Modelado demográfico facial con CNN

Daniel Erick Sanchez Trujillo
Facultad de Ciencias Puras y Naturales
Universidad Mayor de San Andrés
La Paz, Bolivia
daniel.sanchez@agetic.gob.bo

Felipe Roberto Sanchez Saravia
Facultad de Ciencias Puras y Naturales
Universidad Mayor de San Andrés
La Paz, Bolivia
sanchezsaravia@gmail.com

Sergio Alejandro Paucara Saca
Facultad de Ciencias Puras y Naturales
Universidad Mayor de San Andrés
La Paz, Bolivia
sergiopaucara@gmail.com

***Resumen*—Este trabajo presenta el desarrollo de un modelo *multitarea* basado en redes neuronales convolucionales (CNN) para estimar edad (en intervalos de 10 años), género y raza a partir de imágenes faciales del conjunto UTKFace. Se realizó un análisis exploratorio de datos (EDA), un pipeline de preprocesamiento facial (detección con MTCNN, normalización de color y *letterboxing*) .**

***Index Terms*—Visión por Computador, CNN, Aprendizaje Multitarea, Edad, Género, Raza, UTKFace, Validación Cruzada, Google Colab**

I. Introducción

La estimación automática de atributos demográficos (edad, género, raza) a partir de imágenes faciales es una tarea clásica de visión por computador, con aplicaciones en analítica de audiencias, kioscos de autoatención y mejora de experiencias en trámites digitales. En este proyecto se desarrolla un modelo CNN *multitarea* que aprende conjuntamente estas tres variables, apoyado en la etiqueta embebida en el nombre de archivo de UTKFace (`age_gender_race_date&time.jpg`). El entrenamiento se realizó en Google Colab y se documenta tanto el flujo con preprocesamiento facial como un flujo simple sin él, privilegiando reproducibilidad y tiempo de ejecución.

II. Estado del arte

Las CNN han dominado tareas de reconocimiento visual desde VGG y ResNet, y variantes livianas como MobileNetV2 han facilitado despliegues en tiempo real. Para edad, se ha explorado regresión directa, clasificación por *bins* y enfoques ordinales; para género y raza, clasificación multicategoría. El *multitask learning* (MTL) comparte representaciones intermedias y suele mejorar la generalización cuando las tareas están correlacionadas. La detección/alineación facial (p. ej. MTCNN) y el control de iluminación (CLAHE en luminancia) pueden estabilizar la señal visual previa al modelo.

III. Área de aplicación

El sistema apunta a:

- Analítica de atención (kioscos, salas de espera).
- Soporte no intrusivo a perfiles de usuario para adaptación de interfaces.
- Prototipos de verificación auxiliar en trámites digitales.

Se enfatiza el uso responsable: minimización de datos, transparencia, evaluación de sesgos y cumplimiento regulatorio.

IV. Metodología

IV-A. Datos y rotulado

Se empleó el conjunto de datos **UTKFace** [1], el cual contiene imágenes faciales de personas de distintas edades y grupos étnicos. Las etiquetas de cada imagen se extrajeron del nombre del archivo y corresponden a:

- **Edad** — posteriormente agrupada en intervalos de 10 años para estabilizar el aprendizaje.
- **Género** — codificado como:
  - 0: Masculino
  - 1: Femenino
- **Raza** — codificada como:
  - 0: White
  - 1: Black
  - 2: Asian
  - 3: Indian
  - 4: Other

IV-B. Análisis exploratorio de datos

En esta sección se presenta un análisis exploratorio de los datos, con el objetivo de comprender la distribución de las variables clave antes de aplicar modelos de aprendizaje automático. Se examinan las distribuciones de edad, género y raza, así como su relación conjunta mediante mapas de calor (*heatmaps*). Estos análisis permiten identificar posibles desbalances y patrones que podrían influir en el desempeño de los modelos.

Distribución de edades La Tabla I muestra la distribución de la población en intervalos de 10 años.

