

**Identificación de Emociones Básicas en Niños con Dificultades del  
Lenguaje a Partir de sus Expresiones Faciales**

**Manuela Torrente Álvarez**

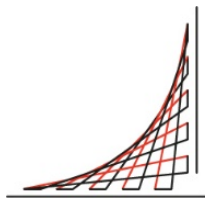
**Trabajo Dirigido**

**Tutor**

**Ing. Oscar Julian Perdomo Charry**



**Universidad del  
Rosario**



**ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO**

**UNIVERSIDAD DEL ROSARIO  
ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ D.C  
2022**

## **Agradecimientos**

Agradezco al apoyo de mis padres, mi hermana, amigos y compañeros. Mi especial y muy reconocido agradecimiento al Ingeniero Jefferson Muñoz por ser un gran soporte incondicional e ímpetu para la realización de este documento, persona del que aprendí y espero seguir haciendo.

Mis más gratas palabras a la Dra. Martha, al Sr. Segura y al Dr. Sánchez, personas que me han permitido conocer y admirar, para así ser la persona en la que me he convertido y del cual son mi inspiración para redactar este documento.

Reconocidos agradecimientos para el estudiante del programa de Ingeniería biomédica José Mateo Barón Restrepo por su gran labor de apoyo y acompañamiento durante el transcurso de estos 7 años de carrera.

Por último, les quiero agradecer a todos los profesores de la Universidad del Rosario y la Escuela Colombiana de Ingeniería Julio Garavito, en primera línea al Ingeniero Oscar Julián Perdomo Charry por su voto de confianza y la grata oportunidad de ser el guía en ésta área de mi interés.

## Resumen

Este proyecto pretende romper ciertas brechas sociales para la identificación de las seis emociones básicas y así tener una mayor asertividad y entendimiento con niños que presentan dificultades en el habla. El desarrollo del lenguaje y el habla en los niños se da en diferentes etapas del crecimiento, que en los subsiguientes años afecta de manera significativa en su comunicación y la expresión de sus emociones con otros individuos. Por otro lado, existen trastornos del lenguaje y el habla que surgen a partir de anormalidades genéticas, traumatismos cerebrales o pérdida de la audición, que no dificulta la comprensión del lenguaje, pero sí la emulación de expresiones coherentes.

A partir de esta premisa, se plantea el desarrollo de un algoritmo que permite identificar las 6 emociones básicas de niños con dificultades en el habla, a partir de un programa de Inteligencia Artificial o IA y una base de datos actualizada con videos cortos que contienen las expresiones espontáneas de 12 niños diversos entre sí, que comprenden las edades de 6 a 12 años.

Actualmente, se tienen muchos avances en cuanto al entrenamiento de algoritmos que permiten determinar y clasificar los gestos característicos para cierta emoción, pero sigue siendo un reto debido a la diversidad cultural, que genera una amplia brecha, por lo que se requiere el aumento de sujetos participantes para una base de datos lo suficientemente sólida.

Este programa, permitirá a los parientes más cercanos de estos niños con condiciones especiales, entender y dirigir la conducta del menor de manera más acertada, por lo que a futuro se espera implementar y validar con niños en situaciones reales o simuladas.

# Índice general

Agradecimientos . . . . .	I
Resumen . . . . .	II
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Historia y Contexto . . . . .	3
1.2. Marco teórico . . . . .	3
1.2.1. Lenguaje . . . . .	3
1.2.1.1. Definición . . . . .	3
1.2.1.2. El valor doble del lenguaje . . . . .	4
1.2.1.3. Competencias del lenguaje y estándares . . . . .	4
1.2.1.4. Cerebro y lenguaje . . . . .	5
1.2.1.5. Desarrollo del lenguaje y el habla . . . . .	6
1.2.1.6. Trastornos específicos del lenguaje . . . . .	6
1.2.2. Habla . . . . .	7
1.2.2.1. Definición . . . . .	7
1.2.2.2. Desarrollo y producción del habla . . . . .	8
1.2.2.3. Producción del habla . . . . .	9
1.2.2.4. Clasificación Trastornos del habla . . . . .	9
1.2.2.5. Tratamientos . . . . .	10
1.2.2.6. Tecnología . . . . .	10
1.2.3. Detección facial de emociones . . . . .	10
1.2.3.1. Producción de expresiones faciales . . . . .	10
1.2.3.2. Clasificación de las emociones . . . . .	11
1.2.4. Inteligencia Artificial . . . . .	11
1.2.4.1. Definición . . . . .	11
1.2.4.2. Machine Learning . . . . .	12
1.2.4.3. Deep Learning . . . . .	12
<b>2. OBJETIVOS</b>	<b>13</b>
2.1. General . . . . .	13
2.2. Específicos . . . . .	13
<b>3. METODOLOGÍA</b>	<b>14</b>
3.1. Base de datos . . . . .	14
3.1.1. Descripción y Composición . . . . .	14
3.1.2. Participantes . . . . .	14

3.1.3. Estimulo de emociones . . . . .	15
3.1.4. Métodos y materiales . . . . .	15
3.1.5. Método de adquisición de videos de la base de datos . . . . .	15
3.2. Programación . . . . .	16
3.2.1. Detección del Rostro . . . . .	16
3.2.2. Segmentación de videos a imágenes . . . . .	16
3.2.3. Entrenamiento red neuronal . . . . .	17
3.2.4. Librería y funciones . . . . .	18
3.3. Resultados . . . . .	18
<b>4. RESULTADOS</b>	<b>19</b>
<b>5. DISCUSIONES</b>	<b>22</b>
<b>6. CONCLUSIONES</b>	<b>26</b>
<b>7. RECOMENDACIONES Y TRABAJOS FUTUROS</b>	<b>27</b>

# Índice de figuras

3.1. Imágen que contiene las 6 emociones básicas de algunos de los 12 participantes que implementó la base de datos de LIRIS-CSE. [40] . . . . .	14
3.2. Implementación del filtro Haar para la detección de rostros mediante clasificadores en cascada basado en Kernel de Convolución. . . . .	16
4.1. Gráfica de precisión, sensibilidad y puntaje F1 . . . . .	21
4.2. Gráfica de precisión, sensibilidad y tamaño de cada una de las muestras. . . . .	21
5.1. Imágenes base de datos para emoción de alegría en diferentes sujetos, participantes del estudio. . . . .	22
5.2. Imágenes base de datos para emoción de miedo de los sujetos S2,S3 y S4 . . . . .	23
5.3. Imágenes base de datos para emoción miedo sujeto S8 . . . . .	24
5.4. Micro expresiones descritas por el estudio de Paul Ekman para emoción de miedo y recreadas en el estudio tomado de [18]. . . . .	24

# Índice de tablas

3.1. Resumen de estímulos implementados para evocar la emoción en los participantes del estudio de la la base de datos de LIRIS-CSE. [40] . . . . .	15
3.2. Parámetros de los videos grabados para la base de datos [40]. . . . .	15
3.3. Características de modelo MobileNet de Keras. . . . .	17
3.4. Frames por emoción capturado en el estudio [40]. . . . .	17
4.1. Matriz de confusión del algoritmo implementado para la detección de las 6 emociones básicas. . . . .	19
4.2. Resultados de precisión, sensibilidad y F1 de Matriz de Confusión para cada emoción. . . . .	20

# Capítulo 1

## INTRODUCCIÓN

El instinto del ser humano es el resultado de una respuesta ante las amenazas del entorno propias de la especie. Gracias a este proceso, en conjunto con los cambios neurobiológicos que este trajo consigo tales como: el desarrollo encefálico, el asentamiento y el origen del sentido de comunidad y convertirse en una especie social, desarrolló el habla, el proceso cognitivo y otros sistemas de comunicación superiores, recursos que se convirtieron indispensables para comunicación entre la especie Homo [1].

Dentro de este sistema complejo de comunicación se encuentran las expresiones faciales, que forman parte del lenguaje no verbal, estas cuentan la historia de la universalidad de las emociones las cuales presentan los mismos patrones sin importar la etnia, raza o posición geográfica. Las expresiones faciales son innatas según lo expresa en su artículo "Lectura de la expresión facial de las emociones": las emociones y sus manifestaciones eran biológicamente naturales, evolutivamente adaptables y con semejanzas existentes las cuales podrían analizarse desde un punto de vista relacional y evolutivo [2]. Se han publicado múltiples estudios en los que se ha demostrado que cuando las emociones se generan de manera espontánea, se logra identificar las mismas expresiones faciales en diferentes individuos [3].

Otras investigaciones suscitan que las expresiones faciales evocadas por las emociones básicas poseen una base genética y se logra demostrar en estudios realizados con personas que presentan ceguera, de las cuales se logra distinguir las mismas expresiones faciales que personas videntes, lo cual permite entender que estas no son producto netamente de la interacción y la mimetización con el medio siendo este un canal y no un medio [4,5,6].

La necesidad de comunicarnos con nuestro entorno se desarrolla a partir de los primeros tres años de vida, mediante la constante exposición a una sociedad que implementa imágenes, sonidos, el habla y el lenguaje de sus parientes o pares para lograr una comunicación efectiva. [7] Por otro lado, existen los trastornos del lenguaje y el habla que se definen como la dificultad de la articulación y verbalización en la comunicación debido a la inhabilidad de comprender, interpretar y/o expresar pensamientos, sentimientos y emociones [8].

