

CLASIFICACIÓN DE CUADROS DE NEUMONÍA A PARTIR DE RADIOGRAFÍAS TORÁCICAS

BAUTISTA
BUYATTI

TOMÁS
DI NAPOLI

JUAN SANTIAGO
GRASSANO



LABORATORIO
DE DATOS
29 DE JUNIO DE 2021



EL PROBLEMA

2 MILLONES

de niños al año mueren al año por neumonía en el mundo



VIRUS Y BACTERIAS

son los dos causantes principales de neumonía,
pero requieren tratamientos distintos



RADIOGRAFÍAS

pueden ayudar a diferenciar entre
diferentes tipos de neumonía

EL OBJETIVO

investigar la efectividad de distintos
MODELOS DE APRENDIZAJE AUTOMÁTICO
para la
CLASIFICACIÓN
de radiografías de tórax pediátricas con el fin de
DETECTAR Y DISTINGUIR
neumonía viral y bacteriana

EL DATA SET

RADIOGRAFÍAS DE TÓRAX

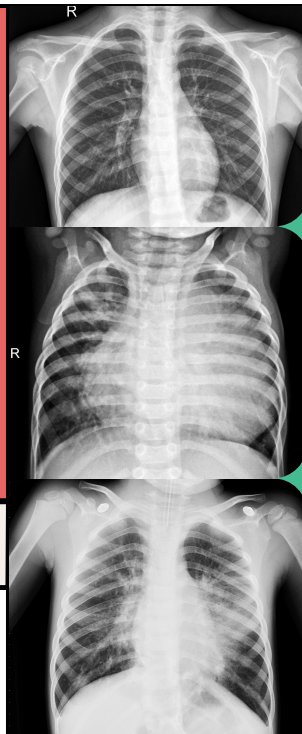
5856 imágenes JPEG

PACIENTES PEDIÁTRICOS

de uno a cinco años

3 CATEGORÍAS

Sin neumonía (1583),
neumonía bacteriana (2780)
y neumonía viral (1493)



