



**UNIVERSIDAD  
DE ANTIOQUIA**

**CLASIFICACIÓN TEMPRANA DE GRUPOS RELACIONADOS  
POR EL DIAGNÓSTICO (GRD) UTILIZANDO TÉCNICAS DE  
INTELIGENCIA COMPUTACIONAL.**

Autor

**Angelower Santana Velásquez**

Universidad de Antioquia

Facultad de ingeniería, Departamento de Sistemas

Medellín, Colombia

2021





Clasificación temprana de Grupos Relacionados por el Diagnóstico (GRD) utilizando técnicas de inteligencia computacional.

**Angelower Santana Velásquez**

Trabajo de investigación presentado como requisito parcial para optar al título de:  
Magister en Ingeniería

Director:  
PhD. John Freddy Duitama Muñoz

Línea de Investigación:  
Machine Learning aplicado a Grupos Relacionados por el Diagnóstico

Grupo de Investigación:  
Intelligent Information Systems Lab - In2Lab

Universidad de Antioquia  
Facultad de Ingeniería, Departamento de Sistemas.  
Medellín, Colombia  
2021.



*A mi familia, por toda su paciencia y apoyo.*



# **Agradecimientos**

Agradezco especialmente a la IPS Universitaria Clínica León XIII de la ciudad de Medellín Colombia por proporcionar su Conjunto Mínimo Básico de Datos y por poner a disposición su personal del área de Gestión de la Información para la realización de este proyecto de investigación.



## Resumen

La optimización de los recursos usados en clínicas y hospitales es un problema clave de la gestión hospitalaria. En particular, el cómo mejorar la eficiencia en los procedimientos y tratamientos a los pacientes, disminuyendo costos, pero sin desmejorar la calidad de la estancia del paciente es uno de sus mayores retos.

Desde los años 60 existe un método de clasificación llamado los Grupos Relacionados por el Diagnóstico (GRD), el cual consiste en clasificar a los pacientes hospitalizados según su nivel de consumo de recursos. Para dicha clasificación se han diseñado softwares comerciales llamados *DRG-grouper* que ubican a los pacientes en GRD con base en información demográfica, de diagnóstico y procedimientos según los estándares internacionales. Esta clasificación se realiza después de dar de alta al paciente, lo que implica que se hace una clasificación a posteriori y un análisis retrospectivo de su estancia: puesto que los recursos ya fueron invertidos en el paciente, si se presenta algún sobrecosto, el hospital o clínica lo detectará en forma tardía, cuando los recursos y servicios ya fueron consumidos.

En el presente trabajo de investigación se propone implementar una metodología alternativa, utilizar técnicas de la Inteligencia computacional, que permitan hacer una clasificación temprana en GRD de pacientes hospitalizados; de tal forma que el médico tratante pueda usar dicha clasificación como apoyo para la toma de decisiones en cuanto a los procedimientos a seguir con los pacientes hospitalizados y que desde el punto de vista administrativo, se puedan proyectar y programar los gastos relacionados con la clasificación de cada paciente. Esta clasificación temprana permitirá a la administración controlar el consumo de los recursos hospitalarios en tiempo real, es decir, durante la estancia del paciente. Para la construcción de la metodología se contó con un conjunto de registros de egresos de pacientes de 3 años y su respectiva clasificación en GRD con

el *DRG-grouper* de la IPS Universitaria Clínica León XIII de la ciudad de Medellín-Colombia.

La contribución que tendrá este trabajo a los servicios de salud es que, a través de la clasificación temprana de los pacientes en los GRD, se puede mejorar la planeación en el uso de recursos y proyectar costos en procedimientos de pacientes.

**Palabras clave:** Grupos Relacionados por el Diagnóstico, GRD, costos de pacientes, “*Case mix*”, Inteligencia computacional, Machine Learning, Clasificación temprana, Predicción, Redes Neuronales Artificiales, RNA, Extreme Gradient Boosting, XGBoost

## Abstract

The optimization of the resources used in clinics and hospitals is a key problem in hospital management. In particular, how to improve the efficiency in procedures and treatments for patients, reducing cost, but without deteriorating the quality of the patient's stay is one of his greatest challenges. In other words, ensure that the patient can be discharged at the times that describe the standard.

Since the 1960s has been used a classification method called Diagnosis Related Groups (DRG), which consists in the classifying of hospitalized patients according to their level of resource consumption. For this classification, commercial software called DRG-grouper has been designed that locate patients in DRG based on demographic, diagnostic and procedural information according to international standards. This classification is done after the patient is discharged, which implies that there is a posteriori classification and a retrospective analysis of their stay: since the resources were already invested in the patient, if there were any over-costs, the hospital or clinic will detect it late, when the resources and services have already been consumed.

In this research work, it is proposed to implement an alternative methodology, use machine learning techniques, which can be used to do an early classification in DRG of hospitalized patients; so that the treating physician can use this classification as support for decision making regarding the procedures to be followed with hospitalized patients and that from the administrative point of view, the expenses related to the classification of patients can be projected and programmed. This early classification will allow the administration to control the consumption of hospital resources in real time, that is, during the patient's stay. For the construction of the methodology, there is a set of patient discharge records of 3 years and their respective classification in DRG with the DRG-grouper from the “*IPS Universitaria Clínica León XIII*” in Medellín- Colombia.

The contribution that this work will have to the health services is, through the early classification of the patients in the DRGs, it is possible to improve the planning in the use of the resources and schedule costs in the patient's procedures.

**Keywords: Diagnosis-Related Groups, DRG, Patient costs, "Case mix", Machine Learning, Early classification, Prediction, Artificial Neural Networks, ANN, Extreme Gradient Boosting, XGBoost.**

# Contenido

	<b>Pág.</b>
<b>Resumen</b> .....	<b>IX</b>
<b>Abstract</b> .....	<b>XI</b>
<b>Lista de figuras</b> .....	<b>XVI</b>
<b>Lista de tablas</b> .....	<b>XVII</b>
<b>Lista de gráficas</b> .....	<b>XIX</b>
<b>Lista de abreviaturas</b> .....	<b>XX</b>
<b>Capítulo 1: Problema de investigación, objetivos del trabajo y metodología seguida.</b>	
<hr/>	
1.1. Introducción. ....	22
1.2. Planteamiento del problema. ....	23
1.3. Preguntas de investigación. ....	26
1.3.1 ¿Qué clasificadores de GRD existen? .....	26
1.3.2 ¿En cuál etapa de la estancia hospitalaria nos vamos a enfocar para la clasificación temprana (predicción) de los GRD? ¿Cuál es el tipo de paciente que va a ser de interés para este proyecto de investigación? .....	26
1.4. Objetivos. ....	27
1.4.1. General. ....	27
1.4.2. Específicos. ....	27
1.5. Alcance de la solución. ....	27
1.6. Metodología usada. ....	29
<b>Capítulo 2: Marco teórico</b>	
<hr/>	
2.1. Introducción. ....	32
2.2. Grupos Relacionados por el Diagnóstico – GRD. ....	33

2.2.1. Clasificación de pacientes en GRD - El método tradicional. ....	33
2.2.2. Contexto global. ....	35
2.2.3. Contexto Latinoamericano. ....	39
2.3. Machine Learning. ....	42
2.3.1. Red Neuronal Artificial - RNA. ....	43
2.3.2. Extreme Gradient Boosting - XGBoost. ....	44
2.4. Conclusiones. ....	45

### Capítulo 3: Estado del arte

3.1. Introducción. ....	46
3.2. Revisión bibliográfica. ....	47
3.3. Machine Learning para clasificación de pacientes en GRD. ....	48
3.4. Machine Learning para predicción de pacientes en GRD. ....	49
3.4.1. Gartner et al. (2015) "Machine learning approaches for Early DRG Classification and Resource Allocation". ....	52
3.4.2. Gartner et al. (2015) "Improving Hospital-Wide Early Resource Allocation through Machine Learning". ....	58
3.4.3. Nakai et al. (2016) "Prediction of Clinical Practices by Clinical Data of the Previous Day Using Linear Support Vector Machine". ....	60
3.4.4. Krämer et al. (2017) "Classification of hospital admissions into emergency and elective care: a machine learning approach". ....	63
3.4.5. Okamoto et al. (2018) "Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days". ....	67
3.4.6. Kuo et al. (2018) "Comparison of Models for the Prediction of Medical Costs of Spinal Fusion in Taiwan Diagnosis-Related Groups by Machine Learning Algorithms". ....	71
3.5. Conclusiones. ....	74

### Capítulo 4: Base de datos

4.1. Introducción. ....	77
4.2. Esquemas de codificación del CMBD. ....	78
4.2.1. Codificación CIE-10. ....	78
4.2.2. Codificación CUPS. ....	80
4.3. Descripción del Conjunto Mínimo Básico de Datos. ....	82
4.3.1. Información demográfica. ....	85
4.3.2. Información de diagnóstico. ....	87
4.3.3. Información de procedimientos. ....	90
4.4. Codificación de características. ....	91
4.5. Cohorte de estudio. ....	92
4.6. Conclusiones. ....	96

## Capítulo 5: Investigación Experimental.

5.1. Introducción. ....	98
5.2. Proceso clínico de admisión y evolución de pacientes. ....	98
5.3. Estrategia de evaluación temprana. ....	102
5.4. Métricas empleadas para el análisis. ....	104
5.5. Entrenamiento y prueba. ....	105
5.6. Técnicas usadas en la investigación experimental. ....	108
5.6.1. Red Neuronal Artificial – RNAs. ....	108
5.6.1.1. Manejo de desbalanceo de las muestras usando la RNA. ....	110
5.6.1.2. Ajuste de parámetros e hiperparámetros de la RNA. ....	112
5.6.1.3. Análisis de sensibilidad del modelo de la RNA.....	113
5.6.1.4. Ejecución de experimentos con la RNA.....	115
5.6.2. Extreme Gradient Boosting – XGBoost. ....	116
5.6.2.1. Manejo de desbalanceo de las muestras usando el XGBoost...	117
5.6.2.2. Ajuste de parámetros e hiperparámetros del XGBoost. ....	120
5.6.2.3. Análisis de sensibilidad del modelo del XGBoost.....	121
5.6.2.4. Ejecución de experimentos con el XGBoost.....	123
5.7. Evaluación de los modelos propuestos en la metodología de predicción. ....	124
5.7.1. Evaluación de modelo M0. ....	124
5.7.2. Evaluación de los modelos M24 al M192. ....	133
5.7.3. Comparación de las técnicas RNA y XGBoost. ....	140
5.8. Importancia de características. ....	143
5.9. Conclusiones. ....	147

## Capítulo 6: Resumen, conclusiones, observaciones y estudios futuros.

6.1. Resumen. ....	150
6.2. Observaciones finales. ....	154
6.3. Limitaciones y estudios futuros. ....	155

<b>7. Referencias bibliográficas. ....</b>	<b>157</b>
<b>A. Anexo A.</b> Matrices de confusión de los diez modelos con la técnica RNA. ....	<b>163</b>
<b>B. Anexo B.</b> Matrices de confusión de los diez modelos con la técnica XGBoost. ....	<b>173</b>
<b>C. Anexo C.</b> Importancia de características en XGBoost. ....	<b>183</b>

## Lista de figuras

	<b>Pág.</b>
<b>Figura 1-1:</b> Diagrama de flujo de la metodología propuesta. ....	29
<b>Figura 2-1:</b> Diagrama de flujo del algoritmo tradicional de clasificación de pacientes en GRD. ....	34
<b>Figura 3-1:</b> Concepto del modelo de predicción. ....	60
<b>Figura 4-1:</b> Estructura del código CIE-10. ....	79
<b>Figura 4-2:</b> Estructura del código CUPS. ....	81
<b>Figura 4-3:</b> Estructura del código GRD. ....	95
<b>Figura 5-1:</b> Flujograma ingreso y hospitalización de paciente. ....	101
<b>Figura 5-2:</b> Metodología propuesta para predicción de GRD. ....	103
<b>Figura 5-3:</b> Arquitectura de Red Neuronal Artificial usada en el experimento. ...	110

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 3-1:</b> Estudios que usan inteligencia computacional para la clasificación temprana de GRD. ....	50
<b>Tabla 3-2:</b> Distribución de los subconjuntos de datos. ....	53
<b>Tabla 3-3:</b> Precisión general (%) de las diferentes técnicas de clasificación antes y durante de la admisión y antes de la selección de atributos. .	54
<b>Tabla 3-4:</b> Precisión general (%) de las diferentes técnicas de clasificación antes y durante de la admisión y después de la selección de atributos. ....	55
<b>Tabla 3-5:</b> Comparación del rendimiento de los diferentes modelos de predicción. ....	57
<b>Tabla 3-6:</b> Ejemplos de categorías de datos. ....	61
<b>Tabla 3-7:</b> Métricas de desempeño por cada <i>folds</i> y general. ....	62
<b>Tabla 3-8:</b> Ejemplo de precisión y <i>recall</i> de infusión intravenosa. ....	62
<b>Tabla 3-9:</b> Ejemplo de resultados obtenidos en la clasificación. ....	66
<b>Tabla 3-10:</b> Códigos DPC y el número de casos. ....	68
<b>Tabla 3-11:</b> Medidas de precisión del experimento 1. ....	69
<b>Tabla 3-12:</b> Comparación del rendimiento de los diferentes modelos de predicción. ....	73
<b>Tabla 4-1:</b> Capítulos en que se divide el CIE-10. ....	79
<b>Tabla 4-2:</b> Ejemplos de códigos CUPS por sección. ....	82
<b>Tabla 4-3:</b> Tamaño de los archivos de CMBD. ....	83
<b>Tabla 4-4:</b> Descripción de atributos resultantes. ....	84
<b>Tabla 4-5:</b> Top 30 de los diagnósticos más frecuentes. ....	88
<b>Tabla 4-6:</b> 47 etiquetas de GRD más frecuentes en el 70% de los datos de 2016 a 2018. ....	93
<b>Tabla 5-1:</b> tamaño de entradas por modelos evaluado. ....	104
<b>Tabla 5-2:</b> Relación GRD estratificados para entrenamiento y prueba. ....	106

<b>Tabla 5-3:</b> pesos $w$ asignados a las 47 clases en la etapa de entrenamiento ( <i>train</i> en la RNA).....	111
<b>Tabla 5-4:</b> pesos $w$ asignados a las 47 clases en la etapa de prueba ( <i>test</i> en la RNA).....	112
<b>Tabla 5-5:</b> Malla de evaluación de parámetros e hiperparámetros de la RNA. ....	112
<b>Tabla 5-6:</b> Parámetros e hiperparámetros determinados de la RNA. ....	113
<b>Tabla 5-7:</b> Arquitecturas para medición de sensibilidad de la RNA. ....	114
<b>Tabla 5-8:</b> Tiempo entrenamiento por modelo – RNA. ....	116
<b>Tabla 5-9:</b> pesos $w$ asignados a las 47 clases en la etapa de entrenamiento ( <i>train</i> en el XGBoost). ....	118
<b>Tabla 5-10:</b> pesos $w$ asignados a las 47 clases en la etapa de prueba ( <i>test</i> en el XGBoost). ....	119
<b>Tabla 5-11:</b> Comparativo de accuracy la RNA en entrenamiento y prueba con pesos de la RNA y el XGBoost.....	119
<b>Tabla 5-12:</b> Comparativo de accuracy del XGBoost en entrenamiento y prueba con pesos del XGBoost y la RNA.....	120
<b>Tabla 5-13:</b> Malla de evaluación de parámetros e hiperparámetros del XGBoost. ....	120
<b>Tabla 5-14:</b> Parámetros e hiperparámetros determinados para el XGBoost. ....	121
<b>Tabla 5-15:</b> Arquitecturas para medición de sensibilidad de la XGBoost. ....	122
<b>Tabla 5-16:</b> Tiempo entrenamiento por modelo – XGBoost. ....	124
<b>Tabla 5-17:</b> Precisión de la RNA y el XGBoost en el modelo M0.....	125
<b>Tabla 5-18:</b> Códigos de diagnóstico comunes para los GRD C-10-15 (clase 11) y P-10-12 (clase 12). ....	129
<b>Tabla 5-19:</b> Códigos de diagnóstico asociados a la clase 46 (DRG C-10-14) que representan el 60%. ....	132
<b>Tabla 5-20:</b> <i>F1-score</i> por modelo por clase en la RNA.....	133
<b>Tabla 5-21:</b> <i>F1-score</i> por modelo por clase en el XGBoost.....	134
<b>Tabla 5-22:</b> Diagnósticos relacionados con el GRD C-08-15 (clase 33). ....	137
<b>Tabla 5-23:</b> Diagnósticos relacionados con el GRD C-09-09 (clase 38). ....	139
<b>Tabla 5-24:</b> Comparativo de mejore <i>F1-score</i> entre la RNA y el XGBoost.....	142
<b>Tabla 5-25:</b> Procedimientos más importantes para clasificación según el XGBoost. ....	146

