



PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para
obtener el Título de INGENIERO DE SISTEMAS

APLICACIÓN DEL APRENDIZAJE PROFUNDO PARA EL RECONOCIMIENTO Y CLASIFICACIÓN DE EMOCIONES FACIALES

Por

Br. Yenifer Rosana Ramírez Rivas

Tutor: Dr. Jonás Arturo Montilva Calderón

Enero 2022

©2022 Universidad de Los Andes, Mérida, Venezuela

APLICACIÓN DEL APRENDIZAJE PROFUNDO PARA EL RECONOCIMIENTO Y CLASIFICACIÓN DE EMOCIONES FACIALES

Br. Yenifer Rosana Ramírez Rivas

Proyecto de Grado — Sistemas Computacionales, 87 páginas

Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2022

Resumen: Las expresiones faciales tienen una alta vinculación con las emociones humanas, siendo una capacidad nata en cualquier persona el poder identificar estas emociones, haciendo posible que se puedan entender unos con otros. Actualmente la computación afectiva ha atisbado el reconocimiento facial de emociones para mejorar la interacción hombre-computador y junto al avance de los algoritmos de Inteligencia Artificial se han incrementado las investigaciones en el área, aunque la mayoría de ellas en espacios totalmente controlados. En este Proyecto de Grado, se implementa un Sistema de Reconocimiento de Emociones Faciales con algoritmos de Aprendizaje Profundo, que permite a los usuarios la clasificación automática de emociones en rostros de personas, mediante el vídeo en tiempo real captado por la cámara del dispositivo donde se ejecute la aplicación web. La tarea de clasificación de este sistema es realizado por un modelo entrenado con Aprendizaje Profundo, implementado con dos algoritmos distintos, por lo que se opta por hacer una validación con un caso de estudio, enfocado en las emociones de influencers en vídeos de reacciones.

Palabras clave: Reconocimiento de emociones, Reconocimiento facial, Aprendizaje profundo

Este trabajo fue procesado en L^AT_EX.

A mis padres ...

Índice general

Introducción	xii
1. Planteamiento del problema	1
1.1. Definición y contextualización del problema	1
1.2. Objetivos	2
1.2.1. Objetivo general	2
1.2.2. Objetivos específicos	3
1.3. Metodología	3
1.4. Alcance	5
1.5. Justificación	6
1.6. Antecedentes	7
1.6.1. Algoritmos de Aprendizaje Profundo aplicados para el Recono- cimiento y Clasificación de Emociones Faciales	8
1.6.2. Herramientas de Software Libre para el Reconocimiento y Clasi- ficación de Emociones Faciales	12
1.7. Estructura del Documento	16
2. Marco Teórico	17
2.1. Inteligencia Artificial	17
2.2. Red Neuronal	18

2.3.	Aprendizaje Automático	19
2.3.1.	Aprendizaje Supervisado	20
2.3.2.	Aprendizaje no Supervisado	20
2.4.	Aprendizaje Profundo	20
2.4.1.	Red Neuronal Prealimentada	21
2.4.2.	Redes Neuronales Convolucionales	22
2.4.2.1.	Capa de convolución	23
2.4.2.2.	Capa de agrupación	24
2.5.	Reconocimiento de Emociones Faciales	26
2.5.1.	Sistema de código de acción facial (FACS)	28
2.6.	Tecnologías	30
2.6.1.	Python	30
2.6.2.	Tensorflow	30
2.6.3.	TensorBoard	31
2.6.4.	Keras	31
2.6.5.	Numpy	31
2.6.6.	Matplotlib	32
2.6.7.	Google Colaboratory	32
2.6.8.	Frameworks	32
2.6.8.1.	Vue.js	32
2.6.8.2.	Vuetify	33
2.6.8.3.	Jest.js	33

3. Ingeniería de Requisitos 34

3.1.	Descripción general del Sistema de Reconocimiento y Clasificación de Emociones Faciales	35
------	---	----

3.1.1.	Funciones del Sistema de Reconocimiento y Clasificación de Emociones Faciales	36
3.1.1.1.	Componente de Pre-procesamiento de Datos	36
3.1.1.2.	Componente de Reconocimiento y Clasificación	37
3.1.1.3.	Interfaz	37
3.2.	Especificación de Requisitos	37
3.2.1.	Requisitos Funcionales	37
3.2.1.1.	Gestión de Entrada de Datos	38
3.2.1.2.	Gestión de Reconocimiento y Clasificación de Emociones Faciales	40
3.2.1.3.	Gestión de Salida de Análisis	42
3.2.2.	Requisitos no Funcionales	43
4.	Diseño	44
4.1.	Diseño del Sistema de Reconocimiento y Clasificación de Emociones Faciales	45
4.1.1.	Diseño Arquitectónico del Sistema	45
4.1.2.	Diseño de Subsistemas	45
4.1.2.1.	Pre-procesamiento	47
4.1.2.2.	Reconocimiento y Clasificación	47
4.1.3.	Diseño de la Interfaz del Sistema	48
4.2.	Diseño de la Base de Datos	49
4.2.1.	FER2013	50
4.2.2.	Base de Datos del Caso de Estudio	50
4.3.	Diseño del Reconocedor de Emociones Faciales	51
4.3.1.	Selección de algoritmo	51

4.3.2. Implementación	52
5. Pruebas	55
5.1. Pruebas del Sistema de Reconocimiento de Emociones Faciales	56
5.1.1. Plan de Pruebas	56
5.1.1.1. Pruebas Unitarias	56
5.1.1.2. Especificación del Diseño de Pruebas	57
5.2. Pruebas del Reconocedor de Emociones Faciales	57
5.2.1. Prueba 1	59
5.2.2. Prueba 2	61
5.2.2.1. Modelo diseñado con un MLP	61
5.2.2.2. Modelo diseñado con una CNN	63
5.2.2.3. AFFDEX SDK	66
5.2.3. Análisis de los resultados	67
6. Conclusiones y Recomendaciones	69
6.1. Conclusiones	70
6.2. Recomendaciones	72
Bibliografía	74
A. FACS	81
A.1. Gestos Faciales	81
A.2. AUs	82
A.3. Acciones Misceláneas	85
A.4. Códigos de Visibilidad	86
A.5. Códigos de Intensidad	86
A.6. Códigos de Comportamiento	87

Índice de figuras

1.1. Modelo general de procesos del Método W_Watch. (bar, 2010)	4
2.1. Representación matemática de una Red Neuronal Prealimentada (Senthilkumar, 2010).	22
2.2. Convolución, mapas de características (Diego Calvo, 2017)	23
2.3. Reducción en capa de agrupación (Diego Calvo, 2017)	25
2.4. Pasos en el sistema de reconocimiento facial (Wati and Abadianto, 2017)	28
3.1. Jerarquía de Usuario del Sistema	35
3.2. Modelo de Componente del Sistema de Reconocimiento y Clasificación de Emociones Faciales	36
3.3. Diagrama General del Sistema de Reconocimiento y Clasificación de Emociones Faciales	38
3.4. Diagrama de la Gestión de Entrada de Datos	40
3.5. Diagrama de Gestión de Reconocimiento y Clasificación de Emociones Faciales	42
3.6. Diagrama de Gestión de Salida de Análisis	42
4.1. Diagrama de Despliegue del Sistema	46
4.2. Árbol de componentes del sistema <i>Vue.js</i>	46
4.3. Mockup del SREF	49

4.4. Distribución de emociones en <i>FER2013</i>	50
4.5. Arquitecturas de los modelos	54
5.1. Pruebas en el modelo MLP.	62
5.2. Pruebas en el modelo MLP.	63
5.3. Pruebas en el modelo CNN.	64
5.4. Pruebas en el modelo CNN.	65
5.5. Pruebas en el modelo CNN en Caso de Estudio.	66
5.6. Matriz de Confusión de <i>Affective</i> evaluado con el Caso de Estudio.	67

Índice de tablas

2.1. Versiones de Herramientas y tecnologías usada en el desarrollo.	30
3.1. <i>CU1</i> : Ingresar al Sistema.	39
3.2. <i>CU2</i> : cargar datos.	39
3.3. <i>CU3</i> : reconocer los AU's.	41
3.4. <i>CU4</i> : clasificación de emociones.	41
3.5. <i>CU5</i> : obtener información.	43
4.1. Tabla de clasificación de emociones para diferentes algoritmos (Soberanes et al., 2019; Gallud Baños, 2019)	51
5.1. Pruebas Unitarias del SREF	58
5.2. Parámetros para <code>ImageDataGenerator</code>	60
5.3. Tasa de aciertos de reconocimiento en modelos	61
5.4. Métricas para el modelo MLP	62
5.5. Métricas para el modelo MLP	63
5.6. Métricas para el modelo CNN	64
5.7. Métricas para el modelo CNN	65
5.8. Métricas para el modelo CNN para Caso de Estudio	65
5.9. Métricas para el modelo CNN para Caso de Estudio con transformación de datos	66

5.10. Métricas para el Software <i>Afectiva</i> con el Caso de Estudio sin transformación de datos	67
A.1. Gestos Faciales en las Emociones	82
A.2. AUs de las Emociones. Traducido de (Kanade et al., 2000)	85
A.3. Acciones Misceláneas. Traducido de (Kanade et al., 2000)	86
A.4. Códigos de Visibilidad	86
A.5. Intensidad de la microexpresión	86
A.6. Códigos de Comportamientos	87

Introducción

La Inteligencia Artificial tiene implicaciones cada vez más amplias en aspectos comunes de la vida del ser humano, la aplicación de técnicas del Big Data y algoritmos de Aprendizaje de Máquina y Aprendizaje Profundo se han ligado fuertemente al crecimiento de las grandes compañías, quienes hacen uso de ellos para poder adaptarse mejor a las necesidades de sus usuarios. Por ejemplo, empresas de la rama médica expresan usar varias técnicas en el apoyo de sus profesionales, compañías de vehículos autónomos respaldan la vida de sus usuarios utilizando algoritmos cada vez más precisos en sus tareas y así muchas otras.

Las técnicas y algoritmos de Aprendizaje Profundo han demostrado un alto rendimiento en la clasificación, si en cuanto a tareas de procesamiento de imágenes se trata, segmentación, eliminación de ruido, reconocimiento de objetos, reconocimiento de rostros, son algunas de ellas. Algunas técnicas han resaltado más que otras, dependiendo de la aplicación en sí. En particular el reconocimiento de emociones en rostros de personas, ha sido un área de interés investigativo debido a su aplicación en los procesos de la Computación Afectiva.

Con la extracción del rostro de las características asociadas a una emoción, se ha expresado en muchos trabajos que las Redes Convolucionales funcionan de

manera sorprendente, logrando un excelente equilibrio entre funcionalidad y la relativa facilidad de implementación (Pérez Estruch, 2017), trabajos como (Calvachi and Alexandra, 2019; Saez De La Pascua, 2019) decantan que con la aplicación de estos tipos de algoritmos se reduce el coste computacional para encontrar soluciones a problemas complejos, como este, gracias a que los algoritmos aprenden y se adaptan rápidamente.

En este proyecto de grado se crea una Red Neuronal Convolutiva para la Detección de Emociones Faciales, para lo que se ha seleccionado una base de datos de imágenes de rostros de personas con emociones, con ella se pretende hacer la clasificación. El entrenamiento de la red se realizará con bibliotecas de *TensorFlow*, una vez finalizado el entrenamiento se convertirá el modelo con sus respectivos pesos en formato *JSON*, con la Biblioteca *TensorFlowjs*.

El objetivo de este trabajo se expande a desarrollar un sistema capaz de reconocer y clasificar emociones faciales de una persona, a través de un vídeo en espacios no controlados, por lo que una vez obtenido un modelo en formato *JSON*, se continuará con el desarrollo de este sistema a través de los Frameworks *Vue.js* y *Vuetify* los cuales facilitarán, entre otras cosas, el uso del modelo Face-landmarks-detection para la detección y rastreo de puntos claves del rostro, que complementado al modelo, clasificará las emociones faciales del rostro entre la 6 expresiones básicas de Ekman Ekman et al. (1969) y la expresión neutral, además de esto permitirá al usuario probar en tiempo real los resultados obtenidos, utilizándolos a través de la aplicación web que se desarrolla.

Capítulo 1

Planteamiento del problema

1.1. Definición y contextualización del problema

Cuando el ser humano intenta comunicarse con otras personas un porcentaje muy alto de esta comunicación es no verbal. Se ha demostrado que las expresiones faciales tienen una vinculación con las emociones humanas y por lo tanto con su intención. La capacidad que tiene el ser humano para detectar e identificar estas emociones es muy buena, y hace posible que se puedan entender los unos a los otros (Saez De La Pascua, 2019). Aunque, a veces esta tarea se torna muy difícil, debido al movimiento no rígido de los músculos faciales (Cordero and Aguilar, 2016) que hace que no se pueda reconocer una emoción espontánea, de una artificial.

Actualmente, y tras la popularización de la Inteligencia Artificial, muchos investigadores del área se han enfocado en resolver los problemas relacionados con el reconocimiento facial de emociones en personas, ya sea para ayudar a otras personas a identificar las emociones reales, o a la Computación Afectiva para mejorar la relación hombre-computador.

Investigaciones realizadas han afirmado que la Inteligencia Artificial y en específico el Aprendizaje Profundo, son herramientas que permiten mejorar los tiempos de reconocimientos faciales automáticos. Para ampliar esta línea de investigación en (Pellejero et al., 2017) se considera necesario desarrollos robustos a variaciones que fácilmente se pueden dar fuera del entorno del laboratorio, como pueden ser variaciones lumínicas, imágenes con ruido, variaciones de traslación, entre otros. Ya que actualmente las investigaciones que atacan el problema de reconocimiento y clasificación de emociones presentan excelentes resultados en entornos controlados, pero no en espacios no controlados, dónde es más común que cualquier dispositivo inteligente o ser humano interactúe.

Los dispositivos que ejercen tareas de reconocimiento de emociones faciales pueden tener diferentes tipos de entradas para obtener información, siendo las más comunes las imágenes, y ahora el vídeo, buscando resultados en tiempo real. En este sentido se considera que se deben obtener alternativas que aborden el problema de reconocimiento y clasificación de emociones faciales de los seres humanos en espacios no controlados, a partir de un vídeo.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo general de este trabajo de grado es desarrollar un sistema capaz de reconocer y clasificar emociones faciales de una persona, a través de un vídeo en espacios no controlados.

Para ello se debe cumplir con los siguientes objetivos específicos:

1.2.2. Objetivos específicos

- Seleccionar los algoritmos para el desarrollo del sistema de reconocimiento y clasificación de emociones en personas.
- Definir la metodología de pre-procesamiento de los datos de entrada utilizados por el sistema.
- Desarrollar un sistema de reconocimiento y clasificación de emociones que determine las emociones de una persona, presentadas en un vídeo en espacios no controlados.
- Entrenar el sistema de reconocimiento y clasificación de emociones de una persona para un caso de estudio.

1.3. Metodología

El siguiente trabajo utilizará parte de la metodología de desarrollo de software basada en el método W-Watch (bar, 2010), definido como “un marco metodológico que describe, el conjunto estructurado de actividades necesarias para desarrollar un producto de software pequeño, baja complejidad y con documentación técnica precisa, es una versión liviana del método Watch”. El método en sí comprende un modelo de procesos que se organiza en dos partes: procesos gerenciales, que incluye los procesos de soporte y los procesos técnicos del desarrollo del producto de software que serán ejecutados de forma cíclica, iterativa y controlada a fin de corregir errores, introducir nuevos requisitos, si es necesario, o mejorar el producto (Méndez Rosales, 2021).



Figura 1.1: Modelo general de procesos del Método W-Watch. (bar, 2010)

En consecuencia y según la metodología W-Watch se detallarán las actividades de los procesos de la Figura 1.1 a seguir en este proyecto:

- **Ingeniería de requisitos:** se identifican y analizan los problemas que tienen los programadores de Redes Neuronales en el desarrollo de Sistemas que permitan identificar las emociones faciales de un rostro humano, en un vídeo en tiempo real. Determinando así los requisitos necesarios para la elaboración del software, organizándolos en orden de prioridad para su posterior ejecución durante la elaboración del prototipo (Méndez Rosales, 2021).
- **Proceso de diseño de software:** se determina la arquitectura que tendrá la aplicación, identificando cuáles serán sus componentes y las interrelaciones entre estos componentes. También, se describe la plataforma e infraestructura física donde esos componentes se desplegarán.
- **Proceso de aprovisionamiento de componentes:** se realiza la instalación del sistema, la adquisición de los componentes, adaptación y desarrollo. Posteriormente se inician las pruebas en cada componente o módulo, y se corrigen errores

de ser necesario.

- **Implementación del sistema de software:** durante este proceso se ensamblan los componente realizados junto a la interfaz del sistema, para pasar al diseño y ejecución de las pruebas.
- **Proceso de pruebas del sistema de software:** se procede a definir los mecanismos de pruebas, se preparan y se realizan distintos tipos de pruebas (Méndez Rosales, 2021). Para este punto el sistema de reconocimiento debe ser probado y depurado.
- **Proceso de entrega del sistema de software:** se planifica la instalación tomando en cuenta costo, tiempo y recursos necesarios. Se debe desplegar o instalar el sistema en la plataforma requerida.

1.4. Alcance

En este proyecto se pretende desarrollar un sistema capaz de gestionar actividades de reconocimiento y clasificación de emociones faciales, a través de la evaluación de vídeos, previamente separados en fotogramas. Este sistema permitirá tener como entrada vídeos en entornos no controlados, semejantes a cualquier caso real de interacción.

El sistema será implementado a través de algoritmos y técnicas de aprendizaje profundo, evaluados previamente. Para el entrenamiento y evaluación del sistema se utilizarán bases de datos conformada por fotogramas de emociones faciales, cuyo uso es de orden libre. Además, como caso de estudio, se utilizarán muestras de vídeos provenientes de influencers de redes sociales.

1.5. Justificación

Las expresiones faciales juegan un papel fundamental en la comunicación humana (Albert, 1972). El lenguaje corporal que expresa una persona tiene un rol importante en la comunicación, estudios han demostrado que las expresiones faciales son útiles para expresar emociones y sentimientos (Zhang et al., 2014).

