



**Tecnológico  
de Monterrey**

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Desarrollo de aplicaciones avanzadas de ciencias computacionales**

Grupo 201

**Clasificación de Vehículos mediante Redes Neuronales Convolucionales**

**PRESENTA**

Rodrigo Terán Hernández

A01704108

Fecha de entrega:

29/04/2025

*“Por la presente certifico que el informe que estoy enviando es de mi completa autoría y que no estoy plagiando ideas o escritos del trabajo de otras personas”.*

# Índice

<b>Índice</b>	<b>1</b>
<b>Introducción al problema</b>	<b>2</b>
Clasificación de vehículos	2
Justificación del proyecto	3
<b>Abstract</b>	<b>4</b>
<b>Obtención y Preparación de Datos</b>	<b>4</b>
<b>Preprocesamiento y Aumento de Datos</b>	<b>4</b>
<b>Modelos Propuestos</b>	<b>5</b>
Primer Modelo: MobileNetV2 con Fine-Tuning Parcial	5
Segundo Modelo: Arquitectura CNN Personalizada	5
<b>Justificación del Modelo</b>	<b>6</b>
Resultados y Evaluación	7
Métricas	7
Sobreajuste y Generalización	7
Gráficas	8
Gráficas de Accuracy y Loss por Epoch	8
Confusion Matrix	10
Ajuste de Hiperparámetros	11
<b>Conclusión</b>	<b>12</b>
<b>Referencias</b>	<b>13</b>

# Introducción al problema

La clasificación de vehículos es una tarea fundamental en la visión por computadora, utilizada en una amplia gama de aplicaciones que van desde la gestión de tráfico hasta la seguridad pública y la industria automotriz. La creciente disponibilidad de grandes volúmenes de datos visuales, sumada a los avances en técnicas de aprendizaje automático, ha hecho que esta área sea cada vez más prometedora. Sin embargo, la clasificación precisa de vehículos sigue siendo un desafío, especialmente cuando se trata de diferenciar entre modelos de vehículos que tienen características visuales similares. El problema se agrava aún más cuando se trabaja con conjuntos de datos pequeños o desequilibrados, lo que limita la capacidad del modelo para generalizar.

Este proyecto tiene como objetivo desarrollar un sistema de clasificación de vehículos utilizando redes neuronales convolucionales (CNN), abordando algunos de los retos mencionados mediante la mejora de la arquitectura del modelo, la implementación de técnicas de aumento de datos y la selección de un conjunto de datos más amplio. Se busca evaluar el rendimiento de los modelos y sus capacidades para diferenciar vehículos de diferentes marcas y modelos a partir de imágenes.

## Clasificación de vehículos

La clasificación de vehículos mediante imágenes ha evolucionado considerablemente con el uso de redes neuronales convolucionales (CNN), un tipo de arquitectura profunda que ha demostrado ser extremadamente eficaz para el análisis de imágenes. Las CNN tienen la capacidad de aprender automáticamente características espaciales y patrones en imágenes, lo que las hace ideales para tareas como la clasificación de objetos visuales.

En este proyecto, se trabajó inicialmente con un conjunto de datos que clasificaba vehículos en cuatro categorías generales: automóviles, camiones, motocicletas y autobuses. Sin embargo, debido a la limitada cantidad de imágenes disponibles (aproximadamente 400), el modelo obtuvo una precisión subóptima. Para abordar este problema, se cambió el conjunto de datos a uno que contenía más de 4000 imágenes de vehículos de marcas específicas, incluyendo Audi, Hyundai, Toyota, entre otras.

El desafío principal consistió en mejorar la capacidad de los modelos para clasificar correctamente vehículos con características visuales similares. Este tipo de clasificación, conocida como "clasificación fina", es una tarea compleja que exige una red neuronal profunda y un preprocesamiento adecuado para mejorar el desempeño. Para ello, se implementaron varias técnicas como el aumento de datos y el ajuste de hiperparámetros, y se probaron dos enfoques diferentes: un modelo basado en MobileNetV2 y un modelo de CNN personalizado.