## Lista de gráficas

	<b>Pág.</b>
<b>Gráfica 3-1:</b> Precisión de clasificación temporal para diferentes enfoques antes de la selección de atributos y optimización de parámetros. ....	56
<b>Gráfica 3-2:</b> La precisión de los 4 modelos para la selección de los códigos DPC en los diferentes escenarios planteados en relación con el # de días de hospitalización de los que se utilizaron los datos. ....	71
<b>Gráfica 4-1:</b> Relación de número de ingresos por cada mes del año. ....	85
<b>Gráfica 4-2:</b> Frecuencia de edad de pacientes. ....	86
<b>Gráfica 4-3:</b> Número de pacientes por rango de edades. ....	87
<b>Gráfica 4-4:</b> Frecuencia de los MCD. ....	90
<b>Gráfica 4-5:</b> Frecuencia de los procedimientos CUPS. ....	91
<b>Gráfica 4-6:</b> Gráfica de frecuencia de GRD. ....	96
<b>Gráfica 5-1:</b> Análisis de Sensibilidad del modelo MBL usando la RNA.....	115
<b>Gráfica 5-2:</b> Análisis de Sensibilidad del modelo MBL usando el XGBoost. ....	123
<b>Gráfica 5-3:</b> Métrica <i>F1-score</i> de la RNA y el XGBoost para todos los GRD. ....	126
<b>Gráfica 5-4:</b> Matriz de confusión del M0 con la RNA. ....	128
<b>Gráfica 5-5:</b> Frecuencia de Diagnósticos asociados a los GRD C-10-15 y P-10-12. ....	129
<b>Gráfica 5-6:</b> Gráfica de frecuencia de diagnósticos la clase 0 (GRD C-09-07)...	131
<b>Gráfica 5-7:</b> Gráfica de frecuencia de diagnósticos la clase 46 (GRD C-10-14)...	132
<b>Gráfica 5-8:</b> Gráfica de frecuencia de diagnósticos la clase 13 (GRD C-08-15)...	138
<b>Gráfica 5-9:</b> Gráfica de frecuencia de diagnósticos la clase 38 (GRD C-09-09)...	139
<b>Gráfica 5-10:</b> Precisión por modelo en las etapas de entrenamiento y prueba para la RNA y el XGBoost. ....	141
<b>Gráfica 5-11:</b> Top 30 de Importancia de características en el modelo M0. ....	144
<b>Gráfica 5-12:</b> Top 30 de Importancia de características en el modelo M24. ....	145

## Lista de abreviaturas

Abreviatura	Término
CIE-9	Clasificación Internacional de Enfermedades, 9. <sup>a</sup> edición
CIE-10	Clasificación Internacional de Enfermedades, 10. <sup>a</sup> edición
CMBD	Conjunto Mínimo Básico de Datos
CUPS	Clasificación Única de Procedimientos en Salud
DPC	<i>Diagnosis Procedure Combination</i> (Combinación de Procedimientos de Diagnóstico)
EPS	Entidades Promotoras de Salud
GRD	Grupos Relacionados por el Diagnóstico
IPS	Institución Prestadoras de Servicios
MDC	<i>Major Diagnostic Categories</i> (Categorías de Diagnóstico Principal)
ML	<i>Machine Learning</i>
OCDE	Organización para la Cooperación y el Desarrollo Económicos
PDPS	<i>Per-Diem Payment System</i> (Sistema de pago por día)
RNA	Red Neuronal Artificial
SVM	<i>Support Vector Machine</i> (Máquinas de Soporte Vectorial)
UCI	Unidad de Cuidados Intensivos
UCE	Unidad de Cuidados Especiales
XGBoost	eXtreme Gradient Boosting



# **Capítulo 1:**

## **Problema de investigación, objetivos del trabajo y metodología seguida.**

### **1.1. Introducción.**

El sector hospitalario ostenta un gran peso en el conjunto del sistema de salud, no sólo por el gran consumo de recursos que genera, sino por su contribución al desarrollo y utilización de nuevas tecnologías y a la mejora de la cualificación profesional. El conocimiento de la actividad hospitalaria se constituye como elemento de gran trascendencia en el desarrollo de la política sanitaria global. [1] Uno de estos elementos son los sistemas de clasificación de pacientes que abarcan todos aquellos sistemas que son utilizados en la gestión de salud para mejorar la calidad de la asistencia, disminuir los costos y elaborar guías y protocolos asistenciales, todos ellos basados en la clasificación de pacientes.[2]

El sistema más utilizado es el sistema “Grupos Relacionados por el Diagnóstico” (GRD), que se utiliza para minimizar la incertidumbre que se genera en la asistencia sanitaria, lo que se logra a través de la clasificación de pacientes como una manera de medir el consumo de recursos y la gestión clínica.[3]

En este capítulo se describe la problemática bajo la cual se planteó este proyecto de investigación, se hace una descripción del uso de los Grupos Relacionados por el

Diagnóstico en el mundo y el impacto que estos tienen en la gestión hospitalaria. Se plantean algunas preguntas que fueron el punto de partida en la investigación y las respuestas que se dieron a medida que se iba avanzando en la solución del problema. Posteriormente, se describen también los objetivos propuestos en este proyecto y el alcance de la solución planteada, es decir, los límites del trabajo, los productos a entregar como resultado de la investigación.

Finalmente, se describe la metodología que se siguió para llevar a cabo el cumplimiento de los objetivos descritos en este mismo capítulo. En esta sección se hace una descripción de las actividades a realizar durante todas las etapas planteadas y que llevan al cumplimiento de los objetivos específicos.

## **1.2. Planteamiento del problema.**

Las presiones por parte de las Entidades Promotoras de Salud (EPS) , los usuarios y el mismo gobierno hacia las Instituciones Prestadoras de Servicios (IPS) están enfocadas en la optimización de los recursos económicos sin bajar la calidad en los servicios prestados, mejorar los servicios actuales y ampliar el portafolio de servicios prestados al público en general, ya sean que pertenezcan al régimen subsidiado o contributivo, y han hecho que las clínicas y hospitales adopten técnicas que le permitan estandarizar procedimientos y optimizar los costos de los pacientes que ingresan a hospitalización.

Bajo este contexto aparecen los GRD como un método de agrupamiento de pacientes basado en la información demográfica, en las características significativas de los diagnósticos y en los procedimientos que se realizan durante toda la hospitalización y que se revelan luego de darlo de alta. Los GRD permiten determinar el nivel de recursos que consumió el paciente con uno o varios diagnósticos específicos y acorde a los tratamientos aplicados hasta el momento del alta. [4][5]

La creación de los GRD como herramienta ha tenido tanto éxito que se ha generalizado su uso; así como se reporta en una serie de artículos que relacionan países europeos, Estados Unidos, Canadá, Australia, entre otros, que en las últimas décadas han usado los GRD como sistema de reembolso por prestación de servicios sanitarios [7] [8] [9] [10] [11] [12] [13] [14]. Actualmente, los datos de salud que describen los fenotipos y los tratamientos de pacientes abarcan múltiples fuentes, incluyendo medicamentos, pruebas de laboratorio, imágenes diagnósticas y datos narrativos [15]. Si se evalúan todas las diferentes opciones que se pueden generar con las variables y sus posibles valores para clasificar un paciente en un GRD, se tienen: 1) cinco grupos etarios (neonatos, niños, adolescentes, adultos, adultos mayores), 2) 8.000 diagnósticos principales (códigos CIE), 3) dos posibles tratamientos: médico (dos opciones) o quirúrgico (tres opciones) y 4) las posibles comorbilidades (pueden llegar a ser cinco), lo que da como resultados 400 mil posibles combinaciones para asignar un GRD a un paciente, lo que implica que la información se hace cada vez más difícil de analizar utilizando el algoritmo tradicional para la clasificación de los GRD. [6]

En un sistema de clasificación de pacientes basado en GRD, implementado en un software llamado *DRG-grouper*, los pacientes son clasificados según sus características clínicas y semejanza en los recursos empleados para su tratamiento durante la estancia hospitalaria; posterior a ser dado de alta, las empresas aseguradoras hacen pagos fijos según la asignación que se haya hecho del GRD para cada paciente. Bajo este contexto, la veracidad de la información se vuelve un papel determinante para una eficaz planeación de los servicios de atención en salud; sin embargo, en las primeras etapas de la estancia hospitalaria, la clasificación del paciente en un GRD no es tan clara debido a faltantes de información, ingreso de información por parte de médico en texto libre no estandarizado, información no parametrizada, etc.[16], lo que convierte este proceso de clasificación en un reto para la gestión de operaciones y recursos en hospitales y clínicas.[17]

Una clasificación posterior al alta del paciente implica que no se puede monitorear y controlar el consumo de recursos durante su estancia, sino que se hace un análisis del

tiempo y los recursos consumidos de más por el paciente, si es el caso, cuando el paciente ya ha salido del hospital. Lo anterior impacta directamente el desempeño de las áreas del hospital, ya que los análisis a posteriori no dan la posibilidad de mejorar sobre la marcha. Es decir, las correcciones, acciones correctivas o de mejora solo se pueden identificar posterior a la salida del paciente. Como alternativa de solución, algunos estudios han comenzado a evaluar la utilización de técnicas de inteligencia computacional que tiene como propósito lograr una clasificación temprana de los pacientes en los GRD [17], es decir, mientras está hospitalizado y no cuando ya se le ha dado de alta.

El presente trabajo parte de la hipótesis que es posible implementar una metodología, utilizando técnicas de Inteligencia computacional, que permita hacer una clasificación temprana de los pacientes hospitalizados en GRD. El propósito es analizar diferentes metodologías de clasificación temprana (predicción) de pacientes en GRD que estén hospitalizados en una institución prestadora de servicios de salud de tercer nivel en Colombia, hasta encontrar aquella que logre un mayor nivel de precisión en la predicción del GRD del paciente. Para lograrlo, en primera instancia, se identificó la población objetivo del estudio acorde a la base de datos de egresos disponible y a las patologías que mayor impacto tuvieron en la frecuencia de ocurrencia en el hospital en el periodo evaluado. Igualmente, el desarrollo del presente trabajo permitió establecer los momentos de la estancia del paciente en que se puede realizar la predicción, de tal manera que se puedan usar sus resultados para proyectar y programar los gastos relacionados con la atención del paciente, con el propósito de lograr mejorar la administración de los recursos hospitalarios. Es de anotar que la clasificación temprana obtenida sólo será un punto de referencia para el médico tratante en el momento de definir los procedimientos acordes al estándar internacional CIE-9 para los pacientes, lo que implica que esta metodología resultante no reemplazará el criterio médico, pero sí puede convertirse en una guía para la gestión administrativa o en una herramienta de apoyo clínico.

### **1.3. Preguntas de investigación.**

Durante el planteamiento del problema y las posibles soluciones a proponer, surgieron algunas preguntas que debían ser contestadas antes de iniciar el proceso de investigación, las cuales ilustraremos en esta sección.

#### **1.3.1 ¿Qué clasificadores de GRD existen?**

En la actualidad existen software llamados “*DRG-grouper*”, que implementan un método simple basado en el diagrama de flujo [17]. Estos software pueden ser comerciales o de libre acceso como los usados en el artículo *Machine Learning Approaches for Early DRG Classification and Resource Allocation*, que usaron los agrupadores de libre acceso GetDRG (<http://www.geosmbh.de/>) y Webgrouper (<https://grouper.swissdrq.org/>) y compararon los resultados con el *DRG-grouper* comercial que el hospital caso de estudio estaba usando durante la investigación.

#### **1.3.2 ¿En cuál etapa de la estancia hospitalaria nos vamos a enfocar para la clasificación temprana (predicción) de los GRD? ¿Cuál es el tipo de paciente que va a ser de interés para este proyecto de investigación?**

Dentro de los procesos de levantamiento de información de este trabajo de investigación, se hizo una revisión de la disponibilidad de la información durante la estancia del paciente hospitalizado en la IPS caso de estudio. Se visitaron las áreas directamente implicadas en el proceso de atención desde el ingreso hasta el alta del paciente. Se identificaron dos aspectos: 1) debido al proceso de admisión del paciente y su evolución clínica por parte del médico tratante, se definió que la clasificación temprana del paciente en un GRD se iba a realizar desde el momento de la admisión hospitalaria y cada 24 horas

hasta que el paciente reciba la alta médica y 2) el paciente de interés para este trabajo de investigación sería el paciente hospitalizado, este tipo de paciente fue seleccionado según el criterio que tiene el hospital caso de estudio.

## **1.4. Objetivos.**

### **1.4.1. General.**

Definir y evaluar una metodología de clasificación temprana de GRD en pacientes hospitalizados en una institución prestadora de servicios de salud de tercer nivel en Medellín-Colombia.

### **1.4.1. Específicos.**

- Seleccionar la población objetivo y los GRD representativos que evidencien el mayor impacto en costos y la mayor frecuencia de ocurrencia entre los pacientes hospitalizados.
- Identificar el conjunto de variables que más aporten a la clasificación temprana de GRD desde el conjunto de datos de registros de egresos seleccionados.
- Proponer y evaluar una estrategia de clasificación temprana de GRD basada en técnicas de Inteligencia Computacional.
- Validar la metodología de clasificación temprana de GRD a partir de criterios de exactitud y desempeño, usando información real de una institución prestadora de servicios de salud de tercer nivel en Medellín- Colombia.

## **1.5. Alcance de la solución.**

La clasificación temprana o predicción de pacientes hospitalizados en GRD en una institución prestadora de servicios de salud, con varios escenarios de disponibilidad de información y en diferentes tiempos de la estancia hospitalaria, permite a las clínicas u

hospitales identificar el consumo de recursos que se debe proyectar para cada caso de pacientes y monitorear en tiempo real los días de estancia hospitalaria.

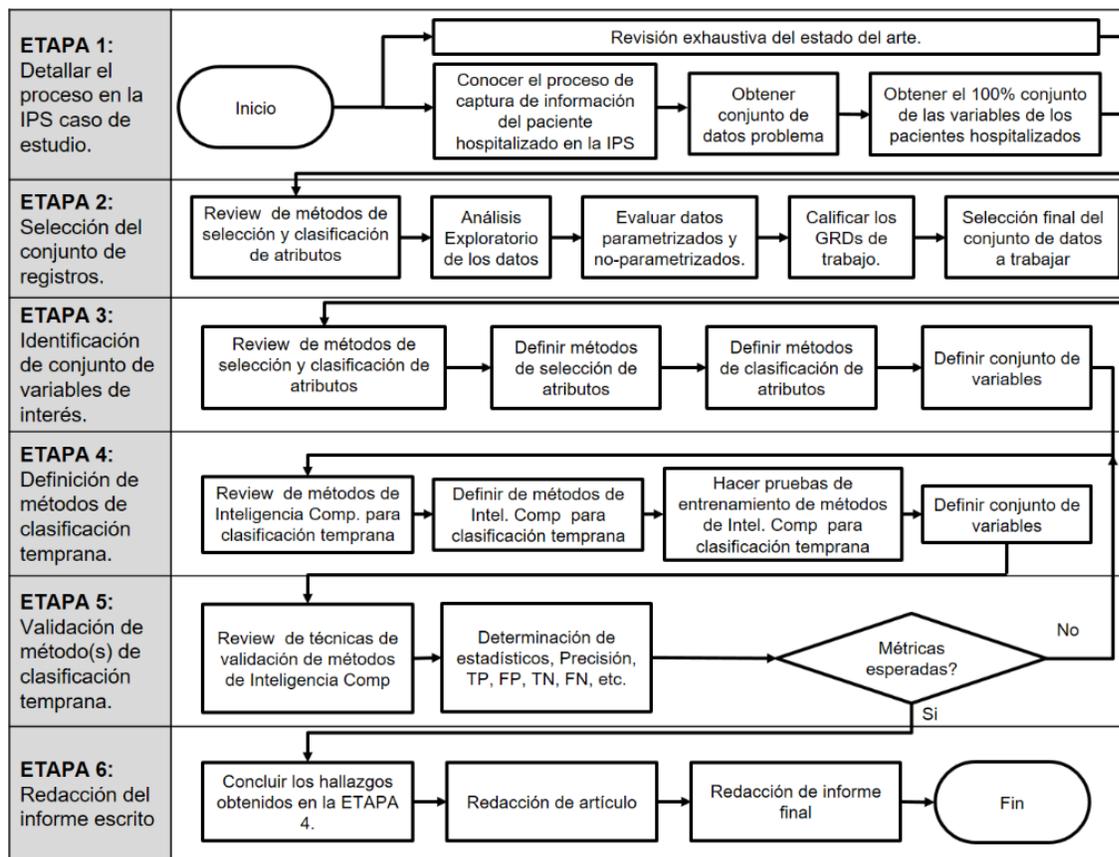
Una correcta clasificación, aporta además a la optimización de la destinación de recursos y a dar prioridad a las áreas más necesitadas de una IPS; debido a que se pueden definir nuevos productos en el portafolio de servicios. Pero lo más importante, será la mejora en los servicios existentes, puesto que los profesionales de la salud pueden disponer de un apoyo para proyectar tratamientos, días de estancia y uso de quirófanos, entre otros recursos.

El producto para entregar con este trabajo es investigación es una metodología que utilice técnicas de inteligencia computacional para hacer una clasificación temprana o predicción de los pacientes que están hospitalizados en la IPS Universitaria Clínica León XIII de la ciudad de Medellín-Colombia en su respectivo GRD. Para definir dicha metodología se debe contar con un conjunto de datos depurado, estandarizado y previamente clasificado en GRD con el algoritmo tradicional que tenga implementando la institución definida como caso de estudio. Este conjunto de datos será la información con la cual se realice el entrenamiento del sistema y posterior prueba para determinar qué metodología propuesta presenta mejor desempeño para este caso particular.

Es de aclarar que para la implementación de la metodología en su etapa de producción, se deben hacer cambios en la cultura organización de la institución, se debe hacer un proceso de sensibilización con los médicos y personal asistencial que reporta información de los pacientes hospitalizados, ya que la confiabilidad de la clasificación temprana de la metodología propuesta está directamente relacionada con la oportunidad y veracidad de la información registrada en el sistema de información clínica que se maneje.

