

Detección de Neumonía Usando Inteligencia Artificial

Un modelo para clasificar radiografías de tórax

Gabriel González Arancibia
Ingeniero Civil Químico

¿Por qué es importante este problema?

- El tratamiento de la neumonía depende de su causa. La **Bacterial** se trata con antibióticos, pero la **Viral** no.
- **El Desafío:** Diferenciar entre un pulmón sano, uno viral y uno bacterial puede ser difícil y lento, incluso para un experto.
- **La Pregunta:** ¿Podemos usar IA para crear un "asistente" que ayude a los médicos a clasificar estas imágenes más rápido?



Nuestro Objetivo: Un Clasificador de 3 Categorías

Construimos un "cerebro" artificial (un modelo) que aprende a mirar una radiografía y la clasifica en una de estas tres categorías.



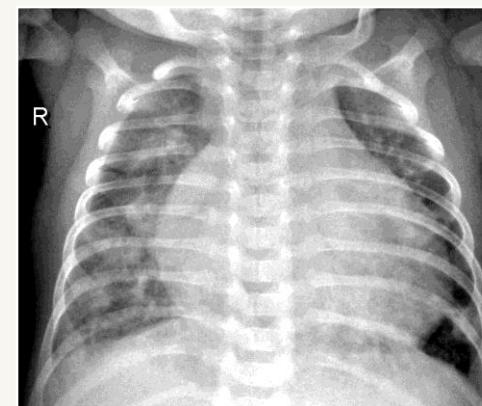
Paciente

Sano



Neumonía

Bacterial



Neumonía

Viral

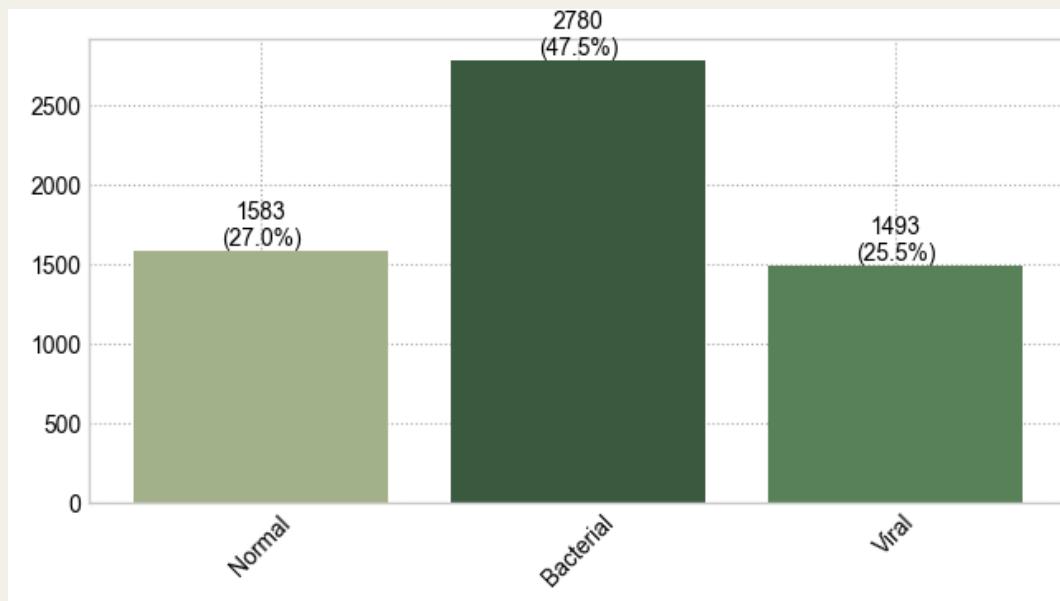


Nuestro Proceso de 4 Pasos

Seguimos el proceso estándar de la ciencia de datos.

Explorar Datos	Analizamos 5,856 imágenes para encontrar patrones y "trampas".
Entrenar Modelo	Le enseñamos a un "cerebro" artificial a "ver" las diferencias.
Optimizar y Mejorar	No nos conformamos con el primer modelo. Buscamos la mejor "receta" (<i>tuning</i>) y creamos un "comité de expertos" (<i>ensamble</i>) para obtener la mejor predicción.
Desplegar	Convertimos el modelo en una herramienta real (una API) que cualquiera puede usar desde internet.

Nº de Imágenes por Clase en el Conjunto de Datos



El Obstáculo: Datos Desbalanceados

Descubrimos que teníamos muchas más imágenes de un tipo ('Bacterial') que de los otros.

