

SEP

TNM

INSTITUTO TECNOLÓGICO DE CULIACÁN



Reconocimiento de emociones centradas en el aprendizaje por medio de expresiones faciales

TESIS

PRESENTADA ANTE EL DEPARTAMENTO ACADÉMICO DE ESTUDIOS DE POSGRADO DEL INSTITUTO TECNOLÓGICO DE CULIACÁN EN CUMPLIMIENTO PARCIAL DE LOS REQUISITOS PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

POR:

Leal Hernández Daniel

LIC. EN INGENIERIA EN INFORMÁTICA

DIRECTOR DE TESIS:

Dr. Ramón Zatarain Cabada

CULIACÁN, SINALOA

SEPTIEMBRE, 2017



DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Culiacán, Sin., 26 de Septiembre del 2017

OFICIO: DEPI/530/09/2017
ASUNTO: **Autorización Impresión**

ING. DANIEL LEAL HERNÁNDEZ
ESTUDIANTE DE LA MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE.

Por medio de la presente y en virtud de que ha completado los requisitos para el examen de grado de la **Maestría en Ciencias de la Computación**, se concede autorización para la impresión de la tesis titulada: **"RECONOCIMIENTO DE EMOCIONES CENTRADAS EN EL APRENDIZAJE POR MEDIO DE EXPRESIONES FACIALES"**, bajo la dirección del(a) **Dr. Ramón Zatarain Cabada**

Sin otro particular reciba un cordial saludo.

ATENTAMENTE
"CON LA TÉCNICA AL PROGRESO"

M.C. MARÍA ARACELY MARTÍNEZ AMAYA
JEFA DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN

SEP TecNM
Instituto Tecnológico
de Culiacán
División de Estudios
de Posgrado e Investigación

C.c.p. archivo
DNU/lucy *



Juan de Dios Bátiz 310 pte. Col. Guadalupe, C.P. 80220
Culiacán, Sinaloa, Tels. 713-17-96, 713-38-04 y 713-86-09 Fax: 716-96-47
www.itculiacan.edu.mx



**“RECONOCIMIENTO DE EMOCIONES
CENTRADAS EN EL APRENDIZAJE POR MEDIO DE
EXPRESIONES FACIALES”**

Tesis presentada por:

ING. DANIEL LEAL HERNÁNDEZ

Aprobada en contenido y estilo por:



Dr. Ramón Zatarain Cabada
Director de Tesis



Dra. María Lucía Barrón Estrada
Secretario



M.C. Rosalío Zatarain Cabada
Vocal -1



M.C. Gloria Ekaterine Peralta Peñuñuri
Vocal -2



M.C. María Aracely Martínez Amaya
Jefe de la División de Estudios de
Posgrado, Investigación

Contenido

Lista de Figuras	viii
Lista de Tablas.....	x
Palabras Clave	xi
Agradecimiento	xii
Dedicatoria.....	xiii
Resumen	1
1. Introducción.....	2
1.1 Descripción del Problema	2
1.2 Objetivo General.....	2
1.3 Objetivos Específicos	2
1.4 Hipótesis	3
1.5 Justificación	3
1.6 Organización de la Tesis.....	3
2. Marco Teórico	5
2.1 Corpus.....	5
2.2 Computación Afectiva	6
2.2.1 Teoría de la emoción	6
2.2.2 Clasificación de emoción.....	7
2.2.3 Expresiones de la emoción	8
2.2.4 Dispositivos de reconocimiento de emoción	8
2.2.5 Emociones centradas en el aprendizaje	9
2.3 Sistemas Reconocedores de Emociones	10
Reconocimiento de expresiones faciales	10

	Reconocimiento de voz	11
	Reconocimiento de señales EEG	11
2.4	Sistemas Tutores Inteligentes	11
2.3.2	Arquitectura de un STI	11
2.3.3		
3.	Estado del Arte	12
2.43.1	Diversos corpus de imágenes o video.....	12
	Presentación y validación de la base de datos de rostros RadBound (RaFD)	12
	DISFA una base de datos espontanea de intensidad de acción facial	14
3.1.1		
3.1.2	Base de datos de rostros PUT	14
3.1.3	Base de datos BP4D espontanea.....	15
3.1.4	Base de datos de NVIE.....	16
3.1.5		
3.1.6	Base de datos Cohn-Kanade extendida (CK+).....	17
3.1.7	Comparación de bases de datos de expresiones faciales.	17
3.2	Sistemas reconocedores de emociones	18
3.2.1	Reconocimiento automático de expresiones faciales para predecir la emoción	
3.2.2	comprometido y frustrado.	18
3.2.3	Detección de emociones en escenarios educativos.....	19
3.2.4	Los rostros de enganchado: Reconocimiento automático de enganchado en	
3.2.5	estudiantes por medio de expresiones faciales	21
	Reconocimiento de expresiones en BDBN.	23
3.2.6	Reconocimiento automático de expresiones faciales utilizando características	
	de regiones faciales.....	24
	Reconocimiento de emociones a partir de características faciales geométricas	
	utilizando mapas propiamente organizados.....	27
3.3	Sistemas tutores afectivos	29

	Predicción de afecto a partir de la mirada durante la interacción con un sistema tutor inteligente.....	29
	TuinLEC, un tutor inteligente para mejorar la competencia lectora.	31
3.3.1	Ambiente inteligente de aprendizaje con manejo afectivo para Java.	32
3.3.2	Construcción de un reconocedor de expresiones faciales y una base de datos de expresiones faciales para un sistema tutor inteligente.....	34
3.3.3		
3.4	Desarrollo del Proyecto	36
4.1	Protocolo de sesión	36
4.2	Equipos de captura de datos.....	37
4.3	Metodología.....	37
4.4	Procesamiento de datos.....	40
4.5	Reconocimiento de emociones centradas en el aprendizaje	43
4.5.1	Análisis	43
4.5.2	Diseño.....	43
4.5.3	Desarrollo	44
5.	Pruebas y resultados	46
5.1	Análisis estadístico de los datos.....	47
5.2	Tabla comparativa de las diferentes versiones del corpus	48
5.3	Resultados de la versión 1 y 2 extendido.....	49
5.4	Tabla comparativa de resultados.....	49
5.5	Otros resultados	50
6	Conclusiones y Trabajos Futuros	50
6.1	Conclusiones.....	50
6.2	Dificultades de la investigación.....	50
6.3	Aportaciones	51

6.4	Trabajo Futuro	51
7	Referencias	52

Lista de Figuras

Figura 1. Metodología del corpus.....	5
Figura 2. Emociones básicas.	7
Figura 3. Modelo de activación y valencia tomada de (Christie & Friedman, 2004).	8
Figura 4. Dispositivos de captura de datos.	9
Figura 5. Modelo de las emociones del aprendizaje tomada de (Kort et al., 2001).	10
Figura 6. Arquitectura de un sistema tutor inteligente tomada de (Akkila & Naser, 2017).	12
Figura 7. Estudio de sesiones tomada de (Langner et al., 2010).	13
Figura 8. Estudio DISFA tomada de (Mavadati, Mahoor, Bartlett, Trinh, & Cohn, 2013).	14
Figura 9. Imágenes de la base de datos PUT tomada de (Kasinski et al., 2008).	15
Figura 10. Diversas vistas tomada de (Zhang et al., 2014).	16
Figura 11. Base de datos CK y CK+ tomada de (Lucey et al., 2010).	17
Figura 12. Área de sesiones tomada de (Grafsgaard et al., 2013).	19
Figura 13. Experimento tomada de (Saneiro et al., 2014).	21
Figura 14. Imágenes del experimento con escala tomada de (Whitehill et al., 2014).	22
Figura 15. Vista general del marco de trabajo BDBN tomada de (P. Liu et al., 2014).	23
Figura 16. Vista general del sistema propuesto basados en el de (Happy & Routray, 2015)	25
Figura 17. Regiones faciales tomada de (Happy & Routray, 2015).	26
Figura 18. Modelo de 23 puntos faciales tomada de (Majumder et al., 2014)	28
Figura 19. Extracción de características tomada de (Majumder et al., 2014)	28
Figura 20. Interfaz Meta tutor tomada de (Jaques et al., 2014).	30
Figura 21. Sistema tutor inteligente TuinLEC. tomada de (Vidal-Abarca et al., 2014).	31
Figura 22. Sistema lógico difuso y red neuronal tomada de (Zatarain-Cabada, Barrón- Estrada, González-Hernández, & Oramas-Bustillos, 2015).	33

Figura 23. Motor de Java Sensei tomada de (Cabada et al., 2015).	34
Figura 24. Puntos faciales tomada de (Zatarain-Cabada et al., 2017).	35
Figura 25. Dispositivos utilizados para las sesiones. Todos los derechos reservados.....	37
Figura 26. Plataforma web de aprendizaje del lenguaje Java Gaming ILE.....	38
Figura 27. Interfaz Emotiv insight.....	39
Figura 28. Presentación de señales EEG de Emociones en Emotiv.	39
Figura 29. señales de emociones centradas en el aprendizaje.	40
Figura 30. Análisis de error en las señales.	41
Figura 31. Ejemplo de imágenes almacenadas en el corpus.....	41
Figura 32. Proceso de etiquetado de imágenes.....	42
Figura 33. Sistema de reconocimiento de emociones.....	44
Figura 34. Aplicación de filtros de grises.....	45
Figura 35. Puntos faciales en el rostro.....	45

Lista de Tablas

Tabla 1. Comparativa de bases de datos.....	18
Tabla 2. Resultados comparados con otras máquinas de aprendizaje tomada de (P. Liu et al., 2014).....	24
Tabla 3. Resultados con JAFFE tomada de (P. Liu et al., 2014).....	24
Tabla 4. Resultados de la base de datos CK + tomada de (Happy & Routray, 2015).....	26
Tabla 5. Resultados de la base de datos JAFFE tomada de (Happy & Routray, 2015).	27
Tabla 6. Resultados obtenidos de (Majumder et al., 2014).....	29
Tabla 7. Resultados obtenidos de (Zatarain-Cabada et al., 2017).....	36
Tabla 8. Número de imágenes obtenidas de cada algoritmo de clasificación.	42
Tabla 9. Matriz de características.	46
Tabla 10. Algoritmo de clasificación número 1.	47
Tabla 11. Algoritmo de clasificación número 2.	47
Tabla 12. Resultados del coeficiente de correlación de Pearson.....	48
Tabla 13. Corpus resultante.....	49
Tabla 14. Resultados de los algoritmos extendidos.....	49
Tabla 15. Resultados de los diferentes corpus.....	49
Tabla 16. Tabla comparativa entre Emotiv y el reconocedor.....	50

Palabras Clave

Computación afectiva

Inteligencia artificial

Detección de emociones en rostro

Expresiones faciales

Señales EEG

Agradecimiento

Agradezco al Instituto Tecnológico de Culiacán por abrirme las puertas de estudiar un posgrado de calidad y pertenecer a una gran institución para seguirme preparando.

Agradezco a mi director de tesis el Dr. Zatarain Cabada Ramón quien me ha apoyado a realizar con éxito mis estudios de maestría.

A la Dra. Barrón Estrada María Lucia, el Dr. Quintero Meza Ricardo Rafael y el M.C.C. Manjarrez Montero Emir Abel quienes me apoyaron en todo momento y compartieron su conocimiento para culminar con éxito mis estudios de maestría.

A mis compañeros de generación Valencia Rodríguez Elvia Minerva, Aispuro Medina Bianca Giovanna, Lara Barrera Ana Cecilia, Quintero Ibarra Fernando, Lindor Valdez Cristian Jair y García Carrillo Noel Andrés, con quienes compartí diferentes momentos durante mis estudios de maestría, su apoyo, comunicación e hice una buena amistad con cada uno de ellos.

Por último, agradezco a la entidad CONACYT quien me brindo el apoyo económico, que hizo posible culminar con mis estudios de posgrado.

Dedicatoria

Dedico este trabajo de tesis principalmente a mi madre la Sra. Hernández Coronado Ana Beatriz quien siempre ha sido mi motor e impulso para seguir superándome y apoyarme en cada una de las decisiones que eh tomado.

A mis hermanos Leal Hernández Ana Karina y Leal Hernández Rafael quienes me han apoyo en todo momento para culminar mis estudios de maestría.

Finalmente agradezco a mi hijo Leal Xaman Diego quien es la fuente de inspiración para seguirme preparando cada día.

Resumen

El presente trabajo de tesis es importante en el ámbito educativo porque ayuda a los ambientes inteligentes de aprendizaje a evaluar aspectos emociones enfocadas al aprendizaje, mediante dos aportaciones: la primera es una base de datos de imágenes (Corpus) enfocadas en emociones del aprendizaje que son: interesado, comprometido, aburrido, enfocado, relajado y emocionado, la cual se construyó mediante la aplicación de un protocolo de sesión donde participaron estudiantes de licenciatura y de posgrado del instituto tecnológico de Culiacán, se siguió una metodología para diseñar y construir la base de datos (Corpus) de imágenes espontaneas, en la cual se aplicaron diversos algoritmos para seleccionar y clasificar cada imagen adquirida. La segunda aportación es un sistema reconocedor de emociones con cualidades de reconocimiento facial mediante técnicas no invasivas, aplicación de técnicas para reconocimiento, procesamiento de características faciales y la clasificación de emociones enfocadas en el aprendizaje, el cual se puede integrar a sistemas tutores inteligentes para que evalúe aspectos emocionales y no solamente aspectos cognitivos. El sistema reconocedor de emociones logra una precisión de 69.93% de precisión para clasificar emociones. El proyecto surge del hecho de que no existe una base de datos de imágenes enfocadas en emociones del aprendizaje que sean espontaneas, la mayoría de las bases de datos de imágenes están enfocadas en emociones básicas y algunas son actuadas, los diversos sistemas reconocedores de emociones están enfocados en estas emociones, como son: felicidad, miedo, sorpresa, tristeza, enojo y disgusto. Otro tipo de reconocedores se enfocan en emociones no básicas también llamadas secundarias como son culpa, vergüenza, celos y orgullo. Sin embargo, existe muy poco trabajo de investigación en lo que respecta a reconocimiento de emociones enfocadas en el aprendizaje. Este trabajo de tesis sienta las bases para continuar con este tipo de emociones específicas en el ámbito educativo y los estudiantes puedan adquirir mejorar en él aprendizaje.

1. Introducción.

En este primer capítulo se presenta la descripción del problema, los objetivos generales y específicos del proyecto, así como el planteamiento de la hipótesis, la justificación y por último una breve descripción de la organización del documento de tesis.

1.1 Descripción del Problema

Hoy en día, la mayoría de los corpus de expresiones faciales están enfocados en las emociones básicas de Paul Ekman (Ekman, 1992), las cuales expresan estados afectivos como felicidad, sorpresa, miedo, tristeza, enojo y disgusto. Es difícil encontrar corpus de imágenes faciales representando emociones secundarias enfocadas en el aprendizaje como aburrido, enganchado, emocionado, enfocado, interesado y relajado.

El principal problema, es que se requiere que los sistemas reconocedores de emociones detecten emociones enfocadas al aprendizaje, con el objetivo de ayudar a los sistemas tutores inteligentes a evaluar aspectos emocionales y no solo procesos cognitivos.

1.2 Objetivo General

Desarrollar un corpus de expresiones faciales enfocado en emociones centradas en el aprendizaje, para entrenar un sistema reconocedor de emociones que permita clasificar emociones y evaluar la precisión del sistema reconocedor.

1.3 Objetivos Específicos

- Diseñar e implementar software para identificar un rostro en una imagen y extraer características importantes de este (e.g. ojos, cejas, boca).
- Identificar, analizar, documentar y construir un corpus de expresiones faciales enfocado en las emociones que se presentan durante el aprendizaje.
- Analizar, diseñar y desarrollar un sistema reconocedor de emociones centradas en el aprendizaje.
- Evaluar el sistema reconocedor de emociones mediante el uso del algoritmo de validación cruzada de diez para obtener una precisión de reconocimiento.

- Evaluar en tiempo real el sistema reconocedor de emociones junto con el reconocimiento de señales cerebrales EEG y procesadas a emociones del aprendizaje mediante el uso de software y dispositivo Emotiv EPOC Insight.