El NIH indica que en Estados Unidos uno de cada doce niños en edades comprendidas de 3 a 17 años, es decir, el 7,7% ha tenido un trastorno en la voz, el habla, del lenguaje o deglución en los últimos 12 meses del año 2016. Además, se estima que los menores más privilegiados



en atención oportuna eran de raza blanca y hacían parte de más de la mitad de los niños que presentaron estas dificultades, esto demuestra las brechas sociales que existen en comparación con la atención a niños hispanos y de raza negra residentes en los EEUU [9].

En contexto, la Universidad Nacional de Colombia realizó una investigación en el año 2014 con niños de uno a tres años de edad y allí determinó que el 30% de los menores presenta un déficit en el lenguaje, además se aproxima que el 7% presenta un trastorno específico del lenguaje, que, a pesar de tratarse con terapias especiales, estos tienden a mantenerse con el tiempo e inclusive a desarrollar otros trastornos para la comunicación en edades más avanzadas [10].

La comunicación en general es más que la habilidad del habla, a pesar de ser la forma de comunicación más utilizada y eficaz para transmitir información con significado, el habla se produce por medio de una secuencia compleja comprendida por la coordinación, planificación, ejecución y secuencias motoras, que, relacionado con el objetivo de población principal, existen personas que a edades tempranas presentan dificultades en alguna de estas habilidades que imposibilitan la producción del habla [11]. Una de las ventajas principales como seres humanos es que nuestra comunicación no sólo se basa en la articulación organizada de sonidos, que producen fonéticamente palabras y frases complejas para expresar lo pensamientos, ya que se data que millones de años antes del uso de la comunicación Verbal, para los inicios del contacto entre semejantes se hacía uso de la comunicación No Verbal [1].

Esta comunicación se dio por medio de las expresiones faciales y las micro expresiones, estas dan respuesta a las emociones que se desarrollan desde el nacimiento, a partir del llanto como medio para expresar las necesidades básicas o sensaciones primarias, a medida que crece el infante, se incrementa la interacción social y la conexión con sus pares, denotando las primeras sonrisas a lo largo de los primeros seis meses. Estas expresiones faciales vendrían siendo una respuesta a estímulos emocionales. La comunicación No Verbal o CNV es aquella paralingüística que implementa sonidos, movimientos corporales, la postura corporal y gestos faciales para acompañar la comunicación verbal, ya sea para reafirmarla o contradecirla. De manera involuntaria, la Comunicación No Verbal constantemente transmite algún tipo de información, que, a diferencia de la Comunicación Verbal, se considera como la más auténtica e instintiva, debido a la aparición cronológica ya que acompaña al origen de la especie Homo, que involucra el uso del cuerpo para comunicarnos con nuestros pares e incluso con otras especies, por ejemplo los gestos, movimientos corporales o sonidos que implementamos para entrenar otro ser vivo del reino animal [12,13].

Las expresiones faciales han sido un debate en cuanto a la universalidad de la locución de las emociones para diferentes culturas, pero para la neuropsicología de las emociones: el aporte de Charles Robert Darwin con la publicación del libro "Expresión de las Emociones en los Animales y en el Hombre" en el año 1872 [14], que concluye con dos afirmaciones, la primera indica que las expresiones de ciertas emociones son innatas y universales, la segunda afirma que nuestras emociones son producto de la evolución y por este motivo ha sido transmitida a través de los años [14,15]. Dos siglos más tarde aparece Paul Ekman que plasma en 1970 su trabajo: "Universal Facial Expressions of Emotions" en dónde además de reafirmar la teoría Darwiniana, crea un esquema de validez universal denominado FACS por sus siglas en inglés

Facial Action Coding System [16].

El objetivo principal de este proyecto es desarrollar un algoritmo que en futuros trabajos se puedan implementar en diferentes dispositivos y aplicativos, para la detección de las seis emociones básicas a partir de los gestos del rostro en niños que presentan dificultades en la expresión oral comunicativa de sus emociones, permitiendo una misma más asertiva. En conjunto con la idea de universalidad y para el reconocimiento de las seis emociones básicas que se manejan en el siguiente trabajo: enojo, disgusto, tristeza, miedo, alegría y asombro, se necesita de diferentes áreas de conocimiento como neurociencias, reconocimiento de patrones y señales biométricas, que, en este caso, son el rostro de niños en edades comprendidas entre los 6 y 12 años. Actualmente el sistema de detección de objetos más utilizado es la librería multiplataforma de OpenCV (Open Computer Vision) que se puede implementar con lenguajes de programación Python [17].

## 1.1 Historia y Contexto

Con el ser humano, también nació la comunicación de este. Por lo tanto, se data que para el año 2'500.000 AC se dio la aparición del lenguaje no verbal como consecuencia de la llegada de la evolución del género Homo en África, en sus inicios el ser humano se comunicaba a través de sonidos y gestos corporales. Existen análisis realizados a cráneos fósiles del género Homo Erectus de hace unos 400.000 años, en los que se logró demostrar que ya contaban con las áreas de Brocca y Wernicke (asociadas al lenguaje) y que su anatomía ya hacía viable la generación de sonidos [1].

Por otro lado, diferentes teorías indican que entre los años 70.000 AC y 30.000 AC se da la revolución cognitiva y la aparición del lenguaje ficticio. Gracias a la necesidad del ser humano de crear el lenguaje Verbal como medio de comunicación, en sus inicios se realizó la creación de nombres de animales a partir del sonido que emitían, para el caso de los objetos inertes que no emitían ningún tipo de sonido, se implementaba el sonido que generaba su uso. Esto permitió la evolución del género Homo Sapiens, para alcanzar avances históricamente importantes como la colonización de continentes, revolución agrícola y domesticación de animales y plantas [1,18].

## 1.2 Marco teórico

### 1.2.1. Lenguaje

#### 1.2.1.1 Definición

El lenguaje se define como la habilidad que tiene el individuo para expresar emociones, sentimientos o pensamientos en su entorno. El lenguaje se compone de dos ramas importantes: El Leguaje Verbal y No Verbal. El primero se enfoca en la gramática, la estructura, generativismo y la lingüista textual, el lenguaje No Verbal se enfoca en la kinésica y la proxémica, esta se enfoca en los gestos faciales, movimientos del cuerpo, especialmente las manos, los ojos y la postura con la que se dirige [19]. Como se menciona anteriormente, el Lenguaje No Verbal

puede confirmar o negar la información dada.

Chomsky, lingüista creador de la teoría del desarrollo del lenguaje, plantea que el ser humano nace con una capacidad natural para el habla. Es capaz de aprender y comprender estructuras lingüísticas y comunicativas. Siendo este un medio por el cual se transmiten pensamientos, también indica que el ser humano tiene la capacidad de aprender el lenguaje sin necesidad de tener las reglas gramaticales, esto se evidencia en niños de cuatro años o menos, que tienen la capacidad de expresar emociones, sentimientos y pensamientos de acuerdo con la estructura de la lengua que los rodea desde la etapa prelingüística que comprende los primeros 10 a 12 meses, a la etapa lingüística que es después del primer año de edad [20,21]. Igualmente, Piaget, padre de la psicología evolutiva, planteaba dentro de su tesis de adquisición del lenguaje social que, el lenguaje se vuelve cada vez más complejo a medida que el desarrollo encefálico y la inmersión en un entorno social se va complejizando en un individuo.

### **1.2.1.2 El valor doble del lenguaje**

El lenguaje representa una capacidad fundamental del ser humano, caracterizado por tener un valor doble. El primero asociado a la individualidad y el segundo al contexto social. Por ende, el lenguaje tiene un valor subjetivo asociado al individuo, debido a que es una herramienta que contribuye a la apropiación de la realidad, al igual que a la diferenciación y representatividad en múltiples contextos, implicando distinción individual pero también integración sociocultural.

### **1.2.1.3 Competencias del lenguaje y estándares**

Los estándares del lenguaje se refieren a sistemas simbólicos entre los que se destacan la lengua castellana y la literatura [22]. También se contemplan como competencias, la lingüística, estudios en cuanto a los procesos de comunicación y exploraciones sobre las funciones del lenguaje [23].

El lenguaje tiene una serie de implicaciones para el desarrollo del ser humano, debido a que le permite la posibilidad de establecerse como persona o individuo, a partir de una caracterización distintiva y diferenciadora que simultáneamente facilita el reconocimiento y la integración socio-cultural al entorno en el que se encuentra o pertenece, favoreciendo la participación, construcción e incluso transformación de su círculo. Con base en lo anterior, se soporta la relevancia del lenguaje como un elemento determinante en la realidad del ser humano, que conlleva a la comprensión, conceptualización y apropiación de la realidad que lo rodea, lo que consecuentemente también favorece la posibilidad de conceptualización y representación. Tal representación parte del relacionamiento conceptual haciendo uso de imágenes, figuras, formas, sonidos o señales corporales para maximizar el entendimiento y la manifestación.

En este sentido, dada la magnitud del relacionamiento entre formas o conceptos y el lenguaje, se ha acrecentado la conceptualización del pensamiento y se ha acelerado su crecimiento,

construyendo un universo conceptual cada vez más amplio. Por lo tanto, es contemplando esto que los seres humanos caracterizan su realidad individual, social y cultural.

