

Reconocimiento de emociones faciales mediante técnicas de aprendizaje profundo aplicado a procesos administrativos de la Universidad Técnica de Manabí

Facial emotion recognition using deep learning techniques applied to administrative processes at the Technical University of Manabí

Catherin Liliana Santana Delgado

Ricardo Ordoñez-Avila



**Investigación
Tecnología e Innovación**



Reconocimiento de emociones faciales mediante técnicas de aprendizaje profundo, aplicado a procesos administrativos de la Universidad Técnica de Manabí

Facial emotion recognition using deep learning techniques applied to administrative processes at the Technical University of Manabí

Catherin Liliana Santana Delgado¹, Ricardo Ordoñez-Avila².

Como citar: Santana Delgado, C.L., & Ordoñez-Avila, R. (2025). Modelado de imágenes faciales con técnicas de aprendizaje profundo en la atención de procesos administrativos de la UTM. *Investigación, Tecnología e Innovación*, 17(24), 94-120. <https://doi.org/10.53591/iti.v17i24.2639>

RESUMEN

Contexto: Las instituciones de educación superior promueven el diseño de soluciones automatizadas para optimizar procesos y servicios; en este estudio se diseñó un prototipo de reconocimiento de emociones faciales aplicado a procesos administrativos de la Universidad Técnica de Manabí (UTM). **Método:** Se utilizaron datos propios en una red neuronal convolucional (CNN) personalizada, optimizada para imágenes en escala de grises de 48×48 píxeles, y se comparó con la arquitectura preentrenada MobileNetV1. El entrenamiento siguió un enfoque reproducible mediante un flujo de preprocessamiento, aumento y validación de seis clases de emociones (felicidad, enojo, miedo, neutral, tristeza y disgusto). **Resultados:** La CNN personalizada alcanzó el mejor rendimiento con una exactitud de 0,88 frente a 0,80 de MobileNetV1. El mejor modelo fue integrado en un prototipo con pruebas en escenarios de matrícula, donde la emoción neutral fue la de mayor frecuencia y confianza. **Conclusiones:** En este contexto, las redes neuronales convolucionales son ratificadas para el reconocimiento de imágenes con enfoque en la gestión de la atención a los usuarios en servicios públicos. En trabajos futuros se pueden mejorar la validación estratificada, la comparación con conjuntos de datos de referencia, así como el establecimiento de criterios éticos para fortalecer la generalización y el uso responsable de datos sensibles.

Palabras clave: Aprendizaje profundo, MobileNetV1, procesos administrativos, reconocimiento de emociones faciales, redes neuronales convolucionales.

ABSTRACT

Context: Higher education institutions promote automated solutions to optimize processes and services; this study designed a facial emotion recognition prototype applied to administrative processes at Universidad Técnica de Manabí (UTM). **Method:** A custom convolutional neural network (CNN) optimized for 48×48 grayscale images was compared against a pretrained MobileNetV1, following a reproducible pipeline of preprocessing, data augmentation, and validation across six emotion classes (happiness, anger, fear, neutral, sadness, and disgust). **Results:** The custom CNN achieved the best performance with an accuracy of 0.88 versus 0.80 for MobileNetV1; the best model was integrated into a prototype tested in enrollment scenarios, where the neutral emotion showed the highest frequency and confidence. **Conclusions:** In this context, convolutional neural networks are confirmed as suitable for image recognition focused on user service management in

¹ Máster en Ciencias de Datos y Máquinas de Aprendizaje, Universidad Técnica de Manabí, Ecuador. Correo electrónico: cathe1097@gmail.com

² Universidad Técnica de Manabí, Ecuador. Correo electrónico: ermenson.ordonez@utm.edu.ec



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

public services; future work should strengthen stratified validation, compare against reference datasets, and establish ethical criteria to support generalization and responsible use of sensitive data.

Keywords: Deep learning, MobileNetV1, administrative processes, facial emotion recognition, convolutional neural networks.

Fecha de recepción: Octubre 9, 2025.

Fecha de aceptación: Noviembre 12, 2025.

INTRODUCCIÓN

La inteligencia artificial, el aprendizaje automático y el aprendizaje profundo han impulsado la automatización de procesos y la transformación digital en instituciones (Mukul & Büyüközkan, 2023; Ocen et al., 2025), mejorando el análisis de datos a escala, la toma de decisiones y la eficiencia operativa (Pise et al., 2022). En procesos administrativos universitarios, la interacción entre la persona y el sistema es constante y exige respuestas oportunas (Akhand et al., 2021). Por lo que captar señales no verbales puede complementar la atención y adaptarla en tiempo real (Díaz-Ramírez & Díaz-Ramírez, 2021; Mukul & Büyüközkan, 2023).

El reconocimiento de emociones faciales utiliza características visuales, estáticas y dinámicas para inferir estados afectivos con fines de interacción humano-computadora (Ocen et al., 2025; Rathour et al., 2022). Su desarrollo reciente se sostiene en arquitecturas de aprendizaje profundo y en conjuntos de datos de referencia que permiten evaluar la generalización de los modelos, dada su estandarización y comprobaciones previas (Pise et al., 2022). Entre los modelos de aprendizaje automático más utilizados se encuentran el perceptrón multicapa (MLP), las redes convolucionales (CNN), las redes neuronales recurrentes de memoria a corto plazo (LSTM) y el aprendizaje por transferencia (DTL).

Estas arquitecturas presentan diferentes ventajas, que van desde la detección visual de patrones en CNN hasta la gestión de secuencias temporales en LSTM o la reutilización de conocimiento preentrenado (Rehman et al., 2025). En escenarios con restricciones de cómputo y resolución, las CNN compactas en escala de grises pueden rendir de forma competitiva frente a líneas base de transferencia entrenadas en RGB y mayor resolución, especialmente cuando se priorizan latencia y robustez en contexto operativo (Tutuianu et al., 2023).

Las CNN se han utilizado para el reconocimiento de emociones faciales en tiempo real. Algunos modelos como AlexNet, CNN FER personalizada y soluciones comerciales como Affdex, han sido probados para reconocer emociones generales como la felicidad, la tristeza, la ira y el miedo (Dhope & Neelagar, 2022; Hussain et al., 2024; Kartali et al., 2018; Pathar et al., 2019; Q. Wang et al., 2023). En el caso de los CNN mejorados con capas adicionales y aumento de datos han mostrado una mayor precisión, alcanzando rendimientos superiores al 75% en el conjunto de datos FER-2013 (Babu et al., 2024).