A pesar de ser una actividad cotidiana en el humano (el reconocer expresiones y sentimientos en sus homólogos), este en ciertas oportunidades no es capaz de diferir una expresión real de una artificial, simplemente por el hecho de encontrarse con un rostro poco emotivo o presentarse con casos donde según la situación, el otro pueda estar fingiendo. Por otro lado, las máquinas siguen buscando realizar esta actividad de manera similar e incluso mejor que el humano, la rama de la Computación Afectiva trata de mejorar la inclusión de los aspectos afectivos en el procesamiento que realizan las máquinas y es por tal que el reconocimiento y clasificación de emociones faciales también se desea mejorar.

La aplicación de técnicas y algoritmos de Aprendizaje Profundo son muy comunes en este tipo de investigación, en trabajos como (Gallud Baños, 2019; Pérez Estruch, 2017; Saez De La Pascua, 2019) se ha demostrado obtener mejores resultados, en comparación a otras técnicas aplicadas para el reconocimiento facial y el reconocimiento de emociones. Aunque, por otra parte, las revisiones bibliográficas demuestran que es poco común el estudio de esta misma actividad pero en entornos no controlados, en estos espacios características como: la variación lumínica, posición del rostro, ángulo de la cámara, entre otros, puede afectar enormemente el acierto de los algoritmos, siendo esto algo negativo, ya que los lugares idóneos donde se ejerce interacción

hombre-hombre u hombre-máquina son estos mismos espacios.

Por tal, en este trabajo se pretende ampliar este campo de investigación, realizando un sistema que permita clasificar y reconocer las emociones faciales de los rostros humanos, en entornos no controlados a través de un vídeo, y así expandir las alternativas de desarrollo en el área, enfocándose en optimizar el uso de recursos computacionales y evitar cambios en el rendimiento de acierto. Esta investigación se pretende desarrollar con la aplicación de técnicas y algoritmos de Aprendizaje Profundo.

1.6. Antecedentes

El reconocimiento facial de la expresión emocional es la capacidad de todos los individuos de reconocer formas básicas de expresión afectiva, la cual aparece en los rostros de las personas y se constituyen por 6 emociones básicas (Russell, 1994; Saracco, 2012). Estas emociones básicas y expresiones faciales son representadas por la forma que manifiesten: la frente o cejas, los ojos y párpados, y la parte inferior de la cara (boca y alrededores). Según la investigación (Ekman and Friesen, 1976), Ekman y Friesen declaran que estas expresiones faciales de las emociones son universales e innatas, sin embargo el desarrollo del reconocimiento facial emocional suele ser un proceso gradual, el cual se ve afectado por múltiples factores que van desde la edad hasta el contexto en donde las personas se desenvuelven, situaciones de desarrollo cognitivo y neuronal, entre otras (Martínez et al., 2010).

El reconocimiento de expresiones faciales permite a los seres humanos identificar los sentimientos y comportamientos de otros, además de ser una clave fundamental para una buena comunicación, por tal razón, el fallo en la interpretación correcta de

información emocional, como lo son las expresiones faciales, pueden ser causantes de conflictos interpersonales y sociales (Arango et al., 2013; Ekman and Oster, 1979).

Es por ello que durante mucho tiempo se ha buscado mejorar esta actividad humana e incluirlas en las máquinas, utilizando los avances, las técnicas y algoritmos de la Inteligencia Artificial, por ello una primera parte de los antecedentes de este proyecto se enfocarán en la aplicación de algoritmos de Aprendizaje Profundo para el reconocimiento y clasificación de emociones humanas y una segunda en softwares de orden libre existentes para la ejecución de esta actividad.

1.6.1. Algoritmos de Aprendizaje Profundo aplicados para el Reconocimiento y Clasificación de Emociones Faciales

Uno de los trabajos más semejantes al concepto de este Proyecto es el expuesto por Pellejero y Grinblat en su artículo (Pellejero et al., 2017) titulado “Análisis semántico en rostros utilizando redes neuronales profundas”, en este trabajo se aplicó la Técnica de Generative Adversarial Networks (GAN) que permite utilizar un gran cúmulo de imágenes no etiquetadas con un estilo de entrenamiento semi supervisado. Sus bases teóricas fueron pautadas por el Facial Action Coding System de Paul Ekman y para el desarrollo del reconocimiento y clasificación de las emociones se enfatiza la selección de los datos y su pre-procesamiento aplicando dos metodologías de procesamiento distintas. La experimentación con clases de emociones se hizo en dos etapas, con el objetivo de estudiar dos metodologías de transferencia de conocimiento y poder compararlas. La primera estrategia fue el enfoque más tradicional, donde todos los pesos de los modelos entrenados para la tarea t son transferidos a la tarea t_0 . Aquí

se transfirieron todas las capas salvo la última, la cual toma la decisión sobre la clasificación final, esto fue así ya que los conjuntos de datos de apoyo estaban pensados para la misma tarea, el reconocimiento de emociones, con lo cual las características de alto contenido semántico aprendidos por las capas más cercanas a la salida de los modelos también podían ser transferidos. La segunda estrategia fue utilizar el modelo GAN, modificado para aceptar un conjunto de datos no etiquetado, y otro etiquetado. De esta forma, se agrega a la función a minimizar un término correspondiente al error de clasificación. Así, el modelo puede aceptar mini-batches de datos con o sin etiquetas y en cada caso la función de error se adaptará.

En el Trabajo Final de Máster titulado “Emotions Recognition using Deep Learning” (Pérez Estruch, 2017) el autor realiza una investigación sobre las Redes Convolucionales Profundas para el reconocimiento de emociones en rostros, evalúa el efecto de cada uno de los parámetros que intervienen, así como aspectos de coste computacional. Para el desarrollo se utilizó software de reconocimiento facial para la extracción de características y la posterior clasificación, una vez se extrajo un modelo facial fue posible extraer los parámetros pertinentes utilizando el sistema de codificación facial de Ekman, que declara unas unidades de movimiento muscular para distintas partes de la cara. Estos parámetros de movimiento como la posición, dirección, entre otros, de músculos se pasaron por diferentes clasificadores. Posteriormente se hizo la misma clasificación con las Redes Neuronales Convolucionales, donde se pudo comprobar que estas funcionan de forma sorprendente, aunque no se logró superar a Yichuan Tang, ganador de la competición de Kaggle.

Siguiendo la línea de revisión bibliográfica se encuentra el artículo científico de Trabajo de Máster (Calvachi and Alexandra, 2019) titulado “Uso de reconocimiento

facial de emociones basado en técnicas de Deep Learning para el mejoramiento de la educación” en el cual se plantean como objetivo utilizar algoritmos de la Inteligencia Artificial que permitan de forma masiva y autónoma identificar sentimientos basados en los rostros de los estudiantes, para aplicar estrategias pedagógicas adecuadas que mejoren su educación. Para el desarrollo se obtuvo un dataset que contiene imágenes de rostros de personas con emociones de tipo alegre, tristeza, sorpresa, enojo, miedo, cansado y normal. Luego se procede con la lectura de la imagen por medio de una cámara digital, utilizando la librería OpenCV se procesaron las imágenes y se procedió a realizar una comparación y análisis para determinar el tipo de emoción en una persona con el software TensorFlow, mediante Redes Neuronales Convolucionales (CNN) para entrenamiento. Como conclusiones de este trabajo se obtuvo que la calidad de la identificación de emociones depende de la extensión de la base de imágenes que se use para el entrenamiento, en este caso, de las CNN. Por medio de estos algoritmos, se reduce el costo computacional para encontrar soluciones optimizadas a problemas complejos gracias a que los algoritmos aprenden y se adaptan rápidamente.

Por otro lado en “Deep learning para el reconocimiento facial de emociones básicas”(Saez De La Pascua, 2019) se plantea como objetivo investigar la utilización de técnicas de Aprendizaje Profundo para el reconocimiento de emociones en imágenes faciales, el trabajo hace una comparación entre los algoritmos de Aprendizaje Automático y Profundo.

En la primera parte del trabajo se utilizan algoritmos de Aprendizaje Automático, se recolectan y preparan los datos, se seleccionan y evalúan los clasificadores y por último se entrenan y evalúan. Los clasificadores seleccionados se validaron con una técnica conocida como “CrossValidation”. Esta técnica consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones. Se

denotó que se obtienen mejores resultados estandarizando los puntos o coordenadas, esto hace que incremente su acierto hasta en 10 % por cada algoritmo. En la segunda parte, se procedió a trabajar con algoritmos de Aprendizaje Profundo siguiendo la metodología de recolección y preparación de datos, y elección y evaluación del modelo. Se trabajó con tres modelos distintos acumulativos de capas, el primer modelo con una capa de convolución y una capa de agrupación, el segundo con dos capas de convolución y dos capas de agrupación y el tercero con tres capas de convolución y tres capas de agrupación. En ambas partes del trabajo se han podido predecir correctamente como mínimo un 50 % de las imágenes de entrada para 7 sentimientos llegando a un 65 % con Aprendizaje Profundo.

También en el Trabajo de Grado titulado (Jaramillo Romero, 2020) “Detección de emociones y reconocimiento facial utilizando aprendizaje profundo” de Romero y Ezequiel se aplicaron CNN utilizando el dataset de Kaggle denominado Learn facial expressions from an image (FER2013), y el entrenamiento de la red convolucional se realizó en Kaggle Notebooks con las bibliotecas Keras, Tensorflow, OpenCV y el algoritmo pre-entrenado Frontal Face Haar Cascade. Con la librería de Matplotlib se genera una gráfica donde se evaluó el modelo, identificando los valores de pérdida y precisión tanto de la fase de entrenamiento como de validación. La matriz de confusión se generó con la librería Sklearn, en la cual se evalúa la precisión de la clasificación. En este trabajo el entrenamiento de la CNN, haciendo referencia a la detección de emociones, terminó en la época 167, debido a la configuración de detención temprana que se realizó, alcanzando hasta esa época una precisión del 67,13 %, el mejor modelo obtuvo una precisión del 71,16 %.

1.6.2. Herramientas de Software Libre para el Reconocimiento y Clasificación de Emociones Faciales

AFFDEX SDK es un conjunto de herramientas de reconocimiento de expresiones faciales en tiempo real, que además detecta automáticamente las expresiones de varias personas simultáneamente, este kit está disponible en las principales plataformas móvil (Android y Ios) y de escritorio (Windows). Como se expone en el artículo (McDuff et al., 2016) titulado “AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit”, los creadores de esta herramienta presentan una aplicación de demostración que proporciona visualización en el tiempo de las expresiones captadas por la cámara, basada, según los autores, en la taxonomía más objetiva y completa para codificar la conducta facial (el Sistema de Codificación de Acción Facial). El sistema de codificación facial automatizado de este proyecto tiene cuatro componentes principales: 1) Detección de puntos de referencia faciales, 2) extracción de características de textura facial, 3) clasificación de acciones faciales y 4) modelado de expresión de emociones. Además de los clasificadores de expresión de emociones y acciones faciales, el SDK tiene clasificadores para determinar el género y si la persona está usando anteojos. Estos clasificadores tienen dos modos de funcionamiento: estático y causal. Los clasificadores estáticos permiten la clasificación de imágenes individuales. Los clasificadores causales aprovechan la información temporal disponible en las secuencias de video para aumentar aún más la precisión de las medidas de expresión facial. En el artículo especifican que se recopiló vídeos de cientos de miles de personas que fueron codificados por expertos en el Sistema de Codificación de Acción Facial para proporcionar un rico conjunto de datos de ejemplos de expresiones faciales, este conjunto de datos se utilizó para el entrenamiento de los clasificadores, mientras que para las pruebas el sistema utilizó un

conjunto independiente de 10000 imágenes para verificar la generalización de algoritmos, en dicho conjunto no se controló ni la iluminación ni la pose de los participantes.

Por otra parte en el artículo (Skiendziel et al., 2019) de Tanja, Andreas y Oliver titulado “Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and Facial Action Coding System Scoring” nos presenta una validación del software FaceReader(FR) en su versión 7.0. FR fue desarrollado en la base de las teorías psicológicas de la emoción, específicamente la investigación de Ekman y el Sistema de Codificación de Acción Facial. El software permite la codificación de las emociones felicidad, tristeza, ira, disgusto, miedo, sorpresa y desprecio además de una expresión neutra, tiene como requisitos mínimos de entrada imágenes de 640 x 480 píxeles, en frontal, primer plano y la visualización completa de la cara, así como con iluminación frontal. El proceso de clasificación de las emociones faciales básicas en este software constan de tres pasos: 1) detección y localización de un rostro; 2) análisis paralelo a través de (a) un clasificador basado en 500 puntos clave en la cara y entrenado con más de 1000 expresiones faciales codificadas manualmente y (b) una red neuronal artificial profunda para patrones de reconocimiento; (3) integración de los resultados de los procesos de análisis paralelos en la clasificación de una emoción. En este estudio se centró el objetivo en la validación del FR como alternativa económica a la gestión de codificación del Sistema de Codificación de Acción Facial. La clasificación automatizada de emociones y la codificación de las Unidades de Acciones Faciales se evaluaron utilizando un conjunto de expresiones emocionales instruidas. Se descubrió que la clasificación de las emociones supera a la codificación de Unidades de Acción. Las Expresiones de emociones se clasificaron correctamente el 80 % de los casos, mientras que la validez agregada para la codificación de las Unidades de Acción estuvo en el rango medio.

Seguindo la investigación se encontró el artículo (Baltrušaitis et al., 2016) nombrado “OpenFace: An open source facial behavior analysis toolkit” donde exponen a OpenFace, una herramienta de código abierto destinado a la visión por computadora y el aprendizaje automático, este software detecta puntos de referencias faciales, estima la postura de la cabeza, reconoce unidades de acción facial y estima la mirada, además permite la detección de múltiples caras en una imagen y seguimiento de múltiples caras en vídeos. OpenFace se puede utilizar a través de su interfaz gráfica de usuario, línea de comandos y el sistema de mensajería en tiempo real horaria (basado en ZeroMQ). OpenFace utiliza el Conditional Lo-Cal Neural Fields (CLNF) para la detección de puntos de referencias faciales y de seguimiento, este realiza la detección de los 68 puntos de referencias faciales cuya capacidad se evaluó en el conjunto de datos de validación facial de 300W. Para la estimación de la postura de la cabeza se utilizaron tres conjuntos de datos disponible publicamente con datos de pose de cabeza (BU, Biwi y ICT-3DHP), mientras que para la estimación de la mirada se evaluó con el conjunto de datos MPII Gaze destinado a evaluar la mirada basada en la apariencia. Por ultimo se realizaron experimentos de reconocimiento de Unidades de Acción Facial con tres conjunto de datos (SEMAINE, DISFA y BP4D), esta ultima evaluación fue la única con mal resultado.

Jin Hyun Cheong en compañía de Xie, Byrne y Chang por su parte publicaron los resultados obtenidos de su investigación en el artículo (Cheong et al., 2021) llamado “Py-Feat: Python Facial Expression Analysis Toolbox”. Py-Feat es una caja de herramientas de Python de código abierto que proporciona soporte para detectar, preprocesar, analizar y visualizar datos de expresiones faciales. El proceso comienza con la grabación de fotografías faciales o vídeos usando un dispositivo de grabación

como cámaras web, videocámaras o cámaras 360, después al capturar la cara, los investigadores pueden usar Py-Feat para detectar rasgos faciales como puntos de referencia faciales, unidades de acción y emociones y verificar los resultados de la detección con superposiciones de imágenes y gráficos de barras. Los datos resultantes se pueden analizar dentro de la caja de herramientas utilizando métodos estadísticos como pruebas t, regresiones y correlaciones intersujetos. Las salidas de visualización pueden generar imágenes faciales a partir de modelos de acción y activaciones de unidades, para mostrar campos vectoriales que representan movimientos de referencia y mapas de calor de activaciones de los músculos faciales. La evaluación comparativa de los modelos de detección de emociones utilizados, se realizó en el conjunto de datos de AffectNet que contiene expresiones no propuestas de emociones, que ocurren naturalmente y no en un entorno de laboratorio cuidadosamente seleccionado. Se seleccionó un subconjunto aleatorio de 500 imágenes para cada una de las siete emociones diferentes para la evaluación comparativa y se calculó la puntuación F1 para cada categoría de emoción. El modelo 83 de la red de enmascaramiento residual logró la puntuación F1 más alta, seguida del modelo FACET-iMotions y los modelos de aprendizaje estadístico entrenados en las características del HOG. Basados en estos resultados, se establece el modelo de red de enmascaramiento residual como el modelo de detección de emociones predeterminado del Py-Feat.

Los antecedentes antes descritos establecen una relación directa con la investigación, puesto que plantean diferentes alternativas para la solución del problema que se presenta y sugieren buenas líneas de metodologías a seguir. Además comparan alternativas indagando en el mismo fin: mejores resultados y reducción de recursos computacionales.

1.7. Estructura del Documento

El documento se constituirá con VI capítulos detallados a continuación:

Capítulo I: da inicio con la definición de problema, continuando con los objetivos, la metodología que se utilizará, el alcance del trabajo, la justificación, los 5 antecedentes de investigación y la estructura del documento.

Capítulo II: se constituirá por el marco teórico, en donde se realizará la descripción detallada de los conceptos del trabajo, considerando el contexto de ejecución.

Capítulo III: se corresponde con la ingeniería de requisitos, detallando la especificación de requisitos del sistema y pasos que sigue la metodología empleada.

Capítulo IV: en este capítulo se procede al diseño del sistema de software y la especificación de algoritmos y técnicas para el sistema de reconocimiento y clasificación de emociones en personas.

Capítulo V: se detallará la implementación del sistema y expondrán los resultados de las pruebas aplicadas.

Capítulo VI: se presentarán las conclusiones generadas tras el proyecto, además de las propuestas de futuros trabajos y recomendaciones.

Capítulo 2

Marco Teórico

2.1. Inteligencia Artificial

La Inteligencia Artificial (IA) es considerado por muchos una disciplina de investigación de la informática que se centra en la creación de programas y mecanismos que pueden considerarse inteligentes (SalesForce Latinoamérica, 2017), ya que emulan características propias del cerebro humano (Benítez et al., 2014).