Contemplando el valor social del lenguaje, este se establece, a través de sus diferentes expresiones, en pilar y argumento de la interacción social. Es por las distintas formas de comunicación como la escritura y el lenguaje que los seres humanos se relacionan e interactúan mutuamente, debaten opiniones, resolucionan problemas, describen eventos, objetos o individuos, cuentan historias y realizan muchas más actividades, lo cual tiene implicaciones a nivel individual y social.

Como mecanismo adaptativo y herramienta del lenguaje el ser humano crea las micro expresiones faciales y corporales, marcadores estandarizados por los cuales existe una identificación y conexión emocional con el otro, esto debido a las neuronas que son activadas al percibir a un individuo realizar una acción o al realizarla, es decir, su activación depende de la acción propia o ajena.

Cuando el individuo ejecuta una acción, las neuronas espejo son activadas y almacenan lo percibido, pero también se activan al observar la realización de una acción pese a la no realización directa. Es decir, estas neuronas se activan para los casos en que una persona corre y o cuando sólo se encuentra como observador de un hombre ejecutando la acción. Las neuronas espejo, al activarse por el comportamiento de otros, permiten que los individuos imiten, aprendan y sientan empatía. El desarrollo del sistema límbico que se especificará a más adelante permitió que se crearan patrones estandarizados de la expresión emocional.

#### **1.2.1.4 Cerebro y lenguaje**

Entendemos en este punto que el lenguaje y el desarrollo encefálico son las dos caras de una misma moneda, ya que no podemos definir con exactitud cuál de ambos procesos es primero. Es muy bien conocido que la especie humana desarrolla el lenguaje mediante el contacto con sus pares y mediado por el ambiente social, sin embargo, esto no es del todo posible sin la predisposición genética para la adquisición de esta.

Es preciso clarificar que las estructuras cerebrales tienen unas etapas de formalización del lenguaje evocativo y articulado las cuales se detallarán a continuación, allí mismo se explica cómo poco a poco el ser humano va afrontando las etapas de desarrollo, hasta que la estructura del lenguaje se forma por completo de manera que el individuo puede desenvolverse en un entorno social complejo.

El estudio que puede explicar el proceso bidireccional del lenguaje y el cerebro es el caso del niño de aproximadamente once años que a finales de 1799 en Francia, es encontrado por un grupo de cazadores en la intemperie y aparentemente se había criado en medio de la naturaleza y sobrevivido por sus propios medios. Tras un periodo de tiempo en el hospital, hasta que se determinó como una oportunidad para diferentes estudios. Congruentemente, el director del asilo decidió examinar al menor de edad y determinó un déficit mental de nacimiento incurable. Sin embargo, el joven médico Jean Marc Gaspard, quien también fue partícipe en

los exámenes, determinó que no se era el caso de una deficiencia mental, sino de un individuo de inteligencia promedio, que debido al total aislamiento de un entorno humano no había tenido la posibilidad de desarrollar las capacidades mentales, entre las que se encuentra el uso y entendimiento del lenguaje. Por lo anterior se proponen tratamientos, para observar si era posible recuperar dichas capacidades del lenguaje que el niño jamás desarrolló.

Las estructuras cerebrales implicadas en el lenguaje se encuentran en las áreas de Broca, Wernicke y Circunvolución angular, las cuales se caracterizan como las 3 zonas de mayor importancia cerebral [24].

**a. Área de Broca:** está relacionada con la habilidad de crear palabras y oraciones mediante el orden lógico de fonemas, por lo que constituye un miembro fundamental de la parte instrumental del lenguaje, y es de suma importancia en los procesos de interacción y comunicación. Consecuentemente, si presenta daños, se genera una dificultad sintáctica.

**b. Área de Wernicke:** contribuye a la manifestación y vocalización de conceptos, conjuntamente con el procesamiento de sonidos para la generación de combinaciones compuestas con significado. Adicionalmente, al igual que área de Broca, también forma parte del sistema instrumental de lenguaje.

#### **c. Circunvolución angular**

La circunvolución angular es uno de los pliegues cerebrales con gran relevancia en la regulación del lenguaje y también se le conoce como giro angular. Sin embargo, no es el único pliegue debido a que el cerebro cuenta con un gran número de estos, los cuales cuentan con funciones de alta importancia. Adicionalmente, el giro angular también es partícipe en las actividades matemáticas, de atención espacial y memoria.

### **1.2.1.5 Desarrollo del lenguaje y el habla**

Como se mencionó anteriormente, en la teoría del desarrollo planteada por Noam Chomsky, afirma que los niños tienen la habilidad innata para la comprensión de la gramática del lenguaje y el desarrollo de esta mediante un sistema implícito de comunicación. Con base en la anterior premisa se determina que el lenguaje es el medio que nos proporciona la capacidad para relacionarnos en una sociedad desde el nacimiento.

### **1.2.1.6 Trastornos específicos del lenguaje**

**Trastornos específicos del desarrollo del lenguaje** alteraciones que dificultan la capacidad comunicativa, tanto en el habla como de comprensión para poder comunicarse con otros seres humanos en su entorno. Estas dificultades pueden ser de tres tipos [25,26,27]:

**Tipo 1:** Del lenguaje receptivo: se trata de la incapacidad de comprensión.

**Tipo 2:** Del lenguaje expresivo: hace referencia a la dificultad en la capacidad de expresar pensamientos e ideas.

**Tipo 3:** Mixtos: comprenden la incapacidad tanto para hablar como para entender correctamente.

## **Trastornos psicolingüísticos**

A continuación se mencionan tres de los más comunes trastornos psicolingüísticos:

### **Trastorno del espectro autista**

Los trastornos del espectro autista o TEA es la discapacidad que se da por causas genéticas que generalmente se ve reflejado en las dificultades para lograr una comunicación e interacciones sociales con el entorno, en ciertas ocasiones se presentan comportamientos represivos o repetitivos [28]. Las personas con TEA por otro lado, suelen tener maneras diferentes para adaptarse y aprender, moverse o atender instrucciones específicas, esto ocasiona dificultades para las personas con TEA la comunicación en su entorno por su forma de interpretar la información, que usualmente suele ser abstracta, efímera y contener alteraciones. Esto incluye el hecho que su condición les dificulte no sólo en la comprensión sino también en la manera de enunciar pensamientos, sentimientos y emociones.

- a. Dificultad en la interpretación de la comunicación no verbal.
- b. Dificultad en la comprensión de formas no literales para comunicar cierta información.
- c. Dificultad en la comprensión y/o expresión de mensajes que se deseen expresar.
- d. Uso de un lenguaje rebuscado y poco común.

Cabe aclarar que algunas personas sin un diagnóstico de TEA pueden presentar alguna de estas condiciones previamente mencionadas.

### **Trastorno de Asperger**

Los individuos con este trastorno presentan aparentemente poco interés por interactuar con otros, lo cual implica complicaciones para estructurar relaciones sociales. Lo anterior se puede deber a la dificultad para sostener el contacto visual, la deficiente articulación de gestos y la no comprensión de emociones ajenas.

**Mutismo selectivo** Se trata de un padecimiento que genera que un individuo deja de hablar de forma repentina, pese a poder hacerlo con normalidad, lo cual se da usualmente en entornos sociales como el colegio u otros ambientes. La causa científica no se ha logrado identificar con claridad, sin embargo, la neuropsicología y psicología clínica, hablan de bases ansiosas e inhibidas genéticamente lo cual produce una vulnerabilidad en quien lo padezca.

## **1.2.2. Habla**

### **1.2.2.1 Definición**

Se define como habla al acto fonético que se produce de manera organizada, equiparado por el conjunto motor mecánico articulatorio [29]. Por medio del habla se transmiten pensamientos, emociones o cualquier otro tipo de información que se desee transmitir.

En el habla intervienen tres partes importantes: la primera es el acto fonético que implica la emisión del sonido, el segundo es el acto fático que permite la transformación de esos sonidos en términos con sentido y el tercero es el acto réctico que pretende la formación de frases coherentes a partir de la estructura gramatical adquirida.

### 1.2.2.2 Desarrollo y producción del habla

El habla como se define anteriormente es el medio verbal que permite la comunicación de un individuo en su entorno y que requiere de factores importantes como la coordinación, planificación y ejecución que involucra la intervención Mecánico-articulatoria. A pesar de que existe un nivel aceptable de variabilidad para el normal desarrollo del habla en niños, la mayoría cumplen con los periodos de desarrollo según la edad de vida [11].

#### **Periodo prenatal (0-1 mes)**

Involucra el desarrollo que se da a partir de la succión de los bebes que permite cerrar el acceso a la cavidad nasal y así evitar la broncoaspiración. Además del aumento en los movimientos mandibulares en donde el recién nacido empieza con la elevación y descenso, hasta llegar a diferentes grados de libertad (adelante-atrás, derecha e izquierda), en cuanto al desarrollo auditivo todavía no es lo suficiente maduro para coordinar las señales que recibe, pero cabe resaltar que ante ciertos estímulos recibidos por parte de la madre, en esta etapa el individuo se expresa mediante el llanto, mecanismos vocálicos.

#### **Periodo posnatal (1-6 mes)**

El aumento del sentido de la vista y el auditivo mejora la comunicación en el entorno, se muestran la evolución de la sonrisa, que se llega a desarrollar en risas por lo que se producen consonantes, con resonancia incompleta. Durante este periodo de tiempo el menor aumenta su capacidad de imitación de gestos, sonidos o balbuceos y el desarrollo del movimiento independiente de los labios que a medida que crece se prolongan y definen aún más estos sonidos.