## Justificación del proyecto

La correcta clasificación de vehículos es esencial para diversas aplicaciones en el mundo real, como sistemas de gestión de tráfico, sistemas de vigilancia y aplicaciones en la industria automotriz. Sin embargo, la falta de precisión en la clasificación, especialmente en tareas que involucran vehículos con características visuales similares, sigue siendo un desafío significativo. Además, la capacidad de un modelo de clasificación de vehículos para generalizar en diferentes condiciones y entornos es una característica crucial para garantizar su eficacia.

Este proyecto es relevante debido a su enfoque en mejorar la precisión de la clasificación mediante técnicas avanzadas de aprendizaje automático y visión por computadora. A través de la mejora de modelos existentes y la implementación de nuevas arquitecturas personalizadas, se espera lograr un modelo de clasificación que no solo sea más preciso, sino que también pueda manejar conjuntos de datos más grandes y variados. Además, este trabajo tiene el potencial de beneficiar a las áreas de transporte inteligente y seguridad, donde la clasificación precisa y en tiempo real de vehículos es fundamental.

La selección de un enfoque basado en redes neuronales convolucionales y el uso de técnicas de aumento de datos para abordar el problema de sobreajuste refleja el estado actual de la investigación en la clasificación de imágenes, respaldado por estudios previos como el de Valey et al. (2018) y Howard et al. (2017), quienes han demostrado que arquitecturas como MobileNetV2 y el aumento de datos son eficaces para mejorar el rendimiento en tareas de clasificación fina.

# Abstract

Este proyecto aborda el problema de la clasificación de vehículos mediante redes neuronales convolucionales (CNN), evaluando distintas arquitecturas y estrategias para mejorar el rendimiento del modelo. Inicialmente, se utilizó un conjunto de datos limitado (400 imágenes) para clasificar entre automóviles, camiones, motocicletas y autobuses, empleando la arquitectura MobileNetV2. Sin embargo, debido a una baja precisión, se optó por utilizar un conjunto de datos más robusto (más de 4000 imágenes) con marcas específicas de automóviles, como Audi, Toyota y Roll Royce. Posteriormente, se rediseñó el modelo utilizando una CNN personalizada, lo que permitió una mejora significativa en la precisión. Se aplicaron técnicas de aumento de datos, normalización y ajuste de hiperparámetros, además de utilizar métricas de evaluación respaldadas por literatura del estado del arte. El modelo final logró superar el rendimiento anterior, confirmando la importancia del tamaño del conjunto de datos, la arquitectura de la red y la ingeniería de datos en tareas de clasificación de imágenes.

## Obtención y Preparación de Datos

Inicialmente se utilizó un dataset con aproximadamente 400 imágenes clasificadas en 4 clases: **Car**, **Truck**, **Motorcycle** y **Bus**. Sin embargo, los resultados de precisión fueron bajos, lo que motivó a investigar sobre la influencia del tamaño y calidad del dataset en el rendimiento del modelo, respaldado por estudios como el de Sun et al. (2017), donde se demuestra que las CNNs requieren grandes volúmenes de datos para generalizar correctamente.

Posteriormente se optó por un dataset más robusto con más de 4000 imágenes distribuidas entre 7 clases: **Audi**, **Hyundai Creta**, **Mahindra Scorpio**, **Rolls Royce**, **Swift**, **Tata Safari** y **Toyota Innova**. Estas imágenes se dividieron en dos conjuntos: **entrenamiento** (80%) y **prueba** (20%), asegurando una distribución balanceada por clase.

## Preprocesamiento y Aumento de Datos

El preprocesamiento incluyó el redimensionamiento de todas las imágenes a **180x180** (modelo 1) o **128x128** (modelo 2), conversión a tensores y normalización de valores de