### 1.6. Metodología usada.

El desarrollo de este proyecto se planteó en seis etapas como se muestra en la figura 1-1: *Diagrama de flujo de la metodología propuesta.*



**Figura 1-1:** Diagrama de flujo de la metodología propuesta.

Es de destacar que las etapas 2, 3, 4 y 5 comparte la actividad de revisión bibliográfica, la cual tiene como objetivo buscar en la literatura trabajos relacionados con la temática de interés e identificar qué métodos podrían ser aplicables al problema aquí planteado.

La ETAPA 1, consiste en conocer y modelar los detalles del proceso que se da en el Hospital caso de estudio; identificar cómo se captura la información del paciente durante su hospitalización, determinar la calidad y completitud de los datos recolectados e

identificar la población objeto de estudio en términos de impacto económico del GRD y frecuencia de casos de hospitalización.

La ETAPA 2, consiste en hacer una revisión detallada del conjunto de datos con el que se cuenta, hacer un análisis exploratorio que permita analizar y valorar la información, identificar información faltante, modos de representación y definir el conjunto de datos de egresos que serán caso de estudio de este trabajo de investigación. Al mismo tiempo se hará un tratamiento estadístico que permitan la medición y la descripción (tales como promedios, desviaciones, correlaciones entre variables, entre otras) de los datos de pacientes que podrían ser usados en etapas posteriores.

Para esta etapa se destaca que ya existe un acercamiento previo con el conjunto de datos caso de estudio. Se cuenta con un conjunto de datos de una institución prestadora de servicios de salud de tercer nivel en Colombia, representados en 27.869 egresos correspondientes al año 2015, que cubren 607 GRD diferentes; de estos el 50% está representado en 22 GRD. Igualmente, se sabe que el 51.51% de la población son hombres, y que la base de datos cuenta con 163 variables de entrada, de las cuales 131 están directamente relacionadas con los GRD; de ser necesario, es posible contar con más variables de entrada. También se conoce que es posible tener acceso a 90.000 registros adicionales, que corresponde a información de los años 2016, 2017 y 2018; sin embargo, estos egresos fueron evaluados con un *DRG-grouper* diferente al utilizado en la clasificación de los datos de 2015, lo que implica una diferencia en la codificación de GRD ya que el *DRG-grouper* de 2015 utiliza una MDC de 23 capítulos, mientras que el *DRG-grouper* de los años 2016 al 2018 usa una MDC de 21 capítulos. Como se verá en los capítulos siguientes, la diferencia en la MDC depende de la variante de clasificación GRD que se use en la implementación del software.

La ETAPA 3, tiene como objetivo evaluar métodos de selección y clasificación de las variables de interés de la base de datos proporcionada por el Hospital, que se filtró previamente y que contiene los datos definitivos para el proceso de investigación. En esta etapa se buscarán y definirán aquellos métodos que evalúen qué características

tienen poco peso, un peso moderado y un alto peso para influir en el proceso de clasificación temprana o predicción de pacientes hospitalizados en los diferentes GRD seleccionados para el caso de estudios.

El propósito de la ETAPA 4 es evaluar métodos de inteligencia computacional encontrados en la revisión bibliográfica y otros no explorados en trabajos previos para la clasificación temprana o predicción de pacientes hospitalizados en los diferentes GRD seleccionados para este caso de estudios. Una vez se tengan los métodos que cumplen con los parámetros de la investigación, se procederá a hacer entrenamiento y prueba con los datos seleccionados en la etapa 2, para determinar cuál tiene mejores resultados de clasificación y predicción.

La ETAPA 5 del proyecto consistirá en validar la metodología implementada usando técnicas reportadas en la literatura y que mejor se adaptan a los objetivos del proyecto de investigación; teniendo como parámetro de comparación los GRD clasificados previamente por el software *DRG-grouper* utilizado en el Hospital caso de estudio, los cuales se tomarán como verdaderos. Dicha validación tendrá como salida métricas tales como precisión, falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos, entre otros, para evaluar la eficiencia de clasificación temprana del o los métodos seleccionados.

Finalmente, en la última etapa (ETAPA 6), la cual será permanente y se desarrollará a lo largo del proceso de investigación, se hará la redacción del informe y el artículo.

# Capítulo 2:

## Marco teórico

### 2.1. Introducción.

En este mundo globalizado, donde todo tiende a cambiar y a ser más propenso a usar recursos tecnológicos para poder innovar y marcar la diferencia frente a los competidores en cada nicho de mercado, el sector salud no es ajeno a este fenómeno. En la actualidad se habla de los sistemas de pago basados en Grupos Relacionados por el Diagnóstico (GRD), como un tipo de mecanismos de pago hospitalario. Aunque los sistemas de pago basados en GRD ahora se entienden principalmente como un mecanismo de reembolso, su propósito original era permitir las comparaciones de desempeño entre los hospitales. Hoy en día, los GRD son utilizados principalmente por las EPS para reembolsar a los proveedores (IPS) por atención médica prestada a pacientes hospitalizados, pero en principio también se pueden utilizar para hacer reembolsos por atención no hospitalaria. [18]

Por otro lado, teniendo en cuenta la tecnología para la clasificación de paciente usando técnicas alternativas, en años recientes los métodos de aprendizaje automático se han utilizado ampliamente en pronóstico y predicción, especialmente en el área de la medicina, más propiamente en el diagnóstico de enfermedades. Un gran número de grupos de investigación están trabajando alrededor del mundo para mejorar las técnicas de apoyo al diagnóstico médico.[46]

Esta sección del marco teórico está dividida en dos partes, la primera hace una descripción del flujograma utilizado en el método tradicional para la clasificación de paciente y aborda los Grupos Relacionados por el Diagnóstico, su creación y utilización del algoritmo tradicional de clasificación de paciente en el contexto a nivel mundial. La segunda parte hace referencia a la descripción de las técnicas ML usadas en el contexto de la medicina, específicamente en la clasificación de pacientes y que serán las técnicas utilizadas en este proyecto de investigación.

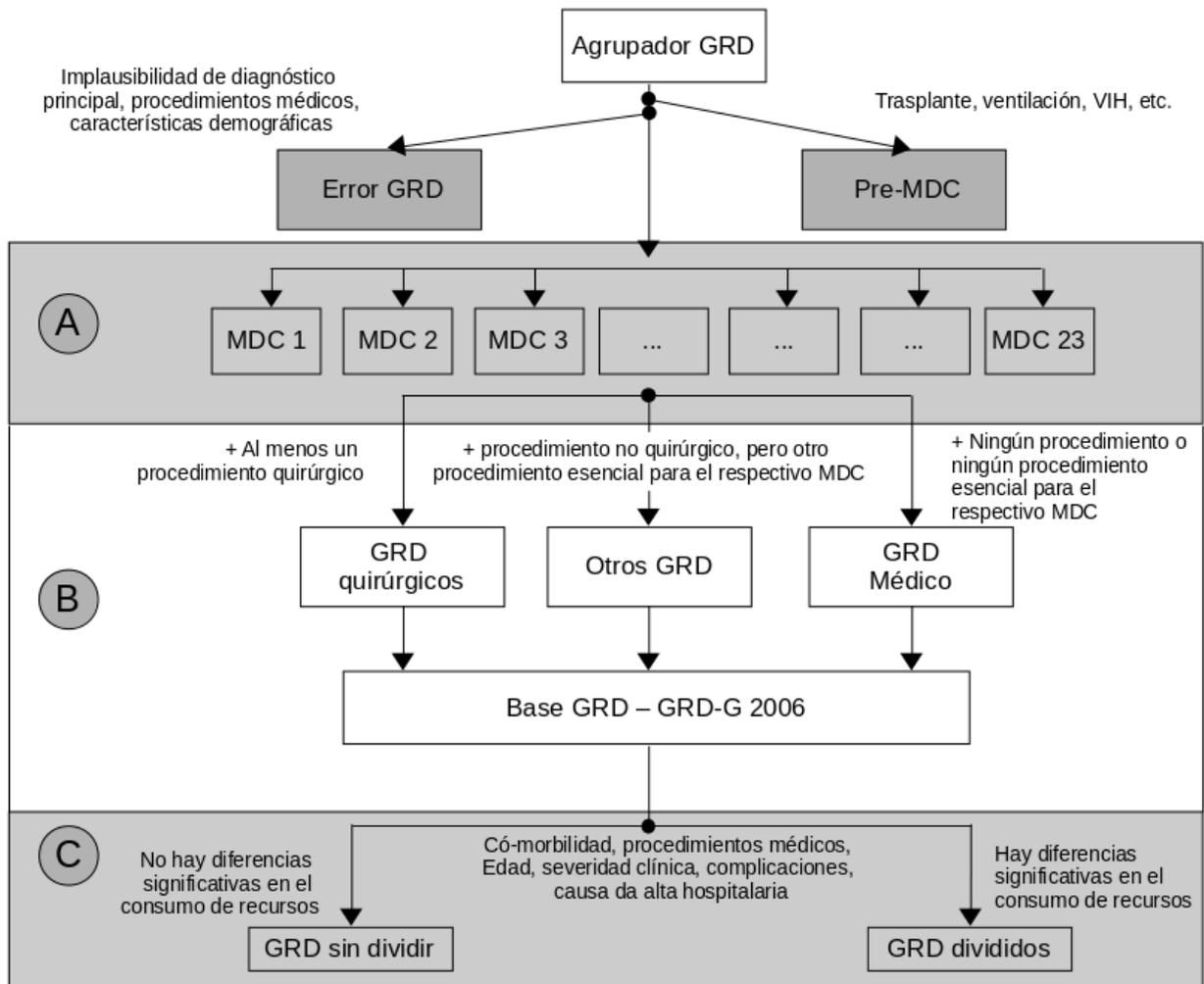
## **2.2 Grupos Relacionados por el Diagnóstico – GRD.**

### **2.2.1 Clasificación de pacientes en GRD - El método tradicional.**

La clasificación de los pacientes se hace en un software comercial o gratuito que implementa el algoritmo tradicional y que se conocen como *DRG-grouper*, el cual está basado en un algoritmo de clasificación que se describe en el Figura 2-1. Como primer paso se ingresan valores de parámetros, tales como diagnóstico principal, otros diagnósticos, procedimientos clínicos, edad, género, así como peso en el caso de recién nacidos, causa de la alta hospitalaria y días de estancia [16]. Los diagnósticos se codifican usando el *International Statistical Classification of Diseases and Related Health Problems (ICD)*.

El algoritmo, en su primera etapa (sección A del diagrama) determina 1 de 22 Categorías de Diagnósticos Principales (*Major Diagnostic Categories – MDC*), los cuales son definidos por el diagnóstico principal. Si no se tiene el diagnóstico principal, se obtiene un GRD errado. [35] Es de destacar en este punto que existen diferentes variantes de los GRD y algunos introducen más o menos MDC, por ejemplo, para el algoritmo que aquí se explica, corresponde a la variante IR-DRG (*International Refined Diagnosis-Related Group*) y que al igual que la variante G-DRG de Alemania contienen 23 MDC; existen otras variantes como AP-DRG, APR-DRG y HCFA-DRG que contienen 25 MDC

o como el caso de la variante GHM-DRG que contiene 26 MDC, entre otras. [36] La base de datos usada en este trabajo de investigación fue etiquetada con el *DRG-grouper* que usa la variante de GRD llamada LAT-GRC, que de acuerdo con su autor, es el único *DRG-grouper* adaptado a la epidemiología y sistemas de información en Latinoamérica.[37]



**Figura 2-1:** Diagrama de flujo del algoritmo tradicional de clasificación de pacientes en GRD.  
 Fuente: Schreyögg J, Tiemann O, Busse R (2006)

Después de determinar la Categoría del Diagnóstico Principal (*Major Diagnostic Categories – MDC*), se encuentran los procedimientos clínicos y las comorbilidades

(sección B del diagrama) que pueden llevar a la clasificación de un paciente en un GRD bien sea quirúrgico o de procedimiento, médico o de otros tipos. En este punto los pacientes se consideran como quirúrgicos si se les practica un procedimiento en el cual es necesario el uso de un quirófano o lugar estéril que sea un recurso importante del hospital. Ahora bien, como la información disponible en el Conjunto Mínimo Básico de Datos (CMBD) no especifica si el paciente ha estado en el quirófano o no, la clasificación del GRD está directamente relacionado con el hecho de que el procedimiento al que fue sometido el paciente fue realizado en un quirófano o en piso. De esta manera, para una valvuloplastia con balón, una biopsia estereotáctica o una extracción de una vesícula biliar enferma conocida como colecistectomías totales, se espera que se haga uso de un quirófano, mientras que procedimientos como toracocentesis, broncoscopias y suturas superficiales no requieren este recurso hospitalario.

Finalmente en la sección C, se tiene que, dentro de esta clasificación inicial e incluyendo la edad del paciente o el peso en caso de recién nacidos, se obtiene una clasificación en diferentes subtipos de GRD; lo cual estará asociado al nivel de consumos de recursos por cada caso particular, es decir, puede ser el mismo diagnóstico para la enfermedad, pero esta puede estar en etapas diferentes, lo que conlleva a un consumo de recursos diferente y una clasificación de GRD específica. [35]

### **2.2.2. Contexto global.**

Los GRD nacieron en la década de las 60's en la Universidad de Yale en Connecticut – Estados Unidos, cuando el gobierno federal pidió desarrollar un modelo de gestión que permitiera 1) medir las diferencias de complejidad y costos de las atenciones en salud y 2) que definiera los “productos” de un hospital. [6] Esta situación se dio al observar que los gastos de los servicios hospitalarios, que en teoría debían ser comparables porque corresponden al mismo diagnóstico de ingreso, difieren notablemente entre un paciente y otro. [19]

Luego de la entrega del prototipo del sistema de gestión por parte de la universidad de Yale, la implementación masiva de esta herramienta se dio a principios de los 80's en el programa *United States Medicare*. Hasta ese momento, los pagos a los hospitales en los Estados Unidos se daban bajo el modelo de reembolso de gastos; esta modalidad de pago afectaba negativamente el sistema, ya que los hospitales no se medían en gastos, pues el estado hacía los reembolsos requeridos. En este punto es en donde los GRD marcaron la diferencia con respecto al sistema tradicional, al permitir determinar los gastos en forma prospectiva; es decir, podían predecir cuánto se iba a gastar un hospital en un procedimiento determinado y así proyectar gastos futuros. Esta nueva forma de costear la salud dio pie para que las entidades hospitalarias encaminaran sus esfuerzos en mejorar la calidad de los servicios y no en el control de los gastos operacionales. [6]

Un sistema GRD es un sistema de clasificación de pacientes que tiene cuatro características principales: (1) los datos recolectados rutinariamente de la alta del paciente (en su mayoría relacionados con el paciente, el tratamiento y las características provistas) se utilizan para clasificar a los pacientes en (2) un número manejable de grupos (es decir, GRD), que pretenden ser (3) clínicamente significativos y (4) económicamente homogéneos. Los GRD consolidan el gran número de pacientes diferentes (individuales) tratados en los hospitales en un número considerable de grupos clínicamente significativos y económicamente homogéneos, proporcionando así una medida concisa de la actividad hospitalaria o, en otras palabras, definen los productos hospitalarios. En consecuencia, la implementación de GRD facilita las comparaciones de costos, calidad y eficiencia de los hospitales, y contribuyen a una mayor transparencia. [20] Cuando se habla de Grupos Relacionados por el Diagnóstico se habla de clasificación de pacientes posterior a que estos sean dados de alta, pues en ese momento ya se tiene la información completa de lo ocurrido durante su estancia. [21]

A continuación, y mediante un ejemplo, ilustramos la manera cómo opera la metodología de los GRD. La información de los pacientes que ingresan a una Clínica u Hospital debe considerar datos como sexo y edad, diagnóstico principal, comorbilidad, tipo de tratamiento (médico o quirúrgico) y complicaciones. [6] Si dos pacientes coinciden en

edad, género, diagnóstico principal y comorbilidades en el ingreso; los GRD parten de la premisa que, durante la estancia en el hospital, tienen un tipo de tratamiento similar. De lo anterior se asume entonces que el gasto para la clínica u hospital es similar y no observar gastos de tratamientos diferentes entre un paciente y otro porque uno tuvo una adición en el procedimiento que se realizó durante su estancia. Es de anotar que el tratamiento de un paciente a otro, aunque presente las similitudes planteadas en el momento del ingreso, de todas maneras, puede variar por imprevistos en el proceso o porque el criterio médico así lo determina. En general, estas situaciones no contradicen el propósito final de los GRD, que buscan ayudar a estandarizar los procedimientos que se realizan a los pacientes de acuerdo con unos parámetros establecidos y de esta forma controlar los gastos que estos generan durante su estancia.