1.4 Hipótesis

En este trabajo se definieron dos hipótesis importantes. Una enfocada al uso del corpus en los clasificadores y otra enfocada al uso de los reconocedores de emociones en ambientes de aprendizaje. Las hipótesis son las siguientes:

- El corpus de emociones centradas en el aprendizaje podrá ser usado para entrenar clasificadores de emociones obteniendo tasas de asertividad mayores de 70 % de éxito.
- El reconocedor de emociones centradas en el aprendizaje podrá integrarse a ambientes de aprendizaje ayudando a mejorar la tasa de aprendizaje de los estudiantes.

1.5 Justificación

El presente trabajo es relevante en ámbitos de aprendizaje porque ayuda a los sistemas tutores inteligentes a entrenar sus respectivos sistemas reconocedores de emociones, mediante el uso del corpus de expresiones faciales espontáneas, enfocado directamente en el aprendizaje y no en emociones básicas.

Además, el desarrollo de un sistema reconocedor de emociones centradas en el aprendizaje, permite integrarlo a los ambientes de aprendizaje que no cuentan con uno, con el objetivo de evaluar los aspectos emocionales del estudiante durante el aprendizaje y en caso de no contar con un sistema reconocedor se pueda implementar este.

1.6 Organización de la Tesis

A través de este documento de tesis, se describe el proceso realizado para la construcción de un corpus de imágenes de expresiones faciales espontáneas enfocada en el aprendizaje, también se describe el desarrollo de un sistema reconocedor de emociones. El documento se organiza en 5 capítulos principales, descritos a continuación.

En el capítulo 2 se presenta el marco teórico, es decir, las bases teóricas utilizadas para diseñar y construir el corpus de expresiones faciales y el sistema reconocedor de emociones

centradas en el aprendizaje, donde se abordan temas como Computación Afectiva, Corpus de Imágenes y Sistemas Reconocedores de Emociones. En el capítulo 3, se expone el estado del arte de trabajos relacionados a este tema de investigación. En el capítulo 4, se detalla el análisis, diseño, desarrollo y estructura del corpus de imágenes faciales y del sistema reconocedor de emociones, además, se describen los componentes más importantes y el modelo del sistema. En el capítulo 5, se muestran las pruebas realizadas, los resultados obtenidos y el análisis de estos, cuyo objetivo es validar las dos hipótesis antes planteadas. Finalmente, en el capítulo 6 se presentan las conclusiones y trabajos futuros de este trabajo de tesis.

2. Marco Teórico

En este capítulo se presentan las bases teóricas que fundamentan el presente trabajo de tesis. Los temas relacionados que se abordan son: corpus de imágenes, computación afectiva, sistemas reconocedores de emociones e inteligencia artificial.

2.1 Corpus

El corpus o base de datos como también se le conoce, han tenido un gran crecimiento en los últimos años en el área de análisis y construcción de corpus. Estos son creados para un propósito específico y representan una colección de grandes volúmenes de datos que pueden ser texto, imágenes o audios, por mencionar algunos. Representan una parte de la infraestructura, para el desarrollo de aplicaciones de procesamiento y reconocimiento de datos (Atkins, Clear, & Ostler, 1992).

La figura 1, muestra la metodología del corpus, la cual consiste en un proceso de extracción, transformación y carga, por sus siglas (ETC). A la primera etapa se le conoce como planeación y captura de datos; en esta etapa se especifican los tipos de datos que se recolectan, el equipo necesario para adquirirlos y el proceso que seguirán los involucrados en caso de que se necesite que participen personas (participante y especialistas). La segunda etapa, llamada procesamiento de datos el cual consiste en reconocer y extraer características de los datos capturados, posteriormente se aplica técnicas de máximos y diferencias de tiempo. La última etapa consiste en almacenar y clasificar la información usando algún título o etiqueta que permita reconocer la información. Finalmente, con este proceso se obtiene un corpus representativo.



Figura 1. Metodología del corpus.

El corpus representativo es sometido a pruebas para identificar sus fortalezas y debilidades, y con base en la experiencia y análisis de las pruebas se agregan y/o eliminan datos del corpus. Este mecanismo se repite continuamente con el objetivo de contar con un corpus balanceado, es decir, que las clases o etiquetas tengan aproximadamente el mismo número

de imágenes. Además, el corpus debe ser representativo, es decir, de imágenes espontáneas y no actuado, también debe ser confiable, que no cuente con datos erróneos o repetidos. Una vez terminado el corpus, es necesario proteger los derechos de autor del corpus, así como difundirlo en medios científicos para la investigación y desarrollo de diversas aplicaciones.

2.2 Computación Afectiva

La computación afectiva es un área de la Inteligencia Artificial (IA) que surge por la necesidad de dotar a los equipos electrónicos de cierta capacidad de interactuar con las personas mediante el uso de visión computacional y técnicas de aprendizaje máquina, con el objetivo de que el sistema sea capaz de producir una respuesta afectiva (Bosch et al., 2015; Picard, 1995). Se dice que la computación afectiva es interdisciplinaria, ya que abarca distintas áreas como informática, psicología y aspectos cognitivos, además representa un papel importante en el desarrollo de interfaces inteligentes aplicada a la educación o al software educativo. Según (Carberry & De Rosis, 2008), la computación afectiva se subdivide en cuatro áreas de investigación, como se detalla a continuación:

1. El análisis y caracterización de los estados afectivos que identifique mediante interacciones naturales las relaciones entre afecto y procesos cognitivos en el aprendizaje.
2. El reconocimiento automático de estados afectivos por medio de analizar expresiones faciales y extraer características de expresiones lingüísticas, postura, seguimiento de la mirada, ritmo cardíaco, entre otros.
3. La adaptación de los sistemas para responder ante un particular estado afectivo de los usuarios.
4. El diseño de avatares que muestren estados afectivos apropiados para una mejor interacción con el usuario.

En la sección 2.2.5 se aborda más sobre esto.

Teoría de la emoción

Cuando se habla de computación afectiva, es necesario definir con precisión el concepto de emoción, ya que se puede confundir con otros conceptos como afecto, sentimiento o motivación. A continuación, se define cada uno de ellos para evitar confusiones.

- El afecto es un proceso de interacción social entre dos o más personas. Dar afecto es algo que se transfiere, es decir, dan afecto y reciben afecto. Por ejemplo, hacer un regalo, visitar a un enfermo, etc.
- Los sentimientos son la expresión mental de las emociones; es decir, se habla de sentimientos cuando la emoción es codificada en el cerebro y la persona es capaz de identificar la emoción específica que experimenta: alegría, pena, rabia, soledad, tristeza, vergüenza, son ejemplos de sentimientos.
- La motivación es un conjunto de procesos implicados en la activación, dirección y persistencia de la conducta lo que nos permite provocar cambios de la vida en general.
- La emoción es un estado de ánimo producido por un suceso o recuerdo y se presenta día a día en nuestra vida diaria el cual juega un papel importante en la comunicación no verbal (Bradley & Lang, 2006).

Clasificación de emoción

2.2.2

Las emociones se clasifican en:

- Primarias o básicas: Paul Ekman (Ekman, 1992), identificó 6 emociones básicas que son: enojo, disgusto, miedo, felicidad, tristeza y sorpresa, además muestra algunas características que se presentan en el rostro de la persona como se muestran en la figura 2.



Figura 2. Emociones básicas.

- Secundarias o alternativas: Son emociones complejas que aparecen después de las emociones primarias o básicas y dependen más de la situación del contexto de la persona. Por ejemplo, una persona que tiene miedo (emoción primaria) puede convertir eso en enfado o enojo (emoción secundaria) y provocar una reacción en su cuerpo de pelear. También se utiliza en este caso un modelo de dimensiones de valencia e intensidad para describir una emoción con mayor precisión (Becker-Asano & Wachsmuth, 2009; Christie & Friedman, 2004), como se muestra en la figura 3.

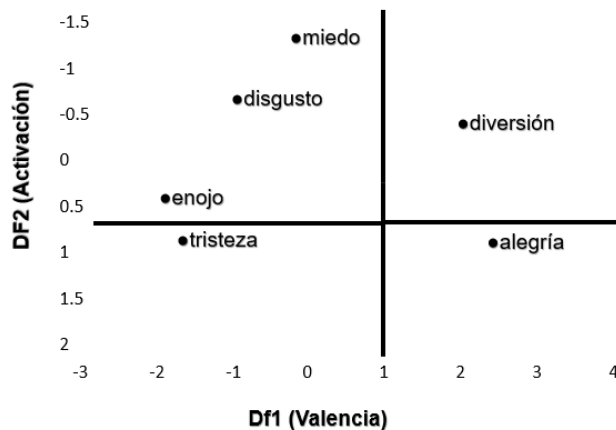


Figura 3. Modelo de activación y valencia tomada de (Christie & Friedman, 2004).

2.2.3

Expresiones de la emoción

Las emociones se pueden clasificar en expresiones internas y externas. Las expresiones internas pueden ser señales generadas por el cuerpo, como la presión sanguínea, sudoración, señales de electroencefalografía y las expresiones externas pueden ser expresiones faciales o

2.2.4 gestos del rostro, el sonido de la voz, la postura del cuerpo y/o movimiento de este (lenguaje corporal).

Dispositivos de reconocimiento de emoción

Las emociones se pueden detectar mediante dispositivos invasivos o no invasivos. Los dispositivos invasivos, son aquellos que están parcial o totalmente sobre el cuerpo de una persona y que permiten extraer cierta información de ésta, por ejemplo, una diadema para captar ondas cerebrales, una banda de medición para detectar ritmo cardiaco, etc. Los dispositivos no invasivos son aquellos que no están en contacto directamente con el cuerpo

de la persona, sino que a través de otros mecanismos captan información de la persona para convertirla en datos que representan las emociones, como fotografías del rostro (captura de imágenes), seguimiento y postura del cuerpo, micrófono para captar la voz, entre otros. Algunos de estos dispositivos usados para captar señales para reconocer emociones se muestran en la figura 4.



Figura 4. Dispositivos de captura de datos.

2.2.5 Emociones centradas en el aprendizaje

Las emociones centradas en el aprendizaje se producen en los estudiantes cuando realizan diferentes actividades, manifestando una variedad de estados afectivos en contextos de aprendizaje. Algunas de estas emociones son: comprometido, aburrido, frustrado, relajado, estresado, enfocado e interesado. Las emociones de confundido, frustrado o aburrido se presentan en los estudiantes cuando realizan ejercicios que requieren ciertas técnicas o información con la que no están familiarizados, y pueden considerarse como negativas para el aprendizaje del estudiante. Las emociones de comprometido, enfocado, relajado o interesado se presentan en estudiantes cuando están familiarizado con los ejercicios, les gusta el tema que están aprendiendo o conocen del tema y resultan positivas para el aprendizaje del estudiante.

Por otra parte, Kort, Reilly y Picard proponen un modelo relacionado con las emociones del aprendizaje (ver figura 5). La cual muestra un modelo dividido en cuatro cuadrantes, donde el cuadrante I muestra una valoración de admiración, satisfacción o curiosidad resultado un aprendizaje constructivo positivo, el cuadrante II expone una valoración de desilusión,

perplejidad o confusión resultando un aprendizaje constructivo negativo, el cuadrante III representa una valoración de frustración, descartar o conceptos erróneos el cual resulta en un desaprendizaje negativo y el ultimo cuadrante IV, muestra una valoración de optimismo, nueva investigación el cual resulta en un desaprendizaje positivo (Bosch et al., 2015; Kort, Reilly, & Picard, 2001; Sabourin, Rowe, Mott, & Lester, 2011).

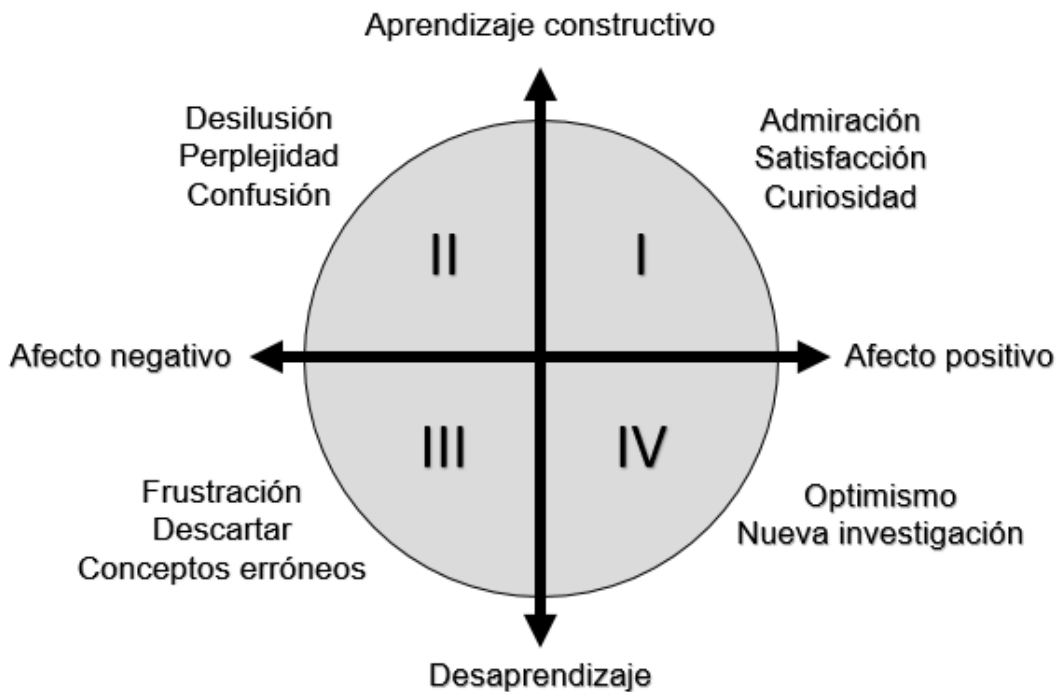


Figura 5. Modelo de las emociones del aprendizaje tomada de (Kort et al., 2001).

2.3 Sistemas Reconocedores de Emociones

Los sistemas reconocedores de emociones utilizan diversos métodos de extracción y clasificación de características. Las características se extraen de las señales generadas por el usuario, por ejemplo, mediante el uso de expresiones faciales, las cuales son características que se presentan en el rostro de una persona, la voz, la cual se obtiene mediante las frecuencias del sonido, el uso de señales cerebrales obtenidas mediante electroencefalografía (EEG), entre otros métodos. La clasificación se efectúa con base en la similitud del conjunto de características obtenidas de las señales.

Reconocimiento de expresiones faciales

El análisis de expresiones faciales se aplica en diferentes áreas de interés como educación, videojuegos, telecomunicaciones, por mencionar algunas. Además, es una de las más

utilizadas en interacciones humano-computadora. El reconocimiento de expresiones faciales es un sistema inteligente que identifica el rostro de una persona y a partir de este obtiene ciertas características que analiza y procesa para conocer el estado afectivo de la persona (Sumathi, Santhanam, & Mahadevi, 2012). En el estado del arte se detalla más información acerca de reconocedores de expresiones faciales.

Reconocimiento de voz

Los reconocedores de voz, son sistemas compuestos de varios componentes, modelos 2.3.2 acústicos, modelos de lenguaje y modelos de pronunciación, los cuales mediante análisis y procesamiento de frecuencia de audio y utilizando redes neuronales transcribe una secuencia de audio a una secuencia de palabras (Chan, Jaitly, Le, & Vinyals, 2016).

Reconocimiento de señales EEG

2.3.3 En los últimos años, más investigadores han realizado trabajos en reconocimiento de emociones por medio de señales electroencefalografías (EEG), teniendo una mayor atención, la tecnología basada en EEG. Esta tecnología se ha enfocado en el desarrollo de nuevas aplicaciones en áreas como entretenimiento, aprendizaje electrónico, mundo virtuales y otras formas de interacciones humano computadora (Y. Liu, Sourina, & Nguyen, 2010).

2.4 Sistemas Tutores Inteligentes

Los sistemas tutores inteligentes (STI) son aplicaciones de software diseñadas con métodos combinados de inteligencia artificial y tutores expertos, se diseña el material para enseñar de acuerdo con los objetivos que estudiante debe aprender y las técnicas con las que adquiere una mayor cantidad de conocimiento. Los STI cumplen con el objetivo de que cada estudiante tenga un profesor virtual, tomando en cuenta las necesidades individuales de los estudiantes.

2.4.1 Los STI usan muchos métodos de inteligencia artificial para generar problemas y sugerencias adecuadas, dependiendo de las necesidades de los estudiantes, y además van ajustando su método de enseñanza conforme va aprendiendo el estudiante (Akkila & Naser, 2017).