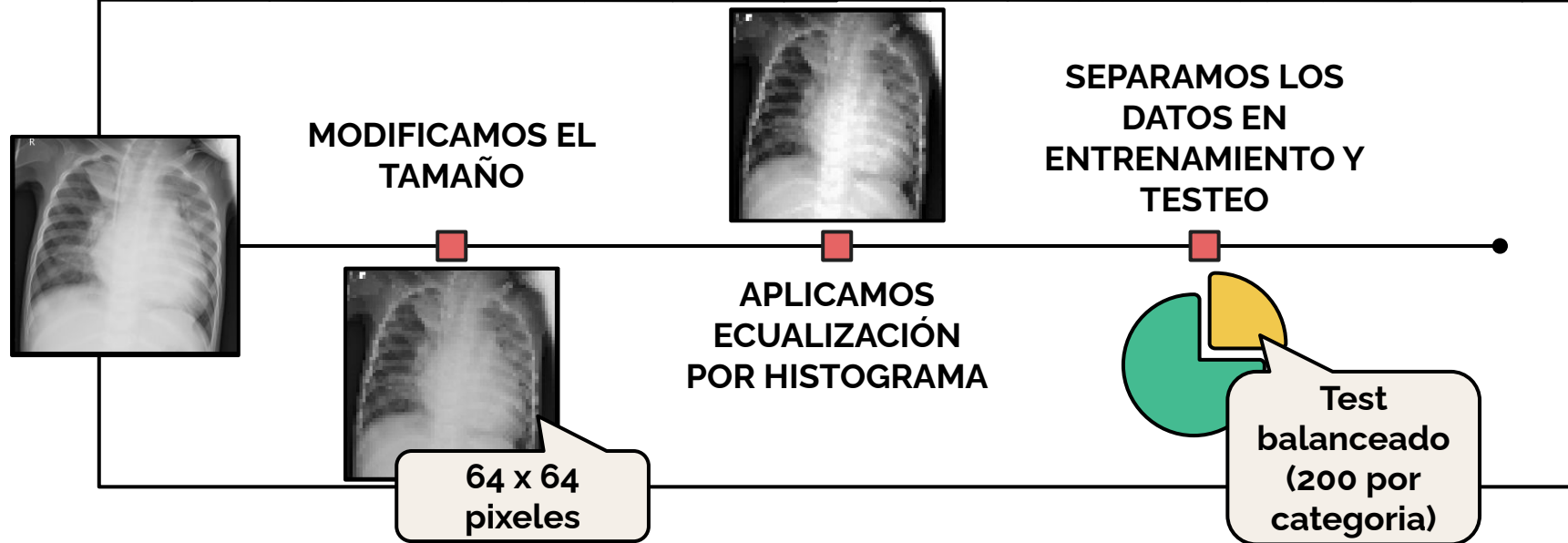
**PULMÓN
NORMAL**

**NEUMONÍA
BACTERIANA**

**NEUMONÍA
VIRAL**

Data set de Kaggle:
Chest X-Ray Images (PNEUMONÍA)