Así mismo, las CNN han sido combinadas con otras técnicas como los Patrones Binarios Locales, y el Histograma de Gradientes Orientados, para identificar señales faciales sutiles (Hassan et al., 2025; Kartali et al., 2018). Por su parte los modelos híbridos se han establecido mediante la combinación de técnicas como CNN y memoria a corto plazo LSTM, y las CNN con memoria a corto y largo plazo CLDNN (Hassouneh et al., 2020; Saxena et al., 2020; Zhe et al., 2025). Ambas combinaciones han alcanzado reconocimiento con tasas superiores al 90% con el empleo de puntos de referencia faciales, como una variante para modelos específicos.



El objetivo de este trabajo fue diseñar un prototipo funcional de reconocimiento de expresiones faciales para procesos administrativos universitarios, comparando una CNN personalizada (entradas 48×48 en escala de grises) con una línea base MobileNetV1 en seis emociones (enojo, neutral, disgusto, miedo, felicidad, tristeza), dentro de un flujo reproducible de preprocesamiento, aumentación y entrenamiento (Zhu et al., 2024). El alcance es evaluar la pertinencia de una CNN específica al dominio bajo restricciones típicas de operación (tiempo de respuesta y recursos), y en documentar hallazgos de uso en un escenario de matrícula en tiempo real. Se abordan tres preguntas Q1. ¿Cuál de las dos arquitecturas evaluadas (CNN personalizada vs. MobileNetV1) ofrece mejor desempeño para seis emociones en el contexto administrativo universitario? Q2. ¿Qué métricas multiclase son más adecuadas para evaluar el desempeño operativo del sistema? Q3. ¿Cuáles son las arquitecturas de reconocimiento de emociones faciales utilizadas en pruebas en tiempo real?

MATERIALES Y MÉTODOS

Se adoptó un diseño experimental comparativo de aprendizaje profundo para reconocimiento de emociones faciales en la Universidad Técnica de Manabí. Se evaluó una CNN personalizada (48×48 , grises) y una línea base MobileNetV1 (224×224 , RGB) en un flujo reproducible de recolección, preprocesamiento, aumentación, entrenamiento y evaluación multiclase, tanto fuera de línea como en un escenario de uso en tiempo real (Shehu et al., 2025). Ambos modelos se implementaron en Python 3.8 con TensorFlow y Keras, además de bibliotecas auxiliares como NumPy, Pandas, Matplotlib, y Seaborn. El entrenamiento se ejecutó localmente en CPU Intel Core i7, y 16 GB de RAM sin aceleración por GPU, empleando semillas, versiones de librerías y parámetros de ejecución para favorecer la reproducibilidad. El flujo metodológico completo se detalla en la Figura 1.

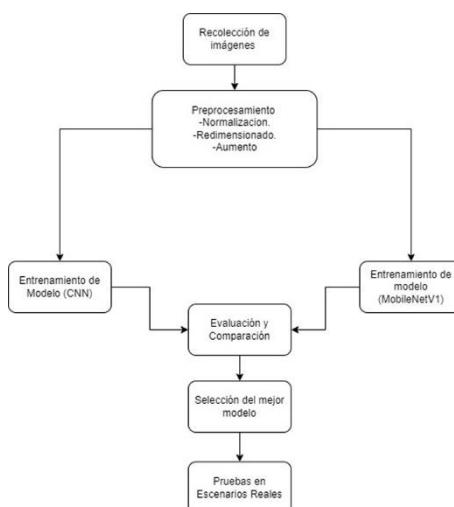


Figura 1. Flujo de trabajo del prototipo de reconocimiento de emociones faciales.

Fuente: Autores.

Recolección de imágenes

Las imágenes se obtuvieron manualmente desde Pinterest para seis emociones (enojo, neutral, disgusto, miedo, felicidad, tristeza), mediante búsquedas en español e inglés. Se aplicaron filtros de calidad (rostro frontal, enfoque, sin occlusiones), duplicación y recorte con detección Haar. El uso de las imágenes es exclusivamente académico y sin redistribución. El entrenamiento se balanceó a 1.000



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

imágenes por emoción mediante aumentación en clases minoritarias (rotación $\pm 30^\circ$ y volteo horizontal), mientras que la validación se mantuvo sin aumentación. Véase la Tabla 1.

Tabla 1. Particionamiento de datos por clase

Clase	Emoción	Entrenamiento	Validación	Total
0	Enojo	1000	55	1055
1	Neutral	1000	55	1055
2	Disgusto	1000	46	1046
3	Miedo	1000	49	1049
4	Felicidad	1000	50	1050
5	Tristeza	1000	45	1045
Total		6000	300	6300

Fuente: Autores.

Preprocesamiento de datos

Todas las imágenes se normalizaron dividiendo los valores de píxel entre 255 para estabilizar la escala de entrada y favorecer la convergencia durante el entrenamiento. Además, se redimensionaron según el modelo: 48×48 en escala de grises para la CNN personalizada y 224×224 en RGB para MobileNetV1, siguiendo referencias de trabajos relacionados (Sajjad et al., 2023).

Para mitigar el sobreajuste, se aplicó aumentación de datos únicamente sobre el conjunto de entrenamiento, usando rotaciones, escalado, zoom y volteo horizontal. Aunque el entrenamiento partió de un balance estricto por clase, se incorporó la ponderación de clases de scikit-learn como salvaguarda ante desbalances efectivos introducidos por la aumentación, preservando contribuciones comparables en la función de pérdida (Alruwais & Zakariah, 2025).

Entrenamiento de modelos

Se entrenaron dos clasificadores multiclas bajo el mismo protocolo experimental. MobileNetV1 se utilizó como backbone preentrenado con una cabeza de clasificación densa y Dropout; recibió imágenes RGB de 224×224 píxeles (Pereira et al., 2024). La CNN personalizada se compuso de bloques Conv2D, BatchNormalization, MaxPooling2D, Dropout y capas densas finales, diseñada para entradas en escala de grises de 48×48 píxeles (Sajjad et al., 2023). En la CNN se aplicó regularización L2 en capas convolucionales y densas para mitigar el sobreajuste. Ambos modelos se optimizaron con Adam ($lr=0,0001$), tamaño de lote 32 y hasta 50 épocas. La Figura 2, muestra las configuraciones de entrenamiento para los dos clasificadores.

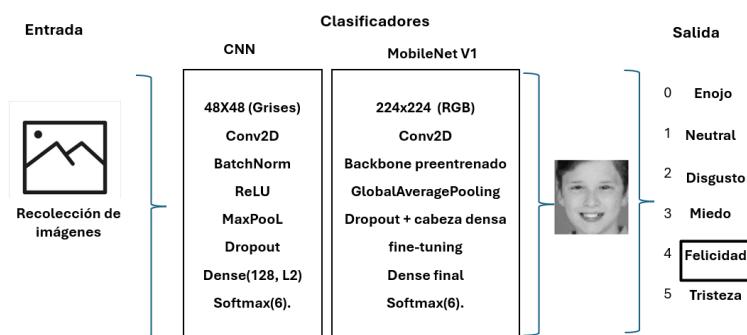


Figura 2. Esquema de clasificadores.