Arquitectura de un STI

La arquitectura de un sistema tutor inteligente comprende el modelo de dominio, el modelo pedagógico, el modelo de estudiante y la interfaz de usuario como se muestra en la figura 6 (Akkila & Naser, 2017).

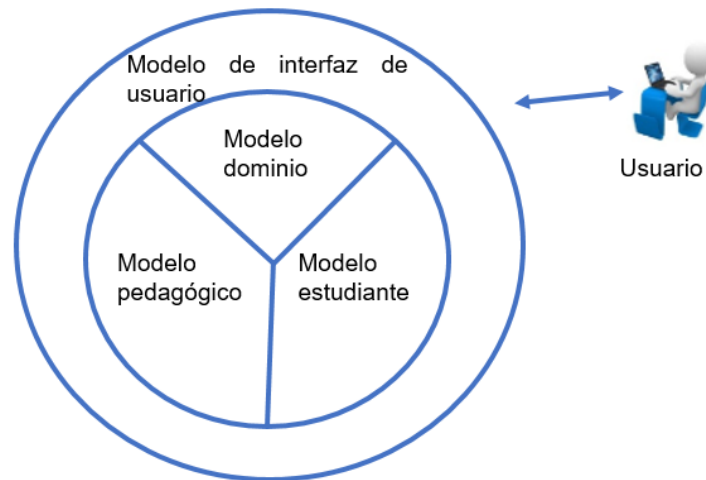


Figura 6. Arquitectura de un sistema tutor inteligente tomada de (Akkila & Naser, 2017).

3. Estado del Arte

Hoy en día, existen muchos corpus con diferentes enfoques como imágenes, texto, sonidos, entre otros. Estos corpus se emplean en los sistemas reconocedores para entrenar algún tipo de inteligencia artificial, que facilite la tarea de reconocimiento y clasificación.

Para crear una buena base de datos de expresiones faciales espontaneas, es necesario resolver algunos problemas, como manejar la efectividad de producir emociones de los sujetos que participan en los experimentos, segmentar y etiquetar imágenes faciales, entre otros.

A continuación, se describen los trabajos e investigaciones relacionadas con los temas de diseño y construcción de un corpus, reconocimiento de expresiones faciales y sistema tutores afectivos.

3.1 Diversos corpus de imágenes o video

3.1.1 Un requisito para el análisis y clasificación de expresiones faciales son los corpus centrados en imágenes o videos, los cuáles pueden ser de expresiones faciales actuadas, no actuadas (espontaneas) o ambas. A continuación, se describen en cada subsección los trabajos relacionados con este tema.

Presentación y validación de la base de datos de rostros RadBound (RaFD)

Procesar el rostro es una tarea compleja, ya que el rostro puede contener mucha información social. Por ejemplo, información acerca de la identidad, estados motivacionales y

emocionales, movimiento de labio y dirección de la vista, los cuáles son importantes en una comunicación exitosa. La base de datos RaFD fue desarrollada en la Universidad de Nijmegen con 49 participantes adultos y niños, controlando factores como condiciones de luz, puntos faciales y un fondo blanco uniforme. Cada modelo muestra ocho expresiones faciales con tres direcciones de la mirada, donde simultáneamente los participantes portan camisa negra, sin lentes, maquillaje, joyería o pelo en el rostro y son fotografiados en cinco ángulos diferentes de izquierda a derecha con diferencia de 45° cada uno. Cada sesión de los participantes tiene una duración de 45 minutos, donde previamente practicaron expresiones faciales con un experto en el Sistema de Codificación de Acciones Faciales (FACS por sus siglas en inglés). Cada imagen fue procesada donde se le aplicó un correcto balance de blancos y se alinearon los puntos faciales por medio de software (Langner et al., 2010).

La Figura 7 muestra el proceso de una sesión para capturar las imágenes de los participantes, la cual contiene un total de 8040 imágenes.

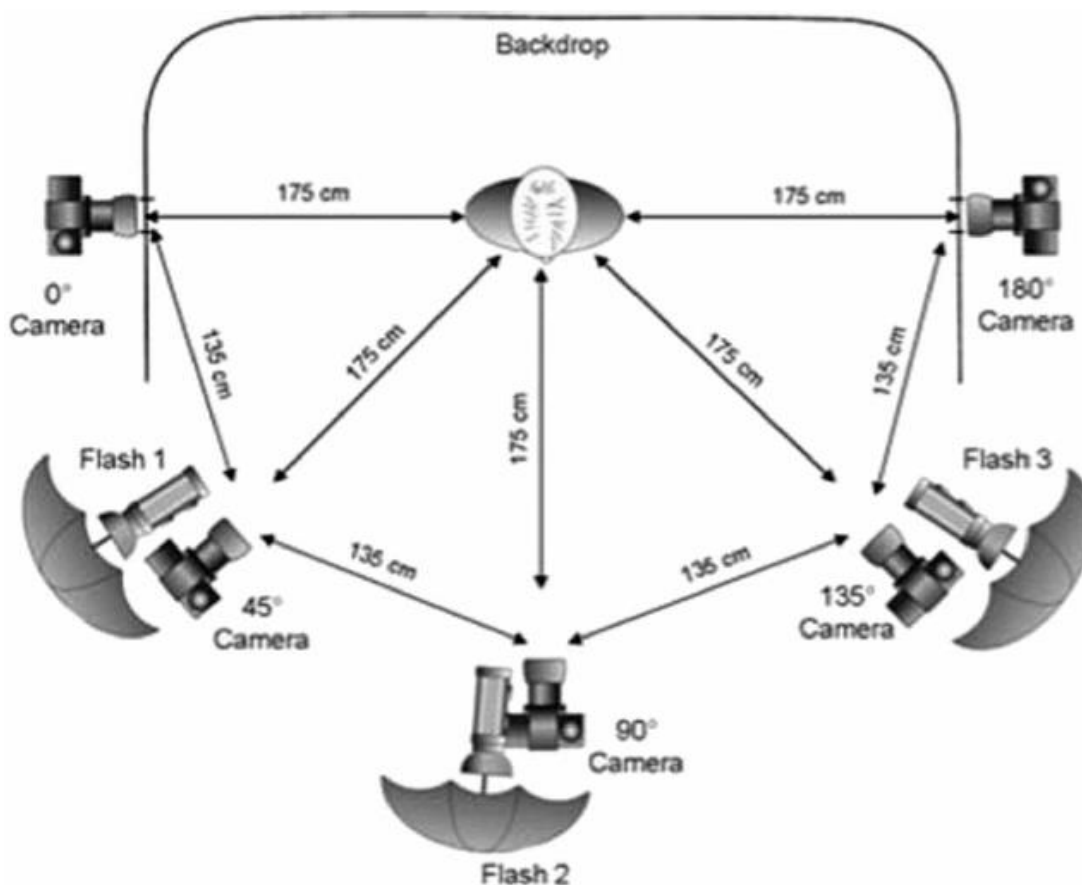


Figura 7. Estudio de sesiones tomada de (Langner et al., 2010).

DISFA una base de datos espontanea de intensidad de acción facial

La base de datos disponible está limitada a comportamiento facial actuado que puede diferir en intensidad y tiempo de un comportamiento espontaneo. Los participantes fueron 27 personas (12 mujeres y 15 hombres) donde su edad varía entre 18 y 50 años, y cada uno observaba un video clip de aproximadamente 4 minutos, con el objetivo de producir expresiones faciales. Cada imagen de comportamiento facial tiene una alta resolución de 1024 x 768 pixeles y bajo iluminación controlada, obteniendo un total de 130,815 marcos de video. La figura 8 muestra un ejemplo del estudio donde se realizaron las sesiones con los participantes.



3.1.3

Figura 8. Estudio DISFA tomada de (Mavadati, Mahoor, Bartlett, Trinh, & Cohn, 2013).

Base de datos de rostros PUT

El rostro humano es un cuerpo no rígido y su apariencia es susceptible a muchos factores como variación de la pose, cambios en la iluminación, oclusiones o perturbaciones estructurales como maquillaje, lentes o bello facial, por mencionar algunos. La base de datos

de rostros PUT, se conforma de un total de 9971 imágenes a color de 2048 x 1536 píxeles, donde participaron alrededor de 100 personas. Las imágenes tienen diferentes poses de la cara que fueron tomadas en ambientes de luz controlados y fondos uniformes. La figura 9 muestra un ejemplo de las imágenes de la base de datos (Kasinski, Florek, & Schmidt, 2008).



Figura 9. Imágenes de la base de datos PUT tomada de (Kasinski et al., 2008).

3.1.4 Base de datos BP4D espontánea.

La mayoría de las bases de datos actuadas, están limitadas a imágenes estáticas de 2D o video, donde el rostro es un objeto tridimensional por lo que un video 2D puede ser insuficiente. BP4D es una base de datos de expresiones faciales espontáneas en video 3D, donde participaron diversos grupos de jóvenes adultos, proporcionando una nueva base de datos 3D que incluye secuencias de modelo 3D y textura de video 2D. El sistema toma videos 3D a una velocidad de 25 marcos por segundo y utiliza dos videocámaras en escala de grises y una videocámara a color. Para crear este corpus participaron 41 personas de las cuales 23 son mujeres y 18 hombres de entre 18 y 29 años. Cada participante realizó diversas actividades como entrevistas, ver video clips, jugar y mantener la mano sumergida en agua helada el mayor tiempo posible, por mencionar algunas, con el objetivo de producir emociones básicas de forma espontánea. La información fue capturada dentro de un laboratorio en condiciones normales de luz. La figura 10, muestra los diferentes tipos de vistas capturadas; estas son: la

vista general, vista video 2D, modelo geométrico dinámico 3D y modelo geométrico dinámico 3D con textura (Zhang et al., 2014).



Figura 10. Diversas vistas tomada de (Zhang et al., 2014).

3.1.5

Base de datos de NVIE

La base de datos NVIE, incluye dos sub-bases de datos que consisten en expresiones espontáneas y de expresiones actuadas. Cada sub-base de datos se compone de imágenes visibles e infrarrojas que fueron registradas por dos cámaras bajo tres diferentes condiciones de iluminación: izquierda, frontal y derecha. Los experimentos fueron registrados utilizando dos dimensiones, valencia e intensidad, y etiquetado de imágenes de acuerdo con las 6 emociones básicas. Participaron 215 personas: 157 hombres y 58 mujeres en un rango de edad de 17 a 31 años. Para estimular las emociones, se utilizaron diferentes tipos de videos emocionales con una duración de entre 3 y 4 minutos cada uno. En el proceso de captura de imágenes, se usaron dos cámaras para registrar la imagen facial, una cámara que captura 30 marcos por segundo con una resolución de 704 x 480 píxeles y una cámara infrarroja que

captura 25 marcos por segundo con una resolución de 320 x 240 píxeles, obteniendo un total de 576 imágenes frontales, 544 de izquierda y 546 de derecha (Wang et al., 2013).

Base de datos Cohn-Kanade extendida (CK+)

3.1.6 En el año 2000, la base de datos Cohn-Kanade fue liberada para promover la investigación en la detección de expresiones faciales. Esta base de datos tiene algunas limitaciones como etiquetas de emociones no bien validadas. La base de datos CK contiene 486 secuencias de 97 sujetos, donde se añadió a 210 adultos de edad entre 18 a 50 años de los cuales el 69% son mujeres. Se obtuvieron secuencias de imágenes de vista frontal y de perfil con una variación de 30 grados. La figura 11, presenta algunas imágenes del corpus donde la parte superior representa las imágenes de la base de datos CK y la parte inferior de la base de datos CK+ (Lucey et al., 2010).



3.1.7

Figura 11. Base de datos CK y CK+ tomada de (Lucey et al., 2010).

Comparación de bases de datos de expresiones faciales.

La tabla 1, presenta una comparación de las diferentes características presentes en las bases de datos de expresiones faciales descritas en esta sección.

Tabla 1. Comparativa de bases de datos.

Base de datos	Actuada/ espontánea	Número de personas	Número de imágenes	Número de videos	Resolución
RadBound (RaFD)	Actuada	49	5,880	No aplica	10 y 12 Mpx
DISFA	Espontánea	27	No aplica	130,815	1024x768 pxs
PUT	Actuada	100	9,971	No aplica	2048x1536 pxs
BP4D	Espontánea	41	No aplica	368,036	25 marcos
Espontanea multimodal (CK+)	Espontánea y actuada	215	1,666	No aplica	704x480 pxs y 320x240 pxs
	Espontánea y actuada	210	822	No aplica	640x490 pxs

3.2 Sistemas reconocedores de emociones

A continuación, se presenta la descripción de diferentes sistemas reconocedores de emociones donde se aplican diversas técnicas de extracción de características, utilizando para ello imágenes, audio o señales EEG.

3.2.1 Reconocimiento automático de expresiones faciales para predecir la emoción comprometido y frustrado.

El reconocimiento automático de expresiones faciales para predecir las emociones comprometido y frustrado (Grafsgaard et al., 2013) es un trabajo de investigación que construyó un corpus de video y utiliza una interfaz de lenguaje de programación llamada Java Tutor; además del uso de una herramienta de reconocimiento de expresiones y una máquina de soporte vectorial para predecir las unidades de acción facial. El corpus de video se compone de un total de 67 estudiantes con un promedio de edad de 18.5 años y una desviación estándar de 1.5 años, capturando alrededor de 4 millones de ventanas de video con una resolución de 640x480 pixeles y con una duración de 37 horas usando la interfaz Java Tutor. El participante realiza una prueba antes de utilizar la interfaz y otra al finalizar esta, con el objetivo de medir diversos aspectos de la emoción enganchado y del aprendizaje cognitivo. En la figura 12 se muestra una imagen donde se aprecia a un estudiante que interactúa con el tutor mientras la cámara graba la sesión en video.



Figura 12. Área de sesiones tomada de (Grafsgaard et al., 2013).

La herramienta de reconocimiento de expresiones faciales permite seguir una amplia variedad de unidades de acción facial ventana por ventana del video, encontrado características faciales y procesándolas mediante el uso de una máquina de soporte vectorial para obtener las expresiones faciales y sus respectivos valores. Por lo tanto, se construyeron algunos modelos predictivos para conocer la relación entre la intensidad y la frecuencia de los movimientos faciales, así como la relación que presentan estos con los estados afectivos enganchado y aburrido. Algunas características presentadas en los gestos de los estudiantes, es que frecuentemente colocaban una mano en el rostro mientras pensaban, o ambas manos cuando estaban aburridos o cansados.

Detección de emociones en escenarios educativos

La detección de emociones en escenarios educativos mediante expresiones faciales y movimientos corporales de los autores (Mar Saneiro, Olga C. Santos, Sergio Salmeron-Majadas, and Jesus G. Boticario, 2014), utiliza múltiples técnicas de análisis de expresiones faciales como detección, seguimiento y reconocimiento del rostro, así como también técnicas avanzadas de computación. Los experimentos se realizaron con el objetivo de detectar emociones de los usuarios en un entorno de aprendizaje electrónico, donde cada sesión tenía una duración de 2 horas y el laboratorio de experimentos tiene una capacidad para realizar 4 sesiones al mismo tiempo. Un total de 75 participantes fueron informados acerca de las tareas a realizar y se obtuvo información de ellos como edad, género, nombre, y otros datos. Se

realizó también un estudio antes y después del experimento para tener una retroalimentación. El objetivo de este fue producir diversas emociones como frustración, concentración, interesado, relajado, nervioso y confundido. Para la captura de información, se utilizaron diversos dispositivos como 2 cámaras web con una resolución de 1280 x 720 pixeles y 15 ventanas por segundo para registrar el rostro y postura del participante, y además se usó el dispositivo Microsoft Kinect para registrar movimientos corporales y procesar el rostro de la imagen. Por último, se utilizaron 2 equipos de cómputo, en uno el participante realiza los ejercicios y en el otro un experto observa la interacción del participante con el uso de un entorno inteligente para identificar características relevantes. La información de las expresiones faciales de los participantes, fueron decodificadas utilizando la herramienta de desarrollo de software Kinect, para localizar la posición de la cara y las expresiones faciales en tiempo real, obteniendo la posición x, y, z en un sistema de coordenadas y ángulos, así como también, el valor de seis unidades de animación y once unidades de forma. En el movimiento del cuerpo, se basaron en la postura, gesto, dirección del movimiento y frecuencia. El proceso de codificación y etiquetado de emociones la realizaron expertos que visualizaban los videos e identificaban los movimientos anotándolo en una hoja de cálculo, para su posterior análisis. Con esto, se detectó una cantidad de movimientos, el promedio y el tiempo que duró cada una de las actividades realizada por el participante. Como resultado, se encontró que si al participante se le limitaba a cierto tiempo de duración de una actividad compleja, se producía cierto nivel de frustración y estrés, así como una gran cantidad de movimientos faciales y corporales. Por otra parte, si el participante se sentía confiado y seguro antes de la tarea a realizar, permitía ayudarlo a terminar la actividad con una sensación de alegría y felicidad, produciendo un bajo nivel de movimientos faciales y corporales. Cuando los participantes se encontraban estresados y/o frustrados se presentaba frecuentemente el movimiento de los ojos. Algunos puntos relevantes fueron que las personas de estatura alta miraban hacia abajo para ver el monitor, haciendo complicado la recolección de información facial. Otro punto fue la sensibilidad del dispositivo Kinect, ya que requiere de por lo menos de un metro de distancia entre el dispositivo y el participante. Por último, algunos participantes que tenían problemas de visión frecuentemente se acercaban al monitor haciendo difícil la captura del rostro. A continuación, en la figura 13, se presenta una imagen del experimento, donde la parte izquierda muestra el video donde se registró el escritorio del

participante, la parte derecha muestra las expresiones faciales detectadas en los participantes, censurado esta por motivos de anonimato.

