

UIBAIFED: Un dataset de expresiones faciales generado por IA para visibilizar la diversidad

UIBAIFED: An AI-generated facial expression dataset to enhance diversity

Esperança Amengual-Alcover

Departament de Ciències
Matemàtiques i informàtica
Universitat de les Illes Balears
Palma de Mallorca, España
eamengual@uib.es

Maria Francesca Roig-Maimó

Departament de Ciències
Matemàtiques i informàtica
Universitat de les Illes Balears
Palma de Mallorca, España
xisca.roig@uib.es

Ramon Mas-Sansó

Departament de Ciències
Matemàtiques i informàtica
Universitat de les Illes Balears
Palma de Mallorca, España
ramon.mas@uib.es

Miquel Mascaró-Oliver

Departament de Ciències
Matemàtiques i informàtica
Universitat de les Illes Balears
Palma de Mallorca, España
miquel.mascaro@uib.cat

Recibido: 15.11.2025 | Aceptado: 09.12.2025
DOI: <https://doi.org/10.65234/interaccion.139>

Palabras Clave

Interacción persona-ordenador
IPO
Aprendizaje profundo
Conjunto de datos de expresiones faciales
FER

Resumen

Este artículo presenta UIBAIFED, un nuevo conjunto de datos de expresiones faciales compuesto por imágenes realistas y de alta calidad, etiquetadas con grupo de edad, género, etnia y 22 microexpresiones basadas en las expresiones universales de Ekman y la taxonomía de Gary Faigin. El conjunto de datos fue generado mediante modelos de difusión y está diseñado para mejorar la investigación en reconocimiento de expresiones faciales, aumentando la representación étnica, generacional y de complejidad física, visibilizando la diversidad existente en el mundo real. La validación inicial mediante una red neuronal convolucional (CNN) alcanzó una precisión del 80% en la clasificación de microexpresiones. Además, se realizó un experimento con participantes humanos con un subconjunto de datos formado por imágenes del grupo de edad de personas mayores (85 años), colectivo tradicionalmente marginado en los conjuntos de datos existentes. Los mejores resultados de clasificación fueron para las emociones de tristeza (94.6%), neutral (92.8%) y alegría (83.8%), mientras que las peores emociones identificadas fueron las emociones de miedo (33.9%) y asco (34.7%), con una precisión global del 74%. UIBAIFED ofrece un nivel de etiquetado más detallado que los conjuntos de datos existentes, lo que facilita el análisis del rendimiento en sistemas de reconocimiento de expresiones faciales en distintos grupos demográficos y contribuye al desarrollo de modelos más robustos y generalizables.

Keywords

Human Computer Interaction
HCI
Machine learning
Facial expression dataset
FER

Abstract

This article presents UIBAIFED, a new facial expression dataset composed of high-quality, realistic images labelled with age group, ethnicity, and 22 micro-expressions based on Ekman's universal expressions and Gary Faigin's taxonomy. The dataset was generated using diffusion models and is designed to support advanced research in Facial Expression Recognitions (FER), increasing ethnical, generational and complexity representativity, enhancing real world diversity. The initial validation using a Convolutional Neural Network (CNN) achieved an accuracy of 80% in micro-expression classification. To complement this automatic evaluation, an experiment with human participants was conducted using a subset of data composed by images corresponding to elder adults (85 years old), a group traditionally underrepresented in existing datasets. The best results were obtained when classifying sadness (94.6%), neutral (92.8%) and joy (83.8%) emotions. We got the worst results when classifying fear (33.9%) and disgust (34.7%) emotions. The overall accuracy was of 74%. UIBAIFED provides a more detailed annotation level than existing facial expression datasets, enabling performance analysis of FER systems across different demographic groups and contributing to the development of more robust and generalizable models.

1. Introducción

El reconocimiento de expresiones faciales (FER, por sus siglas en inglés *Facial Expression Recognition*) ha experimentado avances significativos en los últimos años, impulsados en gran medida por las mejoras técnicas de aprendizaje profundo (Li & Weng, 2022) y la creciente disponibilidad de conjuntos de datos (*datasets*) de alta calidad (Kollias et al., 2020). Estos conjuntos de datos desempeñan un papel crucial en el entrenamiento de modelos capaces de interpretar con precisión expresiones faciales en diferentes contextos. Sin embargo, los conjuntos de datos existentes aún presentan desafíos relacionados con la diversidad demográfica, los desequilibrios entre clases y cuestiones éticas como el sesgo en la representación (Buolamwini & Gebru, 2018). Abordar estos problemas es esencial para desarrollar modelos de FER más robustos y generalizables.

A pesar de la creciente disponibilidad de conjuntos de datos para el reconocimiento de expresiones faciales, *datasets* ampliamente utilizados como Fer2013 (Goodfellow et al., 2013), CK+ (Lucey et al., 2010), RAF-DB (Li et al., 2017) y AffectNet (Mollahosseini et al., 2019) presentan limitaciones. Entre ellas, se incluyen desequilibrios entre clases, donde algunas emociones, como por ejemplo la felicidad, están sobrerrepresentadas, mientras que otras, como el miedo o el asco, permanecen infrarrepresentadas (Fan et al., 2022). Además, muchos conjuntos de datos están compuestos principalmente por poblaciones jóvenes y de origen occidental, lo que limita la capacidad de generalización de los modelos de FER hacia grupos menos representados (Domínguez-Catena et al., 2024).

Para mitigar estas limitaciones, investigaciones previas han explorado enfoques alternativos, como las técnicas de aumento de datos (*data augmentation*) y la generación sintética de expresiones faciales, con el fin de mejorar la diversidad de los datos y el rendimiento de los modelos (Psaroudakis & Kollias, 2022). Sin embargo, hasta donde llega nuestro conocimiento, no existe actualmente ningún conjunto de datos de FER disponible públicamente que haya sido generado íntegramente mediante Inteligencia Artificial (IA).

