



Análisis de bases de datos de expresiones faciales para la identificación automática de emociones centradas en el aprendizaje

Analysis of databases of facial expressions for the automatic identification of learning-centered emotions

Yesenia Nohemí González-Meneses¹ , Josefina Guerrero-García¹ 

¹Benemérita Universidad Autónoma de Puebla, Puebla, México.
yesenia.gonzalezm@alumno.buap.mx, josefina.guerrero@correo.buap.mx

(Recibido: 28 julio 2021; aceptado: 17 septiembre 2021; Publicado en Internet: 1 diciembre 2021)

Resumen. Este trabajo presenta el análisis del estado del arte de bases de datos de expresiones faciales para la identificación automática de emociones centradas en el aprendizaje. La obtención de datos para los procesos de reconocimiento automático en un contexto específico es esencial para su éxito. Así, este tipo de proyectos inician haciendo una revisión de la información disponible para llevar a cabo las etapas de entrenamiento y clasificación de las emociones con las técnicas computacionales que se propongan. Se describen las actividades de búsqueda de las bases de datos de expresiones faciales que capturan emociones centradas en el aprendizaje. Estas actividades formaron parte de las etapas de la metodología del trabajo para reconocer las emociones de estudiantes mientras realizaban actividades de aprendizaje en línea. Esto permitió justificar la creación de la base de datos desde la formalización de un protocolo para su captura hasta su digitalización.

Palabras clave: Bases de datos, Identificación automática de emociones, Expresiones faciales.

Abstract. This work presents the analysis of the state of the art of facial expressions databases for the automatic identification of learning-centered emotions. Obtaining data for automatic recognition processes in a specific context is essential for their success. Thus, this project begins by reviewing the information available to carry out the training and classification stages of emotions with the proposed computational techniques. The search activities of the databases of facial expressions that capture learning-centered emotions are described. These activities were part of the stages of the work methodology to recognize students' emotions while they carried out online learning activities. This allowed justifying the creation of the database, formalizing a protocol from its capture to its digitization.

Keywords: Databases, Automatic identification of emotions, Facial expressions.

Tipo de artículo: Artículo de investigación.

1 Introducción

Este trabajo representa la etapa de investigación, análisis y selección de las bases de datos, de las emociones, y de los algoritmos de aprendizaje computacional necesarios en la metodología para la identificación automática de emociones en ambientes de aprendizaje propuesta en (González Meneses et al., 2019). El inicio de esta etapa consistió en el estudio del marco teórico referente al reconocimiento automático de emociones centradas en el aprendizaje y de las bases de datos de expresiones faciales necesarias para tal propósito. Posteriormente se realizó la recopilación de investigaciones sobre ambos temas lo que permitió describir de manera cronológica y detallada el estado del arte sobre las bases de datos y el reconocimiento automático de emociones centradas en el aprendizaje.

El reconocimiento automático de emociones en la interacción humano-computadora se puede plantear a partir del uso de sensores fisiológicos que permiten la adquisición de datos en forma de señales fisiológicas y a través del análisis de señales de voz, de imágenes de video del rostro, ojos, cabeza, o movimientos corporales de las personas (Fuentes et al., 2017). Dentro de las interfaces humano-computadora para capturar señales fisiológicas que pueden ayudar al reconocimiento de emociones, como lo mencionan en (Cabada et al., 2014), se encuentran las diademas de ondas cerebrales, que miden la actividad cerebral y envían la información a un dispositivo electrónico en forma de electroencefalograma (EEG). También están las pulseras cardiovasculares, que miden el ritmo cardíaco y proporcionan información en forma de

electrocardiograma (ECG). Además, se encuentran los sensores de actividad electrodermal que miden el nivel de conductividad de la piel a través del sudor en las manos. Las cámaras térmicas que permiten medir el cambio de temperatura del cuerpo humano asociada a los diferentes estados emocionales. También existen dispositivos para medir la respuesta muscular o la actividad eléctrica derivada de una estimulación nerviosa de un músculo, en forma de electromiografía (EMG).

Por lo que respecta a dispositivos relacionados a la identificación del comportamiento de las personas – como posturas del cuerpo y gesticulaciones– encontramos las cámaras de video tradicionales, cámaras web o cámaras de realidad aumentada que permiten grabar expresiones faciales y movimientos corporales, así como la mirada de ojos, importante para el reconocimiento de emociones. En este tipo de dispositivos también están las grabadoras de voz, otro medio utilizado para la identificación de emociones, como lo mencionan en (Cowie et al., 2001).

Los datos obtenidos de diferentes dispositivos deben ser procesados y clasificados de acuerdo con los objetivos buscados. Por ejemplo, en este caso son utilizados para el reconocimiento de emociones centradas en el aprendizaje capturados en ambientes educativos reales cuando estudiantes están ejecutando alguna actividad de aprendizaje haciendo uso de algún dispositivo electrónico, ya sea viendo algún video, estudiando, investigando o leyendo.

Desde el enfoque computacional la detección y reconocimiento de emociones es un problema relevante dentro del área de investigación de cómputo afectivo, pues es el punto de partida en el estudio y desarrollo de sistemas de interacción humano-computadora sensibles a las emociones de los seres humanos. La complejidad de los problemas con los que trata el cómputo afectivo radica en el hecho de ser un área interdisciplinaria que abarca las ciencias computacionales, la psicología, y las ciencias cognitivas como lo explica (Picard, 2000) pionera del cómputo afectivo. Así que para desarrollar modelos de reconocimiento de emociones que generen resultados satisfactorios con un grado de precisión aceptable, será necesario el estudio profundo de estas tres áreas. Particularmente, en el área computacional, el reto es seleccionar y probar algoritmos de aprendizaje computacional que puedan ser integrados en un modelo completo de reconocimiento de emociones centradas en el aprendizaje (ECA).

El problema del reconocimiento automático de emociones ha sido un área de investigación altamente activa en los últimos años. Aun así, se está lejos de una solución clara y que a su vez esté al alcance de la mayoría de las personas. Diversos inconvenientes han influido en la construcción de una solución apropiada desde el punto de vista computacional. Por un lado, un factor que afecta el desempeño de los reconocedores de emociones en contextos reales es la dificultad de generar bases de datos con emociones espontáneas. Generalmente se realizan trabajos con bases de datos formadas a partir de actuaciones de personas, las cuales proporcionan retratos de emociones representando expresiones prototípicas e intensas que facilitan la búsqueda de correlaciones y la subsecuente clasificación automática. Este tipo de bases de datos suelen grabarse en un ambiente controlado, lo cual disminuye problemas en el procesamiento de la información (por ejemplo, ruido). Además, se puede garantizar una cantidad balanceada de muestras por clase. Como consecuencia, no se han tenido buenos resultados al trasladar el conocimiento extraído de estas bases de datos a contextos reales (Steidl, 2009).