TRATAMIENTO DEL DATA SET



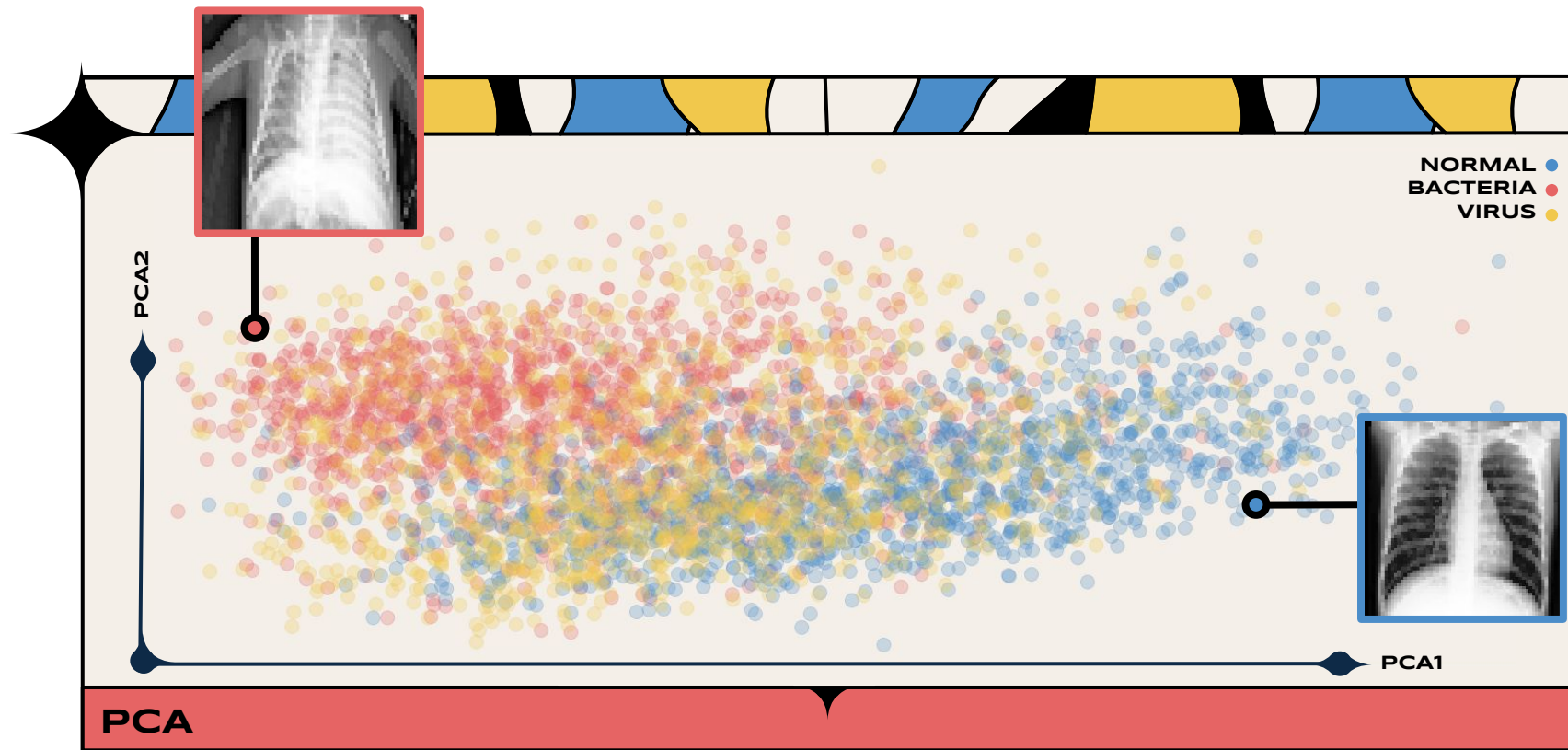
PCA

ANÁLISIS DE
COMPONENTES
PRINCIPALES

EXPLORACIÓN DE LOS DATOS

PCA1

PCA2



01. SVM

OPTIMIZAMOS LA CONSTANTE DE
REGULARIZACIÓN PARA UN KERNEL RBF

**MÁQUINAS DE SOPORTE
VECTORIAL**

		PREDICHO		
		N	B	V
VERDADERO	N	185	6	9
	B	5	156	39
	V	14	55	131

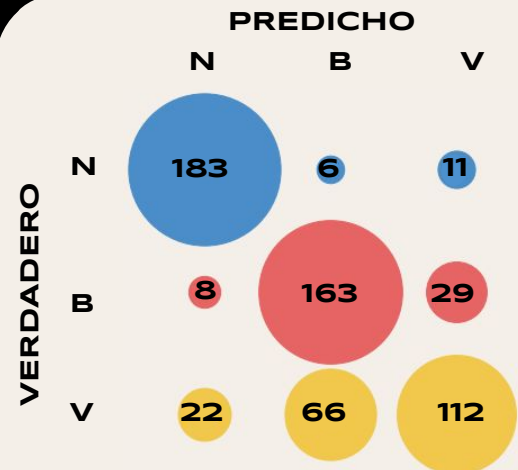
NO SE DETECTÓ EL **4,75%** DE LOS
CASOS CON NEUMONÍA

02.

RF

OPTIMIZAMOS LA PROFUNDIDAD DE LOS
ÁRBOLES Y EL TAMAÑO DEL BOSQUE

BOSQUES ALEATORIOS



NO SE DETECTÓ EL **7,5%** DE LOS CASOS
CON NEUMONÍA

03. KNN

OPTIMIZAMOS LA CANTIDAD DE
VECINOS A COMPARAR

K PRÓXIMOS VECINOS

		PREDICHO		
		N	B	V
VERDADERO	N	171	6	23
	B	9	151	40
	V	18	66	116

NO SE DETECTÓ EL **6,75%** DE LOS
CASOS CON NEUMONÍA

O4.