Fuente: Autores.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

Evaluación de modelos

Para validar el rendimiento de los modelos se emplearon métricas estándar en clasificación multiclase, reportadas a nivel global y por clase: precisión, sensibilidad (recuperación), F1 y exactitud, siguiendo prácticas comunes en el área (Appalanaidu & Kumaravelan, 2023; Mohammed, 2024; Rakesh et al., 2022; Sajjad et al., 2023; Srivastava et al., 2025). La precisión mide la proporción de verdaderos positivos sobre todas las predicciones positivas, mientras que la sensibilidad cuantifica la proporción de verdaderos positivos detectados por el modelo. La puntuación F1 resume el equilibrio entre precisión y sensibilidad, y la exactitud refleja la proporción total de aciertos (Opitz, 2024).

$$\text{Precisión} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos positivos}} \quad (1)$$

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{Falsos Negativos}} \quad (2)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{sensibilidad}}{\text{Precision} + \text{sensibilidad}} \quad (3)$$

$$\text{Exactitud} = \frac{\text{prediciones correctas}}{\text{Total de casos}} \quad (4)$$

Selección del mejor modelo

Tras el entrenamiento, la CNN personalizada y MobileNetV1 se evaluaron en el conjunto de validación. La selección del mejor modelo se basó en la comparación de exactitud, precisión, sensibilidad y F1, eligiendo la arquitectura con mayor rendimiento global y por clase para su integración en el prototipo funcional y las pruebas controladas.

Pruebas en escenarios reales

Se realizaron pruebas en procesos administrativos de la Universidad Técnica de Manabí: matrícula y consulta de pagos. Diez participantes completaron tareas guiadas mientras el sistema ejecutaba detección facial, preprocessamiento e inferencia en tiempo real para registrar frecuencia, confianza y duración de cada emoción observada. Estos resultados se utilizaron posteriormente para análisis y validación operativa del prototipo (Appalanaidu & Kumaravelan, 2023; Mohammed, 2024; Rakesh et al., 2022; Sajjad et al., 2023; Srivastava et al., 2025).

RESULTADOS Y DISCUSIÓN

Esta sección presenta, primero, las transformaciones de preprocessamiento y aumentación aplicadas a las imágenes, y luego los resultados cuantitativos de rendimiento de los modelos, destacando su capacidad de clasificación por emoción y global.

El preprocessamiento incluyó detección y recorte de rostro, conversión a escala de grises y redimensionado según la arquitectura; la aumentación incorporó rotación, escalado, zoom y volteo horizontal para robustecer la generalización.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0. Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra.

La Figura 3 resume el flujo visual: (A) etapas de detección, recorte, grises y cambio de tamaño; (B) ejemplos de aumentación para MobileNetV1 (RGB, 224×224×3) y para la CNN (grises, 48×48×1).

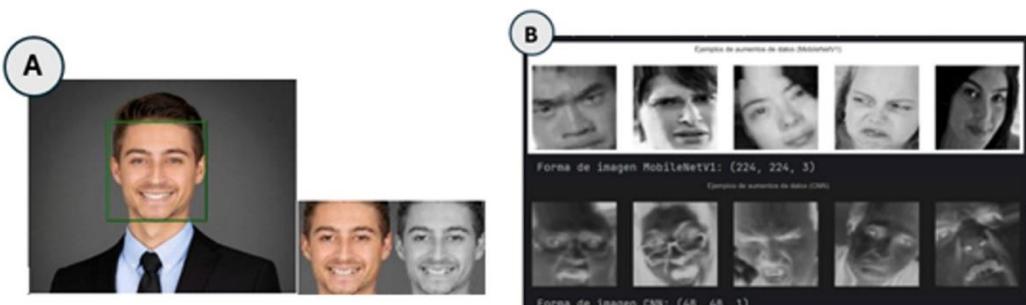


Figura 2. Procesamiento de imágenes faciales y aumento de datos.

Fuente: Autores.

El entrenamiento inicial dio como resultado el 60% de exactitud en validación. Se realizó un ajuste dirigido de hiperparámetros básicos: Adam con lr = 0,0001, batch = 32, L2 y Dropout y se estandarizó el preprocesamiento (normalización x/255 y resize específico por arquitectura), elevando el rendimiento a 0.88 de exactitud para la CNN personalizada y 0.80 para MobileNetV1 bajo las mismas condiciones.

Los resultados comparativos (Figura 4) muestran que la CNN personalizada superó de forma consistente a MobileNetV1 en todas las métricas: exactitud 0,88, precisión 0,89, sensibilidad 0,85 y F1 0,87, frente a exactitud 0,80, precisión 0,81, sensibilidad 0,80 y F1 0,79 de MobileNetV1.