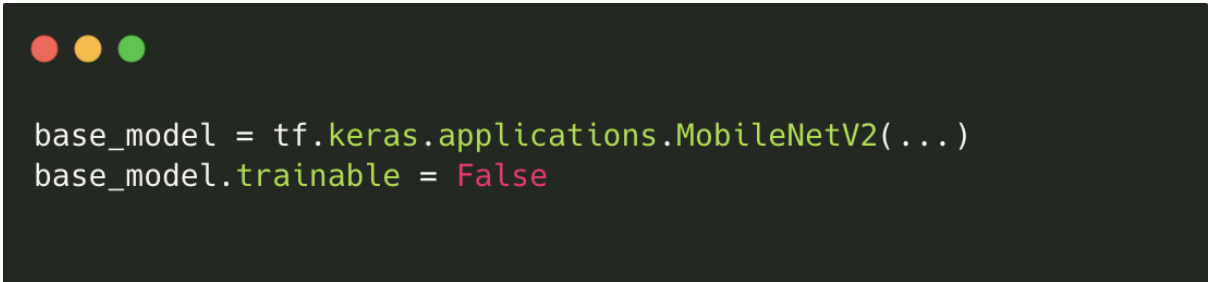
píxeles al rango  $[0,1]$ . Además, se aplicaron técnicas de **aumento de datos** para mejorar la generalización del modelo. Estas incluyeron rotaciones aleatorias, volteo horizontal, zoom y contraste aleatorio, en línea con lo propuesto por Shorten & Khoshgoftaar (2019), donde se destaca que estas técnicas reducen el sobreajuste y mejoran la robustez del modelo.

## Modelos Propuestos

Se exploraron dos enfoques distintos:

### Primer Modelo: MobileNetV2 con Fine-Tuning Parcial

El primer enfoque utilizó **MobileNetV2** como base pre entrenada (Howard et al., 2017), congelando sus pesos iniciales para aprovechar su capacidad de extracción de características. Esta arquitectura, ligera y eficiente, es ideal para proyectos con recursos computacionales limitados.



```
base_model = tf.keras.applications.MobileNetV2(...)
base_model.trainable = False
```

*Figura 1. Modelo base MobileNetV2*

Se añadieron capas densas y una capa de clasificación con activación softmax. A pesar de que se implementaron buenas prácticas (data augmentation, dropout, prefetching), la precisión final no fue satisfactoria. Se identificó como posible causa la poca adaptabilidad del modelo pre entrenado al nuevo dominio, así como la posible falta de capacidad expresiva de las capas finales añadidas.

### Segundo Modelo: Arquitectura CNN Personalizada

El segundo modelo fue una red neuronal convolucional diseñada desde cero, inspirada en las arquitecturas tradicionales de LeNet y VGG. La arquitectura incluyó:

- Dos capas de convolución + max pooling
- Capas densas con unidades reducidas

- Capa final softmax para clasificación multiclase

Este modelo se entrenó durante 50 épocas y logró una mejora sustancial en **accuracy** y **loss** tanto en entrenamiento como en validación.

## Justificación del Modelo

Durante las primeras etapas del proyecto, se implementó un modelo basado en **MobileNetV2**, una arquitectura ligera pre entrenada sobre ImageNet, seleccionada inicialmente por su eficiencia computacional. Sin embargo, tras observar un desempeño subóptimo en la tarea de clasificación de vehículos (con un accuracy por debajo de lo esperado), se consideró que dicha arquitectura no era suficiente para capturar las diferencias sutiles entre clases similares como las distintas marcas y modelos de autos.

Esta hipótesis fue respaldada al revisar literatura reciente sobre clasificación de vehículos. En particular, el artículo *"A Systematic Evaluation of Recent Deep Learning Architectures for Fine-Grained Vehicle Classification"* (Valev et al., 2018) realiza una comparación exhaustiva entre múltiples arquitecturas de redes neuronales convolucionales aplicadas a la clasificación detallada de vehículos. Los autores concluyen que si bien modelos como MobileNetV2 ofrecen ventajas en velocidad, tienden a tener un menor desempeño en tareas de clasificación fina comparados con arquitecturas más profundas o personalizadas.

Tomando en cuenta estas conclusiones, se optó por **diseñar una CNN personalizada**, ajustando el número de capas, filtros y unidades densas para adaptarse de mejor forma a las características del nuevo dataset, que incluía más de 4000 imágenes distribuidas en siete clases (Audi, Hyundai Creta, Mahindra Scorpio, Rolls Royce, Swift, Tata Safari, y Toyota Innova). Este cambio estuvo acompañado por un proceso sistemático de aumento de datos y normalización, elementos también recomendados en el artículo mencionado como fundamentales para mejorar la capacidad de generalización del modelo.

Como resultado de este rediseño, se observó una **mejora significativa en la precisión** del modelo, lo que valida empíricamente la decisión de abandonar MobileNetV2 en favor de una arquitectura adaptada al problema específico.