Otro de los usos de los GRD es estandarizar el estimativo de los gastos en los tratamientos que se realizan a los pacientes. Las Entidades Promotoras de Salud y los Sistemas Nacionales de Salud los toman como base para realizar sus pagos a las Instituciones Prestadoras de los Servicios de Salud; para ello se establecen costos fijos a pacientes agrupados en los GRD iguales. Dicho sistema de reembolso tiene a su vez un peso asociado a los GRD que son el resultado de las rutinas comunes que se llevan a cabo en los diferentes tratamientos de la mayoría de los pacientes, es aquí donde se ve la utilidad de los pesos de los GRD, puesto que tienen una variedad de usos entre los que se encuentran las normas comparativas, la ejecución de pagos y la elaboración de presupuestos, entre otros. [5]

Los Grupos Relacionados por el Diagnóstico empezaron a utilizarse en Francia y Canadá en la década de los ochenta; posteriormente, a principios de los noventa, Bélgica, el Reino Unido, Irlanda, Portugal, Australia, Finlandia y Suecia hicieron uso de los GRD, con ciertas modificaciones y acoplados a cada ámbito local, pero con la misma esencia con la que se diseñaron en principio. [6] En la primera década de los 2000, los alemanes implementaron los GRD para financiar las entidades hospitalarias; por otro lado, Japón, Países Bajos y Austria desarrollaron una variante de este modelo de gestión, con los que se hacen comparaciones basadas en la patología del paciente y los procedimientos

terapéuticos en lugar de sólo utilizar los diagnósticos. Así como estos países mencionados anteriormente, se han implementado variantes de los GRD en Singapur, la República de Corea, México y Sudáfrica, lo que demuestra que esta técnica de gestión es eficiente y se ha convertido en un fenómeno mundial. [19] También se destaca que en algunos países con sistemas tipo SNS (Servicio Nacional de Salud), como Suecia e Italia, han utilizado los GRD para evaluar diferencias transnacionales en la prestación de asistencia [22]. En Colombia solamente ocho de cincuenta hospitales generales que existen de alta complejidad usan GRD, lo que los convierte en instituciones de salud innovadoras para el país. [21]

Es de aclarar que cada caso de implementación del sistema GRD produce efectos diversos según el país que lo implementa. Existen algunos estudios que abordan estas temáticas, pero utilizan las mismas metodologías de cuantificación del impacto, pero en términos generales los hallazgos relacionados con estas consecuencias de los pagos basados en GRD en la eficiencia del hospital son variados. El sistema de pago hospitalario reformado se asoció con una mejor eficiencia técnica en Portugal, Suecia y Noruega. En contraste, no se observó un impacto positivo en los Estados Unidos y hubo mejoras tecnológicas, pero no ganancias de eficiencia técnica en Austria, entre otros. Estos resultados diferentes pueden explicarse debido a los puntos de inicio en la implementación y los contextos específicos de cada país en los que se incorporaron las reformas de pago hospitalario, incluidos los diferentes mecanismos de reembolso, la especificación del pago basado en GRD y/o la introducción simultánea de otras reformas de atención médica. [23]

Teóricamente, los pagos a hospitales basados en GRD brindan incentivos a las IPS para aumentar el número de casos tratados y reducir el número de servicios por caso. En contraste con los sistemas de pago por servicios, los GRD impulsan a los hospitales a limitar su actividad a los servicios necesarios y, a diferencia de los presupuestos globales, los GRD motivan a los hospitales a tratar a más pacientes. En términos de control de gastos, el efecto de los pagos basados en GRD depende, por lo tanto, de qué

efecto prevalece: aumentar el número de casos o reducir el número de servicios por caso. [24]

Es de anotar que, en las primeras etapas de la atención, el GRD de los pacientes no puede ser identificado con claridad, debido a la información incompleta y parámetros de entrada no estandarizados como por ejemplo la descripción que hace el paciente de su estado usando texto libre [17]. Es por esta razón que el algoritmo tradicional concibe la clasificación de los pacientes en GRD después de que son dados de alta, cuando se tiene la totalidad de la información de la estancia de este.

### **2.2.3. Contexto Latinoamericano.**

La implementación de los GRD a lo largo del mundo se ha dado de manera gradual, con el tiempo el número de hospitales ha ido creciendo en los diferentes países y esto debido a la dinámica actual de las enfermedades, las nuevas enfermedades que van surgiendo y sus costos. Actualmente, los GRD se usan principalmente como mecanismo de pago a los hospitales y clínicas, sin embargo tiene otros propósitos como la comparación de desempeño entre estas entidades de salud, entre otros, pero también incluye aumento de la transparencia, la eficiencia y la inducción de apoyo a la gestión clínica hospitalaria. [25][26] Los Sistemas de Salud de Latinoamérica han adoptado el uso de los GRD gradualmente, son los casos de México, Brasil, Costa Rica, Argentina y Uruguay, pero tan solo la implementación de estos en Chile ha sido por política del estado.[3]

Desde finales del 2009, 16 hospitales de alta complejidad de Chile inician el proceso de implementación de los GRD. Esta iniciativa fue introducida por el Ministerio de Salud y necesitó de la voluntad de los equipos directivos de los hospitales ya que necesitaban dar a su personal asistencial clave para que trabajara en el proyecto de implementación. La incorporación de hospitales fue progresiva, en 2011 ingresaron 23 instituciones al programa y en 2013 otras 18.[27]

En México, la adopción del sistema GRD en el Instituto Mexicano de Seguridad Social (IMSS) empezó en 1998 y fue suspendido en 2003. En 2002 se llevó a cabo un proyecto de investigación denominado "Uso de los Grupos Relacionados por el Diagnóstico (GRD) en el manejo clínico en el IMSS". La implementación de sistemas GRD ha promovido la eficiencia en los hospitales sin reducir la calidad de la atención médica. El Centro Médico *American British Cowdray* (ABC) ha utilizado los sistemas de codificación médica y GRD durante varios años, estando a la vanguardia y convirtiéndose en el pionero en México, y ahora ha desarrollado su propio método de codificación para lograr un GRD más preciso y tabular cada GRD con el costo específico del ABC.[28]

En abril del año 2006, el directorio del Sanatorio Americano de Uruguay aprueba incluir en su planificación estratégica la aplicación de un Sistema de Clasificación de Pacientes utilizando los GRD y encomienda a la dirección técnica y gerencia su implementación. [29]

En Argentina, en el marco del Programa Nacional de Garantía de Calidad de la Atención Médica, se habla de los Grupos Relacionados por el Diagnóstico como un instrumento de gestión clínica y de algunos otros trabajos publicados como herramientas para la estimación de costos hospitalarios o desarrollo de indicadores de procesos. Con el propósito de optimizar los sistemas de atención en la salud, brindar los más altos nivel de calidad de atención a un bajo costo, de la manera más equitativa y con una mayor cobertura, las autoridades sanitarias provinciales propusieron la capacitación de un equipo interdisciplinario en los conceptos de GRD. En noviembre de 2010, a solicitud de las autoridades ministeriales, la dirección del Hospital Córdoba convocó a un grupo multidisciplinario de agentes con el fin de capacitarlos en el sistema de clasificación de pacientes, utilizando los GRD como una herramienta de gestión tanto administrativa como clínica. La capacitación del personal se realizó el primer semestre del 2011 y posterior a esto empezó la implementación del sistema. [30]

En 2008 el Banco Mundial adjudicó a una empresa española un contrato para dotar a 29 hospitales de la Caja Costarricense de Seguro Social de herramientas de gestión

basadas en la utilización del GRD [31]. Desafortunadamente, según el reporte de la OCDE de 2017, estas iniciativas fueron abandonadas porque había vencido su licencia de uso y no se hizo un esfuerzo por desarrollar un sistema interno que lo reemplazará. [32]

Los GRD en Brasil aparecieron en 2011 y después de una década de desarrollo constituyen los Grupos Relacionados por el Diagnóstico de Severidad Médica (MS-DRG), modelo creado y usado en la actualidad por Estados Unidos. En vista del buen desempeño económico y de asistencia en la implementación del modelo en 253,000 egresos hospitalarios, se expandió el programa y actualmente 3.8 millones de pacientes, distribuidos en 180 hospitales, están bajo la cobertura del sistema GRD Brasil en todo el país. Esta implementación ha avanzado con tanto éxito, que crearon el *DRG Brasil Refined*, que se refiere al MS-DRG pero reemplazaron los códigos estadounidenses por códigos brasileños. [33]

La implementación de GRD en Colombia fue lenta y demorada. En la primera década del siglo XXI sólo había seis hospitales que contaban con este sistema aplicado de forma autónoma.[21] En 2003, el Centro de Gestión Hospitalaria, junto con ocho hospitales en Colombia, cuatro en Medellín y cuatro en Bogotá, implementaron el sistema GRD para adherirse a un sistema de gestión organizacional que proporcionara una mejora en el sistema financiero y el compromiso clínico de los hospitales. [33]

El Hospital Universitario La Samaritana, ubicado en la ciudad de Bogotá, incorporó el sistema GRD con el objetivo de mejorar el manejo clínico. Esta adopción se consideró una gran innovación para el hospital, puesto que el uso de GRD requiere una cultura y percepción innovadoras por parte de los miembros de la institución de salud. [21] Pero fue el Hospital San Vicente en Medellín, el pionero en la expansión de la gestión administrativa en entidades de salud en Colombia. Esta propuesta de implementación utilizó el Software para el análisis de GRD y según los resultados obtenidos la institución de salud definió el comportamiento de algunas variables que fueron fundamentales para la calidad de la producción. [34].

Finalmente, las instituciones prestadoras de salud en Colombia, implementaron la definición de los productos como grupos homogéneos de patologías, con el objetivo de tener un sistema de clasificación de pacientes, dando lugar a la inclusión de Grupos Relacionados por el Diagnóstico (GRD).[34] Para llamar la atención del Ministerio de Protección Social sobre la importancia estratégica de la contratación de servicios de salud con el objetivo de ofrecer la sostenibilidad de las IPS a largo plazo, se llevó a cabo la implementación de Grupos Relacionados por el Diagnóstico como forma de pago. [33]

### **2.3. Machine Learning.**

En el campo de la Inteligencia Artificial, más concretamente en el aprendizaje automático, existen tres tipos de modelos de aprendizaje; los denominados supervisado, no supervisados y semi-supervisado. En el primer caso, se generan modelos de Machine Learning predictivos, los cuales se basan tanto en los datos de entrada como en los datos de salida, esto consiste en tener un conjunto de datos etiquetado y clasificado, es decir, ya se sabe cuál va a ser la salida de cada una de las muestras en el conjunto de datos. Algunos de los métodos que se pueden usar en este caso son: Redes Neuronales Artificiales, Árboles de Decisión, Máquinas de Soporte Vectorial (SVM), K vecinos más cercanos (K-NN), entre otros.

Para el caso de los denominados modelos No Supervisados, su forma de trabajar es muy similar a la de los modelos supervisados, con la diferencia que estos no utilizan la salida de los datos para ajustar el modelo predictivo, es decir, el conjunto de datos usado no está etiquetado. Para este caso uno de los métodos más usados es el agrupamiento o *clustering* ya que se tienen en cuenta las características de los datos para agruparlos según su similitud. Otras técnicas de uso para estos casos son el *K-means*, agrupamiento jerárquico, reducción de dimensión, detección de datos atípicos (outliers), entre otros.

El aprendizaje semi-supervisado utiliza bases de datos que tienen muestras etiquetadas y otras sin etiquetar. Sin embargo, las muestras etiquetadas a menudo son difíciles, costosas o requieren mucho tiempo, ya que requieren los esfuerzos de anotadores experimentados, es decir, expertos en el tema de interés que etiqueten las muestras. Por otro lado, las muestras sin etiquetar podrían ser más fáciles de conseguir. Por las razones anteriores, el aprendizaje semi-supervisado aborda este problema mediante el uso de una gran cantidad de datos sin etiquetar, junto con, usualmente, pocas muestras etiquetadas, para construir mejores clasificadores. [47]

### **2.3.1. Redes Neuronales Artificiales - RNAs.**

Las RNAs son una de las técnicas más usadas de inteligencia artificial para problemas de clasificación y regresión. Muchas investigaciones mostraron que las RNAs proporcionan una buena precisión en el diagnóstico de muchas enfermedades como el cáncer de mama [48], así como en tareas de clasificación de pacientes en diagnósticos médicos; es el caso de pacientes evaluados para detectar la enfermedad de Parkinson, donde existe evidencia de que presenta mejores resultados que otras técnicas como las Máquinas de Soporte Vectorial y el Naïve Bayes [49]. En los casos citados anteriormente se hace uso de las RNAs para la solución de problemas biclase, por ejemplo, para el caso del cáncer [48], la salida de la red neuronal es “*Tiene*” o “*No Tiene*” cáncer de mama, es decir, diagnostica si el paciente tiene o no la enfermedad, en otras palabras clasifica a los pacientes en dos grupos: los que tienen la enfermedad y los que no la tienen, de igual forma pasa con la enfermedad de Parkinson y todos los proyectos relacionados en la literatura científica que buscan resultados Positivo o Negativo para cualquier enfermedad. Pero hay otros proyectos en los que los modelos a evaluar deben ser multiclase, como es el caso de [50], donde se utilizan Redes Neuronales y otras técnicas de clasificación para identificar pacientes en un triaje (traje 1, 2, 3 y 4).

### 2.3.2. Extreme Gradient Boosting - XGBoost.

Adicional a las redes neuronales, en la sección 3.4 de este trabajo se habla sobre las técnicas usadas en trabajos relacionados con la clasificación temprana de pacientes en GRD y se observa que en varios de estos trabajos se utilizan métodos de árboles, como Random Forest y Árboles de Decisión. Lo anterior sumado también a que el algoritmo tradicional con el que se hace la clasificación del GRD es un árbol, son las razones por las cuales seleccionamos otro método de esta naturaleza, el método *eXtreme Gradient Boosting* conocido como el XGBoost.

El XGBoost obtuvo resultados destacados en diferentes competencias de aprendizaje automático en los últimos años, estos resultados se deben a sus características como la facilidad para usar, para paralelizar y la alta precisión en predicción. Áreas como la bioinformática y la informática médica no son ajenas a estos resultados, se ha demostrado que el XGBoost funciona muy bien en una variedad de tareas en estos campos. Se reportan trabajos donde se han utilizado el método XGBoost para el diagnóstico de enfermedad renal crónica y los resultados fueron destacados, alta precisión, sensibilidad y especificidad. Otro caso de éxito de esta técnica es un trabajo propuesto para la predicción de proteínas, resultados positivos en la clasificación de efectos de las combinaciones de drogas. Estos estudios han demostrado el potencial del XGBoost para la predicción diagnóstica y pronóstica de cáncer.[51]

Al igual que las RNAs, el XGBoost también es usado en modelos de clasificación multiclase. Yoo 2020 expone en su trabajo el uso del XGBoost para la clasificación de pacientes en una de las cuatro clases de cirugía refractiva propuestas que se debe realizar según sus datos de entrada. Para determinar la técnica quirúrgica óptima basada en la evidencia médica y las expectativas del paciente para la cirugía y la recuperación, los cirujanos deben considerar varias mediciones oculares y factores del paciente, como ojo seco, estilo de vida y presupuesto.

## 2.4. Conclusiones.

En la actualidad, las entidades prestadoras de los servicios de salud hacen los cobros de sus servicios prestados a través de diversas modalidades, cada una con ventajas y desventajas. Pero el incentivar a los hospitales para mejorar su rendimiento impulsó el movimiento en la mayoría de los países abanderados en el uso de GRD, hacia sistemas de pago de servicios hospitalarios basados en GRD. Antes de la introducción de los sistemas de pago hospitalarios basados en GRD, los países utilizan dos mecanismos básicos para pagar la atención hospitalaria: los pagos de tarifa por servicio y los presupuestos globales. [24]

Si bien aún se está lejos de que todos los países del mundo implementen sistemas de pago prospectivo como lo son los GRD, existe evidencia de que países del primer mundo y países en vía de desarrollo se preocupan por el uso de herramientas que permitan no solo mejorar la gestión administrativa de los costos hospitalarios, sino también hacer seguimiento a la actividad asistencial con los pacientes, con herramientas de este tipo se puede hacer seguimiento a la evolución de estos, de tal forma se cumplan con los estándares establecidos y no se alarguen las estancias de los pacientes por trámites administrativos o seguimientos asistenciales inadecuados.

Una de las desventajas que tiene la implementación de este sistema de GRD en un país completo es que, si no se convierte en una política de salud pública, es difícil que hospitales o clínicas aisladamente tomen la decisión de hacer dicha implementación debido a que el esfuerzo es alto en tiempo y costos, y la adopción de estas herramientas requieren de cambio en la cultura organizacional de la institución de salud. Además, en algunos casos implica la contratación de más personal, ya sea para implementar o para apoyar las actividades asistenciales y administrativas.

# Capítulo 3:

## Estado del Arte

### 3.1. Introducción.

Existen muchas publicaciones relacionadas con los Grupos Relacionados por el Diagnóstico y su implementación alrededor del mundo, desde su inicio en la Universidad de Yale a finales de los 70's y principios de los 80's en Estados Unidos. Los GRD se han convertido en una de las principales herramientas de gestión hospitalaria, como mecanismo de reembolso de servicios hospitalarios.

El uso de esta herramienta se hace con la implementación de un algoritmo de decisión, el cual consiste en contestar una serie de preguntas que al final dan como resultado la ubicación de un paciente en un grupo determinado según el nivel de consumo de recursos hospitalarios, como es el caso de días de estancia, pruebas diagnósticas, uso de quirófanos, etc.