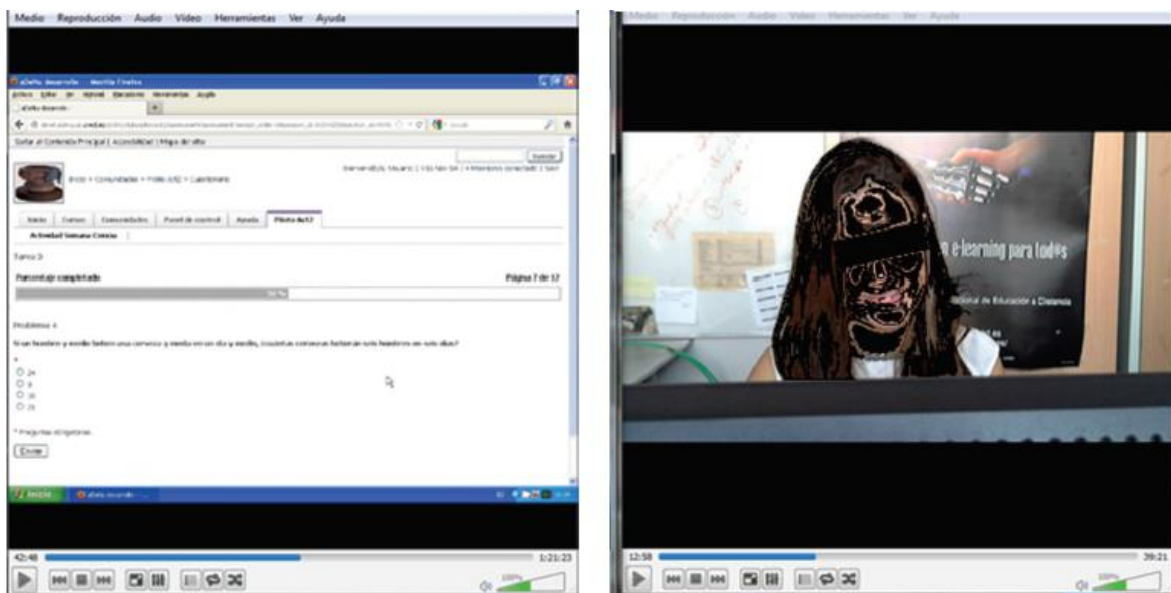


Figura 13. Experimento tomada de (Saneiro et al., 2014).

3.2.3 Los rostros de enganchado: Reconocimiento automático de enganchado en estudiantes por medio de expresiones faciales

El reconocimiento automático de estado afectivo de los autores (Whitehill, Serpell, Lin, Foster, & Movellan, 2014) describen que las técnicas más conocidas para estimar el estado afectivo enganchado en los sistemas tutores inteligentes, están basadas en la precisión y el tiempo de las respuestas de los estudiantes a problemas prácticos y preguntas de examen, también conocida la técnica como seguimiento de enganchado. Los autores calificaron el estado afectivo comprometido de los estudiantes, mediante juicios de expertos, obteniendo datos de 34 estudiantes que participaron en un experimento de entrenamiento de habilidades cognitivas, y mediante el uso de una cámara web para grabar video, un iPad y la colección de datos de un software de habilidades cognitivas. Las actividades del software estaban dirigidas a entrenar la percepción, la lógica, el razonamiento y la habilidad de memoria y consistían en establecer, recordar y sumar; el software fue diseñado para ser mentalmente agotador conforme los estudiantes pasan cada ejercicio. Durante cada sesión, los participantes fueron informados y dieron su consentimiento para realizar el experimento. Después de que se realizaron los experimentos los registros de video fueron etiquetados y

evaluados con base en la apariencia de enganchado, asignándole un valor al video clip, además a cada imagen estática también se le asignó un valor y por último se eliminó el audio. Cada video clip o imagen fue clasificado dentro de una de las siguientes categorías: no enganchado del todo, nominalmente enganchado, enganchado en una actividad y muy enganchado. A continuación, la figura 14 muestra algunas imágenes registradas de los experimentos, donde la escala de valores muestra 1, cuando el participante no está enganchado, subiendo sucesivamente hasta llegar a 4, donde la persona está muy enganchada.



Figura 14. Imágenes del experimento con escala tomada de (Whitehill et al., 2014).

El reconocimiento automático de emociones se realizó por medio de 3 etapas: la primera etapa fue el registro del rostro mediante puntos faciales, localización del rostro y coordenadas obteniendo una imagen de resolución de 36 x 36 y otra de 48 x 48 pixeles. La segunda etapa consistió en clasificar la imagen recortada por medio de cuatro clasificadores binarios donde el resultado de este alimenta a un sistema de regresión para estimar el nivel de comprometido

(tercera etapa). Los resultados de la clasificación binaria se obtuvieron mediante una matriz de confusión donde la precisión baja fue de 0.6525 y la mayor de 0.8434 teniendo un promedio de 0.7340.

Reconocimiento de expresiones en BDBN.

3.2.4 El reconocimiento de expresiones faciales desarrollado por los autores (P. Liu, Han, Meng, & Tong, 2014), presenta un marco de trabajo de redes de confiabilidad profunda impulsada mediante el uso de tres etapas de entrenamiento iterativo caracterizado por expresiones faciales. La primera etapa consiste en la captura de características relacionadas con cambios en la apariencia y/o geometría de las expresiones faciales, extraídas de secuencias de imagen o video. La segunda etapa, es la selección y extracción de las características que distinguen una expresión y en la tercera etapa se construye un clasificador para reconocer expresiones faciales. El marco de trabajo Boosted Deep Belief Network (BDBN por sus siglas en inglés), consiste de dos procesos de aprendizaje interconectados. En un proceso se usa una máquina de aprendizaje no supervisado de abajo hacia arriba que aprende de una representación de características. En el otro proceso se utiliza una máquina de aprendizaje supervisado de arriba hacia abajo que refina las características en conjunto con la máquina de aprendizaje no supervisado. A continuación, se muestra la figura 15 del proceso antes descrito.

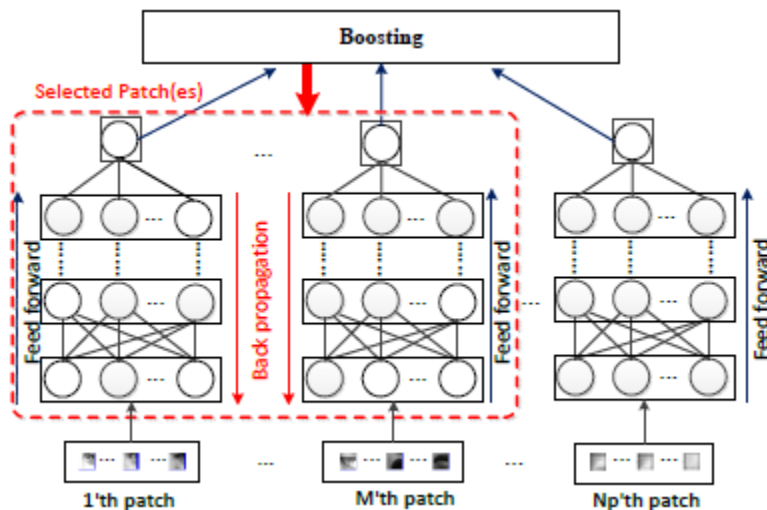


Figura 15. Vista general del marco de trabajo BDBN tomada de (P. Liu et al., 2014).

Las principales contribuciones del marco de trabajo BDBN son, la sistematización unificada de la representación, selección y clasificación de características en un mismo marco de

trabajo, además facilita el trabajo de representaciones basadas en partes y el uso de técnicas de múltiples redes de confiabilidad profunda. Los experimentos se realizaron con el uso de la base de datos extendida Cohn-Kanade (CK+) y la base de datos JAFFE, las cuales son utilizadas ampliamente en sistemas de reconocimiento de expresiones faciales. De cada imagen se obtuvieron 80 imágenes parcialmente enlazadas con un tamaño de 24 x 24 píxeles. Los resultados se obtuvieron mediante dos validaciones cruzadas. La primera fue usando la base de datos CK+ para entrenar; la segunda usando la base de datos JAFFE para clasificar 7 expresiones faciales (6 expresiones básicas y una neutral). La Tabla 2, muestra los resultados de la primera validación usando 4 diferentes clasificadores.

Tabla 2. Resultados comparados con otras máquinas de aprendizaje tomada de (P. Liu et al., 2014).

Ada+SVM (Lineal)	Ada+SVM(Polineal)	Ada+SVM(RBF)	BDBN
0.404	0.404	0.413	0.680

El segundo resultado se obtuvo con el uso de la base de datos JAFFE para entrenar y clasificar 7 expresiones (6 expresiones básicas y una neutral). Los resultados con 4 clasificadores diferentes son mostrados en la tabla 3.

Tabla 3. Resultados con JAFFE tomada de (P. Liu et al., 2014).

SLLE	SFRCS	Ada+SVM(RBF)	BDBN
0.868	0.860	0.810	0.918

3.2.5 Por último, se realizó una prueba usando la base de datos CK+ complementada con la base de datos JAFFE, obteniendo como resultado una precisión de 0.930.

Reconocimiento automático de expresiones faciales utilizando características de regiones faciales.

En este trabajo, el reconocimiento de afecto incluye características del habla, lenguaje corporal, señales electroencefalografías, imágenes de resonancia magnética funcional y expresiones faciales, donde este último se aborda en el trabajo de los autores (Happy & Routray, 2015). En este trabajo se propone un novedoso marco de trabajo para realizar reconocimiento de expresiones faciales mediante el uso de marcadores faciales para reconocer características en algunas regiones del rostro. Para ello, proponen una metodología

basada en cambios faciales la cual involucra contracciones y expansiones de los músculos faciales lo cuales generan características. Primero se detecta el rostro y algunas partes de éste como nariz, cejas, ojos y boca, posteriormente se extraen las regiones de interés activas mediante el uso de puntos faciales y regiones activas. Después se ejecutan dos etapas: la etapa de entrenamiento y la etapa de clasificación. La etapa de entrenamiento incluye preprocesamiento, selección de regiones faciales, extracción de características y aprendizaje de clasificadores de múltiples clases. La etapa de clasificación consiste en detectar los puntos faciales, extraer las características seleccionadas y finalmente clasificar las expresiones faciales. La figura 16 muestra el sistema general propuesto basados en el de los autores.

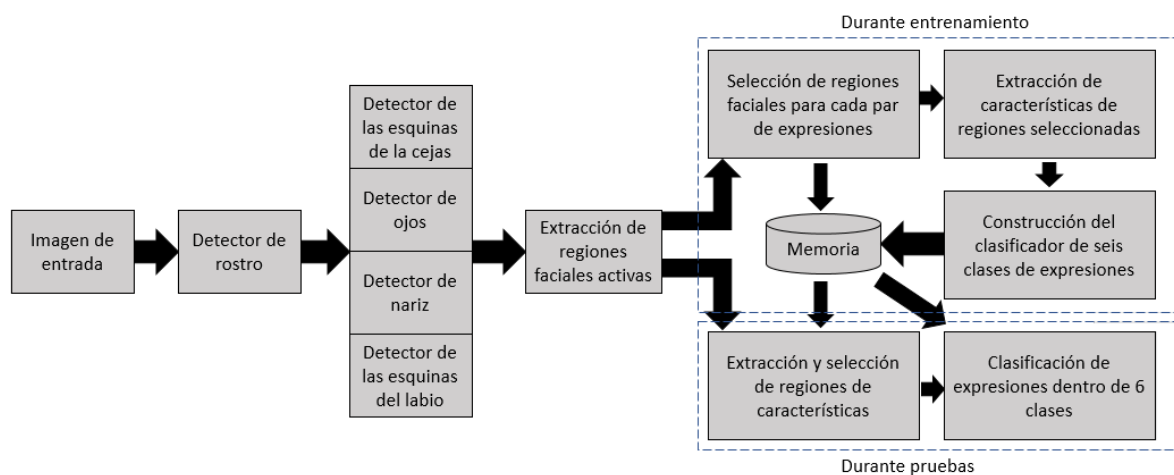


Figura 16. Vista general del sistema propuesto basados en el de (Happy & Routray, 2015)

Los autores aplican un pre procesamiento antes de extraer las regiones faciales activas, donde utilizan técnicas de Viola-Jones de características de Haar con el uso de una máquina de aprendizaje AdaBoost. En cada región de interés seleccionada, se aplican diferentes técnicas como Gausiana Blur, operador horizontal de Sobel para detección de bordes, umbral Otsu, operaciones de dilatación morfológica, y conexión de componentes y posiciones. Para la extracción de características y las variaciones de la iluminación, aplicaron histogramas de patrones binarios locales (LBP por sus siglas en inglés). La figura 17 presenta las diferentes posiciones de las regiones faciales.

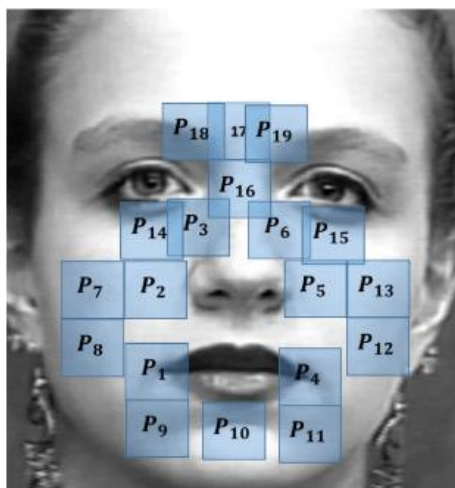


Figura 17. Regiones faciales tomada de (Happy & Routray, 2015).

La evaluación del marco de trabajo propuesto, se realizó con el uso de las bases de datos de expresiones faciales de mujeres japonesas (JAFFE) y Cohn-Kannade (CK+), aplicando una validación cruzada de diez para evaluar el rendimiento. Los resultados se muestran en la tabla 4 en una matriz de confusión con el uso de la base de datos CK+.

Tabla 4. Resultados de la base de datos CK + tomada de (Happy & Routray, 2015).

	Enojo	Miedo	Disgusto	Felicidad	Tristeza	Sorpresa
Enojo	87.8	0	0	0	7.32	4.88
Miedo	0	93.33	0	4.44	0	2.22
Disgusto	0	1.88	94.33	0	1.88	1.88
Felicidad	1.44	2.89	0	94.22	0	1.44
Tristeza	1.78	0	0	1.78	96.42	0
Sorpresa	0	0	0	1.53	0	98.46

Por último, la tabla 5 muestra los resultados obtenidos con la base de datos JAFFE.

Tabla 5. Resultados de la base de datos JAFFE tomada de (Happy & Routray, 2015).

	Enojo	Miedo	Disgusto	Felicidad	Tristeza	Sorpresa
Enojo	100	0	0	0	0	0
Miedo	0	93.75	0	0	0	6.25
Disgusto	6.89	6.89	86.2	0	0	0
Felicidad	0	0	0	96.77	0	3.22
Tristeza	9.67	6.45	0	6.45	77.41	0
Sorpresa	0	3.33	0	0	0	96.66

3.2.6 Reconocimiento de emociones a partir de características faciales geométricas utilizando mapas propiamente organizados

Este trabajo, muestra un modelo de reconocimiento de emociones utilizando un enfoque de sistema de identificación propuesto por los autores (Majumder, Behera, & Subramanian, 2014), que involucra la generación de un esquema automático de características faciales geométricas. Las expresiones faciales son una forma de reconocer emociones y un tipo de comunicación no verbal en forma clara y universal. El análisis de expresiones faciales se enfoca en tres etapas básicas: detección de rostro, extracción de características faciales y la clasificación de estas. La Figura 18, muestra un modelo analítico que consiste en 23 puntos faciales donde los autores describen 6 expresiones faciales básicas.