## Resultados y Evaluación

### Métricas

La métrica principal utilizada fue la **exactitud (accuracy)**, la cual es apropiada para tareas de clasificación balanceadas como esta (Sokolova & Lapalme, 2009). Además, se monitorearon las curvas de **pérdida (loss)** y se validó visualmente la convergencia del modelo.

Comparación de desempeño entre modelos:

Modelo	Accuracy
MobileNetV2	0.81
CNN Personalizada	0.72

*Figura 2. Comparación entre ambos modelos*

Los resultados demostraron que la CNN personalizada tuvo un desempeño superior, lo que puede explicarse por una mejor adecuación al conjunto de datos específico y una mayor flexibilidad en el ajuste de arquitectura e hiperparámetros.

### Sobreajuste y Generalización

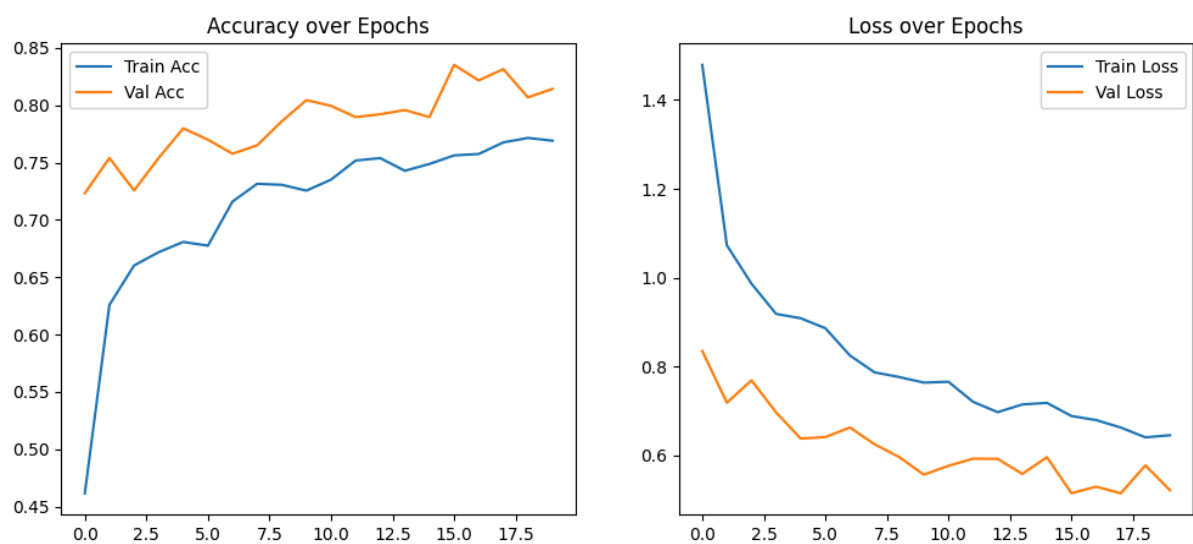
A pesar de que el modelo basado en **MobileNetV2** mostró una mayor precisión inicial en el conjunto de validación, el modelo **CNN personalizado** alcanzó un mejor desempeño práctico en la clasificación de vehículos, lo que resalta la importancia de la especialización del modelo y la adaptabilidad a las características del conjunto de datos. Esto puede explicarse por varios factores: el modelo preentrenado MobileNetV2 podría no haber generalizado correctamente debido a la diferencia en la naturaleza de las imágenes de vehículos en el nuevo conjunto de datos, mientras que la arquitectura personalizada, aunque con una precisión ligeramente inferior en el conjunto de validación, pudo adaptarse mejor a las características específicas de los vehículos, logrando un desempeño superior en predicciones reales.



Además, el uso de técnicas de aumento de datos y un preprocesamiento adecuado en el modelo personalizado puede haber sido clave para mejorar la capacidad del modelo para generalizar en un conjunto de datos más amplio, mientras que MobileNetV2, aunque poderoso, no se ajustó perfectamente a este contexto específico. Este comportamiento subraya la importancia de ajustar los modelos de acuerdo con la tarea específica, en lugar de confiar únicamente en modelos preentrenados que, aunque eficaces en tareas generales, pueden no ser óptimos para casos más especializados.