En este trabajo presentamos UIBAIFED (*UIB Artificial Intelligence Facial Expression Dataset*), el primer conjunto de datos generado mediante inteligencia artificial diseñado para mejorar el entrenamiento y la evaluación de los modelos de reconocimiento de expresiones faciales. A diferencia de los conjuntos de datos tradicionales, UIBAIFED utiliza técnicas de inteligencia artificial generativa para crear un conjunto diverso y equilibrado de expresiones faciales. Este enfoque garantiza un corpus de entrenamiento más

representativo para los sistemas modernos de FER, reduciendo los sesgos demográficos y mejorando la robustez general de los modelos.

2. Trabajo relacionado

2.1 Conjuntos de datos tradicionales de FER

Varios conjuntos de datos han sido ampliamente utilizados en la investigación sobre reconocimiento de expresiones faciales, entre ellos *FER2013*, CK+, RAF-DB y *AffectNet*. Estos conjuntos de datos han contribuido de manera significativa al avance de los modelos de aprendizaje profundo para la clasificación de emociones; sin embargo, con frecuencia presentan limitaciones tales como:

1. **Desequilibrios demográficos:** muchos conjuntos de datos se centran en poblaciones jóvenes y de origen occidental, lo que da lugar a modelos que generalizan de forma deficiente en grupos menos representados (Domínguez-Catena et al., 2024).
2. **Desequilibrios entre clases:** algunas emociones, como la felicidad y la emoción neutra, están representadas con mayor frecuencia que otras, como el miedo o el asco, lo que puede generar un rendimiento sesgado en los modelos (Zhang et al., 2023).
3. **Inconsistencias en el etiquetado:** las diferencias en la forma en que se anotan las emociones entre distintos conjuntos de datos pueden introducir ruido y dificultar la capacidad de generalización de los modelos (Chen & Joo, 2021).

Estas limitaciones han motivado a los investigadores a desarrollar nuevos conjuntos de datos más equilibrados y diversos que permitan una mejor capacidad de generalización de los modelos de FER.

2.2 Reconocimiento de microexpresiones

Las microexpresiones son expresiones faciales breves e involuntarias que revelan emociones genuinas. Su naturaleza efímera las hace difíciles de capturar y clasificar, aunque resultan fundamentales en campos como la psicología, la seguridad y la interacción persona-ordenador (Adegun & Vadapalli, 2020).

Una de las principales carencias de los conjuntos de datos actuales de FER es la ausencia de un etiquetado sistemático de las microexpresiones. A diferencia de los conjuntos de datos estándar, que se centran en categorías emocionales más amplias, las microexpresiones requieren una mayor granularidad y una anotación precisa. Esta limitación dificulta

el desarrollo de modelos capaces de detectar indicios emocionales sutiles en aplicaciones en tiempo real (Guerdelli et al., 2022).

La categorización de expresiones faciales propuesta por Faigin ofrece un marco integral para comprender la dinámica facial más allá de las siete categorías emocionales tradicionales (Faigin, 1990). Esta taxonomía destaca la complejidad de las expresiones, capturando variaciones sutiles que se suelen pasar por alto en los estudios convencionales de FER. Sin embargo, los conjuntos de datos existentes rara vez incorporan este nivel de detalle, lo que limita la capacidad de los modelos actuales para reconocer los matices emocionales. Superar esta brecha requiere conjuntos de datos diseñados explícitamente para alinearse con la categorización de Faigin.

3. Métodos

Para abordar los desafíos mencionados anteriormente, en este trabajo presentamos UIBAIFED, un conjunto de datos generado mediante inteligencia artificial, diseñado para ofrecer una presentación más equilibrada y diversa de las expresiones faciales. Al aprovechar el potencial de los modelos generativos, garantizamos variaciones controladas en cuanto a edad, género y etnia, manteniendo al mismo tiempo diferencias realistas en la postura, la iluminación y la intensidad de la expresión. Este enfoque tiene como objetivo mitigar los sesgos presentes en los conjuntos de datos tradicionales y mejorar la robustez de los modelos de reconocimiento de expresiones faciales.

3.1 Modelos faciales

Para garantizar la calidad del conjunto de datos, las imágenes generadas cumplen con los siguientes criterios: el rostro debe estar centrado y ocupar entre el 40% y el 70% del área de la imagen; la iluminación debe ser suficiente para resaltar con claridad los detalles de la expresión facial, mientras que el fondo debe permanecer uniforme y neutro, a fin de evitar posibles interferencias en la clasificación. Además, las expresiones faciales deben replicar con precisión las descripciones propuestas por Gary Faigin (1990). Asimismo, los defectos visuales deben ser mínimos y no comprometer la expresividad del rostro.

El conjunto de datos UIBAIFED garantiza una distribución equilibrada en cuanto al sexo, cinco grupos de edad diferenciados (véase la figura 1) y tres categorías de composición corporal (véase la figura 2).



Figura 2: Distribución de grupos de edad en el conjunto de datos UIBAIFED.

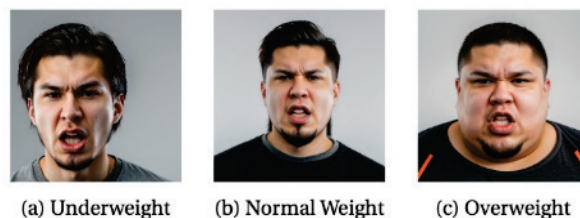


Figura 2: Categorías de composición corporal en el conjunto de datos UIBAIFED.

La diversidad étnica se considera según la clasificación establecida por la OMB (*Office of Management and Budget*) de los Estados Unidos (*Office of Institutional Research*, s.f.), la cual incluye grupos como nativos americanos, asiáticos, personas negras, hispanos, nativos de Hawái u otras islas del Pacífico y personas blancas de ascendencia europea, norteafricana o de Oriente Medio (véase la figura 3).