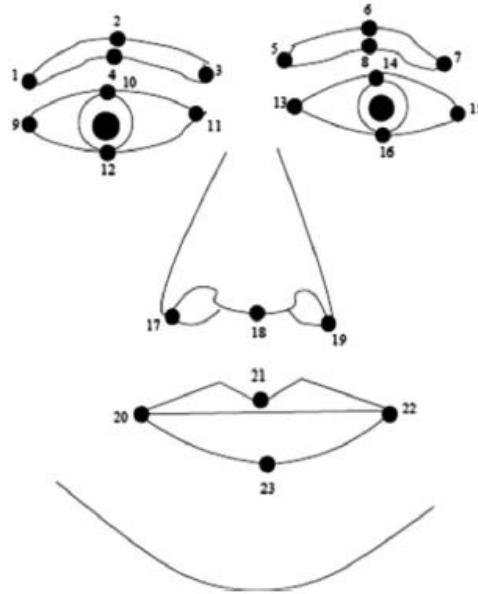


Figura 18. Modelo de 23 puntos faciales tomada de (Majumder et al., 2014)

El desplazamiento o movimiento de las características faciales es calculado utilizando una expresión neutral como referencia, donde la punta de la nariz juega un papel importante en el cálculo de los desplazamientos. En la detección del rostro y algunas partes de éste, los autores usaron un algoritmo de detección de objetos, los métodos de cascada Haar y Viola-Jones. Una vez detectadas, las partes de las cejas, nariz, ojos y boca, se obtiene el contorno de la región de interés y se aplica un histograma de ecualización sobre la imagen plana para eliminar el efecto de variación de intensidad y extraer los puntos clave de la imagen como se muestra en la figura 19.



Figura 19. Extracción de características tomada de (Majumder et al., 2014)

El desplazamiento de cada punto facial con respecto a su localización del punto facial neutral, contiene información acerca del movimiento del musculo facial, el cual indica la expresión mediante sus coordenadas x , y . Dichos desplazamientos, conforman un vector de 26 dimensiones de características geométricas faciales, que se usa para entrenar clasificadores como RBFN, MLP3 y una MSV de múltiples clases. En el entrenamiento se utilizó la base

de datos de videos MMI, que contiene expresiones faciales de seis emociones básicas como felicidad, tristeza, disgusto, enojo, sorpresa y miedo. La tabla 6 muestra en una matriz de confusión los resultados obtenidos, utilizando 26 dimensiones de características geométricas y una máquina de soporte vectorial para clasificar.

Tabla 6. Resultados obtenidos de (Majumder et al., 2014).

	Felicidad	Tristeza	Disgusto	Enojo	Sorpresa	Miedo
Felicidad	91.5	3.66	2.44	2.44	0	0
Tristeza	1.88	86.4	2.24	3.77	5.66	0
Disgusto	0	2.38	90.5	7.14	0	0
Enojo	1.67	0	0	98.3	0	0
Sorpresa	1.88	0	0	0	98.1	0
Miedo	4.76	2.43	2.43	0	0	90.4

La precisión del reconocimiento de la MSV de múltiples clases es del 92.53% con un máximo de 98.33% y un mínimo de 88.10%.

3.3 Sistemas tutores afectivos

A continuación, se presentan diferentes Sistemas Tutores Inteligentes (STI) que ayudan a mejorar el aprendizaje del estudiante mediante el uso de estímulos y métodos de enseñanza electrónica, como pueden ser personajes interactivos, chats de colaboración, ejercicios y técnicas de gamificación. Además, permite a estos adaptarse al estudiante de manera que adquiera la mayor cantidad de conocimiento mediante el uso de ayudas, explicaciones o

3.3.1 tutorías.

Predicción de afecto a partir de la mirada durante la interacción con un sistema tutor inteligente.

El trabajo de los autores (Jaques, Conati, Harley, & Azevedo, 2014), está relacionado con la información del seguimiento de los ojos para predecir emociones relevantes al aprendizaje con enfoque en los estados afectivos aburrimiento y curiosidad. Para capturar la información de los participantes se utilizó un sistema tutor inteligente adaptativo llamado MetaTutor, el cual consiste en un entorno de aprendizaje que enseña conceptos acerca del sistema circulatorio humano, y que cuenta con cuatro agentes pedagógicos que interactúan con el

estudiante para evaluar que este entienda el contexto. La figura 20 muestra una interfaz de MetaTutor.

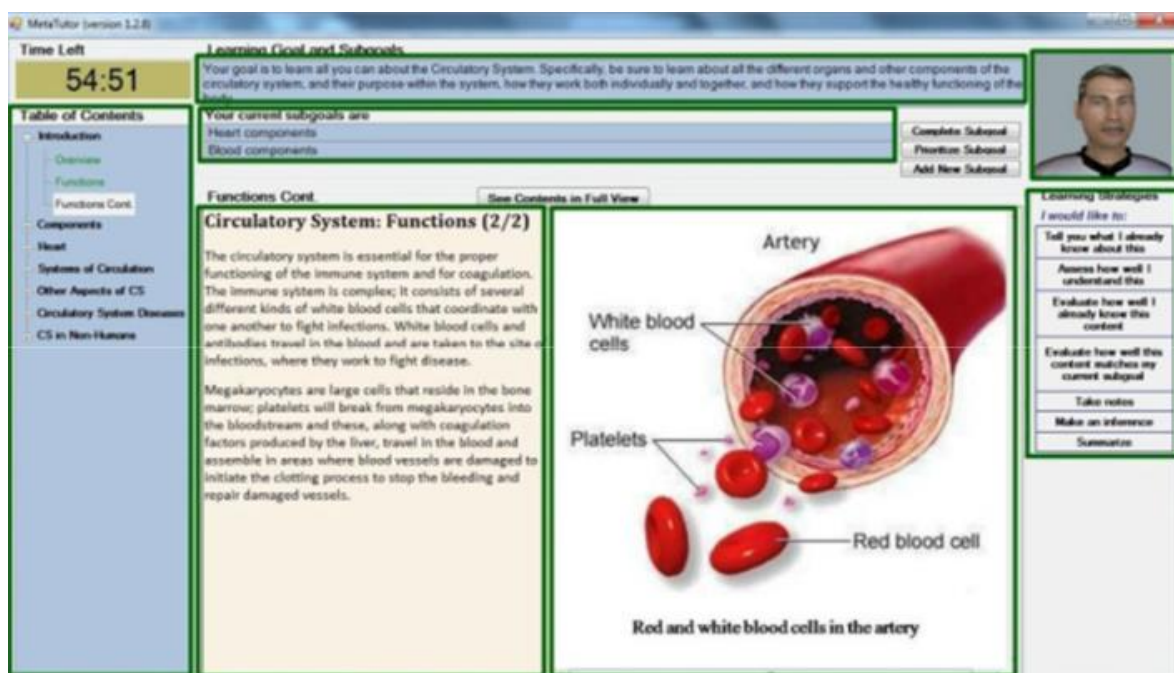


Figura 20. Interfaz Meta tutor tomada de (Jaques et al., 2014).

Los datos fueron capturados de 67 estudiantes de diferentes programas académicos, donde algunos no estaban relacionados con el contenido de MetaTutor. Cada estudiante, utilizaba el sistema alrededor de 90 minutos mientras se registraba el seguimiento de la mirada utilizando diferentes sensores de seguimiento de ojos como Tobii T60. Posteriormente, los estudiantes reportaban sus emociones mediante el uso de un cuestionario de emociones centradas en el aprendizaje. En el seguimiento de la mirada, los autores utilizaron áreas de interés específicas como imágenes, contexto, objetivos, tablas de contenido y agentes dentro de la interfaz MetaTutor. Además, se analizaron otros datos como la duración de la mirada en una región de interés y la transición entre estas. Los experimentos fueron realizados con una herramienta de minería de datos llamada Weka, aplicando una validación cruzada de diez y 4 algoritmos de clasificación que son Random Forest (RF), Naive Bayes, regresión logística y máquina de soporte vectorial (SVM por sus siglas en inglés). Los mejores resultados obtenidos fueron de 69% de precisión para aburrimiento y 73% para curiosidad.

TuinLEC, un tutor inteligente para mejorar la competencia lectora.

El presente trabajo de los autores (Vidal-Abarca et al., 2014), propone un sistema tutor inteligente llamado TuinLEC, el cual tiene como objetivo mejorar la educación mediante estrategias de enseñanza y comprensión lectora de estudiantes de educación primaria. TuinLEC incorpora agentes virtuales para modelar y guiar a los estudiantes. Además, el sistema se adapta y ayuda a mantener la motivación de los estudiantes. Los autores utilizaron una metodología donde participaron 25 estudiantes de sexto año de primaria; la primera mitad tuvo 13 alumnos seleccionados como grupo de control y 12 alumnos como grupo de experimentación, sin diferencias en su capacidad general de comprensión. TuinLEC se compone de ocho lecciones estructuradas de enseñanza explícita y práctica. Las primeras 4 lecciones son: como leer textos continuos y discontinuos, como entender y responder preguntas, cuando decidir releer el texto para responder y como buscar en el texto para responder. Con las otras cuatro lecciones se practican conjuntamente las estrategias anteriormente descritas, donde cada lección contiene dos textos, uno continuo y otro discontinuo, de los cuales cada uno incluye entre seis y ocho preguntas para localizar, relacionar y reflexionar sobre el texto. La figura 21 presenta el sistema tutor inteligente TuinLEC, el cual cuenta con puntuaciones y ayudas e incorpora agentes pedagógicos virtuales.



Figura 21. Sistema tutor inteligente TuinLEC. tomada de (Vidal-Abarca et al., 2014).

TuinLEC recopila información de competencias relevantes, por ejemplo, la cantidad de texto leído inicialmente, decisiones de relectura, búsqueda de información relevante para responder, etc. Los agentes pedagógicos enseñan a los estudiantes durante las lecciones y a

lo largo del proceso proporciona retroalimentación de aciertos y errores. En la fase de práctica el estudiante ya no cuenta con agentes pedagógicos, solamente con retroalimentación la cual incluye información de los estudiantes relacionada con la competencia lectora y ayudas para facilitar el proceso de búsqueda de información textual y la comprensión de los enunciados. El estudiante va ganando puntos que se transforman en estrellas de oro, plata o bronce al final de cada lección, con el fin de mantener la motivación del estudiante. Los resultados obtenidos al usar el STI TuinLEC, fueron realizados mediante el uso de un análisis de covarianza, empleando como variable independiente la instrucción/ausencia de instrucción. Los resultados muestran que el grupo experimental mejoró su rendimiento en comprensión lectora de $M=13.00$, $DS=3.67$ respecto al grupo de control de $M=10.00$, $DS=4.66$. Además, el efecto de la covariable resultó estadísticamente significativo de 13.776 respecto a 4.329, obteniendo un tamaño de efecto medio de 0.665, resultando con esto muy aceptable la eficacia de TuinLEC para enseñar la comprensión lectora.

3.3.3 Ambiente inteligente de aprendizaje con manejo afectivo para Java.

El trabajo de los autores (Zatarain-Cabada, Barrón-Estrada, González-Hernández, & Oramas Bustillos, 2015) presenta un sistema tutor inteligente (STI) con manejo de estados afectivos para el aprendizaje del lenguaje de programación Java llamado Java Sensei. El objetivo del sistema tutor inteligente es ayudar a los estudiantes de programación a comprender y mejorar las diversas áreas del lenguaje Java, evaluando aspectos cognitivos, afectivos e integrando un agente pedagógico. Java Sensei incorpora siete componentes que son: Plataforma afectiva, Módulo del estudiante o usuario, Agente pedagógico, Módulo de adaptación, Módulo de enseñanza (Pedagógico), Módulo del experto y el Sistema de tutoría Inteligente/Afectivo. El módulo Afectivo está basado en un sistema lógico difuso y una red neuronal de propagación hacia atrás, los cuales requieren de variables de entrada y salidas para el reconocimiento de emociones. El modelo del estudiante representa la parte cognitiva, la cual constituye la eficiencia del conocimiento del estudiante respecto al curso y la parte emotiva la cual consiste en la emoción al momento de responder un ejercicio cuyos valores pueden ser feliz, enojado, triste, sorprendido y neutral. El sistema tutor inteligente realiza diversas acciones como retroalimentación, respuesta empática emocional, expresiones faciales que realiza el agente pedagógico e intervenciones para ayudar al estudiante. El sistema lógico difuso y la red neuronal emplean cuatro variables de entrada difusa y tres variables de salida, además cuenta

con 144 reglas difusas que se implementaron con JFuzzyLogic para su creación. El sistema de reconocimiento emocional fue construido en tres etapas: la primera consiste en una implementación de un extractor de características de imágenes faciales, la segunda etapa fue la extracción de características de un corpus de datos llamado Radboud Faces Database (RAFD) para entrenar la red neuronal a partir de algoritmos de propagación hacia atrás y la tercera etapa fue la clasificación de las emociones. La figura 22, presenta el sistema lógico difuso en conjunto con la red neuronal.

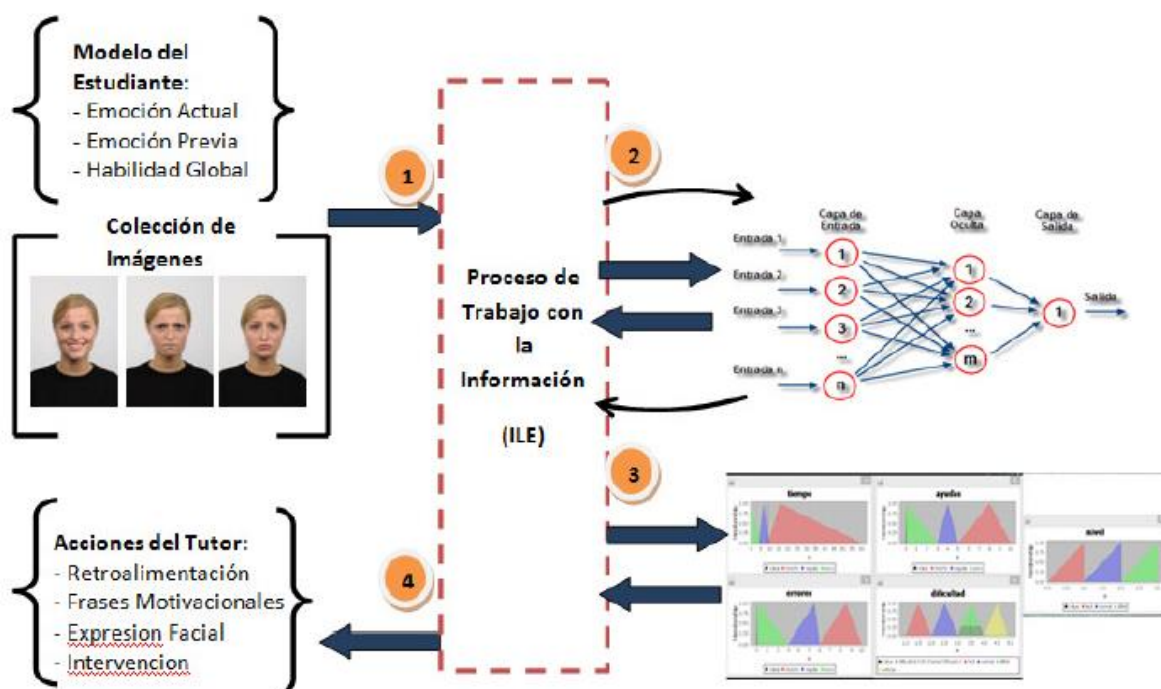


Figura 22. Sistema lógico difuso y red neuronal tomada de (Zatarain-Cabada, Barrón-Estrada, González-Hernández, & Oramas-Bustillos, 2015).