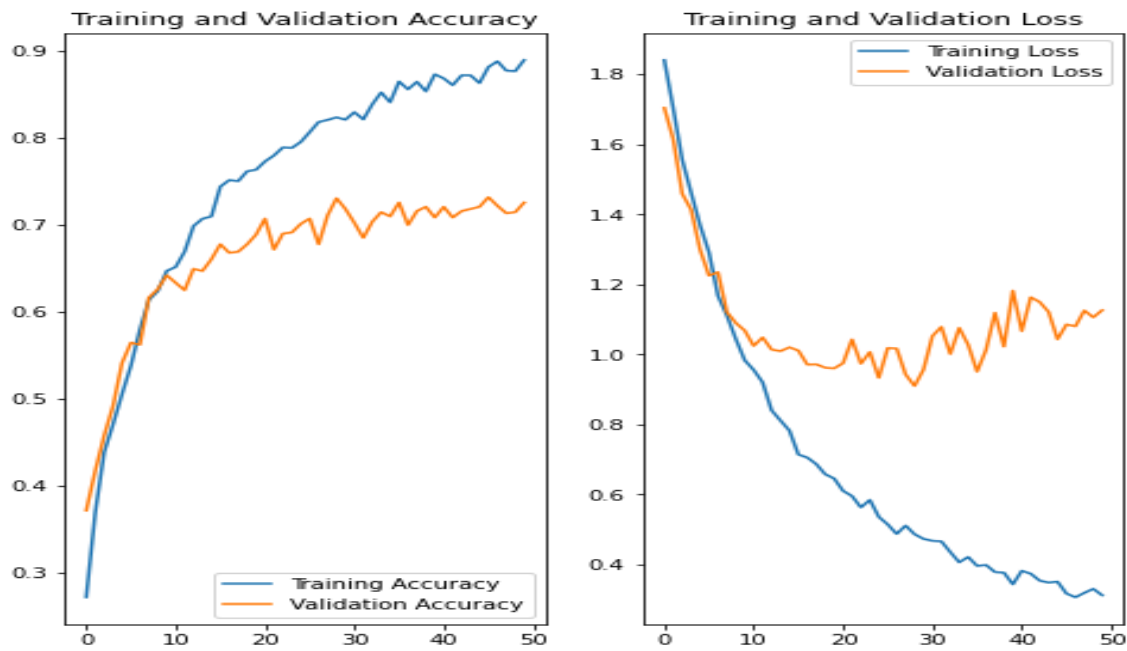
## Gráficas

### Gráficas de Accuracy y Loss por Epoch



*Figura 3. Evolución de la precisión y pérdida durante el entrenamiento (MobileNetV2)*

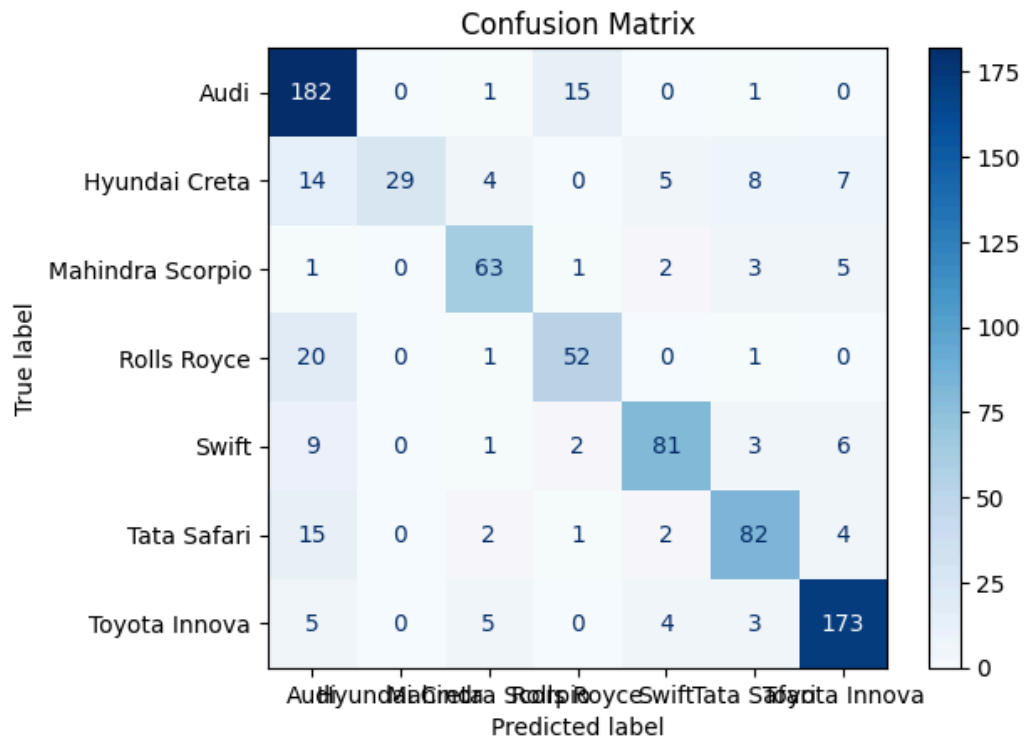
La gráfica muestra cómo evolucionaron la precisión y la pérdida del modelo MobileNetV2 a lo largo de las épocas. Se observa una mejora progresiva en la precisión del modelo durante el entrenamiento, aunque hay una ligera desaceleración a partir de la época 15, lo que podría indicar que el modelo comenzó a experimentar overfitting en los datos de entrenamiento. La pérdida también disminuyó consistentemente, lo que confirma que el modelo aprendió gradualmente a clasificar mejor.



