

# **Evaluación de la experiencia de usuario en teléfonos inteligentes soportada en la medición no invasiva de emociones**



Monografía

**Carlos Andrés Burbano Palechor  
Julián David Sotelo Vargas**

Director:

PhD (c) Andrés Felipe Aguirre Aguirre

Codirector:

PhD. César A. Collazos O.

*Universidad del Cauca*

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Departamento de Sistemas**

**Grupo IDIS – Investigación y Desarrollo en Ingeniería de Software**

**Línea de investigación Interacción Humano Computador**

Popayán, Septiembre de 2021

## Tabla de Contenido

<b>CAPÍTULO 1. INTRODUCCIÓN</b> .....	1
1.1. Planteamiento del Problema.....	1
1.2. Motivación.....	2
1.3. Objetivos .....	2
1.3.1. Objetivo General .....	3
1.3.2. Objetivos Específicos.....	3
1.4. Diseño metodológico.....	3
1.5. Organización del documento.....	4
<b>CAPÍTULO 2. MARCO CONCEPTUAL</b> .....	5
2.1. Métodos de medición no invasiva .....	5
2.1.1. Generalidades de las emociones .....	5
2.1.2. Reconocimiento facial.....	6
2.1.2.1. Enfoques para la representación de expresiones faciales. ....	7
2.1.2.1.1. Enfoque de características geométricas.....	7
2.1.2.1.2. Enfoque basado en la apariencia. ....	7
2.1.3. Reconocimiento del habla .....	8
2.1.3.1. Enfoques para el reconocimiento de emociones mediante la voz .....	8
2.1.3.1.1. Enfoque prosódico.....	9
2.1.3.1.2. Enfoque espectral. ....	9
2.1.3.1.3. Características generales de la voz. ....	9
2.2. Caracterización de la UX en Teléfonos Inteligentes .....	10
2.2.1. Definición del protocolo de caracterización.....	10
2.2.2. Desarrollo de la caracterización.....	11
2.2.2.1. Búsqueda de estudios .....	11
2.2.2.1.1. Formulación de la pregunta de búsqueda .....	11
2.2.2.1.2. Selección de fuentes.....	11
2.2.2.1.3. Elección de estudios.....	12
2.2.2.2. Selección de estudios.....	13
2.2.2.2.1. Criterios de selección .....	13
2.2.2.2.2. Análisis de los estudios .....	14
2.2.2.2.3. Resultado de la selección .....	15
2.2.2.3. Elaboración de la caracterización .....	16

2.2.3. Conclusiones del proceso de caracterización .....	17
2.3. Contextualización de Sistemas Multimodal.....	18
2.4. Inteligencia artificial.....	19
2.5. Machine learning.....	20
2.6. Deep learning.....	21
2.7. Redes neuronales. ....	22
2.8. Aportes de Investigación .....	23
<b>CAPITULO 3. INTEGRACIÓN DE LOS MÉTODOS NO INVASIVOS PARA LA EVALUACIÓN DE LA UX CON TELÉFONOS INTELIGENTES.....</b>	<b>24</b>
3.1. Herramientas para la captura de datos en teléfonos inteligentes. ....	24
3.2. Elección de herramientas para el reconocimiento de emociones.....	25
3.2.1. Herramientas para el reconocimiento de expresiones faciales. ....	25
3.2.2. Herramientas para el reconocimiento de emociones a partir del habla. ....	28
3.3. Arquitectura de prototipo software.....	30
<b>CAPÍTULO 4. DESARROLLO DEL PROTOTIPO SOFTWARE PARA LA EVALUACIÓN DE LA UX EN TELÉFONOS INTELIGENTES.....</b>	<b>32</b>
4.1. Metodología de desarrollo .....	32
4.2. Desarrollo del prototipo software .....	33
4.2.1. Reconocimiento de expresiones faciales .....	34
4.2.1.1. Procesamiento de datos para algoritmo de expresión facial .....	35
4.2.2. Reconocimiento del habla .....	35
4.2.2.1. Procesamiento de datos para algoritmo del habla. ....	36
4.3. Entrada de la información.....	37
4.4. Presentación de la información .....	38
4.5. Preparación de la información de entrenamiento.....	38
4.5.1. Bases de datos.....	38
4.5.2. Selección base de datos. ....	41
4.6. Entrenamiento del prototipo.....	41
<b>CAPÍTULO 5: ESTUDIO DE CASO: EVALUACIÓN DE LA UX CON TELÉFONOS INTELIGENTES UTILIZANDO EL PROTOTIPO SOFTWARE PROPUESTO.....</b>	<b>43</b>
5.1. Metodología .....	43
5.1.1. Diseño del estudio de caso.....	44
5.1.2. Recolección de datos .....	45
5.1.3. Análisis de los datos .....	45

5.2. Ejecución del estudio de caso .....	45
5.3. Análisis de resultados .....	50
5.4. Síntesis y discusión.....	52
5.5. Opinión de expertos .....	52
<b>CAPÍTULO 6. CONCLUSIONES Y TRABAJO FUTURO. ....</b>	<b>55</b>
6.1. Conclusiones.....	55
6.2. Limitaciones .....	56
6.3. Lecciones aprendidas y problemas enfrentados.....	56
6.4. Trabajos futuros.....	57
<b>BIBLIOGRAFÍA.....</b>	<b>58</b>

## LISTA DE FIGURAS

Figura 1. Cadena de búsqueda básica.....	12
Figura 2. Ejemplo diagrama de sistema multimodal.....	19
Figura 3. Vídeo de los cuatro algoritmos corriendo en simultaneo.....	27
Figura 4. Algoritmos procesados en FPS.....	28
Figura 5. Extracción de roles, eventos y artefactos utilizados.....	33
Figura 6. Capas del Prototipo.....	34
Figura 7. Conexión capas red neuronal. ....	35
Figura 8. Pantalla inicio aplicación. ....	37
Figura 9. Grafica de emociones generada por el Prototipo. ....	38
Figura 10. Grafica de resultados de las validaciones con prototipo propuesto. ....	50
Figura 11. Grafica de resultados de las validaciones con el método tradicional. ..	51

## LISTA DE TABLAS

Tabla 1 Criterios cumplidos por cada estudio.....	14
Tabla 2. Listado de características de UX en teléfonos inteligentes resultante de la revisión de la literatura .....	16
Tabla 3. Listado de características con número de estudios donde se considera. 17	
Tabla 4. Comparación de herramientas que permiten la ejecución de otras aplicaciones.....	24
Tabla 5. Comparación de herramientas de grabación en segundo plano. ....	25
Tabla 6. Cuadro comparativo de las herramientas para el reconocimiento de expresiones faciales.....	27
Tabla 7. Comparación herramientas reconocimiento de la voz.....	30
Tabla 8. Bases de datos de expresiones faciales existentes. ....	39
Tabla 9. Relación entre los juegos y la cantidad de niños que lo prefieren.....	46
Tabla 10. Resultado de la validación con prototipo propuesto. ....	50
Tabla 11. Resultado de la validación con método tradicional.....	51
Tabla 12. Perfiles de los expertos. ....	53

# CAPÍTULO 1. INTRODUCCIÓN

## 1.1. Planteamiento del Problema

Debido a su carácter multidisciplinario, la experiencia de usuario (en adelante, UX por sus siglas en inglés *User eXperience*) puede ser abordada desde diferentes perspectivas, lo cual dificulta definirla desde un solo concepto. No obstante, a fin de acercar al lector a su comprensión, se parte de aquella definida por M. Hassenzahl y N. Tractinsky como *“Una consecuencia de estado interno de un usuario (predisposiciones, expectativas, necesidades, motivación, estado de ánimo, etc.), las características del sistema diseñado (por ejemplo, la complejidad, el propósito, la facilidad de uso, funcionalidad, etc.) y el contexto (o el medio ambiente) en el que la interacción se produce”* [1]. En los últimos tiempos la UX ha tomado gran importancia durante el desarrollo y utilización de los sistemas informáticos, dado que ayuda a tener una comprensión profunda del usuario y de su interacción con un sistema, servicio o producto. Por ello, el desarrollo y aplicación de buenas prácticas de evaluación de la UX, permite promover un mejoramiento de la percepción sobre el producto o servicio que se ofrece [2], así como de la propia interacción del usuario.

De otro lado, teniendo en cuenta que la relación entre usuario y producto o servicio evoluciona con el tiempo, desarrollar métodos para medir la UX a largo plazo posibilita hacer inferencias sobre el incremento o no de la lealtad del usuario. Con esto, se entiende que en procura de una UX positiva, como objetivo en el diseño de productos y servicios, su evaluación permanente y el análisis a largo plazo son también fundamentales para su éxito [3].

Entre los productos, se encuentran las aplicaciones móviles, las cuales son utilizadas por los usuarios a través de teléfonos inteligentes, por lo tanto, es necesario que estos dispositivos se tomen en cuenta para realizar una adecuada evaluación de la UX. Actualmente hay una abundante presencia de teléfonos inteligentes que se han instaurado en las prácticas sociales y que han influido en la transformación material y cultural de la vida cotidiana, como sucede en Colombia, donde según el boletín sobre indicadores básicos del Ministerio de Tecnologías de la Información y las Comunicaciones (MINTIC) del Departamento Administrativo Nacional de Estadística, afirma que *“al término del primer trimestre de 2019, el total de líneas de telefonía móvil alcanzo los 65 millones, 2,2 millones de líneas más de las registradas en el mismo trimestre del año inmediatamente anterior, cuando alcanzó una cifra de 62,8 millones. Por su parte, el número de líneas de telefonía móvil por cada 100 habitantes a nivel nacional se situó en 129 al término del primer trimestre de 2019”* [4], lo cual significa que el uso de la telefonía móvil ha ido en aumento, logrando evidenciar el uso progresivo de teléfonos inteligentes, esto supone como reto que se debe responder de manera asertiva a las distintas necesidades de los usuarios a través de la evaluación de la UX.

En este sentido, al evaluar la UX, un aspecto fundamental es el uso de métodos no invasivos, que se identifican por favorecer la comodidad, mitigar la angustia y la presencia de emociones negativas en los usuarios [5], caso contrario a las sensaciones que producen los métodos invasivos como pueden ser indisposición, angustia y sentimientos no reales. Estos métodos están caracterizados por utilizar medidas como: la respuesta galvánica de la piel (GSR), la actividad electrodérmica (EDA), el pulso de volumen de sangre (BVP), la respiración, la frecuencia cardíaca (FC), la variabilidad del ritmo cardíaco (HRV), la resonancia magnética (fMRI) y el electroencefalograma EEG, que tienden a ser medidas obtenidas por medio de dispositivos que pueden generar indisposición en las personas [6]. Una de las consecuencias que produce la utilización de estos métodos es la alteración de los resultados sobre lo que el usuario siente realmente, ya que su forma de captura podría obtener un estado emocional ambiguo [7]. Cabe resaltar que los métodos invasivos y no invasivos generalmente utilizan una sola medida para realizar el reconocimiento de emociones, los cuales se conoce como métodos de medición unimodal. Por el contrario existen métodos que incorporan varias medidas, conocidos como métodos multimodal, los cuales se caracterizan por modelar las emociones de una forma más precisa [8].

Por lo anterior, y teniendo en cuenta que tanto el aspecto no invasivo como la evaluación de la UX en teléfonos inteligentes han sido poco considerados en los estudios literarios, el aporte de la investigación se enfoca en evaluar la UX, mediante un prototipo software apoyado en la integración de métodos no invasivos de medición multimodal, el cual tomará como datos de entrada los rasgos faciales presentes en los usuarios y su variación en el tono de la voz, obtenidos de forma no invasiva al interactuar con la aplicación móvil que se desea evaluar. Como resultado del procesamiento de los rasgos faciales y el tono de voz se obtendrá la evaluación de la UX, la cual considerará las emociones presentes durante la interacción como alegría, tristeza, asombro etc. Los principales aportes, respecto al conocimiento, corresponden a la identificación de las características que describen la experiencia emocional con teléfonos inteligentes y la descripción de los elementos que conforman los métodos no invasivos para la evaluación de la UX. Con esto, se busca dar respuesta a la siguiente pregunta de investigación: ¿Cómo evaluar el factor emocional de la UX en teléfonos inteligentes mediante métodos no invasivos?

## **1.2. Motivación**

En la actualidad, debido al incremento de dispositivos inteligentes en la sociedad, cada vez son más las organizaciones que ofrecen productos o servicios interactivos a través de aplicaciones móviles basados en las necesidades del cliente, en este sentido, es necesario incrementar la calidad, simplicidad de uso, lenguaje y diseño cercano al usuario, lo cual se logra con la evaluación de la UX.

## **1.3. Objetivos**

### **1.3.1. Objetivo General**

Proponer un prototipo software para evaluar la experiencia de usuario en teléfonos inteligentes, el cual utilice la medición no invasiva de emociones.

### **1.3.2. Objetivos Específicos**

- Identificar, a partir de la literatura, las características que describen la experiencia emocional con teléfonos inteligentes.
- Integrar un conjunto de métodos de medición no invasiva de emociones a partir de las características identificadas, para obtener un marco de referencia para el desarrollo del prototipo.
- Desarrollar un prototipo software que utilice los métodos anteriormente integrados, el cual permita la evaluación de la experiencia de usuario en teléfonos inteligentes.
- Realizar un estudio de caso mediante la evaluación de UX con teléfonos inteligentes en un entorno natural<sup>1</sup>, haciendo uso del prototipo software propuesto.

### **1.4. Diseño metodológico.**

La recopilación de información para el presente trabajo de grado se basó en la investigación de documentos investigativos, la cual se caracterizó por la recopilación de artículos, libros, tesis, normas ISO, conferencias, entre otros, que permitieron conocer con mayor profundidad los temas relacionados con el presente proyecto.

Primeramente, es necesario abordar toda la información relevante sobre el objeto de estudio que para este proyecto es la evaluación de UX en los teléfonos inteligentes. Para tal aspecto se debe entender claramente que factores se deben tener en cuenta para poder realizar una evaluación de UX en este tipo de dispositivos, por lo cual se hace una revisión de la literatura acerca del área a tratar y la información recopilada brinda diferentes características que nos permiten conocer sobre el tema además de poder definir un prototipo acorde a lo establecido en los objetivos.

Luego de tener definida la información sobre evaluación de UX en teléfonos inteligentes donde se rescata las características encontradas, se hace necesario buscar información relacionada con la parte no invasiva, a partir de la documentación revisada, se rescatan herramientas que nos permiten obtener información de las interacciones de un usuario con un teléfono inteligente generando la menor incomodidad posible.

---

<sup>1</sup> Entorno natural: Condiciones o espacio normal para un usuario, donde no se vea sujeto a dispositivos que puedan producir alguna clase de interferencia o incomodidad.

Otra parte del problema es el reconocimiento de emociones, por lo tanto, se realiza una investigación acerca de instrumentos, métodos o técnicas de reconocimiento. Luego de analizar la información se determina que técnica es la más adecuada para realizar el reconocimiento y que además se adapte a la información capturada de manera no invasiva.

Por último, se define el instrumento que será utilizado para realizar una evaluación no invasiva de emociones de la interacción de un usuario con un teléfono inteligente, para construirlo se tendrá en cuenta las características de la experiencia emocional en teléfonos inteligentes, las herramientas de captura no invasivo y las técnicas de reconocimiento, todo esto para ser posteriormente analizado en un estudio de caso.

### **1.5. Organización del documento**

El presente documento se encuentra organizado por capítulos de la siguiente manera:

**Capítulo 1:** presenta una introducción al trabajo, se contextualiza y justifica el problema, se define los objetivos y la metodología a seguir.

**Capítulo 2:** en este capítulo se presenta el marco teórico describiendo algunas bases conceptuales relacionadas con: métodos no invasivos, UX, sistema multimodal, teléfonos inteligentes, seguido de los aportes de investigación.

**Capítulo 3:** se presenta la integración de métodos no invasivos seleccionados.

**Capítulo 4:** se presenta el desarrollo del prototipo software para realizar la evaluación de la UX en teléfonos inteligentes, mediante los métodos integrados.

**Capítulo 5:** se presenta los resultados obtenidos a partir de la aplicación del estudio de caso, el cual se hará en un ambiente natural para el usuario utilizando el prototipo software propuesto en la presente investigación.

**Capítulo 6:** presenta las conclusiones, limitaciones, lecciones aprendidas y trabajos futuros.

## **CAPÍTULO 2. MARCO CONCEPTUAL**

Este capítulo contiene una base conceptual acerca de los métodos de medición no invasiva, además se brinda una contextualización de la UX en teléfonos inteligentes y de los sistemas multimodal que permiten incorporar diferentes técnicas sensoriales en el proceso de reconocimiento de emociones, por último, se presenta los aportes de investigación.

### **2.1. Métodos de medición no invasiva**

En esta sección se describe como primer ítem los aspectos generales de las emociones donde se destaca los enfoques dimensional y discreto. Como segundo y tercer ítem se presentan investigaciones que involucran el reconocimiento facial, mediante el cual se logra capturar las expresiones del rostro, y el reconocimiento del habla, que utiliza el tono de la voz, los cuales son insumos para realizar la captura o evaluación de las emociones. Estos enfoques serán la base para la presente investigación, ya que, el estudio presentado por F. Shih, C. Chuang y P. Wang [3], resalta que el reconocimiento facial tiene un porcentaje del 95.71%, y de igual forma, S. Demircan y H. Kahramanli [9], concluyen que el método Support Vector Machines (SVM) utilizado en el reconocimiento del habla presenta una tasa de éxito del 92.86%. Además de lo anterior, estos enfoques comprenden métodos que se pueden utilizar de manera no invasiva frente al usuario, favoreciendo su comodidad, mitigando la angustia y la presencia de emociones negativas, como lo argumentan los autores, D. Girardi, F. Lanubile y N. Novielli [10]. Por último, se presenta el aporte de investigación donde se tendrán en cuenta los métodos anteriormente mencionados. Para realizar la revisión literaria de cada uno de los ítems mencionados anteriormente, se utilizaron las siguientes cadenas de búsqueda, utilizando conectores lógicos como AND y OR: (evaluation OR evaluación OR assessment OR appraisal) AND (ux OR user experience) AND (mobile OR cell phone OR móvil).

#### **2.1.1. Generalidades de las emociones**

La parte emocional tiene una amplia variedad de estudios en diferentes campos de investigación. Sin embargo, en lo referente a la UX, los investigadores buscan ampliar su comprensión en la detección de emociones mediante la tecnología [11]. Partiendo de esta premisa, buscan lograr una detección más objetiva de las emociones con respecto a aquellos métodos tradicionales que no usan tecnología. Esta detección está basada en dos enfoques: el enfoque dimensional y discreto. El primero abarca las emociones positivas, negativas y neutras [12] [13] [14] [15] que agrupan los distintos tipos de emociones sin especificar alguna(s) en particular; y el segundo, el cual destaca la captura o evaluación de emociones específicas tales como: felicidad, enojo, miedo, disgusto, sorpresa, ira, tristeza y aburrimiento [16] [17] [18] [19] [20].

### 2.1.2. Reconocimiento facial

El reconocimiento de expresiones faciales como método para realizar la captura de emociones, ha tomado gran fuerza gracias a los avances de las cámaras fotográficas que han sido integradas a diferentes dispositivos móviles como tablets, portátiles, celulares, entre otros, esto ha provocado que diferentes investigadores adopten este método en sus estudios. El estudio [21], utiliza este método por medio de una red neuronal constituida por tres componentes utilizados para extraer las características del rostro y posteriormente realizar un etiquetado de la emoción que se está presentando. Se destaca que este método puede ser utilizado en diferentes posiciones del rostro y que, además, con un buen entrenamiento, puede mejorar la precisión y obtener buenos resultados. Los autores W. Sun, H. Zhao y Z. Jin [22] presentan dos patrones centrados en el reconocimiento facial, donde el primero, llamado *Maximum Response-based Directional Texture Pattern* (MRDTP) y el segundo, denominado *Maximum Response-based Directional Number Pattern* (MRDNP), funcionan por medio de la división de imágenes para obtener vectores de características, y así realizar una clasificación rápida y eficiente de las emociones. Donde resaltan que sus patrones son efectivos para eliminar el ruido de las imágenes. Por otra parte, es importante mencionar el estudio realizado por H. García, M. Álvarez, y Á. Orozco [23] donde se presenta un método basado en el proceso de Gauss<sup>2</sup>, el cual consiste en un seguimiento de las características faciales con el objetivo de establecer puntos de referencia que serán más relevantes para categorizar algunas emociones.

Para complementar, se debe mencionar que F. Shih, C. Chuang y P. Wang [17], presentan una recopilación de métodos capaces de realizar el reconocimiento de emociones, enfocados en la felicidad, enojo, disgusto, tristeza, miedo, sorpresa y emociones neutrales. Este estudio experimental resalta que los métodos 2D-LDA (Análisis discriminante lineal) y SVM (Maquinas de vectores de soporte) superan en gran medida a otros métodos, teniendo una tasa de reconocimiento del 95.71%. Además, el estudio [24] presenta un análisis de cómo delinear el rostro para poder realizar un reconocimiento adecuado de las emociones llevado a cabo en ambientes controlados, resalta también que zonas como los ojos y boca representan un gran valor para realizar el reconocimiento.

