



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Ingeniería Informática

Trabajo Fin de Grado

**Desarrollo de una Herramienta para
Reconocimiento de Expresiones
Faciales**

Autor: Andrea Velarde Chávez
Tutor(a): Consuelo Gonzalo Martín

Madrid, Julio - 2021

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Ingeniería Informática

Título: Desarrollo de una Herramienta para Reconocimiento de Expresiones
Faciales

Julio - 2021

Autor: Andrea Velarde Chávez

Tutor: Consuelo Gonzalo Martín

Departamento de Arquitectura y Tecnología de Sistemas Informáticos
ETSI Informáticos
Universidad Politécnica de Madrid

Resumen

El reconocimiento de expresiones es un campo de la Inteligencia Artificial cuyas principales finalidades son la contribución a estudios de comportamiento o fines psico-médicos de detección de enfermedades. También está presente en dispositivos o tecnología que utilizamos en nuestra vida cotidiana.

El objetivo principal de este trabajo ha sido la implementación de una herramienta informática para la identificación de expresiones faciales a partir de imágenes o secuencias de imágenes (vídeo). La salida de la herramienta será la categoría (expresión facial) que más se ajuste a la información contenida en la imagen de entrada.

Para lograr dicho objetivo se ha entrenado un modelo clasificador con una BD de imágenes etiquetadas con su categoría correspondiente. En la fase de entrenamiento se establecen relaciones entre la información extraída de las imágenes de entrenamiento y sus respectivas etiquetas. De esta manera, será capaz de predecir la etiqueta de una imagen input que no haya “visto” antes.

Pero ¿quién entrena el modelo?, y ¿cómo lo hace? Aquí entran en escena las redes neuronales convolucionales. Se trata de un conglomerado de “neuronas” encargadas de filtrar las imágenes. Estas neuronas hacen el trabajo de extraer las características que definen y diferencian las expresiones faciales. Para ello se agrupan las neuronas en varias capas que reciben información de las capas anteriores y a su vez la pasan a las siguientes. El modelo aprenderá de todas las características extraídas.

La meta es detectar expresiones faciales por lo que necesitaremos una BD con variedad de imágenes de todas las expresiones faciales que precisemos detectar. Los clasificadores no suelen acertar al 100 % a la hora de devolver el resultado correcto, por lo que se ha recurrido a técnicas como aumentar la variedad de imágenes de la base de datos, para afinar la precisión a la hora de identificar las expresiones.

La metodología que se ha llevado a cabo para la consecución de la herramienta de reconocimiento de expresiones es: en primer lugar, una búsqueda e investigación de toda la información ya existente, en segundo lugar, se han reunido las bases de datos suficientes para la creación de modelos y se ha desarrollado código fuente capaz de procesar dichas bases de datos con redes neuronales, en tercer lugar, se realizaron análisis de los resultados obtenidos y correcciones, en cuarto y último lugar, se hizo una validación final de la eficiencia de la herra-

mienta reflejada en porcentajes de acierto. Además, durante todo el proceso se documentaban en esta memoria final todos los datos obtenidos y el resumen del desarrollo del proyecto.

Los resultados obtenidos de esta herramienta de detección de expresiones han sido buenos, alcanzando un 70,99 % de precisión de aciertos total. Se obtuvieron porcentajes individuales en torno al 80,00 % en algunas de las 6 expresiones. Hay 2 expresiones que ha sido difícil mejorar y que sin embargo siguen teniendo porcentajes bajos, un 32,65 % y un 46,45 %.

Desde el plano personal y profesional de la investigación y realización de este proyecto, he ganado conocimiento en la implementación de sistemas de aprendizaje automático y la implementación de redes neuronales convolucionales. Esto me ha dado la oportunidad de indagar en técnicas utilizadas particularmente para la detección de expresiones faciales. Además de ganar soltura y agilidad en el desarrollo de código fuente en lenguaje Python, también he conseguido mejorar mi organización a la hora de trabajar.

Me siento satisfecha de los resultados del proyecto y de haber puesto mi pequeño grano de arena en uno de los campos de la Inteligencia Artificial que más se está investigando y desarrollando a día de hoy, el Machine Learning.

Abstract

The expression recognition is a field of Artificial Intelligence which main purposes are the contribution to studies on behavior or psycho-medical purposes on disease detection. It's also found in devices and technology we use on our daily life.

The main goal of this project was the implementation of a software tool for the identification of facial expressions from images or image sequences (video). The output of this tool would be the category (facial expression) that best matches the input image information.

To achieve this goal, a classifier model has been trained with a database of images labeled with their corresponding category. In the training phase, the relation between the extracted information from training images and their respective labels are established. In this way, it would be capable of predicting the label of an input image that it has never "seen" before.

But who trains the model?, and how they do it? Here is when convolutional neural networks come on stage. These are a conglomerate of "neurons" in charge of filtering the images. The job of these neurons is to extract the features that define and differentiate the facial expressions. The neurons are grouped into several layers that receive information from previous layers and pass it on to the following layers. The model will learn from all the extracted features.

The purpose is to detect facial expressions, so we need a database with variety of images of all the facial expressions we need to detect. The classifiers do not always get the 100% of the results correct, so we went to techniques such as increasing the variety of images in the database, to improve the accuracy in identifying expressions.

The methodology carried out for the achievement of an expression recognition tool is: first, a search and study of all the existing information, secondly, collect enough databases for the creation of models and development of source code capable of processing these databases with neural networks, thirdly, analysis of the results obtained and corrections, fourthly and finally, a final validation of the efficiency of the tool shown as percentages of success. In addition, during the whole process, all the collected data and the summary of the project were documented in this final report.

The results of this expression recognition tool have been good, reaching a total

accuracy of 70.99%. It has reached individual scores around 80.00% in some of the 6 total expressions. There are 2 expressions that have been difficult to improve and still have low scores, 32.65% and 46.45%.

From the personal and professional point of view of the research and development of this project, I have acquired knowledge in the development of machine learning systems and implementation of convolutional neural networks in. This gave me the opportunity to research techniques used particularly for facial expression recognition. In addition to gaining fluency and agility in source code development in Python language, I have also managed to improve my organization when working.

I'm happy with the project results and to have made my contribution to one of the most researched and developed fields of Artificial Intelligence today, Machine Learning.

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	3
2. Estado del arte	5
3. Desarrollo	13
3.1. Análisis de Bases de Datos públicas	13
3.2. Comparativa de Bases de Datos	27
3.3. Redes Convolucionales	29
3.4. Diseño del prototipo	31
3.5. Implementación del prototipo	32
3.5.1. Herramientas software	32
3.5.2. Fases de desarrollo	33
3.5.3. Revisión y corrección de errores	39
4. Resultados y conclusiones	41
4.1. Resultados	41
4.2. Validación	47
4.2.1. Enfado	47
4.2.2. Repugnancia	48
4.2.3. Neutral	48
4.2.4. Felicidad	49
4.2.5. Tristeza	50
4.2.6. Sorpresa	50
4.3. Conclusiones	51
Bibliografía	54
Anexos	55

Capítulo 1

Introducción

La expresión facial es una de las formas más importantes que tenemos las personas de mostrar nuestros sentimientos. En 1971 el psicólogo Paul Ekman estableció 6 expresiones faciales de emoción humanas que son básicas y universales en cualquier cultura: alegría, enfado, asco, miedo, tristeza y sorpresa [1].

La información recogida a partir de las expresiones revela emociones de una persona que nos sirven para identificar su estado de ánimo. Por lo tanto, las expresiones son un aspecto importante a la hora de comunicar nuestros sentimientos y, de esa manera, interactuar con otras personas. No sólo entendemos la situación de otro a través de una comunicación oral sino también a través de lo que nos transmite su rostro. Con la correcta interpretación de la expresión facial podemos saber cómo se encuentra de ánimo una persona o si tiene alguna molestia o preocupación, y sabiendo esto podemos tratar de ayudarlos.

El campo de la Inteligencia Artificial (IA) estudia el reconocimiento de las expresiones faciales con fines científicos y de salud. Un sistema de reconocimiento de expresiones aprende a relacionar ciertas expresiones con sus características identificativas. Para esto se trabaja con grandes cantidades de imágenes anotadas de las que se extrae la información que necesitamos para entrenar el clasificador de expresiones. Este trabajo tratará del estudio y desarrollo de un sistema automático de reconocimiento de expresión facial (FER) que de forma precisa y eficiente sepa identificar la expresión de un rostro. Dicha herramienta detectará rasgos que determinan y distinguen unas expresiones de otras a partir de fotografías de rostros. Este tipo de sistemas se aplica a campos como la interacción persona-ordenador, análisis de emociones humanas, indexación automática, etc.

Según el artículo [2], el reconocimiento de expresiones faciales se aplica a muy diferentes campos como son: el diagnóstico médico (detectar enfermedades), fines científico-sociológicos (análisis de reacciones ante un producto o evento), animación 3D (videojuegos) o análisis de expresión que facilite actividades de los smartphones (disparo de cámara automático con detección de sonrisa [3]). Además, muchos de nuestros dispositivos móviles disponen de reconocimiento facial, no sólo como opción de autenticación, sino también para ser usado

por otras aplicaciones como Instagram, Facebook o Snapchat. El medio de noticias Quartz [4], expone que Snapchat utiliza la técnica de redes neuronales convolucionales para generar sus filtros, véase la Figura 1.1. Otro ejemplo, es la tecnología, Animoji de los dispositivos de Apple. En 2016 [5], Apple compró Emotient, startup dedicado a reconocer las emociones, con el fin de perfeccionar alguno de sus productos o entrar al campo del ocio y entretenimiento, un año después en 2017 Apple saca la tecnología Animoji [4]. En la Figura 1.2 se muestra la tecnología Animoji de seguimiento facial.



Figura 1.1: Filtro de Snapchat en diferentes rostros.
(Fuente: [6])

Otros usos del reconocimiento de expresiones se dan en el ámbito del marketing. Un ejemplo es la compañía Affective, especializada en análisis de emociones y reacciones humanas. Sus técnicas permiten, mediante el análisis de las emociones de los voluntarios en el primer contacto con el producto, predecir si ese producto va a tener mejor o peor aceptación [7].

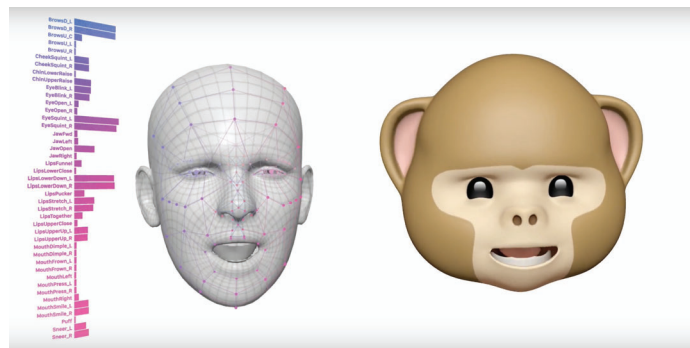


Figura 1.2: Tecnología Animoji. (Fuente: [8])

Como ya se ha comentado, el reconocimiento facial también es utilizado en apli-

Introducción

caciones médicas para cubrir las necesidades de salud mental, cada vez más fuertes debido a la pandemia mundial del Covid-19 a la que nos enfrentamos actualmente. Un claro ejemplo es el Robot Robin, véase Figura 1.3, que el hospital “UCLA Mattel Children’s” planea utilizar a mediados de Julio de 2021 [9]. Robin estará capacitado para dar a los niños atención y compañía. Su sistema de Inteligencia Artificial se basa en reconocer las expresiones faciales, interpretar dichas expresiones y dialogar para responder a las necesidades de los niños.



Figura 1.3: Robot Robin de diseño amigable para captar la atención de los niños. (Fuente: [10])

De todas estas aplicaciones del reconocimiento facial en diversos ámbitos de la investigación y desarrollo de herramientas necesarias, surge la necesidad y motivación de crear un sistema en código abierto capaz de cumplir con la interpretación de emociones y adaptable al análisis de ciertos grupos de población: niños, ancianos, enfermos u otros grupos de interés.

