Análisis de las Emociones Básicas Mediante la Aplicación de Modelos Inteligentes para su Detección y Clasificación

Molina-Reyes, E.[1], Hernández-Hernández, J.C.[2] , Quintero-Flores, P.M.[3], González-Meneses, Y.N.[4]

Datos de Adscripción:

[1] Enrique Molina Reyes. Tecnológico Nacional de México / Instituto Tecnológico de Apizaco. m23370006@apizaco.tecnm.mx
https://orcid.org/0009-0005-7756-6883

[2]✉ José Crispín Hernández Hernández. Tecnológico Nacional de México / Instituto Tecnológico de Apizaco. crispin.hh@apizaco.tecnm.mx
https://orcid.org/0000-0001-7245-9564

[3] Perfecto Malaquías Quintero Flores. Tecnológico Nacional de México / Instituto Tecnológico de Apizaco. perfecto.qf@apizaco.tecnm.mx
https://orcid.org/0000-0001-7651-4364

[4] Yesenia Nohemí González Meneses. Tecnológico Nacional de México / Instituto Tecnológico de Apizaco. yesenia.gm@apizaco.tecnm.mx
https://orcid.org/0000-0003-1034-0204

Resumen **- Este artículo presenta un procedimiento para realizar el reconocimiento de emociones básicas en un grupo de personas. Para ello, se realizó una breve revisión del estado del arte relacionado, considerando los temas más relevantes en el área, los trabajos más destacados, los dispositivos utilizados a lo largo del tiempo, así como su evolución para mejorar esta disciplina. Además, se destaca la importancia de configurar un entorno adecuado para clasificar emociones de manera controlada en un grupo de muestras, mediante la selección y filtrado del material utilizado para predecir dichas emociones. El preprocesamiento aplicado utiliza muestras recolectadas de las bases de datos FER2013, DISFA y BP4D, con el fin de aplicar posteriormente los métodos seleccionados para extraer características de interés que permitan lograr un reconocimiento más preciso. Durante el desarrollo de este trabajo, se seleccionó una muestra de una base de datos compuesta por imágenes y videos que registran reacciones emocionales de personas, con el propósito de utilizarlas de manera efectiva en los momentos requeridos. Con el objetivo de proponer un modelo para el reconocimiento emocional, se implementaron diferentes modelos y algoritmos computacionales, entre ellos la biblioteca de Deep Learning Fastai, así como el análisis de expresiones faciales utilizando el Sistema de Codificación de Acciones Faciales, en conjunto con la implementación de lógica difusa, permitió identificar la coincidencia de una o más Unidades de Acción en dos o más emociones analizadas, obteniendo resultados competitivos en la clasificación del reconocimiento emocional.**

Palabras Clave **- Aprendizaje Profundo, Extracción de Características, Lógica Difusa, Reconocimiento de Emociones, Unidades de Acción**

Abstract **- This article presents a procedure for the recognition of basic emotions in groups of people. To provide the necessary context, a review of the state of the art was carried out, addressing the most relevant topics in the field, highlighting key contributions, and analyzing the devices and methodologies that have been used over time, as well as their evolution to strengthen this discipline. Particular emphasis is placed on the importance of configuring a controlled environment that allows the classification of emotions under reliable conditions, through the careful selection and filtering of the material employed for prediction. For the preprocessing stage, samples were obtained from the FER2013, DISFA, and BP4D databases, to subsequently apply the selected methods to extract features of interest that allow achieving more accurate recognition. During the development of this work, a sample from a database composed of images and videos that capture people's emotional reactions was selected, with the purpose of using it effectively at the required moments. To propose a model for emotion recognition, different computational models and algorithms were implemented, including the Deep Learning library Fastai, as well as facial expression analysis using the Facial Action Coding System, along with the implementation of fuzzy logic, which allowed identifying the coincidence of one or more Action Units in two or more analyzed emotions, obtaining competitive results in recognition emotion classification.**

Keywords **– Action Units, Deep Learning, Emotion Recognition, Feature Extraction, Fuzzy Logic.**

I. INTRODUCCIÓN

La inteligencia artificial junto con el aprendizaje automático ha cambiado diversos sectores como la interacción humano-máquina, la psicología y la seguridad. Dentro de este avance tecnológico, el reconocimiento de emociones se ha establecido como un área clave para predecir las respuestas emocionales a partir de expresiones faciales, voz o comportamiento fisiológico por medio de tecnologías inteligentes (Khanzada et al., 2020).

El reconocimiento de emociones se ha constituido en aplicaciones en múltiples áreas en la investigación científica debido a su impacto en la aplicación de campos como salud, seguridad, educación etc. La recopilación y análisis de los datos emocionales a partir de rostros humanos requiere de metodologías avanzadas para garantizar una clasificación y predicción precisa. Para ello en este contexto, el análisis de emociones por medio de expresiones faciales se ha basado en las Action Units (AU's) y en las Facial Action Coding System (FACS) y con la integración de técnicas como Deep Learning permite predecir emociones transmitidas basadas en los rostros faciales de personas. Estas herramientas han permitido mejorar la fiabilidad en la identificación de emociones otorgando avances significativos por ejemplo Khanzada, Bai y Celepcikay (2020) implementaron una solución para el reconocimiento de emociones utilizando múltiples modelos de aprendizaje profundo a partir de imágenes faciales, con el objetivo de maximizar la precisión de predicción. Utilizando los modelos de aprendizaje

https://doi.org/10.62805/cid2025.CID035

por transparencia ResNet50, SeNet50 y VGG16, basado en redes convolucionales (CNN) y modelo de 5 capas, la argumentación de datos, el balanceo de clases, la adición de datos auxiliares y ensamble de modelos. Los resultados muestran una precisión del 75.8% en el conjunto de pruebas FER2013.

Białek et al. (2023), propone un método eficiente para el reconocimiento de emociones faciales utilizando redes neuronales convolucionales CNN tanto personalizadas como modelos de aprendizaje por transparencia. El modelo se probó inicialmente con el conjunto de datos FER2013, posteriormente, se introdujeron técnicas de filtrado para eliminar imágenes que no mostraban caras humanas o con etiquetas erróneas. Aplicando una clasificación binaria y multiclase para devolver un valor de probabilidad a cada clase de emoción. Se ealizaron ajustes en los hiperparámetros del modelo, tales como el tamaño de lote para entrenamiento, tasas de aprendizaje y métodos de regularización como el dropout para disminuir el sobreajuste. Con un modelo individual de RESNET50 se obtuvo un 72.72% de precisión, y otro con un modelo ensamblado de 4 redes neuronales (2 RESNET50 y 2 VGG16 entrenados en conjuntos de datos filtrados y balanceados) obtuvo una precisión del 76.90% utilizando el conjunto de datos filtrado del FER2013.