En la parte inicial de este capítulo, se hará un recorrido por el estado del arte en cuanto al uso del Machine Learning para la implementación de estos agrupadores, existe literatura científica que habla de los clasificadores alternos; aquí se describe el uso de diferentes técnicas de Machine Learning para la clasificación de pacientes en los GRD que maneja cada entidad caso de estudio. Los objetivos de estos trabajos de investigación han sido evaluar el comportamiento de estas técnicas alternativas en

cuanto a la clasificación de pacientes en los GRD sin tener que pasar por el formulario de preguntas, así como lo hace el algoritmo tradicional y comparar el desempeño para evaluar su implementación. Una última sección se dedica a describir diferentes técnicas de Machine Learning, pero esta vez no para la clasificación de egresos de pacientes en GRD, sino para la predicción del GRD en el cual se agrupará el paciente cuando es admitido en el hospital o clínica. El objetivo de estos últimos trabajos es ayudar a que las entidades de salud sepan el consumo promedio que podría tener un paciente hospitalizado y de ser necesario tomar las medidas pertinentes para que estos gastos no se suban más de lo debido según los estándares.

### 3.2. Revisión bibliográfica.

Para determinar el estado del arte se realizó la revisión bibliográfica en cuanto a clasificación y predicción de pacientes hospitalizados en Grupos Relacionados por el Diagnóstico y que utilizaran técnicas de ML para este propósito. Se hizo una revisión sistemática de artículos científicos en las bases de datos bibliográficas Scopus, Science Direct, Springer Link, IEEE Xplore Digital Library y Google Scholar.

La búsqueda bibliográfica se estructura con las preguntas planteadas a continuación:

- ¿Qué tipos de clasificadores de GRD existen?
- ¿En cuál etapa de la estancia hospitalaria nos vamos a enfocar para la clasificación temprana (predicción) de los GRD?
- ¿Cuál es el tipo de paciente que va a ser de interés para este proyecto de investigación?

En la etapa de búsqueda se hallaron 171 títulos, por lo que fue necesario, ajustar los criterios de búsqueda, se especificó que las palabras claves: “*Diagnosis-Related Groups*”, “*Machine Learning*” e “*Early Classification*” o “*Predict*” solo estuvieran en Título, Resumen o Palabras clave, ya que muchas publicaciones en el ámbito médico tienen en la bibliografía algo relacionado con DRG o el nombre completo “*Diagnosis-Related*

*Group*". Se continuó con la revisión de los resúmenes e introducciones de estos y se encontró que en la línea de investigación que es la clasificación temprana o predicción y clasificación usando técnicas de inteligencia computacional solo quedaron 6 artículos.

### **3.3. Machine Learning para clasificación de pacientes en GRD.**

Los métodos de inteligencia computacional se han utilizado recientemente en diversas aplicaciones en el área de la salud [38]. Los modelos de aprendizaje automático son de naturaleza paramétrica y no paramétrica, estos últimos no necesitan los supuestos que se hacen en las técnicas estadísticas tradicionales [39], a diferencia de la forma en que trabaja el algoritmo tradicional de clasificación de los GRD. A continuación, se comparan varios métodos de inteligencia computacional utilizando conjuntos de datos y características específicas, de los cuales varios tenían un poder de clasificación mayor que las alternativas tradicionales [40].

Las técnicas de Machine Learning han demostrado la capacidad que tienen para resolver los problemas de clasificación. Se han utilizado un amplio conjunto de métodos de inteligencia computacional para desarrollar modelos de clasificación en el ámbito clínico, como es el caso del uso de las técnicas Naïve-Bayesian, Máquinas de Soporte Vectorial (SVM), Regresión Logística, Árbol de Decisión C4.5 y Random Forest, para la comparación de modelos para la predicción de los costos médicos de la fusión espinal en GRD en Taiwán. Los cinco modelos son ejemplos típicos de Machine Learning supervisado [41].

Se han planteado estudios de clasificación de pacientes en GRD, en una etapa temprana de la estancia del paciente. En [17], evalúan si una clasificación temprana de un GRD utilizando técnicas de inteligencia computacional puede mejorar los márgenes de contribución, la asignación de recursos como quirófanos y camas, entre otros. Este estudio se centró en la precisión de la clasificación en el momento en que el paciente busca la admisión, y posterior a esta, es decir, desde la admisión hasta que es dado de

alta. Los resultados mostraron que las técnicas de inteligencia computacional pueden aumentar significativamente la precisión en la clasificación temprana de paciente en GRD, especialmente en aquellos que contactaron a la clínica o al hospital antes de la admisión. Adicionalmente, se demostró en ese estudio que las técnicas de aprendizaje automático combinadas con programación matemática pueden aumentar los márgenes de contribución y mejorar la asignación de recursos escasos.

Debido a su poder predictivo, algunos sistemas de atención médica intentan utilizar algunas técnicas de extracción de datos para descubrir relaciones ocultas, así como las tendencias dentro de la base de datos clínica y convertirla en información útil que puede ser usada en la solución de problemas médicos y otros marcadores de decisión clínica. En general, las técnicas de minería de datos pueden aprender de lo sucedido en eventos anteriores y el modelo resultante proporciona conocimiento formalizado y predicción del resultado. [42]

### **3.4. Machine Learning para predicción de pacientes en GRD.**

En esta sección, se describen los trabajos encontrados que tiene relación con el proyecto de investigación, que usan técnicas de ML para clasificación temprana o predicción de pacientes en GRD, en orden cronológico. Para cada trabajo, se pone especial atención en los siguientes criterios de análisis:

- Objetivo de la investigación.
- Base de datos y características usadas.
- Metodologías de ML utilizadas.
- Resultados obtenidos.
- Faltantes.

La tabla 3-1 contiene el resumen de los artículos encontrados en el estado del arte que tienen una relación directa con este trabajo de investigación. Se describe brevemente los

trabajos, las técnicas de ML utilizada y los resultados obtenidos. Posterior a la tabla en esta misma sección se describe con más detalle cada uno de los trabajos.

	Artículo	¿Qué hace?	Técnicas usadas	Conjunto de datos	Resultados
1	<p>Machine Learning Approaches for Early DRG Classification and Resource Allocation</p> <p>Munich, Alemania, 2015</p> <p>Gartner, D., Kolisch, R., Neill, D. B., &amp; Padman, R. (2015)</p>	<p>Identificación temprana del GRD usando ML: Sus metas:</p> <ol style="list-style-type: none"> <li>1) un mayor margen de contribución,</li> <li>2) mejorar de la asignación de recursos hospitalarios, y</li> <li>3) mejoras potenciales en el número de pacientes electivos no urgentes admitidos</li> </ol> <p>Comparado con el enfoque actual de un hospital usando un <i>DRG-grouper</i>.</p>	<ol style="list-style-type: none"> <li>1. Naïve Bayes (NB)</li> <li>2. Redes Bayesianas (RB)</li> <li>3. Árbol de clasificación</li> <li>4. Combinación de los tres clasificadores básicos por votación</li> <li>5. Combinación de <i>DRG-grouper</i> con RB y Árboles de Clasificación por Promedio de Probabilidad.</li> <li>6. Clasificación de GRD Basado en Descomposición</li> <li>7. Mapeo basado en reglas de decisión de los valores de los atributos a los GRD.</li> <li>8. Asignación aleatoria de los GRD (RND)</li> </ol>	<p>HCE Datos parametrizados y texto libre.</p> <p>Se contó con 3458 pacientes electivos que contactaron al hospital antes de la admisión y 16601 pacientes electivos y no electivos que fueron admitidos durante el 2001.</p>	<p>Se encuentra que con el enfoque Promedio de Probabilidad (PA):</p> <ol style="list-style-type: none"> <li>1) Se mejora el margen de contribución en un 2.9% y la asignación de quirófanos y camas de 66.3% a 67.3% y 70.7% a 71.7%.</li> <li>2) alcanza hasta un 79.5% de clasificación general antes de la admisión y en el momento de la admisión, la clasificación es del 65.5%</li> </ol>
2	<p>Improving Hospital-Wide Early Resource Allocation through Machine Learning</p> <p>Munich, Alemania, 2015</p> <p>Gartner, D., &amp; Padman, R. (2015)</p>	<p>Evaluar cómo ayuda la clasificación temprana de GRD en la mejora de asignación de los recursos hospitalarios (quirófanos y camas).</p>	<ol style="list-style-type: none"> <li>1. <i>DRG-grouper</i></li> <li>2. Naïve Bayes</li> </ol>	<p>HCE Datos parametrizados y texto libre.</p> <p>Se analizaron datos de pacientes hospitalizados de un año que consisten en más de 16,000 registros de un hospital con 350 camas.</p>	<ol style="list-style-type: none"> <li>1) TP Naïve Bayes 25.3%</li> <li>2) 17% de mejora en el margen de contribución</li> </ol>
3	<p>Prediction of Clinical Practices by Clinical Data of the Previous Day Using Linear Support Vector Machine</p> <p>Japón, 2016</p> <p>Nakai, T., Takemura, T., Sakurai, R., Fujita, K., Okamoto, K., &amp; Kuroda, T. (2016)</p>	<p>Sistema de apoyo para los médicos. Se construye un modelo basado en la hipótesis de que las prácticas clínicas de un día podrían determinar los procedimientos del día siguiente para los mismos pacientes.</p>	<ol style="list-style-type: none"> <li>1. Linear SVM</li> </ol>	<p>HCE</p> <p>Se usan datos de los periodos 2006/4 a 2006/6 que suman un total de 6445 hospitalizaciones</p>	<p>Todas las prácticas clínicas tuvieron más del 80% de precisión y el 97.4% de las prácticas tuvieron más del 90% de precisión</p>

4	<p>Classification of hospital admissions into emergency and elective care: a machine learning approach</p> <p>United Kingdom, 2017</p> <p>Krämer, J., Schreyögg, J., &amp; Busse, R. (2017).</p>	<p>El objetivo es desarrollar un modelo para clasificar las admisiones de pacientes hospitalizados según el diagnóstico primario de un paciente, ya sea como atención de emergencia o atención electiva, y prediciendo la urgencia como un valor numérico.</p>	<p>1. Random Forest</p>	<p>Información demográfica, fecha y hora de ingreso y alta por departamento hospitalario, diagnóstico primario y diagnósticos secundarios, procedimientos con fecha y hora y un identificador de hospital.</p> <p>Se usaron datos entre los años 2005-2013 y luego de filtrar, quedaron 119 millones de episodios de pacientes hospitalizados.</p>	<p>El modelo de clasificación final arroja una precisión global de 95.81%. Sensibilidad: 97.0 Especificidad: 93.6.</p> <p>TODOS los AUC estuvieron por encima de 0.99</p>
5	<p>Automatic selection of diagnosis procedure combination codes based on partial treatment data relative to the number of hospitalization days.</p> <p>Japón, 2018</p> <p>Okamoto, K., Uchiyama, T., Takemura, T., Kume, N., Kuroda, T., &amp; Yoshihara, H. (2018)</p>	<p>La combinación de procedimientos de diagnóstico (DPC) es una clasificación de grupo de diagnóstico creada en Japón en imitación de los Grupos Relacionados por el Diagnóstico (GRD) utilizados en los EE. UU.</p> <p>En este artículo, los autores desarrollan un método de selección automática de códigos DPC basado en información de tratamiento estandarizado y evalúan el método desarrollado. Durante la hospitalización, el paciente se somete a una serie de tratamientos y, por lo tanto, los tratamientos son diferentes según las fases de la hospitalización. Además, se solicita a los médicos que seleccionen un código DPC en una fecha temprana de hospitalización. Por lo tanto, se evalúa la precisión con la que se seleccionan los códigos DPC en función de los datos de tratamiento parcial en relación con el número de días de hospitalización.</p>	<p>1. Naïve Bayes 2. SVM 3. Una combinación de las dos anteriores.</p>	<p>La información de tratamiento de los pacientes que se almacenó antes de su alta hospitalaria.</p> <p>Severidad, complicaciones, operaciones requeridas, propósito del ingreso, edad y tratamientos necesarios.</p> <p>Para la investigación se contó con 10.123 casos en 1.572 códigos DPC (GRD) registrados durante el año 2011.</p>	<p>Un método combinado de ambos algoritmos Naïve Bayes y SVM, produce en todas las situaciones el mejor resultado.</p> <p>La precisión de todos los algoritmos es del 85%.</p>

6	<p>Comparison of models for the prediction of medical costs of spinal fusion in Taiwan diagnosis-related groups by machine learning algorithms.</p> <p>Taiwan, 2018</p> <p>Kuo, C. Y., Yu, L. C., Chen, H. C., &amp; Chan, C. L. (2018).</p>	<p>ML para la predicción de los costos médicos asociados con la fusión espinal en términos de ganancia o pérdida con Taiwán Diagnosis-Related Groups</p>	<ol style="list-style-type: none"> <li>1. Naïve-Bayesian,</li> <li>2. VSM,</li> <li>3. Regresión Logística,</li> <li>4. Árbol de decisión C4.5 y</li> <li>5. Random Forest</li> </ol>	<p>La base de datos incluía las características básicas de los pacientes, fechas de ingreso, fechas de alta, diagnósticos primarios, complicaciones o comorbilidad. Códigos ICD-9-CM, órdenes médicas y costos.</p> <p>Contiene información de los años 2010 a 2013, con un total de 532 casos.</p>	<p>El método de Random Forest, el mejor desempeño predictivo, logrando una precisión de 84.30%, una sensibilidad de 71.4%, una especificidad de 92.2% y un AUC de 0.904.</p> <p>El segundo mejor clasificador fue la Regresión Logística y SVM, con desempeños iguales, seguido de Árbol de decisión C4.5 y por último Naïve Bayesian con el rendimiento más bajo.</p>
---	--	--	---	---	--

**Tabla 3-1:** Estudios que usan inteligencia computacional para la clasificación temprana de GRD

### 3.4.1. Gartner et al. (2015): “Machine learning approaches for Early DRG Classification and Resource Allocation”.

Este trabajo de investigación plantea examinar el valor de la planificación inicial al interior de una entidad de salud frente a las decisiones de asignación de recursos en todo el hospital basadas en ML y la programación de enteros mixtos (*MIP – Mixed-Integer Programming*), centrándose en la predicción de Grupos Relacionados por el Diagnóstico (GRD) y el uso de estas predicciones para la asignación recursos hospitalarios como camas y quirófanos.

La base de datos para el desarrollo de este proyecto de investigación proviene de un hospital de 350 camas en las cercanías de Munich, Alemania. El conjunto de datos tuvo dos partes, una primera con información de pacientes que contactaron al hospital antes de la admisión (pacientes electivos), pacientes que programaron su ingreso al hospital o fueron remitidos desde otro centro de salud. Se obtienen 3.458 pacientes *electivos* que contactan con el hospital antes del ingreso y fueron asignados a 413 GRD diferentes. La segunda parte de la base de datos corresponde a los datos de todos los pacientes disponibles al ingreso, en este punto de la admisión se clasificaban los pacientes en *electivos* y *no electivos*, al igual que la primera base de datos los pacientes *electivos* son

pacientes que programaron su ingreso al hospital y los *no electivos* son los pacientes que ingresaron por urgencia. Para esta segunda base de datos, se obtuvieron registros de 16.601 pacientes que ingresaron durante el año 2011, que fueron asignados a 680 GRD diferentes. Adicionalmente, separaron los conjuntos de datos para evaluar diferentes escenarios antes y durante de la admisión, en la tabla 3-2 se describen los subconjuntos creado, donde los subconjuntos de 1 al 9 son pacientes *electivos* que contactaron el hospital antes de la admisión, se evalúan escenarios donde están documentados todos, algunos o ninguno, ya sean diagnósticos o procedimientos. Los subconjuntos 10 al 12 son de los pacientes que contactaron al hospital en el momento de la admisión y representan los escenarios donde se tienen documentados todos, algunos o ninguno de los procedimientos realizados a los pacientes, se asume que estos pacientes al tener el contacto en la admisión tienen todos los diagnósticos documentados.

Códigos de procedimientos	Diagnósticos de admisión	Subconjunto de datos
Todos los códigos	Todos	1
	Solo el Diagnóstico 1	2
	Ninguno	3
Todos los códigos documentados en los dos primeros días	Todos	4
	Solo el Diagnóstico 1	5
	Ninguno	6
Ninguno	Todos	7
	Solo el Diagnóstico 1	8
	Ninguno	9
Todos los códigos		10
Todos los códigos documentados en los dos primeros días		11
Ninguno		12

**Tabla 3-2:** Distribución de los subconjuntos de datos

Esta investigación usó técnicas de clasificación y selección de características, las cuales fueron *Information Gain (IG)* y *Relief-F* con el objetivo de asignar un dato numérico a las características para formar un ranking de estas. Luego se usaron *Correlation-Based*

*Feature Selection* y *Wrapper Attribute Subset Evaluation* para determinar cuáles características de ese ranking tenían una mayor correlación que pudiera aportar al proceso de clasificación. Luego de tener las variables definidas, procedieron a evaluar ocho clasificadores: 1) Naïve Bayes, 2) Redes Bayesiana, 3) Árboles de Clasificación, 4) Clasificación Combinada basada en Votación, 5) Promedio de Probabilidad para combinar el *DRG-Grouper* con enfoques de ML, 6) Clasificación de GRD basado en descomposición (DDC), 7) Mapeo basado en reglas de decisión de Valores de Atributos para GRD y 8) Asignación Aleatoria de GRD (RND).

