2022) recurrió a filtros de Kalman para el análisis vocal. Otros enfoques combinan múltiples clasificadores en forma híbrida o comparativa, como Support Vector Machines, K-Nearest Neighbor, Naïve Bayes, regresión logística, Random Forest y perceptrones multicapa (Siam et al., 2022; Zhang et al., 2020).

Esta diversidad metodológica, aunque enriquecedora, también pone de manifiesto limitaciones

comunes: dependencia de señales invasivas o multimodales, necesidad de grandes volúmenes de datos etiquetados externamente y escasa adaptación a poblaciones infantiles en contextos preventivos. La literatura carece de propuestas unimodales basadas exclusivamente en señales nativas del visor de RVI (cabeza y manos), con etiquetado subjetivo y espontáneo, lo que representa una brecha clave que el módulo de detección emocional de AVAI ASI busca cubrir.

Dada la diversidad de técnicas de clasificación y la influencia de la granularidad en la precisión discutida previamente, la generación de etiquetas emocionales (*ground truth*) representa un componente crítico. El modelo emocional continuo, descrito por las dimensiones de valencia, activación y dominio, destaca como el enfoque más utilizado para representar estados emocionales en la literatura reciente (Mancuso et al., 2024; Wang et al., 2022). El *Self-Assessment Manikin* (SAM), desarrollado por Bradley y Lang (1994), es una técnica pictórica no verbal ampliamente adoptada que mide los niveles autopercebidos en cada una de estas dimensiones en respuesta a distintos estímulos. Magdin et al. (2021) compararon el SAM con detecciones automáticas multimodales (electrocardiograma y software Affdex para análisis facial), obteniendo una correlación de Pearson altamente significativa ($r(138) = 0.92, p < 0.001$). Este resultado respalda la validez de la autoevaluación subjetiva como *ground truth* confiable.

En entornos de RVI, el SAM se ha implementado mediante la digitalización del instrumento (Mancuso et al., 2024; Woodworth & Borst, 2024). Brambilla et al. (2022) conservaron los iconos originales y los presentaron en la interfaz virtual como si fueran una hoja física, aproximación respaldada por evidencia de que las respuestas en instrumentos físicos y virtuales son comparables (Regal et al., 2019). Además, existe una preferencia por responder directamente en el entorno virtual para evitar interrupciones en la experiencia inmersiva.

Otras alternativas en RVI buscan facilitar la evaluación emocional manteniendo las dimensiones de valencia y activación (Toet et al., 2020). Sin embargo, a pesar de su uso extendido, el SAM presenta limitaciones significativas en el grupo etario infantil: los niños suelen requerir explicaciones prolongadas y a veces confusas para interpretar los pictogramas. Esto ha motivado el desarrollo de instrumentos adaptados, como la herramienta pictográfica propuesta por Desmet et al. (2016), basada en caricaturas de personajes que expresan ocho estados de ánimo. Alternativamente, algunos estudios complementan el SAM con instrumentos discretos de emociones categóricas, ajustados al propósito del estudio (Di Tecco et al., 2024), aunque esto limita al participante a un conjunto predefinido de emociones que podrían no capturar fielmente su experiencia real. En conjunto, estos avances destacan la necesidad de instrumentos de autoevaluación válidos, no intrusivos y adaptados a niños en RVI, que minimicen interrupciones y respeten la subjetividad individual; aspectos centrales en el diseño del protocolo de etiquetado emocional del AVAI ASI, donde se emplea una adaptación del SAM para obtener etiquetas auténticas y espontáneas.

El estado del arte revela un panorama consolidado en varios aspectos clave de la computación afectiva y la interacción humano-computadora aplicada a entornos educativos y preventivos, pero también pone en evidencia brechas críticas que justifican y delimitan el aporte de la presente investigación. En primer lugar, los programas de prevención del ASI siguen dependiendo predominantemente de metodologías presenciales tradicionales o de recursos TIC básicos (e.g., videos, tabletas, materiales impresos), con una adopción marginal de la RVI y prácticamente nula de tutores afectivos capaces de regular el estado emocional en tiempo real. El uso de la RVI se limita casi exclusivamente a contextos terapéuticos (e.g., manejo de

traumas, ansiedad, estrés postraumático), desde la perspectiva del especialista y sin enfoque en la prevención primaria infantil. En segundo lugar, los tutores inteligentes y afectivos han avanzado en la detección y regulación emocional, pero predominan enfoques multimodales invasivos (sensores fisiológicos, reconocimiento facial, voz) que comprometen la usabilidad, la privacidad y la portabilidad, especialmente en entornos educativos con niños y multiusuario. Los métodos unimodales basados en comportamiento corporal y postural muestran potencial, particularmente en señales derivadas de la parte superior del cuerpo (cabeza, manos), pero su aplicación en RVI se restringe a la identificación de usuario, atención o estados cognitivos/ansiedad en adultos, sin extenderse a la detección emocional espontánea en niños ni a contextos preventivos sensibles como el ASI. En tercer lugar, los procedimientos de etiquetado emocional siguen dependiendo en gran medida de observadores externos o actuaciones estereotípicas, introduciendo sesgos socioculturales, pérdida de sutilezas individuales y disociación entre expresión y experiencia subjetiva. El modelo dimensional de Russell y el instrumento SAM constituyen estándares válidos y con alta correlación con medidas objetivas, pero presentan limitaciones interpretativas en niños y requieren explicaciones prolongadas. Las adaptaciones pediátricas existentes (pictogramas caricaturescos, soporte categórico discreto) mejoran la usabilidad, pero suelen restringir las emociones posibles o no capturan la espontaneidad real. Finalmente, las técnicas de clasificación son diversas (e.g, CNN, LSTM, Random Forest, DTW, técnicas híbridas), pero su rendimiento se ve afectado por la granularidad emocional, la variabilidad interindividual y el contexto, lo que frecuentemente lleva a reducir categorías o a clasificaciones binarias. Ninguno de los enfoques identificados combina de manera integrada: (I) detección unimodal no invasiva basada exclusivamente en señales nativas del visor de RVI, (II) etiquetado subjetivo y espontáneo mediante autoevaluación adaptada a niños, (III) regulación afectiva en tiempo real, y (IV) aplicación en capacitación de población infantil sobre prevención primaria del ASI. Estas brechas constituyen el núcleo del problema científico que motiva el desarrollo de AVAI ASI. El presente trabajo propone un tutor afectivo inmersivo que aborda simultáneamente estas limitaciones mediante un módulo unimodal de detección emocional basado en comportamiento espacial natural, etiquetado derivado directamente de la experiencia subjetiva del niño y regulación dinámica del contenido en un entorno de RVI diseñado específicamente para la prevención del abuso sexual infantil. De este modo, se avanza hacia soluciones más escalables, respetuosas de la privacidad, ecológicamente válidas y centradas en la población objetivo.

4

Planteamiento del problema

Los programas de prevención del abuso sexual infantil (ASI), problemática de salud pública y social de alcance global, se basan predominantemente en metodologías tradicionales presenciales que dependen de facilitadores humanos y materiales didácticos convencionales (libros, vídeos 2D, dramatizaciones) (Walsh et al., 2019). Aunque las tecnologías de la información y la comunicación han comenzado a incorporarse en contextos educativos, la evidencia sobre el empleo de la realidad virtual inmersiva (RVI) en este ámbito específico sigue siendo escasa. Cuando estas tecnologías disruptivas se utilizan en relación con el ASI, su aplicación se centra casi exclusivamente en el tratamiento terapéutico de víctimas (e.g. traumas o el manejo del trastorno de estrés postraumático), dejando prácticamente inexplorado su potencial como herramienta preventiva.

Otro aspecto crítico radica en la preocupación de padres, madres y docentes respecto del posible impacto emocional negativo que la exposición a un tema tan sensible como el ASI puede generar en los niños y niñas, particularmente en forma de ansiedad, miedo o estrés. En los programas tradicionales, la regulación emocional y el ajuste del contenido durante la sesión se confían implícitamente a la capacidad del facilitador humano, sin que exista evidencia sistemática ni protocolos explícitos que garanticen dicha intervención. Esta limitación se agudiza cuando los grupos son numerosos, pues la atención individualizada se vuelve inviable, comprometiendo tanto la calidad de la experiencia formativa como la eficacia preventiva del programa en su conjunto.

A diferencia de los enfoques tradicionales, los sistemas de tutoría afectiva cuyo fundamento teórico se remonta a los trabajos de Picard (2000) buscan emular el comportamiento de un tutor humano sensible al estado emocional del estudiante. Estos sistemas detectan automáticamente las emociones del aprendiz y ajustan dinámicamente el contenido con el fin de mantener niveles óptimos de atención, motivación e involucramiento (Khadimallah et al., 2020). De este modo, la regulación emocional deja de depender exclusivamente de la intervención humana y pasa a integrarse como una capacidad inherente al sistema.

Un componente esencial de los tutores afectivos es el módulo de detección emocional. En la actualidad, la mayoría de los enfoques existentes se apoyan en el reconocimiento de expresiones faciales, el análisis de voz, el procesamiento del lenguaje natural o, predominantemente, en

señales fisiológicas (e.g. frecuencia cardíaca, respuesta galvánica en piel, actividad cerebral). Este último grupo requiere la colocación de múltiples sensores sobre el usuario, lo que se percibe como invasivo, incómodo y poco privado. Además, por su naturaleza bioeléctrica, dichas señales son altamente susceptibles al ruido y a artefactos generados por movimientos involuntarios o factores ambientales (Marín-Morales et al., 2020)

Un enfoque alternativo considerado menos invasivo es el análisis del comportamiento postural, principalmente mediante sensores de profundidad como Kinect o sistemas de visión artificial que estiman la posición y relaciones geométricas de las articulaciones corporales. No obstante, esta modalidad presenta limitaciones derivadas del propio hardware: campo de visión restringido, dependencia de condiciones de iluminación, sensibilidad a la distancia y oclusiones, así como errores inherentes a la estimación intermedia de las articulaciones. A ello se suma la exposición total del usuario eliminando cualquier garantía de privacidad y el hecho de que prácticamente ninguno de estos estudios se ha llevado a cabo en entornos de RVI, donde tanto la captura de movimientos como el contexto emocional difieren sustancialmente.

Aunque las aproximaciones multimodales han reportado mayor precisión en entornos controlados, su conveniencia práctica y escalabilidad resultan claramente limitadas, especialmente en escenarios multiusuario y en poblaciones infantiles. La necesidad de múltiples sensores externos incrementa significativamente los costos, reduce la portabilidad y compromete la usabilidad, al tiempo que introduce complejidades logísticas y éticas difícilmente asumibles en contextos educativos reales. El conjunto de estas restricciones constituye lo que en el presente trabajo se denomina el problema de la multimodalidad.

La parte complementaria en la detección emocional radica en los procedimientos de obtención y etiquetado de las emociones expresadas. En la mayoría de los estudios, estas emociones se elicitán mediante inducción controlada, elicitación dirigida o mediante la actuación deliberada de los participantes (H. Lu et al., 2025). Aunque la inducción puede ser pasiva o activa (Somarathna et al., 2023) y emplear estímulos contextualizados, el proceso suele seleccionar a priori un conjunto reducido de emociones objetivo y asociarles directamente las señales registradas, ya sea mediante juicio de observadores externos o por instrucción explícita al sujeto. Este enfoque presenta tres limitaciones críticas que comprometen la validez ecológica de los datos:

1. Individualidad: la expresión se orienta hacia patrones estereotípicos esperados, en lugar de capturar manifestaciones espontáneas y naturales del estado afectivo real.
2. Dimensión sociocultural: los gestos y posturas asociados a una misma emoción varían significativamente entre culturas y regiones, lo que introduce sesgos cuando los corpus se construyen en poblaciones homogéneas o bajo instrucciones estandarizadas.
3. Interioridad: se omite la experiencia subjetiva del participante, generando una disociación entre la emoción supuestamente expresada y el estado emocional interno, lo que resulta en etiquetas que no necesariamente reflejan la vivencia auténtica del sujeto.

En consecuencia, los conjuntos de datos en la literatura de computación afectiva adolecen de una representación limitada y potencialmente sesgada de la verdadera diversidad emocional humana. Este problema del etiquetado a priori, junto con el anteriormente descrito problema de la multimodalidad, constituyen las dos principales barreras que restringen la generalización

y el despliegue práctico de tutores afectivos en escenarios reales, especialmente en aplicaciones sensibles y con poblaciones infantiles.

5

Propuesta de solución

5.1. Objetivos

5.1.1. General

Desarrollar e implementar un modelo unimodal de detección y clasificación de estados emocionales en niños y niñas, basado exclusivamente en características espacio-temporales del movimiento natural de la cabeza y las manos, capturadas en un ambiente virtual de aprendizaje inmersivo, utilizando etiquetas derivadas de la autoevaluación subjetiva de los participantes con el propósito de integrarlo en un tutor afectivo orientado a la prevención del abuso sexual infantil.

5.1.2. Específicos

- a) Diseñar, implementar y evaluar un clasificador emocional unimodal basado exclusivamente en características espacio-temporales de la posición y orientación de la cabeza y las manos, capturadas por los sensores nativos del visor de realidad virtual inmersiva, sin utilizar sensores externos adicionales (fisiológicos o de visión artificial).
- b) Adaptar instrumentos de autoevaluación emocional (*Self-Assessment Manikin* y cuestionarios simplificados) para obtener etiquetas subjetivas y continuas de valencia y activación directamente de niños y niñas inmediatamente después de cada escena del AVAI ASI.
- c) Construir un corpus emocional pediátrico ecológicamente válido en el contexto de prevención del abuso sexual infantil, compuesto por trayectorias de la cabeza y las manos, datos de foco de atención, relaciones geométricas espaciales y auto-reportes emocionales, recolectados en entornos de realidad virtual inmersiva con niños y niñas de 8 a 10 años.

5.2. Justificación

El abuso sexual infantil representa una de las formas más graves de violencia y violación de derechos humanos, con una prevalencia nacional en México del 2.5 % y particularmente en el estado de Morelos del 3.6 % en mujeres y 1.9 % en hombres (Valdez-Santiago et al., 2020). Este problema está asociado a secuelas psicológicas e interpersonales que se extienden hasta la edad adulta, incluyendo sentimientos de miedo y ansiedad, baja autoestima, estrés postraumático, abuso de sustancias nocivas, autolesiones y conductas suicidas (Lalor & McElvaney, 2010). Uno de los mecanismos de prevención es la participación en programas educativos que incrementen la información sobre el desarrollo sexual saludable y la identificación temprana de situaciones de riesgo (Kim & Kang, 2017; Tutty et al., 2019). Es imperativo, por tanto, implementar estrategias que optimicen el uso de los recursos disponibles y promuevan una comprensión más profunda, duradera y significativa de los contenidos relacionados con la prevención del abuso sexual infantil y la identificación de situaciones de riesgo. Sin embargo, no atender el manejo emocional durante estos cursos puede comprometer los procesos de aprendizaje, al generar estados afectivos negativos que interfieren con la atención, la motivación y la retención citado en (Jarrell et al., 2017, citado en (Taub et al., 2021)). Esta situación, a su vez, alimenta la preocupación recurrente de padres y profesores acerca del posible impacto emocional adverso en los participantes (Tutty, 2014).

Las emociones desempeñan un rol central en la vida humana al modular funciones cognitivas esenciales, tales como la atención, la memoria y la toma de decisiones, actúan como precursoras de la motivación y tienen impacto en la memoria. En consecuencia, su monitoreo y regulación durante un programa de capacitación no solo resultan altamente deseables, sino necesarios para maximizar los resultados educativos en un tema de alta sensibilidad. La incorporación de tutores afectivos es una oportunidad clave: permite monitorear en tiempo real los estados emocionales de los niños y niñas, ajustar el contenido del curso con el fin de mantener la motivación y ayudar en la contención de emociones que pudieran afectar el proceso de capacitación, superando así las limitaciones de escalabilidad, personalización y seguridad emocional inherentes a los enfoques tradicionales. El presente trabajo se justifica desde múltiples perspectivas:

- Social: busca contribuir a la reducción de la prevalencia del abuso sexual infantil y a visibilizar el porcentaje de casos no denunciados, en gran medida por desconocimiento o desinformación.
- Pedagógica: optimiza los recursos humanos y temporales de los programas preventivos al emplear realidad virtual inmersiva, cuyos beneficios en presencia, aprendizaje y retención a largo plazo están ampliamente documentados, junto con un tutor afectivo capaz de regular el estado emocional del niño o niña, incrementando así la motivación y la eficacia formativa.
- Científico-tecnológica: aborda simultáneamente dos limitaciones críticas del estado del arte en computación afectiva:
 1. el problema de la multimodalidad, al proponer un modelo unimodal que utiliza únicamente las trayectorias naturales de la cabeza y las manos capturadas

exclusivamente por los sensores nativos del visor de RVI, eliminando la necesidad de sensores fisiológicos adicionales, cámaras externas o dispositivos invasivos; y

2. el problema del etiquetado a priori, al derivar las etiquetas emocionales directamente de la autoevaluación subjetiva y espontánea de los participantes, en lugar de depender de actuación, inducción controlada u observadores externos.

Como resultado, esta investigación genera un modelo computacional de clasificación emocional más simple, portátil, y respetuoso de la privacidad que, sumada a la adopción progresiva de los dispositivos de RVI (Olmos et al., 2018), facilita su escalamiento y su integración en programas de capacitación y entrenamiento (Alizadehsalehi et al., 2019) como el AVAI ASI. De este modo se promueve un acceso más equitativo a programas preventivos de alta calidad, fortaleciendo las habilidades de detección temprana y autoprotección en niños y niñas, y contribuyendo directamente a la disminución de una de las formas más graves de violencia infantil.

En síntesis, la investigación justifica su realización por su potencial para generar un impacto social relevante, superar limitaciones técnicas identificadas en la literatura y ofrecer una solución innovadora que combine realidad virtual inmersiva con computación afectiva en un dominio de alta sensibilidad social como es la prevención del abuso sexual infantil.