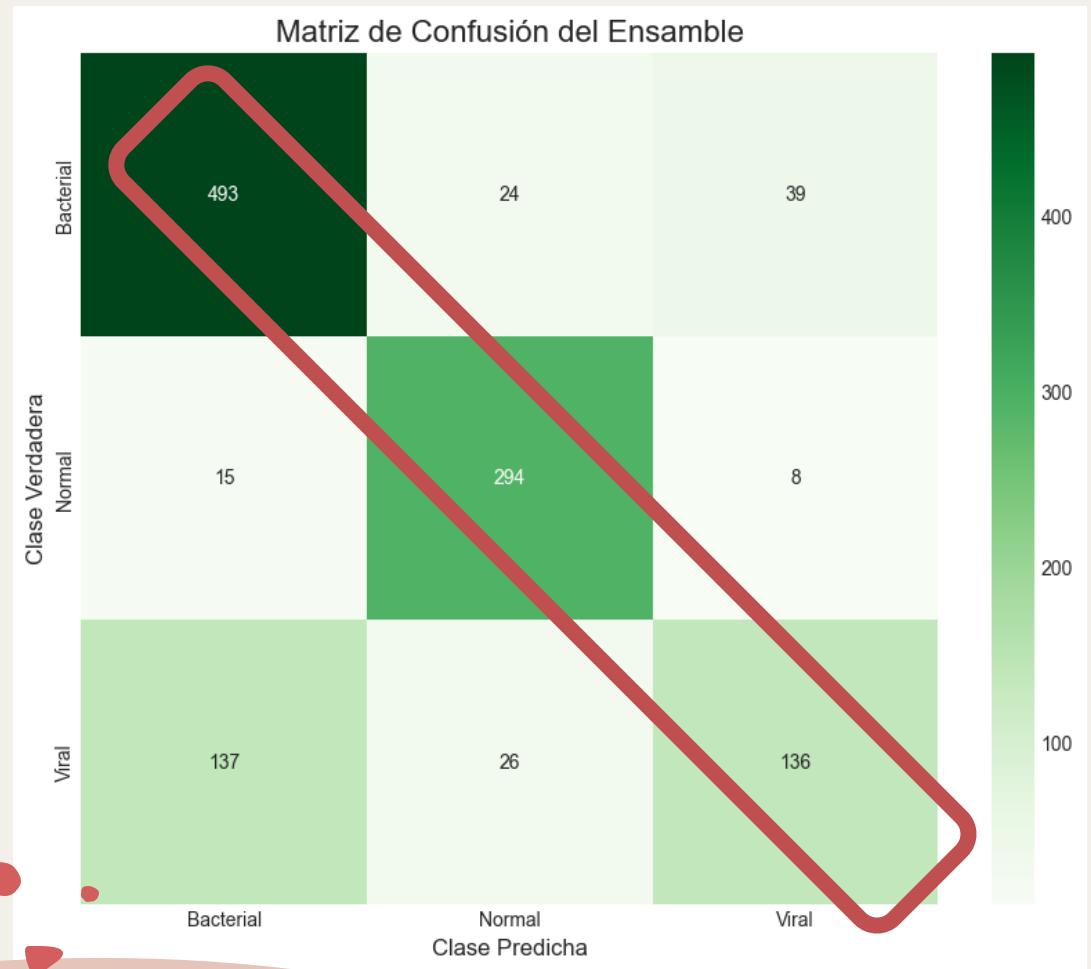
Esto es un problema. Es como intentar enseñar a un niño a reconocer animales mostrándole 100 fotos de perros y solo 50 de gatos. El niño aprenderá a decir 'perro' la mayoría del tiempo.

Para evitar esto, le dijimos al modelo: 'presta el doble de atención cada vez que veas un gato'.

¿Qué tan bien funciona nuestro modelo?

Clase	Puntaje	Qué significa?
Normal	88%	El modelo es excelente para identificar pulmones sanos
Bacterial	80%	Es muy sólido para encontrar la neumonía bacterial.
Viral	60%	Aquí es donde más le cuesta. Confunde 'Viral' con 'Bacterial', lo cual es comprensible porque son visualmente muy parecidas.
Global	78%	Nuestro "comité de expertos" logró una precisión general del 77.4%





Análisis de Aciertos y Errores

Esta es la matriz de confusión, o el 'boletín de calificaciones' del modelo.

La Diagonal: Estos son los **ACIERTOS**. Queremos que estos números sean lo más grandes posible.

Fuera de la Diagonal: Estos son los **ERRORES**.