Por último, estudios como [25] [26] [27] realizan un reconocimiento facial de emociones involucrando la temperatura del rostro. Los autores describen que en el rostro como en cualquier otra zona del cuerpo, existen vasos sanguíneos que pueden producir un cambio en el color de la piel a través de la variación del flujo sanguíneo. Además, C. Benitez-Quiroz [25], presenta una categorización de colores, indicando cuáles de ellos están asociados a emociones negativas y cuales a emociones positivas. Por otro lado, V. Kosonogov *et al.* [26] realiza un estudio por medio de imágenes térmicas, que son capturadas a un grupo de personas a quienes

---

<sup>2</sup> Proceso de Gauss: Generalización del método de reducción, se utiliza para eliminar una incógnita en los sistemas de dos ecuaciones con dos incógnitas.

se les presenta diferentes imágenes, las cuales producen una excitación emocional y a su vez una variación térmica; se resalta que zonas como la punta de la nariz es una importante región para las variaciones térmicas. Por otra parte, K. Lee, H. Hong y K. Park [27] involucran la utilización de cámaras infrarrojas para presentar un método basado en las características simétricas del rostro, por el cual se obtiene un mayor nivel de credibilidad para realizar el reconocimiento del miedo.

### **2.1.2.1. Enfoques para la representación de expresiones faciales.**

En la búsqueda literaria, fueron encontrados dos enfoques para la representación de expresiones faciales, los cuales siguen los principios del reconocimiento de patrones: Uno de ellos se encarga de abordar las características geométricas, y el otro del estudio basado en las apariencias.

#### **2.1.2.1.1. Enfoque de características geométricas.**

Este enfoque se basa en la identificación de rasgos distintivos del rostro como lo son: la boca, nariz, ojos y cejas. Donde aspectos como la distancia entre ellos, el grosor de la nariz, tamaño de la boca, entre otros, son cruciales para el reconocimiento de la emoción [28], porque gracias al aprovechamiento de estos aspectos, los resultados se ven menos afectados por cambios en la orientación, iluminación o sombras que se puedan generar. Razón por la cual el presente proyecto fue encaminado hacia este enfoque.

En el estudio [29] donde se aplica este enfoque, los autores Ingemar J. Cox, Joumana Ghosn y Peter N. Yianilos concluyen que cuantos más puntos característicos son detectados, mayor número de distancias podrán ser calculadas, obteniendo así mejores resultados en el reconocimiento de emociones. Del mismo modo, en estudios como [30] se sugiere que, usando este enfoque, además del algoritmo Viola-jones (usado para identificar rostros), es posible evidenciar una mayor tasa de reconocimiento y por ende una forma de reconocimiento de emociones más efectivo en comparación a otros métodos existentes. En otro estudio Kirby y Sirovich [31], descomponen la imagen facial en varias imágenes, haciendo uso de la expansión de Karhunen y Loeve, además consideran 50 coeficientes de expansión<sup>3</sup>; incluso, aproximando estos parámetros, se logra ejecutar una correcta reconstrucción facial, logrando mejoras en los resultados de reconocimiento.

#### **2.1.2.1.2. Enfoque basado en la apariencia.**

Este enfoque aborda el reconocimiento de una forma holística, donde se considera todo el rostro, como un conjunto de valores de intensidad, valores que son

---

<sup>3</sup> Coeficientes de expansión: Son atributos de la teoría de procesos estocásticos (azar o estadístico) del teorema de Karhunen-Loève generalmente formados por funciones senoidales donde se crean variables aleatorias a partir de un proceso base.

soportados mediante una matriz, que posteriormente, junto con una serie de procesos estadísticos, entrega los posibles resultados. Por este tipo de enfoque se destaca su apreciada aplicabilidad en imágenes con baja resolución y su implementación promete mejoras para usos en sistemas en tiempo real [32].

Uno de los métodos de este enfoque es el método “*Eigenfaces*” [33], construido sobre técnicas de “Análisis de componentes Principales” (PCA), el cual transforma la imagen en un subespacio o conjunto de imágenes de donde es posible obtener una serie de vectores característicos de menor dimensión sin una pérdida importante de información. Además, se destaca que en el estudio [34], se realizó una extensión del PCA convencional a Kernel-PCA, el cual se diferencia y se destaca por ser capaz de extraer características de forma no lineal [35] Aunque el Kernel-PCA logró mejoras, el PCA es uno de los métodos más usados y gracias a la gran acogida por parte de la comunidad científica [36], se desprenden variantes o derivadas que parten de su mismo principio, Uno de ellos es el “*Análisis discriminante Lineal*” (LDA) con su método “*Fisherfaces*” [37], donde demuestra que puede obtener tasas de error más bajas, mejorando su proceso de extracción de características, en comparación al algoritmo PCA con su método “*Eigenfaces*” [38]. Otro método es el “Análisis de componentes independientes” (ICA) que maximiza la información tanto entrante como saliente para encontrar una base de la imagen estadísticamente más independiente y así mejorar la precisión en la detección del rostro [39].

### **2.1.3. Reconocimiento del habla**

El reconocimiento emocional del habla tiene un gran potencial en áreas como psicología, psiquiatría y tecnología computacional [40] [41]. Por lo cual, diferentes investigadores la incluyen en sus estudios sobre el análisis de emociones. Existen estudios como [40] [18] donde se expone que por medio de un conjunto de características relacionadas con el tono y/o balance de la energía espectral generada por la voz, se puede realizar la captura de emociones evocadas por las personas; cabe resaltar que S. Demircan y H. Kahramanli [18] presentan un análisis de diferentes clasificadores como ANN, NB, kNN y SVM, donde se concluye que el clasificador SVM presenta una tasa de éxito del 92,86% con respecto a los demás. Adicionalmente el estudio llevado a cabo por J. B. Alonso *et al.* [41], tiene en cuenta el cambio emocional a largo plazo en el discurso, para lo cual presentan un método basado en la temperatura emocional, el cual consiste en particionar el discurso para ser analizado por tramos, ya que los autores manifiestan que las emociones se pueden presentar en diferentes partes del discurso y además, que pueden ser cambiantes según lo que se esté expresando. Cabe resaltar que estudios como [42] donde se presenta un método basado en DA (Dominio de adaptación) para lograr solventar el problema presentado en diferentes artículos que asumen que las características de los discursos presentados en las etapas de entrenamiento y pruebas van a ser siempre idénticas.

#### **2.1.3.1. Enfoques para el reconocimiento de emociones mediante la voz**

Para abordar el reconocimiento de emociones mediante la voz partimos de dos enfoques: prosódico y espectral, además, de una serie de características que se obtienen de la voz para hacer su asociación con el campo emocional.

#### **2.1.3.1.1. Enfoque prosódico.**

Hace referencia a la frecuencia de tono, el acento, la intensidad y medidas de la calidad de la voz.

#### **2.1.3.1.2. Enfoque espectral.**

Este enfoque incluye una serie de coeficientes de medición entre los que se encuentran:

- Coeficientes cepstrales de frecuencia de Mel (MFCC), estos se enfocan en la representación del habla por medio de señales acústicas, teniendo en cuenta dos parámetros fundamentales: Escala Mel y Espectro de frecuencia Mel. Además, de unos intervalos de tiempo obtenidos de las señales emitidas por la voz [43].
- Coeficientes cepstrales de predicción lineal (LPCC), donde se determina un conjunto único de coeficientes predictorios, realizando operaciones matemáticas entre las muestras de voces reales y las predichas, para obtener la predicción.
- Coeficientes de potencia de frecuencia logarítmica (LFPC).
- Coeficientes de predicción lineal perceptual (PLP).

#### **2.1.3.1.3. Características generales de la voz.**

A continuación, se dará una descripción breve de las características generales que se obtienen de la voz humana para su posterior asociación con el campo emocional.

- El tono: Representa la frecuencia fundamental a la que vibran las cuerdas vocales cuando se habla.
- Los formantes: Son frecuencias de resonancia de los tractos vocales; no se relaciona directamente con la frecuencia fundamental y pueden permanecer constantes en la medida que la frecuencia fundamental cambia. Si la frecuencia fundamental es baja en relación con el rango de los formantes, la calidad de sonido es óptimo. Por el contrario, si la frecuencia fundamental está por encima de los formantes, el sonido será tenue.
- Densidad espectral de Potencia: describe la distribución de potencia de una señal de voz con respecto a la frecuencia. Además, muestra si la energía de esta señal es fuerte o débil según la frecuencia. La energía o potencia proviene de la vibración de las cuerdas vocales, frecuencia del tracto vocal, o del movimiento de los labios y la mandíbula.
- El jitter: se refiere a las perturbaciones generadas por la fuente glotal donde las fonaciones vocales afectan en alguna medida el tono glótico

- Shimer: se refiere a las perturbaciones generadas por la fuente glotal donde las fonaciones vocales afectan la energía glotal.

## 2.2. Caracterización de la UX en Teléfonos Inteligentes

Esta sección presenta el procedimiento que se ha efectuado para realizar la caracterización de la UX en teléfonos inteligentes. Este procedimiento se divide en tres fases: Definición del protocolo de caracterización, desarrollo de la caracterización y conclusiones del proceso de caracterización.

Esta caracterización se realizó con el propósito de identificar y establecer los atributos necesarios que describen la experiencia emocional en teléfonos inteligentes.

Para el desarrollo de esta tarea se tomó como base el protocolo definido por Jhon Bermudez y Juan Vidal [44], el cual permite identificar y examinar diferentes estudios teniendo en cuenta los criterios definidos, luego de identificar los estudios se define una fase de análisis de la literatura, donde se realiza un proceso de abstracción de las características obteniendo como resultado un amplio conjunto de aspectos que se deben tener en cuenta para describir la experiencia emocional en teléfonos inteligentes. Finalmente, el proceso se concluye con la extracción de las características más relevantes que definan la experiencia emocional en un teléfono inteligente.

### 2.2.1. Definición del protocolo de caracterización

Teniendo en cuenta el estudio [44] el protocolo indica establecer diferentes etapas para el proceso de caracterización, con esto se permite buscar, seleccionar y analizar los atributos de la experiencia de usuario en un teléfono inteligente.

El protocolo se divide en tres etapas las cuales se describen a continuación:

- **Búsqueda de estudios**

En esta etapa se procura realizar una búsqueda en la literatura de diferentes estudios los cuales proporcionen información sobre las características de la UX en teléfonos inteligentes, esto se hace con el fin de proporcionar un mayor objetividad y precisión en la caracterización, para hacerlo se toma como referencia el estudio realizado por Francisco J. Pino [45] de donde se extraen las tres primeras fases de una revisión sistemática, con esto se espera encontrar los estudios más relevantes.

Las tres fases tomadas como referencia son:

- ✓ **Formulación de la pregunta de búsqueda:** en esta fase se busca definir la pregunta, la cual será resuelta a partir de los estudios obtenidos, además de identificar las palabras claves.

- ✓ **Selección de fuentes:** en esta fase se genera una cadena de búsqueda partiendo de las palabras claves definidas en la anterior etapa. Además, se busca seleccionar las fuentes donde se realizará la búsqueda y se definen los criterios necesarios.
- ✓ **Elección de estudios:** a partir de los resultados obtenidos en la búsqueda, en esta fase eligen los estudios que se consideren más adecuados y den un mayor valor para estudiarlos.

- **Selección de estudio**

En esta etapa se realiza una definición de los criterios de selección que deben cumplir los estudios encontrados en la literatura, los cuales serán analizados en la etapa de elaboración de la caracterización. Además de los criterios, se busca realizar un análisis de los estudios para determinar los más idóneos para la caracterización

- **Elaboración de la caracterización**

En esta etapa se define el propósito de la caracterización y se realiza el análisis de los estudios que resaltaron de la etapa anterior. A partir de dicho análisis se obtienen unos resultados los cuales ayudarán a la elaboración de la caracterización.

## **2.2.2. Desarrollo de la caracterización**

### **2.2.2.1. Búsqueda de estudios**

En esta etapa se busca examinar los estudios relacionados con la UX en teléfonos inteligentes, por medio del protocolo presentado en la sección anterior. Este proceso permite identificar estudios en los cuales se aborden diferentes características que se deben tener en cuenta respecto a la UX en teléfonos inteligentes.

#### **2.2.2.1.1. Formulación de la pregunta de búsqueda**

Como se planteó, en esta primera fase se definirá la pregunta a resolver, a partir de la cual se obtendrán diferentes palabras claves las cuales ayudarán a realizar la búsqueda.

Como pregunta se plantea ¿Cuáles son las características que describen la UX en teléfonos inteligentes?

Como palabras claves se utilizarán: UX, user experience, smartphone, mobile phone.

#### **2.2.2.1.2. Selección de fuentes**

Por medio de conectores lógicos AND y OR se realizará el enlace de las diferentes palabras claves, con el objetivo de crear una cadena de búsqueda, que represente de manera apropiada la pregunta formulada. Esta cadena será usada en diferentes motores de búsqueda para obtener los resultados acordes. La cadena de búsqueda se puede visualizar en la siguiente **Figura 1**.

```
(characteristics OR property OR attributes)
AND (ux OR user experience)
AND ("smartphone" OR "mobile phone")
```

**Figura 1. Cadena de búsqueda básica**

### **2.2.2.1.3. Elección de estudios**

Después de realizar la búsqueda en las bases de datos Science Direct, IEEEExplore, Springer y de utilizar el buscador Google Scholar para abarcar una mayor cantidad de información, se obtuvieron un total de 105 estudios, con el objetivo de identificar los artículos potencialmente relevantes para responder la pregunta planteada, se aplicaron diferentes criterios de inclusión y exclusión. En el caso de los criterios de inclusión se definieron: un criterio general y tres específicos. El primero (criterio general de inclusión), corresponde a un primer acercamiento de la literatura con un umbral amplio, pero restringido en el tiempo en el cual se ha visto un mayor crecimiento en el uso de las tecnologías móviles. El segundo (criterios específicos de inclusión), concierne al interés de generar filtros que permiten la recopilación de información que responda de forma coherente a las necesidades propias de la presente investigación. Para el caso de los criterios de exclusión, no se consideró ninguna particularidad destacable, salvo la que se describe a continuación:

#### **Criterio de inclusión general:**

- 1) Literatura relevante relacionada con el área de investigación a partir del año 2010 en adelante

#### **Criterios de inclusión específicos:**

- 1) Estudios que reporten características de la UX en teléfonos inteligentes.
- 2) Estudios que describan que características se deben tener en cuenta para evaluar la UX en teléfonos inteligentes.
- 3) Estudios que reporten ventajas y desventajas de evaluar la UX en teléfonos inteligentes.

Por último, se definió un único criterio de exclusión, en el cual se considera el cumplimiento de los mencionados anteriormente, este criterio es:

#### **Criterio de exclusión:**

- 1) Cualquier estudio que no cumpla con ninguno de los criterios mencionados anteriormente.

A partir del proceso de filtrado se obtuvo como resultado un total de 18 estudios:

- Optimization of user experience in mobile application design by using a fuzzy analytic-network-process-based Taguchi method [46]
- An Exploratory Study Using Electroencephalography (EEG) to Measure the Smartphone User Experience in the Short Term [47]
- Identifying Characteristics and Types of Generation Z according to the Behavior of Smartphone Camera Use [48]
- Aesthetic Characteristics and UX Evaluation of Mobile Platforms [49]
- Web browsing on mobile phones: characteristics of user experience [50]
- Developing Elements of User Experience for Mobile Phones and Services: Survey, Interview, and Observation Approaches [51]
- Exploration and optimization of user experience in viewing videos on a mobile phone [52]
- A factor combination approach to developing style guides for mobile phone user interface [53]
- Redesign of User Interface Zakat Mobile Smartphone Application with User Experience Approach [54]
- Evaluation of smartphone user experience: identification of key evaluation items and their relationships [55]
- Product attributes and user experience design: how to convey product information through user-centered service [56]
- Culturability in Mobile Data Services: A Qualitative Study of the Relationship Between Cultural Characteristics and User-Experience Attributes [57]
- Worthwhile Travel Time: Design Challenges of Capturing the User Experience by Smartphone [58]
- A Fuzzy Cognitive Map for Identifying User Satisfaction from Smartphones [59]
- Full-system analysis and characterization of interactive smartphone applications [60]
- Towards Analyzing Mobile App Characteristics for Mobile Software Development [61]
- On methods of designing smartphone interface [62]
- Mobile Phone Usability and its Influence on Brand Loyalty and Re-Purchase Intention: An Empirical [63]

#### **2.2.2.2. Selección de estudios**

La documentación conservada después de la elección, fueron analizados con el fin de identificar cuales cumplen los criterios de selección, determinando así, si serán o no incluidos en la caracterización.

##### **2.2.2.2.1. Criterios de selección**

A continuación, se describen los criterios usados para determinar si un estudio será o no será incluido en la caracterización

- 1) **Características únicamente de la UX en teléfonos inteligentes:** Estás no deben basarse en características del usuario y/o de los entornos. Serán considerados los estudios que planteen o sugieran características que se deben tener en cuenta para describir la UX en teléfonos inteligentes, y serán descartados los estudios que no cumplan esta condición.
- 2) **Características de UX en teléfonos inteligentes relacionado con aplicaciones móviles.** Teniendo en cuenta que la evaluación se va a hacer por a una aplicación móvil por medio del teléfono inteligente, se contemplan aquellos estudios donde se definan características que aborden las aplicaciones móviles. No se consideran estudios que propagan características fuera de este contexto.

#### 2.2.2.2. Análisis de los estudios

Después de aplicar un análisis a los estudios obtenidos de la fase anterior, se logra identificar cuales cumplen con los criterios de selección. La siguiente **Tabla 1** resalta que criterio(s) cumple cada estudio

**Tabla 1 Criterios cumplidos por cada estudio.**

Estudios	Criterios de Selección	
	Primer Criterio	Segundo Criterio
Optimization of user experience in mobile application design by using a fuzzy analytic-network-process-based Taguchi method [46]	✓	✓
An Exploratory Study Using Electroencephalography (EEG) to Measure the Smartphone User Experience in the Short Term [47]	✓	✓
Identifying Characteristics and Types of Generation Z according to the Behavior of Smartphone Camera Use [48]	✓	✓
Aesthetic Characteristics and UX Evaluation of Mobile Platforms [49]	✓	✓
Web browsing on mobile phones: characteristics of user experience [50]	✓	✓
Developing Elements of User Experience for Mobile Phones and Services: Survey, Interview, and Observation Approaches [51]	✓	x
Exploration and optimization of user experience in viewing videos on a mobile phone [52]	✓	x
A factor combination approach to developing style guides for mobile phone user interface [53]	✓	x

Redesign of User Interface Zakat Mobile Smartphone Application with User Experience Approach [54]	✓	✓
Evaluation of smartphone user experience: identification of key evaluation items and their relationships [55]	✓	✓
Product attributes and user experience design: how to convey product information through user-centered service [56]	x	x
Culturability in Mobile Data Services: A Qualitative Study of the Relationship Between Cultural Characteristics and User-Experience Attributes [57]	x	x
Worthwhile Travel Time: Design Challenges of Capturing the User Experience by Smartphone [58]	✓	x
A Fuzzy Cognitive Map for Identifying User Satisfaction from Smartphones [59]	✓	x
Full-system analysis and characterization of interactive smartphone applications [60]	✓	✓
Towards Analyzing Mobile App Characteristics for Mobile Software Development [61]	x	✓
On methods of designing smartphone interface [62]	✓	x
Mobile Phone Usability and its Influence on Brand Loyalty and Re-Purchase Intention: An Empirical [63]	✓	x

### 2.2.2.2.3. Resultado de la selección

Teniendo en cuenta el punto anterior, 8 de los estudios cumplieron con los criterios de selección, por lo tanto, son tomados en cuenta como insumo para la caracterización, a continuación, se muestran los estudios:

- Optimization of user experience in mobile application design by using a fuzzy analytic-network-process-based Taguchi method [46]
- An Exploratory Study Using Electroencephalography (EEG) to Measure the Smartphone User Experience in the Short Term [47]
- Identifying Characteristics and Types of Generation Z according to the Behavior of Smartphone Camera Use [48]
- Aesthetic Characteristics and UX Evaluation of Mobile Platforms [49]
- Web browsing on mobile phones: characteristics of user experience [50]
- Redesign of User Interface Zakat Mobile Smartphone Application with User Experience Approach [54]
- Evaluation of smartphone user experience: identification of key evaluation items and their relationships [55]
- Full-system analysis and characterization of interactive smartphone applications [60]

### 2.2.2.3. Elaboración de la caracterización

A partir de los estudios seleccionados en el apartado anterior, se realizó un análisis en detalle con el fin de identificar y extraer las características relacionadas con la UX en teléfonos inteligentes. Todas las características extraídas del proceso se sometieron a un proceso de análisis comparativo para identificar similitudes entre ellas y lograr unirlas. Como resultado se obtuvo un listado de características con su respectiva descripción. El detalle de este listado se puede ver en el **Anexo A**.