Cuadro I: Distribución de edades en intervalos de 10 años

Grupo de edad	Cantidad
0–9	3217
10–19	1239
20–29	1524
30–39	1038
40–49	679
50–59	946
60–69	664
70–79	393
80–89	325
90–99	103
100–109	4
110–119	3

Distribución de género La Tabla II resume la distribución por género.

Cuadro II: Distribución de género

Género	Cantidad
Femenino	5596
Masculino	4539

Distribución de raza La Tabla III presenta la distribución de la población por raza.

Cuadro III: Distribución de raza

Raza	Cantidad
Blanca	5396
Asiática	1703
India	1493
Otra	1121
Negra	422

Visualización de distribuciones Las Figuras 1 y 2 presentan mapas de calor que muestran la distribución conjunta de la edad con respecto al género y a la raza.

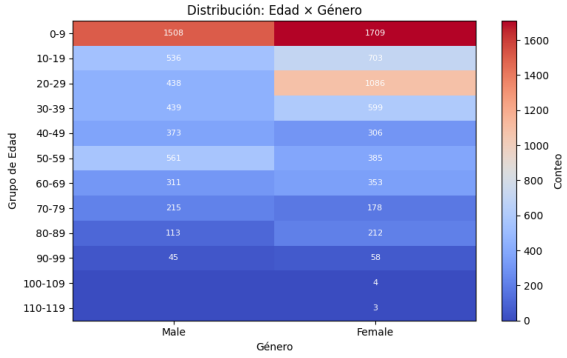

Figura 1: Heatmap de distribución: Edad × Género.

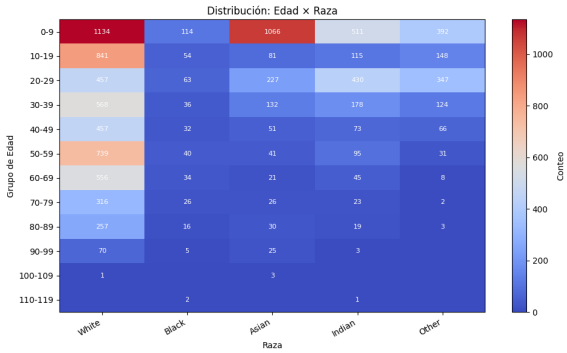

Figura 2: Heatmap de distribución: Edad × Raza.

El conjunto de datos analizado consta de 10 137 imágenes. Se observa un predominio de individuos en el grupo etario de 0–9 años, seguido por los de 20–29 y 10–19 años, lo que indica una mayor presencia de poblaciones jóvenes. Existe un ligero desbalance de género, con mayor proporción de personas femeninas (55 %) frente a masculinas (45 %). En cuanto a la raza, la blanca es la más frecuente, representando más de la mitad de la muestra, mientras que la raza negra presenta la menor participación. Estos patrones sugieren posibles sesgos que deberán considerarse durante el entrenamiento de los modelos y, de ser necesario, aplicar técnicas de balanceo de datos.

IV-C. Preprocesamiento

El objetivo del preprocesamiento es estabilizar el *encuadre*, la *iluminación* y la *escala* para reducir variabilidad que no aporta al aprendizaje. Implementamos tres etapas principales: detección/recorte facial, normalización fotométrica y redimensionado con *letterbox*. Cuando la ruta "rápida" (sin preprocesamiento) fue necesaria por tiempo, se omitieron las etapas 1–2 y se aplicó únicamente redimensionado y normalización a $[0, 1]$.