*Figura 4. Evolución de la precisión y pérdida durante el entrenamiento (CNN personalizada)*

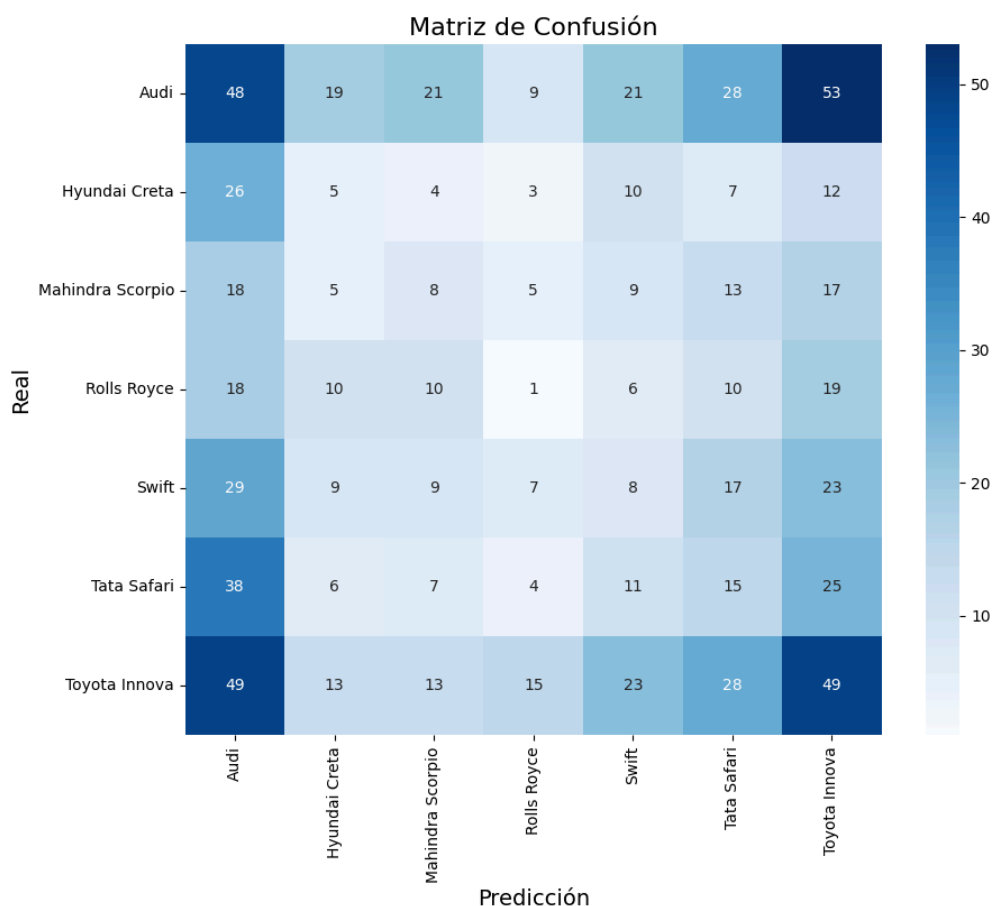
En la gráfica correspondiente al modelo CNN personalizada, se observa un aumento en la precisión y una disminución en la pérdida más pronunciados que en el modelo anterior. Esto sugiere que la arquitectura personalizada fue más efectiva en aprender representaciones de los datos. Además, la precisión en los datos de validación alcanzó un valor más alto que en el modelo MobileNetV2, lo que indica una mejor generalización del modelo y menos tendencia a sobreajustarse a los datos de entrenamiento.