En la **Tabla 2** se observa el listado de características resultantes de la revisión de la literatura.

**Tabla 2. Listado de características de UX en teléfonos inteligentes resultante de la revisión de la literatura**

<b>Característica</b>	<b>Descripción</b>
<b>Simplicidad</b>	Identifica la facilidad que tiene el teléfono inteligente y/o la aplicación móvil de ser utilizado.
<b>Minimalista</b>	Identifica el punto estético, diseñado para facilitar su uso.
<b>Retroalimentación</b>	La relación entre el teléfono inteligente y el usuario debe ser bidireccional, dando información necesaria para que el usuario sepa que está sucediendo.
<b>Conexión</b>	Es un componente esencial principalmente en sistemas que requieren una navegación web.
<b>Pantalla</b>	La resolución y contraste de la pantalla genera una mejor aceptación de los usuarios, ya que permiten una mejor legibilidad de la información.
<b>Sistema operativo</b>	El sistema operativo del teléfono inteligente juega un papel importante para el usuario, ya que un sistema operativo desconocido puede afectar otras características de usabilidad.
<b>Empatía</b>	El teléfono inteligente debe dar impresión de sentir, percibir y responder a lo que el usuario está sintiendo, generando a través de sus interfaces sensaciones de confianza.
<b>Aprendizaje automático</b>	El teléfono inteligente debe ser capaz de aprender de las acciones del usuario.
<b>Satisfacción</b>	Grado en el que un teléfono y/o aplicación móvil proporciona a los usuarios satisfacción consigo mismos.
<b>Franqueza</b>	Grado en el que el usuario percibe que tiene control sobre una aplicación.
<b>Eficiencia</b>	Grado en el que una aplicación y/o teléfono inteligente permite completar tareas sin perder tiempo y esfuerzo.

<b>Flexibilidad</b>	Grado en el que una aplicación se adapta a cambios en tareas o entornos más allá de los inicialmente especificados.
<b>Capacidad de aprendizaje</b>	Tiempo y esfuerzo necesario para que un usuario aprenda a usar el teléfono inteligente y/o una aplicación móvil.
<b>Atractivo</b>	Percepción del usuario sobre el atractivo del teléfono inteligente y/o aplicación móvil.
<b>Placer</b>	Grado con que el usuario experimenta placer al interactuar con el teléfono inteligente o aplicación móvil.

Además, para dar más claridad en la **Tabla 3** se listan las características y se indica el número de estudios en los que fueron considerados.

**Tabla 3. Listado de características con número de estudios donde se considera.**

<b>N°</b>	<b>Característica</b>	<b>Número de estudios</b>
1	Simplicidad	4
2	Minimalista	5
3	Retroalimentación	6
4	Conexión	2
5	Pantalla	4
6	Sistema operativo	2
7	Empatía	3
8	Aprendizaje automático	1
9	Satisfacción	6
10	Franqueza	1
11	Eficiencia	5
12	Flexibilidad	1
13	Capacidad de aprendizaje	3
14	Atractivo	2
15	Placer	2

### 2.2.3. Conclusiones del proceso de caracterización

Durante el proceso de caracterización se observó la relación entre los diferentes estudios analizados y las características identificadas, lo que permite, resaltar que estudios aportaron una mayor cantidad, convirtiéndolos así, en los estudios con mayor relevancia para la presente investigación. Además, se puede destacar aquellas características las cuales los distintos autores coinciden teniendo en cuenta el número de estudios que las consideran.

Cabe resaltar que también se incluyeron características asociadas a la UX en aplicaciones móviles, esto es necesario ya que uno de los focos principales es el análisis de la UX en dispositivos móviles por medio del uso de una aplicación móvil. Con este análisis se logra determinar características potenciales, que se deben considerar durante la presente investigación.

### **2.3. Contextualización de Sistemas Multimodal**

El estado emocional de la mente humana se puede manifestar de diferentes modos en el cuerpo humano: Señales faciales, cerebrales, de voz, gestuales, posturales y biopotenciales. El término multimodal en esta investigación se refiere al reconocimiento de emociones a través de más de una de estas manifestaciones. Aunque las técnicas de reconocimiento que usan un solo modo han sobresalido en el reconocimiento de emociones, no en todos los casos han logrado proporcionar información suficiente, la cual es requerida para abordar el reconocimiento de emociones desde más de una perspectiva. Por esta razón se parte de la posibilidad de mejorar la fiabilidad del proceso de reconocimiento de emociones al contar con más de un parámetro de medición (reconocimiento multimodal).

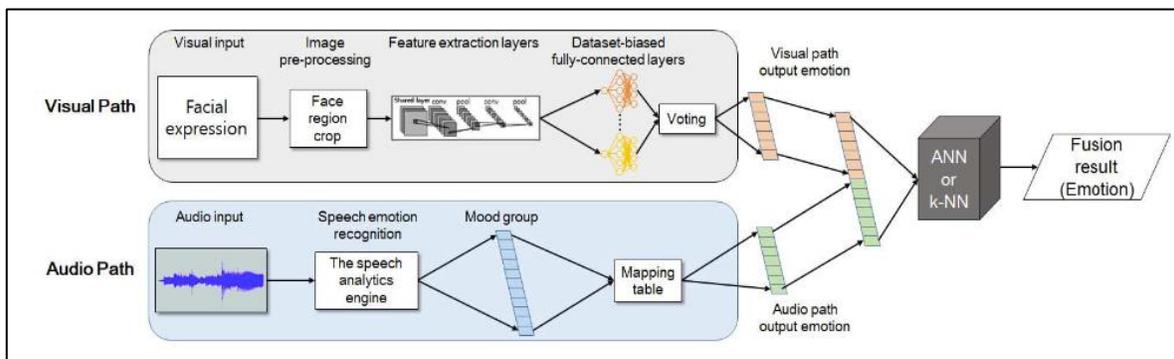
Para evaluar sistemas unimodales, se usa una herramienta muy común, que se denomina matriz de confusión, ésta se usa para evaluar el desempeño de los algoritmos. Dicha matriz, en términos prácticos permite vislumbrar el tipo de aciertos y errores que está teniendo el modelo a la hora de pasar por el proceso de aprendizaje. Sin embargo, a la hora de evaluar sistemas multimodal no se obtuvieron buenos resultados, debido a que, al aplicar una matriz de confusión para cada modalidad independiente, sus resultados muestran unas diferencias significativas con respecto a cada modalidad [64]. Además, que estudios como [65] expresan que al momento de procesar los datos multisensoriales de forma separada y combinarlos al final, se está haciendo de manera incorrecta ya que las personas muestran señales audiovisuales de forma complementaria y redundante.

Por lo anterior, y para mejorar la confiabilidad del reconocimiento de forma multimodal, se van a integrar y combinar las modalidades individuales en una sola representación. Primordialmente, en la literatura se encontraron dos tipos de técnicas de fusión de modalidades, que se han usado para mejorar la confiabilidad del reconocimiento de emociones multimodal:

- Fusión a nivel de características. En el primer paso de esta fusión, se clasifican las emociones de los sujetos de manera individual, o sea por modalidad, para luego fusionar los resultados de la clasificación de cada modalidad con el objetivo de minimizar la incertidumbre. Un método simple de fusión es verificar si los resultados de los clasificadores (emoción 1, emoción 2, emoción 3...) para las múltiples modalidades producen el mismo tipo de emoción para un sujeto. Si coincide en un buen porcentaje, la decisión sobre el tipo de emoción para el sujeto es certera. De otro modo, si los clasificadores difieren en sus resultados y las etiquetas de cada emoción entran en conflicto para un sujeto de prueba, entonces se toma la etiqueta con el mayor porcentaje de los clasificadores.

- Fusión a nivel de decisión. Esta técnica, unifica todas las mediciones obtenidas de los diferentes sensores para todas las modalidades, y posteriormente alimentar un clasificador previamente entrenado, que, a su vez, determina la etiqueta de la emoción desconocida.

Los autores de la referencia [66] aplican en su investigación la fusión tanto a nivel de características como a nivel de decisión, en las modalidades de gestos faciales y corporales, así mismo concluyen que la fusión a nivel de características obtiene un mejor rendimiento que la fusión a nivel de decisión. En otras referencias en donde también se aborda la fusión con rasgos faciales y corporales, como lo es la [64], los autores concluyen que el rendimiento general del clasificador multimodal en este caso bimodal, a nivel de fusión de características como el de fusión de decisión, presentan unos resultados similares; pero al analizar dichos resultados con una matriz de confusión, observaron que la tasa de reconocimiento para cada tipo de emoción muestra discrepancias significativas. En la clasificación bimodal a nivel de fusión de características, la ira y el estado neutro se reconocieron con mayor precisión en comparación con el clasificador unimodal de expresiones faciales. En el clasificador bimodal de nivel de decisión, la felicidad y la tristeza proporcionaron mayor precisión que la fusión de características. Por lo tanto, en cuanto a técnicas de fusión multimodal, se concluye que la mejor técnica de fusión depende de la aplicación. Una conclusión más general, es que haciendo uso de las modalidades audiovisuales es factible reconocer estados afectivos humanos con una alta precisión.



**Figura 2. Ejemplo diagrama de sistema multimodal.**

## 2.4. Inteligencia artificial.

La inteligencia artificial (IA) es una combinación de algoritmos planteados con el propósito de apoyar diferentes aspectos entre ellos la vida social humana [67], para ello se busca que las maquinas puedan hacer la misma clase de cosas que puede hacer la mente humana, por ejemplo: razonar, predecir, planificar, controlar, entre otras [68]. Lograr que una maquina llegue a comportarse como una mente humana, permitirá que pueda ser aplicada en diferentes aspectos de la vida [68], entre ellos encontramos (i) el hogar, donde se resalta el uso de la IA en los vehículos

inteligentes, los cuales pueden ser conducidos sin necesidad de una persona ayudando así a la seguridad vial de las personas, ya que permite tener medidas mayores de seguridad donde se puede reconocer la detección de peatones u otros vehículos [69] [70]; (ii) el internet de las cosas, por medio del cual se logra la conexión de diferentes sensores físicos, que son añadidos a aparatos electrónicos, ropa y el mismo entorno, además autores como Cristian González García *et al* [70] resaltan que este tipo de sistemas ayudan a realizar el reconocimiento de objetos, alarmas con el objetivo de alertar de forma rápida y eficaz cuando una persona mayor sufre un accidente; (iii) animaciones, destacándose el uso de IA para realizar las animaciones de Hollywood y videojuegos, enfocándonos en el uso de IA para videojuegos, los autores Georgios N. Yannakakis y Julian Togelius [71], resaltan que IA está siendo utilizada para el control de personajes no jugados, generando contenido capaz de adaptarse a diferentes tipos de jugadores, con esto se está logrando que los videojuegos sean más diversos y sofisticados, también resaltan que la IA en videojuegos puede apoyar en su diseño y desarrollo, ya que proporciona una perspectiva de lo que un jugador siente mientras está realizando las respectivas pruebas; (iv) la economía, éste contexto es tal vez uno de los campos donde más se ha explotado el uso de la IA, ya que puede contribuir en gran medida al crecimiento sostenible de la economía de un país, como lo resaltan Humin Lu *et al.* [67], que en los últimos años, la IA ha traído un gran impulso económico en países desarrollados como Estados Unidos y países Europeos, esto debido a su avance en el desarrollo de nuevas tecnologías de comunicación de la información de IA y en tecnología robótica. Cabe resaltar que los autores resaltan que este beneficio no ha sido solo para países desarrollados, dado que países como India son pioneros en diferentes tecnologías que involucran la IA. Además se debe mencionar que autores como Thomas H. Davenport y Rajeev Ronanki [72], resaltan que la IA es útil para las empresas, ya que ayuda a soportar tres importantes necesidades comerciales, la primera necesidad relacionado con la automatización de procesos comerciales, la cual se centra en el uso de IA para ejecutar las tareas repetitivas sin la necesidad de la intervención humana, apoyándose en sistemas informáticos o procesos robóticos, entre las tareas más comunes se incluyen: transferencia de datos, leer documentos, conciliación de fallas en el cobro de servicios, etc. La segunda necesidad, el conocimiento cognitivo, el cual se centra en la detección de patrones en grandes volúmenes de datos con el fin de interpretar su significado, este tipo de aplicaciones son conocidas también como aplicaciones de aprendizaje automático, entre las funciones que pueden realizar se encuentran: predecir la probabilidad de que un cliente compre un producto en particular, identificación de fraudes, analizar datos de garantía, automatizar la personalización de anuncios, etc. Por último, la tercera necesidad, compromiso cognitivo, el cual se centra en la interacción con los clientes, las aplicaciones más realizadas de este tipo son: los chatbots y agentes inteligentes, estas aplicaciones se centran en funciones como: ofrecer un servicio al cliente 24/7, sitios para responder preguntas, sistemas de recomendación de productos, entre otras.

## **2.5. Machine learning.**

Machine learning o aprendizaje automático es un subcampo de las ciencias de la computación y una rama de la IA, el cual tiene como objetivo programar equipos para usar datos de ejemplos o experiencias pasadas para resolver un problema determinado [73]. Actualmente, existen muchas aplicaciones exitosas de machine learning, que involucran sistemas que analizan los datos de las ventas anteriores con el fin de predecir el comportamiento de un cliente en sus futuras compras, o sistemas que optimizan el comportamiento del robot para que una tarea se pueda completar con recursos mínimos y logre extraer el conocimiento aprendido [73]. Muchas de estas aplicaciones se centran en la utilización de algoritmos de machine learning, tales como: máquinas de vectores (SVM), K-vecino más cercano (KNN), regresión logística (Logistic Regression), Impulsar, etc. Debido a que proporcionan un gran poder de procesamiento por medio del uso de la unidad de procesamiento gráfico (GPU), apoyando en el proceso de grandes volúmenes de datos.

Los algoritmos de machine learning difieren en su enfoque, por lo cual se clasifican en tres tipos como lo menciona [74], los tres tipos son: (i) aprendizaje supervisado (SL), el cual abarca aquellos algoritmos en los cuales su función de aprendizaje se centra en el mapeo de una entrada a una salida. Para hacerlo el algoritmo requiere que se utilice una cantidad de datos de entrenamiento etiquetados, los cuales son analizados y producen una salida inferida a partir de la información dada, entre los algoritmos más utilizados, se encuentran: regresión lineal, regresión logística (LS), máquinas de vectores (SVM), K-vecinos más cercanos (KNN) y el árbol de decisiones; (ii) aprendizaje semi-supervisado (SSL), es una clase de aprendizaje automático de tareas y técnicas que también hace uso de datos etiquetados, pero con la diferencia de que el conjunto de datos se compone de una pequeña cantidad de datos etiquetados y una gran cantidad de datos sin etiquetar. Este tipo de aprendizaje consiste en realizar un entrenamiento previo con la información etiquetada y posteriormente se hace uso de los datos sin etiquetas, muchos investigadores han encontrado que la utilización de este tipo de datos puede producir considerables mejoras en la precisión del aprendizaje con respecto a uno sin supervisión, pero sin el tiempo y costes necesarios que requiere un supervisado; (iii) por último encontramos el aprendizaje no supervisado (por sus siglas en inglés USL), el cual consiste en los algoritmos de aprendizaje automático que aprenden de los datos que no han sido etiquetados, clasificados o categorizados, para hacerlo realizan un análisis de la información a partir de la búsqueda de las relaciones entre los datos, por medio de patrones o características encontrados en ellos. Algunos de los algoritmos más comunes de este tipo son: K-means, Apriori, self-organizing maps (SOM), análisis de componente principales (PCA), etc.

## **2.6. Deep learning.**

El deep learning o aprendizaje profundo es un enfoque de machine learning que mediante algoritmos permite a los ordenadores aprender de la experiencia y comprender el mundo, a partir de métodos utilizados para el análisis de datos [75]. El deep learning se basa en el uso de redes neuronales artificiales en semejanza al sistema nervioso humano, además cuenta con una serie de niveles o capas, que se

encargan de detectar determinadas características propias del ser humano, en donde cada nivel o capa se encarga de realizar tareas específicas para dicho fin.

El deep learning proporciona una mejora considerable en comparación con algoritmos más tradicionales, ya que puede ser aplicado en la solución de varios problemas que se encuentran en diversos campos, como lo son: utilización de imágenes en lugar de palabras claves para buscar como lo hace el buscador de google; identificar marcas, orientar anuncios, predicciones de preferencias de clientes como lo hace facebook; identificar clientes potenciales de un producto; mejorar la comprensión de enfermedades, mutaciones de enfermedades y terapias genéticas; análisis de imágenes médicas como radiografías, resonancias magnéticas [75]; detección, predicción y prevención de amenazas sofisticadas en tiempo real en el campo de la ciberseguridad [76]; identificación de sentimientos y minería de opiniones en las redes sociales y blogs [77]; localización de caras e identificación de emociones faciales [78]; reconocimiento de voz [79]; entre otras. En una manera general los algoritmos de deep learning al igual que machine learning pueden utilizar el aprendizaje supervisado o aprendizaje no supervisado según la forma de abordar el problema y se encargan primordialmente en el reconocimiento de patrones.

## **2.7. Redes neuronales.**

Al igual que nuestro cerebro está constituido por neuronas interconectadas unas con otras, de manera similar una red neuronal artificial está formada por neuronas artificiales conectadas entre sí y agrupadas en diferentes niveles denominados capas o nodos. Estas, al ser artificiales se describen en un ámbito computacional como modelos que procesan información y que imitando el funcionamiento de una red biológica se encargan de transmitir, recibir, procesar y enviar información entre ellas [80].

Las redes neuronales están conformadas por 3 tipos de nodos: Nodos de entrada (input), encargados de recibir la información desde el exterior de la red; Nodos de salida (output), encargados de enviar información hacia el exterior de la red; y nodos ocultos los cuales transmiten la información entre los nodos de la red, encontrándose estos entre los nodos de entrada y salida [80]. Cabe resaltar que los nodos ocultos no tienen contacto con el exterior y que generalmente las redes neuronales artificiales suelen estar conformadas por múltiples capas de estos nodos [80].

El funcionamiento de estas redes neuronales artificiales se origina en los nodos de entrada donde se reciben los datos del exterior, estos son dirigidos al interior de la red hacia los nodos ocultos los cuales procesan, modifican y transfieren la información de una capa a otra en un proceso denominado como “aprendizaje” en donde cada capa de nodos ocultos va aprendiendo de las capas más externas [80]. Este proceso de aprendizaje es el principio que fundamenta el funcionamiento de deep learning.

Existen diferentes arquitecturas de redes neuronales artificiales en donde se diversifican, no por su número de capas o número de neuronas en cada capa, sino por la conexión que existe entre neuronas o capas, por el tipo de cada neurona presente o incluso por la forma en la cual son entrenadas [81]. Entre las redes neuronales artificiales que resalta la literatura se encuentran:

Las redes convolucionales (CNN) esencialmente usadas en el reconocimiento de patrones, las cuales se dividen en dos secciones: una sección formada por capas convolucionales y de agrupación, encargadas de la identificación de los patrones gráficos y la otra sección tiene como función la clasificación de los datos recibidos [82].

Redes neuronales recurrentes (RNN) las cuales se caracterizan por permitir el procesamiento de datos secuenciales y explotar las dependencias entre dichos datos, son utilizadas ampliamente en tareas de generación de texto, como: traducciones automáticas, generación de subtítulos mediante imágenes y en el reconocimiento de voz [83].

## **2.8. Aportes de Investigación**

A partir de la revisión de la literatura se detectaron tres factores importantes que guiaran la investigación: (i) La mayoría de los estudios encontrados sobre la captura y medición de emociones utilizan métodos invasivos, lo cual ocasiona que las propias emociones experimentadas por los usuarios pueden ser producidas o alteradas por los instrumentos utilizados en los métodos. Por otra parte, se ha identificado que el empleo de métodos invasivos puede generar mucho tiempo para ser resueltos y podrían obtener un estado emocional ambiguo debido a que las emociones son inmediatas y automáticas, lo cual puede alterar los resultados sobre lo que el usuario siente realmente [7], (ii) los métodos que realizan la captura de emociones encontrados en la literatura, generalmente utilizan una medida (unimodal), y aunque cuentan con un alto porcentaje de reconocimiento en sus resultados, dicha información puede ser capturada de manera más completa considerando el uso de varias medidas en forma paralela (medición multimodal) (iii) por último, no se hallaron estudios relacionados a herramientas que realicen la evaluación de la UX utilizando métodos no invasivos y una medición multimodal, siendo una oportunidad para enfocar la investigación.

