Maestría en Ingeniería Electrónica

Herramienta computacional para estimular la imitación y el reconocimiento de expresiones emocionales faciales en niños con Trastorno del Espectro Autista

Sergio David Pulido Castro

Bogotá, D.C., 15 de julio de 2021

ESCUELA COLOMBIANA DE INGENIERÍA
JULIO GARAVITO
VIGILADA MINEDUCACIÓN

Herramienta computacional para estimular la imitación y el reconocimiento de expresiones emocionales faciales en niños con Trastorno del Espectro Autista

Trabajo de grado para optar al título de magíster en Ingeniería Electrónica, con énfasis en Ingeniería Biomédica

Tutor

Juan Manuel López López

Cotutoras

Sandra Liliana Cancino Suarez

Alejandra Rizo Arévalo

Jurados

Manuel Guillermo Forero Vargas

Laura Andrea León Anhuamán

Bogotá, D.C., 15 de julio de 2021

El trabajo de grado de maestría titulado “Herramienta computacional para estimular la imitación y el reconocimiento de expresiones emocionales faciales en niños con Trastorno del Espectro Autista”, presentado por Sergio David Pulido Castro, cumple con los requisitos establecidos para optar al título de Magíster en Ingeniería Electrónica con énfasis en Ingeniería Biomédica.

Tutor:

Juan Manuel López López

Cotutoras:

Sandra Liliana Cancino Suarez

Alejandra Rizo Arévalo

Bogotá, D.C., 13 de agosto de 2021

AGRADECIMIENTOS

Quiero agradecer a mis papás José Agustín y Luz Ángela, quienes han apoyado todas las decisiones que he tomado en mi vida, de todas las formas en las que alguien puede ayudar a otra persona. Sin ustedes, nada de esto habría sido posible. El amor que siempre me han dado me ha impulsado a seguir adelante con los proyectos que me he propuesto, incluso en los momentos más difíciles. Este proyecto y esta maestría no han sido la excepción y yo no podría pedir unos mejores papás, por lo cual nunca me cansaré de agradecerles todo lo que han hecho por mí.

También quiero agradecer a mi hermana Alejandra, quien me apoyó con distintas ideas para el diseño del videojuego y con quien siempre nos divertimos viendo series o hablando de lo que se nos ocurra. El estrés que podía llegar a sentir cada vez que algo no funcionaba se iba cuando hablaba contigo, muchas gracias.

A mi novia, Daya; tu entendiste más que nadie la frustración que sentí en distintas etapas del proyecto, porque tu viviste situaciones similares. Sé que siempre puedo contar contigo en las buenas y en las malas y juntos vamos a cumplir todos nuestros sueños.

La persona que más me ha ayudado en mi vida profesional es mi tutor, Juan Manuel. Se lo he dicho anteriormente, pero más que un tutor o un profesor, usted es la persona que logró renovar mi amor por la ingeniería. Antes de conocerlo, estaba seguro de que estaba destinado a dedicarme a algo que no quería, pero gracias a usted encontré el amor por el procesamiento de señales e imágenes, que se complementa muy bien con mi amor por la programación. A lo largo de estos años, aprendí que la investigación científica es un campo que, aunque puede ser frustrante, también es bastante gratificante. Jamás podré agradecerle lo suficiente por ayudarme a encontrar mi camino.

Otra persona muy importante para mí es mi cotutora, Sandra. Sandy, tú sabes lo especial que eres para mí. Además de enseñarme muchas cosas sobre el procesamiento de imágenes y la visión artificial, sé que puedo contar contigo a nivel personal, porque irradias positividad y alegras a todas las personas que tienen la suerte de conocerte.

Además, quiero agradecer a mi cotutora, Alejandra, y a la psicóloga Carolina, quienes me aconsejaron sobre la forma correcta de llevar a cabo aquellos aspectos de la tesis con los que yo no era familiar. Sin ustedes, este proyecto nunca habría sido lo que es hoy en día.

Finalmente, quiero agradecer a todos mis compañeros durante este proyecto. A todos los estudiantes de psicología de UNIMINUTO que ayudaron en la creación de este proyecto. Especialmente a Katherine, con quien empezamos este proyecto y ayudó enormemente a darle forma y a Michelle; sin toda tu ayuda, conocimiento, esfuerzo y ganas, jamás habríamos podido sacar a tiempo cada etapa del proyecto, muchas gracias, Michelle. También quiero agradecer a Karen y a Santiago, quienes apoyaron en el desarrollo del protocolo experimental. Pero especialmente quiero agradecer a Nubia, ingeniera en todo menos en nombre (aún). Te lo he dicho miles de veces, pero te lo vuelvo a decir: la ayuda que me diste desde el momento que entraste al proyecto fue fundamental para que saliera con la calidad que tiene. Todos tus consejos y trabajo duro fueron esenciales para que nuestro algoritmo sea lo que es hoy en día.

RESUMEN

El Trastorno del Espectro Autista (TEA) es una condición que afecta el neurodesarrollo, caracterizada por déficits en la comunicación e interacción social y patrones comportamentales repetitivos. Según estudios, 15 de cada 1,000 niños sufren de TEA. Varias barreras existen que impiden el acceso a cuidado médico de individuos con TEA, incluyendo la escasez de servicios de salud, un elevado costo en los servicios y un conocimiento deficiente del trastorno, lo que implica una alta necesidad de crear herramientas tecnológicas que aporten en los procesos socio-comunicativos de individuos con TEA. Por este motivo, en este proyecto se decidió hacer un énfasis en la estimulación de los procesos de imitación y reconocimiento de emociones, un aspecto clave para la comunicación no-verbal. Así, el proyecto se dividió en cuatro grandes etapas: el desarrollo de un protocolo experimental, el desarrollo de un algoritmo de reconocimiento de expresiones faciales emocionales en tiempo real, la creación de un videojuego donde se implementa el protocolo experimental y la ejecución de pruebas experimentales con niños con TEA y un grupo control.

El protocolo experimental, diseñado con ayuda de psicólogas y estudiantes de psicología, se dividió en tres etapas fundamentales: (I) Introducción de las partes del rostro, las emociones y las expresiones faciales, (II) imitación de expresiones faciales y (III) reconocimiento de expresiones faciales en situaciones con y sin contexto. La dificultad de los conceptos enseñados aumenta a lo largo de 12 sesiones experimentales, con las cuales se espera que los participantes aprendan los fundamentos de las seis emociones básicas: alegría, tristeza, enojo, asco, sorpresa y miedo. El algoritmo de reconocimiento de expresiones faciales contó con varias etapas en las que se evaluaron las mejores técnicas de preprocesamiento, extracción de características y clasificación. Adicionalmente, se implementó un ensamble de algoritmos de aprendizaje automático que se enfocó en mejorar la exactitud de reconocimiento de ciertas expresiones faciales. Finalmente, se utilizó una técnica de reducción de características, de forma que se lograra mejorar el tiempo de cómputo del algoritmo, habilitando su uso en tiempo real. El videojuego creado que implementó el protocolo experimental y el algoritmo de reconocimiento contó con retos como la generación de bibliotecas dinámicas para permitir el uso de algoritmos de visión artificial en un motor de videojuegos. Finalmente, se realizaron pruebas con la ayuda de tres niños, de los cuales uno tiene TEA.

El algoritmo de reconocimiento de expresiones faciales mostró resultados positivos, de forma que logró separar siete categorías de expresiones faciales con una efectividad satisfactoria. Por otro lado, la herramienta de estimulación demostró tener un impacto positivo en los participantes del estudio, reflejado por medio de puntajes superiores en pruebas psicométricas y registros conductuales después de realizar la intervención. En trabajos futuros, se espera mejorar la efectividad de reconocimiento del algoritmo de expresiones faciales, por medio del uso de otras técnicas de aprendizaje automático, extracción de distintas características e implementación de algoritmos de caracterización del rostro especializados. Finalmente, se espera que más adelante se logre realizar una validación formal de la herramienta desarrollada para que su uso se estandarice en instituciones prestadoras de salud.

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

I. INTRODUCCIÓN

1.1 Trastorno del Espectro Autista (TEA)

1.1.1 Definición y clasificación de TEA

El Trastorno del Espectro Autista, o TEA, es un trastorno del neurodesarrollo, que corresponde a un grupo de condiciones durante el periodo de desarrollo del individuo, caracterizados por déficits que generan discapacidades a nivel personal y social, entre otros. Una persona diagnosticada con TEA puede llegar a tener discapacidad intelectual. En el caso particular de TEA, se debe hacer una serie de estudios rigurosos, ya que los déficits de comunicación característicos del trastorno no son suficientes para realizar un diagnóstico, dado que estos déficits están acompañados por comportamientos repetitivos excesivos e intereses restrictivos, entre otros [1].

Los criterios diagnósticos para TEA, según la quinta edición del Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-5, por sus siglas en inglés), se dividen en cinco grupos [1], los cuales se resumen en este documento, en búsqueda de concisión:

I. Déficits persistentes en la comunicación e interacción social, a través de múltiples contextos, como una disminución en la reciprocidad socioemocional, en comportamientos comunicativos no-verbales utilizados para la interacción social o en el desarrollo, mantenimiento y comprensión de relaciones sociales.

II. Patrones comportamentales repetitivos y restrictivos, manifestados en movimientos motores estereotípicos, adherencia inflexible a rutinas, interés anormal por objetos específicos, entre otros.

III. Los síntomas se deben presentar durante el desarrollo infantil temprano.

IV. Los síntomas generan discapacidades significativas a nivel social y ocupacional, entre otros.

V. Las anomalías anteriormente descritas no son explicadas de mejor manera por un desorden intelectual del desarrollo o por retardos globales en el desarrollo.

Es importante aclarar que el DSM-5 incluye el síndrome de Asperger y el Trastorno Generalizado del Desarrollo como Trastornos del Espectro Autista, a diferencia de su predecesor, DSM-IV. Este cambio se realizó dado que, en estos casos, el sujeto muestra dificultades en la interacción social y en la comunicación al igual que muestra patrones comportamentales restrictivos y estereotípicos, criterios que encajan en el diagnóstico de TEA. Por este motivo, DSM-5 indica que los individuos que presentan un diagnóstico de síndrome de Asperger o Trastorno Generalizado del Desarrollo serán diagnosticados con TEA [1].

Finalmente, en cuanto a la clasificación de TEA, un tema de importancia para este estudio es el nivel de severidad del trastorno. El DSM-5 clasifica la severidad de TEA en tres grandes categorías:

I. Requerimiento de apoyo

II. Requerimiento de apoyo sustancial

III. Requerimiento de apoyo muy sustancial

El DSM-5 detalla los criterios para cada categoría, a nivel de comunicación social y comportamientos repetitivos. Por ejemplo, un individuo con dificultad extrema para lidiar con el cambio estaría en el nivel 3, indicando que requiere apoyo muy sustancial. Por otro lado, si un individuo tiene dificultad para iniciar interacciones sociales y presenta respuestas atípicas cuando otros inician la interacción social, se encontraría en el nivel 1 [1]. Entender estas categorías es esencial para identificar y evaluar mejoras comportamentales en la aplicación de herramientas de estimulación.

Aparte de la clasificación formal que le ha dado el DSM-5 al trastorno del espectro autista, otros autores argumentan que este trastorno no solo compromete las habilidades sociales y afecta el patrón de conducta

del individuo, sino que genera una dificultad de adaptación en diferentes ámbitos de la vida. Estas alteraciones se suelen ver reflejadas en inquietud motora, problemas de sueño, alimentación y agresión. Con relación a esto, se impone mayor carga a cuidadores y profesionales de la salud involucrados en el desarrollo del individuo [2], lo que refleja la necesidad de sistemas de apoyo que faciliten el cuidado de individuos con TEA .

1.1.2 Prevalencia de TEA

Una revisión de literatura que recopiló varios estimados de prevalencia de TEA encontró una alta variabilidad en los resultados de los estudios llevados a cabo en distintos lugares del mundo. Los autores atribuyeron este fenómeno a diferencias metodológicas en la detección de casos y en un incremento consistente en los aproximados de prevalencia dentro de cada área geográfica [3]. Se encontró que, en 11 estudios realizados en Europa, los cuales incluían niños entre 6 y 8 años, 10.49±8.01 de cada 1,000 niños son diagnosticados con TEA. Esta alta variabilidad está dada principalmente por estudios realizados en Suecia, Dinamarca e Islandia, los cuales reportaron que 17.4, 12.6 y 31.3 de cada 1,000 niños, respectivamente, son diagnosticados con TEA.

La revisión previamente mencionada también encontró que aquellos 10 estudios realizados en Asia que incluían niños entre 6 y 8 años, 15.11±27.97 de cada 1,000 niños son diagnosticados con TEA. El caso más excepcional de esta región fue un estudio realizado en Japón, en el cual se estimó la prevalencia de trastornos del neurodesarrollo a partir de una encuesta brindada a padres y profesores. La prevalencia estimada de TEA fue de 19/1,000 basada en los reportes de los padres, mientras que esta fue de 93/1,000 dada en los reportes de los profesores. Adicionalmente, los reportes indicados por profesores fueron mucho mayores a aquellos de cualquier otro estudio utilizado en la revisión de literatura y la tasa de acuerdo entre padres y profesores fue muy baja, lo que sugiere que el estimado de los profesores pudo estar sobreestimado y, por lo tanto, no confiable. Este argumento está sustentado por otras investigaciones, donde se ha observado un desacuerdo general entre padres y profesores de niños con TEA, al evaluar las limitaciones sociales presentadas por los niños [4], [5].

En la revisión, se encontró que los estudios más consistentes respecto a la zona donde se llevaron a cabo fueron en Norteamérica, dado que, de 7 estudios realizados, se encontró que 13.31±3.81 de cada 1,000 niños son diagnosticados con TEA por profesionales de la salud. Cabe notar que el estudio realizado en México mostró una menor prevalencia que aquellos estudios realizados en Canadá y Estados Unidos. Los estudios encontrados en la revisión de literatura para otras regiones (Oceanía, Medio Este, Centro América, Suramérica y África) fueron muy pocos, por lo cual no se incluyen en este texto.

Es interesante observar que aquellos países con mayor índice de desarrollo humano son aquellos con más prevalencia de TEA (en el caso de Europa son los países nórdicos, en el caso de Asia es Japón y en el caso de Norteamérica son Canadá y Estados Unidos). Tal como indica la revisión de literatura, estas variaciones en los resultados pueden darse por un incremento en la detección de casos por un mayor acceso a servicios de salud o por la existencia de distintos protocolos utilizados en cada país para la detección de casos. Por otro lado, dado a que se ha encontrado evidencia directa de la contribución de factores ambientales en la incidencia de TEA [6], como la migración durante el embarazo o la exposición a drogas, es posible que ciertos países sean más propensos a una alta prevalencia de TEA, porque hay una mayor exposición a los factores ya mencionados.

En el caso de Colombia, la prevalencia de TEA es incierta, dada la falta de estudios formales que estimen esta cifra. El Boletín de Prensa No 079 de 2013 del Ministerio de Salud de Colombia indica que el país no cuenta con cifras oficiales que establezcan la prevalencia del trastorno [7]. Aunque estudios posteriores realizados en Colombia tienen en cuenta los trastornos del desarrollo para obtener estadísticas de salud mental, estos trastornos son agrupados, por lo cual no se obtienen cifras de la prevalencia de TEA. El Boletín de Salud Mental – Oferta y Acceso a Servicios en Salud Mental en Colombia de 2018 [8] indica que en el año 2017 se atendieron 116,878 personas por trastornos del desarrollo psicológico en servicios de salud. Es importante resaltar que esta categoría engloba al trastorno generalizado del desarrollo, el cual contiene varios espectros del autismo. Teniendo en cuenta las proyecciones de población del Departamento Administrativo Nacional de Estadísticas (DANE) [9], esto indicaría que 0.2389% de la población fue atendida por trastornos del desarrollo psicológico, lo cual equivale a que 2.39 de cada 1,000

personas son atendidas en centros de salud por estas condiciones. No obstante, estas estadísticas se basan en supuestos y no cuentan con la exactitud requerida para conocer la prevalencia de TEA en Colombia, por lo cual es importante que, en el futuro, existan cifras apropiadas en el país para indicar la prevalencia de este trastorno del neurodesarrollo.

1.1.3 Condiciones familiares y sociales de personas con TEA

Las familias de individuos con TEA deben enfrentar muchos retos al tener que cuidarlos durante todas sus vidas, principalmente si se trata de individuos que requieran apoyo sustancial, según el DSM-5. Los familiares cumplen un rol importante en el tratamiento de sus hijos, ya que deben buscar terapias adecuadas, pagar por el cuidado, tener una comunicación constante con especialistas y dar apoyo durante los tratamientos [10]. Es importante reconocer las necesidades de estas familias, de forma que sea posible entender la magnitud de los beneficios de generar herramientas que aporten en la comunicación de los individuos con TEA. Una investigación analizó las barreras comunes que existen en el acceso al cuidado médico de individuos con TEA, quienes dividieron la problemática en seis pilares principales [11]:

- Escasez de servicios de salud

  La investigación presenta que, en Estados Unidos, hay entre 0.2 y 4 pediatras especializados en desarrollo y comportamiento por cada 100,000 niños, lo cual indica escasez de suministros, desgaste de los médicos y largos tiempos de espera para recibir diagnóstico y tratamiento. De igual forma, los investigadores señalan que aquellos familiares sin acceso a estos servicios especializados dentro de su comunidad son menos propensos a buscar apoyo, porque implica un mayor costo y esfuerzo.

- Conocimiento de los médicos

  Los investigadores del artículo también aseguran que muchos individuos con TEA, sus familiares y médicos reportan que los profesionales de la salud que atienden al paciente no suelen tener conocimiento especializado para diagnosticar y referir sujetos con TEA. De igual forma, estudiantes de medicina de Estados Unidos reportan no recibir entrenamiento suficiente para el tratamiento de niños con TEA. Los programas que se han implementado en distintos países para aumentar la consciencia de los médicos en el adecuado tratamiento de TEA han demostrado ser eficaces, porque logran que los pacientes tengan acceso a servicios de 2 a 6 meses antes, aumentando la efectividad de tratamiento.

- Costo de los servicios

  En el artículo se analizan los costos del tratamiento de TEA en Estados Unidos, indicando que, durante los primeros 5 años de vida, el costo anual de los servicios para tratar TEA es de aproximadamente USD6467, incrementando a USD9053 durante el resto de la infancia y adolescencia. Adicionalmente, los servicios de TEA suelen ser excluidos de planes de seguros brindados a familias de bajos recursos, por lo cual no todos los pacientes pueden recibir un tratamiento adecuado.

- Conocimiento familiar e individual

  El conocimiento respecto a los síntomas de TEA y las posibles opciones de tratamiento son fundamentales para que más personas tengan acceso a los servicios de salud ofrecidos. Sin embargo, este conocimiento está afectado, no solo por factores sociales, sino situacionales. En el artículo se da un ejemplo donde los padres cuyo primogénito tiene TEA son menos propensos a identificar cambios emocionales en sus hijos respecto a padres que ya tengan experiencia en la crianza.

- Idioma

  El estudio encontró que aquellas familias que viven en un país cuya lengua no dominan encuentran una barrera adicional al momento de acceder a servicios médicos, dado que la comunicación con

médicos y el cumplimiento de protocolos administrativos se dificulta. Esto es particularmente cierto en países de Norteamérica y Europa, donde la migración es mayor.

- Estigma

  Un obstáculo general cuando se tiene un trastorno mental es la presencia de estigmas que contribuyen a sentimientos de rechazo entre padres de niños con TEA. La investigación indica que algunos padres inmigrantes de ciertas culturas a Estados Unidos, Canadá y Reino Unido no suelen reconocer que su hijo tiene TEA por estos estigmas, lo que evita que estos individuos tengan acceso al sistema de salud durante la totalidad de su vida.

Aparte de las cifras que se obtienen respecto a las barreras comúnmente conocidas para las personas con TEA, los autores de otro estudio buscaron reconocer las necesidades de niños con TEA y sus familiares a partir de las perspectivas de padres y especialistas [12]. En este estudio cualitativo, se llevaron a cabo entrevistas con 19 especialistas y 23 padres de niños con TEA, en las cuales se hicieron preguntas clave para identificar sus principales necesidades. El estudio encontró que la mayor preocupación de las personas involucradas con el desarrollo de niños con TEA es que muchos padres no tienen el conocimiento necesario para atender a un niño con esta condición, incluyendo su definición, síntomas, servicios requeridos, entendimiento y manejo del comportamiento. Otra gran obligación que tienen los padres es la necesidad de entender las habilidades en las que existe un déficit e intentar transmitirlas de la mejor forma posible; esto incluye destrezas de comunicación social, manejo del comportamiento, técnicas para lidiar el estrés y la ansiedad. Además de esto, los especialistas consideran que una de las competencias que les suele faltar a los padres es saber jugar con ellos, ya que aquellos elementos que les interesan pueden ser muy distintos de los niños neurotípicos. Adicionalmente, especialistas y padres estuvieron de acuerdo en que es necesario que los padres deben tener una presencia activa en la comunidad, de forma que el apoyo brindado también provenga de otros familiares, amigos y la comunidad en general. Finalmente, algunos de los participantes indicaron las competencias educativas (los padres deben enseñarle todo lo que puedan a los niños, teniendo en cuenta la edad de estos), requerimiento de servicios de salud mental (especialistas que apoyen en la salud mental de los cuidadores) y necesidades financieras (apoyo gubernamental y cubrimiento de seguros, dado el alto costo de los servicios de tratamiento). Un punto de interés en esta investigación es que los especialistas entrevistados resaltaron la importancia por parte de los padres de reconocer aquellos tratamientos que utilizan métodos científicos y aquellos que no, dado que es importante que las terapias y estímulos a recibir le ayuden en su desarrollo neuronal para potenciar sus habilidades sociales y comunicativas.

Por otro lado, dado que los individuos con TEA tienen habilidades socio-comunicativas comprometidas, cuentan con retos adicionales al momento de tener interacciones sociales con otras personas. Un estudio en particular exploró el impacto de la complejidad social en el desarrollo de habilidades socio-comunicativas consecuentes [13]. Para realizar esta exploración, el estudio utilizó un paradigma de seguimiento ocular con 63 niños en edad preescolar (26 niños neurotípicos y 37 niños con TEA). En el experimento, se pidió a estos niños observar dos situaciones por medio de videos: En la primera, se mostraba a dos niños jugando de manera independiente con un xilófono (condición paralela), mientras que en la segunda se presentaba a los niños interactuando y jugando con un mismo xilófono (condición interactiva), a la vez que se registraban los puntos de fijación de los participantes en el video. Los resultados mostraron una disminución significativa en el tiempo donde los participantes con TEA mantuvieron la mirada en los niños del video, respecto al tiempo utilizado por los niños neurotípicos. Interesantemente, se encontró que, en ambos grupos, las características de exploración visual son moduladas por el contexto, de manera que la condición que se presentó en los videos afectó tanto a ambos grupos de participantes.

Finalmente, es importante indicar que, aunque la investigación realizada por nuestro equipo de investigación está enfocada en impactar en la imitación y reconocimiento de emociones en niños entre 6 y 8 años, los individuos con TEA suelen tener obstáculos laborales en su vida adulta. Un estudio realizado en Estados Unidos encontró que solo el 53.4% de adultos jóvenes con TEA han tenido un trabajo remunerado sin contar trabajo en casa desde que salieron del colegio, indicando un salario promedio de 8.10 dólares estadounidenses por hora, que es significativamente menor al de un adulto joven neurotípico [14]. Otro estudio realizado en el mismo país encontró que 34.7% de los individuos con TEA han tenido

educación superior y 55.51% han tenido un trabajo remunerado durante los primeros 6 años después del colegio [15]. Estas estadísticas de desempleo son evidentemente mayores a aquellas de individuos neurotípicos.

Un estudio examinó las características de empleo de adultos jóvenes con TEA y los factores que contribuyeron a su estado de empleo a partir de una encuesta en línea y la escala corta de Desequilibrio esfuerzo-recompensa (ERI, por sus siglas en inglés). Este estudio contó con la participación de 254 adultos con TEA, de los cuales 38.58% se encontraban en situación de desempleo. El estudio encontró que más de la mitad de los participantes reportaron una baja recompensa a su esfuerzo según la escala ERI y la gran mayoría no recibe ningún tipo de asistencia en el trabajo. Por otro lado, aquellos participantes que ocultan su diagnóstico de TEA a empleadores son tres veces más propensos a obtener un trabajo que aquellos que no lo ocultan. Una revisión de literatura encontró que 17 factores fueron identificados para facilitar u obstaculizar la obtención de trabajo por parte de individuos con TEA [16]; sin embargo, el único factor consistente a partir de los estudios revisados fue una habilidad cognitiva limitada. No obstante, otros factores influyentes según varios de los estudios incluyen la severidad del trastorno, comportamientos inapropiados y la presencia de discapacidades sociales.

1.2 Emociones

1.2.1 Concepto de las emociones

Las emociones son estados de preparación que varían ampliamente en cuanto a su fisiología y el comportamiento evocado por las mismas [17]. No obstante, todas las emociones se expresan a partir de respuestas motoras somáticas, especialmente movimientos musculares de la cara. Tradicionalmente, los centros neuronales que coordinan las respuestas emocionales se han agrupado dentro del sistema límbico; sin embargo, más recientemente se ha encontrado que varias regiones cerebrales cumplen un rol importante en el procesamiento de emociones, incluyendo la amígdala y áreas corticales del lóbulo frontal. Los cambios fisiológicos en las imágenes incluyen cambios en la frecuencia cardiaca, flujo sanguíneo, sudoración y motilidad gastrointestinal, los cuales se afectan por cambios en la actividad de componentes simpáticos y parasimpáticos del sistema motor [18].

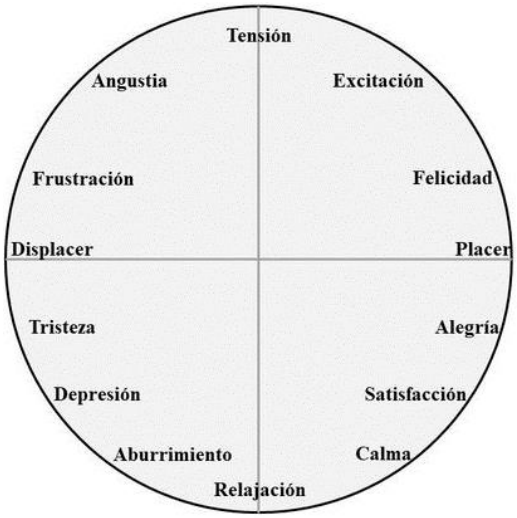

Figura 1. Distintas emociones ubicadas en el modelo circunflejo de Russell. El eje horizontal se refiere a la valencia y el eje vertical se refiere a la activación [19].

Respecto a la clasificación de emociones a partir de reacciones fisiológicas, la comunidad científica concuerda respecto a la forma de clasificar las emociones. Esta forma es el Modelo *Circunflejo de Russell*, el cual se enfoca en dos dimensiones para clasificar las emociones. El autor de este modelo, James Russell, las llamó en su artículo original la dimensión del “agrado-desagrado” y la dimensión de la excitación. Sin embargo, hoy en día se conocen como valencia y activación, respectivamente. La valencia se refiere al

atractivo o aversión hacia una emoción, de forma que la alegría tiene valencia positiva y la tristeza tiene valencia negativa. Por su parte, la activación indica la cantidad de actividad fisiológica que se experimenta con una emoción. Así, aunque el enojo y el aburrimiento tengan valencia negativa, el enojo tendrá activación positiva mientras que el aburrimiento tendrá activación negativa [20]. En la Figura 1 se puede observar el modelo circunflejo con 28 emociones ubicadas en él, según el experimento original de Russell.

Por último, es importante tener en cuenta que, aunque existen un sinnúmero de emociones que pueden ser identificadas a partir de su valencia y su activación, Ekman definió seis emociones básicas y, desde entonces, ha sido un punto de referencia para la investigación en emociones y expresiones faciales. Estas emociones son: alegría, asco, tristeza, enojo, sorpresa y miedo [21].

1.2.2 Emociones y TEA

Tal como se mencionó anteriormente, uno de los principales síntomas de TEA es la reducción en la complejidad en las interacciones sociales y la presencia de interacciones anormales. Una parte importante de este aspecto es el bajo control emocional y las alteraciones en la respuesta emocional. Un fenómeno que puede describir este comportamiento particular de los niños con TEA es una baja regulación emocional. La regulación emocional es la habilidad de afrontar de manera adecuada la experiencia de emociones, de manera que una baja regulación emocional se puede manifestar como la intensificación excesiva de las emociones y la desactivación excesiva de las mismas [22]. Los individuos con TEA pueden llegar a carecer de habilidades para regular sus emociones, de forma que actúen de manera violenta hacia sí mismos o hacia otros, como lo demuestra un estudio sobre el manejo de emociones de niños diagnosticados con síndrome de Asperger [23].

Por otro lado, se ha encontrado que niños con TEA que son expuestos a situaciones levemente frustrantes son menos propensos a mantener la concentración, tener control inhibitorio y mantenerse calmados respecto a niños neurotípicos [24]. Aunque no se tiene un consenso, se considera que una posible explicación para la regulación de emociones inadecuada en niños con TEA es que la coocurrencia de trastornos psiquiátricos es la responsable de este déficit [25]. Sin embargo, otros autores afirman que la regulación de emociones inadecuada es intrínseca de TEA y es la causante de que aumente el riesgo de desarrollar otros trastornos psicológicos, por lo cual consideran necesario que la investigación de emociones en TEA se enfoque en acercamientos multimodales, en los cuales se combinen medidas fisiológicas o neuronales con medidas comportamentales [26]. Finalmente, se ha estudiado la relación del Potencial Positivo Tardío con la habilidad de modular respuestas emocionales; por lo cual, algunos autores lo consideran un marcador neurofisiológico relevante para la regulación emocional en niños [27].

1.2.3 Expresiones faciales de las emociones

Un aspecto de particular importancia para este estudio es la relación de las emociones con las expresiones faciales. El estudio de las expresiones faciales de las emociones inició hace aproximadamente 100 años; sin embargo, fue hasta 1964 donde se demostró que un observador es capaz de reconocer nueve emociones a partir de expresiones faciales presentadas en fotografías. No obstante, en este mismo estudio también se encontró que algunas emociones se pueden confundir con otras de manera sistemática y, de igual forma, algunas emociones podrían ser confundidas dependiendo de los sesgos personales del participante [28].

Por su parte, Ekman estudió el problema de la universalidad en 1969. Allí, analiza si observadores de distintas culturas denominan de la misma forma ciertas expresiones faciales. El estudio buscó categorizar expresiones faciales según su emoción a partir de imágenes tomadas de individuos de seis culturas distintas, de las cuales dos habían tenido mínimo contacto con el mundo occidental. Se encontró que la alegría, el enojo y la tristeza son emociones fácilmente reconocibles independientemente de la cultura de la que provenga la persona que está realizando la expresión facial. En el caso del asco, la sorpresa y la tristeza se observó que al menos dos de cada tres imágenes fueron identificadas correctamente en todas las culturas estudiadas [29]. Ekman argumenta que una limitación de este tipo de estudios transculturales es que las expresiones mostradas no son genuinas, dado que se pide a los sujetos que hagan la expresión

correspondiente a cierta emoción, por lo cual es posible que la universalidad en los juicios sobre la expresión facial se limite a expresiones estereotipadas [21]. Otro aspecto interesante respecto a la universalidad de las emociones se observa en un estudio realizado por Friesen [30], en el cual se observó que sujetos japoneses y estadounidenses mostraban expresiones faciales sin diferencias significativas entre las culturas cuando se les mostraban películas que producían estrés. Sin embargo, se observó que los sujetos japoneses mostraban un mayor control de la expresión facial cuando una persona con más autoridad estaba presente. Por su parte, Matsumoto buscó diferencias culturales en la calificación dada a la intensidad de una emoción a partir de la expresión facial. Para este estudio se contó con la ayuda de participantes estadounidenses y japoneses; en general, se encontró que los sujetos estadounidenses marcan con una mayor intensidad a todas las emociones menos asco. De igual forma, los sujetos estadounidenses indicaron la alegría y el enojo como las emociones con mayor intensidad, mientras que los sujetos japoneses señalaron el asco como la emoción con mayor intensidad. Sin embargo, se encontró que, en general, las diferencias en intensidades eran consistentes para cada emoción entre ambas culturas [31].

Teniendo en cuenta los estudios que confirman la universalidad de las expresiones faciales de las emociones, es posible obtener una descripción detallada de la generación de reacciones musculares a partir de cada una de estas. Ekman detalla estas características a partir de 3 secciones en el rostro: Cejas/frente, ojos/párpados y parte inferior del rostro, de forma que describe las características de estos elementos para cada una de las seis emociones básicas [32]. En la Tabla 1 se puede observar un resumen de las descripciones propuestas por Ekman en su investigación.

Tabla 1. Descripción de la expresión facial en cada una de las emociones básicas según Ekman [32].

Emoción	Parte del rostro	Comportamiento
Alegría	Cejas/frente	Cejas neutrales
	Ojos/párpados	Ojos neutrales o relajados
		Párpado inferior elevado causando que se entrecierran los ojos
	Parte inferior del rostro	Líneas naso labiales de las mejillas
		Bordes de los labios elevados
		Se puede mostrar o no dientes
Miedo	Cejas/frente	Cejas elevadas y juntas o aplanadas
	Ojos/párpados	Ojos abiertos
		Aparente tensión en párpados inferiores
		Puede que se vea la esclera superior de los ojos
	Parte inferior del rostro	La boca puede estar abierta o cerrada
		Borde de la boca estirada, sin hacer curva
Asco	Cejas/frente	Cejas abajo pero no juntas
		Puede haber ceño fruncido o no
		La nariz se arruga
	Ojos/párpados	Párpado inferior elevado y no tensionado
	Parte inferior del rostro	Líneas nasolabiales bastante pronunciadas
		Si la boca está abierta: labio superior hacia arriba, labio inferior hacia afuera
		Si la boca está cerrada: labio superior hacia arriba
		Se puede mostrar o no la lengua
Tristeza	Cejas/frente	Cejas juntas
		Esquina exterior de cejas hacia abajo

		Esquina interior de cejas elevada
		Puede haber ceño fruncido o no
	Ojos/párpados	Ojos brillantes
		Párpados superiores caídos
		Párpados inferiores relajados
		Ojos mirando hacia abajo o con lágrimas
	Parte inferior del rostro	Si la boca está abierta: labios parcialmente estirados
		Si la boca está cerrada: tiene una curvatura hacia abajo
Sorpresa	Cejas/frente	Cejas curvadas hacia arriba
		Arrugas horizontales en la frente
	Ojos/párpados	Ojos bien abiertos, mostrando la esclera arriba y abajo
		Piel estirada en los párpados
	Parte inferior del rostro	Boca abierta sin tensión en las esquinas de los labios
Enojo	Cejas/frente	Cejas hacia abajo y hacia adentro
		Parte interior de las cejas sobresale
		Puede tener arrugas curvas en el centro de la frente
	Ojos/párpados	No se muestra la esclera
		Ojos entrecerrados
		Se puede presentar un movimiento rápido de los ojos
	Parte inferior del rostro	Los labios pueden estar apretados
		Labio superior elevado
		Se puede mostrar o no dientes

1.3 Antecedentes

La elaboración de la herramienta para imitación y reconocimiento descrita en este proyecto se basa en una colaboración multidisciplinaria entre la Escuela Colombiana de Ingeniería Julio Garavito y la Corporación Universitaria Minuto de Dios UNIMINUTO, la cual inició en el año 2019 con el proyecto de grado de la ingeniera Dayana Verdugo. Verdugo desarrolló una interfaz gráfica en el entorno MATLAB como una herramienta para apoyar los procesos de estimulación en el reconocimiento facial de emociones básicas [33]. Para el desarrollo, tomó como base las teorías planteadas por Ekman, y particularmente, la descripción de las expresiones faciales que él brinda [32], para caracterizar cada parte del rostro a partir de técnicas de procesamiento de imágenes, como la detección de rostros y el análisis de distintos espacios de color.

El desarrollo de Verdugo fue de la mano del diseño de un protocolo de estimulación en el proceso de imitación e identificación de expresiones faciales, liderado por integrantes del Semillero de Neurociencias Básica y Clínica [34], [35]. Los resultados de esta investigación multidisciplinaria fueron satisfactorios, dado que fue posible desarrollar un sistema de reconocimiento de expresiones faciales que pudo ser probado con ayuda de un niño con TEA. No obstante, Verdugo argumenta que, para trabajos futuros, es necesario tener en cuenta el uso de transformaciones que permitan analizar la expresión facial independientemente de la escala, posición o rotación del rostro, dado que esto habilitará una evaluación más precisa de las expresiones faciales. Por otro lado, Verdugo considera importante contemplar el uso de aprendizaje automático en el reconocimiento de expresiones faciales durante el desarrollo de la interfaz gráfica, ya que

una implementación adecuada puede permitir una segmentación más coherente y precisa que los algoritmos que se basan únicamente en procesamiento de imágenes clásico.

1.4 Justificación

Teniendo en cuenta los obstáculos a los que se enfrentan los individuos con TEA y sus familias, tal como se ha estudiado en investigaciones previas [11], [12], es claro que la población con TEA requiere herramientas innovadoras que suplan las necesidades expuestas anteriormente; teniendo en cuenta que éstas deben estar dirigidas no únicamente a las necesidades de los niños, sino también a los padres. De igual forma, es importante entender que el desarrollo adecuado de la comunicación a edades tempranas es esencial, ya que las personas que tienen habilidades socio-comunicativas comprometidas tendrán una mayor dificultad para desarrollarlas más adelante, dados los retos adicionales que conlleva mantener una interacción social con otras personas [13]. Adicionalmente, considerando las estadísticas de empleo de adultos jóvenes con TEA [14], [15], es importante que cuenten con herramientas para mejorar la calidad de vida. Considerando que una de las causas principales por las cuales los adultos jóvenes con TEA no logran conseguir o mantener un trabajo es la existencia de déficits en la comunicación [16], la creación de herramientas que permitan el mejoramiento de la comunicación no-verbal aportan en la independencia futura de niños con este trastorno.

Por otro lado, la adecuada expresión de emociones es una parte fundamental de la comunicación, por lo que la correcta regulación de las emociones es necesaria para que se fortalezcan las habilidades socio-comunicativas y sea posible eliminar las conductas violentas hacia otras personas o autoinfligidas, mejorar la concentración y logre un control inhibitorio aumentado [23], [24]. Por último, tal como dice Mazefsky en su investigación sobre el rol de la regulación de emociones en TEA, es necesario que se base en acercamientos multimodales, donde se combinen medidas fisiológicas con medidas comportamentales [26].

Basado en las investigaciones previas realizadas sobre TEA, emociones y expresiones faciales, teniendo en cuenta el contexto y las necesidades de individuos con TEA y sus familiares, además de la experiencia previa que tiene nuestro equipo de investigación en el diseño y desarrollo de herramientas para la estimulación de la imitación y el reconocimiento emocional a partir de expresiones faciales [33]–[35], se elaboró una herramienta didáctica que permite estimular imitación y reconocimiento de emociones faciales para niños con TEA entre 6 y 8 años, tomando como referencia las seis emociones básicas propuestas por Ekman [21]. Para cumplir este objetivo principal, se realizaron tres tareas principales:

- Se diseñó un algoritmo capaz de reconocer la emoción que está evocando o imitando a una persona a partir de la expresión facial de la misma. Esto se desarrolló a partir de métodos de procesamiento de imágenes, extracción de características y aprendizaje automático, utilizando una base de datos como línea base para el reconocimiento de las expresiones faciales. Basado en estudios previos, los cuales indican que las expresiones faciales son universales [21], [29]–[31], se puede concluir que el uso de bases de datos cuyos participantes provengan de distintas culturas es válido para determinar la emoción correspondiente a la expresión facial de una persona. Por otro lado, tomando en cuenta las recomendaciones de Verdugo, el algoritmo de reconocimiento de expresiones faciales se diseñó de forma que el participante tenga una mayor flexibilidad, permitiéndole estar a distintas distancias de la cámara, posiciones y rotaciones.

- Se implementó un protocolo experimental basado en aspectos psicológicos y del desarrollo neuronal de TEA, elaborado en conjunto con integrantes del Semillero de Neurociencias Básica y Clínica de la Corporación Universitaria Minuto de Dios UNIMINUTO. Este protocolo busca estimular el proceso de imitación y reconocimiento de expresiones faciales, favoreciendo el aprendizaje e independencia en la vida de niños con este trastorno. La implementación de este protocolo se basó en el desarrollo de un videojuego que, a partir de su naturaleza interactiva, facilitará el aprendizaje y mejorará la efectividad de este. Este videojuego se desarrolló en la plataforma Unity, que es una mejora respecto a la herramienta desarrollada por Verdugo, dado que se trata de una plataforma de acceso libre. Esto permite la creación de archivos ejecutables que pueden ser utilizados de forma gratuita por cualquier usuario con acceso a ellos, eliminando la barrera económica que impone el uso de herramientas como MATLAB.

- Se evaluó la efectividad de la herramienta desarrollada de manera multimodal. Por un lado, se hizo uso de pruebas psicométricas (particularmente, la subprueba de reconocimiento de expresiones faciales de la Evaluación Neuropsicológica Infantil ENI) para identificar una línea base respecto a las habilidades comunicativas de los participantes y observar su progreso. Por otro lado, se realizó un análisis cuantitativo de los gestos faciales de los sujetos, indicando aquellas emociones cuya imitación se dificulta más y realizando una comparación inter-sujeto que permita analizar diferencias en velocidad de aprendizaje y cambios significativos entre niños con TEA y niños neurotípicos.

1.5 Organización general del documento

En la sección II se plantea el objetivo principal de esta investigación y los elementos particulares que llevaron al cumplimiento de este objetivo. En la sección III se evalúan las investigaciones previas en el área de la neurofisiología y de la ingeniería cuyas contribuciones aportaron en el desarrollo de esta investigación. Estas incluyen el diseño de protocolos experimentales para la estimulación en la imitación y reconocimiento de expresiones faciales, el uso de medidas psicométricas para la evaluación emocional, el uso de técnicas de procesamiento de imágenes y aprendizaje automático en la interacción entre humanos y computadores y el desarrollo de herramientas tecnológicas para el apoyo en los procesos de estimulación de TEA.

En la sección IV se describe la metodología utilizada para cumplir los objetivos de este estudio, teniendo en cuenta de que se trata de una solución multidisciplinaria a una problemática de interés global. Así, se describe el videojuego desarrollado, junto a cada uno de los elementos que lo componen. De igual forma, se detallan las etapas realizadas a partir de técnicas de procesamiento de imágenes, visión artificial y aprendizaje automático para el reconocimiento de expresiones faciales. Adicionalmente, se hace una descripción detallada del protocolo experimental diseñado en colaboración con la Corporación Universitaria Minuto de Dios UNIMINUTO, donde se observan los estímulos presentados para el fortalecimiento de los procesos de imitación y reconocimiento. Por otro lado, se muestran los métodos utilizados para evaluar la efectividad de la herramienta diseñada, dentro de la que se encuentran la implementación de medidas psicométricas y el análisis en la efectividad del reconocimiento de expresiones faciales. Por último, se describe la población objetivo de nuestro estudio, detallando la población objetivo, los criterios de inclusión y exclusión y la muestra utilizada.

En la sección V se detallan los resultados obtenidos en la investigación y se hace un análisis de estos. Los resultados obtenidos se dividen en aquellos obtenidos a partir de aspectos técnicos y aquellos obtenidos a partir de las pruebas de psicometría aplicadas. En cuanto a los aspectos técnicos, se evalúa la efectividad del algoritmo de reconocimiento emocional desarrollado y se hace un análisis de los productos creados dentro del proyecto y lo que esto implica para la investigación en procesos de estimulación y la elaboración de proyectos multidisciplinarios en los que se involucra la ingeniería y las neurociencias. En cuanto a las medidas psicométricas, se realiza un análisis de la evolución de las habilidades de reconocimiento de expresiones faciales de los participantes del proyecto y la comparación en la evolución entre los distintos participantes, haciendo énfasis en las diferencias entre niños con TEA y niños neurotípicos. Finalmente, en la sección VI se hace una recapitulación de cada etapa del trabajo desarrollado, indicando los aspectos sobresalientes de este y los resultados obtenidos. Con base en esta información, se detallan aquellos elementos del proyecto cuya implementación se podría mejorar en un futuro y se exponen las consideraciones que, a juicio propio, se deben tener en cuenta para avanzar en la investigación del desarrollo de herramientas de estimulación para el fortalecimiento de los procesos de imitación y reconocimiento en población con TEA.

II. OBJETIVOS

2.1 Objetivo general

Generar una herramienta didáctica de reconocimiento de emociones para la estimulación emocional en niños con trastorno del espectro autista

2.2 Objetivos específicos

- Diseñar un algoritmo de reconocimiento de gestos faciales en tiempo real por medio del procesamiento digital de imágenes
- Implementar un protocolo experimental basado en los aspectos psicológicos y del desarrollo neuronal del trastorno del espectro autista
- Evaluar la efectividad de la herramienta diseñada por medio del uso de pruebas psicométricas y análisis cuantitativo de los gestos faciales de los sujetos

III. ESTADO DEL ARTE

3.1 Aspectos neurofisiológicos

3.1.1 Protocolos experimentales para el apoyo a individuos con TEA

Como se argumentó en la anterior sección de este documento, los niños con TEA cuentan con habilidades cognitivas sociales distintas a aquellas de niños neurotípicos. A lo largo de los años, varias investigaciones han buscado mejorar estas habilidades cognitivas a partir del uso de protocolos experimentales que sean capaces de adaptar el comportamiento de los individuos. Una de estas investigaciones es la realizada por Laugeson [36], quien describió uno de los programas más utilizados hoy en día para enseñar habilidades sociales a adolescentes con TEA, llamado *The Program for the Education and enrichment of Relational Skills* (PEERS), basado en los principios de la Terapia de Cognitivo-Conductual, o CBT, por sus siglas en inglés. Este programa se basa en ciertos fundamentos:

- Uso de tratamiento grupal, el cual involucra enseñanza a un grupo pequeño de adolescentes con TEA. Laugeson afirma que esta modalidad de enseñanza es al menos tan efectiva como la enseñanza individual.
- Adecuación de lecciones didácticas que hacen énfasis en una serie de reglas y pasos específicos para desarrollar habilidades sociales adecuadas.
- Creación de juegos de roles, los cuales incluyen demostraciones buenas y malas de lo que se considera un comportamiento adecuado.
- Uso de estrategias cognitivas, las cuales incluyen la enseñanza de técnicas de solución de problemas sociales.
- Implementación de prácticas comportamentales, las cuales se enfocan en la práctica constante de situaciones sociales.
- Ofrecimiento de realimentación en el desempeño comunicacional a partir de asesoramiento social.
- Entrega y revisión de tareas.
- Acompañamiento continuo por parte de los padres en el tratamiento.

La efectividad de este programa ha sido evaluada en distintos estudios. Una investigación buscó probar esta efectividad a partir de varias medidas, como el Cociente de Espectro Autista (AQ, por sus siglas en inglés), la Prueba Breve de Inteligencia de Kaufman (K-BIT, por sus siglas en inglés), la Escala de Sensibilidad Social (SRS, por sus siglas en inglés) y el Cuestionario de Calidad de Socialización [37]. El estudio contó con 22 adultos jóvenes con TEA, 12 de ellos fueron asignados a tratamiento por medio del programa PEERS y los otros 10 fueron asignados a tratamiento retardado 16 semanas después (grupo control). Se encontró que los participantes del grupo con tratamiento inmediato tuvieron una mejora significativa en el uso de habilidades sociales adecuadas, frecuencia de participación en interacciones sociales, conocimiento de habilidades sociales y se redujeron sus síntomas relacionados con sensibilidad social. No obstante, los participantes del grupo control también mostraron mejoras significativas en la mayoría de estas habilidades, indicado principalmente por la SRS.

Por su parte, Rabin estudió la posibilidad de implementar el programa PEERS en otro idioma para probar su efectividad transcultural [38]. Particularmente, el programa fue implementado en hebreo, contando con la ayuda de 82 adolescentes de habla hebrea con TEA. De manera similar que en el estudio desarrollado por Laugeson, los participantes fueron ubicados de forma aleatoria en uno de dos grupos: intervención inmediata o intervención retardada. Para medir la efectividad del método, se utilizaron distintas medidas, como la Evaluación Contextual de Habilidades Sociales (CASS, por sus siglas en inglés) y la Prueba de Conocimiento de Habilidades Sociales de Adolescentes (TASSK, por sus siglas en inglés). Los resultados del estudio mostraron que ambos grupos mostraron resultados positivos similares en el incremento de las habilidades sociales de los adolescentes en el ámbito familiar. Sin embargo, los resultados reportados por los profesores no tuvieron diferencias significativas entre las mediciones realizadas antes y después de la intervención. Los investigadores consideran que la intervención puede no ser lo suficientemente efectiva para generalizarse en el ambiente de los colegios o que las mediciones pueden no haber sido las adecuadas para identificar estos cambios. En general, Rabin concluye que la versión adaptada al hebreo

del programa PEERS es una intervención efectiva para adolescentes con TEA, resaltando la importancia del acompañamiento por parte de padres en cualquier intervención en la que se trate TEA.

Al igual que con adolescentes, se han diseñado protocolos experimentales para la intervención de niños con TEA en la primera infancia. Uno de estos es el *Early Start Denver Model* (ESDM), una intervención que busca disminuir los principales déficits observados en niños en estas edades: orientación social, atención, desarrollo del lenguaje e imitación, entre otros [39]. Para probar su efectividad, Dawson diseñó un ensayo controlado y aleatorizado, el cual contó con el apoyo de 48 niños diagnosticados con TEA entre 18 y 30 meses [40]. Los participantes fueron divididos en dos grupos: intervención por parte de terapeutas entrenados a partir del modelo ESDM o intervención por medio de métodos tradicionales, aplicada por parte de proveedores dentro de la comunidad del infante (grupo control). En el estudio se encontró que, al compararse con niños dentro del grupo control, los participantes que recibieron intervención por medio de ESDM mostraron mejoras significativas en coeficiente intelectual, comportamiento adaptativo, lenguaje y diagnóstico de autismo. Además, a lo largo de dos años, el grupo con intervención por medio de ESDM continuó de manera progresiva su mejora, mientras que el grupo control tuvo mayores retardos en el comportamiento adaptativo.

Por otro lado, expertos en el área han estudiado terapias con modalidades distintas a las tradicionales. Una de estas es la equitación terapéutica, en la cual los participantes realizan equitación para mejorar ciertos síntomas de TEA, incluyendo aspectos emocionales de la cognición social. Para probar la efectividad de este método, Gabriels diseñó un ensayo controlado y aleatorizado en el cual 127 niños entre 6 y 16 años diagnosticados con TEA tuvieron uno de dos tratamientos: equitación terapéutica o actividad en una granja sin el uso de caballos (grupo control) [41]. Para medir la efectividad de estos métodos, se midió la irritabilidad, hiperactividad, cognición y comunicación sociales a partir de cuestionarios dados a médicos y padres antes y después de las terapias. De igual forma, se midió el número total de palabras y el número de palabras nuevas dichas por los niños durante una muestra estandarizada de lenguaje. El estudio encontró que ambos grupos mejoran significativamente en todos estos aspectos. Sin embargo, los participantes que se encontraban en equitación terapéutica tuvieron mejoras estadísticamente significativas en estos aspectos respecto a los participantes en el grupo control.

Como se puede observar, tradicionalmente han existido varios tipos de intervenciones para mejorar las habilidades sociales de individuos con TEA, junto a otros síntomas, con resultados variados. Es importante tener en cuenta que la mayoría de estos estudios han sido evaluados con sujetos que no se encuentran dentro de la población objetivo de este estudio, ya que en esta investigación se trabaja con niños entre 6 y 8 años. La dificultad para encontrar estudios enfocados en la edad objetivo está dada por la baja cantidad de investigaciones que no utilizan herramientas tecnológicas; porque hay un mayor enfoque en el uso de herramientas tecnológicas para ayudar en la disminución de síntomas de TEA o mejorar las habilidades socio-comunicativas de los niños. Se ha demostrado que el uso de herramientas tecnológicas en los tratamientos de TEA ayuda a mejorar los procesos de aprendizaje, siendo estas técnicas apoyadas tanto por padres como por profesionales especializados [42]. De igual forma, estas ayudas permiten trabajar habilidades metarrepresentativas como la imitación y el reconocimiento emocional, lo que favorece la solución de problemas ilustrados en situaciones sociales cercanas al contexto donde son desarrolladas [43]. En la subsección 3.2.5 se pueden observar ejemplos de este tipo de herramientas.

