

Modelo Predictor para el Análisis de Emociones

Ernesto Acosta Ruiz

A01364982

ABSTRACT - El principal objetivo de este documento es presentar la implementación del modelo clasificador de emociones con imágenes.

1. INTRODUCCIÓN

Los seres humanos somos seres complejos llenos de múltiples emociones que reflejan nuestro estado de ánimo de acuerdo a la experiencia que estamos viviendo en un dado momento. Aunque los humanos estamos programados para entender de manera intuitiva las emociones de otros humanos, a veces es bueno tener una idea más consciente de cuál es la emoción específica que siente una persona.

2. DATASET

El dataset que se utiliza para entrenar este modelo incluye 7 emociones principales:

- felicidad
- tristeza
- enojo
- miedo
- disgusto
- sorpresa

- neutral

Para cada emoción tenemos múltiples imágenes con el mismo tratado. La resolución es de 96x96 píxeles, todas están a escala de grises y están enfocadas en la cara de las personas. La distribución de la cantidad de imágenes es la siguiente:

- felicidad = 11,398
- tristeza = 6,535
- enojo = 5,920
- miedo = 5,920
- disgusto = 5,920
- sorpresa = 5,920
- neutral = 8,166

Total de imágenes = 49,779

3. LIMPIEZA DE IMÁGENES

Para la primera iteración no hubo limpieza de datos ya que todas las imágenes contaban con el mismo tamaño y formato.

4. DIVISIÓN DE IMÁGENES

La división de imágenes fue la siguiente:

- Train = 80%

Maximiza las fotos disponibles para que pueda aprender de las múltiples categorías presentes en el dataset.

- Validation = 10%

Se usa para la selección de hiperparámetros y ajuste de métricas. Al mantener este conjunto separado evitamos el overfitting por prueba y error.

- Test = 10%

Esta sección se mantiene intacta hasta el final y se utiliza una vez para estimar el desempeño fuera de la muestra total. No se usa para tomar decisiones del modelo

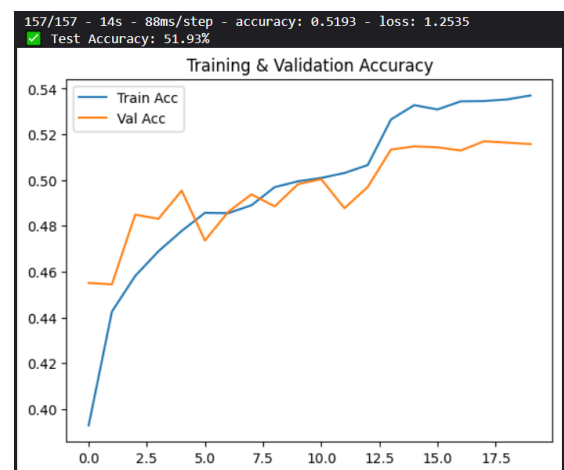
5. MODELO BASE

Para el modelo base usado para la predicción de emociones se utilizó el método de “transfer learning” con el cual usamos un modelo previamente entrenado para ahorrarnos tiempo de crear todo el proceso desde cero. Para eso, se implementó el modelo MobileNetV2. MobileNetV2 es una red convolucional ligera y eficiente que destaca por funcionar en teléfonos inteligentes al mantener alta precisión con pocos parámetros. Su arquitectura se diferencia de una red convolucional tradicional, como lo es una conv2D, al dividir la operación en dos