Por lo tanto, se propone un prototipo software, desarrollado con base a la caracterización e integración de métodos no invasivos, que tomará como insumo, datos basados en el reconocimiento del rostro y habla en un entorno natural, generados a partir de la interacción del usuario con una aplicación ejecutada en un teléfono inteligente. Finalmente, obtendrá como salida el resultado de la evaluación de la UX, donde se considera las emociones presentes durante la interacción. Además, el prototipo software estará basado en un enfoque multimodal, el cual utilizará las medidas de reconocimiento del rostro y habla

## CAPITULO 3. INTEGRACIÓN DE LOS MÉTODOS NO INVASIVOS PARA LA EVALUACIÓN DE LA UX CON TELÉFONOS INTELIGENTES

### 3.1. Herramientas para la captura de datos en teléfonos inteligentes.

Teniendo en cuenta que la evaluación de la UX se realizará de forma no invasiva para el usuario, se hace una búsqueda de las herramientas capaces de capturar los sonidos provocados por la voz, además de los gestos y movimientos del rostro sin que el usuario perciba o se llegue a sentir incómodo en el momento de la captura de dicha información. Estos datos, serán el insumo para la identificación y posterior evaluación de las emociones, proceso que se llevará a cabo a través del prototipo software propuesto.

Entre las herramientas encontradas para realizar la captura de los datos, se resaltan: AutoStart, Automate y MacroDroid capaces de realizar el inicio automático de otras aplicaciones en sistemas Android, las cuales se comparan en la **Tabla 4**.

**Tabla 4. Comparación de herramientas que permiten la ejecución de otras aplicaciones.**

Características	Herramientas		
	MacroDroid	Automate	AutoStart
Requiere permisos de súper usuario	NO	NO	NO
Fácil configuración	SI	NO	SI
Ejecuta otras aplicaciones en el instante deseado	SI	SI	NO
Calificación en Play Store	4.5	4.5	3.3

De acuerdo con la anterior comparación, se elige la herramienta “MacroDroid” por su simplicidad a la hora de realizar su configuración. Además, permite la ejecución de otras aplicaciones en el instante que se desee, consiguiendo de esta manera que se integre con la herramienta de grabación de video “Grabador one touch”.

Por otra parte, con el objetivo de complementar la captura de los datos de manera no invasiva, se encuentran las herramientas ¡Grabadora vídeo, Background vídeo record y Grabador one touch, las cuales permiten realizar la grabación de vídeo en

segundo plano en sistemas Android. A continuación, se presenta un cuadro comparativo de dichas herramientas (**Tabla 5**):

**Tabla 5. Comparación de herramientas de grabación en segundo plano.**

Características	Herramientas		
	¡Grabadora vídeo	Background vídeo record	Grabador one touch
Grabar con cámara trasera y delantera	SI	SI	SI
Fácil configuración	SI	SI	SI
Múltiple resolución de vídeo	SI	SI	SI
Integración con otras aplicaciones	NO	SI	SI
Permite ocultar vídeos	NO	NO	SI
Notificación de inicio de grabación	SI	SI	NO
Calificación Google Drive	4.5	4.4	4.1

De acuerdo con la información mostrada en la Tabla 2, se selecciona la denominada “Grabador one touch” debido a que permite realizar el inicio de la grabación sin que el teléfono inteligente notifique la acción al usuario, característica por la cual se diferencia de las otras herramientas, contribuyendo a la obtención de datos de manera no invasiva. Además, permite ocultar las grabaciones para evitar que sean manipuladas por el usuario y afecte la idoneidad de los resultados.

### **3.2. Elección de herramientas para el reconocimiento de emociones.**

En esta sección se describe la selección de los algoritmos que sirvieron como base para el reconocimiento de emociones; uno de ellos parte de la expresión facial y otro a partir del reconocimiento de la voz.

#### **3.2.1. Herramientas para el reconocimiento de expresiones faciales.**

Para abordar la elección de una herramienta enfocada en el reconocimiento de emociones a partir de expresiones faciales, se hizo una búsqueda literaria de dichas

herramientas y se realizó una comparativa entre ellas para seleccionar la más apropiada. A continuación, se describen las herramientas encontradas en la literatura

**Haar Cascade:** Esta herramienta ha sido utilizada en muchos estudios de reconocimiento facial en tiempo real, fue propuesta por Paul Viola y Michael Jones en su artículo “Rapid Object Detection using a Boosted Cascade of Simple Features” [84] éste es un algoritmo de aprendizaje automático que se entrena a partir de muchas imágenes positivas y negativas, que busca gracias a una secuencia de funciones en cascada, organizar los píxeles de una imagen en grupos que finalmente servirán para detectar un objeto (en este caso el rostro) [85].

**Single Shot Mutibox Detector (SSD):** Esta herramienta es usada en diversos trabajos de reconocimiento de rostros en tiempo real, se basa en una red convolucional de retroalimentación que como su nombre lo dice realiza las tareas de detección y clasificación de objetos en un solo paso por la red. Este algoritmo se apoya de la técnica de regresión “MultiBox” desarrollada por Szegdy *et al.* Que de forma matemática agiliza el proceso en el manejo de coordenadas ayudando a detectar múltiples objetos presentes de una sola imagen [32].

**Dlib:** Es un sistema de alineación facial principalmente desarrollado en c++ que también cuenta con un amplio soporte en Python, y dos componentes principales: (1) subprogramas básicos de álgebra lineal (BLAS) que se usa principalmente para la implementación de redes bayesianas y (2) Herramientas de aprendizaje automático como los son los algoritmos basados en kernel para la clasificación, agrupamiento, detección de anomalías y regresión de características [86].

**The Multi-task Cascaded Convolutional Neural Networks (MTCNN):** Herramienta usada para la detección y alineación de rostros basado en una red neuronal de convolución profunda, este algoritmo consta de tres fases principales, en donde por cada fase se hará uso de una red neuronal [87]. En la primera fase la red neuronal predice las posibles posiciones de las caras junto con sus cuadros delimitadores obteniendo como resultado una gran cantidad de detecciones de rostros en donde muchos serán falsos. En la segunda fase se usan los resultados de la primera predicción para hacer un refinamiento del resultado y eliminar la gran mayoría de las detecciones falsas, en esta fase también se agregan los cuadros delimitadores. En la tercer y última fase se refinan aún más las predicciones obtenidas. Además, la red genera cinco puntos de referencia facial [88].

**MMOD:** Es un algoritmo de detección de rostros, que haciendo uso de “redes neuronales convolucionales” (CNN), reconoce rostros de manera precisa e independiente de la orientación de este. El método de aprendizaje es sencillo ya que no requiere una gran cantidad de datos, pero esto siempre y cuando se apoye por una “Unidad de procesamiento Gráfico” (GPU), de lo contrario muestra una velocidad muy baja de procesamiento, de aproximadamente 0.2 FPS; debido a

estas restricciones, es posible que no sea adecuado para el reconocimiento facial en tiempo real [89].

A continuación, se presenta un cuadro comparativo (**Tabla 6**) de las herramientas mencionadas anteriormente, con criterios o atributos definidos a partir del paralelo realizado entre la comparación de ellas; se evaluaron los pros y los contras de cada una. Además, con los datos de un pequeño experimento, se identifican aspectos cruciales para escoger la herramienta que mejor se adapta a las necesidades del proyecto.

El experimento consiste en poner de entrada para cada algoritmo, un video corto de aproximadamente un minuto, donde se visualiza un rostro en primer plano interactuando con la cámara desde varias perspectivas, además de varios cambios de iluminación. Se toman registros de los tiempos de ejecución y de la cantidad de Fotogramas Por Segundo (FPS) que cada uno es capaz de procesar (ver **Figura 3**). El experimento se realiza en un computador con procesador de 2,5GHz Intel core i5 de dos núcleos, con un video de resolución 1080p en formato mp4



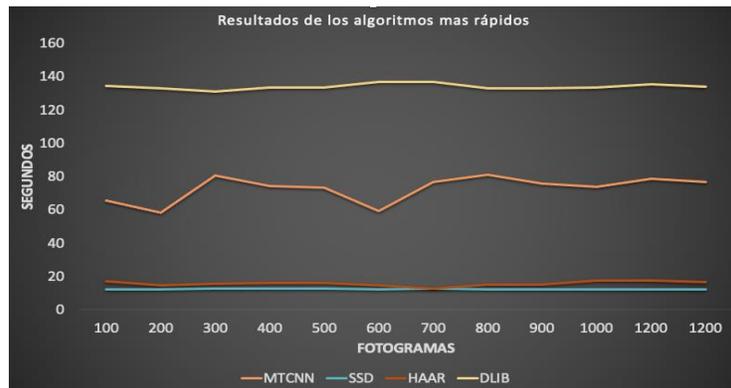
**Figura 3. Vídeo de los cuatro algoritmos corriendo en simultaneo.**

**Tabla 6. Cuadro comparativo de las herramientas para el reconocimiento de expresiones faciales.**

Herramienta	Uso en Tiempo real	Usa (CNN)	Promedio en (FPS)	Consumo de altos Recursos computacionales
<b>SSD</b>	Si	Si	8,05	No
<b>Haar Cascade</b>	Si	No	6,35	No
<b>MTCNN</b>	Si	Si	1,38	Si

<b>Dlib</b>	No	No	0,84	Si
<b>MMOD</b>	No	Si	-	Si

Como se puede apreciar, las herramientas más rápidas son: (1) SSD, ya que procesa en promedio 8,05 FPS haciendo uso de redes neuronales convolucionales (CNN); y (2) Haar Cascade, que procesa en promedio 6,35 FPS, aunque no hace uso de CNN, logra un reconocimiento con una precisión muy cercana a la de SSD solo que usa menos recursos computacionales (Ver Imagen 3). También se resalta que las herramientas MTCNN y Dlib, son soluciones populares en la actualidad por ser robustas y tener un alto grado de precisión, gracias a su alto consumo de recursos. Sin embargo, no es posible considerarlas en el proyecto ya que se requieren computadoras que soporten un alto grado de procesamiento y una GPU, como es el caso del algoritmo MMOD, con el cual no fue posible probar experimentalmente porque requiere un Hardware muy potente para su reconocimiento [89].



**Figura 4. Algoritmos procesados en FPS.**

Finalmente, el criterio decisivo fueron las altas tasas de falsos positivos que se apreciaron en Haar Cascade y dlib, estos algoritmos pueden detectar objetos no faciales como gafas, corbatas o insignias, entre otros. Mientras que las herramientas SSD y MTCNN dan resultados más sólidos respecto a los falsos positivos, por ende, el algoritmo escogido es SSD, por su robustez, precisión y rapidez además de los resultados sólidos.

### 3.2.2. Herramientas para el reconocimiento de emociones a partir del habla.

Con el objetivo de encontrar el complemento adecuado para el reconocimiento de expresiones faciales, se realiza una búsqueda de herramientas capaces de abstraer las diferentes características de la voz y que a su vez se logre integrar de manera adecuada con la información obtenida a partir del reconocimiento de expresiones faciales.

Partiendo de lo anterior y dado que la herramienta de reconocimiento de expresiones faciales utiliza algoritmos de inteligencia artificial, se destacan aquellas

herramientas que utilizan este mismo tipo de algoritmos, ya que tienen como ventaja principal la detección de emociones a partir de datos de entrada complejos y sin la necesidad de ajustarlas de forma manual (lo cual es algo habitual en técnicas tradicionales [90]).

Entre las herramientas encontradas se destaca la realizada por Xuanji He [91], quien implementó una red neuronal convolucional (en adelante Speech emotion recognition) por medio de la utilización de tres estrategias de agrupación (max-pooling, mean-pooling and attention-based weighted-pooling) esto con el fin de realizar un etiquetado de las emociones, resalta también, que hizo uso de la base de datos IEMOCAP para realizar el entrenamiento de la red neuronal, y que la ejecución del algoritmo depende de requisitos software como: Ubuntu 16.04, Tensorflow, Python, Onda, cPickle, Numpy y Sklearn; además de requisitos hardware como: GPU GTX-1080.

Otra herramienta es la realizada por Derek Hung y Mitesh Puthran [92] quienes proponen una red neuronal convolucional (Speech emotion analyzer) la cual es capaz de distinguir entre voces masculinas y femeninas con una precisión del 100%, además de realizar un reconocimiento de la emoción, en un menor grado de precisión. Los autores resaltan que para el entrenamiento de la red usaron los conjuntos de datos RAVDEZZ y SAVEE<sup>4</sup>, los cuales contienen archivos de audio tanto de mujeres como de hombres, resaltan además la utilización de la librería Librosa la cual es un paquete usado para el análisis de música y audio.

Por último, se resalta la herramienta OpenVokaturi [93], la cual realiza la medición de 9 señales por medio del cálculo de los medios y las desviaciones estándar de cinco clases de emociones, para ello hace uso de una red neuronal compuesta por tres capas, las cuales se encargan de procesar la información. Los autores resaltan que para su entrenamiento se hizo uso de los conjuntos de datos EMODB y SAVEE, mencionan además que como único requerimiento software es necesario Python.

Cabe mencionar que las herramientas mencionadas anteriormente fueron validadas con el fin de seleccionar la más adecuada para el presente estudio. Para el análisis se tuvieron en cuenta aspectos como: requerimientos software, hardware, cantidad de emociones que reconocen e idioma en el que puede realizar dichos reconocimientos. A continuación, se presenta un cuadro comparativo de las herramientas mencionadas anteriormente (**Tabla 7**).

---

<sup>4</sup> RAVDEZZ Y SAVEE: Son conjuntos de datos con voces de personas tanto femeninas como masculinas utilizadas en el estudio referenciado para realizar el entrenamiento de redes neuronales.

**Tabla 7. Comparación herramientas reconocimiento de la voz.**

Características	Herramientas		
	Speech emotion recognition	Speech emotion analyzer	OpenVokaturi
<b>Requerimientos software</b>	Ubuntu 16.04, Tensorflow, Python, Onda, cPickle, Numpy, Sklearn y os	Python y Librosa	Python
<b>Requerimiento hardware</b>	GPU GTX-1080	Sin restricción	Sin restricción
<b>Cantidad de emociones</b>	Feliz, Enojo, Triste, Neutral, Miedo, Disgusto y Asco	Feliz, Temeroso, Triste y Enojo	Neutral, Feliz, Triste, Enojo y Miedo
<b>Idioma de reconocimiento</b>	Inglés	Inglés	Español

Teniendo en cuenta la información mostrada en la Tabla 4, se selecciona como herramienta la denominada “OpenVokaturi” ya que puede ser utilizada sin tener tantas restricciones, tanto a nivel software como hardware, además de permitir hacer un reconocimiento con audios en idioma español.

### **3.3. Arquitectura de prototipo software.**

Con el propósito de encontrar la mejor manera de reconocer una emoción por medio de expresiones faciales y reconocimiento del habla, y que a su vez se adecúe a las necesidades del presente proyecto, se realizó un estudio detallado de diversos enfoques, herramientas, programas, APIS, datasets y algoritmos encontrados en la literatura sobre el tema. Con lo anterior se logra encontrar los insumos necesarios para la creación de la herramienta software encargada de realizar el reconocimiento de emociones enfocado en el reconocimiento de expresiones faciales y del habla. A continuación, se describirán los factores que se tuvieron en cuenta para la selección de dichas herramientas:

El primer factor se enfoca en la búsqueda de un lenguaje de programación en el cual se puedan sentar las bases del proyecto, y con la condición de que su curva de aprendizaje fuera aceptable y adecuado para el tiempo estipulado. Además, se buscó que fuera en lo posible un lenguaje multiplataforma para evitar inconvenientes de compatibilidad entre los sistemas operativos más comunes (Mac, Linux o Windows, IOS y Android y que puedan ser usados en dispositivos móviles, y aplicaciones de escritorio sin mayor dificultad. Teniendo en cuenta lo anterior, se seleccionó el lenguaje de programación Python, ya que es un lenguaje sencillo,

robusto, pertinente con un amplio soporte en sus herramientas y un gran número de bibliotecas estadísticas y numéricas, Además, es muy usado por la comunidad científica aportando soluciones prácticas a problemas actuales enfocados en machine learning, redes neuronales, visión por computadora, entre otras, adicionalmente a esto, su curva de aprendizaje es corta y se cuenta con experiencia previa en el manejo de este lenguaje.

El segundo factor, se encarga de explorar herramientas de visión por computadora o visión artificial, ya que es necesario capturar, reconocer y procesar los rasgos faciales por medio de una entrada visual (imagen o video). La elección de la herramienta que se encarga de realizar lo mencionado anteriormente, además de cumplir el primer factor, se denomina OpenCv: la cual es capaz de identificar puntos o rasgos en el rostro que permiten realizar un posterior análisis, además cuenta con soporte en Windows, Linux, Mac, iOS y Android. Otro criterio que se tuvo en cuenta fue que esta herramienta cuenta con una amplia variedad de guías, libros y tutoriales que facilitan su implementación.

El tercer factor consiste en elegir una herramienta que cumpla con los anteriores dos factores y que, además, se encargue de interpretar los algoritmos de redes convolucionales para abordar el reconocimiento de emociones. La elección de esta herramienta se hizo mediante la búsqueda de plataformas de código abierto, donde se encontró que Google, a finales del 2015 con el fin de impulsar la innovación y expandir el conocimiento a multitud de empresas, universidades, ingenieros y científicos, liberó una biblioteca de software denominada TensorFlow, la cual cuenta con una amplia gama de funciones para el lenguaje Python. Además, está compuesta por un sistema integral y flexible de herramientas, bibliotecas, algoritmos y tutoriales que ofrecen un marco de trabajo muy potente, simplificando la complejidad de los procesos que supone implementar los modelos de aprendizaje profundo (Deep learning)

El cuarto y último factor, se enfoca en definir el tipo de aplicación a desarrollar, la cual estará encargada de integrar las herramientas concernientes a los tres factores antes mencionados. En este punto, se opta por escoger una aplicación web, la cual mitiga problemas de versionamiento y compatibilidad que puedan presentarse entre sistemas operativos y es capaz de soportar los factores ya mencionados, además, puede ser ejecutada desde un navegador, facilitando el uso para usuarios futuros. Teniendo en cuenta la anterior elección, se decide trabajar con PHP y JavaScript haciendo uso de los frameworks Angular y Symfony que son herramientas que cuentan con un gran poder operativo y soporte actualizado periódicamente, lo cual hace que estas sean herramientas idóneas para llevar a cabo el prototipo software.

## CAPÍTULO 4. DESARROLLO DEL PROTOTIPO SOFTWARE PARA LA EVALUACIÓN DE LA UX EN TELÉFONOS INTELIGENTES

### 4.1. Metodología de desarrollo

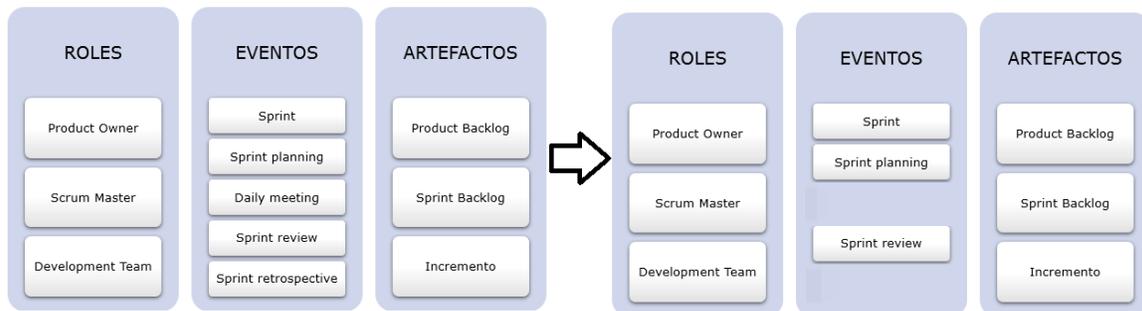
Partiendo de la necesidad del desarrollo de un prototipo software, se requiere seleccionar una metodología de desarrollo con la cual se pueda alcanzar el objetivo propuesto, para hacerlo se define que la metodología cumpla con los siguientes tres puntos: (i) una entrega continua de avances, (ii) una generación mínima de documentación (solo la necesaria) y (iii) por último que se pueda realizar en un lapso corto de tiempo, teniendo en cuenta que el desarrollo se debe hacer en máximo un año. Partiendo de lo anterior, se elige una metodología de desarrollo ágil, donde nos inclinaremos por el marco de desarrollo conocido como Scrum, esta elección se hace porque es una de las metodologías más utilizadas y se encuentra una gran cantidad de documentación de cómo ser aplicada, además el marco define diferentes pasos donde se busca una entrega continua de producto y al definir pasos claro se puede realizar una estimación de la posible duración del proyecto, con esto se logra cubrir los puntos número uno y tres, además este marco provee la generación de artefactos necesarios con los cuales se cubre el punto número dos.