3.1.2 Medidas psicométricas para la evaluación emocional

Como se ha discutido anteriormente, es necesario tener en cuenta las pruebas que se utilizan en los protocolos de TEA para medir progresos en los sujetos. Teniendo en cuenta que la meta de este estudio es brindar ayuda en la estimulación de los procesos de imitación y reconocimiento de expresiones faciales, aspectos fundamentales de la comunicación no-verbal, es necesario analizar medidas psicométricas que se enfoquen en este aspecto. Una de estas es la batería de Reconocimiento de Expresiones Faciales y Corporales (REFyC), diseñada y validada en Argentina [44]. Esta tarea cuenta con cinco pruebas, las cuales se componen de estímulos en forma de videos, con una duración aproximada de cinco segundos. En estos videos se observan personas expresando emociones con el rostro o el cuerpo completo y personas con expresión emocional neutral y son las siguientes:

- Expresiones Corporales de Emociones Básicas

- Expresiones Faciales de Emociones Básicas
- Expresiones Corporales de Emociones Complejas
- Expresiones Faciales de Emociones Complejas
- Movimientos Corporales No-Emocionales

Para el diseño de los videos, se contó con la ayuda de 10 actores (5 hombres y 5 mujeres) con entrenamiento actoral en ámbitos formales de capacitación. La validación de las pruebas se realizó con la ayuda de 101 sujetos (35 hombres y 66 mujeres), a quienes se les presentó cada video de uno en uno y se les explicó que su tarea consistía en identificar qué emoción estaba sintiendo la persona del video a partir de una lista de seis opciones escritas, mostradas en la misma pantalla donde se les mostró los videos. Por cada emoción correctamente clasificada se le daba un punto al participante. Los investigadores argumentan que los resultados mostraron que cada prueba emocional tuvo adecuadas propiedades psicométricas, con adecuados índices de dificultad y de discriminación y buenos indicadores de validez y confiabilidad.

Otra batería desarrollada recientemente enfocada en expresiones faciales es la de Gelder y su equipo de trabajo en 2015, con el principal argumento de identificar con mayor facilidad la prosopagnosia, un trastorno neuropsicológico donde al individuo le cuesta reconocer caras familiares, incluyendo la propia. Esta batería se llamó Facial Expressive Action Stimulus Test (FEAST) y está desarrollada para probar el reconocimiento de la identidad y las expresiones humanas [45]. FEAST incluye las siguientes tareas:

- Tarea de emparejamiento de identidad de cara y zapato: En esta tarea, se le muestra al participante una foto de un zapato o una cara desde una vista frontal. Luego se le muestran dos imágenes: una en la que se observa el mismo objeto desde una vista de perfil y otra en la que se le muestra un objeto similar desde una vista de perfil. El participante debe escoger cuál de estas dos imágenes corresponde al mismo objeto de la primera.

- Tarea de emparejamiento “parte-a-completo” de cara y casa: En esta tarea, se le muestra al participante una parte de un rostro (ojos o boca) o una parte de una casa (ventanas o puertas). Más adelante se le muestra al participante dos imágenes: la casa completa de la imagen anterior o una casa similar. El participante debe escoger cuál de estas dos imágenes corresponde al mismo objeto de la primera.

- Tarea de emparejamiento de expresiones faciales: En esta tarea, se le muestra al participante la imagen de una persona realizando una expresión facial, a partir de las seis emociones básicas (enojo, miedo, alegría, tristeza, sorpresa y asco). Más adelante, se le muestran dos imágenes: Una en la que otra persona está realizando la misma expresión que la persona de la primera imagen y otra en la que se está realizando otra expresión. El participante debe escoger cuál de estas dos imágenes corresponde al mismo gesto facial.

- Tarea de memoria de rostros neutrales: En esta tarea, se le muestra al participante 50 imágenes de rostros neutrales durante tres segundos cada una, indicando que debe recordar esas caras para una etapa posterior. Después, se le muestra al participante dos imágenes, una que ya había visto anteriormente y otra nueva, de forma que el participante debe escoger aquella que ya ha visto anteriormente.

- Tarea de memoria de rostros emocionales: Esta tarea es similar a la anteriormente expuesta; sin embargo, en este caso las personas en las imágenes muestran rostros con miedo, tristeza o felicidad.

Para validar la batería, se reclutaron 58 adultos entre 18 y 62 años sin historial de problemas psiquiátricos o neurológicos, quienes realizaron las pruebas ya mencionadas. Se encontró que la edad fue un factor decisivo en los resultados de las tareas en las que se procesaron caras y objetos, dado que los adultos mayores tuvieron resultados significativamente menores. Los investigadores consideran que la razón para esto no es necesariamente el hecho de que exista una disminución cognitiva con el paso de los años, considerando la posibilidad de que exista una “parcialidad de edad propia”, en la que las personas tienen

una mayor facilidad para reconocer rostros que se encuentran en la misma edad que el participante. Dado que las bases de datos utilizadas contienen principalmente rostros de adultos jóvenes, puede que esto haya afectado los resultados. No se encontraron diferencias significativas en edad o género para otro tipo de tareas. Los autores consideran que FEAST provee a investigadores una batería extensa para habilidades de reconocimiento emocional y memoria de rostros emocionales, entre otros elementos.

Las baterías mencionadas en esta investigación y múltiples otras fueron validadas por medio de métodos rigurosos y con una cantidad significativa de participantes. De igual forma, son de gran utilidad para evaluar la habilidad de una persona de reconocer expresiones faciales. No obstante, varias de ellas no han sido validadas en ambientes transculturales, haciéndolas inapropiadas en la evaluación de individuos pertenecientes a otras culturas. Ardila argumenta que existen varios elementos que son propios de cada cultura en el uso de pruebas psicométricas [46]:

- Relaciones interpersonales: Cuando se realiza una prueba psicométrica, el examinador interactúa con el examinado, y esta relación puede variar entre culturas. Los cambios en este aspecto son particularmente relevantes cuando se comparan culturas “orientadas al individuo” con “culturas orientadas a grupos sociales”, dado que hay varias características en ambientes sociales y lingüísticos que cambian radicalmente entre comunidades.

- Autoridad de antecedentes: Este aspecto se refiere a que los antecedentes de un examinador le deberían dar autoridad para pedirle a un examinado que siga instrucciones; por lo cual, el examinado debe seguir órdenes. Sin embargo, la autoridad real que tenga el examinador puede variar entre culturas, dependiendo de edad, género o clase social, entre otros aspectos.

- Desempeño: El deseo de un individuo de obtener buenos resultados en una prueba psicométrica depende de la cultura de la que provenga, dado que existen sociedades que le dan bastante importancia a la competencia, mientras que hay otras que no le dan relevancia a este aspecto.

- Ambientes aislados: En gran parte de las pruebas psicométricas, se busca tener un ambiente aislado con el examinado. Sin embargo, en algunas culturas, se considera inapropiado tener citas con extraños en ambientes cerrados. Esto indica que el nivel de comodidad de un examinado al participar en una prueba psicométrica puede variar entre culturas.

- Tipo de comunicación especial: En una prueba psicométrica, el lenguaje utilizado debe ser formal para obtener resultados objetivos. Sin embargo, dependiendo de la cultura de un examinado, puede que esta sea o no una interacción natural.

- Velocidad: El concepto del tiempo es entendido de manera distinta en distintas culturas. Puede que una cultura valore realizar tareas en corto tiempo; sin embargo, para otra, puede que se considere que es necesario realizar un trabajo bien, así tome bastante tiempo. Por este motivo, la aplicación de pruebas psicométricas que piden realizar tareas en un tiempo determinado puede no ser apropiada en ciertas sociedades.

- Problemas internos o subjetivos: La privacidad se define de manera distinta entre culturas. Así, puede que ciertas preguntas dentro de una prueba psicométrica se consideran invasivas en culturas en las cuales no fueron diseñadas.

- Uso de elementos y estrategias específicas: En distintas pruebas psicométricas es común el uso de objetos e imágenes. Sin embargo, es posible que estos elementos no sean comunes en culturas distintas en las que la prueba psicométrica fue diseñada. Así, es posible que un examinado que proviene de la misma cultura del examinador obtenga mejores resultados que un examinado que proviene de otra cultura.

Dada la falta de universalidad de las pruebas psicométricas, es importante contar con estudios de validación que demuestren que una prueba psicométrica es apta para ser realizada en una cultura en particular. En el caso de esta investigación, se busca contar con pruebas psicométricas que permitan evaluar el reconocimiento de expresiones faciales en niños entre 6 y 8 años que hayan sido validadas en Colombia.

A conocimiento de nuestro equipo de trabajo, la única batería que cumple con estos criterios es la Evaluación Neuropsicológica Infantil (ENI) [47] en su segunda edición, la cual evalúa características neuropsicológicas de niños y jóvenes en edad escolar. La batería ENI evalúa 13 procesos neuropsicológicos: atención, habilidades construccionales, memoria, percepción, lenguaje oral, lectura, escritura, cálculo, habilidades visoespaciales, organización y conceptuación. Las normas de la batería se obtuvieron a partir de una muestra de 788 niños de 5 a 16 años provenientes de México y Colombia, sin antecedentes de problemas de desarrollo o enfermedades graves. Su validación se realizó a partir de pruebas de confiabilidad y validez [48]. La aplicación de esta batería tiene una duración aproximada de 3 horas; sin embargo, un aspecto de gran importancia para esta investigación es que cada subescala de la misma se puede aplicar por separado [47]. Gracias a esto, es posible aplicar el ítem de «Reconocimiento de expresiones (expresión emocional)» de manera individual, el cual hace parte de la subescala de percepción visual. En este ítem, el examinado debe identificar las expresiones faciales que se muestran en ocho fotografías, de forma que cada una da un punto si la expresión es identificada correctamente, para un total de ocho puntos.

Con el objetivo principal de obtener normas del ENI en la población colombiana entre 5 y 16 años, se realizó un estudio en el cual se contó con la ayuda de 252 niños de la ciudad de Manizales [49]. Los participantes (92 niños y 160 niñas) se separaron según su género, su nivel socioeconómico (medio-alto o medio-bajo) y su edad (cuatro grupos: 5-7 años, 8-10 años, 11-13 años y 14-16 años). Los criterios de exclusión son la presencia de retraso mental o antecedentes neurológicos o psiquiátricos, de acuerdo con las historias clínicas suministradas por padres de los participantes. Con los datos obtenidos a partir de la aplicación de ENI, se creó una base de datos sistematizada, donde se incluyeron los resultados de las subprueba de cada niño. En general, se encontraron diferencias significativas en los resultados de las subpruebas para cada uno de los grupos de edades, al igual que se encontraron diferencias significativas entre niños y niñas para las pruebas de habilidades visuoperceptuales, visuoconstructivas, especiales y numéricas. Finalmente, los investigadores concluyeron que la prueba ENI puede satisfacer la necesidad existente en el mundo hispanoparlante de herramientas neuropsicológicas para evaluar niños y adolescentes.

Dada la validación de la prueba ENI en Colombia, realizada de manera controlada y aleatorizada, es posible concluir que su uso es adecuado para la evaluación del nivel de reconocimiento de expresiones faciales por parte de los participantes en esta investigación.

3.2 Aspectos técnicos

En el desarrollo técnico del proyecto se tuvieron en cuenta dos aspectos principales: el desarrollo de una herramienta para el reconocimiento de expresiones faciales y la elaboración de un videojuego que sirvió como herramienta para la estimulación de procesos de imitación y reconocimiento. Por un lado, el reconocimiento de expresiones faciales utiliza el *pipeline* más común en la creación de herramientas de visión artificial, observado en la Figura 2. Así, es importante analizar las técnicas más utilizadas para cada una de estas etapas, haciendo énfasis en el objetivo específico de reconocer expresiones faciales. Por otro lado, la elaboración del videojuego debe tener en cuenta no solo el entretenimiento del participante; también la implementación de un adecuado protocolo de estimulación, de forma que se fortalezcan los procesos de aprendizaje deseados en este proyecto; por lo cual, se analizarán estudios previos que hayan desarrollado herramientas tecnológicas para apoyar distintos procesos de TEA.

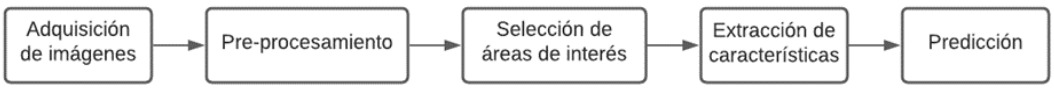

Figura 2. *Pipeline* común para aplicaciones de visión artificial.

3.2.1 Preprocesamiento de imágenes

El objetivo principal de la etapa de pre-procesamiento es facilitar la selección de áreas de interés. Teniendo en cuenta las necesidades de este proyecto, el pre-procesamiento se enfocó en evitar que la efectividad del reconocimiento dependiera del contraste de la imagen original, porque se espera que el algoritmo

funcione en ambientes no controlados. De esta forma, al lograr obtener una imagen con poca dependencia de la luz, es posible identificar rostros y obtener marcadores faciales de manera más sencilla y precisa. Los métodos más comúnmente utilizados son los siguientes:

- Normalización de intensidad:
- Uso de distintos espacios de color:
- Transformada Wavelet:
- Ecualización de histograma:

Además de los métodos tradicionales para mejorar la iluminación de una imagen, se encuentran técnicas basadas en estas y otras técnicas clásicas, buscando mejorar el contraste con algoritmos más complejos. Uno de estos es la Ecualización Adaptativa del Histograma (AHE, por sus siglas en inglés). Como se explicó anteriormente, la ecualización de histograma tradicional puede reducir la información de una imagen al mejorar el contraste de zonas irrelevantes. AHE busca eliminar esta desventaja al aplicar varias ecualizaciones de histograma de manera local en la imagen, de forma que la nueva intensidad de un píxel no dependerá de toda la imagen, sino únicamente de sus vecinos más cercanos.

No obstante, AHE tiene una tendencia a amplificar demasiado el contraste en áreas con intensidades relativamente homogéneas, dado que el histograma en estas zonas estará muy concentrado en un rango del espectro lumínico [50]. Una alternativa a AHE es la Ecualización Adaptativa del Histograma Limitada por Contraste (CLAHE, por sus siglas en inglés). Esta técnica reduce los problemas generados en AHE al limitar la variación en intensidad que puede tener un píxel, evitando cambios excesivos en el contraste [51].

De igual forma, algunos autores han buscado mejorar el contraste de una imagen por medio de análisis morfológico, tomando en consideración que varios sistemas buscan realizar este proceso para la posterior segmentación de objetos. Una de las técnicas creadas con este propósito toma como supuesto que un objeto se podrá segmentar mejor si su contraste es mejorado de manera local, de forma que la escala a la cual se realiza la mejora es proporcional al tamaño del objeto a segmentar [52]. Aunque los autores de este método encontraron una relación entre la escala de segmentación y su efectividad, consideran que su algoritmo es más efectivo cuando es utilizado como complemento de otro método de mejora de contraste.

Finalmente, en los últimos años se han desarrollado técnicas de pre-procesamiento basadas en el uso de aprendizaje profundo. Aunque varios algoritmos se han creado en este aspecto, los cambios entre ellos suelen enfocarse en la arquitectura utilizada para diseñar el modelo de aprendizaje profundo utilizado. Una de estas arquitecturas es la Red Neuronal Convolucional de Luz Baja (LLCNN, por sus siglas en inglés), diseñada por Tao en el 2017 [53], la cual se basa en una serie de cinco capas convolucionales y cinco unidades lineales rectificadas (ReLU) para mejorar el contraste de una imagen. Al igual que con todos los algoritmos de redes neuronales, las operaciones internas que modifican la imagen de entrada no son intuitivas, ya que son elegidas automáticamente por el modelo. Esto reduce la interpretabilidad de este, por lo cual, solo es posible juzgarlo a partir de su exactitud. En el caso de LLCNN, se evaluó la pérdida de información respecto a otros modelos similares, obteniendo mejoras estadísticamente significativas.

3.2.2 Detección de rostros

La detección automática de rostros es la habilidad de un computador de encontrar rostros humanos en una imagen. Esta habilidad es de importancia para esta investigación, ya que se trata del primer paso para reconocer expresiones faciales. Es importante reconocer la diferencia entre detección y reconocimiento faciales; la detección busca caras dentro de una imagen, mientras que el reconocimiento identifica a una persona a partir de su rostro y la vincula con una base de datos [54]. Dado que solo se requiere saber la expresión facial que está realizando el individuo, no es importante identificar quién la está haciendo, teniendo en cuenta que se espera que la cámara web solo encuentre un rostro.

Una de las primeras técnicas para detección de rostros creada fue la diseñada por Paul Viola y Michael Jones en 2001 [55], la cual se volvió popular por su robustez y velocidad. Esta técnica hace uso de una técnica de aprendizaje automático llamada *boosting*, en la cual se utilizan varios clasificadores con mala exactitud en cascada para crear un clasificador con buena exactitud. En este caso, los clasificadores de mala exactitud son características *Haar*, las cuales se basan en promediar la intensidad de distintas zonas en la imagen y obtener la diferencia entre ellas, como se observa en la Figura 3. La principal desventaja de este algoritmo es que no detecta rostros que no estén mirando directamente hacia la cámara [56].

Figura 3. Uso de características *Haar* para la detección de rostros [55].

Otro método popular y de uso libre es la detección de caras a partir de características obtenidas a partir del Histograma de Gradientes Orientados (HOG) y Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) [57]. Por su parte, HOG es un descriptor que se basa en obtener los gradientes orientados de una imagen a partir de diferencias de intensidad en la misma. Esto genera dos histogramas: uno que indica la distribución de los ángulos de los gradientes y otro que indica la distribución de su magnitud. Por otro lado, SVM es un algoritmo de aprendizaje automático que busca separar linealmente dos categorías al modificar la cantidad de dimensiones que se introducen. Para el modelo que se generó por los autores al momento de diseñar esta técnica, quienes utilizaron una base de datos en la cual los participantes miraban hacia el frente, la izquierda y la derecha. Por este motivo, una ventaja de este algoritmo es que los sujetos pueden mirar levemente hacia un lado y aun así ser detectados por el algoritmo. Sin embargo, el detector suele ignorar la frente y las mejillas de las personas detectadas [56].

Además de los métodos que utilizan técnicas de procesamiento de imágenes tradicional, algunos investigadores han diseñado modelos de aprendizaje profundo diseñados con el principal propósito de detectar rostros. En este documento se incluyen dos de ellos para su posterior prueba: Single Shot MultiBox Detector (SSD) [58] y Max-Margin Object Detection (MMOD) [59]. Por un lado, SSD utiliza como base la arquitectura Res-Net10, una red neuronal convolucional con 10 capas. No obstante, modifica las capas de entrada y salida para que las imágenes que se ingresen tengan una resolución de 300x300 pixeles, a diferencia de su valor base de 32x32 pixeles. Por otro lado, MMOD está diseñado para mejorar la efectividad de otros modelos, como el filtro HOG mencionado anteriormente, y busca maximizar el margen de los rostros detectados, permitiendo incluir partes del rostro que HOG no incluye normalmente, como frente y mejillas. La principal desventaja de estos algoritmos de aprendizaje profundo es que los modelos que necesitan para funcionar pueden llegar a ser archivos muy grandes, que pueden o no necesitar una tarjeta gráfica (GPU, por sus siglas en inglés) para funcionar correctamente.

3.2.3 Características faciales

La investigación en las características faciales que describen una expresión se ha estudiado desde hace más de un siglo. Sin embargo, el interés por este tema ha aumentado con la investigación que ha realizado Ekman a lo largo de su trayectoria, quien describió en 1971 las características que tienen las cejas, ojos y boca al expresar una de las seis emociones básicas, como se observa en la Tabla 1 [32]. A partir de este artículo y publicaciones posteriores, Ekman desarrolló en 1978 un sistema que llamó Sistema de Codificación de Acciones Faciales (FACS, por sus siglas en inglés), el cual, a partir de bases anatómicas,

describió las expresiones faciales en componentes individuales de movimiento muscular, los cuales llamó Unidades de Acción (AUs por sus siglas en inglés). El manual original de FACS no está disponible y las técnicas utilizadas para ubicar las AUs se han modificado desde entonces. No obstante, en 2002 se publicó la segunda edición de este manual, donde se indica a un investigador observar y codificar cada AU y describe las combinaciones existentes entre AUs. La relevancia de FACS es tal que se han convertido en una piedra angular para investigadores que buscan profundizar en el área de reconocimiento facial. De igual forma, ha sido utilizado por compañías como Disney, Pixar y EA Games para desarrollar animaciones donde los personajes creados por estas compañías expresan emociones humanas [60].

El principal aporte de FACS y la investigación de Ekman no es analizar la habilidad de las personas de reconocer expresiones faciales, sino demostrar que es posible medir cuantitativamente distintos descriptores del rostro, los cuales trabajan en conjunto para categorizar diferentes expresiones. Este estudio analiza principalmente el rostro para categorizar emociones, pero el grupo de trabajo de Ekman afirma que es posible describir la personalidad a partir de sus expresiones, o si hay indicios de psicopatología, por ejemplo [61].

Existe un total de 64 AUs, divididas en tres grandes grupos: AUs principales, AUs basadas en movimiento de cabeza y AUs basadas en movimientos de ojos. Cada una de estas AUs indica distintos movimientos faciales y describe los músculos involucrados; por ejemplo, AU1 indica la elevación de la parte interna de las cejas, involucrando los músculos *frontalis y pars medialis*, mientras que AU12 indica el estiramiento de las esquinas de los labios, involucrando el músculo *zygomatic major* [62].

Distintos investigadores han hecho uso de las AUs para el reconocimiento de expresiones faciales. Uno de estos casos es el de Affectiva, una compañía de desarrollo de software enfocada en la construcción de inteligencia artificial que identifica emociones humanas, estados cognitivos y actividades por medio del reconocimiento de expresiones faciales y voz [63]. La combinación de AUs que utiliza Affectiva para identificar las 6 emociones básicas se puede observar en la Tabla 2 [62].

Tabla 2. AUs utilizadas por Affectiva para identificar las seis emociones básicas [62].

Emoción	**Unidades de Acción**	**Descripción**
Alegría	AU6	Elevación de mejillas
	AU12	Estiramiento de las esquinas de los labios
Tristeza	AU1	Elevación de la parte interna de las cejas
	AU4	Depresión de las cejas
	AU15	Depresión de las esquinas de los labios
Sorpresa	AU1	Elevación de la parte interna de las cejas
	AU2	Elevación de la parte externa de las cejas
	AU5	Elevación de los párpados superiores
	AU26	Caída de la mandíbula
Miedo	AU1	Elevación de la parte interna de las cejas
	AU2	Elevación de la parte externa de las cejas
	AU4	Depresión de las cejas
	AU5	Elevación de los párpados superiores
	AU7	Apriete de los párpados
	AU20	Estiramiento de los labios
	AU26	Caída de la mandíbula
Enojo	AU4	Depresión de las cejas
	AU5	Elevación de los párpados superiores
	AU7	Apriete de los párpados
	AU23	Apriete de los labios
Asco	AU9	Arrugamiento de la nariz
	AU15	Depresión de las esquinas de los labios
	AU16	Depresión del labio inferior

En la Figura 4 se observa un ejemplo en el que una actriz imita las AUs de alegría y tristeza [62]. A partir de estos descriptores, es posible reconocer distintas expresiones faciales. Sin embargo, es necesario hacer uso de herramientas que permitan encontrar las AUs en la práctica por medio de técnicas de procesamiento de imágenes.

Figura 4. AUs correspondientes a la alegría y a la tristeza [62].

3.2.4 Reconocimiento de expresiones faciales

Al analizar las AUs más utilizadas en la identificación de las expresiones faciales de las emociones, se puede observar que se suele hacer un énfasis en la posición de los ojos, las cejas, la boca, la nariz y las mejillas. Una técnica que soluciona parte del problema de describir las AUs de manera automática en tiempo real es la identificación de marcadores faciales. Esta técnica busca localizar puntos clave en un rostro y ha sido utilizada en el pasado para aplicaciones en las que se reconoce el estado mental de una persona o para el reconocimiento biométrico. Este es un tema de investigación que sigue en construcción continua, ya que identificar marcadores faciales de manera automática puede ser un reto computacional grande, por la cantidad de variables involucradas, como la pose de la persona, la rotación del rostro, la oclusión y la iluminación. Una revisión de literatura describió los algoritmos de ubicación de marcadores faciales más populares desarrollados en los últimos años [64]. En este documento se describen algunos de ellos.

Kazemi y Sullivan desarrollaron un algoritmo de ubicación de marcadores faciales basado en un conjunto de árboles de regresión, cuya principal ventaja es la selección de características invariantes a la forma del rostro [65]. Además de esto, el método implementado permite ubicar marcadores faltantes a partir de extrapolar la ubicación de los marcadores que si fueron encontrados y logra ubicar los marcadores con una velocidad de 1000 recuadros por segundo (FPS, por sus siglas en inglés). El modelo diseñado fue entrenado a partir de la base de datos HELEN, la cual cuenta con 2330 fotos encontradas en el servició web flickr.com, las cuales fueron tomadas en ambientes no controlados y cuentan con una amplia variedad de poses, iluminación y oclusión. Cada una de las fotos se etiquetó con 194 marcadores faciales, de forma que esta información se utilizó como verdad absoluta para el entrenamiento del algoritmo. Para probar la efectividad del algoritmo desarrollado, se comparó la efectividad de este con otros ya desarrollados, los cuales también se entrenaron a partir de la base de datos HELEN. Para medir el error, se halló el promedio de la distancia normalizada de cada marcador respecto a su ubicación real según la verdad absoluta. Se encontró que el error del algoritmo desarrollado fue menor que el de todos los otros algoritmos observados, con un valor de 0.049.

Por su parte, Tzimiropoulos y Pantic desarrollaron un marco de referencia que se basa en dos problemas de optimización para mejorar la velocidad y la exactitud de algoritmos de ubicación de marcadores faciales [66]. A partir de esto, desarrollaron algoritmos cuyo principal beneficio es su aplicabilidad en 3D. Los algoritmos desarrollados fueron llamados Fast-SIC y Fast-forward. Los algoritmos desarrollados fueron

evaluados en dos bases de datos: La ya mencionada anteriormente HELEN, y LFPW, una base de datos que contiene 1287 imágenes descargadas de varios sitios de internet, las cuales contienen una alta variedad en pose, expresiones, iluminación y oclusiones. Las anotaciones de verdad absoluta consisten en la ubicación de 35 marcadores faciales. Al comparar la efectividad de estos algoritmos con otro algoritmo desarrollado anteriormente llamado POIC, los autores encontraron que en la base de datos HELEN ambos algoritmos desarrollados tienen un error significativamente menor que POIC y comparable entre ellos. Al evaluar este mismo error en la base de datos LFPW, los autores encontraron que Fast-SIC tiene un error significativamente menor que FAST-Forward, el cual tiene un error significativamente menor a POIC.

En la Tabla 3 se puede observar un resumen de las bases de datos más utilizadas para el entrenamiento de modelos que buscan ubicar marcadores faciales para cada una de las técnicas descritas y en la Figura 5 se observan ejemplos de cada una de ellas [64]. Así, es posible en un futuro desarrollar un modelo propio que haga énfasis en los AUs que se buscan, aquellos que están relacionados con las expresiones faciales de las emociones.

Tabla 3. Resumen de los hallazgos de la revisión de literatura de Ouanan [64].

Base de datos	**Ambiente**	**Número de marcadores**	**Pose**
Multi-PIE	Controlado	68	[-45°, 45°]
XM2VTS	Controlado	68	0°
FRGC-V2	Controlado	5	0°
AR	Controlado	22	0°
KFOW	No-controlado	35	[-45°, 45°]
HELEN	No-controlado	194	[-45°, 45°]
AFW	No-controlado	6	[-45°, 45°]
AFLW	No-controlado	21	[-45°, 45°]

Aunque los marcadores faciales brindan una gran cantidad de información para la descripción de las AUs, existen ciertos detalles que no contemplan, como el ceño fruncido o arrugas que pueda haber en las mejillas. Para combinar estos aspectos de las AUs, es necesario hacer uso de otras técnicas.

Una forma común de hallar aquellos cambios no visibles a partir de marcadores faciales es el análisis de texturas dentro de una imagen. Una forma de analizar la textura en una imagen utiliza un método ya mencionado anteriormente: los patrones binarios locales (LBP, por sus siglas en inglés). En los LBP, se analizan los vecinos más cercanos a un píxel y se da un valor binario, dependiendo de si su valor es menor o mayor que aquel del píxel que se está evaluando. Si se analizan ocho vecinos, se obtienen ocho números binarios, los cuales se organizan secuencialmente para obtener un número patrón. Posteriormente, se acumulan estos patrones y se observan en un histograma. Esta metodología permite encontrar formas específicas como puntos, esquinas y bordes, dando una aproximación de distintas texturas encontradas en la imagen. Una investigación analizó el uso de LBP para el reconocimiento de expresiones faciales, haciendo uso de una serie de SVMs para clasificar rostros [67]. Para analizar la efectividad de este método, se utilizó la base de datos JAFFE, la cual cuenta con 213 imágenes de 10 personas japonesas que expresan seis emociones básicas y un rostro neutro, obteniendo una exactitud de clasificación del 93.8%.

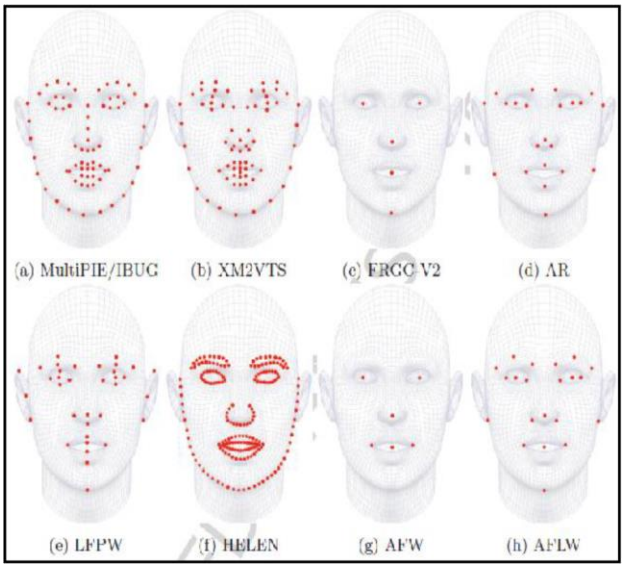

Figura 5. Ubicación de los marcadores faciales por cada algoritmo encontrado en la revisión de literatura [64].

Otra técnica comúnmente utilizada para reconocer rasgos que no se pueden hallar por medio de marcadores faciales es HOG, una técnica mencionada anteriormente. Una de las investigaciones que ha analizado el uso de HOG en el reconocimiento de expresiones faciales es la de Carcagnì [68]. En esta investigación, los autores dividieron los rostros de imágenes de entrada en varias secciones y obtuvieron los histogramas de orientación de los gradientes orientados, los cuales utilizaron luego para entrenar una SVM de clasificación. La evaluación de la efectividad de este algoritmo se realizó con la ayuda de la base de datos Cohn-Kanade (CK+), la cual cuenta con 593 videos de 123 personas realizando cada una de las seis expresiones faciales de las emociones, más una expresión neutra. Por un lado, los autores analizaron la efectividad del método propuesto con aquella de otros tres métodos, entre los que se encontraba LBP. Se encontró que el método con mayor exactitud para reconocer expresiones faciales fue HOG, con una exactitud del 95.8%. Por otro lado, se analizó la configuración adecuada para HOG, teniendo como variables de entrada el número de secciones en las que se dividieron las imágenes y el número de compartimientos a usar en los histogramas generados. Se encontró que la mayor efectividad se consigue cuando las secciones de la imagen tienen un tamaño de 7x7 píxeles y cuando se utilizan 7 compartimientos en el histograma.

Finalmente, también se ha hecho uso de técnicas de aprendizaje profundo para el reconocimiento de expresiones faciales. Al igual que con otros modelos de aprendizaje profundo ya discutidos, la principal diferencia entre investigaciones que utilizan aprendizaje profundo es la arquitectura utilizada para entrenar el modelo deseado. Una investigación desarrolló un clasificador basado en redes neuronales convolucionales para categorizar las emociones de alegría, tristeza, sorpresa, enojo y miedo [69]. Sin embargo, para hacer esto, primero se pasaron las imágenes de entrada por una red neuronal que clasificaba cada una en dos categorías: alegría o tristeza; a esto, los autores lo llamaron *la emoción primaria*. Dependiendo de la emoción escogida, luego se hacía una segunda clasificación entre sorpresa y estado neutro o enojo y asco; a esto, los autores lo llamaron *la emoción secundaria*. Al evaluar los resultados con la base de datos JAFFE, se obtuvo una exactitud del 93.45%. No obstante, los autores no indican cómo se escoge entre las *emociones primarias* y las *emociones secundarias*.

3.2.5 Desarrollo herramientas tecnológicas para TEA

En los últimos años, el número de investigaciones enfocadas en la ayuda a personas con TEA que están apoyadas por herramientas tecnológicas ha aumentado. A continuación, se incluirán algunos artículos en los cuales se haya estudiado la imitación o el reconocimiento de expresiones faciales en niños con TEA cuyas edades abarcan las de interés para este estudio (6 a 8 años). Algunos de estos estudios analizan las diferencias entre niños neurotípicos y niños con TEA, mientras que otros se enfocan en la mejora de habilidades emocionales de niños con TEA.

En el primero de estos estudios, se exploraron las diferencias en la imitación de expresiones faciales entre niños neurotípicos y niños con TEA, a partir de calificaciones dadas por jurados y por clasificadores de *Random Forest* (RF) [70]. Para lograr esto, se pidió a 157 niños neurotípicos y a 36 niños con TEA entre 6 a 12 años que imitaran distintas expresiones faciales, entre las que se encontraban alegría, enojo, tristeza y expresión neutra. Para lograr esto, los niños realizaron dos tareas: una donde imitaban expresiones faciales que les pedía imitar de manera verbal y otra en que imitaban las expresiones faciales de un avatar. Esta imitación se hacía de dos maneras distintas: visual o audiovisual. Se grabaron videos de todas las sesiones para su posterior análisis por parte de jurados, quienes utilizaron una escala de Likert para calificar la expresión de emociones; en esta escala, se evalúan dos dimensiones: reconocimiento y credibilidad. Sin embargo, dado que no es posible tener credibilidad sin reconocimiento, los jurados evaluaron cada expresión en una escala de 0 a 10: 0 indica que no hay reconocimiento, 5 indica que el reconocimiento es máximo, pero no hay credibilidad y 10 indica que tanto el reconocimiento como la credibilidad son máximas. Los resultados de esta prueba indicaron que los niños con TEA tienen más dificultad produciendo expresiones faciales que los niños neurotípicos.

En cuanto al entrenamiento del modelo RF, se utilizó el algoritmo creado por Viola y Jones mencionado anteriormente (cascadas Haar) y se utilizó un modelo que ubica 49 marcadores faciales. A partir de los marcadores faciales, los autores decidieron calcular la distancia euclidiana entre cada par de marcadores (1176 características geométricas) y generar histogramas de gradientes orientados en distintas regiones de la cara, utilizando cada compartimiento como una característica (441 características de apariencia) para un total de 1617 características. Para la clasificación, los autores decidieron usar RF dado que en el pasado ha brindado buenos resultados en problemas con alta dimensionalidad. Los modelos generados utilizaron 500 árboles con una profundidad máxima de 16 capas. Los autores crearon tres modelos para observar diferencias entre niños neurotípicos y niños con ASD. El primer modelo fue entrenado con datos de niños neurotípicos y probado con esta misma población. El segundo modelo fue entrenado con datos de niños con TEA y probado con esta misma población. El tercer modelo fue entrenado con niños neurotípicos y probado con niños con TEA. Cada modelo fue evaluado por medio de una validación cruzada. La exactitud global del primer modelo fue de 82.05%, la del segundo modelo fue de 66.53% y la del tercer modelo fue de 69.3%. En general, los autores encontraron que la emoción más difícil de imitar para ambos grupos fue la tristeza y el enojo se suele confundir con la alegría. Los autores argumentan que, para clasificar mejor las expresiones faciales de niños con TEA, se necesita una mayor cantidad de marcadores faciales. Una preocupación que surge respecto a esta investigación es la cantidad de características extraídas. Lamentablemente, los autores de esta investigación no indican el tiempo tardado en clasificar cada recuadro. Teniendo en cuenta que el proyecto a realizar debe funcionar en tiempo real, probablemente no es viable extraer más de mil características, ya que esto reduciría considerablemente la velocidad de clasificación.

Otro estudio interesante indicó la dificultad de sujetos neurotípicos para interpretar las expresiones faciales emocionales de personas con TEA, por lo cual investigó las diferencias entre los patrones de movimiento de expresiones faciales de personas neurotípicas y personas con TEA, haciendo uso del análisis de videos para detectar estas diferencias [71]. Para lograr esto se contó con la ayuda de 19 participantes con TEA y 18 participantes neurotípicos, todos niños o adolescentes, quienes observaron 36 videos cortos cada uno donde se observan actores representando expresiones y posteriormente se pidió imitar las expresiones realizadas en ellos. Para cuantificar las expresiones faciales realizadas por los participantes, se les puso 32 marcadores físicos en el rostro, los cuales cubrieron gran parte del rostro y se registraron sus movimientos con una cámara Vicon. Para caracterizar los movimientos, se calcularon distancias entre marcadores, para identificar acciones como la apertura de la boca o la distancia entre las cejas. El análisis principal se realizó con el promedio de las distancias escogidas entre los marcadores. En general, los investigadores encontraron que este promedio fue significativamente mayor en participantes con TEA que aquel en participantes neurotípicos. De igual forma, la variabilidad en las distancias entre marcadores fue significativamente menor en el grupo participantes neurotípicos, indicando una mayor inconsistencia en la imitación de expresiones faciales en participantes con TEA. Un resultado interesante encontrado en el estudio fue que las expresiones faciales con mayor intensidad mostraban una mayor variabilidad en la distancia promedio de marcadores faciales. De igual forma, se observó que, para participantes neurotípicos, las emociones con valencia positiva generan una mayor variabilidad en la distancia de los marcadores. Dada la baja variabilidad en las distancias entre marcadores para el grupo de participantes neurotípicos,

los autores argumentan que es posible predecir cuánto movimiento realizará un niño neurotípico para una expresión deseada, dado tanto por su intensidad como por su valencia. Finalmente, concluyen que esta predicción también se puede realizar para niños con TEA cuando se tiene en cuenta la intensidad de la expresión facial, más no a partir de su valencia.

Por último, se buscaron investigaciones con propósitos similares al propuesto en este proyecto. A partir de esta búsqueda, se encontró que existen muy pocas herramientas interactivas para la estimulación de los procesos de imitación y reconocimiento en niños con TEA. Se han creado varias herramientas con el fin de identificar diferencias entre niños neurotípicos y niños con TEA, al igual que se han creado herramientas que brindan instrucciones sobre la adecuada imitación y reconocimiento de expresiones faciales. Sin embargo, existen pocas investigaciones que brinden un proceso de aprendizaje interactivo con realimentación. Una de estas investigaciones se basó en el uso de robótica social para enseñar a los niños reconocimiento emocional y comprensión [72]. En esta investigación se hace uso de robótica social dado que anteriores investigaciones han demostrado la efectividad de que personajes virtuales y robots acompañen los procesos de aprendizaje de niños con TEA. Para validar esto, los investigadores contaron con la ayuda de 14 niños para un protocolo experimental. 7 de ellos asistieron a una intervención tradicional de TEA para mejorar habilidades sociales (grupo control), mientras que los otros 7 asistieron a la misma intervención, pero con asistencia robótica. A partir de pruebas psicométricas, como la Prueba de Comprensión Emocional (TEC, por sus siglas en inglés) y la Prueba de Léxico Emocional (ELT, por sus siglas en inglés), se encontró una mejora substancial en las habilidades de reconocimiento emocional y comprensión del grupo experimental. Aunque se encontró que las mejoras también fueron importantes en el grupo control, los cambios fueron estadísticamente mejores en el grupo experimental.

Otra investigación que cumple estos criterios buscó validar los argumentos presentados por la herramienta *FaceSay* [73]. *FaceSay* brinda varios juegos diseñados para la enseñanza de habilidades de procesamiento de rostros para la cognición social. Dentro de los juegos desarrollados en esta herramienta se encuentran *Amazing Gazing*, el cual enseña a niños a fijar la mirada en situaciones sociales, *Band Aid Clinic*, el cual les pide a los niños completar una imagen de un rostro a partir de una plantilla creada y *Follow the Leader*, el cual les pide a los niños indicar si dos rostros están realizando la misma expresión facial [74]. La realimentación que brinda este juego se enfoca en indicarle visual y verbalmente al participante si realizó una acción correctamente, dependiendo del juego que se esté jugando.

La investigación que validó esta herramienta contó con la ayuda de 31 niños entre 5 y 11 años, elegibles para servicios de educación especial según las normas de California, por lo cual no estaban necesariamente diagnosticados con TEA. Por un lado, 16 de los niños utilizaban *FaceSay* durante 10 semanas, 25 minutos semanalmente. Por otro lado, los 15 niños restantes utilizaron la herramienta *SuccessMaker*, diseñada como una serie de juegos que suplementan clases regulares, sin ningún tipo de estímulo social. Para medir la efectividad de las herramientas, los participantes se sometieron a la subprueba *Emotion/Affect Recognition* de NEPSY-II, la cual les pide a los niños identificar y reconocer rostros de niños en imágenes. El estudio encontró que los niños sometidos a *FaceSay* tuvieron una mejora significativa en sus habilidades sociales respecto a su línea base, mientras que los niños en el grupo control no obtuvieron mejoras significativas. El programa *FaceSay* es una herramienta interesante y útil para mejorar las habilidades sociales de niños con TEA, por lo cual es posible aprovechar las ideas brindadas por los autores y mejorarlas, de forma que sea posible darles a los niños una realimentación más adecuada, teniendo en cuenta su habilidad para imitar y reconocer expresiones faciales por medio de una cámara web. De igual forma, si se amplía la cantidad de juegos que se brindan, abarcando una mayor variedad de habilidades, es posible brindar una enseñanza progresiva que les facilite desenvolverse adecuadamente en situaciones sociales más complejas.

IV. METODOLOGÍA

Las tareas realizadas en este proyecto se basaron en cada uno de los objetivos específicos establecidos en la subsección 2.2. La subsección 4.1 describe la implementación de un protocolo experimental basado en los aspectos psicológicos y del desarrollo neuronal del síndrome del espectro autista. El diseño y posterior desarrollo de un algoritmo de reconocimiento de gestos faciales en tiempo real se describe en la subsección 4.2. El desarrollo de la herramienta interactiva diseñada en la cual se implementa el algoritmo de reconocimiento de gestos faciales se describe en la subsección 4.3. Los métodos con los cuales se evalúa la efectividad de la herramienta diseñada se detallan en la subsección 4.4. Finalmente, se describen los aspectos muestrales del experimento en la subsección 4.5.

4.1 Implementación del protocolo experimental

4.1.1 Introducción a la herramienta de estimulación

La herramienta de estimulación está diseñada para que la imitación y el reconocimiento de expresiones faciales se dé de manera progresiva. Por este motivo, se crearon dos etapas para la estimulación: imitación y reconocimiento, con una transición clara entre ellas. En esta subsección se describen las características que tuvo la aplicación creada para potenciar su utilidad como una herramienta para estimular la imitación y el reconocimiento de expresiones faciales. Es importante notar que el diseño del protocolo experimental se desarrolló en conjunto con psicólogas y estudiantes de psicología de la Corporación Universitaria Minuto de Dios UNIMINUTO.

Uno de los primeros aspectos que se tuvieron en cuenta para el desarrollo del protocolo experimental fue la creación de un avatar virtual que acompaña los procesos de estimulación durante la totalidad del experimento, dado que se ha encontrado que el uso de personajes virtuales en programas computacionales y robots potencia el aprendizaje en niños con TEA [72]. Se buscó que los participantes del experimento se identificaran con el avatar, por lo cual, se optó por un aspecto humano realista y una edad similar a la de los participantes, con un nombre fácil de pronunciar y con pocas sílabas, Emma. En la Figura 6 se puede observar a Emma, quien está presente en todos los juegos desarrollados en la aplicación.

Figura 6. Emma, el avatar que acompaña al participante a lo largo de la aplicación.

Un aspecto importante de Emma es que cuenta con una gran cantidad de acciones faciales, las cuales se pueden usar para la creación de las expresiones faciales de las emociones y para mover sus labios de distintas formas. Mediante un guion, se les indica a los participantes las instrucciones de cada juego y les

explica las partes del rostro y las emociones. La voz de Emma es realizada por Michelle Ballén, estudiante de la Corporación Universitaria Minuto de Dios UNIMINUTO, miembro del Semillero de Neurociencias Básica y Clínica e investigadora del proyecto. Para aportar al realismo y lograr que los participantes se identifiquen con Emma, a cada audio donde Emma está presente durante la aplicación se le añadieron animaciones únicas para cada fonema expresado. A la herramienta de estimulación se le dio el nombre «Emmaciones» y este será el nombre con el que se referirá en el documento.

Otro aspecto que se consideró en el diseño de la interfaz, siguiendo el ejemplo de otras herramientas similares como *FaceSay* [74], fue la presencia de colores brillantes, tipografías «alegres» y tamaños grandes de letra, para capturar la atención de los niños con mayor facilidad. Todos los elementos utilizados en la interfaz gráfica que no fueron creación propia son de dominio público, lo que permite su uso en producciones comerciales y no-comerciales. Para ejemplificar el diseño de la interfaz gráfica, en la Figura 7 se observa la escena principal que observan los participantes al iniciar Emmaciones.

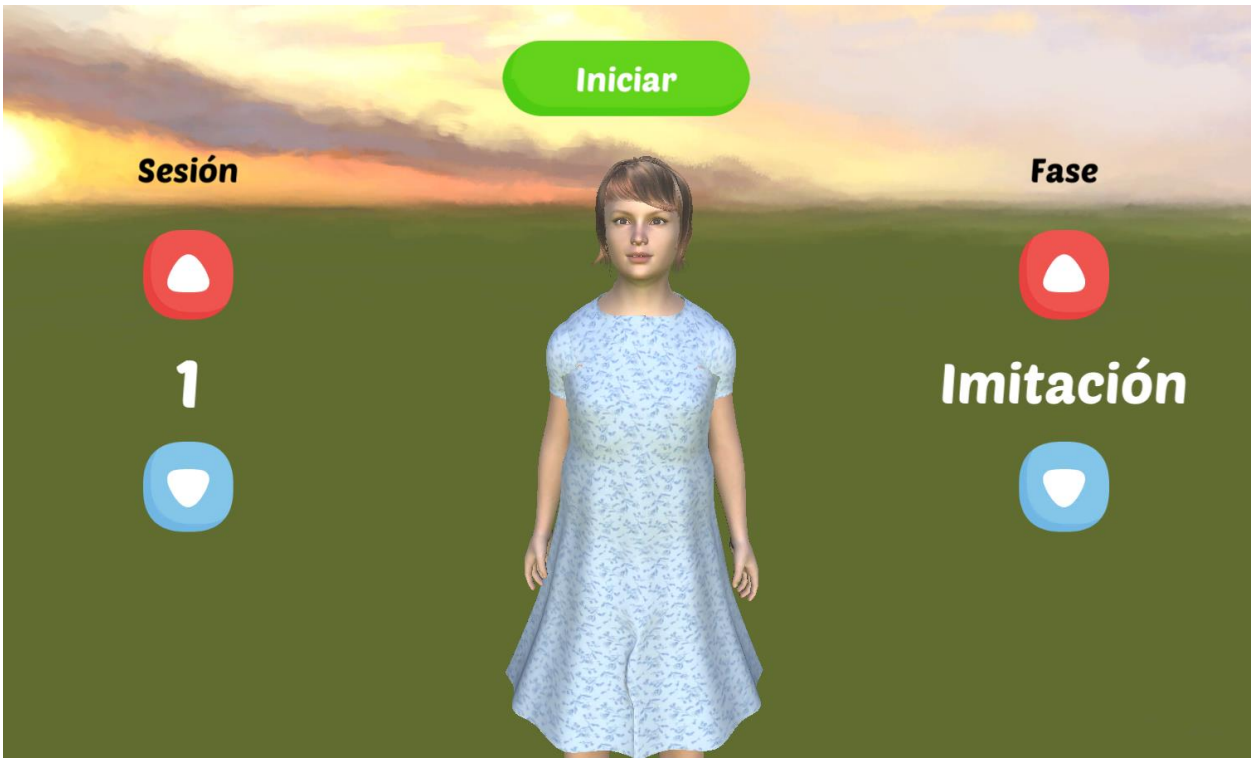

Figura 7. Escena principal de Emmaciones, en la que se observa el tipo de fuente, la paleta de colores y algunos elementos gráficos del juego.

En cuanto a la disposición del espacio requerido para llevar a cabo las pruebas, lo único que necesita el participante es un computador sin muchos requisitos técnicos y una cámara web. Se requiere del participante ubicarse en un lugar en el que tenga suficiente espacio para mover su cuerpo, aproximadamente un área de 1.5m^2. De igual forma se le pide que su rostro sea el único que se vea en la cámara durante la aplicación, para que el sistema no detecte a otras personas. Finalmente, se le solicita que se ubique en un lugar con iluminación adecuada y estable, haciendo prioridad en no encontrarse a contraluz.

A lo largo de la subsección 4.1 se describirán más detalles relacionados con el diseño de Emmaciones, los cuales son específicos para distintos momentos en el proceso de estimulación. En la Figura 8 se observa un diagrama con las actividades que se realizan en cada sesión de la etapa de imitación, para un total de seis sesiones. La sesión 1 se enfoca en introducir las actividades y en el reconocimiento de las partes del rostro, mientras que las sesiones 2-6 hacen énfasis en los procesos de imitación de expresiones faciales, con una introducción al reconocimiento de estas. En el diagrama se puede observar que los bloques indican cuántas emociones se reforzarán en cada sesión, dado que ciertas sesiones solo evalúan

tres emociones. El protocolo se elaboró de esta forma para que los participantes tengan una mayor facilidad reforzando conceptos; así, no deben recordar las características de todas las emociones al tiempo durante las primeras sesiones del proceso. A partir de pruebas piloto, las cuales se detallan en la subsección 4.5.3, se encontró que cada sesión dura aproximadamente 20 minutos. Teniendo en cuenta que el consentimiento informado que se le brindó a los representantes legales de los participantes (ver en el Anexo 1), indica que cada sesión experimental dura aproximadamente 60 minutos, se decidió que, si el tiempo lo permitía, se realicen múltiples sesiones en un día. Así, en la Figura 8 se observa que existen pausas activas entre las sesiones 1 y 2 y las sesiones 4 y 5. Estas pausas activas, las cuales duran aproximadamente 15 minutos cada una, están diseñadas para que los participantes descansen de las actividades propuestas y realicen actividades físicas. Teniendo en cuenta que varias de las actividades propuestas exigen que el participante active varios músculos faciales, las pausas activas sirven para relajarlos y evitar la fatiga. De igual manera, antes de iniciar actividades que requieran imitación constante de expresiones faciales, se incluyen secciones de calentamiento. Los detalles de las pausas activas y los calentamientos se detallan en la subsección 4.1.5.

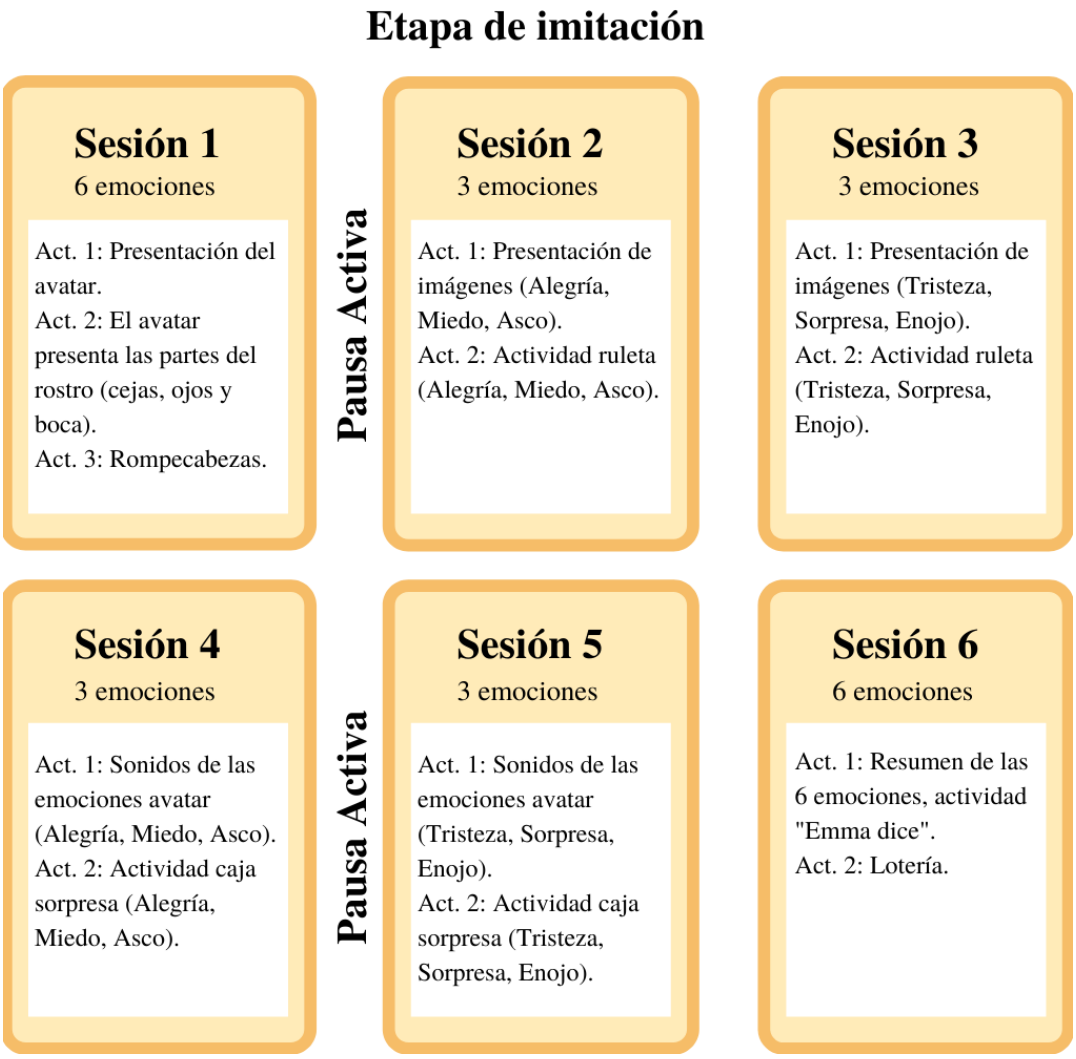

Figura 8. Resumen de las actividades realizadas durante la etapa de imitación.

