



TECNOLÓGICO  
NACIONAL DE MÉXICO



VI CONGRESO  
Nacional de Investigación en  
Ciencia e Innovación de  
Tecnologías Productivas

## ESTUDIO Y ENTRENAMIENTO DE UN MODELO DE APRENDIZAJE PROFUNDO PARA EL RECONOCIMIENTO DE EMOCIONES HUMANAS.

José Crispín Hernández Hernández<sup>1</sup>; María del Rocío Ojeda López<sup>1</sup>; Eduardo Sánchez Lucero<sup>1</sup>; Edmundo Bonilla Huerta<sup>1</sup>;

José Federico Ramírez Cruz<sup>1</sup>

### Resumen

En los últimos años, el reconocimiento de emociones faciales (FER, Face Expression Recognition) incrementó su interés de estudio en diversas disciplinas, por citar: la industria, la medicina, la psicología cognitiva, y en especial, en la computación. A pesar de existir nuevas técnicas de aprendizaje automático y de aprendizaje profundo que abordan dicho FER, el medir las emociones es una tarea difícil en ambientes de la vida real, en expresiones débilmente posadas, y los atributos faciales. En este trabajo se presenta el uso de la biblioteca de aprendizaje profundo fastai, la cual cuenta con una biblioteca de visión por computadora optimizada para GPU. Específicamente se utilizó una red de aprendizaje profundo ya entrenada, aplicando la técnica de aprendizaje por transferencia en los modelos: ResNet50, ResNet34, y ResNet18; que se caracterizan por una mejor comprensión de las características discriminativas de baja dimensión a partir de patrones faciales complejos de alta dimensión para reconocimiento automático de emociones faciales. Para mitigar el problema de sobreajuste debido a la disponibilidad inadecuada de datos de entrenamiento, se utilizaron 2 bases de datos públicas de fotografías de rostros etiquetadas con las 6 etiquetas: ira, asco, miedo, felicidad, tristeza y sorpresa; más la neutral (los conjuntos de datos Kaggle's Facial Expression Recognition Challenge y CK+). Este trabajo abarca dos objetivos principales. El primero es debatir el escenario actual de los enfoques FER y el segundo es presentar algunas reflexiones acerca de los obstáculos y las perspectivas sobre las direcciones futuras del reconocimiento automático de emociones faciales.

Palabras clave: Fastai, GPU, Aprendizaje por transferencia, Reconocimiento de emociones faciales, Aprendizaje automático, Aprendizaje profundo

### INTRODUCCIÓN

El reconocimiento automático de emociones faciales que alguna vez se consideró de ciencia ficción, hoy en día, se está convirtiendo en una parte esencial de la vida cotidiana de las personas. En lo general, se utiliza para comprender los comportamientos emocionales humanos. Por tal motivo, en diferentes campos: industria, seguridad, medicina (Deramgozin et al., 2022), psicología cognitiva, y en especial en la computación (Goodfellow et al., 2013), se incrementó su interés de estudio. Por ejemplo, en la computación se estudia desde diferentes áreas de estudio: visión por computadora, robótica (Gokul Prasad et al., 2023), sistemas de gestión de aprendizaje (Zhang et al., 2021), computación afectiva, ciencia de datos, aprendizaje máquina, minería de datos. Por lo que el reconocimiento de emociones faciales reconocimiento (FER) es un campo dinámico y en constante evolución al estudiarse desde diferentes campos de estudio, y áreas de conocimiento. A continuación se describen 5 áreas de interés de los sistemas FER. Las redes neuronales convolucionales profundas (CNN) se han convertido en el eje de los sistemas ER, dado que pueden aprender automáticamente características relevantes a partir de datos de imágenes sin procesar. Además de las expresiones faciales, se pueden combinar con otras entradas modales como: audios, textos, voz, y señales

<sup>1</sup> <sup>1</sup>Tecnológico Nacional de México / Instituto Tecnológico de Apizaco. [crispin.hh@apizaco.tecnm.mx](mailto:crispin.hh@apizaco.tecnm.mx), [rocio.ol@apizaco.tecnm.mx](mailto:rocio.ol@apizaco.tecnm.mx),

[eduardo.sl@apizaco.tecnm.mx](mailto:eduardo.sl@apizaco.tecnm.mx), [edmundo.bh@apizaco.tecnm.mx](mailto:edmundo.bh@apizaco.tecnm.mx), [federico.rc@apizaco.tecnm.mx](mailto:federico.rc@apizaco.tecnm.mx)

Reserva de Derechos al Uso Exclusivo No. 04-2023-091910490900-102, ISSN: En trámite. Año 1, No. 1, Diciembre 2023- Marzo 2024.