5.3. Hipótesis

“ Existe una relación significativa y útil entre las características espacio-temporales del movimiento natural de cabeza y manos capturadas por los sensores nativos de un visor de realidad virtual inmersiva, y el estado emocional autopercebido de niños y niñas de 8 a 10 años durante sesiones de prevención del abuso sexual infantil, permitiendo desarrollar un modelo unimodal de detección emocional que habilite la regulación dinámica del contenido del programa para mejorar la experiencia afectiva y pedagógica de los participantes. ”

5.4. Alcance y limitaciones

5.4.1. Alcances

El presente trabajo tiene como alcance principal el diseño, implementación y evaluación experimental de un modelo unimodal de detección emocional basado exclusivamente en la información de ubicación y orientación de la cabeza y las manos capturada por los sensores nativos de un visor de realidad virtual inmersiva. Específicamente, la investigación abarca:

- El desarrollo de un clasificador emocional unimodal entrenado y validado con datos reales de niños y niñas de 8 a 10 años.
- La construcción de un corpus emocional en contexto de RVI con autoevaluación subjetiva.
- La adaptación de instrumentos de autoevaluación emocional para población infantil en entornos inmersivos. El estudio se realizará en un entorno semi-controlado con participantes de escuelas públicas, garantizando el cumplimiento de todos los estándares éticos y de protección infantil con el apoyo del Instituto Nacional de Salud Pública.

5.4.2. Limitaciones

La investigación presenta las siguientes delimitaciones:

- Población: Se limita a niños y niñas de 8 a 10 años de edad escolar, estudiantes de escuelas públicas de municipios del estado de Morelos.
- Contexto de aplicación: El tutor afectivo AVAI ASI se diseña específicamente para programas de prevención primaria del abuso sexual infantil y no se evalúa en contextos terapéuticos ni con sobrevivientes de ASI.
- Modalidad tecnológica: El modelo de detección se restringe al uso exclusivo de señales nativas del visor de RVI. No se incluyen sensores fisiológicos externos, reconocimiento facial ni análisis de voz.
- Dimensiones emocionales: Se centra principalmente en las dimensiones de valencia y activación del modelo circunplejo de Russell. No se abordan emociones discretas complejas ni estados cognitivos relacionados al proceso de aprendizaje.
- Alcance temporal: El estudio se limita a sesiones experimentales semi controladas y no incluye seguimiento longitudinal a mediano o largo plazo de los participantes.
- Entorno de implementación: Las pruebas se realizarán en condiciones escolares semi controladas, por lo que los resultados deben interpretarse considerando las limitaciones propias del entorno.

6

Metodología

6.1. Enfoque y tipo de investigación

La presente investigación adopta un enfoque cuantitativo y se enmarca principalmente en un estudio de desarrollo tecnológico con un componente observacional y correlacional. Su objetivo central es diseñar, implementar y validar un modelo unimodal de detección emocional que permita extender las capacidades del sistema AVAI ASI hacia un tutor afectivo. El estudio busca determinar si es posible estimar los niveles de valencia y activación emocional de niños y niñas de 8 a 10 años durante su exposición a un entorno de realidad virtual inmersiva, mediante el análisis de información espacial obtenida de los dispositivos de RVI (posición, orientación y movimiento de cabeza y manos) y características derivadas de dichas señales. Se analizan las respuestas emocionales reportadas por los participantes mediante los instrumentos SAM y REN ante las ocho escenas del programa AVAI ASI en su versión inmersiva. Aunque no se manipulan variables de forma experimental, se examina la relación entre las características espacio-temporales extraídas y los estados emocionales autopercebidos, con el fin de evaluar la viabilidad del modelo propuesto. Los datos analizados fueron recolectados durante las pruebas realizadas los días 24, 25 y 26 de marzo de 2025 por el grupo conjunto del Centro Nacional de Investigación y Desarrollo Tecnológico (Cenidet) y el Instituto Nacional de Salud Pública (INSP).

6.2. Definición de la población muestra

Esta investigación emplea técnicas de detección emocional en realidad virtual dentro del sistema AVAI ASI para la capacitación en prevención de abuso sexual infantil. La selección se realizó mediante un muestreo no probabilística por conveniencia, en colaboración con el INSP. Dado que el programa de prevención busca aplicarse en las edades más tempranas posibles, se definió como población objetivo a niños y niñas de educación primaria, considerando que las características ergonómicas del visor de realidad virtual no representaran un obstáculo

significativo para su participación. El rango etario se determinó siguiendo las observaciones encontradas en (J.-E. Velázquez-Cano, 2021).

6.2.1. Criterios de inclusión

- Niños y niñas de nacionalidad mexicana, cursando educación primaria, con edades comprendidas entre 8 y 10 años. Este rango etario se estableció a partir de observaciones y experiencia previa durante el desarrollo del AVAI ASI.
- Distribución balanceada entre niños y niñas.
- Contar con consentimiento informado firmado por los padres o tutores legales, así como con el asentimiento verbal y voluntad expresa del niño o niña.
- Poca o nula experiencia con dispositivos de RVI.

6.2.2. Criterios de exclusión

- Participantes diagnosticados con padecimientos que afecten significativamente su atención, nivel de actividad física o capacidad para interactuar con el entorno de realidad virtual (por ejemplo, epilepsia fotosensible, trastornos vestibulares graves o dificultades motoras importantes).
- Niños o niñas que no pudieran comprender o seguir las instrucciones básicas del estudio.

Debido a la naturaleza del muestreo por conveniencia y al tamaño moderado de la muestra, los resultados no pretenden ser estadísticamente representativos de la población infantil mexicana en su totalidad. Sin embargo, la muestra refleja las características demográficas del grupo etario objetivo y del contexto escolar en el que se pretende aplicar el programa de prevención.

6.3. Materiales e instrumentos

Para llevar a cabo esta investigación se utilizaron los siguientes materiales e instrumentos:

6.3.1. Equipo de Realidad Virtual Inmersiva

Se emplearon visores Meta Quest 2, equipados con un procesador Snapdragon XR2, 6GB de RAM y pantallas LCD con resolución de 1832×1920 píxeles por ojo. El visor ofrece una

frecuencia en pantalla de hasta 90 Hz, y cuenta con controladores Oculus Touch que proporcionan seguimiento nativo de 6 grados de libertad (6-DoF) permitiendo una captura precisa de las trayectorias espaciales sin necesidad de sensores externos adicionales.

6.3.2. Plataforma tecnológica

El AVAI «Pinta tu raya ASI» fue desarrollado en colaboración con el Instituto Nacional de Salud Pública para servir como material didáctico en la capacitación en prevención del Abuso Sexual Infantil en niños y niñas de edad primaria. La plataforma fue desarrollada en Unity Engine, y es una plataforma diseñada para explorar técnicas de detección emocional y el desarrollo de tutores afectivos. En este trabajo se presenta como un mecanismo unimodal para la recolección de datos sobre el comportamiento postural del usuario, con el objetivo de identificar su estado emocional mediante técnicas de inteligencia artificial. El AVAI ASI consta de ocho escenas didácticas de las cuales cinco son interactivas. Aunque las escenas no fueron diseñadas con el propósito de inducir emociones específicas en el usuario, su contenido sensible permite anticipar diferentes reacciones emocionales en los participantes, particularmente durante las interacciones clave. Por esta razón, el propio AVAI ASI se considera también como el estímulo emocional principal del estudio.

6.3.3. Instrumentos

Con el propósito de registrar el estado emocional de los participantes durante la interacción con el AVAI ASI, se diseñó una adaptación específica del *Self-Assessment Manikin* para su uso en RVI, denominada Emoji-SAM RV. Asimismo, se desarrolló un instrumento post-intervención para capturar las palabras asociadas a las emociones experimentadas por los niños. Los instrumentos utilizados durante la prueba fueron los siguientes:

- **Demográfico básico (ad-hoc).** Recolecta información elemental sobre sexo y edad. Se presenta al participante como la primera escena de interacción con el AVAI ASI (Figura 6.1).

Figura 6.1

Interfaz del cuestionario demográfico en RV. Por defecto tiene señaladas las opciones de hombre y 8 años respectivamente.

- **Emoji-SAM RV (*Self-Assessment Manikin de emojis para Realidad Virtual*.)** Se realizó una revisión de instrumentos existentes orientados a entornos de realidad virtual y población infantil. Se identificaron como referencias principales el *Affective Slider* (Betella & Verschure, 2016) y el Emoti-SAM (Hayashi et al., 2016). De este último se tomaron como base sus representaciones gráficas enfocadas a niños. Con el objetivo de aumentar la familiaridad y comprensión por parte de los participantes, se rediseñó toda la iconografía utilizando emojis conocidos de plataformas digitales (Facebook, WhatsApp, entre otras). Se crearon versiones emoji consistentes para aquellos iconos del Emoti-SAM que no tenían una correspondencia exacta, manteniendo siempre elementos presentes en emojis existentes para garantizar coherencia visual. La interfaz del Emoji-SAM RV sigue una secuencia vertical (Figura 6.2):

- Como título, una pregunta sobre la dimensión explorada.
- Como subtítulo, una instrucción breve de interacción.
- Iconografía correspondiente (emojis).
- Control tipo slider.
- Botón de confirmación (“Ok”).

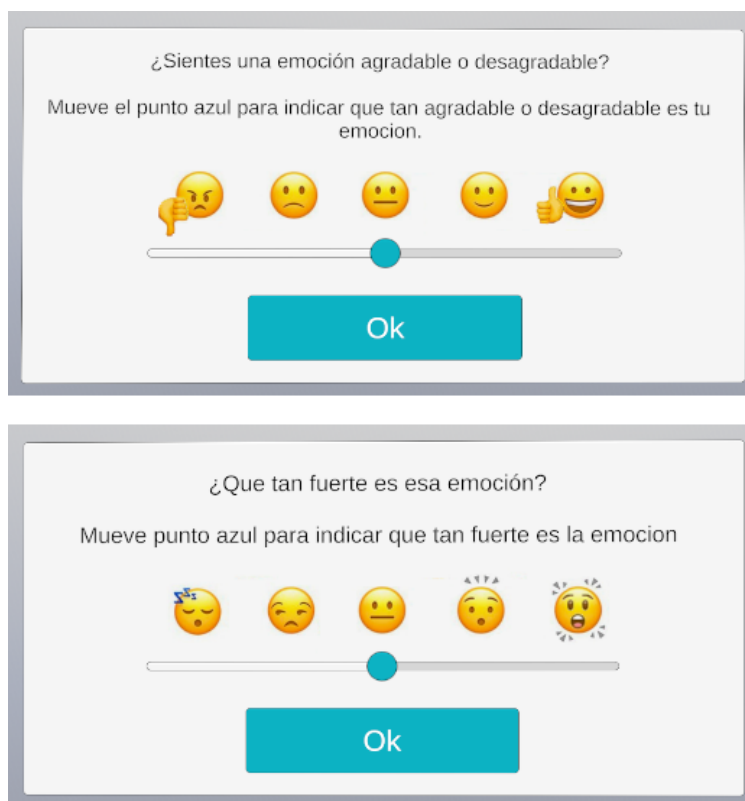


Figura 6.2

Interfaz del Emoji-SAM RV. Arriba, la interfaz correspondiente a la dimensión valencia. Abajo, la interfaz correspondiente a la dimensión activación.

Se utilizó una paleta de colores neutra con bajos contrastes para el fondo y elementos de la interfaz, empleando escalas de grises cálidos, mientras que los emojis mantienen colores brillantes para facilitar su identificación. Las palabras utilizadas en las preguntas e instrucciones fueron seleccionadas y simplificadas según observaciones de una prueba piloto, siguiendo tres criterios principales:

- a) Minimizar la cantidad de palabras para reducir el tiempo de lectura y procesamiento.
 - b) Utilizar vocabulario simple y concreto, evitando conceptos abstractos (por ejemplo, “emoción sentida” en lugar de “estado emocional experimentado”).
 - c) Preferir lenguaje coloquial y familiar en lugar de términos técnicos.
- **Registro Emocional Nominal (REN) Versión 2.2.** Es un instrumento diseñado para obtener un etiquetado nominal complementario y asociado a los valores de valencia y activación registrados mediante el Emoji-SAM RV. Su propósito es explorar la relación entre las respuestas dimensionales del SAM y las emociones específicas asociadas por los participantes a cada escena del programa. El instrumento consta de dos componentes principales:

- Elemento de respuestas: Contiene ocho frases asociativas a las escenas del AVAI ASI, cada una acompañada de 13 emoticonos diferentes como opciones de respuesta. Cada emoticono representa una emoción específica y va acompañado de un breve subtexto descriptivo. Para evitar sesgos de orden, los emoticonos se presentan en secuencia aleatoria en cada ítem del instrumento.
- Elemento de inducción emocional: Consiste en un conjunto de tarjetas visuales que muestran una imagen a color representativa de cada escena del AVAI ASI. Estas tarjetas guardan correspondencia directa con las frases del elemento de respuestas, facilitando el recuerdo y la asociación emocional por parte del niño.

Es importante destacar que la escena 8 del AVAI ASI contiene dos sub-escenas claramente diferenciadas. Aunque el REN incluye frases asociativas para ambas sub-escenas, para el análisis se consideró únicamente la frase correspondiente a la sub-escena “Memorama”, por ser la inmediatamente anterior a la aplicación del Emoji-SAM RV y la que se consideró de mayor influencia en los valores de valencia y activación registrados.

6.4. Procedimiento

La recolección de datos se realizó en colaboración con el Instituto Nacional de Salud Pública (INSP). Se contactaron dos escuelas primarias en el estado de Morelos: la escuela primaria urbana federal “Héroe de Nacozari” en Puente de Ixtla, y la escuela primaria matutina “Jaime Torres Bodet” en Jiutepec. Adicionalmente, se participó en la 3ª Jornada de Mujeres y Niñas en la Ciencia, Humanidades y Artes, organizada por la Universidad Autónoma del Estado de Morelos (UAEM), donde se reclutaron participantes del público general que cumplieran con el rango etario objetivo y hubiera consentimiento por parte de los padres. Una prueba piloto se llevó a cabo en la escuela “Héroe de Nacozari”, la cual permitió obtener observaciones iniciales y realizar mejoras en los instrumentos y en el protocolo de aplicación. Los datos recolectados en la UAEM sirvieron como primer dataset para pruebas preliminares de modelos de regresión. La prueba principal, sobre la cual se basa el análisis y los resultados presentados en este trabajo, se realizó en la escuela “Jaime Torres Bodet”. El procedimiento general de la prueba principal fue el siguiente:

1. Integración de grupos: Los participantes se organizaron en grupos de máximo 6 niños, cantidad determinada por el número de visores de realidad virtual disponibles y el espacio asignado.
2. Discurso introductorio y preentrenamiento: Se explicó el propósito de la actividad, el funcionamiento de los dispositivos de RVI, se ejemplificó el uso de los controles y las interacciones dentro del AVAI ASI.
3. Colocación de los dispositivos: Se ajustó el visor en la cabeza de cada participante, verificando la claridad de la imagen y la comodidad según lo expresado por el usuario. Se realizó un *reset* de ubicación central y se entregaron los controladores en ambas manos.

4. Aplicación del cuestionario demográfico: Los participantes respondieron el cuestionario básico como primera interacción con el sistema. Este paso permitió verificar también la comprensión del uso de los dispositivos y atender dudas sobre los mismos.
5. Interacción con el AVAI ASI: Cada participante interactuó durante aproximadamente 15 minutos con las ocho escenas del programa.
6. Autoevaluación emocional: Después de cada escena, los participantes completaron el Emoji-SAM RV para registrar sus niveles de valencia y activación.
7. Aplicación del REN: Al finalizar todas las escenas, se aplicó el Registro Emocional Nominal versión 2.2 de manera individual, pero recibiendo guía de manera grupal.
8. Cierre de la sesión: Se solicitó retroalimentación cualitativa voluntaria y se agradeció la participación. Este espacio ayuda además al participante a relajarse después de la interacción con el AVAI y «reconectar» con el mundo real.

La prueba principal se llevó a cabo de forma presencial y grupal durante los días 24, 25 y 26 de marzo de 2025. Cada sesión tuvo una duración aproximada de 25 a 30 minutos. Se brindó apoyo constante a los participantes para resolver dudas sobre las interacciones o problemas con los dispositivos. En total participaron 65 niños y niñas (36 niños y 29 niñas). Durante la depuración del *dataset* se excluyeron 7 participantes: dos por registros nulos o incompletos de información espacial, dos por no registrar respuestas del SAM, dos por respuestas inadecuadas del SAM y uno por estar fuera del rango etario objetivo. La muestra final quedó conformada por 58 participantes (31 niños y 27 niñas).

6.5. Consideraciones éticas

El presente estudio se diseñó y ejecutó conforme a los principios éticos establecidos en la Declaración de Helsinki. El protocolo recibió la aprobación del Comité de Ética en INSP, institución que además fungió como observadora del cumplimiento de los estándares y directrices éticas durante todo el proceso. Se obtuvieron dos niveles de autorización para la participación de los niños:

- Consentimiento informado por escrito de los padres o tutores legales, en el cual se explicaba de manera clara y accesible el propósito del estudio, el uso de realidad virtual, las actividades que realizarían los niños, la confidencialidad de los datos y su derecho a retirar la participación en cualquier momento sin consecuencias.
- Asentimiento verbal de cada niño, obtenido de forma previa a cada sesión y reafirmado antes de iniciar la interacción con el visor de realidad virtual. Se explicó a los participantes, con lenguaje adecuado a su edad, que podían detener la actividad en cualquier momento si así lo deseaban.

Se garantizó en todo momento la confidencialidad y anonimato de los datos. Los registros se codificaron con identificadores únicos, y no se almacenó ninguna información personal que permitiera la identificación de los participantes o pudiera asociarse a los datos espaciales o emocionales. Dado el tema sensible del programa, se contó con la presencia de personal capacitado del INSP durante todas las sesiones para brindar apoyo emocional inmediato en caso de que algún participante manifestara malestar, ansiedad o cualquier reacción adversa. No se registraron incidentes de este tipo durante las pruebas. Adicionalmente, se respetaron los horarios escolares y los recesos, evitando interferir con las actividades académicas regulares de los niños. Las sesiones se realizaron en espacios adecuados, seguros y supervisados en todo momento por investigadores del INSP y personal docente.