El método aplicado para el reconocimiento de expresiones faciales por los autores Halim et al. (2023), propone una solución realizando pruebas con tres modelos diferentes de aprendizaje profundo para visualizar su nivel de comprensión en estudiantes y revaluados con las bases de datos de macro-expresiones FER2013 y AffectNet-8. El modelo de enfoque utilizado por CNN utilizado para identificar las emociones en las expresiones faciales alcanzó un 68.51% con el conjunto de datos FER2013. El modelo ResNet50-CBAM utiliza un módulo de atención en bloque convolucional logró una precisión del 64.8% con el conjunto de FER2013 y el modelo DCNN-CBAM basado en aprendizaje profundo y con atención en bloque convolucional para mejorar el rendimiento alcanzó una precisión del 72.28% con el conjunto FER2013 y 66.09% con AffectNet-8.

Aplicando un modelo diferente de CNN para el reconocimiento de emociones que propone el autor Zhu et al. (2024) sobre una mejora a la red neuronal MobileNetV2, denominada como I-MobileNetV2 que está específicamente diseñada para el reconocimiento de expresiones faciales por medio de un mecanismo de atención en canales de Squeeze and Exitation Networks o SE-Net para mejorar la capacidad de extracción de características. Emplea una técnica de fusión inversa para conservar las características negativas en el proceso de convolución minimizando la pérdida de información y mejorando las tasas de clasificación. Los resultados demostraron que el modelo propuesto mejoró la precisión del reconocimiento facial en un 0.72% con la base de datos FER2013 y CK+ con un 6.14% y en comparación con la versión base MobileNetV2 logró reducir el número de parámetros en un 83.8%.

Además de diversas investigaciones sobre las capacidades de modelos basados en redes neuronales convolucionales (CNN) siempre se busca un punto para procesar la extracción de características reduciendo la cantidad de datos de entrada, por lo que el autor Gursesli et al. (2024) desarrolla una mejora de rendimiento para una CNN denominada "Custom Lightweight CNN-Based Model" (CLCM). Está basada en la arquitectura MobileNetV2 con la diferencia que ha sido optimizada para funcionar eficientemente en dispositivos con capacidades limitadas. Emplea un enfoque de aprendizaje por transparencia y una estructura de bloques individuales completamente conectadas con unidades de activación para mejorar la extracción de características y reducir la dimensionalidad de los datos de entrada. Ha sido evaluada utilizando cuatro conjuntos de datos públicos FER-2013 alcanzando 63%, RAF-DB con un 84%, AffectNet 54% y CK+ con un 78% de precisión.

Sin embargo, no solo se busca aplicar una única metodología basadas en CNN para predecir emociones, se desarrollaron pruebas donde se integraron otras metodologías como algoritmos que ayuden en el procesamiento de información y predicción de las emociones humanas, como el caso de este artículo por los autores Khaireddin y Chen (2021), proponen una solución centrada en el mejoramiento de precisión de predicción y clasificación de emociones faciales con el conjunto de datos FER2013 utilizando CNN con arquitectura VGGNet diseñando experimentos de optimización en los hiperparámetros del modelo. Así como la exploración de diferentes algoritmos de optimización como Adam y su variante AMSGrad y programadores de tasas de aprendizaje como la Reducción de la Tasa de Aprendizaje en el Plateau (RLRP), enfriamiento coseno entre otros. También aplicaron una normalización de datos durante el entrenamiento como el tamaño siendo de 40x40 pixeles. Con las técnicas combinadas alcanzan una precisión de clasificación hasta el 73.28% con el conjunto de datos FER2013.

II. PARTE TÉCNICA DEL ARTÍCULO

2.1 Bases de Datos

Para el desarrollo e investigación del reconocimiento de emociones enfocadas en el rostro humano, se requiere del uso de bases de datos que contengan imágenes de rostros humanos con expresiones emocionales diversas. Se consideraron las siguientes bases de datos principalmente para la fase de entrenamiento y evaluación de los modelos seleccionados. Los datos están estructurados y etiquetados para las emociones básicas de enojo, disgusto, miedo, alegría, sorpresa, tristeza y neutralidad.

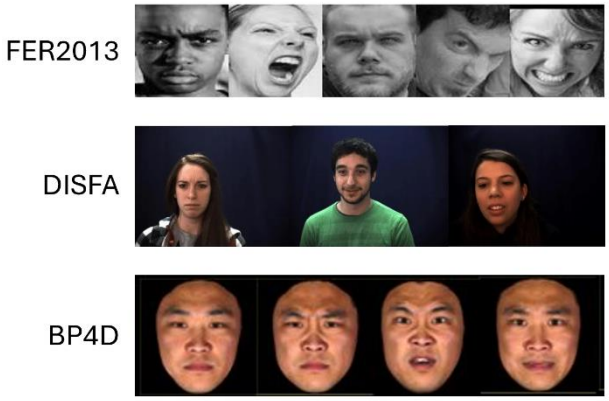
FER2013: La base de datos Facial Expression Recognition (FER2013) es un conjunto de datos comúnmente utilizado para el campo del reconocimiento de emociones faciales mediante técnicas de visión por computadora o implementación de inteligencia artificial. Contiene una cantidad aproximadamente de 35887 imágenes en escala de grises, con una resolución estándar de 48x48 pixeles, clasificadas en las emociones: Enojo, Disgusto, Miedo, Felicidad, Neutralidad, Tristeza y Sorpresa. El conjunto de datos está dividido en subconjuntos: entrenamiento con 28709 imágenes y prueba(validación) 7178 imágenes. La base de datos es de acceso libre y disponible en Kaggle en la página web: https://www.kaggle.com/deadskull7/fer2013

https://doi.org/10.62805/cid2025.CID035

DISFA: El conjunto de datos Denver Intensity of Spontaneous Facial Action (DISFA) consta de 27 vídeos de 4.844 fotogramas cada uno, con 130.788 imágenes en total. Las anotaciones de las unidades de acción están en distintos niveles de intensidad, etiquetas donde marcan las diferentes unidades de acción activadas en el rostro. Se seleccionó de entre una extensa variedad de bases de datos populares en el campo del reconocimiento de la expresión facial debido al elevado número de sonrisas, es decir, la unidad de acción 12. En concreto, 30.792 imágenes tienen definida esta unidad de acción, 82.176 imágenes tienen definida alguna unidad de acción y 48.612 imágenes no tienen definida ninguna unidad de acción. Es necesario solicitar autorización para acceder a la base de datos en el sitio oficial.