Fecha de Recepción: 03/12/2023| Fecha de Aceptación: 11/12/2023



biométricas (Sistemas de entradas multimodal), para mejorar los sistemas FER. Indudablemente, las condiciones en tiempo real y escenarios naturales sin restricciones hacen difícil el uso adecuado de estos sistemas FER por lo que se ha considerado útil la detección de expresiones faciales en la naturaleza, que incluyen imágenes con oclusiones y cambios de pose. También se ha aplicado el aprendizaje por transferencia para el aprovechamiento de modelos preentrenados al aplicarlo al problema de reconocimiento de emociones faciales y su perfeccionamiento de estos. Por último, se han estudiado las consideraciones étnicas, dado que juegan un rol importante en el desarrollo de los sistemas FER.

Algunos trabajos recientes y relevantes fueron analizados a profundidad. Dado su aportación de conocimiento para alcanzar los objetivos de esta investigación se resumen a continuación.

(Depuru et al., 2022) proponen el uso de la Inteligencia artificial para reconocer emociones faciales utilizando las diferentes capas de una red convolucional, CNN. Se examina a fondo el reconocedor de expresiones faciales (FER) con aprendizaje profundo, incluyendo conjuntos de datos y métodos que clarifican sobre estas dificultades subyacentes. Es decir, se implementa el esquema FER, que incluye información de fondo pertinente, para buscar directrices en cada nivel de la red.

(Mukhopadhyay et al., 2021) plantean un nuevo enfoque para detectar expresiones faciales básicas a partir de imágenes texturales utilizando el modelo de una red neuronal convolucional con aprendizaje profundo. En su propuesta se generan imágenes texturales de rostros humanos utilizando el método de patrón binario local (LBP) y se entrena el modelo CNN con las imágenes LBP.

(Dwijayanti et al., 2022) en su trabajo proponen una combinación de los reconocimientos: facial y emocional para su aplicación en tiempo real en un robot humanoide. En concreto, se desarrollan simultáneamente sistemas de reconocimiento de rostros y emociones utilizando arquitecturas de redes neuronales convolucionales AlexNet y VGG16. Demuestran resultados competitivos con otros trabajos del estado del arte.

(Karnati et al., 2023) muestran un estudio exhaustivo de los métodos basados en Aprendizaje profundo más significativos que han contribuido notablemente al campo de los Sistemas de reconocimiento de emociones faciales. También describen sistemáticamente distintos componentes de los métodos, como el preprocesamiento, la extracción de características y la clasificación de las emociones faciales. Además, se analizan los enfoques discutidos para comparar su rendimiento junto con sus ventajas y limitaciones.

(R et al., 2022) Discutieron que normalmente, las redes neuronales convolucionales estándar tienen algoritmos de aprendizaje sencillos con capas de extracción de características finitas para extraer información. su inconveniente del enfoque anterior era que sólo validaban la vista frontal de las fotos aunque la imagen se obtuviera desde diferentes ángulos. Este trabajo de investigación propone utilizar una Red Neuronal Convolucional profunda junto con un modelo preentrenado DenseNet-169 como red troncal para el reconocimiento de emociones faciales.

## OBJETIVOS

Objetivo general: desarrollar un modelo para el reconocimiento de emociones faciales rápido y confiable.