En contraparte, los repositorios con datos de emociones espontáneas muestran información con contenido emocional no perteneciente a una sola clase, sino que son una mezcla de emociones. En otros casos, existen muestras con una carga emocional muy ligera, cercana a un estado emocional neutro. Además, las bases de datos con emociones espontáneas suelen grabarse en ambientes ruidosos como salones de clase, cuartos de estudio, áreas de entretenimiento, oficinas, fábricas, o en conversaciones telefónicas, lo que conlleva la inclusión de ruido. Finalmente, por la naturaleza misma del problema, no es posible asegurar una cantidad balanceada de ejemplos por clase.

Otro reto por resolver es la extracción y selección de un conjunto de características que permitan reconocer emociones en los datos capturados de manera espontánea. A pesar de que los avances en el área han sido importantes, es evidente que en contextos realistas aún falta bastante por hacer y uno de los principales retos a vencer es la captura de datos en ambientes educativos reales para conformar una base de datos de calidad que permita obtener mejores resultados en las etapas de entrenamiento y clasificación de los algoritmos utilizados para la identificación de ECA. Para conocer y entender el problema, en la sección 2 se describen los conceptos relacionados con el proceso de reconocimiento automático de ECA, presentando un análisis de los trabajos más recientes al respecto. La sección 3 presenta un estado del arte sobre el reconocimiento automático de emociones centradas en el aprendizaje. La sección 4 presenta un estado del arte sobre bases de datos de expresiones faciales. Finalmente, en la sección 5, son presentadas las conclusiones del artículo.

2 Marco teórico

2.1 Cómputo afectivo y reconocimiento de emociones

El cómputo afectivo es el estudio y desarrollo de sistemas y dispositivos que pueden reconocer, interpretar, procesar, y simular el afecto humano. Se relaciona directamente con las emociones, cómo y cuándo son producidas, y lo que ellas generan. El cómputo afectivo incluye interpretar emociones, y por lo tanto puede ayudar a desarrollar y probar nuevas y viejas teorías. También incluye muchas otras cosas, como dar a la computadora la habilidad de reconocer y expresar emociones, desarrollando su habilidad para responder inteligentemente a la emoción del humano, y hacer posible regularlas y utilizarlas (Picard, 2003).

Uno de los retos del cómputo inteligente es la habilidad para reconocer emociones, inferir un estado emocional desde la observación de expresiones, y razonar sobre la situación que genera la emoción. La computadora debe de tratar de reconocer la emoción del usuario. El reconocimiento puede requerir habilidades de visión y oído para recolectar expresiones faciales, gestos, y entonación de la voz. Adicionalmente, la computadora puede usar otras entradas de datos como las lecturas de temperatura infrarroja, medidas de respuestas electrodermales, entre otras. Una vez que las expresiones emocionales son censadas y reconocidas, el sistema puede usar su conocimiento sobre la situación y sobre la generación de la emoción para inferir el estado emocional subyacente el cual muy probablemente da origen a las expresiones. Dar a una computadora estas habilidades perceptivas e interpretativas puede potencialmente darle la posibilidad de reconocer emociones tal como lo hacen las personas. Para saber cuándo una computadora ha logrado esta habilidad, se debe de evaluar la precisión del reconocimiento comparándola con el reconocimiento que un humano hace después de observar las expresiones de la cara, los gestos, o voz de una persona.

Como lo menciona Rosalind Picard (2000) en su libro, el reconocimiento de emociones es un problema de reconocimiento de patrones, que incluso las mismas personas no pueden identificar de forma totalmente correcta, por lo tanto, la expresión “reconocimiento de emociones” debe ser interpretada como inferir un estado emocional de la observación de expresiones y comportamiento, y del razonamiento de la situación que genera la emoción (Picard, 2000). Para lograr esto existen numerosos requerimientos técnicos, por ejemplo, para capturar la entrada de datos, se requiere de tecnología precisa para digitalizar señales fisiológicas, de audio o video, así como también investigar cuáles señales son más importantes. En el reconocimiento de patrones, las características informativas de las señales necesitan ser identificadas – estadística, estructura, no linealidad– junto con variables que influyen en el significado de estas características.

2.2 Modelos de emociones

Cowie et al. (2001) define a la emoción como un episodio de cambios sincronizados e interrelacionados en los estados de todos o la mayoría de los cinco subsistemas del organismo en respuesta a la evaluación de un evento de estímulo externo o interno relevante para las principales preocupaciones del organismo. Hill (2010) las define como sucesos espontáneos que se desarrollan dentro de nosotros. Ayudan a movilizar al cuerpo para salir de una situación de urgencia. Tienen un inicio, una cúspide, y por lo general se atenúan en pocos segundos. Son sumamente propensas a detonar estímulos que conducen a la acción. Son mucho más intensas que los estados de ánimo.

Al examinar las opiniones expuestas por diversos expertos en la materia (Cornelius, 1996), surge un consenso psicofisiológico (mente/cuerpo). Existen tres cualidades universales que caracterizan las emociones: Un componente de sentimiento: sensaciones físicas, incluyendo cambios químicos en el cerebro. Un componente de pensamiento: apreciaciones “racionales” conscientes o intuitivas. Un componente de acción: reacciones expresivas (sonrisas o fruncir el ceño), al igual que conductas de afrontamiento (pelea o huida). A veces existe opcionalmente: Un componente sensorial; vista, sonido, etcétera, que se inmiscuye y sirve como detonador de la respuesta emocional.

El reconocimiento de emociones en los seres humanos tiene sus inicios con el análisis gestual haciendo observaciones detalladas, ya que las expresiones faciales son uniformes y universales. Charles Darwin fue el primero en descubrir la sorprendente verdad acerca de la naturaleza innata y preprogramada de las expresiones faciales. Por desgracia, no fue sino hasta mediados de 1960 que el Doctor Paul Ekman, profesor de la Universidad de California San Francisco y su colega Wally Friesen crearon el Sistema de Codificación

de Acción Facial (*Facial Action Coding System – FACS*), haciendo posible cuantificar las emociones y definir siete emociones básicas: sorpresa (neutra), temor, enojo, tristeza, repugnancia, desprecio (negativas), y felicidad (positiva) (Ekman, 2004). El FACS es un sistema basado en cambios de músculos faciales, y puede caracterizar acciones de la cara para expresar emociones humanas individuales. El FACS codifica los movimientos de músculos del rostro específicos llamados unidades de acción (*AUs, Action Units*), las cuales reflejan distintos cambios momentáneos en la apariencia facial (Hjortsj et al., 2019). Las unidades de acción facial (AUs) codifican las acciones fundamentales (hay 46 AUs básicas) de un músculo o de un grupo de músculos típicamente vistos cuando la expresión facial de una emoción en particular es producida. Para reconocer la emoción facial, la AU individual es detectada y el sistema de clasificación facial la categoriza de acuerdo con la combinación de AUs.