Aunque el marco de desarrollo Scrum define una serie de roles, eventos y artefactos como se muestra en la **Figura 5**, se extrajeron los necesarios con el objetivo de cumplir con el desarrollo del prototipo, para hacerlo se hizo una asignación de roles, donde el director de trabajo de grado será el Product Owner quien es la persona que conoce del tema y además tiene un claro panorama de que se debería desarrollar con el prototipo; se cuenta también con el equipo de desarrollo, el cual está integrado por las dos personas que hacen parte de la investigación; y por último, al ser un equipo tan reducido, se define que el Scrum Master será uno de los integrantes del equipo de desarrollo. Además de la definición de roles, se establecen diferentes eventos que puedan ayudar a realizar una entrega continua de los avances. Entre los eventos se encuentran: (i) la planeación del Sprint<sup>5</sup>, donde se define las historias de usuario (HU) que serán implementadas durante las dos semanas que dura cada Sprint, esto se hace con el objetivo de generar un entregable funcional al finalizar cada Sprint; (ii) la revisión del Sprint, en la cual se realizará la entrega del avance al Product Owner con el objetivo de que sea validado y aprobado el entregable. Por último, teniendo en cuenta la necesidad de la generación de documentación, se utilizan los diferentes artefactos que plantea Scrum, entre ellos se encuentra el Product Backlog, el cual contiene una lista de las funcionalidades necesarias para el correcto funcionamiento del prototipo. Cada funcionalidad se encuentra detallada por las diferentes HU realizadas por el equipo de desarrollo, además se define la utilización del Sprint Backlog con el cual se hará el seguimiento de las funcionalidades a realizar durante cada Sprint. Por último, se

---

<sup>5</sup> Sprint: Se refiere al nombre que va a recibir cada uno de los ciclos o iteraciones que se van a tener dentro de un proyecto ejecutado bajo la marco Scrum.

utiliza el artefacto incremento del producto, el cual se hace entrega al finalizar cada Sprint en el evento de revisión del Sprint.



**Figura 5. Extracción de roles, eventos y artefactos utilizados.**

## 4.2. Desarrollo del prototipo software

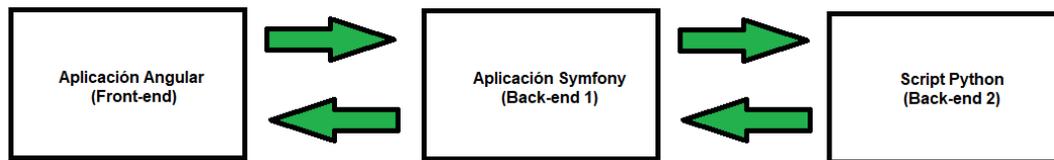
Con el objetivo de realizar un prototipo software que permita su acceso desde cualquier ubicación geográfica, se orienta el desarrollo hacia una aplicación web, para lo cual se utilizará una arquitectura vista-controlador, omitiendo la parte del modelo, dado que la aplicación no almacenará información, si no que funcionará como una aplicación de carga y ejecución de los algoritmos implementados.

Partiendo de lo anterior, se establece la necesidad de contar con un framework front-end para realizar toda la parte gráfica (vista) de la aplicación, donde se tendrá en cuenta: (i) la interacción con el usuario, principalmente en la carga de los datos (vídeo o audio) y representación de la información, (ii) una conexión fácil y funcional con la parte back-end. Para abarcar los puntos mencionados se seleccionó el framework Angular, ya que ofrece una funcionalidad por componentes, lo que proporciona una reutilización de código y la capacidad de crear aplicaciones sin importar el destino (web, web móvil, móvil nativo, etc.), también se debe mencionar que este framework proporciona una mayor velocidad y rendimiento lo que ayudaría en la presentación de la información procesada por el back-end, además se debe resaltar que el framework provee componentes con diferentes funcionalidades, entre los cuales se resalta el componente de carga de información, lo cual contribuye con una de las necesidades del desarrollo [94].

Además de la parte front-end, es necesaria la creación de los controladores utilizados para realizar el procesamiento y análisis de la información, para ello se utilizará el framework PHP Symfony, el cual es conocido por el equipo de desarrollo y proporciona diferentes funcionalidades para la ejecución de scripts. Se plantea que el back-end se divida en dos partes (ver detalle de capas en **Figura 6**):

- La primera parte encargada de recibir las peticiones realizadas por el front-end y realizar el procesamiento de la información para que pueda ser suministrada a la segunda parte del back-end, además se enfocará en retornar la información al front-end para que pueda ser visualizada de forma gráfica.

- La segunda parte estará asociada a la ejecución de los algoritmos de inteligencia artificial, de manera que se encargarán de realizar el análisis de la información procesada por la primera parte del back-end, y tendrán como finalidad determinar las emociones presentes en los datos cargados desde el front-end. Para la implementación de los algoritmos de reconocimiento de emociones, se utilizó como lenguaje de programación Python, el cual ayuda en la generación de scripts que son fáciles de manejar y controlar a nivel de PHP. Además, la segunda parte del back-end estará enfocada en el análisis de datos con el objetivo de determinar las emociones presentes. Para hacerlo la aplicación puede utilizar uno o dos algoritmos, el primero de ellos enfocado en un reconocimiento de emociones a partir de expresiones faciales y el segundo en un reconocimiento a partir de la voz. Para determinar los algoritmos más idóneos se realizó una búsqueda en la literatura y se hizo una clasificación de la información como se muestra en la sección [3.2.1](#) y [3.2.2](#).



**Figura 6. Capas del Prototipo**

A continuación, se dará más detalle del funcionamiento de los algoritmos seleccionados, donde se mostrará su arquitectura y como está construido:

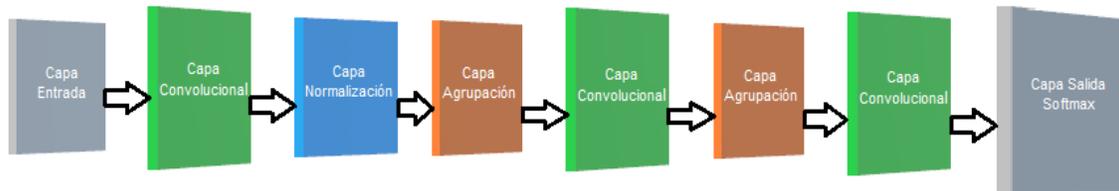
#### **4.2.1. Reconocimiento de expresiones faciales**

Para realizar la detección de emociones por medio del reconocimiento facial se realiza un Script, el cual toma como base el trabajo propuesto por Correa Enrique *et al.* [95] donde presentan una red neuronal construida por las siguientes capas: (i) tres capas convolucionales utilizadas para reducir el tamaño de la imagen por medio de la función de convolución<sup>6</sup>; (ii) dos capas de agrupación o Max Pooling que se encargan de realizar la agrupación de los valores máximos obtenidos de la capa anterior; (iii) una capa de normalización destinada a normalizar<sup>7</sup> los resultados de la capa anterior, aplicando una transformación que mantiene la activación media cerca de 0 y la desviación estándar de activación cerca de 1; (iv) una capa de entrada de tamaño 48x48 destinada a recibir la información de las imágenes; y por último, (v) una capa de salida Softmax utilizada para realizar la asignación de probabilidades tomando como insumo los resultados de la capa anterior. Asimismo, se debe

<sup>6</sup> Función de convolución. Operación matemática que realiza el filtro de imágenes utilizando máscaras, las máscaras representan las conexiones entre neuronas de capas anteriores.

<sup>7</sup> Normalización. Se utiliza para reducir el tiempo de ejecución de entrenamiento de la red neuronal.

mencionar que estas capas se encuentran conectadas de la siguiente forma: inicia con una capa de entrada, en seguida se encuentra una capa convolucional, una capa de normalización y una capa de agrupación, además, conectada a estas se encuentra una segunda capa convolucional, seguida de una segunda capa de agrupación, y, por último, se encuentra una tercera capa convolucional completamente conectada a la capa de salida Softmax, como se muestra en la **Figura 7**.



**Figura 7. Conexión capas red neuronal.**

Cabe resaltar, que la red neuronal funciona por medio del procesamiento de datos visuales donde identifica gestos, disposición y forma de, por ejemplo, las cejas y los labios, con lo cual logra determinar la expresión facial, y, por lo tanto, la emoción de la persona. La red neuronal es capaz de determinar 7 emociones entre las cuales se encuentran: enojo, disgusto, miedo, felicidad, tristeza, sorpresa y neutral, lo que determina que el enfoque utilizado para el reconocimiento de expresiones faciales será un enfoque discreto.

#### **4.2.1.1. Procesamiento de datos para algoritmo de expresión facial**

Partiendo del hecho que la aplicación tendrá como entrada un video, producto de la interacción de una persona con un videojuego, donde se evidencia que este tipo de interacción permite que la persona mantenga su mirada fija a la pantalla durante un rango de tiempo que puede variar entre 8 y 24 minutos [96]. El Script hace uso de la librería OpenCv para el procesamiento del vídeo, donde obtiene imágenes por cada 20 segundos, se define este tiempo tomando como referencia el estudio [97] quienes demuestran que la expresión facial tiene respuestas muy rápidas, definiendo 20 segundos como el rango de tiempo más idóneo, a las imágenes obtenidas se realiza un cambio de color a escala de grises y de ser necesario un cambio de su orientación, luego de realizar el proceso de extracción de imágenes, éstas son suministradas a la red neuronal con el objetivo de realizar el análisis de la información, dicho análisis es almacenado en un archivo para servir como insumo a la hora de realizar la presentación de la información.

#### **4.2.2. Reconocimiento del habla**

Con el objetivo de integrar el reconocimiento del habla en el prototipo, se desarrolla un Script en Python el cual hace uso del clasificador OpenVokaturi [93], este clasificador realiza una medición de 9 señales, las cuales, según el estudio realizado por Kienast y Sendlmeir (2000) se encuentran relacionadas con las clases de emociones. Partiendo de lo anterior, el clasificador calcula los medios y las

desviaciones estándar de las cinco clases de emociones a lo largo de las nueve señales, el resultado es almacenado en diferentes variables que serán utilizadas por el clasificador.

A partir de las 9 señales, realizan un cálculo de las probabilidades de emoción, haciendo uso de una red neuronal con tres niveles de conexión lineal. Los desarrolladores del clasificador mencionan que la red fue entrenada en EmoDB y Savee. Además, se menciona que la entrada consta de nueve nodos que contienen las fortalezas de las nueve señales, convertidas en diferentes valores. Para la transformación realizan una resta de los medios divididos por las desviaciones estándar calculadas (misma transformación que utilizaron para el entrenamiento). Luego de este proceso la información es dirigida a la primera capa de la red neuronal, en esta capa los 100 nodos tienen un sesgo y un peso para cada uno de los nueve nodos de entrada, posteriormente la información es transmitida a la segunda capa de nodos, como en la capa anterior, los 20 nodos contenidos en esta capa tienen un sesgo y un peso para cada uno de los nodos anteriores. Por último, la información es dirigida a la capa de salida, la cual contiene cinco nodos, es decir un nodo por cada clase de emoción, cada uno de estos nodos tienen un sesgo y un peso para cada uno de los nodos de entrada. Además, con el objetivo de convertir la salida en probabilidades, realizan una transformación Softmax; dichas probabilidades se calculan por medio de su valor exponenciado, con estos valores las sopesan con las probabilidades anteriores.

#### **4.2.2.1. Procesamiento de datos para algoritmo del habla.**

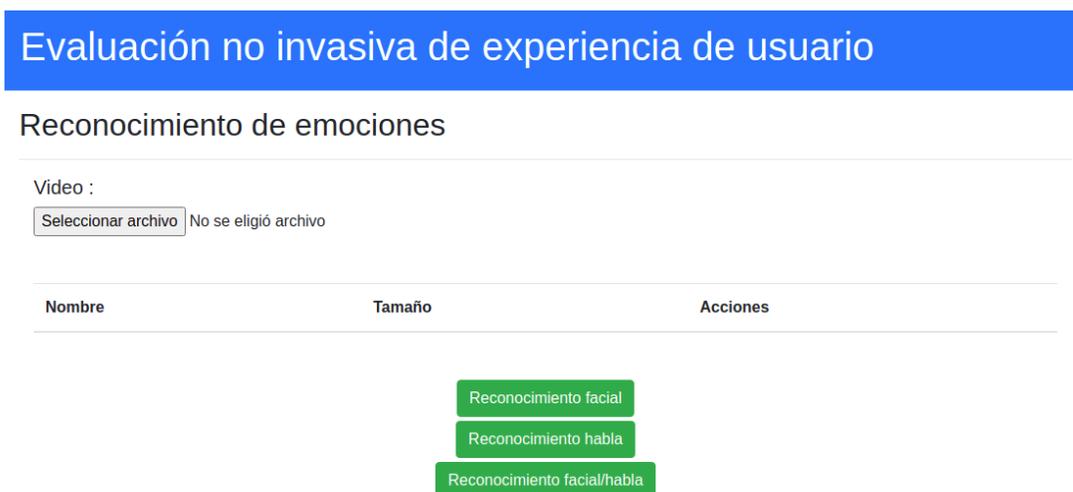
Teniendo en cuenta que se va a realizar el procesamiento de un vídeo, es necesario abstraer el audio con el fin de realizar el análisis de la información, para hacerlo nos apoyamos de la librería MoviePy la cual se encuentra soportada en el lenguaje Python y es utilizada para realizar edición de vídeo. Lo que se busca con la utilización de esta librería es realizar la extracción del audio en unos periodos cortos de tiempo. Teniendo en cuenta el clasificador, se hicieron diferentes pruebas y el mínimo tiempo permitido que se obtuvo fue de 20 segundos el cual se definió como periodo.

Para realizar la extracción del audio en estos periodos de tiempo, se incluyó en el Script de reconocimiento del habla, una sección donde la librería realiza la apertura del vídeo cargado. Posteriormente, se itera por todo el vídeo teniendo en cuenta la constante de tiempo, en este caso 20 segundos, mientras se está realizando la iteración se extrae en una carpeta cada una de las partes del audio, para que posteriormente sean analizadas por el clasificador. Este proceso genera un archivo con cada una de las emociones presentes en el instante de tiempo dado, el cual será el insumo para presentar la información.

Para más detalle del código realizado revisar **Anexo B**.

### 4.3. Entrada de la información

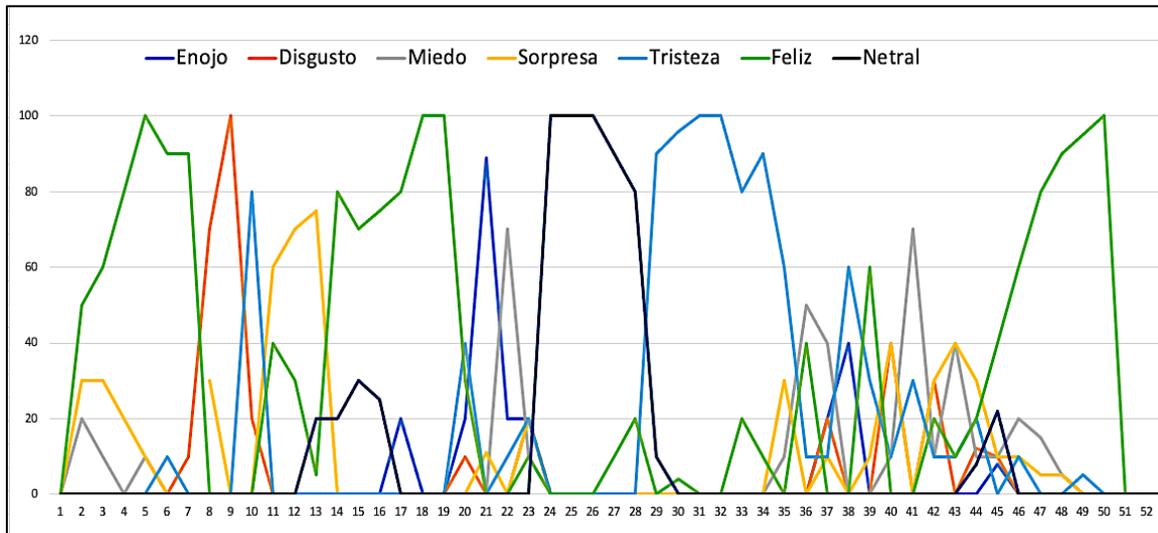
Partiendo del hecho que se busca es realizar el análisis de información de un vídeo generado a partir de la interacción entre el usuario con una aplicación móvil usada por medio del teléfono inteligente, se hace necesario en primera instancia que el vídeo sea cargado a la plataforma, para hacerlo se provee una opción de seleccionar la información desde el equipo y que este sea cargado al servidor de la aplicación web. Una vez esta información ha sido suministrada, se permite realizar un análisis de la información de tres maneras: (i) análisis por medio de reconocimiento facial, el cual consiste en que la aplicación extrae del vídeo una imagen cada segundo, esta imagen es procesada para obtener una representación en blanco y negro, la cual se analiza por medio de la red neuronal y se genera información de la emoción presente en ese instante; (ii) análisis por medio de reconocimiento de la voz, para este tipo de análisis se hace necesario extraer la información del audio que contiene el vídeo, esta extracción se hace cada 5 segundo por medio de un script de Python, cada parte del audio extraído es suministrado al algoritmo de reconocimiento el cual lo analiza y presenta la emoción que más predomina en la parte del audio dado; (iii) por último se tiene un análisis por medio de reconocimiento facial y de la voz, en este tipo de análisis se realiza la extracción de las imágenes cada segundo, además de realizarse una extracción del audio cada 5 segundo, esta información es suministrada al algoritmo correspondiente con el objetivo de realizar el análisis de la información y extraer información de la emoción que más predomina según cada método, una vez se tiene la información de todo el análisis tanto facial como de la voz, se hace un proceso de merge entre los dos datos, donde se toma como prioridad la emoción que más predomine en el instante dado, en este caso se hace cada segundo que es el mínimo expresión de tiempo que se maneja como medida de extracción, en la **Figura 8** se muestra la pantalla de inicio de la aplicación, donde se puede cargar el vídeo además de las tres opciones descritas anteriormente.



**Figura 8. Pantalla inicio aplicación.**

#### 4.4. Presentación de la información

Con el objetivo de presentar la información de manera agradable al evaluador de UX, se genera una gráfica donde se muestra la emoción que más predomina durante cada veinte segundos. Esta gráfica se genera utilizando la información producida por el Back-end, después de realizar el procesamiento y análisis de los datos suministrados por el front-end. En la **Figura 9** puede observar cómo se mostrará la información:



**Figura 9. Grafica de emociones generada por el Prototipo.**

Para más información del funcionamiento del prototipo, revisar manual de usuario en el **Anexo C**.

#### 4.5. Preparación de la información de entrenamiento.

Teniendo en cuenta que la red neuronal requiere de un proceso de entrenamiento para realizar un adecuado reconocimiento de las emociones, en esta sección se da a conocer los aspectos más importantes que se tuvieron en cuenta para la búsqueda y selección de una base de datos adecuada para dicho entrenamiento.

##### 4.5.1. Bases de datos.

En el transcurso de la investigación se logró abstraer y clasificar información relacionada a bases de datos de expresiones faciales, utilizadas en los artículos de investigación revisados, algunas de estas de uso libre y otras de acceso restringido. Dicha información, se encuentra organizada en la **Tabla 8**, la cual cuenta con: un identificador de la base de datos, nombre de la base de datos, información incluida (video, fotos, cantidad de sujetos), expresiones faciales estudiadas, tipo de captura de los datos (de forma actuada o espontanea) y finalmente la disponibilidad de las bases de datos.

**Tabla 8. Bases de datos de expresiones faciales existentes.**

Identificador	Base de datos	Información incluida	Expresiones	Tipo	Disponibilidad
<b>BD1</b>	Cohn-Kanade Database[98]	486 secuencias de 97 sujetos	alegría, sorpresa, ira, miedo, asco y tristeza	Actuada	Gratuito
<b>BD2</b>	MMI Database[99]	2900 videos e imágenes de 75 sujetos	feliz, triste, enojado y neutral	Actuada / Espontánea	Gratuito
<b>BD3</b>	FER-2013[100]	Más o menos 32000 imágenes	enojo, disgusto, miedo, felicidad, tristeza, sorpresa y neutral	Espontánea	Gratuito
<b>BD4</b>	CMU Facial Expression Database (Cohn-Kanade) [101]	200 secuencias de 97 sujetos	Alegría, sorpresa, ira, miedo, asco y tristeza	Actuada	Pago
<b>BD5</b>	BU-3DFE Database[102]	100 sujetos	alegría, sorpresa, ira, miedo, asco y tristeza	Actuada	Gratuito
<b>BD6</b>	Extended Cohn-Kanade Dataset (CK+) [103]	593 secuencias	Neutral, tristeza, sorpresa, felicidad, miedo, ira y asco	Actuada / (risa)Espontánea	Gratuito
<b>BD7</b>	Japanese Female Facial Expressions (JAFFE) [104]	213 imágenes por 10 modelos femeninas	Neutral, tristeza, sorpresa, felicidad, miedo, ira y asco	Actuada	Gratuito
<b>BD8</b>	FERG (Facial Expression	55767 imágenes 6 sujetos	Enojado, asco, miedo,	Actuada	Gratuito

	Research Group Database)-DB [105]	hechos en computadora	alegría, neutral, triste y sorpresa.		
<b>BD9</b>	Radboud Faces Database (RaFD) [106]	8040 imágenes de 67 sujetos	Neutral, tristeza, desprecio, sorpresa, felicidad, miedo, ira y disgusto	Actuada	Pago
<b>BD10</b>	Beihang University facial expression database[107]	32 sujetos de 7 secuencias cada sujeto	Desesperación, pena, preocupación, sorpresa, ráfaga, horror, asco, furia, miedo, duda, impaciencia, odio, desprecio, menosprecio, burla, Sonrisa, ruego, carcajada	Actuada	Gratuito
<b>BD11</b>	NVIE (Natural Visible and Infrared facial Expression)[108]	100 sujetos infrarojo	Expresiones básicas	Actuada /Espontánea	Gratuito
<b>BD12</b>	The Yale Face Database [109]	15 sujetos con 165 imágenes	Triste, somnoliento, sorprendido	Actuada	Pago
<b>BD13</b>	The Psychological Image Collection at Stirling [110]	352 imágenes de 99 sujetos	Sonrisa, sorpresa, disgusto	Actuada	Pago

#### 4.5.2. Selección base de datos.

En esta sección se presentan los diferentes aspectos que se tuvieron en cuenta para seleccionar la base de datos que será utilizada durante el desarrollo del experimento, se tendrá como punto de partida la información descrita en la Tabla 4.