Objetivos particulares:



- Contar con un conjunto de datos de imágenes de emociones faciales
- Estudiar diferentes alternativas de algoritmos para el reconocimiento de emociones faciales
- Desarrollar y evaluar diferentes modelos para el reconocimiento de emociones faciales

### MATERIALES Y MÉTODOS

En este trabajo se utilizaron dos conjuntos de datos de imágenes de rostros ampliamente utilizados para el reconocimiento de expresiones faciales. El primer conjunto de datos, preparado por [] para el concurso de reconocimiento de expresiones faciales Kaggle's 2013, consta de imágenes de rostros en escala de grises de 48x48 píxeles. Las caras se han registrado automáticamente para que estén más o menos centradas y ocupen aproximadamente el mismo espacio en cada imagen. Se utilizaron para categorizar cada rostro en función de la emoción mostrada en la expresión facial en una de siete categorías (0=ira, 1=asco, 2=miedo, 3=alegría, 4=tristeza, 5=sorpresa y 6=neutra). Las imágenes se encuentran separados en tres grupos: el conjunto de entrenamiento: consta de 28.709 imágenes, el conjunto de pruebas público utilizado para la clasificación consta de 3.589 imágenes, y el conjunto de pruebas final, que se utilizó para determinar el ganador del concurso, que consta de otras 3.589 imágenes. El segundo conjunto de datos propuesto por [] consiste de 593 secuencias de video de 123 diferentes tipos de personas entre 18 y 50 años. De 327 secuencias de video, se etiquetaron con al menos, una de las siguientes emociones: (ira, desprecio, asco, miedo, alegría, tristeza y sorpresa) y se obtuvieron 3589 imágenes para entrenamiento.

Para el uso de una red neuronal convolucional con aprendizaje por transferencia existen modelos preentrenados. La Fig. 1 muestra algunos de los modelos más ampliamente utilizados (Krizhevsky et al., 2012), (Szegedy et al., 2015), (Simonyan & Zisserman, 2015). En este trabajo se utilizaron los modelos Resnet: ResNet-18, ResNet-34, y ResNet-50 por su tiempo relativamente corto para dar una predicción.

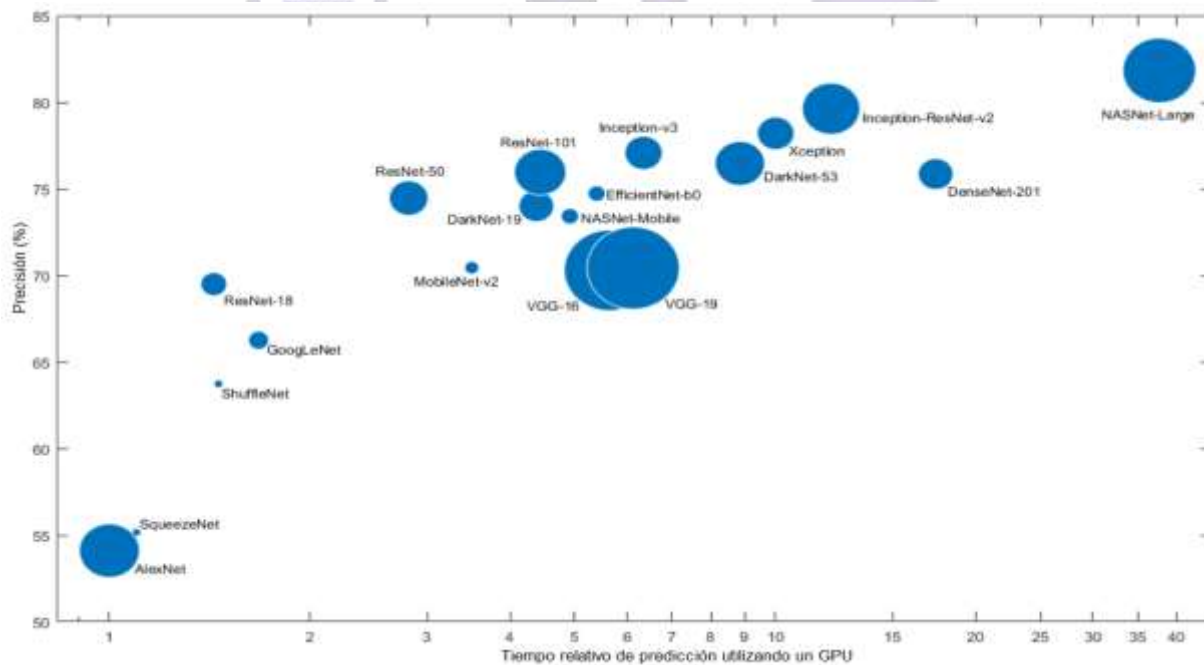


Figura 1. Modelos preentrenados utilizados en una red neuronal convolucional con aprendizaje por transferencia.