7

Diseño y desarrollo del modelo unimodal de detección emocional.

Este capítulo presenta el diseño, desarrollo e implementación del modelo unimodal de detección emocional propuesto como componente central del tutor afectivo AVAI ASI. Se describen la arquitectura general del sistema, el preprocesamiento de las señales, la extracción y selección de características, y los algoritmos utilizados. El enfoque se centra en utilizar exclusivamente información espacial nativa del visor de RVI, con el objetivo de superar las limitaciones de invasividad y multimodalidad identificadas en el estado del arte.

7.1. Arquitectura general del modelo

El modelo tiene como objetivo estimar los dos valores dimensionales del estado emocional del usuario: *valencia* y *activación*, registrados mediante el instrumento Emoji-SAM RV. Dado que la salida del modelo consiste en dos valores de punto flotante, el problema se abordó como una tarea de regresión. La Figura 7.1 presenta la arquitectura general del modelo unimodal de detección emocional.

El sistema toma como entrada la ubicación y orientación de la cabeza y las manos capturadas por los sensores nativos del visor de RVI durante las interacciones del participante con las escenas del AVAI ASI. El flujo del modelo se compone de las siguientes etapas principales:

1. Adquisición de señales: Captura en tiempo real de las posiciones y orientaciones de cabeza y manos.
2. Extracción de características: Cálculo de un conjunto de variables espacio-temporales que incluyen tanto características básicas de ubicación y orientación, como características derivadas que capturan relaciones espaciales entre articulaciones.

3. Selección de características: Aplicación de Análisis de Componentes Principales (PCA) para reducir dimensionalidad, reteniendo aquellos componentes que explican el 95 % de la varianza.
4. Balanceo de datos: Debido al desbalance natural en la distribución de valencia y activación, se aplicó la técnica *Synthetic Minority Over-sampling Technique* (SMOTE) después de discretizar ambas dimensiones en cinco categorías con intervalos de 0.2.
5. Modelado: Entrenamiento de algoritmos de regresión para estimar los valores continuos de valencia y activación.

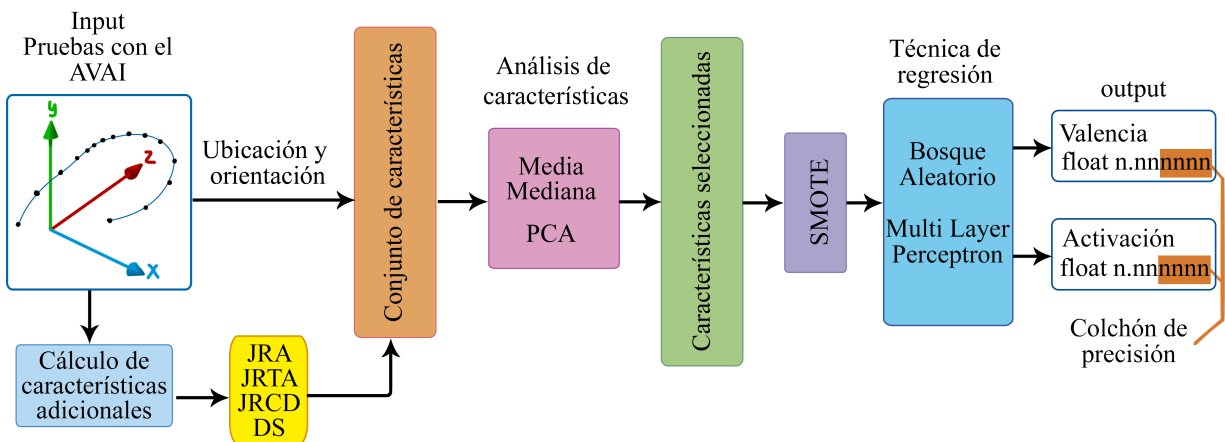


Figura 7.1
Arquitectura general para la detección de estados emocionales.

El análisis de componentes principales y las etapas posteriores de modelado se implementaron en entornos de *Colab Notebook*, utilizando principalmente las librerías *Scikit-learn* de *Python* y paquetes estadísticos en *R*.

7.2. Análisis de los datos

Los datos obtenidos del visor de realidad virtual inmersiva fueron sometidos a un preprocesamiento con el objetivo de garantizar su calidad, integridad y prepararlos para los entrenamientos de los modelos. El proceso incluyó las siguientes etapas:

- Limpieza e imputación: Eliminación de registros incompletos junto con imputación de valores faltantes cuando fue necesario.
- Cálculo de características derivadas: Generación de nuevas variables a partir de las trayectorias de cabeza y manos.
- Agregación estadística: Cálculo de vectores de medias y medianas por escena y participante.

- Integración de etiquetas: Concatenación de los valores de valencia y activación del Emoji-SAM RV, las respuestas del Registro Emocional Nominal (REN) y los datos demográficos en archivos consolidados para su posterior análisis.

Se generaron archivos de respaldo en formato Excel para facilitar la verificación manual y la inclusión de fórmulas adicionales. Cada registro fue asociado al instrumento REN correspondiente mediante un identificador aleatorio único y la hora de inicio de la sesión, asegurando un correcto emparejamiento entre los datos espaciales y las etiquetas emocionales. Este preprocesamiento aseguró la calidad y trazabilidad de los datos.

7.3. Cálculo de características

Durante la interacción con el AVAI ASI, se registraron de forma continua (50 hz) las variables básicas de posición y orientación de los tres dispositivos (HMD, mando derecho y mando izquierdo), junto con su marca de tiempo correspondiente. A partir de estas señales se calcularon en tiempo real características espaciales y relaciones geométricas que han sido reportadas en la literatura como relevantes para la detección de emociones mediante el análisis postural. Adicionalmente, se registró el foco de atención como proxy del seguimiento ocular. Este proceso generó un archivo de datos por cada escena y por cada participante. Posteriormente, a partir de estos registros se derivaron características adicionales, tales como el delta de simetría distal y versiones transformadas de las características originales bajo diferentes marcos de referencia. De cada característica se calcularon la media y la mediana, conformando así el vector final de características por escena y participante. A continuación se detalla el proceso de generación de cada una de ellas.

7.3.1. Posición

Coordenadas en X, Y, Z de cada dispositivo. Las variables de cada eje se registraron de manera separada. Los valores fueron proporcionados directamente por los sensores nativos de cada dispositivo.

7.3.2. Rotación Real

Rotación sobre los ejes X, Y, Z (extensión flexión o cabeceo, inclinación lateral o alabeo, guiñada o rotación) de cada dispositivo. Los valores se registraron tal y como los dispositivos y Unity los proporcionan. Sin embargo, una limitante interna registra valores de rotación comprendidos solo entre 0 a 359 grados, esto generó artefactos en los registros cuando se pasa de 0 a 359 en una fracción de segundo y viceversa.

7.3.3. Rotación

Rotación sobre los ejes X, Y, Z con un ajuste realizado en tiempo real durante el uso del AVAI ASI que permitió registrar grados por encima de los 360 y en dirección negativa (e.g. -10 grados) generando una gráfica continua sobre del comportamiento del usuario.

7.3.4. Velocidad

Velocidad lineal sobre los ejes X, Y, Z. Mide el cambio de posición por unidad de tiempo. Se calcularon para cada dispositivo (7.1).

$$V_i = \frac{Pi_2 - Pi_1}{\Delta t} \quad (7.1)$$

Donde:

- Pi_1 = Es la posición en el eje i en el momento t-1.
- Pi_2 = Es la posición en el eje i en el momento t.
- Δt = El tiempo transcurrido entre el momento t y t-1, Este valor es fijo en 0.02 segundos correspondiente a 50Hz.

7.3.5. Velocidad angular

Velocidad angular respecto a cada eje X, Y, Z. Mide la rapidez con la que un objeto gira o rota alrededor de un eje, definida como el cambio de desplazamiento angular por unidad de tiempo (7.2).

$$VA_i = \frac{\theta_{i_2} - \theta_{i_1}}{\Delta t} \quad (7.2)$$

Donde:

- θ_{i_1} = Es la orientación en el eje i en el momento t-1.
- θ_{i_2} = Es la orientación en el eje i en el momento t.
- Δt = El tiempo transcurrido entre el momento t y t-1, Este valor es fijo en 0.02 segundos correspondiente a 50Hz.

7.3.6. Aceleración

Es la tasa de variación de la velocidad de un objeto por unidad de tiempo. Se define como el cambio de velocidad dividido por el tiempo transcurrido. Se calculó y registró para cada eje en

cada dispositivo (7.3).

$$a_i = \frac{Vi_2 - Vi_1}{\Delta t} \quad (7.3)$$

Donde:

- Vi_1 = Es la aceleración en el eje i en el momento t-1.
- Vi_2 = Es la velocidad en el eje i en el momento t.
- Δt = Es el tiempo transcurrido entre el momento t y t-1. Este valor es fijo en 0.02 segundos correspondiente a 50Hz.

7.3.7. Aceleración angular

Mide el cambio de la velocidad angular por unidad de tiempo. Indica qué tan rápido un objeto en rotación gira más rápido o más lento (7.4).

$$\alpha Ai = \frac{VAi_2 - VAi_1}{\Delta t} \quad (7.4)$$

Donde:

- VAi_1 = Velocidad angular sobre el eje i en el momento t-1.
- VAi_2 = Velocidad angular sobre el eje i en el momento t.
- Δt = Es el tiempo transcurrido entre el momento t y t-1. Este valor es fijo en 0.02 segundos correspondiente a 50Hz.

7.3.8. Curvatura de la trayectoria

La curvatura (k) mide la tasa con la que un vector tangente gira, conforme la trayectoria se curva (Dibris et al., 2013). Una trayectoria que siga curvas cerradas (i.e. un círculo pequeño) girará abruptamente significando poca fluidez y una curvatura grande; por el contrario, una trayectoria en línea recta (i.e. técnicamente un círculo muy grande) tendrá una curvatura casi nula y más fluidez. En este trabajo la curvatura no fue computada sobre la trayectoria completa del dispositivo, sino en forma de proxy para capturar cambios abruptos en el movimiento a lo largo de cada eje y dirección de forma independiente:

$$ki = \begin{cases} \frac{Vi \cdot Ai}{Vi^3} & \text{si } Vi \neq 0 \text{ y } Ai \neq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (7.5)$$

Donde:

- Vi = La velocidad en el eje i.
- Ai = La aceleración en el eje i.

Cuando el dispositivo permanece estático, no existe una trayectoria y la curvatura es indefinida (NaN). El condicional mostrado en (7.5) se implementó para manejar estos casos. Este ajuste pudiera influenciar significativamente la forma en que la curvatura contribuye en la estimación de los valores de valencia y activación.

7.3.9. Vector HR

Distancia entre la ubicación del visor y el mando derecho (7.6).

$$HR = \sqrt{(X_H - X_R)^2 + (Y_H - Y_R)^2 + (Z_H - Z_R)^2} \quad (7.6)$$

Donde:

- X_H, Y_H, Z_H = son las coordenadas X, Y, Z, del visor.
- X_R, Y_R, Z_R = son las coordenadas X, Y, Z, del mando derecho.

7.3.10. Vector HL

Distancia entre la ubicación del visor y el mando izquierdo (7.7).

$$HL = \sqrt{(X_H - X_L)^2 + (Y_H - Y_L)^2 + (Z_H - Z_L)^2} \quad (7.7)$$

Donde:

- X_H, Y_H, Z_H = son las coordenadas X, Y, Z, del visor.
- X_L, Y_L, Z_L = son las coordenadas X, Y, Z, del mando izquierdo.

7.3.11. Vector RL

Distancia entre la ubicación del mando derecho y el mando izquierdo (7.8).

$$RL = \sqrt{(X_R - X_L)^2 + (Y_R - Y_L)^2 + (Z_R - Z_L)^2} \quad (7.8)$$

Donde:

- X_R, Y_R, Z_R = son las coordenadas X, Y, Z, del mando derecho.
- X_L, Y_L, Z_L = son las coordenadas X, Y, Z, del mando izquierdo.

7.3.12. Joint Relative Triangle Area (JRTA)

Representa la geometría de los movimientos de las articulaciones del cuerpo calculadas respecto a un punto (Bari & Gavrilova, 2019), en este caso la cabeza. El area del triángulo en el espacio 3D está determinados considerando las coordenadas de los tres dispositivos. El area del paralelogramo es el producto cruz de los vectores entre cabeza y manos, la mitad del área del paralelogramo es el área del triángulo. En (7.9) $\|\cdot\|$ representa la norma del vector.

$$\triangle ABC = \frac{\|\vec{RH} \times \vec{HL}\|}{2} \quad (7.9)$$

7.3.13. Joint Relative Angle (JRA)

Este es el ángulo formado entre los vectores de la cabeza a las manos. Esta característica es representativa de los movimientos angulares entre los tres dispositivos (Bhatia et al., 2022) usando la cabeza como punto central (7.10).

$$\theta_{RL} = \cos^{-1} \left(\frac{\vec{RH} \cdot \vec{HL}}{\|\vec{RH}\| \|\vec{HL}\|} \right) \quad (7.10)$$

Donde:

- $\vec{HR}$ = Es el vector desde la cabeza a la mano derecha.
- $\vec{HL}$ = Es el vector desde la cabeza a la mano izquierda.

De esta característica se calcularon variantes usando la mano derecha e izquierda como punto central, denominandose JRA_HL y JRA_HR.

7.3.14. Joint relative cosine dissimilarity (JRCD)

Es la distancia coseno entre dos puntos en el espacio vector (Bari & Gavrilova, 2019). La JRCD identifica de forma única los movimientos direccionales recurrentes (7.11).

$$\begin{aligned} \vec{R} \cdot \vec{L} &= \|\vec{R}\| \|\vec{L}\| \cos \theta \\ \cos \theta &= \frac{\vec{R} \cdot \vec{L}}{\|\vec{R}\| \|\vec{L}\|} \\ \delta_{\coseno}(R, L) &= 1 - \frac{\vec{R} \cdot \vec{L}}{\|\vec{R}\| \|\vec{L}\|} \end{aligned} \quad (7.11)$$

7.3.15. Foco de atención

Para determinar el foco de atención del usuario se implementó el mecanismo descrito por la figura (7.2). En el escenario 3D se incorporaron áreas de interés etiquetadas siguiendo el contenido didáctico: atención específica, atención general, evasión y distracción (AE, AG, E y D respectivamente). Cada área de interés se compone de un elemento invisible para el usuario pero que puede ser detectado por un vector que se origina en el punto c ubicado entre “los ojos” del usuario (la cámara en el escenario 3D) y termina a una distancia d que garantiza que el vector siempre entre en contacto con las áreas. Esta aproximación es una adaptación del trabajo de (Lin et al., 2023).

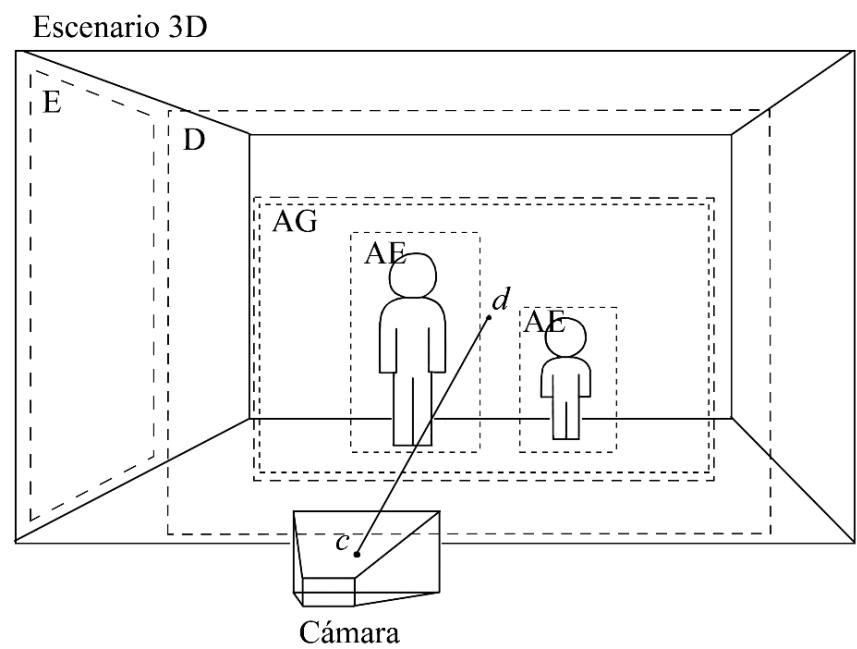


Figura 7.2
Mecanismo de detección del foco de atención del usuario. Se muestra el etiquetado de las áreas de acuerdo al contenido didáctico

Se registran valores enteros correspondiente al nivel de atención del objeto observado, la tabla muestra las equivalencias.

Tabla 7.1
Valores registrados de acuerdo al nivel de atención detectado.

Nivel de atención	Valor registrado
Atención específica	1
Atención general	2
Distracción	3
Evasión	4

7.3.16. Simetría distal simple

Es una razón de proporción entre las distancias euclidianas de la cabeza a la mano derecha y de la cabeza a la mano izquierda. Un condicional permite mantener el valor central en cero y la direccionalidad de la asimetría en lo positivo o negativo del valor según lo descrito en (7.12).

$$Sd = \begin{cases} \left(\frac{\overrightarrow{HL}}{\overrightarrow{HR}} \right) - 1 & \text{si } \overrightarrow{HR} \leq \overrightarrow{HL} \\ - \left(\frac{\overrightarrow{HR}}{\overrightarrow{HL}} \right) + 1 & \text{si } \overrightarrow{HL} < \overrightarrow{HR} \end{cases} \quad (7.12)$$

Donde:

- $\overrightarrow{HL}$ = la distancia entre la cabeza y la mano izquierda.
- $\overrightarrow{HR}$ = la distancia entre la cabeza y la mano derecha.