El módulo del agente pedagógico, es el encargado de realizar la interacción con el estudiante mediante la comunicación de mensajes, diálogos y expresiones faciales, las cuales pueden ser encantado, escéptico, sorprendido o neutral. El sistema tutor inteligente, emplea técnicas seguimiento de ejemplos, guiando al estudiante a resolver los ejercicios paso a paso, y mostrando empatía hacia este. La figura 23, muestra cómo funciona el motor de comportamiento del sistema Java Sensei a partir de la conducta del estudiante y su estado afectivo.

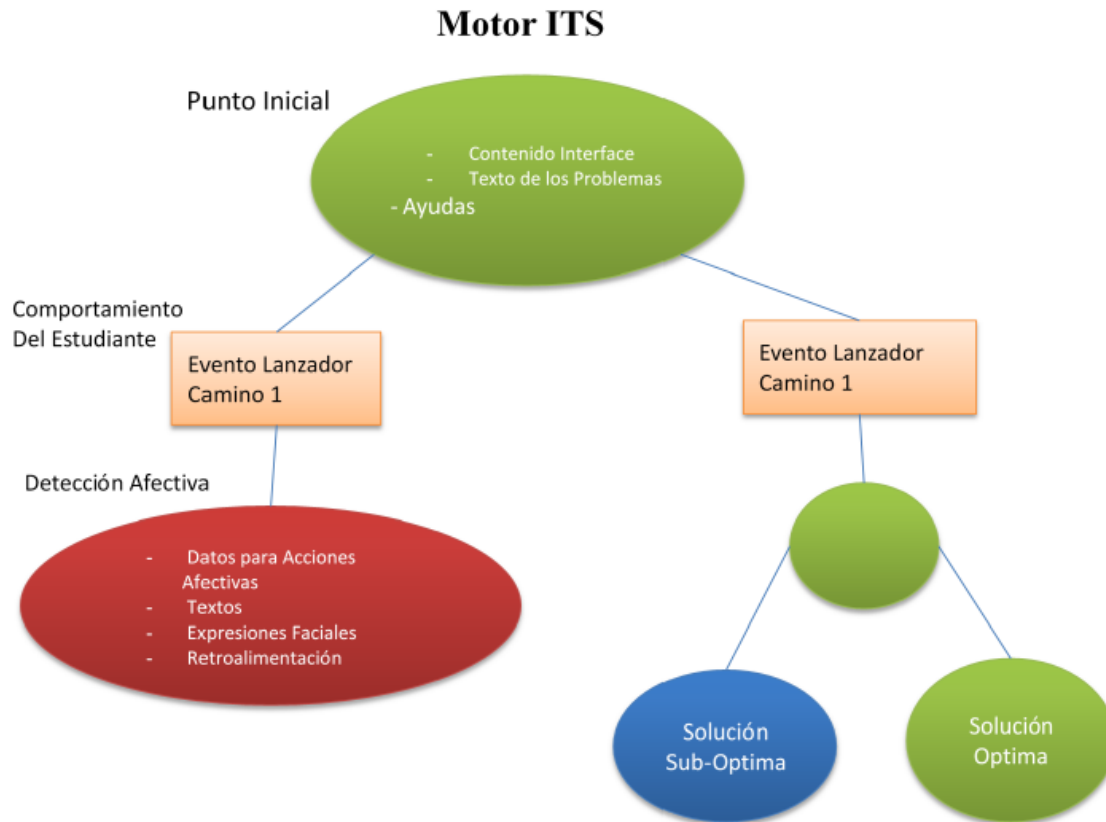


Figura 23. Motor de Java Sensei tomada de (Cabada et al., 2015).

La evaluación del sistema Java Sensei examinó la funcionalidad y la utilización de este en un grupo de estudiantes donde primero se aplicó un examen de diagnóstico antes de utilizar Java Sensei, posteriormente se les impartió un curso de la plataforma web Java Sensei para que los estudiantes se familiarizaran con este, y para finalizar se les aplicó un examen para medir la eficiencia del sistema. El grupo fue dividido en uno de control y otro de experimentación. Los resultados del examen diagnóstico fueron 80% aprobados y 20% no aprobados. Por otra parte, los resultados del examen después de utilizar Java Sensei mostraron una mejora del 5%, dando como resultado en el grupo 85% aprobados y 15% no aprobados.

Construcción de un reconocedor de expresiones faciales y una base de datos de expresiones faciales para un sistema tutor inteligente.

El trabajo de los autores (Zatarain-Cabada et al., 2017), presenta la construcción y validación de una base de datos de expresiones faciales, así como también un reconocedor de

expresiones faciales basados en técnicas de geometría. En este trabajo se construyó una propia base de datos de expresiones faciales relacionadas con las emociones en contextos educativos frustrado, aburrido, enganchado y emocionado. Cada imagen del corpus contiene una etiqueta de la emoción relacionada, la cual se obtuvo mediante el uso de la tecnología Emotiv Epoc de señales electroencefalográficas (EEG). La metodología para capturar las imágenes se definió para que los estudiantes solucionaran ejercicios de programación en Java mientras el dispositivo Emotiv Epoc capturaba su actividad cerebral y con el uso de una cámara web se obtenía cada 5 segundos una imagen del rostro del estudiante. Posteriormente la imagen era etiquetada con la emoción obtenida del dispositivo Emotiv y la imagen era almacenada en una base de datos de expresiones faciales. Finalmente se obtuvo una base de datos de 7019 fotografías de expresiones faciales donde se aplicó un filtro para eliminar los registros incorrectos, obteniendo un total de 730 fotografías. La construcción del reconocedor de expresiones faciales implementó una técnica basada en geometría y apariencia, la cual iniciaba extrayendo características del rostro con el uso de una librería Dlib y localizaba un conjunto de 68 puntos faciales del rostro. Cada punto facial contiene coordenadas (x,y) de localización en el rostro y puede cambiar su valor respecto a la posición del punto facial. Después se obtiene el centro de gravedad de todos los puntos faciales y posteriormente se realizan cálculos del punto central (centro de gravedad) a cada punto facial localizado en el rostro. La figura 24 presenta dos imágenes, la imagen de lado izquierdo muestra el centro de gravedad y la derecha muestra todos los puntos faciales encontrados.

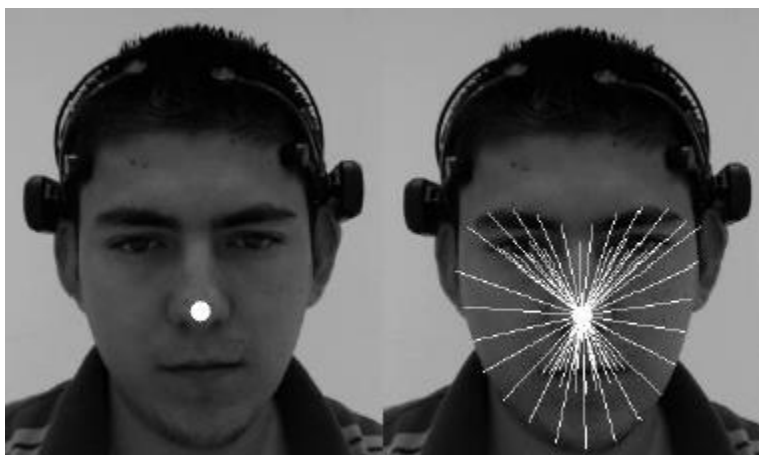


Figura 24. Puntos faciales tomada de (Zatarain-Cabada et al., 2017).

Las pruebas fueron realizadas mediante una validación cruzada de diez, donde el clasificador se entrenó con el 90% de la base de datos y el 10% restante se utilizó para clasificar. Un total de 655 fotografías fueron procesadas en el entrenamiento. La tabla 7 presenta los resultados obtenidos de las pruebas y el número de imágenes en cada clase de emoción.

Tabla 7. Resultados obtenidos de (Zatarain-Cabada et al., 2017).

Emoción	Numero de clases	Precisión de este trabajo	Precisión de otro trabajo
Aburrido	17	64%	76%
Enganchado	430	64%	91%
Emocionado	114	83%	68%
Frustración	94	62%	55%

4 Desarrollo del Proyecto

En esta sección se presenta la metodología y construcción de un corpus de imágenes de expresiones faciales, además de un sistema de reconocimiento de emociones enfocadas al aprendizaje a partir de expresiones faciales. A continuación, se describen a detalle el desarrollo de los proyectos mencionados anteriormente.

4.1 Protocolo de sesión

Las sesiones se realizaron con estudiantes de nivel licenciatura en el laboratorio de la Maestría en Ciencias de la Computación del Instituto Tecnológico de Culiacán, donde a cada uno se le explicó la descripción, los objetivos del corpus, así como también las herramientas que se utilizarían y se les pidió su consentimiento para poder capturar imágenes faciales con fines solamente de investigación. La captura de datos se realizó en dos etapas. En la primera participaron 25 estudiantes de los cuales 18 son hombres y 7 mujeres, todos ellos estudiantes del instituto tecnológico de Culiacán con un rango de edad de 18 a 47 años. En la segunda participaron 13 estudiantes de los cuales 10 son hombres y 3 mujeres en el mismo rango de edades.

4.2 Equipos de captura de datos

El equipo que se utilizó para capturar los datos fue una diadema Emotiv Epoc Insight, la cual cuenta con 5 electrodos para extraer señales EEG aplicadas a métricas de emociones enfocadas al aprendizaje y 2 electrodos de referencia para colocar la diadema, una cámara web Logitech HD Pro C920, para capturar imágenes de los participantes y dos equipos de cómputo; uno donde el estudiante resolvía ejercicios en el lenguaje Java en una plataforma web de aprendizaje y el otro equipo en donde se ejecutaban dos tareas al mismo tiempo. En uno se procesaban y almacenaban las señales EEG de la diadema con su fecha y hora precisa y en el otro se realizaba la captura de imágenes y se almacenaba con un nombre específico, el cual consistía en un identificador, un número, la fecha y hora exacta, como se muestra en el siguiente patrón: identificador_263_2016-11- 08_13-30- 11. La figura 25, muestra la diadema Emotiv Epoc Insight y de la cámara web Logitech HD Pro C920, utilizadas para las sesiones de creación de la base de datos de expresiones faciales.



Figura 25. Dispositivos utilizados para las sesiones. Todos los derechos reservados.

4.3 Metodología

Las sesiones se realizaron dividiendo los estudiantes en dos grupos, el grupo A con un total de 18 estudiantes realizaron diversas actividades que consistían en leer un problema, codificar el programa en Java, compilar y ejecutar el código en una plataforma web de aprendizaje llamada Java Gaming ILE. La figura 26, muestra la interfaz principal de la plataforma web de aprendizaje.

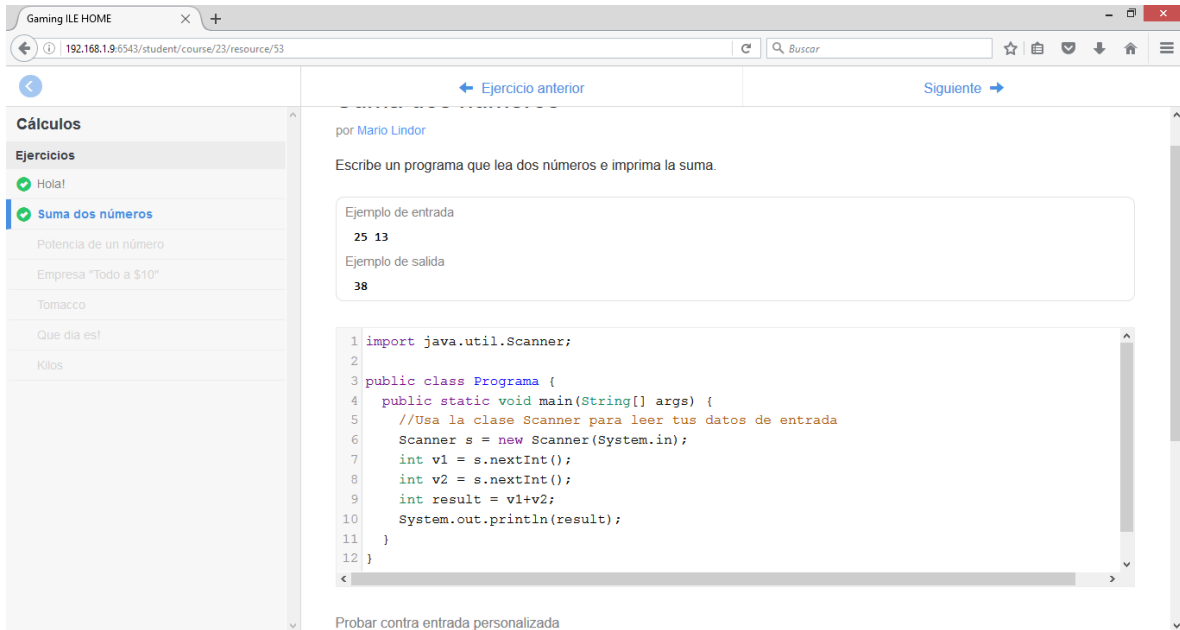


Figura 26. Plataforma web de aprendizaje del lenguaje Java Gaming ILE.

El objetivo del experimento en el grupo A fue inducir emociones en los estudiantes tales como comprometido, interesado, emocionado y enfocado. El grupo B con un total de 13 estudiantes, realizó actividades de observar videos cortos para inducir las emociones de aburrido y relajado.

Con respecto a la sesión, primero se aplicó solución salina a los electrodos de la diadema emotiv con el objetivo de obtener una señal óptima. Posteriormente, se colocó la diadema en el cráneo del estudiante comenzando por los electrodos de referencia y luego los restantes, verificando que todos los electrodos estuvieran en su posición y recibieran una señal óptima. Las señales en la interfaz de software de la diadema, son representadas con varios colores como son negro, que significa que el electrodo no hace buen contacto con el cuero cabelludo del cráneo, rojo que significa que existe una señal débil o con interferencia, amarilla que representa una señal con poca interferencia y por último verde que quiere decir que existe una señal óptima para obtener datos. La figura 27 muestra la interfaz del software de Emotiv.

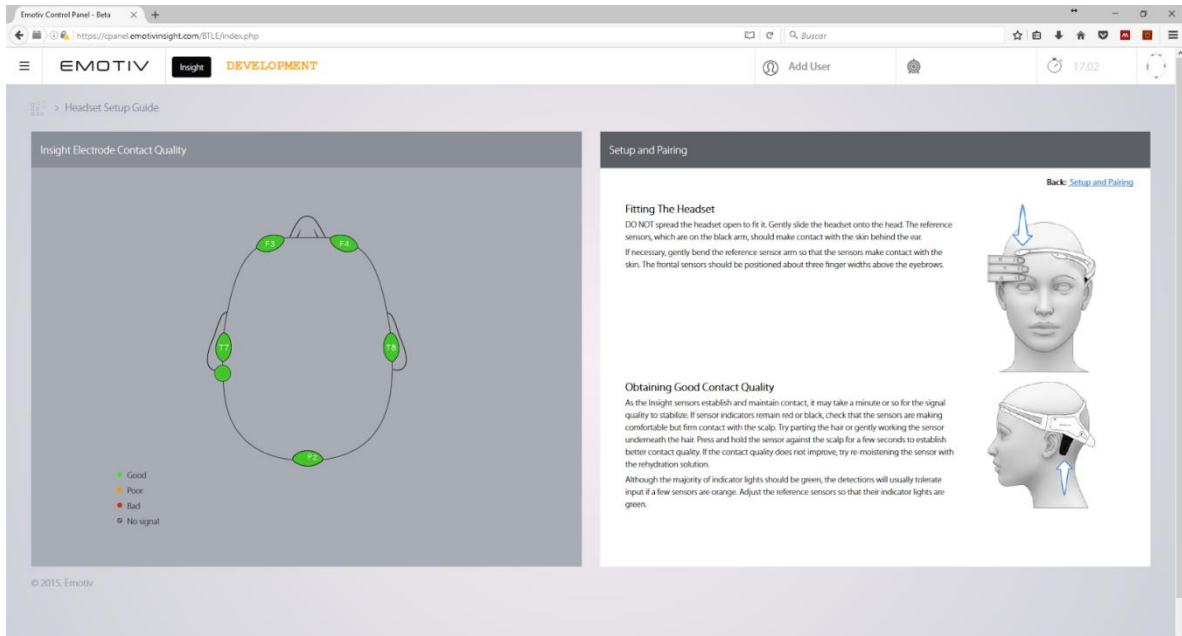


Figura 27. Interfaz Emotiv insight.