BP4D: El conjunto de datos BP4D-Spontaneous es una base de datos expresiones faciales en 3D a partir de video de un grupo diverso de adultos jóvenes. Se utilizaron inducciones de emociones bien validadas para provocar expresiones de emoción y comunicación paralingüística. El sistema FACS permitió obtener datos reales de las acciones faciales a nivel de fotograma. Los rasgos faciales se rastrearon tanto en 2D como en 3D utilizando enfoques genéricos y específicos para cada persona. La base de datos incluye cuarenta y un participantes (23 mujeres, 18 hombres). De diferentes edades y rasgos étnicos Se diseñó un protocolo de elicitación de emociones para elicitar eficazmente ocho emociones. La base de datos está estructurada por participantes. Los metadatos incluyen unidades de acción anotadas manualmente (FACS, AU), pose de la cabeza rastreada automáticamente y puntos de referencia faciales 2D/3D. Para obtener su acceso se requiere una solicitud de permiso en el sitio oficial. Ver Figura 1.

Figura 1
Ejemplo de imágenes de cada base de datos utilizada.

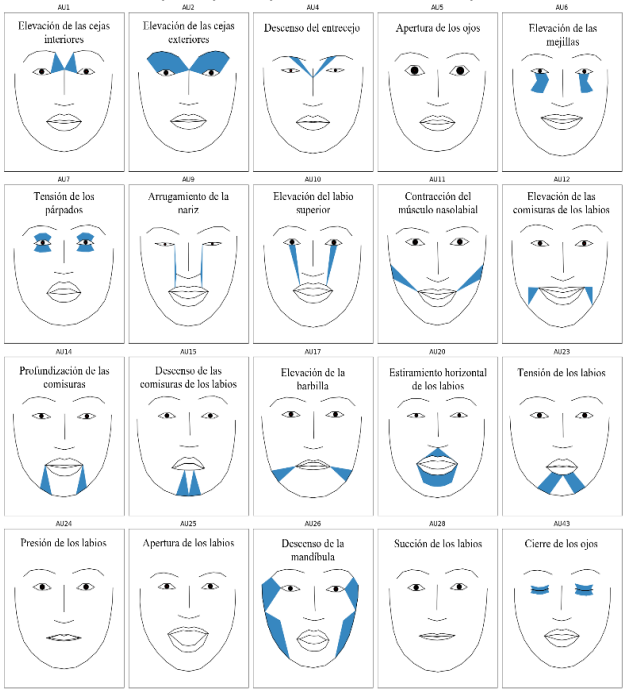


2.2 Marco Conceptual

A. Unidades de Acción

El sistema de Codificación de Acciones Faciales (FACS) desarrollado por Paul Ekman y Wallace Friesen consiste en un sistema basado en la anatomía facial donde busca medir todos los movimientos faciales discernibles y éstas se constituyen en 44 unidades de acción (AU). Cada AU tiene un código numérico la cual sus asignaciones son completamente arbitrarias. Donde se asocia con la contracción o el movimiento de uno o más músculos específicos, lo que permite analizar el rostro ante expresiones que puede mostrar una persona. Sin embargo, cada Acción Facial no tiene una asociación 1:1 entre los grupos musculares con cada AU, es decir que un músculo determinado puede responder de formas distintas para hacer acciones visibles que transmitan algo diferente. En la Figura 2 se muestran las 20 unidades de acción más utilizadas para el análisis e investigación sobre las emociones humanas (Ekman y Rosenberg, 1997).

Figura 2
Unidades de acción principales para el análisis de expresiones faciales.



Diversos investigadores han propuesto mapeos de AUs para la detección de emociones, se plantearon diversos enfoques que establezcan una relación entre las AU y las emociones humanas ya sea a través de combinaciones de algunas unidades o mediante asociación directa y determinar con mayor acercamiento que emoción estaría transmitiendo una persona. Como se muestra en la Tabla 1, se categorizan ciertas AU para la determinación de cada emoción correspondiente según la investigación de los diversos autores.

Tabla 1
Asociaciones propuestas entre las unidades de acción y emociones.

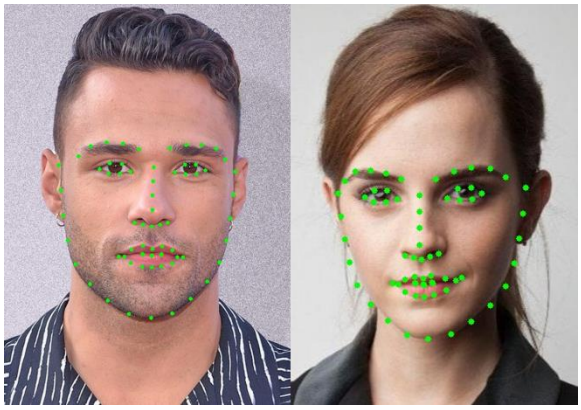
Emoción/ AUS	(Du et al., 2014)	(Yao et al., 2021)	(Contreras et al., 2010)	(Morales-Vargas et al., 2019)
Enojo	4,7,24	4,5,7,23	4,7	4,7
Disgusto	9,10,17	-	9,10	-
Miedo	1,4,20,25	1,2,4,5,7,20, 26	1,2,4	1,2,4,5,20, 27
Alegría	12,25	6,12	12	12,7
Tristeza	4,15	1, 4,15	1, 4,15	1,4,15
Sorpresa	1,2,25,26	1,2,5,26	1,2,5,27	1,2,5,27

https://doi.org/10.62805/cid2025.CID035

B. Landmarks

Los puntos de referencia (Landmarks) como se observa en la Figura 3, son puntos específicos anatómicos del rostro humano como el contorno de las cejas, la nariz, boca, mandíbula o las comisuras de los ojos. Estos puntos permiten capturar tanto deformaciones rígidas (como el movimiento de cabeza) o no rígidas (como las expresiones faciales), pueden ser detectados o localizados mediante algoritmos de aprendizaje automático o librerías especializadas en el área de visión por computadora con el objetivo de realizar análisis faciales como la geometría, estimación de pose, reconstrucción tridimensional y las expresiones faciales humanas (Wu y Ji, 2019).

Figura 3
Visualización de landmarks faciales mediante la librería face_recognition.