7.3.17. Simetría vertical

Es una razón de proporción entre las distancias euclidianas de la mano derecha al suelo y de la mano izquierda al suelo. Se calculó igual que en (7.12) Pero usando los valores de la posición Y de ambas manos.

7.3.18. Volumen

Es el espacio máximo ocupado por los tres dispositivos en cualquier momento t . Se calculó utilizando los valores máximos y mínimos de cada eje entre las posiciones de la cabeza y manos (7.13).

$$\begin{aligned} MaxX &= \max(XH, XR, XL) \\ MinX &= \min(XH, XR, XL) \\ MaxY &= \max(YH, YR, YL) \\ MinY &= \min(YH, YR, YL) \\ MaxZ &= \max(ZH, ZR, ZL) \\ MinZ &= \min(ZH, ZR, ZL) \end{aligned} \quad (7.13)$$

$$Volumen = (MaxX - MinX) * (MaxY - MinY) * (MaxZ - MinZ)$$

Donde:

- XH, XR, XL = son las coordenadas en el eje X de la cabeza, mano derecha y mano izquierda.
- YH, YR, YL = son las coordenadas en el eje Y de la cabeza, mano derecha y mano izquierda.
- ZH, ZR, ZL = son las coordenadas en el eje Z de la cabeza, mano derecha y mano izquierda.

7.3.19. Delta de la simetría distal

Describe los cambios en la simetría distal debido al movimiento de las manos.

$$\Delta Sd = Sd_2 - Sd_1 \quad (7.14)$$

Donde:

- Sd_1 = Es la simetría distal simple en el momento t-1.
- Sd_2 = Es la simetría distal simple en el momento t.

7.3.20. Variables con prefijo A

Las variables con el prefijo “A” (e.g. AH_x) no forman parte del registro automático durante la interacción con el AVAI ASI. Estas fueron calculadas manualmente con el objetivo de cambiar el marco de referencia espacial. Para ello, se utilizó la media de los primeros 49 registros (correspondientes al primer segundo de interacción) de la variable original como punto de anclaje (7.15).

$$AV_n = V_n - \left(\frac{\sum_1^{49} V}{49} \right) \quad (7.15)$$

Donde:

- AV_n = Es la variable ajustada de la variable V en el registro n al marco de referencia nuevo.
- V_n = Es el valor de la variable a ajustar en el mismo registro n .

De esta forma, AH_x representa la posición de la cabeza relativa al promedio inicial de la posición en el eje X, siguiendo un razonamiento análogo para el resto de las variables con prefijo “A”. Usando estas características se aplicaron las formulas (7.1), (7.3), (7.5), (7.6), (7.7), y (7.8) para generar nuevas variables bajo este marco de referencia también con el prefijo “A”. Así se conservan solo los desvíos desde un punto origen definido al inicio de cada escena y que puede variar entre escenas y/o sujetos de prueba.

7.3.21. Variables categorizadas

Los valores de Valencia y Activación se incluyeron también de forma categorizada siguiendo la tabla 7.2. Estos valores se usaron como ground truth para la generación de las matrices de confusión y el posterior cálculo de las métricas asociadas.

Tabla 7.2*Categorización de valencia y activación.*

Rango de valor	Categoría correspondiente
$Valor < 0,2$	1
$0,2 \leq Valor < 0,4$	2
$0,4 \leq Valor < 0,6$	3
$0,6 \leq Valor < 0,8$	4
$0,8 \leq Valor$	5

7.3.22. Estructura del vector de características

Una vez calculadas y definidas todas las características descritas anteriormente, se conformó el vector final de características para cada escena y participante. Este vector quedó integrado por un total de 124 características por escena y por participante. Se generaron dos conjuntos de datos independientes: uno basado en las medias y otro en las medianas de las características extraídas. Ambos vectores fueron utilizados posteriormente para explorar y modelar su relación con los niveles de valencia y activación reportados por los participantes a través del instrumento Emoji-SAM RV.

7.4. Selección de características

Una vez conformados los vectores de características, se realizó el proceso de selección de variables con el objetivo de reducir la dimensionalidad, eliminar redundancias y mejorar la eficiencia computacional de los modelos posteriores. Se aplicó el Análisis de Componentes Principales (PCA) de forma independiente para cada conjunto de datos. En el caso del vector basado en medias, los dos primeros componentes principales fueron suficientes para explicar el 96.7 % de la varianza acumulada. El primer componente estuvo dominado principalmente por la curvatura de la trayectoria en el eje X de la mano derecha (R_Kx). El segundo componente fue explicado por la curvatura de la trayectoria en el eje Y de la mano izquierda (L_Ky) y su versión con referencia relativa (AL_Ky). Dado que ambas variables aportaban información prácticamente equivalente, se consideró la posibilidad de omitir la versión relativa. El tercer componente destacó por la curvatura de la trayectoria de la cabeza en el eje Z. Finalmente, el conjunto se redujo a solo 3 características principales: R_Kx, L_Ky y AL_Ky.

Por su parte, el PCA aplicado al vector basado en medianas requirió un umbral más flexible. Se determinó retener componentes que explicaran al menos el 80 % de la varianza, ya que alcanzar el 95 % demandaría 23 componentes, mientras que con 14 componentes se alcanzó el 80 %. Las variables con mayor contribución incluyeron principalmente medidas de orientación de la cabeza y ambas manos, así como características de curvatura. Esta aproximación resultó en un conjunto de 25 características, representando una dimensionalidad considerablemente mayor que la obtenida con el vector de medias.

La diferencia en el número de características retenidas entre ambos enfoques (3 contra 25) evidencia la distinta distribución y redundancia presente en las medias y medianas de las señales espaciales. La tabla 7.3 detalla sobre las características seleccionadas para cada vector.

Tabla 7.3

Características seleccionadas para los vectores de medias y medianas.

Característica	Mediana	Media	Característica	Mediana	Media
H_RealRoll	*		R_Pitch	*	
H_Roll	*		R_RealYaw	*	
H_RealPitch	*		R_Yaw	*	
H_Pitch	*		H_ky	*	
H_RealYaw	*		AH_ky	*	
H_Yaw	*		H_kz	*	
L_RealYaw	*		L_ky	*	*
L_Yaw	*		AL_ky	*	*
L_RealPitch	*		L_kx	*	
L_Pitch	*		L_kz	*	
R_RealRoll	*		R_kz	*	
R_Roll	*		R_ky	*	
R_RealPitch	*		R_kx		*

7.5. Algoritmos de regresión

Para estimar de forma continua los niveles de valencia y activación emocional a partir de las características seleccionadas, se implementaron y compararon tres algoritmos de aprendizaje automático. La selección de estos algoritmos se basó en su capacidad para manejar datos estructurados con posible ruido, su interpretabilidad y su buen desempeño reportado en tareas similares de detección emocional (Leong et al., 2023; Younis et al., 2024).

- *Random Forest* (RF): Empleado tanto para tareas de regresión como de clasificación. Este algoritmo fue seleccionado por su robustez ante datos ruidosos y su capacidad para proporcionar medidas de importancia de características.
- *XGBoost*: Algoritmo de gradient boosting utilizado principalmente para regresión. Destaca por su alto rendimiento y eficiencia en el manejo de datos estructurados con posibles desbalances.
- *Redes neuronales*: Utilizadas con un modelo de perceptrón multicapa para regresión, con el fin de capturar posibles relaciones no lineales complejas entre las características espaciales y las dimensiones emocionales.

La salida esperada corresponde a dos valores continuos (valencia y activación) que el *ground truth* registra con 6 valores decimales. A pesar de que para lograr una buena estimación del estado emocional del participante se requieren solo dos valores decimales, se usaron los 4 decimales extra como un colchón de precisión, de tal forma que si los modelos fallan en este nivel, no tenga un impacto significativo. Así, el problema se abordó principalmente como una tarea de regresión. No obstante, también se discretizaron ambas dimensiones en cinco categorías cada una para la evaluación de los modelos mediante matriz de confusión.

7.6. Entrenamiento y validación del modelo

El entrenamiento de los modelos se realizó de forma independiente para las dimensiones de valencia y activación. Para mitigar el desbalance en la distribución de las etiquetas emocionales, se aplicó la técnica SMOTE utilizando como referencia la categorización de cada dimensión. Los datos se dividieron mediante validación cruzada de 10 folds. Los conjuntos de datos se separaron en 80 % para entrenamiento y 20 % para prueba. La optimización de hiperparámetros se llevó a cabo mediante una estrategia híbrida que combinó *GridSearchCV* y *Optuna*. Inicialmente, se utilizó *GridSearchCV* para explorar un espacio amplio de parámetros y analizar las tendencias de cada hiperparámetro, identificando tanto sus efectos positivos como negativos. Con base en estos resultados, se definió un espacio de búsqueda más acotado para *Optuna*, donde se realizaron múltiples iteraciones progresivas con refinamiento del espacio de solución tras cada ejecución para localizar de manera más precisa los valores óptimos. Como métrica principal de optimización se empleó el Error Cuadrático Medio (ECM). Un procedimiento análogo se siguió para el algoritmo *XGBoost*. En el caso de las redes neuronales multicapa (MLP), se evaluaron diferentes topologías uniformes y en forma de embudo, probando exhaustivamente las combinaciones de neuronas por capa dentro de los rangos definidos. Los resultados detallados del desempeño de los modelos, incluyendo las métricas de evaluación, análisis por edad, género y escena, así como las matrices de confusión, se presentan en el siguiente capítulo.

8

Resultados

Aquí se presentan los principales resultados obtenidos del entrenamiento y evaluación del modelo unimodal de detección emocional propuesto. Se incluyen el desempeño de los algoritmos implementados, las métricas de evaluación, así como algunos análisis sobre la muestra por edad, género y escena del AVAI ASI. Los hallazgos se presentan de manera estructurada para facilitar la comprensión del comportamiento del modelo en el contexto de prevención del abuso sexual infantil.

8.1. Desempeño de los modelos

8.1.1. Resultado del modelo Random Forest

Se implementaron modelos de Random Forest de regresión de forma independiente para las dimensiones de valencia y activación, utilizando los vectores de características basados en medias y medianas, sin aplicar el escalado de variables (*StandardScaler*). Los mejores hiperparámetros obtenidos para cada modelo se presentan en la Tabla 8.1.

Tabla 8.1

Mejores hiperparámetros.

Aproximación	Dimensión	Estimadores	Max Depth	Max feat.	MSL	MSS
Media	Valencia	330	37	2	1	2
	Activación	380	50	1	1	2
Mediana	Valencia	300	17	4	1	2
	Activación	103	25	9	1	2

Nota. feat: features; MSL: Min samples leaf; MSS: Min samples split

La aproximación basada en medianas generó modelos más simples, con menor profundidad y un menor número de estimadores, siendo esta diferencia más notable en la dimensión activación. Los modelos se evaluaron utilizando un conjunto de prueba correspondiente al 20 % de los datos después de aplicar SMOTE. Cada modelo resultante fue evaluado con las siguientes métricas: Error Cuadrático Medio (ECM), Error Absoluto Medio (EAM), Raíz del Error Cuadrático Medio (RECM) y Coeficiente de Determinación (R^2). Los resultados se muestran en la tabla 8.2.

Tabla 8.2

Resultados de los modelos para el 20 % de los datos de prueba.

Aproximación	Dimensión	ECM	EAM	RECM	R^2
Media	Valencia	0.039433	0.130334	0.198578	0.589
	Activación	0.052401	0.165562	0.228914	0.414
Mediana	Valencia	0.015695	0.08216	0.12528	0.833
	Activación	0.016884	0.089824	0.12993	0.811

Nota. ECM: Error Cuadrático Medio; EAM: Error Absoluto Medio; RECM: Raíz del Error Cuadrático Medio; R^2 : Coeficiente de Determinación

Adicionalmente, se realizó una prueba con una muestra aleatoria del 10 % de los datos previos al SMOTE. Esta aproximación mantiene la correspondencia real entre usuario, escena y etiquetas emocionales, simulando de manera más fiel una posible implementación final del sistema. Esta prueba generó resultados notablemente distintos a los obtenidos con la partición post-SMOTE (Tabla 8.3) mostrando desempeños más similares entre los modelos y entre las dimensiones de valencia y activación.

Tabla 8.3

Resultados de los modelos para el 10 % de los datos previos al SMOTE.

Aproximación	Dimensión	ECM	EAM	RECM	R^2
Media	Valencia	0.017241	0.096613	0.131305	0.689
	Activación	0.018882	0.102374	0.137412	0.586
Mediana	Valencia	0.017107	0.097072	0.130794	0.691
	Activación	0.012986	0.087536	0.113959	0.715

Nota. ECM: Error Cuadrático Medio; EAM: Error Absoluto Medio; RECM: Raíz del Error Cuadrático Medio; R^2 : Coeficiente de Determinación

Para evaluar el desempeño en términos categóricos, se discretizaron las dimensiones emocionales y se calcularon matrices de confusión. La Figura 8.1 presenta las matrices de confusión para la dimensión de valencia. Ambas aproximaciones (medias y medianas) mostraron un desempeño similarmente aceptable.

Las métricas derivadas de las matrices de confusión, junto con la precisión de cada modelo, se detallan en la tabla 8.4 En esta dimensión, el modelo basado en medianas mostró un mejor

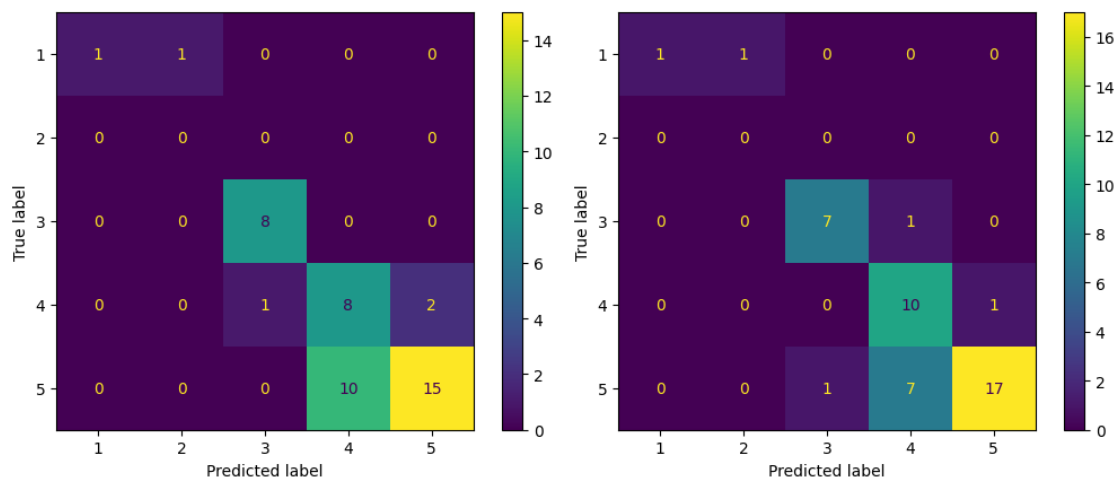


Figura 8.1

Matrices de confusión de los resultados categorizados de la dimensión valencia. A la izquierda la aproximación por medias y a la derecha la aproximación por medianas.

desempeño, particularmente en las clases 4 y 5.

Tabla 8.4

Métricas de las matrices de confusión de la dimensión valencia

Clase	Medianas				Medias			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
1	1	0.5	0.67	2	1	0.5	0.67	2
2	0	0	0	0	0	0	0	0
3	0.88	0.88	0.88	8	0.89	1	0.91	8
4	0.56	0.91	0.69	11	0.44	0.73	0.55	11
5	0.94	0.68	0.79	25	0.88	0.6	0.71	25
Macro avg	0.68	0.59	0.6	46	0.64	0.57	0.57	46
Weighted avg	0.84	0.76	0.78	46	0.78	0.7	0.71	46
Accuracy			0.76	46			0.7	46

La Figura 8.2 muestra las matrices de confusión correspondientes a la dimensión de activación. En esta dimensión, la superioridad del vector basado en medianas es aún más evidente.

La Tabla 8.5 resume las métricas de las matrices de confusión para activación. Al comparar la precisión global de ambos vectores, se observa que la activación fue estimada con mayor exactitud que la valencia en ambas aproximaciones.

Finalmente, aunque el vector basado en medianas sigue mostrando un mejor desempeño general, la diferencia entre ambos vectores es menor que la observada en las particiones post-SMOTE. Este hallazgo resalta la importancia de considerar el balance entre precisión y costo

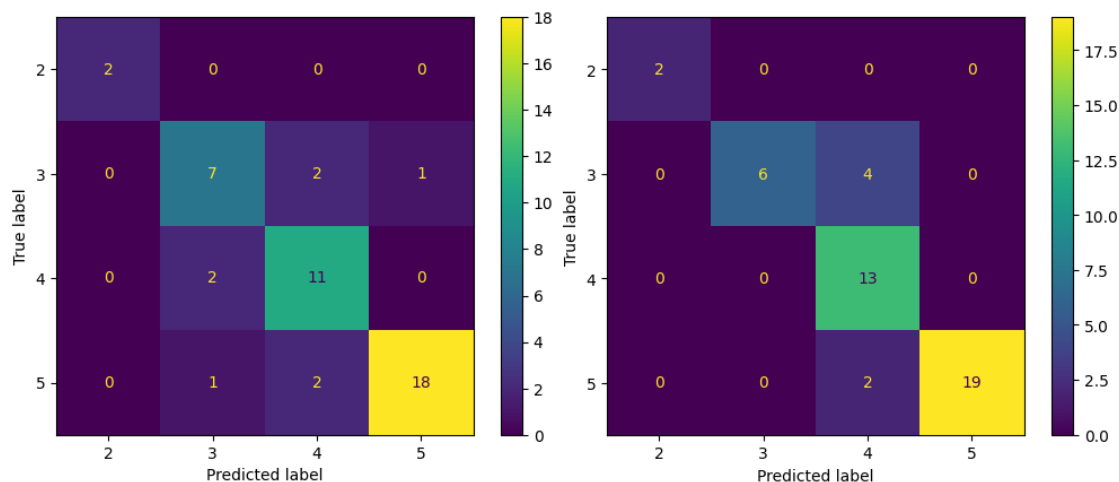


Figura 8.2

Matrices de confusión de los resultados categorizados de la dimensión activación. A la izquierda la aproximación por medias y a la derecha la aproximación por medianas.