Tabla 1. Modelos ResNet que se utilizaron.



Red neuronal	Profundidad	Tamaño	Tamaño de entrada de imagen	GFLOPs
ResNet18	18	44 MB	224 por 224	233 GFLOPs
ResNet34	34	83 MB	224 por 224	470 GFLOPs
ResNet50	50	96 MB	224 por 224	527 GFLOPs

Las redes neuronales residuales, o ResNets son un tipo de modelo de aprendizaje profundo que permite un mejor rendimiento en el reconocimiento de emociones faciales en imágenes (He et al., 2016). La idea clave de las ResNets es el uso de las llamadas "conexiones de salto" o "atajos" que permiten que la información eluda una o más capas de la red. Esto facilita que la red aprenda y mejore su rendimiento con el tiempo. Son ampliamente usadas porque pueden ayudar a reducir el problema del "gradiente disipado", que se produce cuando los gradientes de error de un modelo de aprendizaje profundo son tan pequeños que el modelo tiene problemas para aprender. Esto hace que las ResNets sean especialmente útiles para entrenar redes muy profundas, lo que puede ser difícil de hacer con otros tipos de modelos. Otras de sus elementos se resumen en la tabla 2.

Tabla 2. Elementos de una red neuronal convolucional ResNet.



Nombre de capa	Tamaño de salida	ResNet- 18 capas	ResNet - 34 capas	ResNet - 50 capas
conv 1	112 x 112	7 x 7, 64, atajo de 2 capas		
conv 2, x	56 x 56	3 x 3 agrupación máxima, atajo de 2 capas		
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv 3, x	28 x 28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv 4, x	14 x 14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv 5, x	7 x 7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1 x 1	agrupación promedio, conexión a 1000N , función softmax		
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$

La arquitectura ResNet sigue dos reglas de diseño básicas. Primero, la cantidad de filtros en cada capa es la misma según el tamaño del mapa de características de salida. En segundo lugar, si el tamaño del mapa de características se reduce a la mitad, tiene el doble de filtros para mantener la complejidad temporal de cada capa. La red neuronal ResNet - 50 capas utiliza un diseño de cuello de botella para el bloque de construcción. Un bloque residual de cuello de botella utiliza convoluciones 1x1, conocidas como "cuello de botella", lo que reduce el número de parámetros y multiplicaciones de matrices. Esto permite un entrenamiento mucho más rápido de cada capa. Utiliza una pila de tres capas en lugar de dos capas. Sus elementos son: una convolución de núcleos de 7x7 junto con otros 64 núcleos con un atajo de 2 capas; una capa de agrupación máxima con un atajo de 2 capas; 9 capas más: convolución de núcleo de 3x3, 64, otra con 1x1, 64 núcleos y una tercera con 1x1, 256 núcleos. Estas 3 capas se repiten 3 veces; 12 capas más con 1x1, 128 núcleos, 3x3, 128 núcleos y 1x1, 512 núcleos, repetidas 4 veces; 18 capas más con 1x1, 256 núcleos, y 2 núcleos 3x3, 256 y 1x1, 1024, repetidos 6 veces; 9 capas más con 1x1, 512 núcleos, 3x3, 512 núcleos y 1x1, 2048 núcleos, repetidos 3 veces y una última capa de agrupación promedio, seguida de una capa conectada con 1000 nodos, utilizando la función de activación Softmax.