C. Lógica Difusa

La lógica difusa consiste en una extensión de la lógica tradicional que facilita el manejo de la información imprecisa tomando valores intermedios de pertenencia representados en un rango de 0 al 1 que a diferencia de la lógica booleana que solo toma un valor 0 o 1. Esta lógica se compone de un conjunto difuso que es una extensión de los conjuntos clásicos en aquellos elementos que no poseen una pertenencia absoluta, sino que puede pertenecer parcialmente a un conjunto en función de su grado de pertenencia. Mientras que una función de pertenencia de un grupo señala el nivel en el que cada elemento de un universo específico pertenece a un grupo en un valor entre 0 y 1. Existen diferentes tipos de funciones de pertenencia como: triangular, trapezoidal, sigmoidal, gaussiana y gamma, que dependen del contexto del problema. En la Figura 4 se muestra un ejemplo sobre un sistema difuso aplicado que determina la velocidad de un ventilador aplicando una función de pertenencia triangular para variable de salida (Ross, 2010).

Las funciones de pertenencia en el contexto de la lógica permiten representar matemáticamente el grado en que un elemento pertenece a un conjunto difuso. A diferencia de conjuntos clásicos donde la pertenencia es binaria, estos tienen un grado de pertenencia gradual y se expresa mediante valores entre el 0 al 1. Existen diversos tipos de funciones de pertenencia como se aprecia en la Tabla 2 algunas de las funciones de pertenencia más comunes, donde cada tipo cuenta con características específicas que se seleccionan en función al tipo de problema, el contexto de cada universo, el comportamiento o en la precisión requerida para representar el universo del problema y la toma de las decisiones en entornos de incertidumbre.

Figura 4
Ejemplo de aplicación de lógica difusa para control de velocidad.

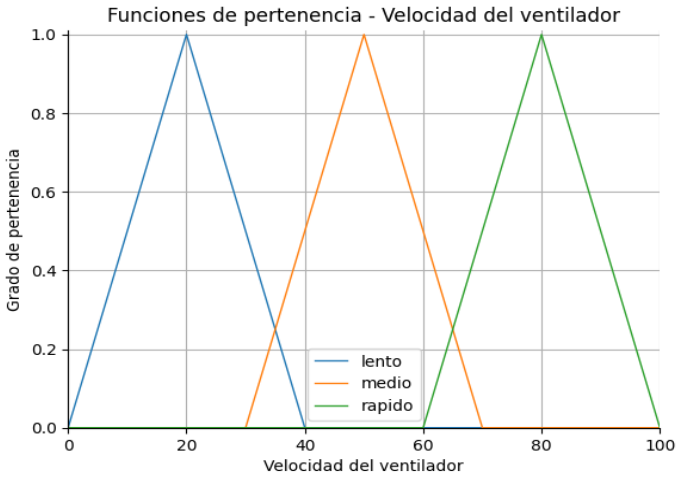
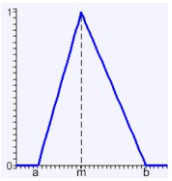
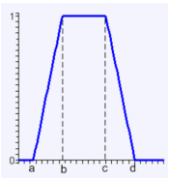
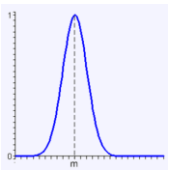
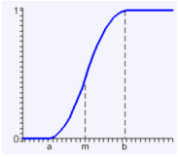


Tabla 2
Funciones de membresía utilizadas frecuentemente en la lógica difusa.

Nombre de la función	*Fórmula*	*Gráfica*
Triangular	$\mu(x) = \{0, \quad si\ x \leq a \ \frac{x-a}{m-a} \quad si\ a < x \leq m \ \frac{b-x}{b-m} \quad si\ m < x < b\ 0, \quad si\ x \geq b$	
Trapezoidal	$\mu(x) = \{0, \quad si\ (x < a)\ ó\ (x > d)\ \frac{x-a}{m-a}, \quad si\ a \leq x \leq b\ 1, \quad si\ b < x < c\ \frac{d-x}{d-c}, \quad si\ c \leq x \leq d$	
Gaussiana	$\mu(x) = e^{-k(x-m)^2}$	
Sigmoidal	$\mu(x) = \frac{1}{1+e^{-a(x-c)}}$	

https://doi.org/10.62805/cid2025.CID035

Año: 2025. Volumen: 1. Numero:11 ISSN: 2448-623X

2.3 Metodología Implementada

La metodología desarrollada para el análisis del reconocimiento de emociones humanas tiene un enfoque basado en la identificación de puntos de referencia mediante el uso de herramientas especializadas en el procesamiento y análisis de imágenes. El objetivo es reconocer y asociar a que AU pueden corresponder a partir de puntos clave utilizando el cálculo de la distancia entre dos puntos en el espacio mediante la distancia euclidiana. Posteriormente, esta información será utilizada como entrada en un prototipo de sistema difuso para el análisis y predicción de la emoción está transmitiendo el rostro. Esto es necesario porque la expresión del rostro humano es altamente variable y difícil de predecir debido a la gran variedad de rasgos que poseen las caras para comunicar emociones. Este método permitió llevar a cabo el análisis a través de tres etapas principales para alcanzar la evaluación y la predicción de emociones utilizando imágenes de rostros humanos.

1.- Procesamiento de Imagen y Obtención de Landmarks

Cuando se trata de análisis de imágenes digitales es fundamental aplicar un preprocesamiento adecuado que permite optimizar la calidad y estandarizar las imágenes a analizar. En esta primera fase consiste en la lectura de las imágenes que serán analizadas, una por una será convertida a escala de grises con el propósito de reducir la complejidad de los datos. Posteriormente se lleva a cabo un recorte de la imagen que permite enfocarse en el área de interés que es el rostro y eliminando información irrelevante, por medio de la biblioteca *Dlib*. Se trata de una herramienta desarrollada en C++ que proporciona un amplio número de algoritmos enfocados para el aprendizaje profundo, visión por computadora y procesamiento de imágenes, entre sus características más destacadas se encuentra en el soporte de detección facial y regresión estructurada y clasificación (King, 2009).