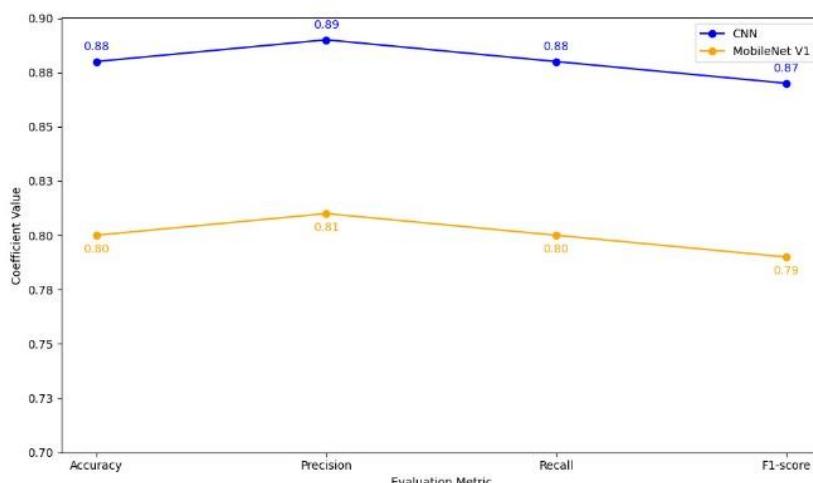


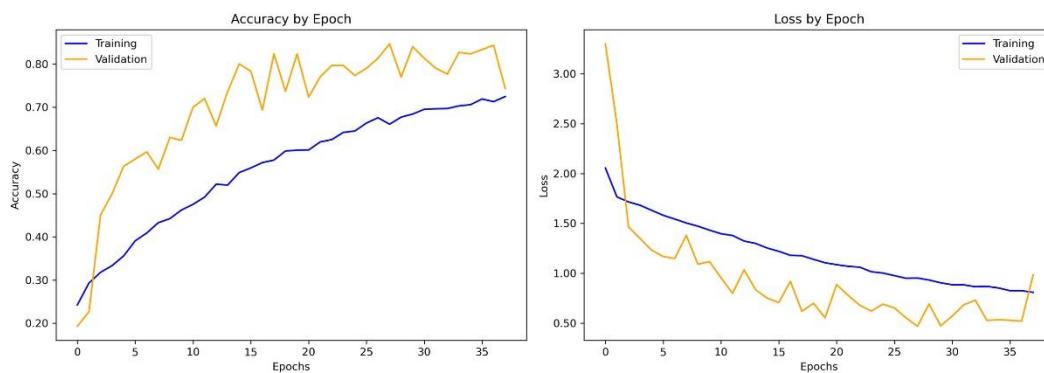
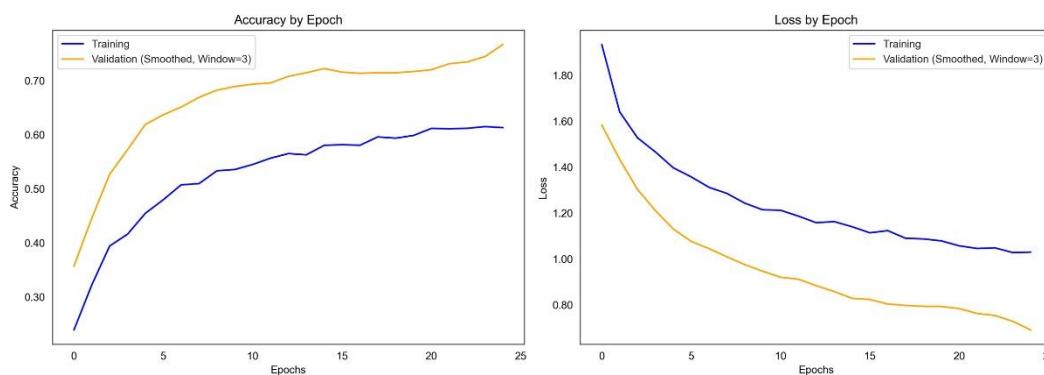
Figura 3. Métricas de evaluación para CNN y MobileNet V1 en reconocimiento de emociones.

Fuente: Autores.

Las curvas de entrenamiento (Figuras 5 y 6) confirman el patrón: la CNN alcanzó 88% de exactitud de validación y pérdida estable cercana a 0,44 con detención en la época 38; MobileNetV1 progresó de 20% a 72% de exactitud de validación y estabilizó la pérdida en 0,76, sin sobreajuste, pero con rendimiento final inferior.

Estos hallazgos respaldan la idoneidad de una CNN específica al dominio y con entradas en grises y baja resolución para escenarios con restricciones operativas.

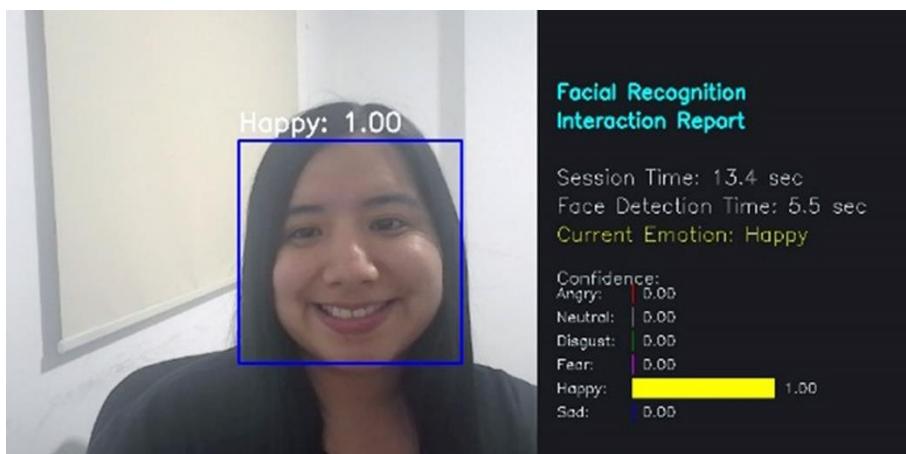


**Figura 4.** Curvas de entrenamiento de la CNN personalizada durante el entrenamiento.**Fuente:** Autores.**Figura 5.** Curvas de entrenamiento del MobileNet V1 personalizado durante el entrenamiento.**Fuente:** Autores.

Para evaluar el comportamiento del prototipo en condiciones cercanas a la operación, se empleó el modelo CNN personalizado por su mejor desempeño en fases previas, y se probó en dos procesos administrativos de la Universidad Técnica de Manabí: matrícula y consultas de valores pendientes. Participaron 10 estudiantes (cinco mujeres y cinco hombres), quienes, durante interacciones guiadas, generaron más de 2.000 imágenes con diversas expresiones; el sistema evaluó automáticamente las emociones detectadas durante las interacciones, lo que permitió observar el rendimiento en un contexto verosímil e identificar oportunidades de mejora.

El algoritmo, implementado en Python, utiliza Haar Cascade para detectar y delimitar rostros en tiempo real; sobre cada rostro detectado aplica el modelo entrenado para inferir una de seis emociones: enojo, neutral, disgusto, miedo, felicidad y tristeza junto con su nivel de confianza (Ver Figura. 7). La interfaz superpone etiquetas y barras de confianza en el video y muestra alertas visuales cuando se identifican emociones negativas, lo que facilita una interpretación inmediata en el contexto administrativo.



**Figura 6.** Interfaz de usuario para pruebas de reconocimiento de emociones.**Fuente:** Autores.

Al finalizar cada sesión, se genera un reporte por participante en formato editable con estadísticas de duración, frecuencia y confianza promedio por emoción, además de gráficos para interpretación; luego, los reportes se consolidan para la visión global del sistema. La Tabla 2 presenta los resultados agregados, incluyendo la duración en segundos, la frecuencia media de ocurrencia por clase con su desviación estándar y la confianza global del modelo para cada emoción.