De igual forma, en la Figura 9 se observa el diagrama de actividades realizadas durante la etapa de reconocimiento. En esta etapa, existe un énfasis en el reconocimiento de expresiones faciales, de forma que la mayoría de los juegos le piden al niño identificar una emoción a partir de imágenes, videos o relatos. Sin embargo, para reforzar el proceso de imitación, varias de estas actividades también le solicitan al niño imitar expresiones faciales una vez se hayan reconocido. De manera similar que la etapa de imitación, en

algunas sesiones de la etapa de reconocimiento se refuerza el aprendizaje sobre tres de las emociones básicas y en otras, tres emociones restantes. Al igual que con la etapa de imitación, existen pausas activas como método de transición entre las sesiones 1 y 2 y las sesiones 4 y 5. En la sesión 3 se incluyó la actividad adicional *Foto en Vivo* para que el participante imite las emociones en el orden que él desee, reforzando el proceso de imitación. En las sesiones 4 y 5 se introduce el concepto del reconocimiento de emociones a partir de videos sin contexto y, finalmente, en la sesión 6 se hace reconocimiento de emociones con contexto.

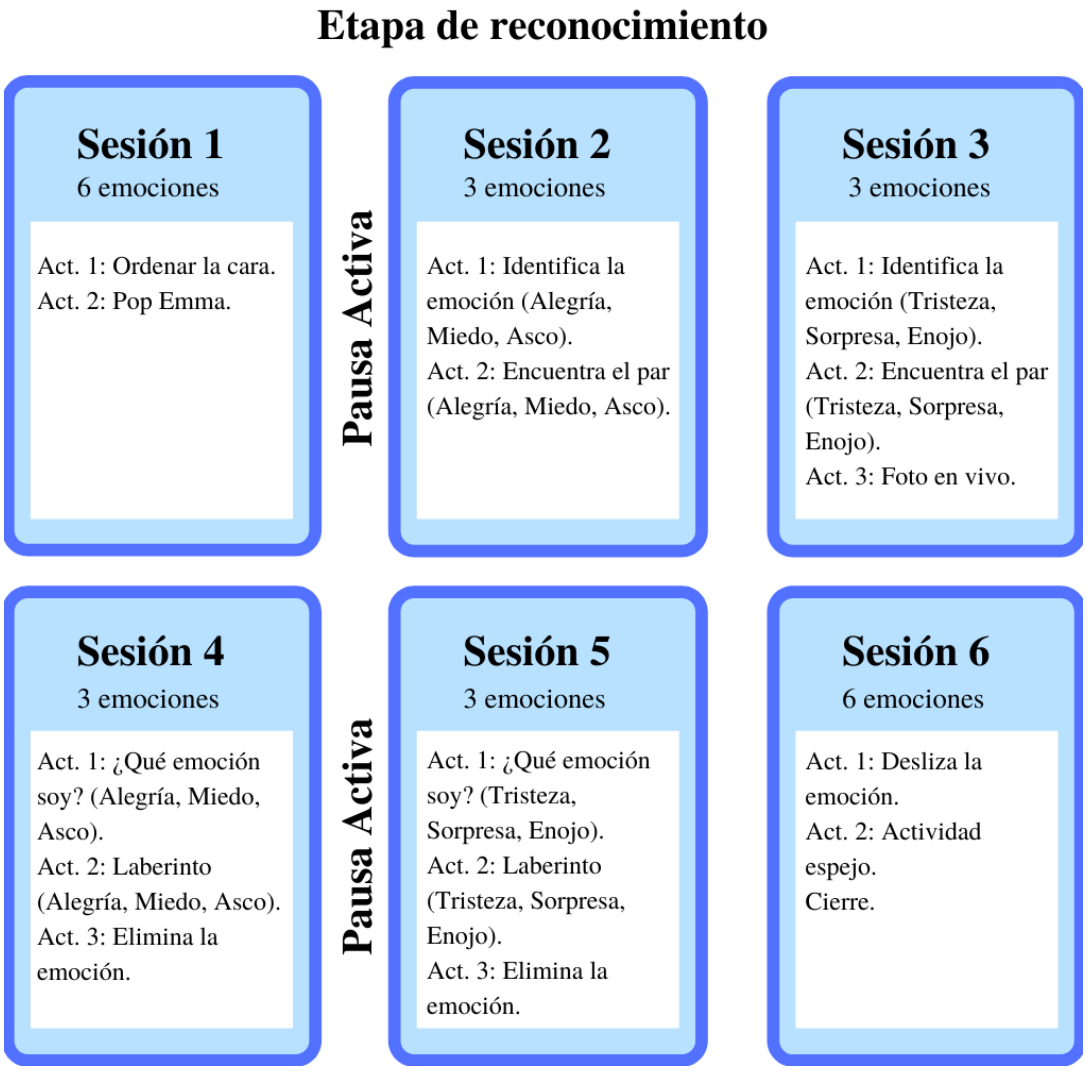

Figura 9. Resumen de las actividades realizadas durante la etapa de reconocimiento.

4.1.2 Aprendizaje de las partes del rostro

El aprendizaje de las partes del rostro se basó en los estudios de Ekman [21], [32] e, indirectamente, en el sistema FACS [62]. Como se indicó antes, estas investigaciones indican que las expresiones faciales de las emociones se pueden identificar a partir de tres partes principales en el rostro: cejas/frente, ojos/párpados y sección inferior. Algunos conceptos en esta teoría pueden ser difíciles de entender para niños de 6 a 8 años, particularmente si cuentan con déficits en comunicación no verbal. Por este motivo, para la elaboración de las instrucciones que brinda Emma y el posterior desarrollo de Emmaciones se simplificaron estas tres secciones en cejas, ojos y boca, respectivamente. Adicionalmente, para incluir descripciones que no encajan en estas tres partes del rostro como «ceño fruncido» o «pronunciación de

líneas nasolabiales» se le pide al niño imitar las expresiones que hacen Emma y las personas que se muestran en imágenes y videos a lo largo del proceso.

La enseñanza de las partes del rostro se hace en la sesión 1 de la etapa de imitación, por medio de los juegos *Partes del Rostro* y *Rompecabezas*. Adicionalmente, en *Partes del Rostro* se hace una introducción a las emociones, que incluye sinónimos, o palabras con significados similares, de cada una, dado el caso que los participantes las conozcan con otro nombre. Posteriormente, se describen las características de cada expresión facial, haciendo énfasis en las partes del rostro ya mencionadas y teniendo en cuenta los estándares establecidos por el grupo de investigación de Ekman. El guion de Emma para este primer juego se observa en el Anexo 2:

En este guion, dicho por Emma, se hace énfasis en las palabras en negrilla para que los participantes comprendan con mayor facilidad la descripción de las expresiones faciales. Existen guiones similares para cada una de las actividades desarrolladas en Emmaciones; sin embargo, en aras de la brevedad, no se incluirán en este documento.

Después que el participante escucha estas instrucciones, se le presenta un rostro de Emma. Al pasar el cursor por cada una de las partes del rostro descritas anteriormente, esta parte se iluminará y Emma indicará de cual parte se trata.

En el *Rompecabezas*, los participantes pueden ver el mismo rostro de la actividad anterior; sin embargo, las partes del rostro están mezcladas. El objetivo es reordenar la cara hasta que todas las partes del rostro estén en el lugar indicado. Esta actividad es particularmente importante para reforzar el aprendizaje sobre las partes del rostro. Los detalles de estas actividades y figuras explicándolas pueden ver en la subsección 4.3.2.

4.1.3 Actividades para la imitación de expresiones faciales

La imitación de expresiones faciales se refiere a la capacidad de los participantes de reproducir expresiones faciales presentes en Emmaciones. Esto se hace a través de texto, imágenes o videos. Este proceso inicia en la segunda sesión de la etapa de imitación. La primera actividad de esta sesión, llamada *Presentación de Imágenes*, busca que los participantes comprendan cómo se ven las expresiones faciales de las emociones en ambientes reales o simulados. Así, se le presentan tres botones al participante, uno para la alegría, uno para el miedo y uno para el asco. Al oprimir cada uno, se observará una imagen aleatoria de la emoción escogida, seleccionada entre un grupo de imágenes de uso público obtenidas a partir de herramientas como *Pixabay* o *Shutterstock*. En la primera actividad de la sesión 3 de imitación se repite esta actividad, pero con las expresiones faciales de la tristeza, la sorpresa y el enojo. En la Figura 10 se observan ejemplos de las imágenes utilizadas para cada una de las emociones. Este grupo de imágenes también se utiliza en actividades posteriores y ayuda a los participantes a entender cómo se ven estas emociones en personas reales.

Figura 10. Imágenes utilizadas en Emmaciones para ejemplificar cada emoción. Fila superior: alegría, miedo, asco. Fila inferior: tristeza, sorpresa, enojo.

En la segunda actividad de las sesiones 2 y 3, llamada la *Ruleta*, se inicia la primera tarea real de imitación, donde los participantes giran una ruleta en la cual se encuentran tres emociones, dependiendo de la sesión

realizada. Cuando la ruleta se detiene, un rostro de Emma con la emoción que se está indicando aparece. En este momento, se enciende la cámara del participante y éste debe imitar la emoción que indique Emma durante un tiempo determinado. Esta actividad continúa hasta que todas las emociones de la sesión se hayan imitado. En esta y en todas las siguientes actividades que requieran imitación, se les presenta a los participantes un botón, con el cual pueden cambiar la emoción a imitar si llevan suficiente tiempo tratando de imitar una. Posteriormente, el juego vuelve a pedirle al participante que imite la emoción que no logró imitar anteriormente. Los rostros de Emma que salen en esta actividad y en múltiples actividades posteriores se observan en la Figura 11.

Figura 11. Rostros de Emma con cada expresión facial de las emociones básicas. Fila superior: alegría, miedo, asco. Fila inferior: tristeza, sorpresa, enojo.

En la actividad 1 de las sesiones 4 y 5 de imitación, llamada *Sonidos de las Emociones*, se explica al participante que las expresiones faciales de las emociones pueden estar acompañadas por sonidos. De manera similar a *Presentación de Imágenes*, se presenta al participante tres botones, uno para cada emoción, dependiendo de la sesión que se encuentre realizando. Al oprimir cada botón, Emma explica el sonido que puede acompañar a cada expresión facial. Una vez se hayan oprimido todos los botones, el juego permitirá que se pase a la siguiente actividad.

En la actividad 2 de las sesiones 4 y 5 de imitación, llamada *Caja Sorpresa*, los participantes deben abrir una caja que contiene una imagen aleatoria donde se muestra una expresión facial. La imagen se muestra durante cinco segundos, en los cuales, el participante puede observar detenidamente. Después, se

remueve la imagen y se activa la cámara del participante. En este momento, Emma le pide que imite la expresión facial de la emoción que se observaba en la imagen. Esto se hizo porque en las pruebas piloto se observó que los participantes intentaban imitar la imagen a la perfección y no la expresión facial de la emoción que les enseñó Emma. Por ejemplo, en la imagen de miedo de la Figura 10 los participantes podrían tapar su boca con las manos, imitando a una de las personas en la imagen. Esto interfiere con el algoritmo de reconocimiento de expresiones faciales, por lo cual se decidió pedirles a los participantes que imitaran la emoción de la imagen, una vez la imagen no se muestre en la pantalla.

Finalmente, en la sesión 6 de la etapa de imitación, se llevan a cabo dos actividades que hacen uso de las seis emociones básicas, aumentando la dificultad en la imitación de las expresiones faciales. En la actividad 1, llamada *Emma Dice*, Emma le pide al participante que imite las expresiones faciales y los sonidos de las emociones, escogidos de manera aleatoria. Para la imitación de las expresiones faciales, se enciende la cámara del participante y se da realimentación visual y auditiva una vez el sistema detecte que imitó correctamente la expresión seleccionada. Para la imitación de los sonidos, no se creó un algoritmo de reconocimiento de sonidos, ya que se trata de una adición al proyecto y no su objetivo principal. En este caso, se habilita un botón y se solicita al participante oprimirlo una vez haya imitado correctamente el sonido de la emoción. Este proceso continúa durante 12 iteraciones: 6 relacionadas a expresiones faciales y 6 relacionadas a sonidos.

La última actividad de la etapa de imitación, llamada *Lotería*, refuerza el aprendizaje de la etapa de imitación al preguntarle al participante si una expresión facial corresponde a una emoción. En esta actividad, se observan seis cartas bocabajo en la izquierda de la ventana y se le pide al participante sacar otra carta de una baraja en la derecha de la pantalla. En esta carta, se indica el nombre de una emoción. Posteriormente, el participante debe seleccionar una de las seis cartas bocabajo; cuando la voltee, puede observar un rostro de Emma y el participante debe indicar si la expresión mostrada en la carta corresponde al nombre de la carta en la sección de la derecha. De hacerlo incorrectamente, se le indica visual y auditivamente que lo intente nuevamente. Si lo hace correctamente, Emma lo felicita y la actividad continúa hasta que se hayan seleccionado correctamente las cartas de las seis emociones. Al finalizar esta actividad, se vuelve al menú principal de Emmaciones y se obtiene acceso a la etapa de reconocimiento.

4.1.4 Actividades para el reconocimiento de expresiones faciales

El reconocimiento de expresiones faciales se refiere a la capacidad de los participantes de distinguir la emoción general presentada en una escena, la cual puede o no tener contexto. El reconocimiento de expresiones faciales se hace de cuatro maneras distintas: presentación de imágenes, relatos narrados, videos sin contexto y videos con contexto. La etapa de reconocimiento está enfocada en este aspecto; sin embargo, retoma la imitación de expresiones faciales y, en algunos casos, combina ambos procesos, de forma que aumenta la dificultad de manera progresiva en el aprendizaje de estos. La primera actividad de la sesión 1 de la etapa de reconocimiento se llamada *Ordenar la Cara*. En esta actividad se refuerza la visualización de partes particulares de la cara para identificar una expresión facial. En este juego, se muestran dos mitades de rostros de Emma, como los observados en la Figura 11. Las dos divisiones son: parte inferior, la cual incluye la nariz y la boca y parte superior, la cual incluye los ojos y las cejas. Así, la parte inferior del rostro es característica de una emoción y la parte superior del rostro es característica de otra. El objetivo del juego es que los participantes cambien estas mitades hasta que el rostro de Emma corresponda a la emoción que se está pidiendo. Una vez se hace esto, se indica al participante imitar la expresión facial que está haciendo Emma. A partir de esta actividad, se refuerza el concepto de observar la posición y orientación de cejas, ojos y boca. Dado que se utilizan las seis emociones, la dificultad del juego es mayor.

La segunda actividad de la primera sesión de la etapa de reconocimiento, llamada *Pop Emma*, está diseñada para que los participantes sean capaces de diferenciar la expresión facial entre dos emociones, dado que anteriormente solo se les había pedido seleccionar la emoción correcta. Este proceso se considera importante, dado que existen características semejantes en la expresión facial de distintas emociones. Por ejemplo, las cejas en el asco y en el enojo se ven de manera similar, pero hay características distintivas en las líneas nasolabiales y en la boca. *Pop Emma* consiste en presentarle al participante dos rostros de Emma escogidos de manera aleatoria. El nombre de la emoción correspondiente a uno de ellos aparece en la pantalla y se indica al participante escoger el rostro adecuado.

Para hacer más interesante la actividad, el participante escoge el rostro, no seleccionándolo con el cursor, sino moviendo su cabeza en dirección al rostro adecuado. Si mueve la cabeza hacia la derecha, un rostro se hará más grande y si la mueve a la izquierda, el otro rostro se hará más grande. Cuando uno de los rostros sea suficientemente grande, el juego asumirá que esta es la decisión del participante. De responder de manera incorrecta, se inicia nuevamente el juego con otros rostros y de manera correcta, debe imitar el rostro de Emma. Este proceso continúa hasta contestar correctamente seis veces, una por cada emoción.

En la primera actividad de las sesiones 2 y 3, llamada *Identifica la Emoción*, se le pide al participante por primera vez identificar una emoción a partir de la descripción de una escena. Al igual que en otros pares de sesiones, en la sesión 2 se debe identificar la alegría, el miedo o el asco mientras que en la sesión 3 se debe identificar la tristeza, la sorpresa y el enojo. En esta actividad, el participante escoge la narración que quiere escuchar a partir de imágenes, como las que se observan en la Figura 12. Estas imágenes fueron hechas o modificadas por estudiantes de psicología de la Corporación Universitaria Minuto de Dios UNIMINUTO con ayuda de los programas *Canva* y *Storybird*. Una vez escogida una imagen, iniciará una descripción corta de una situación emocional y el participante debe seleccionar la emoción que se representa en esa descripción. Dependiendo de su selección, se da realimentación visual y auditiva. El proceso continúa durante seis iteraciones: dos por cada emoción, de las correspondientes a la sesión escogida.

Figura 12. Ejemplos de imágenes utilizadas en la actividad *Identifica la Emoción*.

A continuación, se muestra un ejemplo de una de las descripciones brindadas a los participantes:

> *La mamá de David lo ha regañado por no hacerle caso cuando le dijo que recogiera sus juguetes. Cuando la mamá regaña a David él baja sus cejas y sus ojos se hacen pequeños ¿Sabes qué gesto tiene David?*

Las narraciones de estas descripciones se hicieron por miembros del proyecto, incluyendo estudiantes y profesores de la Escuela Colombiana de Ingeniería Julio Garavito y la Corporación Universitaria Minuto de Dios UNIMINUTO.

En la segunda actividad de las sesiones 2 y 3 de reconocimiento, llamada *Encuentra el Par*, se introduce un nuevo concepto: relacionar emociones a partir de dos contextos distintos. El juego se trata de un juego común de memoria, en el cual, teniendo varias cartas bocabajo, se deben seleccionar dos de ellas. Si son pareja, se eliminan del tablero. El juego termina cuando no haya más cartas en el tablero. No obstante, en este caso, se refuerzan los procesos de imitación, ya que, cuando se selecciona un par correcto, el participante debe imitar la emoción que escogió. Adicionalmente, se tienen tres niveles en el juego:

- Nivel 1: El participante debe encontrar rostros idénticos de Emma para completar pares.
- Nivel 2: El participante debe encontrar imágenes idénticas, como las observadas en la Figura 10, para completar pares.
- Nivel 3: El participante debe encontrar imágenes distintas que muestran la misma emoción para completar pares.

El nivel 1 y el nivel 2 ayudan a reforzar los conocimientos previos en reconocimiento de imágenes. Sin embargo, el nivel 3 es la parte donde más se progresa en el aprendizaje en esta actividad, ya que permite a los niños relacionar situaciones distintas, donde observan varios contextos en los cuales se puede presentar una misma emoción.

La última actividad de la sesión 3 de reconocimiento, llamada *Foto en Vivo*, está diseñada como una actividad de práctica, donde los participantes pueden escoger las emociones en el orden que quieran y tomarse una foto imitando la expresión facial correspondiente. Así, Emma le indica al participante si lo hizo bien o no, permitiéndole intentarlo nuevamente.

La primera actividad de las sesiones 4 y 5 de reconocimiento, llamada *¿Qué emoción soy?*, permite profundizar en el reconocimiento de emociones a partir de una narración, adquirida en *Identifica la Emoción*. En este caso no se presenta a los participantes una descripción de una escena, sino un cuento corto. En estos cuentos, los protagonistas son niños que se encuentran en situaciones con las cuales se podrían llegar a relacionar los participantes. Cada cuento dura aproximadamente tres minutos y, al igual que en *Identifica la Emoción*, son narrados por miembros del proyecto. Al final de cada cuento, se hace al participante una pregunta relacionada con el estado emocional de alguno de los personajes. Finalmente, se brinda realimentación visual y auditiva, dependiendo de la respuesta dada.

En la segunda actividad de las sesiones 4 y 5 de reconocimiento, llamada *Laberinto*, se prueban las habilidades de imitación y reconocimiento ya adquiridas por los participantes, añadiendo destreza por medio de comandos de movimiento. Así, los participantes refuerzan la habilidad de realizar varias tareas simultáneamente. En este juego, los participantes controlan uno de cuatro personajes: Sandy, Caro, Ale o Juan, llamados así en dedicación a los profesores que dirigieron este proyecto. Estos personajes se pueden controlar con las flechas del teclado y deben atravesar un laberinto generado automáticamente, de forma que nunca es igual. El propósito principal es encontrar rostros de Emma a lo largo del laberinto. Una vez se encuentra un rostro, el participante debe imitar la expresión facial que está haciendo Emma y llevar el rostro hasta la sección que tenga el nombre de la emoción correspondiente. El juego termina una vez se lleven todas las expresiones faciales de Emma a su correspondiente ubicación.

La última actividad de las sesiones 4 y 5 de reconocimiento, llamada *Elimina la Emoción*, introduce el concepto de aprender a reconocer emociones a partir de videos. En este juego, se muestra a los participantes videos sin contexto de personas expresando emociones. Esto significa que se trata de videos donde, sin que se conozca el motivo, la persona muestra una emoción específica. Los videos empiezan con una expresión facial emocionalmente neutral y progresivamente va cambiando a la emoción objetivo. De esta forma, los participantes reconocen la transición entre estados emocionales. Una vez se termina el video, el cual tiene una duración aproximada de 5 segundos, se visualizan botones con las emociones correspondientes a la sesión escogida. En este momento, se indica al participante que escoja todas las emociones que no aparecieron en el video, por lo cual, se requiere concentración de su parte.

La primera actividad de la sesión 6, llamada *Desliza la Emoción*, vuelve a requerir destreza por parte del participante. Aquí, se observa una imagen con el rostro de Emma, la cual se está moviendo de manera aleatoria por toda la ventana. Una vez alcanza esta imagen con el cursor, debe ubicarla encima del nombre de la emoción correcta. Un aspecto interesante de esta actividad es que, cuando se acerca el rostro de Emma a alguno de los nombres de las emociones, un avatar de Emma imita la emoción indicada en el nombre. Así, los participantes pueden observar la transición de expresiones faciales entre cada una de las emociones cuando él lo desee.

Finalmente, la última actividad de la sesión 6 de reconocimiento, llamada *Espejo*, introduce el concepto más complejo de la herramienta: el reconocimiento de emociones a partir de videos con contexto. Estos videos, similares a los de *Elimina la Emoción*, tienen contenido emocional que no se adquiere únicamente observando la expresión facial de las personas en ellos, sino también entendiendo la situación en la que se encuentran. Por ejemplo, en uno de ellos, una persona entra a la habitación de una niña, quien está viendo un video en un computador. Aunque la niña no abre la boca, si abre los ojos y la situación indica que está sintiendo sorpresa. Estos videos, con una duración aproximada de 10 segundos, buscan enseñarles a los niños que es posible reconocer una emoción a partir de situaciones y del reconocimiento de algunos de los aspectos de la expresión facial, incluso si la expresión facial no es exagerada. En la actividad, después que se muestran los videos, se pide al participante escoger la emoción apropiada e imitarla. La actividad termina una vez ha identificado e imitado las seis emociones.

4.1.5 Calentamientos y pausas activas

Como se mencionó anteriormente, al inicio de algunas sesiones y como transición entre estas, se desarrollaron ejercicios de calentamiento y pausas activas. Los calentamientos sirven para preparar al participante para las sesiones donde debe imitar expresiones faciales constantemente, así como para generar confianza entre el niño y el equipo de trabajo. Por su parte, las pausas activas sirven para que los participantes hagan actividades distintas entre sesiones y relajen sus músculos faciales. Todas las actividades que serán descritas a continuación fueron diseñadas por estudiantes de psicología, miembros del proyecto de investigación.

La primera sesión de calentamiento se da antes de iniciar la sesión 1 de imitación. Esta consiste en dos partes:

- Masaje del rostro: Se le recomienda al participante lavarse las manos para iniciar las actividades e imitar las acciones del investigador. Así, el participante debe cerrar ambas manos en forma de puño y pasar sus nudillos por todas las partes del rostro, realizando movimientos circulares. Luego, debe abrir las manos y realizar movimientos horizontales con sus palmas, acariciando zonas como mejillas y frente.

- Movimientos de cabeza: Nuevamente, el participante debe imitar al investigador, moviendo la cabeza lentamente hacia los lados y de arriba hacia abajo, lentamente. Luego, se le pide mantener la mirada en un punto fijo y mover el cuerpo hacia los lados sin mover la cabeza. Finalmente, debe abrir y cerrar lentamente la boca, evitando lastimarse.

Al finalizar esta sesión, se realiza la primera pausa activa, la cual consta de dos actividades:

- Imitar movimiento de manos: Durante esta actividad el investigador debe realizar de forma aleatoria movimientos con los pulgares de sus manos y el participante debe imitarlos. Estos movimientos no se realizan con ningún patrón específico y la velocidad aumenta progresivamente.

- Reflejo: El participante observa e imita los movimientos realizados por el investigador que involucraran libremente las partes del cuerpo. Esta actividad es similar a la a anterior; sin embargo, involucra un mayor movimiento por parte del participante.

Antes de iniciar la sesión 4 de la etapa de imitación se realiza el segundo calentamiento de Emmaciones, que consiste en:

- Juego de miradas: El participante debe realizar movimientos con sus ojos y cejas de tal modo que pueda abrir los ojos lo que más pueda sin lastimarse. Luego, debe cerrar sus ojos durante dos

segundos. Más adelante, debe intentar mover cada ceja hacia arriba de manera independiente y con ambas al tiempo.

- Jugando con las vocales: El participante menciona las vocales en el orden que desee, pero debe hacerlo lo más exagerado posible, evitando lastimarse. Esta actividad la realiza en compañía del investigador.

Al finalizar esta sesión y, en preparación para la sesión 5, se realiza otra pausa activa, que consta de las siguientes actividades:

- Estatuas: El investigador le pregunta al participante una canción que quiera bailar. Si el participante no elige, se reproduce una canción elegida por los investigadores. Tanto el investigador como el niño deben bailar al ritmo de la canción. El investigador pausa la reproducción y en este momento el niño debe quedarse quieto en la posición en la que esté al momento de pausarse. A los 10 segundos, el investigador debe volver a reproducir la canción. Este proceso se hace cinco veces.

- Copión: El investigador realiza diferentes movimientos y desplazamientos en un espacio pequeño. A su vez, el participante debe observar. Después de finalizar las acciones, el participante imita las acciones que recuerde haber observado.

En la etapa de reconocimiento también se tienen dos secciones de calentamiento y dos secciones de pausas activas, distribuidas de igual manera que en la etapa de imitación. El calentamiento previo a la sesión 1 de reconocimiento contiene las siguientes actividades:

- Masaje del rostro, actividad ya descrita anteriormente, en el calentamiento previo a la sesión 1 de imitación.

- Estiramientos: El participante mueve su cabeza hacia la izquierda y hacia la derecha, durante dos minutos. Después de terminar, el participante mueve la cabeza en círculos y, al finalizar, hace lo mismo con las manos y muñecas hacia adelante y atrás durante un minuto.

Al finalizar esta sesión, inicia la primera pausa activa de la etapa de reconocimiento:

- Imitar movimiento de las manos, actividad ya descrita anteriormente, en la primera pausa activa de la etapa de imitación.

- Robot: el investigador mueve las partes del cuerpo como si fuera un robot. Luego de realizar estos movimientos, el participante los imita.

Antes de iniciar la sesión 4 de la etapa de reconocimiento, se realizan los siguientes ejercicios de calentamiento:

- Movimientos de cabeza, actividad ya descrita anteriormente, en el calentamiento previo a la sesión 1 de imitación.

- Fotografías: Se le pide al participante prender la cámara de su computador y tomarse fotos realizando distintos gestos de forma libre. Es importante que se implemente el uso de las manos y otras partes del cuerpo. Más adelante, el investigador propone gestos con el fin de que el participante los imite.

Finalmente, al finalizar esta sesión, se da la última pausa activa del proceso, que contiene las actividades Estatuas y Copión, descritas en la segunda pausa activa de la etapa de imitación.

4.2 Algoritmo de reconocimiento de expresiones faciales

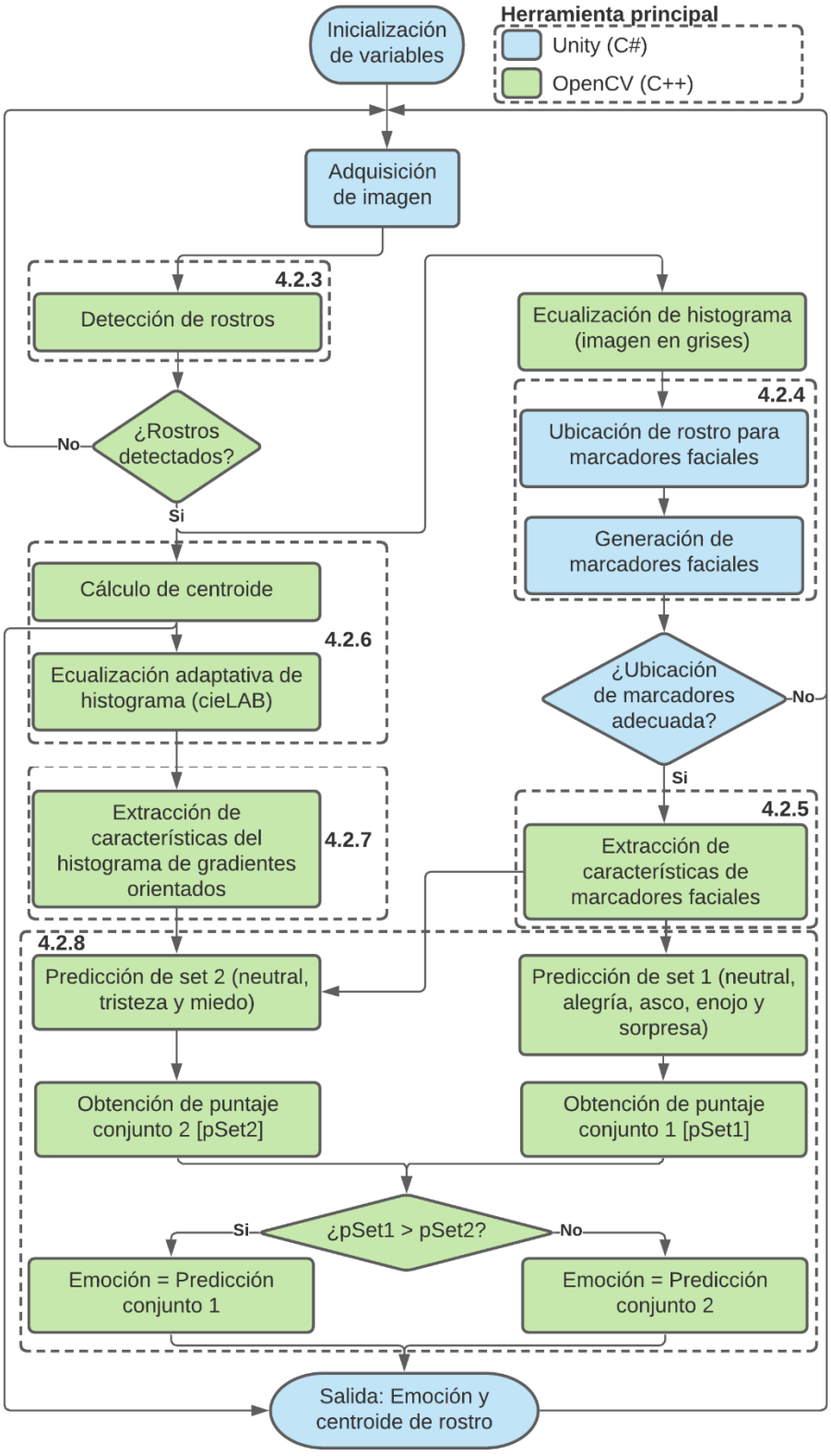

Figura 13. Diagrama general del algoritmo de reconocimiento de las expresiones faciales de las emociones.

4.2.1 Diagrama general para reconocimiento de expresiones faciales

Para el desarrollo de este proyecto, el reconocimiento de expresiones faciales fue el mayor reto técnico, dado que, a conocimiento de nuestro equipo de investigación, no existen antecedentes con código abierto de videojuegos que reconocen expresiones faciales en tiempo real, por lo cual existe un alto componente de innovación. Para lograr este objetivo, se diseñó el plan de desarrollo que se observa en el diagrama de la Figura 13. Este proceso incluye adecuación de las imágenes para su posterior procesamiento, la detección de rostros, el uso de modelos de ubicación de marcadores faciales, la extracción de características y finalmente la categorización de las expresiones faciales. Dado que en el juego *Pop Emma* es necesario conocer la ubicación del rostro del participante, también se incluyó la ubicación de estas coordenadas, como adición al *pipeline* del reconocimiento de expresiones faciales. En la figura se observa la subsección correspondiente a cada paso dentro del diagrama. Es importante tener en cuenta que el diagrama observado en la Figura 13 es una versión modificada del diseño original, ya que las técnicas utilizadas en cada bloque de este alteraron el orden de los pasos y la complejidad del algoritmo. Por ejemplo, la técnica utilizada para detectar rostros es muy robusta en comparación a otros métodos, por lo cual fue posible eliminar secciones de la etapa de pre-procesamiento.

Dos aspectos importantes del algoritmo desarrollado son, que fue planeado para funcionar con código libre y ser compacto. Esto significa que no es necesario pagar licencias ni instalar programas externos para utilizarlo. Fue posible lograr este objetivo por medio del uso de aplicaciones libres, como *Unity*, *OpenCV* y *Spyder*. Unity es un motor de videojuegos de código libre que le brinda a desarrolladores una interfaz gráfica con todas las herramientas necesarias para elaborar un videojuego. En este proyecto, es el programa principal utilizado para desarrollar la herramienta de estimulación. En la subsección 4.3 se describe el uso de Unity, en su versión 2019.4 (LTS). OpenCV es una biblioteca libre de visión artificial implementada en C++ y Python que brinda funciones para que, entre otros objetivos, sea posible generar soluciones a problemas computacionales por medio de procesamiento de imágenes. En este proyecto, se utiliza como fuente principal de funciones de procesamiento de imágenes y de aprendizaje automático. A lo largo de esta subsección se describirán los instantes en los que fue utilizado OpenCV, en su versión 4.3. Spyder es un entorno de desarrollo para Python de acceso libre que facilita la programación científica en Python, brindando herramientas como la exploración de variables y la integración de distintas bibliotecas de ciencia computacional, de forma que facilita la depuración en el desarrollo de aplicaciones. En este proyecto, su uso principal es la generación de modelos de aprendizaje automático para el reconocimiento de expresiones faciales, aprovechando la ventaja que los modelos de aprendizaje automático de OpenCV desarrollados en Python son compatibles con C++ y viceversa. El uso de Spyder se hace más claro en la subsección 4.2.8.

Para el desarrollo de este proyecto, fue necesaria el uso y la generación de bibliotecas dinámicas (DLL, por sus siglas en inglés). Los DLLs son un tipo de archivo similar a los archivos ejecutables. Sin embargo, no se pueden ejecutar de manera directa; se tratan de una implementación de bibliotecas compartidas. En general, un DLL puede contener funciones, variables y recursos de la misma forma que son contenidos en un archivo ejecutable. Su principal ventaja es que, a diferencia de las librerías estáticas, no se encuentran vinculadas a un ejecutable durante la compilación, sino que se vinculan durante la ejecución. Esto permite que los DLLs se puedan modificar en cualquier momento, sin necesidad de volver a compilar el ejecutable al cual se vincula. En el caso de este proyecto, los DLLs son fundamentales, dado que las librerías externas a Unity, como OpenCV, funcionan a partir de DLLS, brindando funciones de uso general que se pueden acomodar a las necesidades de distintos proyectos. Así, cuando se hace uso de OpenCV, es necesario contar con los DLLs que implementan distintas funciones. Por ejemplo, hay un DLL para la implementación de algoritmos de detección de rostros y otro para la implementación de algoritmos de ubicación de marcadores faciales. Adicionalmente, los DLLs cuentan con un atributo llamado *dllexport*, el cual permite transmitir datos entre lenguajes de la familia C. Este atributo es esencial para el desarrollo del proyecto, ya que permite desarrollar los algoritmos de visión artificial en C++ y transmitir sus resultados a C#, para que Unity los pueda implementar en los juegos desarrollados.

En la leyenda de la Figura 13 se puede observar que el diseño del algoritmo de reconocimiento de expresiones faciales se divide entre las partes que se desarrollan directamente desde Unity y aquellas que se crean a partir de DLLs generados en C++, para habilitar el uso de OpenCV. Principalmente, esto se hace porque no es posible que dos procesos utilicen la misma cámara web simultáneamente,

independientemente del sistema operativo utilizado. En ciertas actividades de Emmaciones es importante que se active la cámara web del participante, por lo cual es necesario que un proceso de Unity sea el indicado para utilizar los recursos de la cámara web. Posteriormente, la imagen adquirida por la cámara web es transferida a un DLL de C++ para su procesamiento. En la figura se observa que la ubicación de marcadores faciales se hace directamente desde Unity; la razón de esto se explica en la subsección 4.2.4. Al finalizar todas las etapas del algoritmo de reconocimiento, se retorna a Unity el índice codificado de la emoción elegida y el centroide del rostro detectado por la cámara.

En las próximas subsecciones se describirán cada uno de los pasos contemplados en el diagrama y, en casos específicos, se mostrarán los argumentos utilizados para la selección de distintas técnicas, al compararlas con aquellas que tienen objetivos similares. Es importante resaltar que algunas técnicas fueron descartadas dado que no se implementan en OpenCV o directamente desde Unity. Aunque el uso de bibliotecas adicionales es posible, esto requeriría de un mayor tiempo de desarrollo y depuración, por lo cual, su uso es considerado para versiones posteriores de Emmaciones o para nuevas herramientas.

Dado que en esta subsección se consiguieron varios resultados parciales, a continuación, se explicará la metodología utilizada para tomar distintas decisiones en el desarrollo del algoritmo de reconocimiento de emociones faciales. En la subsección 5.1, correspondiente a los resultados correspondientes a este algoritmo, se mostrarán los datos que llevaron a estas decisiones.

4.2.2 Preprocesamiento de imágenes

Como se explicó previamente, el preprocesamiento para el algoritmo de reconocimiento se enfocó en la corrección de contraste. Esto se consideró en tres partes del algoritmo: antes de la detección de rostros, antes de la generación de marcadores faciales y antes de la extracción de características del histograma de gradientes orientados (HOG). La efectividad de cada técnica no se midió de manera cuantitativa durante esta etapa del desarrollo, ya que esto hubiera requerido de bases de datos con etiquetas manuales que indiquen una verdad absoluta. Por lo tanto, se habría necesitado de una base de datos etiquetada para la detección de rostros y otra para la ubicación de marcadores faciales. En cambio, la efectividad de cada método se probó de manera empírica, a partir de inspección visual. Dado que en estos casos es fácil reconocer si un algoritmo está realizando su tarea correctamente y las diferencias entre distintas técnicas es notoria, se considera que esta es una forma válida de discriminar técnicas de preprocesamiento, considerando su uso en varias aplicaciones previas [75], [76].

Las técnicas de preprocesamiento probadas para el desarrollo del algoritmo son la ecualización de histograma tradicional y la ecualización adaptativa de histograma limitada en el contraste (CLAHE). De igual forma, cada técnica se probó con distintos espacios de color: grises, RGB, HSL y cieLAB. Las técnicas encontradas en la revisión de literatura que no se pudieron implementar para preprocesamiento fueron la transformada Wavelet y técnicas de aprendizaje profundo. Por su parte, la transformada Wavelet no se utilizó ya que no se encuentra en OpenCV y su implementación manual tiene una alta complejidad. Por otro lado, no se encontraron modelos entrenados de aprendizaje profundo de uso libre para la corrección de contraste.

Aunque estos métodos se combinaron con múltiples métodos para detectar rostros y ubicar marcadores faciales, en este documento solo se mostrará la efectividad de cada uno con los métodos elegidos. Por un lado, en el caso de la detección de rostros, se decidió utilizar el algoritmo de aprendizaje profundo SSD, explicado en la subsección 3.2.2. En la subsección 4.2.3 se explica esta decisión en detalle. Por otro lado, en el caso de la ubicación de marcadores faciales, se decidió utilizar el algoritmo desarrollado por Kazemi [65], el cual se explica en la subsección 3.2.4. En la subsección 4.2.4 se explica esta decisión en detalle.

4.2.3 Detección facial

Para la detección de rostros, se consideraron cuatro algoritmos, ya explicados previamente en la subsección 3.2.2. Los métodos evaluados fueron los siguientes:

- Cascadas Haar [55]
- Detección a partir de características HOG [57]

- Single Shot MultiBox Detector (SSD) [58]
- Max-Margin Object Detection (MMOD) [59]

Para evaluar cuál de estos métodos es el más efectivo, se siguió un procedimiento similar de la subsección 4.2.2. En este caso, se probó la efectividad de los algoritmos para detectar rostros en distintas condiciones donde se podrían encontrar los participantes durante el uso de Emmaciones. Idealmente, los participantes se encuentran quietos mirando hacia la cámara; sin embargo, dado que los participantes son niños, se busca desarrollar un algoritmo capaz de detectar emociones en distintas poses.

Se evaluaron tres poses en particular: Mirando directamente hacia la cámara con la cabeza derecha, mirando directamente hacia la cámara con la cabeza girada y una imagen de perfil. No se incluyó el cambio de iluminación dado que esto ya se detalló en la subsección 4.2.2. Dos condiciones adicionales que se suelen tener en cuenta para probar estos algoritmos es el reconocimiento de rostros durante oclusión y el reconocimiento de múltiples rostros. Sin embargo, dado que los participantes no se encuentran en estas condiciones según las normas del protocolo experimental, no se tomaron en cuenta para decidir el algoritmo utilizado. Es importante tener en cuenta que los algoritmos seleccionados no deben únicamente encontrar el rostro, sino cubrirlo lo mejor posible, incluyendo frente y mejillas, dado que se obtienen características de texturas a partir de estas zonas.

Para realizar una selección objetiva, se evaluó la efectividad del algoritmo a partir de tres criterios: La selección correcta del rostro (lo cual incluye falsos positivos y falsos negativos como posibles errores), la inclusión de la frente y la inclusión de las mejillas. En la subsección 5.1.4 se muestra que dos de los cuatro algoritmos evaluados cumplen con efectividad estos criterios: SSD y MMOD. No obstante, al realizar una prueba de tiempo de cómputo, se observó que la velocidad de SSD es considerablemente mayor, lo cual es útil al momento de reconocer expresiones faciales en tiempo real. Por este motivo, se decidió utilizar SSD para detectar rostros.

Por otro lado, dado que para el juego *Pop Emma* es necesario conocer la posición del rostro del participante, una vez se realiza la detección del rostro se calcula el centroide de este. Este es un proceso sencillo: El algoritmo de detección genera cuatro variables: La posición horizontal X de la esquina superior izquierda, la posición vertical Y de la esquina superior izquierda, el ancho W del recuadro y el alto H del recuadro. Para facilitar los cálculos en Unity, se obtiene la posición normalizada; lo que significa que el centroide se compone de dos coordenadas C_x y C_y, las cuales estarán en el rango [0, 1] cada una. Para lograr esta normalización, se utiliza el ancho W_{cam} y el alto H_{cam} de la cámara del participante. En la ecuación 1 se observa el cálculo del centroide a partir de las variables mencionadas.

$$C_x = \frac{2X + W}{2W_{cam}}, C_y = \frac{2Y + H}{2H_{cam}} \quad (1)$$

4.2.4 Ubicación de marcadores faciales

Para decidir el algoritmo de ubicación de marcadores faciales, y todos los algoritmos posteriores, se utilizó el modelo de detección de rostros SSD y corrección de contraste por medio de ecualización de histograma en la imagen en grises. Dentro de los algoritmos de ubicación de marcadores faciales encontrados en la revisión de literatura, solo uno de ellos está implementado de forma nativa en OpenCV. Este algoritmo se basa en características binarias locales (LBF) para ubicar los marcadores [77]. No obstante, luego de una búsqueda más profunda, se encontró que el algoritmo creado por Kazemi, el cual está basado en conjuntos de árboles de regresión [65], está implementado en Dlib. Dlib es una biblioteca que contiene algoritmos de aprendizaje automático, la cual está implementada en C++ [78] y contiene *wrappers* para Python. Para probar la funcionalidad del algoritmo de Kazemi, se creó un proyecto, el cual se desarrolló en su totalidad en C++. Las pruebas mostradas a continuación se hicieron a partir de la implementación de ambos algoritmos (LBF y Kazemi) en C++. En la Figura 14 se observan los índices estándar dados a cada uno de los marcadores en un modelo de 68 marcadores faciales. A lo largo de esta subsección, se hará referencia a estos índices.

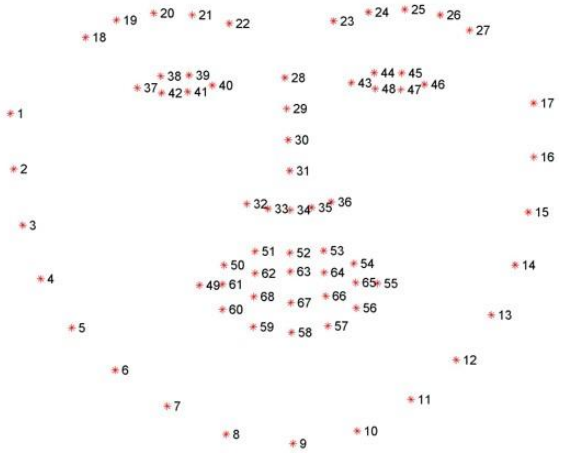

Figura 14. Índices estándar para la ubicación de marcadores faciales en el modelo de 68 puntos.

Una forma adecuada de haber comparado los algoritmos es por medio de imágenes etiquetadas, calculando el error entre los puntos calculados y los puntos etiquetados, para cada uno de los algoritmos. Sin embargo, las dos bases de datos de acceso libre que contienen imágenes etiquetadas con la ubicación de 68 marcadores faciales no son aptas para estas pruebas. Por un lado, la base de datos XM2VTS solo contiene rostros mirando hacia el frente, por lo que no se podrían probar poses deseadas. Por otro lado, la base de datos MultiPie incluye personas observadas desde distintas poses, más no realizando expresiones deseadas, como apertura de boca o pronunciación de líneas nasolabiales.

Por este motivo, al igual que en la detección de rostros, los algoritmos de ubicación de marcadores faciales explorados fueron probados en tiempo real, observando las ventajas y desventajas de cada uno en distintas situaciones. Las imágenes que se mostrarán a continuación buscan replicar los casos donde se observaron resultados interesantes en la ubicación de marcadores faciales que pueden suceder durante las pruebas. Al igual que en la detección de rostros, para estas pruebas no se incluyeron imágenes con distintas iluminaciones, ya que este tema fue abordado en la subsección 4.2.2. De igual forma, no se incluyó la capacidad del algoritmo de ubicar marcadores faciales para múltiples rostros o imágenes donde se observa oclusión.

Se realizaron varias pruebas para evaluar la efectividad en la ubicación de marcadores faciales de estos dos algoritmos: condiciones ideales de pose (mirando hacia el frente, expresión neutra), mirando levemente hacia un lado con expresión neutra, girando levemente la cabeza hacia un lado, realizando la expresión facial de la sorpresa y realizando la expresión facial del asco. Aunque en realidad se realizaron más pruebas en tiempo real, en este documento solo se exponen aquellas con resultados más interesantes, dado que, en otros casos, los algoritmos tienen efectividad similar. Como se muestra en la subsección 5.1.5, se encontró que el algoritmo de Kazemi logra una mayor efectividad en la ubicación de marcadores faciales; sin embargo, su implementación no era sencilla.

Como se mencionó en la subsección 4.2.1, las funciones de OpenCV se implementan a través de la generación de DLLs en el lenguaje de programación C++, para permitir la comunicación con Unity. No obstante, la generación de un DLL no fue posible en el caso de Dlib, ya que las bibliotecas que se utilizaron para las herramientas de Dlib son estáticas y no dinámicas. Esto implica que éstas se vinculan al ejecutable generado en C++ en el momento de compilación y no es posible hacer uso del atributo *dllexport.* Después de una extensa búsqueda en foros y de consultar a varios desarrolladores, no fue posible encontrar una implementación de Dlib en un DLL. No obstante, unos desarrolladores habían logrado generar el DLL de esta biblioteca manualmente y crearon un producto enfocado en el uso de Dlib directamente desde Unity, en el cual crearon todas las funciones y clases necesarias para permitir la funcionalidad de Dlib en C#. A este producto (comúnmente llamado *asset* en el desarrollo de videojuegos) lo llamaron *Dlib FaceLandmark Detector* y actualmente se vende a USD40 como un servicio en *Asset Store*, la tienda de Unity [79]. A falta de otras opciones y observando la drástica diferencia en la efectividad de los algoritmos de ubicación de marcadores faciales analizados, se decidió hacer esta compra.

Aunque se encontró que este algoritmo funciona como esperado, los creadores del asset asumen que el usuario ha comprado anteriormente *OpenCV for Unity*, un asset con costo de USD95 que implementa OpenCV en Unity [80], evitando la necesidad de generar librerías dinámicas. Dado que ya se había logrado vincular OpenCV con Unity sin necesidad de este asset, no tenía sentido hacer esta compra. Por este motivo, se decidió no hacer uso de las funciones prefabricadas de *Dlib FaceLandmark Detector* y, en vez, se exploró el paquete, extrayendo aquellas funciones de utilidad para el proyecto y descartando aquellas que podían ser implementadas externamente por medio de un DLL. Por este motivo, el diagrama de la Figura 13 indica que la ubicación de marcadores faciales es realizada directamente en Unity. Esto implica que el DLL generado en C++ debe contener tres funciones de exportación, a las cuales se llamaron *Init*, *Preprocessing* y *DetectEmotion*:

- *Init* inicializa todas las variables necesarias para el preprocesamiento y el reconocimiento de expresiones faciales, además de cargar todos los archivos necesarios para su correcto funcionamiento, como el modelo de aprendizaje profundo para detección de rostros y el modelo de ubicación de marcadores faciales.
- *Preprocessing* realiza todas las etapas del diagrama hasta la detección del rostro, información que es enviada a Unity para que Dlib pueda ubicar los marcadores faciales.
- En este momento, Unity devuelve las coordenadas de los marcadores faciales a *DetectEmotion*, función encargada de las etapas restantes del diagrama, hasta el reconocimiento de expresiones faciales.

4.2.5 Extracción de características de marcadores faciales

Aunque una parte importante del algoritmo de reconocimiento de emociones es la ubicación de marcadores faciales, por sí solos no ayudan a caracterizar una emoción. Por este motivo, se generaron una serie de características que relacionan la información brindada por cada marcador para determinar la presencia de distintas AUs. Adicionalmente, se evaluó el uso de características independientes de las AUs para analizar su utilidad en la caracterización de expresiones faciales. Es importante tener en cuenta que no todas las características que se describen a continuación fueron utilizadas en el entrenamiento final de los modelos de reconocimiento de expresiones faciales, dado que hubo un proceso de filtrado de características, eliminando aquellas que no mostraron tener relevancia para la clasificación.