1. **Detección y recorte facial** Se emplea MTCNN sobre la imagen en RGB para localizar el rostro.
   - Si hay múltiples detecciones, se selecciona la de mayor área.
   - Se descartan cajas con confianza $< 0{,}90$ o menores a 64×64 píxeles.
   - Para capturar atributos de cabello (útiles en género/edad), se expande la caja de forma **asimétrica** del alto/ancho original.
   - Se recorta a los límites de la imagen.
   - Con los puntos de ojos provistos por MTCNN puede alinearse el rostro mediante rotación del eje ocular; en este trabajo se omite para reducir costo computacional.
   - Si no se detecta rostro, se aplica un *center-crop* cuadrado como *fallback*.
2. **Normalización fotométrica** Busca reducir variaciones de color e iluminación que afectan la textura de piel.
   - **Balance de blancos (gray-world):** escala cada canal RGB para igualar medias, limitando las ganancias a $[0{,}8,\ 1{,}2]$.
   - **CLAHE en luminancia:** conversión a YCrCb y aplicación en el canal Y, mejorando contraste local sin saturar tonos.
   - **Corrección gamma:** ajuste para suavizar sombras o brillos.

   Para entrenamiento, la imagen se normaliza a $[0, 1]$ y se convierte de BGR a RGB en la tubería de datos.
3. **Redimensionado con *letterbox*** Se preserva la razón de aspecto al ajustar la imagen a un tamaño objetivo:
   - Se aplica *padding* centrado con color neutro (p. ej., 114) o promedio del borde.
   - Se registran *scale* y *padding* para mapear coordenadas de vuelta si es necesario.
4. **Robustez y rendimiento**
   - Control de determinismo mediante *seeds*.
   - Exclusión de archivos mal formados.
   - Cacheo de imágenes preprocesadas (`.npz`) para reutilización en los 5 folds.

- En la variante rápida, se omiten MTCNN, CLAHE y corrección gamma, manteniendo la misma partición (10 % test, 5-fold en el 90 % restante).

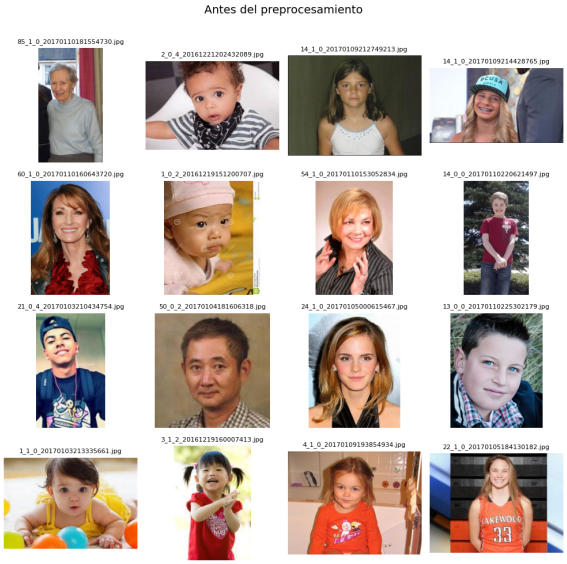

Figura 3: Ejemplo antes del preprocesamiento.

Figura 4: Ejemplo después del preprocesamiento.

IV-D. Arquitectura del modelo

El modelo es una *CNN* multitarea que comparte un tronco para extraer rasgos faciales y se divide en tres salidas que, de forma conjunta, estiman edad por rangos, género y raza. Emplea bloques sencillos de convolución y agrupamiento para capturar patrones visuales, los condensa en un vector compacto y aplica una ligera regularización para evitar sobreajuste. Sobre ese resumen se sitúan tres clasificadores independientes (uno por objetivo). Con un esquema de entrenamiento estándar, resulta liviano y rápido, y el aprendizaje conjunto favorece predicciones más estables al reutilizar información entre las tres tareas.

Cuadro IV: Configuración de entrenamiento.

Elemento	**Valor**
Optimizador	Adam (por defecto de Keras)
Pérdidas	CCE por cabeza: age, gender, race
Métricas	Accuracy por cabeza
Tamaño de imagen	`IMG_SIZE` (p.ej., 100–128)
Normalización	Reescalado a $[0, 1]$
Regularización	Dropout(0.4)
Callbacks	EarlyStopping, ReduceLROnPlateau
Salidas	`age`(10), `gender`(2), `race`(5)

IV-E. Esquema de entrenamiento y particiones

Se adoptó:

- **10 % Test** fijo y estratificado para evaluación final imparcial.
- **5-fold CV** sobre el **90 % restante**: en cada fold ∼80 % entrenamiento y ∼20 % validación (promedios globales: 72 %/18 %/10 %).