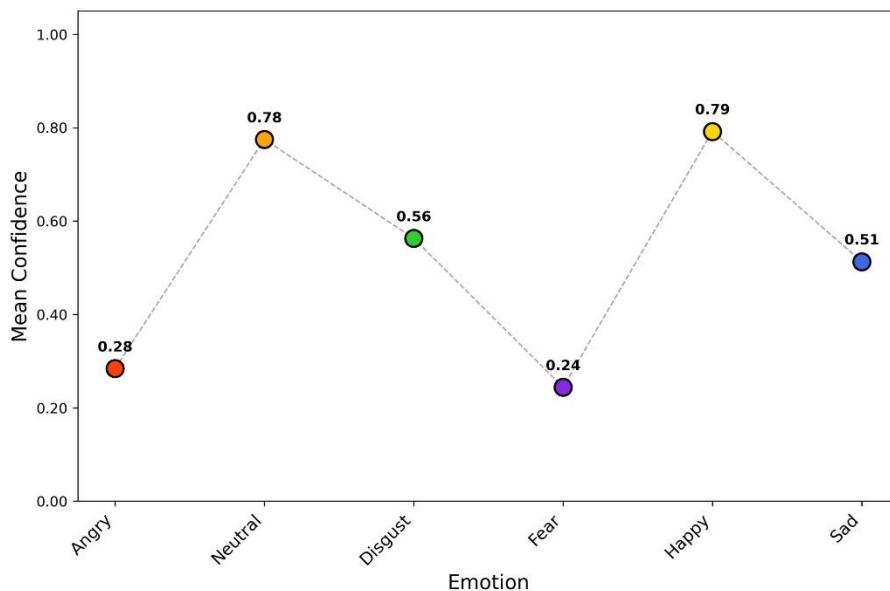
Tabla 2. Estadísticas consolidadas de emociones detectadas durante la evaluación del prototipo.

Emoción	Duración		Frecuencia		Confianza		Participantes
	Media	Sd	Media	Sd	Media	Sd	
Enojo	6.99	14.70	12.6	20.17	0.28	0.21	10
Neutral	172.59	87.53	82.8	34.39	0.78	0.12	10
Disgusto	3.26	4.47	10.6	9.03	0.57	0.07	10
Miedo	2.90	5.24	9.8	17.07	0.24	0.21	10
Felicidad	29.00	25.89	29.1	11.35	0.79	0.17	10
Tristeza	20.63	17.95	64.3	39.93	0.51	0.06	10

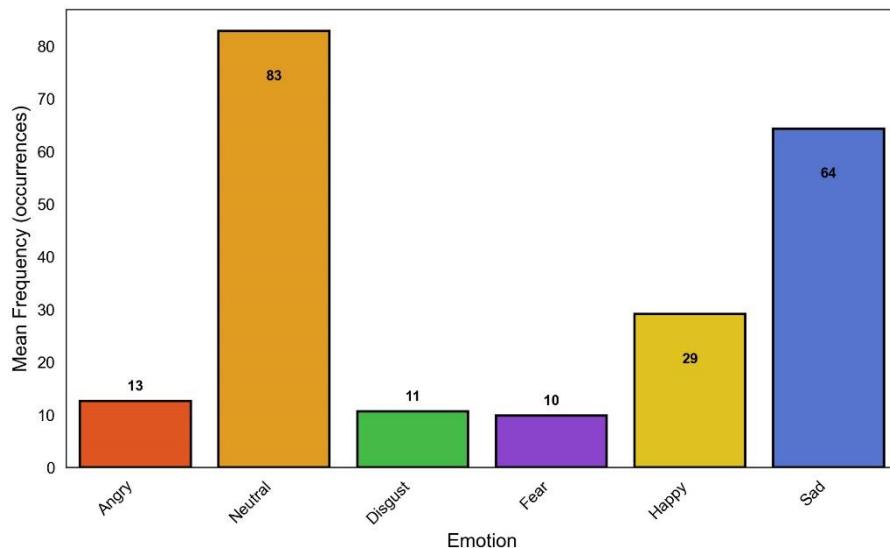
Fuente: Autores.

Para complementar la información cuantitativa, los gráficos siguientes resumen de forma visual los resultados principales sobre confianza, frecuencia y duración promedio de las emociones detectadas durante las pruebas. Se observa mayor confianza del modelo al identificar las emociones “felicidad” (media 0.79) y “neutral” (0.78), lo que sugiere un mejor rendimiento en estados emocionales frecuentes y positivos en el contexto administrativo. En contraste, “enojo” (0.28) y “miedo” (0.24) registran las menores confianzas, lo que indica mayor dificultad o ambigüedad para interpretar expresiones más intensas o menos frecuentes en estas interacciones (Ver Figura. 8).



**Figura 7.** Perfil de confianza promedio según la emoción del modelo durante la prueba.**Fuente:** Autores.

La Figura. 9 muestra la frecuencia media de las emociones detectadas: “neutral” fue la más recurrente, con 83 apariciones, seguida de “tristeza” con 64 y “felicidad” con 29. En cambio, “enojo”, “disgusto” y “miedo” fueron menos frecuentes, con 13, 11 y 10 apariciones, respectivamente.

**Figura 8.** Frecuencia media de emociones detectadas.**Fuente:** Autores.

Finalmente, la Figura. 10 complementa la frecuencia con la duración media de las emociones detectadas. La emoción “neutral” registró la mayor permanencia, con un promedio de 172.59 segundos (aprox. 2 minutos y 53 segundos), lo que sugiere estados sostenidos de neutralidad durante las interacciones. También destacaron “felicidad” con 29.00 segundos (0.48 minutos) y “tristeza” con 20.63 segundos (0.34 minutos), reflejando momentos positivos y reflexivos de duración intermedia.



En el extremo opuesto, “enojo” (7.00 segundos; 0.12 minutos), “disgusto” (3.26 segundos; 0.05 minutos) y “miedo” (2.90 segundos; 0.05 minutos) mostraron las duraciones más breves, coherentes con expresiones más intensas, pero menos sostenidas en este contexto administrativo.