A partir del modelo de emociones básicas es posible ampliar el conjunto de emociones para anticipar más situaciones y considerar emociones específicas para un contexto. Cada emoción secundaria es una combinación de dos emociones primarias, por ejemplo, indignación es una combinación de enojo y sorpresa. Cuando tomamos en cuenta todas las combinaciones de la lista de Ekman, se crean en total 30 emociones. Así las emociones secundarias se pueden tomar como referencia universal mensurable, contra las cuales se pueden cuantificar metas específicas en forma científica (Hill, 2010). Dentro de este grupo de emociones secundarias se encuentra la clasificación de las emociones centradas en el aprendizaje como interés, aburrimiento, confusión, frustración, y sorpresa.

2.3 ¿Cómo medir las emociones?

Los psicólogos han usado dos métodos para obtener reportes personales de la experiencia emocional: (1) el enfoque de emociones discreto, y (2) el enfoque dimensional. El primero, describe las expresiones como estados claramente separados. El enfoque de emociones discreto se basa en la categorización que es reflejada en la organización de los campos semánticos en los lenguajes naturales. La justificación de aceptar la estructura provista por el lenguaje es el hecho de que las categorías basadas en el lenguaje parecen corresponder a patrones de respuesta únicos, por ejemplo, los patrones específicos categoría-emoción de las expresiones faciales y de la boca, así como de respuestas fisiológicas. El método de evaluación que los investigadores adoptan con el enfoque de emociones discreto es el uso de escalas nominales, ordinales o intervalos de características, en una o varias listas de emociones que varían de acuerdo con la categorización o escalas que manejen. Aunque existen algunos instrumentos estandarizados de este tipo, muchos investigadores prefieren crear categorías de emociones que son relevantes para un contexto de investigación específico (Scherer, 2005, p. 717).

En el segundo método, el enfoque dimensional, se sugiere que los sentimientos subjetivos pueden ser descritos por su posición en un espacio tridimensional formado por la dimensión de valencia (positiva-negativa), incitación (calma-excitación), y tensión (tenso-relajado). Estas tres dimensiones se usan para describir el fenómeno mental de los sentimientos y varían con estados medibles del cuerpo tales como la incitación fisiológica. Dada la dificultad de identificar consistentemente una tercera dimensión -como incitación o tensión- muchos teóricos dimensionales se limitan a la dimensión de valencia e incitación. Algunos otros sugieren estructuras circulares más adaptadas para mapear los sentimientos emocionales en el espacio de dos emociones (Scherer, 2005, p. 718).

2.4 Reconocimiento de expresiones faciales

Las expresiones faciales son de los canales de comunicación no verbal más importantes para expresar las emociones internas y las intenciones. Como lo describimos anteriormente, Ekman define seis expresiones básicas (angustia, disgusto, miedo, felicidad, tristeza, y sorpresa) que son universales entre los seres humanos. El reconocimiento de expresiones faciales (*Facial Expression Recognition – FER*) automatizado ha sido un tema de estudio por décadas. Aunque hay muchos avances en el desarrollo de sistemas de FER, muchos de ellos muestran comportamiento inadecuado en aplicaciones prácticas o carencia de generalización debido a las condiciones controladas en las cuales fueron desarrollados.

El reconocimiento de expresiones es un proceso muy retador dividido en tres fases: inicio, ápice, y culminación. El ápice describe la expresión en su máxima intensidad y la culminación describe la expresión desvaneciéndose. La mayoría de las veces la entrada del evento de la expresión facial desde el inicio hasta su culminación es muy rápido, lo cual hace el proceso de reconocimiento muy complejo. En el

reconocimiento de emociones ya sean primarias o secundarias (como las ECA) el elemento primordial del proceso es la base de datos de expresiones faciales con la que se trabajará. De ella dependerá la calidad de los resultados de reconocimiento, entonces en su creación recae uno de los esfuerzos principales en un proceso de reconocimiento automático de emociones.

Muchos métodos han sido propuestos para el FER. Los enfoques tradicionales principalmente consideran imágenes independientes e ignoran las relaciones temporales de los *frames* consecutivos en una secuencia lo cual es esencial para reconocer cambios sutiles en la apariencia de imágenes faciales especialmente en la transición de emociones entre *frames*. Ya sea considerando imágenes individuales o una secuencia de *frames*, en los enfoques tradicionales el primer paso consiste en la extracción de características bajo diferentes enfoques.

2.5 Bases de datos de expresiones de emociones

Una base de datos (BD) de expresiones de emociones es una colección de imágenes, videos, voz, y señales fisiológicas relacionadas con un amplio rango de emociones. Su contenido corresponde a expresiones de emociones relacionadas al contexto en donde fueron capturadas y en base al cual son etiquetadas, lo cual es esencial para el entrenamiento, prueba, y validación de algoritmos para el desarrollo de sistemas de reconocimiento de expresiones. El etiquetado de emociones puede ser hecho en escala discreta o continua. Muchas de las bases de datos usualmente se basan en la teoría de emociones que asume que existen en una escala discreta seis emociones básicas y una variedad aproximada de 22 de emociones secundarias. Sin embargo, en algunas bases de datos las emociones son etiquetadas en la escala continua de excitación-valencia. Otras bases de datos incluyen las AUs del FACS (Ekman et al., 2002)

Las bases de datos de expresiones de emociones en su mayoría están formadas solo por expresiones faciales y se clasifican en actuadas y espontáneas. En las bases de datos de expresiones actuadas, se le pide a los participantes que muestren diferentes expresiones emocionales, mientras que, en las bases de datos espontáneas las expresiones son naturales. Las expresiones espontáneas difieren de las actuadas notablemente en términos de intensidad, configuración y duración. En la mayoría de los casos, las expresiones actuadas son exageradas, mientras que las espontáneas son sutiles y difieren en apariencia. Además de esto, la síntesis de algunas AUs apenas es alcanzable sin experimentar el estado emocional asociado, por lo que no es posible capturar datos fisiológicos pues estos no pueden ser controlados por las personas y por lo tanto no corresponden a la emoción actuada. De ahí la importancia de contar con bases de datos que, primero correspondan a capturas realizadas en el contexto real donde existe la necesidad de identificar emociones de expresiones faciales y segundo que estas sean capturadas de expresiones generadas de manera natural.