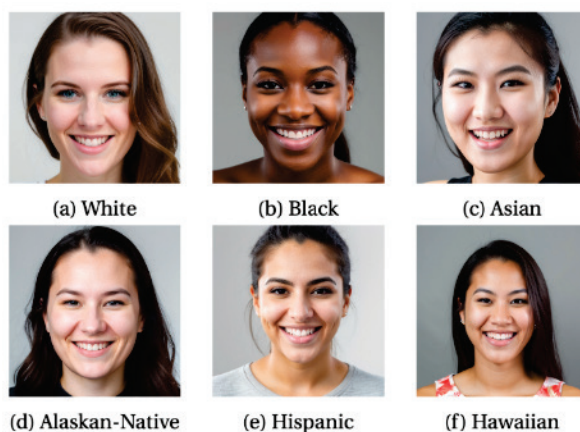


Figura 3: Distribución de la diversidad étnica en el conjunto de datos UIBAIFED.

3.2 Proceso de generación y filtrado de imágenes

Para la generación de imágenes de expresiones faciales en el conjunto de datos UIBAIFED, se utilizó el modelo *Stable Diffusion* (Stable Diffusion AI, s.f.) Esta tecnología de código abierto puede ejecutarse localmente, lo que ofrece la ventaja de generar un número ilimitado de imágenes. Su naturaleza flexible y las continuas contribuciones de la comunidad han permitido el desarrollo de versiones mejoradas, aumentando la variedad y la calidad de los resultados y garantizando que las imágenes cumplan con los criterios establecidos para el análisis de expresiones faciales.

Los *checkpoints* de *Stable Diffusion* son modelos pre-entrenados diseñados para generar imágenes a partir de descripciones textuales, donde se emplean grandes conjuntos de datos para aprender las correlaciones entre palabras y elementos visuales. La selección de un *checkpoint* requiere considerar su capacidad para generar una amplia variedad de imágenes con todas las características requeridas, al mismo tiempo que se minimiza el tiempo de generación. En base a hallazgos empíricos, se determinó que el *checkpoint Realistic Vision* (Civitai, s.f.-b) es el que mejor se adapta a las necesidades del conjunto de datos.

Para optimizar el modelo en la generación de expresiones faciales, se empleó la técnica *Low-Rank Adaptation (LoRA)* (Hu et al., 2022; Zeng and Lee, 2023). *LoRA* permite adaptar rápidamente modelos de aprendizaje automático a nuevos contextos mediante la incorporación de componentes ligeros al modelo original, en lugar de modificar toda su estructura. En el caso de *Stable Diffusion*, se utilizaron *LoRAs* específicamente entrenadas para la generación de expresiones faciales, obtenidas de *CivitAI* (Civitai, s.f.-a). La tabla 1 muestra las *LoRAs* utilizadas para la generación del conjunto de datos UIBAIFED.

Tabla 1: LoRAs utilizadas para la generación del conjunto de datos UIBAIFED

LoRA	Categoría emocional
Sad - Facial Expression	Tristeza
Look alive! Excited Facial Expression/Emotion	Entusiasmo
Ashamed - Embarrassed Facial Expression	Vergüenza
Look alive! Smirking Facial Expression/Emotion	Alegría
Disgusted - Disapproving Facial Expression	Asco
Angry - Facial Expression	Ira
Shocked - Facial Expression	Sorpresa
Pleading Eyes - Facial Expression	Miedo
Scared - Facial Expression	Miedo
Expressions Helper Realistic	De uso general

Además, se desarrollaron los *prompts* necesarios para la generación de las micro expresiones que conforman el conjunto de datos. De las 33 micro expresiones descritas por Gary Faigin, solo se logró reproducir un subconjunto de 22, debido a la dificultad de describir ciertas sutilezas mediante modelos generativos.

A continuación, se muestra un ejemplo de los *prompts* generados (positivos y negativos):

--prompt "White Man, 15y.o,
(AngryShouting:0),(angry!!),
(((shouting!!!))),
<lora:l_ang_ae_sd_64_32:0.9
>, Underweight, ((looking at
the camera)), hyperrealistic,
professional photo, studio
lighting, sharp focus,
centered on the image,
vertical alignment, face, plain
grey background"

--negative_prompt
"((Deformed)), disfigured,
hat,(artifacts in eyes, bad iris),
((artifacts in face)), hawaiian
clothes, worse quality, low
quality, jpeg, pixelated,
anime, ((poorly illuminated
face)), red eyes, ((bad teeth)),
((body, arms, hands, legs,
naked))"

El *prompt* descrito anteriormente genera la imagen mostrada en la figura 4, que representa a un varón de 15 años de complexión delgada expresando la emoción de ira; concretamente, la micro expresión *Angry Shouting*, según la taxonomía de Faigin.



Figura 4: Ejemplo de una imagen generada correspondiente a la expresión de ira (varón blanco de 15 años, complexión delgada).

La estructura de los distintos *prompts* se mantiene de forma consistente, siguiendo el formato que se indica a continuación:

*"Ethnicity, gender, age,
<description of the
expression>"*

Dentro de la expresión, se incluye la referencia a la *LoRA* utilizando la siguiente nomenclatura:

<lora: (LoRA name):(weight)>

En esta estructura, el término "*weight*" se refiere a la intensidad de la expresión. Algunas micro expresiones se generan utilizando la misma *LoRA*, pero con descriptores diferentes. Por ejemplo, las microexpresiones *NearlyCrying* y *Sad*, ambas representando la tristeza, se generan con los siguientes dos *prompts* que producen las imágenes mostradas en la figura 5, utilizando la misma *LoRA*.