#### **Periodo primer año (7-12 mes)**

Durante esta etapa de crecimiento se incrementa la composición y emisión de sílabas repetitivas y los balbuceos que evolucionan a vocalizaciones que son acompañados por gestos corporales. Ya acercándose al primer año de vida este es capaz de prestar atención a las conversaciones de pares o padres que involucren el enunciado de su nombre.

#### **Periodo segundo año (12-24 meses)**

En este largo periodo de tiempo se evidencia una marcada diferencia entre el desarrollo del lenguaje y el habla de niños y niñas, siendo estas últimas las que manejan un aproximado de aprendizaje entre 4 y 6 palabras, el aumento del vocabulario se da gracias a la necesidad de comunicar lo que desean alcanzando un total de 300 a 600 palabras, todavía persiste el uso de gestos para señalar objetos. A finales de los 24 meses, el infante desarrolla mejor la composición de frases gracias a la mejora en la captación de información de la memoria de corto plazo.

#### **Periodo infancia prematura (3-5 años)**

El aumento progresivo en el vocabulario se da de las 1.000 a 2.200 palabras, de las cuales ya el infante es capaz de implementar por lo menos 12.000 palabras diarias con una sintaxis y gramática correctamente compuesta de por lo menos 3 o 4 palabras. Se explora e implementan con mayor frecuencia los interrogantes, frases negativas e imperativas. Ya para los últimos meses antes de cumplir los 5 años de vida, el infante es capaz de comprender e implementar el 80 % de las estructuras sintácticas que implementarán en la etapa adulta.

### 1.2.2.3 Producción del habla

Para lograr la producción del habla, en primer lugar, se pueden clasificar los órganos que intervienen para que se dé la producción del habla y se puede dividir en dos grupos: Los órganos activos que son los labios, la lengua y el paladar blando, por otro lado, tenemos los órganos pasivos como los dientes, el alvéolo y el paladar duro. Por otro lado, se tiene el sistema respiratorio que es el impulsador para que se produzca como tal el sonido mediante la expulsión de aire.

### 1.2.2.4 Clasificación Trastornos del habla

Como se indica anteriormente la diferencia entre las dificultades del habla y el lenguaje radica en que la primera puede estar contenida dentro de la segunda, en donde las dificultades en el habla se limitan al impedimento que se presenta a la hora de emitir fonéticamente y sin interrupciones la información, emociones o sentimientos que deseen expresar. Por otro lado, el lenguaje puede incluir además la falta de comprensión, desarrollo de las ideas y emisión de la información a expresar.

Esto puede afectar de manera significativa el comportamiento y el desarrollo social en el que se desenvuelve un individuo, así como el lenguaje verbal y no verbal que expresa.

A continuación, se muestra cómo se clasifica cada una de estas dificultades basado en Chevrie-Muller y Narbona [30].

#### **Deficit Auditivo**

Es la pérdida parcial o completa de la audición debido a la perturbación o daño en los órganos del oído interno o inclusive en el nervio auditivo.

#### **Deficit Mecánico Articulatorio**

**I.** Disglosias Existen de diferentes tipos y depende de la parte articuladora o de los órganos buco fonatorios afectados, que pueden ser algunos de los que se muestran a continuación, como: Disglosia Labial, disglosia Mandibular, disglosia Lingual y disglosia Palatina [31,32].

#### **Los trastornos adquiridos de la comunicación**

**I.** Disartrias: se refiere a un trastorno específico en el habla debido a la parálisis, falta de fuerza o coordinación en la musculatura que constituyen los órganos que intervienen en el habla.

**II.** Afasias: son el deterioro adquirido por lesiones en zonas cerebrales que afectan la receptividad y expresividad del lenguaje hablado.

**III.** Apraxia: es un trastorno específico para la coordinación motora y fonética que resulta en la segmentación intra e inter articuladora, sin deficiencias primarias en el procesamiento de la información sensorial o del lenguaje

#### **Otros trastornos**

**Dislalias** es una dificultad que se presenta en la pronunciación debido a algún problema articulatorio o errores en dicción de ciertos fonemas.

**Trastornos del ritmo y la fluencia** es una dificultad en el habla en que una persona repite



sonidos, sílabas o palabras y en casos más severos se alargan estos mismos. Esto causa una disrupción en la fluencia normal del lenguaje.

#### **1.2.2.5 Tratamientos**

En la actualidad se cuenta con diferentes intervenciones a temprana edad para el correcto manejo de dificultades del habla y el lenguaje, entre ellas está el tratamiento mediante especialistas como que incluyen a pediatra, neuropediatría, logopeda, psicólogo y maestros escolares.

#### **1.2.2.6 Tecnología**

En el medio tecnológico se tiene una amplia gama de dispositivos (de apoyo auditivo, interactivo y aprendizaje), aplicativos e inclusive robots con aplicaciones interactivas que permiten el acompañamiento en la mejora del desarrollo del habla y el lenguaje en edades tempranas. El propósito de este proyecto radica en el desarrollo de un algoritmo que en futuros trabajos se puede implementar en trabajos como los que realizan actualmente, en donde se implementan robots para el acompañamiento, como lo es el robot Aiko, considerado como el robot “afectivo” . Aiko está basado en un algoritmo de IA emocional y su propósito es el acompañamiento emocional, permitiendo mejorar la comunicación. Este dispositivo tuvo un impulso acrecentado después del año 2020 y las consecuencias de la pandemia mundial lo cual generó el incremento de problemas derivados de la salud mental, en el proyecto propuesto se pretende implementar este algoritmo para poder aplicarlo a futuro como un sistema de retroalimentación a la emoción generada en el paciente.

### **1.2.3. Detección facial de emociones**

#### **1.2.3.1 Producción de expresiones faciales**

En primera instancia se detalla que la manifestación de las micro expresiones se da de modo semejante en los individuos. Consecuentemente, todas las personas, independientemente de su interiorización cultural, su crianza, desarrollo o la forma en que se haya dado su niñez, presentan gestos comunes, como elevación de comisura de labios y mejillas para denotar una sonrisa, apertura del contorno visual y la boca para denotar asombro [18].

Contemplando la manifestación de micro expresiones, se afirma que existe un grupo de emociones generalizadas que se relacionan estrechamente con leves gestos musculares del rostro, generalmente sostenidos en el tiempo, como elevación simultánea de las cejas como asombro en respuesta a un estímulo de sorpresa o el declive que se presenta en la comisura de los labios en manifestación de tristeza. Cada uno de los gestos suele ser involuntario e incluso difícil de percibir y se encuentran altamente relacionados con la demostración de las emociones de los individuos [33].

Congruentemente, considerando la existencia de emociones básicas y características extrapolables para cada una, es posible aseverar la viabilidad para detectar o reconocer las

emociones a partir de la gesticulación facial de los individuos.

### 1.2.3.2 Clasificación de las emociones

El entendimiento de las emociones se hace necesario para la comunicación efectiva y asertiva, la toma de decisiones e incluso para fines reproductivos. Considerando la importancia de las emociones y su comprensión, Paul Ekman logró distinguir incluso diez mil gestos diferentes. Conjuntamente Paul Ekman y Wallace Friesen, clasificaron dichas expresiones en un sistema denominado FACS (Sistema de codificación facial), contemplando el comportamiento anatómico de los músculos del rostro [34].

Ahora bien, resulta indispensable relacionar las micro expresiones con las seis emociones básicas [18], con el fin de poder reconocer cada una de las emociones a partir de su caracterización en términos de la gesticulación facial. A continuación, se detallan las micro expresiones relacionadas o correspondientes con las seis emociones básicas:

**Alegría:** Se elevan la comisura de los labios y mejillas, se arruga la piel que se encuentra en el contorno de los ojos, entre ellos el surco orbito palpebral inferior y la comisura lateral de los ojos. Por último, micro expresiones en arrugas en nariz y labio superior.

**Disgusto:** Se genera de manera asimétrica un marcado ascenso del labio superior y se presentan arrugas en el entrecejo, párpados inferiores, nariz y áreas próximas al labio superior. En conjunto movimiento de elevación de mejillas y declive en las cejas

**Enojo:** Las cejas realizan un movimiento hacia abajo y contraídas hacia el entrecejo. la boca cerrada o entreabierta para denotar ira y parte ocular prominente.

**Miedo:** Se presenta una suave elevación y contracción de las cejas hacia el centro, ambos párpados elevados, tensión en labio y/o presenta boca entreabierta.

**Sorpresa:** Se presenta un marcado ascenso de las cejas, que presentan una forma semicircular y se presenta un alargamiento del rostro de manera vertical, de manera que los párpados y la mandíbula se extienden.

**Tristeza:** el contorno de los ojos desciende, así como la comisura de ambos lados del labio. En ciertas ocasiones se presenta un estiramiento en la mandíbula.

## 1.2.4. Inteligencia Artificial

### 1.2.4.1 Definición

Se considera la Inteligencia Artificial como un campo de la computación que se encarga del entrenamiento y aprendizaje automático de las máquinas mediante la programación de algoritmos que reciben información de entrada clasificada de manera ordenada y así la máquina tenga la capacidad de identificar y ordenar los datos mediante los patrones adquiridos en la

entrada a partir de los grupos información que hay en común entre ellos. En pocas palabras se pretende que las máquinas tengan la capacidad autonomía en toma de decisions como lo es la clasificación de una característica.