La IA es uno de los términos más generales de la informática y engloba al Aprendizaje Automático (en inglés Machine Learning), que a su vez engloba al Aprendizaje Profundo (en inglés Deep Learning). También incluye otras técnicas como los algoritmos de búsqueda, el razonamiento simbólico, el razonamiento lógico y la estadística.

Según Gallud (Gallud Baños, 2019) la IA sigue siendo por todo esto, un concepto ambiguo, y una posible definición sería: el esfuerzo por automatizar las tareas intelectuales que normalmente realizan los seres humanos.

2.2. Red Neuronal

Una Red Neuronal es un sistema de programas y estructuras de datos que se aproxima al funcionamiento del cerebro humano (Ahmed Banafa, 2016), las neuronas en lugar de estar dispuestas de forma arbitraria, están dispuestas en capas formando un grafo acíclico (Juan Zambrano, 2018). En (Pérez Estruch, 2017) Pérez explica que para un vector de entrada x ofrece un vector de salida distinta $o(x)$, constituida por tres tipos de capas:

- **Capa de entrada:** es el vector de datos de entrada antes de realizarse ninguna operación. Si se tiene un vector x de datos, cada valor desde $(x_1...x_n)$ constituye una neurona de entrada.
- **Capa oculta:** está formado por las neuronas ocultas y contiene todos los cálculos intermedios de la red. Cada neurona oculta contiene un peso y bias que son los elementos que modificarán los datos de entrada a “algo distinto”. El número y tamaño de capas ocultas varia y se determina según el tipo de necesidad.
- **Capa de salida:** es la capa en la que se realiza la clasificación y tiene tantas neuronas como posibles clases que presenta el problema asociado.

Al conocer la estructura y la división en capas es importante conocer el funcionamiento interno, el vector de datos en cada capa oculta se transforma a través de los pesos que las neuronas tienen entre si (conexiones) y, en conjunto a la función lineal presente, actúa como extractor de características. La capa de salida representa las posibles clases en las que se separa el conjunto de datos del problema.

Matemáticamente las transformaciones que se producen a través de la red se

modelan de la siguiente forma $h(x) = s(wx + b)$, donde b es el vector de bias, w es la matriz de pesos y s la función de activación.

La función de activación define el tipo de neurona con la que se trabaja, se utiliza para dar no linealidad al problema y por tanto que la red sea capaz de resolver problemas complejos. (Pellejero et al., 2017)

2.3. Aprendizaje Automático

El Aprendizaje Automático, conocido en inglés como Machine Learning, se refiere a un amplio conjunto de técnicas informáticas que permiten dar a una máquina la capacidad de aprender sin ser explícitamente programada (Gallud Baños, 2019), estudiando el reconocimiento de patrones y el aprendizaje por parte de las computadoras.

Considerando que una de las descripciones principales del Aprendizaje Automático es el imitar la capacidad del ser humano para procesar la información y así poder crear modelos que generalicen la información obtenida para realizar predicciones (Saez De La Pascua, 2019), podemos conseguir que el Aprendizaje Automático es aplicado en diferentes tareas, tanto generales como específicas, y en un sin fin de áreas que van desde nuestro hogar, hasta el campo industrial.

Esta área de la IA se trabaja con algoritmos que aplicados a buenos conjuntos de datos pueden realizar hallazgos sorprendentes, con muy poca intervención del ser humano, son muchas categorías en las que podemos clasificar estos algoritmos, sin embargo la más aceptada es: aprendizaje supervisado y aprendizaje no supervisado.

2.3.1. Aprendizaje Supervisado

Este tipo de aprendizaje consiste en hacer predicciones a futuro, basadas en comportamientos o características que se han visto en un conjunto de datos ya almacenado (Juan Zambrano, 2018). Para obtener estas predicciones se debe entrenar un sistema con el conjunto de datos ya etiquetados que servirá como entrada. Conociendo las entradas (características) y también sus correspondientes salidas (etiquetas), se van corrigiendo los parámetros de la función designada para el sistema, por lo general definida como $y = f(x)$, en un proceso iterativo. Tras el entrenamiento se predice con una determinada precisión las etiquetas de nuevos datos de entrada cuyos valores de salida sean desconocidos (Gallud Baños, 2019).

2.3.2. Aprendizaje no Supervisado

Este aprendizaje trabaja con sistemas que no cuentan con un conjunto de datos clasificados, su función es la agrupación, por lo que el algoritmo debería catalogar por similitud y poder crear grupos, sin tener la capacidad de definir cómo es la individualidad de cada uno de los integrantes del grupo (Juan Zambrano, 2018).

2.4. Aprendizaje Profundo

El Aprendizaje Profundo, conocido en inglés como Deep Learning, es un concepto muy amplio, incluso conlleva a que no tenga una única definición veraz (Deng and Yu, 2014), aunque, se generaliza que son todas aquellas técnicas de IA que hacen uso de arquitecturas de Redes Neuronales. Mientras que Restrepo (Restrepo Arteaga et al., 2015) por su parte indica que el aprendizaje profundo es un concepto que surge de la idea de imitar el cerebro a partir del uso de hardware y software, para crear una IA pura, utilizando una capacidad de abstracción jerárquica.

El Aprendizaje Profundo también se puede definir como un aspecto de la IA que se ocupa de emular el enfoque de aprendizaje que utilizan los humanos para obtener conocimiento. Sáez (Saez De La Pascua, 2019) indica que puede considerarse una forma de automatizar el análisis predictivo.

Los algoritmos de Aprendizaje Profundo se apilan en una jerarquía de creciente complejidad y extracción (Saez De La Pascua, 2019), de acá deriva el termino de “Profundo”. Estos algoritmos están englobados dentro del Aprendizaje Automático, cada uno en la jerarquía aplica una transformación no lineal en sus datos de entrada y utiliza lo que aprende para crear un modelo estadístico como salida. Las iteraciones continúan hasta que la salida ha alcanzado un nivel de precisión aceptable. A diferencia del Aprendizaje Automático, en el Aprendizaje Profundo el modelo aprende por sí mismo y puede así descubrir las relaciones entre las variables, evitando la intervención del ser humano.

2.4.1. Red Neuronal Prealimentada

Conocida también como Perceptrones Multicapa (MLP), es un algoritmo que consiste en una serie de unidades de procesamiento simples similares a neuronas, organizadas en capas y cada unidad de una capa está conectada con todas las unidades de la capa anterior. Estas conexiones no son todas iguales, ya que cada conexión puede tener una fuerza o peso diferente. Los pesos de estas conexiones codifican el conocimiento de una red. A menudo, las unidades de una Red Neuronal se denominan nodos (Senthilkumar, 2010).

Los datos ingresan por la entrada y pasan a través de la red, capa por capa, hasta llegar a la salida, como se muestra en la figura 2.1. La capa de entrada tiene solo las entradas a la red, continua con una capa oculta, que consta de cualquier número de

neuronas o unidades ocultas colocadas en paralelo. Cada neurona realiza una suma ponderada de las entradas (ecuación de la figura 2.1), que luego pasa a una función de transferencia/activación (Senthilkumar, 2010).

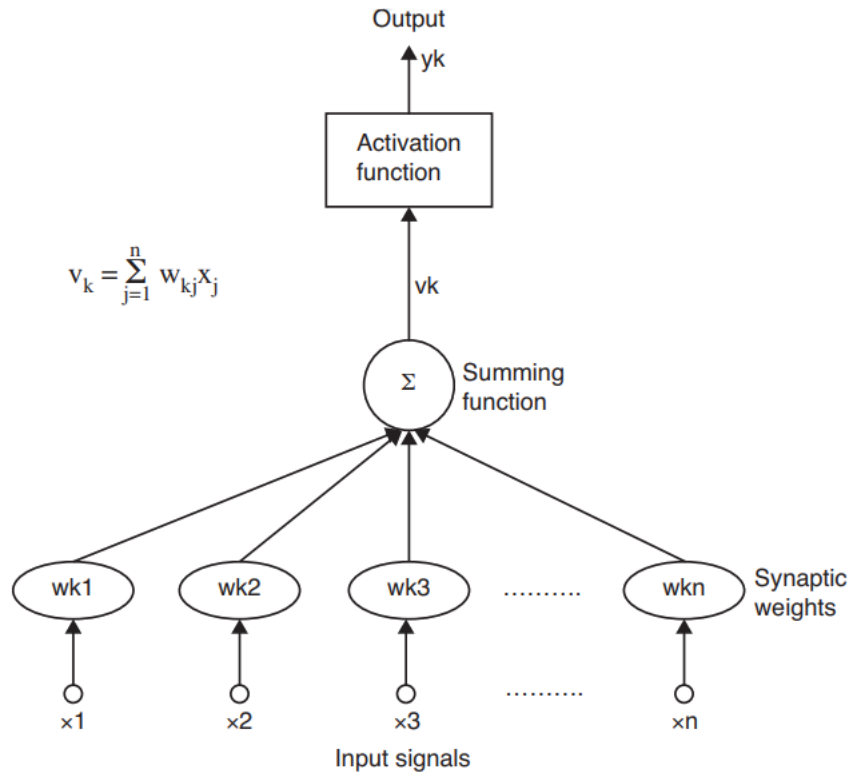


Figura 2.1: Representación matemática de una Red Neuronal Prealimentada (Senthilkumar, 2010).

2.4.2. Redes Neuronales Convolucionales

Las CNN, conocidas en inglés como Convolutional Neural Network, son redes muy efectivas para tareas de detección y clasificación de imágenes, estas técnicas hacen uso de Redes Neuronales (Saez De La Pascua, 2019). Una CNN es una modificación de una Red Neuronal formadas de pesos y sesgos que se pueden aprender y con las mismas capas de una Red Neuronal básica, adicional poseen capas ocultas que están especializadas en dos operaciones: convolución y agrupación (Saez De La Pascua, 2019).

2.4.2.1. Capa de convolución

En (Saez De La Pascua, 2019), el autor define una capa convolucional como capas ocultas que se encargan de aprender patrones locales en ventanas pequeñas de dos dimensiones, alto por ancho. Estas capas sirven para detectar características visuales en las imágenes, y además aprenden una propiedad de la imagen en un punto específico, siendo capaz de reconocerla en cualquier lugar de la propia imagen.

Las CNN's operan sobre 3 ejes: ancho, alto y otro eje que sería el canal (podría ser de 1 si la imagen de entrada es en escala de grises o bien 3 para una imagen en color RGB).

Las capas convolucionales son capaces de aprender diferentes elementos cada vez más complejos, en función de lo aprendido en la capa anterior, aunque esto no implique que todas las neuronas de entradas estarán conectadas con todas la neuronas de una capa convolucional.

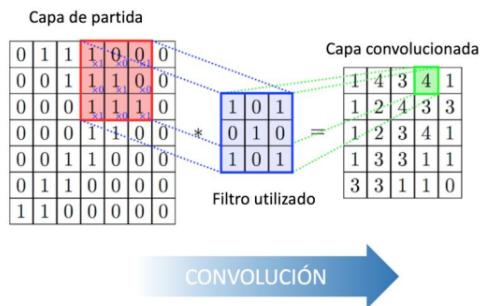


Figura 2.2: Convolución, mapas de características (Diego Calvo, 2017)

Cuando existe una conexión entre ellas se utiliza la matriz de Kernal (ver figura 2.2) que recorre todas las neuronas de entrada. Por cada posición de la matriz hay una neurona. Se inicia el recorrido por la parte superior izquierda y se desplaza hacia la derecha. Cuando ha acabado una fila, sigue con la fila de abajo.

Con este método se puede observar que se define un espacio de neuronas a la salida inferior al de la entrada, para ser exactos se reduce a: $(Neuronas\ alto * Neuronas\ alto) Salida = (Neuronas\ ancho - ancho\ matriz - 1) * (Neuronas\ alto - alto\ matriz - 1) Entrada$. Esto puede ocasionar pérdida de información, por lo que en las Redes Neuronales se aplica una técnica de relleno de ceros para no perder neuronas.

En esta capa de convolución se aplica una serie de filtros, donde cada filtro detecta una propiedad de la imagen. Por lo que a la salida se podrá observar un tamaño de: $Neuronas\ Salida = Neuronas\ Ancho * Neuronas\ Alto * Filtros$.

2.4.2.2. Capa de agrupación

Las capas de agrupación también llamadas capas de reducción, agrupan las neuronas en grupos (Saez De La Pascua, 2019), para ayudar a obtener las propiedades más significativas de los datos de entrada y obviar los datos que pueden inducir a un error.

Según (Saez De La Pascua, 2019) existen varias formas de agrupación, una de ellas es la agrupación promedio, la cual obtiene el valor promedio del grupo de neuronas. También existe la agrupación por valor máximo donde se escoge el valor máximo de cada grupo.

En las capas de agrupaciones la entrada se ve reducida, Saéz de la Pascua (Saez De La Pascua, 2019) indica que se reduce en función del tamaño de las agrupaciones. Si los grupos son de un tamaño $2 * 2$ como se ve en la Figura 2.3, en la salida obtendremos la mitad del tamaño de los datos de entrada. Estas agrupaciones se aplican tanta veces como profundidad tengan los datos, el cual viene determinado por el número de filtros

aplicados en la capa anterior.

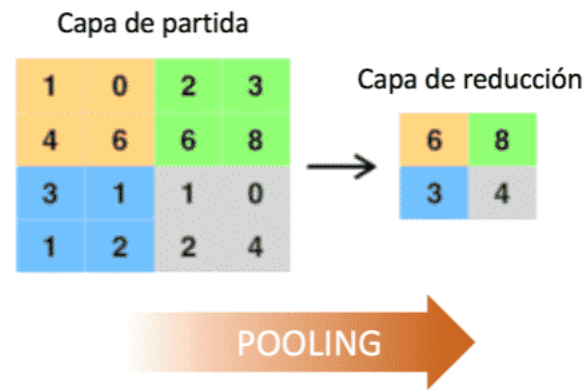


Figura 2.3: Reducción en capa de agrupación (Diego Calvo, 2017)

2.5. Reconocimiento de Emociones Faciales

El cerebro humano es un complejo órgano, que durante años ha sido fuente de investigación para científicos, psicólogos y un sin fin de investigadores, en particular los fisiólogos. Estos últimos han explorado las funciones del cerebro limitando sus intentos de explicación a las ejecuciones cognitivas y volitivas de éste, ignorando la esfera estética de la mente, sus anhelos, sus placeres y penas, y sus emociones, a tal punto que si se les solicitara una teoría de dichos hechos mentales basada en términos del cerebro, muchos responderían o bien que todavía no se ha reflexionado sobre el tema o bien que su complejidad no ha permitido fijar hipótesis precisas.

Sin embargo, lo concerniente a las emociones hace pensar de manera natural que la percepción mental de algún hecho provoca la disposición mental llamada emoción, y que este estado mental da lugar a la expresión corporal (James, 1985), es por ello que Rosetti (Rosetti, 2017) expresa: no somos seres racionales, somos seres emocionales que razonan, afirmando con su analogía que la evolución de las estructuras cerebrales se asemejan a la construcción de un edificio. Las funciones más elementales y básicas (las emociones) aparecen en la especie humana al igual que la planta baja de un edificio, mucho antes que el piso número diez. Solo después y con la lentitud del tiempo de la evolución, la especie desarrolla el sistema nervioso central, con nuevas estructuras cerebrales que agregan funciones más complejas. A pesar de esto James (James, 1985) expone todo lo contrario, considerando que los cambios corporales siguen directamente a la percepción del hecho desencadenante y que nuestra sensación de esos cambios, según se van produciendo, es la emoción.

Disconforme a cualquiera de las definiciones sobre la emoción, es claro que el

ser humano puede expresar a través de diferentes formas lo que siente al experimentar cualquiera de ellas, según (Iglesias-Hoyos et al., 2016) estos medios incluyen componentes fisiológicos y conductuales, como lo son los cambios faciales, gestuales y del lenguaje, siendo los cambios faciales los que se resaltaran en esta investigación.

El cerebro humano cuenta con una capacidad para extraer diferentes tipos de información del rostro de otras personas; por ejemplo, se puede identificar el estado emocional, determinar si una persona es atractiva a la vista de quien la percibe, o incluso determinar la identidad de la persona observada. El rostro ha resultado ser una de las claves más distintivas y ampliamente utilizadas por las personas, convirtiéndola en la parte del cuerpo humano que más cerca y detalladamente se observa durante cualquier tipo de interacción social, considerándolo así el principal sistema de señales para mostrar las emociones (Martínez et al., 2010; Caballo, 1993).

Las expresiones faciales de emociones, junto con la mirada, es el medio más rico e importante para expresar emociones y estados de ánimo además de resultar de gran importancia cuando se trata de la interacción con otras personas, esto debido a que revelan estados mentales complejos, los cuales son transmitidos a los otros. La imprecisión o fallo en la interpretación correcta de información emocional, pueden ser causantes de conflictos interpersonales y sociales (Arango et al., 2013; Fasel and Luetttin, 2003; Ekman and Oster, 1979).

El reconocimiento facial de las expresiones emocionales se define como la capacidad de todos los individuos de reconocer formas básicas de expresión afectiva, la cual se muestra en los rostros de las personas (Russell, 1994). El desarrollo de esta capacidad suele ser un proceso gradual, el cual se ve afectado por múltiples factores que van

desde la edad hasta el contexto en donde las personas se desenvuelven, situaciones de desarrollo cognitivo y neuronal entre otras (Martínez et al., 2010). El reconocimiento facial de emociones consta de 3 pasos: detección de rostro; extracción de características faciales; y clasificación de expresión facial (Wati and Abadianto, 2017), ver figura 2.4. El conjunto de expresiones faciales que se detectan, forman como conjunto, expresiones afectivas, estas constituyen 6 emociones básicas universales e innatas, que fueron definidas por Ekman (Ekman et al., 1999) cada una de ellas con expresiones faciales diferentes. La frente/cejas, ojos/párpados y la parte inferior de la cara (boca) son responsable de su manifestación (Caballo, 1993; Ekman and Friesen, 1976, 1971; Ekman et al., 1969).