El primer aspecto relevante para la selección de la base de datos está relacionado con su disponibilidad, donde se encuentran bases de datos gratuitas o pagas, teniendo en cuenta que el desarrollo de esta investigación no cuenta con una financiación, se optó por bases de datos gratuitas, lo cual deja un total de 9 posibles bases de datos para la selección, que corresponden a: BD1, BD2, BD3, BD5, BD6, BD7, BD8, BD10 y BD11. Como segundo aspecto relevante se define la cantidad de expresiones, y como tercer aspecto, el tipo (*actuado o espontáneo*). Teniendo en cuenta estos criterios, la base de datos con mayor cantidad de expresiones de los 9 restantes, es la B10 (Beihang University facial expression database); sin embargo, las imágenes de esta base de datos son de tipo *actuada*, y teniendo en cuenta que la investigación se llevará en un ambiente no controlado, este tipo de expresiones no se adaptan muy bien al contexto de la investigación. Las bases de datos identificadas como BD3, BD6, BD7 y BD8, se descartan por ser también de tipo *actuada*, además de tener pocas expresiones. Las 2 restantes identificadas con BD3 y BD6, cuentan con un total de 7 expresiones, siendo las bases de datos más prometedoras para realizar el experimento (de acuerdo con los criterios establecidos). Como criterio final para elegir una de estas dos bases de datos, se toma en cuenta nuevamente la cantidad de imágenes que poseen cada una, lo que deja como base de datos seleccionada la identificada como BD3 “FER-2013”, ya que tiene un aproximado de 32000 imágenes, comparado con las 593 que tiene la base de datos BD6.

#### 4.6. Entrenamiento del prototipo.

Teniendo en cuenta que en la sección 4.2 se definió como base de datos FER-2013, se realiza un primer entrenamiento de la red neuronal, obteniendo un 57% de precisión, lo que lleva a realizar una investigación de cómo lograr que dicha precisión mejore, se encontraron diferentes estudios de los cuales se resaltan: el realizado por Correa Enrique *et al.* [95] quien menciona que una forma de mejorar la precisión es la utilización de más datos (imágenes) para realizar el entrenamiento. Y el estudio de Sefik Ilkin Serengil quien resalta que es posible mejorar la precisión del reconocimiento aplicando Auto-keras<sup>8</sup> al conjunto de datos; sin embargo, esto ocasiona que la estructura de la red se vuelve más compleja, originando así un costo computacional mucho mayor.

Partiendo de los estudios revisados, se toma como referencia el estudio de Correa Enrique *et al.* [95], quien propone incluir imágenes de entrenamiento de la base de datos RaFD ya que proporciona imágenes de alta resolución y a su vez son

---

<sup>8</sup> Auto-keras. Librería de código abierto que ofrece herramientas de Deep Learning

imágenes limpias (es decir, “sin ningún tipo de ruido”). Con lo anterior, se logra un aumento sustancial de la precisión a un 66%.

## **CAPÍTULO 5: ESTUDIO DE CASO: EVALUACIÓN DE LA UX CON TELÉFONOS INTELIGENTES UTILIZANDO EL PROTOTIPO SOFTWARE PROPUESTO**

En este capítulo se presenta el desarrollo del estudio de caso, haciendo uso del prototipo software diseñado para el presente estudio. En este caso, el propósito es realizar la obtención de las emociones presentes en niños durante la interacción con una aplicación móvil, para hacerlo la ejecución se dividirá en dos soportes, el primero llamado “CAPTURA” y el segundo “PROCESAMIENTO”, todo esto se llevará acabo teniendo en cuenta los aspectos identificados en la contextualización y caracterización realizadas en este proyecto.

Los resultados obtenidos se utilizan como insumo para el desarrollo de un estudio comparativo, que permita determinar si el prototipo propuesto sería el apropiado para realizar una evaluación de la UX en teléfonos inteligentes de forma no invasiva. Esto se hace con el fin de revelar los aportes significativos que brinda la propuesta de este trabajo.

### **Elección del Mecanismo Tradicional**

Debido a que se pretende usar como fuente un de video desde el cual se pueda realizar un reconocimiento de emociones, es necesario definir un mecanismo con el cual se pueda reconocer diferentes características de los usuarios y puedan ser analizadas y posteriormente representadas como emociones, partiendo de esto se pudo determinar que el método más apropiado es utilizar las 10 heurísticas emocionales, ya que permiten utilizar pautas sencillas con las cuales se puede realizar de forma eficiente, fácil y económica la valoración de las reacciones expresivas durante un proceso de evaluación de un producto, teniendo esto en cuenta, realizar esta identificación de emociones a partir de las pautas resulta más eficiente al contar con una fuente de datos como lo son los vídeos de las interacciones. Estas razones son suficientes para considerar que este método es el apropiado para realizar la comparación con la herramienta propuesta.

### **5.1. Metodología**

Las investigaciones que involucran a los humanos y a sus interacciones con la tecnología hacen que los paradigmas de investigación analítica no sea suficiente para abordar cuestiones complejas de la vida real, por esto, el estudio de las situaciones contemporáneas en su contexto natural hace que un estudio de caso sea una metodología de investigación adecuada para afrontar esta investigación, además de sus crecientes contribuciones en el aumento de conocimiento como lo han mostrado los análisis de enfoques cualitativos [111], las investigaciones empíricas en ingeniería de software [112], y guías enfocadas exclusivamente en casos de estudio como las directrices de JM Verne *et al.* [109] para casos de estudio

en ingeniería del software. Desde este punto de vista es necesario definir una estructura para el estudio de caso, está será soportada en el trabajo de P. Runeson y M. Höst [113], la cual se observará a continuación.

### 5.1.1. Diseño del estudio de caso

El estudio de caso diseñado para evaluar la UX con el prototipo software, consta de dos partes que se ejecutaran de forma independiente: la primera parte que llamaremos “*CAPTURA*”, se basa en obtener de forma no invasiva, información audiovisual de la interacción de una persona (niño), con una aplicación móvil (vídeo juego). Se procura que el niño no se percate que está siendo monitoreado; todo esto haciendo uso de los sensores del mismo dispositivo: cámara delantera y micrófono; para esta tarea se cuenta con la aplicación “Grabador one touch”; Como segunda parte está el que llamaremos “*PROCESAMIENTO*”, donde se soporta en una aplicación propia, encargada de recibir y procesar la información (audio y video) obtenida por el primer soporte “*CAPTURA*”, esta información será comparada con un método de captura de emociones de movimientos corporales (llamado de aquí en adelante CBME por sus siglas en inglés Capturing Body Movement Emotions ) [114] [115].

#### Objetivo del estudio de caso

Determinar el nivel de precisión del prototipo software para identificar adecuadamente las emociones de manera no invasiva en niños entre 8 y 14 años, haciendo uso de un video juego en un teléfono inteligente.

#### Pregunta de investigación

¿Qué tan preciso es el prototipo software para identificar las emociones de manera no invasiva en niños entre 8 y 14 años que hagan uso de un video juego móvil en un teléfono inteligentes?

#### Métrica

Para dar respuesta a la pregunta de investigación de este estudio de caso se hace uso de una métrica definida para medir la precisión del prototipo, teniendo en cuenta la percepción del usuario que realizó la interacción. Esta métrica consiste en realizar una validación de las emociones generadas por el prototipo con el usuario, en ella se valida si la emoción determinada es la que el usuario experimento, la métrica consiste en la siguiente:

$$x = \sum \frac{P_i}{N}$$

Pi: Cantidad de emociones confirmadas

N: Número de emociones por validar

### **5.1.2. Recolección de datos**

Partiendo que se pretende recoger la percepción de los usuarios frente a una aplicación móvil ejecutada sobre un teléfono inteligente, se determina que el método más apropiado para realizar esta captura de datos y que a su vez se haga de manera no invasiva, es a través de un vídeo con el cual se recolecte información tanto del rostro como de la voz. Para realizar dicha recolección de información, se hace uso de algunas aplicaciones ya existentes dentro de las diferentes plataformas de distribución de aplicaciones, las cuales se describen en la sección 3.1. Con la utilización de estas aplicaciones, se busca realizar la captura de video en segundo plano, ayudando a que la persona no perciba que está siendo observada y que su interacción no sea alterada por ningún aspecto externo. Posteriormente la información será analizada por el prototipo propuesto, el cual indicará cada una de las emociones presentes en el usuario durante la interacción con la aplicación móvil. Además, con el objetivo de realizar el análisis por medio de CBME se realiza una captura de video donde se registre todos sus movimientos corporales y así lograr determinar que emoción está sintiendo. Por último, con el objetivo de validar con el usuario como se sintió se realizará una encuesta para obtener información relacionada con su percepción frente al experimento realizado.

### **5.1.3. Análisis de los datos**

El tratamiento de los datos recolectados durante el proceso de estudio de caso se hará con un enfoque de análisis cuantitativo.

## **5.2. Ejecución del estudio de caso**

En primera medida, se establece la o las aplicaciones de prueba adecuadas para poder realizar el estudio de caso, el cual se inclina por buscar aplicaciones que puedan generar estímulos emocionales fuertes, espontáneos, notables, enérgicos o impetuosos e intempestivos en los usuarios, ya que el instrumento es sensible a la presencia de este tipo de estímulos, permitiendo así que el nutrirlo sea más sencillo, además de lo anterior es pertinente que se haga un uso continuo de la aplicación (de al menos unos 15 minutos de uso prolongado y mínimo 3 veces por semana). Teniendo en cuenta que los juegos y aplicaciones de entretenimiento cuentan con las características antes mencionadas, y que el estudio [116] resalta que los usuarios lúdicos compuestos por jóvenes y niños entre 8 y 17 años cuentan con un 30% de preferencia por este tipo de aplicaciones, se seleccionan los juegos de video como la aplicación adecuada para realizar el presente estudio de caso.

Ahora la tarea es decidir qué juego de video es el adecuado, para esto, se va a tener en cuenta al usuario, que para este caso y en paralelo con la elección de los juegos de video, se seleccionaron niños por su considerable actividad e interacción

con este tipo de aplicaciones lúdicas [117], en lo posible, niños de entre 8 y 14 años de edad. Se descartan a niños mayores de 14 años, porque según [117] éstos ya se encuentran en etapa de evolución del uso lúdico del dispositivo móvil al uso comunicativo (potenciado por las redes sociales). En tal caso se escogieron a 44 niños de entre 8 y 14 años. De esos 44 niños se seleccionaron 28 niños, a los cuales, con base en las respuestas a una encuesta contestaron afirmativamente al uso de juegos de video en dispositivos móviles, además de jugar al menos 3 veces por semana y al menos 15 minutos de interacción continua (la mayoría de las veces). Para más detalle de la encuesta y respuesta obtenida revisar **Anexo D**.

Para la elección del video juego se hizo una encuesta abierta (para más detalle revisar **Anexo E**), para identificar los 4 video juegos con los que más tienen interacción y que cuente con las características antes mencionadas (interacción emocional fuerte, notable, enérgica o impetuosa) y de estos 4 juegos se les pidió seleccionar el que más les gusta. En la **Tabla 9** se muestra los cuatro juegos identificados que cantidad de niños tienen preferencia con ellos.

**Tabla 9. Relación entre los juegos y la cantidad de niños que lo prefieren.**

Cantidad (niños)	Edad (años)	Juegos de video			
		Free fire	Call of duty Mobile	Among US	Minecraft
7	10	1	2	3	1
9	11	3	3	1	2
9	12	3	5	1	0
6	13	2	4	0	0
<b>TOTAL:</b>		9	14	5	3

Con base en los resultados obtenidos por la encuesta, se selecciona el juego de Call of duty y a 14 participantes que eran quienes lo jugaban. De los 14 participantes, tres de ellos no pudieron participar en el experimento. Por lo tanto el estudio se completó con 11 participantes, que finalmente fueron los sujetos que participaron en el un estudio para evaluar su experiencia de usuario mediante la captura de sus emociones con los dispositivos móviles. Para permitir su participación se solicitó por medio de una carta el consentimiento a los padres, para más detalle del consentimiento revisar el **Anexo F**.

En el transcurso del experimento surgieron una serie de limitaciones circunstanciales, pero sobre todo sobresalió una situación de orden social presentada durante todo el desarrollo del mismo, a causa del virus covid-19, por lo cual no fue posible realizar la prueba de forma presencial, optando por hacerlo de forma remota, eso sí, concertando un mayor esfuerzo por parte de los padres de los participantes, en donde estos consintieron la situación, y entendieron que para el

buen desarrollo del experimento era necesario este aumento en sus responsabilidades a lo largo de este.

Se tuvieron en cuenta cuatro (4) momentos que estructuran la “CAPTURA”:

1) Verificación del funcionamiento de la aplicación de prueba (Call of duty), 2) Instalación de las aplicaciones de monitoreo, 3) Configuración y prueba de las aplicaciones de monitoreo y 4) prueba longitudinal (por 5 días).

A continuación, se presenta la descripción de cada uno de los momentos antes mencionados.

1) Verificación del funcionamiento de la aplicación de prueba (Call of duty): Se verifica que, al ejecutar la aplicación, ésta abra correctamente y esté corriendo en su última versión.

Dado que no se pudo realizar la verificación en conjunto con todos los participantes, fue necesario transportarse a las residencias de cada uno de ellos, contando con los protocolos de salud necesarios para esta visita.

En general la mayoría de los participantes contaban con la última versión de la aplicación en sus dispositivos, salvo por 2, que se tuvieron que actualizar a la última versión con su consentimiento.

2) Instalación de las aplicaciones de monitoreo: se necesitan dos aplicaciones. La primera encargada de encender y capturar lo que transmite la cámara delantera del dispositivo, además del audio del exterior (Grabador one touch). Y la segunda encargada de la programación de la ejecución de la primera aplicación, es decir, programa en qué momento se enciende la primera aplicación (MacroDroid). En nuestro caso se programa para que active la primera aplicación, justo cuando se inicia la aplicación (Call of duty) y se cierre justo cuando (call of duty) se cierre. De esta forma se podrá obtener la información de la cámara delantera y del audio solo cuando la aplicación (call of duty) esté en ejecución.

Algunos teléfonos contaban con un almacenamiento limitado y para poder realizar la instalación de estas, fue necesario eliminar información del dispositivo. Otros participantes habían cambiado su dispositivo móvil a uno diferente de Android por lo que no fue posible hacer la instalación de las aplicaciones.

En general se logró instalar las aplicaciones de monitoreo a los participantes, salvo a 2 de ellos que contaban con dispositivos IOS, por lo que se tuvo que excluir a esos participantes del experimento.

3) Configuración y prueba de las aplicaciones de monitoreo: En lo que respecta a la aplicación Grabador one touch, se configura de tal manera que encienda los sensores de la cámara delantera y el micrófono, y que almacene todo lo que se captura en el dispositivo (la resolución con la que se configura la grabación del video capturado es de 720 píxeles) este archivo audiovisual se guarda en el almacenamiento interno del dispositivo, pero la aplicación proporciona la opción de agregar un correo Gmail para que sea enviado a un servidor Google drive y sea subido apenas cuente con una conexión wifi, además, la aplicación elimina el archivo del dispositivo una vez se haya subido al servidor google Drive.

En lo que respecta a la aplicación MacroDroid es más sencillo ya que al ser una aplicación para la configuración y automatización de tareas, permitió fácilmente que

se configurara de tal manera que encendiera la aplicación Grabador one touch cada vez que se ejecute la aplicación call of duty y se cierre cada vez que (call of duty) se cierre.

Se encontraron algunos casos donde la aplicación Grabador one touch funcionó de forma inadecuada y se cerraba inesperadamente después de unos segundos en ejecución o no se ejecutaba.

Se modificó la configuración de Grabador one touch, disminuyendo la resolución del video a 480 pixeles resolviendo así un posible problema de almacenamiento con el que contaban algunos dispositivos.

4) Prueba longitudinal (por 5 días): Para esta parte fue necesario establecer 3 jornadas de prueba de 5 días cada una, escogiendo las primeras dos jornadas para resolver los posibles fallos, desaciertos o fracasos que puedan acontecer en el proceso del experimento y tomando la tercera jornada para realizar la recolección de datos con la menor cantidad de incidentes posibles.

Para cada jornada fue necesario hacer un seguimiento de los participantes por medio de los padres, donde se pactó con ellos, les permitan jugar al menos una hora diaria, durante los 5 días de prueba.

Para la primera jornada se esperaba que los archivos audiovisuales empezaran a llegar al servidor de google Drive mínimo 3 archivos por participante de 9 participantes, pero al cabo de la primera jornada solo se recibieron 5 archivos de todos los esperados, además, de esos 5, 4 habían capturado solo unos segundos de la interacción, haciendo a la mayoría de estos datos defectuosos e inadecuados para continuar con el proceso, por ello se tomó la decisión de acudir al lugar de vivienda de los participantes para verificar los dispositivos y saber que sucedió, encontrando que en la mayoría de los casos el problema radicaba en que algunos dispositivos seguían teniendo problemas por el limitado almacenamiento. Con base en esto se pactó con los padres el tratar de mantener un almacenamiento mínimo disponible de 5 a 10 GB mientras duraba el proceso. (lo cual conllevó a un decrecimiento en la disposición de los participantes y de los padres para con el proyecto).

Una vez establecido el nuevo pacto con los padres y configurados cada uno de los dispositivos, se esperaba que los archivos audiovisuales empezaran a llegar al servidor de Google Drive, pero al cabo de la segunda jornada solo se recibieron 13 archivos de todos los esperados 9 participantes x 3 interacciones = 27 archivos, además de que gran parte de los archivos seguían capturando solo unos segundos de cada interacción. Aunque la solución al problema de almacenamiento incrementó el número de archivos recibidos, aun hacían falta varios de ellos. Dado lo anterior se tomó la decisión de ir nuevamente a verificar cada uno de los dispositivos de los participantes para saber qué había sucedido, encontrando que en la mayoría de los casos los problemas de almacenamiento habían sido resueltos, en otros casos los padres no habían acatado lo pactado, y los participantes no habían ni si quiera ejecutado la aplicación y en otros casos los padres intempestivamente habían decidido que los participantes (sus hijos) y ellos no continuarían más con el proceso. Además de esto, se encontró que la aplicación Grabador one touch continuaba teniendo problemas de cierres inesperados y envío de datos incompletos.

Con base en todo lo anterior se tomó la decisión de poner en pausa el proceso del experimento y se pactó una reunión con todos los padres de los participantes donde se discutieron los posibles problemas circunstanciales de cada uno, encontrando a modo general que las clases y los exámenes de los niños interferían con el buen desarrollo del experimento, por lo que se concertó dejar pasar un tiempo para que se liberaran las cargas de los participantes y así poder dedicarle el tiempo necesario y exclusivo al proceso del experimento. Mientras el proceso se encontraba en pausa, se logró resolver el problema con la aplicación Grabador one touch gracias a una actualización que reparó el mal funcionamiento de esta y resolvió en gran medida los problemas de ejecución

Después de un tiempo prudencial donde los participantes y sus padres contaban con un poco más de tiempo libre y dispuestos para el proceso, se inicia nuevamente el proceso con los 6 participantes restantes, se dispuso a actualizar y reconfigurar la aplicación Grabador one touch de cada uno de los dispositivos. En esta tercera jornada se esperaba que los archivos audiovisuales empezaran a llegar al servidor de Google Drive (en este caso 6 participantes x 3 interacciones = 18 archivos mínimo). Al final de la tercera jornada se recibieron 23 archivos (un poco más de los mínimos esperados).

Con los resultados satisfactorios de la tercera jornada se procede a analizar cada uno de los archivos para verificar si son o no aptos para continuar con el proceso y culminar la parte de “*CAPTURA*”, de estos 23 archivos se seleccionaron los vídeos con mejor imagen para cada uno de los participantes, obteniendo al final un total de 6 vídeos (para más detalle de los vídeos seleccionados ver **Anexo G**) donde se muestra una interacción completa con lo cual se decide culminar la parte de “*CAPTURA*” y continuar con la segunda parte del proceso del experimento “*PROCESAMIENTO*”.