Dentro de los conocimientos de IA se encuentra el Aprendizaje automático o machine learning cómo técnica especializada y derivada de la inteligencia artificial específicamente para el entrenamiento de estas máquinas, que subsecuentemente el deep learning es un subconjunto del Machine Learning o entrenamiento por redes neuronales.

#### **1.2.4.2 Machine Learning**

Conjunto de herramientas estadísticas y matemáticas que aportan valor con datos mediante la exploración, interpretación y diferentes algoritmos, como en este caso, el entrenamiento mediante una base de datos que contiene las imágenes de niños con las expresiones faciales de las 6 emociones básicas de entrada, de tal manera que se pueda obtener la predicción.

#### **1.2.4.3 Deep Learning**

Son algoritmos específicos que utilizan Redes Neuronales para la predicción de eventos mediante la construcción de árboles de decisión o modelos para el aprendizaje supervisado. Para los casos en los que no se sabe cuál es el output por la cantidad desmesurada de datos en el estudio, se implementa el aprendizaje no supervisado.

## Capítulo 2

# OBJETIVOS

### 2.1 General

Desarrollar un algoritmo que permita identificar las 6 emociones básicas en niños con dificultad en el habla, a partir de una base de datos de expresiones faciales auténticas de 12 niños diversos entre sí.

### 2.2 Específicos

1. Objetivo Específico 1: Examinar y establecer la información bibliográfica para la construcción del estado del arte basado en los criterios de inclusión y exclusión.
2. Objetivo Específico 2: Desarrollar un programa que permita identificar las expresiones faciales en niños no patológicos durante la emisión de emociones.
3. Objetivo Específico 3: Interpretar y comparar los resultados obtenidos del algoritmo implementado y los resultados de la bases de datos.

## Capítulo 3

# METODOLOGÍA

### 3.1 Base de datos

#### 3.1.1. Descripción y Composición

La base de datos en la que se basa este documento es LIRIS-CSE es un conjunto de 208 videos que contienen las expresiones auténticas de las 6 emociones básicas de 12 niños en edades comprendidas de los 6 a 12 años, que son étnicamente diversos. Además, se considera la primera base de datos que contiene las expresiones no estandarizadas de niños expresando en 208 videos clasificados algunas de las 6 emociones básicas, con un total de 26.000 frames de video [40].

#### 3.1.2. Participantes

Es un total de 12 niños (cinco niños y siete niñas), étnicamente diversos, en edades comprendidas entre los 6 y 12 años, con una edad promedio de 7,3 años. El 60% de los videos fueron realizados en el aula de clase y el restante desde casa, esto con el fin de obtener imágenes con diferentes fondos y condiciones de iluminación como se muestra en la Figura 3.1, en la que se evidencian cada una de las emociones de izquierda a derecha, en la primera se muestra alegría, en la segunda asombro, tercera tristeza, la cuarta corresponde a enojo, la quinta miedo y la sexta y última disgusto.



Figura 3.1: Imágen que contiene las 6 emociones básicas de algunos de los 12 participantes que implementó la base de datos de LIRIS-CSE. [40]

### 3.1.3. Estimulo de emociones

Se implementaron videos animados o cortos de películas animadas para inducir las emociones a los niños, considerando las razones éticas y la edad de los participantes (6-12 años) se seleccionaron cuidadosamente los videos, de manera que no generara ningún impacto negativo o adverso, por este motivo no se incluyen videos que inducen enojo y se colocaron unos cuantos que inducen tristeza y miedo.

Expresión que se induce	Cantidad	fuelle	Tiempo
Disgusto	2	Youtube	92 segundos
Miedo	3	Youtube y película	290 segundos
Felicidad	5	Youtube	186 segundos
Tristeza	3	Youtube	100 segundos
sorpresa	1	Pelicula	32 segundos
Miedo y sorpresa	2	Youtube y Película	171 segundos
Felicidad y sorpresa	3	Youtube y Película	108 segundos
Felicidad y sorpresa	1	Youtube	40 segundos

Tabla 3.1: Resumen de estímulos implementados para evocar la emoción en los participantes del estudio de la la base de datos de LIRIS-CSE. [40]

### 3.1.4. Métodos y materiales

Para la elaboración de la recopilación de videos, se implementaron dos ambientes diferentes ya previamente mencionados que fue en un salón de clases y en casa, en ambos espacios se implementó una silla para que el menor de edad se encontrara en posición y mirada hacia el escritorio en el que se encontraba la pantalla por la que se colocaban los fragmentos de videos de YouTube o películas, finalmente se implementa una cámara de la cual se resumen sus características técnicas (FPS y resolución) de video en la Tabla 3.2.

Sujeto	Ambiente	FPS	Resolución de video
S1-S7	Salón de clases	25	800*600
S8-S10	Casa	25	720*480
S11-S12	Casa	25	1920*1080

Tabla 3.2: Parámetros de los videos grabados para la base de datos [40].

### 3.1.5. Método de adquisición de videos de la base de datos

Para obtener los videos en los que se recopilan las emociones evocadas en los niños en edades comprendidas entre los 6 y los 12 años, en la Tabla 3.1 se evidencia la cantidad de videos y películas implementadas para evocar cada una de las 6 emociones básicas en los participantes, adicional otras tres emociones combinadas, en la segunda columna se muestra la cantidad de videos que fueron reproducidos por cada emoción, la tercera es la fuente de la que se tomó el

video y la última columna muestra la duración en segundos.

## 3.2 Programación

### 3.2.1. Detección del Rostro

Este es un método que se basa en el aprendizaje automático para la detección del rostro, se utilizó la librería OpenCV, específicamente el filtro HaarCascadeClassifier que permite esta detección mediante clasificadores ubicados en cascada que se entrenan a partir de muchas imágenes positivas y negativas, en este caso las imágenes positivas son imágenes con rostros de personas y en el caso contrario, las imágenes negativas como las que no contienen rostros. Esto es con el fin de entrenar el clasificador, en la Figura 3.2 se muestra cómo actúan las características Haar que se asemejan al kernel convolucional, que se encarga de obtener un valor único al restar la suma de los píxeles debajo del rectángulo blanco de la suma de los píxeles ubicados debajo del rectángulo negro.

Este modelo previamente entrenado, para poder implementarlo en nuestro programa de Python se utiliza el método `cv::CascadeClassifier::load`, para ello primero se crea un clasificador mediante `cv::CascadeClassifier` y allí se carga un archivo XML. Luego, para la detección se implementa el método `cv::CascadeClassifier::detectMultiScale` que permite crear un contorno para el rostro a detectar. En su defecto y para un mejor manejo de las imágenes se realiza un recorte del rostro a la medida del contorno que se obtiene del MultiScale.

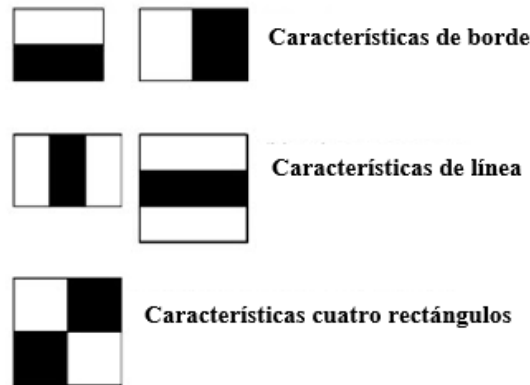


Figura 3.2: Implementación del filtro Haar para la detección de rostros mediante clasificadores en cascada basado en Kernel de Convolución.

### 3.2.2. Segmentación de videos a imágenes

Para la segmentación de los 208 videos que tiene la base de datos se utilizó de la librería OpenCV, el comando `imwrite` para guardar los frame de video. Luego se revisaron las carpetas y se evidenció que los primeros y últimos frames de los videos, contenían imágenes que no

contenían la emoción mencionada, por lo que se decidió eliminar las imágenes que pudieran afectar en el modelamiento predictivo, entre ellas las que se encuentran al comienzo y al final del video, en donde las micro expresiones no se mostraban tan evidentes y más bien neutras que es el momento exacto entre antes y después de expresar la emoción.

### 3.2.3. Entrenamiento red neuronal

Para el entrenamiento se implementa Tensorflow y Keras, que permite crear el modelo predictivo de aprendizaje profundo, a partir de una base de datos de entrada y con los parámetros con las características que se implementan para crear el modelo, como se muestran en la Tabla 3.3, en este caso se ingresan 6 clases de entrada que corresponden a la seis emociones básicas en el siguiente orden:

1. Disgusto
2. Alegría
3. Tristeza
4. Sorpresa
5. Miedo
6. Enojo

Modelo	Peso de la Red	Exactitud	Profundidad
MobileNet	16 MB	70.4 %	55

Tabla 3.3: Características de modelo MobileNet de Keras.

Después de la preparación y exportación del modelo de predicción, se descargan tres archivos que corresponden primero al modelo keras en un archivo ".h5z el modelo para el lenguaje de programación que se está implementando, en este caso Python. En la Tabla 3.4 se muestra a detalle la cantidad de frames por emoción y los sujetos por emoción para la preparación del modelo, estas se adquieren mediante la segmentación de los videos por emociones de todos los sujetos y las imágenes que se descartan por esos pequeños episodios efímeros en donde el sujeto no muestra emoción alguna, es decir que se encuentra neutro ya sea segundos previos a la ejecución de las emociones en los momentos subsecuentes después de emular la emoción.