1.1. Objetivos

El principal objetivo de este TFG es la creación de una herramienta capaz de identificar de forma automática expresiones faciales básicas a partir de imágenes o secuencias de imágenes(vídeo) de rostros. Este objetivo global se puede desglosar en los siguientes objetivos específicos:

- Familiarización con el manejo y uso de imágenes digitales.
- Documentar en un estado del arte la información que se tiene hasta día de hoy del reconocimiento de expresiones faciales.
- Diseñar y desarrollar una herramienta capaz de reconocer expresiones faciales básicas.
- Realizar una validación del sistema realizado y analizar resultados obtenidos.

Capítulo 2

Estado del arte

A lo largo de los años, en el reconocimiento de expresiones faciales se han investigado y desarrollado diferentes técnicas de detección de rasgos faciales capaces de clasificar los rostros según su emoción.

En un estudio, realizado en el año 2000, [11] se propuso analizar los movimientos faciales, a partir de secuencias de imágenes, y su posterior descomposición en unidades de acción. Las unidades de acción son todos los posibles movimientos faciales detectados visualmente. Véase la Figura 2.1 donde se muestran unidades de acción de la parte superior e inferior del rostro. Una o varias unidades de acción componen la expresión de una emoción. El objetivo de la investigación era el reconocimiento de unidades de acción con las que poder identificar fácilmente una expresión. Se trataron las siguientes técnicas:







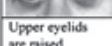
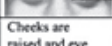
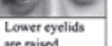

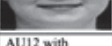
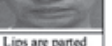
Upper Face Action Units			Lower Face Action Units		
AU4	AU1+4	AU1+2	AU25	AU26	AU27
					
Brows lowered and drawn together	Medial portion of the brows is raised and pulled together	Inner and outer portions of the brows are raised	Lips are relaxed and parted	Lips are relaxed and parted; mandible is lowered	Mouth is stretched open and the mandible pulled down
AU5	AU6	AU7	AU12	AU12+25	AU20+25
					
Upper eyelids are raised	Cheeks are raised and eye opening is narrowed	Lower eyelids are raised	Lip corners are pulled obliquely	AU12 with mouth opening	Lips are parted and pulled back laterally

Figura 2.1: Unidades de acción de la parte superior (cuadro a la izquierda) e inferior (cuadro a la derecha) del rostro. (Fuente: en el artículo [11])

1. Obtención de información a partir de las diferencias de intensidad en secuencias de imágenes, es decir, información a partir del movimiento y cambio de expresión en un rostro. Para ello se utiliza un pequeño número de fotogramas, suficientes para apreciar unidades de acción en las zonas de las cejas y ojos. Como se puede ver en la Figura 2.2, la primera y tercera

fila son imágenes que muestran el movimiento facial del bebé desde una expresión neutral a una expresión alegre, en la segunda y cuarta fila se resaltan las zonas faciales que se han movido. El problema de esta técnica era que, las imágenes en las que se recoge el movimiento o diferencia de intensidad degradan la calidad de los píxeles y con ello el reconocimiento de unidades de acción.

2. Extracción de líneas de expresión o arrugas en zonas como las cejas, párpados, labios y surcos naso-labiales.

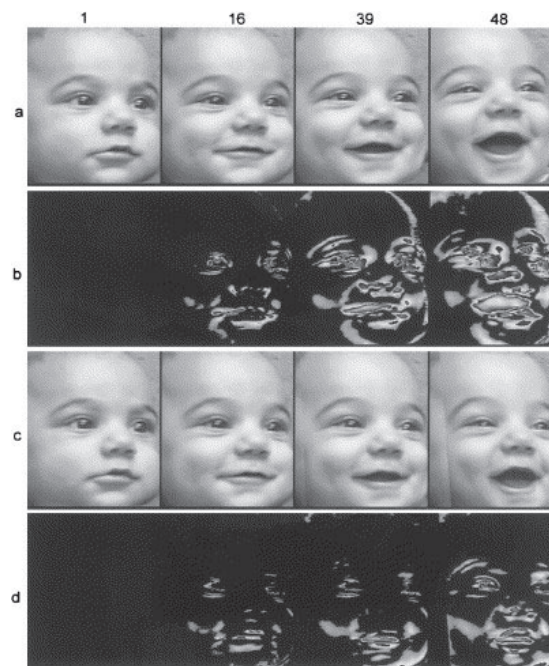


Figura 2.2: Secuencia de imágenes que graban el movimiento facial desde una expresión neutral a alegre.
(Fuente: en el artículo [11])

En la bibliografía se pueden encontrar un número considerable de trabajos en esta línea de investigación basados en los clasificadores Haar. Para una adecuada descripción de estos trabajos, se va a presentar a continuación una breve descripción de las características Haar.

Las características Haar [12] se puede definir como la diferencia de intensidad de zonas adyacentes de la imagen con diferentes configuraciones que permiten detectar cambios de intensidad en diferentes orientaciones. En la Figura 2.3, se presentan 2 características: la primera a la derecha se refiere a que la nariz suele ser una zona más iluminada que los ojos por eso la característica Haar serán 3 rectángulos verticales siendo el de en medio el rectángulo blanco. La imagen situada en el centro de la Figura 2.3, indica la característica de que los ojos son una zona más oscura que las mejillas por lo tanto se utilizan 2 rectángulos tumbados, el blanco en las mejillas y el oscuro en los ojos.



Figura 2.3: Características Haar. (Fuente [13])

Se suman los píxeles del rectángulo blanco y se restan a la suma de píxeles del negro para obtener un solo valor, ese valor será una característica. Puede variar los tamaños de estos rectángulos y la posición, por ejemplo, en un cuadro de 24x24 píxeles se pueden encontrar hasta 180.000 características posibles. Las características serán distintas según los rectángulos y su colocación respecto al cuadro de búsqueda. Las características que nos interesan son las que se repitan y nos aporten patrones de búsqueda. Para detectar partes clave del rostro se unen características individuales. Por ejemplo, las características de la boca mostradas en la Figura 2.4 definen puntos clave de una sonrisa, junto con las características que se obtengan de otras imágenes de personas que sonrían, formarían un conjunto de filtros que con un algoritmo de aprendizaje automático (por ejemplo, una red neuronal convolucional) servirían como entrenamiento de un clasificador de detección de una sonrisa.



Figura 2.4: Características Haar en la boca. (Fuente: [14])

Se podrían probar todos los tamaños de ventana de búsqueda de características y ubicaciones en la imagen que sea posible, sin embargo, esto supondría un coste computacional demasiado elevado, la complejidad sería $O(N^2)$. Las características Haar se pueden calcular rápidamente mediante la técnica de las

imágenes integrales para reducir el coste computacional.

Las imágenes integrales se crean a partir de la imagen original, desde la que se hace la búsqueda de características Haar. Para cada píxel, se suman los valores de todos los píxeles que queden encima y a su izquierda. De esta manera se obtiene el valor de un rectángulo de manera más simple. Gracias a las imágenes integrales se reduce la complejidad del algoritmo de $O(N^2)$ a $O(1)$.

En la Figura 2.5 se muestra una simulación de como se obtiene la imagen integral a partir de una imagen original, en la imagen original el valor de la suma de los píxeles dentro del rectángulo naranja se puede obtener de manera más rápida gracias a la imagen integral con una simple resta de valores.

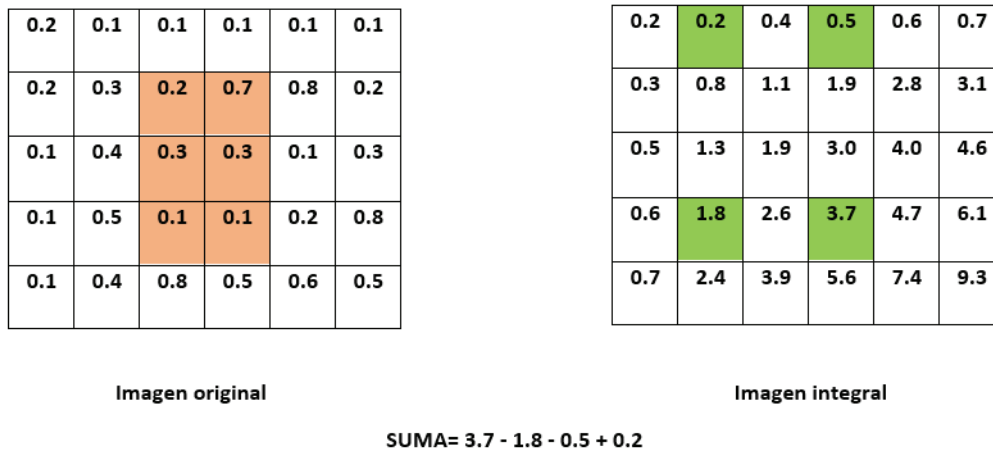


Figura 2.5: Cálculo de la zona naranja en imagen original a partir de la imagen integral.

Una vez obtenidas las características mediante técnicas rápidas como la imagen integral, se necesitará asignar las características a una clase o categoría (en nuestro caso las clases son las expresiones), de modo que agrupemos características que identifiquen si una imagen es de una clase u otra, es decir, que expresión representa: feliz, triste, asustada, etc.

Se introducen ahora los conceptos de boosting o clasificadores en cascada. El algoritmo AdaBoost (“Adaptive Boosting”), creado por Freund y Schapire, es un clasificador que aprende de la combinación de clasificadores más débiles [15]. El algoritmo entrena por separado cada clasificador débil y cada clasificador trabajará con los datos clasificados erróneamente por el anterior. En cada iteración se calcula la tasa de error de todas las características y se escoge la característica con menor tasa de error ya que son las ideales para la detección del rostro. De esta manera el algoritmo alcanza una alta tasa de acierto.

Se toman un conjunto de características seleccionadas entre muchísimas que previamente se obtienen de cada imagen de entrenamiento. Los clasificadores débiles del AdaBoost se centrarán en una única característica, por lo que en cada

iteración se escogerá una característica nueva como nuevo clasificador débil. Cuando todos los clasificadores débiles no detecten un rostro en una imagen, esta imagen será negativa. En cambio cuando un clasificador débil reconoce una región de la imagen como rostro, esa región pasa al siguiente clasificador débil, de esta manera se seleccionan las características.

En 2001 se presenta un nuevo algoritmo, denominado “Viola-Jones” [16] [17], éste usa el algoritmo AdaBoost para su fase de entrenamiento con el fin de elegir, entre un conjunto de filtros, las características más convenientes para el clasificador. Los conceptos explicados de características Haar e imágenes integrales son usados en el Viola-Jones para preparar los datos con los que trabajaremos. El algoritmo consta de 2 fases:

1. **Aprendizaje:** Esta fase consiste en entrenar el algoritmo para que sea capaz de identificar un objeto (en este caso un rostro) en una imagen. Se utilizan imágenes en las que aparecen rostros, consideradas como positivas, e imágenes sin rostros, consideradas negativas. De estas imágenes aprende qué características son más probables de un rostro y cuáles no.
2. **Detección:** Primero se trata la imagen convirtiendo los colores a blanco y negro y, si fuera necesario, se recorta la imagen para seleccionar la zona donde se encuentra el rostro, la razón de esto es que reduce el número de datos a procesar. Tras adaptar la imagen, se procede a extraer las características Haar en la imagen. Con todas las características Haar encontradas, el algoritmo determina si es el objeto que estamos buscando. Como último paso se remarcará el área de interés.

En la Figura 2.6 se muestra el proceso de la clasificación en cascada, el input o imagen de entrada pasa por el primer clasificador y en caso de que detecte un rostro continúa al siguiente clasificador hasta pasar por el último.

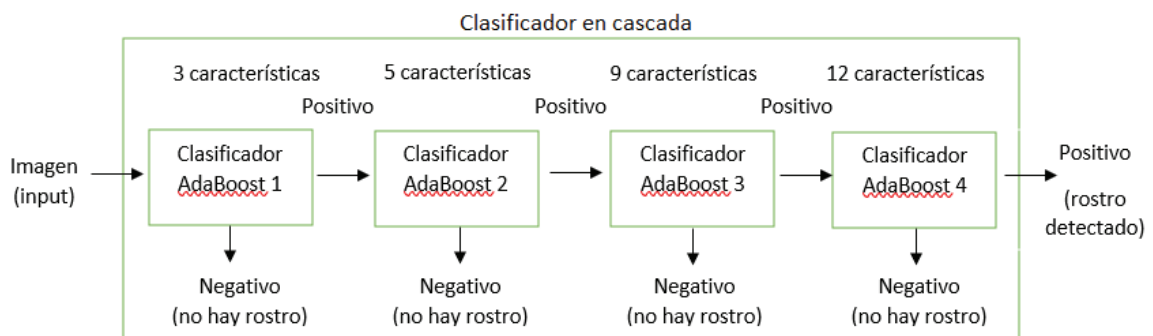


Figura 2.6: Clasificador en cascada formado por 4 clasificadores más débiles. Los 4 clasificadores seleccionan las características Haar con el algoritmo AdaBoost (p. ej. en el “Clasificador AdaBoost 1” son 3 características).