Las características que se crearon tienen en cuenta ángulos y distancias entre marcadores, además de curvaturas generadas entre ellos. Las características fueron divididas en cuatro grupos principales: ojos, cejas, boca y características generales, entre las que se encuentran apertura de fosas nasales e información del ángulo de cabeceo. En la Tabla 4 se describe cada característica obtenida a partir de la ubicación de marcadores faciales, en la cual se tienen en cuenta el grupo al que pertenece cada característica, el tipo de característica (distancia vertical, distancia horizontal, ángulo relativo o valor absoluto) y la ecuación utilizada para calcularla, para un total de 23 características relacionadas con marcadores faciales. La importancia del tipo de característica yace en la normalización de la característica, de forma que cada una sea independiente de la posición, rotación y escala del rostro que se está analizando. La normalización realizada para cada tipo de característica se detalla más adelante. En las ecuaciones de la tabla, X_n es la coordenada *x* del marcador con índice *n* y Y_n es la coordenada *y* del marcador con índice *n*, según los índices establecidos por la Figura 14.

Tabla 4. Descripción de las características extraídas que están relacionadas con la ubicación de marcadores faciales. Azul: distancia vertical. Verde: distancia horizontal. Amarillo: ángulo relativo. Naranja: valor absoluto.

Grupo	Nombre	Ecuación
Ojos	Altura del ojo izquierdo	$\sqrt{\left(\frac{X_{44}+X_{45}}{2}-\frac{X_{47}+X_{48}}{2}\right)^2+\left(\frac{Y_{44}+Y_{45}}{2}-\frac{Y_{47}+Y_{48}}{2}\right)^2}$
	Ancho del ojo izquierdo	$\sqrt{(X_{43}-X_{46})^2+(Y_{43}-Y_{46})^2}$

	Altura del ojo derecho	$\sqrt{\left(\frac{X_{38}+X_{39}}{2}-\frac{X_{41}+X_{42}}{2}\right)^2+\left(\frac{Y_{38}+Y_{39}}{2}-\frac{Y_{41}+Y_{42}}{2}\right)^2}$
	Ancho del ojo derecho	$\sqrt{(X_{37}-X_{40})^2+(Y_{37}-Y_{40})^2}$
Cejas	Rotación interna de la ceja izquierda	$\tan^{-1}\left(\frac{Y_{23}-Y_{25}}{X_{23}-X_{25}}\right)$
	Radio de curvatura de la ceja izquierda	$R_{cur}([X_{23},Y_{23}],[X_{25},Y_{25}],[X_{27},Y_{27}])..Detalles\ en\ ecuación\ 2.$
	Cercanía entre ceja izquierda y tabique nasal	$\sqrt{(X_{23}-X_{28})^2+(Y_{23}-Y_{28})^2}$
	Altura de la ceja izquierda	$\frac{\lvert mX_{25}-Y_{25}+Y_1-mX_1\rvert}{\sqrt{m^2+1}}, m=\left(\frac{Y_{17}-Y_1}{X_{17}-X_1}\right)$
	Rotación interna de la ceja derecha	$\tan^{-1}\left(\frac{Y_{22}-Y_{20}}{X_{22}-X_{20}}\right)$
	Radio de curvatura de la ceja derecha	$R_{cur}([X_{22},Y_{22}],[X_{20},Y_{20}],[X_{18},Y_{18}]).Detalles\ en\ ecuación\ 2.$
	Cercanía entre ceja derecha y tabique nasal	$\sqrt{(X_{22}-X_{28})^2+(Y_{22}-Y_{28})^2}$
	Altura de la ceja derecha	$\frac{\lvert mX_{20}-Y_{20}+Y_1-mX_1\rvert}{\sqrt{m^2+1}}, m=\left(\frac{Y_{17}-Y_1}{X_{17}-X_1}\right)$
Boca	Apertura de la boca	$\sqrt{(X_{52}-X_{58})^2+(Y_{52}-Y_{58})^2}$
	Apertura de la parte interna de la boca	$\sqrt{(X_{63}-X_{67})^2+(Y_{63}-Y_{67})^2}$
	Ancho de la boca	$\sqrt{(X_{49}-X_{55})^2+(Y_{49}-Y_{55})^2}$
	Grosor de labio superior	$\sqrt{(X_{63}-X_{52})^2+(Y_{63}-Y_{52})^2}$
	Grosor de labio inferior	$\sqrt{(X_{67}-X_{58})^2+(Y_{67}-Y_{58})^2}$
	Ángulo de curvatura del labio superior	$\cos^{-1}\left(\frac{(X_{55}-X_{52})(X_{49}-X_{52})+(Y_{55}-Y_{52})(Y_{49}-Y_{52})}{\sqrt{((X_{55}-X_{52})^2+(Y_{55}-Y_{52})^2)((X_{49}-X_{52})^2+(Y_{49}-Y_{52})^2)}}\right)$
	Ángulo de curvatura del labio inferior	$\cos^{-1}\left(\frac{(X_{55}-X_{58})(X_{49}-X_{58})+(Y_{55}-Y_{58})(Y_{49}-Y_{58})}{\sqrt{((X_{55}-X_{58})^2+(Y_{55}-Y_{58})^2)((X_{49}-X_{58})^2+(Y_{49}-Y_{58})^2)}}\right)$
General	Apertura de fosas nasales	$\sqrt{(X_{32}-X_{36})^2+(Y_{32}-Y_{36})^2}$
	Cercanía entre las cejas	$\sqrt{(X_{22}-X_{23})^2+(Y_{22}-Y_{23})^2}$
	Valor relativo a la guiñada de la cabeza	$\frac{\sqrt{(X_1-X_{31})^2+(Y_1-Y_{31})^2}}{\sqrt{(X_{17}-X_{31})^2+(Y_{17}-Y_{31})^2}}$
	Valor relativo al cabeceo de la cabeza	$\frac{Y_{17}+Y_1}{2}-Y_{31}$

Una de las características utilizadas es el radio de curvatura, el cual se puede observar en la ecuación 2. Este es obtenido a partir de la curvatura de Menger, y se expone en una ecuación aparte para una mayor facilidad en la lectura de la Tabla 4.

$$R_{cur}(P_1,P_2,P_3) = \frac{2\left|P_{1_x}P_{2_y}+P_{2_x}P_{3_y}+P_{3_x}P_{1_y}-P_{1_y}P_{2_x}-P_{2_y}P_{3_x}-P_{3_y}P_{1_x}\right|}{\sqrt{\left(\left(P_{1_x}-P_{2_x}\right)^2+\left(P_{1_y}-P_{2_y}\right)^2\right)\left(\left(P_{2_x}-P_{3_x}\right)^2+\left(P_{2_y}-P_{3_y}\right)^2\right)\left(\left(P_{3_x}-P_{1_x}\right)^2+\left(P_{3_y}-P_{1_y}\right)^2\right)}} \quad (2)$$

Se buscó que las características de la Tabla 4 fueran independientes a la posición, rotación y escala del rostro en la imagen, dado que los participantes podrían encontrarse en distintas poses al momento de

imitar las expresiones faciales. Por un lado, todas las características son inherentemente independientes a la posición, ya que en ninguna se está registrando las coordenadas de un marcador. Sin embargo, si se registra la posición de un marcador con relación a otro. Así, por ejemplo, la apertura de las fosas nasales tiene el mismo valor si el participante se encuentra a la derecha o a la izquierda de la cámara.

Así, solo es necesario modificar las características para que sean independientes de la rotación y de la escala del rostro. En cuanto a la rotación, se busca normalizar las características señaladas en amarillo en la Tabla 4, ya que se trata de características cuyo valor puede cambiar con la rotación del rostro. Para normalizar los ángulos obtenidos en estas características se toma como punto de referencia la orientación del rostro, a partir de marcadores encontrados en el contorno de este, como se muestra en la ecuación 3, donde θ es la orientación del rostro en el eje del alabeo. Así, si θ es positivo, el rostro está orientado hacia la izquierda y si es negativo, está orientado hacia la derecha. Para normalizar las características correspondientes, se le resta θ al valor obtenido en la característica.

$$\theta = \tan^{-1}\left(\frac{Y_{17} - Y_1}{X_{17} - X_1}\right) \tag{3}$$

En cuanto a la escala, se busca normalizar las características señaladas en azul y en verde en la Tabla 4. Por un lado, las características en azul varían cuando el largo del rostro se modifica, mientras que las características en verde varían cuando el ancho del rostro se modifica. Una forma sencilla de normalizar estas características es por medio del recuadro del rostro detectado por SSD. Este recuadro brinda la altura *H* y el ancho *W* del rostro. Así, las características resaltadas en azul en la Tabla 4 se normalizan por medio de la ecuación 4 y las características en verde se normalizan por medio de la ecuación 5, donde C_N es la característica normalizada y *C* es la característica original.

$$C_N = \frac{C}{H} \tag{4}$$

$$C_N = \frac{C}{W} \tag{5}$$

Por último, las características de la Tabla 4 resaltadas en naranja se refieren a características con valores absolutos que son inherentemente independientes de posición, escala y rotación. Por un lado, la curvatura de los labios son valores de ángulos, por lo cual no dependen de la posición ni de la escala. Sin embargo, también es independiente a la rotación del rostro, porque cuando el rostro rota, los puntos que se tienen en cuenta para la curvatura también rotan de la misma manera. Por otro lado, el valor relativo a la guiñada de la cabeza se refiere a un valor adimensional que indica la razón entre la distancia del contorno derecho del rostro a la nariz y del contorno izquierdo del rostro a la nariz. Así, la distancia más pequeña entre estas dos indicará el lado hacia el cual está mirando el participante. Dado que se trata de una razón entre distancias, esta característica es independiente a la escala. De igual forma, las distancias no tienen en cuenta la posición ni la rotación del rostro, por lo cual la característica también es independiente a estos dos aspectos del rostro.

4.2.6 Preparación de los datos para el cálculo de HOG

La principal razón para utilizar HOG es para poder cubrir características importantes de las expresiones faciales de las emociones que no se pueden cubrir por medio del uso de marcadores faciales, teniendo en cuenta las unidades de acción descritas en la subsección 3.2.3. A partir de la Tabla 1 y la Tabla 2 se llega a la conclusión que hay tres zonas del rostro donde es relevante tener información sobre la textura de este:

- Mejillas: La información de más relevancia en las mejillas son las líneas nasolabiales, las cuales se pronuncian en las expresiones faciales de alegría, tristeza y asco.
- Ceño: El ceño muestra bordes más pronunciados en las expresiones faciales de asco, enojo, tristeza y miedo.

- Frente: La frente muestra bordes en distintas direcciones dependiendo de la expresión facial que se esté haciendo. Por ejemplo, en la sorpresa, se generan bordes curvos, mientras que en enojo se observan líneas horizontales.

Dado que solo es importante obtener información de la textura en esas tres áreas, se decide no obtener información de los gradientes orientados en toda la imagen brindada sino solamente en esas zonas. Adicionalmente, teniendo en cuenta que la expresión facial de las emociones es simétrica y para evitar información redundante y, consecuentemente, reducir tiempos de cómputo, solo se consideró un lado del rostro para las áreas de las mejillas y la frente.

Para decidir el lado del rostro que se consideraría en el análisis de texturas, se tomó en cuenta la rotación de este, dando preferencia a la mejilla y mitad de frente que tuvieran una mayor área de cobertura. Para lograr esto y la ubicación de las zonas de interés, se utilizó la información brindada por los marcadores faciales y el algoritmo de detección de rostros. La rotación del rostro se calculó por medio de los marcadores 1 y 17, correspondientes a los extremos derecho e izquierdo del contorno del rostro. Si la altura del marcador 1 es mayor a la del marcador 17, se considera que el rostro está rotado hacia la izquierda, de lo contrario, se considera que el rostro está rotado hacia la derecha. Así, se escoge la mejilla y la mitad de la frente opuestas a la dirección a la que esté rotado el rostro, lo que, indirectamente, selecciona las zonas con mayor área.

En la Tabla 5 se observan los criterios que se tuvieron en cuenta para estas tres zonas, donde X_n se refiere a la coordenada x en el marcador con índice n, Y_n se refiere a la coordenada y en el marcador con índice n y T se refiere al borde superior del rostro según SSD. En los casos en los que se observan dos opciones de coordenadas, la primera opción se utiliza cuando el rostro está rotado a la derecha y la segunda opción se utiliza cuando el rostro está rotado a la izquierda.

Tabla 5. Coordenadas que limitan las zonas de las mejillas, ceño y frente para características HOG.

Zona	Borde izquierdo	Borde derecho	Borde superior	Borde inferior
Mejilla	X_4 o X_{29}	X_{29} o X_{14}	Y_{29}	Y_4 o Y_{14}
Ceño	X_{21}	X_{24}	T	Y_{28}
Frente	X_{18} o X_{28}	X_{28} o X_{27}	T	Y_{20} o Y_{25}

Para una mejor visualización de estas zonas, en la Figura 15 se observa la ubicación de cada una para las expresiones faciales de la tristeza y del miedo, sobrepuestas en una imagen preprocesada por medio de CLAHE en el espacio de color cieLAB. Como se explicará en la subsección 0, las características HOG solo se tuvieron en cuenta para estas dos emociones. El recuadro amarillo indica la zona de la mejilla escogida, el recuadro magenta indica la zona del ceño y el recuadro rojo indica la zona de la mitad de la frente escogida. En la figura se puede ver con detalle el cambio de la textura de cada zona para estas expresiones faciales. Por un lado, en la tristeza, se observa una mayor pronunciación de la línea nasolabial, apertura de las fosas nasales y falta de líneas de expresión horizontales en la zona de la frente. De igual forma, se observa un ceño fruncido notable. Por otro lado, en el miedo, la línea nasolabial no se pronuncia de igual forma, se observan líneas de expresión horizontales en la frente y el ceño fruncido es menos notorio.

Figura 15. Representación visual de la ubicación de las zonas de mejilla, ceño y frente, enmarcadas en amarillo, magenta y rojo, respectivamente.

Por último, teniendo en cuenta que la ubicación de las zonas de mejilla, ceño y frente se hacen principalmente con marcadores faciales, se tomó en cuenta casos excepcionales en los cuales no fuera posible generar estos marcadores. Por ejemplo, cuando la persona mira hacia un lado, existe una pequeña probabilidad que el tabique nasal se encuentre más alejado del centro del rostro que el contorno. Esto causaría que el recuadro de la mejilla tenga área negativa, dado que el borde derecho estaría a la izquierda del borde izquierdo, causando errores en el algoritmo de reconocimiento. Por este motivo, cuando se dan estos casos excepcionales, se ignora el recuadro y el ciclo de reconocimiento vuelve a empezar, como está establecido en el diagrama de la Figura 13.

En la subsección 4.2.7 se detalla el uso que se le da a estas zonas para caracterizar los cambios en textura cuando se realizan distintas expresiones faciales.

4.2.7 Extracción de características de HOG

A partir de los recuadros de mejilla, ceño y frente obtenidos se calcula la magnitud y el ángulo del gradiente de estas zonas. El cálculo del gradiente ∇f se observa en la ecuación 6, donde g_x y g_y son el componente horizontal y vertical de la imagen, respectivamente.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{6}$$

A partir del gradiente, se obtiene la magnitud y la dirección de este, como se observa en las ecuaciones 7 y 8, respectivamente.

$$M = \sqrt{g_x^2 + g_y^2} \tag{7}$$

$$\theta = \tan^{-1}\left(\frac{g_y}{g_x}\right) \tag{8}$$

A partir de estas dos matrices, se obtienen varias características relacionadas con el gradiente de la imagen, como se observa en la Tabla 6. Se puede observar que la gran mayoría de características son iguales para las matrices de dirección y de magnitud; no obstante, no se calcula el índice del valor máximo del HOG de la magnitud. Esto es así porque todas las zonas evaluadas tienden a tener baja magnitud, porque la magnitud de los gradientes es nula cuando no se observan bordes en la imagen. Por este motivo, el compartimento con un mayor número de gradientes siempre es aquel de las magnitudes más bajas, lo que implica que la característica no brinda información sobre la imagen. No obstante, a partir del valor máximo del HOG de la magnitud, es posible saber indirectamente si una zona muestra más bordes, ya que el compartimento de las magnitudes bajas se disminuye cuando hay más bordes.

De la Tabla 6 se observa que se obtienen 13 características; sin embargo, teniendo en cuenta que se evalúa el gradiente en tres zonas distintas (mejilla, frente y ceño), se tiene un total de 39 características

basadas en HOG. Añadiendo las características obtenidas por medio de la ubicación de marcadores faciales, se obtienen un total de 62 características utilizadas para clasificar expresiones faciales. No obstante, como se mostrará en la subsección 4.2.8, este número de características es muy alto para la cantidad de imágenes de entrenamiento con las que se cuenta, causando sobreajuste. Por ese motivo, se hace una evaluación de las características más relevantes para clasificar expresiones faciales y se analiza la posibilidad de generar múltiples modelos de aprendizaje automático para facilitar la diferenciación entre expresiones faciales similares.

Tabla 6. Descripción de las características extraídas que están relacionadas con el histograma de gradientes orientados de la imagen.

Característica	**Descripción**
Media de la dirección del gradiente	Momentos estadísticos que caracterizan la matriz de direcciones del gradiente.
Desviación estándar de la dirección del gradiente	
Asimetría de la dirección del gradiente	
Curtosis de la dirección del gradiente	
Media de la magnitud del gradiente	Momentos estadísticos que caracterizan la matriz de magnitudes del gradiente.
Desviación estándar de la magnitud del gradiente	
Asimetría de la magnitud del gradiente	
Curtosis de la magnitud del gradiente	
Índice del valor máximo del HOG de la dirección	Tendencia direccional del gradiente, a partir de un histograma con ocho compartimentos.
Valor máximo del HOG de la dirección	Cantidad normalizada de elementos en el compartimento de mayor tendencia para la dirección del gradiente.
Valor máximo del HOG de la magnitud	Cantidad normalizada de elementos en el compartimento de mayor tendencia para la dirección del gradiente.
Entropía de la dirección del gradiente	Medida cuantitativa de la cantidad de información en la matriz de dirección del gradiente.
Entropía de la magnitud del gradiente	Medida cuantitativa de la cantidad de información en la matriz de magnitud del gradiente.

4.2.8 Predicción de expresiones faciales

La predicción de expresiones faciales se realizó por medio de algoritmos de aprendizaje automático, los cuales son capaces de aprender las características de una categoría de datos a partir de imágenes de prueba, para clasificar posteriormente más imágenes. En esta sección del desarrollo, se utilizó Spyder, un IDE de Python que facilita herramientas como la exploración de variables y la sencilla visualización de gráficas, de forma que se facilitara el análisis de aspectos como la obtención de matrices de confusión y la implementación de algoritmos que reflejan la relevancia de las características utilizadas. Como resultado, los modelos de aprendizaje automático fueron entrenados por medio de Python, aprovechando que los algoritmos de OpenCV generados a partir de Python son compatibles con C++. Así, la predicción en tiempo real es lograda a partir de la versión de C++ de OpenCV.

Para lograr esta predicción, es necesario contar con bases de datos adecuadas. En este caso, se busca que las bases de datos utilizadas cuenten con los siguientes criterios:

- Imágenes donde se observan rostros expresando distintas emociones.
- Como mínimo, las bases de datos deben contar con siete etiquetas: miedo, enojo, asco, alegría, tristeza, sorpresa y expresión neutral.
- Rostros con una resolución mínima de 128*128px.
- Una cantidad significativa de imágenes.

A partir de estos criterios se contó con cuatro bases de datos que cumplían parcial o totalmente los criterios establecidos. A continuación, se exponen las bases de datos cuyo uso se estudió:

- Karolinska Directed Emotional Faces (KDEF)

  KDEF es un conjunto de 4900 imágenes en las que se muestran expresiones humanas [81]. Estas imágenes fueron tomadas con la ayuda de 70 actores aficionados (35 hombres y 35 mujeres) entre 20 y 30 años, a quienes se les pidió no tener vello facial, aretes, gafas o maquillaje durante las sesiones de fotografía. La base de datos no especifica la etnia de los actores; sin embargo, por inspección visual, tienen características caucásicas. Cada actor imitó siete expresiones emocionales: miedo, enojo, asco, alegría, expresión neutral, tristeza y sorpresa, desde cinco ángulos distintos: perfil completamente a la derecha, perfil parcialmente a la derecha, de frente, perfil parcialmente a la izquierda y perfil completamente a la izquierda. A cada actor se le pidió practicar las expresiones faciales durante una hora antes de la sesión fotográfica y que, en lo posible, evocaran la emoción que estaban expresando. Todos los actores utilizaron el mismo una camiseta gris especial y la distancia a la cual fue tomada la foto fue la misma en todos los casos. Las imágenes resultantes tienen una resolución de 562*762px. Para este estudio, solo se tuvieron en cuenta las fotos tomadas de frente, dado que esta es la pose requerida de los participantes.

- The Japanese Female Facial Expression Dataset (JAFFE)

  JAFFE es una base de datos que consta de 213 imágenes en las que se muestran expresiones humanas [82]. Un aspecto particular de esta base de datos es que las participantes del estudio fueron 10 mujeres japonesas, quienes posaron las 7 expresiones básicas requeridas por nuestro equipo de investigación. Cada participante realizó expresiones faciales múltiples veces. Las imágenes resultantes están a grises, una resolución de 256*256px.

- The Child Affective Facial Expression Set (CAFE)

  CAFE es un conjunto de fotografías tomadas a 154 niños entre 2 y 8 años [83], quienes posaron las 7 expresiones deseadas. Un aspecto interesante de esta base de datos es que los modelos pertenecen a 5 etnias distintas (afroamericanos, asiáticos, caucásicos, latinos y del sudeste asiático). Se les pidió a los participantes posar cada expresión, a excepción de la sorpresa, con la boca abierta y cerrada, mientras que la sorpresa solo se posó con la boca abierta. Todos los participantes que no lograron imitar correctamente las siete expresiones fueron eliminados del estudio. Las 1192 imágenes resultantes tienen resoluciones variadas, aproximadamente entre 2000*2000px y 3000*3000px.

- FER-2013

  Aunque no existe mucha información sobre esta base de datos, se sabe que se trata de una base de datos con 28,709 imágenes etiquetadas con las 7 categorías de expresiones emocionales básicas. No obstante, la resolución de estas es de 48*48px [84].

Tres de las cuatro bases de datos fueron descartadas eventualmente, por no cumplir con los requisitos mínimos para una correcta ubicación de marcadores faciales. La primera base de datos descartada fue FER-2013. Aunque la detección de rostros funcionó correctamente, la ubicación de los marcadores faciales no fue precisa en la gran mayoría de imágenes. En la Figura 16 se observan dos ejemplos de imágenes a las cuales se les aplicó la ubicación de marcadores faciales; en la columna de la izquierda se ven las imágenes originales y en la columna de la derecha se ven las imágenes con marcadores faciales ubicados. En el primer caso, se puede ver que la ubicación es errónea, ya que la persona en la imagen está mirando a la derecha y los marcadores se ubican asumiendo que la persona está mirando hacia la izquierda. En el segundo caso, se observa que la ubicación de marcadores asume la pose correctamente, sin embargo, la ubicación de marcadores en la boca se hace de manera errónea. Adicionalmente, teniendo en cuenta la cercanía entre marcadores, dada la resolución de la imagen, no es posible identificar la expresión facial visualmente a partir de los marcadores faciales.

Figura 16. Muestras de la base de datos FER-2013, a las cuales se les intentó detectar el rostro. Las imágenes están amplificadas para mayor visibilidad.

Las otras bases de datos descartadas, JAFFE y CAFE, fueron removidas de los datos de entrenamiento por los resultados obtenidos al validar la exactitud de los algoritmos, como se mostrará a continuación. A partir de las tres bases de datos restantes, se entrenaron varios modelos, en los cuales se utilizaron todas las características expuestas anteriormente, con las etiquetas de las siete expresiones faciales. Para exponer las pruebas realizadas, se entrenaron redes neuronales artificiales (ANNs) con una capa oculta con 5 neuronas. De manera previa al entrenamiento, se normalizaron las características para que todas tuvieran la misma relevancia durante el entrenamiento. Esto se hizo obteniendo la media y la desviación de cada característica y modificándose por medio de la ecuación 9, donde C_N es la característica normalizada, C es la característica original, μ es la media de la característica y σ es la desviación estándar de esta.

$$C_N = \frac{C - \mu}{\sigma} \tag{9}$$

Así, analizaron las matrices de confusión generadas al entrenar las siguientes combinaciones de bases de datos:

- Solo KDEF
- Solo JAFFE
- Solo CAFE
- KDEF y JAFFE
- KDEF y CAFE
- JAFFE y CAFE
- Todas las bases de datos

Para realizar las pruebas que indicaron la relevancia de cada base de datos, las imágenes fueron divididas de manera aleatoria entre un conjunto de entrenamiento (70%) y un conjunto de prueba (30%), asegurando que hubiera una cantidad similar de imágenes entre etiquetas. Así, las ANNs utilizadas se entrenaron con las imágenes del conjunto de entrenamiento, cuyo número muestral varió dependiendo de las bases de datos utilizadas. Las matrices de confusión se generaron por medio del conjunto de prueba, de forma que dieron un estimado de la capacidad de generalización del algoritmo de predicción al utilizar las distintas combinaciones de bases de datos. Una matriz de confusión indica en las filas las etiquetas reales de una imagen y en las columnas las etiquetas predichas; por lo cual, una matriz de confusión con buenos

resultados es aquella cuya diagonal principal está altamente poblada, indicando que la clase real y la clase predicha son iguales en varias muestras.

Como se observa en la subsección 5.1.6, las combinaciones que mostraron exactitudes generales aceptables fueron: KDEF, JAFFE, KDEF+JAFFE y KDEF+JAFFE+CAFE. Sin embargo, KDEF+JAFFE y KDEF+JAFFE+CAFE mostraron exactitudes reducidas en el miedo, el enojo y la tristeza, por lo cual se descartaron. Por su parte, JAFFE tenía el principal problema de tener un bajo número de muestras, reduciendo su generalización. De igual forma, teniendo en cuenta que se trata de una base de datos de mujeres japonesas, era posible que no detectara con facilidad la expresión facial de niños latinoamericanos. Por estos motivos, se eligieron las imágenes de KDEF para las pruebas posteriores.

Una vez decididas las imágenes de muestra, se probaron distintas arquitecturas de algoritmos de aprendizaje automático y se incluyó en la evaluación la posibilidad de utilizar un algoritmo de aprendizaje profundo. Primero, se detallarán las pruebas realizadas con el algoritmo de aprendizaje profundo, ya que estas no tienen en cuenta los pasos previos de ubicación de marcadores faciales, histogramas de gradientes orientados y extracción de características. El entrenamiento de una arquitectura de aprendizaje profundo se hizo de manera previa al desarrollo de estos algoritmos y, al observar que los resultados no fueron satisfactorios, se decidió abarcar el acercamiento más tradicional, aprendizaje automático por medio de extracción de características.

Uno de los principales obstáculos del entrenamiento de algoritmos de aprendizaje profundo es la necesidad de altas especificaciones técnicas en el computador en el que se realiza. Como ejemplo, AlexNet, una popular red que clasifica imágenes en una de mil distintas categorías, fue entrenada durante seis días simultáneos por medio de dos tarjetas gráficas Nvidia Geforce GTX 580 [85]. Esto evita que se puedan realizar pruebas extensas de aprendizaje profundo por medio de modificaciones a la arquitectura o cambios en el conjunto de muestras de entrenamiento.

Para reducir los tiempos de cómputo, se utilizó la arquitectura de redes neuronales convolucionales (CNN, por su nombre en inglés) ResNet18, la cual se reentrenó para clasificar expresiones faciales [86]. El nombre ResNet18 es corto para Residual Network – 18 y se trata de una red convolucional inspirada en AlexNet que introduce un concepto novedoso, el uso de bloques residuales. Un bloque residual es un elemento dentro de una arquitectura de aprendizaje profundo que es capaz de ignorar capas convolucionales. Al añadir capas convolucionales a una CNN se disminuye drásticamente el gradiente, reduciendo la retropropagación del error. Esto, a su vez, reduce los cambios en los pesos de las neuronas en la CNN y evita la mejora del desempeño de esta. Así, un bloque residual tiene el principal propósito de evitar la supresión del gradiente, permitiendo mayores cambios en el desempeño de la red

Otra desventaja importante en el uso de aprendizaje profundo es la necesidad de contar con un conjunto de entrenamiento extenso, ya que requiere de muchas más muestras que las técnicas de aprendizaje automático tradicional. Al momento de desarrollar este elemento del proyecto, solo se contaba con acceso a la base de datos KDEF, por lo cual las imágenes de ésta conformaron el conjunto de entrenamiento. Además de necesitar un conjunto extenso, las técnicas de aprendizaje profundo son más efectivas cuando hay una alta variación en la pose e iluminación de este conjunto, dado que le permite una mayor generalización. Esta no es una característica con la que cuente KDEF, ya que todas las imágenes son tomadas con iluminación similar y con tan solo 5 variaciones en la pose de los actores. Sin embargo, para reducir la uniformidad de la base de datos, no se utilizó la totalidad de las imágenes como conjunto de entrenamiento, se ingresó únicamente el rostro de los participantes, el cual fue detectado por medio de SSD.

Para entrenar la arquitectura seleccionada, se hizo uso del sistema embebido Jetson TX1, el cual fue brindado por tiempo limitado por la Escuela Colombiana de Ingeniería Julio Garavito. Este sistema cuenta con 256 núcleos de GPU Nvidia CUDA, de la familia de GPUs Nvidia Maxwell y 4GB de RAM [87]. Las capacidades técnicas de la Jetson TX1 la hacen ideal para el entrenamiento de algoritmos de aprendizaje profundo. En la subsección 5.1.6 se observan los resultados obtenidos al reentrenar ResNet para cumplir con los objetivos del proyecto. Sin embargo, se encontró que este método no era viable para el reconocimiento de expresiones faciales con los recursos con los que se contaban.

Teniendo en cuenta que se pueden realizar múltiples entrenamientos de algoritmos de aprendizaje automático tradicional con menos recursos que en el caso de algoritmos de aprendizaje profundo, se estudió la posibilidad de separar las categorías deseadas en conjuntos, de forma que los algoritmos fueran capaces de discriminar características con mayor facilidad. A partir del promedio de exactitudes y de la exactitud máxima por emoción de la Tabla 17 se puede observar que las emociones con mayor dificultad para detectar por las ANN entrenadas fueron miedo y tristeza. Por este motivo, se decidió que se crearían dos modelos de aprendizaje automático, a partir de dos conjuntos de emociones:

- Conjunto 1: Alegría, asco, enojo, sorpresa y expresión neutra
- Conjunto 2: Miedo, tristeza y expresión neutra

Dado que en este caso se obtendrían dos resultados (uno para la predicción del conjunto 1 y otro para la predicción del conjunto 2), se decidió hacer uso del puntaje que ofrecen los algoritmos de aprendizaje automático, los cuales indican el nivel de certeza que una predicción sea correcta. Así, se comparó cual puntaje es más alto y, a partir de esto, se seleccionó la emoción predicha final, como se observa en el diagrama de la Figura 13. Dado que la expresión neutra se encuentra en ambos conjuntos, una vez se selecciona la expresión facial, se considera que la categoría neutra final equivale a cualquiera de las etiquetas de expresión neutra.

De esta forma, se analizó cuantitativamente si la separación por conjuntos resultaba ser más efectiva que el entrenamiento de un solo modelo de aprendizaje automático. Para cada caso, se estudió el uso de varios modelos de aprendizaje automático y, posteriormente, se observó la relevancia de cada característica en la predicción. Los algoritmos de aprendizaje automático tradicional que se analizaron son los más utilizados en investigaciones, teniendo en cuenta aquellos cuya implementación fuera sencilla, con la ayuda de OpenCV. Cada uno de estos algoritmos cuenta con parámetros de entrenamiento. Inicialmente, para observar la efectividad de separar las categorías de expresiones faciales en dos grupos, se utilizaron los parámetros por defecto que brinda OpenCV para cada algoritmo de aprendizaje automático. Los algoritmos utilizados, con sus respectivos parámetros de entrenamiento, se observan en la Tabla 7.

Tabla 7. Parámetros utilizados para evaluar efectividad de separar las expresiones faciales en dos conjuntos.

Técnica	**Parámetro**	**Valor**
K-Nearest Neighbors (k-NN)	*K* vecinos más cercanos	3
Support Vector Machine (SVM)	Tipo de kernel	lineal
	Iteraciones máximas	100
	Error mínimo entre iteraciones	10^{-6}
Artificial Neural Network (ANN)	Número de capas ocultas	1
	Número de neuronas	5
	Función de activación	Sigmoidal simétrica
	Método de entrenamiento	Retropropagación
Random Forest (RF)	Profundidad máxima de los árboles	15
	Mínimo de muestras por nodo	3
	Número de árboles	100

Al igual que en los casos anteriores, se fijó una semilla para que todos los algoritmos de la misma familia fueran entrenados con los mismos parámetros. Así, la exactitud de cada uno depende únicamente de su efectividad y no de fenómenos aleatorios propios de cada algoritmo. Teniendo en cuenta que se evalúan cuatro algoritmos y para cada uno se entrenan tres modelos (uno en el que no se separan los conjuntos y dos para cuando estos se separan), se entrenaron un total de 12 modelos, cuya exactitud se muestra en la subsección 5.1.6. Se encontró que la exactitud de ANN y de RF fueron notoriamente mejores que aquellas de otras técnicas, por lo cual se profundizó en su uso, modificando parámetros de entrenamiento en ambas. En la Tabla 8 se observan las modificaciones realizadas a los parámetros de cada algoritmo. Los valores escogidos son cercanos a los valores por defecto para evitar resultados deficientes y tiempos de cómputo excesivos. Se puede notar que no se consideró el uso de varias capas ocultas en el uso de

ANN. La razón principal de esto es porque las redes neuronales con una sola capa oculta, que cuenten con funciones de activación no lineales, pueden ser aproximadores universales a cualquier función que se busque. Por esta razón, añadir más capas ocultas únicamente brinda más parámetros en la clasificación, lo que significa que se requieren más datos de entrada para evitar el sobreajuste. Así, aunque en un principio se crea que añadir más capas puede ser beneficioso para el modelo entrenado, la realidad es que suele ser un procedimiento perjudicial [88].

Tabla 8. Parámetros modificados a ANN y RF para la selección del mejor modelo.

Técnica	**Parámetro**	**Valor**
Artificial Neural Network (ANN)	Número de neuronas en la capa oculta	{2, 3, 5, 10, 20, 50, 100}
Random Forest (RF)	Profundidad máxima de los árboles	{10, 15, 20}
	Mínimo de muestras por nodo	{2, 3}
	Número de árboles	{50, 100, 1000}

A partir de estos cambios, se encontró que para el conjunto 1, el mejor modelo fue el de ANN con 5 neuronas en la capa oculta. De igual forma, se encontró que el mejor modelo para el conjunto 2 fue el de RF con 15 niveles de profundidad de los árboles, un mínimo de 2 muestras por nodo y 50 árboles. No obstante, varios modelos de esta arquitectura mostraron exactitudes similares. Estos modelos se escogieron tanto por la exactitud como por el tiempo de cómputo.

El último paso que se realizó en el desarrollo del algoritmo de reconocimiento de expresiones faciales es mejorar los tiempos de cómputo, a partir de la eliminación de características que no son relevantes para la predicción. Aunque existen varios métodos para lograr esto, se decide aprovechar una de las ventajas de los algoritmos de RF, la Importancia de Gini. La importancia de Gini es una medida que indica cuanta importancia tiene una característica. Esto lo hace analizando aquellas variables que, al ser utilizadas para crear un nodo en cada uno de los árboles, lograron obtener predicciones correctas. De esta forma, es posible saber cuáles variables dentro de las generadas son las más relevantes para la correcta predicción de expresiones faciales.

En la Figura 17 se observa la importancia de Gini de cada característica ordenada para el conjunto 1. El índice de la característica es el dado en este documento: Primero se extraen las características de la Tabla 4 en el orden expuesto y luego se extraen las características de la Tabla 6, en el orden expuesto, iniciando en la zona de la mejilla, luego el ceño y luego la frente. En la Figura 17 se observa que después de la 18va característica, la importancia de Gini no cambia de manera significativa. Las 18 características con mayor importancia de Gini para el conjunto 1 se pueden observar en la Tabla 9, ordenadas de manera descendente. Es interesante notar que ninguna de las características extraídas a partir de HOG tuvieron relevancia para el conjunto 1. Por su parte, 78.26% de las características relacionadas a los marcadores faciales hacen parte de las características más relevantes.

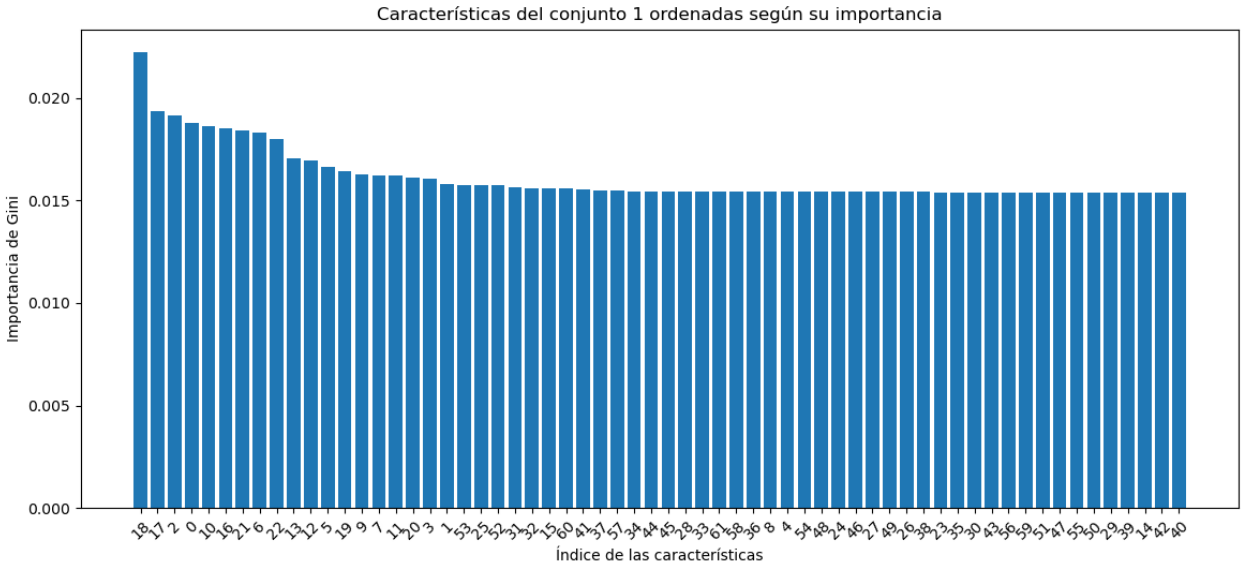

Figura 17. Importancia de Gini de cada característica en el entrenamiento del modelo de RF para las expresiones faciales del conjunto 1.

Tabla 9. Características más relevantes en la predicción de expresiones faciales para el conjunto 1, ordenadas de manera descendiente.

Índice	**Nombre**
18	Ancho de la boca
17	Apertura de la parte interna de la boca
2	Apertura del ojo derecho
0	Apertura del ojo izquierdo
10	Cercanía entre ceja derecha y tabique nasal
16	Apertura de la boca
21	Ángulo de curvatura del labio superior
6	Cercanía entre ceja izquierda y tabique nasal
22	Ángulo de curvatura del labio inferior
13	Cercanía entre las cejas
12	Apertura de fosas nasales
5	Radio de curvatura de la ceja izquierda
19	Grosor de labio superior
9	Radio de curvatura de la ceja derecha
7	Altura de la ceja izquierda
11	Altura de la ceja derecha
20	Grosor del labio inferior
3	Ancho del ojo derecho

Se decidió entrenar el modelo de ANN correspondiente únicamente con las características expuestas en la tabla y se comparó su efectividad y tiempo de cómputo para la extracción respecto al entrenamiento con la totalidad de las características. Es importante tener en cuenta que, en este caso, el tiempo de cómputo no se reduce únicamente en la extracción de características, sino en el cálculo del gradiente orientado, dado que ninguna característica utiliza HOG. Los valores de efectividad y tiempo de cómputo se pueden observar en la subsección 5.1.6. Con estos resultados, se considera que la reducción de características para el conjunto 1 genera resultados positivos.

A continuación, se detalla un proceso similar al anterior, aplicado al conjunto 2. En la Figura 18 se observa la importancia de Gini de cada característica ordenada para el conjunto 2. El índice de la característica es el mismo indicado en el caso del conjunto 1. Se observa que, después de la 15va característica, la importancia de Gini no cambia de manera significativa. Las 15 características con mayor importancia de Gini para el conjunto 2 se pueden observar en la Tabla 10, ordenadas de manera descendente. En total, se incluyeron 10 (39.1%) de las características relacionadas a marcadores faciales y 6 (15.4%) de las características relacionadas con gradientes orientados.

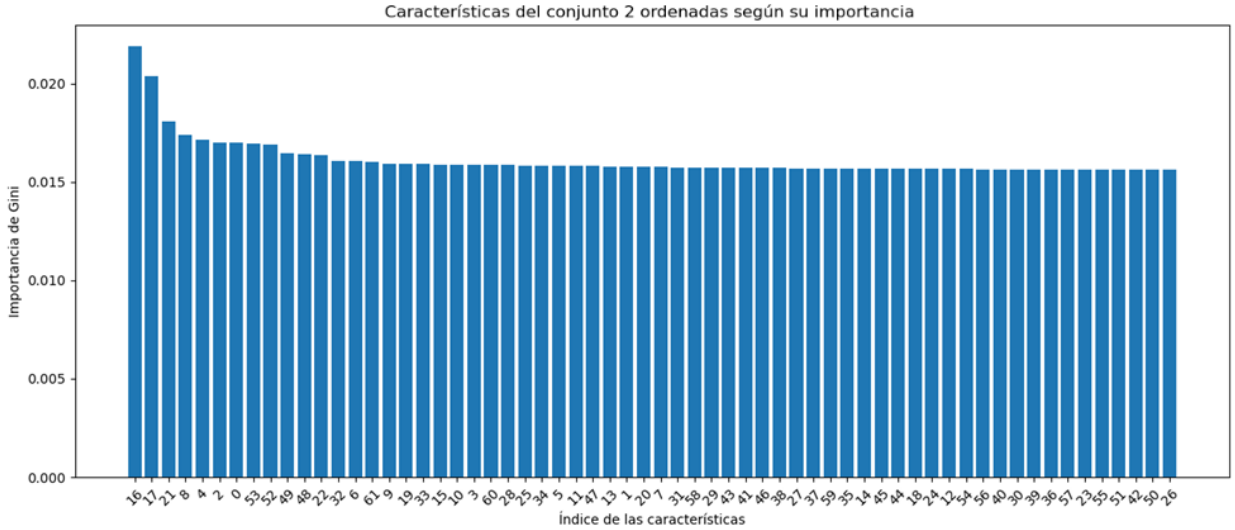

Figura 18. Importancia de Gini de cada característica en el entrenamiento del modelo de RF para las expresiones faciales del conjunto 2.

Tabla 10. Características más relevantes en la predicción de expresiones faciales para el conjunto 2, ordenadas de manera descendiente.

Índice	**Nombre**
16	Apertura de la boca
17	Apertura de la parte interna de la boca
21	Ángulo de curvatura del labio inferior
8	Rotación interna de la ceja derecha
4	Rotación interna de la ceja izquierda
2	Apertura del ojo derecho
0	Apertura del ojo izquierdo
53	Curtosis de la magnitud del gradiente del ceño
52	Asimetría de la magnitud del gradiente del ceño
49	Curtosis de la magnitud del gradiente de la frente
48	Asimetría de la magnitud del gradiente de la frente
22	Ángulo de curvatura del labio superior
32	Desviación estándar de la magnitud del gradiente del ceño
6	Cercanía entre ceja izquierda y tabique nasal
61	Entropía de la dirección del gradiente del ceño

Se decidió entrenar el modelo de RF correspondiente únicamente con las características expuestas en la tabla y se comparó su efectividad y tiempo de cómputo para la extracción respecto al entrenamiento con la totalidad de las características. Estos valores se pueden observar en la subsección 5.1.6.

Por medio de esta reducción de características, se finaliza el proceso propuesto en el diagrama de la Figura 13. Finalmente, en la subsección 5.1.6 se detallan los resultados finales obtenidos por el algoritmo de reconocimiento de expresiones faciales, incluyendo exactitud y tiempo de cómputo de todo el proceso.

4.2.9 Preparaciones finales del algoritmo de reconocimiento

En la subsección 5.1.6 se observa la exactitud final del algoritmo de reconocimiento y el tiempo de cómputo para realizar una predicción, lo que incluye todas las etapas desde el preprocesamiento hasta la predicción final. De ser necesario, para permitir que los minijuegos explicados en las subsecciones 4.1 y 4.3 corrieran de manera fluida, se implementó un algoritmo que permite fijar un retardo, en segundos, para el algoritmo de reconocimiento de expresiones faciales. Así, si el retardo escogido es de 0s, el algoritmo buscará reconocer expresiones faciales siempre que pueda, mientras que cualquier otro tiempo *t* fijado indicará que el algoritmo esperará *t* segundos antes de reconocer nuevamente expresiones faciales, permitiendo la renderización de otros aspectos del juego durante este tiempo. Por defecto, el algoritmo de reconocimiento está fijado en 0.1 segundos de retardo en cada minijuego. En cuanto al correcto funcionamiento del algoritmo de reconocimiento, se recopilaron y generaron archivos necesarios a lo largo de las distintas etapas del algoritmo. En la Tabla 11 se observa un resumen de estos archivos.

Tabla 11. Resumen de los archivos necesarios para el correcto funcionamiento del algoritmo de reconocimiento de expresiones faciales.

Archivo	**Descripción**
opencv_core430.dll	Biblioteca base de OpenCV 4.3.0. Brinda tipos de variables para el tratamiento de imágenes y permite el funcionamiento de otras bibliotecas de OpenCV.
opencv_imgproc430.dll	Biblioteca que permite procesamiento básico de imágenes en OpenCV 4.3.0. Incluye funciones como la conversión entre espacios de color y la ecualización de histograma.
opencv_dnn430.dll	Biblioteca que implementa la detección de rostros a partir de aprendizaje profundo en OpenCV 4.3.0. En esta biblioteca, se encuentran las funciones necesarias para utilizar el algoritmo SSD.
opencv_ml430.dll	Biblioteca que recopila algoritmos de aprendizaje automático en OpenCV 4.3.0. Esto incluye arquitecturas como ANN, RF, SVM, K-NN, entre otras.
deploy.prototxt	Uno de los archivos necesarios para el funcionamiento de SSD en OpenCV. Da información sobre la arquitectura de la CNN utilizada para la detección, como el número de capas convolucionales y las neuronas encontradas en cada una de ellas.
res10_300x300_ssd_iter_140000_fp16.caffemodel	Uno de los archivos necesarios para el funcionamiento de SSD en OpenCV. Indica los pesos dados a cada nodo en la CNN. El nombre indica que la red fue desarrollada a partir de ResNet10, requiere una entrada de 300*300px y utilizó 140,000 iteraciones para su entrenamiento.
sp_human_face_68.dat	Archivo con todos los datos de la arquitectura utilizada por Kazemi en su algoritmo de ubicación de marcadores faciales.
Conj1ModelKdef.xml	Modelo generado personalmente para la predicción de expresiones faciales del conjunto 1. Entrenador a partir de la base de datos KDEF.
Conj2ModelKdef.xml	Modelo generado personalmente para la predicción de expresiones faciales del conjunto 2. Entrenador a partir de la base de datos KDEF.
normConj1Kdef.csv	Archivo generado que brinda información para normalizar las características utilizadas en el modelo de predicción del conjunto 1. Contiene la media y la desviación estándar de cada característica para la aplicación de la ecuación 9.
normConj2Kdef.csv	Archivo generado que brinda información para normalizar las características utilizadas en el modelo de predicción del conjunto 2. Contiene la media y la desviación estándar de cada característica para la aplicación de la ecuación 9.
EmotionRecognition.dll	Biblioteca generada personalmente para permitir la comunicación entre C++ y C#. Incluye funciones que cargan archivos mencionados anteriormente en esta tabla, preprocesan imágenes provenientes de C# y predicen expresiones faciales a partir de marcadores faciales y HOG.

Todos los DLLs se ubicaron en una carpeta oculta al usuario a la cual puede acceder Unity para leer los archivos. Había dos opciones para almacenar los demás archivos: la primera ubicación es un directorio persistente, el cual permite almacenar archivos ocultos al usuario. Sin embargo, se decidió no utilizar esta opción, dado que no se generó un instalador para el archivo ejecutable y, una vez el usuario desee eliminar el juego de su computador, estos archivos seguirán almacenados en este. Esto es considerado una mala práctica en programación, dado que estos archivos no tendrían ningún uso, pero ocupan espacio en el computador, sin que un usuario no experimentado sepa cómo eliminarlos.

Por este motivo, se decidió utilizar una segunda ubicación para los archivos: en la misma dirección en la que se encuentra el ejecutable del juego. La principal desventaja de esto es que un usuario podría eliminar accidentalmente un archivo necesario para el reconocimiento de expresiones faciales; sin embargo, al eliminar la carpeta que contiene el ejecutable, el usuario habrá eliminado todo rastro del juego de su computador. En el archivo EmotionRecognition.dll hay una sección en la inicialización de variables para búsqueda y carga de los archivos de la Tabla 11; si no se encuentra un archivo en particular, se le envía una alerta al usuario indicándole que el reconocimiento de expresiones faciales no es posible, pidiéndole que contacte a los desarrolladores para arreglar este problema.

4.3 Desarrollo de la herramienta interactiva

4.3.1 Aspectos generales

La herramienta interactiva diseñada, Emmaciones, recompila el trabajo realizado en el diseño del protocolo experimental para la estimulación de imitación y reconocimiento de emociones y el algoritmo de reconocimiento de expresiones faciales, por medio del desarrollo de un videojuego donde los participantes aprenden claves de comunicación de manera didáctica. Aunque varias de las actividades planteadas en la herramienta interactiva ya fueron explicadas en la subsección 4.1, esta subsección se encargará de indicar detalles técnicos para cada una de las actividades desarrolladas.

Un aspecto importante en el desarrollo de Emmaciones es el uso de variables globales, las cuales se comparten a lo largo del protocolo. Principalmente, se registra de manera global la etapa (imitación o reconocimiento), la sesión y la actividad, dado que estas tres variables modifican la manera con la que se cargan distintas escenas dentro del juego desarrollado. Esto es particularmente cierto para escenas que se utilizan para distintas actividades; aunque en el protocolo se realizan 28 actividades a lo largo de 12 sesiones, únicamente 17 escenas fueron desarrolladas, las cuales pueden o no ser utilizadas múltiples veces en Emmaciones, dadas similitudes entre actividades, permitiendo un menor tiempo de desarrollo.

Las escenas creadas se pueden observar en la Tabla 12; cada una tiene un nombre código en inglés y en la tabla se expone los momentos en donde es utilizada y detallando variaciones. Esta tabla se utilizará como referencia a lo largo de la subsección 4.3.2 para describir detalles técnicos propios de cada una al implementarlas en Unity. Finalmente, un último aspecto general de todas las escenas de Emmaciones es que fueron creadas para ajustarse a una resolución de 1280x720px, la cual es suficientemente pequeña para ajustarse a la gran mayoría de computadores, pero suficientemente grande para que los estímulos visuales de interés sean notorios.

Tabla 12. Resumen de las escenas creadas y su uso.

Etapa	**Nombre**	**Sesión**	**Actividad**	**Descripción**
Ambas	Main	[1,6]	1	Menú principal, donde se puede escoger la etapa y la sesión deseada. Cada sesión inicia en la actividad 1. Emma se presenta al seleccionar la primera sesión de imitación, antes de pasar a las demás actividades.
Imitación	FaceParts	1	2	Presentación del minijuego *Partes del Rostro*.
	FacePuzzle	1	3	Presentación del minijuego *Rompecabezas*.