3 Revisión del estado del arte del reconocimiento automático de emociones centradas en el aprendizaje

Partimos de consultas generales sobre el reconocimiento automático de emociones empleando los descriptores *automatic recognition*, *identification recognition*, y *learning centered emotions*. La búsqueda fue realizada en Web of Science dentro del rango 2010 a 2020 encontrando 525 trabajos relacionados al tema. La última consulta fue realizada en diciembre de 2020. Los descriptores empleados para la búsqueda de documentos fueron. De este total de publicaciones aproximadamente el 60% de ellas corresponden a investigaciones en el campo de las ciencias computacionales, y el 40% corresponden a investigaciones de otras áreas como educación, psicología, neurología, neuropsicología, psiquiatría, entre otras. Así que tomamos este 60% para iniciar nuestro análisis estado del arte. La mayor parte de estas investigaciones (aproximadamente el 80%) se enfocan en el reconocimiento de las emociones básicas: alegría, tristeza, enojo, miedo, asco, sorpresa, y neutro; y sólo algunos pocos (aproximadamente el 20%) intentan reconocer emociones centradas en el aprendizaje: interés, aburrimiento, frustración, confusión, excitación, y sorpresa.

Así, el análisis del estado del arte que se realizó consiste en una recopilación de artículos enfocados al reconocimiento de emociones en actividades de aprendizaje y a la identificación de la relación emoción-aprendizaje. Hay una clara diferencia en la literatura encontrada. Existen trabajos que, aunque realizan una identificación de emociones le dan mayor relevancia al análisis de la relación emoción-aprendizaje (aproximadamente 37% = 20 *publicaciones*), mientras que otros hacen énfasis en los algoritmos para la

identificación automática de emociones y la precisión de reconocimiento alcanzada dando menor importancia al análisis de la relación emoción-aprendizaje (aproximadamente 63% = 30 *publicaciones*). Es en estos trabajos en los que nos enfocamos en este artículo.

En la [Tabla 1](#) presentamos las principales características analizadas en la revisión de la literatura sobre trabajos del reconocimiento automático de ECA. En la tabla describimos los trabajos más recientes de un total de 30. En todos los trabajos se analizaron las siguientes características: emociones que reconocen, bases de datos que utilizan, tecnologías de captura de datos, el método de etiquetamiento de las emociones, los algoritmos de selección de características y de clasificación, la actividad de aprendizaje que utilizan los alumnos en la captura de datos y finalmente las métricas con las que evalúan sus resultados.

Tabla 1. Resumen de trabajos sobre el reconocimiento automático de emociones centradas en el aprendizaje

Trabajo	(Mohamad Nezami et al., 2020)	(Gupta et al., 2016)	(González-Hernández et al., 2018)	(Mehmood & Lee, 2017)
<i>Emociones</i>	Interesado.	Aburrido, confundido, interesado, y frustrado.	Frustración, interés, excitación, aburrimiento, y relajación.	Felicidad, calma, tristeza, y miedo.
<i>Base de datos</i>	BD FER de emociones básicas con 35,887 imágenes de video y BD propia de ECA de 22 alumnos de secundaria.	BD propia DAiSEE de ECA con 9068 videos de 112 estudiantes.	BD RaFD de emociones básicas actuadas con 1146 fotos y dos bases de datos propias de ECA: BD E con 730 y la BD I con 5056 fotogramas de 38 estudiantes.	No utilizan una base de datos, la identificación la hacen en tiempo real por estudiante.
<i>Tecnologías utilizadas</i>	Video cámara.	Cámara web HD.	Cámara web y Diadema Emotiv.	Diademas para EEG (<i>Emotiv-EPOC System</i>).
<i>Método de etiquetamiento</i>	No aplica.	Observadores humanos (psicólogos).	Utilizando el software de la diadema Emotiv.	Observadores humanos.
<i>Método de extracción de características</i>	Capas convolucionales.	Capas convolucionales.	Red convolucional con 3 capas.	Seleccionan tres tipos de características: 1) Datos de frecuencia filtrados, 2) Parámetros Hjort, con 14 señales cerebrales, y 3) Parámetros Hjort con 6 señales cerebrales.
<i>Algoritmo de clasificación</i>	CNN tipo VGG-B.	CNN <i>Inception v3</i> , CNN 3D y CNN <i>Long-Term</i> .	CNN.	SVM y KNN.
<i>Tasa de reconocimiento</i>	Exactitud para interesado: 72.38%	Exactitud por emoción. Interesado: 51.07%, aburrido: 35.89%, confundido: 57.45%, y frustrado: 73.09%	Con la BD RaFD exactitud 95% y con las otras BDs exactitud de 88% y 74% respectivamente.	Presenta valores de pertenencia para cada una de las emociones identificadas.
<i>Actividad de aprendizaje</i>	Utilizan el software OMOSA para aprendizaje de habilidades de investigación.	Video educativo y video recreativo en línea.	Escribir, compilar, y ejecutar programas en JAVA.	Sistema de educación asistido por computadora para estudiantes con discapacidad física.
Trabajo	(Zatarain-Cabada et al., 2017a)	(Zatarain-Cabada et al., 2017b)	(Arana-Llanes et al., 2018)	(Botelho et al., 2017)
<i>Emociones</i>	Frustración, aburrimiento,	Frustración, aburrimiento,	Atención y concentración.	Frustrado, confundido,