Los clasificadores “Haar Cascade” [18] aplican los conceptos de clasificadores en cascada usando características Haar y selección con AdaBoost. La librería OpenCV contiene clasificadores de detección de rostro ya entrenados, funciona mediante ficheros XML que contienen los datos del objeto a detectar [19].

Sin embargo, nuestro objetivo es, no sólo la detección del rostro, sino también el reconocimiento de su expresión. Por lo que, una vez detectado el rostro en una imagen, habrá un segundo proceso para detectar su expresión. En este proyecto se van a considerar 8 tipos de expresiones: felicidad, tristeza, sorpresa, ira, miedo, disgusto, neutral y desprecio. Al ampliar la variedad de posibles expresiones nos encontramos que algunas son muy parecidas, lo que puede reducir la precisión de la clasificación. Según [20], donde se estudia la clasificación de expresiones por medio de imágenes usando librerías desarrolladas en Python, las expresiones de “repugnancia” se clasificaron erróneamente como “triste” y las expresiones de “sorpresa” se clasificaron como “feliz”. Su solución fue incrementar el número de imágenes para las expresiones con baja tasa de aciertos, el resultado fue el aumento desde un 70 % de aciertos a un 80 %.

Las redes neuronales convolucionales (CNNs) están tomando protagonismo en el reconocimiento de expresiones faciales [21]. Este tipo de redes se usan especialmente para el tratamiento de imágenes o vídeos. Las redes neuronales en general, son un conjunto estructurado de neuronas cuya información se transmite de unas a otras. Se agrupan en capas, la salida de unas será la entrada de otras evitando caer en bucles infinitos. En [22], se propone una CNN para la detección de expresiones. Esta CNN tiene una capa de entrada que tiene tantas entradas como píxeles tenga la imagen. Luego se encontrarán otras capas llamadas ocultas, a las que llegará la información que pasen las capas de entrada. En las capas ocultas se podrán hacer 2 operaciones: pooling, para redimensionar la imagen a un tamaño más pequeño, y las convoluciones, que filtran la imagen para determinar patrones de interés.

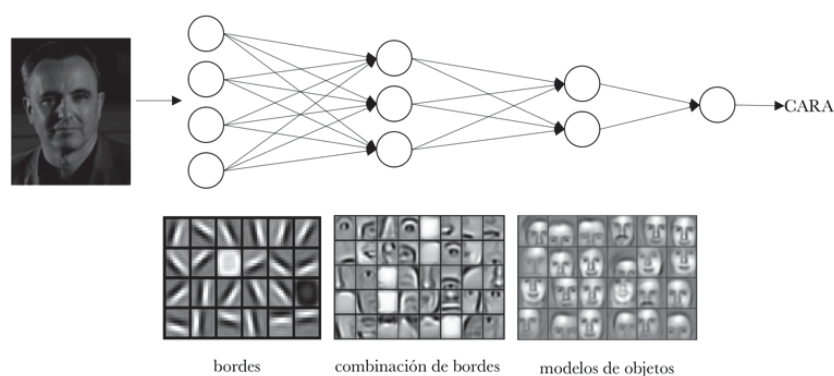


Figura 2.7: Ejemplo de red neuronal convolucional de 4 capas. (Fuente: en el artículo [23])

En cada capa la imagen o la salida de la capa anterior, depende del nivel de la capa, es convolucionada con diferentes filtros que identificarán diferentes obje-

tos. En la Figura 2.7 se muestra una red neuronal convolucional, mucho más pequeña de la que será necesaria para llevar a cabo el objetivo por lo que ésta sólo se presenta para aportar una explicación. A la izquierda en la Figura 2.7, nos encontramos con la capa de entrada, las imágenes de salida de esta primera capa no nos aportan información significativa. Después de las convoluciones, de las siguientes 2 capas ocultas, obtendremos características (ojos, nariz o boca) más próximas a lo que buscamos hasta obtener los rostros. La última capa es la capa de salida final para clasificar la imagen, con la misma cantidad de neuronas que de categorías a detectar.

Gracias a los estudios e investigaciones nos damos cuenta de rasgos a los que prestar más atención a la hora de elegir una expresión y cómo esa expresión nos da información de los sentimientos de la persona. Ejemplo de esto es un estudio del año 2018, “Survey on Human Face Expression Recognition Techniques” (“Investigación sobre técnicas de reconocimiento de expresión facial humana”) [24], en el que se tomaron en cuenta patrones como, por ejemplo:

- La sonrisa junto con ojos ligeramente curvados son un rasgo característico de la felicidad. Esta expresión es la más fácil de reconocer [25].
- La expresión de tristeza se caracteriza por cejas caídas y fruncimiento del ceño.
- La expresión de enfado se reconoce por cejas inclinadas hacia el centro del rostro, aletas de la nariz hinchadas, surcos nasogenianos notorios, forma del párpado estirada. El enfado es parte del sentimiento de ira o furia.
- La expresión de repugnancia se reconoce por cejas hacia abajo y nariz arrugada. Característico de sentimientos de desagrado hacia algo visible, algún olor o sabor. También puede deberse al rechazo a alguna idea o acto.
- La expresión de sorpresa se reconoce por boca y ojos abiertos. Puede deberse a acontecimientos que la persona no se espera.
- La expresión de susto se puede confundir con la de sorpresa. El susto se puede diferenciar de la sorpresa por las cejas, en el susto las cejas se ven más apretadas en cambio en la sorpresa son más curvadas hacia el exterior del rostro. Es una expresión ligada a sentimientos de alarma y sobresalto ante un peligro real o imaginario.

La mayoría de los sistemas software de detección de emociones trabajan con empresas que quieran lanzar un producto y quieran hacer una investigación o seguimiento previos sobre la reacción e interacción de las personas con el producto o servicio. También hay mucho código disponible en Github de desarrolladores de diferentes partes del mundo, que aportan su grano de arena a la detección de emociones sin depender de herramientas hardware complejas. Algunos de estos sistemas software son:

ParallelDots [26]: herramienta online que permite reconocer las emociones de un sujeto, se recomienda usar para monitorizar emociones respecto a un producto o contenido en una red social. Tiene usos en el diseño de videojuegos (adaptarse a la experiencia del jugador), industria automóvil (mantener despierto

al conductor si se encuentra somnoliento) o en entrevistas laborales (estrategias de reclutamiento).

Noldus FaceReader [27]: herramienta que detecta la expresión del rostro de manera muy detallada: expresiones básicas (véase la Figura 2.8), expresiones personalizadas, orientación de la cabeza, dirección de la mirada, características de la persona, valencia(sentimiento de aversión o atracción que un individuo siente hacia algo) y excitación, unidades de acción, frecuencia cardíaca y variabilidad de la frecuencia cardíaca, audio y comportamiento de consumo. Según un estudio de validación reciente, FaceReader 6 muestra el mejor rendimiento de entre las principales herramientas de software para la clasificación de emociones disponibles actualmente, con una media del 88 %.

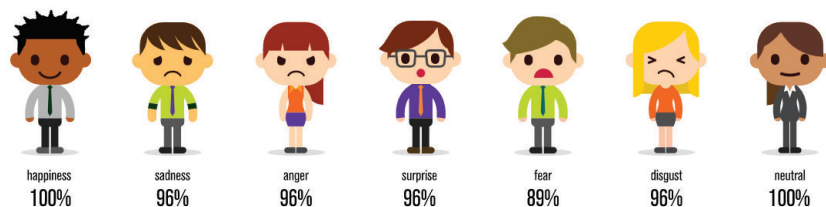


Figura 2.8: Expresiones detectables con Noldus FaceReader. (Fuente: [27])

IMOTIONS [28]: software que sirve como plataforma de análisis del comportamiento humano. Integra sensores biométricos para investigación de comportamiento humano entre los que está el reconocimiento de expresión facial.

FaceAnalysis [29]: detecta rostros en imágenes o vídeos y con análisis facial y unidades de acción consigue devolver el género, emoción y edad del rostro. Las expresiones faciales que maneja son las 6 emociones universales: felicidad, tristeza, enfado, miedo, sorpresa, repugnancia y añade la expresión neutral. Puede analizar más de un rostro a la vez, las expresiones se muestran como diagramas de barras mostrando la estimación de cada expresión.

Affectiva [30]: software que detecta emociones, estados cognitivos, comportamientos, interacciones, etc. . . . Utiliza grandes cantidades de información con bases de datos compuestas de más de 4 billones de fotogramas tomados de 7.5 millones de rostros de 87 países diferentes. Hecho sobre una arquitectura de aprendizaje supervisado, de redes neuronales convoluciones (imágenes) y recurrentes (texto).

OpenFace [31]: herramienta destinada a investigadores y personas interesadas en construir aplicaciones interactivas basadas en el análisis del comportamiento facial. OpenFace es el primer conjunto de herramientas capaz de detectar puntos de referencia faciales, estimar la postura de la cabeza, reconocer la unidad de acción facial y estimar la mirada, con el código fuente disponible para ejecutar y entrenar los modelos. Es capaz de funcionar en tiempo real y puede ejecutarse desde una simple cámara web sin ningún hardware especializado.

Capítulo 3

Desarrollo

3.1. Análisis de Bases de Datos públicas

Para poder entrenar la mayoría de los clasificadores y en particular las CNN se requiere de la disponibilidad de un gran número de imágenes variadas, de personas de diferente edad, raza o género. Por lo que el acceso a grandes “Datasets” que faciliten estas imágenes es un factor determinante en la generación de modelos robustos para la identificación de emociones. En la siguiente tabla, Tabla 3.1, se ha incluido un resumen de las principales bases de datos de imágenes públicas.

Tabla 3.1: Bases de Datos públicas.

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
fer2013 [26]	Datos divididos en conjuntos de entrenamiento (80 %), prueba (10 %) y validación (10 %)	28.709	Enfado, disgusto, miedo, felicidad, tristeza, sorpresa y neutral	https://www.kaggle.com/astraszab/facial-expression-dataset-image-folders-fer2013?