	EmotionImages	2	1	Presentación del minijuego *Presentación de Imágenes*, para las expresiones faciales de alegría, miedo y asco.
		3	1	Presentación del minijuego *Presentación de Imágenes*, para las expresiones faciales de tristeza, sorpresa y enojo.
		4	1	Presentación del minijuego *Sonidos de las emociones*, para las emociones alegría, miedo y asco.
		5	1	Presentación del minijuego *Sonidos de las emociones*, para las emociones tristeza, sorpresa y enojo.
	EmotionWheel	2	2	Presentación del minijuego *Ruleta*, para las expresiones faciales de alegría, miedo y asco.
		3	2	Presentación del minijuego *Ruleta*, para las expresiones faciales de tristeza, sorpresa y enojo.
	SurpriseBox	4	2	Presentación del minijuego *Caja Sorpresa*, para las expresiones faciales de alegría, miedo y asco.
		5	2	Presentación del minijuego *Caja Sorpresa*, para las expresiones faciales de tristeza, sorpresa y enojo.
	EmmaSays	6	1	Presentación del minijuego *Emma Dice.*
	Lottery	6	2	Presentación del minijuego *Lotería.*
Reconocimiento	SortFace	1	1	Presentación del minijuego *Ordenar la Cara.*
	Pop	1	2	Presentación del minijuego *Pop Emma*.
	IdentifyEmotion	2	1	Presentación del minijuego *Identifica la Emoción*, para las emociones alegría, miedo y asco.
		3	1	Presentación del minijuego *Identifica la Emoción*, para las emociones tristeza, sorpresa y enojo.
		4	1	Presentación del minijuego *¿Qué Emoción Soy?*, para las emociones alegría, miedo y asco.
		5	1	Presentación del minijuego *¿Qué Emoción Soy?*, para las emociones tristeza, sorpresa y enojo.
	Pairs	2	2	Presentación del minijuego *Encuentra el Par*, para las expresiones faciales de alegría, miedo y asco.
		3	2	Presentación del minijuego *Encuentra el Par*, para las expresiones faciales de tristeza, sorpresa y enojo.
	LivePhoto	3	3	Presentación del minijuego *Foto en Vivo*.
	FindEmotions	4	2	Presentación del minijuego *Laberinto*, para las expresiones faciales de alegría, miedo y asco.
		5	2	Presentación del minijuego *Laberinto*, para las expresiones faciales de tristeza, sorpresa y enojo.

	DeleteEmotion	4	3	Presentación del minijuego *Elimina la Emoción*, para las expresiones faciales de alegría, miedo y asco.
		5	3	Presentación del minijuego *Elimina la Emoción*, para las expresiones faciales de tristeza, sorpresa y enojo.
	SlideEmotion	6	1	Presentación del minijuego *Desliza la Emoción.*
	Mirror	6	2	Presentación del minijuego *Espejo*.

4.3.2 Escenas desarrolladas

A. Main

Esta escena sirve como menú principal del juego, desde el cual se puede entrar a cualquier sesión. De igual forma, en esta escena los participantes conocen por primera vez a Emma, quien se presentará antes de mostrar las actividades si el participante escoge la primera sesión de imitación. Dado que se espera que los participantes logren completar por lo menos una sesión por día, no se les da la opción de entrar a cualquier actividad, dado que el orden está establecido para que el aprendizaje sea progresivo. En la Figura 7 se observa la escena *Main*, diseñada de forma que no hubiera una gran cantidad de estímulos visuales y el participante se pudiera concentrar en Emma y lo que está diciendo.

B. *FaceParts*

Esta escena está desarrollada para brindar una introducción a las partes del rostro, las cuales toman relevancia en escenas posteriores, donde los participantes deben observarlas e imitarlas para jugar los distintos juegos. En este juego, cuya imagen se observa en la Figura 19, el participante puede seleccionar con el cursor distintas partes del rostro de Emma, quien le indica qué parte del rostro se está seleccionando. Este efecto se logró a partir de la separación del rostro en varias imágenes, correspondientes a cada parte del rostro mencionada por Emma. Dado que cada parte es un objeto individual dentro de la escena, la colisión del cursor con cada una se pudo lograr de manera individual, lo que permite a su vez una interacción única; en este caso, una disminución en la transparencia de la imagen y un texto relacionado a la imagen seleccionada. Internamente, un contador registra aquellas partes del rostro que ya fueron activadas; el juego no continúa hasta que el participante haya seleccionado cada parte.

Figura 19. Imagen correspondiente a la escena *FaceParts*.

C. FacePuzzle

FacePuzzle implementa el juego *Rompecabezas*, en el cual se aumenta el conocimiento obtenido en la actividad anterior al pedirle al participante que construya un rompecabezas con la cara de Emma, para reforzar el conocimiento sobre las partes del rostro. En este caso, cada parte, incluyendo cuatro secciones de la cara, es una imagen aparte, la cual se puede arrastrar a lo largo de la escena. Un ejemplo de este juego se observa en la Figura 20. Si el participante remueve una pieza de la escena, ésta vuelve automáticamente, de forma que siempre sea posible completar el juego. Cada vez que se abre el juego, las piezas están ubicadas en distintos lugares, de forma que los participantes no puedan memorizar su ubicación y deban comprender el concepto de ubicación de las partes del rostro. Una vez el centroide de la pieza se encuentre a menos de una unidad (unidad de Unity, nombrada de ahora en adelante con la sigla U) de la ubicación que debería tener, la pieza se ubica exactamente donde debería estar y se inhabilita, de forma que el jugador no la pueda desplazar una vez ya se encuentre en posición. Cada vez que una pieza se ubica de manera correcta, Emma brinda realimentación positiva auditiva al participante. Cuando todas las piezas se encuentran en la posición adecuada, Emma felicita al participante y automáticamente se devuelve al menú principal, donde se encuentra seleccionada por defecto la sesión 2 de imitación.

D. *EmotionImages*

Esta es la primera escena utilizada en el protocolo experimental que se utiliza en múltiples actividades. En particular, se utiliza en los juegos *Presentación de Imágenes* y *Sonidos de las Emociones*, en dos sesiones aparte para cada uno de estos, dando un total de 4 actividades en las que se utiliza esta escena. A nivel técnico, se trata de una escena muy sencilla en la que, a la vez que Emma explica las actividades a realizar, internamente se cargan ciertas imágenes o sonidos por medio de una corrutina, de forma que el discurso y la animación de Emma se vean fluidos. Esta escena hace uso de las variables globales, dado que: si se selecciona la sesión 2 de imitación, únicamente se cargan las imágenes correspondientes a alegría, miedo y asco, si se selecciona la sesión 3 de imitación, únicamente se cargan las imágenes correspondientes a tristeza, sorpresa y enojo, si se selecciona la sesión 4 de imitación, únicamente se cargan los sonidos correspondientes a alegría, miedo y asco y si se selecciona la sesión 5 de imitación, únicamente se cargan los sonidos correspondientes a tristeza, sorpresa y enojo. Al finalizar la presentación del juego, se visualizan 6 botones, correspondientes a cada una de las emociones; aquellos botones que correspondan a emociones que no se trabajan en la actividad se encuentran deshabilitados.

Figura 20. Imagen correspondiente a la escena *FacePuzzle.*

Si el juego seleccionado es *Presentación de Imágenes*, una imagen aleatoria correspondiente a la emoción escogida se muestra en la pantalla. La primera vez que se selecciona una imagen para una emoción,

Emma realiza una breve explicación en la que describe las características físicas del rostro observado. Una vez se presentan cinco imágenes de cada emoción correspondiente a la sesión escogida, se activa un botón que permite que el jugador avance al siguiente juego. Si el juego seleccionado es *Sonido de las Emociones*, el rostro de Emma de la Figura 11 correspondiente a la emoción escogida se muestra en la pantalla y se activa un segmento de audio que indica cual es el sonido correspondiente a la emoción escogida. Una vez se escuchan todos los sonidos, se activa un botón que permite que el jugador avance al siguiente juego. En la Figura 21 se observa un ejemplo de esta escena, en la cual se muestra una imagen correspondiente a sorpresa en el juego *Presentación de Imágenes*.

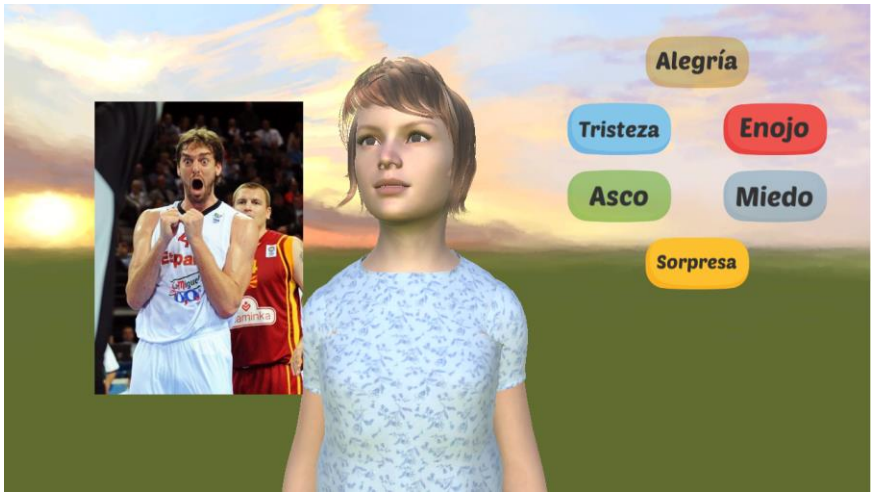

Figura 21. Imagen correspondiente a la escena *EmotionImages.*

E. EmotionWheel

Esta escena corresponde al juego *Ruleta*, el cual se usa dos veces en el protocolo: una vez para las emociones alegría, miedo y asco y otra para las emociones tristeza, sorpresa y enojo. Esta es la primera escena en la que se utiliza el algoritmo de reconocimiento de expresiones faciales, dado que los participantes deben imitar la emoción seleccionada en la rueda. En todos los juegos en los que se utiliza la cámara del participante, esta se activa al inicio de la escena, así el participante no pueda observar su imagen. Esto permite activar y desactivar rápidamente la imagen de la cámara, sin necesidad de apagar esta. Un aspecto interesante de esta escena es que, aunque el participante solo sabe cuál emoción debe imitar después de que la rueda para, el número aleatorio se genera antes de esto, ya que se calcula el movimiento de la rueda de manera correspondiente a la emoción obtenida. Así, es posible controlar la emoción que se visualiza en la rueda. Esto se hace principalmente para que todas las emociones salgan de manera uniforme y no sea necesario esperar a que la rueda pare en la emoción que uno desee. Si una emoción ya ha sido seleccionada en la rueda, esta no vuelve a salir en el conjunto de posibles opciones, lo que obliga a que cada emoción salga exactamente una vez.

Una vez la rueda para, se activa la cámara del participante y este debe imitar la emoción correspondiente. Se le brinda realimentación visual por medio de un molinillo: Si el molinillo se mueve rápidamente, está imitando correctamente la emoción; si se mueve lentamente, la está imitando incorrectamente. Emma felicita al participante una vez el molinillo llegue a cierta velocidad y se le pide al participante volver a girar la rueda. El juego termina una vez el participante haya imitado todas las expresiones faciales correspondientes a la sesión. En la Figura 22 se observa un ejemplo de esta escena, en el cual la rueda puede parar en una de seis casillas, obteniendo uno de tres posibles resultados (en este caso, tristeza, enojo o sorpresa).

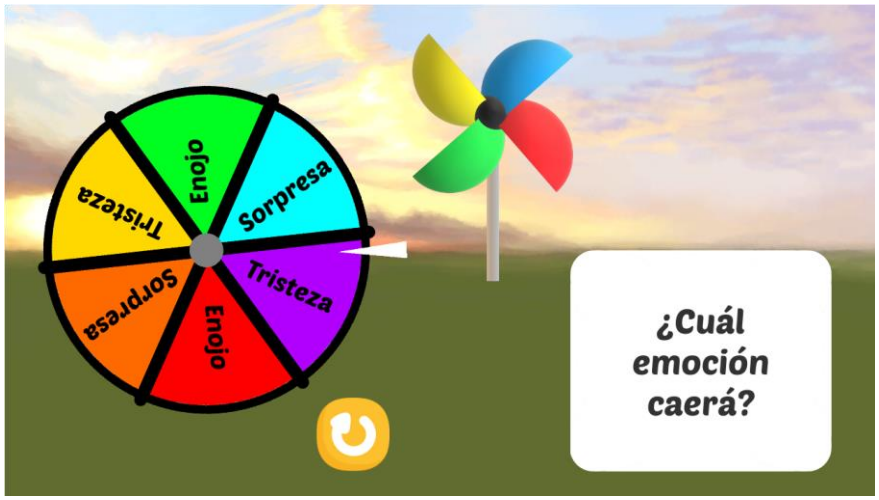

Figura 22. Imagen correspondiente a la escena *EmotionWheel*.

F. SurpriseBox

Esta escena corresponde al juego *Cada Sorpresa* y se trata de otro juego en el cual se utiliza el algoritmo de reconocimiento de emociones faciales. Este juego, el cual se utiliza en dos sesiones separadas para tres emociones distintas, hace uso de imágenes seleccionadas de manera aleatoria para que el participante reconozca la expresión facial mostrada en ellas y la imite. El juego inicia con una caja animada, la cual debe ser abierta por el participante para encontrar una imagen. Se le muestra la imagen durante cinco segundos, pidiéndole que observe muy bien las expresiones faciales. Posteriormente, la imagen se vuelve a guardar en la caja, se activa la cámara del participante y se le pide que imite la emoción que se expresa en la imagen. Se decidió que no se pediría imitar la emoción mientras se mostraba la imagen, porque durante las pruebas piloto se observó que los participantes intentaban imitar la pose exacta de la persona en la imagen y no la expresión facial que estaba realizando. Se le brinda realimentación al participante por medio de una barra que se va llenando cada vez que el sistema reconoce que está imitando correctamente la emoción seleccionada. Una vez se llena la barra, se solicita volver a abrir la caja. Este proceso continúa hasta que el participante haya logrado imitar las expresiones faciales de todas las emociones correspondientes a la sesión activa. Un ejemplo de esta actividad se observa en la Figura 23, donde se está imitando el asco después de observar una imagen correspondiente a esta emoción. Una vez se llena la barra, se le da realimentación visual y auditiva al participante.

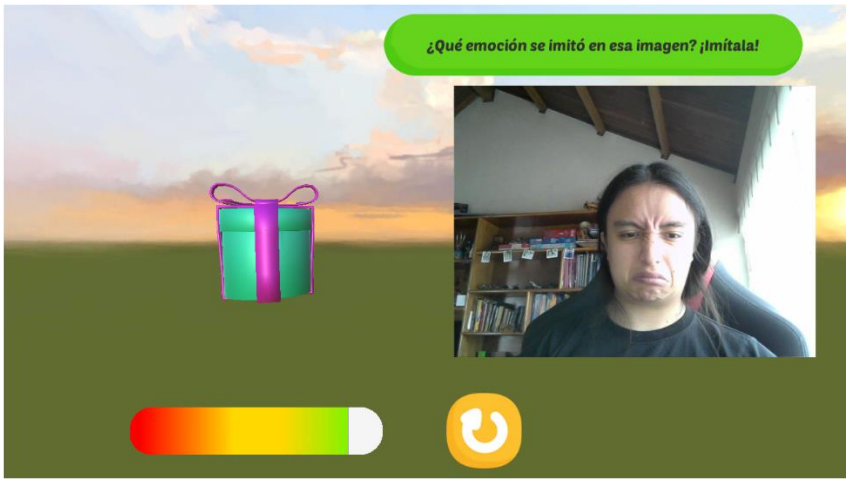

Figura 23. Imagen correspondiente a la escena *SurpriseBox*.

G. *EmmaSays*

Esta escena, correspondiente al juego *Emma Dice*, sirve para reforzar el conocimiento obtenido a lo largo de la etapa de imitación, ya que requiere de los participantes imitar todas las expresiones faciales y sonidos que se aprendieron en sesiones anteriores. A nivel técnico, esta escena es bastante sencilla, porque solo se trata de 12 frases que expresa Emma de manera aleatoria: 6 de ellas donde pide imitar los sonidos de las emociones y 6 donde pide imitar las expresiones faciales de las emociones. Se decidió que esto se haría de manera aleatoria para evitar que los participantes deben memorizar algún orden en particular para imitar las emociones. Una vez un comando es dicho por Emma, este se elimina del conjunto de posibles comandos, de forma que solo sea posible que Emma diga comandos que no ha dicho anteriormente. Para la imitación de sonidos, se implementó un botón que se pide presionar a los participantes una vez los investigadores indican que ha reconocido el sonido correctamente, dado que el desarrollo de Emmaciones no contó con un algoritmo de reconocimiento de sonidos. El juego termina una vez se haya logrado imitar correctamente todos los sonidos y todas las expresiones faciales. En la Figura 24 se observa un ejemplo de esta escena, donde se observa que el mismo tipo de realimentación que en *EmotionWheel*.

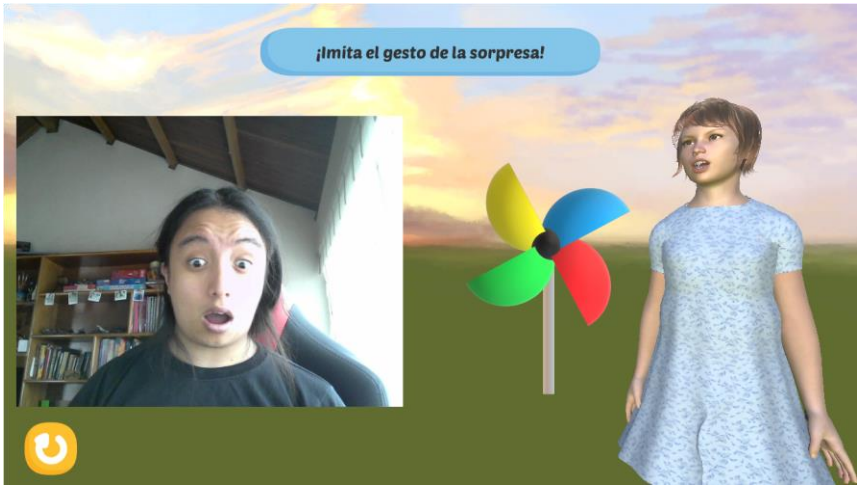

Figura 24. Imagen correspondiente a la escena *EmmaSays*.

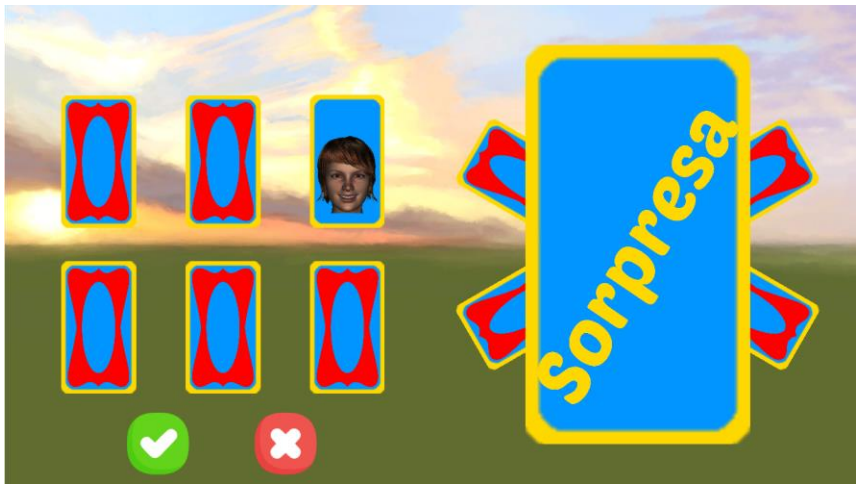

Figura 25. Imagen correspondiente a la escena *Lottery*.

H. *Lottery*

Esta escena, correspondiente al juego *Lotería*, busca que los participantes puedan relacionar el nombre de una emoción con su expresión facial. Se observan una serie de cartas en la derecha de la pantalla y se

le pide al participante presionar un botón para que seleccione una carta de manera aleatoria, que tiene el nombre de una emoción específica. En este momento, se solicita al participante seleccionar una carta a partir de una serie de cartas ubicadas en la izquierda de la pantalla. Cada carta contiene un rostro de Emma; así, el participante debe indicar si la carta seleccionada corresponde al nombre que se ve en la derecha. Si indica correctamente que la carta no corresponde, Emma lo felicita. Si escoge incorrectamente que la carta corresponde, Emma le pide que pruebe nuevamente con otra. Finalmente, si indica correctamente que la carta corresponde, se agranda la carta seleccionada a partir de una animación, se prende su cámara y Emma le pedirá que imite la emoción indicada. Se muestra una barra similar a la de *SurpriseBox* para realimentar que está imitando la emoción correctamente. El juego termina una vez que el participante haya imitado correctamente todas las emociones. En la Figura 25 se observa un ejemplo de esta escena.

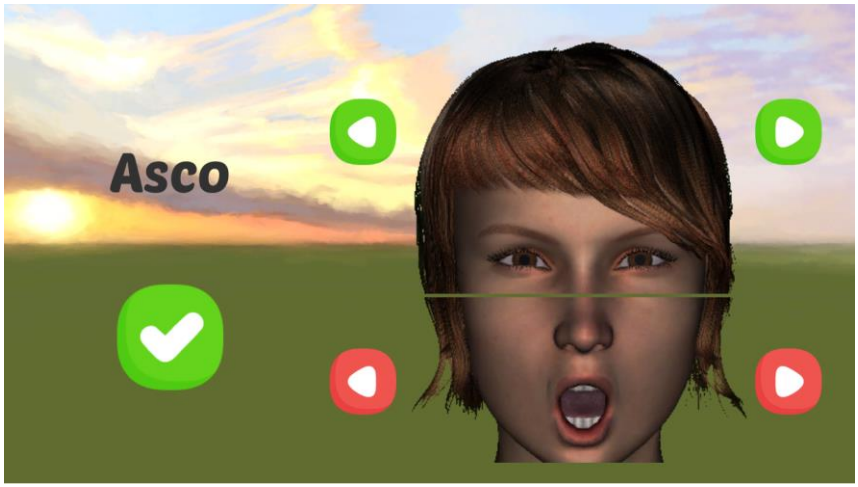

Figura 26. Imagen correspondiente a la escena *SortFace.*

I. SortFace

Esta escena corresponde al juego *Ordenar la Cara* y busca que los participantes logren diferenciar con mayor detalle las características de las partes del rostro al expresar distintas emociones, iniciando la etapa de reconocimiento. Para este juego, se crearon imágenes donde se dividió el rostro de Emma en tres partes, cortadas horizontalmente: una para las cejas, otra para los ojos y finalmente otra para la boca. En el juego, se les pide a los participantes ordenar la cara de Emma para que coincida con la emoción que se muestra a la izquierda. No obstante, en pruebas piloto realizadas se observó que este juego tenía una gran dificultad, dado que había casos donde la forma de las partes del rostro era muy similar en distintas expresiones faciales, por lo cual era difícil identificar la expresión facial correcta. Para mejorar esta escena, se disminuyó la dificultad de dos maneras distintas. Primero, se combinaron las imágenes de las cejas y los ojos, de forma que solo fuera necesario ordenar dos elementos. Segundo, de forma transparente para el participante, solo se incluyeron tres posibles combinaciones de emociones para cada intento, de manera que el participante no tuviera que analizar seis distintas expresiones faciales. Una vez el participante cree tener la combinación correcta, debe indicarlo por medio de un botón. En este momento, Emma le da realimentación correspondiente, dependiendo de si realizó la actividad de manera correcta. Si el participante acierta, se muestra su cámara y se le pide que imite la emoción. Se le da realimentación visual por medio de una estrella, la cual crece cuando el participante imita correctamente la emoción. El juego termina una vez el participante haya logrado ordenar e imitar todas las expresiones faciales. En la Figura 26 se observa un ejemplo de esta escena.

J. *Pop*

Esta escena, correspondiente al juego *Pop Emma*, tiene una particularidad respecto a todas las otras escenas de Emmaciones y es que hace uso de la posición relativa del rostro en la imagen de la cámara. Al iniciar el juego, Emma le explica al participante que el juego se basa en mover la cabeza para ganar y

le muestra un ejemplo, en el cual se ven dos caras de Emma en la pantalla: una de alegría y otra de tristeza. Emma le pide al participante que intente mover las cabezas de Emma con su cabeza. Cuando el participante mueve su cabeza hacia un lado, el rostro de Emma correspondiente al lado en que se movió se hará más grande y la otra se hará más pequeña. Esto se logra a partir de dos ecuaciones lineales cuyas entradas son la posición horizontal de la cabeza del usuario y su salida es la escala correspondiente para cada rostro de Emma.

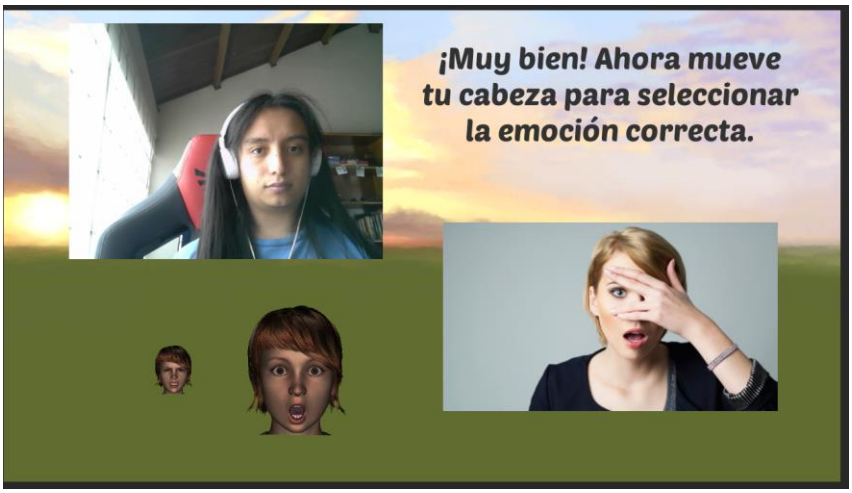

Figura 27. Imagen correspondiente a la escena *Pop.*

Una vez el niño haya logrado que uno de los rostros de Emma doble su escala, Emma le indica que lo ha hecho bien e inicia el juego. En el juego, se le muestran imágenes al participante y este debe imitarlas. Cuando las imite correctamente, Emma le pide que mueva la cabeza de la misma manera que lo hizo anteriormente, haciendo grande el rostro de Emma correspondiente a la emoción que se observa en la imagen. Nuevamente, Emma da distinta realimentación, dependiendo de si realizó la actividad correctamente o no. El juego termina una vez el participante haya logrado imitar y seleccionar las seis emociones básicas. En la Figura 27 se observa un ejemplo de este juego, donde se pidió imitar la sorpresa y luego se pidió mover la cabeza para seleccionar la emoción correcta. Se puede ver que, al estar posicionado hacia la derecha según la cámara, el rostro de la derecha, correspondiente a sorpresa, se hace más grande. De igual manera, el rostro de la izquierda, correspondiente a enojo, se hace más pequeño.

K. *IdentifyEmotion*

Esta escena corresponde a otra en la cual se combinan múltiples juegos. En este caso, *IdentifyEmotion* se utiliza para los juegos *Identifica la Emoción* y *¿Qué Emoción Soy?*, los cuales se usan en dos sesiones cada uno, para un total de cuatro actividades en las que se utiliza esta escena. Estos dos juegos se utilizan de manera muy similar, ya que en ambos se relatan situaciones. Por un lado, en *Identifica la Emoción*, se realizan descripciones cortas de distintas situaciones. Por otro lado, en *¿Qué Emoción Soy?*, se narran cuentos cortos donde el ambiente general corresponde a una de las seis emociones básicas. En esta escena se hace uso de la variable global de la sesión, que se utiliza de la siguiente manera:

- Sesión 2: Se muestran botones con las imágenes observadas en la Figura 12, correspondientes a alegría, miedo y asco.
- Sesión 3: Se muestran botones con las imágenes observadas en la Figura 12, correspondientes a tristeza, sorpresa y enojo.
- Sesión 4: Se muestran botones con las imágenes correspondientes a *¿Qué Emoción Soy?*, de forma que únicamente se habilitan aquellos correspondientes a alegría, miedo y asco.

- Sesión 5: Se muestran botones con las imágenes correspondientes a *¿Qué Emoción Soy?*, de forma que únicamente se habilitan aquellos correspondientes a tristeza, sorpresa y enojo.

Dependiendo del botón seleccionado, se le muestra al participante un cuento. Éste es narrado por uno de los miembros del equipo de investigación y se muestra la parte que se está leyendo en la pantalla, de forma que la velocidad con que avanza el texto en la pantalla es relativa a la longitud del segmento de audio del cuento. Al finalizar cada situación o cuento, se pide al participante indicar la emoción que se estaba evocando. Si el participante responde de manera incorrecta, Emma le pide intentar nuevamente y si responde de manera correcta, Emma lo felicita. Cada actividad termina una vez se hayan leído todas las situaciones o cuentos correspondientes a la sesión que esté activa y el participante haya logrado indicar la emoción que se siente durante el cuento. En la Figura 28 se observan dos imágenes de esta escena, la cual se activó durante la sesión 4. En la primera imagen, el participante puede escoger un botón para que inicie la narración del cuento y, en la segunda, el cuento está siendo narrado.

L. *Pairs*

Esta escena corresponde al juego *Encuentra el Par*, donde los participantes deben encontrar pares de cartas, los cuales corresponden a imágenes con la misma emoción. Puede que las imágenes sean iguales o únicamente tengan en común la emoción evocada. Este juego tiene tres niveles, como se detalla en la subsección 4.1.4. Esta escena se utiliza en las sesiones 2 y 3 de reconocimiento, correspondientes a tres distintas emociones cada una.

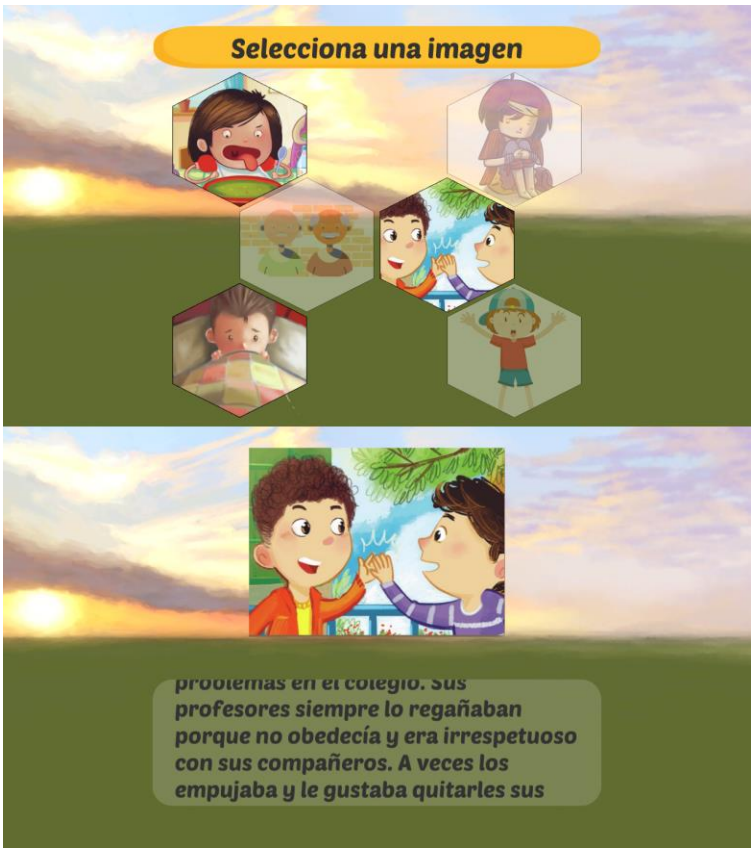

Figura 28. Imagen correspondiente a la escena *IdentifyEmotion.*

A nivel técnico, esta escena es bastante sencilla de implementar, ya que, en ella, se ubican seis cartas (3 pares) bocabajo de manera aleatoria en la pantalla. Cuando se selecciona la primera carta, esta queda bocarriba hasta que se seleccione otra. Si la segunda carta no hace pareja con la primera, ambas son volteadas nuevamente bocabajo. Si hacen pareja, se muestra al participante la imagen de su cámara y se pide imitar la emoción que se observa en las cartas. Una vez se combinan los tres pares, el juego avanza

al siguiente nivel. Una vez se complete el tercer nivel, el juego termina. En la Figura 29 se observa un ejemplo de esta escena.

Figura 29. Imagen correspondiente a la escena *Pairs*.

M. LivePhoto

Esta escena, correspondiente al juego *Foto en Vivo*, permite al participante practicar todas las emociones en el orden que él quiera. Así, en la parte izquierda de la pantalla se observan seis botones, correspondientes a las seis emociones básicas. Cuando el participante selecciona uno de los botones, el juego le pide posar para tomar una foto realizando la expresión facial correspondiente, de manera que se activa un botón en el cual puede tomar la foto. Todas las otras escenas funcionan en tiempo real, de manera que se detectan varias expresiones faciales al tiempo y se facilita más la imitación de estas emociones. No obstante, dado que en este juego solo se analizaría la expresión facial para un recuadro, la dificultad aumenta. Para disminuir levemente esta dificultad, de manera transparente para el jugador, se toman 10 fotos simultáneamente y se toma como predicción la emoción que se haya predicho más veces. Si el participante imita la emoción incorrectamente, Emma le pide que intente de nuevo. Si la imita correctamente, Emma lo felicita y se deshabilita el botón correspondiente a esa emoción. El juego continúa hasta que todos los botones estén deshabilitados. En la Figura 30 se observa un ejemplo de esta escena.

Figura 30. Imagen correspondiente a la escena *LivePhoto.*

N. *FindEmotions*

Esta escena corresponde al juego *Laberinto*, y se trató de la escena que implicó un mayor reto técnico, dado que se generó de manera procedimental. En *FindEmotions*, escena activa durante dos actividades,

se les pide a los participantes navegar por un laberinto utilizando un personaje. Para navegar por el laberinto, los participantes deben utilizar las flechas del teclado. Dentro de este, deben buscar rostros de Emma, los cuales expresan distintas emociones. Una vez encuentren uno de los rostros, deben imitar la expresión facial que está mostrando Emma, para poder obtener el rostro. Una vez lo obtengan, deben ubicarlo en el lugar donde se observe el nombre de la emoción. Una vez hayan ubicado correctamente todos los rostros, el juego termina.

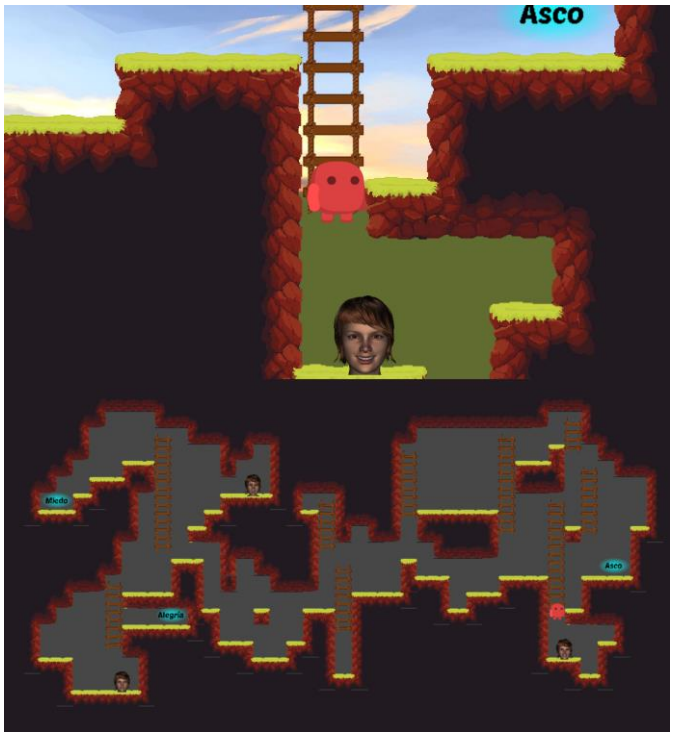

Figura 31. Imagen correspondiente a la escena *FindEmotions*.

La generación procedimental del laberinto se basa en modificar de manera aleatoria un mapa binario, en el cual, cada valor corresponde a un espacio dentro del juego. Un valor de 1 indica que en ese espacio hay obstáculos, mientras que un valor de 0 indica un espacio libre. La primera etapa para lograr esto es la generación de una plantilla, dada a partir del modelo automático llamado *autómata celular* [89], desde el cual se generan una serie de regiones sin conexión entre ellas. Para identificar qué regiones se generaron por medio de este algoritmo (espacios en los que hay múltiples valores en 0 unidos), se utiliza un mecanismo de procesamiento de imágenes llamado la conexión de componentes: para cada componente que se ha etiquetado, se aplica un algoritmo de inundación, de forma que sea posible identificar la región a la que pertenece. Este proceso continúa hasta que todos los componentes hayan sido etiquetados.

Dado que, para este juego, es necesario que todos los componentes hagan parte de una misma región, el siguiente paso en el algoritmo es conectar las regiones. Así, se buscan componentes de regiones semejantes que no estén conectadas y se realizan conexiones horizontales, verticales o diagonales, de manera aleatoria, momento en el que ambas regiones se consideran parte de una misma. Si se encuentra una región con pocos componentes (*N<5)*, se elimina esta región. Este proceso continúa hasta que solo haya una región dentro del mapa.

Finalmente, dado que los personajes del juego están afectados por gravedad, se generan sistemáticamente escaleras, para que los personajes sean capaces de llegar a zonas altas. Una vez terminado el mapa, se categoriza cada componente dependiendo de sus vecinos. Así, a cada componente se le da un nivel, dependiendo de cuantas paredes haya cerca de él. De esta manera, aquellos con una gran cantidad de paredes como vecinos se consideran candidatos para la ubicación de elementos en el

mundo. En este caso en particular, estos elementos son de tres tipos: Ubicación inicial del jugador, ubicación de los rostros de Emma y ubicación de los nombres de las emociones. Al ubicar así estos elementos, hay la seguridad que se distribuirán a lo largo del mapa, ya que no es posible que dos elementos sean vecinos si hay pocos obstáculos en cada uno de ellos.

En la Figura 31 se visualizan dos imágenes de *FindEmotions*. En la primera, se observa la escena tal como la está viendo el jugador. Se observa que abajo del jugador, se ubica el rostro de Emma correspondiente a la alegría y en la esquina superior derecha se observa el nombre de asco, lugar a donde el jugador debe llevar la expresión de asco. En la segunda imagen, se observa la totalidad de este laberinto, mapa que no puede observar el jugador, obligándolo a explorar completamente la escena. Como se puede detallar, una vez el jugador obtenga el rostro de la alegría, deberá llevarlo a la izquierda del mapa, donde se encuentra el letrero de alegría. En la segunda imagen también se observa que cada elemento fue ubicado en lugares con varios obstáculos alrededor, forzando al jugador a explorar el laberinto en su totalidad.

O. *DeleteEmotion*

Esta escena corresponde al juego *Elimina la Emoción* y es la primera donde se utilizan videos para estimular el reconocimiento de expresiones faciales. Esta escena se reproduce dos veces, para cada conjunto de emociones, dependiendo de la sesión de la etapa de reconocimiento que esté activa: 4 o 5. Al inicio de la escena, Emma le explica al participante las instrucciones del juego, en el cual se presiona un botón para reproducir un video sin contexto de duración aproximada de 7 segundos. Una vez termina el video, se muestran tres rostros de Emma, realizando expresiones faciales y se pide al jugador seleccionar aquellas que no se mostraron en el video. Cada vez que se seleccione correctamente una expresión facial, el rostro de Emma se vuelve pequeño hasta desaparecer. Cada vez que se seleccione incorrectamente una expresión facial, Emma brinda realimentación al participante, pidiéndole que lo vuelva a intentar. El juego termina una vez el participante haya logrado identificar correctamente las emociones evocadas en los videos correspondientes a las tres emociones presentes en la sesión activa. En la Figura 32 se observa un ejemplo de esta escena, donde se le pregunta al participante las emociones que no se mostraron en el video que se muestra anteriormente.

Figura 32. Imagen correspondiente a la escena *DeleteEmotion*.

P. *SlideEmotion*

Esta escena, correspondiente al juego *Desliza la Emoción*, busca que los participantes observen con mayor detalle las características de cada expresión facial. En el juego, se observan los nombres de tres emociones en distintas ubicaciones. Adicionalmente, se puede observar un rostro de Emma moviéndose de manera aleatoria por la pantalla, a una velocidad establecida. Primero, los participantes deben seleccionar al rostro de Emma para que deje de moverse y, posteriormente, ubicarlo en el nombre correspondiente. En una esquina de la pantalla se puede ver a Emma, quien imita expresiones faciales

dependiendo de la cercanía del rostro de Emma a cada uno de los nombres de las emociones. Así, los participantes pueden ver cómo cambia el rostro de Emma al pasar la imagen por distintos lugares, haciendo más fácil observar cambios entre emociones y transiciones entre ellas. El juego termina una vez el participante haya logrado ubicar correctamente las seis emociones básicas. En la Figura 33 se observa un ejemplo de esta escena, donde se observa que el rostro de Emma está expresando principalmente alegría, pero tiene rasgos de enojo, dado que está cerca a estos dos nombres.

Q. *Mirror*

Finalmente, esta escena corresponde al juego *Espejo*, en el cual se presentan videos con contexto a los participantes. Esta escena funciona de manera similar a *DeleteEmotion*, ya que se presenta un video y se debe indicar al final la emoción evocada en ese video. No obstante, en este caso, los participantes deben seleccionar la emoción que sí se vio en el video después de que este finalice, a partir de la selección de uno de seis botones. Al seleccionar correctamente la emoción, el participante debe imitarla a partir de la emoción que muestra Emma.

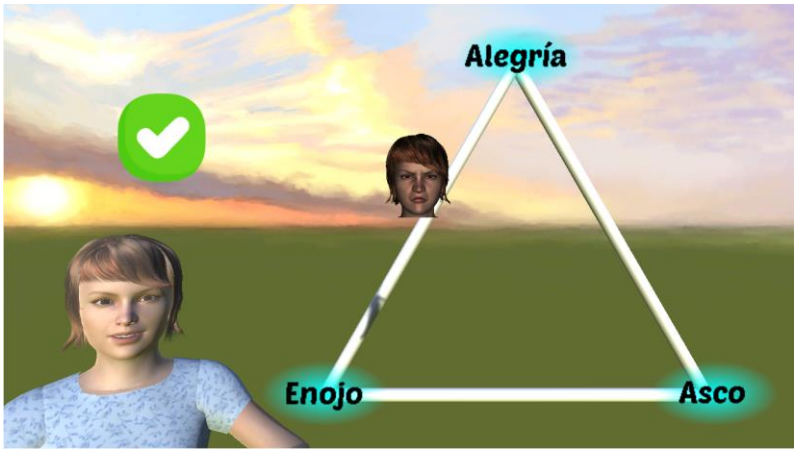

Figura 33. Imagen correspondiente a la escena *SlideEmotion*.

Al igual que en actividades anteriores, se realiza realimentación visual por medio de una estrella que crece en tamaño cuando el participante imita correctamente la emoción seleccionada. La actividad termina una vez el participante haya identificado correctamente las seis emociones evocadas en los videos con contexto. En la Figura 34 se observa un ejemplo de esta escena, donde se observa que los participantes deben imitar la expresión realizada por Emma, lo que le da al juego el nombre de “espejo”.

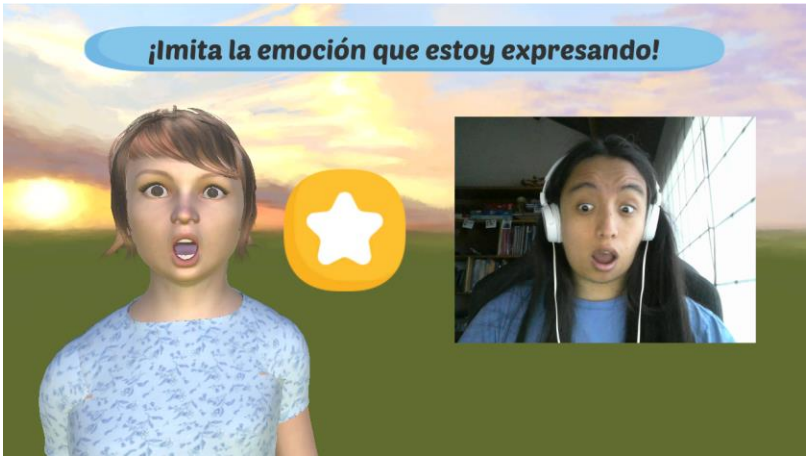

Figura 34. Imagen correspondiente a la escena *Mirror*.

4.4 Evaluación de la efectividad

4.4.1 Implementación de medidas psicométricas

Como se mencionó en la subsección 3.1.2, a conocimiento de nuestro equipo de trabajo, la única batería validada en Colombia que permita evaluar el reconocimiento de expresiones faciales en niños entre 6 y 8 años es la Evaluación Neuropsicológica Infantil (ENI). No obstante, esta batería evalúa varios procesos neuropsicológicos, de los cuales algunos no tienen relevancia para este proyecto. Teniendo en cuenta que los desarrolladores de ENI indicaron que es posible aplicar subpruebas de esta batería de manera independiente, se decidió aplicar el ítem «Reconocimiento de expresiones (expresión emocional)», el cual hace parte de la subescala de percepción visual. Esta evaluación consiste en realizar ocho preguntas a los participantes, que tienen respuestas objetivamente correctas o incorrectas. De esta forma, se evalúa la capacidad del participante de reconocer expresiones emocionales en una escala entre 0 y 8, dependiendo del número de respuestas correctas.

Figura 35. Ejemplo de las imágenes mostradas en el ítem de reconocimiento de expresiones de ENI.

Cada pregunta evalúa la capacidad del niño de reconocer la expresión facial que se observa en imágenes previamente escogidas por los autores de ENI. Un ejemplo de estas imágenes se observa en la Figura 35. Específicamente, al iniciar la prueba se les da la siguiente instrucción a los participantes: «Te voy a enseñar fotos de niños con distintas expresiones en sus caras y quiero que me digas que expresión tienen; por ejemplo, si están alegres o están tristes» y se les pregunta «¿Qué expresión tiene esta cara?» al mostrarles cada imagen. Aunque las respuestas esperadas son una de las seis emociones básicas, se aceptan sinónimos como respuestas correctas. Por ejemplo, se acepta que el niño responda «bravo» cuando la respuesta correcta es «enojado». De igual forma, tienen un tiempo máximo de 20 segundos para responder la pregunta; de no hacerlo, se continúa con la siguiente imagen y se considera que contestó mal. Para este proyecto, se utilizó ENI en dos instantes: Como línea base antes de iniciar el proceso de estimulación y al finalizar la última sesión de reconocimiento, de forma que se comparan los resultados obtenidos en estos dos instantes. Dado que la prueba únicamente evalúa el reconocimiento de expresiones, no se consideró necesario realizar otro instante de la prueba entre las etapas de imitación y reconocimiento.

4.4.2 Registros conductuales

Uno de los aspectos más importantes de este proyecto es el reconocimiento de expresiones faciales en tiempo real. No obstante, esta medida no se obtiene de manera trivial y no es posible obtenerla de manera que sea exacta, precisa y objetiva, ya que depende de múltiples variables no controlables, que llevan a los siguientes problemas:

- Contando únicamente con el programa, no es posible saber si los participantes no están imitando bien una emoción o si la están imitando bien pero el algoritmo no reconoce adecuadamente la imitación.
- Hay un retardo entre el momento en que se pide al participante imitar una emoción y el momento en que el participante inicia la imitación; sin embargo, sin dispositivos especializados no es posible conocer el tiempo exacto de este retardo.

Para disminuir la incertidumbre causada por estos dos problemas, se decidió evaluar la efectividad del algoritmo en tiempo real por medio de un registro conductual, cuya plantilla se muestra en el Anexo 5, donde los investigadores marcan el cumplimiento de signos de una expresión facial e indican el tiempo en el que inicia la imitación de esta, teniendo en cuenta la descripción de los rasgos faciales de las emociones que indica Ekman [32]. De esta manera, ya que un humano puede identificar con mayor certeza los rasgos faciales, se puede indicar si un retardo alto en el reconocimiento de expresiones faciales es causado por una falla del algoritmo de reconocimiento o por la imitación errónea por parte de los participantes. Para reducir la subjetividad de este método, todos los registros conductuales de todos los participantes fueron diligenciados por la misma persona, una estudiante de psicología de la Corporación Universitaria Minuto de Dios UNIMINUTO.

Cabe resaltar que en el futuro es posible realizar esta medición de manera más exacta y precisa. Actualmente, el tiempo indicado en el registro conductual depende de la exactitud con la cual el observador registra distintos instantes y lo que se considera el instante en el que inicia la imitación, lo que aumenta los errores aleatorios. En trabajos futuros donde se quiera profundizar en la efectividad de reconocimiento de expresiones faciales en tiempo real, se puede hacer uso de dispositivos de electroencefalografía, con los cuales se pueden detectar cambios en la actividad cerebral de la corteza visual y en el área fusiforme ubicada en el lóbulo temporal inferior, encargada del reconocimiento de rostros con mayor exactitud, que indiquen el instante en que el participante es consciente que debe imitar cierta expresión facial. Tomando este valor como verdad absoluta, se puede cuantificar con mayor exactitud la efectividad de reconocimiento del algoritmo, ya que es posible aislar el tiempo de reconocimiento del algoritmo de aquel causado por el comportamiento de los participantes. No obstante, el registro conductual sigue siendo fundamental en esta medición, porque se trata del sistema que permite separar errores humanos de errores causados por el algoritmo de reconocimiento de expresiones faciales.

4.5 Selección muestral

4.5.1 Población objetivo

Como se ha mencionado a lo largo de este documento, la herramienta de estimulación busca ayudar a los procesos de comunicación de niños con TEA, por medio de los procesos de imitación y reconocimiento. No obstante, también se utiliza con un grupo control, con el cual se pueden observar diferencias en los resultados ENI y en la imitación de expresiones faciales.

4.5.2 Criterios de inclusión y exclusión

A. Grupo experimental

a. Criterios de inclusión

- Niños diagnosticados con TEA por un profesional de la salud.
- Edades entre 6 y 8 años.
- Acceso a un computador, internet y una cámara web.

b. Criterios de exclusión

- Problemas auditivos, visuales o motores que no permitan la interacción con el computador de manera independiente.

B. Grupo control

a. Criterios de inclusión

- Niños sin diagnósticos de trastornos del neurodesarrollo, psiquiátricos o neurológicos.

- Edades entre 6 y 8 años.
- Acceso a un computador, internet y una cámara web.

b. Criterios de exclusión

- Problemas auditivos, visuales o motores que no permitan la interacción con el computador de manera independiente.

4.5.3 Pruebas piloto

Antes de realizar las pruebas que se han mencionado a lo largo de esta sección, se realizaron pruebas piloto de índole técnico, de forma que niños dentro de las edades deseadas, sin trastornos del neurodesarrollo, psiquiátricos o neurológicos, probaran la interfaz. Así, se realizaron pruebas piloto con dos niñas, cada una con 7 años, donde se aplicaron todas las sesiones de Emmaciones. Un resultado importante de las pruebas piloto es que se encontró que cada sesión dura aproximadamente 20 minutos en realizarse, motivo por el cual se evaluó la posibilidad de incluir múltiples sesiones por día.

Adicionalmente, a partir de las pruebas piloto, se realizaron varios ajustes a la interfaz, los cuales se mencionan en la subsección 4.5.3. Entre estos ajustes, se encuentran la reubicación de piezas en el juego *Rompecabezas* y la disminución de dificultad en el juego *Ordena la Cara*. Las participantes indicaron que les gustaron las pausas activas realizadas. Una de las participantes indicó que su juego favorito fue *Laberinto* y que no le gustaron las actividades *Identifica la Emoción* y *¿Qué Emoción Soy?*, ya que le parecieron actividades muy lentas. De igual forma, por medio de inspección visual, se identificó que ambas participantes mostraron facilidad para la imitación de alegría, miedo, sorpresa y asco, mientras que hubo mayor dificultad en la imitación de tristeza y enojo.

4.5.4 Muestra utilizada

Los niños que participaron en las pruebas finales del experimento fueron hallados por medio de contactos personales de los investigadores, de forma que no se involucró ninguna organización en el reclutamiento. Se logró contar con el apoyo de tres participantes, cuyos detalles se observan en la Tabla 13.

Tabla 13. Información de los participantes del estudio experimental.

Sujeto	**Edad**	**Género**	**Estado**	**Curso académico**
1	6	Masculino	TEA	Primero de primaria
2	6	Masculino	Neurotípico	Desescolarizado
3	6	Masculino	Neurotípico	Primero de primaria

Los representantes legales de cada uno de los participantes firmaron el consentimiento informado mostrado en el Anexo 1. De igual forma, se llevaron a cabo reuniones donde se aclararon dudas de los experimentos y se pudo conocer a cada uno de los participantes, preguntándoles sus intereses para tenerlos en cuenta durante la realización de las pausas activas.

V. RESULTADOS Y DISCUSIÓN

5.1 Algoritmo de reconocimiento facial

Algunos de los resultados técnicos que se van a mostrar a continuación se refieren a tiempos de ejecución. Teniendo en cuenta que estos pueden cambiar dependiendo del poder computacional con el que se cuente, todas las pruebas se hicieron con un mismo computador, de forma que son comparables. Las especificaciones técnicas de este computador se observan en la Tabla 14.

Tabla 14. Especificaciones técnicas del computador en el que se realizaron las pruebas.