	interés, y excitación.	interés, y excitación.		interesado, aburrido e imposible de codificar.
<i>Base de datos</i>	BD propia de ECA con 730 fotogramas.	BD propia de ECA con 730 fotogramas.	Base de datos propia con aproximadamente 23,072 instancias.	Base de datos propia con datos de 646 alumnos.
<i>Tecnologías utilizadas</i>	Webcam C920 Logitech HD Pro y diadema de EEG (Emotiv-EPOC).	Webcam C920 Logitech HD Pro y diadema de EEG (Emotiv-EPOC).	Video cámara y diadema de EEG (Emotiv-EPOC).	Ningún tipo de tecnología.
<i>Método de etiquetamiento</i>	Sistema Emotiv y el juicio humano.	Sistema Emotiv y el juicio humano.	No aplica.	Observadores humanos.
<i>Método de extracción de características</i>	Filtrado de las imágenes, operador uniforme del patrón local binario (PLB).	Técnica basada en geometría (coordenadas de los puntos relevantes, distancias y ángulos relativos).	Extracción de ondas Alpha y Beta.	51 características nivel-acción obtenidas de la interacción con el sistema y funciones estadísticas.
<i>Algoritmo de clasificación</i>	SVM.	SVM.	<i>K-means</i> y <i>clustering</i> .	Redes neuronales recurrentes tradicionales (RNN), red neuronal de unidad recurrente cerrada (GRU), y red de memoria de término largo-corto (LSTM).
<i>Tasa de reconocimiento</i>	Exactitud general: 80%. Exactitud por emoción: Aburrido: 64%, interesado: 64% frustración: 62%, excitación: 83%.	Exactitud por emoción: Aburrido: 64%; interesado: 64%, frustración: 62%, excitación: 83%.	Porcentaje total de inducción de atención/concentración del 84% al 99%.	Área bajo la curva <i>ROC</i> = 0.78% con RNN, para <i>Cohen's kappa</i> = 0.21% con LSTM y para <i>Fleiss kappa</i> = 0.27% con LSTM.
<i>Actividad de aprendizaje</i>	Medio ambiente de aprendizaje para programar en Java (Java Sensei).	Medio ambiente de aprendizaje para programar en Java (Java Sensei).	Test psicológicos para inducir estados mentales de atención y concentración.	Plataforma de aprendizaje en línea <i>ASSISTments</i> .
Trabajo	(Bosch et al., 2016a)	(Bosch et al., 2016b)	(Monkaresi et al., 2017)	(Barrón-Estrada et al., 2016)
<i>Emociones</i>	Aburrimiento, confusión, placer, frustración, interés.	Aburrimiento, confusión, placer, frustración, interés.	Interés.	Frustración, aburrimiento, interés, y meditación.
<i>Base de datos</i>	Base de datos propia de 137 estudiantes.	Base de datos propia de 137 estudiantes.	Base de datos propia de 22 estudiantes.	Base de datos propia de 8 estudiantes.
<i>Tecnologías utilizadas</i>	Webcam.	Webcam.	Video del Kinect para windows y sistema de adquisición de señal de electrocardiograma BIOPAC MP150.	Webcam C920 Logitech HD Pro y diadema de EEG (Emotiv-EPOC).
<i>Método de etiquetamiento</i>	Método de observación BROMP por etiquetadores humanos.	Método de observación BROMP por etiquetadores humanos.	Auto etiquetamiento cada dos minutos y una semana después.	Sistema Emotiv.
<i>Método de extracción de características</i>	Algoritmo RELIEF-F sobre: AUs, orientación y posición de la cabeza y	Algoritmo RELIEF-F sobre: AUs, orientación y posición de la cabeza y	Algoritmo RELIEF-F sobre: movimiento de labios, cejas y cabeza y, LBP-TOP (patrones binarios en tres planos	No lo especifica.

	estimación de los movimientos del cuerpo.	estimación de los movimientos del cuerpo.	ortogonales) sobre la textura y movimiento de ojos, boca, y de la frecuencia cardíaca.	
Algoritmo de clasificación	Árboles C4.5, <i>k-means</i> , y el clasificador Bayesiano.	Utilizan 14 clasificadores entre ellos: bayesianos, regresión logística, <i>clustering</i> (con <i>k-means</i>), árboles C4.5, etc.	Naïve Bayes, red bayesiana, regresión lineal, clusterización, bosque de rotación y ensamble de algoritmos.	Algoritmo de Patrón Local Binario.
Tasa de reconocimiento	Aburrimiento 64% (<i>k-means</i>), confusión 74% (Bayes net), placer 83% (Naïve Bayes), interés 64% (Bayes net), frustración 62% (Bayes net).	Aburrimiento 64% (<i>k-means</i>), confusión 74% (Bayes net), placer 83% (Naïve Bayes), interés 64% (Bayes net), frustración 62% (Bayes net).	Área bajo la curva ROC de 0.758, para anotaciones concurrentes y 0.733 para anotaciones de retrospectiva.	Precisión general de 86.95%.
Actividad de aprendizaje	Juego educativo <i>Physics Playground</i> .	Juego educativo <i>Physics Playground</i> .	Escritura de un resumen.	Sistema Tutorial Inteligente.

Como conclusión podemos observar que la mayoría de estos trabajos tratan de reconocer las principales emociones centradas en el aprendizaje: frustración, interés, aburrimiento, confusión, excitación, y sorpresa. Los algoritmos más utilizados para este propósito son: máquina de vectores de soporte (SVM), regresión lineal, y ensambles de algoritmos. El porcentaje de exactitud de reconocimiento varía desde el 62% al 88%, dependiendo de las métricas utilizadas para evaluar sus resultados.

La gran mayoría de trabajos generan sus propias bases de datos con un número de alumnos que van desde 20 hasta 646 (en una investigación libre de sensores), algunos otros utilizan bases de datos actuadas como RAFD, JAFFEE, o Grimace. Las tecnologías de adquisición de datos que predominan son las cámaras web y las diademas Emotiv, algunos otros trabajos hacen uso del Kinect de Windows, aplicación de detección del ritmo cardíaco, sensores de silla, y pulseras cardiovasculares.

4 Revisión del estado del arte de bases de datos de expresiones faciales

En el análisis del estado del arte sobre bases de datos de expresiones faciales presentamos dos apartados divididos por el tipo de expresiones faciales que utilizan para su creación. Primero nos enfocamos en las bases de datos de expresiones faciales de emociones básicas y posteriormente abordamos las de emociones secundarias, específicamente de ECA.

4.1 Bases de datos de expresiones faciales de emociones básicas

Aunque nuestro interés recae en bases de datos híbridas que contengan imágenes faciales y datos fisiológicos, hasta el momento solo hemos encontrado disponible públicamente una base de datos de este tipo. La base de datos DEAP (Koelstra et al., 2012), contiene grabaciones fisiológicas (de EEG) y video facial de un experimento donde 32 voluntarios vieron un subconjunto de 40 videos musicales. En este experimento también se les pide a los participantes calificar cada video de acuerdo con la emoción que provocaba en ellos.

El análisis de las bases de datos mostrado en la [Tabla 2](#) incluye solamente bases de datos de expresiones faciales disponibles al público. Estas bases de datos corresponden a expresiones faciales de emociones básicas, que se han capturado mientras las personas realizan diversas actividades. La mayoría de las bases de datos se componen de espectadores viendo contenido de medios (es decir, anuncios, avances de películas, programas de televisión, gifs animados, y campañas virales en línea) y de expresiones faciales

actuadas. De ellas la base de datos más robusta por la cantidad de datos que almacena es *Afectiva* (el Kaliouby & Picard, 2019).