Emoción	Sujetos	Cantidad de sujetos	Frames
Disgusto	S1 - S5 - S6 - S8 - S11 - S12	6	525
Alegría	S1 - S3 a S9 - S11 - S12	10	2.384
Tristeza	S1 a S12	12	3.163
Asombro	S1 a S12	12	1.566
Miedo	S2 a S16 - S8 a S12	10	1.025
Enojo	S11 y 7 adicionales	8	70

Tabla 3.4: Frames por emoción capturado en el estudio [40].



### 3.2.4. Librería y funciones

Las librerías y métodos implementados para la detección de rostros, segmentación de videos, y entrenamiento e implementación del modelo de predicción son los siguientes:

OpenCV, OS, Imutils, Keras, Pandas.

## 3.3 Resultados

Para la obtención de resultados se ejecutó el programa con un rostro femenino de edad 23 años y se realizaron 4 intentos por emoción en donde se toman 100 frames de video en tiempo real. De allí se obtiene un total de 400 imágenes por emoción con cada una de las etiquetas (la emoción que detecta el algoritmo) se realiza un promedio de la cantidad de veces que se ejecuta la emoción y se acierta para poder construir la matriz de confusión que se mostrará en la sección de resultados del Capítulo 4. Con la matriz de confusión se obtiene la exactitud de todo el algoritmo, la precisión de cada parámetro (de cada emoción) para que el programa detecte cada emoción correctamente y finalmente el F1 Score que hace referencia a la relación que hay entre la precisión y la sensibilidad.

## Capítulo 4

# RESULTADOS

El reto del presente proyecto se evidencia en la base de datos implementada, que realiza la adquisición de los videos en ambientes e iluminación no tan controlados, cumpliendo con el mayor propósito de lograr detectar cada una de las seis emociones básicas con ambientes más realistas, adaptados a la vida cotidiana y que se obtuviesen expresiones faciales lo más natural posibles. Para ello se realizaron por cada una de las 6 emociones, 4 pruebas, cada una de a 100 frames de imágenes. Como resultado se obtuvo 400 predicciones por emoción. Se realizó la media aritmética por emoción y se eligió el intento más parecido al valor de la media para poder elaborar la matriz de confusión como se muestra en la Tabla 4.1 y así evaluar el desempeño del algoritmo desarrollado.

A partir de la matriz de confusión, se obtiene la exactitud de todo el sistema con un valor de 74,17% como se muestra en la Ecuación 4.1.

$$Exactitud = \frac{VP + VN}{VP + FP + FN + VN} = 74,17\% \quad (4.1)$$

Además se agrupan los resultados de la precisión, sensibilidad y el puntaje F1 en la tabla 4.2, para cada una de las clases (emociones).

	Clase 1 Disgusto	Clase 2 Alegría	Clase 3 Tristeza	Clase 4 Asombro	Clase 5 Miedo	Clase 6 Enojo
Clase 1: Disgusto	<b>71</b>	9	0	1	16	0
Clase 2: Alegría	1	<b>96</b>	0	0	3	0
Clase 3: Tristeza	0	2	<b>87</b>	2	9	0
Clase 4: Asombro	0	0	0	<b>87</b>	13	0
Clase 5: Miedo	2	0	17	36	<b>45</b>	0
Clase 6: Enojo	8	12	0	16	8	<b>56</b>

Tabla 4.1: Matriz de confusión del algoritmo implementado para la detección de las 6 emociones básicas.

Clase	Emoción	n (verdadero)	n (clasificado)	Precisión	Sensibilidad	F1
Clase 1	Disgusto	85	100	0,74	0,87	0,8
Clase 2	Alegría	119	100	0,96	0,81	0,88
Clase 3	Tristeza	104	100	0,87	0,84	0,85
Clase 4	Sorpresa	142	100	0,87	0,61	0,72
Clase 5	Miedo	94	100	0,45	0,48	0,46
Clase 6	Enojo	56	100	0,56	1	0,72

Tabla 4.2: Resultados de precisión, sensibilidad y F1 de Matriz de Confusión para cada emoción.

Para recopilar la información contenida en la tabla 4.2 se utiliza la figura 4.1 que muestra la relación que hay entre la precisión, la sensibilidad y el puntaje F1 en cada una de las emociones, que se calculan con las Ecuaciones 4.2, 4.3 y 4.4. El último parámetro (Puntaje F1), útil para el caso de las emociones de disgusto, miedo y enojo que tienen una distribución desigual, en el caso de la clase de disgusto, tiene una pequeña distribución con las emociones de alegría y miedo, para la clase miedo se evidencia una distribución significativa con las emociones de tristeza y asombro, y para la emoción de enojo, una distribución significativa con las emociones de disgusto, alegría, asombro y miedo.

$$Precisión = \frac{VP}{VP + FP} \quad (4.2)$$

$$Sensibilidad = \frac{VP}{VP + FN} \quad (4.3)$$

$$F1Score = \frac{2 \times Precisión \times Sensibilidad}{Precisión + Sensibilidad} \quad (4.4)$$

Por último, para las pruebas de rendimiento que realizan al algoritmo para la detección de las seis emociones básicas en rostros de niños, se realiza una gráfica de precisión en comparación con el tamaño de la muestra (n) en la figura 4.2 y obteniendo un porcentaje de correlación del 70 %.

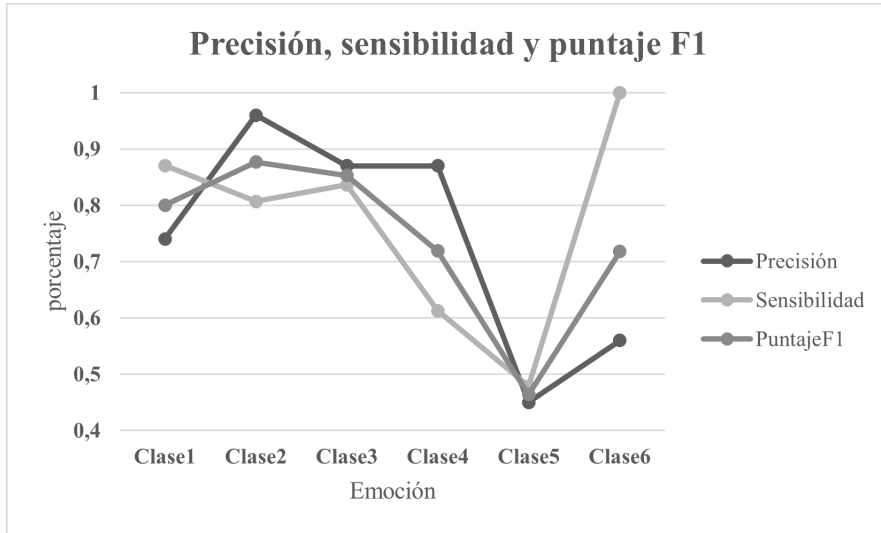


Figura 4.1: Gráfica de precisión, sensibilidad y puntaje F1

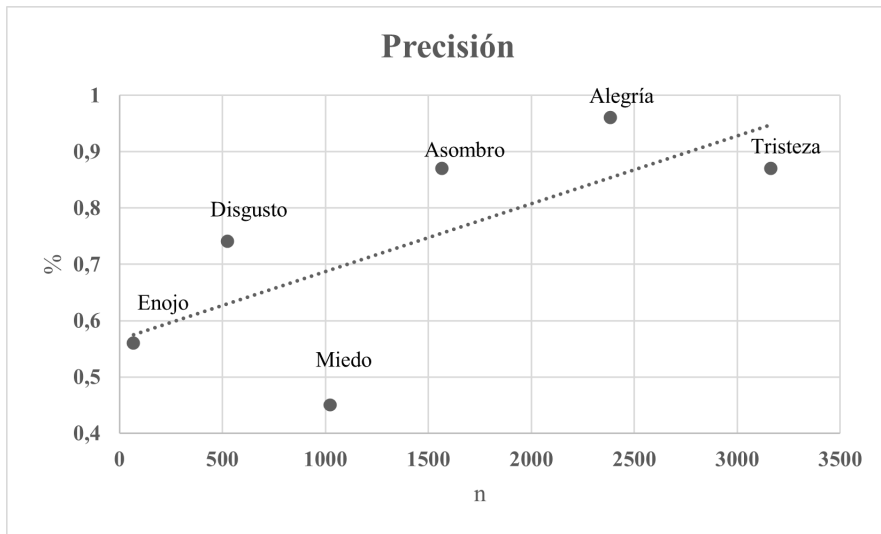


Figura 4.2: Gráfica de precisión, sensibilidad y tamaño de cada una de las muestras.

## Capítulo 5

# DISCUSIONES

Se declara todo un reto el hecho de trabajar con la subjetividad que implican las emociones en las expresiones faciales, que, a pesar de considerarse expresiones universales, pueden variar dependiendo de la cultura y a pesar de que en un individuo en su primer año se logran identificar las seis emociones básicas, las que contienen una connotación negativa tienden a variar con el tiempo.