*--prompt "Black Woman,
25y.o, (NearlyCrying:0), ((sad
mouth)), miserable face,
(sad:1.2),
<lora:l_sad_se_sd_64_32:1>,
Overweight, ((looking at the
camera)), hyperrealistic,
professional photo, studio
lightning, sharp focus,
centered on the image,*

*vertical alignment, face, plain
grey background"*

*--prompt "Black Woman,
25y.o, (Sad:0), (sad),
(melancholic face), closed lips,
small mouth,
<lora:l_sad_se_sd_64_32:1>,
Overweight, ((looking at the
camera)), hyperrealistic,
professional photo, studio
lightning, sharp focus,
centered on the image,
vertical alignment, face, plain
grey background"*

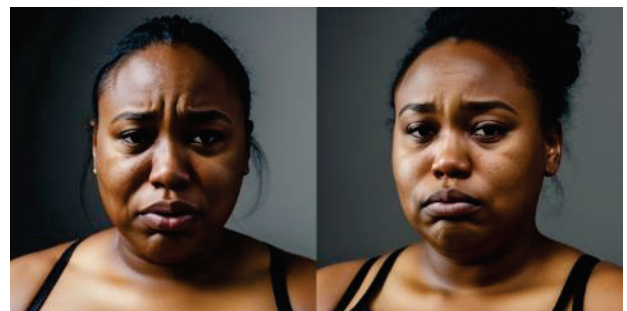


Figura 5: Expresiones *NearlyCrying* y *Sad* generadas con la misma *LoRA*.

Los paréntesis y los valores numéricos se utilizan para destacar palabras o frases específicas.

La figura 6 muestra las 22 expresiones generadas para un varón blanco de 15 años. Se desarrolló un *script* automatizado para generar 3960 *prompts*, resultado de la combinación de 2 géneros, 6 etnias, 5 grupos de edad y 3 tipos de complexión corporal, todos organizados según las seis expresiones universales definidas en la clasificación de Ekman (1999).

Las imágenes correspondientes a los *prompts* generados se produjeron utilizando la aplicación *Automatic111* (2022). Debido a la naturaleza aleatoria del proceso de generación de imágenes, no todas las imágenes resultan precisas en el primer intento. Por ello, para cada micro expresión, se

generaron entre 15 y 30 imágenes, con el fin de garantizar la calidad y la coherencia deseadas.

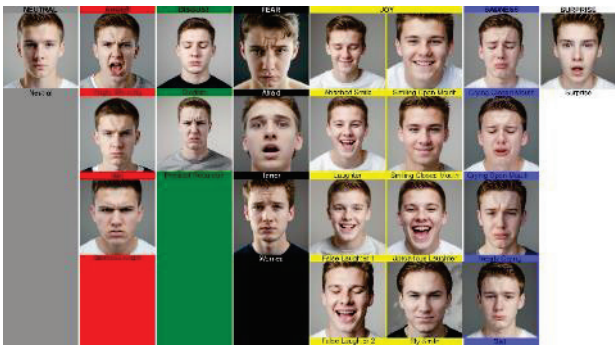


Figura 6: Ejemplos de las 22 expresiones generadas para un varón blanco de 15 años.

Las imágenes generadas a partir de los *prompts* especificados fueron seleccionadas manualmente en función de su grado de similitud con las descripciones y representaciones gráficas proporcionadas por Gary Faigin (1990).

La figura 7 ilustra el proceso de comprobación manual para la micro expresión *SlySmile*. La imagen de la izquierda representa la expresión generada a partir del conjunto de datos UIBAIFED, mientras que la imagen de la derecha corresponde a la ilustración de referencia del trabajo de Gary Faigin. Este proceso de selección garantizó que cada imagen representara con precisión la expresión facial prevista y cumpliera con los criterios establecidos.

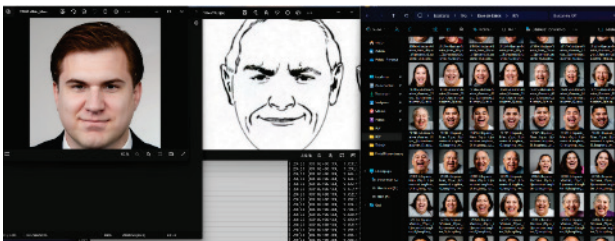


Figura 7: Proceso de comprobación manual para la expresión *SlySmile*.

4. El conjunto de datos UIBAIFED

El número total de imágenes que conforman el conjunto de datos es de 2948. Las imágenes generadas a partir de los *prompts* con diferentes tipos de complexión corporal fueron eliminadas debido a las mínimas diferencias observadas entre aquellas etiquetadas como *peso normal* y *bajo peso*.

Se mantuvieron un mayor número de representaciones para las expresiones más complejas, lo que dio lugar a la

distribución de imágenes por microexpresión que se muestra en la tabla 2.

Tabla 2: Distribución de imágenes por microexpresión

Expresión Universal	Microexpresión	Número de imágenes
Neutralidad	Neutral	123
Ira	AngryShouting	121
	Mad	165
	SternnessAnger	126
Asco	Disdain	161
	PhysicalRepulsion	151
Miedo	Afraid	121
	Terror	124
	Worried	135
Alegria	AbashedSmile	129
	Laughter	121
	FalseLaughter1	120
	FalseLaughter2	122
	SmilingOpenMouth	120
	SmilingClosedMouth	121
	UproariousLaughter	119
	SlySmile	120
Tristeza	CryingClosedMouth	132
	CryingOpenMouth	158
	NearlyCrying	158
	Sad	155
Sorpresa	Surprise	146

La base de datos está organizada en carpetas, cada una de las cuales contiene imágenes con una resolución de 512×512 píxeles. Existe una carpeta para cada una de las seis expresiones universales según la clasificación de Ekman (1999).

Es importante señalar que la expresión *Desdén* (*Comptent*), que Ekman añadió posteriormente a su clasificación original, está etiquetada en nuestro conjunto de datos como la micro expresión *Disdain*. Esta expresión se incluye dentro de la carpeta *Disgust*, siguiendo el enfoque de clasificación propuesto por Gary Faigin.

Dentro de cada carpeta, las imágenes se nombran siguiendo el siguiente formato:

*Num_ethnicity_gender_age_
microexpression.png*

El elemento "Num" representa el número de generación asignado por *Stable Diffusion*, indicando el orden de las imágenes dentro de cada carpeta.

Las imágenes se organizan primero por microexpresión y, posteriormente, por etnia, género y edad.

5. Validación del conjunto de datos UIBAIFED

Para validar inicialmente el conjunto de datos UIBAIFED, se empleó un modelo simple de red neuronal convolucional (*CNN*, *Convolutional Neural Network*) para la clasificación de expresiones faciales.