Propiedad	**Valor**
Sistema operativo	Windows 10 – 64 bits
Procesador	Intel Core i7-6700HQ
Memoria	16GB de RAM
DirectX	Versión 12
Gráficos	Nvidia GeForce GTX 1060 6GB

5.1.1 Corrección de contraste previa a la detección de rostros

En la Figura 36 se observan las imágenes utilizadas para probar la corrección de contraste previa a la detección de rostros. Parte de estas pruebas también son utilizadas para probar la corrección de contraste previa a la ubicación de marcadores faciales, la cual se describe más adelante.

Figura 36. Imágenes de prueba para analizar la efectividad de los algoritmos de corrección de contraste.

En la Figura 37 se puede observar la combinación la efectividad del algoritmo de detección de rostros al aplicarse con cada combinación de técnicas de ecualización de histograma y espacios de color mencionadas anteriormente. Las imágenes de la izquierda corresponden a modificaciones realizadas a la imagen con alta iluminación, mientras que las imágenes de la derecha corresponden a las modificaciones realizadas a la imagen con baja iluminación. Teniendo en cuenta esta división, las columnas impares corresponden a la ecualización de histograma tradicional, mientras que las columnas pares corresponden a CLAHE. Las filas indican el espacio de color donde se hizo la ecualización de histograma, en el siguiente orden de arriba hacia abajo: imagen original, grises, cieLAB, HSL y RGB.

Figura 37. Prueba del algoritmo de detección de rostros con cada combinación de técnicas de ecualización y espacios de color.

Se puede observar que, independientemente del espacio de color utilizado o de la técnica de corrección de contraste utilizada, SSD puede ubicar el rostro de la imagen, sin eliminar secciones como la frente o las mejillas. Esto es particularmente interesante en combinaciones que distorsionan completamente la imagen, como la ecualización tradicional aplicada en HSL o CLAHE aplicado en RGB. Estos resultados pueden no indicar cual técnica o cual espacio de color son los más adecuados para mejorar el contraste de una imagen, pero son un buen ejemplo de la robustez de SSD, ya que es un algoritmo que puede detectar rostros bajo distintas condiciones de luz. Esto es cierto, no solamente para estas imágenes, sino para las pruebas empíricas que se realizaron con una cámara en vivo, las cuales trataron distintas condiciones de luz. Es importante resaltar que el protocolo contempla que el participante se encuentre en un ambiente con iluminación adecuada, por tanto, no es necesario probar ciertas condiciones, como estar a contraluz, porque esto no puede pasar en las pruebas. En la subsección 4.2.3 se puede observar con más detalle la robustez de este algoritmo, donde se muestran cambios distintos a la iluminación. A partir de los resultados obtenidos, se decidió que no se aplicaría ninguna técnica de corrección de contraste para la detección de rostros, ya que estas pueden ralentizar el algoritmo de reconocimiento de expresiones faciales y no aportan beneficios visualmente notorios.

5.1.2 Corrección de contraste previa a la ubicación de marcadores faciales

En aras de brevedad y de una mayor visualización de los detalles, solo se mostrarán las pruebas realizadas en la imagen con menor iluminación, porque corresponde al caso con problemas en la imagen sin modificar. En la Figura 38 se observa la ubicación de marcadores faciales para la imagen original y cada una de las modificaciones realizadas a esta. Aunque no se observa en la figura, este algoritmo requiere de la detección previa del rostro, la cual se hizo con la imagen original sin modificar. En la primera columna se observan todas las modificaciones realizadas con ecualización de histograma tradicional y en la segunda columna se observan todas las modificaciones realizadas con CLAHE. De arriba hacia abajo, los espacios de color utilizados en la modificación son: imagen original, grises, cieLAB, HSL y RGB.

Se observa en la Figura 38 que, en la imagen original, la ubicación de marcadores faciales no es adecuada, porque no se reconoce la quijada correctamente, lo que modifica de igual forma los marcadores de los

labios. Así mismo, hay un leve desfase en la ubicación de los marcadores de los ojos. En cuanto a las modificaciones de la imagen original, se observan errores despreciables en la ubicación, particularmente en el contorno del rostro y las cejas. Al realizar pruebas con una cámara en vivo, se observó que este patrón se repetía en distintos tipos de iluminación. Generalmente, cualquier corrección de contraste realiza un buen trabajo para mejorar la ubicación de los puntos. En la subsección 4.2.4 se observan errores más notables en esta ubicación de los marcadores faciales, los cuales no están relacionados con la iluminación de la habitación.

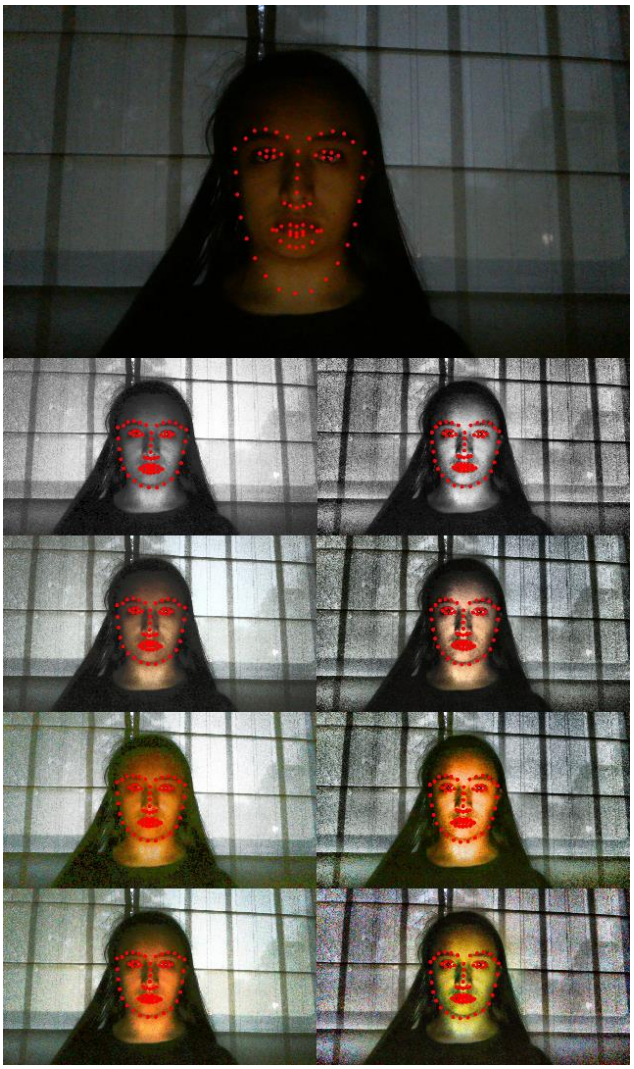

Figura 38. Prueba del algoritmo de ubicación de marcadores faciales con cada combinación de técnicas de ecualización y espacios de color.

Teniendo en cuenta que una prioridad de este proyecto es que el reconocimiento de expresiones faciales se haga en tiempo real, se decidió seleccionar el método más eficiente entre los explorados. En la Tabla 15 se observan los tiempos de cada uno de los métodos, los cuales son comparables, dado que fueron calculados en el mismo computador, cuyas especificaciones técnicas se observan en la Tabla 14.

Tabla 15. Tiempos de ejecución de técnica de corrección de contraste.

Método	**Tiempo de ejecución (ms)**
Ecualización tradicional en grises	1.962

CLAHE en grises	1.994
Ecualización tradicional en cieLAB	12.964
CLAHE en cieLAB	11.457
Ecualización tradicional en HSL	6.981
CLAHE en HSL	7.979
Ecualización tradicional en RGB	2.991
CLAHE en RGB	7.979

Como se puede ver en la tabla, estos tiempos parecen despreciables; sin embargo, teniendo en cuenta que las distintas etapas del proceso acumulan el tiempo que se demora el reconocimiento, es importante aprovechar al máximo los recursos con los que se cuenta. Así, se decide realizar una ecualización tradicional en grises. Las razones por las cuales este método es más ágil que los demás son dos principalmente: Por un lado, la ecualización tradicional hace cálculos de manera global, no local, por lo cual no es necesario hacer un recorrido por la imagen para mejorar el contraste. Por otro lado, teniendo en cuenta que una imagen en grises tiene un solo canal de intensidad, solo es necesario hacer una vez la ecualización y no es necesario combinar nuevamente los canales para reconstruir la imagen.

5.1.3 Corrección de contraste previa a la extracción de características de HOG

En la Figura 39 se observa la aplicación de distintas técnicas de corrección de contraste para adaptar los datos a la extracción de características HOG. También se observa una representación visual de los gradientes orientados. De arriba hacia abajo: imagen original, ecualización tradicional en grises, CLAHE en RGB y CLAHE en cieLAB.

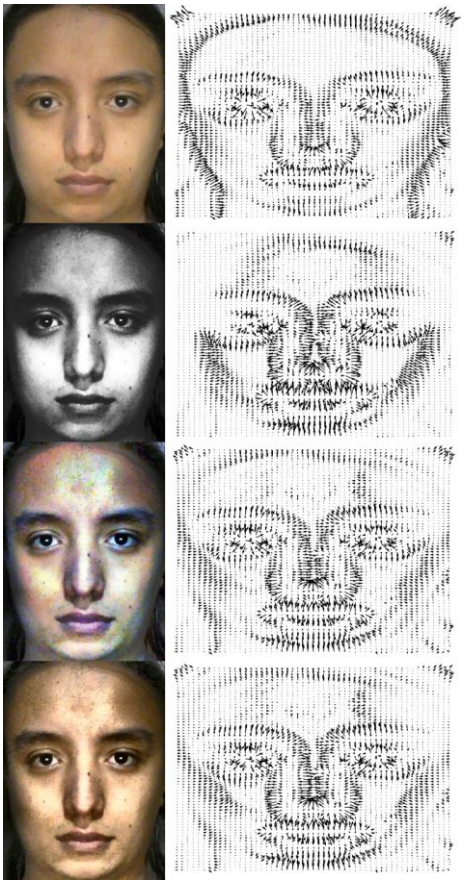

Figura 39. Representación visual del efecto de cada técnica de preprocesamiento en el cálculo de HOG.

Se pueden observar las ventajas con las que cuenta el método CLAHE en cieLAB respecto a otros. Por un lado, en la imagen original no se detallan mucho las expresiones faciales; por ejemplo, en la frente no hay cambios notorios, por lo cual la magnitud del gradiente es nula. En la ecualización tradicional en grises, se puede ver que se uniformiza la intensidad en el contorno del rostro y se pierde información en estas áreas. En la aplicación de CLAHE en RGB se observa que se añade ruido a la imagen, ya que se distorsionan los colores. Finalmente, en la aplicación de CLAHE en cieLAB se resaltan detalles que no se observan fácilmente en la imagen original, particularmente en la frente, el ceño y las mejillas. Sin embargo, se mantiene la naturalidad de los colores, lo que evita que se añada información que no pertenece a la imagen.

5.1.4 Detección facial

En la Figura 40 se observan ejemplos de imágenes utilizadas para decidir el algoritmo a utilizar.

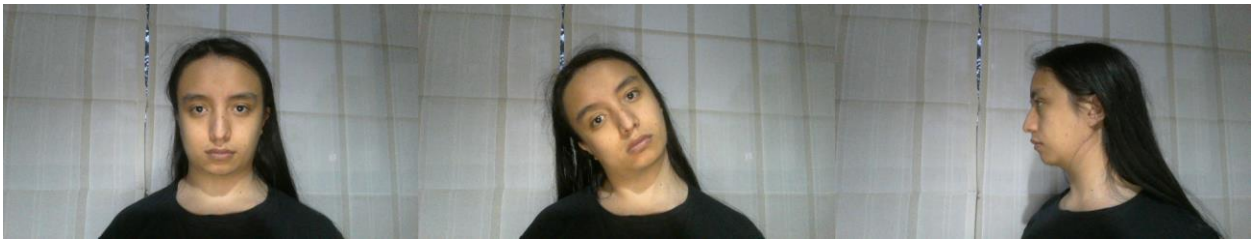

Figura 40. Imágenes de prueba para los algoritmos de detección de rostros.

En la Figura 41 se muestra la detección de rostros de cada algoritmo para la primera condición.

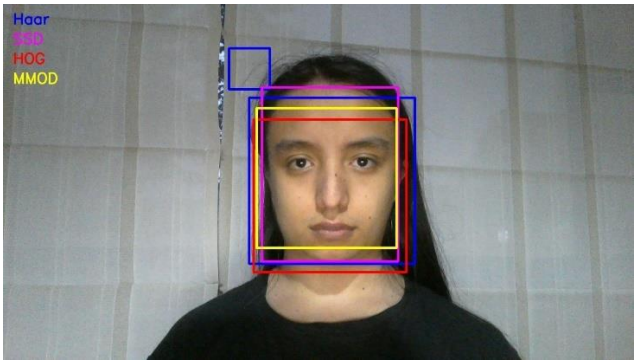

Figura 41. Implementación de detección de rostros para cada algoritmo en la primera condición.

En la Figura 41 se pueden apreciar muchas diferencias entre cada uno de los algoritmos de detección de rostros. El primer algoritmo, Haar, logra cumplir con el objetivo de detectar el rostro, incluyendo frente y mejillas. Sin embargo, también se encuentra un falso positivo en la esquina superior izquierda del rostro por medio de este método. En el caso de SSD, también se incluye todo el rostro en la detección del algoritmo, incluso una mayor área de la frente que Haar. HOG hace un buen trabajo al contener un mayor rango horizontal del rostro respecto a SSD; sin embargo, no incluye gran parte de la frente. MMOD, por su parte, incluye gran parte de la frente y la totalidad de las mejillas.

En la Figura 42 se observa la aplicación de cada uno de estos algoritmos para la segunda condición. En este caso, es notorio que Haar no detecta el rostro correctamente, ya que no logra ubicarlo y genera dos falsos positivos. SSD, por su parte, cubre la totalidad del rostro, incluyendo la frente y las mejillas. HOG nuevamente logra abarcar la totalidad de las mejillas, sin embargo, ignora parte de la frente. Finalmente, MMOD tiene una exactitud similar a la de SSD, pero su rango es levemente menor tanto en frente como en mejillas.

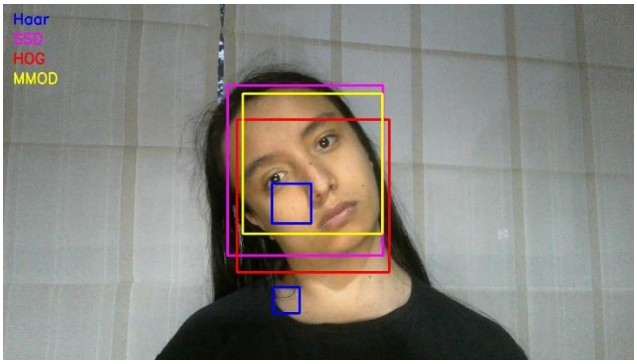

Figura 42. Implementación de detección de rostros para cada algoritmo en la segunda condición.

Finalmente, en la Figura 43 se observa la efectividad de cada algoritmo para detectar el rostro de la tercera condición. En este caso, Haar vuelve a generar un falso negativo y no detecta el rostro en la imagen. Similarmente, HOG no detecta ningún rostro en esta situación. SSD se enfoca en el rostro e ignora el resto de la cabeza. MMOD por su parte, cubre un mayor rango que SSD, incluyendo la oreja que se visualiza.

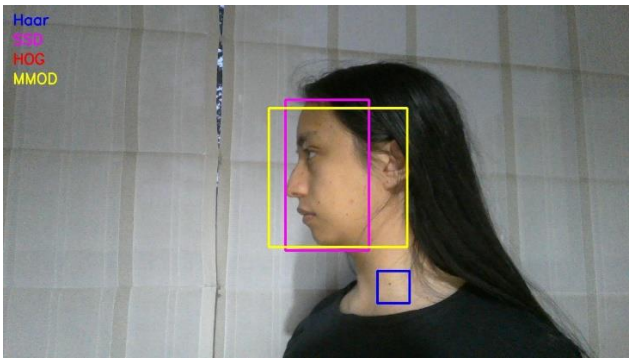

Figura 43. Implementación de detección de rostros para cada algoritmo en la tercera condición.

Además de evaluar la exactitud de los algoritmos, se calculó el tiempo que se demora cada algoritmo en detectar el rostro, con el computador que se observa en la Tabla 14. Esta medición se encuentra en la Tabla 16. Para decidir de manera cuantitativa cuál es el mejor algoritmo de detección de rostros, dentro de los analizados, se genera una puntuación objetiva dependiendo de la calidad de detección. Los criterios que se tuvieron en cuenta fueron: la detección del rostro, la inclusión de mejillas, la inclusión de frente y el tiempo de cómputo. Las áreas utilizadas en estos criterios se escogieron dado que son zonas de interés para el análisis de los gradientes orientados del rostro. Cada uno de los primeros tres criterios suman un punto por cada pose y el criterio de tiempo de cómputo se utiliza para tomar una decisión final. Dado que algunos algoritmos cumplen parcialmente con los criterios, se acumulan 0.5 puntos por detección parcial de mejillas y frente. De igual forma, si un algoritmo muestra falsos positivos, no obtiene puntos por detección de rostro. Así, el algoritmo con una mayor cantidad de puntos y tiempo de cómputo adecuado para detección en tiempo real es elegido para el reconocimiento de expresiones faciales. Esto también se puede observar en la Tabla 16.

Tabla 16. Comparación de la efectividad de los algoritmos de detección de rostros.

Pose	**Criterio**	**Haar**	**SSD**	**HOG**	**MMOD**
Tiempo (ms)		67.8	40.9	439.8	6842.6
Pose 1	**Rostro**	0	1	1	1
	Frente	1	1	1	1
	Mejillas	1	1	0	0.5

Pose 2	**Rostro**	0	1	1	1
	Frente	0	1	0.5	0.5
	Mejillas	0	1	1	1
Pose 3	**Rostro**	0	1	0	1
	Frente	0	1	0	1
	Mejillas	0	1	0	1
Total		2	9	4.5	8

En la tabla, se puede observar que ambos algoritmos de aprendizaje profundo (SSD y MMOD) son bastante robustos. Por un lado, SSD tuvo un puntaje perfecto, mientras que MMOD falló en algunos casos por no incluir en su totalidad a la frente o a las mejillas. No obstante, la velocidad de SSD es drásticamente superior a la de MMOD, porque se puede realizar a una velocidad de 14.7 recuadros por segundo (FPS), a comparación de 0.15FPS de MMOD, una velocidad 100 veces mayor, aproximadamente. Autores indican que, utilizando una tarjeta gráfica, la velocidad de MMOD puede ser aproximadamente 7 veces mayor a la de SSD [56]. Sin embargo, no se puede esperar que todos los participantes cuenten con una tarjeta gráfica en sus computadores, por lo cual se descartó realizar esa prueba. Un aspecto interesante de SSD es que es más rápido que Haar y HOG, por lo cual, la única desventaja apreciable con la que cuenta respecto a estos algoritmos es que hace uso de un modelo de aprendizaje profundo, esto hace que Emmaciones ocupe más espacio de disco duro. Por las razones expuestas, se decidió utilizar SSD como algoritmo de detección de rostros para el algoritmo de reconocimiento de expresiones faciales.

5.1.5 Ubicación de marcadores faciales

Como primera prueba, se muestra la ubicación de marcadores en condiciones ideales de pose, dado que la persona mira directamente hacia la cámara sin ninguna rotación aparente. En la Figura 44 se observan los resultados para cada algoritmo en este caso. En esta prueba, se empiezan a observar las primeras fallas del algoritmo LBF, en el cual, los marcadores faciales correspondientes a la quijada se ubican mucho más debajo de lo requerido, posiblemente por elementos como sombras o la camiseta en la imagen. Dado que el algoritmo busca que los marcadores se ubiquen de manera natural, eso hace que los marcadores correspondientes a los labios y a la nariz también se encuentren en posiciones inadecuadas. No obstante, el algoritmo es capaz de ubicar correctamente los puntos correspondientes al contorno del rostro, a los ojos y a las cejas. En cuanto al algoritmo de Kazemi, no hay fallas visualmente notables para esta pose.

Figura 44. Ubicación de los marcadores faciales para la primera situación. Izquierda: LBF. Derecha: Kazemi.

La segunda prueba realizada representa un caso donde el participante mira levemente hacia un lado. En la Figura 45 se observan los resultados de cada algoritmo para este caso. En LBF, se observa que la ubicación general de los marcadores de la parte derecha del rostro no es adecuada, porque no se encontraron elementos claves como el ojo y la ceja derechos, a la vez que se distorsionaron elementos como los labios y la nariz. En el caso del algoritmo de Kazemi, se observa que la ubicación de puntos está hecha de manera más precisa. Sin embargo, el algoritmo falla levemente en la ubicación de los marcadores

correspondientes a la ceja derecha, especialmente los puntos 19-22. De igual forma, se observa un error en los puntos 24 y 25, correspondientes a la ceja izquierda.

Un caso donde ninguno de los algoritmos dio resultados deseados es cuando se gira la cabeza hacia un lado, observado en la Figura 46. Ambos algoritmos distorsionan la ubicación de los puntos de la parte derecha del rostro. En ambos algoritmos, los puntos correspondientes al ojo derecho se ubican en la ceja y esto sitúa los puntos de la ceja en la frente. Es razonable que esto pase, porque ninguno de los algoritmos fue entrenado con imágenes donde los participantes rotaran el rostro. No obstante, se puede observar que LBF ubica de mejor los puntos 1-12, localizados en el contorno del rostro. Sin embargo, el algoritmo de Kazemi ubica mejor los puntos de los labios y de la nariz. En general, aunque ninguno de los algoritmos cumple con el objetivo propuesto, se puede decir que el algoritmo de Kazemi es más efectivo para los fines de la investigación, dado que ubica los puntos más relevantes de mejor forma.

Figura 45. Ubicación de los marcadores faciales para la segunda situación. Izquierda: LBF. Derecha: Kazemi.

Figura 46. Ubicación de los marcadores faciales para la tercera situación. Izquierda: LBF. Derecha: Kazemi.

Las últimas dos pruebas se hicieron teniendo en cuenta expresiones faciales que se realizan comúnmente en Emmaciones. Aunque se podrían mostrar pruebas para cada una de las seis expresiones faciales de las emociones, se van a presentar únicamente aquellas donde hay diferencias visuales más notorias entre los algoritmos, como la sorpresa y el asco. Para las expresiones faciales de la alegría, la tristeza, el enojo y el miedo, ambos algoritmos brindan resultados similares y satisfactorios.

En la Figura 47 se observa la ubicación de los marcadores faciales para cada uno de los algoritmos cuando se imita la expresión facial de la sorpresa. En este caso, ambos algoritmos logran ubicar satisfactoriamente los marcadores correspondientes a las cejas, los ojos y el contorno del rostro. Sin embargo, se observa que LBF no logra colocar exitosamente los labios, de forma que los marcadores 61-65 quedan ubicados en el labio inferior cuando deberían estar en el labio superior. Como en otros casos, esto distorsiona la ubicación de otros marcadores, como los de la nariz, los cuales quedan localizados en el labio superior. Por su parte, no se observan errores visualmente notorios en la ubicación de los marcadores del algoritmo de Kazemi.

Figura 47. Ubicación de los marcadores faciales la expresión facial de la sorpresa. Izquierda: LBF. Derecha: Kazemi.

En la Figura 48 están los marcadores faciales cuando se imita la expresión facial del asco. Este caso se escogió porque se puede observar que ninguno de los algoritmos es capaz de ubicar correctamente los marcadores 56-60, correspondientes al labio inferior. Esto causa que se distorsione la posición de los marcadores del contorno del rostro. No obstante, LBF parece hacer un mejor trabajo en este aspecto. Es posible que los marcadores faciales se ubiquen incorrectamente en este caso porque los algoritmos esperan que el labio inferior tenga una curvatura definida, y al no presentarse esto, como es el caso del asco, los puntos se desubican.

Figura 48. Ubicación de los marcadores faciales la expresión facial del asco. Izquierda: LBF. Derecha: Kazemi.

Con los resultados obtenidos con estos algoritmos, no pareció relevante realizar una comparación cuantitativa entre ellos, porque el algoritmo de Kazemi demostró ser visualmente superior en casos comunes. Aunque LBF es ligeralmente superior que el algoritmo de Kazemi en el caso del asco, este parece ser el único caso donde se ve esto. Por otro lado, es importante que la apertura de boca se detecte correctamente, dado que varias expresiones como la sorpresa, el miedo y el enojo pueden tener esta característica. Finalmente, debido a que se extraen características de textura, los fallos que presenta el algoritmo de Kazemi, por ejemplo, en el asco, se balancean por una detección de líneas nasolabiales por medio de histogramas de gradientes orientados.

5.1.6 Predicción de expresiones faciales

A continuación, se expondrán las matrices de confusión obtenidas para cada combinación de bases de datos, de forma que se pueda tomar una decisión respecto a las bases de datos a utilizar. Dado que existen parámetros aleatorios dentro de una red neuronal, como la inicialización aleatoria de pesos en las neuronas, se fija una semilla, de forma que los resultados no dependan de suerte y únicamente de la habilidad de la arquitectura de discriminar categorías. En las matrices de confusión, las categorías están ordenadas alfabéticamente según su nombre en inglés, porque que este es el orden que reciben en todas las bases de datos analizadas, de la siguiente manera:

1. Miedo
2. Enojo
3. Asco

4. Alegría
5. Expresión neutra
6. Tristeza
7. Sorpresa

Primero, el Anexo 3.1 se observan los resultados obtenidos al solo utilizar la base de datos KDEF para entrenar el modelo de ANN. En general, se puede ver que los resultados obtenidos son satisfactorios, particularmente para la alegría y para la expresión neutra. Un caso excepcional es el enojo, cuya expresión facial no fue predicha en ningún caso por la ANN. También, existen claras confusiones entre expresiones faciales. Por un lado, gran parte de las veces que enojo fue clasificado incorrectamente, se clasificó como asco. Otro caso notorio fue la sorpresa, clasificada 12 veces como miedo.

En el Anexo 3.2 se observan los resultados de predicción al solo utilizar la base de datos JAFFE, en la cual participan mujeres japonesas. En este caso, se cuenta con pocas imágenes de prueba y entrenamiento. Es difícil saber con certeza si los resultados obtenidos son buenos en este caso, dado que no hay suficientes datos para realizar un análisis apropiado. No obstante, se observa que hubo pocos errores en la predicción del enojo, la alegría y la expresión neutra, mientras que las demás expresiones faciales fueron fácilmente confundidas entre ellas.

El Anexo 3.3 contiene la matriz de confusión de la ANN entrenada únicamente con la base de datos CAFE, la cual está compuesta por imágenes de niños pertenecientes a cinco etnias distintas. Al igual que en los dos casos anteriores, la alegría es fácilmente reconocible por esta ANN. No obstante, en este caso, son dos las expresiones faciales que no son reconocidas en lo absoluto por el algoritmo: el asco y la tristeza. De igual forma, se observa una confusión consistente entre el miedo y el enojo, tanto para imágenes cuya expresión real es enojo como para aquellas cuya expresión real es miedo. Finalmente, se observa que la tristeza es comúnmente predicha como expresión neutra.

Respecto al Anexo 3.4, se aprecian los resultados generados por la ANN cuando se combinan las bases de datos KDEF y JAFFE. Al igual que en los dos casos anteriores, la tristeza no se clasifica con facilidad, con solo 4 muestras clasificadas correctamente en esta categoría. La alegría y la sorpresa se clasifican correctamente en la mayoría de los casos y, aunque las imágenes que muestran expresión neutra también se clasifican correctamente, varias imágenes de tristeza igualmente se clasifican como expresión neutra, tendencia que se observa en la mayoría de los casos. Por su parte, el miedo y el enojo obtienen resultados parcialmente positivos.

En el Anexo 3.5 se observa la matriz de confusión que se obtiene cuando se combinan las bases de datos KDEF y CAFE. En este caso, con una gran cantidad de imágenes, la alegría es fácilmente reconocible por el algoritmo, con leves confusiones con el miedo. El miedo, por su parte, no es predicho correctamente en ningún caso y la tristeza se clasifica correctamente un número insignificante de las veces. En este caso, el miedo es reconocido erróneamente por otras emociones, incluyendo asco, alegría y expresión neutral, pero particularmente por enojo y sorpresa. Por último, el asco es reconocido gran parte de las veces como enojo.

Con relación al Anexo 3.6 se observan los resultados obtenidos cuando la ANN es entrenada por imágenes de las bases de datos JAFFE y CAFE. En este caso, se obtiene una gran cantidad de errores en la clasificación de emociones. La expresión del enojo no es predicha en ningún momento por la ANN. De igual forma la tristeza se suele confundir con el enojo y con la expresión neutra. Por su parte, el enojo tiende a ser elegido erróneamente como miedo. No obstante, la alegría y la sorpresa muestran buenos resultados en su clasificación.

Por último, en el Anexo 3.7 se observan los resultados obtenidos cuando se combinan las tres bases de datos para entrenar una ANN. En este caso, las emociones suelen ser reconocidas correctamente por el algoritmo gran parte de las imágenes, con la notable excepción de tristeza, la cual se clasifica erróneamente en la mayoría de los casos. Por su parte el enojo se clasifica como asco comúnmente, al

igual que en la ANN que utiliza únicamente la base de datos KDEF. El miedo tiende a ser reconocido correctamente; sin embargo, también es comúnmente clasificado como enojo, alegría, tristeza o sorpresa.

En la Tabla 17 se resumen las exactitudes de cada ANN, correspondiente a una combinación de bases de datos. En esta tabla, se hace énfasis en la exactitud individual de cada expresión facial al igual que en la exactitud general obtenida por cada ANN. Aquí se observa que, en general, las expresiones faciales de la alegría y de la sorpresa, además de la expresión neutra son fáciles de reconocer, independientemente de las bases de datos utilizadas para entrenar los datos. Por su parte, la tristeza se reconoce con bastante dificultad en todas las combinaciones de bases de datos, pero se reconoce con mayor facilidad cuando KDEF o JAFFE se utilizan de manera individual. Para la selección de las bases de datos, se tomaron en cuenta las exactitudes promedio al igual que las exactitudes individuales por emoción. Los conjuntos de bases de datos con mejor exactitud que se evaluaron fueron:

- KDEF
- JAFFE
- KDEF+JAFFE
- KDEF+JAFFE+CAFE

KDEF+JAFFE y KDEF+JAFFE+CAFE se descartaron rápidamente, porque mostraron exactitudes deficientes en el miedo, el enojo y la tristeza. Por su parte, JAFFE solo mostró bajas exactitudes en el miedo y en la expresión neutral, mientras que KDEF solo mostró baja exactitud en el asco. Se decidió seleccionar a KDEF, ya que JAFFE contaba con dos problemas fundamentales: El primero: la cantidad de imágenes en la base de datos es reducida, por lo cual, la generalización no se puede obtener con facilidad. El segundo: la base de datos consta únicamente de rostros de mujeres japonesas; teniendo en cuenta que, inicialmente, Emmaciones se probó con un público latinoamericano, el reconocimiento de emociones se podía dificultar. Aunque se ha probado que las expresiones faciales son universales, los marcadores faciales de Kazemi fueron entrenados a partir de HELEN, una base de datos que tiene en cuenta múltiples culturas, lo que puede hacer que la ubicación de marcadores faciales sea levemente distinta en rostros asiáticos respecto a rostros latinoamericanos.

Tabla 17. Resumen de la efectividad de cada ANN para clasificar las expresiones faciales de las emociones, correspondientes a combinaciones de bases de datos. Naranja: Promedios de exactitud de cada ANN. Naranja oscuro: Exactitud más alta de ANN. Azul: Promedio de exactitud de cada emoción. Azul oscuro: Exactitudes más altas de las emociones.

Emoción	KDEF	JAFFE	CAFE	KDEF + JAFFE	KDEF + CAFE	JAFFE + CAFE	KDEF + JAFFE + CAFE	Promedio	Máximo
Miedo	0.61	0.40	0.49	0.47	0.00	0.63	0.31	0.42	0.63
Enojo	0.00	0.71	0.67	0.39	0.28	0.00	0.35	0.42	0.71
Asco	0.74	0.90	0.00	0.61	0.68	0.21	0.60	0.43	0.90
Alegría	0.98	0.50	0.76	0.73	0.93	0.79	0.84	0.79	0.98
Neutral	0.84	0.33	0.86	0.91	0.83	0.88	0.79	0.78	0.91
Tristeza	0.68	0.71	0.00	0.08	0.10	0.18	0.18	0.28	0.71
Sorpresa	0.68	0.67	0.75	0.89	0.91	0.73	0.80	0.77	0.91
Promedio	0.65	0.60	0.50	0.58	0.53	0.49	0.55		

En cuanto a la selección del modelo de aprendizaje automático, primero se realizó el reentrenamiento de ResNet18 con los parámetros de entrenamiento que se observan en la Tabla 18. Estos parámetros fueron escogidos a partir de revisar investigaciones previas que recomiendan estos valores [90], [91]. El tiempo

de entrenamiento fue de 1 día, 7 horas, 27 minutos y 33 segundos. Dado que otros investigadores requerían el uso de la Jetson TX1, esta fue la única prueba que se realizó.

Tabla 18. Parámetros de entrenamiento para ResNet18.

Propiedad	**Descripción**	**Valor**
Número de capas épocas	Cantidad de ciclos a realizar para entrenar el modelo	120
Tamaño de lote	Número de entradas para tener en cuenta para el entrenamiento en cada iteración	50
Tasa de aprendizaje inicial	Tasa que indica la importancia dada al error obtenido en la anterior iteración	10^{-4}

Se calculó que la velocidad promedio de predicción de ResNet18 fue de 42ms, un tiempo aceptable para la predicción de expresiones faciales en tiempo real. Por otro lado, en la Tabla 19 se puede observar la matriz de confusión obtenida al validar la efectividad del algoritmo entrenado por medio de un conjunto de prueba, el cual se trata de un subconjunto de KDEF con una cantidad uniforme de imágenes por categoría. Los índices de las categorías son iguales a los utilizados en los modelos de ANN mostrados anteriormente.

Tabla 19. Matriz de confusión del algoritmo de ResNet18 reentrenado para reconocimiento de expresiones faciales.

	Clase predicha							**Exactitud por emoción**
Clase real	0	109	0	5	89	0	0	0.00
	0	109	0	12	75	0	0	0.56
	0	114	0	11	90	0	0	0.00
	0	94	0	33	97	0	0	0.15
	0	104	0	4	106	0	0	0.50
	0	122	0	9	90	0	0	0.00
	1	97	0	11	113	0	0	0.00
Promedio de exactitud								0.17

En la Tabla 19 se observa que este método no es viable para el reconocimiento de expresiones faciales, dado que la mayoría tienen una exactitud despreciable e incluso las exactitudes más efectivas correspondientes a la expresión neutra y al enojo, no cuentan con valores satisfactorios. Existen varias razones por las cuales estos resultados pudieron no ser efectivos. La primera de ellas es la cantidad de imágenes utilizadas para entrenar el modelo; como se mencionó anteriormente, los algoritmos de aprendizaje profundo requieren más muestras que aquellos de aprendizaje automático tradicional. Por otro lado, es posible que el uso de otras arquitecturas más profundas o una variación en los parámetros de entrenamiento mejoren la efectividad en el reconocimiento. Finalmente, es posible que la reducida variación en iluminación y pose en las imágenes afecten la capacidad de generalización del algoritmo, por lo cual, su uso en tiempo real se puede ver afectado negativamente. Es importante tener en cuenta esta herramienta en trabajos futuros, donde se cuente con herramientas y tiempo suficientes para entrenar una CNN con efectividad viable.

En cuanto al uso de otras técnicas de aprendizaje profundo, se tuvieron en cuenta los modelos K-NN, SVM, ANN y RF. En aras de brevedad, no se incluirán todas las matrices de confusión, pero si se incluirá la exactitud en la predicción de cada expresión facial, como se observa en la Tabla 20. La exactitud reportada de la expresión neutral es el promedio de la exactitud obtenida en el conjunto 1 con la obtenida en el conjunto 2. La exactitud para los casos con dos modelos se obtiene al combinarlos y analizar los puntajes obtenidos, tal como se describió anteriormente.

Tabla 20. Exactitud de cada arquitectura de aprendizaje automático al entrenar uno o dos modelos. Naranja: Exactitud promedio para cada arquitectura.

	K-NN		SVM		ANN		RF	
Emoción	**Un modelo**	**Dos modelos**	**Un modelo**	**Dos modelos**	**Un modelo**	**Dos modelos**	**Un modelo**	**Dos modelos**
Miedo	0.39	0.53	0.64	0.67	0.61	0.84	0.57	0.82
Enojo	0.40	0.35	0.64	0.69	0.00	0.77	0.68	0.58
Asco	0.38	0.62	0.64	0.47	0.74	0.72	0.78	0.87
Alegría	0.84	0.90	0.88	0.96	0.98	0.87	0.88	0.96
Neutral	0.50	0.85	0.82	0.85	0.84	0.85	0.91	0.91
Tristeza	0.16	0.41	0.51	0.56	0.68	0.73	0.59	0.71
Sorpresa	0.66	0.69	0.70	0.91	0.68	0.91	0.84	0.97
Promedio conjunto 1		0.69		0.78		0.83		0.85
Promedio conjunto 2		0.58		0.69		0.81		0.83
Promedio	0.48	0.63	0.69	0.73	0.65	0.82	0.75	0.84

En la Tabla 20 se observa que, para cada arquitectura analizada, la versión con dos modelos cuenta con una mayor exactitud que la versión con un modelo, lo que demuestra que este método de separar categorías es viable para obtener una mayor exactitud en la detección de emociones. De igual forma, la exactitud obtenida por ANN y RF es notoriamente mejor que la obtenida por otros métodos. Así, evaluó con mayor profundidad estas dos arquitecturas, analizando la exactitud de los conjuntos 1 y 2 por separado y modificando parámetros de estas arquitecturas, de forma que se mejore la exactitud obtenida por cada una.

A continuación, se mostrarán los resultados obtenidos en ambos conjuntos de etiquetas para varios modelos de ANN y RF. Al igual que en casos anteriores, se fija una semilla, de forma que todos los modelos de aprendizaje automático son entrenados con el mismo set de entrenamiento y son validados con el mismo set de prueba.

- Conjunto 1: Alegría, asco, enojo, sorpresa y expresión neutra

En el Anexo 4.1 se observa la exactitud obtenida por los modelos de ANN entrenados con imágenes del conjunto 1 al modificar el número de neuronas en la capa oculta, teniendo en cuenta que, para todos los casos, solo se utilizó una capa oculta. Se observa que la versión con 5 neuronas en la capa oculta obtuvo el mejor promedio de exactitud. De igual forma, para ninguna de las emociones la exactitud es menor a 75%. Considerando que se trata de un problema de 5 etiquetas, en el cual el azar tiene una exactitud del 20%, es un resultado satisfactorio. El caso con 2 neuronas clasifica con buena exactitud la mayoría de las emociones, pero no logra detectar enojo en ningún caso, por lo cual no es un clasificador robusto. Sucede un fenómeno similar en el caso de 100 neuronas, donde asco no es detectado correctamente en la mayoría de los casos. La exactitud promedio del caso de 20 neuronas también es satisfactorio; sin embargo, al utilizar más neuronas, pero el tiempo de cómputo para la predicción aumenta.

En el Anexo 4.2 se observa la efectividad de los RF al modificar la profundidad y número de los árboles del algoritmo, manteniendo un mínimo de dos muestras por nodo. Por su parte, en el Anexo 4.3 se observa la efectividad de los RF al modificar la profundidad y número de los árboles del algoritmo, manteniendo un mínimo de tres muestras por nodo. En estas dos tablas se observa un aspecto interesante sobre RF, y es que se trata de un método con efectividad bastante estable, independientemente de los parámetros que se cambien. Se puede observar que la exactitud de cada uno de los 18 modelos está entre 82% y 84%. De igual forma, la exactitud individual de cada emoción no tiene cambios significativos.

En general, se observa que el modelo con mayor efectividad para el conjunto 1 es la ANN con 5 neuronas en su capa oculta. Otras opciones viables son la ANN con 20 neuronas ocultas, y los RF con un mínimo de 3 muestras por nodo, un máximo de 50 árboles y una profundidad de árboles de 10 y 15 niveles. Se decidió cual opción entre estas es la más viable con base en el tiempo que tardan en predecir un resultado a partir de una imagen escogida al azar. Estas pruebas se realizaron en el computador con especificaciones técnicas mostradas en la Tabla 14, y sus resultados se observan en la Tabla 21. Para el caso de la ANN con 5 neuronas en la capa oculta, el tiempo de cómputo fue menor que la resolución mínima ofrecida por el compilador, de aproximadamente 1µs. Dado que esta arquitectura ofrece la mejor predicción y cuenta con tiempo de cómputo despreciable, es la elegida para predecir las emociones del conjunto 1. No obstante, es importante resaltar que el tiempo de cómputo de las cuatro arquitecturas es aceptable.

Tabla 21. Tiempo de cómputo de la predicción de las arquitecturas con mayor exactitud para el conjunto 1.

Arquitectura	**Tiempo de cómputo para la predicción (ms)**
ANN con 5 neuronas	0.00
ANN con 20 neuronas	0.99
RF con 10 niveles de profundidad	3.01
RF con 15 niveles de profundidad	4.97

- Conjunto 2: Miedo, tristeza y expresión neutra

Para el conjunto 2, se siguió el mismo procedimiento que en el conjunto 1. En el Anexo 4.4 se observa la exactitud obtenida por los algoritmos de ANN para clasificar las emociones de este conjunto. En este caso, el mejor resultado es obtenido por la red neuronal con 50 neuronas en la capa oculta; sin embargo, todas las ANNs tienen exactitudes similares, con una diferencia de tan solo 9% entre la mejor y la peor. Se puede ver que la emoción más difícil de reconocer es la tristeza y que hay una mayor facilidad cuando se detectan expresiones neutrales. Otro resultado satisfactorio es el obtenido con la ANNs con 100 neuronas en la capa oculta.

Tabla 22. Efectividad de las técnicas de ANN para clasificar las expresiones faciales del conjunto 2. Naranja: Exactitud promedio. Naranja Oscuro: Mejor exactitud promedio.

Número de neuronas	**2**	**3**	**5**	**10**	**20**	**50**	**100**
Neutral	0.73	0.88	0.76	0.80	0.76	0.85	0.83
Miedo	0.83	0.76	0.83	0.81	0.69	0.69	0.79
Tristeza	0.60	0.56	0.65	0.63	0.58	0.77	0.67
Promedio	0.72	0.73	0.75	0.75	0.68	0.77	0.76

En el Anexo 4.5 se observan los resultados de los árboles de decisión con un mínimo de dos muestras por nodo y en el Anexo 4.6 se observan los resultados cuando el número mínimo de muestras por nodo se cambia a tres. Al igual que en el caso anterior, la efectividad de RF es bastante estable, independientemente de cambios en sus parámetros. Aunque también existe una diferencia del 2% entre la exactitud del mejor y del peor modelo, en este caso, este valor es comparable con el de ANN. Al igual que con los modelos de ANN, hay una mayor dificultad para reconocer tristeza en los modelos de RF respecto a otras expresiones faciales. No obstante, en este caso, la detección de miedo se logra de manera más sencilla.

Al igual que en el caso anterior, se toma en consideración el tiempo de cómputo que requiere cada algoritmo para tomar la decisión final del algoritmo a utilizar, lo cual se puede observar en la Tabla 23. Dado que todos los algoritmos cuentan con tiempos de cómputo despreciables, se decidió seleccionar uno

de los modelos de RF, ya que son los que brindan una mayor exactitud. Particularmente, se escogió el modelo con 15 niveles de profundidad, un máximo de 50 árboles y un mínimo de 2 muestras por nodo.

Tabla 23. Tiempo de cómputo de la predicción de las arquitecturas con mayor exactitud para el conjunto 2.

Arquitectura	**Tiempo de cómputo para la predicción (ms)**
ANN con 50 neuronas	0.97
ANN con 100 neuronas	0.99
RF con 15 niveles de profundidad, un máximo de 50 árboles y un mínimo de 2 muestras por nodo	2.99
RF con 15 niveles de profundidad, un máximo de 100 árboles y un mínimo de 2 muestras por nodo	2.99
RF con 20 niveles de profundidad, un máximo de 50 árboles y un mínimo de 3 muestras por nodo	2.99

Finalmente, en la Tabla 24 se observa un resumen de los dos modelos de aprendizaje automático utilizados para reconocer expresiones faciales, considerando la separación de etiquetas en dos conjuntos.

Tabla 24. Resumen de los modelos de aprendizaje automático utilizados para reconocer expresiones faciales.

Modelo	**Parámetros**	**Exactitud**							
		Miedo	**Enojo**	**Asco**	**Alegría**	**Neutra**	**Tristeza**	**Sorpresa**	**Promedio**
ANN	1 capa oculta 5 neuronas		0.75	0.80	0.94	0.87		0.94	0.86
RF	2 muestras por nodo 15 niveles de profundidad 50 árboles	0.93				0.80	0.65		0.79

Para la reducción de características según su relevancia en el reconocimiento de expresiones faciales, a continuación, se muestran los resultados obtenidos para el conjunto 1. En la Tabla 25 se observa que la exactitud general se reduce únicamente un 2%, mientras que el tiempo de cómputo se redujo un 92.9% al reducir características. Esto implica que, a partir del uso de únicamente 18 características, es posible realizar el proceso de extracción de características a 1010.1FPS. Se considera que este aumento en la velocidad de procesamiento compensa la pérdida del 2% de la exactitud del algoritmo, por lo cual, únicamente se toman en cuenta estas características para el algoritmo final.

Tabla 25. Exactitud y tiempo de cómputo en la extracción de características para el modelo con todas las características y el modelo con las características más relevantes de las emociones del conjunto 1. Naranja: Exactitud promedio.

Emoción	**ANN entrenada con todas las características**	**ANN entrenada con las 18 características más relevantes**
Expresión neutra	0.87	0.79
Enojo	0.75	0.65
Asco	0.80	0.82
Alegría	0.94	0.96
Sorpresa	0.94	1.00
Promedio	0.86	0.84
Tiempo de extracción (ms)	13.93	0.99

Este proceso también se observa para el conjunto 2. En la Tabla 26 se observa que la exactitud general se reduce únicamente un 3%, mientras que el tiempo de cómputo se redujo un 42.9%, al utilizar únicamente las características más relevantes. Esto implica que, a partir del uso de únicamente 15 características, es posible realizar el proceso de extracción de características a 125.6FPS. Se considera que este aumento en la velocidad de procesamiento compensa la pérdida del 3% de la exactitud del algoritmo, por lo cual, únicamente se toman en cuenta estas características para el algoritmo final.

Tabla 26. Exactitud y tiempo de cómputo en la extracción de características para el modelo con todas las características y el modelo con las características más relevantes de las emociones del conjunto 2. Naranja: Exactitud promedio.

Emoción	**RF entrenado con todas las características**	**RF entrenado con las 15 características más relevantes**
Expresión neutra	0.80	0.78
Miedo	0.93	0.90
Tristeza	0.65	0.58
Promedio	0.79	0.76
Tiempo de extracción (ms)	13.93	7.96

Por último, se muestran los resultados obtenidos por el algoritmo final de reconocimiento de expresiones faciales. Para realizar la prueba, se predice la expresión facial de aquellas imágenes en KDEF que no fueron utilizadas para entrenar el algoritmo, evaluando su nivel de generalización. Es importante tener en cuenta que la exactitud en este caso es menor a la obtenida en pruebas anteriores, porque en este caso no se está evaluando de manera separada cada conjunto de expresiones faciales, sino que se realiza la predicción por medio de los dos modelos entrenados y se predice una expresión facial al escoger la emoción con mayor puntaje, a partir de las dos escogidas por los modelos de aprendizaje automático. En la Tabla 27 se muestra la matriz de confusión generada a partir de realizar este proceso, siguiendo los índices con los que se ha trabajado en pruebas anteriores.

Tabla 27. Matriz de confusión del set de prueba para el algoritmo final de reconocimiento de expresiones faciales.

		Clase predicha						**Exactitud por emoción**
Clase real	35	2	2	2	6	2	3	0.6731
	8	28	3	1	3	1	2	0.6087
	9	3	19	0	1	5	0	0.5135
	5	1	0	28	2	0	0	0.7778
	2	3	1	0	39	0	1	0.8478
	5	1	3	1	9	24	3	0.5217
	15	0	0	0	0	0	25	0.6250
Promedio de exactitud								0.6525

Esta matriz de confusión brinda resultados interesantes, dado que algunas emociones cuya exactitud era mejor en pruebas anteriores obtuvieron exactitudes menores. Un caso particularmente notorio es el de sorpresa, emoción cuya exactitud fue del 100% al evaluar únicamente la efectividad del modelo entrenado para el conjunto 1. Al combinar ambos modelos, la exactitud de esta categoría se redujo a 63%. La razón por la que se dio esto es porque varias imágenes fueron confundidas con miedo al comparar los puntajes obtenidos por ambos modelos de predicción. Visualmente, estos resultados tienen sentido, ya que miedo y sorpresa comparten varias características faciales, como apertura de ojos y boca.

Otro caso interesante es la expresión neutra, cuya exactitud aumentó respecto a los dos casos donde los modelos se entrenaron de manera individual. Dado que esta fue la única categoría presente en ambos modelos, tiene sentido que la exactitud aumente, porque cada modelo tiene en cuenta distintas

características, por lo cual hay un apoyo mutuo en la detección de esta expresión facial. Por ejemplo, si el conjunto 1 confunde la expresión neutra por otra expresión, pero el conjunto 2 detecta correctamente la expresión neutra, es probable que la decisión final sea escoger la expresión neutra.

Teniendo en cuenta que se trata de un problema con siete categorías, se considera que los resultados son satisfactorios. En un problema como este, la probabilidad del azar es del 14.29%, mientras que la expresión facial del asco, la cual fue aquella con mayor dificultad para ser reconocida, contó con una exactitud de 51.35%, lo cual la vuelve 3.59 veces mejor que el azar. Por su parte, la predicción de la expresión neutra, la cual es de 84.78%, es 5.93 veces mejor que el azar. Interesantemente, el modelo final obtenido es mejor que el mejor modelo de la Tabla 17, en la cual se analizan las imágenes de entrada a utilizar para el entrenamiento de los modelos de aprendizaje automático. Aunque ambos modelos cuentan con una exactitud final del 65%, el modelo final logra tener una exactitud aceptable para todas las categorías, mientras que el algoritmo inicial de KDEF no detectaba enojo correctamente. Esto implica que las etapas adicionales realizadas en este proyecto fueron efectivas, dado que la separación de categorías y la eliminación de características lograron mejorar la exactitud y tiempo de cómputo, respectivamente.

No obstante, en un trabajo futuro se puede buscar mejorar la efectividad de este algoritmo. Por un lado, se puede profundizar más en el uso de técnicas de aprendizaje profundo para el reconocimiento de expresiones faciales, dado que el proyecto no se enfocó en este tema. Así, se puede buscar hacer uso de bases de datos como FER-2013 u obtener imágenes de uso libre en distintos motores de búsqueda, haciendo énfasis en que las personas de las imágenes se encuentren en distintas poses y con distintas iluminaciones, para aumentar la generalización del algoritmo generado. También se puede explorar el uso de arquitecturas de CNN como AlexNet, ResNet y GoogLeNet, evaluando aquellas que den mayor efectividad.

En cuanto a la exploración de técnicas de aprendizaje automático tradicional, se pueden realizar muchas más pruebas en la etapa de extracción de características, dado que existen otras opciones para caracterizar un rostro más allá de los marcadores faciales y de los gradientes orientados. Se puede evaluar el uso de técnicas como la transformada Wavelet y la matriz de co-ocurrencias. No obstante, estas técnicas no fueron utilizadas en este proyecto por su difícil implementación y sus altos tiempos de cómputo, por lo cual sería necesario explorar cómo implementarlas de manera óptima.

También, teniendo en cuenta la matriz de confusión obtenida, se puede considerar para un trabajo futuro la creación de distintos conjuntos de emociones, haciendo énfasis en aquellas que se confunden entre si normalmente, como es el caso del miedo y la sorpresa o el miedo y el asco. Adicionalmente, teniendo en cuenta que el método de selección de modelos causó falsos negativos en el reconocimiento de la sorpresa, es importante reevaluar cómo se interpreta el puntaje brindado por los modelos de aprendizaje profundo.

A continuación, en la Tabla 28, se observan los tiempos de cómputo de cada uno de los subprocesos para lograr el reconocimiento de expresiones faciales, los cuales fueron evaluados con el computador con las especificaciones técnicas indicadas en la Tabla 14. Cada valor incluye el tiempo necesario para realizar una etapa de preprocesamiento; por ejemplo, para ubicar marcadores faciales, se incluye la ecualización de histograma realizada a la imagen en grises. De igual forma, los tiempos indicados son calculados a partir del promedio de 100 registros en tiempo real con una persona realizando expresión neutra, dado que cambios como la adición de otros rostros pueden alterar estos valores.

Tabla 28. Tiempo de cómputo para cada una de las etapas del algoritmo de reconocimiento de expresiones faciales.