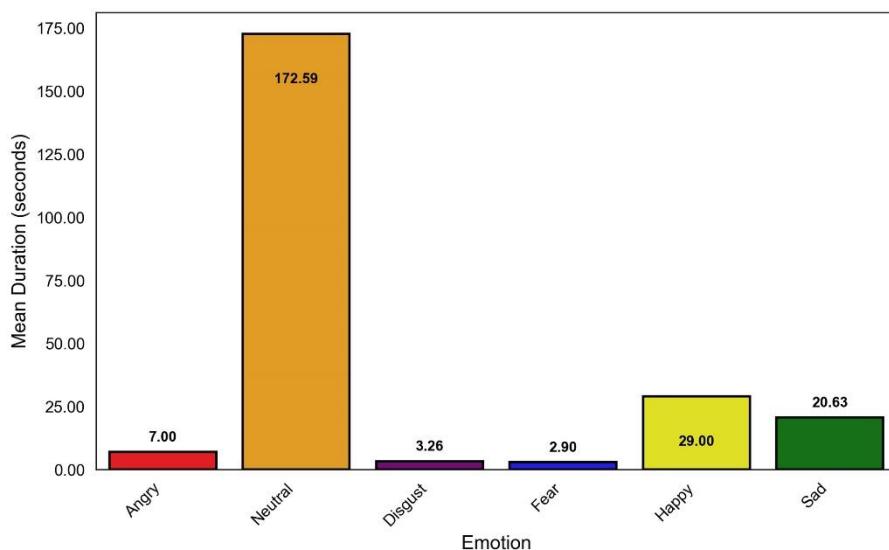


Figura 9. Duración media de emociones detectadas.

Fuente: Autores.

A partir de la evidencia presentada, se precisan a continuación sus implicaciones teóricas y prácticas en el entorno evaluado.

En relación con la pregunta de investigación ¿qué arquitectura resulta más efectiva en entornos administrativos?, la evidencia experimental muestra que la CNN personalizada supera de forma consistente a MobileNetV1 en exactitud y F1-score bajo el mismo protocolo experimental, en línea con revisiones recientes que señalan ventajas de las CNN ajustadas a las características del conjunto de datos y de la tarea concreta, especialmente cuando se controlan entradas, preprocesamiento y régimen de entrenamiento. Adicionalmente, estudios comparativos y aplicaciones reales con variantes eficientes como MobileNet, demuestran que su desempeño depende fuertemente de la configuración y del dominio, lo que explica que, en este escenario con entradas 48×48 en escala de grises, la CNN adaptada haya mostrado mayor adecuación que un backbone preentrenado en RGB de propósito general (Haq et al., 2024).

Las curvas de entrenamiento corroboran estas observaciones: la CNN mostró mejoras sostenidas en validación y convergencia estable de la pérdida, mientras que MobileNetV1 progresó con menor pendiente, en consonancia con resultados donde modelos diseñados para la tarea superan transferencias genéricas cuando el espacio de entrada y las restricciones computacionales se alinean cuidadosamente con el problema (Agung et al., 2024).

En cuanto a las métricas más adecuadas para la evaluación operativa, se emplearon exactitud, precisión, sensibilidad y F1, priorizando macro-F1 y balanced accuracy para mitigar sesgos por desbalance, conforme a recomendaciones que advierten las limitaciones de la exactitud global y promueven análisis por clase en contextos no uniformes (Kaur & Kumar, 2024).

Entre las técnicas que favorecieron la generalización destacan la normalización de imágenes, el redimensionado acorde a cada arquitectura y la aumentación (rotaciones, escalado, zoom, volteo), junto con ponderación de clases y callbacks (parada temprana y guardado del mejor modelo),



prácticas documentadas en implementaciones de reconocimiento en tiempo real y estudios de arquitecturas ligeras (Haq et al., 2024).

En relación con Q3, ¿Cuáles son las arquitecturas de reconocimiento de emociones faciales que han realizado pruebas en tiempo real? Las redes neuronales convolucionales han mostrado buenos niveles de precisión en reconocimiento de emociones en tiempo real, con resultados cercanos al 75%, que pueden acercarse al 95% en modelos personalizados, cuando se ha utilizado el conjunto de datos FER-2013 para el entrenamiento del modelo (Babu et al., 2024; Sanli & Saran, 2024). Otras arquitecturas como VGG16 y AlexNet han alcanzado rendimientos del 100% y del 85% respectivamente, para el reconocimiento facial probados con un humanoide. Mientras que, en trabajos de reconocimiento de emociones en tiempo real, VVG16 ha alcanzado el 73% de precisión, y la arquitectura AlexNet alcanzó el 64% (Dwijayanti et al., 2022). Otras arquitecturas como Xception, la integración de EfficientNet B3 y ResNeT-50, y YOLOv9, han sido reportadas con reconocimiento de emociones, incluso en interacciones en línea, con rendimiento eficiente a partir de videos en tiempo real (Dwijayanti et al., 2022; Poongodai et al., 2025; Singh & Nand, 2024; Y. Wang et al., 2024).

En este estudio se probaron modelos CNN personalizados y MobileNetV1, que han reportado un buen rendimiento en tareas de reconocimiento de emociones faciales. Los patrones emocionales reportados en las pruebas en el contexto administrativo son de prevalencia “neutral”, presencia intermedia de “felicidad” y “tristeza”, y baja ocurrencia de “enojo” y “miedo”, lo que coincide con hallazgos en literatura que vinculan el desempeño y la confiabilidad del sistema a la distribución real de emociones y a la representación de clases raras durante el entrenamiento (Gursesli et al., 2024; Sharif et al., 2022; Sharma et al., 2025; Zhang et al., 2023; Zhu et al., 2024).

Ante las limitaciones de entrenar modelos con datos demográficos de diversidad cultural y geográfica, los resultados se remiten a un caso de estudio en particular sobre datos de una población local, por tanto, en trabajos futuros es conveniente ampliar la validación con conjuntos públicos y protocolos cross-dataset, así como explorar variantes eficientes como la arquitectura MobileNetV2 mejorada, que emplea aprendizaje por transferencia y ajuste fino, para analizar el compromiso entre precisión, latencia y robustez en condiciones operativas (Zhu et al., 2024).

CONCLUSIONES

Este trabajo abordó el reconocimiento de emociones faciales en procesos administrativos mediante un prototipo basado en aprendizaje profundo. En la comparación directa, la CNN personalizada resultó más efectiva que MobileNetV1, alcanzando exactitud de 0,88 frente a 0,80 bajo el mismo protocolo, lo que respalda el diseño de modelos ajustados a datos y entorno de aplicación, especialmente con entradas 48×48 en escala de grises y restricciones de tiempo real. Para lograrlo, se usaron métricas estándar (exactitud, sensibilidad, F1, precisión) y técnicas como aumentación, parada temprana y guardado del mejor modelo, reduciendo sobreajuste y mejorando generalización.

Las pruebas en la Universidad Técnica de Manabí evidencian utilidad práctica del prototipo para mejorar la interacción en trámites administrativos y orientar trabajos futuros. La emoción “neutral” fue la más frecuente y prolongada, consistente con la naturaleza rutinaria de estas interacciones, mientras que la detección de otras emociones en situaciones espontáneas sugiere potencial para personalizar la atención. En conjunto, se aporta un prototipo validado en entorno universitario, estableciendo bases para implementaciones que optimicen servicios mediante comprensión emocional automatizada.