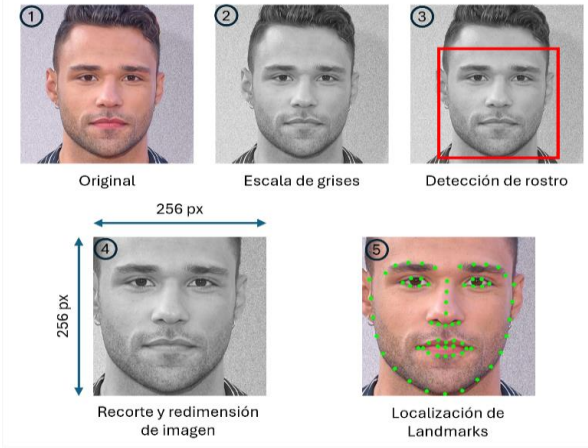
Utilizando la función *dlib.get_frontal_face_detector()* detecta rostros frontales y guarda la información por medio de coordenadas de localización de contorno de imagen que será enviado para recortar la imagen y únicamente analizar el área de interés. Finalmente, se realiza la estandarización del tamaño de imagen a una resolución de 256x256 pixeles, garantizando que todos los datos de entrada cuenten con las mismas dimensiones para posteriormente procesar la información en la obtención de las landmarks. Se tomó esta medida de imagen normalizada debido a que la detección de landmarks se basa en coordenadas bidimensionales, si el tamaño de una imagen es grande o pequeño las coordenadas serán variables según sea la imagen, por lo que el proceso de análisis se vuelve complejo y se decidió que todas las imágenes se analicen bajo esta resolución.

Una vez la imagen procesada, se procede a localizar los landmarks a través de la misma biblioteca con "*shape_predictor_68_face_landmarks.dat*" para la localización de los 68 puntos de referencia faciales otorgando una matriz de la localización en las coordenadas (x,y) de cada landmark en base a la imagen analizada. En la Figura 5 se muestra el proceso mencionado que pasa una imagen tomada de una de las bases de datos para la obtención de sus correspondientes landmarks.

2.- Asociación de Landmarks y Cálculo de Distancias Euclidianas para Activación de AUs

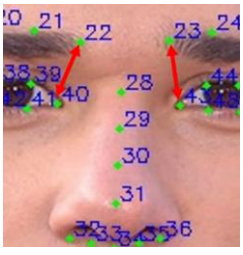
Para la segunda fase, con la información de las coordenadas de cada landmark se realizan dos procesos para obtener una salida sobre a que AU estarán presentes en el rostro. Al tratarse de información basada en coordenadas bidimensionales, el cálculo para determinar que AU's se encuentran en el rostro, se basará en distancias de determinados landmarks que estén asociadas a cada AU correspondiente para establecer a qué punto se encuentran presentes. Lo cual el primer proceso es la determinación y extracción de landmarks específicos para calcular su distancia entre estos dos puntos. Ver Figura 5.

Figura 5
Procesamiento de imagen para la obtención de las landmarks.



Primeramente, se realizó un análisis para identificar que landmarks pueden asociarse con las AU, por ejemplo, en la Figura 6 se muestra como las landmarks 22 y 23 son puntos clave para determinar la activación de la AU1 referente a la "Elevación de cejas inferiores" y se estableció como punto de referencia para calcular su distancia un punto fijo como es el landmark 40 y 43 referentes al contorno del ojo interno. Por lo que para cada AU se determinaron diferentes puntos específicos dependiendo de qué parte del rostro se realiza la activación de determinadas AU, como el caso de la boca, se puede ahí extraer información para la AU25 "Apertura de labios" y AU12 "Elevación de comisuras de los labios" por lo que el landmark 52 y 58 otorgan información a la distancia de apertura de la boca si es muy abierta, se determina que la AU25 está activada. En la Tabla 3 se muestran algunas asignaciones de landmarks para asociar las AUs más utilizadas.

Figura 6
Selección de landmarks para determinar la activación de la AU1.



https://doi.org/10.62805/cid2025.CID035

Año: 2025. Volumen: 1. Numero:11 ISSN: 2448-623X

Tabla 3
Asignación de landmarks y tipo de activación para algunas AUs.

AU	Nombre	Landmarks asociadas	Activación (Distancia)
1	Elevación de cejas interiores	(22,40) (23,43)	Creciente
4	Descenso del entrecejo	(21,28) (24,28)	Decreciente
5	Apertura de ojos	(38,42) (48,46)	Creciente
9	Arrugamiento de la nariz	(27,31) (27,35)	Decreciente
12	Elevación de comisuras de los labios	(48,54)	Creciente
23	Tensión en los labios	(67,58), (65,56)	Decreciente
25	Apertura de labios	(61,68), (63,65)	Creciente

Se calcularon las distancias de la AU1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 17, 20, 23, 24, 25, 28 y 43 que son las más relevantes para análisis y reconocimiento de emociones. Obtenido las distancias de las AUs, el segundo proceso es el cálculo de sus distancias con respecto a las landmarks establecidas para cada AU asociada mediante el cálculo de la distancia euclidiana, en (1) se muestra la obtención de la distancia entre dos puntos en un espacio bidimensional.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Se colocan las coordenadas de cada landmark que se referenció para obtener la distancia entre estos dos puntos y determinar en una escala de activación para la AU correspondiente su intensidad. Mediante pruebas realizadas de una muestra de imágenes de las bases de datos se determinaron unas reglas de activación como en la Figura 7 se muestra algunas de reglas de activación, estableciendo su escala de activación entre una distancia mínima y máxima, % de activación y tipo de intensidad. Por ejemplo, si es creciente, la distancia tendrá una intensidad de activación más alta.

Figura 7
Reglas de activación para la determinación de intensidad de las AUs.

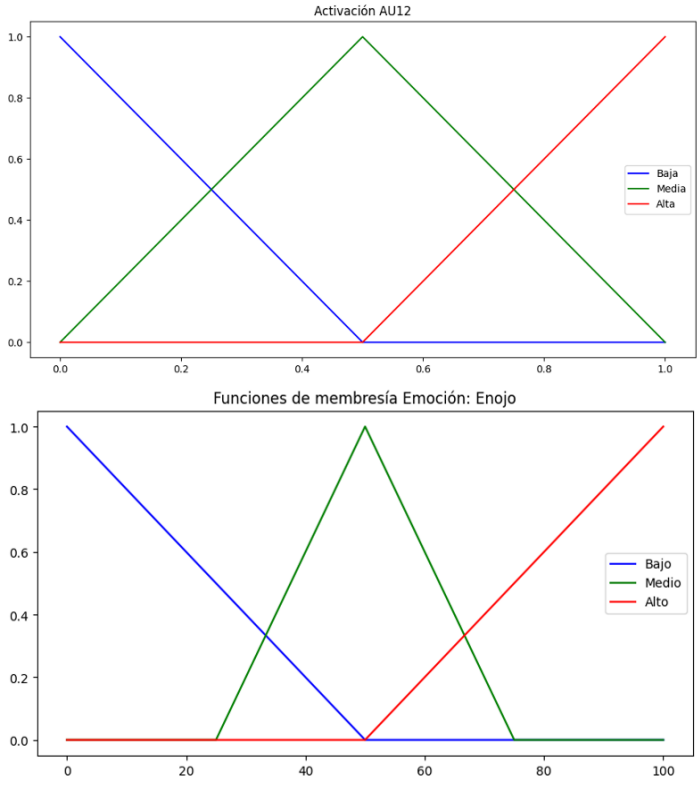
```
Reglas de activación de AUs (Escala 0 - 1):
--------------------------------------
AU1:
  Rango de distancia: 15.0 a 70.0
  Intensidad creciente con una distancia de:
  0.0 de intensidad a: 15.0
  1.0 de intensidad a: 70.0

AU2:
  Rango de distancia: 20.0 a 65.0
  Intensidad creciente con una distancia de:
  0.0 de intensidad a: 20.0
  1.0 de intensidad a: 65.0

AU4:
  Rango de distancia: 30.0 a 80.0
  Intensidad decreciente con una distancia de:
  0.0 de intensidad a: 80.0
  1.0 de intensidad a: 30.0
```