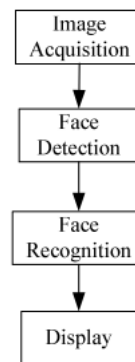


Figura 2.4: Pasos en el sistema de reconocimiento facial (Wati and Abadianto, 2017)

2.5.1. Sistema de código de acción facial (FACS)

Conocido en inglés como el Facial Action Coding System (FACS), fue desarrollado por Paul Ekman, Wallace V. Friesen y Joseph C. Hager, como guía para reconocer la clasificación de expresiones de emociones a partir de gestos faciales (ver tabla A.1), Ekman clasifica las expresiones en Unidades de Acción específica, Action Units (AU), que son acciones fundamentales que realizan los músculos individualmente o grupos de músculos que se concentran en un área específica para poder ejecutar la acción de

expresión emocional y los segmentos temporales que se pueden producir dentro de una expresión (Mallitasig Arellano, 2018). Existe una lista de AUs (ver tabla A.2), que no se encuentran enumeradas secuencialmente (por ejemplo la AU 3 no existe), dichos AUs combinados entre ellos permiten la identificación de cualquiera de las 6 emociones básicas, a continuación se detallan ciertos conceptos que sirven, junto a la lista de AUs, para la reconocimiento de las emociones:

- **Unilateralidad y Bilateralidad:** se refiere a cuando se realizan los gestos con un lado de la cara o con los dos. No se debe confundir la unilateralidad, hay microexpresiones que se reflejan más en un lado que en otro, esto quiere decir que si se nota un atisbo de bilateralidad, posiblemente se marque como ella. La unilateralidad viene reflejada por el cerebro. El hemisferio derecho del cerebro controla las emociones, mientras que el izquierdo la lógica, el hemisferio izquierdo controla el lado derecho del cuerpo humano y viceversa. Por tal se tiene que el lado derecho va a tener la cara controlada por la razón, o la que el ser humano quiere que vean, mientras que en la parte izquierda se tendrá la cara emocional, la que seguramente diga la verdad.
- **La intensidad:** es muy importante determinar la intensidad de la microexpresión, ya que cuanto más intensa sea, más se sentirá esa emoción. Ekman determinó ciertos grados de intensidad en las AUs: A,B,C,D y E (ver tabla A.5).
- **Códigos de Visibilidad:** simplemente son códigos para indicar que ciertas partes de la cara, o la cara entera, no son visibles, o no se pueden determinar por cualquier motivo (la cabeza puede estar girada, agachada, puede no haber luz, entre otros). No les precede ninguna letra como AU, AD o M (ver tabla A.4).
- **Códigos de comportamientos:** proveen información adicional que puede ayudar a las AUs que se catalogan (ver tabla A.6).

2.6. Tecnologías

La mayoría de herramientas y bibliotecas para la clasificación de imágenes están escritas en Python, por lo que se hará referencia a este lenguaje y a las bibliotecas que se utilizaron. Debido a los inconvenientes presentados al utilizar este lenguaje, referente a incompatibilidad entre versiones (las versiones de las herramientas se detallan en la Tabla 2.1), y además de la exigencia de recursos al trabajar con una CNN, se inclinó por utilizar herramientas on-line, como Google Colab. Por ultimo se expondrá los Frameworks utilizados en el desarrollo y pruebas del sistema.

Tecnologías	
Python	3.7.12
TensorFlow	2.6
TensorFlowjs	3.9
Numpy	1.19.5
Vue.js	2.6
Vuetify	2.3

Tabla 2.1: Versiones de Herramientas y tecnologías usada en el desarrollo.

2.6.1. Python

Python es un lenguaje de programación orientado a objetos, con facilidades imperativas y funcionales, por lo que se considera un lenguaje multi-paradigmas. Esta basado en el lenguaje ABC e influenciado por C, Algol 60, Modula-3 e Icon (Challenger-Pérez et al., 2014).

2.6.2. Tensorflow

Es un sistema de Aprendizaje Automático y de código abierto, desarrollado por Google que funciona a gran escala y en entornos heterogéneos. Tensorflow tiene una

arquitectura flexible donde se asigna los nodos de un gráfico de flujo de datos en muchas máquinas de un clúster y dentro de una máquina en varios dispositivos computacionales, incluidas CPU multinúcleo, GPU de uso general y ASIC de diseño personalizado conocidos como Unidades de procesamiento de tensor (TPU) (y Paul Barham y Jianmin Chen y Zhifeng Chen y Andy Davis y Jeffrey Dean y Matthieu Devin y Sanjay Ghemawat y Geoffrey Irving y Michael Isard y Manjunath Kudlur y Josh Levenberg y Rajat Monga y Sherry Moore y Derek G. Murray y Benoit Steiner y Paul Tucker y Vijay Vasudevan y Pete Warden y Martin Wicke y Yuan Yu y Xiaoqiang Zheng, 2016).

2.6.3. TensorBoard

Es una biblioteca de código abierto para Aprendizaje Automático, desarrollado por Google. TensorBoard proporciona la visualización y las herramientas necesarias para experimentar junto a la biblioteca TensorFlow (Google, b).

2.6.4. Keras

Es una biblioteca de Redes Neuronales escrita en Python, capaz de ejecutarse en Tensorflow, CNTK o Theano. Caracterizada por un fácil y rápido prototipado, gracias a su modularidad, minimalismo y extensibilidad (Antona Cortés, 2017).

2.6.5. Numpy

Término reducido de Numerical Python, es una biblioteca de Python de código abierto para informática científica. Permite trabajar con arreglos y matrices de forma natural, debido a la larga lista de funciones matemáticas útiles que contiene, incluidas algunas para álgebra lineal, transformación de Fourier y rutinas de generación de números aleatorios (Idris, 2013).

2.6.6. Matplotlib

Es una biblioteca de Python para trazado 2D que genera gráficos con calidad de producción. Admite el trazado interactivo y no interactivo, y puede guardar imágenes en varios formatos de salida. Proporciona una amplia variedad de tipos de gráficos, siendo altamente personalizable, flexible y fácil de usar (Tosi, 2009).

2.6.7. Google Colaboratory

También llamado Colab, es una herramienta de Google que permite escribir y ejecutar código de Python en un navegador sin configuraciones previas, acceso gratuito a GPU y TPU, y facilidad para compartir (Google, a).

2.6.8. Frameworks

Es un conjunto de conocimientos y técnicas cuya aplicación permite la utilización racional de los materiales y de los recursos naturales, mediante invenciones, construcciones u otras realizaciones provechosas para el hombre. (Degiovannini, 2011).

En la programación es definido como un entorno de trabajo para el desarrollo de aplicaciones, sean web o de escritorio, que ofrecen componentes, como bibliotecas de funciones, uso de plantillas, administración de recursos en tiempo de ejecución y reutilización de código.

2.6.8.1. Vue.js

Es un Framework progresivo creado por Evan You de código abierto, desarrollado para la construcción de interfaces de usuario. La biblioteca principal se centra en la capa de vista, siendo perfectamente capaz de impulsar aplicaciones sofisticadas de una

sola página (Sharo, 2020).

2.6.8.2. Vuetify

Vuetify es un framework de interfaz de usuario de Vue.js. Esta diseñado desde cero con cientos de componentes prediseñados bajo la especificación de Material Design (LLC, 2020).

2.6.8.3. Jest.js

Es un framework de prueba adaptable a cualquier biblioteca o framework de Javascript, diseñado por Christoph Nakazawa con código abierto, mayormente aplicado en testing de aplicaciones web (Docusaurus, 2021).

Capítulo 3

Ingeniería de Requisitos

La Ingeniería de Requisitos (IR) es un proceso clave para desarrollar un sistema de software de calidad, debido a que en este proceso se analiza lo que debe hacer una aplicación para satisfacer necesidades de información y automatización, convirtiendo cada requisito “en una propiedad que debe exhibir (cumplir o satisfacer) un sistema de software desarrollado o adaptado para resolver un problema particular” (Sawyer P., 2001).

Considerando la metodología planteada para el desarrollo del sistema de reconocimiento presentado en este Proyecto de Grado, este capítulo exhibe de manera resumida lo que concierne al primer proceso. Previamente a la IR se realiza una descripción general del sistema y de las funciones que deben cumplir sus componentes, continuando con la especificación de requisitos, donde se definen los requisitos funcionales y no funcionales del sistema a desarrollar.

3.1. Descripción general del Sistema de Reconocimiento y Clasificación de Emociones Faciales

El Sistema de Reconocimiento y Clasificación de Emociones Faciales es un sistema complejo de IA orientado a la actividad humana de reconocimiento de emociones. Este sistema está dividido, de acuerdo a sus tareas principales, en dos componentes: el componente de pre-procesamiento de datos y el componente de reconocimiento y clasificación, entendiéndose cada uno de ellos como el módulo donde se normaliza el conjunto de entrada del siguiente componente, y el módulo donde se reconocerá y clasificará cada emoción, respectivamente. El usuario de estos componentes puede ser uno o más, con dos roles claves, el primer rol quien ejecutará el papel del sujeto al cual se le clasifica sus emociones, identificado como evaluado y el segundo rol quien percibirá los resultados del análisis de sentimientos, llamado evaluador (para el caso de estudio se trabajará como lo estipula la Jerarquía de usuario de la figura 3.1).

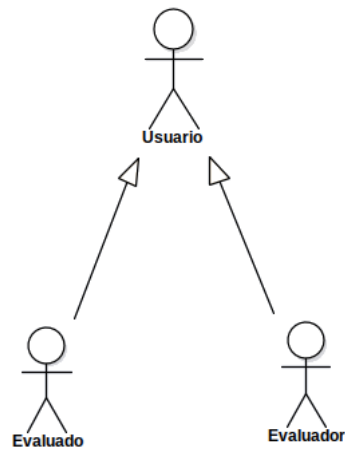


Figura 3.1: Jerarquía de Usuario del Sistema

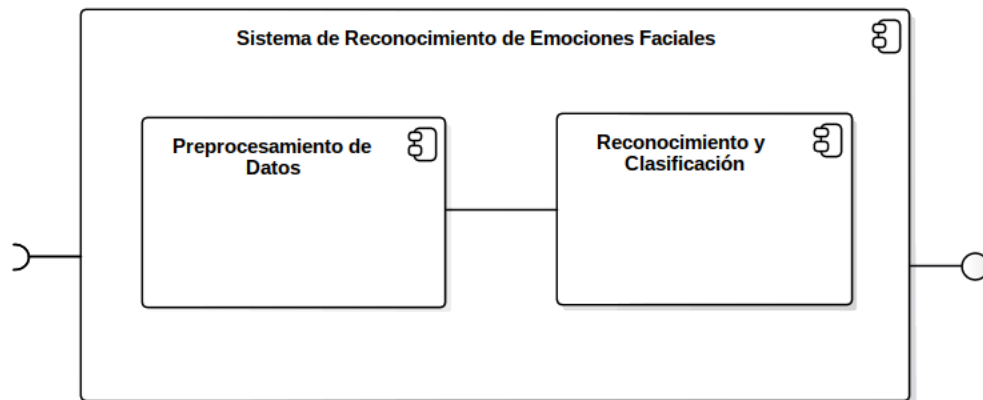


Figura 3.2: Modelo de Componente del Sistema de Reconocimiento y Clasificación de Emociones Faciales

3.1.1. Funciones del Sistema de Reconocimiento y Clasificación de Emociones Faciales

El Sistema de Reconocimiento se encarga de la identificación de las emociones faciales de una persona, según las 6 emociones básicas universales (felicidad, rabia, tristeza, sorpresa, miedo, asco) y la expresión neutral, a continuación se detallan las funciones de cada módulo que se desarrolla.

3.1.1.1. Componente de Pre-procesamiento de Datos

Este componente debe realizar el pre-procesamiento de los datos que alimentan el siguiente modulo, el componente de pre-procesamiento debe normalizar los conjuntos, de manera que permita en lo posible controlar la iluminación, seleccionar solo el área a evaluar (rostro), cambiar tamaño y modificar la escala de colores a grises.

3.1.1.2. Componente de Reconocimiento y Clasificación

La función principal del sistema la ejerce este componente, puesto que a través de este, se pretende reconocer los AU's del rostro humano que luego son clasificadas a través de técnicas y algoritmos de IA, entre las 7 emociones.

3.1.1.3. Interfaz

El Sistema de Reconocimiento y Clasificación de Emociones Faciales maneja una interfaz, la cual unifica los componentes antes mencionados, además de que permite a los usuarios el ingreso del vídeo en tiempo real para su análisis, y a su vez señala al segundo usuario qué emoción se esta presentando en el vídeo.

3.2. Especificación de Requisitos

La especificación de requisitos consiste en determinar las necesidades que debe cubrir el espacio de la solución al problema, así como las restricciones que se planteen en el mismo. En la especificación de requisitos se debe describir con claridad, sin ambigüedad y de manera precisa, cuál debe ser el comportamiento del sistema para brindar solución. Para el Sistema de Reconocimiento y Clasificación de Emociones Faciales se expondrán los requisitos funcionales de usuarios, los cuales se detallarán utilizando los Casos de Usos, además se muestran los requisitos no funcionales del software.

3.2.1. Requisitos Funcionales

Los requisitos funcionales, describen una actividad, comportamiento o función que el sistema deba realizar (Méndez Rosales, 2021), determinando así la funcionalidad del software. Para el sistema de este proyecto se especificaron tres requisitos funcionales: gestión de entrada de datos, gestión de reconocimiento y clasificación de emociones

faciales y gestión de salida del análisis, cada uno es descrito mediante casos de uso y sus respectivos diagramas.

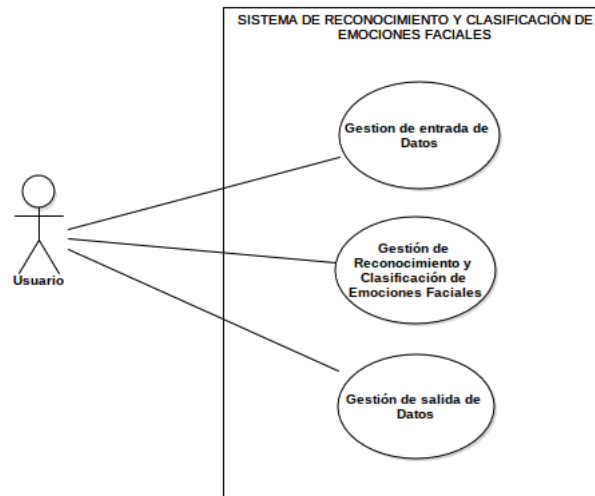


Figura 3.3: Diagrama General del Sistema de Reconocimiento y Clasificación de Emociones Faciales

3.2.1.1. Gestión de Entrada de Datos

Esta enfocado en la funcionalidad para dar inicio al sistema y enviar datos de entrada, el usuario en el rol del Evaluado puede ingresar al sistema como se observa en su Diagrama de Caso de Uso (ver figura 3.4), posterior a ello debe cargar los datos de vídeo en tiempo real a través de la cámara del dispositivo desde dónde ingrese. El ingreso al sistema no será restringido, al igual que la carga de datos. La descripción de los Casos de Uso se pueden observar en las Tabla 3.1 y 3.2

Usuarios:	evaluador.
Condiciones de entrada:	evaluador
Condiciones de salida:	el usuario ingresa exitosamente al sistema.
Flujo básico:	<ol style="list-style-type: none"> 1. El usuario selecciona el botón →, para ingresar. 2. El sistema solicita permiso para iniciar la cámara en el dispositivo. 3. El usuario accede a los permisos.
Flujo alternativo:	<ol style="list-style-type: none"> 2.1 El sistema no solicita permisos porque previamente ya fueron otorgados o el dispositivo no requiere darlos. 3.1 El usuario no accede a los permisos, el sistema no abre la cámara del dispositivo por lo que no se puede continuar con el proceso.
Requisitos especiales:	la activación de la cámara es necesaria..

Tabla 3.1: *CU1*: Ingresar al Sistema.

Usuarios:	evaluador.
Condiciones de entrada:	cámara del dispositivo activa.
Condiciones de salida:	el usuario brinda el vídeo tiempo real al sistema.
Flujo básico:	<ol style="list-style-type: none"> 1. El usuario espera la carga del sistema. 2. El sistema recepciona el vídeo en tiempo real.
Flujo alternativo:	1.2 El usuario interrumpe la carga del sistema, el proceso no continua.
Requisitos especiales:	ninguno.

Tabla 3.2: *CU2*: cargar datos.

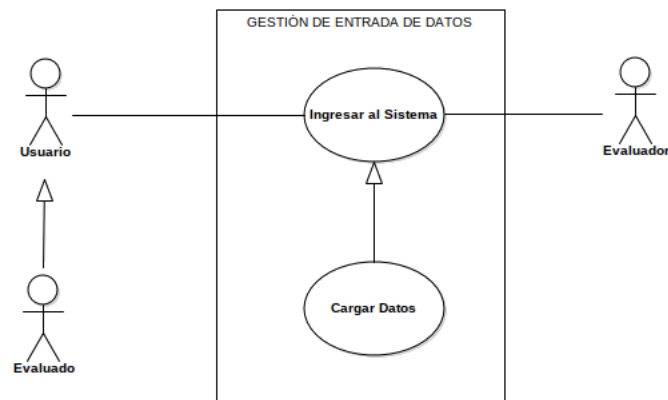


Figura 3.4: Diagrama de la Gestión de Entrada de Datos

3.2.1.2. Gestión de Reconocimiento y Clasificación de Emociones Faciales

Una vez ingresado el vídeo se debe iniciar la extracción de los datos necesarios para la clasificación de emociones, teóricamente se debe proceder a reconocer el rostro humano de dónde se extraerán los AU's expresadas en el rostro. Para que se pueda realizar una buena identificación los datos de entrada deben ser normalizados acorde a servirle al sistema. Todos los datos extraídos serán utilizados por el sistema para la clasificación de la emoción, en el rango de las 6 emociones básicas universales (felicidad, tristeza, miedo, sorpresa, ira, asco) y la expresión neutral. Los pasos sugeridos para este requerimiento se detallan en la imagen 3.5 y las tablas 3.3, 3.4.