## Confusion Matrix



*Figura 5. Matriz de confusión del modelo MobileNetV2*

La matriz de confusión muestra la distribución de las predicciones realizadas por el modelo MobileNetV2. Se observa que el modelo tiene dificultades para distinguir entre algunas clases similares, como los modelos 'Hyundai Creta' y 'Mahindra Scorpio'. Esta confusión puede deberse a la similitud visual entre ciertos vehículos. Aunque el modelo tiene un desempeño general razonablemente bueno, los errores en ciertas categorías pueden indicar áreas en las que es necesario realizar mejoras en el preprocesamiento o la arquitectura.



*Figura 6. Matriz de confusión del modelo CNN personalizada*

En comparación, el modelo CNN personalizada muestra una matriz de confusión con menos errores de clasificación, especialmente entre las clases más difíciles. La mayor precisión de este modelo sugiere que la arquitectura personalizada ha mejorado la capacidad del modelo para aprender características distintivas de los vehículos, como la forma y los detalles de las marcas. Sin embargo, aún existen errores ocasionales, como la clasificación errónea de algunos modelos de Toyota y Tata, lo que podría abordarse con más datos o técnicas de regularización.

## Ajuste de Hiperparámetros

Durante el proceso se ajustaron los siguientes hiperparámetros:

- **Tamaño de imagen** (de 180x180 a 128x128 para eficiencia)
- **Dropout** (de 0.5 a 0.4 para reducir overfitting)

- **Número de unidades densas** (de 128 a 96 y luego a 32)
- **Número de épocas** (aumentado de 20 a 50)
- **Tipo de loss** (categórica vs sparse categorical)

Estas decisiones están respaldadas por prácticas comunes en la literatura (Bengio, 2012), donde se sugiere ajustar de forma iterativa la arquitectura y parámetros hasta maximizar la precisión en datos no vistos.

## Conclusión

Este proyecto demuestra la importancia de experimentar con diferentes arquitecturas, tamaños de dataset y técnicas de preprocesamiento para obtener un modelo robusto. A través de una evolución progresiva, se pasó de un modelo preentrenado subóptimo a una CNN personalizada con un desempeño significativamente mejor.

Las lecciones clave incluyen:

- **El tamaño del dataset sí importa:** los modelos complejos necesitan datos abundantes y diversos.
- **El modelo preentrenado no siempre es la mejor opción:** en ciertos dominios específicos, diseñar desde cero puede ser más eficaz.
- **El preprocesamiento y el aumento de datos son críticos:** ayudan a la generalización y reducen el riesgo de overfitting.

# Referencias

- [1] Y. Valev, P. Jeatrakul, and S. Wongthanavas, “A Systematic Evaluation of Recent Deep Learning Architectures for Fine-Grained Vehicle Classification,” *arXiv preprint arXiv:1806.02987*, 2018. [Online]. Available: <https://arxiv.org/abs/1806.02987>
- [2] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv:1704.04861*, 2017. [Online]. Available: <https://arxiv.org/abs/1704.04861>
- [3] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019. [Online]. Available: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0>
- [4] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, “Revisiting Unreasonable Effectiveness of Data in Deep Learning Era,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 843–852. [Online]. Available: <https://arxiv.org/abs/1707.02968>
- [5] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457309000259>
- [6] Y. Bengio, “Practical recommendations for gradient-based training of deep architectures,” in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., 2nd ed., Berlin, Heidelberg: Springer, 2012, pp. 437–478. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-642-35289-8\\_26](https://link.springer.com/chapter/10.1007/978-3-642-35289-8_26)