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
IMPA-FACE 3D [27]	Se pueden elegir las personas de las que quieras las fotos, la expresión y el formato de descarga. Las fotos están tomadas desde varios ángulos. También permite la descarga de la BD entera.	A elegir	Neutral, alegría, tristeza, sorpresa, enfado, asco y miedo	http://app.visgrafimpa.br/database/faces/download-with-email/

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Radbound DB [28]	Esta BD contiene imágenes de 8 tipos de expresiones modeladas por 67 personas (hombres y mujeres de diferentes edades y rasgos caucásicos, y hombres de origen marroquí - alemán). Cada emoción está capturada con 3 direcciones de la mirada y 5 ángulos fotográficos distintos. Se necesita registrarse en la web para que permita la descarga.	231	Enfado, disgusto, miedo, felicidad, tristeza, sorpresa, desprecio y neutral	http://www.socsci.ru.nl:8180/RaFD2/RaFD
JAFFE [29]	Varias imágenes de expresiones de 10 modelos de origen japones.	213	Neutral, tristeza, sorpresa, felicidad, miedo, enfado, y disgusto	https://zenodo.org/record/3451524#.X3o8-GgzZPY

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
FERG-3D-DB [30]	Es una base de datos en la que los modelos son personajes creados bajo diseño 3D, estilizados con expresiones faciales. La base de datos contiene 39574 ejemplos de cuatro personajes diseñados (2 mujeres y 2 hombres): Mery, Bonnie, Ray y Malcolm. Cada ejemplo es una lista de valores de parámetros de plataforma que cuando se transfieren a la plataforma 3D crea una expresión facial particular.	39574	Ira, disgusto, miedo, alegría, neutral, tristeza y sorpresa.	http://grail.cs.washington.edu/projects/deepexpr/ferg-3d-db.html

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
RAVDESS [31]	Es una base de datos audiovisual. Contiene 7356 archivos (tamaño total: 24,8 GB). La base de datos contiene audios de 24 actores profesionales con un acento norteamericano neutral (12 mujeres, 12 hombres). El material está disponible en tres formatos: solo audio, audio y video y solo video (sin sonido).	7356	Calma, felicidad, tristeza, enojo, miedo, sorpresa y disgusto	https://zenodo.org/record/1188976#.X3pAlmgzZPY

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Base de datos de expresión facial de MMI [32]	Esta base de datos se concibió como un recurso para construir y evaluar algoritmos de reconocimiento de expresiones faciales. En particular, contiene grabaciones del proceso completo hasta que un rostro hace una expresión facial. En segundo lugar, contiene expresiones con una sola Unidad de acción (AU) de FACS activada, para todas las AU existentes y muchos otros descriptores de acción.	2900	—	http://mmifacedb.eu/
Belfast Database [33]	Cuenta con 3 sets de videoclips.	3 sets (video): 570, 650 y 280	Disgusto, miedo, diversión, frustración, sorpresa y enfado	https://belfast-naturalistic-db.sspnet.eu/

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
DISFA [34]	Los modelos son en total 27 (12 hombres y 15 mujeres) de distinta etnia. El material de esta base de datos está valorado por 2 expertos en FACS (Sistemas de desarrollo de acción facial). La base de datos consta de más de 2900 videos e imágenes fijas de alta resolución de 75 sujetos. Está completamente anotado para la presencia de AU en videos (codificación de eventos) y parcialmente codificado a nivel de cuadro, lo que indica para cada cuadro si un AU está en la fase neutra, de inicio, vértice o de desplazamiento.	4.845	Expresiones faciales espontáneas	http://www.engr.du.edu/mahoor/DISFA.htm

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Indian Spontaneous Expression Database (ISED) [35]	Se grabó un video casi frontal para 50 participantes de la India mientras veían videoclips emocionales. Contiene emociones. Los videoclips fueron anotados cuidadosamente por cuatro decodificadores entrenados, que fueron validados por la naturaleza de los estímulos utilizados.	480 vídeos	Felicidad, sorpresa, tristeza y repugnancia	https://sites.google.com/site/iseddatabase/
Multi media Understanding Group (MUG) [36]	Secuencias de imágenes de 86 modelos posando.	1462 secuencias de imágenes	Neutral, tristeza, sorpresa, felicidad, miedo, enfado y disgusto	https://mug.ee.auth.gr/

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Oulu-CASIA NIR&VIS [37]	La base de datos de expresiones faciales Oulu-CASIA NIR & VIS contiene videos de seis expresiones de 80 sujetos capturados con dos sistemas de imágenes, NIR (infrarrojo cercano) y VIS (luz visible), bajo tres condiciones diferentes de iluminación: iluminación interior normal, iluminación débil (solo la pantalla del ordenador está encendida) e iluminación oscura (todas las luces están apagadas).	2880 secuencias de imagen	Felicidad, tristeza, sorpresa, ira, miedo y disgusto	http://www.cse.oulu.fi/wsgi/CMV/Downloads/Oulu-CASIA

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
FERG (Facial Expression Research Group Database) [38]	Basse de datos de 55767 imágenes en pose frontal de 6 sujetos en 2D diseñados por ordenador.	55767	Enfado, repugnancia, miedo, diversión, neutral, tristeza y sorpresa	http://grail.cs.washington.edu/projects/deepexpr/ferg-2d-db.html

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Base de Datos facial FEI [39]	<p>La base de datos de rostros de FEI es una base de datos de rostros brasileña. Hay 14 imágenes para cada uno de los 200 individuos, un total de 2800 imágenes. Todas las imágenes son coloridas y se toman sobre un fondo blanco homogéneo en una posición frontal vertical con una rotación de perfil de hasta aproximadamente 180 grados. Todos los rostros están representados principalmente por estudiantes y personal de FEI, entre 19 y 40 años con apariencia, peinado y adornos distintos. El número de sujetos masculinos y femeninos es exactamente el mismo e igual a 100.</p>	2800	Neutral y sonriendo	https://fei.edu.br/~cet/facedatabase.html
		23		

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
AffectNet [40]	AffectNet contiene más de 1 millón de imágenes faciales recopiladas de Internet mediante la consulta de tres motores de búsqueda principales utilizando 1250 palabras clave relacionadas con las emociones en seis idiomas diferentes. Es, con mucho, la base de datos más grande de expresiones faciales, valencia y excitación en la naturaleza, lo que permite la investigación en el reconocimiento automático de expresiones faciales en dos modelos de emociones diferentes. Se utilizan dos redes neuronales profundas de referencia para clasificar las imágenes en el modelo categórico.	450.000 manual-mente tomadas. 500.000 automá-ticamente tomadas. 534 imágenes estáticas	Neutral, felicidad, tristeza, sorpresa, miedo, repugnancia, enfado y desprecio.	http://mohammadmahoor.com/affectnet/

Desarrollo

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Aff-Wild [41]	Los fotogramas de la base de datos Aff-Wild muestran sujetos en diferentes estados emocionales, de diferentes etnias, en una variedad de poses de cabeza, condiciones de iluminación y oclusiones.	1.250.000	—	https://ibug.doc.ic.ac.uk/resources/first-affect-wild-challenge/

3.1. Análisis de Bases de Datos públicas

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
Cohn-Kanade (CK+) [42]	La base de datos de expresión facial codificada por AU de Cohn-Kanade ofrece un banco de pruebas para la investigación en el análisis automático de imágenes faciales. Los datos de imágenes constan de aproximadamente 500 secuencias de imágenes de 100 sujetos. Incluyen la anotación de unidades de acción FACS y expresiones específicas de emoción. Los sujetos tienen edades comprendidas entre los 18 y los 30 años. El sesenta y cinco por ciento eran mujeres; El 15 por ciento eran afroamericanos y el 3 por ciento asiáticos o latinos.	981	Alegría, sorpresa, ira, miedo, disgusto y tristeza	https://www.kaggle.com/shawon10/ckplus

Nombre	Descripción	Nº imágenes	Lista de expresiones	Link
The Karolinska Directed Emotional Faces (KDEF) [43]	Desarrollada por motivos psico-médicos. 70 individuos modelan 7 expresiones diferentes desde 5 ángulos fotográficos.	4.900	Ira, repugnancia, miedo, felicidad, neutral, tristeza y sorpresa	https://www.emotionlab.se/kdef/

3.2. Comparativa de Bases de Datos

La mayoría de las bases de datos idóneas para este campo requieren hacer una petición que provenga de un organismo oficial y aceptar términos de confidencialidad y normas para el almacenamiento de esos datos. Por lo que de todas las bases de datos expuestas en la Tabla 3.1, se ha trabajado con las que tienen permisos menos estrictos o inexistentes en cuanto a términos de uso y requisitos de afiliación, o su respuesta a mi petición de descarga era rápida.

Contamos con 4 bases de datos: Fer2013 (Tabla 3.2), Cohn-Kanade (Tabla 3.3), JAFFE (Tabla 3.4) y KDEF (Tabla 3.5). Las expresiones que todas manejan son: “enfado”, “repugnancia”, “miedo”, “felicidad”, “tristeza” y “sorpresa”. Tres de ellas tienen la expresión “neutral”, y sólo una maneja la expresión “desprecio”. A continuación, se pueden ver el número de imágenes que componen los sets dedicados a entrenamiento (“train”) y validación (“test”), que al mismo tiempo están subdivididos en las expresiones de su respectiva base de datos. Únicamente la base de datos Fer2013 viene ya dividida en dichas carpetas. Para las restantes hice una repartición aproximada del 80 % para “training” y el 20 % para “testing”.

Tabla 3.2: Distribución de imágenes de base de Datos "Fer2013".

Expresión	Entrenamiento	Validación
Enfado	267	72
Repugnancia	42	12
Miedo	167	51
Felicidad	819	206
Neutral	478	127
Tristeza	169	54
Sorpresa	182	48
Total	2.124	570

3.2. Comparativa de Bases de Datos

Tabla 3.3: Distribución de imágenes de Base de datos “Cohn – Kanade”.

Expresión	Entrenamiento	Validación
Enfado	108	27
Desprecio	43	11
Repugnancia	142	35
Miedo	60	15
Felicidad	166	41
Tristeza	68	16
Sorpresa	200	49
Total	787	194

Tabla 3.4: Distribución de imágenes de Base de datos “JAFPE”.

Expresión	Entrenamiento	Validación
Enfado	24	6
Repugnancia	23	6
Miedo	25	7
Felicidad	24	7
Neutral	24	6
Tristeza	24	7
Sorpresa	24	6
Total	168	45

Tabla 3.5: Distribución de imágenes de Base de datos “KDEF”.

Expresión	Entrenamiento	Validación
Enfado	112	28
Repugnancia	112	28
Miedo	112	28
Felicidad	112	28
Neutral	112	28
Tristeza	112	28
Sorpresa	112	28
Total	784	196

3.3. Redes Convolucionales

Para extraer las características que diferenciarán unas expresiones de otras entra en juego la red neuronal convolucional. Esta red neuronal se trata de una red de capas por las que irán pasando las imágenes de entrenamiento, de una o más bases de datos, y devolviendo unas características de salida(output) que serán la entrada(input) de las siguientes capas.

Cabe aclarar que el diseño de la red neuronal convolucional presentado a continuación no es un diseño propio sino un diseño para modelos de bases de imágenes como las que usaremos, disponible en la plataforma Github [44]. La red neuronal, que hemos utilizado para nuestro proyecto, usa la biblioteca de redes neuronales “Keras” y la API Sequential para crear su secuencia de capas.

Como se puede observar, en la Figura 3.1, la red se construye de hasta 4 capas convolucionales, una capa oculta y la capa de salida. Las capas mencionadas son las principales, pero por detrás se van haciendo otras operaciones como la reducción de espacio o dilución. El orden de capas por el que irán pasando los inputs es el siguiente:

- 1ª capa convolucional (Conv2D): se utilizan 32 neuronas, reciben inputs de dimensión 48x48 pixels y los filtros/kernels que se usarán son de dimensión 3x3 pixels.
- 2ª capa convolucional (Conv2D): se utilizan 64 neuronas, reciben como inputs los outputs de la primera capa y los filtros/kernels que se usarán son de dimensión 3x3 pixels.
- 1ª capa de reducción del espacio (MaxPooling2D): se recogen los outputs y se extraen los valores máximos sobre una ventana de dimensión 2x2 pixels.
- 1ª Capa de dilución (Dropout): se reducen los pesos con el fin de que no haya un sobreajuste el modelo.
- 3ª capa convolucional (Conv2D): se utilizan 128 neuronas, reciben como inputs los outputs de la capa anterior y los filtros/kernels que se usarán son de dimensión 3x3 pixels.
- 2ª capa de reducción del espacio (MaxPooling2D): se recogen los outputs y se extraen los valores máximos sobre una ventana de dimensión 2x2 pixels.
- 4ª capa convolucional (Conv2D): se utilizan 128 neuronas, reciben como inputs los outputs de la capa anterior y los filtros/kernels que se usarán son de dimensión 3x3 pixels.
- 3ª capa de reducción del espacio (MaxPooling2D): se recogen los outputs y se extraen los valores máximos sobre una ventana de dimensión 2x2 pixels.
- 2ª capa de dilución (Dropout): se reducen los pesos con el fin de que no haya un sobreajuste el modelo.
- 1ª capa de clasificación (Dense): se utilizan 1024 neuronas encargadas de llevar los outputs de la capa anterior.

3.3. Redes Convolucionales

- 3ª capa de dilución (Dropout): se reducen los pesos con el fin de que no haya un sobreajuste el modelo.
- 2ª capa de clasificación (Dense): se utilizan 6 neuronas encargadas de llevar los outputs de la capa anterior. Se dejan los pesos de las 6 expresiones en formato de un vector de probabilidades.
- Compilación del modelo
- Fit del modelo: se pasan las imágenes de la base de datos que se usa para entrenar el modelo y actualiza los pesos en cada epoch (Epoch: término usado en aprendizaje automático para referirnos a un procesamiento completo del set dedicado a “train” por parte del algoritmo de aprendizaje o algoritmo de extracción de características).
- Guardado del modelo: se guardan los pesos de las categorías/expresiones del modelo entrenado.