Tabla 8.5

Métricas de las matrices de confusión de la dimensión activación

Clase	Medianas				Medias			
	Precision	Recall	F1-score	Support	Precision	Recall	F1-score	Support
2	1	1	1	2	1	1	1	2
3	1	0.6	0.75	10	0.7	0.7	0.7	10
4	0.68	1	0.81	13	0.73	0.85	0.79	13
5	1	0.9	0.95	21	0.95	0.86	0.9	21
Macro avg	0.92	0.88	0.88	46	0.85	0.85	0.85	46
Weighted avg	0.91	0.87	0.87	46	0.84	0.83	0.83	46
Accuracy			0.87	46			0.83	46

computacional, especialmente dada la gran diferencia en dimensionalidad entre los dos enfoques (3 características para medias contra 25 para medianas).

8.1.2. Resultado del modelo Redes Neuronales

Se implementaron redes neuronales en configuración de perceptrón multicapa (MLP) para tareas de regresión, con el fin de evaluar su desempeño en comparación con los modelos de *Random Forest*. El entrenamiento se realizó sobre los datos balanceados mediante SMOTE para la dimensión de valencia.

Se utilizaron los siguientes hiperparámetros: función de activación **ReLU**, optimizador **adam**, alpha de **0.001**, tamaño de batch automático, tasa de aprendizaje constante con valor inicial

de **0.001**, máximo de **1000** iteraciones, *shuffle* **activado**, parada temprana con umbral de 50 iteraciones sin cambio.

Se formularon y evaluaron exhaustivamente diferentes topologías de red, variando tanto la profundidad y como el número de neuronas por capa. La tabla 8.6 muestra los resultados de cada configuración para el vector basado en medias. En comparación con Random Forest, los modelos

Tabla 8.6

Resultados de los modelos MLP para la dimensión valencia usando el vector de medias

Anchura	Modelo	Mejor topología	ECM	EAM	R^2
Constante	1 [1-250]	1 [61]	0.09492	0.26542	0.012
Constante	10 [1-100]	10 [8]	0.09562	0.2674	0.005
Constante	5 [1-100]	5 [9]	0.09573	0.130794	0.691
Embudo	3 [1-20]	3 [15, 13, 2]	0.0889	0.2558	0.074
Variable	4 [1-15]	4 [9, 3, 8, 3]	0.071	NC	NC

Nota. Los modelos estan descritos por el numero de capas seguido del número de neuronas por capa entre corchetes, los rangos implican que se probó cada variación comprendida entre los extremos definidos. ECM: Error Cuadrático Medio; EAM: Error Absoluto Medio; R^2 : Coeficiente de Determinación; NC: No calculado

MLP utilizando el vector de medias mostraron un desempeño inferior. Sin embargo, al aplicar el mismo procedimiento al vector basado en medianas, se obtuvieron resultados notablemente superiores, alcanzando valores de Coeficiente de Determinación mayores a 0.9, junto con mejoras significativas en el ECM (tabla 8.7). Sin embargo, al evaluar estos modelos con un conjunto

Tabla 8.7

Resultados de los modelos MLP para la dimensión valencia usando el vector de medianas

Anchura	Modelo	Mejor topología	ECM	EAM	R^2
Constante	1 [1-250]	1 [201]	0.01996	0.08914	0.78
Constante	10 [1-100]	10 [86]*	0.00923	0.0502	0.901
Constante	5 [1-100]	5 [95]	0.00888	0.0549	0.905
Embudo	3 [1-201]	3 [201, 173, 24]	0.00826	0.0516	0.912

Nota. Los modelos estan descritos por el numero de capas seguido del número de neuronas por capa entre corchetes, los rangos implican que se probó cada variación comprendida entre los extremos definidos.

*Aunque se encontró que 82 y 78 neuronas dan mejor ECM, el EAM comienza a aumentar por lo que se descartaron. ECM: Error Cuadrático Medio; EAM: Error Absoluto Medio; R^2 : Coeficiente de Determinación; NC: No calculado

de datos recolectado en la UAEM, el desempeño se deterioró considerablemente. Tras estas observaciones se atribuyeron dos posibles explicaciones: un sobreajuste del modelo o la presencia de factores adicionales no registrados en el *dataset* actual ni capturados por los instrumentos implementados en su versión actual, los cuales estarían influyendo de manera relevante en los resultados.

8.1.3. Resultado del modelo XGBoost

Se exploró el algoritmo *XGBoost* (*Gradient Boosting*) para la dimensión activación y el vector de medias. Se probó con meta parámetros similares a los de *random forest*, especificando el número de estimadores, la profundidad máxima y la tasa de aprendizaje. La tabla 8.8 presenta los hiperparámetros y los resultados obtenidos de las pruebas realizadas.

Tabla 8.8

Hiperparámetros y resultados de las pruebas con XGBoost

Max depth	Subsample	Colsample by tree	Min child weight	Learning rate	NE	ECM	EAM	R^2
25	SE	SE	SE	1	257	0.06028	0.14557	0.381
75	0.9918	0.741	23	2	355	0.07020	0.18916	0.279

Nota. NE: Numero de estimadores; ECM: Error Cuadrático Medio; EAM: Error Absoluto Medio; R^2 : Coeficiente de Determinación; SE: Sin especificar

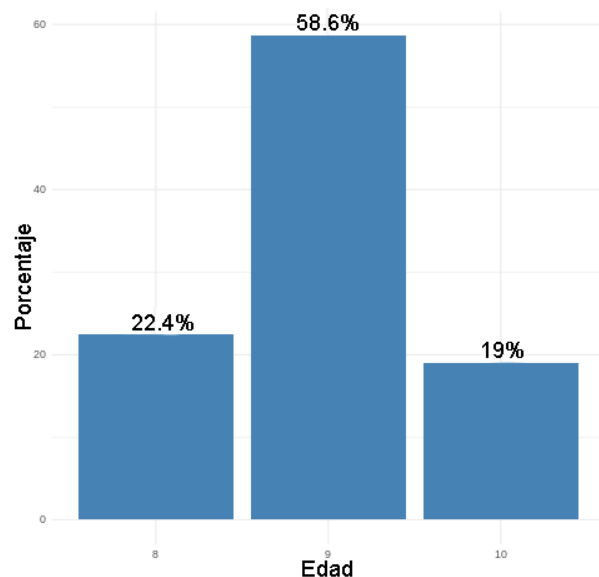
Los resultados mostraron una consistencia limitada. Aunque el EAM fue inferior al obtenido con *Random Forest*, el ECM resultó más alto. Además, se observó una notable inestabilidad en los resultados: a pesar de fijar la semilla aleatoria (*random state*), modelos entrenados con los mismos hiperparámetros generaron valores diferentes en las métricas de ECM, EAM y R^2 . Esta falta de reproducibilidad sugirió una alta sensibilidad del modelo a la inicialización y partición de los datos. Debido a estas inconsistencias, se decidió abortar el desarrollo adicional con esta estrategia.

8.2. Análisis exploratorio

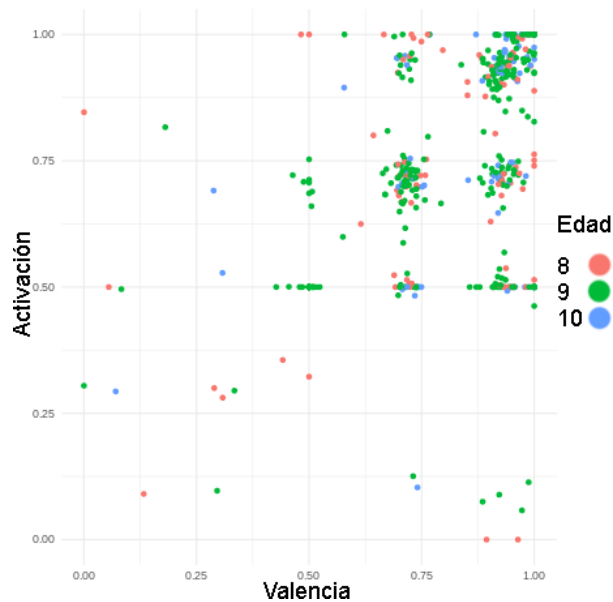
8.2.1. Análisis por edad

La muestra final del estudio estuvo compuesta por niños y niñas con edades comprendidas entre 8 y 10 años, sin establecer una proporción específica entre los grupos etarios. La prueba de Shapiro-Wilk indicó que la distribución de edades no sigue una distribución normal ($W = 0,78807, p < 0,001$), principalmente debido a la mayor prevalencia de participantes de 9 años y al rango etario reducido del estudio. La distribución de edades se presenta en la Figura 8.3.

Para evaluar si la composición etaria se ajustaba a una distribución teórica de 20 % (8 años), 60 % (9 años) y 20 % (10 años), se aplicó una prueba de χ^2 de bondad de ajuste. El resultado fue no significativo ($\chi^2(2) = 1,7471, p = ,4175$), lo que indica que la distribución observada se ajusta adecuadamente al modelo propuesto. Posteriormente, se analizaron las dimensiones de valencia y activación según el grupo de edad. La prueba de Kruskal-Wallis reveló diferencias

**Figura 8.3**

Distribución porcentual por edad. Sigue un modelo de 20-60-20.

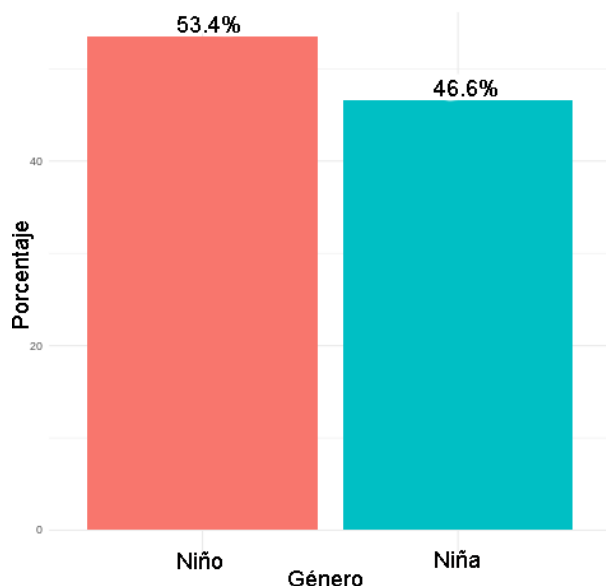
**Figura 8.4**

Distribución de los valores de valencia activación con respecto a la edad

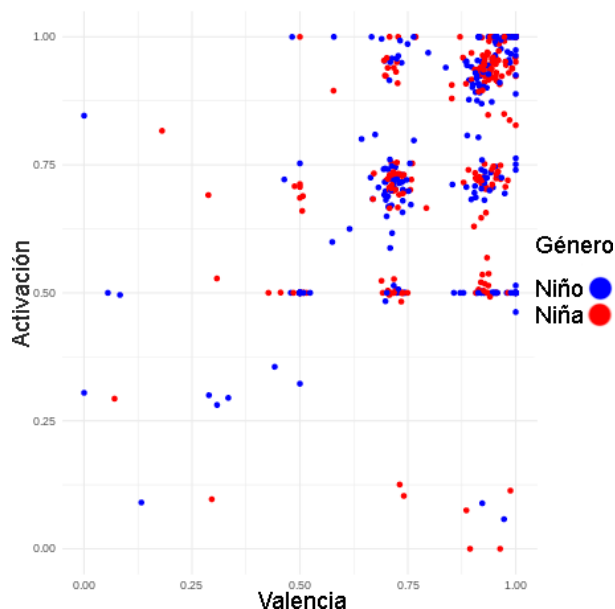
estadísticamente significativas tanto en valencia ($\chi^2(2) = 14,07$, $p < ,001$) como en activación ($\chi^2(2) = 11,21$, $p = ,0037$). En ambos casos se realizó un análisis post hoc mediante la prueba de Dunn con corrección de Bonferroni. Para la dimensión de valencia se encontraron diferencias significativas entre los grupos de 10 y 8 años ($Z = 2,40$, $p_{ajustado} = ,024$) y entre los grupos de 10 y 9 años ($Z = 3,75$, $p_{ajustado} < ,001$). De manera similar, en la dimensión de activación se observaron diferencias significativas entre los niños de 10 y 8 años ($Z = 2,21$, $p_{ajustado} = ,040$) y entre los de 10 y 9 años ($Z = 3,35$, $p_{ajustado} = ,001$). Aunque visualmente no destacan diferencias marcadas en la Figura 8.4 debido a la aglomeración de los datos, el análisis inferencial confirma la existencia de patrones etarios diferenciados, particularmente en los niños de 10 años.

8.2.2. Análisis por género

La distribución de los participantes por género se presenta en la Figura 8.5. Para evaluar si esta distribución se ajustaba a una proporción esperada de 50/50, se aplicó una prueba de χ^2 de bondad de ajuste, obteniendo un resultado no significativo ($\chi^2(1) = 2,207$, $p = ,1374$). Por lo tanto, no existe evidencia estadística para rechazar la hipótesis de que la muestra se encuentra balanceada por género. Posteriormente, se analizaron las dimensiones de valencia y activación según el género de los participantes. La Figura 8.6 muestra la distribución de los valores reportados. De forma visual, se observa una mayor concentración de registros de niños en el cuadrante de baja valencia y baja activación, mientras que las niñas tienden a concentrarse en el cuadrante de alta valencia y alta activación. Dado que las distribuciones de valencia y activación no siguieron una distribución normal (prueba de Shapiro-Wilk, $p < 0.001$ para ambos géneros en las dos dimensiones), se utilizó la prueba no paramétrica de Mann-Whitney U para

**Figura 8.5**

Distribución porcentual por género. De acuerdo a la prueba de bondad de ajuste, puede considerarse una distribución 50/50

**Figura 8.6**

Distribución de los valores de valencia activación con respecto a género

comparar las medianas entre niños y niñas. Los resultados indicaron que no existen diferencias estadísticamente significativas ni en valencia ($p = ,820$) ni en activación ($p = ,3678$) entre ambos géneros. Esto significa que ambos registraron valores similares en ambas dimensiones.

8.2.3. Análisis por escena

Se realizó un análisis de las dimensiones de valencia y activación con respecto a las ocho escenas del AVAI ASI, bajo la hipótesis de que cada escena induciría estados emocionales diferenciados, los cuales podrían ser identificados con mayor claridad mediante el REN. Sin embargo, como se observó en los análisis anteriores, no se aprecia una agrupación clara ni patrones altamente discriminantes entre escenas (Figura 8.7).

La prueba de Kruskal-Wallis reveló diferencias estadísticamente significativas en la dimensión de valencia entre las escenas ($\chi^2(7) = 17,551$, $p = 0,0142$). El análisis post hoc de Dunn con corrección Bonferroni mostró que la escena 6 difiere significativamente de la escena 8 ($Z = -3,74$, $p_{ajustado} = ,003$), siendo esta última la que reportó valores más altos de valencia. De manera similar, para la dimensión de activación se obtuvo un resultado significativo en la prueba de Kruskal-Wallis ($\chi^2(7) = 14,312$, $df = 7$, $p = ,046$). El análisis post-hoc indicó nuevamente una diferencia estadísticamente significativa entre la escena 6 y la escena 8 ($Z = -3,584$, $p_{ajustado} = ,0047$). En general, los niveles de valencia y activación reportados fueron relativamente similares entre la mayoría de las escenas, destacando únicamente la diferencia entre las escenas 6 y 8.

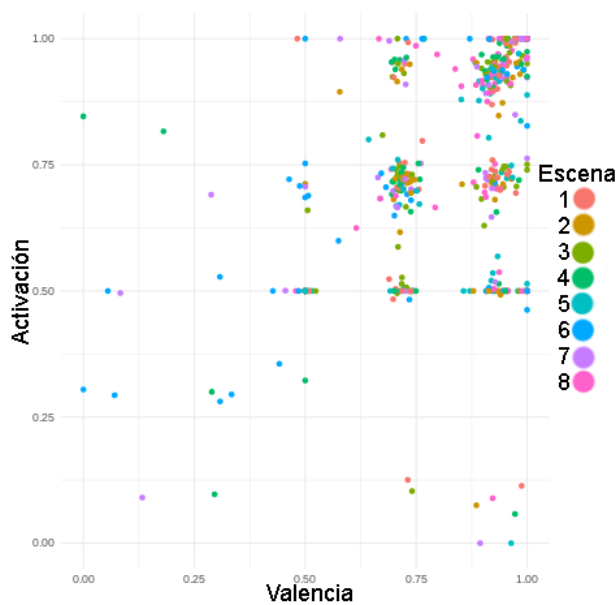


Figura 8.7
Distribución de los valores de valencia activación con respecto a las escenas

8.2.4. Análisis del REN por escena

Se realizó un análisis de las respuestas nominales del REN con el objetivo de identificar patrones emocionales específicos por escena y evaluar su consistencia con los resultados dimensionales del Emoji-SAM RV. Se esperaba que cada escena generara perfiles emocionales diferenciados y claramente distinguibles. La prueba de Kruskal-Wallis reveló diferencias altamente significativas en la distribución de valencia nominal entre las ocho escenas ($\chi^2(7) = 118,48$, $p < ,001$). El análisis post hoc de Dunn con corrección Bonferroni mostró que la escena 7 se diferencia significativamente de todas las demás escenas, registrando consistentemente emociones de menor valencia, principalmente “apenado”, “molesto”, “enojado” y “triste”, como se observa en la Figura 8.8.