Usuarios:	evaluado y evaluador (ambos no de forma directamente)
Condiciones de entrada:	el evaluado debe ingresar el video en tiempo real.
Condiciones de salida:	el sistema extrae los AU's del rostro humano identificado.
Flujo básico:	<ol style="list-style-type: none"> 1. El sistema reconoce el Rostro humano. 2. El sistema identifica los AU's. 3. El sistema retorna el conjunto de AU's identificados.
Flujo alternativo:	1.1 No se identifica la imagen o no es un rostro humano, el evaluado debe ingresar nuevamente la data de entrada.
Requisitos especiales:	gestión de entrada de datos.

Tabla 3.3: *CU3*: reconocer los AU's.

Usuarios:	evaluado y evaluador (ambos no de forma directamente)
Condiciones de entrada:	AU's.
Condiciones de salida:	Análisis de las emociones extraída.
Flujo básico:	<ol style="list-style-type: none"> 1. El sistema recibe los AU's e inicia la comparación con las 7 emociones. 2. El sistema identifica las emociones del rostro de entrada. 3. El sistema envía el resultado de más alto porcentaje en precisión.
Flujo alternativo:	
Requisitos especiales:	AU's.

Tabla 3.4: *CU4*: clasificación de emociones.

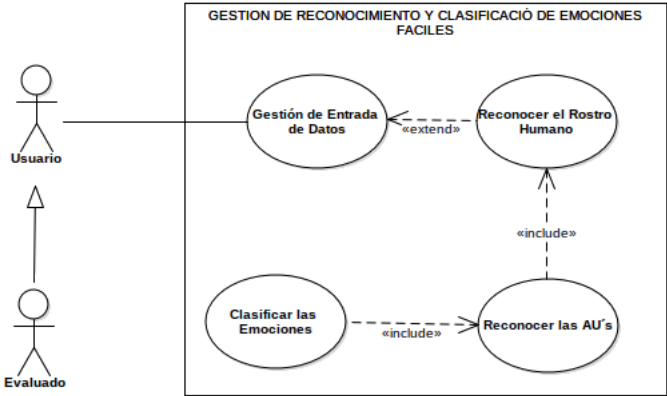


Figura 3.5: Diagrama de Gestión de Reconocimiento y Clasificación de Emociones Faciales

3.2.1.3. Gestión de Salida de Análisis

La gestión de salida del análisis esta enfocada en la forma en que visualizará los resultados el evaluador, este debe obtener una muestra de la identificación del rostro y la emoción clasificada para ese momento, además debe presentar el porcentaje de seguridad de la emoción identificada. Si el sistema no arroja ningún resultado o falla durante el análisis se le notificará al evaluador. El caso de uso de este requisito se detalla en la figura 3.6 y la tabla 3.5.

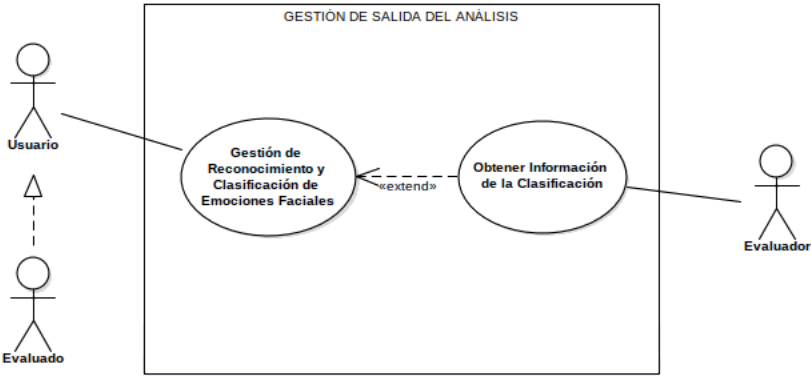


Figura 3.6: Diagrama de Gestión de Salida de Análisis

Usuarios:	evaluador.
Condiciones de entrada:	análisis de reconocimiento.
Condiciones de salida:	el evaluador debe recibir los resultados.
Flujo básico:	1. El evaluador, mediante el sistema, observa los resultados del análisis obtenido. 2. El evaluador presiona el botón atrás para ir al inicio del sistema y finalizar con la clasificación.
Flujo alternativo:	
Requisitos especiales:	Gestión de Reconocimiento y Clasificación de Emociones Faciales..

Tabla 3.5: *CU5*: obtener información.

3.2.2. Requisitos no Funcionales

Los requisitos no funcionales no se relacionan directamente con el comportamiento de la aplicación, sino que se encargan de restringir el diseño de la aplicación, además de definir la calidad que el software debe tener.

1. El sistema debe contener una interfaz gráfica con mínima complejidad para la entrada y salida de datos.
2. El sistema debe tener una confiabilidad igual o mayor al 50 %.
3. Los datos recopilados por el sistema deben ser estrictamente autorizados por voluntad propia del evaluado.
4. La salida del sistema debe presentar datos porcentuales por emoción.

Capítulo 4

Diseño

El capítulo anterior presentó la IR del Sistema de Reconocimiento y Clasificación de Emociones Faciales (SREF), el cual permitió la claridad del problema, identificando los componente principales, objetivos y actores del mismo. Basándose en este análisis se presenta en detalle la arquitectura del sistema, la cual esta ligado con las metas establecidas del funcionamiento del mismo.

El SREF funciona con dos componentes principales, uno de ellos (el componente de Reconocimiento y Clasificación) se subdivide en dos fases, cada una trabaja con algoritmos de Aprendizaje Profundo, por lo que se precisa la revisión de los algoritmos seleccionados y adicionalmente se justificará la selección de los mismos.

Para el entrenamiento de estos algoritmos se requirió la selección de datos, por lo que se presentan 2 bases de datos con diferentes criterios de selección, una de ella aplicada solo para el reconocimiento de emociones en el caso de estudio.

4.1. Diseño del Sistema de Reconocimiento y Clasificación de Emociones Faciales

El diseño de un sistema se encarga de realizar el modelado del mismo, desde la identificación de los subsistemas que conforman su estructura, hasta llegar a las interfaces que visualizarán los usuarios. La definición de la arquitectura, los módulos que la componen y el modelado de los datos, son fundamentos vitales de un buen diseño.

4.1.1. Diseño Arquitectónico del Sistema

La arquitectura es el modelo conceptual que comprende las principales propiedades físicas, el estilo, la estructura, las interacciones y el objetivo de un sistema (Hatley et al., 2013). En la figura 4.1, se observa el diagrama de despliegue que regirá el desarrollo de esta investigación, el cual describe una arquitectura de dos niveles y tres capas, basada en el patrón de diseño Modelo Vista Vista- Modelo (MVVM). El MVVM estructura de manera eficaz todos los componentes de la aplicación, la capa Vista-Modelo actualizará automáticamente los cambios en las otras dos capas, es decir, si se produce un cambio en la Vista, los datos son actualizados de forma automática en la parte del Modelo, y viceversa. El entorno de desarrollo *Vue.js* está centrado en el componente Vista-Modelo del patrón MVVM Herrera Carrillo (2020).

4.1.2. Diseño de Subsistemas

Un subsistema puede definirse como un conjunto de partes e interrelaciones estructurales y funcionales que forman parte del sistema (Méndez Rosales, 2021), la identificación de los subsistemas es de utilidad para identificar las partes que componen al software, permitiendo un estudio más detallado por separado. El Sistema en

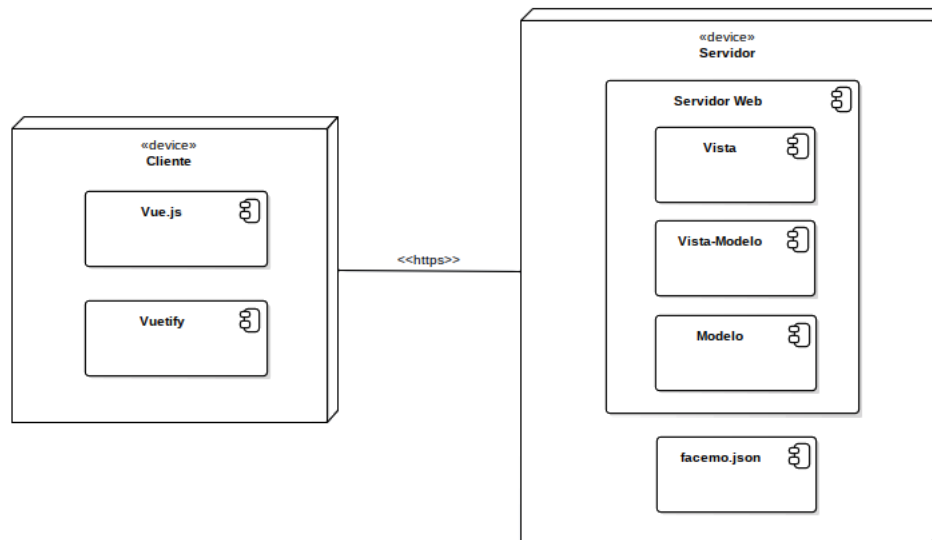
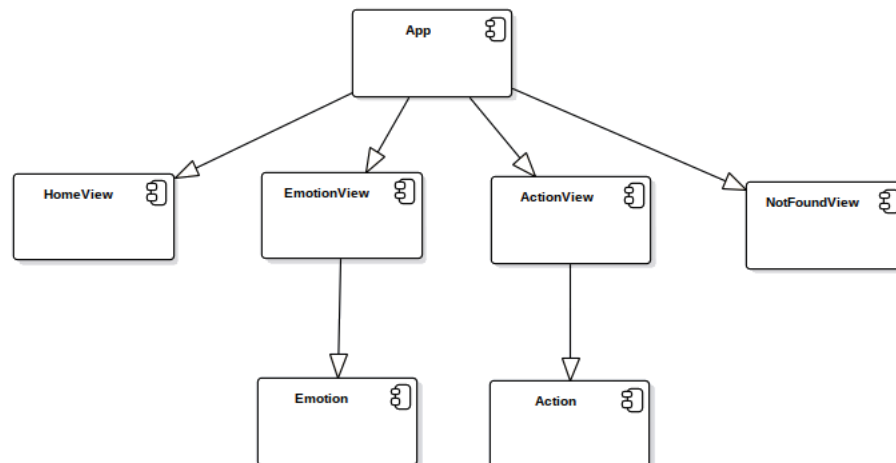


Figura 4.1: Diagrama de Despliegue del Sistema

general fue implementado mediante los Frameworks *Vue.js* y *Vuetify*, identificando (de acuerdo al análisis de los casos de usos expuestos en la IR) dos subsistemas o componentes generales, que se presentarán haciendo énfasis en su diseño con los Frameworks nombrados, de ser el caso se expandirá la explicación de su implementación si se utiliza otra tecnología.

Figura 4.2: Árbol de componentes del sistema *Vue.js*

4.1.2.1. Pre-procesamiento

Es el primer proceso que se aplica a las muestras de vídeos de entrada para realizar la clasificación con las etiquetas de emociones (felicidad, rabia, tristeza, sorpresa, miedo, asco y neutral). Los Casos de Usos relacionados son el CU1 y CU2 que son los utilizados para la recepción de los datos de entrada del sistema. En este componente se hace una extracción de fotogramas del vídeo de entrada, el número de fotogramas puede variar de 5 a 10, se aplican filtros de conversión de colores, intensidad, rotación, recorte y cambio de tamaño y ecualización de la imagen, estos parámetros son definidos por los modelos pre-entrenados de *face-landmarks-detection*.

4.1.2.2. Reconocimiento y Clasificación

Las emociones que se detectan serán extraídas de la manifestación de gestos en el rostro, la cual se identifica para esta investigación como el proceso de reconocimiento, relacionando con el caso de uso CU3. Posterior a la fase de reconocimiento se debe clasificar los AU's extraídos en una emoción cumpliendo con los dos últimos casos CU4 y CU5. Para la facilidad en el desarrollo se decidió hacer una subdivisión en este subsistema para estos dos procesos.

- **Reconocimiento de Rostro:** en este proceso es fundamental detectar el rostro y obtener características para su posterior procesamiento y clasificación. Muchos investigadores utilizan muchos algoritmos de detección de rostros como OpenCV, DLIB, Eigenfaces, LBPH y Viola-Jones (Al-Sumaidae et al., 2015). Los algoritmos convencionales extraen los aspectos destacados del rostro, llevando la teoría de la detección de AU's de Ekman a la práctica.

En este proyecto se utiliza el *face-landmarks-detection* para detectar esos puntos de referencias faciales. En el sistema se utiliza este paquete en el componente

Action (ver figura 4.2), para seleccionar los puntos de: punta de nariz, mejilla izquierda, mejilla derecha, parte inferior y superior del ojo izquierdo, parte inferior y superior del ojo derecho, ceja inferior izquierda, ceja inferior derecha, labio inferior y superior interno. Los cuales se muestran en forma de malla encima del rostro detectado, en la captura de vídeo en tiempo real del sistema.

- **Clasificación de Emociones:** luego de ser extraídos los AU's se procede a la clasificación de las emociones, en el desarrollo se aplica este proceso en el componente **Emotion** (ver figura 4.2), el cual al igual que el subsistema anterior, realiza una extracción de los AU's para luego procesar estos puntos y clasificar según la emoción, esta clasificación se lleva a cabo a través de uno de los modelos entrenados en esta investigación, `mlp.json`. Este modelo recibe el array de puntos extraídos por el *face-landmarks-detection* y predice cual es el emoción resultante tomada de la captura de vídeo del sistema, además de resaltar el contorno del rostro del evaluado.

4.1.3. Diseño de la Interfaz del Sistema

La interfaz del usuario se diseñó en función del cumplimiento de los requisitos no funcionales, enfatizando la mínima complejidad en su manejo. El mockup se desarrolló en colores cálidos, y estructurados de manera tal que se aprovecharan los componentes del Framework *Vuetify*. Los elementos más importantes se detallan a continuación:

- **Header:** a su lado izquierdo muestra el logo del sistema que permite el redireccionamiento a la vista **Home**, al lado derecho muestra un pequeño menú con enlaces importantes para el proyecto.
- **Área de contenido:** muestra los contenidos de los componentes **HomeView**, **EmotionView**, **ActionView** y **NotFoundView** (ver figuras 4.3a, 4.3b, 4.3c, 4.3d).

- **Footer:** contiene el pie del sitio, muestra información referente a la Universidad de Los Andes.

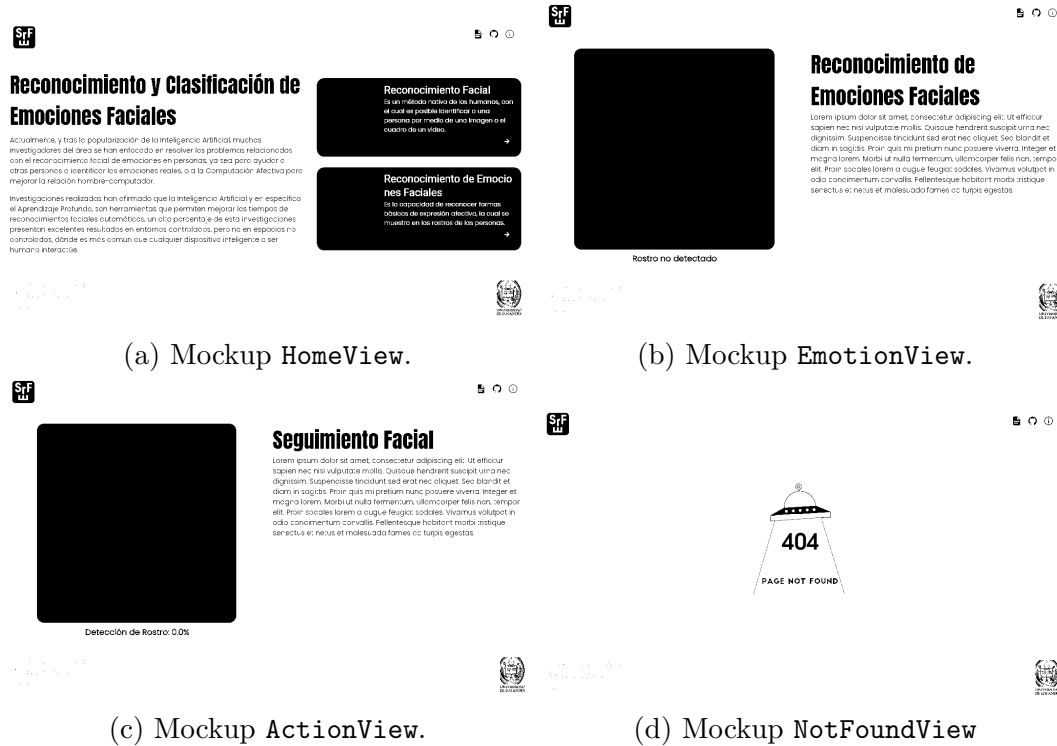


Figura 4.3: Mockup del SREF

4.2. Diseño de la Base de Datos

El diseño de la base de datos en esta investigación se centra en los datos que se utilizaron para el entrenamiento y validación de los modelos y el caso de estudio. Este proceso es uno de los pasos críticos en la construcción de un modelo, los datos involucrados deben ser eficaz en términos de detalles y usabilidad.

4.2.1. FER2013

Para el reconocimiento de emociones faciales en personas, existen muchos datos disponibles, es el caso de la base de datos abierta FER2013, la cual contiene 35887 imágenes de rostros en escala de grises de 48*48 pixeles, recolectadas de forma semi automática, mediante una metodología basada en la API del motor de búsqueda de Google, que hace que el rostro esté más o menos centrado y ocupe aproximadamente la misma cantidad de espacio en cada imagen.

Este conjunto se encuentra etiquetado con las 6 emociones básicas de Ekman y la expresión neutral (ver figura 4.4), dividido en 28709 imágenes para entrenamiento y 7178 para validación.