LR

OPTIMIZAMOS LA CONSTANTE DE
REGULARIZACIÓN DE RIDGE






REGRESIÓN LOGÍSTICA

		PREDICHO		
		N	B	V
VERDADERO	N	195	1	4
	B	29	90	81
	V	22	63	115

NO SE DETECTÓ EL **12,75%** DE LOS
CASOS CON NEUMONÍA

AUMENTO DE DATOS

DATA AUGMENTATION

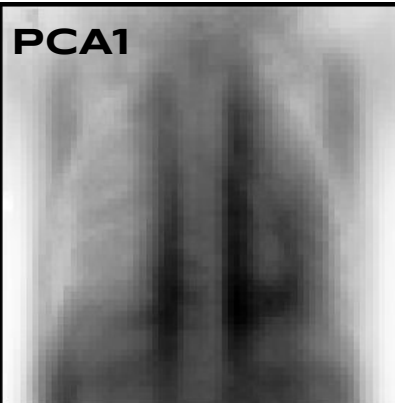
BLUR	CLAHE	NOISE	EMBOSS	SHARPEN
				
NORMAL		BACTERIA		VIRUS
1203 IMÁGENES		1200 IMÁGENES		1293 IMÁGENES

PCA

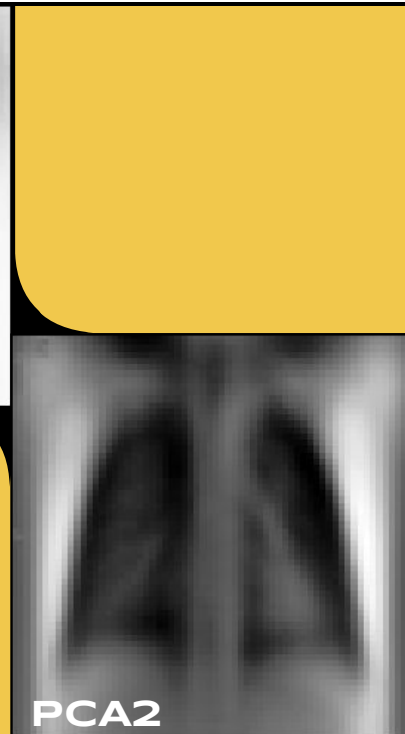