En el AVAI ASI, la escena 7 destaca por un perfil emocional más negativo que puede explicarse por su contenido temático: mientras que la mayoría de las escenas mantienen un tono predominantemente preventivo e informativo, la escena 7 involucra elementos de mayor intensidad emocional, pues presenta una situación más directa, confrontativa y cercana a la posibilidad real de abuso. Estos hallazgos son relevantes en el contexto del tutor afectivo, ya que permiten identificar escenas con mayor carga emocional negativa, reforzando la necesidad de implementar regulación afectiva. Mas allá de este hallazgo, en la Figura 8.9, no se aprecia una agrupación clara o altamente discriminante entre la mayoría de las escenas. La evidencia señala que los usuarios asignan la misma etiqueta nominal a diferentes binomios de valencia y activación.

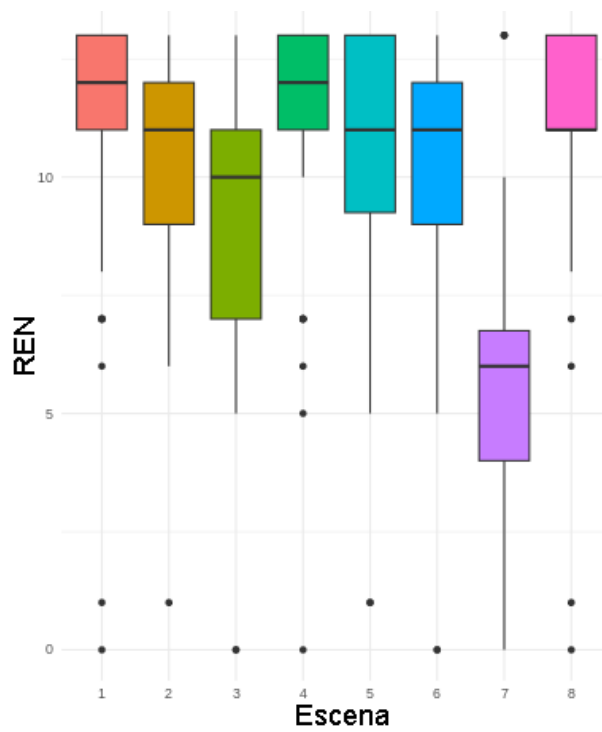


Figura 8.8
Boxplot de las respuestas del REN por escena. La escena 7 registra emociones con menor nivel de valencia.



Figura 8.9
Distribución de los valores de valencia activación con respecto al REN

8.2.5. Resumen de resultados

Los resultados obtenidos confirman la hipótesis planteada, demostrando que existe una relación significativa entre las características espacio-temporales del movimiento natural de cabeza y manos capturadas por los sensores nativos de un visor de RVI, y el estado emocional autopercebido de niños y niñas de 8 a 10 años, haciendo posible estimar su estado emocional a partir de esta información. Los modelos de regresión implementados, particularmente Random Forest, mostraron los mejores resultados generales, logrando niveles aceptables de Error Cuadrático Medio (ECM) y Error Absoluto Medio (EAM). Tanto los vectores basados en medias como en medianas mostraron capacidad predictiva para estimar los niveles de valencia y activación, aunque con diferencias notables en su desempeño. El vector basado en medianas presentó un desempeño superior al de medias, especialmente en la dimensión de activación, aunque el vector basado en medias requirió considerablemente menos características (3 frente a 25), lo que subraya la importancia del balance entre dimensionalidad, precisión y costo computacional. El instrumento Emoji-SAM RV propuesto se desempeñó adecuadamente como herramienta de autoevaluación emocional en realidad virtual, mostrando buena comprensión por parte de los participantes. Los análisis exploratorios revelaron que no existen diferencias estadísticamente significativas en las respuestas emocionales según género. Sin embargo, la edad sí resultó un factor relevante: los niños de 10 años reportaron valores de valencia y activación significativamente diferentes a los de 8 y 9 años. Respecto al análisis por escena, la mayoría generaron respuestas emocionales similares. Sin embargo, la escena 7 se distinguió claramente por inducir emociones de menor valencia (apenado, molesto, enojado y triste), lo que sugiere una mayor carga emocional negativa asociada a su contenido. El instrumento REN demostró potencial como herramienta complementaria al SAM, permitiendo explorar la relación entre etiquetas dimensionales y nominales en población infantil. Su diseño adaptable sugiere que puede ser utilizado más allá del contexto de realidad virtual. Finalmente, aunque se observaron áreas de agrupación en el espacio emocional, no se encontró una correspondencia clara entre las etiquetas dimensionales y las nominales, lo que refleja la complejidad de la expresión emocional en población infantil. En general, los hallazgos confirman el potencial del enfoque unimodal propuesto, aunque resaltan la importancia de considerar el tipo de agregación estadística (media vs. mediana), la dimensionalidad de las características y las diferencias individuales relacionadas con la edad.

Discusión

Los resultados de este estudio confirmaron la viabilidad técnica de estimar el estado emocional de niños y niñas mediante un modelo unimodal basado exclusivamente en información espacial nativa de un visor de realidad virtual inmersiva. Este enfoque representó una contribución relevante al campo de la computación afectiva, al evitar los problemas de multimodalidad invasiva y etiquetado a priori identificados en el estado del arte. Hasta donde se conoce, no existían trabajos previos que abordaran la detección emocional utilizando únicamente información de cabeza y manos en entornos de RVI con población infantil.

Entre las ventajas más destacadas de la propuesta se encuentran su portabilidad, no invasividad y no intrusividad, al eliminar la necesidad de sensores fisiológicos o de reconocimiento facial. Esto favorece una mayor confianza en los participantes y, en consecuencia, una mayor naturalidad en sus movimientos. Asimismo, el modelo aprovecha la ergonomía del visor y presenta un bajo costo computacional, facilitando su escalabilidad en entornos educativos.

Aunque en la literatura se afirma frecuentemente que los enfoques multimodales ofrecen mayor precisión, estos suelen implicar un costo computacional significativamente más alto, mayor complejidad de implementación y la necesidad de sensores y software adicionales. En contraste, el modelo unimodal propuesto utiliza exclusivamente las señales nativas del visor de RVI y sus controles. En términos económicos, el uso de un visor de gama media como el Meta Quest 2 (cuyo costo aproximado al momento del estudio rondaba los 300–400 USD) evita la adquisición de sensores fisiológicos adicionales, como sistemas de electroencefalograma (EEG) o de respuesta galvánica de la piel, cuyo precio suele ser considerablemente superior y requiere equipo de cómputo complementario. Esto reduce no solo el costo de adquisición, sino también la complejidad de implementación, el mantenimiento y la invasividad del sistema, facilitando su escalabilidad en entornos educativos con recursos limitados y multiusuario.

Los resultados mostraron que el vector basado en medias redujo drásticamente la dimensionalidad (solo 3 características), pero su poder predictivo fue inferior al vector basado en medianas durante el entrenamiento (ECM: .039 vs .015 en valencia; .052 vs .016 en Activación). Sin embargo, en escenarios más cercanos a la realidad, el desempeño de ambos enfoques fue similar (*accuracy*: 76 % vs 70 % en valencia, 87 % vs 83 % en activación), valores que contrastan favorablemente con la precisión humana promedio en la detección de emociones (que ronda el 62 %), y con otros

resultados reportados por otros enfoques unimodales como la de Hofmann et al. (2021) que obtuvo una precisión del 60.76 % al 60.83 % con electroencefalograma; Sapiński et al. (2019) reportó 64 % con *Random Forest* en 4 clases; Pfeuffer et al. (2019) analizó los movimientos de la cabeza y obtuvo un 63 %; Si bien Kaza et al. (2016) reportó 93 %, su trabajo aplica el etiquetado apriori abordando finalmente el problema como detección de posturas mas que de emociones.

Los resultados obtenidos permiten realizar varias interpretaciones sobre el comportamiento de los modelos y las características utilizadas. En primer lugar, el mejor desempeño del vector basado en medianas puede explicarse por la mayor robustez de la mediana ante valores atípicos y ruido. Las trayectorias de cabeza y manos en RVI suelen presentar variaciones abruptas; la mediana reduce el impacto de estos valores capturando de forma más estable la tendencia central del comportamiento. Aunque el vector de medias logró una reducción extrema de la dimensionalidad, esta simplificación también eliminó información potencialmente útil, que se tradujo en una menor capacidad predictiva durante los entrenamientos. Respecto a los algoritmos, *Random Forest* mostró consistentemente el mejor desempeño. Esto puede atribuirse a su naturaleza de ensamble, que le permite manejar de forma efectiva datos estructurados con ruido y relaciones no lineales moderadas, además de ofrecer una buena capacidad de generalización incluso con tamaños de muestra moderados. En contraste, *XGBoost* presentó resultados inconsistentes e inestables. Una posible explicación es su mayor sensibilidad a la calidad y al balance de los datos, así como a la configuración de hiperparámetros. Dado el tamaño moderado del conjunto de datos y la presencia de posible ruido principalmente en los valores autoevaluados, es probable que haya tendido al sobreajuste o haya mostrado una alta variabilidad en su desempeño a lo largo de las distintas particiones de la validación cruzada, a diferencia de *Random Forest*, que resultó más robusto en estas condiciones. Asimismo, el hecho de que ciertas familias de variables (e.g., orientaciones o características derivadas) no hayan resultado tan discriminativas sugiere que no toda la información espacial disponible aporta de manera equivalente a la estimación emocional. Es probable que las variables relacionadas con la curvatura de trayectoria y ciertas orientaciones de cabeza y manos capturen mejor los patrones asociados a cambios en valencia y activación, además que su cálculo las vuelve independientes del marco espacial de referencia, mientras que otras características presentan mayor redundancia o menor relación con el estado emocional autopercebido. Estos hallazgos refuerzan la importancia de considerar no solo la cantidad de características, sino su naturaleza estadística y las particularidades del conjunto de datos al diseñar modelos de detección emocional unimodales en entornos inmersivos. Aun así, se identificó una limitación importante (Pfeuffer et al., 2019): en actividades repetitivas o con herramientas especializadas, los usuarios tienden a optimizar sus movimientos con la experiencia, lo que puede causar que los patrones se vuelvan similares, limitando la robustez del modelo a largo plazo. Este aspecto deberá evaluarse en estudios longitudinales futuros.

La autoevaluación con Emoji-SAM RV permitió capturar diferencias individuales bajo un mismo estímulo, producto de las experiencias personales y condiciones específicas de cada niño en la expresión emocional. Los estados emocionales negativos fueron de particular interés, dada la naturaleza sensible del contenido didáctico y el objetivo de lograr regulación emocional hacia estados positivos. No obstante, los resultados mostraron un predominio de estados afectivos positivos (emoción y alegría), sin que esto sea menos interesante.

Una ventaja relevante del modelo circunplejo de Russell es la flexibilidad que ofrece para categorizar los resultados bajo diferentes granularidades sin necesidad de reformular las pruebas. Esto resulta especialmente útil en población infantil, donde la conceptualización emocional varía

significativamente con la edad (Posner et al., 2005). Los niños tienden a agrupar emociones en categorías más amplias (positivas y negativas), mientras que su capacidad para discriminar rangos intermedios se desarrolla progresivamente. Este fenómeno se reflejó en las respuestas del Emoji-SAM RV, que mostraron agrupamientos en áreas que no tienen una definición explícita dentro del modelo circunplejo de Russell.

Originalmente se buscó omitir variables contextuales explícitas, sin embargo durante la investigación se determinó que el diseño del AVAI ASI, sus mecánicas de interacción y el contenido de las escenas constituyeron un contexto implícito que influyó en los movimientos de los usuarios. Este hallazgo otorgó al contexto un aspecto ineludible y de suma importancia en la detección emocional. Estudios como el de Can et al. (2023), indican que los modelos construidos con emociones elicitadas artificialmente tienen limitada generalización. Sin embargo, la RVI reduce esta brecha al ser capaz de elicitar respuestas emocionales más cercanas a las del mundo real.

En conjunto, este trabajo representó un avance significativo en la integración de computación afectiva y realidad virtual en el campo de la prevención del abuso sexual infantil. No obstante, también evidenció los retos técnicos, metodológicos y contextuales que aún deben abordarse para lograr una implementación robusta, generalizable y efectiva en entornos reales. Aunque la metodología es generalizable, las variables seleccionadas están sujetas a las condiciones específicas del experimento, incluyendo las características propias de la población estudiada, tales como aspectos socioculturales no capturados en el *data set* ni por los instrumentos implementados. No obstante, estas condiciones pueden adaptarse a otros grupos y entornos virtuales mediante el reentrenamiento de los modelos o la exploración de características más relevantes para el nuevo entorno.

Conclusiones y trabajos futuros

Esta investigación logró desarrollar un modelo unimodal de detección emocional basado exclusivamente en información espacial nativa de un visor de realidad virtual inmersiva, utilizando únicamente las trayectorias de cabeza y manos. Los resultados demostraron que es posible estimar de forma razonable los niveles de valencia y activación en niños y niñas de 8 a 10 años durante un programa de prevención del abuso sexual infantil, sin necesidad de sensores adicionales ni reconocimiento facial.

El enfoque propuesto representó un avance respecto al estado del arte al abordar simultáneamente dos limitaciones principales: la multimodalidad invasiva y el etiquetado a priori. El modelo de *Random Forest* basado en medianas mostró el mejor desempeño (76 % de precisión en valencia y 87 % de precisión en activación), mientras que el basado en medias destacó por su extrema reducción de dimensionalidad (3 características únicamente de las manos), aunque con menor precisión. Estos hallazgos confirmaron la viabilidad técnica del enfoque unimodal en contextos de entornos virtuales inmersivos y con población infantil.

Adicionalmente, se diseñó e implementó el Emoji-SAM RV, una adaptación del *Self-Assessment Manikin* que demostró buena comprensión y usabilidad por parte de los niños. El Registro Emocional Nominal (REN) complementó el análisis, permitiendo explorar la relación entre etiquetas dimensionales y nominales. Sin embargo, los resultados no mostraron la correlación esperada. Esta discrepancia entre el REN y el SAM pudiera deberse a dos factores:

1. La aplicación post intervención del REN. El usuario evaluó en retrospectiva, analizando con más detenimiento lo sucedido y desarrollando una nueva emoción más enfocada al contenido que a la experiencia en RVI.
2. La formulación de la pregunta para cada instrumento. El Emoji-SAM RV no especificó si la reacción emocional debía asociarse a algo (i.e. un causal al cual orientar la emoción) sino lo preguntó de forma abierta y general, lo que podría interpretarse como más dirigido a la experiencia en RVI que al contenido (contextualización). El REN, por otro lado, preguntó directamente sobre lo que se observó y presencié en cada escena, claramente enfocando las respuestas al contenido mostrado, y hasta cierto punto, independiente del medio en el que se observó.

Desde el punto de vista aplicado, el AVAI ASI logró inducir predominantemente estados emocionales positivos, lo cual resulta favorable para su objetivo preventivo. Sin embargo, también se identificaron escenas con mayor carga emocional negativa (particularmente la escena 7), lo que resaltó la importancia de incorporar regulación afectiva.

10.1. Limitaciones

Entre las principales limitaciones se encuentran:

- El tamaño moderado de la muestra final (58 participantes). Además, al tratarse de un muestreo no probabilístico por conveniencia, la diversidad sociocultural de la muestra es limitada. La mayoría de los participantes provenían de una sola escuela primaria ubicada en el estado de Morelos, lo que restringe la representatividad de los resultados respecto a otras regiones, contextos educativos o poblaciones con diferentes características socioculturales. Lo observado durante las pruebas de los modelos apuntaron a la posible existencia de elementos de dichas áreas que influyeron en la manera de registrar los valores de valencia y activación bajo un mismo estímulo emocional.
- El tamaño de la muestra también limitó una evaluación más robusta del Registro Emocional Nominal (REN). Una muestra más amplia y diversa permitiría analizar con mayor profundidad la relación entre las etiquetas dimensionales y nominales.
- La ausencia de seguimiento longitudinal. La exposición repetida a la misma experiencia y estímulos, podría acentuar o atenuar algunas respuestas emocionales.
- La dependencia de autoevaluación subjetiva. Aunque el instrumento SAM y su aproximación pictórica facilitan su uso intuitivo y es usado ampliamente en el grupo infantil, la conceptualización y consistencia para reportar la activación puede significar un desafío para algunos niños, potencialmente introduciendo ruido y variabilidad en las etiquetas.
- Los resultados quedaron condicionados al contexto específico del AVAI ASI, por lo que su generalización a otros entornos de realidad virtual debe validarse en futuros estudios. Estos efectos permanecen como parte intrínseca de las condiciones del experimento y no debilitan en lo general los patrones observados.
- La validación de los modelos se realizó mediante validación cruzada de 10 *folds* utilizando la función *cross_val_score* de *scikit-learn*, junto con una partición inicial mediante *train_test_split*. Es importante señalar que esta división se realizó a nivel de registro y no a nivel de participante. Por lo tanto, existe la posibilidad de que registros pertenecientes a un mismo niño o niña hayan aparecido tanto en el conjunto de entrenamiento como en el de prueba, aunque dichos registros pertenezcan a escenas (estímulos) diferentes. Esta situación representa una limitación del estudio, ya que podría conducir a una sobreestimación del desempeño del modelo debido a la correlación intra-sujeto. En trabajos futuros se recomienda implementar estrategias de validación a nivel de participante (por ejemplo *Leave-One-Subject-Out*) para obtener una evaluación más realista de la generalización del modelo.

- El grado de confiabilidad del REN. Si bien el instrumento arrojó resultados prometedores aun es un instrumento en proceso de mejora.

10.2. Aportaciones

10.2.1. AVAI ASI extendido

Se desarrolló e implementó mecanismos de recolección y cálculo de características en tiempo real durante la interacción con las escenas del AVAI ASI. Además se diseñó, desarrolló e implementó el instrumento de autoevaluación emocional Emoji SAM RV. En conjunto estas características extienden las capacidades del AVAI ASI.