Etapa	**Tiempo de cómputo (ms)**
Adquisición de imagen	0.273±0.054
Detección de rostro	40.784±3.755
Ubicación de marcadores faciales	5.075±0.559

Extracción de características relacionadas con marcadores faciales	0.868±0.396
Predicción de expresión facial para conjunto 1 de emociones	0.101±0.068
Extracción de características relacionadas con HOG	12.519±1.631
Predicción de expresión facial para conjunto 2 de emociones	0.504±0.092
Proceso completo	**56.685±4.541**

En la tabla se puede observar que el mayor tiempo de cómputo se realiza en tres procesos principalmente: la detección de rostros, la extracción de características relacionadas con HOG y la ubicación de marcadores faciales. Una forma de reducir tiempo de cómputo puede ser utilizando otras técnicas para detectar rostros que tomen menos tiempo para realizar este proceso. No obstante, como se observó en la subsección 4.2.3, a conocimiento de nuestro equipo de investigación, no existen métodos de fácil implementación con robustez similar a la de SSD. MMOD cuenta con una exactitud similar; sin embargo, su implementación no es viable, dado que los tiempos de cómputo son mucho mayores.

Por otro lado, se podría mejorar el tiempo de cómputo al utilizar otro método de ubicación de marcadores faciales. Dado que este proceso es esencial para el correcto reconocimiento de expresiones faciales, no es recomendable el uso del algoritmo LBF, ya que se observó en la subsección 4.2.4 que su exactitud no es adecuada. Sin embargo, haciendo uso de las bases de datos mencionadas en la subsección 3.2.3, sería posible crear un modelo de ubicación de marcadores faciales, buscando que la exactitud no disminuya, pero reduciendo tiempos de cómputo.

Finalmente, para reducir el tiempo de cómputo en la etapa de extracción de características relacionadas con HOG, se puede intentar reducir la resolución de la imagen de entrada para que el cómputo de los gradientes orientados no requiera de tantos cálculos. Igualmente, se puede evaluar la importancia de las tres zonas escogidas para analizar el gradiente del rostro, dado que, como se mostró anteriormente, ninguna característica de HOG fue relevante para el reconocimiento de 5 de las 7 expresiones faciales utilizadas en este proyecto después de realizar la reducción de características.

No obstante, el tiempo de cómputo obtenido es aceptable para el reconocimiento en tiempo real, porque la velocidad de reconocimiento es de 17.64FPS. Dado que, en ambientes naturales, un humano no cambia su expresión facial tan rápidamente, el tiempo de cómputo obtenido es viable. Sin embargo, si en una investigación futura se deseara evaluar una microexpresión en tiempo real, es necesario que este tiempo de cómputo sea mayor. Las microexpresiones son expresiones faciales involuntarias que indican la emoción real que está sintiendo una persona, así intente disimularla más adelante. El grupo de investigación de Ekman indica que una microexpresión puede durar entre 0.04s y 0.067s [92]. Por lo cual se necesitaría un sistema que logre reconocer microexpresiones a 25FPS como mínimo.

5.2 Aspectos psicométricos

5.2.1 Medidas psicométricas

La medida psicométrica utilizada, ENI, se registró antes y después de la aplicación del protocolo experimental. Los detalles de esta prueba psicométrica se exponen en la subsección 4.4.1. En la Tabla 29 se observan los resultados de la prueba ENI para el sujeto 1, participante con TEA.

Tabla 29. Resultados de la prueba ENI para el sujeto 1.

Emoción	**Pre-intervención**		**Pos-intervención**	
	Respuesta	**Puntaje**	**Respuesta**	**Puntaje**

1. Alegría (niña).	Feliz	1	Alegría	1
2. Enojo (niño).	Bravo	1	Bravo	1
3. Tristeza (niño).	Serio	0	Serio	0
4. Enojo (niña).	Bravo	1	Brava	1
5. Alegría (niño).	Feliz	1	Feliz	1
6. Tristeza (niña).	Seria	0	Tristeza	1
7. Miedo o asombro (niño).	Sorprendida	1	Asustado	1
8. Miedo o asombro (niña).	Sorprendido	1	Miedo	1
Total (8)	**6**		**7**	

A partir de la Tabla 29 se observa que el puntaje del sujeto 1 mejoró en un 12.5%. Principalmente, se observa que una diferencia clave se presenta en la pregunta 6, ya que inicialmente el participante había contestado que la niña con rostro de tristeza estaba seria y después de la intervención cambió su respuesta a tristeza. De igual forma, se observó que el participante cambió su respuesta en la pregunta 1, de «feliz» a «alegría», indicando que los términos utilizados durante la intervención pudieron tener un efecto duradero en él. No obstante, este mismo efecto no se observó en las preguntas 2-5, en las que utilizó la misma respuesta inicial. No obstante, dado que, por ejemplo, «bravo» se considera un sinónimo de «enojado», se obtuvieron resultados positivos en algunas de estas preguntas. Otro resultado interesante se observa en las preguntas 7 y 8, el participante modificó las respuestas originales de «sorpresa» por «miedo». Aunque ambas respuestas son válidas según las normas establecidas del ENI, en estas imágenes se observa que las descripciones faciales establecidas por Ekman concuerdan más con el miedo que con la sorpresa, lo cual puede indicar un resultado favorable en el reconocimiento de emociones. Las imágenes en cuestión no se muestran en este documento, dado que su publicación no está permitida. En general, el participante mostró un reconocimiento de emociones más rápido después de la intervención, con dificultad en el reconocimiento de la tristeza.

En la Tabla 30 se observan los resultados de la prueba ENI para el sujeto 2, participante neurotípico actualmente desescolarizado. En este caso, también se observa una mejora del 12.5% del puntaje obtenido para ENI. Es interesante observar que se modificó de manera positiva la respuesta a dos preguntas, pero se modificó de manera negativa la respuesta de otra. No obstante, es claro que la intervención realizada tuvo un efecto en el participante, evidenciado por los términos utilizados al describir las fotografías presentadas. En la aplicación de ENI de manera previa a la intervención, el participante utilizó términos como «pensativo» y «aburrido», términos que no son propios de las seis emociones básicas y que no se mencionaron a lo largo de la intervención. En la prueba posterior a la intervención se observa que, sin contar por el término «bravo», todos los términos utilizados por el participante se mencionan durante la intervención. Particularmente, se observa que el participante cambió los términos «pensativo» y «aburrido» por «tristeza», lo que permitió que aumentara su puntaje en la prueba. Los resultados indican una mejora en su velocidad de reconocimiento; sin embargo, se evidencia dificultad en el reconocimiento de la emoción de enojo, debido a que su tiempo de respuesta es más prolongado. Así mismo se observa que la dificultad al reconocer la emoción de miedo persiste.

Tabla 30. Resultados de la prueba ENI para el sujeto 2.

Emoción	**Pre-intervención**		**Pos-intervención**	
	Respuesta	**Puntaje**	**Respuesta**	**Puntaje**
1. Alegría (niña).	Feliz	1	Feliz	1
2. Enojo (niño).	Bravo	1	Bravo	1
3. Tristeza (niño).	Pensativo	0	Tristeza	1
4. Enojo (niña).	Bravo	1	Asco	0

5. Alegría (niño).	Feliz	1	Feliz	1
6. Tristeza (niña).	Aburrido	0	Tristeza	1
7. Miedo o asombro (niño).	Sorprendido	1	Sorprendido	1
8. Miedo o asombro (niña).	No sabía	0	No sabía	0
Total (8)	**5**		**6**	

No obstante, en la pregunta 4 se observa que el término «bravo», etiquetado correctamente, fue cambiado por el término «asco», lo que redujo su puntaje. En la imagen, se observa que, aunque el rostro correspondiente demuestra rasgos propios del enojo, algunas características, como el pronunciamiento de las líneas nasolabiales, son similares a aquellas del asco, razón por la cual se pudo generar esta confusión. Al igual que con el caso anterior, la publicación de esta imagen no está permitida.

En la Tabla 31 se observan los resultados de la prueba ENI para el sujeto 3, participante neurotípico actualmente cursando primero de primaria. Un aspecto importante para tener en cuenta para los resultados del sujeto 3 es que el día en que se realizaron las últimas tres sesiones de la etapa de reconocimiento fue su primer día de clases en el colegio después de vacaciones de mitad de año. El participante mostró estar cansado durante este día e indicó que se había sentido confundido porque no sabía en qué salón debería estar. Puede que esto haya afectado los resultados de estas sesiones, dado el estado emocional en el que se encontraba. Se observa que, en este caso, se generó una disminución del 25% del puntaje obtenido para ENI. Considerando los resultados de las dos aplicaciones se observa que el sujeto 3 alcanza un puntaje esperado de acuerdo con su edad y nivel de escolaridad. Durante la primera aplicación, se evidencia dificultad para reconocer la expresión facial del miedo. No obstante, a partir de la segunda aplicación se identifica que se presenta dificultad para reconocer la expresión facial de tristeza. Así mismo, persiste en la expresión facial del miedo. Por lo tanto, se infiere que esta afectación puede estar relacionada a la fatiga física y cognitiva que presenta el participante al momento de la aplicación.

El manual de ENI indica que el puntaje final del sujeto 1 lo ubica en el percentil 75, el puntaje final del sujeto 2 lo ubica en el percentil 50 y el puntaje final del sujeto 3 lo ubica en el percentil 37. No obstante, todos los participantes son considerados dentro de los valores de reconocimiento de expresiones faciales promedio según su edad [49].

Tabla 31. Resultados de la prueba ENI para el sujeto 3.

Emoción	**Pre-intervención**		**Pos-intervención**	
	Respuesta	**Puntaje**	**Respuesta**	**Puntaje**
1. Alegría (niña).	Feliz	1	Alegría	1
2. Enojo (niño).	Furioso	1	Furioso	1
3. Tristeza (niño).	Triste	1	Asco	0
4. Enojo (niña).	Furioso	1	Furioso	1
5. Alegría (niño).	Alegre	1	Alegre	1
6. Tristeza (niña).	Triste	1	Asco	0
7. Miedo o asombro (niño).	Asombrado	1	Sorpresa	1
8. Miedo o asombro (niña).	Asco	0	Asco	0
Total (8)	**7**		**5**	

5.2.2 Registros conductuales

Los registros conductuales llevados a cabo durante la totalidad de las sesiones realizadas, observados en el Anexo 5, se componen de 13 documentos por participante, uno por cada actividad en la que se utilizó el algoritmo de reconocimiento de expresiones faciales. La única excepción para esto fue la actividad *Foto*

en Vivo, porque, en este caso, el sistema no busca reconocer en tiempo real la expresión facial de los participantes, sino que detecta la expresión facial que se tiene durante un momento específico. En este documento se resumen los 13 registros conductuales de cada participante a partir de tablas, donde se buscó incluir la información más relevante de estos y se estandarizaron las observaciones realizadas por la estudiante que los diligenció. Los espacios en gris de estas tablas indican que, para una sesión dada, los participantes no imitaron expresiones faciales de algunas emociones, de acuerdo con el protocolo experimental.

En cada una de las siguientes tablas expuestas, **R.I.** es el retardo de imitación; una medida que indica el tiempo, en segundos, desde el instante que se pidió al participante imitar la emoción hasta el instante en que empezó a imitarla. **D.I.** es la duración de imitación; una medida que señala el tiempo, en segundos, desde el instante en que el participante empezó a imitar la emoción hasta el instante en que Emmaciones le indicó que lo hizo correctamente. Finalmente, **Obs.** Son observaciones realizadas por la estudiante de psicología. Más adelante en el documento se realizará un análisis comparativo de los resultados obtenidos para cada sujeto en los registros conductuales.

Un resultado importante de esta investigación es que, independientemente de la emoción o de la actividad planteada, todos los participantes lograron imitar todas las expresiones faciales, lo que da una exactitud del 100%. Las actividades están diseñadas para no avanzar hasta que los participantes logren imitar cada expresión facial; no obstante, en ningún momento durante las pruebas fue necesario parar una actividad y continuarla más adelante, lo que indica que el algoritmo de reconocimiento de expresiones faciales es suficientemente efectivo para ser utilizado adecuadamente en tiempo real.

Para los resultados del participante 1, es importante tener en cuenta que éste no contaba con mucha ayuda técnica, porque, de manera previa a las pruebas realizadas, nadie en su casa conocía lo suficiente de computación para realizar acciones como compartir pantalla y prender la cámara. Así, es importante resaltar que los resultados pueden estar afectados por el nivel de frustración del participante, ya que las primeras sesiones tomaron más tiempo del estimado. Como ejemplo, la primera sesión con el sujeto 1 duró 30 minutos más, aproximadamente, dado que fue necesario dar instrucciones remotas para que el participante compartiera su pantalla y ejecutara el juego. En la Tabla 32 se observa el resumen de los registros conductuales correspondientes al sujeto 1 para las expresiones faciales de la alegría, el miedo y el asco. A partir de las observaciones realizadas en el registro conductual, es interesante notar que, durante el juego *Ruleta*, correspondiente a la sesión 2 de imitación, se presentaron varios errores conceptuales en la expresión de las emociones, los cuales no se observaron en sesiones posteriores. Esto puede implicar que el participante aprendió de sus errores a partir de la realimentación brindada en el juego, lo que mejoró su habilidad de imitación.

En cuanto al juego *Caja Sorpresa*, correspondiente a la sesión 4 de imitación, se detallaron errores de otra índole. Por un lado, se observó que el participante se distrajo al imitar la alegría; este comportamiento se repite de manera recurrente durante la imitación de la tristeza, como se observa más adelante. También, durante la segunda sesión, al participante le costó imitar la expresión facial del miedo, dado que fue necesario realizar múltiples intentos, reflejado en la duración de imitación.

Tabla 32. Resumen de los registros conductuales aplicados en las pruebas del sujeto 1 para la alegría, el miedo y el asco.

Actividad		**Alegría**	**Miedo**	**Asco**
Ruleta	**R.I. (s)**	1	2	1
	D.I. (s)	2	13	3
	Obs.	No se observaban líneas nasolabiales.	No abría mucho los ojos.	No se observaban líneas nasolabiales.
Caja sorpresa	**R.I. (s)**	1	1	1
	D.I. (s)	3	21	2
	Obs.	Se distraía.	Múltiples intentos.	
	R.I. (s)	0	0	4

Emma dice	**D.I. (s)**	3	4	3
	Obs.			
Lotería	**R.I. (s)**	0	0	0
	D.I. (s)	2	3	3
	Obs.			
Ordena la cara	**R.I. (s)**	1	1	0
	D.I. (s)	1	1	1
	Obs.			
Pop Emma	**R.I. (s)**	1	0	0
	D.I. (s)	2	9	4
	Obs.			
Encuentra el par	**R.I. (s)**	0	0	0
	D.I. (s)	1	4	1
	Obs.			
Laberinto	**R.I. (s)**	0	0	0
	D.I. (s)	1	4	2
	Obs.			
Espejo	**R.I. (s)**	0	1	1
	D.I. (s)	3	9	1
	Obs.			

En la Tabla 33 se observa resumen de los registros conductuales correspondientes al sujeto 1 para las expresiones faciales de la tristeza, la sorpresa y el enojo. A partir de las observaciones realizadas, se puede notar que estas expresiones faciales demostraron ser más difíciles de imitar para el sujeto 1, especialmente la tristeza. Para esta emoción, es claro que al participante se le dificultaba realizar los movimientos faciales necesarios; particularmente, fruncir el ceño y curvar la boca. Como observación personal, este problema aumentaba progresivamente: dado que se dificultó la imitación de la tristeza durante las primeras sesiones, el participante mostraba frustración durante la imitación de esta emoción en sesiones posteriores, reflejado en un bajo nivel de atención durante las actividades. Esto se corrobora por observaciones realizadas por la estudiante de psicología, quien indicó que durante los juegos *Caja Sorpresa* y *Ordena la Cara*, el participante se movía demasiado y no mantenía el gesto de la emoción. Una observación interesante es aquella de *Laberinto*, la cual indica que, cuando el participante realizó el gesto de manera correcta, tal como le decía Emma, el sistema reconoció su expresión facial. Para el caso de la sorpresa, se observó un patrón similar al de las emociones expuestas en la Tabla 32; ya que, durante las primeras sesiones, se presentaron dificultades en el movimiento de los músculos faciales, las cuales no se notaron en sesiones posteriores. Por su parte, las observaciones del enojo no reflejan una mayor dificultad en la imitación de esta expresión facial, exceptuando problemas técnicos causados durante la primera sesión, donde no se observaba la totalidad del rostro, problema que se arregló durante sesiones siguientes al pedirle al participante que se recogiera el cabello.

Tabla 33. Resumen de los registros conductuales aplicados en las pruebas del sujeto 1 para la tristeza, la sorpresa y el enojo.

Actividad		**Tristeza**	**Sorpresa**	**Enojo**
Ruleta	**R.I. (s)**	1	1	1
	D.I. (s)	16	3	38
	Obs.	No fruncía el ceño.		No se observaban las cejas debido al cabello.

Caja sorpresa	**R.I. (s)**	1	0	0
	D.I. (s)	25	9	5
	Obs.	Se movía mucho, la cámara no lo reconocía.	No se veía su frente debido al cabello.	
Emma dice	**R.I. (s)**	0	1	1
	D.I. (s)	48	3	22
	Obs.		Poca pronunciación en las arrugas de la frente.	
Lotería	**R.I. (s)**	1	0	0
	D.I. (s)	32	2	9
	Obs.		Poca pronunciación en las arrugas de la frente.	
Ordena la cara	**R.I. (s)**	1	1	1
	D.I. (s)	64	3	8
	Obs.	No mantenía el gesto.		
Pop Emma	**R.I. (s)**	1	1	1
	D.I. (s)	30	1	1
	Obs.	Imitaba la pose de la imagen, no la emoción.		
Encuentra el par	**R.I. (s)**	1	0	1
	D.I. (s)	30	1	5
	Obs.	No curvaba la boca ni fruncía el ceño.		
Laberinto	**R.I. (s)**	0	1	0
	D.I. (s)	19	6	5
	Obs.	El sistema lo detectó cuando curvó la boca.		
Espejo	**R.I. (s)**	1	1	0
	D.I. (s)	46	2	2
	Obs.	Se le dificultaba curvar la boca. No se veía su frente.		

Una mejor visualización de los datos temporales de la Tabla 32 y la Tabla 33 se ubica en la Figura 49. A partir de esta figura, se observa que existe una diferencia importante en la dificultad de imitación de expresiones faciales para el sujeto 1. Por un lado, emociones como la alegría, el asco y la sorpresa obtuvieron tiempo de reconocimiento consistentemente corto, en comparación con aquel de emociones como el enojo y la tristeza. Aunque no es posible saber si la duración de la imitación depende más del participante o del algoritmo, hay una relación entre las observaciones y la duración para algunas emociones. El caso más claro es el de miedo, donde aquellos juegos donde se realizaron observaciones negativas respecto a la habilidad de imitación del participante fueron aquellas cuya duración de imitación fue mayor. La tristeza demostró ser una emoción consistentemente difícil de imitar para el participante, dado que no se observa un patrón claro en la duración hasta o más de un minuto.

El enojo, por su parte, también mostró una alta dificultad de imitación; llegando incluso a una duración de 38 segundos en la primera sesión. No obstante, este caso particular puede estar afectado por la baja visualización de las cejas, como se indica en las observaciones. A diferencia de la tristeza, en el enojo se observa un patrón en la duración de imitación, ya que, sin contar las actividades *Caja Sorpresa* y *Pop Emma*, cuyo tiempo de imitación fue considerablemente bajo con relación a las demás actividades, cada tiempo de imitación fue menor al de la sesión anterior. Aunque sería relevante contar con más datos para

realizar un análisis adecuado, se puede considerar que el participante logró mejorar su imitación de expresiones faciales, teniendo que en cuenta que el retardo en la imitación es similar en todas las sesiones. Finalmente, dados los valores bajos en duración de la imitación para las emociones de alegría, asco y sorpresa, en comparación con la resolución, no se puede indicar con certeza si se lograron mejoras en la imitación. No obstante, los valores de todas las sesiones fueron considerablemente bajos para estas expresiones faciales, por lo cual se considera que el participante logra imitar estas emociones con consistencia. Finalmente, se observa que el retardo en iniciar la imitación de expresiones faciales es bastante consistente para este participante, porque, en la gran mayoría de casos, inicia la imitación con un segundo o menos de retardo, siendo la imitación de asco en *Emma Dice* la excepción.

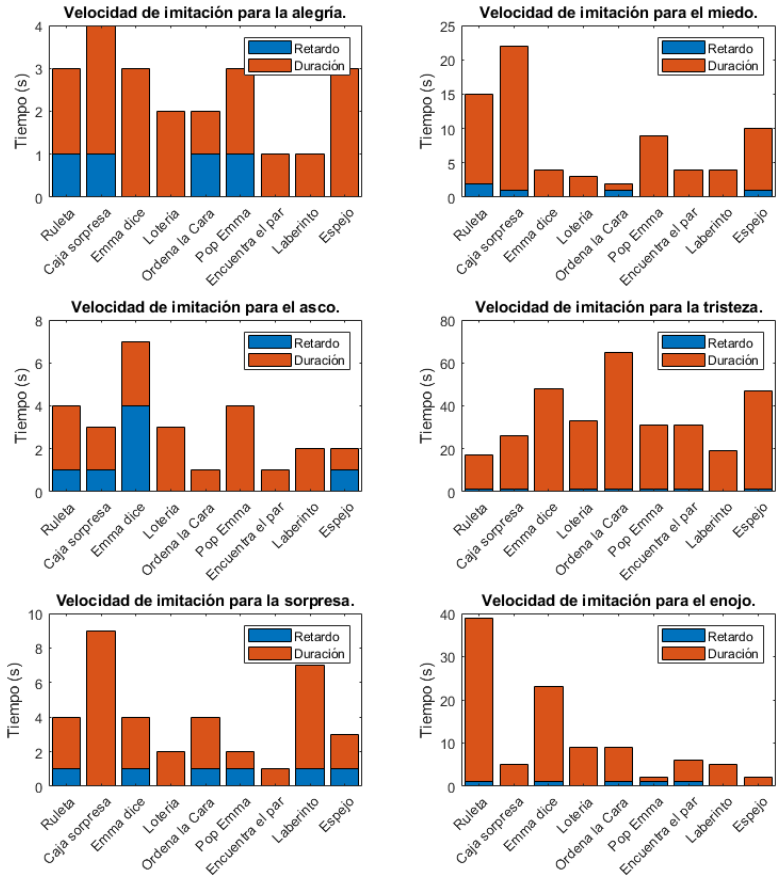

Figura 49. Resultados temporales del registro conductual del sujeto 1.

En la Tabla 34 se observa un resumen de los registros conductuales correspondientes a las expresiones faciales de la alegría, el miedo y el asco para el sujeto 2. Al igual que para el sujeto 1, la cantidad de observaciones para estas tres emociones es baja, con dos observaciones en asco y una en miedo. En interesante que las observaciones indican que en algunas sesiones el participante no frunció el ceño durante la imitación de asco, mas no se trató patrón recurrente. En cuanto al miedo, se observó que el participante no abría los ojos los suficiente durante la actividad *Encuentra el Par*, lo que también se notó para el participante 1 con esta misma emoción. Las observaciones para este participante no indican cambios particularmente positivos, dado que solo se realizaron observaciones en dos actividades de la etapa de reconocimiento. Sin embargo, teniendo en cuenta que las actividades que se realizaron observaciones tuvieron lugar en el mismo día, puede que estas deficiencias sean producto del estado emocional del participante.

Tabla 34. Resumen de los registros conductuales aplicados en las pruebas del sujeto 2 para la alegría, el miedo y el asco.

Actividad		Alegría	Miedo	Asco
Ruleta	**R.I. (s)**	3	3	4
	D.I. (s)	2	2	2
	Obs.			
Caja sorpresa	**R.I. (s)**	3	5	3
	D.I. (s)	4	11	5
	Obs.			
Emma dice	**R.I. (s)**	2	2	3
	D.I. (s)	4	5	5
	Obs.			
Lotería	**R.I. (s)**	1	2	1
	D.I. (s)	2	7	2
	Obs.			
Ordena la cara	**R.I. (s)**	0	2	1
	D.I. (s)	2	9	1
	Obs.			
Pop Emma	**R.I. (s)**	1	2	2
	D.I. (s)	17	5	4
	Obs.			No fruncía el ceño recurrentemente.
Encuentra el par	**R.I. (s)**	1	1	1
	D.I. (s)	1	5	1
	Obs.		No abría los ojos lo suficiente.	No fruncía el ceño.
Laberinto	**R.I. (s)**	1	1	1
	D.I. (s)	1	3	1
	Obs.			
Espejo	**R.I. (s)**	0	1	0
	D.I. (s)	1	4	2
	Obs.			

En la Tabla 35 se observa el resumen de los registros conductuales correspondientes a las expresiones faciales de tristeza, sorpresa y enojo para el sujeto 2. Se puede ver que el participante tuvo más problemas en la imitación de estas emociones, a partir de las observaciones realizadas. Esto es particularmente cierto para las emociones de tristeza, donde se le dificultaba fruncir el ceño y mantener la boca curvada, y el enojo, fruncir el ceño es difícil para el participante. A partir de las observaciones, se puede inferir que el niño se frustraba con los resultados obtenidos, porque se movía mucho cuando imitaba la sorpresa y se cansaba de imitar el enojo. No obstante, también es posible que esto se deba a un bajo nivel de atención propio del participante.

Tabla 35. Resumen de los registros conductuales aplicados en las pruebas del sujeto 2 para la tristeza, la sorpresa y el enojo.

Actividad	Tristeza	Sorpresa	Enojo

Ruleta	**R.I. (s)**	3	10	1
	D.I. (s)	25	12	51
	Obs.	Movía demasiado la boca.	Movía la boca de manera incorrecta.	Movía demasiado la boca.
Caja sorpresa	**R.I. (s)**	1	1	1
	D.I. (s)	9	2	6
	Obs.			Movía demasiado la boca
Emma dice	**R.I. (s)**	3	1	2
	D.I. (s)	12	3	8
	Obs.	No fruncía el ceño ni mantenía la boca curvada.	Se movía demasiado.	No fruncía el ceño. Se cansaba de imitar.
Lotería	**R.I. (s)**	2	2	2
	D.I. (s)	11	5	82
	Obs.	No fruncía el ceño ni mantenía la boca curvada.		No fruncía el ceño. Se cansaba de imitar.
Ordena la cara	**R.I. (s)**	2	1	1
	D.I. (s)	3	2	407
	Obs.	No fruncía el ceño.		No fruncía el ceño. Se cansaba de imitar.
Pop Emma	**R.I. (s)**	1	0	1
	D.I. (s)	18	9	12
	Obs.			No fruncía el ceño. Mueve mucho la boca.
Encuentra el par	**R.I. (s)**	0	1	2
	D.I. (s)	2	2	63
	Obs.		No se veía su frente debido al cabello.	No fruncía el ceño. Se cansaba de imitar.
Laberinto	**R.I. (s)**	1	1	0
	D.I. (s)	6	2	52
	Obs.	No fruncía el ceño.	No se veía su frente debido al cabello.	No fruncía el ceño.
Espejo	**R.I. (s)**	4	0	1
	D.I. (s)	31	2	31
	Obs.			No fruncía el ceño.

Una mejor visualización de los datos temporales de la Tabla 34 y la Tabla 35 se observa en la Figura 50. La figura indica que este participante no logró imitar de manera consistente las expresiones faciales, dada una mayor variabilidad en la duración de imitación de todas las expresiones faciales respecto a los resultados del sujeto 1. No obstante, se observa que, sin contar notables excepciones, la duración de imitación para la alegría, el asco y la sorpresa son menores respecto a las demás emociones. La imitación del enojo fue excepcionalmente difícil para este participante, dado que en 55.5% de las actividades, se demoró más de 50 segundos en imitar esta emoción, sin que se observara una clara mejora. Al igual que el participante 1, tuvo dificultad en imitar la tristeza. Sin embargo, en este caso, los malos resultados en la imitación de estas dos emociones son coherentes con las observaciones realizadas: la tristeza y el enojo son dos emociones cuya expresión facial depende en gran parte del ceño fruncido, el cual no es fácilmente evocado por este participante. Por su parte, la duración de imitación del miedo aumentó a lo largo de la etapa de imitación y disminuyó nuevamente a lo largo de la etapa de reconocimiento. Al igual que en el

caso del participante 1, este comportamiento puede ser causado por múltiples motivos, incluyendo el estado emocional del participante durante la aplicación de la prueba. Por tanto, las observaciones y la alta variación en la duración de imitación indican que, en este caso en particular, esta inconsistencia se puede deber a un bajo nivel de atención por parte del participante. Esto es debido a que existen retardos excepcionales en la imitación de emociones, con un promedio de 1.7 segundos en imitar la emoción y un caso excepcional de 10 segundos.

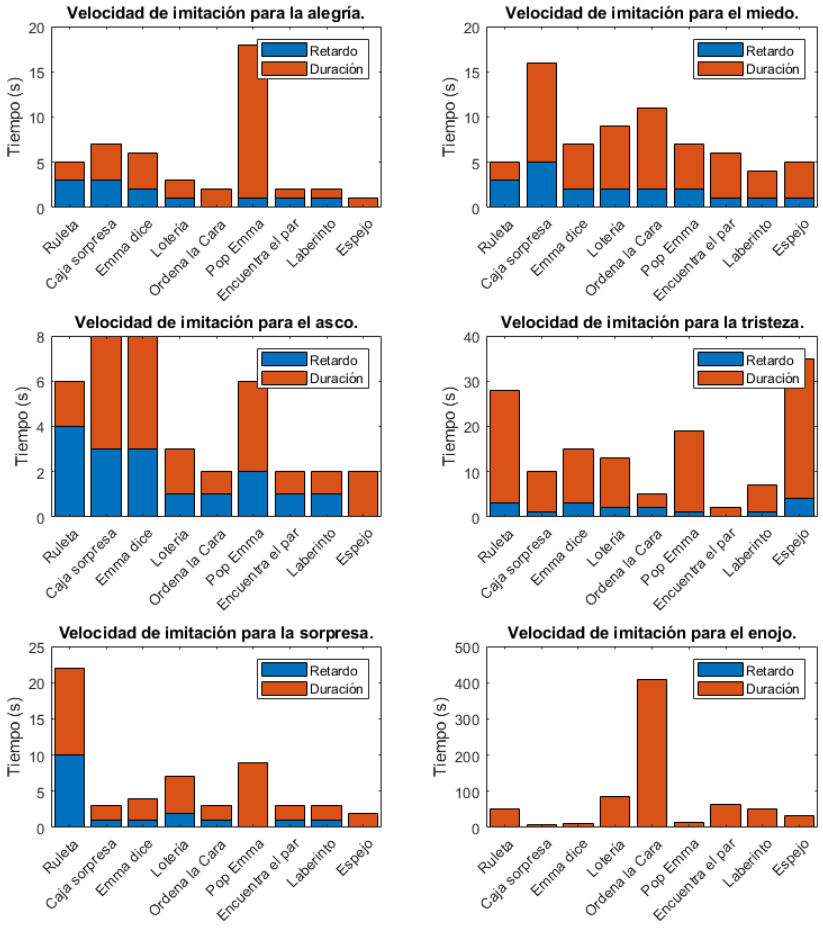

Figura 50. Resultados temporales del registro conductual del sujeto 2.

En la Tabla 36 se muestra el resumen de los registros conductuales para la alegría, el miedo y el asco, correspondientes al sujeto 3. En este caso se realizaron observaciones durante la primera etapa de imitación, más no en las siguientes, lo que puede indicar que los resultados del participante para esta primera sesión se debían a que estaba acostumbrándose al funcionamiento de la interfaz. Principalmente, esto puede ser cierto dado que las observaciones realizadas, para alegría y miedo, no indicaban errores en la imitación de las expresiones faciales, sino errores posturales. Como se señala por la falta de observaciones y, como se mostrará más adelante, en una reducción en la duración de imitación, se puede concluir que el participante fue capaz de asemejar lo que Emmaciones requería de él.

Tabla 36. Resumen de los registros conductuales aplicados en las pruebas del sujeto 3 para la alegría, el miedo y el asco.

Actividad		**Alegría**	**Miedo**	**Asco**
Ruleta	**R.I. (s)**	1	3	2
	D.I. (s)	15	13	7

	Obs.	Giraba la cabeza.	No se ubicaba correctamente.	
Caja sorpresa	**R.I. (s)**	1	1	1
	D.I. (s)	2	13	8
	Obs.			
Emma dice	**R.I. (s)**	4	1	1
	D.I. (s)	3	7	2
	Obs.			
Lotería	**R.I. (s)**	1	2	1
	D.I. (s)	1	2	2
	Obs.			
Ordena la cara	**R.I. (s)**	1	1	1
	D.I. (s)	3	3	2
	Obs.			
Pop Emma	**R.I. (s)**	1	1	1
	D.I. (s)	1	1	2
	Obs.			
Encuentra el par	**R.I. (s)**	1	1	2
	D.I. (s)	1	6	7
	Obs.			
Laberinto	**R.I. (s)**	0	3	1
	D.I. (s)	1	2	1
	Obs.			
Espejo	**R.I. (s)**	0	1	1
	D.I. (s)	6	6	3
	Obs.			

En la Tabla 37 se muestra el resumen de los registros conductuales para la tristeza, la sorpresa y el enojo, correspondientes al sujeto 3. Al igual que en el caso de los otros dos participantes, se puede ver que, a partir de las observaciones realizadas, hay una mayor dificultad para imitar tristeza y enojo respecto a las demás emociones. No obstante, en este caso se observaron pocas fallas en la imitación de la expresión facial del participante, con fallas principalmente del tipo postural. Una excepción a esto es la imitación de enojo durante la primera actividad de la etapa de imitación, donde se observó que el participante no fruncía el ceño. Dado que esto no se observa en actividades posteriores, es posible indicar que se logró una mejora en la imitación de esta emoción para este participante en particular.

Tabla 37. Resumen de los registros conductuales aplicados en las pruebas del sujeto 3 para la tristeza, la sorpresa y el enojo.

Actividad		**Tristeza**	**Sorpresa**	**Enojo**
Ruleta	**R.I. (s)**	0	0	1
	D.I. (s)	6	3	29
	Obs.			No levantaba la cara. No fruncía el ceño.
Caja sorpresa	**R.I. (s)**	1	1	1
	D.I. (s)	29	1	9

	Obs.	No se ubicaba correctamente.		
Emma dice	**R.I. (s)**	2	3	3
	D.I. (s)	3	8	9
	Obs.			No se ubicaba correctamente.
Lotería	**R.I. (s)**	1	0	1
	D.I. (s)	5	3	17
	Obs.			No se ubicaba correctamente.
Ordena la cara	**R.I. (s)**	0	1	1
	D.I. (s)	3	1	26
	Obs.			No se ubicaba correctamente.
Pop Emma	**R.I. (s)**	5	1	6
	D.I. (s)	45	1	7
	Obs.	No se ubicaba correctamente.		
Encuentra el par	**R.I. (s)**	1	0	1
	D.I. (s)	1	1	2
	Obs.			
Laberinto	**R.I. (s)**	1	0	0
	D.I. (s)	4	1	1
	Obs.			
Espejo	**R.I. (s)**	1	1	1
	D.I. (s)	5	4	16
	Obs.			La interfaz lo reconoció cuando se quitó gafas.

Una mejor visualización de los datos temporales de la Tabla 36 y la Tabla 37 se observa en la Figura 51. Se observa una alta consistencia en la duración de imitación del participante para la gran mayoría de emociones. En los casos de alegría, miedo, asco, sorpresa y tristeza, la imitación duró menos de 10 segundos para todos los casos, sin contar notables excepciones, las cuales se dieron principalmente en las primeras sesiones de imitación. El nivel de asimilación del participante se observa a partir de esta figura, concordando con las observaciones realizadas en el registro conductual, porque en varias emociones, la duración de la imitación fue mayor durante las primeras actividades. Un caso interesante es el del miedo, se observa que la duración en imitación se reduce progresivamente durante la intervención, llegando a durar 1 segundo en la imitación de esta emoción durante el juego *Pop Emma*. Este valor vuelve a aumentar durante las últimas sesiones. Esto se puede deber al cambio en el estado emocional del participante durante el día de aplicación de estas sesiones, como se mencionó anteriormente. Esto es reforzado por los resultados observados en otras emociones: la duración de imitación de la actividad *Espejo* es considerablemente mayor respecto a aquella de las actividades inmediatamente anteriores para distintas emociones como la alegría, el asco, la sorpresa y el enojo. El progreso en la imitación del enojo también es interesante, porque se observa que la dificultad en imitar esta emoción se da en las primeras actividades; no obstante, desde el juego *Ordena la Cara*, se observó una notable disminución en el tiempo de imitación, lo que puede indicar que el participante entendió durante las últimas sesiones la expresión facial correcta de esta emoción. Finalmente, se observa que el retardo en imitación para este participante es bastante consistente, con un valor promedio de 1.27 segundos.

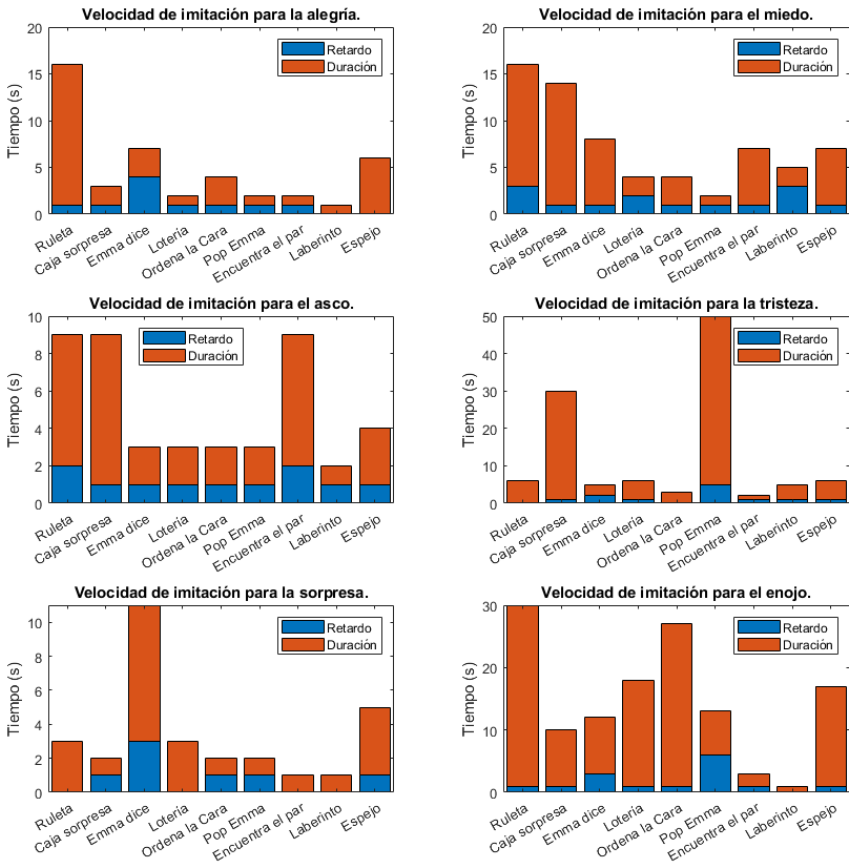

Figura 51. Resultados temporales del registro conductual del sujeto 3.

En la Figura 52 se observa un resumen del tiempo que le tomó a cada participante imitar las emociones durante las pruebas experimentales, donde **D.I.** se refiere a la duración de imitación del sujeto, definida por medio de su media y su desviación estándar. El rango temporal del enojo para el sujeto 2, dado por su desviación estándar, fue recortado de la figura para una mejor visualización de esta, dado que la desviación estándar tuvo un valor de 118.56 segundos, por una actividad cuya duración fue de más de 6 minutos, como se observa en la Tabla 35. Por medio de la Figura 52 se observa con más detalle aquellas emociones con mayor dificultad para imitar por cada participante. En general, el sujeto 1 logró imitar las expresiones faciales en 10.018±14.005 segundos, el sujeto 2 lo logró en 18.074±56.396 segundos y el sujeto 3 en 6.667±8.600 segundos. Sin embargo, como se mencionó anteriormente, este tiempo no está influenciado únicamente por la habilidad de cada participante, dado que también se ve afectado por la habilidad del algoritmo de reconocimiento de expresiones faciales de reconocer el rostro que están realizando, porque la exactitud del algoritmo no es del 100%. Teniendo en cuenta que varios registros conductuales indicaron que los participantes realizaban correctamente las expresiones faciales durante un tiempo considerable antes que el sistema los reconociera, se puede concluir que la influencia en la exactitud del algoritmo es significativa.

A partir de la Figura 52 se puede observar un patrón, en el cual el reconocimiento de las expresiones faciales de la tristeza y el enojo por parte del algoritmo es inferior a las otras emociones, asumiendo que el tiempo de imitación está inversamente relacionado con la exactitud del algoritmo. Si se comparan estos resultados con los obtenidos en la Tabla 27, la cual indica la matriz de confusión del set de prueba del algoritmo de reconocimiento, se puede ver que el enojo y la tristeza tienen unas de las peores exactitudes, con 60.87% y 52.17% respectivamente, lo que puede relacionar directamente la dificultad en la imitación con la efectividad del algoritmo. No obstante, en la Tabla 27 se observa que la categorización del asco tiene una exactitud incluso menor a la de estas dos emociones, siendo de 51.35%, lo cual no se relaciona inversamente con los resultados de la Figura 52. No obstante, este fenómeno se podría explicar por una

alta cantidad de falsos positivos para asco: el sistema reconoce asco en el participante incluso cuando no está evocando esta expresión facial. Aunque no se realizaron observaciones de este estilo en los registros conductuales para el asco, también se puede deber a la alta velocidad en el diligenciamiento de estos documentos que debió tener la estudiante de psicología, razón por la cual pudo haber omitido algunos errores.

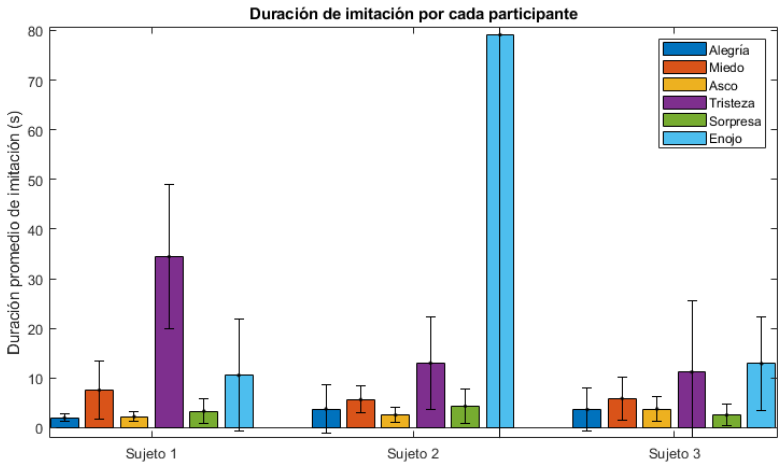

Figura 52. Duración de imitación de cada participante para todas las emociones.

La Figura 52 no indica diferencias importantes entre el participante con TEA (sujeto 1), el participante neurotípico desescolarizado (sujeto 2) y el participante neurotípico escolarizado (sujeto 3). Aunque el sujeto 1 mostró una mayor dificultad consistente en la imitación de la tristeza, su habilidad para imitar el enojo fue mayor que aquella de los participantes neurotípicos, lo mismo para la alegría y para el asco, obteniendo no solo duración menor, sino también menor variabilidad. Por su parte, la gran dificultad del sujeto 2 para imitar el enojo, dada por los resultados de los registros conductuales y la prueba ENI, puede deberse a varias razones. Aunque se podría pensar que la desescolarización afectó las habilidades sociales de este participante, no es posible realizar esta conclusión sin una muestra población más grande, dado que esto también se puede dar por múltiples factores contextuales del participante. Finalmente, el sujeto 3 mostró gran consistencia en la imitación de todas las emociones, ya que ninguna media de duración fue mayor a 13 segundos para este sujeto. Aunque el valor de la duración de imitación fue reducido para todas las emociones, las que más se le dificultaron, al igual que los otros dos participantes, fueron la tristeza y el enojo.

En la Figura 53 se observa el promedio inter-sujeto de duración de imitación para cada emoción. Al igual que en el caso anterior, se ignoró el valor de la desviación estándar del enojo para una mejor visualización de los datos. Este valor fue de 51.037 segundos, considerablemente mayor a aquel de las otras emociones. Esta figura indica que efectivamente existe un patrón en este grupo de participantes en cuanto a la dificultad para reconocer emociones. Se observa que todos los participantes tuvieron gran dificultar para imitar la tristeza, con una duración de 19.556±2.41 segundos. Teniendo en cuenta que la desviación estándar obtenida, se observa una reducida variabilidad en este valor, lo que indica que si se trata de una expresión facial de difícil imitación. Esto no se puede decir del enojo; porque, aunque el tiempo promedio de imitación es mayor al de tristeza, la variabilidad obtenida es muy grande, debido a una baja efectividad por parte del sujeto 2 para imitar esta emoción respecto a los otros dos participantes. La alegría, por su parte, muestra una variabilidad alta en comparación a la media obtenida, por lo cual tampoco se puede indicar que se trate de una emoción consistentemente fácil de imitar; sin embargo, se observa que la imitación no se dificulte en ningún momento. Las otras emociones (miedo, asco y sorpresa) muestran una baja variabilidad de imitación.

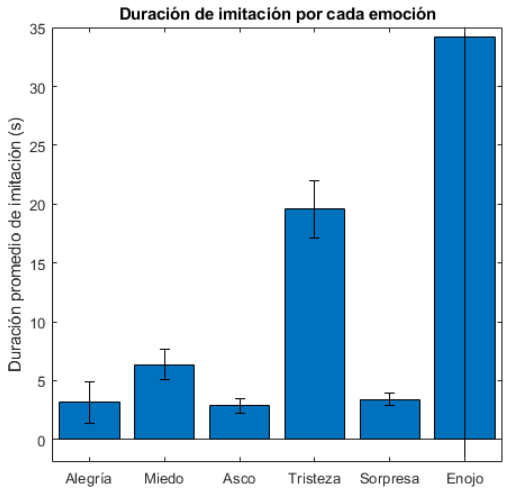

Figura 53. Duración de imitación promedio para todas las emociones.

Aunque se observe una baja variabilidad en la duración de imitación de gran parte de las emociones, esto no se puede relacionar directamente con la efectividad del algoritmo de reconocimiento, ya que hay otras variables para tener en cuenta. La principal es la habilidad general de un niño para reconocer distintas expresiones faciales. Un estudio analizó la habilidad de reconocer emociones a partir de los ojos y la boca por parte de adultos mayores, adultos jóvenes y niños indicó que la expresión facial que más se le dificulta reconocer a un niño es la tristeza, seguido por la sorpresa. No obstante, este mismo estudio también indicó que la expresión facial con mayor facilidad de reconocer por parte de un niño es el enojo [93]. Teniendo en cuenta que nuestro estudio consideró una muestra poblacional muy reducida y cada participante cuenta con condiciones académicas y del neurodesarrollo distintas, no es posible realizar un análisis conclusivo. Esto vuelve necesario, una vez más, la realización de un estudio aleatorizado y controlado con un gran número de participantes que permita afirmar la efectividad de la herramienta de estimulación y del algoritmo de reconocimiento de expresiones faciales.

Finalmente, es importante tener en cuenta que este estudio estuvo limitado de manera considerable en cuanto al número de participantes, ya que únicamente se contó con un participante en el grupo experimental y dos participantes en el grupo control. En cuanto a la utilidad que tiene el estudio para la enseñanza y refuerzo de la imitación y reconocimiento de expresiones faciales, tanto las medidas psicométricas como los registros conductuales indican que los participantes del estudio lograron mejorar sus capacidades comunicativas relacionadas a las emociones. Sin embargo, esto no indica de manera certera que la herramienta diseñada mejoraría las habilidades socio-comunicativas de todos los niños que la usen. Para tener una mayor certeza, es importante realizar una validación rigurosa, como se mostrará más en detalle en la sección VI.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

El proyecto desarrollado buscó apoyar los procesos de estimulación en la imitación y reconocimiento de expresiones faciales emocionales en niños con Trastorno del Espectro Autista (TEA). No obstante, dada la complejidad del proyecto, se logró generar varios desarrollos, relevantes en múltiples campos científicos, como el desarrollo de juegos serios, el reconocimiento de expresiones faciales y el diseño de protocolos experimentales. El producto final fue una herramienta de estimulación que permite a niños con distintas condiciones aprender sobre las emociones a la vez que se divierten, por medio de actividades interactivas con realimentación.

El protocolo experimental diseñado en conjunto con psicólogas y estudiantes de psicología de la Corporación Universitaria Minuto de Dios UNIMINUTO demuestra ser efectivo por varios motivos. Por un lado, los resultados de la prueba psicométrica utilizada, la Evaluación Neuropsicológica Infantil, mostraron cambios después de la intervención realizada, de forma que dos de los participantes lograron mejorar su habilidad para reconocer expresiones faciales emocionales, según esta prueba, siendo el participante con TEA aquel con mayor puntaje en la subprueba al finalizar la intervención. Adicionalmente, un resultado satisfactorio secundario observado en esta prueba psicométrica fue el cambio positivo con el cual los participantes se referían a las emociones, indirectamente indicando un mayor conocimiento sobre ellas. Un ejemplo de esto se observa en el sujeto 2, quien, a partir de las enseñanzas dadas por Emma, el avatar virtual de Emmaciones, cambió sus respuestas originales donde indicaba que los niños en las imágenes de ENI estaban «pensativos» o «aburridos» por el término «tristes», teniendo en cuenta las repetidas ocasiones que recordaron a los participantes las características faciales para la expresión de cada emoción básica. El manual ENI indica que todos los participantes se encuentran dentro del rango normal para el reconocimiento de expresiones faciales. Finalmente, en los registros conductuales se observa que, a nivel general, los participantes mejoraron progresivamente la imitación de expresiones faciales, dado que las observaciones negativas realizadas en la imitación fueron más prevalentes durante las primeras actividades de la etapa de imitación y eran prácticamente inexistentes durante las últimas actividades de la etapa de reconocimiento, con la notable excepción de observaciones que indicaban una mala postura por parte de los participantes.

A partir de las experiencias obtenidas por nuestro equipo de trabajo, se considera necesaria la creación y validación en Colombia de pruebas psicométricas enfocadas en las habilidades de reconocimiento e imitación de expresiones faciales. Aunque la prueba ENI fue de utilidad durante el desarrollo de este proyecto, las pruebas utilizadas se trataron de un ítem dentro de la subprueba de percepción de la batería, por lo cual no se realiza una evaluación detallada del reconocimiento de expresiones faciales. Para trabajos futuros, se considera útil el desarrollo de una batería en la cual se evalúe la capacidad de los participantes de reconocer expresiones faciales a partir de distintas imágenes y videos con y sin contexto, donde los actores tengan distintas edades y se encuentren en distintas poses y ángulos. De igual forma, se considera útil la inclusión de una sección de imitación, donde se le pida a los participantes realizar distintas expresiones faciales, lo cual sería evaluado a partir de criterios objetivos como los mostrados en los registros conductuales de este proyecto.

El algoritmo de reconocimiento de expresiones faciales elaborado es uno de los resultados más importantes de este proyecto, dado que se innovó en el desarrollo e implementación de este. Dos puntos clave en el desarrollo del algoritmo fueron el uso en ensambles de algoritmos de aprendizaje automático para mejorar la efectividad de reconocimiento y la reducción de características para mejorar la velocidad de reconocimiento. Por un lado, utilizar modelos que se entrenen con distintas imágenes, correspondientes a distintas categorías, demostró mejorar la efectividad de reconocimiento de expresiones faciales, ya que fue posible que cada modelo diferenciara con mayor detalle aquellas expresiones faciales que se confunden con facilidad. Como trabajo futuro, es importante explorar las distintas combinaciones de categorías que se pueden generar para lograr una mayor efectividad. Por ejemplo, las matrices de confusión mostradas en los resultados indican que el miedo y la sorpresa se confunden consistentemente. De esta forma, se puede considerar en un futuro crear un modelo especializado en diferenciar entre miedo

y sorpresa, lo que hace más fácil la diferenciación respecto a un modelo que debe diferenciar entre siete categorías distintas. Eventualmente, es posible generar un algoritmo que evalúe las mejores combinaciones de categorías para la creación de ensambles de modelos de aprendizaje automático, de manera que no sea necesario realizar este proceso manualmente.