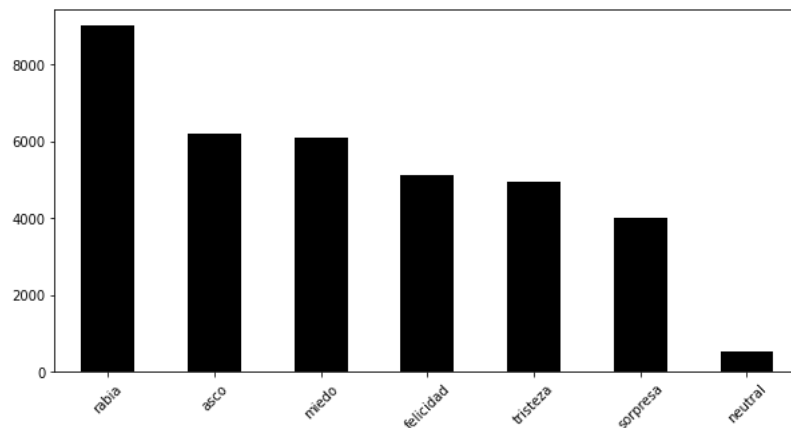


Figura 4.4: Distribución de emociones en *FER2013*

4.2.2. Base de Datos del Caso de Estudio

Para el caso de estudio se decidió trabajar con vídeos de reacciones en las Redes Sociales (se considera un vídeo de reacción en los cuales los “influencers” expresan emociones ante vídeos de terceros). Para esta base de datos se realizó una recopilación de videos en la red social Facebook, luego se procedió con la extracción de fotogramas con

el Reproductor de Vídeo VLC, posteriormente se recortaron, ajustaron y clasificaron de manera similar a FER2013.

4.3. Diseño del Reconocedor de Emociones Faciales

Investigaciones relacionadas al reconocimiento de emociones faciales mediante imágenes/vídeo concluyen que la tasa de éxito depende de la configuración que tenga el sistema en cada una de sus etapas (pre-procesamiento, reconocimiento de rostro, extracción de AU's y clasificación) (García Mariño, 2017), aunque, gran parte de estas centran su exploración en la comparación de la tasa de acierto de reconocimiento, entre los diferentes algoritmos. Algoritmos como Redes Bayesianas Simples, SVM, Árbol de Decisiones, Bosque Aleatorio, AdaBoost, MLP, CNN, entre otros, son utilizados. En la tabla 4.1 se muestran algunas de las tasas de acierto, bases de datos y algoritmos estudiados en algunas investigaciones relevantes.

Algoritmo	Emociones	Exactitud	Base de Datos
Redes Bayesianas Simples	8	82,05 %	Lucey
SVM	8	38,46 %	Lucey
Árbol de Decisiones	8	82,05 %	Lucey
Bosque Aleatorio	8	82,05 %	Lucey
AdaBoost	8	38,46 %	Lucey
MLP	8	88,03 %	Lucey
CNN+MMOD	7	87,91 %	FER2013

Tabla 4.1: Tabla de clasificación de emociones para diferentes algoritmos (Soberanes et al., 2019; Gallud Baños, 2019)

4.3.1. Selección de algoritmo

Son muchos los algoritmos propuestos para la tarea de clasificación de emociones, pero como se expone en la tabla 4.1 el MLP y el CNN destacan por su tasa de acierto.

Investigaciones han resaltado que con estos algoritmos se reduce el costo computacional, sobre todo con el CNN, gracias a que los algoritmos aprenden y se adaptan rápidamente (Calvachi and Alexandra, 2019). En concordancia a las afirmaciones en estos y otros antecedentes (Gallud Baños, 2019; Pérez Estruch, 2017; Jaramillo Romero, 2020) y motivado a profundizar la investigación, se ha decidido realizar una comparación entre los dos algoritmos más acertados de esta tabla. Utilizando el MLP para desarrollar el modelo del SREF y una CNN para aplicar pruebas en el caso de estudio.

4.3.2. Implementación

Para la implementación de los modelos con los algoritmos MLP y CNN se utiliza la biblioteca Keras dentro de la herramienta Colab. Para la elaboración de los modelos se ejecutan dos procesos uno de entrenamiento y el otro de prueba, al igual que en el SREF, para el modelo se ejecuta un pre-procesamiento de datos, la cual hace una extracción de características de las muestras de las bases de datos, obteniendo un conjunto de vectores etiquetados con las 7 clases de emociones y posteriormente se lleva a cabo el entrenamiento con los algoritmos de aprendizaje profundo, en la fase prueba se realizó el mismo enfoque pero con el modelo obtenido se clasifican las muestras.

Como ya se ha mencionado el Aprendizaje Profundo trabaja con Redes Neuronales, donde cada una tiene un número de capas (Saez De La Pascua, 2019). Con el algoritmo MLP se ha decidido desarrollar un modelo con tres capas ocultas (ver figura 4.5a), al clasificador entran imágenes de 48x48 proveniente de la Base de Datos FER2013, compuestas por un canal en escala de grises, mientras que la salida tendrá 7 neuronas de salida equivalentes a las emociones a clasificar. Para el caso del modelo con una CNN se compone de 4 capas de agrupación y convolución, con las mismas capas de

entrada y salida que el MLP (ver figura 4.5b).

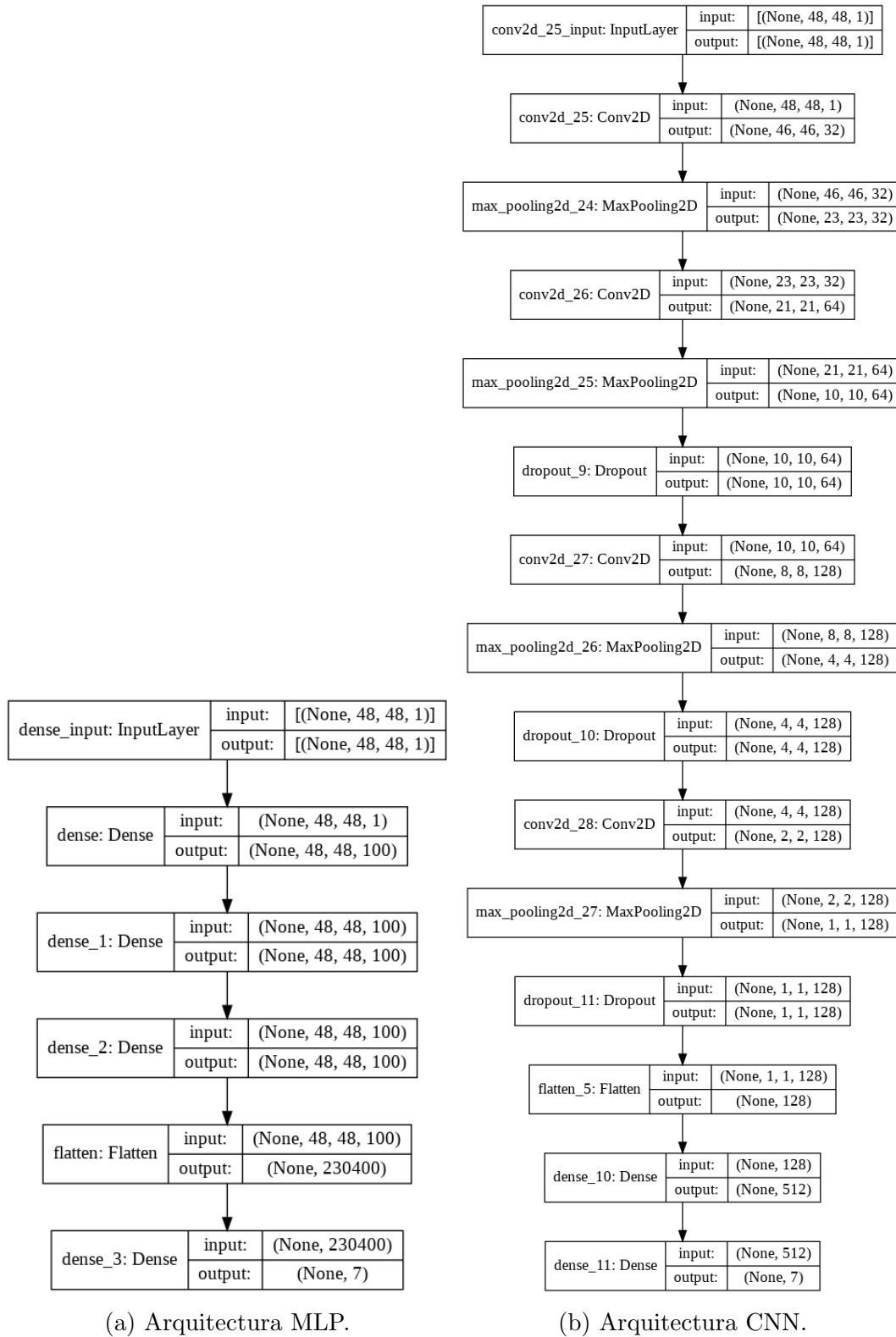


Figura 4.5: Arquitecturas de los modelos

Capítulo 5

Pruebas

Las pruebas comprueban la calidad de un software, reducen fallos y además verifican el cumplimiento de los requisitos. Para el SREF se estructura una serie de pruebas unitarias para cada uno de los componentes diseñados, y algunas pruebas entre ellos, debido a la dependencia entre sus subsistemas. Debido a que el sistema fue implementado mediante los frameworks *Vue.js* y *Vuetify*, se utilizó el framework *Jest.js* para ejecutar las pruebas en conjunto, la mayoría de estas pruebas consisten en renderización.

De igual forma se exponen las pruebas de exactitud de los modelos de Aprendizaje Profundo evaluados para la implementación de la tarea de Reconocimiento de Emociones Faciales, estas pruebas fueron realizadas bajo el conjunto de datos **FER2013** y el caso de estudio, utilizando los entornos de Google Colab y TensorBoard.

5.1. Pruebas del Sistema de Reconocimiento de Emociones Faciales

El SREF se compone de subsistemas: el componente de pre-procesamiento y el componente de reconocimiento y clasificación los cuales al integrarse ofrece un sistema funcional. Para comprobar la calidad del sistema se realizarán varias pruebas que permiten verificar el correcto funcionamiento de manera cualitativa.

5.1.1. Plan de Pruebas

Debido a que el sistema se realizó con los frameworks *Vue.js* y *Vuetify* se adoptó las pruebas unitarias por componente, implementadas en conjunto a través del framework *Jest.js*, a continuación se explicará de manera gradual las pruebas realizadas en cada componente

5.1.1.1. Pruebas Unitarias

Las pruebas unitarias son parte esencial del software, estas pruebas ejecutan las unidades más pequeñas de código, de manera independiente, aumentando la facilidad de rastrear errores. La estructura por componente de los frameworks en este desarrollo ayuda a la escritura de pruebas unitarias (Tes, 2021).

Estas pruebas serán aplicadas a cada uno de los subsistemas del SREF, esto se debe a que cada uno de ellos esta representado o es parte de un componente y además ofrece una funcionalidad específica.

5.1.1.2. Especificación del Diseño de Pruebas

El sistema de este proyecto esta compuesto por dos subsistemas, cada uno de ellos de manera integrada en los componentes **Action** y **Emotion** en la implementación (ver figura 4.2), por lo que se procedió a realizar pruebas de renderizado a cada característica posible de ellos y a las vistas que permiten su funcionalidad (ver tabla 5.1).

- **Pruebas del componente de preprocesamiento:** a nivel de implementación este componente se encuentra codificado en los componentes **Action** y **Emotion** del SREF, permite la entrada de los datos (vídeo) para su posterior procesamiento. Las pruebas unitarias realizadas se enfocaron en la comprobación de los elementos `canvas` y `video` del código.
- **Pruebas del componente de reconocimiento y clasificación:** de la misma manera que el subsistema anterior, este se encuentra codificado en los dos componentes de la implementación, encargándose del reconocimiento facial y extracción de las AU's y la clasificación de las emociones, cuyos resultados se expresan a través del dibujo de la malla de los Au's y el lienzo del rostro, por tal las pruebas unitarias también se aplicaron en la renderización de los elementos de salida.

5.2. Pruebas del Reconocedor de Emociones Faciales

Por lo general los algoritmos de Aprendizaje son probados con datos diferentes a los que se entrena el modelo, midiendo la capacidad de predecir con datos antes no vistos. Estas pruebas ayudarán a las decisiones durante el diseño de la Red Neuronal, como el cambio de función de pérdida, optimización, el número de épocas, entre otras.

Componente	Entrada	Test	Salida	Salida Esperada
Action	Action	expect(shallowMount(Action));	True	Render de Componente
Action	title	expect(wrapper.vm.title).toMatch('Reconocimiento de Rostro y Aus');	True	“Reconocimiento de Rostro y Au”
Action	h3	expect((wrapper.find('h3')).exists()).toBe(true);	True	True
Action	canvas	expect((wrapper.find('canvas')).exists()).toBe(true);	True	True
Action	video	expect((wrapper.find('video')).exists()).toBe(true);	True	True
Action	ActionView	expect(shallowMount(Action));	True	Vista Action
Emotion	Emotion	expect(shallowMount(Emotion));	True	Render de Componente
Emotion	title	expect(wrapper.vm.title).toMatch('Reconocimiento de Emociones');	True	“Reconocimiento de Emociones”
Emotion	h3	expect((wrapper.find('h3')).exists()).toBe(true);	True	True
Emotion	canvas	expect((wrapper.find('canvas')).exists()).toBe(true);	True	True
Emotion	video	expect((wrapper.find('video')).exists()).toBe(true);	True	True
Emotion	EmotionView	expect(shallowMount(Emotion));	True	Vista Emotion
Home	HomeView	expect(shallowMount(Home));	True	Vista Home
Home	RouterLinkStub	expect(wrapper.getComponent(RouterLinkStub).props().to).toBe('/action' or '/emotion')	True	http://localhost:8000/action o http://localhost:8000/emotion

Tabla 5.1: Pruebas Unitarias del SREF

El Aprendizaje Profundo necesita de grandes bases de datos, lo que exige definir una metodología para la separación de datos que ayude a evaluar las diferentes configuraciones, es típico separar los datos en un conjunto de datos para entrenamiento y otro conjunto para prueba, *Keras* brinda dos maneras de evaluar los algoritmos: usando un conjunto de verificación automática y uno de verificación manual. La verificación automática utiliza el argumento `validation.split` dentro de la función `fit()` para segmentar el conjunto de datos, según un porcentaje indicado por el programador. Mientras que en la verificación manual ya existe una separación previa de los datos en los dos conjuntos, aunque se puede utilizar los argumentos de `fit()` para las pruebas durante el ajuste del diseño.

Además de estas verificaciones existe una técnica considerada un método estadístico, llamado validación cruzada donde los conjuntos de entrenamiento deben cruzarse en rondas sucesivas de modo que cada punto de datos tenga la oportunidad de ser validado (Refaeilzadeh et al., 2009), a pesar de ser una de las técnicas más complejas y certeras, la validación cruzada no se usa generalmente para evaluar modelos de Aprendizaje Profundo debido a la exigencia computacional, ya que se deben construir y evaluar los modelos k veces, por esto y sumando que el entrenamiento de los modelos de esta investigación tardaron en promedio un tiempo de entrenamiento de 3 a 11 horas, se descartó utilizar esta verificación.

Para la evaluación del Reconocedor de Emociones Faciales se utiliza la técnica de validación o verificación manual, para estimar el rendimiento de los modelos con cada algoritmo (MLP y CNN), y así definir el mejor. A continuación se presentarán cada una de las pruebas que se realizaron a los algoritmos de Aprendizaje Profundo utilizados para el desarrollo del modelo:

5.2.1. Prueba 1

En la prueba 1 se utilizará el método de validación manual sobre la base de datos FER2013 y el caso de estudio, esta prueba permitirá evaluar la independencia de la particiones de datos entre entrenamiento y prueba.

Para la primera parte de estas pruebas en el algoritmo de MLP se utilizaron los parametros de $\text{batch_size} = 256$ y $\text{epochs} = 30$, el diseño de la arquitectura no fue modificado, ni la tasa de verificación fue afectada notablemente como para modificar algún parámetro de configuración. Mientras que para el modelo con CNN se variaron

los parámetros y la inclusión de capas según las tasas de aciertos obtenidas, se incluyeron algunas capas de dropout para mejorar el sobreajuste (ver figura 4.5b), este parámetro se ajustó de acuerdo a lo indicado en el trabajo de investigación (Park and Kwak, 2016) quién señala que la aplicación de la deserción en capas convolucionales a niveles bajos como 0.1, 0.2 y 0.3 mejora considerablemente el sobreajuste, los mejores resultados se obtuvieron con los parámetros `batch_size = 256` y `epochs = 50`.

Debido a que el objetivo de esta investigación es clasificar emociones faciales de una persona a través de un vídeo en espacios no controlados, se aplicó una segunda fase de pruebas haciendo un tratamiento de imágenes usando `ImageDataGenerator` de *Keras*, este tratamiento puede asemejar más los datos de pruebas a datos de la vida real no controlados. Los parámetros utilizados se detallan en la tabla 5.2.

Parámetro	Valor
<code>rotation_range</code>	30
<code>width_shift_range</code>	0.2
<code>height_shift_range</code>	0.2
<code>shear_range</code>	15
<code>zoom_range</code>	[0.7, 1.4]
<code>horizontal_flip</code>	True
<code>vertical_flip</code>	True

Tabla 5.2: Parámetros para `ImageDataGenerator`

Las tasas de rendimiento calculadas mediante la verificación manual fueron expresadas en la tabla 5.3, los resultados obtenidos muestran que para el CNN se consigue mejores resultados para ambas base de datos, pero el aprendizaje combinado que se realizó del modelo MLP empleado en el SREF junto a la librería `face-landmarks-detection` (FLD) tienen la tasa de exactitud más alta.

Algoritmo	Base de Datos	Exactitud	ImageDataGenerator
MLP	FER2013	36,14 %	No
MLP	FER2013	27,83 %	Sí
CNN	FER2013	60,07 %	No
CNN	FER2013	54,53 %	Sí
CNN	Caso de Estudio	31,25 %	No
CNN	Caso de Estudio	17,62 %	Sí
MLP+FLD	FER2013	67,10 %	No

Tabla 5.3: Tasa de aciertos de reconocimiento en modelos

5.2.2. Prueba 2

En la prueba 2 se llevaron a cabo las pruebas orientadas a realizar estimaciones de la capacidad de generalización de cada algoritmo. Para el algoritmo MLP se realizaron las pruebas con las muestras de la base de datos FER2013, mientras que para el modelo con CNN se utilizó además la base de datos del Caso de Estudio, con este mismo conjunto de datos se evaluó la API de uso libre AFFDEX SDK para obtener una línea de comparación. Adicional a la matriz de confusión generada en estas pruebas, se extrajeron las métricas de: exactitud, precisión, sensibilidad, puntaje f1 y curva ROC de los modelos diseñados, a continuación se detallan los resultados:

5.2.2.1. Modelo diseñado con un MLP

En la figura 5.1a podemos observar las cantidades de emociones acertadas con este algoritmo en el modelo 4.5a, utilizando la base de datos FER2013 sin aplicar ninguna transformación en los datos. Para este algoritmo las mayores emociones reconocidas fueron para la expresión felicidad y la menor, rabia. En la tabla 5.4 se pueden observar las métricas obtenidas para este modelo, las emociones asco, felicidad y sorpresa tienen la mayor precisión, sensibilidad y puntaje F1, mientras que la rabia es la emoción con menor puntaje en todas las métricas.