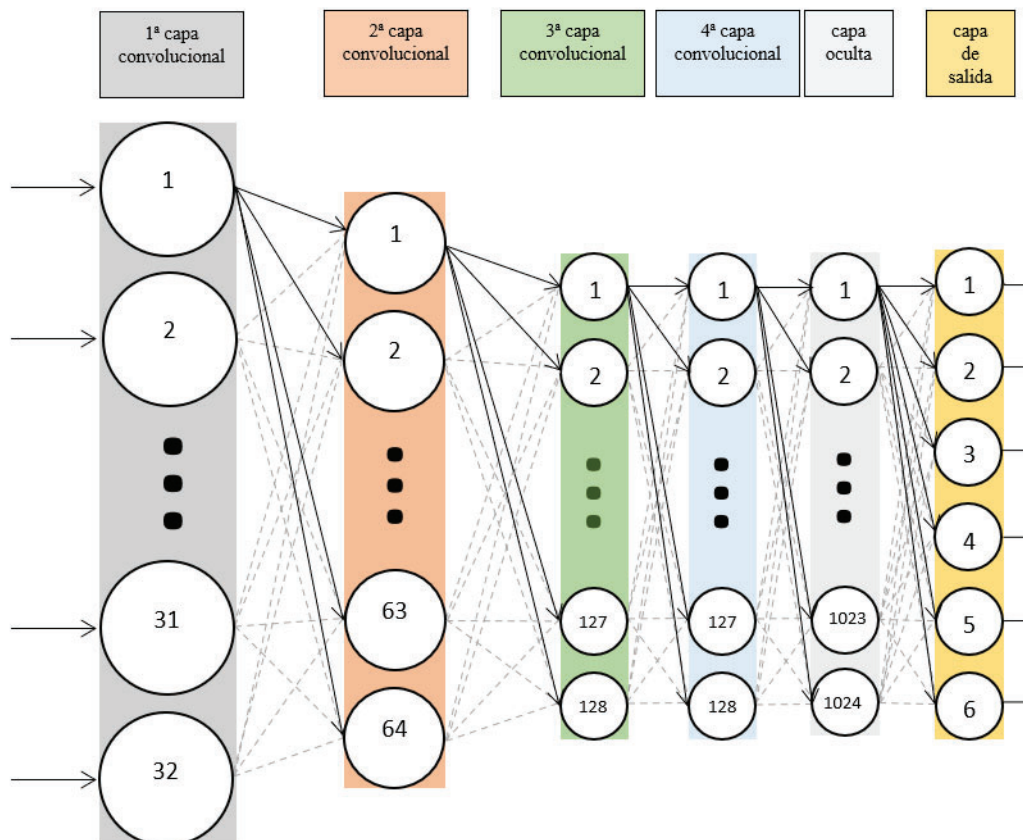


Figura 3.1: Arquitectura de red neuronal convolucional utilizada para el reconocedor de expresiones.

3.4. Diseño del prototipo

En esta parte se establece el diseño arquitectónico del software. Las fases o procesos que se cubren son 2 básicamente:

- Preparación del modelo o clasificador: entrenamiento y testeo de un modelo de clasificación de expresiones.
- Aplicación del clasificador o reconocimiento de expresiones con el modelo previamente entrenado

En la Figura 3.2, se pueden visualizar los pasos y elementos del entrenamiento y validación del modelo. Se preparan los sets de entrenamiento y validación para suministrarlos a la red neuronal, ésta se encarga de extraer las características de la imagen que diferencian unas expresiones de otras. La salida del modelo de clasificación es un vector de tantas posiciones como expresiones a clasificar, que representa la probabilidad de cada una de las expresiones detectables. Al acabar el entrenamiento se guardan los pesos de los filtros usados en todo el entrenamiento. Tras procesar todas las imágenes, el modelo habrá terminado de aprender a reconocer las expresiones faciales. Se hace, para terminar, una validación de su eficiencia con el set de validación.

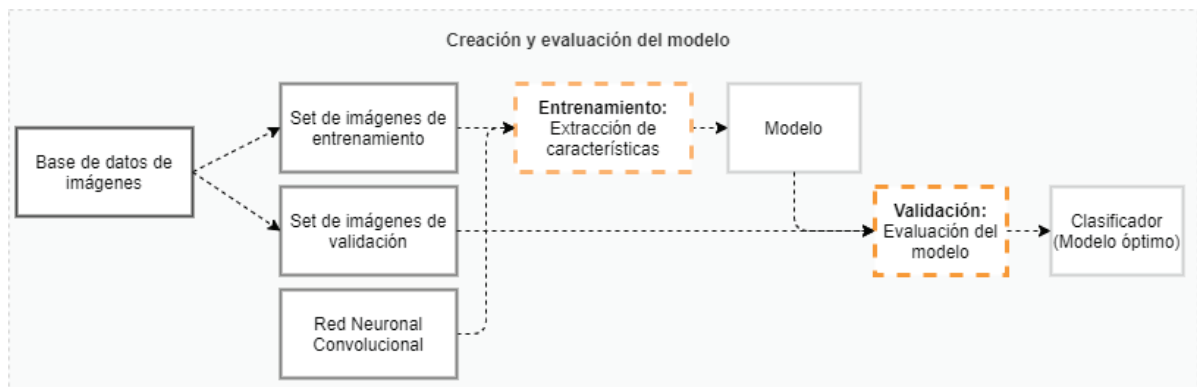


Figura 3.2: Esquema conceptual que explica en qué momento está ubicada la fase de entrenamiento y validación en la obtención del modelo.

El modelo previamente entrenado y validado será el sistema clasificador de imágenes o detector de expresiones faciales. Para la aplicación de este clasificador necesitaremos como objeto de entrada las imágenes de las que se quiera hacer una detección facial, como se explica en la Figura 3.3 y Figura 3.4. El clasificador puede recibir los inputs de 2 maneras: por cámara webcam o un directorio de imágenes. Para entrenar el modelo se leyeron lotes de imágenes desde directorios, por lo que también se podría hacer la detección con un lote de imágenes nuevo, leyendo una a una y devolviendo el resultado por consola. La otra opción, y más entretenida, es mediante una webcam, se tratará como una secuencia de imágenes y devolverá la expresión en texto en la misma ventana de vídeo donde se muestra nuestro rostro.

3.5. Implementación del prototipo

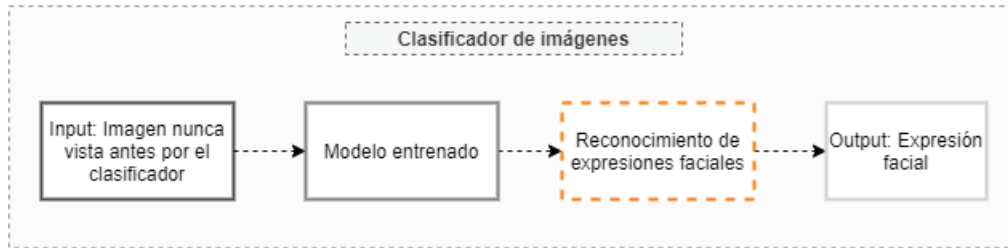


Figura 3.3: Esquema conceptual de la ejecución del clasificador.

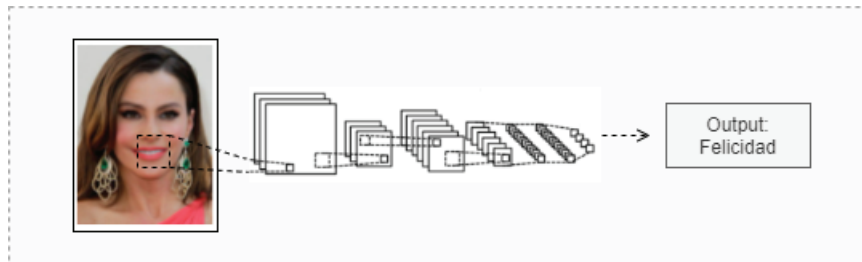


Figura 3.4: Ejemplo visual de recepción de input y salida de output del sistema de reconocimiento de expresiones faciales.

La ejecución de la clasificación comprende la carga del modelo que vayamos a usar y la lectura de fotogramas a partir de nuestra cámara. Detectamos la estructura facial en la imagen y nos quedamos con sus coordenadas para dibujar un recuadro que muestre por pantalla el área que se ha detectado. Llegados a este punto se emplea el modelo para predecir la expresión de área del rostro. Esta predicción viene en forma de vector de probabilidades por lo que tomaremos la posición del mayor índice de probabilidad e identificamos dicha posición con la expresión que representa. Por último, se muestra cerca al recuadro la expresión en texto. De esta manera es más visible.

3.5. Implementación del prototipo

3.5.1. Herramientas software

En el desarrollo del proyecto he utilizado el lenguaje de programación Python. Para tareas de aprendizaje automático (machine learning), Python es un lenguaje de programación que permite usar y combinar librerías dedicadas específicamente al desarrollo y evaluación de modelos de redes neuronales. El entorno de desarrollo elegido fue PyCharm, debido al conocimiento personal de la herramienta. Se hizo uso también de la interfaz gráfica de usuario Anaconda Navigator. Con esta interfaz se gestionaron los entornos virtuales en los que instalamos paquetes auxiliares como Tensorflow, que era de gran ayuda para trabajar de forma fácil con modelos de entrenamiento.



Figura 3.5: Logos de herramientas Software: Python, Anaconda, Tensorflow y Pycharm.

Para el almacenamiento del código fuente y la gestión de versiones se utilizó la plataforma Github. Se creó un repositorio con el nombre “Sentimental-recognition-from-images” donde el código fuente estará a disposición de la tutora en cualquier momento, como se puede ver en la Figura 3.6.

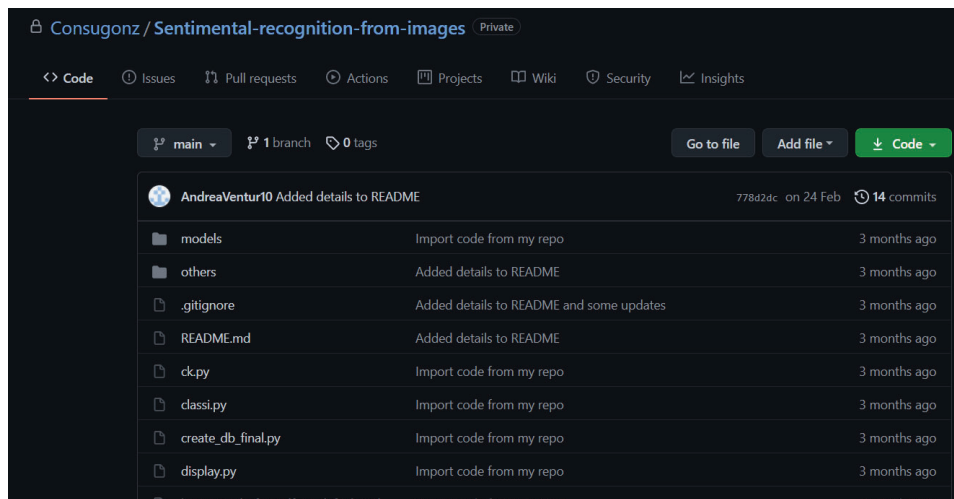


Figura 3.6: Repositorio Github en el que se almacena el código fuente del proyecto.

Por último, Microsoft Teams, como medio de comunicación para hacer las reuniones semanales. Hay un apartado destinado a los documentos (memoria, artículos, entregables, etc...) y tablón de tareas.

3.5.2. Fases de desarrollo

El desarrollo de la herramienta se ha dividido en las fases que se explican a continuación. Todas las fases se pueden contrastar con los ficheros Python del código fuente subido al repositorio Github: “Sentimental-recognition-from-images”.

Organización de los datos: esta fase inicial comprende la distribución de las imágenes de las bases de datos en un orden de directorios por expresiones. No todas las bases de datos vienen almacenadas de la misma manera, lo que

3.5. Implementación del prototipo

supone crear distintos programas Python para cada caso. Por lo general suelen venir distribuidas por directorios para cada modelo o actor. Dentro del directorio de un modelo o actor encontraremos las imágenes sin clasificar y con siglas distintivas, a continuación véase la Figura 3.7.