ANÁLISIS DE
COMPONENTES
PRINCIPALES

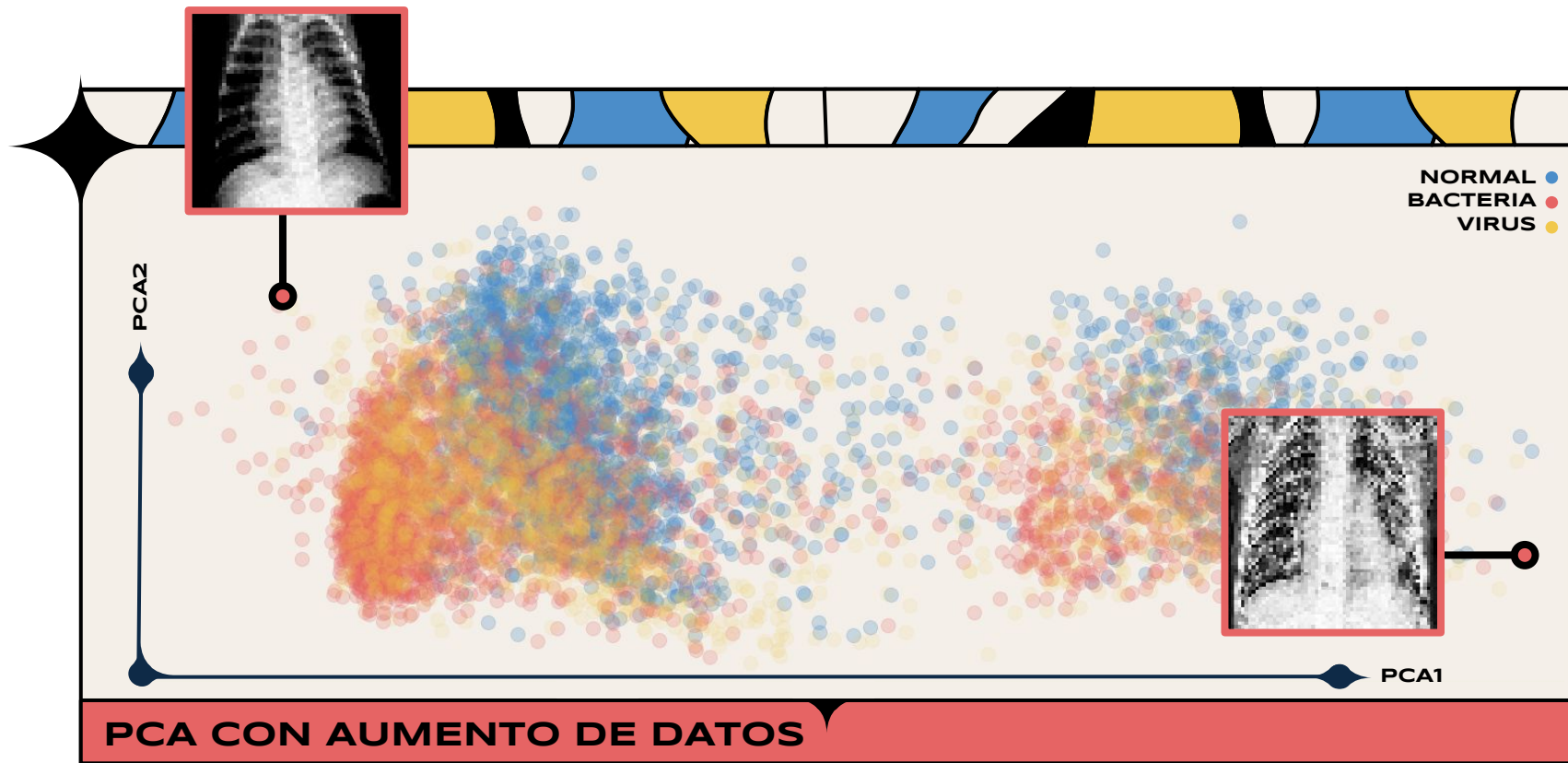
CON DATOS AUMENTADOS

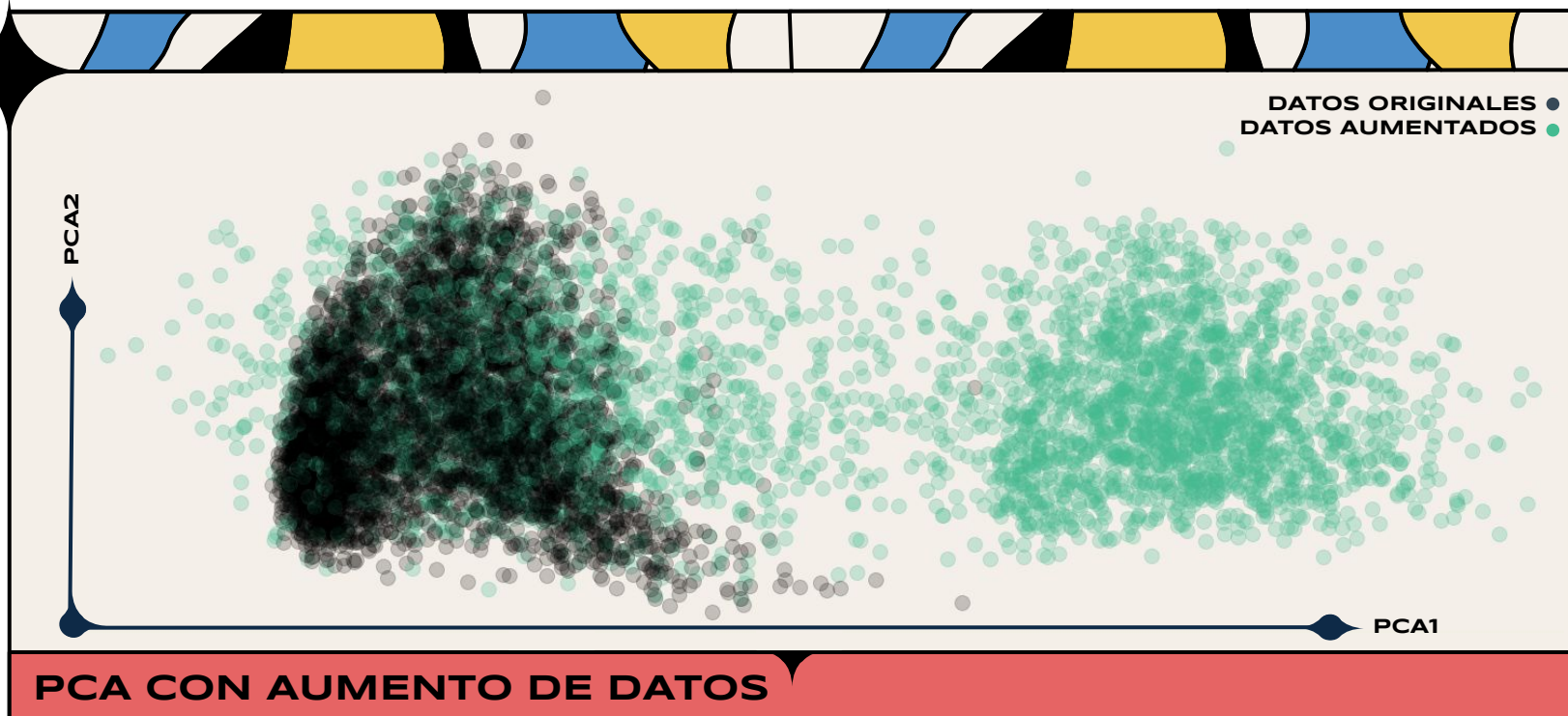
PCA1



PCA2







01.

SVM CON DATOS AUMENTADOS

OPTIMIZAMOS LA CONSTANTE DE
REGULARIZACIÓN PARA UN KERNEL RBF

**MÁQUINAS DE SOPORTE
VECTORIAL**

		PREDICHO		
		N	B	V
VERDADERO	N	187	4	9
	B	4	166	30
	V	14	57	129

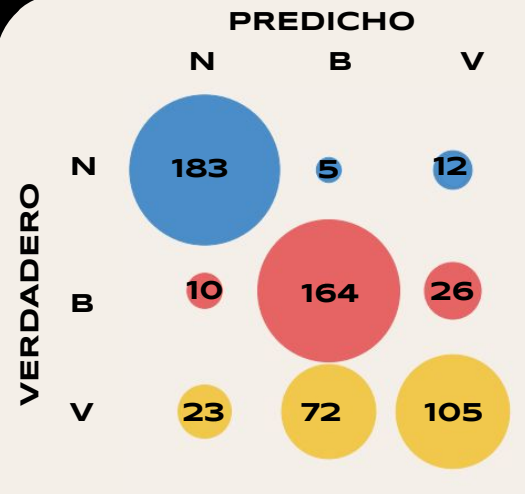
NO SE DETECTÓ EL **4,5%** DE LOS CASOS
CON NEUMONÍA

02.

RF CON DATOS AUMENTADOS

OPTIMIZAMOS LA PROFUNDIDAD DE LOS ÁRBOLES Y EL TAMAÑO DEL BOSQUE

BOSQUES ALEATORIOS



NO SE DETECTÓ EL **8,25%** DE LOS CASOS CON NEUMONÍA

03.

KNN CON DATOS AUMENTADOS

OPTIMIZAMOS LA CANTIDAD DE VECINOS A COMPARAR

K PRÓXIMOS VECINOS

		PREDICHO		
		N	B	V
VERDADERO	N	173	4	23
	B	12	151	37
	V	16	66	118

NO SE DETECTÓ EL 7% DE LOS CASOS CON NEUMONÍA

O4.