10.2.2. Modelo unimodal de detección emocional en Realidad Virtual Inmersiva

Se diseñó y validó un modelo unimodal de detección emocional que utiliza exclusivamente información espacial nativa del visor de realidad virtual. Los resultados basados en *Random Forest* de regresión, demostraron su viabilidad para estimar valencia y activación en niños de 8 a 10 años, ofreciendo ventajas significativas en portabilidad, privacidad y bajo costo computacional.

10.2.3. Corpus

Se generó un corpus emocional con niños mexicanos de 8 a 10 años en un entorno real de prevención del abuso sexual infantil utilizando realidad virtual inmersiva. Este corpus incluye más de 465 registros sincronizados que contienen:

- Datos espacio-temporales de cabeza y manos, información de foco de atención y las características descritas en 7.3.
- Etiquetas emocionales subjetivas de valencia y activación obtenidas mediante el Emoji-SAM RV.
- Etiquetas nominales complementarias recolectadas a través del Registro Emocional Nominal (REN).

A diferencia de la mayoría de los corpus existentes en computación afectiva, que suelen basarse en actuaciones inducidas, estímulos artificiales, o instrucciones estereotípicas, este corpus se caracteriza por su validez ecológica al haber sido recolectado durante una intervención educativa real con niños en edad escolar. Además, al utilizar exclusivamente señales nativas del visor de RVI, representa un recurso valioso para el avance de enfoques unimodales y no invasivos

en detección emocional. Este corpus constituye una contribución significativa al campo, ya que proporciona datos reales y contextualizados que pueden ser utilizados para entrenar, validar y mejorar modelos de detección emocional en población infantil, especialmente en aplicaciones educativas y de prevención.

10.2.4. Metodología

Se diseñó e implementó una metodología completa de etiquetado emocional adaptada a población infantil en entornos de RVI. Esta incluye dos instrumentos principales: el Emoji-SAM RV, una adaptación pictórica y culturalmente familiar del *Self-Assessment Manikin* para medir valencia y activación, y el Registro Emocional Nominal, que permite capturar etiquetas nominales complementarias. La metodología combina datos objetivos con datos subjetivos obtenidos inmediatamente después de cada escena, minimizando la interrupción de la experiencia inmersiva y reduciendo el sesgo de memoria. Se estableció un protocolo que asegura la sincronización precisa entre las señales espaciales y las autoevaluaciones, respetando la subjetividad individual y evitando el uso de etiquetado externo o actuación estereotípica. Esta aproximación representó una contribución metodológica relevante, al proponer un esquema de etiquetado que equilibra validez ecológica, practicidad y sensibilidad cultural para niños entre 8 y 10 años en contextos educativos sensibles.

10.2.5. Publicaciones

- Unimodal Model for Emotion Detection in Immersive Virtual Learning Environments Using Spatial Analysis of Hands and Head (J. E. Velázquez-Cano et al., [2026](#)). Este artículo presentó los resultados más relevantes de la presente investigación en la revista IEEE Latin America Transactions.
- Method to Identify Emotions in Immersive Virtual Learning Environments Using Head and Hands Spatial Behavioral Information (J. E. Velázquez-Cano et al., [2024](#)). Este trabajo se presentó en 16th Workshop on Intelligent Learning Environments (WILE), ganando el premio al mejor paper de dicha edición.
- Rumbo a la detección de estados emocionales en ambientes virtuales de aprendizaje inmersivo. Este trabajo mostró la metodología desarrollada junto con resultados preliminares de la misma. Se presentó en el marco del 3er Congreso Internacional de Tecnología y Ciencias Aplicadas (CITCA), celebrado en el Tecnológico Nacional de México, campus Centro Nacional de Investigación y Desarrollo Tecnológico. (TecNM/CENIDET), del 19 al 22 de noviembre de 2024.

10.3. Trabajos futuros

Como trabajo futuro se consideran relevantes las siguientes líneas de investigación:

- Integrar los modelos de detección emocional en el AVAI ASI para obtener el AVAI ASI clasificador. Esta versión no solo reportaría el estado emocional del usuario, sino que estaría preparada para incorporar escenas de regulación emocional basadas en evidencia, con el fin de obtener un AVAI ASI afectivo.
- Explorar la transferibilidad del modelo a otros entornos y aplicaciones de realidad virtual. El modelo desarrollado puede servir como mecanismo base para otros entornos virtuales, permitiendo continuar la exploración de la detección emocional unimodal.
- Investigar la inclusión de variables como la lateralidad, o condiciones socioculturales que podrían influir en las características espaciales. Esto tiene el potencial de mejorar la generalización del modelo.
- Mejorar los instrumentos de autoevaluación emocional (Emoji-SAM RV y REN) para aumentar su sensibilidad, reducir la carga cognitiva en niños y niñas, y mejorar su usabilidad. Se recomienda aplicar el REN también de forma virtual, inmediatamente después del Emoji-SAM RV.
- Explorar otros modelos de detección emocional basado en redes *Long-Short Term Memory* para la extracción de características mediante el análisis de segmentos temporales.
- Realizar estudios longitudinales para evaluar la estabilidad del modelo con el uso repetido del mismo contenido.
- Evaluar el impacto real del tutor afectivo sobre variables pedagógicas (aprendizaje, retención y cambio de comportamiento preventivo) en intervenciones a mayor escala.

En conclusión, este trabajo aporta una base sólida para el desarrollo de tutores afectivos no invasivos en realidad virtual, con potencial aplicación, no solo en la prevención del abuso sexual infantil, sino en diversos campos educativos incluidos los que involucren material sensibles.

11

Cierre

Es importante contextualizar el alcance de los resultados de esta investigación respecto a la prevención real del abuso sexual infantil. En primer lugar, debe resaltarse que el propósito principal de este trabajo no fue el desarrollo integral del AVAI ASI, sino la implementación de mecanismos de detección emocional en un entorno de realidad virtual inmersiva, orientados a su futura incorporación en un tutor afectivo. Se trata, por tanto, de una contribución metodológica y tecnológica dentro de un sistema más amplio.

Asimismo, es fundamental comprender que el AVAI ASI no pretende ser una herramienta capaz de detener de forma directa la incidencia de eventos de abuso sexual infantil, ni de erradicarlos por sí misma. Afirmar un porcentaje específico de reducción resultaría irreal e imposible de comprobar de manera rigurosa, especialmente considerando el bajo índice de denuncia y la falta de unificación en los criterios legales que clasifican estos casos en México. Como señalan Gubbels, van der Put et al. (2021), una ganancia en conocimientos y habilidades no constituye un indicador directo de que los participantes sean menos propensos a experimentar abuso. Para evaluar de manera confiable el impacto preventivo real se requieren estudios longitudinales de gran escala, los cuales exceden el alcance del presente proyecto, pero que pueden consultarse en los esfuerzos de Rivera-Rivera et al. (2026).

La prevención del abuso sexual infantil involucra múltiples variables sobre las cuales la víctima no siempre tiene control, incluyendo diferencias de poder, manipulación (*grooming*) y el hecho de que la mayoría de los agresores suelen ser personas cercanas o familiares (Valdez-Santiago et al., 2020). No obstante, como argumenta Finkelhor et al. (2007), considerar que las acciones de un niño o niña no pueden influir en absoluto constituye una sobresimplificación. Muchos agresores seleccionan a sus víctimas en función de su vulnerabilidad y de la facilidad de manipulación; en este sentido, cualquier forma de resistencia (incluso decir “no”) puede marcar una diferencia.

Desde esta perspectiva, el AVAI ASI se orienta a fortalecer las competencias de los niños y niñas en lo que se ha denominado las 3R (Wurtele 2009, como fue citado en Tutty et al. (2019)): *Reconocer* una situación de riesgo, *Remove* de ella lo más rápido posible y *Reportar* lo ocurrido a una persona de confianza. El contenido didáctico es principalmente informativo y procedimental, con el objetivo de incrementar las habilidades de autoprotección, incluyendo la autorrevelación de experiencias previas (Gubbels, van der Put et al., 2021). Aunque el impacto no siempre se refleje

de forma inmediata en las estadísticas, la participación en este tipo de programas puede generar un efecto significativo en la vida de los individuos, favoreciendo la reducción de incidentes incluso cuando estos no son reportados (Kenny et al., [2020](#)).

En este marco, el modelo de detección emocional desarrollado en la presente investigación constituye un componente fundamental, pero no el único, de un sistema más amplio orientado a la regulación afectiva y al acompañamiento educativo. Su valor radica en contribuir a mantener un estado emocional favorable para el aprendizaje, sin pretender ser, por sí mismo, la solución al complejo problema del abuso sexual infantil.

Referencias

- Ahmed, F., Paul, P. P., & Gavrilova, M. L. (2015). *DTW-based kernel and rank-level fusion for 3D gait recognition using Kinect*. The Visual Computer, 31(6), 915-924. <https://doi.org/10.1007/s00371-015-1092-0>
- Alizadehsalehi, S., Hadavi, A., & Huang, J. C. (2019). Virtual Reality for Design and Construction Education Environment. En AEI 2019 (pp. 193-203). <https://doi.org/10.1061/9780784482261.023>
- Allen, K. P., Livingston, J. A., & Nickerson, A. B. (2020). *Child Sexual Abuse Prevention Education: A Qualitative Study of Teachers' Experiences Implementing the Second Step Child Protection Unit*. American Journal of Sexuality Education, 15(2), 218-245. <https://doi.org/10.1080/15546128.2019.1687382>
- Alqahtani, F., Katsigiannis, S., & Ramzan, N. (2021). *Using Wearable Physiological Sensors for Affect-Aware Intelligent Tutoring Systems*. IEEE Sensors Journal, 21(3), 3366-3378. <https://doi.org/10.1109/JSEN.2020.3023886>
- Argasiński, J. K., & Węgrzyn, P. (2019). *Affective patterns in serious games*. Future Generation Computer Systems, 92, 526-538. <https://doi.org/10.1016/j.future.2018.06.013>
- Azevedo, R., Taub, M., Mudrick, N. V., Millar, G. C., Bradbury, A. E., & Price, M. J. (2017). Using Data Visualizations to Foster Emotion Regulation During Self-Regulated Learning with Advanced Learning Technologies. En J. Buder & F. W. Hesse (Eds.), Informational Environments : Effects of Use, Effective Designs (pp. 225-247). Springer International Publishing. https://doi.org/10.1007/978-3-319-64274-1_10
- Baloch, S., Bakar, S. A. R. S. A., Mokji, M. M., Wassem, S., & Hafeezallah, A. (2023). *Affect recognition using simplistic 2D skeletal features from the upper body movement*. Proceedings of the 2022 5th Artificial Intelligence and Cloud Computing Conference, 96-102. <https://doi.org/10.1145/3582099.3582115>
- Bari, A., & Gavrilova, M. (2019). *Artificial Neural Network Based Gait Recognition Using Kinect Sensor*. IEEE Access, PP, 1-1. <https://doi.org/10.1109/ACCESS.2019.2952065>
- Behera, A., Matthew, P., Keidel, A., Vangorp, P., Fang, H., & Canning, S. (2020). *Associating Facial Expressions and Upper-Body Gestures with Learning Tasks for Enhancing Intelligent Tutoring Systems*. International Journal of Artificial Intelligence in Education, 30(2), 236-270. <https://doi.org/10.1007/s40593-020-00195-2>

- Betella, A., & Verschure, P. F. M. J. (2016). *The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions*. PLOS ONE, 11(2), 1-11. <https://doi.org/10.1371/journal.pone.0148037>
- Bhatia, Y., Bari, A. H., Hsu, G.-S. J., & Gavrilova, M. (2022). *Motion Capture Sensor-Based Emotion Recognition Using a Bi-Modular Sequential Neural Network*. Sensors, 22(1). <https://doi.org/10.3390/s22010403>
- Bradley, M. M., & Lang, P. J. (1994). *Measuring emotion: The self-assessment manikin and the semantic differential*. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49-59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Brambilla, S., Boccignone, G., Borghese, N. A., & Ripamonti, L. A. (2022). *Between the Buttons: Stress Assessment in Video Games using Players' Behavioural Data*. International Conference on Computer-Human Interaction Research and Applications. <https://api.semanticscholar.org/CorpusID:253212942>
- Can, Y. S., Mahesh, B., & André, E. (2023). *Approaches, Applications, and Challenges in Physiological Emotion Recognition—A Tutorial Overview*. Proceedings of the IEEE, 111(10), 1287-1313. <https://doi.org/10.1109/JPROC.2023.3286445>
- Capece, N., Erra, U., & Mirauda, D. (2019). StreamFlowVR: A Tool for Learning Methodologies and Measurement Instruments for River Flow Through Virtual Reality. En L. T. De Paolis & P. Bourdot (Eds.), *Augmented Reality, Virtual Reality, and Computer Graphics* (pp. 456-471). Springer International Publishing.
- Cebeci, B., Celikcan, U., & Capin, T. K. (2019). *A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments* [e1893 cav.1893]. Computer Animation and Virtual Worlds, 30(3-4), e1893. <https://doi.org/10.1002/cav.1893>
- Chirico, A., Ferrise, F., Cordella, L., & Gaggioli, A. (2018). *Designing Awe in Virtual Reality: An Experimental Study*. Frontiers in Psychology, Volume 8 - 2017. <https://doi.org/10.3389/fpsyg.2017.02351>
- Cisneros-Brito, E.-J. (2018). *Tutor Virtual Inteligente para Ambiente Virtual de Aprendizaje Inmersivo mediante Tecnología de Cognición Aumentada* [Tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico].
- Corno, G., & Bouchard, S. (2015, enero). *An Innovative Positive Psychology Vr Application For Victims Of Sexual Violence: A Qualitative Study*. Nova Science.
- Coto, M., Mora, S., Grass, B., & Murillo-Morera, J. (2022). *Emotions and programming learning: systematic mapping*. Computer Science Education, 32(1), 30-65. <https://doi.org/10.1080/08993408.2021.1920816>
- Cummings, J. J., & Bailenson, J. N. (2016). *How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence*. Media Psychology, 19(2), 272-309. <https://doi.org/10.1080/15213269.2015.1015740>
- Desmet, P. M., Vastenburg, M. H., & Romero, N. (2016). *Mood measurement with Pick-A-Mood: review of current methods and design of a pictorial self-report scale*. Journal of Design Research, 14(3), 241-279. <https://doi.org/10.1504/JDR.2016.079751>
- Devon, A., & von Mühlenen Adrian. (2018). *Learning in virtual reality: Effects on performance, emotion and engagement*. Research in Learning Technology, 26. <https://doi.org/10.25304/rlt.v26.2140>

- Di Tecco, A., Foglia, P., & Prete, C. A. (2024). *Video Quality Prediction: An Exploratory Study With Valence and Arousal Signals*. IEEE Access, 12, 36558-36576. <https://doi.org/10.1109/ACCESS.2024.3374056>
- Dibris, S. P., Dibris, A. S., Dibris, A. C., & Dibris, F. O. (2013). *A set of Full-Body Movement Features for Emotion Recognition to Help Children affected by Autism Spectrum Condition*. <https://api.semanticscholar.org/CorpusID:232216129>
- D'Mello, S., & Graesser, A. (2012). *Dynamics of affective states during complex learning*. Learning and Instruction, 22(2), 145-157. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Endendijk, J. J., Tichelaar, H. K., Deen, M., & Deković, M. (2021). *Vil Du?! incorporation of a serious game in therapy for sexually abused children and adolescents*. Child and Adolescent Psychiatry and Mental Health, 15(1), 25. <https://doi.org/10.1186/s13034-021-00377-3>
- Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). *Poly-victimization: a neglected component in child victimization*. Child Abuse Neglect, 31(1), 7-26.
- Gerdemann, S. C., Vaish, A., & Hepach, R. (2025). *Body posture as a measure of emotional valence in young children: a preregistered validation study*. Frontiers in Developmental Psychology, 3. <https://doi.org/10.3389/fdpys.2025.1536440>
- Gubbels, J., Assink, M., Prinzie, P., & Put, C. (2021). *What Works in School-Based Programs for Child Abuse Prevention? The Perspectives of Young Child Abuse Survivors*. Social Sciences, 10, 404. <https://doi.org/10.3390/socsci10100404>
- Gubbels, J., van der Put, C. E., Stams, G.-J. J. M., & Assink, M. (2021). *Effective Components of School-Based Prevention Programs for Child Abuse: A Meta-Analytic Review*. Clinical Child and Family Psychology Review, 24(3), 553-578. <https://doi.org/10.1007/s10567-021-00353-5>
- Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2015). *The Indian Spontaneous Expression Database for Emotion Recognition*. IEEE Transactions on Affective Computing, 8, 1-1. <https://doi.org/10.1109/TAFFC.2015.2498174>
- Hasan, M. A., Noor, N. F. M., Rahman, S. S. B. A., & Rahman, M. M. (2020). *The Transition From Intelligent to Affective Tutoring System: A Review and Open Issues*. IEEE Access, 8, 204612-204638. <https://doi.org/10.1109/ACCESS.2020.3036990>
- Hassan, S. Z., Salehi, P., Røed, R. K., Halvorsen, P., Baugerud, G. A., Johnson, M. S., Lison, P., Riegler, M., Lamb, M. E., Griwodz, C., & Sabet, S. S. (2022). *Towards an AI-driven talking avatar in virtual reality for investigative interviews of children*. Proceedings of the 2nd Workshop on Games Systems, 9-15. <https://doi.org/10.1145/3534085.3534340>
- Hayashi, E. C. S., Posada, J. E. G., Maíke, V. R. M. L., & Baranauskas, M. C. C. (2016). *Exploring new formats of the Self-Assessment Manikin in the design with children*. Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems. <https://doi.org/10.1145/3033701.3033728>
- Herrera, A. (s.f.). *Ambiente virtual de aprendizaje inmersivo para realizar intervenciones educativas para la prevención del Abuso Sexual Infantil* [Tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico].
- Hofmann, S. M., Klotzsche, F., Mariola, A., Nikulin, V., Villringer, A., & Gaebler, M. (2021). *Decoding subjective emotional arousal from EEG during an immersive virtual reality experience*. Elife, 10.