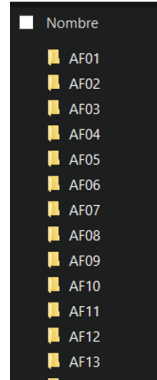


Figura 3.7: Distribución por defecto de la base de datos KDEF: una carpeta por actor o modelo.

Diferenciando los títulos de cada imagen logramos la distribución creando carpetas para cada expresión existente. En la Figura 3.8 podemos ver como aparecen revueltas las imágenes dentro de una de las carpetas de la Figura 3.7.

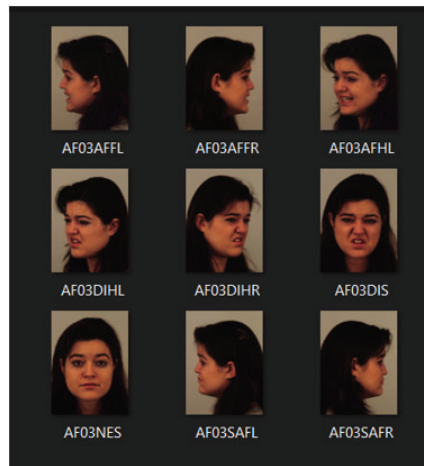


Figura 3.8: Contenido de una carpeta de un actor o modelo de la base de datos KDEF.

De forma que de un directorio como la Figura 3.7 pasamos a un directorio de sets, véase Figura 3.9, que a su vez contiene directorios por cada expresión. Queda una distribución totalmente entendible y manejable.



Figura 3.9: A la izquierda, directorios de sets de train y test. A la derecha, directorios de expresiones.

Preprocesamiento de imágenes: comprende todas las operaciones de ajuste de características de las imágenes. Ejemplo de esto son las imágenes de la Figura 3.10, las 2 imágenes a color de los extremos son las imágenes en el formato por defecto que trae la BD. Las imágenes del centro son las imágenes anteriormente mencionadas tras pre-procesarlas cambiando su escala de color, contraste, tamaño y encuadre.



Figura 3.10: Imágenes de la base de datos KDEF. Las imágenes en blanco y negro son las de los extremos tras hacer el pre-procesado de la base de datos (Fuente: KDEF BD)

En primer lugar, se redimensionan las imágenes recortando únicamente la zona que nos interesa, es decir, el rostro. Como se puede ver en la Figura 3.11, aunque los rostros de la parte izquierda se vean lo suficientemente cerca, todavía se puede cuadrar más la zona de interés, la diferencia tras pre-procesarlas se puede observar en las imágenes que quedan en la parte derecha del cuadro de imágenes. Entre las mejoras del pre-procesado también se hace un ajuste del histograma o “stretching de histograma” del contraste, que se resume en el ajuste de la gama de valores de píxeles en caso de no utilizar todo el rango $[0, 255]$.

3.5. Implementación del prototipo

Esto mejora la iluminación de algunas imágenes que se veían con demasiada iluminación o, por el contrario, muy apagadas u oscuras como se puede apreciar sobre todo en la imagen de la segunda fila a la izquierda, que tras realizar el stretching de histograma se obtiene la imagen de la derecha.



Figura 3.11: En la primera fila, imágenes de expresión "feliz", y en la de abajo expresión "neutral". En el lado izquierdo: imágenes sin pre-procesar. En el lado derecho : resultado tras pre-procesado. (Fuente: BD Fer2013)

Después, se hace un cambio de la escala de color RGB a escala de GRAY (escala de grises). El resultado será la obtención de imágenes con las mismas propiedades y totalmente aptas para procesarlas y mezclarlas en caso de tener que juntar 2 o más bases de datos, el formato será el mismo. Esto facilita toda la fase de creación de modelos.

Además de correcciones respecto a las características de las imágenes, surgen problemas con imágenes que no tienen que ver con las expresiones faciales. Al revisar las imágenes de cada base de datos encontramos un problema con Fer2013, contiene imágenes que no son de un rostro real, de formas irreconocibles o incluso imágenes que albergan texto, véase Figura 3.12.



Figura 3.12: 8 Imágenes de Fer2013 que no contienen un rostro humano. (Fuente: BD Fer2013)

Desarrollo

Estas imágenes se deberían excluir de la base de datos ya que empeorarían nuestro entrenamiento del modelo aportando información inútil. Para evitarlo se le aplica un proceso de reconocimiento facial en cada imagen de los sets de entrenamiento y test, en caso de no detectar un rostro se descarta dicha imagen para el set de imágenes final. En este caso para Fer2013 se reduce el número de imágenes considerablemente, véase la Tabla 3.7 resultado de pre-procesar la base de datos de la Tabla 3.6.

Tabla 3.6: Distribución de imágenes de base de Datos "Fer2013".

Expresión	Entrenamiento	Validación
Enfado	3.995	958
Repugnancia	436	111
Miedo	4.097	1.024
Felicidad	7.215	1.774
Neutral	4.965	1.233
Tristeza	4.830	1.247
Sorpresa	3.171	831
Total	28.709	7.178

Tabla 3.7: Distribución de imágenes de base de Datos "Fer2013"(Pre-procesada)

Expresión	Entrenamiento	Validación
Enfado	267	72
Repugnancia	42	12
Miedo	167	51
Felicidad	819	206
Neutral	478	127
Tristeza	169	54
Sorpresa	182	48
Total	2.124	570

Creación y entrenamiento de modelos: para la creación de un modelo por cada base de datos, necesitaremos las imágenes organizadas en directorios como se explica en la fase de "Organización de los datos". Para crear el modelo se pasan lotes de imágenes con la ruta de los directorios donde se albergan los sets de entrenamiento y validación. Dicha ruta se introduce como argumento al crear la red neuronal convolucional, y ésta realizará la extracción de características. Posteriormente, se realiza una etapa de validación donde también se leen las imágenes del set de validación con la ruta de su respectivo directorio, en esta etapa se pone en práctica el modelo con lo que ha aprendido de la etapa de entrenamiento. El modelo se almacena en la carpeta "models" como un archivo ".h5" y con un nombre identificativo de la base de datos a la que pertenece, como se puede ver en la Figura 3.13. En total tendremos 1 modelo por cada base de datos, es decir 4 modelos.

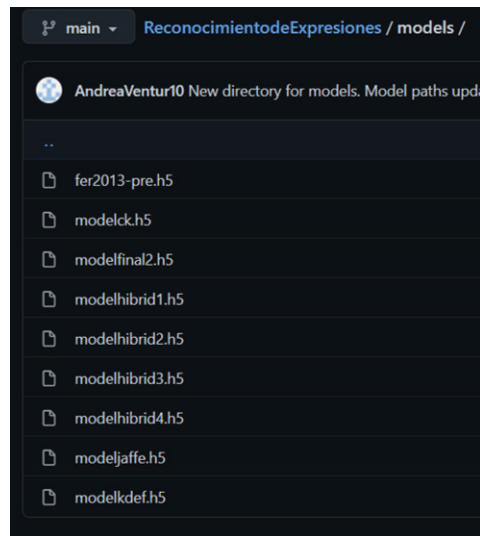


Figura 3.13: Carpeta “models” donde se almacenan os modelos de las bases de datos del proyecto.

Testings: una vez obtenidos los modelos pasamos a la fase de pruebas o “testings”. Para esta fase se analiza la eficiencia del modelo creado, por lo que se prueba su funcionamiento con las imágenes de los sets de validación del resto de bases de datos. La manera de medir la eficiencia será haciendo el porcentaje de aciertos por cada expresión. Se pasa al modelo una a una las imágenes de los sets de validación y este devolverá, de cada una, la expresión que ha interpretado como más probable. Se calcula el número de resultados correctos e incorrectos por cada expresión reconocible por el modelo clasificador y se obtiene el porcentaje de acierto. Se proporciona también un porcentaje total de aciertos sobre el total de imágenes usadas para “testing”, con el que luego se hará comparaciones y análisis respecto de los “testings” de otra base de datos.

Híbridos: tras la creación de los 4 modelos y “testings” se replantea la solución de una base de datos híbrida. Se reagrupan las bases de datos en parejas con lo que obtenemos tres modelos nuevos, más uno que reúne todas las bases de datos. También se hará una fase de Testing que permita una evaluación de sus respectivos aciertos. Como consecuencia se alcanza una precisión mayor, aunque aún insuficiente.

Modelo final: esta parte del desarrollo se orienta a la creación del último híbrido, el modelo definitivo que será el instrumento principal para nuestra herramienta de detección. Se creará a partir de los sets de imágenes correspondientes a las expresiones, de los híbridos creados anteriormente, que mejor desempeño hayan ofrecido en las detecciones hechas en los “testings”. Por lo tanto, se escoge la información (sets “train” y “test”) de las expresiones que han alcanzado los valores máximos de precisión como se observa en la Tabla 4.3. Una vez creado y rellenado el directorio de sets de imágenes del nuevo modelo, se procede a la creación y entrenamiento del modelo con la estructura de la red neuronal

convolucional usada hasta ahora para todos los entrenamientos.

Display: por último, nos encontramos en la fase de monitorización de la herramienta. Este desarrollo completa todas las fases anteriores aportando una vista del mecanismo de reconocimiento. Se retransmite por ventana la grabación en vivo desde la cámara del usuario, dicha retransmisión muestra el rostro encuadrado con un texto que identifica la expresión que nuestro reconocedor de expresiones ha detectado, como se puede observar en la Figura 3.14.

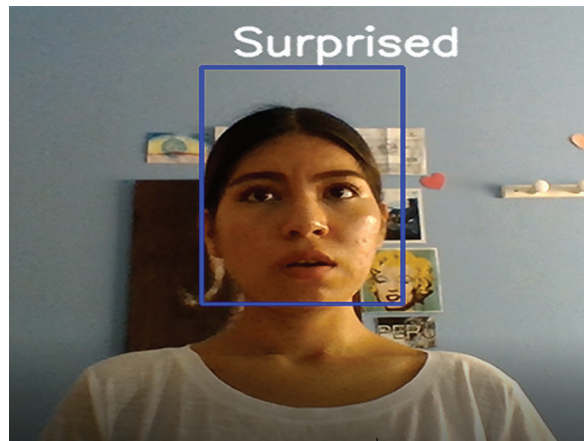


Figura 3.14: Detector de expresiones en modo display reconociendo expresión Sorpresa("Surprised").

3.5.3. Revisión y corrección de errores

Los errores más comunes han sido por motivos de desarrollo del código fuente y la distribución de imágenes y directorios de las bases de datos. Al principio hubo que hacer mucho preprocesamiento de cada una de las bases de datos para dejarlas en una distribución de carpetas que fuese cómoda para poder trabajar y manipular las imágenes identificando a que categoría/expresión pertenecen. Para solucionar estos errores de organización se identificó cada expresión con un "Id": expresiones = 0: enfado("angry"), 1: repugnancia("disgusted"), 2: miedo("fear"), 3: felicidad("happy"), 4: neutral("neutral"), 5: tristeza("sad"), 6: sorpresa("surprised").

En la fase de Testing también se presentaron errores a la hora de interpretar el acierto de los modelos, no todas las bases de datos tienen el mismo número de expresiones. Por lo que para el testing de un modelo con 6 expresiones y un set de imágenes test de 7 expresiones sólo se tomaban en cuenta las imágenes de las expresiones que ambos tenían. Por ejemplo, si tenemos muchas imágenes test de la expresión "neutral" y nuestro modelo no se ha entrenado teniendo en cuenta ésta, entonces todas las imágenes de "neutral" serán contadas como fallos porque nuestro modelo no las reconocerá como "neutral" nunca y supondrán una precisión final más baja y menos real.

Tras revisar nuestra herramienta final de detección de expresiones, nos queda-

3.5. Implementación del prototipo

mos con valores de acierto altos para todas las expresiones excepto para “tristeza”. Como corrección se podría buscar y aplicar más material de entrenamiento (bases de datos de imágenes) que aporten más información a “tristeza”, aunque tras volver a entrenar el modelo con esos cambios puede ser que otra expresión baje su porcentaje de acierto. Esto se podría dar cuando el modelo haga una predicción para imágenes de las otras expresiones que antes devolviesen el resultado correcto, pero ahora al tener más información de “tristeza”, puede encontrar por error más coincidencia y detectarlas como “tristeza”.