En la figura 5.2a podemos observar las cantidades de emociones acertadas con el

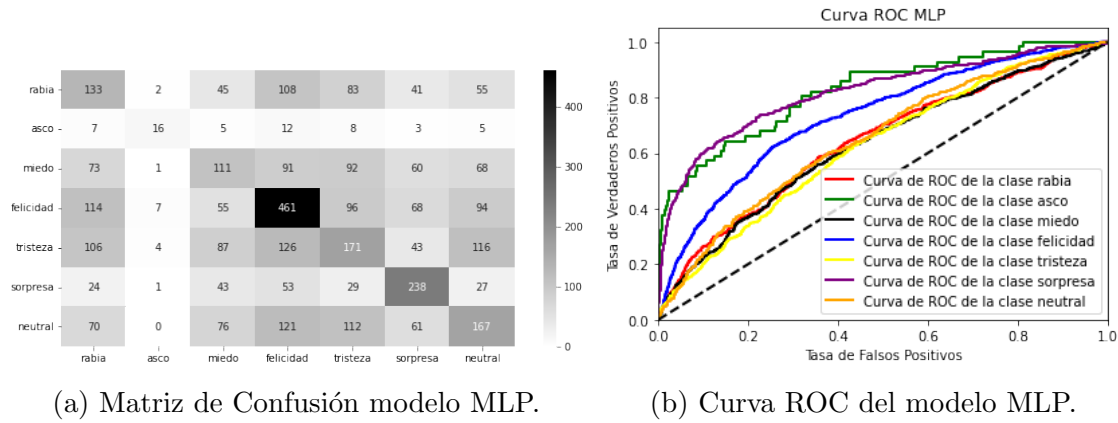
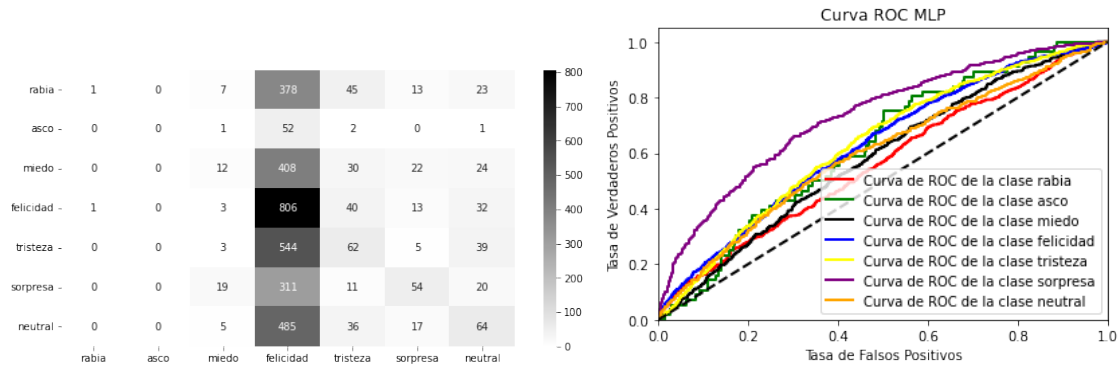


Figura 5.1: Pruebas en el modelo MLP.

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	25 %	28 %	27 %
Asco	52 %	29 %	97 %
Miedo	26 %	22 %	24 %
Felicidad	47 %	52 %	49 %
Tristeza	29 %	26 %	27 %
Sorpresa	46 %	57 %	51 %
Neutral	31 %	28 %	29 %

Tabla 5.4: Métricas para el modelo MLP

algoritmo MLP, utilizando la base de datos FER2013 aplicando la transformación de la tabla 5.2 en los datos. Para este algoritmo la emoción felicidad tuvo los mayores verdaderos positivos pero también los mayores falsos positivos. En la tabla 5.5 se detallan las métricas obtenidas para este modelo con la transformación de datos. Se obtienen métricas con muy bajos porcentajes en todas las emociones, aunque la sensibilidad y Puntaje F1 de felicidad resaltan como las más altas.



(a) Matriz de Confusión modelo MLP.

(b) Curva ROC del modelo MLP.

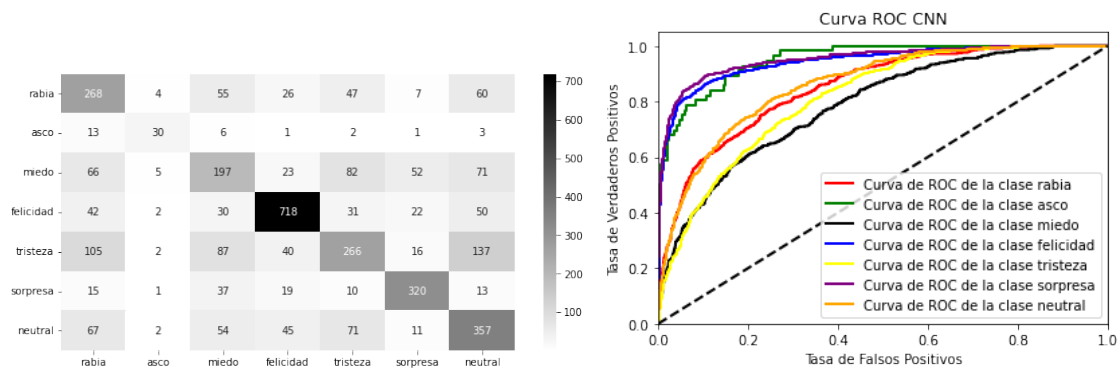
Figura 5.2: Pruebas en el modelo MLP.

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	50 %	0 %	0 %
Asco	0 %	0 %	0 %
Miedo	24 %	2 %	4 %
Felicidad	27 %	90 %	42 %
Tristeza	27 %	9 %	14 %
Sorpresa	44 %	13 %	20 %
Neutral	32 %	11 %	16 %

Tabla 5.5: Métricas para el modelo MLP

5.2.2.2. Modelo diseñado con una CNN

Para el modelo diseñado con una CNN (ver figura 4.5b) se diseñaron varias pruebas, en la figura 5.3a podemos observar la matriz de la primera de ellas, utilizando la base de datos FER2013 y sin aplicar ninguna transformación en las imágenes. Para este algoritmo las mayores emociones reconocidas fueron para la expresión felicidad al igual que en su homólogo con MLP y la menor el miedo. En la tabla 5.6 se pueden observar las métricas obtenidas, las emociones felicidad y sorpresa tienen la mayor precisión, sensibilidad y puntaje F1, mientras que el miedo es la emoción con menor puntaje en las métricas.



(a) Matriz de Confusión modelo CNN.

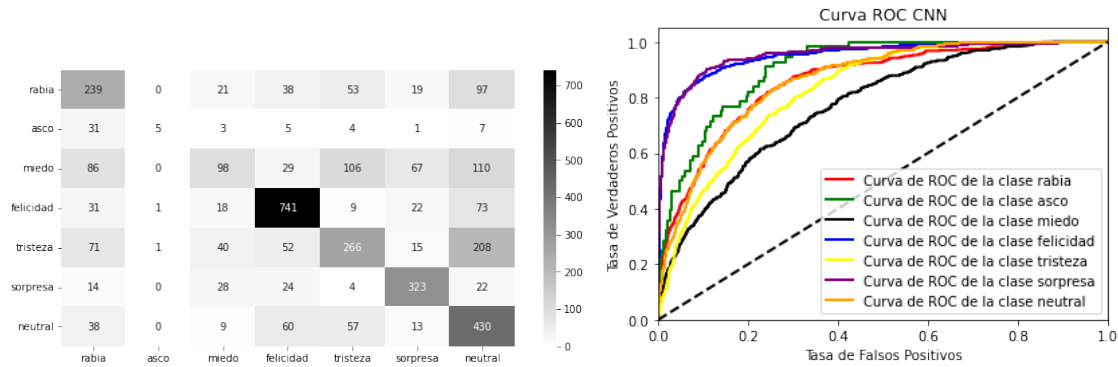
(b) Curva ROC del modelo CNN.

Figura 5.3: Pruebas en el modelo CNN.

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	47 %	57 %	51 %
Asco	65 %	54 %	59 %
Miedo	42 %	40 %	41 %
Felicidad	82 %	80 %	81 %
Tristeza	52 %	41 %	46 %
Sorpresa	75 %	77 %	76 %
Neutral	52 %	59 %	55 %

Tabla 5.6: Métricas para el modelo CNN

Para la segunda prueba de este modelo aplicando la transformación en las imágenes de FER2013, se obtuvieron resultados más bajos, pero las tendencias de acierto tanto en 5.4a como en las métricas de la tabla 5.7 son similares, aunque a la baja el asco presentó los más altos falsos positivos.



(a) Matriz de Confusión modelo CNN.

(b) Curva ROC del modelo CNN.

Figura 5.4: Pruebas en el modelo CNN.

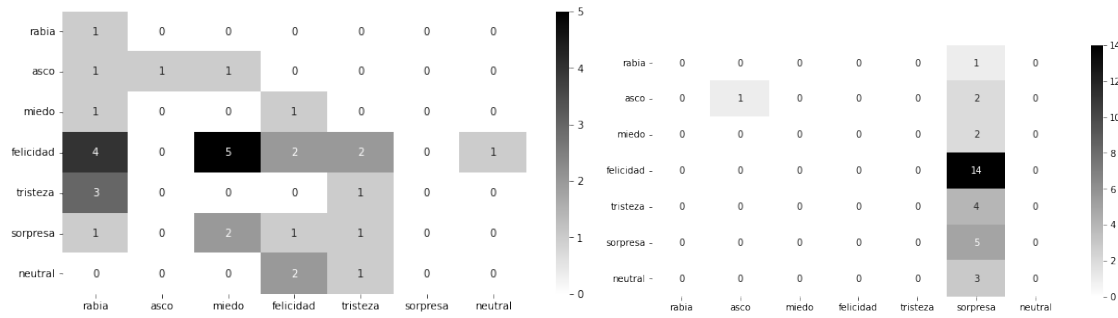
Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	47 %	51 %	49 %
Asco	71 %	9 %	16 %
Miedo	45 %	20 %	27 %
Felicidad	78 %	83 %	80 %
Tristeza	53 %	41 %	46 %
Sorpresa	70 %	78 %	74 %
Neutral	45 %	71 %	55 %

Tabla 5.7: Métricas para el modelo CNN

En la última fase de esta prueba se hicieron análisis con la base de datos del caso de estudio, tanto con imágenes transformadas como sin ellas, pero como se observa los resultados obtenidos fueron muy bajos, acertando en muy pocos casos la emoción y con baja precisión (ver figura 5.5 y tablas 5.8 y 5.9).

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	9 %	100 %	17 %
Asco	100 %	33 %	50 %
Miedo	0 %	0 %	0 %
Felicidad	33 %	14 %	20 %
Tristeza	20 %	25 %	22 %
Sorpresa	0 %	0 %	0 %
Neutral	0 %	0 %	0 %

Tabla 5.8: Métricas para el modelo CNN para Caso de Estudio



(a) Matriz de Confusión modelo CNN para Ca- de Estudio. (b) Matriz de Confusión del modelo CNN para Caso de Estudio.

Figura 5.5: Pruebas en el modelo CNN en Caso de Estudio.

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	0 %	0 %	0 %
Asco	100 %	33 %	50 %
Miedo	0 %	0 %	0 %
Felicidad	0 %	0 %	0 %
Tristeza	0 %	0 %	0 %
Sorpresa	16 %	100 %	28 %
Neutral	0 %	0 %	0 %

Tabla 5.9: Métricas para el modelo CNN para Caso de Estudio con transformación de datos

5.2.2.3. AFFDEX SDK

AFFDEX SDK conocida actualmente como Affective, permite a los usuarios probar parte de sus herramientas a través de una pequeña aplicación web, dónde se identifican los porcentajes obtenidos por cada expresión exceptuando la neutral, esto a través de un archivo de imagen o vídeo previamente cargado. Además de identificar las expresiones obtenidas del rostro muestra también los porcentajes encontrados de algunos AU's.

Las pruebas de este software se llevaron a cabo con la base de datos del Caso de Estudio sin transformación de datos, se obtuvo un archivo en formato *JSON* por cada rostro evaluado, según el manual de la API cuando se obtiene un archivo vacío significa que la aplicación no detectó ningún rostro, caso tal se presentó para 18 imágenes de la

base de datos. Luego de obtener todos los archivos *JSON* se procedió a combinarlos y exportarlos en un archivo *CSV* para el cálculo de las métricas con librerías de *Python* (ver tabla 5.10). La matriz de confusión se calculó manualmente por lo que no se obtuvo la Curva de ROC 5.6.

Emoción	Precisión	Sensibilidad	Puntaje F1
Rabia	0 %	0 %	0 %
Asco	33 %	100 %	50 %
Miedo	100 %	100 %	100 %
Felicidad	25 %	50 %	33.3 %
Tristeza	0 %	0 %	0 %
Sorpresa	0 %	0 %	0 %

Tabla 5.10: Métricas para el Software *Afectiva* con el Caso de Estudio sin transformación de datos

rabia	0	0	0	0	0	1
asco	1	1	0	1	0	0
miedo	0	0	1	0	0	0
felicidad	0	0	0	1	0	3
tristeza	0	0	0	0	0	0
sorpresa	0	0	0	0	0	0
	rabia	asco	miedo	felicidad	tristeza	sorpresa

Figura 5.6: Matriz de Confusión de *Afectiva* evaluado con el Caso de Estudio.

A pesar de que no fueron consideradas las imágenes dónde el software no detecto ningún rostro para los cálculos, se evidencia claramente que al igual que en las pruebas realizadas con los modelos MLP y CNN bajo las mismas condiciones y con los mismos datos, los resultados fueron vagamente precisos.

5.2.3. Análisis de los resultados

Considerando los resultados que se obtuvieron en las pruebas mostradas en las secciones anteriores realizadas sobre los modelos 4.5b, 4.5a y con las bases de datos FER2013 y Caso de Estudio, se realizan las siguientes observaciones:

- En las pruebas de verificación manual se permitió estimar la exactitud en datos no vistos, como se detalla en 5.3, se obtuvieron buenos resultados para el modelo con CNN, aunque la mayor tasa de reconocimiento se obtuvo al realizar la combinación del modelo MLP y FLD, por lo que para el SREF se aplicó esta combinación.
- En las pruebas de la segunda fase se observó que la mayor tasa de reconocimiento fue para las emociones de felicidad y sorpresa, aunque se atisbaron gran cantidad de falsos positivos para el Caso de Estudio. De manera general las emociones neutral, miedo y rabia obtuvieron menor cantidad de precisión en las métricas.
- Considerando las pruebas realizadas con la base de datos de Caso de Estudio, se observa que los resultados son pésimos, por lo que se deberá considerar a futuro crear una base datos más amplia, para corroborar si los resultados de estas pruebas son certeros.

Capítulo 6

Conclusiones y Recomendaciones

En esta investigación se llevó a cabo el desarrollo de 2 productos: el Reconocedor de Emociones Faciales y el SREF, cuyo objetivo de desarrollo, es clasificar emociones faciales de una persona a través de un vídeo en espacios no controlados, dónde es más común que cualquier dispositivo inteligente o ser humano interactúe.

Para la creación de ambos productos se realizó una profunda indagación bibliográfica en áreas de IA, desarrollo web, y reconocimiento de emociones faciales, permitiendo así, no solo obtener un resultado palpable, sino un resultado de calidad con fundamentos teóricos reales. Estos fundamentos ayudaron a definir criterios de diseño e implementación que permitieron organizar pruebas de comparación críticas que evaluaron rígidamente a los productos.

Luego de probado ambos productos se obtuvo un sistema completo como aplicación web, que permite a sus usuarios tener a la mano una sencilla herramienta para el reconocimiento de emociones faciales en humanos, a través de vídeos con entrada en tiempo real.

6.1. Conclusiones

El desarrollo de la sociedad en los últimos años ha ido de la mano con el crecimiento de grandes tecnologías, en gran medida con la IA, por lo general este crecimiento ha impactado de manera positiva en las personas, ya que los mayores avances buscan maximizar la interacción con sus usuarios, brindar facilidad de uso y prestar funcionalidades en las tareas comunes.

Se considera que para que un dispositivo inteligente, tal como un robot, un televisor o incluso un sistema de domótica impacte positivamente en la sociedad, es beneficioso que sea capaz de detectar las emociones predominantes de sus usuarios (Pellejero et al., 2017), también es importante diferenciar entre las expresiones faciales que demuestran emociones artificiales y espontáneas, actividad que para al ser humano es compleja, ampliando aun más las posibilidades donde las tareas de reconocimientos son necesarios.

La dependencia de estas tecnologías y actividades con la tarea de reconocimiento de emociones faciales, motivó el inicio de esta investigación, por lo que se consideró fuera un gran aporte desarrollar un sistema para el reconocimiento y clasificación de emociones faciales, con entradas de vídeos en tiempo real en espacios no controlados. Este sistema se llevó a cabo con la utilización del método W_Watch, el cual permitió con facilidad la iteración entre las fases del desarrollo del sistema. Teóricamente, una de las partes del reconocimiento de emociones faciales es la detección de los AUs, por lo que una de las iteraciones se dedicó a la comprensión del FACS. Posterior a ello se iteró entre el desarrollo del modelo, el cual a través del entrenamiento de algoritmos de Aprendizaje Profundo permitió la investigación y comparación para seleccionar el

mejor algoritmo y diseñar así un sistema certero. Una de las iteraciones se refirió a la construcción de una base de datos para el reconocimiento de emociones con el caso de estudio: emociones expresadas por influencers en vídeos de reacciones, el cual ayudó a realizar pruebas entre los algoritmos. Estas pruebas definieron que el modelo con MLP combinado al modelo pre-entrenado FLD tiene una tasa de acierto de 67,10 % por lo que se decidió utilizar esta combinación en el SREF.