En la tabla 3-3 se muestran los resultados de precisión de las ocho técnicas (columnas) de clasificación más los resultados del *DRG-grouper*, el cual es el clasificador tradicional que usaban el hospital en el momento del proyecto de investigación, las filas de la tabla relacionan los subconjuntos de registros de pacientes en que fueron divididos el conjunto de datos según el momento de contacto del paciente. Esta tabla muestra que el enfoque llamado PA que combina técnicas de aprendizaje automático con el agrupador DRG, siempre supera al enfoque actual empleado en el hospital (*DRG-grouper*). Además, en comparación con los otros enfoques de aprendizaje automático, el enfoque de PA puede superarlos en cuatro de los nueve conjuntos de datos. Ninguno de los otros enfoques (BN, NB, reglas de decisión, árboles de clasificación o votación) supera el enfoque actual del hospital para todos los conjuntos de datos. Sin embargo, los métodos de aprendizaje automático tienden a superar consistentemente al *DRG-grouper* en los subconjuntos de datos 7 a 9, que representan los escenarios más comunes, donde los códigos de procedimiento y/o diagnósticos no están disponibles en el momento en que se realiza la clasificación de GRD.

k	RND	BN	Grouper	DDC	PA	NB	Rules	Tree	Vote
1	0.1 (0.0)	70.0 (1.8)	77.9 (2.9)	10.0 (5.7)	<b>78.7 (3.3)</b>	35.5 (1.3)	75.8 (1.7)	75.8 (1.7)	75.9 (1.2)
2	0.1 (0.0)	70.2 (2.2)	77.2 (2.5)	10.3 (6.1)	<b>77.9 (3.3)</b>	37.4 (1.1)	75.2 (1.8)	75.3 (2.0)	75.5 (1.4)
3	0.1 (0.0)	55.9 (5.2)	0.8 (0.4)	9.8 (5.3)	58.3 (4.6)	30.5 (1.9)	20.1 (0.3)	<b>62.1 (3.1)</b>	52.1 (1.9)
4	0.1 (0.0)	66.4 (2.2)	71.9 (3.3)	10.1 (5.7)	<b>72.8 (3.1)</b>	35.1 (2.8)	70.2 (2.3)	70.2 (2.3)	70.4 (2.7)
5	0.1 (0.0)	66.5 (2.5)	71.3 (3.4)	11.0 (4.9)	<b>72.0 (3.3)</b>	38.2 (2.0)	69.7 (2.4)	69.7 (2.4)	69.9 (2.6)
6	0.1 (0.0)	52.0 (7.8)	0.8 (0.4)	9.8 (5.3)	53.5 (5.7)	31.1 (1.9)	20.1 (0.3)	<b>56.5 (1.8)</b>	48.4 (1.0)
7	0.1 (0.0)	52.3 (5.2)	18.3 (0.5)	11.9 (4.3)	52.5 (3.9)	39.5 (0.5)	45.2 (3.3)	52.9 (4.1)	<b>54.2 (5.4)</b>

8	0.1 (0.0)	51.5 (3.3)	18.3 (0.5)	13.9 (5.7)	52.3 (4.8)	44.4 (1.5)	45.2 (3.3)	52.7 (6.6)	<b>53.6 (4.7)</b>
9	0.1 (0.0)	31.1 (6.0)	0.1 (0.0)	10.6 (5.2)	28.3 (3.6)	33.4 (1.7)	20.1 (0.3)	<b>38.7 (4.4)</b>	35.3 (2.7)
Media	0.1 (0.0)	57.3 (4.0)	37.4 (1.5)	10.8 (5.4)	60.7 (4.0)	36.1 (1.6)	49.1 (1.7)	<b>61.5 (3.2)</b>	59.5 (2.6)
10	0.2 (0.0)	60.8 (2.1)	60.8 (1.2)	20.4 (2.4)	<b>65.5 (1.7)</b>	32.1 (0.3)	61.7 (1.2)	63.7 (1.5)	63.8 (1.2)
11	0.2 (0.0)	56.5 (0.9)	56.6 (1.5)	20.8 (2.7)	<b>59.7 (1.2)</b>	32.7 (0.2)	57.3 (1.4)	58.4 (1.4)	58.6 (1.2)
12	0.2 (0.0)	49.1 (1.7)	37.1 (0.6)	26.0 (2.3)	49.7 (1.4)	39.7 (0.8)	45.2 (0.9)	49.3 (0.9)	<b>50.4 (0.9)</b>
Media	0.2 (0.0)	55.5 (1.6)	51.5 (1.1)	22.4 (2.5)	<b>58.3 (1.4)</b>	34.8 (0.4)	54.7 (1.2)	57.1 (1.3)	57.6 (1.1)

**Tabla 3-3:** Precisión general (%) de las diferentes técnicas de clasificación antes y durante de la admisión y antes de la selección de atributos

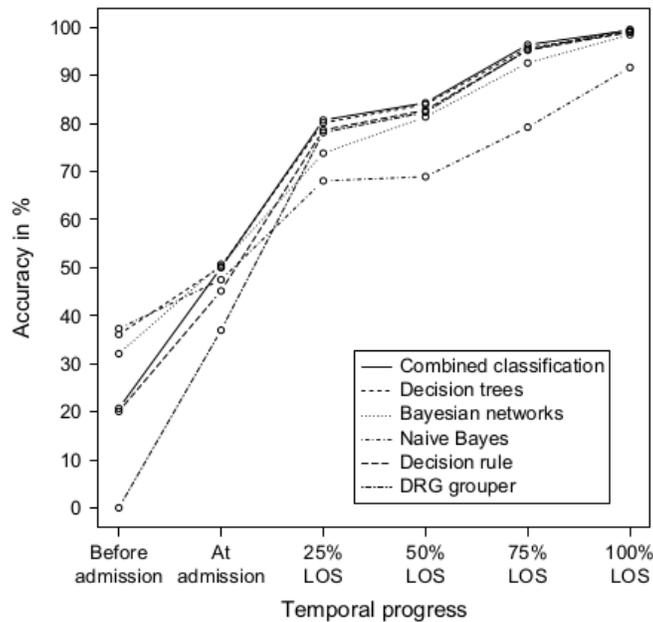
k	RND	BN	PA	NB	Rules	Tree	Vote
1	0.1 (0.0)	72.8 (1.4)	<b>79.5 (2.1)</b>	60.0 (1.3)	75.8 (1.7)	76.3 (1.8)	76.3 (1.6)
2	0.1 (0.0)	73.5 (3.1)	<b>78.9 (3.5)</b>	65.6 (1.6)	75.2 (1.8)	76.0 (1.8)	76.0 (2.4)
3	0.1 (0.0)	38.7 (4.0)	33.5 (2.2)	43.8 (2.8)	20.1 (0.3)	<b>45.4 (2.9)</b>	43.3 (3.7)
4	0.1 (0.0)	69.4 (1.4)	<b>73.2 (4.0)</b>	61.3 (2.0)	70.2 (2.3)	70.8 (3.1)	70.9 (3.8)
5	0.1 (0.0)	68.9 (3.0)	<b>72.4 (3.0)</b>	62.5 (2.2)	69.7 (2.4)	70.5 (3.3)	70.1 (2.9)
6	0.1 (0.0)	37.1 (4.7)	31.1 (1.3)	41.2 (2.3)	20.1 (0.3)	<b>44.0 (2.5)</b>	41.6 (2.8)
7	0.1 (0.0)	53.2 (6.1)	53.4 (3.8)	48.1 (3.7)	45.2 (3.3)	53.5 (1.8)	<b>54.5 (3.0)</b>
8	0.1 (0.0)	52.7 (5.2)	52.6 (4.2)	49.7 (3.1)	45.2 (3.3)	52.9 (1.8)	<b>54.4 (5.0)</b>
9	0.1 (0.0)	32.1 (2.2)	22.6 (2.1)	37.3 (5.2)	20.1 (0.3)	<b>36.1 (2.1)</b>	36.6 (2.7)
Media	0.1 (0.0)	55.4 (4.0)	55.2 (2.9)	52.2 (2.7)	49.1 (1.7)	58.4 (2.3)	58.2 (3.1)
10	0.2 (0.0)	59.8 (3.2)	<b>64.9 (2.1)</b>	55.5 (2.2)	61.7 (1.2)	64.5 (2.0)	64.1 (2.4)
11	0.2 (0.0)	55.7 (1.2)	<b>60.0 (1.5)</b>	52.7 (1.7)	57.3 (1.4)	59.7 (2.0)	59.4 (1.9)
12	0.2 (0.0)	50.7 (1.0)	49.8 (1.2)	47.5 (0.9)	45.2 (1.5)	50.2 (2.1)	<b>51.0 (1.8)</b>
Media	0.2 (0.0)	55.4 (1.8)	<b>58.2 (1.6)</b>	51.9 (1.6)	54.7 (1.4)	58.1 (2.0)	58.2 (2.0)

**Tabla 3-4:** Precisión general (%) de las diferentes técnicas de clasificación antes y durante de la admisión y después de la selección de atributos.

En la tabla 3-4 relaciona los resultados de precisión luego de hacer el proceso de selección de atributos para siete de las ocho técnicas de clasificación, en este análisis se omite los clasificadores Clasificación de GRD Basado en Descomposición (DDC) debido al bajo rendimiento del experimento anterior y *DRG-grouper* porque es superado por varios de los clasificadores propuestos como; el enfoque de PA, los árboles de decisión y la votación superan al clasificador tradicional en cada uno de los conjuntos de datos dados.

Otro de los resultados obtenidos en el proyecto de investigación fue evaluar la clasificación temprana (predicción) de pacientes en GRD, para esto evaluaron los

modelos de clasificación planteados en diferentes momentos de la estancia hospitalaria del paciente, antes de la admisión, en el momento de la admisión, en el 25%, 50%, 75% y 100% de los días de estancia hospitalaria del paciente. Por ejemplo, para un paciente que haya estado 12 días en hospitalización, evaluaron la precisión de los clasificadores en seis momentos de su estancia hospitalaria; antes de que el paciente ingrese (para el caso de paciente electivo), durante la admisión, en los días 3, 6, 9 y finalmente el día 12 cuando fue dado de alta.



**Gráfica 3-1:** Precisión de clasificación temporal para diferentes enfoques antes de la selección de atributos y optimización de parámetros.

En la gráfica 3-1 se muestra un comparativo de la precisión (eje *y*) de seis clasificadores en los seis momentos de la estancia hospitalaria del paciente (eje *x*). Como es de esperarse, los resultados de predicción son más pobres antes de la admisión, ya que la información utilizada por el clasificador es más escasa, mientras que en los últimos días de la estancia hospitalaria la predicción mejora ya que la información usada es mayor. En la tabla 3-5 se muestra un comparativo del desempeño del *DRG-grouper* y el enfoque PA usando los conjuntos de datos 9 y 12 (ver tabla 3-2) que son los escenarios más frecuentes en los registros de pacientes, para cada uno de los seis momentos de

evaluación de la estancia hospitalaria del paciente. Uno de los grandes logros de esta investigación fue demostrar que el enfoque de ML, con información escasa, permite predecir en cierta medida el GRD de un paciente. Antes de la admisión el enfoque PA predice el GRD de 20 de cada 100 pacientes, mientras que el enfoque tradicional (*DRG-grouper*) no tiene la capacidad de predecir ningún paciente. En la tabla 3-5 también se observa que cuando se evalúan los modelos en el alta del paciente, es decir, cuando se tiene la información del 100% de la estancia hospitalaria, los resultados entre ambos enfoques son prácticamente iguales.

	antes de la admisión	en la admisión	25% *	50% *	75% *	100% *
<i>DRG-grouper</i>	0,1	37,0	78,1	82,3	95,3	99,5
PA	20,8	50,0	82,7	84,3	96,3	99,4

\* porcentaje de la estancia hospitalaria del paciente

**Tabla 3-5:** Comparativo de precisión entre *DRG-grouper* y el enfoque PA en el tiempo

En términos generales, el enfoque de Promedio de Probabilidad (PA – Probability Averaging) logra hasta un 79,5% de precisión en la clasificación temprana (predicción) de pacientes antes de la admisión, es decir, cuando los pacientes electivos se comunican con el hospital para la admisión. En el momento de la admisión hospitalaria, es decir, cuando se ingresa un paciente electivos y no electivos, este enfoque lograr un máximo de 65,5% de precisión de clasificación general.

Durante la etapa de entrenamiento se identifica también que incluso con el enfoque de clasificación con peor desempeño y un mínimo de información, son necesarias clases con más de 2500 registros de pacientes hospitalizados para superar el desempeño del *DRG-grouper* en el momento de la admisión de paciente. Finalmente, conectar la información de los GRD de los métodos de ML en un modelo de asignación de recursos reveló dos resultados principales: el margen de contribución se puede aumentar y los recursos hospitalarios, como quirófanos y camas, se pueden asignar de manera más efectiva en comparación con el uso de la información de un *grouper*.

Uno de los logros obtenidos en este proyecto de investigación fue que introdujeron técnicas de selección y clasificación de características con el fin de realizar una clasificación temprana de los Grupos Relacionados por el Diagnóstico para los pacientes hospitalizados. Demostraron que el conjunto de características del paciente se puede reducir a un conjunto de características muy relevantes y las redundantes se pueden filtrar.

### **3.4.2. Gartner et al. (2015): “Improving Hospital-Wide Early Resource Allocation through Machine Learning”.**

En este artículo, se investiga si los métodos de ML pueden aumentar la precisión de la clasificación del GRD de un paciente hospitalizado en comparación con el enfoque actual que consiste en utilizar un agrupador de GRD (*DRG-grouper*). En particular, el objetivo es clasificar a los pacientes en un GRD antes de la admisión, es decir, predecir el GRD, para mejorar las decisiones de admisión mediante modelos estadísticos novedosos.

La base de datos de este trabajo contiene datos de pacientes hospitalizados de un año que constan de más de 16.000 registros de un hospital de 350 camas. Los resultados muestran que, en general, los enfoques de ML pueden aumentar sustancialmente la precisión de la clasificación temprana de GRD, especialmente para los pacientes electivos (ingreso programado) que se comunican con el hospital antes de la admisión. Además, se demostró que las técnicas de ML combinadas con la programación matemática pueden conducir a mejores decisiones de asignación de recursos.

Para evaluar el modelo utilizaron datos de un hospital de tamaño medio en las cercanías de Munich, Alemania. Para el entrenamiento del predictor se utilizaron datos de alta de pacientes desde enero a junio de 2011 y para la etapa de prueba (test), se evaluó con datos desde el 1 de julio de 2011 hasta el resto del año, donde cada día se resuelve de forma independiente.

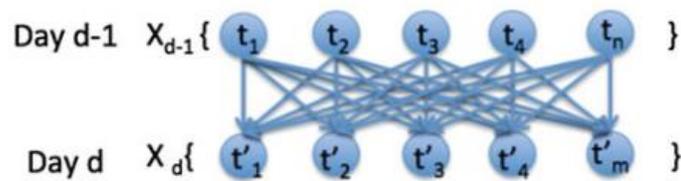
Los resultados de los experimentos revelan que el agrupador tradicional (*DRG-grouper*) no pudo clasificar ninguno de los GRD de los pacientes antes de la admisión. Una explicación de este fenómeno es que el agrupador no puede manejar información de texto libre. Un análisis más detallado reveló que el 960Z siempre se asignaba a los pacientes ya que no podía ser clasificados. Por el contrario, la tasa de verdaderos positivos del enfoque de Naïve Bayes fue, en promedio, del 25,3%, es decir, con el enfoque de Naïve Bayes se observa una mejora en la tarea de predicción en esta misma etapa de clasificación temprana.

Los resultados muestran que, si los hospitales persiguen el objetivo de tomar decisiones de admisión basadas en los GRD, las técnicas básicas de ML como Naïve Bayes pueden conducir a decisiones de asignación de recursos más eficientes en comparación con el uso del enfoque actual de un hospital que utiliza un *DRG-grouper*. El análisis del margen de contribución reveló que programar pacientes usando GRD, según la clasificación temprana (predicción) de Naïve Bayes puede aumentar el margen de utilidad del hospital, según los parámetros de entrada del modelo y su incertidumbre. Los resultados revelaron una mejora de hasta un 17% en este margen de utilidad.

Un logro importante de este proyecto de investigación es el uso de ML para la predicción de GRD en pacientes que no han sido hospitalizados, cuando en la forma tradicional este proceso no es de predicción sino de clasificación, es decir, se asigna el GRD al paciente cuando este ha sido dado de alta. Aunque la precisión de la predicción se puede considerar baja (25,3%), se debe tener en cuenta que esta predicción se hizo con la escasa información proporcionada del paciente antes de la admisión, mientras que la clasificación tradicional se hace con el 100% de la historia clínica del paciente. Y más evidente el logro cuando se observa que la predicción del *DRG-grouper* es del 0%, que en todos los casos evaluados el resultado fue una clasificación de “*error-DRG*”.

### 3.4.3. Nakai et al. (2016) “Prediction of Clinical Practices by Clinical Data of the Previous Day Using Linear Support Vector Machine”.

El objetivo de este trabajo es proponer un modelo basado en la hipótesis de que las prácticas clínicas pueden predecirse a partir de los datos de los procedimientos clínicos realizadas el día anterior para un mismo paciente, la figura 3-1 ilustra esta descripción del modelo planteado. Concretamente, se crearon predictores aplicando ML a un conjunto de prácticas clínica de todos los datos de acción clínica  $d$  y  $d-1$ .