Capítulo 4

Resultados y conclusiones

4.1. Resultados

Para hacer un análisis y observación iniciales de los resultados que se obtienen de cada base datos, contamos con 4 modelos de los que necesitamos saber la precisión de acierto total e individual para cada expresión, esta información es útil para comparar unos con otros y elegir lo mejor de cada uno. He utilizado los modelos de las bases de datos y sus respectivos sets de imágenes dedicados a test. El modelo previamente entrenado analiza cada imagen devolviendo como resultado la expresión que más encaja, dependerá de las características aprendidas en el entrenamiento. Se contabilizan los aciertos y fallos al detectar la expresión de todas las imágenes. Con dichos datos se calculan los porcentajes de acierto respecto al número total de imágenes de cada expresión. Véase la Figura 4.1 donde se muestran los resultados de ejecutar un Testing del modelo de la BD Fer2013 con el set test de la BD KDEF.

```
Resultados Testing modelo fer2013 con BD kdef

ENFADO Id 0: 64.29 % | correct : 18  fail: 10
REPUGNANCIA Id 1: 0.00 % | correct : 0  fail: 28
MIEDO Id 2: 7.14 % | correct : 2  fail: 26
FELICIDAD Id 3: 100.00 % | correct : 28  fail: 0
NEUTRAL Id 4: 96.43 number % | correct : 27  fail: 1
TRISTEZA Id 5: 7.14 % | correct : 2  fail: 26
SORPRESA Id 6: 10.71 % | correct : 3  fail: 25

PRECISION TOTAL DE BD: 40.82 %
```

Figura 4.1: Ejemplo de resultado de Testing con el modelo de Fer2013 y el set test de KDEF.

4.1. Resultados

Los 4 modelos obtenidos a partir de los sets de entrenamiento se ponen a prueba con cada set de imágenes test. En total se sacan 16 Testings (Tabla 4.1):

Tabla 4.1: Comparativa Fer2013, Cohn-Kanade, JAFFE y KDEF BD.

Modelo BD	BD set test		Expresión	Acierto por expresión	Acierto total
CK	CK	27	Enfado	55.56 %	85.49 %
		11	Desprecio	40.00 %	
		35	Repugnancia	97.14 %	
		15	Miedo	100.00 %	
		41	Felicidad	87.80 %	
		-	Neutral	-	
		16	Tristeza	87.50 %	
		49	Sorpresa	95.92 %	
	JAFFE	6	Enfado	0 %	18.60 %
		-	Desprecio	-	
		6	Repugnancia	0 %	
		6	Miedo	0 %	
		6	Felicidad	16.67 %	
		7	Neutral	-	
		6	Tristeza	16.67 %	
		6	Sorpresa	100.00 %	
	FER2013	72	Enfado	4.17 %	15.96 %
		-	Desprecio	-	
		12	Repugnancia	58.33 %	
		51	Miedo	25.49 %	
		206	Felicidad	21.84 %	
		127	Neutral	-	
		54	Tristeza	20.37 %	
		48	Sorpresa	25.00 %	
	KDEF	28	Enfado	10.71 %	53.57 %
		-	Desprecio	-	
		28	Repugnancia	96.43 %	
		28	Miedo	53.57 %	
		28	Felicidad	89.29 %	
		28	Neutral	0.00 %	
		28	Tristeza	35.71 %	
		28	Sorpresa	89.29 %	
JAFFE	CK	27	Enfado	0.00 %	21.24 %
		11	Desprecio	-	
		35	Repugnancia	0.00 %	
		15	Miedo	0.00 %	
		41	Felicidad	100.00 %	
		-	Neutral	-	
		16	Tristeza	0.00 %	
		49	Sorpresa	0.00 %	

Resultados y conclusiones

Modelo BD	BD set test	Expresión	Acierto por expresión	Acierto total
	JAFPE	6	Enfado	0.00 %
		-	Desprecio	-
		6	Repugnancia	0.00 %
		6	Miedo	0.00 %
		6	Felicidad	100.00 %
		7	Neutral	0.00 %
		6	Tristeza	0.00 %
		6	Sorpresa	0.00 %
	FER2013	72	Enfado	0.00 %
		-	Desprecio	-
		12	Repugnancia	0.00 %
		51	Miedo	0.00 %
		206	Felicidad	100.00 %
		127	Neutral	0.00 %
		54	Tristeza	0.00 %
		48	Sorpresa	0.00 %
	KDEF	28	Enfado	0.00 %
		-	Desprecio	-
		28	Repugnancia	0.00 %
		28	Miedo	0.00 %
		28	Felicidad	100.00 %
		28	Neutral	0.00 %
		28	Tristeza	0.00 %
		28	Sorpresa	0.00 %
FER2013	CK	27	Enfado	25.93 %
		11	Desprecio	-
		35	Repugnancia	0.00 %
		15	Miedo	6.67 %
		41	Felicidad	100.00 %
		-	Neutral	-
		16	Tristeza	6.25 %
		49	Sorpresa	20.41 %
	JAFPE	6	Enfado	0.00 %
		-	Desprecio	-
		6	Repugnancia	0.00 %
		6	Miedo	0.00 %
		6	Felicidad	83.33 %
		7	Neutral	100.00 %
		6	Tristeza	16.67 %
		6	Sorpresa	33.33 %
	FER2013	72	Enfado	8.33 %
		-	Desprecio	-
		12	Repugnancia	0.00 %
		51	Miedo	33.33 %
		206	Felicidad	80.58 %
		127	Neutral	70.87 %

4.1. Resultados

Modelo BD	BD set test		Expresión	Acierto por expresión	Acierto total
		54	Tristeza	9.26 %	
		48	Sorpresa	16.67 %	
	KDEF	28	Enfado	64.29 %	40.82 %
		-	Desprecio	-	
		28	Repugnancia	0.00 %	
		28	Miedo	7.14 %	
		28	Felicidad	100.00 %	
		28	Neutral	96.43 %	
		28	Tristeza	7.14 %	
		28	Sorpresa	10.71 %	
KDEF	CK	27	Enfado	14.81 %	43.00 %
		11	Desprecio	-	
		35	Repugnancia	5.71 %	
		15	Miedo	6.67 %	
		41	Felicidad	80.49 %	
		-	Neutral	0.00 %	
		16	Tristeza	50.00 %	
		49	Sorpresa	71.43 %	
	JAFPE	6	Enfado	0.00 %	27.90 %
		-	Desprecio	-	
		6	Repugnancia	0.00 %	
		6	Miedo	0.00 %	
		6	Felicidad	16.67 %	
		7	Neutral	71.43 %	
		6	Tristeza	16.67 %	
		6	Sorpresa	83.33 %	
	FER2013	72	Enfado	4.17 %	31.05 %
		-	Desprecio	-	
		12	Repugnancia	0.00 %	
		51	Miedo	0.00 %	
		206	Felicidad	28.16 %	
		127	Neutral	79.53 %	
		54	Tristeza	24.07 %	
		48	Sorpresa	4.17 %	
	KDEF	28	Enfado	53.57 %	55.10 %
		-	Desprecio	-	
		28	Repugnancia	67.86 %	
		28	Miedo	0.00 %	
		28	Felicidad	85.71 %	
		28	Neutral	89.29 %	
		28	Tristeza	39.29 %	
		28	Sorpresa	50.00 %	

Resultados y conclusiones

Es importante aclarar que las bases de datos que cuentan con pocas imágenes no ofrecen suficiente información para establecer diferencias o patrones para detectar las expresiones. Tampoco es útil que en las imágenes no se modele la expresión correctamente. Esto sucede en imágenes de JAFFE, en las que el lenguaje facial de imágenes correspondientes a diferentes expresiones era muy similar, por lo que no se cuenta con material bueno para conseguir patrones discriminantes de expresiones, por esta razón los porcentajes de acierto son sumamente bajos.

Tras la primera comparativa de aciertos de cada modelo con cada set de imágenes test de las bases de datos, se plantea un segundo análisis de resultados. Se propone un híbrido como solución a la base de datos final que usaremos. Debido a que la mayoría de los resultados de acierto no superaron el 60 % de precisión, se hacen grupos de 2 con las bases de datos para aumentar estos resultados y uno que será la mezcla de todos. Se reduce la lista de expresiones detectables a 7 expresiones: [enfado, repugnancia, miedo, alegría, neutral, tristeza y sorpresa], eliminando la expresión “desprecio” debido a que no hay suficientes bases de datos que compartan dicha expresión. Se decide la mezcla de las bases de datos con mayor tasa de acierto por lo que la base de datos JAFFE queda descartada para estos grupos de prueba y por lo tanto para nuestra base de datos final. Con estas decisiones se quería buscar el equilibrio de la precisión para todas las expresiones detectables.

Los grupos creados son:

- Híbrido 1: mezcla de CK BD y KDEF BD.
- Híbrido 2: mezcla de CK BD y Fer2013 BD.
- Híbrido 3: mezcla de Fer2013 BD y KDEF BD.
- Híbrido 4: mezcla de CK BD, KDEF BD y Fer2013 BD.

Como se puede apreciar en la Tabla 4.2, con los datos de precisión conseguidos a partir de los híbridos se observa una considerable mejora del resultado para casi todo el conjunto de expresiones. No obstante, el máximo valor conseguido para cada expresión no pertenece al mismo híbrido, por lo que se eligió para cada expresión el set de imágenes del híbrido que mejor resultado haya logrado.

Tabla 4.2: Comparativa de Híbridos (a color el valor máximo de cada expresión).

EXPRESIÓN	ACIERTO H1	ACIERTO H2	ACIERTO H3	ACIERTO H4
ENFADO	25,98 %	47,24 %	28,35 %	38,50 %
REPUGN.	78,67 %	78,68 %	34,67 %	80,00 %
MIEDO	20,21 %	29,79 %	20,21 %	37,23 %
FELICIDAD	67,64 %	87,64 %	86,91 %	86,54 %
NEUTRAL	33,55 %	54,19 %	74,84 %	69,03 %
TRISTEZA	47,96 %	39,80 %	23,47 %	44,89 %
SORPRESA	58,40 %	82,40 %	56,00 %	76,80 %

En la Tabla 4.3 y Figura 4.3 se muestra la elección de los mejores valores obte-

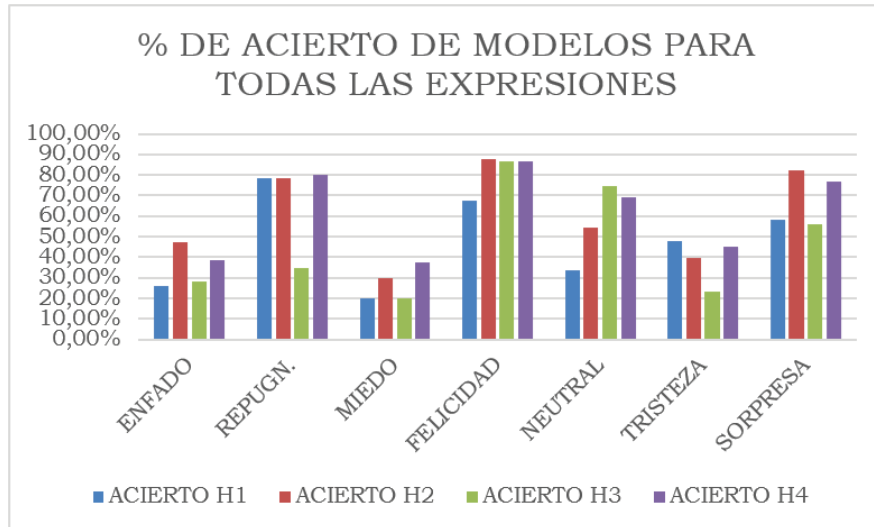


Figura 4.2: % de acierto de los 4 modelos híbridos para todas las expresiones.