El modelo recibe como entrada imágenes en escala de grises de 128×128 píxeles, que se procesan a través de tres capas convolucionales. Estas capas van seguidas de una unidad lineal rectificadora (*ReLU*, *Rectified Linear Unit*) y una capa de *max pooling*, utilizadas para extraer las características más relevantes.

La arquitectura también incluye cuatro capas totalmente conectadas que se usan para clasificar las expresiones faciales en una de las 22 micro expresiones objetivo descritas en el conjunto de datos.

La estructura general de la red se muestra en la figura 8.

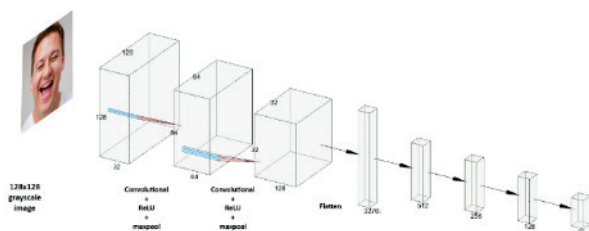


Figura 8: Estructura del modelo de red neuronal para la validación inicial del conjunto de datos UIBAIFED.

Para mejorar la capacidad de generalización y evitar el sobreajuste, se aplica una capa *dropout* entre las capas totalmente conectadas. Esta técnica ayuda a que la red aprenda características más robustas, eliminando aleatoriamente algunas unidades durante el entrenamiento, lo que mejora la capacidad del modelo para generalizar a datos no vistos.

El conjunto de datos se divide en subconjuntos de entrenamiento y prueba, utilizando el 67% de los datos para el entrenamiento y el 33% restante para las pruebas.

La distribución de los datos se equilibra únicamente en función de los tipos de microexpresiones, sin considerar otros factores como género, tipo de complejión corporal, etnia o edad en esta etapa de validación. Estos factores serán explorados en estudios futuros.

6. Resultados

Tras completar el proceso de entrenamiento, se obtuvo un valor de *loss* cercano a 0.5 y una precisión global (*accuracy*) del 82%. Estos resultados se alcanzaron utilizando el 67% de las imágenes para el entrenamiento, lo que equivale a un total de 1975 imágenes.

El modelo *CNN* entrenado fue evaluado con el conjunto de prueba (compuesto por 1975 imágenes), alcanzando una precisión global del 85.71%.

La matriz de confusión resultante se presenta en la figura 9, mientras que la tabla 3 detalla las métricas de rendimiento obtenidas para cada una de las 22 micro expresiones. La tabla 4 presenta un resumen de las métricas globales de clasificación.

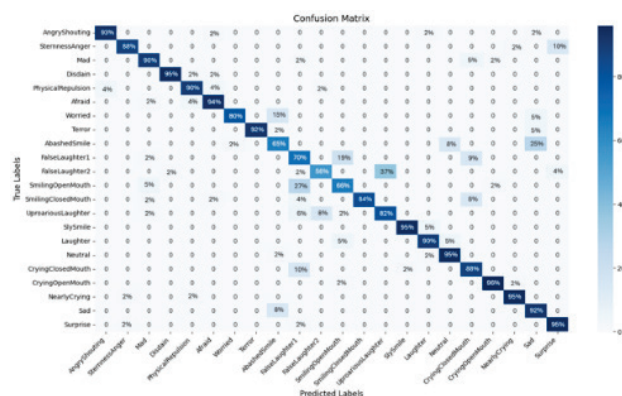


Figura 9: Matriz de confusión de los resultados de clasificación del modelo CNN.

Tabla 3: Métricas de rendimiento para cada microexpresión

Microexpresión	Precisión	Recall	F1-Score	Soporte
Neutral	0.92	0.82	0.87	40
AngryShouting	0.84	0.95	0.89	43
Mad	0.77	0.82	0.80	40
SternnessAnger	0.86	0.90	0.88	40
Disdain	0.97	0.86	0.91	43
PhysicalRepulsion	0.88	0.87	0.87	52
Afraid	0.94	0.85	0.89	53
Terror	0.93	0.97	0.95	40
Worried	0.73	0.80	0.76	40
AbashedSmile	0.68	0.65	0.67	40
Laughter	0.93	0.97	0.95	40
FalseLaughter1	0.57	0.72	0.63	54
FalseLaughter2	0.67	0.27	0.38	52
SmilingOpenMouth	0.61	0.61	0.61	41
SmilingClosedMouth	1.00	0.82	0.90	50
UproariousLaughter	0.51	0.76	0.61	51
SlySmile	0.97	0.97	0.97	40
CryingClosedMouth	0.72	0.69	0.71	42
CryingOpenMouth	0.87	0.98	0.92	48
NearlyCrying	0.97	0.93	0.95	41
Sad	0.82	0.79	0.81	39
Surprise	0.85	0.80	0.82	44

Tabla 4: Resumen de las métricas globales de clasificación

Métrica	Precisión	Recall	F1-Score
accuracy		0.80	
macro avg	0.82	0.81	0.81
weighted avg	0.81	0.80	0.80

Los resultados de la prueba indican que el modelo *CNN* ha aprendido y generalizado correctamente la mayoría de las expresiones faciales en un conjunto de validación de más de 900 imágenes que no se utilizaron durante el entrenamiento.

La mayoría de las expresiones fáciles alcanzan una precisión superior al 75%. Sin embargo, las expresiones relacionadas con la alegría presentan mayores dificultades de clasificación. En concreto, *AbashedSmile* se clasifica erróneamente en ocasiones como *Sad* o *Worried*, mientras que *FalseLaughter1* se confunde con frecuencia con *CryingOpenMouth*.

Un patrón recurrente observado en todos los ciclos de entrenamiento y prueba es la confusión entre *FalseLaughter2* y *UproariousLaughter*. La principal dificultad para distinguir estas expresiones radica en su gran semejanza visual: ambas representan una boca ampliamente abierta y ojos cerrados o casi cerrados. Este problema ya se había previsto durante el proceso de filtrado de imágenes, donde se observó que las diferencias visuales entre ambas expresiones eran mínimas (véase la figura 10).