3.- Modelo Difuso de AUs para la Predicción de Emociones

Para esta última fase, se propone el uso de un modelo basado en lógica difusa, se decidió este modelo debido a que los datos faciales son imprecisos al momento de realizar una interpretación de qué emoción puede estar transmitiendo el rostro. A través de los niveles de activación de las AU del proceso anterior se genera un conjunto de reglas difusas que permite inferir la emoción predominante de la imagen. Para la implementación del modelo, se emplea la biblioteca *scikit-fuzzy,* una herramienta en Python diseñada para el desarrollo de sistemas difusos mediante funciones de pertenencia, operadores lógicos y técnicas de inferencia difusa (Warner et al., 2024).

El universo de entrada está configurado en base al rango de 0.0 a 1.0 basado en la activación de las AU en la fase previa. En la Figura 8 se muestra las funciones de membresía para las entradas del sistema, donde se estableció en 3 niveles de activación (bajo, medio y alto) para cada AU con una función de tipo triangular para cada nivel, lo mismo para el universo de las salidas donde se predecirá la emoción que más domine en la imagen, se han establecido como salidas seis: enojo, disgusto, miedo, felicidad, tristeza y sorpresa con 3 funciones de membresía de la misma que las entradas en un rango de 0 a 100 para manejar las salidas en %.

Figura 8
Gráficas de los universos de entrada y salida del sistema difuso.



Para la determinación de las reglas se basaron en las referencias de las AU para detectar emociones como se muestra en la Tabla 2, ajustando como mínimo 3 AUs por emoción, sin embargo,

https://doi.org/10.62805/cid2025.CID035

existen AUs, que ayudan a determinar a más de una emoción, pero no con el mismo valor de peso que con otras. Ejemplo, la emoción de enojo se utilizan las AU 4, 7 y 23 pero también la emoción de tristeza comparte con la AU 4 pero no con el mismo valor ya que para esta emoción el AU15 tiene mayor relevancia para determinar un estado de tristeza. En la Tabla 4 se muestran algunas reglas establecidas y un peso de relevancia para determinar las emociones de enojo, tristeza y felicidad.

Tabla 4
Reglas y establecimiento de AUs para la predicción de emociones.

Emoción	*AU*	*Peso de relevancia*	*Nivel de relevancia*
Enojo	4	0.5	Alta
	7	0.3	Alta
	23	0.2	Baja
Felicidad	12	0.5	Alta
	6	0.35	Alta
	25	0.15	Baja
Miedo	1	0.2	Baja
	4	0.1	Baja
	5	0.3	Alta
	20	0.4	Alta
Sorpresa	5	0.3	Alta
	2	0.1	Baja
	25	0.5	Alta
	26	0.1	Baja
Tristeza	15	0.6	Alta
	1	0.2	Baja
	4	0.2	Baja
Disgusto	9	0.6	Alta
	10	0.3	Alta
	17	0.1	Baja

Con esta información buscamos darle un peso importante para establecer que algunas AUs tienen una relevancia más importante para predecir una emoción, pero requiere de otras para complementar o dar soporte que se pueda tratar de esa emoción. Por lo que se establecieron las reglas difusas en base a la relevancia de cada AU que tanto influye en cada emoción y otorgar una salida de predicción

III. RESULTADOS Y DISCUSIÓN

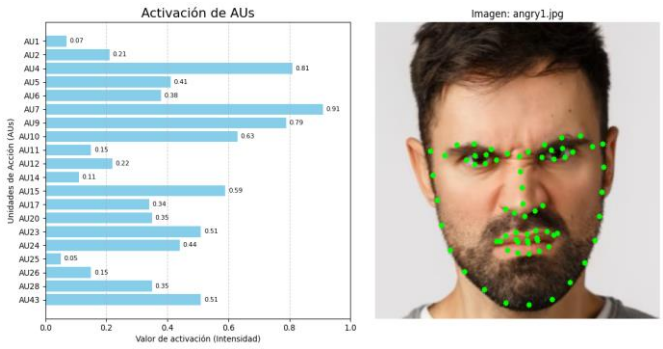
Los resultados del análisis se llevaron a cabo integrando una pequeña muestra de imágenes que transmitían la emoción correspondiente a su etiqueta. Para ello con un algoritmo de aprendizaje profundo implementado con *Fastai* una biblioteca principalmente para tareas de Deep Learning, fue entrenado con una de las bases de datos y que determine qué imágenes son clasificadas correctamente con la emoción descrita para que sean analizadas y realizar la prueba para el sistema (Howard y Gugger, 2020). Al realizar las pruebas correspondientes con una toma de imágenes de las diversas bases de datos obtenidas, se colocan en una carpeta para ser evaluadas y obtener los resultados correspondientes. En la Tabla 5 se muestran datos de las primeras dos imágenes de cada emoción al pasar por la primera fase obteniendo las distancias para cada AU correspondiente y posteriormente calcular el nivel de intensidad de cada una.