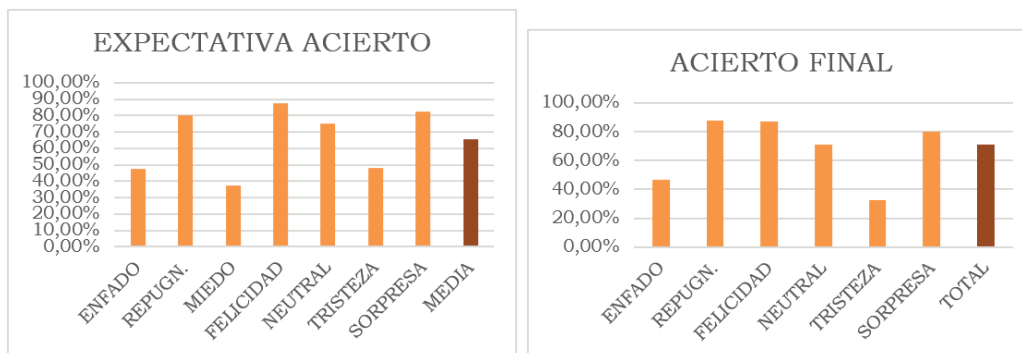


Figura 4.3: % de expresiones de BD final. A la izquierda, teniendo en cuenta la expresión miedo. A la derecha, sin expresión "miedo".

nidos y la posible media del nuevo híbrido. Aunque un 65,33 % es la precisión total más alta de los modelos hasta ahora vistos, se siguió buscando una mejor calificación. La expresión “miedo” no ha aumentado lo suficiente el resultado de acierto en ninguna de las mejoras planteadas. Nos quedaremos con un conjunto de 6 expresiones: [“enfado”, “repugnancia”, “felicidad”, “neutral”, “tristeza”, “sorpresa”] que tras todos los procesos de elección se ha demostrado que aumentan la precisión total de nuestro modelo de entrenamiento. La tasa de acierto final tras entrenar el nuevo modelo de 6 expresiones queda en un 70,99 % .

Resultados y conclusiones

Tabla 4.3: % de expresiones de BD final. A la izquierda, teniendo en cuenta la expresión "miedo". A la derecha, sin expresión "miedo".

EXPRESIÓN	ACIERTO FINAL	EXPRESIÓN	ACIERTO FINAL
ENFADO	47,24 %	ENFADO	46,45 %
REPUGN.	80,00 %	REPUGN.	88,00 %
MIEDO	37,23 %	—	—
FELICIDAD	87,64 %	FELICIDAD	87,27 %
NEUTRAL	74,84 %	NEUTRAL	70,96 %
TRISTEZA	47,96 %	TRISTEZA	32,65 %
SORPRESA	82,40 %	SORPRESA	80,00 %
MEDIA	0,6533 %	TOTAL	70,99 %

4.2. Validación

4.2.1. Enfado

La expresión de enfado se manifiesta como en la Figura 4.4. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Cejas bajas y rectas
- Fruncimiento en el entrecejo
- Surcos nasogenianos marcados

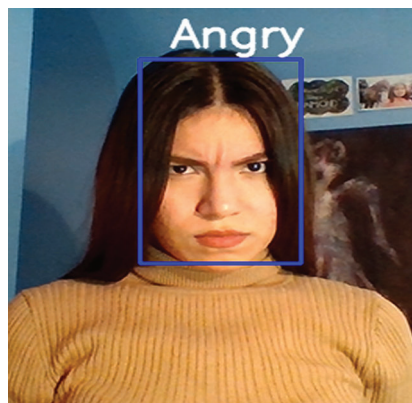


Figura 4.4: Captura de detector en vivo reconociendo expresión "enfado" ("angry").

4.2.2. Repugnancia

La expresión de repugnancia se manifiesta como en la Figura 4.5. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Cejas bajas
- Fruncimiento en el entrecejo
- Nariz arrugada
- Labios apretados y ligeramente elevados

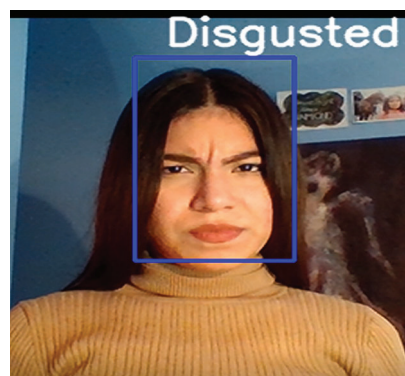


Figura 4.5: Captura de detector en vivo reconociendo expresión “repugnancia” (“disgusted”).

4.2.3. Neutral

La expresión de neutralidad se manifiesta como en la Figura 4.6. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Ojos abiertos y cejas sin tensión
- Pómulos relajados
- Labios relajados

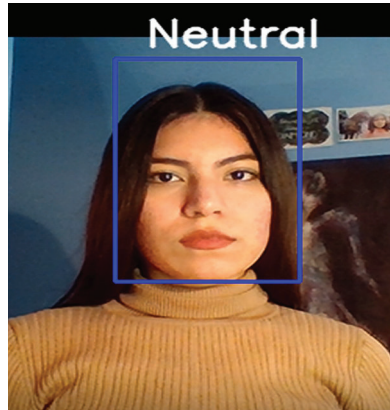


Figura 4.6: Captura de detector en vivo reconociendo expresión "neutral" ("neutral").

4.2.4. Felicidad

La expresión de felicidad se manifiesta como en la Figura 4.7. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Sonrisa amplia y con las comisuras elevadas
- Pómulos elevados
- Ojos ligeramente achinados

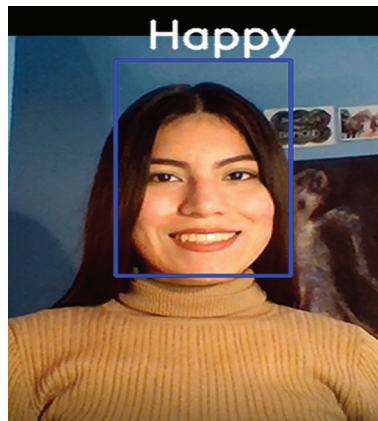


Figura 4.7: Captura de detector en vivo reconociendo expresión "felicidad" ("happy").

4.2.5. Tristeza

La expresión de tristeza se manifiesta como en la Figura 4.8. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Entrecejo ligeramente fruncido
- Mentón ligeramente elevado
- Labios caídos

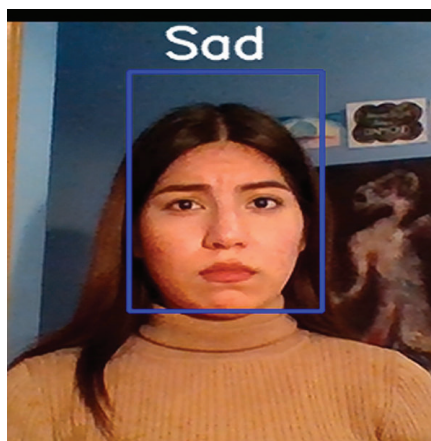


Figura 4.8: Captura de detector en vivo reconociendo expresión “tristeza” (“sad”).

4.2.6. Sorpresa

La expresión de sorpresa se manifiesta como en la Figura 4.9. Algunas de las características identificativas que se han comprobado tras la prueba de esta expresión son:

- Cejas elevadas
- Ojos muy abiertos y párpados subidos
- Boca abierta en forma de O

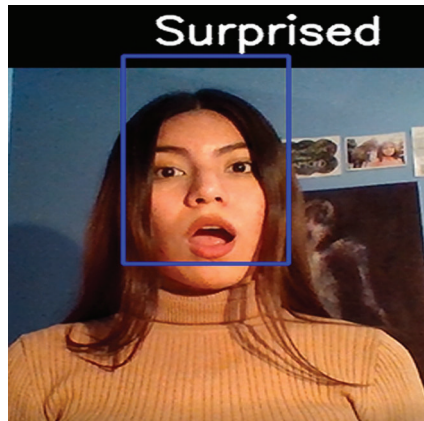


Figura 4.9: Captura de detector en vivo reconociendo expresión “sorpresa” (“surprised”).

4.3. Conclusiones

Con este trabajo se ha conseguido alcanzar el objetivo de obtener una herramienta capaz de establecer relaciones de características de imágenes, distinguir expresiones y realizar una predicción en vivo de nuestra expresión, en definitiva, un sistema FER. Tras todo el proceso de construcción del programa he podido concluir en la importancia de la calidad y variedad del material con el que se trabaja, en nuestro caso las bases de imágenes. Aspectos tan simples como la iluminación, color o tamaño son importantes para los procesos de extracción de información de imágenes.

Revisando la lista de objetivos específicos planteados podemos comprobar que se han desarrollado exitosamente, desde la búsqueda y estudio de información, la experiencia y agilidad ganada en el lenguaje Python y en especial con el procesamiento de grandes cantidades de imágenes, hasta el desarrollo de un sistema que aplique todo ese conocimiento aprendido y otorgarle autonomía para interpretar las imágenes y reconocer una expresión humana en ellas. Otro objetivo específico muy importante era la validación y búsqueda de un resultado más que aceptable, para el que se ha alcanzado un resultado bueno pero que podría mejorarse en caso de disponer de más bases de datos.

En definitiva, ha sido un proyecto que me ha permitido entrar en contacto con el campo de la Inteligencia Artificial, especialmente en el Reconocimiento de emociones, pero también me ha ayudado a desarrollarme a nivel personal, he ganado capacidad de autodisciplina y también autodidacta para adquirir nuevos conocimientos.

Bibliografía

- [1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. of Personality and Social Psychology*, vol. 17, No. 2, pp. 124–129, 1971.
- [2] "Reconocimiento de expresiones faciales," *Wikipedia, la enciclopedia libre*.
- [3] C. Sánchez, "¿cómo sabe una máquina que sonríes? reconocimiento facial para principiantes," *ElDiario.es*.
- [4] D. Gershgorn, "Snapchat quietly revealed how it can put ai on your phone," *Quartz*.
- [5] "Apple compra emotient, una startup de inteligencia artificial," *Hipertextual*.
- [6] "Why snapchat filters make them more attractive ?," *Medium*.
- [7] "Así es como las marcas utilizan el reconocimiento facial para analizar las reacciones ante sus anuncios," *PuroMarketing*.
- [8] "iphone x después de unas semanas de uso," *OBG888*.
- [9] "Robin, el robot de ia para apoyar las necesidades emocionales de los niños," *Consalud*.
- [10] J. Kart, "Robin the robot comforts kids in hospitals, can help with covid-19," *Forbes*.
- [11] J. Lien, J. F Cohn, T. Kanade, and C.-C. Li, "Detection, tracking, and classification of action units in facial expression," *Researchgate*, 1999.
- [12] H. Gonzalez, "Recocimiento facial utilizando viola-jones," *Academia.edu*.
- [13] *Siret.ms.mff.cuni.cz*.
- [14] F. Merchán, S. Galeano, and H. Poveda, "Mejoras en el entrenamiento de esquemas de detección de sonrisas," *Revistas.utp.ac.pa*.
- [15] "Boosting - wikipedia, la enciclopedia libre," *Es.wikipedia.org*.
- [16] "The intuition behind facial detection: The viola-jones algorithm," *towards data science*.
- [17] "Understanding face detection with the viola-jones object detection framework," *towards data science*.

- [18] “Detección de rostros, caras y ojos con haar cascad,” *Unipython*.
- [19] *GitHub*.
- [20] R. Puri, A. Gupta, M. Sikri, M. M. Tiwari, D. N. Pathak, and D. S. Goel, “Emotion detection using image processing in python,” *Researchgate*, 2018.
- [21] M. Al-Shabi, W. P. Cheah, and T. Connie, “Facial expression recognition using a hybrid cnn-sift aggregator,” *Researchgate*, 2016.
- [22] “Convolutional neural networks: La teoría explicada en español | aprende machine learning,” *Aprendemachinelearning.com*, 2018.
- [23] “El transfer learning y las redes convolucionales,” *Viewnext*.
- [24] I. Revina and S. E. W.R., “A survey on human face expression recognition techniques,” *ScienceDirect*, 2018.
- [25] K. Guo, “Holistic gaze strategy to categorize facial expression of varying intensities,” *Plos one*, 2012.
- [26] *Paralleldots*.
- [27] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, “Facial expression analysis with affdex and facet: A validation study,” *Springer Link*, 2018.
- [28] *IMOTIONS*.
- [29] *Visage Technologies - Face Analysis*.
- [30] *:) Afectiva - Deep Learning*.
- [31] *Github - OpenFace 2.2.0: a facial behavior analysis toolkit*.

Anexos

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=Facultad de Informatica - UPM, C=ES
Fecha/Hora	Tue Jun 22 11:03:25 CEST 2021
Emisor del Certificado	EMAILADDRESS=camanager@fi.upm.es, CN=CA Facultad de Informatica, O=Facultad de Informatica - UPM, C=ES
Numero de Serie	630
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)