- Holzwarth, V., Schneider, J., Handali, J., Gisler, J., Hirt, C., Kunz, A., & vom Brocke, J. (2021). *Towards estimating affective states in Virtual Reality based on behavioral data*. Virtual Reality, 25(4), 1139-1152. <https://doi.org/10.1007/s10055-021-00518-1>
- Jiménez, S., Juárez-Ramírez, R., Castillo, V. H., & Ramírez-Noriega, A. (2018). *Integrating affective learning into intelligent tutoring systems*. Universal Access in the Information Society, 17(4), 679-692. <https://doi.org/10.1007/s10209-017-0524-1>
- Jones, C., Scholes, L., Rolfe, B., & Stieler-Hunt, C. (2020). *A serious-game for child sexual abuse prevention: An evaluation of Orbit*. Child Abuse & Neglect, 107, 104569. <https://doi.org/10.1016/j.chiabu.2020.104569>
- Kaza, K., Psaltis, A., Stefanidis, K., Apostolakis, K., Thermos, S., Dimitropoulos, K., & Daras, P. (2016, junio). Body Motion Analysis for Emotion Recognition in Serious Games. https://doi.org/10.1007/978-3-319-40244-4_4
- Kenny, M. C., Helpingstine, C., & Long, H. (2020). *College students' recollections of childhood sexual abuse prevention programs and their potential impact on reduction of sexual victimization*. Child Abuse & Neglect, 104, 104486.
- Khadimallah, R., Abdelkefi, M., & Kallel, I. (2020). *Emotion Regulation in Intelligent Tutoring Systems: A Systematic Literature Review*. 2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 363-370. <https://doi.org/10.1109/TALE48869.2020.9368372>
- Kim, S.-J., & Kang, K.-A. (2017). *Effects of the Child Sexual Abuse Prevention Education (C-SAPE) Program on South Korean Fifth-Grade Students' Competence in Terms of Knowledge and Self-Protective Behaviors* [PMID: 27573418]. The Journal of School Nursing, 33(2), 123-132. <https://doi.org/10.1177/1059840516664182>
- Lalor, K., & McElvaney, R. (2010). *Child Sexual Abuse, Links to Later Sexual Exploitation/High-Risk Sexual Behavior, and Prevention/Treatment Programs* [PMID: 20679329]. Trauma, Violence, & Abuse, 11(4), 159-177. <https://doi.org/10.1177/1524838010378299>
- Leong, S. C., Tang, Y. M., Lai, C. H., & Lee, C. (2023). *Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing*. Computer Science Review, 48, 100545. <https://doi.org/10.1016/j.cosrev.2023.100545>
- Lin, Y., Lan, Y., & Wang, S. (2023). *A method for evaluating the learning concentration in head-mounted virtual reality interaction*. Virtual Reality, 27(2), 863-885.
- Lipp, N., Dużmańska-Misiarczyk, N., Strojny, A., & Strojny, P. (2021). *Evoking emotions in virtual reality: schema activation via a freeze-frame stimulus*. Virtual Reality, 25(2), 279-292. <https://doi.org/10.1007/s10055-020-00454-6>
- Lu, H., Jiuyi, Zhang, Z., Liu, R., Zeng, R., & Hu, X. (2025). *Emotion Recognition from skeleton Data: A Comprehensive Survey*. ArXiv, abs/2507.18026. <http://api.semanticscholar.org/CorpusID:280018004>
- Lu, M., Barlow, J., Meinck, F., & Neelakantan, L. (2023). *Unpacking School-Based Child Sexual Abuse Prevention Programs: A Realist Review* [PMID: 35544032]. Trauma, Violence, & Abuse, 24(4), 2067-2081. <https://doi.org/10.1177/15248380221082153>
- Lu, M., Barlow, J., Meinck, F., Walsh, K., & Wu, Y. (2023). *School-based child sexual abuse interventions: A systematic review and meta-analysis*. Research on Social Work Practice, 33(4), 390-412. <https://doi.org/10.1177/10497315221111393>
- Madera-Carrillo, H., Zarabozo, D., Ruiz, M., & Nanclares, P. (2015, febrero). *El Sistema Internacional de Imágenes Afectivas (IAPS) en población mexicana. Autoevaluación con*

- maniqués y etiquetas*. Instituto de Neurociencias del Centro Universitario de Ciencias Biológicas y Agropecuarias de la Universidad de Guadalajara. <https://doi.org/10.13140/RG.2.1.3220.3683>
- Magdin, M., Balogh, Z., Reichel, J., Francisti, J., Koprda, Š., & György, M. (2021). *Automatic detection and classification of emotional states in virtual reality and standard environments (LCD): comparing valence and arousal of induced emotions*. *Virtual Reality*, 25(4), 1029-1041. <https://doi.org/10.1007/s10055-021-00506-5>
- Makransky, G., & Lilleholt, L. (2018). *A structural equation modeling investigation of the emotional value of immersive virtual reality in education*. *Educational Technology Research and Development*, 66(5), 1141-1164. <https://doi.org/10.1007/s11423-018-9581-2>
- Malamsha, M. P., Sauli, E., & Luhanga, E. T. (2021). *Development and Validation of a Mobile Game for Culturally Sensitive Child Sexual Abuse Prevention Education in Tanzania: Mixed Methods Study*. *JMIR Serious Games*, 9(4). <https://doi.org/https://doi.org/10.2196/30350>
- Mancuso, V., Borghesi, F., Chirico, A., Bruni, F., Sarcinella, E. D., Pedroli, E., & Cipresso, P. (2024). *IAVRS—International Affective Virtual Reality System: Psychometric Assessment of 360° Images by Using Psychophysiological Data*. *Sensors*, 24(13). <https://doi.org/10.3390/s24134204>
- Mann, S., Furness, T., Yuan, Y., Iorio, J., & Wang, Z. (2018). *All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multim mediated Reality*. <https://arxiv.org/abs/1804.08386>
- Marcos Pablos, S., Lobato Alejano, F., & García Peñalvo, F. J. (2022). *Integrating Emotion Recognition Tools for Developing Emotionally Intelligent Agents*. *International Journal of Interactive Multimedia and Artificial Intelligence*, 7(6), 69-76. <https://doi.org/10.9781/ijimai.2022.09.004>
- Marín-Morales, J., Llinares, C., Guixeres, J., & Alcañiz, M. (2020). *Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing*. *Sensors*, 20(18). <https://doi.org/10.3390/s20185163>
- Mehrabian, A. (2017). *Nonverbal communication*. Routledge. <https://doi.org/https://doi.org/10.4324/9781351308724>
- Mikropoulos, T. A., & Natsis, A. (2011). *Educational virtual environments: A ten-year review of empirical research (1999–2009)*. *Computers & Education*, 56(3), 769-780. <https://doi.org/https://doi.org/10.1016/j.compedu.2010.10.020>
- Noroozi, F., Corneau, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2021). *Survey on Emotional Body Gesture Recognition*. *IEEE Transactions on Affective Computing*, 12(2), 505-523. <https://doi.org/10.1109/TAFFC.2018.2874986>
- Olmos, E., Cavalcanti, J. F., Soler, J.-L., Contero, M., & Alcañiz, M. (2018). *Mobile Virtual Reality: A Promising Technology to Change the Way We Learn and Teach*. En S. Yu, M. Ally & A. Tsinakos (Eds.), *Mobile and Ubiquitous Learning: An International Handbook* (pp. 95-106). Springer Singapore. https://doi.org/10.1007/978-981-10-6144-8_6
- Pekrun, R. (2006). *The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice*. *Educational Psychology Review*, 18(4), 315-341. <https://doi.org/10.1007/s10648-006-9029-9>
- Petrovica, S., Anohina-Naumeca, A., & Ekenel, H. K. (2017). *Emotion Recognition in Affective Tutoring Systems*. *Procedia Computer Science*, 104(C), 437-444. <https://doi.org/10.1016/j.procs.2017.01.157>

- Pfeuffer, K., Geiger, M. J., Prange, S., Mecke, L., Buschek, D., & Alt, F. (2019). *Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality*. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1-12. <https://doi.org/10.1145/3290605.3300340>
- Piana, S., Staglianò, A., Odone, F., & Camurri, A. (2016). *Adaptive Body Gesture Representation for Automatic Emotion Recognition*. ACM Trans. Interact. Intell. Syst., 6(1). <https://doi.org/10.1145/2818740>
- Piana, S., Staglianò, A., Odone, F., Verri, A., & Camurri, A. (2014). *Real-time Automatic Emotion Recognition from Body Gestures*. CoRR, abs/1402.5047. <http://arxiv.org/abs/1402.5047>
- Picard, R. W. (2000). *Affective Computing*. The MIT Press.
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). *Foundations of Game-Based Learning*. Educational Psychologist, 50(4), 258-283. <https://doi.org/10.1080/00461520.2015.1122533>
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). *The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology*. Dev Psychopathol, 17(3), 715-734.
- Regal, G., Voigt-Antons, J.-N., Schmidt, S., Schrammel, J., Kojić, T., Tscheligi, M., & Möller, S. (2019). *Questionnaires embedded in virtual environments: reliability and positioning of rating scales in virtual environments*. Quality and User Experience, 4(1), 5. <https://doi.org/10.1007/s41233-019-0029-1>
- Reichenberger, J., Pfaller, M., & Mühlberger, A. (2020). *Gaze Behavior in Social Fear Conditioning: An Eye-Tracking Study in Virtual Reality*. Frontiers in Psychology, Volume 11 - 2020. <https://doi.org/10.3389/fpsyg.2020.00035>
- Rivera-Rivera, L., Séris-Martínez, M., González-Serna, G., Castro-Sánchez, N. A., González-Hernández, D., Jiménez-Tapia, A., & Treviño-Siller, S. (2026). *Pinta tu Raya ASI [Set your Limit LIKE THIS]: an educational intervention using immersive reality to prevent Child Sexual Abuse in Mexico*. Frontiers in Public Health, Volume 13 - 2025. <https://doi.org/10.3389/fpubh.2025.1686104>
- Russell, J. (1980). *A circumplex model of affect*. Journal of Personality and Social Psychology, 39, 1161-1178. <https://doi.org/10.1037/h0077714>
- Sapiński, T., Kamińska, D., Pelikant, A., & Anbarjafari, G. (2019). *Emotion Recognition from Skeletal Movements*. Entropy (Basel), 21(7).
- Scherer, K. R. (2005). *What are emotions? And how can they be measured?* Social Science Information, 44(4), 695-729. <https://doi.org/10.1177/0539018405058216>
- Scholes, L., Jones, C., Stieler-Hunt, C., & Rolfe, B. (2014). *Serious games for learning: games-based child sexual abuse prevention in schools*. International Journal of Inclusive Education, 18(9), 934-956. <https://doi.org/10.1080/13603116.2013.860195>
- Schweiger, M., Wimmer, J., Chaudhry, M., Siegle, B., & Xie, D. (2022). *Lernerfolg in der Schule durch Augmented und Virtual Reality?: Eine quantitative Synopse von Wirkungsstudien zum Einsatz virtueller Realitäten in Grund- und weiterführenden Schulen*. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, 47, 1-25. <https://doi.org/10.21240/mpaed/47/2022.04.01.X>
- Siam, A. I., Soliman, N. F., Algarni, A. D., Abd El-Samie, F. E., & Sedik, A. (2022). *Deploying Machine Learning Techniques for Human Emotion Detection*. Computational Intelligence and Neuroscience, 2022(1), 8032673. <https://doi.org/https://doi.org/10.1155/2022/8032673>

- Somarathna, R., Bednarz, T., & Mohammadi, G. (2023). *Virtual Reality for Emotion Elicitation – A Review*. IEEE Transactions on Affective Computing, 14(4), 2626-2645. <https://doi.org/10.1109/TAFFC.2022.3181053>
- T S, A., & Guddeti, R. R. (2020). *Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks*. Education and Information Technologies, 25. <https://doi.org/10.1007/s10639-019-10004-6>
- Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2021). *How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system?* [Temporal and adaptive process of regulated learning -what can multimodal data tell?]. Learning and Instruction, 72, 101200. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2019.04.001>
- Toet, A., Heijn, F., Brouwer, A.-M., Mioch, T., & van Erp, J. B. F. (2020). *An Immersive Self-Report Tool for the Affective Appraisal of 360° VR Videos*. Frontiers in Virtual Reality, Volume 1 - 2020. <https://doi.org/10.3389/frvir.2020.552587>
- Tutty, L. M. (2014). *Listen to the children: kids' impressions of Who Do You Tell™*. Journal of Child Sexual Abuse, 23(1), 17-37.
- Tutty, L. M., Aubry, D., & Velasquez, L. (2019). *The "Who Do You Tell?"™ Child Sexual Abuse Education Program: Eight Years of Monitoring*. Journal of Child Sexual Abuse, 29(1), 2-21.
- Universidad de Alicante. (2007). *Tema 8. La Emoción* [Accedido: 2025-02-03]. <https://rua.ua.es/server/api/core/bitstreams/ce5616f3-c0ff-4e7f-8efb-c65b43907f30/content>
- Valdez-Santiago, R., Villalobos, A., Arenas-Monreal, L., Flores-Celis, K., & Ramos-Lira, L. (2020). *Abuso sexual infantil en México: conductas de riesgo e indicadores de salud mental en adolescentes*. Salud Pública de México, 62(6, Nov-Dic), 661-671. <https://doi.org/10.21149/11924>
- Van Bennekom, M. J., de Koning, P. P., Gevonden, M. J., Kasanmoentalib, M. S., & Denys, D. (2021). *A Virtual Reality Game to Assess OCD Symptoms*. Frontiers in Psychiatry, Volume 11 - 2020. <https://doi.org/10.3389/fpsy.2020.550165>
- Van Rooij, M., Lobel, A., Harris, O., Smit, N., & Granic, I. (2016). *DEEP: A Biofeedback Virtual Reality Game for Children At-risk for Anxiety*. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 1989-1997. <https://doi.org/10.1145/2851581.2892452>
- Velázquez-Cano, J. E., González-Serna, G., Rivera-Rivera, L., González-Franco, N., López-Sánchez, M., & Reyes-Ortiz, J. A. (2026). *Unimodal Model for Emotion Detection in Immersive Virtual Learning Environments Using Spatial Analysis of Hands and Head*. IEEE Latin America Transactions, 24(6), 560-569. <https://latamt.ieee9.org/index.php/transactions/article/view/10344>
- Velázquez-Cano, J. E., Gonzáles-Serna, J. G., Rivera-Rivera, L., Gonzáles-Franco, N., Reyes-Ortiz, J. A., López-Sánchez, M., & Valenzuela-Robles, B. D. (2024). Method to Identify Emotions in Immersive Virtual Learning Environments Using Head and Hands Spatial Behavioral Information. En H. Calvo, L. Martínez-Villaseñor, H. Ponce, R. Zatarain Cabada, M. Montes Rivera & E. Mezura-Montes (Eds.), Advances in Computational Intelligence. MICAI 2023 International Workshops (pp. 21-28). Springer Nature Switzerland.
- Velázquez-Cano, J.-E. (2021). *Entorno Virtual de Aprendizaje para evaluar niveles de inmersión, interacción y estados emocionales durante un proceso de intervención educativa de prevención*

- del Abuso Sexual infantil en niños y niñas de edad escolar* [Tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico].
- Vilhunen, E., Turkkila, M., Lavonen, J., Salmela-Aro, K., & Juuti, K. (2022). *Clarifying the Relation Between Epistemic Emotions and Learning by Using Experience Sampling Method and Pre-posttest Design*. *Frontiers in Education, Volume 7 - 2022*. <https://doi.org/10.3389/feduc.2022.826852>
- Wallbott, H. G. (1998). *Bodily expression of emotion*. *European Journal of Social Psychology*, 28, 879-896. <https://api.semanticscholar.org/CorpusID:145268858>
- Walsh, K., Berthelsen, D., Hand, K., Brandon, L., & Nicholson, J. M. (2019). *Sexual Abuse Prevention Education in Australian Primary Schools: A National Survey of Programs*. *Journal of Interpersonal Violence*, 34(20), 4328-4351. <https://doi.org/10.1177/0886260519869246>
- Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., & Zhang, W. (2022). *A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances*. <https://arxiv.org/abs/2203.06935>
- Woodworth, J. W., & Borst, C. W. (2024). *Design and Validation of a Library of Active Affective Tasks for Emotion Elicitation in VR*. 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR), 398-407. <https://doi.org/10.1109/VR58804.2024.00061>
- Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. B., Affal, H. B., & Hussin, N. B. (2019). *Affective computing in education: A systematic review and future research*. *Computers & Education*, 142, 103649. <https://doi.org/https://doi.org/10.1016/j.compedu.2019.103649>
- Yaremych, H. E., & Persky, S. (2019). *Tracing Physical Behavior in Virtual Reality: A Narrative Review of Applications to Social Psychology*. *Journal of Experimental Social Psychology*, 85.
- Younis, E. M. G., Mohsen, S., Houssein, E. H., & Ibrahim, O. A. S. (2024). *Machine learning for human emotion recognition: a comprehensive review*. *Neural Computing and Applications*, 36(16), 8901-8947. <https://doi.org/10.1007/s00521-024-09426-2>
- Yu, M., Xiao, S., Hua, M., Wang, H., Chen, X., Tian, F., & Li, Y. (2022). *EEG-based emotion recognition in an immersive virtual reality environment: From local activity to brain network features*. *Biomedical Signal Processing and Control*, 72, 103349. <https://doi.org/https://doi.org/10.1016/j.bspc.2021.103349>
- Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). *The impact of gamification on learning and instruction: A systematic review of empirical evidence*. *Educational Research Review*, 30, 100326. <https://doi.org/https://doi.org/10.1016/j.edurev.2020.100326>
- Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). *Emotion recognition using multi-model data and machine learning techniques: A tutorial and review*. *Information Fusion*, 59, 103-126. <https://doi.org/https://doi.org/10.1016/j.inffus.2020.01.011>