**Figura 3-1:** Concepto del modelo de predicción

En este artículo, se utilizan datos de DPC (*Diagnosis Procedure Combination*) como datos de acciones clínicas. Los datos de DPC se acumulan mediante la combinación de procedimientos de diagnóstico / sistema de pago por día (*DPC / PDPS - Per-Diem Payment System*). El sistema de combinación de DPC / PDPS es el sistema de pago de tasa fija basado en la combinación de procedimientos de diagnóstico (DPC). El DPC se implementó considerando las circunstancias clínicas en Japón según el programa de GRD de EE.UU. Los propósitos de DPC y la relación DPC/PDPS son la mejora de los costos y servicios médicos. DPC/PDPS se introdujo por primera vez en 82 hospitales de tratamiento avanzado que tenían 66.497 camas en total en abril de 2003, y se introdujo en 1.585 hospitales que tenían 492.206 camas en total en abril de 2014.

La información proviene de unos archivos (Archivo  $F$ ) para la comparación de servicios médicos, donde se acumulan las acciones clínicas para los pacientes, estos archivos incluyen detalles de las acciones clínicas. En este archivo, las acciones clínicas a los

pacientes se representan como el código de categoría de datos y el código de recibo electrónico unificado. La Tabla 3.6 muestra ejemplos de códigos de categorías de datos.

No	Código de categoría de datos	Nombre de categoría de datos
1	11	Tarifa de primera consulta
2	13	Educación
3	14	Cuidado en casa
4	21	Medicamentos para uso interno
5	31	Inyección intramuscular

**Tabla 3-6:** Ejemplos de categorías de datos

Los datos usados para este proyecto de investigación corresponden a registros de pacientes entre el periodo de abril a junio de 2006 de 6.445 hospitalizaciones representadas en 118.150 días de hospitalizaciones y 2.543.205 números de registros del archivo *F*. Por otra parte, si dos consultas tienen el mismo código de recibo electrónico unificado y un código de categoría de datos diferente, es posible considerar una práctica clínica diferente. Por lo tanto, se consideraron 5.279 combinaciones de “código de categoría de datos” y “código de recibo electrónico unificado” como objetivos de predicción.

En este trabajo de investigación se utiliza una SVM lineal como técnica de Machine Learning, el cual maneja los datos como un vector, por lo tanto, en este experimento se manejaron los datos como vectores de 5.279 dimensiones. Cada vector se asigna a una de 5.279 prácticas clínicas.

La Tabla 3.7 muestra el resultado del experimento con "*CT simple*". Se calcula la precisión y el recall para los 10 conjuntos de datos de prueba construidos. Se halla el promedio para determinar las métricas finales, los resultados fueron 0,9711 y 0,4910 respectivamente.

Target test data	True positive (TP)	False Positive (FP)	False negative (FN)	True negative (TN)	Precision = TP/(TP + FP)	Recall = TP/(TP + FN)
1	95	4	127	10945	0.9596	0.4279
2	101	7	118	10945	0.9352	0.4612
3	76	4	136	10955	0.9500	0.3585
4	98	4	127	10942	0.9608	0.4356
5	97	5	120	10949	0.9510	0.4470
6	122	1	106	10941	0.9919	0.5351
7	104	1	100	10965	0.9905	0.5098
8	105	2	105	10958	0.9813	0.5000
9	103	1	117	10949	0.9904	0.4682
10	106	0	107	10957	1.0000	0.4977
The average precision and recall (result of 10-folds cross validation)					0.9711	0.4910

**Tabla 3-7:** Métricas de desempeño por cada folds y general

El proceso contempla “50 operaciones”, “60 exámenes”, “70 diagnósticos de radiología e imagen”, “80 otros”, “31 infusión subcutánea”, “32 infusión intravenosa”, “21 medicamentos para uso interno”, “22 pociones” y “23 medicina externa” como categorías de práctica clínica. Se muestra un ejemplo de prácticas clínicas y su tasa de precisión y recall de “32 infusiones intravenosas” ver Tabla 3.8.

Receipt computerized system code	Medical specification name	Number of appearance	Precision	Recall
646130282	Thienam 500 mg	182	1.0000	0.9970
642450054	Decadron 1.65 mg	85	1.0000	0.8429
646130137	Pentacillin 2 g	81	1.0000	1.0000
646130267	Modacin 1 g	74	1.0000	0.9687
640451009	Omepral Injection 20	68	1.0000	0.9774
640433045	Cefamezin $\alpha$ 2 g	58	1.0000	1.0000
640454012	Gemzar 1 g	53	1.0000	0.5897
640454012	Gemzar 200 mg	46	1.0000	0.6368
646130369	Carbenin 0.5 g	203	0.9982	0.9451
646130265	Sulperazon 1 g	185	0.9967	0.9636
646130269	Rocephin 1 g	182	0.9957	1.0000
620000339	Lasix 20 mg	167	0.9951	0.9786

**Tabla 3-8:** Ejemplo de precisión y *recall* de infusión intravenosa.

Uno de los logros importantes de este proyecto de investigación es la utilización de procedimientos para la predicción como operaciones, exámenes, diagnósticos de radiología e imagen, entre otros y no solo los procedimientos que requieren el uso de recursos estériles (Por ejemplo: quirófanos) para la predicción y/o clasificación de

pacientes como lo hace la clasificación de GRD. Otro logro para resaltar es la codificación *One-hot-Encoding* de las variables de conjunto de datos que va a ser la entrada al clasificador, se hacen los conjuntos más grandes, pero es una forma clara de asignar números a las variables, sobretodo para los clasificadores que no admiten texto en los datos de entrada.

#### **3.4.4. Krämer et al. (2017) “Classification of hospital admissions into emergency and elective care: a machine learning approach”.**

El objetivo de este trabajo de investigación fue desarrollar un modelo de clasificación temprana de los ingresos hospitalarios según el diagnóstico principal del paciente como atención de emergencia o atención electiva y predecir la urgencia como un valor numérico. Un valor de urgencia mayor a 50, le corresponderá la etiqueta atención de emergencia (“*emergency*”) y si por el contrario, el valor de la urgencia es menor o igual 50, la etiqueta será atención electiva (“*elective*”). Esta última corresponde a los casos donde la situación en que el ingreso del paciente puede ser claramente programado.

El conjunto de datos contiene cada episodio hospitalario en Alemania para los años 2005–2013. Cada egreso contiene información demográfica, fecha y hora de cada admisión y alta por departamento hospitalario, diagnóstico primario y secundarios (hasta 90 códigos CIE-10), procedimientos con fecha y hora (hasta 100 códigos OPS, que son la versión alemana de los códigos de procedimiento CIE-9 para pacientes hospitalizados) y un identificador de hospital. Se excluyeron maternas, pacientes recién nacidos y todos los egresos de pacientes con diagnósticos principales que se codifican menos de una vez por mes en promedio en toda Alemania. Luego de aplicar las restricciones anteriores, el conjunto de datos completo comprende 119 millones de egresos de pacientes. Además, se utilizaron datos sobre la población alemana para diferentes grupos de edad de 2007 a 2011. Finalmente, el conjunto de datos contó con características de datos administrativos, episodios de paciente hospitalizado, juicios médicos (“*emergency*” o “*elective*”) que corresponde a las etiquetas que va a tener cada registro de egreso según

sea el caso. Para la predicción del paciente en atención se utiliza la técnica de ML Random Forest (RF).

Para la definición de “*emergency*”, se usan dos fuentes: 1) El algoritmo de la Universidad de New York desarrollado por Billings, es el método más común y está basado en el diagnóstico principal. Un panel de expertos del departamento de emergencias y médicos de atención primaria resumió y clasificó 5.700 historias clínicas. 2) El sistema de clasificación de gravedad (SCS – *Severity Classification System*), basado en diagnósticos para servicio médico de emergencia para niños, el cual usan igual un panel de expertos.

En general, los datos de entrenamiento comprenden el 35,3% de los diagnósticos primarios con la etiqueta “*emergency*” y el 64,7% la “*elective*”, se contó con aproximadamente 24 millones de registros hospitalarios individuales para generar las variables predictoras.

El objetivo fundamental de la definición de las variables predictoras era extraer características que contribuyen a la clasificación. Se determinaron 82 predictores entre los cuales se encontraron variables como: tiempos de admisión, departamentos del hospital, duración de la estadía, operación y código de procedimiento, retraso en la operación y en el código de procedimiento, diagnósticos secundarios, entre otros. Todos los predictores son continuos y estandarizados (centrados y escalados). Se utilizó la imputación de  $K$  vecinos más cercanos para las variables predictoras faltantes.

Para evaluar la importación de las variables predictoras se utiliza la función del Random Forest que evalúa la importancia de las variables, se utiliza la permutación aleatoria de las variables (utilizando muestras OOB – *Out of Bag*) para determinar cómo varía la precisión de la predicción en la ausencia de variables. Si un predictor es importante, su permutación debería degradar notablemente la precisión de la clasificación. En general, aunque la técnica RF es muy robusta frente a predictores redundantes y altamente

correlacionados, una cantidad baja de predictores podría conducir a un mejor rendimiento de predicción OOB.

Para la codificación de los diagnósticos, se utilizó el CIE-10-GM, el cual es el CIE-10 adaptado a la terminología y epidemiología alemana. Cuando intentaron extrapolar la información al CIE-10 para dar los resultados más globales, 28 de los 7.481 diagnósticos principales del CIE-10-GM no se pudieron asignar al CIE-10. Esto representó 37.439 registros de los 119 millones.

La evaluación indica que utilizar el algoritmo de ML Random Forest para la clasificar de la atención de “*emergency*” y “*elective*” funciona muy bien. Se hace un estudio de sensibilidad variando los parámetros, generando 12 modelos y comparando los valores de AUC se encuentra que, entre el modelo con el peor desempeño y el modelo con el mejor, solo hay una diferencia de 0,12%.

El mejor modelo obtiene una sensibilidad de 97,0 y una especificidad de 93.6. Por lo tanto, el clasificador final detecta muy bien tanto la atención de emergencia “*emergency*” como la atención electiva “*elective*”. Este modelo también obtiene una precisión general del 95,81% y un intervalo de confianza que oscilan entre el 95,49 y el 96,11% de precisión.

Usando la propiedad del RF, importación de características, se detecta que las variables que más aportan a la clasificación están relacionadas con los tiempos en los que se realiza una admisión, por ejemplo; la admisión entre las 6 AM y las 12 M es la variable predictora con mayor importancia, esto se pueden analizar como: El ingreso de un paciente típico para una admisión planificada o cirugía programada se puede realizar en este rango de hora, mientras que, una admisión fuera del horario laboral, por ejemplo, en la noche o durante el fin de semana, está relacionado con admisiones no planificadas y una demanda alta de los servicios de urgencia hospitalaria que no pueden esperar hasta el día siguiente.

Otra característica que es un predictor importante son los códigos de procedimiento. Así, el momento en que se realiza el primer procedimiento después de la admisión indican atención de emergencia o electiva. Por ejemplo, si el primer procedimiento quirúrgico se realiza fuera de las horas consideradas típicamente como laborales o en la noche, donde en ambos casos es notable la escasez de recurso, es probable que la urgencia sea alta (“*emergency*”). Por el contrario, si es en horas consideradas laborales, por ejemplo, entre las 7 y las 11 de la mañana sugiere planificación y baja urgencia.

En vista que no se pudo correlacionar el 100% de la información del CIE-10-GM con el CIE-10 para entregar resultados más generales y no solo para la epidemiología alemana, los modelos entregan resultados para ambas nomenclaturas. Así, el modelo de clasificación arroja clase (“*elective*” y “*emergency*”) y probabilidades de clase ( $0 \leq \text{urgencia} \leq 100$ ) para 7.481 diagnósticos de CIE-10-GM y 5.659 diagnósticos del CIE-10, para los cuales la precisión fue del 99,8% y 99,5%, respectivamente, del número total de pacientes hospitalizados.

No	ICD-10	Descripción	Admisión (Alemania 2005 a 2013)	Clase (1: “ <i>emergency</i> ” 0: “ <i>elective</i> ”)	Urgencia (0% a 100%)
1	I501	Insuficiencia ventricular izquierda	> 1.000.000	1	0,6401792
2	I48	Fibrilación y aleteo auricular	> 1.000.000	0	0,3457627
3	I10	Hipertensión esencial (primaria)	> 1.000.000	1	0,7875431
4	I702	Aterosclerosis de las arterias de los miembros	> 1.000.000	0	0,2452723
5	K4090 *	Hernia inguinal, Unilateral o sin indicación del lado, sin atrapamiento y sin gangrena: no se denomina hernia recurrente	> 500.000 - 1.000.000	0	0,02

\* Código CIE-10-GM

**Tabla 3-9:** Ejemplo de resultados obtenidos en la clasificación.

En la tabla 3-9 se pueden observar algunos ejemplos de los resultados obtenidos por el clasificador. Se observa en la tabla que los códigos de los diagnósticos CIE-10 se utilizaron códigos de 2 y 3 caracteres con el fin de mejorar la clasificación de los

pacientes. En el numeral 5 de la tabla se observa un ejemplo de código CIE-10-GM, del cual no hubo equivalencia con el CIE-10.

Uno de los logros importantes de este trabajo fue la utilización de los tiempos de realización de los procedimientos como características de clasificación, se observa que el orden cronológico de la realización de procedimiento aporta considerablemente a la clasificación de los pacientes. Un segundo logro de este estudio fue la utilización de los códigos CIE-10 en diferentes niveles, por ejemplos, utilizar códigos con 3, 4 y 5 caracteres para hacer más específico el diagnóstico del paciente hospitalizado. Esto es de suma importancia ya que el reembolso de los hospitales alemanes para pacientes hospitalizados incluye pagos basados en casos de GRD y estos están directamente relacionados con la clasificación basada en los códigos CIE.

#### **3.4.5. Okamoto et al. (2018) “Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days”.**

En este artículo, los autores desarrollaron y evaluaron un método para seleccionar códigos de combinación de procedimientos de diagnóstico (*DPC - Diagnosis Procedure Combination*) precisos basado en la información del tratamiento que se almacena antes del alta del paciente. La DPC es una clasificación de grupos de diagnóstico creada a imitación de los Grupos Relacionados por el Diagnóstico (GRD) utilizados en los EE.UU. Los GRD y DPC son clasificaciones según el diagnóstico principal, la presencia de complicaciones, operaciones y tratamientos, y otros criterios relevantes. Los GRD y DPC se han utilizado para analizar enfermedades y estados de gestión hospitalaria.

En este estudio, los autores utilizan las técnicas de ML Naïve Bayes, SVM y la combinación de ambas técnicas. Varios estudios previos han demostrado que la precisión de clasificación puede aumentar cuando se combinan métodos de ML. El método combinado elimina los niveles de confianza de las técnicas de ML a través del

análisis de Regresión Logística, entonces para obtener la clasificación del método combinado se utiliza las puntuaciones de clasificación de la primera y segunda clase más probable.

El conjunto de datos para el proyecto contó con los casos de todos los pacientes que ingresaron después del 1 de abril de 2008 y fueron dados de alta antes del 31 de marzo de 2009 en el Hospital Universitario de Kyoto. El número de casos presentados al sistema DPC / PDPS fue de 14.460 y el número de códigos DPC relacionados con estos casos fue de 1.133. Al generar modelos de clasificación, las técnicas de ML requieren una cantidad mínima de datos correspondientes a cada clase, por esta razón, solo se eligieron códigos DPC que tuvieran al menos 20 muestras en el conjunto de datos. Finalmente, el número de códigos DPC fue 166 y el número de casos asignados a estos códigos fue de 10.123 registros. La Tabla 3-10 muestra la lista de los 10 principales códigos DPC en términos del número de casos.

DPC code	DPC name	# of cases
020200xx970xxx	Macula and posterior pole degeneration, presence of an operation.	352
060050xx97100x	Cancer of liver and intrahepatic bile duct, presence of another operation, presence of operation and treatment 1.	291
050050xx9910xx	Angina and chronic ischemic heart disease, presence of operation and treatment 1.	242
020110xx97x0x0	Cataract and disease of the lens, presence of an operation, one eye.	215
100070xxxxxx0x	Type 2 diabetes except for diabetic ketoacidosis.	201
070560xx99x0xx	Autoimmune disease with generalized organ dysfunction.	198
020110xx97x0x1	Cataract and disease of the lens, presence of an operation, both eyes.	194
120020xx99x41x	Cancer of cervix and uterine body, presence of operation and treatment 2.4, presence of a complication.	180
030250xx991xxx	Sleep apnea, presence of operation and treatment 1.	179
120010xx99x31x	Cancer of ovary and uterine adnexa, presence of operation and treatment 2.3, presence of a complication.	174

**Tabla 3-10:** Códigos DPC y el número de casos.

Se hicieron dos experimentos, el primero consistió en evaluar la factibilidad de la precisión en la selección automática del código DPC aplicando Naïve Bayes, SVM y la combinación de ambas técnicas. Se midió la precisión de la selección automática de los 14 dígitos de los códigos DPC y la precisión de los primeros seis dígitos, que indican el nombre de la enfermedad. Por otro lado, analizaron que un sistema que selecciona cinco posibles candidatos de código DPC puede ser más efectivo que uno que solo seleccione

un código de uno posible. Hicieron variaciones a los métodos para que generaran cinco posibles códigos DPC de 14 dígitos y midieron su capacidad para producir un código DPC correcto de los cinco. Como parte de la experimentación, también hicieron que los métodos generaran cinco posibles candidatos de los primeros seis dígitos de los códigos DPC y midieron su capacidad para producir una secuencia correcta de esos seis dígitos. Sin embargo, el método combinado no se puede utilizar para seleccionar códigos DPC adicionales, porque el método combinado seleccionado para este proyecto de investigación solo puede escoger un código DPC mediante la selección de un método de ML. Por lo tanto, el método combinado selecciona los cuatro restantes de los cinco posibles códigos DPC de acuerdo con el método Naïve Bayes. El segundo experimento consistió en hallar las mismas mediciones del experimento 1, pero en los días 1, 2, 3, ..., 19, 20 de la estancia hospitalaria del paciente.

