

PROYECTO 1: PREDICCIÓN DEL GRD CINF04 – APRENDIZAJE DE MÁQUINA

FACULTAD DE INGENIERÍA
INGENIERÍA CIVIL INFORMÁTICA



Universidad
Andrés Bello®
Conectar • Innovar • Liderar

INTEGRANTES:

- MARTÍN ALEJANDRO IBACETA HENRÍQUEZ (NRC: 8229)
- ALEJANDRO IVÁN MARCO HANNUS (NRC: 8229)
- LUIS ALBERTO NÚÑEZ ACUÑA (NRC: 8230)
- LUIS NICOLÁS NÚÑEZ CASTRO (NRC: 8230)
- FELIPE IGNACIO RODRÍGUEZ PROBOSTE (NRC: 8229)

REVISIÓN BIBLIOGRÁFICA

- EN LA ACTUALIDAD, EL AUGE DE LOS MODELOS PREDICTIVOS DE INTELIGENCIA ARTIFICIAL HA COMENZADO A FAVORECER SU APLICACIÓN EN DIVERSAS ÁREAS.
- ESPECÍFICAMENTE, EN EL SECTOR DE LA SALUD, SE ATRIBUYE UN ÍNDICE DE ADOPCIÓN POR DEBAJO DEL PROMEDIO.
- EN CONSECUENCIA, SE OBSERVA UN ELEVADO MARGEN DE ERROR EN LOS GRUPOS RELACIONADOS DE DIAGNÓSTICO (GRD), DADA UNA SOBRECARGA LABORAL, LA DEFICIENTE DOCUMENTACIÓN O LA FALTA DE HERRAMIENTAS COMPUTACIONALES DE APOYO.

REFERENCIAS:

[HTTPS://PMC.NCBI.NLM.NIH.GOV/ARTICLES/PMC8701302/](https://PMC.NCBI.NLM.NIH.GOV/ARTICLES/PMC8701302/)

[HTTPS://WWW.NATURE.COM/ARTICLES/s41746-023-00989-3](https://WWW.NATURE.COM/ARTICLES/s41746-023-00989-3)

ANÁLISIS DE DATOS

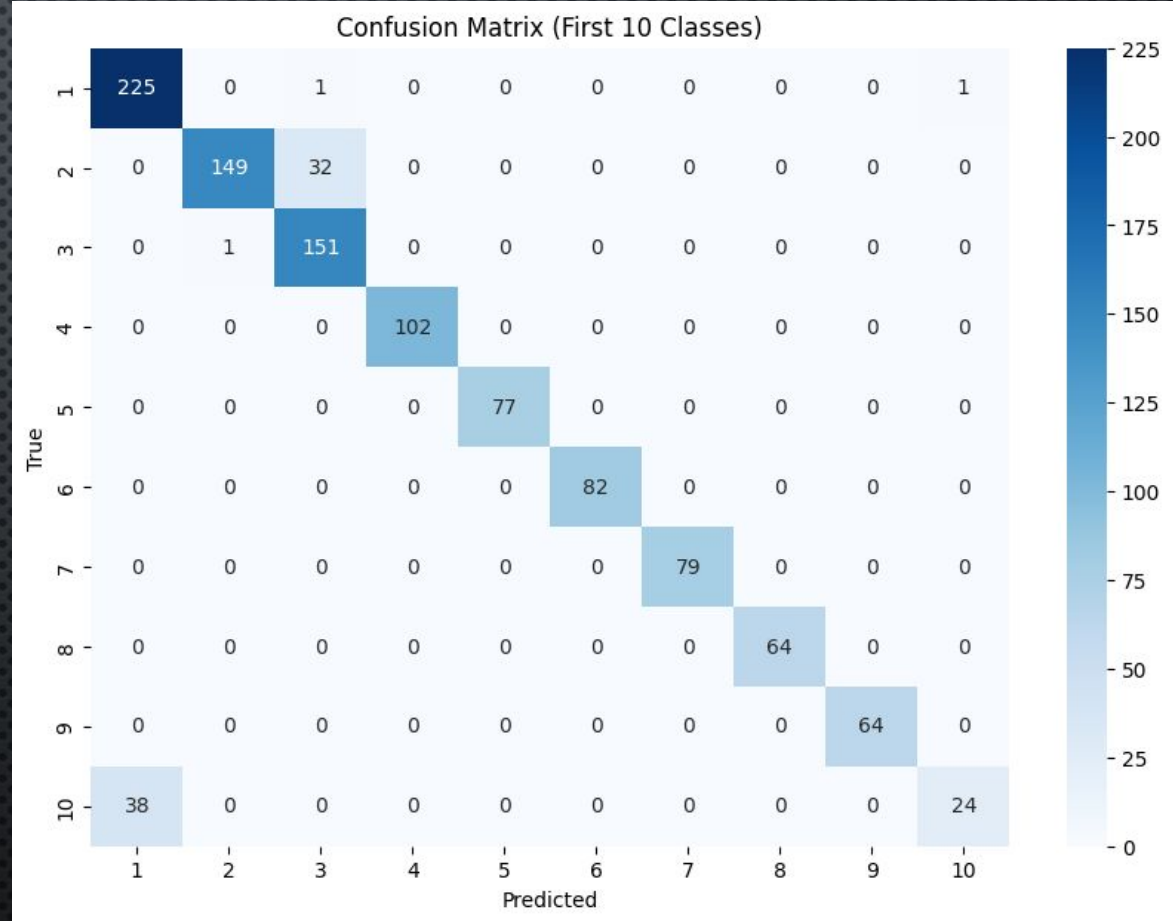
Compleitud: Altamente completa en las primeras columnas (*Diag01* a *Diag05*), decreciente después de *Diag. 06*.