De manera general se puede concluir que la utilización de algoritmos de Aprendizaje Profundo para el reconocimiento de emociones faciales tiene tasas de acierto altas, como se evidenció la aplicación de CNNs tiene mayor precisión por emoción, aunque la combinación de algoritmos o modelos mejora considerablemente los porcentajes de validación. Con el caso de estudio se demuestra que para este tipo de algoritmos es necesario contener un datasets amplio, ya que se genera un sobreajuste en el modelo, haciendo que las métricas de validación sean muy bajas.

Por otro parte es importante aclarar que la utilización de las herramientas como Google Colaboratory y TensorBoard fueron fundamentales para el desarrollo del reconocedor de emociones, puesto que no se disponía de la capacidad de computo física necesaria para este desarrollo. Además la evaluación de los logs en tiempo real creados por TensorBoard, agilizó la variación de parámetros de entrenamiento, permitiendo rapidez en el diseño del modelo. También, la utilización de los frameworks *Vue.js* y *Vuetify* permitieron el cambio ligero de enfoque del sistema, puesto que durante las iteraciones de investigación se concluyó, que para la mayoría de usuarios es más representativo visualizar una aplicación con interfaz que les permita probar, que los resultados de las métricas de una validación.

Finalmente, se concluye que los objetivos planteados en el capítulo 1 fueron alcanzados tras haber realizado la integración del modelo del reconocedor de emociones faciales y el SREF y probado con el caso de estudio.

6.2. Recomendaciones

Se considera conveniente, que a partir de este estudio se continúe la línea de investigación, tomando en cuenta las siguientes observaciones para mejorar el Reconocedor de Emociones y el SREF que se ha desarrollado:

- El SREF esta desarrollado para utilizarse en cualquier dispositivo con un navegador pero debido a que usarlo requiere gran parte de recursos del lado del cliente, es probable que algunos dispositivos tengan problemas de rendimiento al abrir la aplicación, por lo que para trabajos futuros se recomienda optimizar el sistema.
- Por lo general, las condiciones del ambiente donde se requiere aplicar la tarea de reconocimiento son poco controladas, por tal se considera que se debe enfatizar, aún más, la fase de pre-procesamiento de los datos de entrenamiento, con mejoras de amplitud de tamaño, énfasis de luz, entre otras.
- Debido a la exigencia computacional por parte de los algoritmos de Aprendizaje Profundo, se considera, que para el desarrollo de mejores modelos se inicie por la investigación de herramientas que permitan aprovechar al máximo los recursos del equipo que utilizará para el entrenamiento.
- Ya que esta investigación demostró que con la combinación de algoritmos o modelos se obtienen mayores tasas de acierto, se recomienda expandir el área de investigación con esta técnica.

- Considerando la investigación realizada en el FACS, se puede optar por un cambio de enfoque, aplicando el reconocimiento de emociones pero con indicio a la detección de mentiras, tal como expone Guillen en (Guillén, 2019).

Bibliografía

(2021). Pruebas unitarias de componentes Vue.

Ahmed Banafa (7 de Agosto de 2016). ¿Qué es el aprendizaje profundo?

Al-Sumaidae, S., Dlay, S., Woo, W., and Chambers, J. (2015). Facial expression recognition using local gabor gradient code-horizontal diagonal descriptor. In *2nd IET International Conference on Intelligent Signal Processing 2015 (ISP)*, pages 1–6. IET.

Albert, M. (1972). Nonverbal communication.

Andrés González (2021). ¿Aprendizaje supervisado o no supervisado?

Antona Cortés, C. (2017). Herramientas modernas en redes neuronales: la librería keras. B.S. thesis.

Arango, I., Brüne, M., Fresán, A., Ortega, V., Villanueva, J., Saracco, R., et al. (2013). Recognition of facial expression of the emotions and their relation to attachment styles and psychiatric symptoms. *Salud Mental*, 36:95–100.

Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). Openface: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10.

- Barrios Judith, M. J. (2010). W_watch: Método white_watch para el desarrollo de proyectos pequeños de software.
- Benítez, R., Escudero, G., Kanaan, S., and Rodó, D. M. (2014). *Inteligencia artificial avanzada*. Editorial UOC.
- Caballo, V. E. (1993). *Manual de evaluación y entrenamiento de las habilidades sociales*.
- Calvachi, A. and Alexandra, M. (2019). Uso de reconocimiento facial de emociones basado en técnicas de deep learning para el mejoramiento de la educación.
- Cebrián Chuliá, L. (2016). *Reconocimiento de Emociones mediante técnicas de aprendizaje profundo*. PhD thesis, Universitat Politècnica de València.
- Challenger-Pérez, I., Díaz-Ricardo, Y., and Becerra-García, R. A. (2014). El lenguaje de programación python. *Ciencias Holguín*, 20(2):1–13.
- Cheong, J. H., Xie, T., Byrne, S., and Chang, L. J. (2021). Py-feat: Python facial expression analysis toolbox. *arXiv preprint arXiv:2104.03509*.
- Cordero, J. and Aguilar, J. (2016). Reconocimiento multimodal de emociones en un entorno inteligente basado en crónicas. *Gráficas El Portatítulo*, pages 525–541.
- Degiovannini, M. (2011). Comparativa de frameworks web. *JavaHispano. com*, 2:3.
- Deng, L. and Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, 7(3–4):197–387.
- Diego Calvo (20 de Julio de 2017). Red Neuronal Convolutacional CNN.
- Docusaurus (2021). Jest-Delightful Javascript Testing, url = <https://jestjs.io/>.

- Ekman, P., Dalglish, T., and Power, M. (1999). Handbook of cognition and emotion. Chichester, UK: Wiley.
- Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.
- Ekman, P. and Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1):56–75.
- Ekman, P. and Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*, volume 10. Ishk.
- Ekman, P. and Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology*, 30(1):527–554.
- Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.
- Fasel, B. and Luetttin, J. (2003). Automatic facial expression analysis: a survey. *Pattern recognition*, 36(1):259–275.
- Gallud Baños, G. (2019). Reconocimiento de emociones humanas y su aplicación a la robótica social.
- García Mariño, E. (2017). *Diseño e implementación de un sistema de reconocimiento facial para la caracterización de personas*. PhD thesis.
- Google. Te damos la bienvenida a Colaboratory.
- Google. TensorBoard: el kit de herramientas de visualización de TensorFlow .

- Guillén, D. A. S. (2019). *Diseño de una Arquitectura de Aprendizaje Automático que Brinde Soporte para la Detección de Mentiras Mediante el Análisis de Video*. PhD thesis, Pontificia Universidad Católica del Perú-CENTRUM Católica (Peru).
- Hatley, D., Hruschka, P., and Pirbhai, I. (2013). *Process for system architecture and requirements engineering*. Addison-Wesley.
- Herrera Carrillo, M. (2020). Desarrollo de una aplicación web para la gestión de una asociación scout. B.S. thesis.
- Idris, I. (2013). *NumPy Beginner's Guide*. Packt Publishing Ltd.
- Iglesias-Hoyos, S., Castillo Arreola, A. d., and Muñoz-Delgado, J. I. (2016). Reconocimiento facial de expresión emocional: diferencias por licenciaturas. *Acta de investigación psicológica*, 6(3):2494–2499.
- James, W. (1985). What is an emotion? *Studies in Psychology*, 6(21):57–73.
- Jaramillo Romero, S. E. (2020). Detección de emociones y reconocimiento facial utilizando aprendizaje profundo.
- Juan Zambrano (31 de Marzo de 2018). **Conceptos Básicos de Machine Learning**.
- Kanade, T., Cohn, J. F., and Tian, Y. (2000). Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53. IEEE.
- LLC (2016-2020). **Material design component framework**.
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). 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*, pages 94–101. IEEE.
- Mallitasig Arellano, H. W. (2018). Paul ekman y las microexpresiones faciales de las emociones.
- Martínez, G. M., Ramírez, E. O. L., and Aguilar, M. I. H. (2010). *La psicología de las emociones: La expresión facial como una revelación de la emoción y el pensamiento*. Trillas.
- McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., and Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pages 3723–3726.
- Méndez Rosales, M. A. (2021). Certificación de documentos de los estudiantetes egresados de la universidad de los andes mediante blockchain.
- Park, S. and Kwak, N. (2016). Analysis on the dropout effect in convolutional neural networks. In *Asian conference on computer vision*, pages 189–204. Springer.
- Pellejero, N., Grinblat, G. L., and Uzal, L. (2017). Análisis semántico en rostros utilizando redes neuronales profundas. In *XVIII Simposio Argentino de Inteligencia Artificial (ASAI)-JAIIO 46 (Córdoba, 2017)*.
- Pérez Estruch, C. (2017). Emotions recognition using deep learning.
- Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 5:532–538.

- Restrepo Arteaga, G. J. P. et al. (2015). Aplicación del aprendizaje profundo (deep learning) al procesamiento de señales digitales. B.S. thesis, Universidad Autónoma de Occidente.
- Rosetti, D. L. (2017). Emoción y sentimientos.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? a review of the cross-cultural studies. *Psychological bulletin*, 115(1):102.
- Saez De La Pascua, A. (2019). Deep learning para el reconocimiento facial de emociones básicas. B.S. thesis, Universitat Politècnica de Catalunya.
- SalesForce Latinoamérica (22 de Junio de 2017). **Inteligencia Artificial: ¿Qué es?**
- Saracco, R. (2012). Reconocimiento facial de las emociones básicas en pacientes con esquizofrenia, hermanos y controles.
- Sawyer P., K. G. (2001). Software requirements. in the trial version (0.95) of swelok, guide to the software engineering body of knowledge. *Software Computer Society Press Tutorial*, pages 1–26.
- SearchDataCenter (Abril de 2017). **Aprendizaje profundo (deep learning)**.
- Senthilkumar, M. (2010). Use of artificial neural networks (anns) in colour measurement. In *Colour Measurement*, pages 125–146. Elsevier.
- Sharo, Z. (2020). *Maintenance Management System*. PhD thesis, University of Zakho.
- Skiendziel, T., Rösch, A. G., and Schultheiss, O. C. (2019). Assessing the convergent validity between the automated emotion recognition software noldus facereader 7 and facial action coding system scoring. *PloS one*, 14(10):e0223905.

- Soberanes, J. D., Gómez, C. N. S., Pato, V. M. Á., Calderón, J. S. G., and Guerrero, R. V. (2019). Estudio comparativo de clasificadores para el reconocimiento de expresiones faciales (a comparative study of classifiers for facial expression recognition). *Pistas Educativas*, 41(134).
- Tosi, S. (2009). *Matplotlib for Python developers*. Packt Publishing Ltd.
- Wati, D. A. R. and Abadianto, D. (2017). Design of face detection and recognition system for smart home security application. In *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pages 342–347. IEEE.
- y Paul Barham y Jianmin Chen y Zhifeng Chen y Andy Davis y Jeffrey Dean y Matthieu Devin y Sanjay Ghemawat y Geoffrey Irving y Michael Isard y Manjunath Kudlur y Josh Levenberg y Rajat Monga y Sherry Moore y Derek G. Murray y Benoit Steiner y Paul Tucker y Vijay Vasudevan y Pete Warden y Martin Wicke y Yuan Yu y Xiaoqiang Zheng, M. A. (2016). Tensorflow: un sistema para el aprendizaje automático a gran escala. pages 265–283.
- Zhang, L., Hossain, A., and Jiang, M. (2014). Intelligent facial action and emotion recognition for humanoid robots. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 739–746. IEEE.

Apéndice A

FACS

A.1. Gestos Faciales

Emoción	Lenguaje Corporal
Miedo	Cejas levantadas y contraídas al mismo tiempo. Las arrugas de la frente se sitúan en el centro y no extendidas por toda la frente. Párpado superior levantado, mostrando la esclerótica, con el párpado inferior en tensión y alzado. Boca abierta y labios o bien tensos y ligeramente contraídos hacia atrás, o bien estrechados y contraídos hacia atrás.
Sorpresa	Cejas levantadas, colocándose curvas y elevadas. Piel estirada debajo de las cejas. Arrugas horizontales en la frente. Párpados abiertos, párpado superior levantado, y párpado inferior bajado; el blanco del ojo suele verse por encima del iris. La mandíbula cae abierta, de modo que los labios y los dientes quedan separados, pero no hay tensión ni estiramiento de la boca.

Ira	Cejas bajas y contraídas al mismo tiempo. Líneas verticales entre las cejas. Párpado inferior tenso; puede estar levantado o no. Párpado superior tenso, puede estar bajo o no por la acción de las cejas. Mirada dura en los ojos, que pueden parecer prominentes. Labios en una de estas dos posiciones: continuamente apretados, con las comisuras rectas o bajas, o abiertos, tensos y en forma cuadrangular, como si gritaran. Las pupilas pueden estar dilatadas.
Tristeza	Los ángulos interiores de los ojos hacia arriba. La piel de las cejas forma un triángulo. El ángulo interior del párpado superior aparece levantado. Las comisuras de los labios se inclinan hacia abajo, o los labios tiemblan.
Asco	La nariz se arruga y los labios superiores se elevan, mientras que las comisuras descienden. Cuando el asco es muy fuerte, la lengua sale de la boca ligeramente.
Felicidad	Comisuras de los labios hacia atrás y arriba. La boca puede estar abierta o no, con o sin exposición de los dientes. El pliegue naso-labial, baja desde la nariz hasta el borde exterior por fuera de la comisura de los labios. Mejillas levantadas. Aparecen arrugas debajo del párpado inferior. Las arrugas denominadas 'patas de gallo' van hacia afuera, desde el ángulo externo del ojo.

Tabla A.1: Gestos Faciales en las Emociones

A.2. AUs

AU	Músculo Facial	Descripción del Movimiento del Músculo
1	Ceja interior	Esquina interior de la ceja levantada.
2	Ceja exterior	Esquina exterior de la ceja levantada.
4	Músculo Superciliar	Cejas dibujadas medialmente y hacia abajo.
5	Elevador del párpado superior	Ojos Abiertos, subir párpado superior.
6	Orbicular de los ojos, porción orbital	Mejillas levantadas, ojos entrecerrados.
7	Orbicular de los ojos, porción del párpado.	Párpado inferior levantado y estirado medianamente.
9	Elevador del labio superior y del ala de la nariz.	Labio superior levantado e invertido; parte superior de la surco nasolabial más profundo; Fosa nasal dilatada.
10	Elevador del labio superior	Labio superior levantado; surco nasolabial profundizado, se producen surcos cuadrados alrededor de las fosas nasales.
11	Elevador del ángulo de la boca	Parte inferior a media del surco nasolabial profundizada.
12	Cigomático Mayor	Esquinas de los labios hacia arriba y lateralmente.
13	Cigomático Menor	Ángulo de la boca elevado, solo capa de músculos que abre los labios.
14	Buceador	Esquinas de los labios tensas, Mejillas comprimidas contra los dientes.
15	Depresor del ángulo de la boca	Esquina de la boca hacia abajo y hacia adentro.

AU	Músculo Facial	Descripción del Movimiento del Músculo
16	Depresor del labio inferior	Labio interior tirado hacia abajo y lateralmente.
17	Mentoniano	Piel de la barbilla elevada.
18	Orbicular superior de los labios e inferior	Labios fruncidos.
20	Risorio, platisma	Esquina de los labios tiradas lateralmente.
22	Orbicular de la boca	Labios evertidos (canalizados).
23	Orbicular de la boca	Labios apretados.
24	Orbicular de la boca	Labios apretados juntos
25	Depresor del labio inferior, o relajación del mentoniano, o orbicular de la boca	Labios entreabiertos.
26	Masetero, tempotal relajado, pterigoideo interno	Boca abierta.
27	Pterigoideo y digástrico	Boca estirada abierta.
28	Orbicular de la boca	Labios Chupados.
41	Relajación del elevador del párpado superior	Caída del párpado superior.
42	Orbicular de los ojos	Hendidura del párpado.
43	Relajación del Elevador Párpado superior, orbicular de los ojos, porción Palpedral	Ojos cerrados.

AU	Músculo Facial	Descripción del Movimiento del Músculo
44	Orbicular de los ojos, porción Palpedral	Ojos entrecerrados.
45	Relajación del Elevador Parpado superior, orbicular de los ojos, porción Palpedral	Parpadeo.
46	Relajación del Elevador Parpado superior, orbicular de los ojos, porción Palpedral	Guiño.

Tabla A.2: AUs de las Emociones. Traducido de (Kanade et al., 2000)

A.3. Acciones Misceláneas

AU	Descripción del Movimiento
8	Labios tocándose
19	Mostrar lengua
21	Apretar el cuello
29	Forzar la mandíbula.
30	Mandíbula hacia los lados
31	Apretar mandíbula
32	Morder el labio
33	Exhalar aire
34	Exhalar aire

35	Succionar mejillas
36	Protuberación de la lengua
37	Frotar lengua
38	Dilatar las fosas nasales
39	Comprimir las fosas nasales

Tabla A.3: Acciones Misceláneas. Traducido de (Kanade et al., 2000)

A.4. Códigos de Visibilidad

70	Cejas y frente no visibles.
71	Ojos no visibles
72	La región inferior de la cara no se ve
73	La cara entera no se ve
74	No marcable. No se pueden determinar AUs.

Tabla A.4: Códigos de Visibilidad

A.5. Códigos de Intensidad

A	Rastro
B	Ligero
C	Marca Pronunciada
D	Grave
E	Máximo Extremo

Tabla A.5: Intensidad de la microexpresión

A.6. Códigos de Comportamiento

40	Aspirar
50	Discurso
80	Tragar
81	Masticar
82	Encogerse de hombro.
84	Sacudir la cabeza hacia adelante y hacia atrás
85	Cabecear hacia arriba y hacia abajo

Tabla A.6: Códigos de Comportamientos