En la segunda parte del estudio “*PROCESAMIENTO*” se enfoca sobre el fenómeno emocional que surge de la interacción del participante (niños entre 8 a 14 años) con el video juego (Call of duty), con ayuda de la información obtenida durante la primera parte se realiza un análisis del video por medio del prototipo desarrollado. En primera instancia el video es cargado a la plataforma, una vez la información ha sido suministrada, se ejecuta el prototipo tomando como opción un reconocimiento facial y de la voz, después de esto el prototipo se encarga de procesar y analizar los datos, con esta información se genera una gráfica con todas las emociones presentes por el usuario durante la interacción con el video juego. Esta grafica sirve de insumo para los evaluadores de UX y tener una apreciación de la parte emocional que tiene el participante con el juego.

Con el objetivo de poder realizar una revisión de los movimientos corporales del usuario se realizar una captura de vídeo donde se permita ver todos sus movimientos, para este proceso se inicia la grabación de vídeo una vez el participante comienza la interacción con el teléfono inteligente, para esta última captura de información se realizó un acompañamiento para validar la captura correcta de la información.

### 5.3. Análisis de resultados

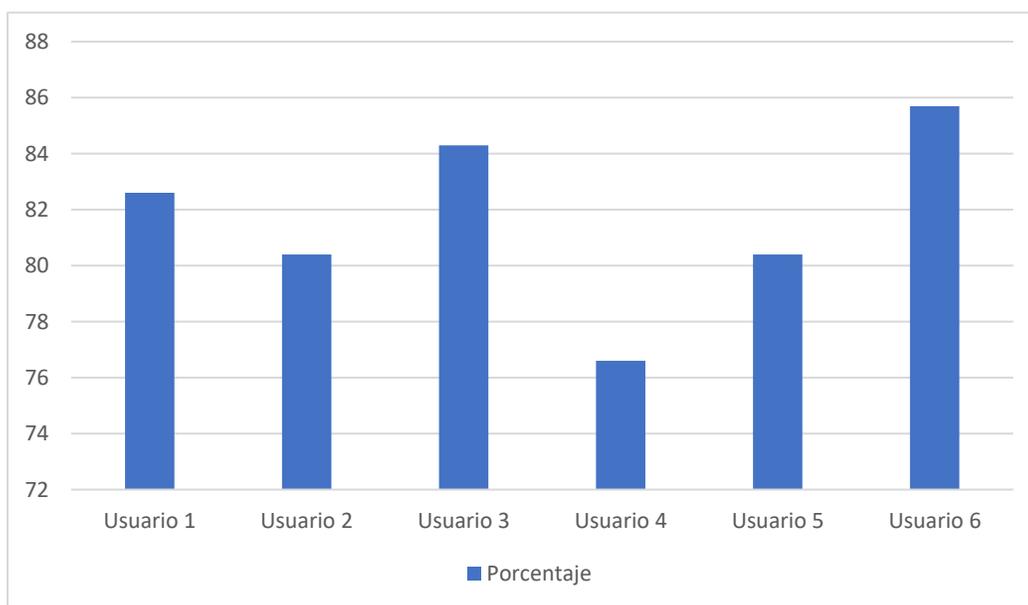
Después de realizar el estudio de caso se obtuvieron los siguientes resultados para cada uno de los métodos de reconocimiento de emociones utilizados como objeto de estudio.

En la **Tabla 10** se muestra la información relacionada con la validación obtenida a partir de las iteraciones realizadas con el prototipo propuesto.

**Tabla 10. Resultado de la validación con prototipo propuesto.**

Participante	Número de emociones por validar	Cantidad de emociones confirmadas
1	46	38
2	51	41
3	51	43
4	47	36
5	46	37
6	42	36

En la anterior **Tabla 10** se aprecian los valores resultantes al aplicar la métrica de reconocimiento a los resultados obtenidos por el prototipo propuesto, cabe resaltar que esta validación se hizo con la apreciación del usuario, la cual nos ayudó a determinar la precisión del prototipo. Los resultados también se pueden apreciar a continuación en la **Figura 10**



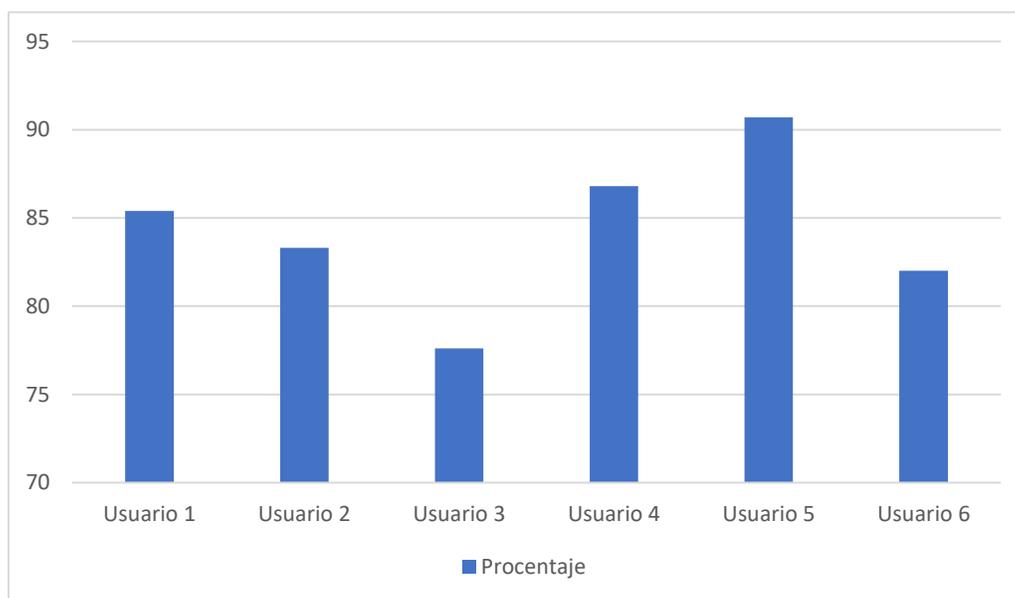
**Figura 10. Grafica de resultados de las validaciones con prototipo propuesto.**

A continuación, en la **Tabla 11** se pueden observar los valores obtenidos a partir de las iteraciones realizadas con el segundo método donde se aplicó las 10 heurísticas emocionales. Para más detalle de los resultados obtenidos del método tradicional ir al **Anexo H**.

**Tabla 11. Resultado de la validación con método tradicional.**

Participante	Número de emociones por validar	Cantidad de emociones confirmadas
1	41	35
2	48	40
3	49	38
4	38	33
5	43	39
6	50	41

Para dar una mejor interpretación de la información se presenta la **Figura 11**, donde se puede observar los resultados obtenidos con el método tradicional aplicando las 10 heurísticas, esta información al igual que con el prototipo propuesto fue generada a partir de la validación con el usuario, lo que nos ayudó a determinar la precisión de aplicar este método en el caso de estudio.



**Figura 11. Grafica de resultados de las validaciones con el método tradicional.**

Cabe resaltar que la información que se provee se limita únicamente a realizar el reconocimiento de emociones de la interacción de un usuario con una aplicación ejecutada en un teléfono inteligente, con lo cual no se hace una identificación de posibles falencias en los sistemas de donde se obtiene la interacción.

## **5.4. Síntesis y discusión**

Después de analizar los resultados de este estudio de caso, uno de los aspectos más importantes a resaltar es que se pudo observar que la precisión del reconocimiento obtenido de la interacción del usuario con el teléfono inteligente a través de ambos métodos fue muy similar, no obstante, esta apreciación requiere de un análisis más allá de los números para identificar las ventajas o aportes que brinda el prototipo propuesto sobre el método que aplica las 10 heurísticas. A pesar de que para este estudio de caso en particular los resultados numéricos de ambos métodos tienen grandes similitudes, no sería correcto concluir que son métodos equivalentes, ya que no se puede garantizar que en estudios futuros los resultados continúen siendo similares.

Una de las grandes ventajas que se puede destacar del prototipo propuesto a la hora de realizar la captura de información, sobre los otros métodos es el nivel de comodidad que ofrece al usuario, esto debido a que se utilizó la revisión realizada sobre métodos de captura no invasiva, lo que nos permitió encontrar herramientas las cuales favorecieran este aspecto. Durante el proceso de validación fue posible identificar como se sintieron los participantes del caso de estudio a la hora de realizar la captura de información, donde se puede resaltar que nuestro instrumento generó menor incomodidad incluso en algunos casos el participante no se percató que estaba siendo monitoreado para el presente caso de estudio.

Otro aspecto que vale la pena mencionar es el tema del reconocimiento y análisis de información, con el objetivo de realizar una captura de emociones generadas de la interacción del usuario con el teléfono inteligente, esta característica del instrumento permite mitigar en gran medida la subjetividad que se pueda presentar durante el proceso de evaluación que dependan de las apreciaciones y/o percepciones de un individuo, las cuales se pueden ver afectadas por los prejuicios y el estado emocional de cada individuo entre otras cosas. La mitigación de la subjetividad antes mencionada se logra a partir del uso de algoritmos de reconocimiento de emociones, los cuales analizan la información y determinan la que más predomine, esto permite que se genere la emoción sin ninguna intervención externa.

## **5.5. Opinión de expertos**

Como apoyo al proceso de validación del prototipo desarrollado en la actual investigación, se realizó una encuesta para expertos en el área de UX, con el objetivo de obtener sus opiniones, apreciaciones y/o sugerencias a cerca del prototipo.

El proceso de validación se realizó de la siguiente manera:

### 1) Detalle perfil de los expertos

Los expertos elegidos para llevar a cabo el proceso de validación cuentan con experiencia en el campo de la UX, siendo más específicos, se han desempeñado como desarrolladores de software en empresas, donde han pertenecido al área de experiencia de usuario.

La información más detallada del perfil de los expertos se presenta a continuación

**Tabla 12. Perfiles de los expertos.**

<b>Nombre</b>	<b>Profesión</b>	<b>Perfil</b>
<b>Miguel Ángel Daza</b>	Ingeniero de Sistemas	Desarrollador UX
<b>Sebastián Rosero García</b>	Ingeniero de Sistemas	Desarrollador UX
<b>Miguel Andrés Morcillo</b>	Ingeniero de Sistemas	Líder Técnico

### 2) Preparación de la información

Con el objetivo de poner en contexto a los encuestados, se incluye información relacionada con las técnicas, métodos y herramientas utilizadas para conformar el prototipo y la forma como se unen para permitir realizar la captura y análisis de la información de una manera no invasiva, esta información permite tener un panorama del prototipo desarrollado.

Además, se plantea una serie de preguntas relacionadas con la información entregada, que buscan conocer la percepción de los expertos frente a las bases que se utilizaron para desarrollar el prototipo.

### 3) Ejecución de la encuesta

Teniendo en cuenta la emergencia sanitaria que vive actualmente el país, se optó por aplicar la encuesta a los expertos a través de Encuestas de Google, en la cual se transcribió cada una de las preguntas realizadas en el **Anexo I** y se generó opciones de respuesta acorde a cada una de ellas. La encuesta fue distribuida por diferentes medios electrónicos, a cada uno de los expertos que participaron en la validación.

### 4) Análisis de la información

En esta etapa se revisa la información obtenida a partir de las encuestas realizadas, donde se identifica una serie de recomendaciones relacionadas con las bases utilizadas para el desarrollo del prototipo, entre algunas recomendaciones se resaltan las siguientes:

- Se debería explorar otra aplicación que permita una interacción más sana e interactiva por parte de los niños de la edad propuesta.
- Que no se usen tools diferentes, que sea un solo aplicativo capaz de detectar las emociones durante un periodo de tiempo con los mecanismos descritos.
- Se debe considerar otros algoritmos y tipo de entrenamiento. Igualmente, del especialista que brinda dicho entrenamiento.
- Hasta donde veo solo se tiene en cuenta rostro y voz no se tiene un contexto de que paso en ese momento, se asume que todas las emociones son por interacción del usuario con una aplicación.

Además de lo anterior se destaca que el instrumento propuesto supera los 4 puntos dentro de la escala de Likert en cada una de las preguntas planteadas respecto a las bases tomadas para desarrollar el instrumento de la investigación, lo que nos permite concluir que tiene un grado de acuerdo alto entre los encuestados.

Para más detalle de la encuesta y la información recolectada revisar **Anexo I**.

## CAPÍTULO 6. CONCLUSIONES Y TRABAJO FUTURO.

A continuación, se detalla información de las conclusiones, limitaciones, lecciones aprendidas, problemas enfrentados y trabajos futuros.

### 6.1. Conclusiones

- A pesar de que incluir la medición no invasiva en la captura de emociones puede favorecer a que un usuario se sienta más cómodo, muchas veces no es tomado en cuenta debido a lo complicado que puede ser implementar técnicas que favorezcan este aspecto. En este estudio se puede observar que existen diferentes alternativas para ayudar en la comodidad del usuario, herramientas que unidas pueden ayudar en la captura de información y que sea lo menos invasivo posible para el usuario.
- Considerando que el prototipo desarrollado tiene un enfoque multimodal en el que interviene dos métodos que soportan el reconocimiento de emociones, es importante resaltar que estas técnicas pueden ser ajustadas teniendo en cuenta el contexto. En el presente, la unión de dos métodos ha permitido cubrir más características fisiológicas, obteniendo un reconocimiento emocional más completo en el cual se mitiga la pérdida de información.
- Aunque el prototipo fue creado a partir de dos mecanismos de reconocimiento de emociones que toman como base la inteligencia artificial, es importante resaltar que su precisión puede mejorar ya que gran parte de esta depende del entrenamiento a la que es sometida la red neuronal y de la cantidad de capas que pueda tener. Se debe resaltar también que, al aumentar el número de capas, puede generar que estos procesos requieran un mayor tiempo de procesamiento.
- Con base en la caracterización que describe la experiencia emocional en teléfonos inteligentes, obtenida a partir de la revisión de documentación, se genera información que puede llegar a ser utilizada para el desarrollo de aplicaciones para teléfonos inteligentes. Además, se debe resaltar que estas características pueden ser utilizadas en etapas tempranas del desarrollo y así incluir en el mismo, aspectos relacionados a la UX de los usuarios.
- La generación de un prototipo de reconocimiento de emociones capaz de detectarlas sin necesidad del acompañamiento de un tercero permite mitigar la subjetividad e identificar las emociones a partir de características que han sido prestablecidas a partir del entrenamiento riguroso de los algoritmos de inteligencia artificial utilizados para dicho fin.
- Aunque la presente investigación está enfocada en la evaluación de la experiencia de usuario en un juego móvil específico, los elementos que conforman el instrumento, podrían ser utilizados como soporte para evaluar

diversos tipos de aplicaciones móviles, ya que el instrumento permite capturar la interacción del usuario de manera no invasiva en una gran diversidad de escenarios, además que las técnicas utilizadas para el reconocimiento facial y de voz, se pueden replicar para cualquier otra aplicación que lleve al usuario a exteriorizar sus emociones a través de expresiones faciales y/o de la voz.

- El tratar con entornos no controlados de forma no invasiva, se obtiene información más fidedigna respecto a la interacción de los seres humanos con un teléfono inteligente. En ese sentido, el presente trabajo logra cubrir este tipo de entornos sin generar perturbaciones en el usuario al momento de capturar información de su interacción usando un teléfono inteligente.
- Teniendo en cuenta el resultado de la encuesta hecha a los expertos, se determina que el prototipo desarrollado cuenta con los requerimientos necesarios para realizar la evaluación de la UX de manera no invasiva en teléfonos inteligentes.

## **6.2. Limitaciones**

- Es posible que surja la necesidad de usar un prototipo como el que se propone en este proyecto, para aplicarlo en teléfonos inteligentes con sistema operativo iOS. Pero cabe aclarar que este prototipo fue concebido a partir de herramientas del sistema operativo Android, lo cual significa que intentar usarla sobre otro tipo de sistema operativo, podría implicar un nuevo proceso de adecuación, pero puede emerger como un punto de partida para prototipos de esta naturaleza.

## **6.3. Lecciones aprendidas y problemas enfrentados**

- Validar los resultados de una investigación a través de un estudio de caso aplicado en un entorno no controlado, requiere de un diseño muy detallado y de un análisis a priori de los posibles inconvenientes que pueden surgir al momento de ponerlo en práctica, esto nos llevó a generar estrategias que permitieron reducir la pérdida de información, entre estas estrategias se encuentra la captura de información de manera longitudinal, donde se permite hacer captura durante varios días para al final seleccionar la información que mejor se adapte a nuestras necesidades.
- Al final del proyecto surgió un inconveniente circunstancial de orden social, que modificó en gran medida la manera con la que se llevó a cabo el caso de estudio. Por causa de la emergencia sanitaria generada por el virus covid-19, no fue posible realizar la prueba presencial, por lo que se decidió hacerlo de forma remota, adecuándose así a la situación, lo cual hizo que se la estimación del tiempo que se tenía para esta tarea fuera más amplia.

## 6.4. Trabajos futuros

Luego de dar cumplimiento con los objetivos propuestos para el desarrollo de la presente investigación, se han analizado algunos puntos que pueden ser tenidos en cuenta para trabajo futuro, entre ellos se incluye:

- Las emociones que fueron consideradas por el prototipo para realizar el reconocimiento es una propuesta de la presente investigación, las cuales se encuentran limitadas teniendo en cuenta los estudios y algoritmos de reconocimiento utilizados. Por esta razón, se ve conveniente poder realizar investigaciones donde se permitan incluir otros aspectos emocionales, ayudando así a tener un mayor rango de emociones que se puedan capturar.
- Con los posteriores avances tecnológicos en cuanto a algoritmos más precisos de reconocimiento de patrones, bases de datos más robustas, sensores de los teléfonos inteligentes mejorados, es posible que con lo mencionado se logre optimizar el instrumento en aspectos como eficiencia, eficacia, rapidez y precisión, lo que prolongará la vida del instrumento.
- En cuanto a las bases de datos, utilizadas para el reconocimiento de emociones, se podría ir nutriendo al instrumento con información más completa donde se tenga en cuenta variaciones culturales y étnicas, partiendo de que los estudios psicológicos interculturales recientes [118], revelan que las expresiones faciales tiene una variación diferente según el origen geográfico del sujeto o la cultura, se podría enfocar esfuerzos en discriminar el reconocimiento facial y ajustar el instrumento para que se pueda adaptar según la cultura de los sujetos.
- Dentro del presente proyecto se realizó un prototipo de reconocimiento de emociones multimodal centrado en dos aspectos el rostro y la voz, sería conveniente agregar otros aspectos con el fin de cubrir la mayor parte de datos emocionales que surgen de una interacción y disminuir así la pérdida de información.