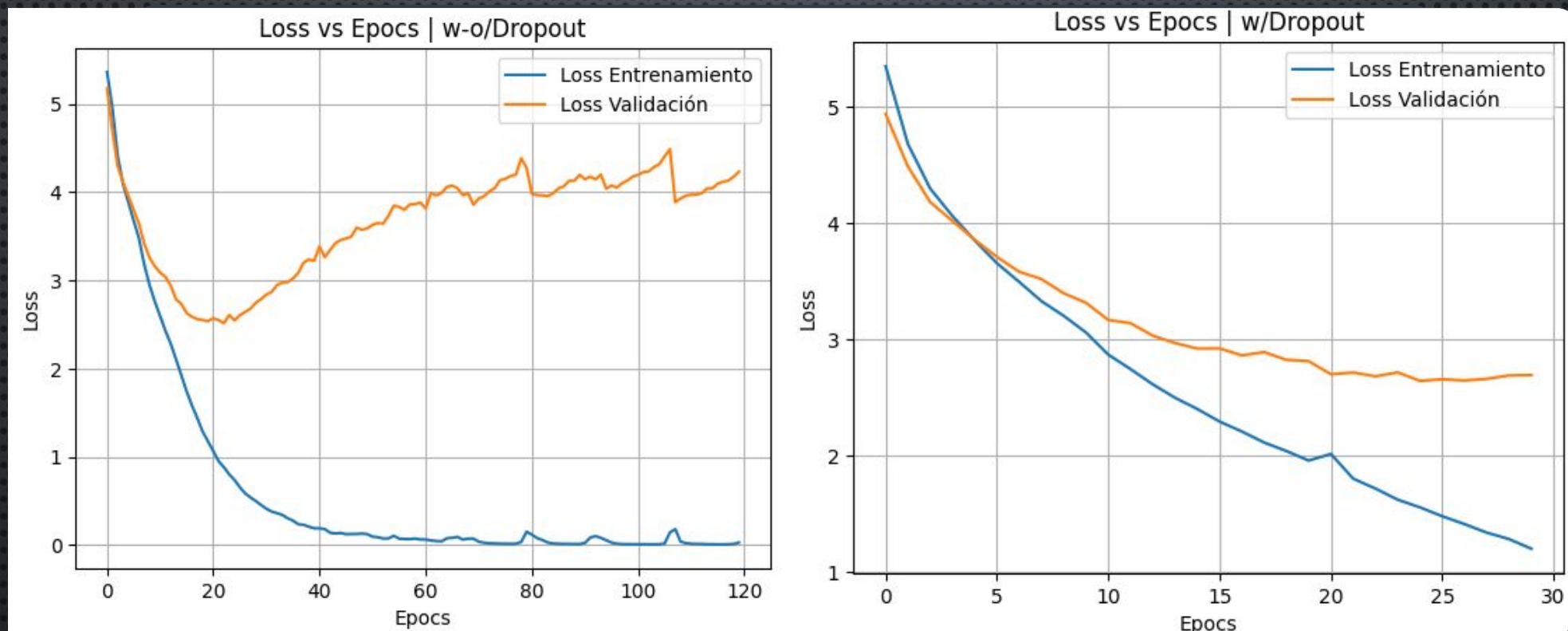
Correctitud: Generalmente buena, aunque hay muchos valores únicos que podrían investigarse si hay errores de tipeo del médico.