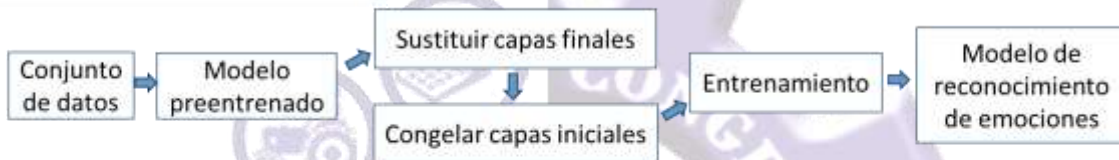
A continuación se describen los paquetes utilizados para el desarrollo del modelo. Se utilizó Google Colab por el requerimiento de equipo para procesar la gran cantidad de imágenes. Es un producto gratuito de Google, que permite



ejecutar código Python en formato de servicio de cuaderno Jupiter alojado. No requiere configuración y proporciona acceso a las unidades de procesamiento gráfico, especialmente adecuado para el aprendizaje automático, y el análisis de datos. Igualmente, se utilizó el paquete de fastai (Howard, 2017/2023). Es una biblioteca de aprendizaje profundo que tiene una API y componentes de bajo y alto nivel para visión por computadora. Esto brinda la posibilidad de utilizar esta API y componentes de última generación, que se pueden mezclar y combinar para construir nuevos enfoques. Por obviedad, se utilizaron los paquetes: Numpy, Pandas, Matplotlib, PIL, y Os. Los paquetes anteriores ayudaron a realizar el manejo de datos, operaciones matemáticas, gráficas de resultados, procesamiento de imágenes y operaciones en el sistema operativo.

La fig. 2 muestra el diagrama a bloques del modelo que se utilizó.

### Fase de entrenamiento



### Fase de predicción y evaluación del modelo

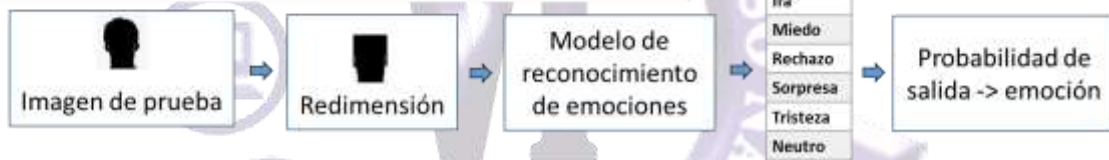


Figura 2. Modelo utilizado para el reconocimiento de emociones faciales.

Dentro de la fase de aprendizaje, se utilizó cualquiera de los dos conjuntos de datos previamente descritos. Posteriormente, se usó de manera separada cualquiera de los tres modelos ResNet como modelo preentrenado, descritos anteriormente.

## RESULTADOS Y DISCUSIÓN

La Fig. 3 muestra los resultados obtenidos en la fase de entrenamiento con una red convolucional Resnet-18. Para la fase entrenamiento, se ejecutó por 20 épocas y se ajustó el modelo usando la tasa de aprendizaje óptima con los valores [1e-5, 1e-2]. En seguida, para volver a entrenar la red preentrenada para clasificar nuevas imágenes, se sustituyó las últimas capas totalmente conectadas, se congelaron las capas iniciales, para adaptarlo al nuevo conjunto de datos. En la Fig. 3, se ve la matriz de confusión obtenida. La emoción con la que se obtiene mejores resultados en la clasificación es la de alegría (Happy). En la parte baja de la Fig. 3, se muestra los casos con las imágenes más difíciles de clasificar.

```
interp.plot_confusion_matrix()
```

Confusion matrix

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	306	3	20	25	51	46	16
Disgust	24	21	1	3	0	7	0
Fear	66	3	171	21	66	121	48
Happy	21	1	10	780	51	18	14
Neutral	48	2	22	55	390	85	5
Sad	80	2	49	32	134	342	14
Surprise	17	0	27	18	15	9	329

```
# Counting the values of losses
interp.top_losses(largest=True)
```

```
(TensorBase[[8.5432e+00, 8.1342e+00, 7.9010e+00, ..., 1.1169e-04,
6.8901e-05, 5.6384e-05]],
TensorBase[[3900, 581, 524, ..., 3447, 162, 828]])
```

```
# Comparing the predicted and actual output
interp.plot_top_losses(20,figsize = (20,15))
```



Figura 3. Resultados obtenidos después de reusar una red ResNet-18 con nuevas imágenes.