Tabla 5
Datos de las distancias de las primeras AU de algunas imágenes.

imagen	*AU1*	*AU2*	*AU4*	*AU5*	*AU6*
angry1.jpg	19.07	29.24	39.37	12.265	68.875
angry2.jpg	18.93	30.515	45.67	8.5625	76.585
disgust1.jpg	24.05	28.745	42.5	9.085	75.7375
disgust2.jpg	38.28	35.54	55.655	9.035	88.03
fear1.jpg	37.93	23.145	47.27	18.285	77.635
fear2.jpg	50.4	31.145	61.885	16.765	64.51
happy1.jpg	33.29	28.885	51.655	9.0425	75.1675
happy2.jpg	36.87	35.34	57.285	15.105	84.0125
sad1.jpg	41.73	29.43	53.745	15.2975	69.39
sad2.jpg	33.37	30.495	48.975	9.7625	77.855
surprise1.jpg	47.36	40.835	63.855	18.5675	81.7525
surprise2.jpg	50.72	38.775	71.86	24.3025	86.74

En la Figura 9 se muestra la activación de las AU en función de las distancias calculadas para las imágenes de prueba, ejemplo en una imagen catalogada como Enojo se estima una activación de las AU4, 7 o del 23 para representar esta emoción, en este análisis se demuestra la presencia de las AU4 y 7 con mayor presencia significando que las landmarks elegidas para la activación de esas AU proporcionan la información más cercana para demostrar la relación con la AU correspondiente y también a la emoción de enojo.

Figura 9
Análisis de activación de AU para una imagen catalogada como Enojo.

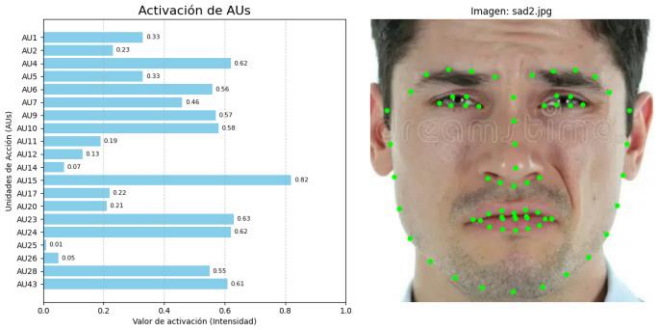


En la Figura 10 se muestra otro análisis sobre una imagen etiquetada como Triste y en su gráfica de activación promete corresponder con las AU descritas para determinar una emoción de Tristeza, donde la AU15 tiene un nivel de activación considerable sin embargo autores también describen que la AU1 o AU4 son indispensables para predecir esta emoción, más sin embargo el hecho de detonar en esta investigación que para cada emoción se encuentra una AU que tiene mayor peso que otras para determinar con la emoción correspondiente y en esta caso el AU15 si tiene una mayor activación puede darle un mayor peso que se trata de una emoción de tristeza.

h t t p s : / / d o i . o r g / 1 0 . 6 2 8 0 5 / c i d 2 0 2 5 . C I D 0 3 5

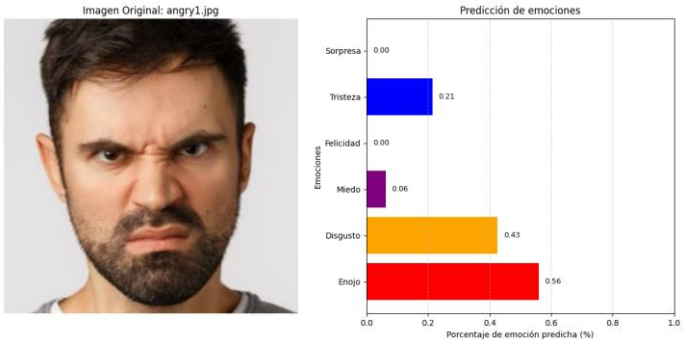
Año: 2025. Volumen: 1. Numero:11 ISSN: 2448-623X

Figura 10
Análisis de activación de AU para una imagen catalogada como Triste.



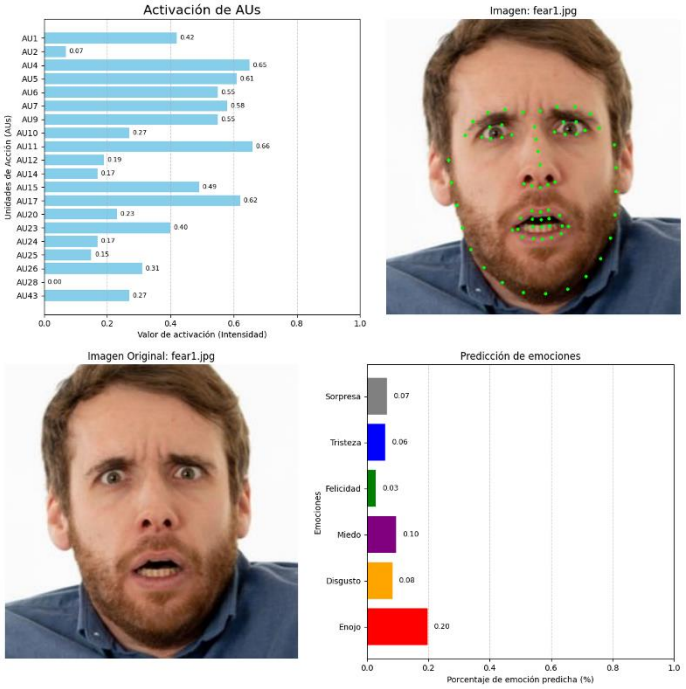
Para la última fase, con la información de activación de las AU de cada imagen, pasa a las entradas del sistema difuso en base a las reglas diseñadas y condiciones para predecir de posibles emociones se están transmitiendo en cada imagen. En la Figura 11 se muestra el resultado final aplicando la lógica difusa y prediciendo en este caso para la imagen con etiqueta de enojo donde se observa que la emoción predominante es el enojo pero tiene un valor cercano que es el de disgusto, esto es debido a que en la Figura 9 se aprecia que en la AU9 tiene un alto nivel de intensidad y normalmente esta unidad tiene un mayor peso para la detección de la emoción de disgusto por lo que hace una presencia en esta gráfica que el sistema infiere que también puede presentarse la emoción de disgusto, sin embargo como las AU definidas para el enojo están más presentes hacen que el valor de presencia sea más alta que la de disgusto.

Figura 11
Análisis de la salida del sistema difuso para predecir emociones.



Sin embargo, de todas las emociones se detectó que para las imágenes de Miedo hay mucha variabilidad entre ellas y no son detectadas a la emoción que corresponde por lo que aún se está analizando que AU podrían ser relevantes para la detección de esta emoción. En la Figura 12 se muestra el análisis de una imagen etiquetada como miedo y donde se visualiza que tiene en su mayoría las AU presentes mayor a 40% lo que dificulta el análisis de detección de emoción prediciendo una emoción no correctamente predicha a la que demuestra. Esto crea un desafío en detectar que AUs serían relevantes e ideales para la detección de emoción al igual y determinar cuál de ellas influye más para determinar que se una emoción en particular.