Tabla 2. Bases de datos de expresiones de emociones básicas disponibles públicamente

Base de datos	Expresión facial	Número de muestras	Número de imágenes / videos	Tipo
DEAP (Koelstra et al., 2012).	Análisis de emociones del modelo continuo.	32	Señales fisiológicas periféricas de EEG y 22 videos de la cara.	Espontánea.
RAVDESS (Livingstone & Russo, 2018).	Voz y canto: Calma, Felicidad, tristeza, miedo, enojo, sorpresa, disgusto, y neutral.	24	7356 archivos de video y audio.	Posada.
F-M FACS 3.0 (EDU, PRO & XYZ versions) (Freitas-Magalhães, 2021).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, disgusto, desprecio.	10	4877 videos y secuencias de imágenes.	Espontánea / Posada.
CK ++ (Lucey et al., 2010).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, y disgusto.	123	593 secuencias de imágenes (327 con etiquetas de emociones secundarias).	Posada / sonrisa espontánea.
JAFFE (Lyons et al., 2014).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, y disgusto.	10	213 imágenes estáticas.	Posada.
MMI (Valstar & Pantic, 2010).	Disgusto, felicidad, y tristeza.	43	1280 videos y 250 imágenes.	Espontánea / Posada.
BELFAST (Sneddon et al., 2012)	Conjunto 1 (disgusto, miedo, frustración, sorpresa, diversión).	114	570 video clips.	Emociones naturales.
	Conjunto 2 (disgusto, miedo, frustración, sorpresa, diversión, enojo, y tristeza).	82	650 video clips.	
DISFA (Mavadati et al., 2013).	Intensidad de las AUs.	27	4845 frames de video.	Espontánea.
MUG (Aifanti et al., 2010)	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, y disgusto.	82	1462 secuencias.	Posada.
ISED (Happy et al., 2017)	Tristeza y felicidad.	50	428 videos.	Espontánea.
RaFD (Langner et al., 2010).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, disgusto, y desprecio.	67	3 direcciones diferentes de la mirada y 5 ángulos de la cámara (8,040 imágenes).	Actuada.
Oulu-CASIA NIR-VIS (Zhao et al., 2011).	Tristeza, sorpresa, felicidad, miedo, y disgusto.	80	3 condiciones de iluminación diferentes: normal, clara, y oscura (2880 videos)	Actuada.
FERG (Aneja et al., 2017).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, disgusto, y alegría.	6	55,767 imágenes.	Pose frontal.
Affectnet (Mollahosseini et al., 2019).	Neutral, tristeza, sorpresa, felicidad, miedo, enojo, disgusto, y desprecio.	-	~450,000 anotaciones manuales ~500,000 anotaciones automáticas.	Espontánea.

Base de datos	Expresión facial	Número de muestras	Número de imágenes / videos	Tipo
IMPA-FACE3D (Mena-Chalco et al., 2008).	Neutral, tristeza, sorpresa, diversión, enojado, disgustado, abierto, cerrado, y beso.	38	534 imágenes estáticas.	Actuada.
FEI (Thomaz, 2012).	Neutral y sonriente.	200	2850 imágenes estáticas.	Actuada.
Aff-Wild (Zafeiriou et al., 2017).	-	200	~ 1,250,000 etiquetadas manualmente.	Espontánea.
Afectiva (el Kaliouby & Picard, 2019).	Frustración, sorpresa, diversión, enojo, tristeza, miedo, disgusto, y desprecio.	-	7,860,463 caras analizadas.	Espontánea.

4.2 Bases de datos de expresiones faciales de emociones centradas en el aprendizaje

En la [Tabla 3](#) se muestran los detalles de bases de datos de expresiones faciales espontáneas. Estas corresponden a emociones centradas en el aprendizaje que han sido capturadas mientras los estudiantes realizan alguna actividad de aprendizaje en un entorno natural. Algunas de ellas contienen datos fisiológicos, pero desafortunadamente no están disponibles al público, ni proporcionan información sobre características de los datos, ni mucho menos detalles del procesamiento que hacen con ellos.

Tabla 3. Bases de datos de expresiones de ECA.

Base de datos	Expresión facial	Número de muestras	Número de imágenes / videos	Tipo/Software
<i>Engagement Recognition</i> (ER), base de datos propia (Mohamad Nezami et al., 2020).	Interés.	20 estudiantes de secundaria (11 mujeres, 9 hombres).	20 videos de una hora cada uno.	Espontánea/Mundo virtual OMOSA para aprendizaje de habilidades de investigación.
DAiSEE, base de datos propia (Gupta et al., 2016).	Interés, aburrimiento, frustración, confusión.	112 estudiantes.	9068 videos.	Espontánea/Aprendizaje en línea.
Base de datos propia (Nye et al., 2017).	Frustración, confusión, diversión, tristeza, miedo, angustia.	39 estudiantes.	30 a 60 minutos de video por estudiante.	Espontánea/Sistema tutor inteligente.
Base de datos propia (Xiao et al., 2017).	Frustración, confusión, complacencia, interés, aburrimiento, sorpresa, curiosidad, felicidad, y neutral.	22 estudiantes universitarios.	Imágenes.	Espontánea/Sistema administrador de aprendizaje.
Base de datos propia (Barrón-Estrada et al., 2016; Zatarain-Cabada et al., 2017a, 2017b).	Interés, aburrimiento, frustración, y excitación.	8 estudiantes.	Video y señales de EGG (etiquetas de emociones discretas).	Espontánea/Sistema tutor inteligente.
Base de datos propia (Bosch & D'Mello, 2017).	Interés, aburrimiento, frustración, y confusión.	29 estudiantes.	Video y auto reportes.	Espontánea/Ambiente para programar en Python.

Base de datos	Expresión facial	Número de muestras	Número de imágenes / videos	Tipo/Software
Base de datos propia (Harley et al., 2013).	Emociones básicas y ECA (19 emociones y 1 un estado neutral).	67 estudiantes.	Video y auto reportes.	Espontánea/Sistema tutor inteligente multi agente para enseñar el sistema circulatorio humano.
Base de datos propia (Graesser & D'Mello, 2012).	Interés, aburrimiento, frustración, y confusión.	28–30 estudiantes.	Video de movimiento de ojos, notas de diálogos con el tutor, expresiones faciales, posturas del cuerpo, sensor de presión de mouse y teclado, y voz.	Espontánea/Sistema tutor inteligente para reparar computadoras.
Base de datos propia (Arana-Llanes et al., 2018).	Atención y concentración.	23,072 instancias.	Video y señales de EGG (Emotiv-EPOC). Extracción de ondas alfa y beta.	Espontánea/Test psicológico para inducir los estados mentales de atención y concentración.
Base de datos propia (Bosch et al., 2016, 2016; Bosch & D'Mello, 2017).	Complacencia, frustración, confusión, interés, y aburrimiento.	137 estudiantes.	Video y método de observación BROMP.	Espontánea/Juego educativo de física <i>Playground</i> .
Base de datos propia (Monkaresi et al., 2017).	Interés.	22 estudiantes.	Video de Kinect y Sistema de adquisición de señales BIOPAC MP150 de ECG.	Espontánea/ Escritura de un resumen.
Base de datos propia (Almohammadi et al., 2017).	Interés.	30 estudiantes.	Video de Kinect V2.	Espontánea/ curso de excel.
Base de datos propia (Bixler & D'Mello, 2013).	Interés, aburrimiento, neutral.	44 estudiantes.	Análisis de teclado y video.	Espontánea/ Escritura de un ensayo.
Base de datos propia (Arroyo et al., 2009).	Interés, excitación, confianza, y frustración.	2 grupos de 38 y 29 estudiantes de nivel medio superior.	Sensores fisiológicos (cámara de video, mouse, silla y brazalete).	Espontánea/Sistema tutor multimedia adaptivo de geometría.