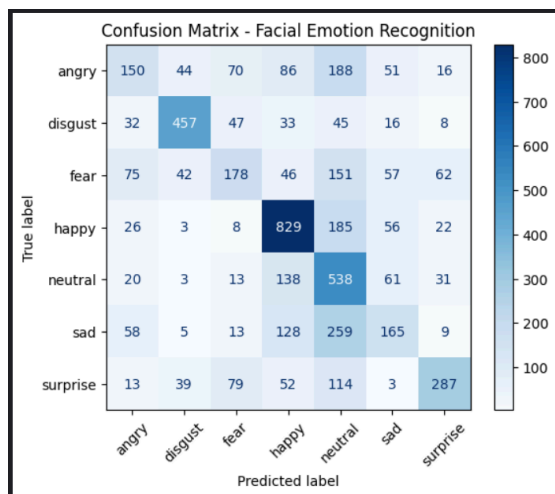
pasos: aplica un filtro diferente a cada canal del RGB por separado y combina esos resultados con filtros 1x1 mezclando la información de cada canal. Esto le permite la misma capacidad de extraer patrones y realiza 9 veces menos operaciones.

Después del procesado realizado por MobileNetV2, tenemos una capa con max pooling para mantener las características más fuertes del mapa de características, luego pasamos por una capa densa de 128 neuronas como clasificador intermedio para traducir las features extraídas por MobileNetV2, luego pasa por una capa de dropout con la cual apagamos aleatoriamente el 40% de las neuronas de la capa anterior para evitar el overfitting y finalmente tenemos la última capa densa con el número de clases siendo el número de neuronas de esta como clasificador final activado por softmax.

6. RESULTADOS MODELO BASE



Lo primero que podemos notar de los resultados es el accuracy del modelo es del 51.93% con el cual podemos concluir que es suficiente para tener un modelo con potencial para tener un clasificador de emociones. Cabe resaltar que este modelo está adecuadamente ajustado ya que no podemos notar overfitting gracias a que el accuracy de train y validation son muy similares.



La matriz de confusión de las clases del modelo nos muestra una situación importante: la emoción “neutral” está confundiendo significativamente al modelo ya que las emociones de enojo, miedo, y tristeza tiene como primera o segunda opción la emoción previamente mencionada. Por otro lado, la felicidad parece ser la emoción con mejor resultados pero muy probablemente se deba a que tiene casi el doble de muestras que la mayoría de las demás emociones.

	precision	recall	f1-score	support
angry	0.40	0.25	0.31	605
disgust	0.77	0.72	0.74	638
fear	0.44	0.29	0.35	611
happy	0.63	0.73	0.68	1129
neutral	0.36	0.67	0.47	804
sad	0.40	0.26	0.32	637
surprise	0.66	0.49	0.56	587
accuracy			0.52	5011
macro avg	0.52	0.49	0.49	5011
weighted avg	0.53	0.52	0.51	5011

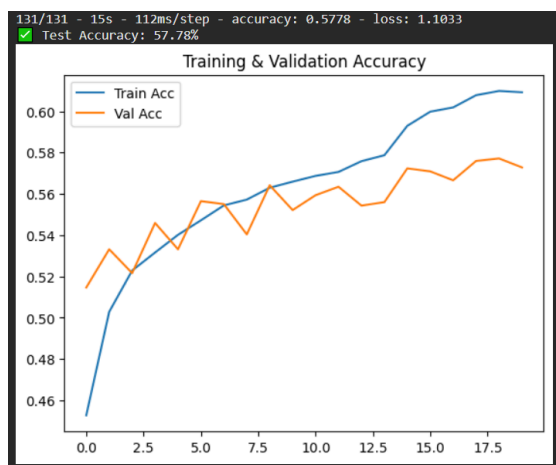
Viendo el f1-score de cada emoción, podemos comprobar que en efecto enojo, miedo y tristeza tienen puntajes significativamente bajos y este modelo no aprende lo suficientemente bien para tener un clasificador de emociones.