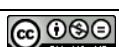
TRABAJOS FUTUROS

Se propone ampliar la evaluación a mayor escala y en diversos contextos administrativos, incorporando conjuntos de datos con balance y cobertura demográfica y cultural para incrementar la precisión e interpretar el impacto práctico (Tutuiaru et al., 2024). La generalización puede potenciarse integrando variables demográficas (p. ej., género, edad) y datos contextuales (p. ej., voz) hacia aplicaciones multimodales; asimismo, conviene aumentar la diversidad de datos para emociones poco frecuentes, explorar arquitecturas preentrenadas mejor alineadas con el dominio y usar aceleración por GPU para mejorar latencia y robustez, promoviendo un sistema más estable, eficiente y equitativo en operación real (FER2013, RAF-DB, AffectNet).

Respecto a consideraciones éticas, en los escenarios de prueba se aplicaron procedimientos de consentimiento informado, minimización de datos, anonimización/pseudonimización y resguardo seguro, alineados con marcos regulatorios y guías institucionales de investigación (Wang et al., 2024).

REFERENCIAS BIBLIOGRÁFICAS

- Agung, E. S., Rifai, A. P. & Wijayanto, T. (2024). Image-based facial emotion recognition using convolutional neural network on emognition dataset. *Scientific Reports*, 14(1), 1–22. <https://doi.org/10.1038/S41598-024-65276-X>;SUBJMETA=117,258,2811,477,631,639,705;KWRD=COMPUTER+SCIENCE,HUMAN+BEHAVIOUR,INFORMATION+TECHNOLOGY
- Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S. & Shimamura, T. (2021). Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics 2021, Vol. 10, Page 1036*, 10(9), 1036. <https://doi.org/10.3390/ELECTRONICS10091036>
- Alruwais, N. & Zakariah, M. (2025). Detecting Student Engagement with Convolution Neural Network and Facial Expression Recognition. *Traitemet Du Signal*, 42(2). <https://doi.org/10.18280/TS.420229>
- Appalanaidu, M. V. & Kumaravelan, G. (2023). Rice plant nutrient deficiency classification using modified MOBILENET convolutional neural network. *International Journal of Modeling, Simulation, and Scientific Computing*, 14(1). <https://doi.org/10.1142/S1793962322430036>;REQUESTEDJOURNAL:JOURNAL:IJMSSC;PAGEGROUP:STRING:PUBLICATION
- Babu, T., Ebin, P. M. & Nair, R. R. (2024). Real-Time Facial Expression Recognition Using Deep Learning for Enhanced Human-Machine Interactions. *2024 IEEE Flagship International BIT Conference: Next Generation Applications in Green Energy Technology, BITCON 2024*. <https://doi.org/10.1109/BITCON63716.2024.10984634>
- Dhope, P. & Neelagar, M. B. (2022). Real-Time Emotion Recognition from Facial Expressions using Artificial Intelligence. *2022 2nd International Conference on Artificial Intelligence and Signal Processing, AISIP 2022*. <https://doi.org/10.1109/AISIP53593.2022.9760654>
- Díaz-Ramírez, J. & Díaz-Ramírez, J. (2021). Aprendizaje Automático y Aprendizaje Profundo. *Ingeniare. Revista Chilena de Ingeniería*, 29(2), 180–181. <https://doi.org/10.4067/S0718-33052021000200180>
- Dwijayanti, S., Iqbal, M. & Suprapto, B. Y. (2022). Real-Time Implementation of Face Recognition and Emotion Recognition in a Humanoid Robot Using a Convolutional Neural Network. *IEEE*



Access, 10, 89876–89886. <https://doi.org/10.1109/ACCESS.2022.3200762>

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>

Haq, H. B. U., Akram, W., Irshad, M. N., Kosar, A. & Abid, M. (2024). Enhanced Real-Time Facial Expression Recognition Using Deep Learning. *Acadlore Transactions on AI and Machine Learning*, 3(1), 24–35. <https://doi.org/10.56578/ATAIML030103>

Hassan, E., Bhatnagar, R., El-Hafeez, T. A. & Shams, M. Y. (2025). Detection of Suicide and Depression for Early Intervention and Initiative-taking Mental Healthcare. *Proceedings of IEEE International Conference on Signal Processing, Computing and Control*, 99–104. <https://doi.org/10.1109/ISPCC66872.2025.11039547>

Hassouneh, A., Mutawa, A. M. & Murugappan, M. (2020). Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, 20, 100372. <https://doi.org/10.1016/J.IMU.2020.100372>

Hussain, S. A., Reddy, N. Y., Srivardhan, J., Sharma, A., Sharma, S. & Gochhait, S. (2024). Automated Emotion Recognition from Facial Expressions using Convolutional Neural Network. *2024 5th International Conference on Data Analytics for Business and Industry, ICDABI 2024*, 139–143. <https://doi.org/10.1109/ICDABI63787.2024.10800158>

Kartali, A., Roglic, M., Barjaktarovic, M., Duric-Jovicic, M. & Jankovic, M. M. (2018). Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches. *2018 14th Symposium on Neural Networks and Applications, NEUREL 2018*. <https://doi.org/10.1109/NEUREL.2018.8587011>

Kaur, M. & Kumar, M. (2024). Facial emotion recognition: A comprehensive review. *Expert Systems*, 41(10), e13670. <https://doi.org/10.1111/EXSY.13670>; WGROUP:STRING:PUBLICATION

Mohammed, O. N. (2024). Enhancing Pulmonary Disease Classification in Diseases: A Comparative Study of CNN and Optimized MobileNet Architectures. *Journal of Robotics and Control (JRC)*, 5(2), 427–440. <https://doi.org/10.18196/JRC.V5I2.21422>

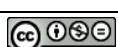
Mukul, E. & Büyüközkan, G. (2023). Digital transformation in education: A systematic review of education 4.0. *Technological Forecasting and Social Change*, 194, 122664. <https://doi.org/10.1016/J.TECHFORE.2023.122664>

Ocen, S., Elasu, J., Aarakit, S. M. & Olupot, C. (2025a). Artificial intelligence in higher education institutions: review of innovations, opportunities and challenges. *Frontiers in Education*, 10, 1530247. [https://doi.org/10.3389/FEDUC.2025.1530247/BIBTEX](https://doi.org/10.3389/FEDUC.2025.1530247)

Ocen, S., Elasu, J., Aarakit, S. M. & Olupot, C. (2025b). Artificial intelligence in higher education institutions: review of innovations, opportunities and challenges. *Frontiers in Education*, 10, 1530247. <https://doi.org/10.3389/FEDUC.2025.1530247/BIBTEX>