De estas bases de datos cuatro incluyen datos fisiológicos. De ellas la más completa es la mencionada en Arroyo et al. (2009), en la que participan 67 estudiantes y utilizan tres sensores de datos fisiológicos y una cámara de video. En general la mayoría de estas bases de datos han sido creadas con datos de muy pocos participantes.

A partir de este análisis se identificó como área de oportunidad la creación de una base de datos fisiológicos y de comportamiento respaldada por un protocolo formal para la captura de los datos. La definición del protocolo nos permitió ejecutar un experimento controlado en un ambiente natural y que puede replicarse el número de veces necesarias para crear una base de datos robusta.

5 Conclusiones

Un análisis del estado del arte como el presentado en este trabajo da un panorama amplio sobre algún tema de interés a investigar. En este caso el objetivo es contribuir en el área de cómputo afectivo, específicamente en el reconocimiento de ECA. Los trabajos analizados nos permiten plantear estudios similares utilizando ambientes de aprendizaje bajo configuraciones diferentes priorizando la naturalidad y comodidad de los

estudiantes dando la posibilidad de generar nuevas propuestas para abordar problemas similares. Aunque existen avances significativos en el área aún hay mucho por hacer. Después de este análisis, podemos identificar áreas de oportunidad con el objetivo de lograr mejores precisiones de reconocimiento utilizando otros dispositivos de captura de datos fisiológicos como las cámaras térmicas, la captura del ritmo cardíaco con sensor de dedo en combinación con grabaciones de video. Otro aspecto que falta cubrir en la creación de las bases de datos de ECA es poder capturar un número balanceado de muestras por clase con la finalidad de evitar sesgos durante los procesos de entrenamiento y clasificación. También el poder capturar datos de un mayor número de estudiantes, que garantice la existencia de todas las ECA.

Este análisis es útil también para discriminar y elegir entre diferentes aspectos a considerar en el reconocimiento automático de emociones como por ejemplo decidir los algoritmos de selección de características y clasificación, los tipos de emociones que se reconocerán, las formas de etiquetarlas, entre otros, esto con la finalidad de avanzar en el reconocimiento automático de emociones centradas en el aprendizaje y contribuir en esta área. Así con este análisis se identificó como área de oportunidad la creación de una base de datos fisiológicos y de comportamiento respaldada por un protocolo formal para la captura de los datos. La definición del protocolo deberá permitir la ejecución de un experimento controlado en un ambiente natural y que pueda replicarse el número de veces necesarias para crear una base de datos robusta.

Declaración de conflicto de intereses

Los autores declaran no tener conflicto de intereses con respecto a la investigación, autoría o publicación de este artículo.

Financiación

Consejo Nacional de Ciencia y Tecnológica, CONACYT México, Becas Nacionales para estudios de posgrado en Instituciones de Alta Calidad.

ORCID iD

Yesenia Nohemí González-Meneses  <https://orcid.org/0000-0003-1034-0204>

Josefina Guerrero-García  <https://orcid.org/0000-0002-3393-610X>

Referencias

- Aifanti, N., Papachristou, C., & Delopoulos, A. (2010). The MUG facial expression database. *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*, 1–4.
- Almohammadi, K., Hagrass, H., Yao, B., Alzahrani, A., Alghazzawi, D., & Aldabbagh, G. (2017). A type-2 fuzzy logic recommendation system for adaptive teaching. *Soft Computing*, 21(4). <https://doi.org/10.1007/s00500-015-1826-y>
- Aneja, D., Colburn, A., Faigin, G., Shapiro, L., & Mones, B. (2017). Modeling Stylized Character Expressions via Deep Learning. In *Computer Vision – ACCV 2016. ACCV 2016. Lecture Notes in Computer Science* (Vol. 10112). Springer, Cham. https://doi.org/10.1007/978-3-319-54184-6_9
- Arana-Llanes, J. Y., González-Serna, G., Pineda-Tapia, R., Olivares-Peregrino, V., Ricarte-Trives, J. J., & Latorre-Postigo, J. M. (2018). EEG lecture on recommended activities for the induction of attention and concentration mental states on e-learning students. *Journal of Intelligent & Fuzzy Systems*, 34(5). <https://doi.org/10.3233/JIFS-169517>
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *Artificial Intelligence in Education*, 17–24.
- Barrón-Estrada, M. L., Zatarain-Cabada, R., Aispuro-Medina, B. G., Valencia-Rodríguez, E. M., & Lara-Barrera, A. C. (2016). Building a Corpus of Facial Expressions for Learning-Centered Emotions. *Research in Computing Science*, 129, 45–52.