Se ejecutó en Google Colab, utilizando su entorno de ejecución con GPU NVIDIA T4 para el entrenamiento del modelo.

V. Resultados

A nivel de validación por *fold* se recopilaron métricas de `accuracy` para cada cabeza de salida (*age*, *gender*, *race*) y la pérdida total. Finalmente, se reentrenó con el conjunto completo *Train+Val* (90 %) y se evaluó sobre el conjunto *Test* (10 %).

Los siguientes ejemplos muestran predicciones en las que el modelo estimó correctamente la edad aproximada, el género y la raza, demostrando un rendimiento exitoso en datos no vistos.

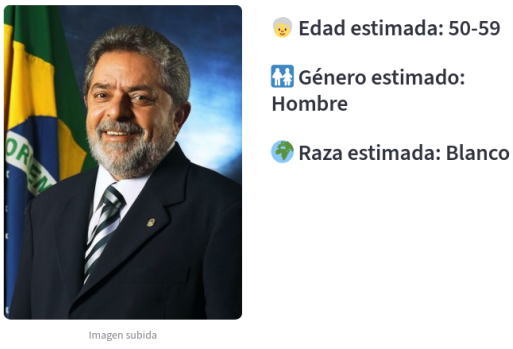

Figura 5: Predicción de una aspectos de una persona

Figura 6: Predicción de una aspectos de una persona

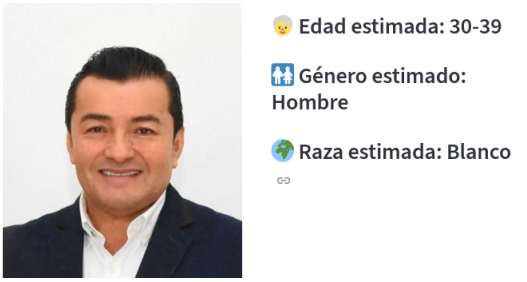

Figura 7: Predicción de una aspectos de una persona

El entrenamiento final se guardó como `modelo25.h5` para despliegue.

VI. Discusión

El multitask learning favoreció género y raza, mientras que la edad resultó más desafiante (etiquetas ruidosas y superposición entre bins adyacentes). El preprocesamiento (detección/alineación, normalización de luminancia) puede mejorar estabilidad, aunque incrementa el tiempo de cómputo; la ruta simple sin preprocesamiento acelera notablemente con un costo potencial en precisión y robustez. El desbalance por edad/raza influye en la generalización; *class weights*, *data augmentation* y curación de datos locales (p. ej. población objetivo) son palancas de mejora. En producción, deben considerarse aspectos de equidad y privacidad.

VII. Conclusiones

El modelo mostró un desempeño consistente en la predicción de género y raza, mientras que la estimación de la edad mejoró mediante el uso de agrupación en intervalos (*binning*) y técnicas ordinales. Como trabajo futuro, se propone aplicar *fine-tuning* con *backbones* preentrenados (p. ej., MobileNetV2), incorporar *class weights* y ampliar la validación con datos locales. Asimismo, se recomienda reentrenar el modelo con un conjunto de datos de personas bolivianas para mejorar su capacidad de generalización en contextos nacionales y optimizar el proceso de entrenamiento.

Referencias

[1] S. Zhang. "UTKFace Dataset." Disponible en: https://susanqq.github.io/UTKFace/. Accedido: ago. 2025.
[2] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks," *IEEE SPL*, 2016.
[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *CVPR*, 2018.
[4] D. Sanchez, "Repositorio del proyecto," GitHub, (https://github.com/destandroid/Proyecto-2-Modulo-10).