En la fig. 4, se muestran los resultados obtenidos en la fase de predicción y evaluación del modelo desarrollado. En la predicción, seis de las nueve imágenes son predichas correctamente dado la emoción de su categoría respectiva. Para la emoción de neutral que fue predicha como tristeza, los umbrales entre ambas emociones son muy cercanos. En cuanto a la evaluación del modelo, se obtuvo una tasa de 67.51%

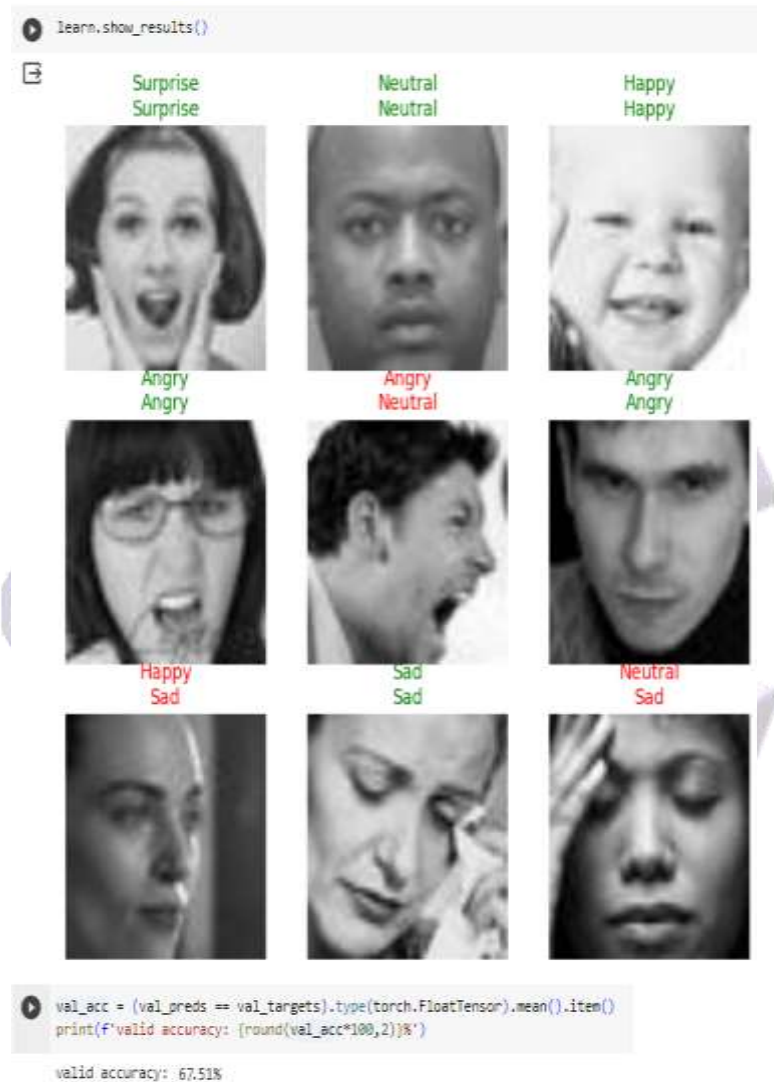


Figura 4. Resultados obtenidos de la predicción y evaluación del modelo desarrollado.

### CONCLUSIONES

Al término de este trabajo, se llegó a las siguientes conclusiones. Los modelos preentrenados hacen posible el reconocimiento de emociones faciales, partiendo de la extracción y selección de características ocultas de las imágenes ya aprendidas. También, la reutilización de un modelo previamente entrenado para el propósito de esta investigación comenzó eliminando el clasificador original, luego se agregó un nuevo clasificador que se ajustara a nuestro propósito y, finalmente, se ajustó el modelo, para finalmente predecir y evaluar el modelo obtenido con Reserva de Derechos al Uso Exclusivo No. 04-2023-091910490900-102, ISSN: En trámite. Año 1, No. 1, Diciembre 2023- Marzo 2024.