LR CON DATOS AUMENTADOS

OPTIMIZAMOS LA CONSTANTE DE
REGULARIZACIÓN DE RIDGE

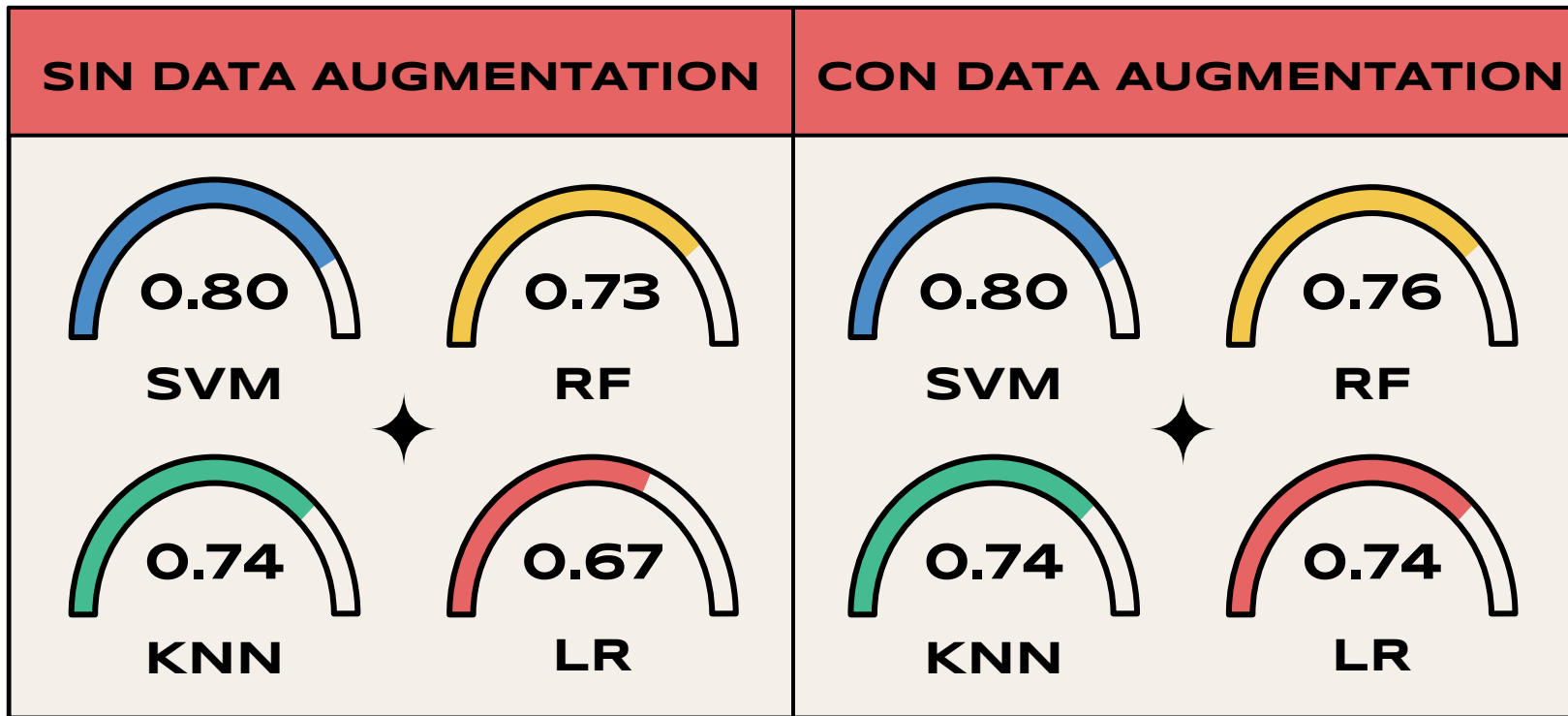
REGRESIÓN LOGÍSTICA

		PREDICHO		
		N	B	V
VERDADERO	N	183	4	13
	B	12	146	42
	V	22	62	116

NO SE DETECTÓ EL **8,5%** DE LOS CASOS
CON NEUMONÍA

EXACTITUD

ACCURACY



CONCLUSIONES Y APRENDIZAJE

- El preprocesamiento de los datos es tan importante como el modelo elegido
- Al trabajar con imágenes, hay que tener en cuenta el tiempo de cómputo
- Organizar la estructura del modelo antes de empezar a predecir
- Comprobar que las transformaciones a los datos sean efectivas
- Garantizar que los datos generados sean “parecidos” a los datos reales

PERSPECTIVAS Y MEJORAS

- Buscar una técnica de augmentation específica para estas imágenes
- Probar modelos de redes neuronales convolucionales
- Aumentar el rango etario de los datos para generalizar el modelo



¡GRACIAS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