Figura 12
Análisis completo para una imagen etiquetada como Miedo.



IV. CONCLUSIONES

En los resultados del caso de estudio apunta que el proceso de detección de emociones por medio de modelos inteligentes basadas en el análisis y procesamiento de imágenes faciales ofrece una vía efectiva para interpretar las emociones humanas de manera automatizada en base a rasgos o puntos de interés del rostro que puedan interpretar y predecir la emoción a transmitir. Con la implementación de metodologías de extracción de landmarks y las Unidades de Acción permitió establecer una base de análisis sobre los puntos de interés que pueden dar la suficiente información en cuestión para la detección de emociones. Con la implementación de un sistema basado en lógica difusa resultó adecuado para modelar la variabilidad que presenta el rostro humano, debido a que cada persona transmite sus emociones de manera diferente, complicando detectar las emociones generalizando reglas que en algunos casos no pueda aplicar.

No obstante, este sistema aún se encuentra en desarrollo, por lo que la mejora en el establecimiento de nuevas o modificaciones de reglas para la detección de algunas emociones que resultan ser complicadas de detectar como el caso de la emoción de Miedo que requiere de un amplio número de AU para lograr detectarlo. Como también se identificaron áreas de oportunidad orientadas a la mejora de la precisión del sistema.

Asimismo, sería conveniente incorporar nuevas metodologías o modelos inteligentes que mejoren el procesamiento para la detección a través de aprendizaje automático en cuestión de optimización de las reglas y funciones de pertenencia.

https://doi.org/10.62805/cid2025.CID035

Finalmente, se está investigando a fondo sobre la implementación de integración de información de los rostros mediante secuencias de vídeo, lo cual podría enriquecer la detección de emociones abarcando más áreas que solo la parte frontal del rostro y explorar áreas laterales o en mismo ángulo del rostro pero que en segundos puede presentar muecas o algún movimiento que ayude a la detección de la emoción en secuencias más amplias de información.

V. AGRADECIMIENTOS

Queremos expresar un sincero agradecimiento a todas las personas que hicieron posible el desarrollo de esta investigación. En particular a mi asesor por su tiempo, dedicación, conocimiento y apoyo durante el tiempo de desarrollo de este trabajo y para la publicación de este artículo. También se agradece profundamente a los colaboradores y profesores en la orientación y aportaciones de su conocimiento para la correcta ejecución de esta investigación. Al Tecnológico Nacional de México / Instituto Tecnológico de Apizaco por brindar el entorno académico, la infraestructura y los recursos necesarios para llevar a cabo este estudio. Su compromiso con la formación y la investigación fue pilar clave durante el proceso de este trabajo. Y a SECIHTI por el apoyo financiero recibido a través del Programa de Becas de Posgrado Nacionales lo cual hizo posible llevar a cabo esta investigación.

VI. REFERENCIAS

Białek, C., Matiolański, A., & Grega, M. (2023). An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks. *Electronics*, *12*(12), 2707. https://doi.org/10.3390/electronics12122707

Contreras, R., Starostenko, O., Alarcon-Aquino, V., & Flores-Pulido, L. (2010). Facial Feature Model for Emotion Recognition Using Fuzzy Reasoning. J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, & J. Kittler (Eds.), *Advances in Pattern Recognition* (Vol. 6256, pp. 11–21). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15992-3_2

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, *111*(15), 1454–1459. https://doi.org/10.1073/pnas.1322355111

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179644.001.0001

Gursesli, M. C., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini, A., & Lanata, A. (2024). Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. *IEEE Access*, *12*, 45543–45559. https://doi.org/10.1109/ACCESS.2024.3380847

Halim, A., El-Manfy, A., Badr, A. E.-R., El-Khatib, A., El-Basir, M. A., El-Tabee, S., El-Den, Z. A., & El-Khouly, A. (2023). Facial expressions analysis to evaluate the level of students' understanding. *2023 Intelligent Methods, Systems, and Applications (IMSA)*, 424–429. https://doi.org/10.1109/IMSA58542.2023.10217489

Howard, J., & Gugger, S. (2020). *Deep learning for coders with fastai and PyTorch: AI applications without a PhD*. O'Reilly Media.

Khaireddin, Y., y Chen, Z. (2021). *Facial Emotion Recognition: State of the Art Performance on FER2013* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2105.03588

Khanzada, A., Bai, C., & Celepcikay, F. T. (2020). *Facial expression recognition with deep learning* [Preprint]. *arXiv*. https://doi.org/10.48550/arXiv.2004.11823

King, D. E. (2009). Dlib-ml: A machine Learning Toolkit. *Journal of Machine Learning Research*, *10*(60), 1755-1758. https://www.jmlr.org/papers/volume10/king09a/king09a.pdf)

Morales-Vargas, E., Reyes-García, C. A., & Peregrina-Barreto, H. (2019). On the use of Action Units and fuzzy explanatory models for facial expression recognition. *PLOS ONE*, *14*(10), e0223563. https://doi.org/10.1371/journal.pone.0223563

Ross, T. J. (2010). *Fuzzy Logic With Engineering Applications* (3rd ed.). Wiley.

Warner Josh, Jason Sexauer, Wouter Van den Broeck, Bruno P. Kinoshita, Jakub Balinski, scikit-fuzzy, Christian Clauss, twmeggs, alexsavio, Aishwarya Unnikrishnan, Marco Miretti, Guilherme Castelão, Felipe Arruda Pontes, Tobias Uelwer, phme283, pd2f, laurazh, Fernando Batista, moetayuko, … Thomas Germer. (2024). *JDWarner/scikit-fuzzy: Scikit-Fuzzy 0.5.0* (*Versión v0.5.0*) [Software]. Zenodo. https://doi.org/10.5281/ZENODO.802396

Wu, Y., & Ji, Q. (2019). Facial Landmark Detection: A Literature Survey. *International Journal of Computer Vision*, *127*(2), 115–142. https://doi.org/10.1007/s11263-018-1097-z

Yao, L., Wan, Y., Ni, H., & Xu, B. (2021). Action unit classification for facial expression recognition using active learning and SVM. *Multimedia Tools and Applications*, *80*(16), 24287–24301. https://doi.org/10.1007/s11042-021-10836-w

Zhu, Q., Zhuang, H., Zhao, M., Xu, S., & Meng, R. (2024). A study on expression recognition based on improved mobilenetV2 network. *Scientific Reports*, *14*(1), 8121. https://doi.org/10.1038/s41598-024-58736-x