Outliers: Aquellos códigos que tienen una frecuencia muy baja, es decir, que aparecen muy pocas veces comparado con el resto. Esto tiene sentido porque códigos muy poco frecuentes pueden ser errores de ingreso, diagnósticos extremadamente raros, o simplemente irrelevantes para la predicción del GRD.

EXPERIMENTOS: ACCURACY, PRECISIÓN, RECALL, F1 Y MATRIZ DE CONFUSIÓN

Reporte clasificacion:					
	precision	recall	f1-score	support	
1	0.85	0.97	0.91	231	
2	0.89	0.82	0.85	182	
3	0.56	0.93	0.70	162	
4	0.69	0.84	0.76	122	
5	0.78	0.79	0.78	98	
6	0.98	0.86	0.92	95	
7	0.95	0.91	0.93	87	
8	0.98	0.98	0.98	65	
9	0.89	0.89	0.89	72	
10	0.67	0.37	0.48	65	
11	0.82	0.94	0.88	79	
12	0.87	0.88	0.88	78	
13	0.69	0.81	0.74	59	
14	0.83	0.84	0.84	64	
15	0.00	0.00	0.00	54	
16	0.80	0.73	0.76	44	
17	0.86	0.76	0.81	42	
18	0.88	0.91	0.89	46	
19	0.06	0.02	0.03	49	
20	0.53	0.70	0.60	50	
21	0.67	0.66	0.67	44	
...					
accuracy			0.56	4369	
macro avg	0.18	0.19	0.18	4369	
weighted avg	0.52	0.56	0.53	4369	





EXPERIMENTOS: GRÁFICOS LOSS VS. EPOCHS

DISCUSIÓN DE RESULTADOS

MEJOR MODELO:

Accuracy	72%
Precision	67%
Recall	21%
F1-score	18%

CONCLUSIÓN Y SIGUIENTES PASOS

A PARTIR DE LOS EXPERIMENTOS Y RESULTADOS OBTENIDOS, SE EVIDENCIA UN APRENDIZAJE Y MEJORA EN FUTURAS ITERACIONES DEL MODELO. SIN EMBARGO, SE PRESENTAN CIERTOS PROBLEMAS DURANTE EL ENTRENAMIENTO DE ESTOS, AL NO ENCONTRARSE BALANCEADOS LOS DATOS.

POSIBLES MEJORAS:

- IMPLEMENTAR DIFERENTES MODELOS PARA ENFRENTAR DE MEJOR FORMA EL DESBALANCE: RANDOMFOREST, XGBOOST Y SMOTE (OVERSAMPLING).
- AJUSTE DE LOS PESOS DE CLASE.
- UTILIZAR OTROS MÉTODOS COMO REDES NEURONALES.

GRACIAS POR SU ATENCIÓN!