Por otro lado, la reducción de características por medio de medidas de relevancia demostró tener una gran utilidad, de forma que se logró reducir el tiempo de cómputo del algoritmo de reconocimiento sin tener pérdidas relevantes en la efectividad de este. Por ejemplo, el proceso de reducción de características para el conjunto 1 logró que el proceso completo de extracción fuera aproximadamente 14 veces más rápido de lo que era originalmente, dado que no fue necesario hacer un análisis de los gradientes orientados de la imagen. Esta técnica es extrapolable a otros algoritmos de clasificación, dado que cualquier algoritmo que utilice aprendizaje profundo tradicional debe realizar un proceso de extracción de características, lo que implica que hay características más relevantes que otras. Esto se demuestra en una investigación propia realizada anteriormente, en la cual se observó que la reducción de características permite una veloz clasificación de imágenes de dermatoscopia, con una reducción mínima en la efectividad [94].

Adicionalmente, existen varias etapas dentro del desarrollo del algoritmo de reconocimiento de expresiones faciales que se pueden mejorar. Por un lado, una solución muy utilizada en la clasificación automática es el uso de aprendizaje profundo, lo que permite eliminar las etapas de preprocesamiento y extracción de características del algoritmo, brindando a la vez mejores resultados en varios casos. Aunque en este proyecto se exploró el uso de aprendizaje profundo, no se contó con suficiente tiempo para hacer varias pruebas con redes neuronales convolucionales, dado que es una técnica cuyo entrenamiento requiere de bastante tiempo y recursos especiales. No obstante, en el futuro, se puede volver a explorar el uso de aprendizaje profundo, teniendo en cuenta la necesidad de contar con bases de datos heterogéneas que mejoren el nivel de generalización del algoritmo. Un aspecto importante para tener en cuenta es que estos algoritmos deben ser capaces de reconocer expresiones faciales en tiempo real, lo que significa que se debe buscar un balance entre la exactitud obtenida y la velocidad de procesamiento.

Otro cambio importante que se puede realizar es el desarrollo de un modelo propio de marcadores faciales, el cual sea entrenado a partir de un conjunto heterogéneo de imágenes, en las cuales los rostros de las personas presentes se encuentren en distintas poses y ángulos, a la vez que realicen distintas expresiones faciales, ya que las bases de datos con las cuales se entrenan los algoritmos de ubicación de marcadores faciales actualmente suelen mostrar poses y expresiones faciales reducidas. Al tener conjuntos más variados, se puede, por un lado, evitar ciertas secciones de la etapa de preprocesamiento, las cuales son implementadas para facilitar la ubicación de marcadores faciales. Por otro lado, si la base de datos cuenta con expresiones faciales variadas, se pueden evitar errores como los mostrados en resultados, donde, por ejemplo, si se abre bastante la boca, la ubicación de marcadores faciales se hace de manera errónea.

Otro aspecto en el que se puede realizar una mayor exploración para mejorar el algoritmo de reconocimiento es el uso de otras características que logren definir de mejor manera la expresión facial de una persona. Aunque este proyecto se enfocó en datos geométricos de los marcadores faciales y en los gradientes orientados, existen otras características expuestas en el estado del arte que podrían mostrar utilidad para cumplir este objetivo, como las matrices de coocurrencia y la transformada Wavelet en dos dimensiones. Estas técnicas no se utilizaron por su dificultad de implementación, con relación al tiempo de desarrollo esperado para el proyecto. No obstante, en un futuro, se puede desarrollar una biblioteca compatible con OpenCV que incluya estas técnicas no implementadas actualmente en la herramienta; una biblioteca con buena calidad puede ser eventualmente incluida dentro de los módulos extra de OpenCV. Sin embargo, al igual que en casos anteriores, es importante realizar un estudio profundo de la viabilidad de implementación de estas técnicas, ya que sus tiempos de cómputo pueden ser excesivamente altos para la categorización en tiempo real. De igual forma, expertos en el área que evaluaron el proyecto recomiendan evaluar el uso del Análisis de Componentes Principales (PCA) para lograr una reducción en las características y mejorar el tiempo de cómputo. También consideran pertinente su uso en la etapa de preprocesamiento de las imágenes para descartar aspectos de las imágenes que no son pertinentes para

el estudio de las expresiones faciales. Estos cambios se tendrán en cuenta en etapas posteriores del proyecto.

Una investigación futura relacionada con las expresiones faciales que puede ser de gran interés para distintos investigadores es la validación de los hallazgos de Ekman que indican que las expresiones faciales son universales, por medio de medios tecnológicos. De esta manera, en un futuro se puede analizar si el algoritmo desarrollado tiene la misma efectividad al reconocer expresiones faciales en personas con diferencias culturales o, en caso contrario, evaluar formas en las que se pueda mejorar la universalidad en el reconocimiento, a partir del uso de bases de datos culturalmente heterogéneas, pero con condiciones lumínicas y de posición y orientación similares.

El algoritmo de reconocimiento de expresiones faciales y aquellos elementos que lo componen se pueden aplicar en un futuro en otros contextos. Por ejemplo, al aplicar ciertas modificaciones a los modelos de aprendizaje automático, es posible saber qué tan a gusto se siente una persona con la terapia que le están realizando, mejorando la calidad de la intervención y brindando realimentación al terapeuta. Por otro lado, su uso en la interacción humano-robot puede ser importante para que un robot sea capaz de modificar su comportamiento a partir de la emoción que esté sintiendo la persona con la cual esté interactuando.

De igual forma, si se desea utilizar el algoritmo de reconocimiento de expresiones faciales en otras aplicaciones, hay varios elementos de interés en los que se puede profundizar más en un futuro. Uno de estos es el reconocimiento de microexpresiones, expresiones faciales con una duración limitada que se muestran durante los primeros instantes después de la evocación de una emoción. Este tema es de interés, ya que puede indicar la emoción real que está sintiendo una persona, incluso si luego la disimula. Para profundizar en este tema, es importante una reducción en tiempo de cómputo de todas las etapas de procesamiento. De igual forma, es necesario contar con un dispositivo de adquisición de imágenes con alta resolución, que sea capaz de identificar cambios leves en los músculos faciales Para este caso, el número de características extraídas debe ser reducido y su obtención no debe tener gran complejidad. El uso de técnicas de reducción de características es aconsejado para este caso.

Un aspecto que le puede dar más robustez a los resultados que indican la efectividad del algoritmo de reconocimiento en tiempo real es el uso de marcadores biológicos. Un candidato relevante para esto es el uso de señales de electroencefalografía, que permitan indicar a partir de la actividad cerebral en la corteza visual el instante en el que una persona responde a un estímulo visual, de forma que este tiempo sea comparable con aquel en el que el algoritmo reconoce la emoción, reduciendo la incertidumbre causada por no saber si la falta de reconocimiento es una falla del algoritmo o que el participante no ha reaccionado a la instrucción dada.

Finalmente, se desarrolló un juego donde se implementó un protocolo experimental para la estimulación de procesos de imitación y reconocimiento de expresiones faciales. El aspecto más importante de este juego es que se trata de un producto que únicamente utiliza herramientas de dominio público; por lo cual, cualquier persona con acceso a él no debe pagar ninguna licencia externa. Adicionalmente, dado que el despliegue de Emmaciones se hizo con ayuda de las herramientas de Unity, es posible ejecutar el juego sin necesidad de instalar programas adicionales. Actualmente, Emmaciones funciona en Windows 10; sin embargo, en un futuro se pueden desarrollar versiones para macOS y Linux, sin necesidad de hacer grandes cambios. De igual forma, más adelante se pueden crear versiones móviles para iOS y Android, teniendo en cuenta que los controles del juego se deben cambiar para que funcionen con una pantalla táctil y que es posible que las bibliotecas de visión artificial utilizadas que son compatibles con otros sistemas operativos pueden no serlo con las plataformas móviles. Por último, un experto en el área de procesamiento de imágenes considera pertinente el uso de caras felices para la realimentación visual que se le brinda a los participantes, en vez de los estímulos que brinda Emma actualmente, dado que es un elemento con el que ha trabajado con anterioridad. En trabajos futuros se evaluará esta sugerencia.

Al observar la totalidad del proyecto, es necesario realizar en un futuro una validación del sistema desarrollado, en la que se demuestre la efectividad de este. Teniendo en cuenta investigaciones previas,

mencionadas en el estado del arte, una validación adecuada consistiría en contar con la participación de aproximadamente 100 niños con TEA [13], [14], divididos en dos grupos: intervención por medio de Emmaciones (grupo experimental) e intervención tradicional (grupo control). Si se observa que los resultados obtenidos en el grupo experimental son estadísticamente mejores a aquellos obtenidos con el grupo control, se puede demostrar que el sistema diseñado es efectivo. Así, en un futuro, la meta final del proyecto es construir alianzas con instituciones prestadoras de salud, donde se implemente el uso de Emmaciones como parte de las terapias de comunicación para niños con TEA.

VII. REFERENCIAS

[1] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] S. E. Piñeros-Ortiz y S. M. Toro-Herrera, «Conceptos generales sobre ABA en niños con trastorno del espectro autista», *Rev. Fac. Med.*, vol. 60, n.º 1, Art. n.º 1, ene. 2012.

[3] F. Chiarotti y A. Venerosi, «Epidemiology of Autism Spectrum Disorders: A Review of Worldwide Prevalence Estimates Since 2014», *Brain Sci.*, vol. 10, n.º 5, may 2020, doi: 10.3390/brainsci10050274.

[4] C. Lopata, J. P. Donnelly, A. K. Jordan, M. L. Thomeer, C. A. McDonald, y J. D. Rodgers, «Brief Report: Parent-Teacher Discrepancies on the Developmental Social Disorders Scale (BASC-2) in the Assessment of High-Functioning Children with ASD», *J. Autism Dev. Disord.*, vol. 46, n.º 9, pp. 3183-3189, sep. 2016, doi: 10.1007/s10803-016-2851-0.

[5] S. Levinson, J. Neuspiel, A. Eisenhower, y J. Blacher, «Parent–Teacher Disagreement on Ratings of Behavior Problems in Children with ASD: Associations with Parental School Involvement Over Time», *J. Autism Dev. Disord.*, vol. 51, n.º 6, pp. 1966-1982, jun. 2021, doi: 10.1007/s10803-020-04675-1.

[6] P. Chaste y M. Leboyer, «Autism risk factors: genes, environment, and gene-environment interactions», *Dialogues Clin. Neurosci.*, vol. 14, n.º 3, pp. 281-292, sep. 2012.

[7] «MinSalud incluye en el estudio de salud mental el autismo». https://www.minsalud.gov.co/Paginas/salud-mental-el-autismo.aspx (accedido may 06, 2021).

[8] Ministerio de Salud, «Boletín de Salud Mental Oferta y Acceso a Servicios en Salud Mental en Colombia». jul. 2018. [En línea]. Disponible en: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ENT/Boletin-6-salud-mental-2018.pdf

[9] «Proyecciones de población». https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion (accedido may 06, 2021).

[10] X. Li, Y. Yan, y W. Wei, «Identifying Patients with Poststroke Mild Cognitive Impairment by Pattern Recognition of Working Memory Load-Related ERP», *Comput. Math. Methods Med.*, vol. 2013, 2013, doi: 10.1155/2013/658501.

[11] N. Malik-Soni *et al.*, «Tackling healthcare access barriers for individuals with autism from diagnosis to adulthood», *Pediatr. Res.*, 2021, doi: 10.1038/s41390-021-01465-y.

[12] R. Khara, L. R Kalankesh, H. Shahrokhi, S. Dastgiri, K. Gholipour, y M.-R. Feizi-Derakhshi, «Identifying the Needs of Families of Children with Autism Spectrum Disorder from Specialists and Parents' Perspectives: A Qualitative Study», *Iran. J. Psychiatry Behav. Sci.*, vol. 14, n.º 4, Art. n.º 4, dic. 2020, doi: 10.5812/ijpbs.107203.

[13] F. Robain, N. Kojovic, S. Solazzo, B. Glaser, M. Franchini, y M. Schaer, «The impact of social complexity on the visual exploration of others' actions in preschoolers with autism spectrum disorder», *BMC Psychol.*, vol. 9, n.º 1, p. 50, mar. 2021, doi: 10.1186/s40359-021-00553-2.

[14] A. M. Roux, P. T. Shattuck, B. P. Cooper, K. A. Anderson, M. Wagner, y S. C. Narendorf, «Postsecondary employment experiences among young adults with an autism spectrum disorder», *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 52, n.º 9, pp. 931-939, sep. 2013, doi: 10.1016/j.jaac.2013.05.019.

[15] P. T. Shattuck, S. C. Narendorf, B. Cooper, P. R. Sterzing, M. Wagner, y J. L. Taylor, «Postsecondary education and employment among youth with an autism spectrum disorder», *Pediatrics*, vol. 129, n.º 6, pp. 1042-1049, jun. 2012, doi: 10.1542/peds.2011-2864.

[16] A. Holwerda, J. J. L. van der Klink, J. W. Groothoff, y S. Brouwer, «Predictors for work participation in individuals with an Autism spectrum disorder: a systematic review», *J. Occup. Rehabil.*, vol. 22, n.º 3, pp. 333-352, sep. 2012, doi: 10.1007/s10926-011-9347-8.

[17] P. J. Lang, «The emotion probe. Studies of motivation and attention», *Am. Psychol.*, vol. 50, n.º 5, pp. 372-385, may 1995, doi: 10.1037//0003-066x.50.5.372.

[18] *Neuroscience, 3rd ed*. Sunderland, MA, US: Sinauer Associates, 2004, pp. xix, 773.

[19] M. Scandar, «Emoción y Comprensión Lectora: Relación entre Niveles de Valencia, Activación y Dominancia y la Comprensión de Textos Expositivos y Argumentativos», 2019. doi: 10.13140/RG.2.2.34321.38246.

[20] J. A. Russell, «A circumplex model of affect», *J. Pers. Soc. Psychol.*, vol. 39, n.º 6, pp. 1161-1178, 1980, doi: 10.1037/h0077714.

[21] P. Ekman y H. Oster, «Facial expressions of emotion», *Annu. Rev. Psychol.*, vol. 30, pp. 527-554, 1979, doi: 10.1146/annurev.ps.30.020179.002523.
[22] R. L. Leahy, D. Tirch, y L. A. Napolitano, *Emotion Regulation in Psychotherapy: A Practitioner's Guide*. Guilford Press, 2011.
[23] K. Sofronoff, T. Attwood, S. Hinton, y I. Levin, «A randomized controlled trial of a cognitive behavioural intervention for anger management in children diagnosed with Asperger syndrome», *J. Autism Dev. Disord.*, vol. 37, n.º 7, pp. 1203-1214, ago. 2007, doi: 10.1007/s10803-006-0262-3.
[24] M. M. Konstantareas y K. Stewart, «Affect regulation and temperament in children with Autism Spectrum Disorder», *J. Autism Dev. Disord.*, vol. 36, n.º 2, pp. 143-154, feb. 2006, doi: 10.1007/s10803-005-0051-4.
[25] M. Berking y P. Wupperman, «Emotion regulation and mental health: recent findings, current challenges, and future directions», *Curr. Opin. Psychiatry*, vol. 25, n.º 2, pp. 128-134, mar. 2012, doi: 10.1097/YCO.0b013e3283503669.
[26] C. A. Mazefsky *et al.*, «The Role of Emotion Regulation in Autism Spectrum Disorder RH: Emotion Regulation in ASD», *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 52, n.º 7, pp. 679-688, jul. 2013, doi: 10.1016/j.jaac.2013.05.006.
[27] T. A. Dennis y G. Hajcak, «The late positive potential: a neurophysiological marker for emotion regulation in children», *J. Child Psychol. Psychiatry*, vol. 50, n.º 11, pp. 1373-1383, 2009, doi: 10.1111/j.1469-7610.2009.02168.x.
[28]S. S. Tomkins y R. McCarter, «What and Where are the Primary Affects? Some Evidence for a Theory», *Percept. Mot. Skills*, vol. 18, n.º 1, pp. 119-158, feb. 1964, doi: 10.2466/pms.1964.18.1.119.
[29] P. Ekman, E. R. Sorenson, y W. V. Friesen, «Pan-Cultural Elements in Facial Displays of Emotion», *Science*, vol. 164, n.º 3875, pp. 86-88, abr. 1969, doi: 10.1126/science.164.3875.86.
[30] W. V. Friesen, «Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules», ProQuest Information & Learning, US, 1973.
[31] D. Matsumoto y P. Ekman, «American-Japanese cultural differences in intensity ratings of facial expressions of emotion», *Motiv. Emot.*, vol. 13, n.º 2, pp. 143-157, jun. 1989, doi: 10.1007/BF00992959.
[32] P. Ekman, «Universals and cultural differences in facial expressions of emotion», *Nebr. Symp. Motiv.*, vol. 19, pp. 207-283, 1971.
[33] D. Verdugo, «Implementación de un sistema de identificación de emociones a partir de gestos para el apoyo en la terapia de niños con trastorno del espectro autista». 2019.
[34] K. Arenas, D. Verdugo, J. M. López, A. Rizo, S. Linares, y C. Cárdenas, «Creación de un Sistema de Software para estimular la imitación y el reconocimiento de expresiones emocionales faciales en niños con Trastorno del Espectro Autista entre los 6 a 8 años». 2019.
[35] A. Rizo *et al.*, «Interfaz gráfica para la estimulación en el reconocimiento e imitación de expresiones emocionales faciales en niños con trastorno del espectro autista». 2020.
[36] E. Laugeson y M. N. Park, «Using a CBT Approach to Teach Social Skills to Adolescents with Autism Spectrum Disorder and Other Social Challenges: The PEERS® Method», *undefined*, 2014, Accedido: may 20, 2021. [En línea]. Disponible en: /paper/Using-a-CBT-Approach-to-Teach-Social-Skills-to-with-Laugeson-Park/a1634b9d07594f75e66a32cc02559a52c27c85de
[37] E. A. Laugeson, A. Gantman, S. K. Kapp, K. Orenski, y R. Ellingsen, «A Randomized Controlled Trial to Improve Social Skills in Young Adults with Autism Spectrum Disorder: The UCLA PEERS® Program», *J. Autism Dev. Disord.*, vol. 45, n.º 12, pp. 3978-3989, dic. 2015, doi: 10.1007/s10803-015-2504-8.
[38] S. J. Rabin, E. A. Laugeson, I. Mor-Snir, y O. Golan, «An Israeli RCT of PEERS®: Intervention Effectiveness and the Predictive Value of Parental Sensitivity», *J. Clin. Child Adolesc. Psychol.*, 2020, doi: 10.1080/15374416.2020.1796681.
[39] S. J. Rogers, «Early Start Denver Model», en *Encyclopedia of Autism Spectrum Disorders*, F. R. Volkmar, Ed. New York, NY: Springer, 2013, pp. 1034-1042. doi: 10.1007/978-1-4419-1698-3_1821.
[40] G. Dawson *et al.*, «Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model», *Pediatrics*, vol. 125, n.º 1, pp. e17-e23, 2010, doi: 10.1542/peds.2009-0958.
[41] R. L. Gabriels, Z. Pan, B. Dechant, J. A. Agnew, N. Brim, y G. Mesibov, «Randomized Controlled Trial of Therapeutic Horseback Riding in Children and Adolescents With Autism Spectrum Disorder», *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 54, n.º 7, pp. 541-549, jul. 2015, doi: 10.1016/j.jaac.2015.04.007.

[42] M. R. Pazmiño y I. Harari, «Uso de nuevas tecnologías TICS -realidad aumentada para tratamiento de niños TEA un diagnóstico inicial», *CienciAmérica Rev. Divulg. Científica Univ. Tecnológica Indoamérica*, vol. 6, n.º 3, pp. 131-137, 2017.
[43] J. L. Martínez, J. B. Pagán, S. A. García, y M. del C. C. Máiquez, «Las tecnologías de la información y comunicación (TIC) en el proceso de enseñanza y aprendizaje del alumnado con trastorno del espectro autista (TEA)», *Rev. Fuentes*, n.º 14, pp. 193-208, 2014.
[44] S. Leiva, «Validación de una batería para evaluar el reconocimiento de emociones a través del rostro y del cuerpo utilizando estímulos dinámicos», *Rev. Argent. Cienc. Comport.*, vol. 9, n.º 3, Art. n.º 3, dic. 2017, doi: 10.32348/1852.4206.v9.n3.17186.
[45] B. de Gelder, E. M. J. Huis in 't Veld, y J. Van den Stock, «The Facial Expressive Action Stimulus Test. A test battery for the assessment of face memory, face and object perception, configuration processing, and facial expression recognition», *Front. Psychol.*, vol. 6, 2015, doi: 10.3389/fpsyg.2015.01609.
[46] A. Ardila, «Cultural values underlying psychometric cognitive testing», *Neuropsychol. Rev.*, vol. 15, n.º 4, pp. 185-195, dic. 2005, doi: 10.1007/s11065-005-9180-y.
[47] E. Matute, M. Rosselli, A. Ardila, y F. Ostrosky-Solís, «Evaluación neuropsicológica infantil», *México Man. Mod.*, 2007.
[48] E. Matute, O. Inozemtseva, A. L. González-Reyes, y Y. Chamorro, «La Evaluación Neuropsicológica Infantil (ENI): Historia y fundamentos teóricos de su validación. Un acercamiento práctico a su uso y valor diagnóstico», *Rev. Neuropsicol. Neuropsiquiatría Neurocienc.*, vol. 14, n.º 1, pp. 68-95, 2014.
[49] M. Roselli *et al.*, «Evaluación Neuropsicológica Infantil (ENI): una batería para la evaluación de niños entre 5 y 16 años de edad. Estudio normativo colombiano», *Neurología*, vol. 38, 2004.
[50] V. G. Rangarajan, *Effectiveness of Contrast Limited Adaptive Histogram Equalization on Multispectral Satellite Imagery*. GRIN Verlag, 2018.
[51] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, y K. E. Muller, «Contrast-limited adaptive histogram equalization: speed and effectiveness», en *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, may 1990, pp. 337-345. doi: 10.1109/VBC.1990.109340.
[52] A. Zadorozny y H. Zhang, «Contrast enhancement using morphological scale space», en *2009 IEEE International Conference on Automation and Logistics*, ago. 2009, pp. 804-807. doi: 10.1109/ICAL.2009.5262814.
[53] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, y X. Xie, «LLCNN: A convolutional neural network for low-light image enhancement», en *2017 IEEE Visual Communications and Image Processing (VCIP)*, dic. 2017, pp. 1-4. doi: 10.1109/VCIP.2017.8305143.
[54] «Face Detection vs Facial Recognition – what's the difference? - NEC NZ», *NEC*, feb. 12, 2020. https://www.nec.co.nz/market-leadership/publications-media/face-detection-vs-facial-recognition-whats-the-difference/ (accedido jun. 28, 2021).
[55] P. Viola y M. Jones, *Rapid object detection using a boosted cascade of simple features*. 2001.
[56] «Face Detection – OpenCV, Dlib and Deep Learning (C++ / Python)», oct. 22, 2018. https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/ (accedido jun. 01, 2021).
[57] R. Thaware, «Real-Time Face Detection and Recognition with SVM and HOG Features», *EEWeb*, may 28, 2018. https://www.eeweb.com/real-time-face-detection-and-recognition-with-svm-and-hog-features/ (accedido jun. 01, 2021).
[58] W. Liu *et al.*, «SSD: Single Shot MultiBox Detector», *ArXiv151202325 Cs*, vol. 9905, pp. 21-37, 2016, doi: 10.1007/978-3-319-46448-0_2.
[59] D. E. King, «Max-Margin Object Detection», *ArXiv150200046 Cs*, ene. 2015, Accedido: jun. 01, 2021. [En línea]. Disponible en: http://arxiv.org/abs/1502.00046
[60] «Facial Action Coding System», *Paul Ekman Group*. https://www.paulekman.com/facial-action-coding-system/ (accedido jun. 02, 2021).
[61] *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press. Accedido: jun. 02, 2021. [En línea]. Disponible en: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195179644.001.0001/acprof-9780195179644
[62] «Facial Action Coding System (FACS) - A Visual Guidebook - iMotions», *Imotions Publish*, ago. 18, 2019. https://imotions.com/blog/facial-action-coding-system/ (accedido jun. 02, 2021).

[63] «New Affectiva cloud API helps machines understand emotions in human speech», *TechCrunch*. https://social.techcrunch.com/2017/09/13/new-affectiva-cloud-api-helps-machines-understand-emotions-in-human-speech/ (accedido jun. 02, 2021).
[64] H. Ouanan, M. Ouanan, y B. Aksasse, «Facial landmark localization: Past, present and future», en *2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, oct. 2016, pp. 487-493. doi: 10.1109/CIST.2016.7805097.
[65] V. Kazemi y J. Sullivan, «One millisecond face alignment with an ensemble of regression trees», en *2014 IEEE Conference on Computer Vision and Pattern Recognition*, jun. 2014, pp. 1867-1874. doi: 10.1109/CVPR.2014.241.
[66] G. Tzimiropoulos y M. Pantic, «Optimization Problems for Fast AAM Fitting in-the-Wild», en *2013 IEEE International Conference on Computer Vision*, dic. 2013, pp. 593-600. doi: 10.1109/ICCV.2013.79.
[67] M. Pietikäinen y A. Hadid, «Texture Features in Facial Image Analysis», en *Advances in Biometric Person Authentication*, Berlin, Heidelberg, 2005, pp. 1-8. doi: 10.1007/11569947_1.
[68] P. Carcagnì, M. Del Coco, M. Leo, y C. Distante, «Facial expression recognition and histograms of oriented gradients: a comprehensive study», *SpringerPlus*, vol. 4, oct. 2015, doi: 10.1186/s40064-015-1427-3.
[69] G. Verma y H. Verma, «Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions», *Rev. Socionetwork Strateg.*, vol. 14, ago. 2020, doi: 10.1007/s12626-020-00061-6.
[70] C. Grossard *et al.*, «Children with autism spectrum disorder produce more ambiguous and less socially meaningful facial expressions: an experimental study using random forest classifiers», *Mol. Autism*, vol. 11, n.º 1, p. 5, ene. 2020, doi: 10.1186/s13229-020-0312-2.
[71] E. Zane, Z. Yang, L. Pozzan, T. Guha, S. Narayanan, y R. B. Grossman, «Motion-Capture Patterns of Voluntarily Mimicked Dynamic Facial Expressions in Children and Adolescents With and Without ASD», *J. Autism Dev. Disord.*, vol. 49, n.º 3, pp. 1062-1079, mar. 2019, doi: 10.1007/s10803-018-3811-7.
[72] F. Marino *et al.*, «Outcomes of a Robot-Assisted Social-Emotional Understanding Intervention for Young Children with Autism Spectrum Disorders», *J. Autism Dev. Disord.*, vol. 50, n.º 6, pp. 1973-1987, 2020, doi: 10.1007/s10803-019-03953-x.
[73] L. M. Rice, C. A. Wall, A. Fogel, y F. Shic, «Computer-Assisted Face Processing Instruction Improves Emotion Recognition, Mentalizing, and Social Skills in Students with ASD», *J. Autism Dev. Disord.*, vol. 45, n.º 7, pp. 2176-2186, jul. 2015, doi: 10.1007/s10803-015-2380-2.
[74] «FACESAY™ SOCIAL SKILLS SOFTWARE GAMES», *FACESAY™ SOCIAL SKILLS SOFTWARE GAMES*. https://www.facesay.com/ (accedido jun. 03, 2021).
[75] S. Adatrao y M. Mittal, *An analysis of different image preprocessing techniques for determining the centroids of circular marks using hough transform*. 2016, p. 115. doi: 10.1109/ICFSP.2016.7802966.
[76] A. Gholizadeh, L. Borůvka, M. M. Saberioon, J. Kozák, R. Vašát, y K. Němeček, «Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features», *Soil Water Res.*, vol. 10 (2015), n.º No. 4, pp. 218-227, oct. 2015, doi: 10.17221/113/2015-SWR.
[77] S. Ren, X. Cao, Y. Wei, y J. Sun, «Face Alignment at 3000 FPS via Regressing Local Binary Features», en *2014 IEEE Conference on Computer Vision and Pattern Recognition*, jun. 2014, pp. 1685-1692. doi: 10.1109/CVPR.2014.218.
[78] «dlib C++ Library». http://dlib.net/ (accedido jun. 11, 2021).
[79] «Dlib FaceLandmark Detector | Integration | Unity Asset Store». https://assetstore.unity.com/packages/tools/integration/dlib-facelandmark-detector-64314 (accedido jun. 11, 2021).
[80] «OpenCV for Unity | Integration | Unity Asset Store». https://assetstore.unity.com/packages/tools/integration/opencv-for-unity-21088 (accedido jun. 11, 2021).
[81] E. Goeleven, R. De Raedt, L. Leyman, y B. Verschuere, «The Karolinska Directed Emotional Faces: A validation study», *Cogn. Emot.*, vol. 22, pp. 1094-1118, sep. 2008, doi: 10.1080/02699930701626582.
[82] M. Kamachi, M. Lyons, y J. Gyoba, «The japanese female facial expression (jaffe) database», *Availble Httpwww Kasrl Orgjaffe Html*, ene. 1997.
[83] V. LoBue y C. Thrasher, «The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults», *Front. Psychol.*, vol. 5, 2015, doi: 10.3389/fpsyg.2014.01532.
[84] «FER-2013». https://kaggle.com/msambare/fer2013 (accedido jun. 15, 2021).

[85] «AlexNet - ImageNet Classification with Convolutional Neural Networks», oct. 29, 2018. https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/ (accedido jun. 16, 2021).
[86] K. He, X. Zhang, S. Ren, y J. Sun, «Deep Residual Learning for Image Recognition», *ArXiv151203385 Cs*, dic. 2015, Accedido: jun. 16, 2021. [En línea]. Disponible en: http://arxiv.org/abs/1512.03385
[87] «Jetson TX1 Module», *NVIDIA Developer*, ago. 01, 2016. https://developer.nvidia.com/embedded/jetson-tx1 (accedido jun. 16, 2021).
[88] M. Sahay, «Neural Networks and the Universal Approximation Theorem», *Medium*, mar. 13, 2021. https://towardsdatascience.com/neural-networks-and-the-universal-approximation-theorem-8a389a33d30a (accedido oct. 01, 2021).
[89] F. Berto y J. Tagliabue, «Cellular Automata», en *The Stanford Encyclopedia of Philosophy*, Spring 2021., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2021. Accedido: jul. 08, 2021. [En línea]. Disponible en: https://plato.stanford.edu/archives/spr2021/entries/cellular-automata/
[90] J. Kim, J. K. Lee, y K. M. Lee, «Accurate Image Super-Resolution Using Very Deep Convolutional Networks», *ArXiv151104587 Cs*, nov. 2016, Accedido: jun. 16, 2021. [En línea]. Disponible en: http://arxiv.org/abs/1511.04587
[91] X. Guo, L. Chen, y C. Shen, «Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis», *Measurement*, vol. 93, pp. 490-502, nov. 2016, doi: 10.1016/j.measurement.2016.07.054.
[92] «F.A.C.E. Training | About», feb. 09, 2010. https://web.archive.org/web/20100209065736/http://face.paulekman.com/aboutmett2.aspx (accedido jun. 22, 2021).
[93] M. Guarnera, P. Magnano, M. Pellerone, M. I. Cascio, V. Squatrito, y S. L. Buccheri, «Facial expressions and the ability to recognize emotions from the eyes or mouth: A comparison among old adults, young adults, and children», *J. Genet. Psychol.*, vol. 179, n.º 5, pp. 297-310, oct. 2018, doi: 10.1080/00221325.2018.1509200.
[94] S. Pulido Castro, Á. Bocanegra, J. López, M. Forero, y S. Cancino, *Feature relevance in dermoscopy images by the use of ABCD standard*. 2020. doi: 10.1117/12.2567946.

ANEXOS

Anexo 1. Consentimiento informado de los experimentos (Espacio para firmas eliminado).

Consentimiento Informado
Corporación Universitaria Minuto de Dios- UNIMINUTO - Escuela Colombiana de Ingeniería Julio Garavito

La información aquí descrita propende por el cumplimiento de los principios generales del Artículo 2 del código deontológico y bioético del psicólogo de la ley 1090 de 2006. En este se plantea que desde el ejercicio profesional como psicólogas y psicólogos se deberá velar por el bienestar e integridad de las y los participantes de cualquier proyecto de investigación. Así mismo, por el cumplimiento del artículo 17 de la ley 842 de 2003 en el cual se menciona el ejercicio de la profesión de Ingeniería Biomédica que debe ser guiado por fines que exalten su profesión, en el marco de este proyecto las y los profesionales de ingeniería contribuyen con sus conocimientos, capacidad y experiencia, servir a la humanidad, proteger la vida y salud de los miembros de la comunidad, evitando riesgos innecesarios en la ejecución sus investigaciones.

Por tanto, el proyecto **piloto** denominado **Herramienta computacional para estimular la imitación y el reconocimiento de expresiones emocionales faciales en niños con Trastorno del Espectro Autista** dirigido por profesores investigadores y estudiantes del Programa Ingeniería Biomédica y Maestría de Ingeniería Electrónica de la Escuela Colombiana de Ingeniería Julio Garavito (Escuela) y del Programa de Psicología de la Corporación Universitaria Minuto de Dios (UMD), tiene como *objetivo* estimular los procesos de imitación e identificación de las expresiones faciales emocionales por medio de herramientas tecnológicas, que serán supervisadas por practicantes de Psicología de la UMD y un estudiante de maestría de la Escuela.

La participación se llevará a cabo de la siguiente forma: *Primero* se realizará la aplicación de la subprueba Reconocimiento de Expresiones Faciales de la Evaluación Neuropsicológica Infantil (ENI-2). *Segundo*, se implementará la herramienta computacional con el participante en compañía de un investigador(a) de la Escuela y uno de la UMD-SP, realizando diferentes actividades en cada sesión, estas tendrán una duración de 60 minutos y se realizarán dos por encuentro, para un total de 12 sesiones y 6 encuentros. *Tercero,* una vez finalizadas las sesiones con la herramienta se aplicará nuevamente la subprueba de Reconocimiento de Expresiones Faciales de la ENI-2. *Cuarto,* se hará una devolución verbal o escrita del proceso realizado.

De acuerdo con lo anterior, el proceso a desarrollar durante esta investigación corresponde a la categoría "con riesgo mínimo" que según el artículo 11 de la resolución número 8430 de 1993, se caracteriza por emplear registros de datos a través de procedimientos comunes, exámenes físicos o psicológicos que no buscan manipular la conducta del participante con medicamentos de uso común o de uso terapéutico. La aplicación de la subprueba y la Herramienta Computacional NO tiene como fin generar un **diagnóstico o intervención clínica, ni tampoco tiene validez jurídica**.

Sí durante el estudio usted o el niño tiene preguntas o incomodidades, no dude en informar a las y los investigadores, y si persisten durante el proceso recuerde que puede retirarse sin dar explicación ni perjuicio alguno para usted, el niño o las y los investigadores del proyecto.

Los resultados y hallazgos encontrados en el proyecto serán presentados en trabajos de grado, posiblemente publicados y comunicados de forma *anónima* en eventos académicos y científicos, revistas científicas y otros espacios académicos, con la finalidad de divulgar el conocimiento obtenido y no se espera recibir beneficio económico por los productos publicados.

Así mismo, los resultados serán consolidados en una base de datos que podrá ser consultada por éste y/o futuros proyectos relacionados con la expresa autorización de las y los investigadores de este proyecto, dicha autorización y manejo de información será *anónima* que se acoge a lo establecido en el artículo 3 de la ley 1581 de 2012 sobre el tratamiento de los datos.

Se garantiza la confidencialidad de su nombre y datos personales, que serán custodiados por las y los investigadores del proyecto de conformidad con el decreto 1377 de 2013 de la ley 1581 de 2012 sobre el tratamiento de los datos. Sin embargo, si durante el proceso o en otro momento del proyecto se identifica que la vida del niño o de un tercero se encuentra en inminente riesgo, es deber del psicólogo y practicantes informar al personal pertinente de acuerdo con el artículo 2 numeral 5 de la ley 1090 de 2006.

DECLARACIÓN DE LOS PADRES O ACUDIENTES

Nosotros, _____ _____ obrando en calidad de representantes del/ la menor _____ identificado(a) con la T.I. N°_____ certificamos que hemos recibido la información pertinente del proyecto, se nos indicó el procedimiento a seguir de manera clara y sencilla, el cual consta de la aplicación de una subprueba, ejecución de una serie de actividades en una Herramienta Computacional con el objetivo de estimular el proceso de imitación y reconocimiento facial de las emociones, se nos indicó el número de sesiones y el tiempo estipulado para cada una de ellas.

Adicionalmente, certificamos que hemos comprendido que la información recibida por los investigadores es de carácter confidencial y la podrán revelar en caso de que el profesional detecte un evidente daño para el participante durante el procedimiento. También se nos indicó que podemos revocar el consentimiento cuando lo consideremos pertinente.

Una vez leído y comprendido el procedimiento que se llevará a cabo, se firma el siguiente consentimiento el día _____ del mes _____ del año _____, en la ciudad de _____.

DECLARACIÓN DE LOS INVESTIGADORES

Nosotros los profesores investigadores Dr. Juan Manuel López, Ps. Alejandra Rizo Arévalo, Ing. Sandra Liliana Cancino Suarez y estudiantes Sergio David Pulido Castro, Nubia Jasbleidy Palacios Quecan, Michelle Paola Ballen Cárdenas, Angie Katherine Arenas Pérez y Santiago López Sanchez certificamos que le hemos informado al acudiente de _____ acerca del objetivo del proyecto, que la participación del niño o niña y los resultados son confidenciales, se publicarán tanto en revistas como eventos académicos de forma anónima, que el procedimiento no representa ningún riesgo para él/ella. También resolvimos todas las inquietudes respecto a la participación del niño en este proyecto piloto.

Anexo 2. Ejemplo de uno de los guiones aplicados en la herramienta de estimulación.

¡Hola! Antes de iniciar con nuestras actividades te voy a presentar las emociones. Tenemos 6 emociones importantes, que son: Tristeza. Alegría, que seguro la habrás escuchado como felicidad. Miedo también conocido como susto o temor. Ira o enojo. Sorpresa, que puede ser llamada asombro y, finalmente, asco al que otras personas le dicen desagrado o disgusto. En total son 6 emociones importantes y las vamos a conocer ***todas*** *a lo largo de nuestros juegos ¡espero te gusten mucho! [Pausa].*

¡Empecemos! En nuestro primer juego quiero que mires la foto de mi rostro y te fijes en tres partes ***muy*** *importantes: mis cejas, mis ojos y mi boca. ¿Las ves? [Pausa para que responda] ¡Genial! Ahora te toca a ti. Muéstrame ¿dónde están tus cejas? ¿Dónde están tus ojos? y finalmente ¿Dónde está tu boca? [Pausa para que las muestre] ¡Muy bien!*

Ahora, en la foto de mi rostro que estás viendo selecciona cada una de las partes que vimos: mis cejas, mis ojos y mi boca. No las olvides, repite conmigo: Cejas, Ojos y Boca. [Pausa] ¡Muy bien, continuemos! Quiero que mires nuevamente mi rostro muy atentamente porque te voy a mostrar los gestos para cada emoción, ¿Las recuerdas? [Pausa para que responda] ¡Muy bien!

Empecemos con la ***alegría****: mira que las cejas* ***no*** *se mueven, los ojos se hacen más* ***pequeños*** *y la boca hace una* ***sonrisa*** *donde puedes o no mostrar los dientes. ¿Viste lo que hice con mi rostro en la alegría? [Pausa] Ahora quiero que lo hagas tú. [Pausa para que lo haga] ¡Súper!*

En el gesto de la ***tristeza****, las cejas permanecen hacia* ***abajo****, los ojos se hacen* ***pequeños*** *y la boca hace una* ***curva*** *hacia abajo. Ahora inténtalo tú. [Pausa para que lo haga] ¡Genial!*

Mira muy bien la ***sorpresa****, porque las cejas se* ***suben*** *y los ojos y la boca se abren, haciendo ¡Aaaah! con la boca. Ahora hazlo tú. [Pausa para que lo haga] ¡Súper!*

Con el ***asco****, las cejas se* ***fruncen*** *así, los ojos se* ***cierran*** *un poco y la boca se abre porque la nariz se arruga. Algunas personas incluso muestran la lengua. Dale, ahora imítame tú. [Pausa para que lo haga] ¡Muy bien!*

Así como en el asco, en el ***enojo*** *las cejas también se* ***fruncen****, pero los ojos se vuelven* ***pequeños*** *y la boca se* ***cierra*** *quedando recta. Es tu turno de intentarlo. [Pausa para que lo haga] ¡Eso es!*

Por último, en el ***miedo****, las cejas se* ***alzan*** *y los ojos se hacen* ***grandes****, pero a diferencia de la sorpresa la boca no se abre, sino que se mantiene cerrada, ahora hazlo tú [Pausa para que lo haga].*

¡Felicitaciones, hemos terminado todos los gestos de las emociones, qué buen trabajo!

Anexo 3. Matrices de confusión correspondientes a la predicción de expresiones faciales por medio de la combinación de distintas bases de datos.

1. KDEF

Clase real \ Clase predicha							
	20	0	1	0	3	4	5
	2	0	24	3	9	7	0
	4	0	31	1	1	5	0
	0	0	0	42	1	0	0
	1	0	0	1	37	3	2
	4	0	3	0	4	25	1
	12	0	1	1	2	0	34

2. JAFFE

Clase real \ Clase predicha							
	4	0	0	2	0	4	0
	0	5	2	0	0	0	0
	1	0	9	0	0	0	0
	0	0	0	4	2	1	1
	1	1	0	0	4	3	3
	0	1	1	0	0	5	0
	1	1	0	0	0	1	6

3. CAFE

Clase real \ Clase predicha							
	24	18	0	4	1	0	2
	10	28	0	4	0	0	0
	0	0	0	0	4	0	8
	2	7	0	45	0	0	5
	1	3	0	1	38	0	1
	1	5	0	3	12	0	0
	4	0	0	0	3	0	21

4. KDEF+JAFFE

Clase real \ Clase predicha							
	20	1	3	1	8	1	9
	4	18	6	0	8	10	0
	2	8	34	2	4	6	0
	0	1	4	40	5	0	5
	1	1	0	0	50	0	3
	7	6	3	1	25	4	2
	6	0	0	0	0	0	48

5.KDEF+CAFE

Clase real \ Clase predicha							
	0	24	11	11	11	1	27
	0	25	44	9	6	3	2
	0	0	28	1	4	4	4
	0	4	0	103	0	1	3
	0	2	4	0	72	8	1
	0	5	11	5	34	6	0
	2	1	0	1	3	0	68

6. JAFFE+CAFE

Clase real \ Clase predicha							
	33	0	2	6	2	3	6
	25	0	0	2	1	24	1
	3	0	3	1	4	1	2
	10	0	2	58	0	2	1
	2	0	2	1	50	2	0
	5	0	0	1	16	5	1
	1	0	2	4	4	0	29

7. KDEF+JAFFE+CAFE

Clase real \ Clase predicha							
	34	18	8	12	6	16	15
	4	31	43	3	3	4	1
	2	5	36	5	4	1	7
	7	4	0	96	4	1	2
	0	5	4	0	72	6	4
	1	14	19	5	16	12	1
	9	0	1	1	1	4	64

Anexo 4. Exactitud de distintos modelos de aprendizaje automático cambiando los hiperparámetros de entrenamiento.

1. Exactitud de ANN para conjunto 1

Número de neuronas	2	3	5	10	20	50	100
Neutral	0.85	0.46	0.87	0.94	0.85	0.96	0.83
Enojo	0.00	0.74	0.75	0.71	0.74	0.47	0.74
Asco	0.85	0.62	0.80	0.60	0.74	0.26	0.43
Alegría	0.95	0.93	0.94	0.83	0.90	0.88	0.95
Sorpresa	0.91	0.97	0.94	0.89	0.94	0.83	0.97
Promedio	0.71	0.74	0.86	0.79	0.84	0.68	0.78

2. Exactitud de RF (2 muestras) para conjunto 1

Profundidad de los árboles	10			15			20		
Número máximo de árboles	50	100	1000	50	100	1000	50	100	1000
Neutral	0.92	0.92	0.89	0.92	0.90	0.90	0.90	0.90	0.90
Enojo	0.71	0.68	0.71	0.68	0.74	0.68	0.68	0.68	0.71
Asco	0.62	0.64	0.64	0.62	0.62	0.64	0.64	0.62	0.60
Alegría	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98
Sorpresa	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
Promedio	0.83	0.83	0.83	0.83	0.83	0.83	0.82	0.82	0.82

3. Exactitud de RF (3 muestras) para conjunto 1

Profundidad de los árboles	10			15			20		
Número máximo de árboles	50	100	1000	50	100	1000	50	100	1000
Neutral	0.92	0.92	0.92	0.92	0.90	0.90	0.90	0.90	0.90
Enojo	0.74	0.68	0.68	0.71	0.74	0.68	0.68	0.68	0.71
Asco	0.62	0.62	0.62	0.64	0.62	0.62	0.62	0.62	0.60
Alegría	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Sorpresa	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
Promedio	0.84	0.83	0.83	0.84	0.83	0.82	0.82	0.82	0.82

4. Exactitud de ANN para conjunto 2

Número de neuronas	2	3	5	10	20	50	100
Neutral	0.73	0.88	0.76	0.80	0.76	0.85	0.83
Miedo	0.83	0.76	0.83	0.81	0.69	0.69	0.79
Tristeza	0.60	0.56	0.65	0.63	0.58	0.77	0.67
Promedio	0.72	0.73	0.75	0.75	0.68	0.77	0.76

5. Exactitud de RF (2 muestras) para conjunto 2

Profundidad de los árboles	10			15			20		
Número máximo de árboles	50	100	1000	50	100	1000	50	100	1000
Neutral	0.78	0.78	0.78	0.80	0.80	0.78	0.78	0.78	0.78
Miedo	0.90	0.90	0.90	0.93	0.93	0.90	0.90	0.90	0.90
Tristeza	0.65	0.63	0.63	0.65	0.65	0.63	0.65	0.63	0.63
Promedio	0.78	0.77	0.77	0.79	0.79	0.77	0.78	0.77	0.77

6. Exactitud de RF (3 muestras) para conjunto 2

Profundidad de los árboles	10			15			20		
Número máximo de árboles	50	100	1000	50	100	1000	50	100	1000
Neutral	0.78	0.78	0.78	0.78	0.78	0.78	0.80	0.78	0.78
Miedo	0.90	0.93	0.90	0.90	0.90	0.90	0.93	0.90	0.90
Tristeza	0.63	0.63	0.63	0.63	0.65	0.65	0.65	0.63	0.65
Promedio	0.77	0.78	0.77	0.77	0.78	0.78	0.79	0.77	0.78

Anexo 5. Registro conductual de los participantes al utilizar Emmaciones.

Nombre del participante:

Sesión:

Actividad:

Emoción	Parte del rostro	Comportamiento	No	Si	Inicio de imitación (tiempo)	Final de imitación (tiempo)	Obser-vaciones
Alegría	Cejas/ frente	Neutrales					
	Ojos/ párpados	Ojos neutrales o relajados					
		Párpado inferior elevado causando que se entrecierren los ojos*					
	Parte inferior	Líneas nasolabiales de las mejillas [1]*					
		Bordes de los labios elevados					
		Mostrando dientes*					
		Sin mostrar dientes*					
Miedo	Cejas/ frente	Elevadas y juntas					
		Cejas aplanadas /horizontales					
	Ojos/ párpados	Abiertos					
		Aparente tensión en párpados inferiores					
		Puede que se vea la parte blanca superior de los ojos*					
	Parte inferior	Boca cerrada*					
		Boca abierta*					
		Borde de la boca y labios estirados sin hacer curva					
Asco	Cejas/ frente	Cejas abajo pero no juntas [2]					
		Ceño fruncido*					
		Ceño no fruncido*					
		Nariz se arruga					
	Ojos/ párpados	Parpado inferior elevado y no tensionado					
	Parte inferior	Líneas nasolabiales bastante pronunciadas					
		Boca abierta: labio superior hacia arriba, labio inferior hacia afuera*					
		Boca cerrada: labio superior hacia arriba*					
		Mostrando lengua*					
		Sin mostrar lengua*					

Tristeza	Cejas/ frente	Cejas juntas					
		Esquina exterior de cejas hacia abajo					
		Esquina interior de cejas elevada					
		Centro de cada ceja hacia abajo					
		Arrugas horizontales y verticales en el centro de la frente*					
		Ceño fruncido*					
	Ojos/ párpados	Ojos brillantes*					
		Párpados superiores caidos					
		Parpados inferiores relajados					
		Ojos mirando hacia abajo o con lágrimas*					
	Parte inferior	Boca abierta con labios parcialmente estirados*					
		Boca cerrada con curvatura hacia abajo*					
Sorpresa	Cejas/ frente	Curvadas hacia arriba					
		Arrugas horizontales en la frente					
	Ojos/ párpados	Ojos bien abiertos mostrando la parte blanca de los ojos arriba y abajo					
		Piel estirada en los párpados					
	Parte inferior	Boca abierta sin tensión en las esquinas de los labios					
Enojo	Cejas/ frente	Cejas hacia abajo y hacia adentro					
		Parte interior de las cejas sobresale					
		Puede tener arrugas curvas en el centro de la frente*					
	Ojos/ párpados	No se muestra la parte blanca de los ojos					
		Ojos entrecerrados					
		Puede producir una apariencia arqueda debajo de los ojos*					
		Puede que se presente movimiento rápido de lo ojos*					
	Parte inferior	Los labios pueden estar presionados entre ellos*					
		Boca cuadrada con los labios hacia arriba*					
		Mostrando dientes*					
		Sin mostrar dientes*					

Anexo 6. Comprobante de sometimiento de artículo al congreso UruCon 2021.

URUCON2021
CONSTRUYENDO FUTURO

IEEE URUCON 2021
ONLINE
MONTEVIDEO, URUGUAY
24 – 26 NOVEMBER

IEEE

URUCON

#27 (1570741636): Ensemble of Machine Learning Models for an Improved Facial Emotion Recognition

#27 (1570741636): *Ensemble of Machine Learning Models for an Improved Facial Emotion Recognition*

bib

Property	Change Add	Value
Conference and *track*		**2021 IEEE URUCON** - *2021 IEEE URUCON*

	Drag to change order	Name	ID	Edit	Flag	Affiliation (edit for paper)	Email	Country	Email	Delete	Register
Authors		Sergio D. Pulido	1655161	not creator		Escuela Colombiana de Ingeniería Julio Garavito & Universidad del Rosario, Colombia	sergio.pulido@mail.escuelaing.edu.co	Colombia			
		Nubia Palacios-Queca	1879708			Escuela Colombiana de Ingeniería Julio Garavito, Colombia	nubia.palacios@mail.escuelaing.edu.co	Colombia			
		Michelle Ballen-Cárdenas	1879709			Corporación Universitaria Minuto De Dios, Colombia	mballencard@uniminuto.edu.co	Colombia			
		Sandra Liliana Cancino	1655221	not creator		Escuela Colombiana de Ingenieria, Colombia	sandra.cancino@escuelaing.edu.co	Colombia			
		Alejandra Rizo-Arévalo	1879710			Corporación Universitaria Minuto De Dios, Colombia	arizoareval@uniminuto.edu.co	Colombia			
		Juan M. López López	624715			Escuela Colombiana de Ingeniería Julio Garavito, Colombia	juan.lopezl@escuelaing.edu.co	Colombia			

Title	*Ensemble of Machine Learning Models for an Improved Facial Emotion Recognition*
Abstract	The creation of algorithms that predict emotional recognition is a subject that has been of particular interest by researchers around the world for the last few years, as many computer vision-based systems make use of this information to get an approximation of the emotional state of an individual. This study aims to develop a real-time emotional recognition algorithm based on the facial expression. This algorithm was tested in a computational tool designed to stimulate the imitation and recognition of emotions in faces in children with Autism Spectrum Disorder. By designing an ensemble of machine learning models which separates emotions into different sets, we are able to improve the recognition accuracy. Additionally, the selection of relevant features greatly reduces the execution time of the algorithm, making it feasible for real-time recognition. Testing of different label combinations is yet to be performed in order to further improve the recognition accuracy.
Topics	Biomedical; Signal Processing
Status	Active (has manuscript)
	Can upload 4 pages (track) until Jul 31, 2021 23:59:59 EDT.

	Document (show)	Pages	File size	Changed	Check format / Report problem	Delete
Review manuscript		4	304,268	Jul 8, 2021 16:39:53 America/New_York	checked Jul 8, 2021 16:39:53 EDT	

pdf	authorinitials	Author Ms. Michelle Ballen-Cárdenas should appear with their full name, not just initials, in PDF file. -
pdf	gutter	The gutter between columns is 0.166 inches wide (on page 1), but should be at least 0.18 inches. -

Video Presentation: Can upload 16 minutes (track) until Nov 10, 2021 23:59:59 EST.

Personal notes

You are the creator and an author for this paper.

Reviews

Reviews are not yet visible to authors.

EDAS at 172.30.1.76 for 186.155.24.123 (Thu, 08 Jul 2021 16:41:57 -0400 EDT) [User 624715 using Win10:Firefox 89.0 cached 0.064/0.971 s] Request help