7. MEJORAS A MODELO

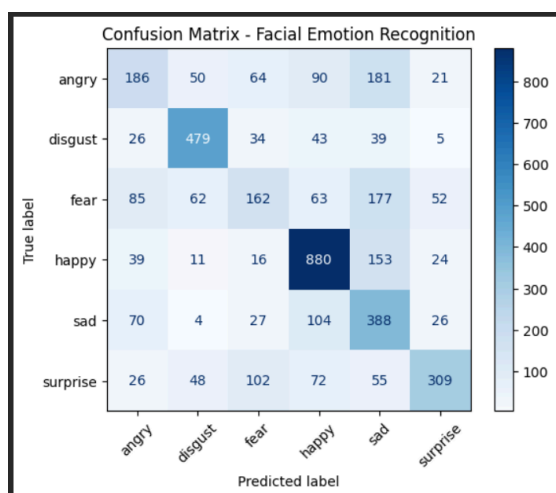
Hubo varias iteraciones con las cuales se intentó mejorar el modelo, por ejemplo, se cambió el modelo de transfer learning a EfficientNetB0, ResNet50V2, DenseNet121 y Xception. Ninguno de estos modelos lograron mejorar el accuracy del modelo base con MobileNetV2. Lo que terminó haciendo un cambio significativo fue un ajuste en el dataset. Cuando vemos la matriz de confusión del modelo base, podemos notar que para todos los sentimientos, el sentimiento que más se predice aparte del true label es la emoción neutral. Esto causa que parte del error de las predicciones se va porque la emoción neutral realmente puede llegar a confundirse con cualquiera de las demás ya que no hay características fuertes definidas. Este efecto también

puede apreciarse con los seres humanos, así que para que el modelo pueda aprender a identificar las emociones con precisión, quitaremos las imágenes con la emoción neutral.

8. RESULTADOS MEJORAS A MODELO



En este caso tenemos un accuracy del 57.78% en el conjunto de test, un incremento del 5.85% comparado con el modelo base. De la misma manera, no contamos con overfitting ya que la diferencia de accuracy entre train y validation es mínima.



Ahora, todas las clases menos miedo clasifican en su mayoría la emoción correcta, sin embargo ahora tristeza es la que tomó el rol de la emoción neutral haciendo esta emoción la segunda mejor opción para las demás emociones.

	precision	recall	f1-score	support
angry	0.42	0.30	0.35	596
disgust	0.73	0.76	0.75	629
fear	0.40	0.32	0.35	596
happy	0.71	0.80	0.75	1130
sad	0.40	0.58	0.48	609
surprise	0.69	0.54	0.60	613
accuracy			0.58	4173
macro avg	0.56	0.55	0.55	4173
weighted avg	0.58	0.58	0.57	4173

Las mejoras de f1-scores de las emociones son las siguiente:

- enojo = +0.04
- disgusto = +0.01
- miedo = 0.00
- felicidad = +0.07
- tristeza = +0.16
- sorpresa = +0.04

Tristeza tuvo un brinco considerable pero también se debe a que el modelo tiende a clasificar como tristeza a las emociones con las que no está seguro. Enojo y miedo siguen teniendo puntajes muy bajos como para ser útil como clasificador.

9. CONCLUSIONES

El reconocimiento de emociones faciales puede llegar a ser complicado para los seres humanos, entonces es aún más difícil

para los modelos de deep learning. Una de las razones principales por la cual es difícil hacer un clasificador de emociones es que varias emociones pueden verse igual a primera vista, lo que las diferencia son microexpresiones que pueden no ser muy visibles pero que podemos notar de manera intuitiva. Así mismo puede que el etiquetado del dataset no sea perfecto, haciendo que el modelo no aprenda a distinguir estas diferencias sutiles. Posiblemente, para mejorar el accuracy del modelo se tenga que usar imágenes de mayor resolución para que se pueda alcanzar a detectar estas diferencias sutiles entre emociones similares.

emociones (clasificación y descripción). *Psicología y Mente*.

<https://psicologiaymente.com/psicologia/tipos-de-emociones>

10. REFERENCIAS

- Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (s. f.). A survey of modern deep learning based object detection models. *Digital Signal Processing*.
<https://doi.org/10.1016/j.dsp.2022.103514>
- Corbin, J. A. (2025, 16 septiembre). Los 8 tipos de