En caso de no tener una señal óptima en cierto electrodo, se remueve la diadema y se aplica nueva solución salina hasta que todos los electrodos estén en color verde para obtener la señal perfecta, como se muestra en la figura 27. Después de haber colocado la diadema, se verifica que la interfaz este procesando y mostrando las señales cerebrales EEG. La figura 28 muestra un ejemplo de la presentación de las señales EEG en la diadema Emotiv.

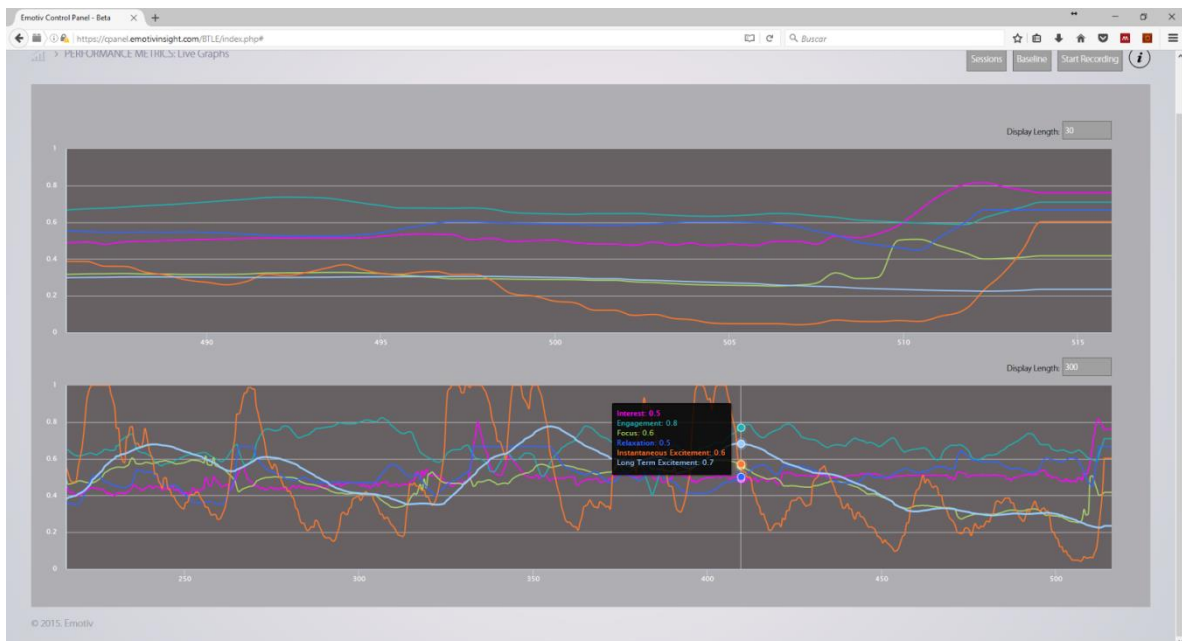


Figura 28. Presentación de señales EEG de Emociones en Emotiv.

Después de colocar bien la diadema en el estudiante, se ejecutan dos tareas que se describen a continuación. La primera consistió en capturar imágenes del rostro del estudiante por medio de la cámara web para después procesar las imágenes usando la librería OpenCV. El software de esta librería se utilizó para detectar el rostro de una persona. En caso de no encontrar el rostro del estudiante en la imagen ya sea por problemas de iluminación o porque la persona no se encontraba mirando hacia el frente, la imagen era eliminada. Si el rostro era detectado en la imagen, se almacenaba con el patrón descrito en la sección 4.2. La segunda tarea consistió, en procesar y transformar las señales EEG que se obtienen en tiempo real a valores de emociones centradas en el aprendizaje, para después registrar en un archivo log cada señal ya etiquetada con su respectiva fecha y hora precisa. Los valores de las emociones son representados con valores entre 0 y 1. A continuación la figura 29 muestra un ejemplo de un archivo log almacenado. Los valores de las señales de la emoción se describen de izquierda a derecha y representan las emociones de interesado, enganchado o aburrido, enfocado, emocionado en un periodo corto, emocionado en un periodo más largo y finalmente muestran el tiempo en que se capturó.

```

0.385181,0.73643,0.246374,0.495917,0.301858,0.228694,11:56:32,659PerformanceMetrics.js:380:6
0.391949,0.746621,0.248644,0.495754,0.324567,0.231643,11:56:33,297PerformanceMetrics.js:380:6
0.395279,0.751098,0.245559,0.500882,0.292957,0.234021,11:56:33,864PerformanceMetrics.js:380:6
0.401655,0.755117,0.238855,0.515663,0.292534,0.236351,11:56:34,445PerformanceMetrics.js:380:6
0.407724,0.758722,0.230721,0.533417,0.295985,0.238698,11:56:35,5PerformanceMetrics.js:380:6
0.410268,0.79022,0.221736,0.549422,0.298074,0.241039,11:56:35,593PerformanceMetrics.js:380:6
0.407898,0.914511,0.198072,0.584677,0.294232,0.24545,11:56:36,219PerformanceMetrics.js:380:6
0.40688,0.977534,0.189155,0.596549,0.293938,0.247599,11:56:36,737PerformanceMetrics.js:380:6
0.405765,1,0.184586,0.604304,0.292224,0.249679,11:56:37,326PerformanceMetrics.js:380:6
0.403962,1,0.183946,0.60936,0.313684,0.252075,11:56:37,875PerformanceMetrics.js:380:6
0.37697,1,0.186528,0.616692,0.403432,0.255924,11:56:38,433PerformanceMetrics.js:380:6
0.383077,1,0.191286,0.625003,0.512012,0.261526,11:56:38,983PerformanceMetrics.js:380:6
0.391049,1,0.198633,0.632432,0.632271,0.269062,11:56:39,550PerformanceMetrics.js:380:6
0.403606,1,0.199259,0.65507,0.705619,0.277732,11:56:40,140PerformanceMetrics.js:380:6
0.429214,1,0.208413,0.666667,0.754918,0.295966,11:56:40,662PerformanceMetrics.js:380:6
0.433043,1,0.21584,0.666667,0.784348,0.305644,11:56:41,213PerformanceMetrics.js:380:6
0.428782,1,0.222175,0.666667,0.812687,0.315686,11:56:41,756PerformanceMetrics.js:380:6
0.420484,1,0.226628,0.666667,0.809314,0.325559,11:56:42,344PerformanceMetrics.js:380:6
0.411589,1,0.23052,0.666667,0.79231,0.335036,11:56:42,910PerformanceMetrics.js:380:6
0.475658,1,0.318085,0.666667,0.752826,0.343745,11:56:43,415PerformanceMetrics.js:380:6
0.449289,1,0.237296,0.666667,0.752826,0.352349,11:56:43,966PerformanceMetrics.js:380:6
0.418727,1,0.238657,0.665177,0.684181,0.359704,11:56:44,544PerformanceMetrics.js:380:6
0.395334,0.967569,0.256378,0.631606,0.597848,0.365526,11:56:45,95PerformanceMetrics.js:380:6
0.396394,0.925649,0.273456,0.601417,0.528341,0.370106,11:56:45,609PerformanceMetrics.js:380:6

```

Figura 29. señales de emociones centradas en el aprendizaje.

4.4 Procesamiento de datos

Una vez finalizado el proceso de captura de datos, las señales fueron analizadas y graficadas para encontrar señales nulas o errores en la lectura de la diadema EEG. Esto se puede producir

cuando la diadema se mueve de su posición por ejemplo que el participante se moviera cuando se estaba realizando el experimento. Un ejemplo de señales nulas o con errores se muestra en la figura 30.

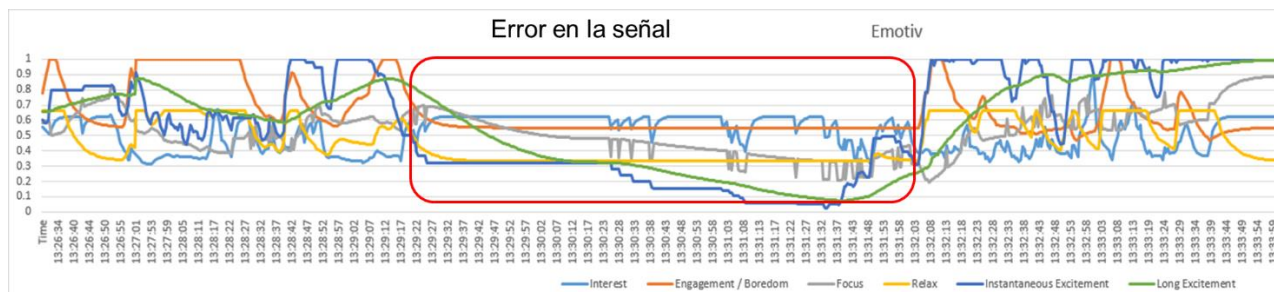


Figura 30. Análisis de error en las señales.

Como resultado del análisis de los datos, cinco sesiones fueron descartadas, obteniendo finalmente con éxito un total de 4286 imágenes en 25 sesiones. La figura 31 muestra algunos ejemplos de las imágenes de los rostros capturados durante el experimento.



Figura 31. Ejemplo de imágenes almacenadas en el corpus.

El proceso de etiquetado de imágenes consistió en asignar una emoción a cada imagen dependiendo de los valores de las señales EEG, donde mediante dos algoritmos de máximos y diferencias de tiempo se clasificaron las imágenes.

Dos diferentes versiones del corpus fueron el resultado de aplicar los dos algoritmos antes mencionados. El primer algoritmo consiste en obtener la fecha y hora de cada imagen del rostro y encontrar la coincidencia de la fecha y hora en el archivo log de señales EEG. Posteriormente se obtienen los valores de las señales EEG en donde la señal con el valor más alto determina la categoría de la imagen seleccionada la cual es almacenada con su respectiva etiqueta. El segundo algoritmo está basado en obtener las relaciones entre fecha y hora a

través del tiempo, es decir, el algoritmo utiliza un rango específico de valores EEG para determinar de forma precisa la emoción en cada expresión facial. Primero se obtiene la fecha y hora de la imagen y se busca la señal del archivo log de señales EEG para compararla con la fecha y hora de la señal EEG anterior, dando como resultado una sola señal. Después se utiliza el valor absoluto y aquella señal que resulte con un mayor cambio determina la emoción actual de la imagen. La ecuación siguiente es la que determina la emoción resultante.

$$\text{Emoción resultante} = \text{Máximo} (\text{Valor absoluto} (\text{Señal actual} - \text{señal anterior})).$$

La figura 32 muestra el proceso completo para capturar, etiquetar y almacenar las imágenes en una base de datos.

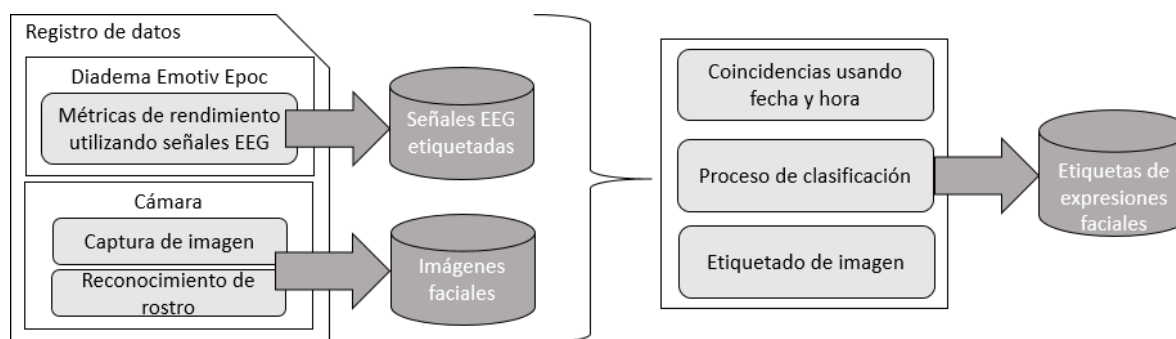


Figura 32. Proceso de etiquetado de imágenes.

La tabla 8 muestra el número de imágenes faciales obtenidas en cada uno de los dos algoritmos explicados anteriormente.

Tabla 8. Número de imágenes obtenidas de cada algoritmo de clasificación.

Numero de imágenes clasificadas		
Emoción	Algoritmo 1	Algoritmo 2
Aburrido	122	123
Enganchado	1995	461
Emocionado	1661	2953
Enfocado	222	356
Interesado	150	45
Relajado	28	333
Total	4178	4271

Los resultados de la evaluación y otras versiones generadas del corpus se muestran en el capítulo 5 “Pruebas y resultados”.

4.5 Reconocimiento de emociones centradas en el aprendizaje

En esta sección, se describe el sistema reconocedor de emociones centradas en el aprendizaje que se desarrolló para evaluar la precisión del corpus mediante la detección de expresiones faciales y la clasificación de estas en emociones. A continuación, se abordan las diferentes etapas en que se desarrolló el reconocedor.

Análisis

4.5.1 En esta etapa, basados en el conocimiento adquirido en el estado del arte y otras investigaciones de diversos sistemas reconocedores de expresiones faciales se determinó aplicar la ejecución de una secuencia de tareas que se describen. En el análisis del sistema reconocedor de emociones se decidió utilizar para desarrollar el sistema reconocedor de emociones, el lenguaje de programación Python con su compilador versión 2.7 junto con un complemento llamado miniconda, un entorno de desarrollo integrado llamado JetBrains Pycharm versión comunitaria, la librería OpenCV para analizar y procesar imágenes o videos y dos librerías llamadas scikit-learn y Dlib para entrenar máquinas de soporte vectorial.

4.5.2

Diseño

En esta etapa de Diseño, se planeó el proceso completo que seguiría nuestro reconocedor de emociones. Primero se determinó validar que la imagen que recibe el reconocedor contenga un rostro y que en caso de no encontrarlo descartar la imagen. Posteriormente, en el proceso se aplica un filtro de grises para eliminar los distintos canales de colores para después detectar algunas partes importantes del rostro como nariz, ojos, cejas y boca. Posteriormente, se aplica una técnica de puntos faciales a las partes antes detectadas, donde se localiza un punto inicial de referencia en el centro de la nariz y se identifican diversos puntos faciales en las partes del rostro que anteriormente se mencionaron. Después se realizan cálculos geométricos de distancia entre el punto inicial de referencia y cada punto facial detectado en el rostro, obteniendo como resultado una matriz de características del rostro que es enviada a una máquina de soporte vectorial con su etiqueta de emoción para que esta aprenda a clasificar expresiones faciales mediante el uso de la técnica de aprendizaje supervisado. Finalmente,

una vez entrenada la máquina de soporte vectorial se le envían más vectores de características faciales para probar si la máquina aprendió a clasificar. A continuación, la figura 33 muestra el proceso completo que seguirá el sistema reconocedor de emociones.

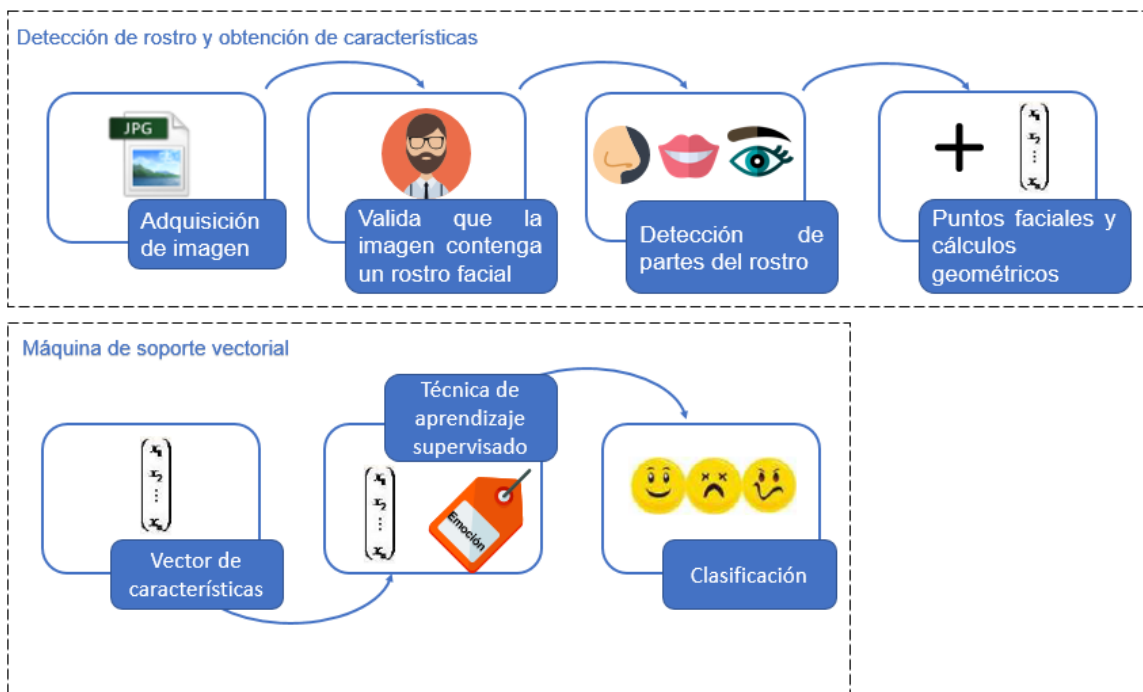


Figura 33. Sistema de reconocimiento de emociones.