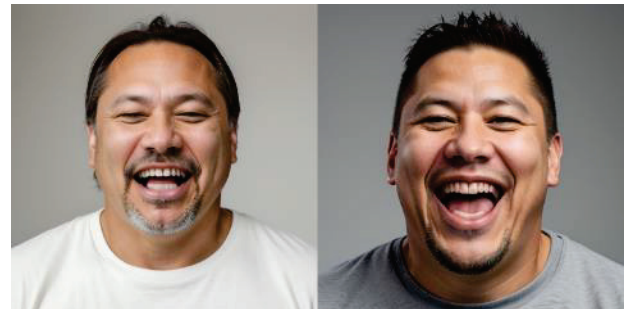


Figura 10: Comparación entre *FalseLaughter2* y *UproariousLaughter*

7. El experimento con humanos

Una vez comprobado que el conjunto de datos generado puede ser utilizado para entrenar modelos de inteligencia artificial capaces de clasificar imágenes en micro expresiones, se realizó una validación desde la perspectiva del reconocimiento humano para comprobar hasta qué punto el conjunto de datos sintético preserva las características esenciales de las expresiones y verificar que los rasgos generados son reconocibles. La validación consistió en un experimento basado en el reconocimiento, por parte de un grupo de participantes humanos, de un subconjunto de expresiones faciales de UIBAIFED centrado en las imágenes del grupo de edad de 85 años. La elección del subconjunto de datos se basó en el criterio de que, tradicionalmente, el colectivo de personas mayores se encuentra infrarrepresentado en los conjuntos de datos de FER. Además, debido a las modificaciones físicas provocadas por el envejecimiento (como la aparición de arrugas, la flacidez de la piel o la pérdida de firmeza), la externalización de las expresiones faciales en este colectivo puede verse alterada, dificultando, en consecuencia, el reconocimiento de las emociones desde el punto de vista del observador.

7.1 Participantes

37 participantes (22 mujeres) fueron reclutados entre el personal del campus de la universidad local y conocidos. Las edades oscilaron entre 20 y 81 con una media de 52 años ($SD = 13.69$). No hubo requisitos de experiencia previa para participar en el estudio.

7.2 El cuestionario

El cuestionario estaba formado por 263 preguntas. Cada pregunta consistía en una imagen de un personaje escogido de forma aleatoria entre las 576 imágenes generadas para el grupo de edad de 85 años del conjunto de datos UIBAIFED. El número final de imágenes correspondientes a cada emoción se mantuvo proporcional al número de imágenes generados por cada emoción en este subconjunto de edad. Junto a la expresión facial se mostraron 7 opciones de respuesta correspondientes a las 6 emociones universales más la neutra: miedo, tristeza, ira, alegría, sorpresa, asco y neutral. Para cada una de las emociones también se proporcionaban algunos de sus sinónimos más habituales (véase la figura 11).

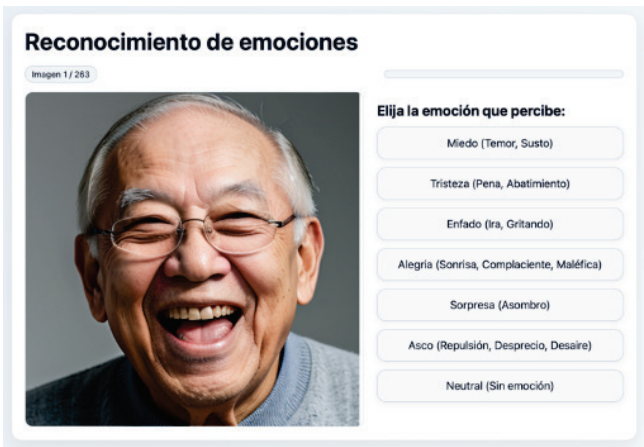


Figura 11: Ejemplo de una pregunta del cuestionario

Para cada una de las 263 imágenes se obtuvo una respuesta por participante con el nombre de la emoción percibida. La tabla 5 muestra el número de imágenes por cada emoción utilizadas en el cuestionario. El orden de las preguntas presentadas a los participantes fue aleatorio.

Tabla 5: Imágenes por emoción utilizadas en el cuestionario

Emoción	Número de imágenes
Alegría	96
Tristeza	47
Ira	36
Miedo	36
Asco	24
Sorpresa	12
Neutral	12

7.3 Procedimiento

El cuestionario fue completado online por los participantes. Después de explicar los objetivos del experimento, se

instruyó a los participantes para que se sentaran y respondieran el cuestionario con las 263 preguntas. Se les indicó que observaran la imagen presentada en la pregunta e intentaran identificar la emoción correspondiente a la expresión facial mostrada por el personaje. Después de presentar las instrucciones, se proporcionó a los participantes el enlace para poder completar el cuestionario.

Responder al cuestionario tuvo una duración media de unos 18 minutos.

7.4 Resultados

En esta sección presentamos los resultados de clasificación obtenidos en términos de precisión (*accuracy*) y matriz de confusión.

La Figura 12 muestra la matriz de confusión obtenida con la clasificación de emociones realizada por los participantes humanos. Los mejores resultados de clasificación fueron para las emociones tristeza (94.6%), neutral (92.8%) y alegría (83.8%), mientras que las peores emociones identificadas fueron las emociones de miedo (33.9%) y asco (34.7%). Las emociones más habitualmente confundidas fueron asco con tristeza (24.9%), miedo con sorpresa (33.5%) y tristeza con sorpresa (23.1%), y sorpresa con miedo (21.4%).