- Bixler, R., & D'Mello, S. (2013). Towards Automated Detection and Regulation of Affective States During Academic Writing. In *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science* (Vol. 7926). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_142
- Bosch, N., & D'Mello, S. (2017). The Affective Experience of Novice Computer Programmers. *International Journal of Artificial Intelligence in Education*, 27(1). <https://doi.org/10.1007/s40593-015-0069-5>
- Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2016a). Detecting student emotions in computer-enabled classrooms. *IJCAI International Joint Conference on Artificial Intelligence*, 4125–4129.
- Bosch, N., D'Mello, S. K., Ocumpaugh, J., Baker, R. S., & Shute, V. (2016b). Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms. *ACM Transactions on Interactive Intelligent Systems*, 6(2). <https://doi.org/10.1145/2946837>
- Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving Sensor-Free Affect Detection Using Deep Learning. In *Artificial Intelligence in Education. AIED 2017. Lecture Notes in Computer Science* (Vol. 10331). https://doi.org/10.1007/978-3-319-61425-0_4
- Cabada, R., Barrón, M., & Olivares, J. M. (2014). Reconocimiento automático y aspectos éticos de emociones para aplicaciones educativas. *Inteligencia Artificial: Una Reflexión Obligada*.
- Cornelius, R. R. (1996). *The science of emotion: Research and tradition in the psychology of emotions*. Prentice-Hall, Inc.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1). <https://doi.org/10.1109/79.911197>
- Ekman, P. (2004). *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Henry Holt and Company.
- Ekman, P., Friesen, W., & Hager, J. (2002). *Facial action coding system: Research nexus network research information*.
- el Kaliouby, R., & Picard, R. W. (2019). *Affectiva Database*. MIT Media Laboratory. <https://www.affectiva.com/>
- Freitas-Magalhães, A. (2021). *Facial Action Coding System 4.0-Manual of Scientific Codification of the Human Face*. Leya.
- Fuentes, C., Herskovic, V., Rodríguez, I., Gereá, C., Marques, M., & Rossel, P. O. (2017). A systematic literature review about technologies for self-reporting emotional information. *Journal of Ambient Intelligence and Humanized Computing*, 8(4). <https://doi.org/10.1007/s12652-016-0430-z>
- González Meneses, Y. N., Guerrero García, J., Reyes García, C. A., Olmos Pineda, I., & González Calleros, J. M. (2019). Methodology for Automatic Identification of Emotions in Learning Environments. *Research in Computing Science*, 148(5), 89–96. <https://doi.org/10.1088/0031-9112/29/6/013>
- González-Hernández, F., Zatarain-Cabada, R., Barrón-Estrada, M. L., & Rodríguez-Rangel, H. (2018). Recognition of learning-centered emotions using a convolutional neural network. *Journal of Intelligent & Fuzzy Systems*, 34(5). <https://doi.org/10.3233/JIFS-169514>
- Graesser, A. C., & D'Mello, S. (2012). Emotions During the Learning of Difficult Material. *Psychology of Learning and Motivation*, 57. <https://doi.org/10.1016/B978-0-12-394293-7.00005-4>
- Gupta, A., D'Cunha, A., Awasthi, K., & Balasubramanian, V. (2016). DAiSEE: Towards User Engagement Recognition in the Wild. *ArXiv Preprint*.
- Happy, S. L., Patnaik, P., Routray, A., & Guha, R. (2017). The Indian Spontaneous Expression Database for Emotion Recognition. *IEEE Transactions on Affective Computing*, 8(1). <https://doi.org/10.1109/TAFFC.2015.2498174>
- Harley, J. M., Bouchet, F., & Azevedo, R. (2013). Aligning and Comparing Data on Emotions Experienced during Learning with MetaTutor. In *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science* (Vol. 7926). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_7
- Hill, D. (2010). *Emotionomics: Leveraging emotions for business success*.
- Hjortsj, C. H., Ekman, P., Friesen, W. v., Hager, J. C., Facs, F., & Facs, F. M. (2019). *Sistema de Codificación Facial*.
- Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1). <https://doi.org/10.1109/T-AFFC.2011.15>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8). <https://doi.org/10.1080/02699930903485076>
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDSS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, 13(5). <https://doi.org/10.1371/journal.pone.0196391>
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., & Ambadar, Z. (2010). The extended Cohn-Kanade dataset (CK+): a complete facial expression dataset for action unit and emotion-specified expression Conference on. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference On, July*, 94–101.
- Lyons, M. J., Kamachi, M., & Gyoba, J. (2014). *Japanese Female Facial Expressions (JAFFE), Database of digital images*. http://www.kasrl.org/jaffe_info.html

- Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). DISFA: A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing*, 4(2). <https://doi.org/10.1109/T-AFFC.2013.4>
- Mehmood, R., & Lee, H. (2017). Towards Building a Computer Aided Education System for Special Students Using Wearable Sensor Technologies. *Sensors*, 17(317), 1–22. <https://doi.org/10.3390/s17020317>
- Mena-Chalco, J. P., Cesar-Jr, R., & Velho, L. (2008). Banco de dados de faces 3D: IMPA-FACE3D. *IMPA-RJ, Tech. Rep.*
- Mohamad Nezami, O., Dras, M., Hamey, L., Richards, D., Wan, S., & Paris, C. (2020). Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression. In *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2019. Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-030-46133-1_17
- Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10(1). <https://doi.org/10.1109/TAFFC.2017.2740923>
- Monkaresi, H., Bosch, N., Calvo, R. A., & D’Mello, S. K. (2017). Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate. *IEEE Transactions on Affective Computing*, 8(1). <https://doi.org/10.1109/TAFFC.2016.2515084>
- Nye, B., Karumbaiah, S., Tokel, S. T., Core, M. G., Stratou, G., Auerbach, D., & Georgila, K. (2017). Analyzing Learner Affect in a Scenario-Based Intelligent Tutoring System. In *Artificial Intelligence in Education. AIED 2017. Lecture Notes in Computer Science* (Vol. 10331). Springer, Cham. https://doi.org/10.1007/978-3-319-61425-0_60
- Picard, R. W. (2000). *Affective Computing*. MIT Press.
- Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1–2). [https://doi.org/10.1016/S1071-5819\(03\)00052-1](https://doi.org/10.1016/S1071-5819(03)00052-1)
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4). <https://doi.org/10.1177/0539018405058216>
- Sneddon, I., McRorie, M., McKeown, G., & Hanratty, J. (2012). The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, 3(1). <https://doi.org/10.1109/T-AFFC.2011.26>
- Steidl, S. (2009). *Automatic classification of emotion related user states in spontaneous children’s speech*. Logos-Verlag.
- Thomaz, C. E. (2012). *FEI Face Database*. <https://fei.edu.br/~cet/facedatabase.html>
- Valstar, M., & Pantic, M. (2010). Induced disgust, happiness and surprise: an addition to the MMI facial expression database. *Proc. 3rd Intern. Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect*.
- Xiao, X., Pham, P., & Wang, J. (2017). Dynamics of Affective States During MOOC Learning. In *Artificial Intelligence in Education. AIED 2017. Lecture Notes in Computer Science* (Vol. 10331). Springer, Cham. https://doi.org/10.1007/978-3-319-61425-0_70
- Zafeiriou, S., Kollias, D., Nicolaou, M. A., Papaioannou, A., Zhao, G., & Kotsia, I. (2017). Aff-wild: valence and arousal In-the-Wild challenge. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 34–41.
- Zatarain-Cabada, R., Barrón-Estrada, M. L., González-Hernández, F., Oramas-Bustillos, R., Alor-Hernández, G., & Reyes-García, C. A. (2017b). Building a Corpus and a Local Binary Pattern Recognizer for Learning-Centered Emotions. In *Advances in Soft Computing. MICAI 2016. Lecture Notes in Computer Science* (Vol. 10062). https://doi.org/10.1007/978-3-319-62428-0_43
- Zatarain-Cabada, R., Barron-Estrada, M. L., Gonzalez-Hernandez, F., & Rodriguez-Rangel, H. (2017a). Building a Face Expression Recognizer and a Face Expression Database for an Intelligent Tutoring System. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. <https://doi.org/10.1109/ICALT.2017.141>
- Zhao, G., Huang, X., Taini, M., Li, S. Z., & Pietikäinen, M. (2011). Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9). <https://doi.org/10.1016/j.imavis.2011.07.002>