La exactitud que se obtiene del algoritmo implementado es del 74,17 %, lo cual indica que se tuvo una buena aproximación y muy cercana a la realizada en el estudio de la base de datos. Las emociones con una mayor precisión fueron Alegría, seguida de Asombro y Tristeza.



Figura 5.1: Imágenes base de datos para emoción de alegría en diferentes sujetos, participantes del estudio.

Empezando con la emoción de alegría se obtuvo una precisión y sensibilidad superiores al 80 %, ofreciendo una distribución mínima, confiabilidad alta y la estabilidad de video en la detección de la emoción del programa mientras se realizaban movimientos (arriba, abajo y hacia los lados), se mantienen estables en el tiempo. Esta alta confiabilidad en el modelo, específicamente para la emoción de alegría, se da porque esta es una expresión que contiene micro expresiones marcadas que contienen la elevación de las comisuras de los labios y mejillas, en conjunto con la apertura del contorno visual. Además es importante notar en la Figura 4.2 que esta emoción tiene una base de datos con una muestra de más de 2.000 imágenes con expresiones homogéneas, en donde la variación de las micro expresiones es mínima como se

muestra en la Figura 5.1, como lo puede ser con boca cerrada o abierta, parpados arriba o hacia abajo, entre otros.

La emoción de tristeza por otro lado obtuvo una precisión del 87% y una mejora en el porcentaje de sensibilidad con un 84% en comparación con la emoción de alegría, esto se evidencia en la matriz de confusión de la Tabla 4.1 ya que la porción de falsos positivos de esta clase fue de 17% del 100% de la clase miedo, caso que se discutirá más adelante. La tristeza puede llegar a ser una consecuencia a alguna emoción intensa fuerte como lo puede ser miedo, ya que la edad de la población de estudio maneja un nivel de interpretación diferente ante estímulos fuertes y poco comunes en su cotidianidad, debido al fuerte impacto que puede generar en su salud mental, así como el desarrollo del lenguaje y el habla en los subsecuentes años en su vida y su entorno.

La clase de asombro obtuvo el mismo porcentaje de precisión que la emoción de tristeza, pero presenta una disminución al 61% en la sensibilidad por lo que presenta una baja tasa para encontrar casos positivos que fueran correctamente identificados por el algoritmo implementado, esto se debe a la dispersión existente en las clases de Disgusto, Tristeza, Enojo y Miedo. La principal explicación que se puede obtener está relacionada con las micro expresiones que se pueden ejecutar en las emociones de Miedo, Enojo y tristeza.

La emoción que presentó una precisión menor al 50% fue miedo y en la matriz de confusión de la Tabla 4.1 se evidencia los porcentajes que recae sobre la detección de otras emociones como disgusto con un 2%, tristeza 17% y Asombro 36%, se puede atribuir a dos razones principales:



Figura 5.2: Imágenes base de datos para emoción de miedo de los sujetos S2,S3 y S4

1. La base de datos contenida para el entrenamiento del programa y la creación del modelo, fue de 1.025 imágenes y contenía 10 de los 12 participantes del estudio, al contener los videos de los cambios en la expresión gestual y facial, se logra identificar la combinación de las emociones de tristeza, asombro y miedo, así como la expresión de miedo en algunos sujetos es sostenida y en otros casos se desvanece rápidamente, esto se puede evidenciar en la Figura 5.2.

Del sujeto S8 se captaron 8 videos para la emoción de miedo y como se puede ver en la figura 5.3, las diferentes transiciones en los movimientos gestuales del rostro, en los que evidentemente se muestran las emociones combinadas entre disgusto, asombro, tristeza y el mismo miedo.



Figura 5.3: Imágenes base de datos para emoción miedo sujeto S8

2. En el trabajo "Paul Ekman y las micro expresiones faciales de las emociones" de Henry Wladimir [18], se obtienen cada una de las micro expresiones de cada una de las emociones y la emoción de miedo evidencia 7 unidades de acción, como se muestra en la Figura 5.4., se pueden observar cambios significativos en la expresión en el contorno de los ojos y en comparación con las otras 5 emociones básicas, que contienen mínimo 2 y máximo 4 micro expresiones.

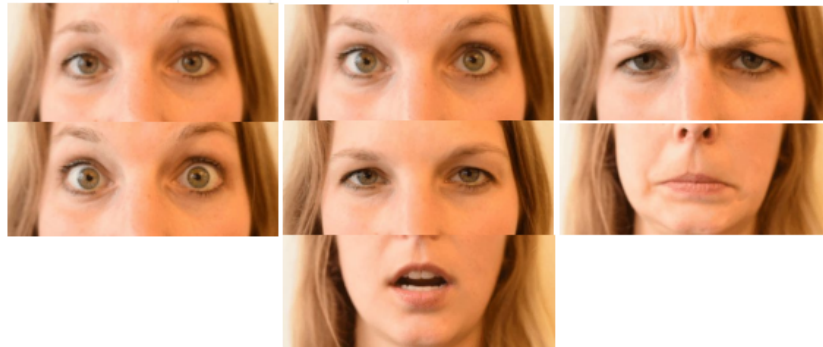


Figura 5.4: Micro expresiones descritas por el estudio de Paul Ekman para emoción de miedo y recreadas en el estudio tomado de [18].

Disgusto presenta una precisión del 74% pero aumenta su sensibilidad, teniendo el valor más alto con un 87% por la mínima dispersión existente en la clase de enojo, este hecho se discute con la justificación de que éticamente no se implementaron videos para generar enojo en los participantes por la edad que los comprendía obteniendo solo un total de 70 imágenes para esta emoción, esto se debe a dos razones principales, no se logró el balanceo del tamaño de la muestra de esta clase ya que no se logró obtener otra base de datos con este tipo de imágenes con niños enojados en donde se tuviese una expresión auténtica, es por este motivo que enojo maneja la peor tasa de precisión y la justificable elevada tasa de sensibilidad, ya que

al no tener las suficientes imágenes en comparación con las otras emociones, no se generaron falsos positivos en las pruebas de las otras clases (emociones) precisando la emoción de enojo.



## Capítulo 6

# CONCLUSIONES

A pesar del arduo esfuerzo, se logra realizar un algoritmo que al ejecutar detecta en video las 6 emociones básicas a partir del entrenamiento de un modelo de IA por medio de una base de datos que contiene las expresiones faciales auténticas de los doce participantes que son niños en edades comprendidas de los 6 a 12 años y diversos entre sí, del cual se obtiene una exactitud del 74,17%. La precisión y sensibilidad más baja se da en la emoción de enojo debido a la ética para generar este tipo de emoción en niños con las edades comprendidas del estudio. Por otro lado, el algoritmo obtiene muy buenos resultados para las emociones de alegría, asombro y tristeza, por lo que a mitad de camino, para las tres emociones restantes, se puede construir una base de datos más sólida conservando la autenticidad y naturalidad en la obtención de estos videos en trabajos futuros.

La construcción del estado del arte permite justificar los resultados obtenidos y del mismo modo, la propuesta de nuevos trabajos y complementos para la mejora del sistema, esto se justifica ya que a pesar de la universalidad de los gestos que se realizan en la comunicación no verbal, específicamente en las expresiones faciales, dentro de esta misma categoría, existen emociones (generalmente las que tienen una connotación negativa) que contienen diferentes micro expresiones, que al ser combinadas entre ellas, se puede expresar la misma emoción, caso que aplica para las emociones de miedo, disgusto y enojo, como se pudo evidenciar en los resultados y su respectiva discusión.

Se espera lograr en un futuro probar el algoritmo con pacientes patológicos, por el momento con sujetos que presenten dificultades en el habla y que no se les dificulte en primera instancia el entendimiento, mientras se aplican mejoras en el algoritmo para detectar las emociones con una precisión y sensibilidad baja.

## Capítulo 7

# RECOMENDACIONES Y TRABAJOS FUTUROS

En primer lugar, se puede mejorar la exactitud del programa y por consiguiente la precisión en la emoción de miedo, mediante el estudio de la variación en la activación de la actividad de la amígdala, que resulta tener un incremento ante expresiones faciales de miedo.

La posición corporal, por otro lado, que también hace parte de la comunicación no verbal, puede reflejar movimientos de acercamiento o alejamiento ante cierto estímulo y que puede generar alguna respuesta positiva o negativa respectivamente. Esto se puede estudiar mediante la distancia que se pueda medir entre la cámara y el sujeto objetivo.

Por otro lado, existen estudios que profundizan en cuanto a la medición de la dilatación de la pupila ante estímulos emocionales, por lo que para este caso se requiere un dispositivo de medición más preciso para lograr complementar el trabajo presente con esta posibilidad.