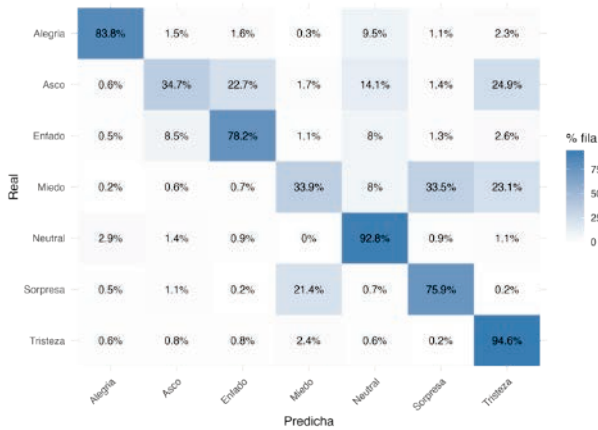


Figura 12: Matriz de confusión de la clasificación de las emociones realizada por los participantes humanos

La clasificación de emociones realizada por los participantes obtuvo una precisión global de 0.74. Analizando el efecto del sexo sobre la clasificación de emociones, no se obtuvo significación estadística para el efecto del sexo sobre la precisión ($F_{1,35} = 0.05$, ns).

El análisis de los resultados de clasificación de emociones en función de las micro expresiones, tal como se presenta en la figura 13, permite interpretar con mayor precisión las confusiones observadas en la clasificación.

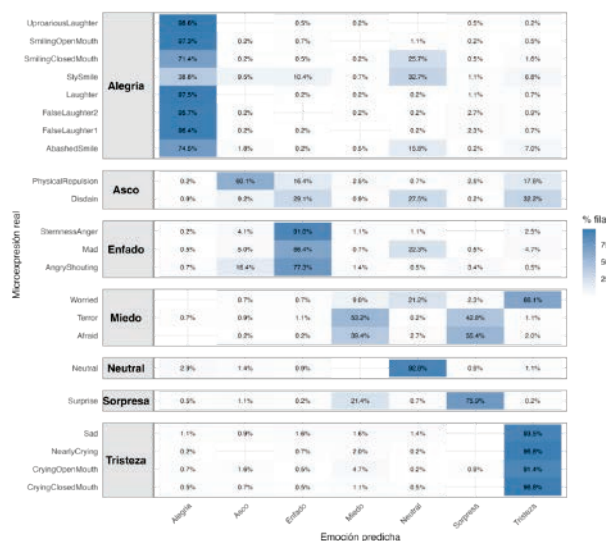


Figura 13: Matriz de confusión de la clasificación de cada una de las micro expresiones en emociones realizadas por los participantes humanos

En el caso de la emoción de alegría, las micro expresiones se caracterizan por la activación del músculo cigomático. En aquellas categorizaciones en las que dicha activación es más intensa (*UproariousLaughter*, *SmilingOpenMouth*, *Laughter*, *FalseLaughter*, ...), la tasa de confusión es mínima. En cambio, las expresiones de baja intensidad, como la risa maliciosa o irónica (*Slysmile*), en las que la elevación de las comisuras labiales es apenas perceptible, muestran una tendencia recurrente a ser clasificadas como expresión neutral.

La emoción de asco incluye dos micro expresiones diferenciadas: la de repulsión física (*PhysicalRepulsion*) y la de desdén (*Disdain*). Los datos indican una confusión clara entre la repulsión y las emociones de enfado y tristeza. Esta confusión puede atribuirse a la relación estrecha entre dichas emociones, así como a la coincidencia en la actividad muscular de la región de las cejas y de la boca, especialmente en expresiones vinculadas al llanto. La expresión de desdén, incorporada más recientemente al conjunto de emociones consideradas universales (Ekman, 1999), presenta una intensidad muscular reducida, lo cual dificulta su reconocimiento en ausencia de un contexto situacional. La confusión con la tristeza puede considerarse, en este sentido, coherente desde un punto de vista perceptivo.

En las expresiones de enfado, la mayor parte de las confusiones se concentra en la micro expresión de menor

intensidad (*Mad*), que tiende a ser categorizada como neutral debido a la limitada activación muscular observable.

Las expresiones de miedo constituyen la categoría con menor precisión de clasificación, hecho atribuible principalmente a la falta de contexto situacional. Las confusiones con sorpresa y tristeza resultan coherentes con esta limitación. De manera similar, la confusión entre sorpresa y terror puede explicarse por la similitud en los patrones de activación muscular propios de ambas expresiones. La expresión de preocupación (*Worried*) fue identificada como tristeza por el 66.1% de los participantes, lo que sugiere, no tanto una confusión expresiva como una diferencia en la apreciación subjetiva de la emoción representada.

Por último, las micro expresiones de tristeza alcanzan de media una precisión superior al 94%.

8. Conclusiones y trabajo futuro

En este estudio se ha presentado y evaluado UIBAIFED, un nuevo conjunto de datos de expresiones faciales generado por IA compuesto por imágenes en color de alta calidad y realismo, etiquetadas según grupo de edad, género, etnia y expresión facial. El etiquetado sigue las expresiones universales, abarcando un total de 22 micro expresiones basadas en la terminología propuesta por Gary Faigin.

UIBAIFED supone un avance respecto a los conjuntos de datos tradicionales al mejorar la representación étnica, generacional y de complejión física, contribuyendo así a visibilizar la diversidad presente en la población real.

Para validar el conjunto de datos se utilizó una red neuronal convolucional, que alcanzó una precisión del 80% y mostró un rendimiento adecuado en la mayoría de las expresiones faciales. Adicionalmente, se diseñó un experimento con participantes humanos con el fin de evaluar hasta qué punto el conjunto de datos sintético preserva los rasgos distintivos de las expresiones. El experimento se llevó a cabo utilizando un subconjunto de imágenes del grupo de edad de 85 años, un colectivo tradicionalmente marginado en los conjuntos de datos existentes y que, debido a los cambios físicos en la piel provocados por el envejecimiento, puede dificultar el reconocimiento de las expresiones faciales. Los mejores resultados de reconocimiento correspondieron a las emociones de tristeza (94.6%), neutral (92.8%) y alegría (83.8%), mientras que las emociones de miedo (33.9%) y asco (34.7%) presentaron las tasas más bajas, alcanzándose una precisión global del 74%.