Opitz, J. (2024). A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice. *Transactions of the Association for Computational Linguistics*, 12, 820–836. https://doi.org/10.1162/tacl_a_00675

Pathar, R., Adivarekar, A., Mishra, A. & Deshmukh, A. (2019). Human Emotion Recognition using



Convolutional Neural Network in Real Time. *Proceedings of 1st International Conference on Innovations in Information and Communication Technology, ICIICT 2019.* <https://doi.org/10.1109/ICIICT1.2019.8741491>

Pereira, R., Mendes, C., Ribeiro, J., Ribeiro, R., Miragaia, R., Rodrigues, N., Costa, N. & Pereira, A. (2024). Systematic Review of Emotion Detection with Computer Vision and Deep Learning. *Sensors*, 24(11). <https://doi.org/10.3390/S24113484>,

Pise, A. A., Alqahtani, M. A., Verma, P., Purushothama, K., Karras, D. A., Prathibha, S. & Halifa, A. (2022). Methods for Facial Expression Recognition with Applications in Challenging Situations. *Computational Intelligence and Neuroscience*, 2022, 9261438. <https://doi.org/10.1155/2022/9261438>

Poongodai, A., Bhavitha, V., Hemanth, D., Jahnavi, G. & Huzma, S. (2025). Human Emotion Detection Through Real-Time Facial Expressions Using Deep Learning. *Proceedings of the 6th International Conference on Inventive Research in Computing Applications, ICIRCA 2025*, 1298–1303. <https://doi.org/10.1109/ICIRCA65293.2025.11089747>

Rakesh, K. R., Namita, G. R. & Kulkarni, R. (2022). Image Recognition, Classification and Analysis Using Convolutional Neural Networks. *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*. <https://doi.org/10.1109/ICEEICT53079.2022.9768474>

Rathour, N., Singh, R., Gehlot, A., Vaseem Akram, S., Kumar Thakur, A. & Kumar, A. (2022). The decadal perspective of facial emotion processing and Recognition: A survey. *Displays*, 75, 102330. <https://doi.org/10.1016/J.DISPLA.2022.102330>

Rehman, A., Mujahid, M., Elyassih, A., AlGhofaily, B. & Bahaj, S. A. O. (2025). Comprehensive Review and Analysis on Facial Emotion Recognition: Performance Insights into Deep and Traditional Learning with Current Updates and Challenges. *Computers, Materials and Continua*, 82(1), 41–72. <https://doi.org/10.32604/CMC.2024.058036>

Sajjad, M., Ullah, F. U. M., Ullah, M., Christodoulou, G., Alaya Cheikh, F., Hijji, M., Muhammad, K. & Rodrigues, J. J. P. C. (2023). A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. *Alexandria Engineering Journal*, 68, 817–840. <https://doi.org/10.1016/J.AEJ.2023.01.017>

Sanli, A. T. & Saran, M. (2024). Application of a Voting-Based Ensemble Method for Recognizing Seven Basic Emotions in Real-Time Webcam Video Images. *2024 IEEE 9th International Conference for Convergence in Technology, I2CT 2024*. <https://doi.org/10.1109/I2CT61223.2024.10543506>

Saxena, S., Tripathi, S. & Tsb, S. (2020). Deep Robot-Human Interaction with Facial Emotion Recognition Using Gated Recurrent Units & Robotic Process Automation. *Frontiers in Artificial Intelligence and Applications*, 332, 115–126. <https://doi.org/10.3233/FAIA200773>

Sharif, M. S., Afolabi, M. O., Zorto, A. & Elmedany, W. (2022). Enhancement Techniques for Improving Facial Recognition Performance in Convolutional Neural Networks. *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2022*, 494–499. <https://doi.org/10.1109/3ICT56508.2022.9990811>

Sharma, K. P., Nagpal, T., Raja Praveen, K. N., Yadav, A., Tham, J., Bhosle, N., Palla, S. R. & Chauhan, M. (2025). Evaluating MobileNetV2 Architecture for Resource-Efficient Facial Emotion Recognition. *National Academy Science Letters* 2025, 1–5.



<https://doi.org/10.1007/S40009-025-01671-W>

- Shehu, H. A., Browne, W. N. & Eisenbarth, H. (2025). Emotion categorization from facial expressions: A review of datasets, methods, and research directions. *Neurocomputing*, 624, 129367. <https://doi.org/10.1016/J.NEUROCOMPUTING.2025.129367>
- Singh, E. & Nand, P. (2024). Efficient Multi-Class Facial Emotion Recognition using YOLOv9: A Deep Learning Approach for Real-Time Applications. *International Journal of Performativity Engineering*, 20(9), 581. <https://doi.org/10.23940/IJPE.24.09.P6.581590>
- Srivastava, S., Ali, P., Yadav, S., Khan, A. & Pandey, P. (2025). Sign Ease: AI-Based Gesture Recognition. *2025 International Conference on Pervasive Computational Technologies, ICPCT 2025*, 645–649. <https://doi.org/10.1109/ICPCT64145.2025.10940959>
- Tutuijanu, G. I., Liu, Y., Alamäki, A. & Kauttonen, J. (2023). *Benchmarking Deep Facial Expression Recognition: An Extensive Protocol with Balanced Dataset in the Wild*. [https://arxiv.org/pdf/2311.02910](https://arxiv.org/pdf/2311.02910.pdf)
- Wang, Q., Yan, X. & Wang, Y. (2023). Research on deep learning-based facial expression recognition and its application in online learning state monitoring. *Https://Doi.Org/10.1117/12.3005824*, 12799, 459–464. <https://doi.org/10.1117/12.3005824>
- Wang, Y., Jiang, P., Wang, C. & Hachisuka, S. (2024). Facial Recognition: Decoding Emotions in Online Collaboration. *2024 12th International Conference on Information and Education Technology, ICIEET 2024*, 333–337. <https://doi.org/10.1109/ICIEET60671.2024.10542815>
- Zhang, Z., Fort, J. M. & Giménez Mateu, L. (2023). Facial expression recognition in virtual reality environments: challenges and opportunities. *Frontiers in Psychology*, 14, 1280136. [https://doi.org/10.3389/FPSYG.2023.1280136/BIBTEX](https://doi.org/10.3389/FPSYG.2023.1280136)
- Zhe, J., Ouyang, L., Pan, S.-T. & Wu, H.-J. (2025). FPGA Chip Design of Sensors for Emotion Detection Based on Consecutive Facial Images by Combining CNN and LSTM. *Electronics 2025, Vol. 14, Page 3250*, 14(16), 3250. <https://doi.org/10.3390/ELECTRONICS14163250>
- 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>