imágenes propias. Los resultados obtenidos y mostrados en este artículo, aún se pueden mejorar a medida que las imágenes pasan del entorno controlado del laboratorio a las circunstancias reales. Por último, desafortunadamente los recursos de Colab no están garantizados ni son ilimitados y, en ocasiones, los límites de uso fluctúan. Esto afectó negativamente en este trabajo, ya que solo se pudieron usar los modelos de red Resnet-18, ResNet34 y Resnet-50, debido a los recursos que consumen estos modelos preentrenados y que una cuenta estándar de Colab permite. Se buscará otras alternativas para probar otros modelos preentrenados existentes, o implementar nuestros propios modelos de red convolucional.

## REFERENCIAS BIBLIOGRÁFICAS

- Depuru, S., Nandam, A., Sivanantham, S., Amala, K., Akshaya, V., & Saktivel, M. (2022). Convolutional Neural Network based Human Emotion Recognition System: A Deep Learning Approach. *2022 Smart Technologies, Communication and Robotics (STCR)*, 1-4. <https://doi.org/10.1109/STCR55312.2022.10009123>
- Deramgozin, M. M., Jovanovic, S., Arevalillo-HerrÁez, M., & Rabah, H. (2022). An Explainable and Reliable Facial Expression Recognition System for Remote Health Monitoring. *2022 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 1-4. <https://doi.org/10.1109/ICECS202256217.2022.9971040>
- Dwijayanti, S., Iqbal, M., & Suprpto, 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>
- Gokul Prasad, C., Gowtham, S., Mahesh, M., Sainudeen, A. B., Prabhakaran, P., & Kumareshan, N. (2023). A DT Framework based AI Robot For Facial Recognition using IoT. *2023 International Conference on Computer Communication and Informatics (ICCCI)*, 1-5. <https://doi.org/10.1109/ICCCI56745.2023.10128633>
- Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shave-Taylor, J., Milakov, M., Park, J., ... Bengio, Y. (2013). Challenges in Representation Learning: A Report on Three Machine Learning Contests. En M. Lee, A. Hirose, Z.-G. Hou, & R. M. Kil (Eds.), *Neural Information Processing* (pp. 117-124). Springer. [https://doi.org/10.1007/978-3-642-42051-1\\_16](https://doi.org/10.1007/978-3-642-42051-1_16)
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Howard, J. (2023). *Fastai* [Jupyter Notebook]. fast.ai. <https://github.com/fastai/fastai> (Obra original publicada en 2017)
- Karnati, M., Seal, A., Bhattacharjee, D., Yazidi, A., & Krejcar, O. (2023). Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey. *IEEE Transactions on Instrumentation and Measurement*, 72, 1-31. <https://doi.org/10.1109/TIM.2023.3243661>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25. [https://proceedings.neurips.cc/paper\\_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html)
- Mukhopadhyay, M., Dey, A., Shaw, R. N., & Ghosh, A. (2021). Facial emotion recognition based on Textural pattern and Convolutional Neural Network. *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, 1-6. <https://doi.org/10.1109/GUCON50781.2021.9573860>
- R, S., G, S., & V, A. (2022). Facial Emotion Recognition using Deep Learning Approach. *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 1064-1069. <https://doi.org/10.1109/ICACRS55517.2022.10029092>
- Reserva de Derechos al Uso Exclusivo No. 04-2023-091910490900-102, ISSN: En trámite. Año 1, No. 1, Diciembre 2023- Marzo 2024.
- Fecha de Recepción: 03/12/2023| Fecha de Aceptación: 11/12/2023



TECNOLÓGICO  
NACIONAL DE MÉXICO®



**VI** CONGRESO Nacional de Investigación en  
Ciencia e Innovación de  
Tecnologías Productivas

Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. <http://arxiv.org/abs/1409.1556>

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>

Zhang, B., Wei, D., Zhang, Q., Si, W., Li, X., & Zhu, Q. (2021). Classroom monitoring system based on facial expression recognition. *2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 108-111. <https://doi.org/10.1109/DCABES52998.2021.00034>