El análisis detallado del reconocimiento de micro expresiones permitió identificar las principales fuentes de confusión: la baja intensidad muscular de alguna de las micro expresiones, el solapamiento en los patrones de activación muscular de la región de las cejas y la boca, y la ausencia de un contexto situacional claro.

En comparación con los conjuntos de datos de expresiones faciales existentes, UIBAIFED introduce una contribución fundamental al ofrecer un nivel de etiquetado notablemente más detallado. Hasta donde alcanza nuestro conocimiento, no existe actualmente ninguna base de datos que proporcione un grado similar de granularidad en la anotación de micro expresiones. Además, el uso de la inteligencia artificial generativa facilita la generación de un conjunto de datos más diverso, incluyendo colectivos que tradicionalmente se encuentra infrarrepresentados en los conjuntos de datos de FER.

Este nuevo conjunto de datos abre nuevas oportunidades para la investigación futura, particularmente en el análisis del reconocimiento de expresiones faciales en distintos grupos de edad y etnia. Abordar estos desafíos contribuirá de forma significativa al avance del campo del reconocimiento automático de expresiones faciales.

Agradecimientos

Este trabajo forma parte del Proyecto PID2022-136779OB-C32 (PLEISAR), financiado por el MICIU/AEI /10.13039/501100011033/ y FEDER, UE, y del Proyecto PID2023-149079OB-I00 (EXPLAINME), financiado por el MICIU/AEI/10.13039/501100011033/ y FEDER, UE.

Declaración sobre el uso de la IA generativa

Durante la preparación de este trabajo, los autores utilizaron *Stable Diffusion 1.5* y *Realistic Vision V6.0 B1* para la generación de las imágenes que componen el conjunto de datos UIBAIFED.

Asimismo, se empleó *ChatGPT* (versión GPT-4, octubre de 2025) para asistir en la revisión gramatical, ortográfica y de estilo del lenguaje.

Tras el uso de estas herramientas, los autores revisaron y editaron el contenido según lo consideraron necesario y asumen la plena responsabilidad del contenido final de la publicación

Referencias

- Adegun, I. P., & Vadapalli, H. B. (2020). Facial micro-expression recognition: A machine learning approach. *Scientific African*, 8, e00465. <https://doi.org/10.1016/j.sciaf.2020.e00465>
- AUTOMATIC1111. (2022). *Stable diffusion web UI* [Software]. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. En *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91). PMLR. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Chen, Y., & Joo, J. (2021). Understanding and mitigating annotation bias in facial expression recognition. *arXiv*. <https://doi.org/10.48550/arXiv.2108.08504>
- Civitai. (s. f.-a). *LoRA Stable Diffusion & Flux AI models*. <https://civitai.com/tag/lora>
- Civitai. (s. f.-b). *Realistic Vision V6.0 B1 – V5.1 Hyper (VAE)*. <https://civitai.com/models/4201/realistic-vision-v60-b1>
- Dominguez-Catena, I., Paternain, D., & Galar, M. (2024). Metrics for dataset demographic bias: A case study on facial expression recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8), 5209–5226. <https://doi.org/10.1109/TPAMI.2024.3361979>
- Ekman, P. (1999). *Handbook of cognition and emotion*. Wiley.
- Faigin, G. (1990). *The artist's complete guide to facial expression*. Watson-Guption.
- Fan, J., Zhou, J., Deng, X., Wang, H., Tao, L., & Kwan, H. K. (2022). Combating uncertainty and class imbalance in facial expression recognition. En *TENCON 2022 – IEEE Region 10 Conference* (pp. 1–4). IEEE. <https://doi.org/10.1109/TENCON55691.2022.9977693>
- Goodfellow, I. J., et al. (2013). Challenges in representation learning: A report on three machine learning contests. *arXiv*. <https://arxiv.org/abs/1307.0414>
- Guerdelli, H., Ferrari, C., Barhoumi, W., Ghazouani, H., & Berretti, S. (2022). Macro- and micro-expressions facial datasets: A survey. *Sensors*, 22(4), 1524. <https://doi.org/10.3390/s22041524>
- Hu, E., et al. (2022). LoRA: Low-rank adaptation of large language models. *arXiv*. <https://arxiv.org/abs/2106.09685>
- Kollias, D., Schulc, A., Hajiyeve, E., & Zafeiriou, S. (2020). *Analysing affective behavior in the first ABAW 2020 competition* (p. 643). IEEE. <https://doi.org/10.1109/FG47880.2020.00126>

- Li, S., & Deng, W. (2022). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(3), 1195–1215. <https://doi.org/10.1109/TAFFC.2020.2981446>
- Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. En *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2584–2593). IEEE. <https://doi.org/10.1109/CVPR.2017.277>
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. En *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 94–101). IEEE. <https://doi.org/10.1109/CVPRW.2010.5543262>
- Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18–31. <https://doi.org/10.1109/TAFFC.2017.2740923>
- Office of Institutional Research. (s. f.). *Race/ethnicity FAQs*. Tufts University. <https://provost.tufts.edu/institutionalresearch/race-ethnicity-faq/>
- Psaroudakis, A., & Kollias, D. (2022). MixAugment & Mixup: Augmentation methods for facial expression recognition. *arXiv*. <https://doi.org/10.48550/arXiv.2205.04442>
- Stable Diffusion AI. (s. f.). *Stable diffusion online – Free AI image generator*. <https://stablediffusion.com>
- Yu, J., Liu, Y., Fan, R., & Sun, G. (2024). MixCut: A data augmentation method for facial expression recognition. *arXiv*. <https://doi.org/10.48550/arXiv.2405.10489>
- Zeng, Y., & Lee, K. (2023). The expressive power of low-rank adaptation. *arXiv*. <https://doi.org/10.48550/arXiv.2310.17513>
- Zhang, Y., Li, Y., Qin, L., Liu, X., & Deng, W. (2023). Leave no stone unturned: Mine extra knowledge for imbalanced facial expression recognition. *arXiv*. <https://doi.org/10.48550/arXiv.2310.19636>