# Bibliografía

- [1] Y. Harari, *De animales a dioses: breve historia de la humanidad*, 1ra ed. Penguin Random House, 2014.
- [2] D. Matsumoto, R. M. López, M. Ángel, and P. Nieto, “Lectura de la Expresión Facial de las Emociones: Investigación básica en la mejora del reconocimiento de emociones,” 2013. [Online]. Available: <https://www.researchgate.net/publication/258119111>
- [3] D. Matsumoto, D. Keltner, M. N. Shiota, M. O’Sullivan, and M. Frank, “Facial expressions of emotion.,” in *Handbook of emotions*, 3rd ed., New York, NY, US: The Guilford Press, 2008, pp. 211–234.
- [4] P. M. Cole, P. A. Jenkins, and C. T. Shott, “Spontaneous Expressive Control in Blind and Sighted Children,” *Child Dev*, vol. 60, no. 3, p. 683, Jun. 1989, doi: 10.2307/1130733.
- [5] D. Galati, R. Miceli, and B. Sini, “Judging and coding facial expression of emotions in congenitally blind children,” *Int J Behav Dev*, vol. 25, no. 3, pp. 268–278, May 2001, doi: 10.1080/01650250042000393.
- [6] D. Galati, B. Sini, S. Schmidt, and C. Tinti, “Spontaneous Facial Expressions in Congenitally Blind and Sighted Children Aged 8–11,” *J Vis Impair Blind*, vol. 97, no. 7, pp. 418–428, Jul. 2003, doi: 10.1177/0145482X0309700704.
- [7] National Institute on Deafness and Other Communication Disorders, “Etapas del desarrollo del habla y del lenguaje,” National Institute on Deafness and Other Communication Disorders. <https://www.nidcd.nih.gov/es/espanol/etapas-del-desarrollo-del-habla-y-el-lenguaje> (accessed Nov. 25, 2022).
- [8] Center for Parent Information and Resources, “Trastornos del Habla y Lenguaje,” Center for Parent Information and Resources, May 2014. <https://www.parentcenterhub.org/lenguaje/> (accessed Nov. 25, 2022).
- [9] National Institute on Deafness and Other Communication Disorders, “Quick Statistics About Voice, Speech, Language,” National Institute on Deafness and Other Communication Disorders. <https://www.nidcd.nih.gov/health/statistics/quick-statistics-voice-speech-language> (accessed Nov. 25, 2022).

- [10] Agencia UNAL, “30 % de preescolares tienen algún retraso en el lenguaje,” Universidad Nacional de Colombia, Mar. 2014. <http://agenciadenoticias.unal.edu.co/detalle/30-de-preescolares-tienen-algun-retraso-en-el-lenguaje> (accessed Nov. 25, 2022).
- [11] N. Laza Gutiérrez, R. de La, and R. Ramírez, “Desarrollo del lenguaje.” [12] E. Corrales, N. el Lenguaje, and E. C. Navarro, “El lenguaje no verbal: un proceso cognitivo superior indispensable para el ser humano,” *Revista Comunicación*, vol. 20, no. 1, 2011.
- [13] P. Ekman, *¿Qué dice ese gesto?* Barcelona RBA, 2004.
- [14] C. Darwin, *La expresión de las emociones en el hombre y en los animales*. John Murray, 1872.
- [15] A. V. Fernández and Y. H. R. Garzón, “Neuropsicología de las emociones: el aporte de Charles Darwin,” *Cuadernos de neuropsicología*, vol. 3, pp. 225–233, 2009.
- [16] P. Ekman, *Darwin and facial expression: A century of research in review*. Oxford, England: Academic Press, 1973.
- [17] Q. Abigail, P. Montes, S. L. Salas, H. Rafael, O. Aguirre, and M. Q. López, “Detección automatizada de emociones a través del rostro, una revisión del estado del arte,” *Coloquio de Investigación Multidisciplinaria 2020 Journal CIM*, vol. 8, no. 1, 2020.
- [18] H. Mallitasig, “Paul Ekman y las microexpresiones faciales de las emociones.”
- [19] H. Pérez, *Lenguajes Verbales Y No Verbales: Reflexiones Pedagógicas Sobre Las Competencias Lingüística Y Comunicativa*, 2da ed. Cooperativa editorial magisterio, 2007.
- [20] P. Castañeda, *El Lenguaje verbal del niño: ¿cómo estimular, corregir y ayudar para que aprenda a hablar bien?* Lima: Universidad Nacional Mayor de San Marcos, 1999.
- [21] A. Maestre, “DESARROLLO DEL LENGUAJE VERBAL,” *Innovación y experiencias educativas*, 2010.
- [22] L. Tobón de Castro, *La lingüística del lenguaje: estudios entorno a los procesos de significar y comunicar*, 2nd ed. Universidad Pedagógica Nacional, 2007.
- [23] M. A. K. Halliday, “Learning to mean—Explorations in the development of language,” in *Foundations of Language Development*, E. H. Lenneberg and E. Lenneberg, Eds. Academic Press, 1975, pp. 239–265. doi: <https://doi.org/10.1016/B978-0-12-443701-2.50025-1>.
- [24] J. Narbona and C. Chevie-Muller, *El lenguaje del niño - desarrollo normal, evaluación y trastornos*, 2nd ed. Masson, 2001.

- [25] S. Aguilera Albesa and O. B. Crespo, “Trastornos del lenguaje,” *Pediatría Integral*, vol. 16, no. 9, 2012.
- [26] R. Castro-Rebolledo, M. Giraldo-Prieto, L. Hincapié-Henao, F. Lopera, and D. A. Pineda, “Trastorno específico del desarrollo del lenguaje: una aproximación teórica a su diagnóstico, etiología y manifestaciones clínicas,” *Rev Neurol*, pp. 111504–12974, 2004.
- [27] A. Arboleda-Ramírez et al., “Trastorno específico del desarrollo del lenguaje: problema selectivo o generalizado de la cognición,” *Rev Neurol*, 2007.
- [28] V. M. Acosta Rodríguez, G. M. Ramírez Santana, A. D. Cruz, and Y. N. del Valle Hernández, “Trastorno Específico del Lenguaje y Trastorno del Espectro Autista: similitudes y diferencias en el discurso narrativo,” *Revista de Investigación en Logopedia*, vol. 12, no. 1, Jan. 2022, doi: 10.5209/rlog.76069.
- [29] B. Jaramillo-Valencia, L. Castaño-Barreneche, and K. Ossa-Álvarez, “El habla y su desarrollo a través de la didáctica,” *Revista Docencia Universitaria*, vol. 23, no. 1, Jun. 2022, doi: 10.18273/revdu.v23n1-2022001.
- [30] F. Peinado, “Disglosia labial y palatal en Educación Primaria,” *Revista Internacional de Apoyo a la Inclusión, Logopedia, Sociedad y Multiculturalidad.*, vol. 3, no. 1, 2016.
- [31] M. L. Cera, T. P. P. Romeiro, P. P. Mandrá, and M. T. H. Fukuda, “Variables associated with speech and language therapy time for aphasia, apraxia of speech and dysarthria,” *Dement Neuropsychol*, vol. 13, no. 1, pp. 72–77, Mar. 2019, doi: 10.1590/1980-57642018dn13-010007.
- [32] Á. Loeches, F. Carvajal, J. M. Serrano, and S. Carriba, “Neuropsicología de la percepción y la expresión facial de emociones: Estudios con niños y primates no humanos,” *Anales de Psicología*, vol. 20, no. 2, Dec. 2004.
- [33] I. L. Mieles Toloza, F. O. Rojas Delgado, O. E. Torrado Duarte, L. J. Plata Osma, and E. L. Prada Sarmiento, “Reconocimiento e intensidad emocional de la expresión facial – Presentación de una versión corta de la prueba ‘Picture of Facial Affect’ (POFA).,” *Revista de Psicología Universidad de Antioquia*, vol. 12, no. 1, pp. 89–110, Dec. 2020, doi: 10.17533/udea.rp.v12n1a05.
- [34] E. Yaire and A. Marte, “Neuropsicología de la Percepción y la Expresión Facial de Emociones.”
- [35] A. M. O. de Lima, M. R. dos A. Medeiros, P. D. P. Costa, and C. A. S. Azoni, “Analysis of softwares for emotion recognition in children and teenagers with autism spectrum disorder,” *Revista CEFAC*, vol. 21, no. 1, 2019, doi: 10.1590/1982-02162019/21112318.

- [36] A. Piedrahíta-Carvajal, P. A. Rodríguez-Marín, D. F. Terraza-Arciniegas, M. Amaya-Gómez, L. Duque-Muñoz, and J. D. Martínez-Vargas, “Aplicación web para el análisis de emociones y atención de estudiantes,” *TecnoLógicas*, vol. 24, no. 51, p. e1821, May 2021, doi: 10.22430/22565337.1821.
- [37] M. Fernández Ríos, R. Redolat, E. Serra, and G. González-Alcaide, “Una revisión sistemática acerca del reconocimiento facial de las emociones en la Enfermedad de Alzheimer: una perspectiva evolutiva y de género,” *Canales de Psicología*, vol. 37, no. 3, pp. 478–492, Aug. 2021, doi: 10.6018/analesps.439141.
- [38] A. P. Canazas, J. J. Ramos Blaz, P. D. Torres Martínez, and X. J. Mamani, “Sistema de identificación de emociones a través de reconocimiento facial utilizando inteligencia artificial Emotion identification system through facial recognition using artificial intelligence,” *Revista Innovación y Software*, vol. 3, no. 2, 2022.
- [39] L. Aballay, S. Aciar, and E. Reategui, “Propuesta de un Método para Detección de Emociones en E-Learning,” *ASAI 2015, 16o Simposio Argentino de Inteligencia Artificial*.
- [40] R. A. Khan, C. Arthur, A. Meyer, and S. Bouakaz, “A novel database of Children’s Spontaneous Facial Expressions (LIRIS-CSE),” Dec. 2018, doi: 10.1016/j.imavis.2019.02.004.