## BIBLIOGRAFÍA

- [1] M. Hassenzahl and N. Tractinsky, "User experience - a research agenda," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 91–97, Mar. 2006.
- [2] S. Kujala, V. Roto, K. Väänänen-Vainio-Mattila, E. Karapanos, and A. Sinnelä, "UX Curve: A method for evaluating long-term user experience," *Interact. Comput.*, vol. 23, no. 5, pp. 473–483, 2011.
- [3] T. Walsh, S. Kujala, H. Petrie, and C. Power, "Axe UX : Exploring Long-Term User Experience with iScale and AttrakDiff," *Acad. '14 Proc. 18th Int. Acad. MindTrek Conf. Media Business, Manag. Content Serv.*, pp. 32–39, 2014.
- [4] MinTIC, "Boletín Trimestral de las TIC," pp. 1–52, 2019.
- [5] *Eye Tracking in User Experience Design*. Elsevier, 2014.
- [6] C. L. B. Maia and E. S. Furtado, "Using psychophysiological measures to estimate dimensions of emotion in hedonic experiences," *Comput. Electr. Eng.*, vol. 71, no. March, pp. 431–439, 2018.
- [7] Y. Méndez, "Marco de Referencia para la Captura Semi-Invasiva y Análisis de las Emociones en el Proceso de Evaluación con Usuarios de Sistemas Interactivos.," *Tesis Dr. Univ. del Cauca. Colomb.*, 2015.
- [8] J. Cordero and J. Aguilar, "Reconocimiento multimodal de emociones en un entorno inteligente basado en crónicas," 2016.
- [9] L. Appleton, "User experience (UX) in libraries: let's get physical (and digital)," *Insights UKSG J.*, vol. 29, no. 3, pp. 224–227, 2016.
- [10] D. Girardi, F. Lanubile, and N. Novielli, "Emotion detection using noninvasive low cost sensors," *2017 7th Int. Conf. Affect. Comput. Intell. Interact. ACII 2017*, vol. 2018-Janua, no. August, pp. 125–130, 2018.
- [11] M. Hassenzahl, S. Diefenbach, and A. Göritz, "Needs, affect, and interactive products – Facets of user experience," *Interact. Comput.*, vol. 22, no. 5, pp. 353–362, Sep. 2010.
- [12] R. Jenke and A. Peer, "A cognitive architecture for modeling emotion dynamics: Intensity estimation from physiological signals," *Cogn. Syst. Res.*, vol. 49, pp. 128–141, Jun. 2018.
- [13] E. Kanjo, D. J. Kuss, and C. S. Ang, "NotiMind: Utilizing Responses to Smart Phone Notifications as Affective Sensors," *IEEE Access*, vol. 5, pp. 22023–22035, 2017.
- [14] Wei-Long Zheng and Bao-Liang Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
- [15] W. Sato, T. Fujimura, T. Kochiyama, and N. Suzuki, "Relationships among Facial Mimicry, Emotional Experience, and Emotion Recognition," *PLoS One*, vol. 8, no. 3, p. e57889, Mar. 2013.
- [16] J.-J. Wong and S.-Y. Cho, "A local experts organization model with application to face emotion recognition," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 804–819, Jan. 2009.
- [17] F. Y. Shih, C.-F. Chuang, and P. S. P. Wang, "Performance comparisons of facial expression recognition in jaffe database," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 22, no. 03, pp. 445–459, May 2008.
- [18] S. Demircan and H. Kahramanli, "Application of fuzzy C-means clustering algorithm to spectral features for emotion classification from speech," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 59–66, Apr. 2018.
- [19] S. Bao *et al.*, "Mining Social Emotions from Affective Text," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 9, pp. 1658–1670, Sep. 2012.
- [20] M. D. Pell and S. A. Kotz, "On the Time Course of Vocal Emotion Recognition," *PLoS One*, vol. 6, no. 11, p. e27256, Nov. 2011.
- [21] W. Sun, H. Zhao, and Z. Jin, "A visual attention based ROI detection method for facial expression recognition," *Neurocomputing*, vol. 296, pp. 12–22, Jun. 2018.
- [22] A. S. Alphonse and D. Dharma, "Novel directional patterns and a Generalized Supervised

- Dimension Reduction System (GSDRS) for facial emotion recognition," *Multimed. Tools Appl.*, vol. 77, no. 8, pp. 9455–9488, Apr. 2018.
- [23] H. F. García, M. A. Álvarez, and Á. A. Orozco, "Dynamic facial landmarking selection for emotion recognition using Gaussian processes," *J. Multimodal User Interfaces*, vol. 11, no. 4, pp. 327–340, Dec. 2017.
- [24] M. Wegrzyn, M. Vogt, B. Kireclioglu, J. Schneider, and J. Kissler, "Mapping the emotional face. How individual face parts contribute to successful emotion recognition," *PLoS One*, vol. 12, no. 5, p. e0177239, May 2017.
- [25] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Facial color is an efficient mechanism to visually transmit emotion," *Proc. Natl. Acad. Sci.*, vol. 115, no. 14, pp. 3581–3586, Apr. 2018.
- [26] V. Kosonogov *et al.*, "Facial thermal variations: A new marker of emotional arousal," *PLoS One*, vol. 12, no. 9, p. e0183592, Sep. 2017.
- [27] K. Lee, H. Hong, and K. Park, "Fuzzy System-Based Fear Estimation Based on the Symmetrical Characteristics of Face and Facial Feature Points," *Symmetry (Basel)*, vol. 9, no. 7, p. 102, Jun. 2017.
- [28] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1701–1708, 2014.
- [29] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, "Classical and modern face recognition approaches: a complete review," *Multimed. Tools Appl.*, vol. 80, no. 3, pp. 4825–4880, Jan. 2021.
- [30] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
- [31] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, Present, and Future of Face Recognition: A Review," *Electronics*, vol. 9, no. 8, p. 1188, Jul. 2020.
- [32] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," 2016, pp. 21–37.
- [33] V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using Eigenfaces," in *2011 3rd International Conference on Computer Research and Development*, 2011, pp. 302–306.
- [34] D. Lopez-Paz, S. Sra, A. J. Smola, Z. Ghahramani, and B. Schölkopf, "Randomized nonlinear component analysis," *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 4, pp. 3196–3204, 2014.
- [35] W. Yang, C. Sun, J. Yang, H. S. Du, and K. Ricanek, "Face Recognition Using Kernel UDP," *Neural Process. Lett.*, vol. 34, no. 2, pp. 177–192, Oct. 2011.
- [36] H. Moon and P. J. Phillips, "Computational and Performance Aspects of PCA-Based Face-Recognition Algorithms," *Perception*, vol. 30, no. 3, pp. 303–321, Mar. 2001.
- [37] M. Anggo and La Arapu, "Face Recognition Using Fisherface Method," *J. Phys. Conf. Ser.*, vol. 1028, p. 10, Jun. 2018.
- [38] G. M. Zafaruddin and H. S. Fadewar, "Face Recognition Using Eigenfaces," 2019, pp. 855–864.
- [39] M. N. Patil, Brijesh Iyer, and Rajeev Arya, "Performance Evaluation of PCA and ICA Algorithm for Facial Expression Recognition Application," 2016, pp. 965–976.
- [40] J. B. Alonso, J. Cabrera, M. Medina, and C. M. Travieso, "New approach in quantification of emotional intensity from the speech signal: emotional temperature," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9554–9564, Dec. 2015.
- [41] J. B. Alonso, J. Cabrera, C. M. Travieso, K. López-de-Ipiña, and A. Sánchez-Medina, "Continuous tracking of the emotion temperature," *Neurocomputing*, vol. 255, pp. 17–25, Sep. 2017.
- [42] Q. Mao, G. Xu, W. Xue, J. Gou, and Y. Zhan, "Learning emotion-discriminative and domain-invariant features for domain adaptation in speech emotion recognition," *Speech Commun.*, vol. 93, pp. 1–10, Oct. 2017.
- [43] F. Z. Chelali and A. Djeradi, "Text dependant speaker recognition using MFCC, LPC and

- DWT,” *Int. J. Speech Technol.*, vol. 20, no. 3, pp. 725–740, Sep. 2017.
- [44] B. John and J. Vidal, “Integración de técnicas de medición de emociones explícitas en la evaluación de la satisfacción de uso de Agentes Pedagógicos Virtuales,” pp. 124–132, 2014.
- [45] F. J. Pino, F. García, and M. Piattini, “Software process improvement in small and medium software enterprises: a systematic review,” *Softw. Qual. J.*, vol. 16, no. 2, pp. 237–261, Jun. 2008.
- [46] Y. Li and L. Zhu, “Optimization of user experience in mobile application design by using a fuzzy analytic-network-process-based Taguchi method,” *Appl. Soft Comput.*, vol. 79, pp. 268–282, Jun. 2019.
- [47] Y. Ding, Y. Cao, Q. Qu, and V. G. Duffy, “An Exploratory Study Using Electroencephalography (EEG) to Measure the Smartphone User Experience in the Short Term,” *Int. J. Human-Computer Interact.*, vol. 36, no. 11, pp. 1008–1021, Jul. 2020.
- [48] M. Kang *et al.*, “Identifying Characteristics and Types of Generation Z according to the Behavior of Smartphone Camera Use,” *Arch. Des. Res.*, vol. 33, no. 3, pp. 155–175, Aug. 2020.
- [49] D. Chung, “Aesthetic Characteristics and UX Evaluation of Mobile Platforms,” *Korean Soc. Emot. Sensib.*, vol. 18, no. 3, pp. 71–80, Sep. 2015.
- [50] V. Roto, *WEB BROWSING ON MOBILE PHONES - CHARACTERISTICS OF USER EXPERIENCE WEB BROWSING ON MOBILE PHONES - CHARACTERISTICS OF USER EXPERIENCE Doctoral Dissertation*, vol. 152, no. 3. 2006.
- [51] J. Park, S. H. Han, H. K. Kim, Y. Cho, and W. Park, “Developing Elements of User Experience for Mobile Phones and Services: Survey, Interview, and Observation Approaches,” *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 23, no. 4, pp. 279–293, Jul. 2011.
- [52] W. SONG, D. TJONDRONEGORO, and M. DOCHERTY, “EXPLORATION AND OPTIMIZATION OF USER EXPERIENCE IN VIEWING VIDEOS ON A MOBILE PHONE,” *Int. J. Softw. Eng. Knowl. Eng.*, vol. 20, no. 08, pp. 1045–1075, Dec. 2010.
- [53] W. Park, S. H. Han, S. Kang, Y. S. Park, and J. Chun, “A factor combination approach to developing style guides for mobile phone user interface,” *Int. J. Ind. Ergon.*, vol. 41, no. 5, pp. 536–545, Sep. 2011.
- [54] G. Shabrina, L. A. Lestari, B. M. Iqbal, and D. H. Syaifullah, “Redesign of User Interface Zakat Mobile Smartphone Application with User Experience Approach,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 505, p. 9, Jul. 2019.
- [55] K. J. Kim *et al.*, “Evaluation of smartphone user experience: identification of key evaluation items and their relationships,” *Int. J. Mob. Commun.*, vol. 16, no. 2, pp. 167–189, 2018.
- [56] C. J. Lin and L.-Y. Cheng, “Product attributes and user experience design: how to convey product information through user-centered service,” *J. Intell. Manuf.*, vol. 28, no. 7, pp. 1743–1754, Oct. 2015.
- [57] B. Choi, I. Lee, and J. Kim, “Culturability in Mobile Data Services: A Qualitative Study of the Relationship Between Cultural Characteristics and User-Experience Attributes,” *Int. J. Hum. Comput. Interact.*, vol. 20, no. 3, pp. 171–203, Jul. 2006.
- [58] Y. Cornet, M. J. Barradale, J. Bernardino, and G. Lugano, “Worthwhile Travel Time: Design Challenges of Capturing the User Experience by Smartphone,” in *2019 Smart City Symposium Prague (SCSP)*, 2019, pp. 1–9.
- [59] V. C. Gerogiannis, S. Papadopoulou, and E. I. Papageorgiou, “A Fuzzy Cognitive Map for Identifying User Satisfaction from Smartphones,” in *2012 16th Panhellenic Conference on Informatics*, 2012, pp. 156–161.
- [60] A. Gutierrez *et al.*, “Full-system analysis and characterization of interactive smartphone applications,” in *2011 IEEE International Symposium on Workload Characterization (IISWC)*, 2011, pp. 81–90.
- [61] A. Patidar and U. Suman, “Towards Analyzing Mobile App Characteristics for Mobile Software Development,” *2021 8th Int. Conf. Comput. Sustain. Glob. Dev.*, pp. 786–790, 2021.
- [62] Z. Linghao and L. Ying, “On methods of designing smartphone interface,” in *2010 IEEE International Conference on Software Engineering and Service Sciences*, 2010, pp. 584–587.

- [63] K. V. Sriram, H. M. Prabhu, and A. A. Bhat, "Mobile Phone Usability and its Influence on Brand Loyalty and Re-Purchase Intention: An Empirical," in *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2019, pp. 1–4.
- [64] C. Busso *et al.*, "Analysis of emotion recognition using facial expressions, speech and multimodal information," *ICMI'04 - Sixth Int. Conf. Multimodal Interfaces*, pp. 205–211, 2004.
- [65] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang, "Multimodal Approaches for Emotion Recognition: A Survey," *Internet Imaging VI*, vol. 5670, pp. 56–67, 2005.
- [66] J. YAN, G. LU, X. BAI, H. LI, N. SUN, and R. LIANG, "A Novel Supervised Bimodal Emotion Recognition Approach Based on Facial Expression and Body Gesture," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E101.A, no. 11, pp. 2003–2006, Nov. 2018.
- [67] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain Intelligence: Go beyond Artificial Intelligence," *Mob. Networks Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.
- [68] V. Lope Salvador, X. Mamaqi, and J. Vidal Bordes, "La inteligencia artificial," *Rev. ICONO14 Rev. científica Comun. y Tecnol. emergentes*, vol. 18, no. 1, pp. 58–88, Jan. 2020.
- [69] D. S. Manoharan, "AN IMPROVED SAFETY ALGORITHM FOR ARTIFICIAL INTELLIGENCE ENABLED PROCESSORS IN SELF DRIVING CARS," *J. Artif. Intell. Capsul. Networks*, vol. 2019, no. 2, pp. 95–104, Dec. 2019.
- [70] C. González García, E. Núñez-Valdez, V. García-Díaz, C. Pelayo G-Bustelo, and J. M. Cueva-Lovelle, "A Review of Artificial Intelligence in the Internet of Things," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 5, no. 4, p. 9, 2018.
- [71] G. N. Yannakakis and J. Togelius, *Artificial Intelligence and Games*. Cham: Springer International Publishing, 2018.
- [72] T. H. Davenport and D. D. D'Ercole, "Artificial intelligence for the real world," *Harv. Bus. Rev.*, 2017.
- [73] J. Hurwitz and D. Kirsch, *Machine Learning For Dummies, IBM Limited Edition*. 2018.
- [74] P. Amaral, J. Dinis, P. Pinto, L. Bernardo, J. Tavares, and H. S. Mamede, "Machine Learning in Software Defined Networks: Data collection and traffic classification," in *2016 IEEE 24th International Conference on Network Protocols (ICNP)*, 2016, pp. 1–5.
- [75] and A. C. Ian Goodfellow, Yoshua Bengio, "Deep learning book review," *Nature*, vol. 29, no. 7553, pp. 1–73, 2016.
- [76] F. Jiang *et al.*, "Deep Learning Based Multi-Channel Intelligent Attack Detection for Data Security," *IEEE Trans. Sustain. Comput.*, vol. 5, no. 2, pp. 204–212, 2020.
- [77] L. M. Rojas-Barahona, "Deep learning for sentiment analysis," *Lang. Linguist. Compass*, vol. 10, no. 12, pp. 701–719, 2016.
- [78] S. Tripathi, S. Tripathi, and H. Beigi, "Multi-Modal Emotion recognition on IEMOCAP Dataset using Deep Learning," 2018.
- [79] H. S. Bae, H. J. Lee, and S. G. Lee, "Voice recognition based on adaptive MFCC and deep learning," *Proc. 2016 IEEE 11th Conf. Ind. Electron. Appl. ICIEA 2016*, pp. 1542–1546, 2016.
- [80] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer (Long. Beach. Calif.)*, vol. 29, no. 3, pp. 31–44, 1996.
- [81] X. Yao, "Evolving artificial neural networks," *Proc. IEEE*, vol. 87, no. 9, pp. 1423–1447, 1999.
- [82] S. Albawi, T. A. M. Mohammed, and S. Alzawi, "Layers of a Convolutional Neural Network," *Ieee*, 2017.
- [83] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," pp. 1–38, 2015.
- [84] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2018, vol. 1, pp. I-511-I-518.
- [85] P. N. Astya, M. Singh, Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers. Uttar Pradesh Section, and Sharda University. School of Engineering and Technology, "Face Detection and Recognition Using OpenCV," pp. 116–119, 2019.

- [86] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, "FAREC - CNN based efficient face recognition technique using Dlib," *Proc. 2016 Int. Conf. Adv. Commun. Control Comput. Technol. ICACCCT 2016*, no. 978, pp. 192–195, 2017.
- [87] H. Ku and W. Dong, "Face Recognition Based on MTCNN and Convolutional Neural Network," *Front. Signal Process.*, vol. 4, no. 1, pp. 37–42, 2020.
- [88] E. Jose, M. Greeshma, T. P. Mithun Haridas, and M. H. Supriya, "Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2," *2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019*, pp. 608–613, 2019.
- [89] S. Rhyou and H. Kim, "Development of Access Management System based on Face Recognition using ResNet," vol. 22, no. 8, pp. 823–831, 2019.
- [90] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech Emotion Recognition Using Deep Learning Techniques: A Review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.
- [91] M. Chen, X. He, J. Yang, and H. Zhang, "3-D Convolutional Recurrent Neural Networks With Attention Model for Speech Emotion Recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1440–1444, Oct. 2018.
- [92] K. M. Kudiri, A. M. Said, and M. Y. Nayan, "Emotion Detection through Speech and Facial Expressions," in *2014 International Conference on Computer Assisted System in Health*, 2014, pp. 26–31.
- [93] V. B.V, "Vokaturi," 2021. [Online]. Available: <https://vokaturi.com/>.
- [94] M. Hevery, "Angular." [Online]. Available: <https://angular.io/>.
- [95] E. Correa, A. Jonker, M. Ozo, and R. Stolk, "Emotion Recognition using Deep Convolutional Neural Networks," p. 12, 2016.
- [96] S. Tobin and S. Grondin, "Video games and the perception of very long durations by adolescents," *Comput. Human Behav.*, vol. 25, no. 2, pp. 554–559, 2009.
- [97] U. Dimberg, M. Thunberg, and S. Grunedal, "Facial reactions to emotional stimuli: Automatically controlled emotional responses," *Cogn. Emot.*, vol. 16, no. 4, pp. 449–471, 2002.
- [98] H. Hassan and S. A. Suandi, "Analysis of Local Binary Pattern for Facial Expression Recognition Using Patch Local Binary Pattern on Extended Cohn Kanade Database," 2019, pp. 633–639.
- [99] V. Mokaya, S. Gyan, and V. Univeristy, "PERFORMANCE EVALUATION OF MMI DATASET ON SIX MACRO EXPRESSIONS USING 2D-LDA AND," no. August 2020, 2021.
- [100] L. Zahara, P. Musa, E. Prasetyo Wibowo, I. Karim, and S. Bahri Musa, "The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–9.
- [101] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 94–101.
- [102] C. Fan, F. Li, Y. Jiao, and X. Liu, "A novel lossless compression framework for facial depth images in expression recognition," *Multimed. Tools Appl.*, vol. 80, no. 16, pp. 24173–24183, Jul. 2021.
- [103] A. Mahmood, S. Hussain, K. Iqbal, and W. S. Elkilani, "Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion," *Math. Probl. Eng.*, vol. 2019, pp. 1–12, Aug. 2019.
- [104] H.-S. Lee and B.-Y. Kang, "Continuous emotion estimation of facial expressions on JAFFE and CK+ datasets for human–robot interaction," *Intell. Serv. Robot.*, vol. 13, no. 1, pp. 15–27, Jan. 2020.
- [105] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Learning Stylized Character Expressions from Humans," Nov. 2019.
- [106] I. A. M. Verpaalen, G. Bijsterbosch, L. Mobach, G. Bijlstra, M. Rinck, and A. M. Klein,

- “Validating the Radboud faces database from a child’s perspective,” *Cogn. Emot.*, vol. 33, no. 8, pp. 1531–1547, Nov. 2019.
- [107] J. Yi, Y. Sima, M. Zhou, and J. Yang, “Facial Expression Sequence Interception Based on Feature Point Movement,” in *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*, 2019, pp. 58–62.
- [108] A. K. Prabhakaran, J. J. Nair, and S. Sarath, “Thermal Facial Expression Recognition Using Modified ResNet152,” 2021, pp. 389–396.
- [109] xinzheng Chen, L. Song, and C. Qiu, “Face Recognition by Feature Extraction and Classification,” in *2018 12th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, 2018, pp. 43–46.
- [110] M. Azimi and A. Pacut, “The effect of gender-specific facial expressions on face recognition system’s reliability,” in *2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, 2018, pp. 1–4.
- [111] C. B. Seaman, “Qualitative methods in empirical studies of software engineering,” *IEEE Trans. Softw. Eng.*, vol. 25, no. 4, pp. 557–572, 1999.
- [112] B. A. Kitchenham *et al.*, “Preliminary Guidelines for Empirical Research in Software Engineering,” *Ieee Trans. Softw. Eng.*, vol. 28, no. 8, pp. 721–734, 2002.
- [113] J. M. Verner, J. Sampson, V. Tomic, N. A. Abu Bakar, and B. A. Kitchenham, “Guidelines for industrially-based multiple case studies in software engineering,” *Proc. 2009 3rd Int. Conf. Res. Challenges Inf. Sci. RCIS 2009*, pp. 313–324, 2009.
- [114] M. Karg, A.-A. Samadani, R. Gorbet, K. Kuhlentz, J. Hoey, and D. Kulic, “Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation,” *IEEE Trans. Affect. Comput.*, vol. 4, no. 4, pp. 341–359, Oct. 2013.
- [115] M. Coulson, “Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence,” *J. Nonverbal Behav.*, vol. 28, no. 2, pp. 117–139, 2004.
- [116] T. Toa and R. José, “Diseño de un Videojuego para Dispositivos Móviles que Fomente el Reciclaje en los Niños y Niñas,” pp. 107–113, 2016.
- [117] G. Besoli, N. Palomas, and A. Chamorro, “Uso del movil en padres, niños y adolescentes,” *Rev. Psicol. Ciències l’Educació i l’Esport*, vol. 36, no. 1, pp. 1–12, 2018.
- [118] C. Saha, W. Ahmed, S. Mitra, D. Mazumdar, and S. Mitra, “Facial Expressions: A Cross-Cultural Study,” *Emot. Recognit. A Pattern Anal. Approach*, no. July 2018, pp. 69–87, 2015.