4.5.3

Desarrollo

Basados en los resultados de los pasos de análisis y diseño, se procedió a implementar los algoritmos en el lenguaje de programación Python. A continuación, se muestran algunas imágenes de los diferentes procesos que se van ejecutando. La figura 34 muestra el resultado (derecha) de aplicar un filtro de grises a la figura (izquierda).



Figura 34. Aplicación de filtros de grises.

Posteriormente se detectan las partes del rostro y se encuentra el punto central inicial que será usado como referencia para los diferentes puntos faciales (ver figura 35).



Figura 35. Puntos faciales en el rostro.

Después de obtener todos los puntos faciales se realizan diversos cálculos geométricos basados en la distancia entre un punto inicial de referencia (centro de la nariz) hacia los otros puntos faciales y sus respectivas coordenadas de posición (x, y), obteniendo una matriz de características como la siguiente.

Tabla 9. Matriz de características.

Matriz				
78.45	89.21	66.54	101.86	148.23
123.51	246.32	183.71	200.56	230.75
23.74	182.34	121.72	120.65	149.45
147.35	215.43	204.65	198.54	176.29
67.22	210.34	129.28	89.45	90.35
72.09	99.25	190.56	130.45	115.63
59.25	201.33	202.53	173.56	192.34
36.33	158.63	92.57	105.67	109.65
45.65	22.86	104.34	63.56	67.23
90.23	140.65	152.37	193.67	201.27

Dicha matriz se envía a la máquina de soporte vectorial entrenada para que reconozca el tipo de emoción.

5. Pruebas y resultados

Una vez finalizado el procesamiento y clasificación del corpus, se utilizó el sistema reconocedor de expresiones faciales con el objetivo de conocer la precisión de cada clase del corpus. Este reconocedor utiliza un método basado en puntos faciales y utiliza la distancia entre estos en vez de utilizar coordenadas para obtener características del rostro en la imagen y una máquina de soporte vectorial para entrenar y clasificar (Sariyanidi, Gunes, & Cavallaro, 2015).

La tabla 10 muestra la precisión obtenida por el sistema de reconocimiento de expresiones faciales utilizando el corpus resultante del algoritmo número 1 previamente explicado y una validación de 10 cruzadas. La columna derecha muestra el promedio de precisión de cada clase de emoción.

Tabla 10. Algoritmo de clasificación número 1.

Clasificación	Precisión (Promedio)
Aburrido	68.57%
Enganchado	71.83%
Emocionado	70.99%
Enfocado	75.82%
Interesado	92.07%
Relajado	100%
Información estadística	
Precisión media	52.78%
Desviación estándar	13.70%

La tabla 11, muestra la precisión obtenida por el mismo sistema reconocedor de expresiones faciales, pero utilizando el corpus resultante del algoritmo número 2. Como se muestra en la tabla se obtuvieron mejores resultados con este algoritmo, el cual incrementa la precisión media en un 13.81%.

Tabla 11. Algoritmo de clasificación número 2.

Clasificación	Precisión (Promedio)
Aburrido	67.05%
Enganchado	79.07%
Emocionado	72.84%
Enfocado	81.48%
Interesado	100%
Relajado	86.21%
Información estadística	
Precisión media	66.59%
Desviación estándar	1.24%

5.1 Análisis estadístico de los datos

Se realizó un análisis estadístico con el objetivo de encontrar una posible correlación entre el género de los participantes y las emociones que ellos presentaban, con el uso de pruebas de

Pearson. Para este propósito cada señal de EEG fue categorizada ordinariamente. Como resultado no se encontró evidencia de alguna relación entre el género de los participantes con sus respectivas emociones. Posteriormente se utilizó un procedimiento de correlación bivariado para calcular el coeficiente de correlación Pearson, entre el promedio de las emociones presentado por cada participante. La tabla siguiente muestra el resultado de estas pruebas. En esta se observa una correlación negativa (-.471) entre las emociones interesado, enganchado/aburrido y también negativa (-.407) entre las emociones enfocado y relajado. Esto significa que cuando la emoción del estudiante enfocado incrementa, la emoción de relajado tiende a disminuir. Además, también fueron encontradas correlaciones positivas en enganchado/aburrido y emocionado (.435), en enfocado y emocionado (.806). Esto significa que cuando el valor enfocado incrementa, el valor emocionado también se intensifica.

Tabla 12. Resultados del coeficiente de correlación de Pearson

Clasificación	Comprometido/Aburrido	Relajado	Emocionado
Interesado	-.471	-.316	-.225
Comprometido/Aburrido	1	.223	.435
Enfocado	.134	-.407	.806

5.2 Tabla comparativa de las diferentes versiones del corpus

Después de analizar los datos se concluyó que las clases del corpus estaban desbalanceadas con lo cual se procedió a realizar una mayor cantidad de sesiones, con el objetivo de incrementar el número de expresiones faciales de la emoción aburrido. Los participantes realizaron actividades de lecturas de artículos literarios para producir dicha emoción. Participaron en este experimento un total de 13 estudiantes entre los cuales 10 fueron hombres y 3 mujeres. Se obtuvieron un total de 918 imágenes capturadas, las cuales fueron anexadas formando las versiones 3 y 4 del corpus. La tabla 13 muestra el total de imágenes del corpus anexando las sesiones donde se capturaba la emoción aburrimiento.

Tabla 13. Corpus resultante.

Emoción	Algoritmo 1	Algoritmo 2	Algoritmo 1 extendido	Algoritmo 2 extendido
Aburrido	122	123	1040	1041
Enganchado	1955	461	461	1955
Emocionado	1661	2953	2953	1661
Enfocado	222	356	356	222
Interesado	150	45	45	150
Relajado	28	333	333	28

5.3 Resultados de la versión 1 y 2 extendido

Con las nuevas imágenes del corpus se procedió a realizar pruebas de validación cruzada para conocer si se adquiere una mayor precisión. Los resultados obtenidos se presentan en la tabla 14.

Tabla 14. Resultados de los algoritmos extendidos.

Clasificación	Algoritmo 1 extendido	Algoritmo 2 extendido
Precisión media	65.74%	69.93%
Desviación estándar	2.23%	1.13%

5.4 Tabla comparativa de resultados

A continuación, la tabla 15 muestra una comparación de las 4 diferentes versiones generadas del corpus donde se muestran la precisión media y la desviación estándar de cada una de ellas.

Tabla 15. Resultados de los diferentes corpus.

Clasificación	Algoritmo 1	Algoritmo 2	Algoritmo 1 extendido	Algoritmo 2 extendido
Precisión media	52.78%	66.59%	65.74%	69.93%
Desviación estándar	13.70%	1.24%	2.23%	1.13%

5.5 Otros resultados

Por último, con el fin de validar de otra forma nuestros resultados se realizó una comparativa del reconocedor de emociones centradas en el aprendizaje y el uso de la diadema Emotiv Epoc Insight en tiempo real. Se realizaron 5 sesiones en las que participaron cuatro hombres y una mujer, de edad entre 23 a 47 años. Los participantes realizaron dos actividades que consistían en leer un artículo científico y realizar 3 ejercicios de programación básica en lenguaje de programación Java. Los resultados se obtuvieron al registrar y contar el número de veces en que Emotiv y el reconocedor coincidían en la clasificación de la emoción.

Tabla 16. Tabla comparativa entre Emotiv y el reconocedor.

Clasificación	Participante 1	Participante 2	Participante 3	Participante 4	Participante 5	Totales
Coincidencia	35	45	52	47	52	231
Total de datos	46	77	59	83	77	342
Promedio	0.760869	0.584415	0.881355	0.566265	0.675324	69.36%

6 Conclusiones y Trabajos Futuros

6.1 Conclusiones

En primer lugar, se debe trabajar en realizar más experimentos para balancear las diferentes clases del corpus y que permita capturar una mayor cantidad de imágenes, con el objetivo de contar con un corpus con una precisión mayor al 80%. Los resultados presentados anteriormente muestran un avance significativo desde la primera versión hasta la última generada, logrando una precisión cercana al 70%. Por otra parte, se necesita realizar pruebas con otro tipo de técnicas de reconocimiento y clasificación de imágenes, que permita comparar los resultados contra el reconocedor utilizado en este trabajo de tesis para tener un punto de comparación en los resultados.

6.2 Dificultades de la investigación

Durante el desarrollo del corpus se enfrentó a diversas dificultades que se exponen a continuación, la diadema Emotiv Epoc+ de 14 canales y 2 de referencia que cuenta el laboratorio de investigación del instituto tecnológico de Culiacán sufrió una deformación la

cual su diámetro creció y no hacía buen contacto con el cuero cabelludo del participante, por lo que se procedió a adquirir una nueva diadema, la Emotiv Epoc+ de 14 canales y 2 canales de referencia de modelo reciente y también se adquirió otra diadema llamada Emotiv Epoc Insight con la cual se realizaron los experimentos. El kit de desarrollo de software Emotiv (SDK por sus siglas en inglés), con el que se cuenta para interconectar el software del proyecto y la diadema Emotiv para adquirir los datos no era compatible, porque la diadema nueva es de modelo reciente. Se intentó adquirir el nuevo SDK pero no se obtuvieron resultados favorables. Finalmente se utilizó la diadema Emotiv Epoc Insight para la realización de la investigación.

6.3 Aportaciones

En este trabajo se tuvieron principalmente dos aportaciones: Una fue el diseño y construcción de un corpus de expresiones faciales espontáneas a partir de señales EEG que puede ser utilizado en entornos de aprendizaje, sistemas tutores inteligentes u otros sistemas educativos. Otra aportación fue el diseño y desarrollo de un sistema reconocedor de emociones a partir de expresiones faciales y una máquina de soporte vectorial.

6.4 Trabajo Futuro

- ✓ Realizar un mayor número de sesiones para incrementar la cantidad de imágenes de expresiones faciales del corpus, con el objetivo de incrementar la precisión.
- ✓ Aplicar algoritmos de clasificación alternativos e información adicional que permita realizar una mayor cantidad de pruebas para obtener mejores resultados.
- ✓ Aplicar diferentes técnicas de reconocimiento y clasificación para comparar diversos resultados.
- ✓ Desarrollar una plataforma web de reconocimiento de emociones para simplificar su uso a cualquier investigador y/o desarrollador con la finalidad de que se incorporen estados emocionales en distintos sistemas.

7 Referencias

- Akkila, A. N., & Naser, S. S. A. (2017). Teaching the right letter pronunciation in reciting the holy Quran using intelligent tutoring system, 2(1), 64–68.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1–16. <https://doi.org/10.1093/lc/7.1.1>
- Becker-Asano, C., & Wachsmuth, I. (2009). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1), 32–49. <https://doi.org/10.1007/s10458-009-9094-9>
- Bosch, N., Mello, S. D., Hall, F., Baker, R., Shute, V., & Wang, L. (2015). Automatic Detection of Learning - Centered Affective States in the Wild. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 379–388. <https://doi.org/10.1145/2678025.2701397>
- Bradley, M., & Lang, P. (2006). Emotion and motivation. *Handbook of Psychophysiology*, 581–607. <https://doi.org/10.1017/CBO9780511546396>
- Cabada, R. Z., Lucía, M., & Estrada, B. (2015). Ambiente inteligente de aprendizaje con manejo afectivo para Java. *Research in Computing Science*, 92, 111–121.
- Carberry, S., & De Rosis, F. (2008). Introduction to special Issue on “Affective modeling and adaptation.” *User Modeling and User-Adapted Interaction*. <https://doi.org/10.1007/s11257-007-9044-7>
- Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2016–May*, 4960–4964. <https://doi.org/10.1109/ICASSP.2016.7472621>
- Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, 51(2), 143–153. <https://doi.org/10.1016/j.ijpsycho.2003.08.002>
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3), 169–200. <https://doi.org/10.1080/02699939208411068>

- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., Lester, J. C., & Joseph Grafsgaard, Joseph B. Wiggins, Kristy Elizabeth Boyer, Eric N. Wiebe, and J. L. (2013). Automatically recognizing facial expression: Predicting engagement and frustration. *The 6th International Conference on Educational Data Mining EDM 2013*, 43–50.
- Happy, S. L., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing*, 6(1), 1–12. <https://doi.org/10.1109/TAFFC.2014.2386334>
- Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8474 LNCS, pp. 29–38). https://doi.org/10.1007/978-3-319-07221-0_4
- Kasinski, A., Florek, A., & Schmidt, A. (2008). The PUT face database. *Image Processing and ...*, (Table 1), 1–8. Retrieved from http://www.researchgate.net/publication/232085001_The_put_face_database/file/79e4150d0bf1e5080f.pdf
- Kort, B., Reilly, R., & Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. *Proceedings - IEEE International Conference on Advanced Learning Technologies, ICALT 2001*, 43–46. <https://doi.org/10.1109/ICALT.2001.943850>
- 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), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Liu, P., Han, S., Meng, Z., & Tong, Y. (2014). Facial Expression Recognition via a Boosted Deep Belief Network. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1805–1812. <https://doi.org/10.1109/CVPR.2014.233>
- Liu, Y., Sourina, O., & Nguyen, M. K. (2010). Real-time EEG-based human emotion recognition and visualization. In *Proceedings - 2010 International Conference on*

Cyberworlds, CW 2010 (pp. 262–269). <https://doi.org/10.1109/CW.2010.37>

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kande dataset (CK+): A complete facial expression dataset for action unit and emotion specified expression. *Cyprw*, (July), 94–101. <https://doi.org/10.1109/ISIEA.2010.5679500>

Majumder, A., Behera, L., & Subramanian, V. K. (2014). Emotion recognition from geometric facial features using self-organizing map. In *Pattern Recognition* (Vol. 47, pp. 1282–1293). <https://doi.org/10.1016/j.patcog.2013.10.010>

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), 151–160. <https://doi.org/10.1109/T-AFFC.2013.4>

Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6738 LNAI, pp. 534–536). https://doi.org/10.1007/978-3-642-21869-9_93

Saneiro, M., Santos, O. C., Salmeron-Majadas, S., & Boticario, J. G. (2014). Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches. *Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/484873>

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2014.2366127>

Sumathi, C. P., Santhanam, T., & Mahadevi, M. (2012). AUTOMATIC FACIAL EXPRESSION ANALYSIS A SURVEY. *International Journal of Computer Science & Engineering Survey (IJCSSES)*, 3(6). <https://doi.org/10.5121/ijcses.2012.3604>

Vidal-Abarca, E., Gilabert, R., Ferrer, A., Ávila, V., Martínez, T., Mañá, A., ... Serrano, M.-Á. (2014). TuinLEC, an intelligent tutoring system to improve reading literacy skills / TuinLEC, un tutor inteligente para mejorar la competencia lectora. *Infancia Y*

Aprendizaje, 37(1), 25–56. <https://doi.org/10.1080/02103702.2014.881657>

- Wang, S., Liu, Z., Wang, Z., Wu, G., Shen, P., He, S., & Wang, X. (2013). Analyses of a multimodal spontaneous facial expression database. *IEEE Transactions on Affective Computing*, 4(1), 34–46. <https://doi.org/10.1109/T-AFFC.2012.32>
- Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1), 86–98. <https://doi.org/10.1109/TAFFC.2014.2316163>
- Zatarain-Cabada, R., Barrón-Estrada, M. L., González-Hernández, F., & Rodríguez-Rangel, H. (2017). Building a Face Expression Recognizer and a Face Expression Database for an Intelligent Tutoring System. *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on*, 391–393.
- Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., ... Girard, J. M. (2014). BP4D-Spontaneous: A high-resolution spontaneous 3D dynamic facial expression database. *Image and Vision Computing*, 32(10), 692–706. <https://doi.org/10.1016/j.imavis.2014.06.002>