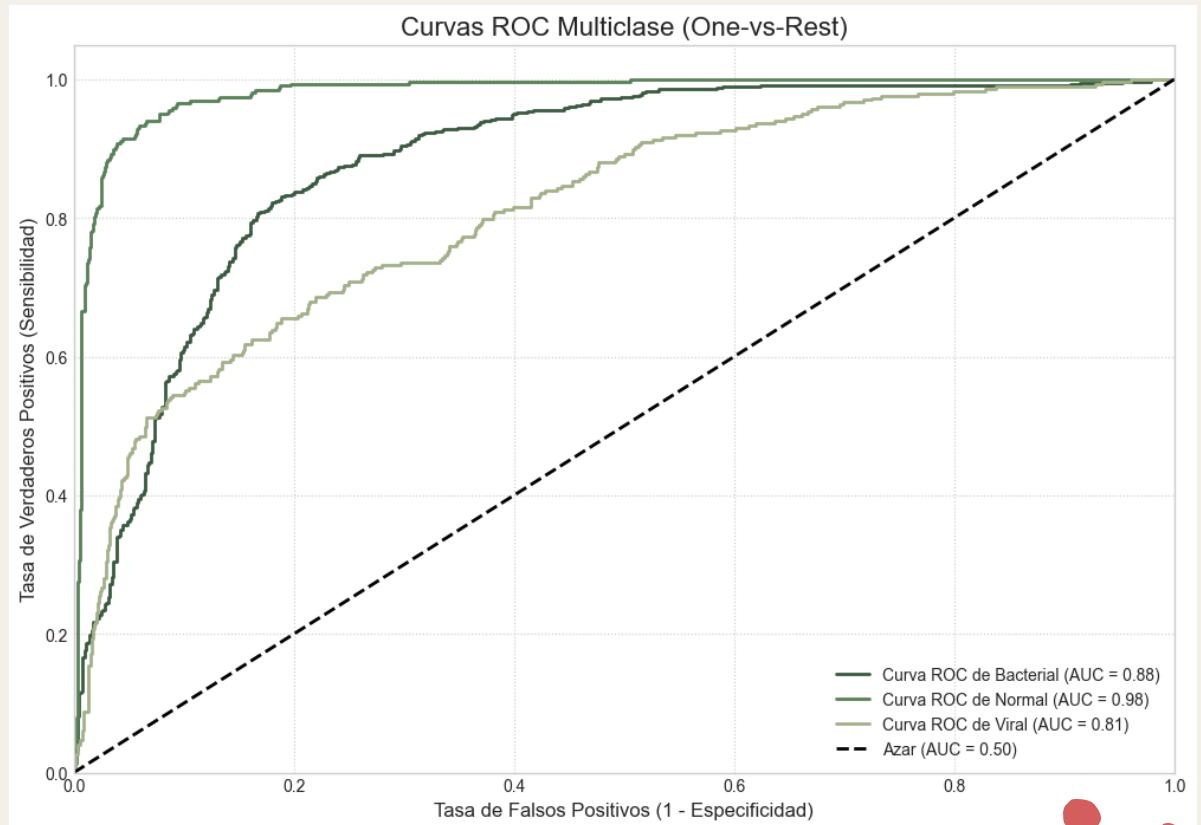
El Hallazgo Clave: El error más grande es que 137 casos 'Virales' fueron confundidos con 'Bacterial'. Esto confirma lo que vimos: diferenciar los dos tipos de neumonía es el mayor desafío.

¿Qué tan "seguro" está el modelo de sus respuestas?

La Línea Punteada (Azar): Esta línea es si el modelo adivinara lanzando una moneda (50/50).

Nuestras Curvas: Mientras más 'esquinadas' y más arriba estén nuestras curvas, mejor.

El Hallazgo Clave: La curva de '**Normal**' es casi perfecta. Esto significa que el modelo está *extremadamente seguro* cuando dice que un pulmón está sano. La curva de '**Viral**' es la más baja, lo que confirma que es la clase que le genera más dudas.



Convertir el Modelo en una Herramienta Real

El modelo no es solo un archivo; es un servicio funcional disponible 24/7 en internet (gracias a Hugging Face).

Probamos la API enviándole una imagen a la URL pública.

El servidor nos devolvió la respuesta correcta:
Respuesta del servidor: {'confianza': '87.13%',
'prediccion': 'Normal'}

El Producto Final (La API)

Convertir el Modelo en una Herramienta Real



El modelo no es solo un archivo; es servicio funcional disponible 24/7 en internet.

Conclusiones

Lo que logramos:

Se construyó un modelo de IA con una precisión fiable del **77.4%**.

El modelo es **excelente para descartar casos normales** (Sensibilidad del 96%), lo que es de gran valor clínico.

El modelo está **desplegado en una API pública**, lista para ser integrada.

Próximos Pasos (Cómo mejorar):

El principal desafío es la confusión Viral/Bacterial. Para mejorar, el siguiente paso sería **probar con arquitecturas de modelo más potentes** (como *Transfer Learning* con ResNet) y conseguir más datos de entrenamiento, especialmente de la clase 'Viral'.