Los resultados del experimento 1 se muestran en la tabla 3-11, en la columna dos se muestra la precisión en la selección de código DPC utilizando los 14 dígitos del código para los tres clasificadores y la columna tres muestra la precisión para la selección de los primero seis dígitos de código DPC. En las columnas cuatro y cinco se observan los resultados de precisión para la predicción de uno de los cinco escogidos por la técnica.

Machine Learning Methods	Precision (%)		Prec. of top 5 (%)	
	14 digits	6 digits	14 digits	14 digits
Naïve Bayes method	63.2	71.5	87.6	90.6
SVM	72.1	75.7	86.4	89.8
Combined method	73.8	78.7	89.9	93.2

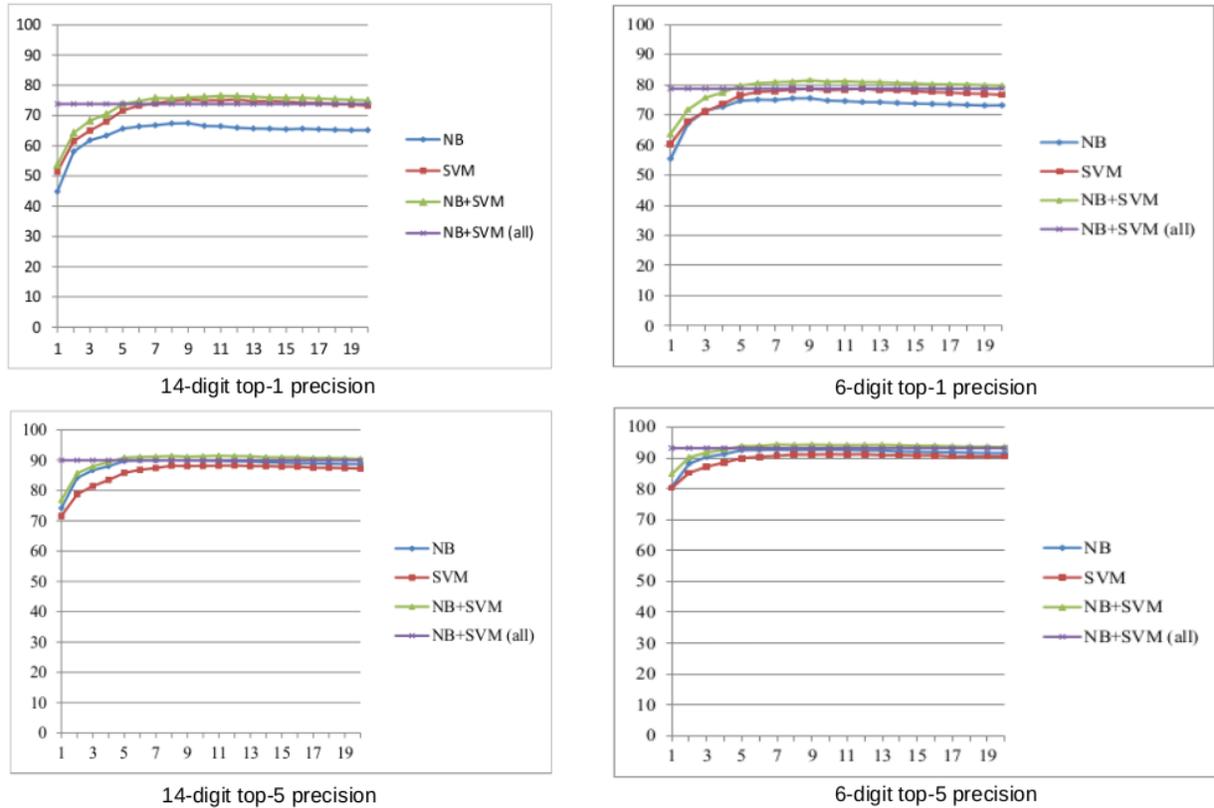
**Tabla 3-11:** Medidas de precisión del experimento 1.

El método que combinó las dos técnicas Naïve Bayes y SVM arrojó mejor desempeño que cada una de ellas por separado. La precisión del método combinado en el modelo que evalúa un código de 14 dígitos entre un top cinco y la precisión del modelo que evalúa un código de seis dígitos entre un top cinco fueron del 89,9% y 93,2%, respectivamente; estos niveles de precisión demuestran que el método combinado es útil.

En el experimento 2, como se ilustra en la gráfica 3-2, los valores de precisión de los cuatro modelos se maximizan a medida que el número de días de hospitalización se acerca a 10, lo que demuestra que la selección del código DPC basada en datos de tratamiento parcial en relación con el número de días de hospitalización es ideal y que los tratamientos característicos tienden a administrarse al inicio de la hospitalización. En realidad, los tratamientos característicos, como operaciones y medicamentos costosos, se administran al inicio de la hospitalización y parece que ayudan a determinar con precisión los códigos DPC en función de los tratamientos característicos.

Uno de los logros importantes de este trabajo de investigación es determinar la importancia que tienen en la precisión de clasificación los procedimientos y exámenes que se realizan a los pacientes hospitalizados a medida que avanzan los días de estancia clínica, lo cual está en coherencia con la clasificación tradicional de los GRD (DPC en este trabajo), en la medida en que más información se tenga del paciente mejor es la precisión de clasificación temprana (predicción).

También es de destacar que utilizan la separación de los códigos para evaluar la precisión, se utilizan los códigos DPC completos, que son los 14 dígitos y dan cuenta de la enfermedad, los tratamientos y la información del paciente como género y edad, y el código DPC pero solo con 6 dígitos, que son los que dan cuenta de la enfermedad del paciente, sin especificar más información. Esto demuestra que si bien los resultados son muy similares cuando se evalúan por separado, es posible que utilizarlos en conjunto podría mejorar los resultados de precisión.



**Gráfica 3-2:** Precisión de los 4 modelos para la selección de los códigos DPC en relación con el # de días de hospitalización.

### 3.4.6. Kuo et al. (2018) “Comparison of Models for the Prediction of Medical Costs of Spinal Fusion in Taiwan Diagnosis-Related Groups by Machine Learning Algorithms”.

Los objetivos de este estudio fueron comparar el rendimiento de técnicas de ML para la predicción de los costos médicos asociados con la fusión espinal en términos de ganancias o pérdidas en los Grupos Relacionados por el Diagnóstico de Taiwán (Tw-DRG) y aplicar estos métodos para explorar los factores importantes asociados con los costos médicos de la fusión espinal.

Se obtuvo un conjunto de datos de un hospital regional en la ciudad de Taoyuan en Taiwán, que al momento del estudio contaba 972 camas. Los datos contenían información de egresos de pacientes entre los años 2010 y 2013, clasificados en el GRD

Tw-DRG49702 (fusión espinal posterior y de otro tipo sin complicaciones o comorbilidades). Así, como en la modalidad de reembolso de los GRD, estos eran pacientes a los que el hospital se dispone a pedir reembolso a la aseguradora. El conjunto de datos incluye características básicas de pacientes (edad, género, obesidad, estado de diabetes y presión), fecha de admisión, fecha de alta, diagnóstico principal, códigos CIE-9 de complicaciones o comorbilidades, órdenes médicas, y costos. Finalmente, las variables de predicción utilizadas fueron: género, edad, diagnóstico principal, complicaciones o comorbilidades, número de complicaciones y comorbilidades, número de jaulas intervertebrales y días de estancia.

La etiqueta de clase se definió como “*loss*” para aquellos pacientes cuyos costos médicos excedieron el pago fijo Tw-DRG49702, lo que significa que el hospital sufrió una pérdida monetaria, y “*non-loss*” para los pacientes cuyos costos médicos cayeron por debajo del pago fijo del GRD Tw-DRG49702, en estos casos el hospital no sufrió pérdidas económicas. El conjunto de datos contó con 532 muestras, de las cuales 124 (23,3%) etiquetados con “*loss*” y 408 (76,7%) etiquetados como “*non-loss*”. Para tratar el desbalance de muestras usaron *Synthetic Minority Over-sampling TEchnique* (SMOTE). Para la predicción se emplearon técnicas Naïve-Bayesian, Máquinas de Soporte Vectorial, Regresión Logística, Árbol de Decisión C4.5 y Random Forest utilizando WEKA 3.8.1. Las medidas utilizadas para determinar el rendimiento de las técnicas fueron: Precisión, Sensibilidad, Especificidad y curva ROC.

La duración de la estancia hospitalaria, el número de cajas intervertebrales, el desplazamiento del disco lumbar, la fractura por atresia lumbar y la escoliosis entre los grupos etiquetados como “*loss*” y “*non-loss*”, fueron factores importantes asociados con los costos médicos del Tw-DRG49702. Sin embargo, no se observaron diferencias significativas en términos de género, edad, número de complicaciones o comorbilidades, estenosis lumbar, presión arterial alta, ciática, espondilolistesis, diabetes, espondilolistesis adquirida y osteoporosis entre los grupos con pérdida y sin pérdida monetaria.

Evaluando los resultados de precisión en las diferentes técnicas de ML, los valores obtenidos fueron de 76,68% para Naïve-Bayesian hasta 84,30% para el Random Forest. El RF fue el mejor en la predicción logrando también una sensibilidad de 71,40%, una especificidad de 92,20%, y un AUC de 0.904, seguido por la Regresión Logística, SVM, Árbol de Decisión C4.5 y el peor desempeño para Naïve-Bayesian. En la tabla 3-12 se observan los resultados completos de los desempeños de las diferentes técnicas usadas.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Naïve Bayesian	76.68	56.90	88.70	0.815
SVM	82.16	66.10	91.90	0.790
Logistic regression	82.16	69.80	89.70	0.860
C4.5 decision tree	78.51	68.50	84.60	0.770
Random forest	84.30	71.40	92.20	0.904

**Tabla 3-12:** Comparación del rendimiento de las diferentes técnicas de predicción.

El estudio demostró que el RF se puede utilizar para predecir los costos médicos de Tw-DRG49702 (fusión espinal posterior y otra fusión espinal sin complicaciones o comorbilidades), y con base en los factores importantes identificados, este estudio puede informar la estrategia del hospital en términos de aumento o disminución en la eficiencia de gestión de este tipo de operaciones en términos económicos. Además, técnicas de este tipo también se pueden utilizar para abordar problemas relacionados, como predecir los costos de otros GRD.

Hubo algunas limitaciones en el estudio. Primero, la precisión de este modelo fue del 84,30%, lo que significa que todavía existen otros factores potenciales que podrían afectar los costos médicos de la fusión espinal. En segundo lugar, el estudio solo se realizó en un único hospital y con un tamaño de muestra pequeño. Una de las recomendaciones que hace el estudio, es implementar estas técnicas en hospitales más grandes.

### 3.5. Conclusiones.

En la tabla 3-1 se evidencia que el tema de clasificación temprana (predicción) de pacientes en GRD usando técnicas de ML lleva algunos años con escasos artículos reportados en la literatura científica. Si bien no todos los trabajos previos hablan explícitamente de predicción o clasificación temprana, todos tienen en común el uso de técnicas de ML para la clasificación de pacientes en GRD o su equivalente según el país donde se desarrolle y en todos los casos expuestos el objetivo es contribuir a la mejora en el uso de los recursos sanitarios o mejorar las prácticas clínicas.

Si bien los GRD es un sistema de clasificación de pacientes que se usa a nivel mundial, los países que reportan avances en la combinación de GRD y ML son Japón, Taiwán y algunos europeos como Alemania y Reino Unido. En la literatura científica no se reportan avances en este ámbito en países desarrollados de Occidente como Estados Unidos y Canadá y mucho menos en países en vía de desarrollo como los de Latinoamérica. Esto último puede deberse a que como se evidenció en los artículos anteriores, en la mayoría de los trabajos de investigación, los países caso de estudio cuentan con su sistema GRD adaptado a la epidemiología y terminología de cada país, como es el caso de Alemania, Japón y Taiwán que tienen sus equivalentes G-DRG, CPD y Tw-DRG, respectivamente. Latinoamérica no cuenta con un equivalente de GRD que pueda ser análogo al resto del mundo, pero conservando su propia epidemiología y terminología.

En cuanto a la estrategia de selección de características se puede evidenciar que:

- No se requiere el 100% de las características para obtener resultados aceptables en la clasificación y predicción de GRD.
- Es crucial el uso del diagnóstico principal, es decir, con el que se da el alta del paciente, sin este dato, se observó que los resultados de clasificación y predicción eran pobres.
- En la mayoría de los casos utilizaron procedimientos secundarios, las comorbilidades del paciente para este proceso, pero también hubo casos en los que se utilizó esta codificación de diagnósticos completa y/o parcial, como es el

caso de la codificación CIE, algunos trabajos usaron el código de diagnóstico completo y otros parcial, sin especificación de severidad o el diagnóstico exacto.

- Si bien una de las premisas del ML es que no se desechen registros ni características ya que pueden aportar información que ayuda a la precisión en los modelos evaluados. Los estudios aquí relacionados, usan técnicas de clasificación y selección de características y se detectan varias que pueden no aportar o ser redundantes en el proceso y por ende tampoco aportan. En este punto es importante analizar que no es necesario usar el 100% de las variables que contiene el Conjunto Mínimo Básico de Datos.

En cuanto a las metodologías seleccionadas para los trabajos, se destaca que:

- En su mayoría se usaron técnicas de árboles o ensamble de árboles, esto se puede deber a que el algoritmo tradicional es un simple árbol de decisión.
- No solo se usan las técnicas de ML, se proponen metodologías de trabajo, se contempla la posibilidad de implementar las técnicas, pero los datos de entrada van variando según el tiempo en el que se evalúe el modelo.
- La mayoría de los trabajos están diseñados para predecir usando metodologías alternas o clasificación temprana de una enfermedad en particular, es decir, problemas biclase.

En cuanto al rendimiento, en todos los resultados de los estudios se observa una mejora usando las técnicas de ML en comparación con la metodología tradicional implementada en los llamados *DRG-groupers*. Pero es de resaltar que en términos de predicción (clasificación temprana) ni en la metodología tradicional, ni en las alternativas que usan ML se obtienen rendimientos cercanos al 100%, sobretodo cuando la información de entrada es escasa.

Una de las consideraciones a resaltar en los trabajos relacionados, es que la mayoría tienen en cuenta el factor tiempo para hacer el proceso de predicción de pacientes en algún GRD. Por ejemplo, el trabajo de *Gartner. D, 2015* plantea la posibilidad de predecir el GRD de un paciente utilizando diferentes momentos de la estancia hospitalaria como

lo son 0%, 25%, 50%, 75% y 100% de la estancia hospitalaria de los pacientes. *Okamoto. K, 2018* expone en el trabajo que los valores de precisión se maximizan a medida que el número de días de hospitalización a partir de los cuales se usaron los datos se acercó a 10, o sea, entre más días, más alta la precisión en la clasificación temprana (predicción) de pacientes en un GRD. *Nakai. T, 2016* presenta la opción de evaluar los procedimientos a realizar a un paciente evaluando la gestión sanitaria de lo realizado en los días anteriores.

En ninguno de los estudios descritos en esta sección se aborda un enfoque en que se evalúe una predicción de GRD eliminando la variable de diagnósticos secundarios. Recordemos que en *Kuo. C, 2018* concluye que las comorbilidades no tienen alto impacto en los resultados de predicción, esto puede acercarse mucho a la realidad de los hospitales ya que como se reporta en *Gartner. D, 2015*, los escenarios más comunes es donde los registros de pacientes no cuentan con información de diagnósticos y procedimientos.

En términos de predicción, en todos los trabajos descritos, el planteamiento es la realización de la clasificación temprana utilizando múltiples atributos de entrada, entre los que se encuentran información demográfica, diagnósticos y procedimientos. Pero, teniendo en cuenta que en muchas ocasiones la única información con la que se cuenta es el diagnóstico principal, ninguno de los estudios plantea la posibilidad de predecir utilizando solo este atributo como información sanitaria.

Solo uno de los trabajos de investigación descritos anteriormente utiliza un enfoque de predicción multiclase, es un estudio realizado con la epidemiología alemana y con resultados que pueden ser mejores, sobretodo con la predicción del GRD en el momento de la admisión del paciente. Es posible evaluar otros atributos o utilizar otras técnicas para mejorar los resultados de precisión en la clasificación temprana.

# Capítulo 4:

## Base de datos

### 4.1. Introducción.

En este capítulo se presenta una descripción de las bases de datos con las que se contó para el diseño de la metodología desarrollada, se presentan los diferentes atributos proporcionados en las bases de datos originales entregadas por la entidad caso de estudio. Los conjuntos de datos estuvieron representados en dos grandes grupos; el primero llamado el Conjunto Mínimo Básico de Datos (CMBD), el cual contenía toda la información de egresos de los pacientes que estuvieron hospitalizados entre los años 2016, 2017 y 2018. Esta base de datos estaba debidamente anotada con los GRD a los que pertenecía cada paciente, el etiquetado fue hecho por el *DRG-grouper* comercial con el que cuenta el Hospital caso de estudio. El segundo conjunto de datos era un complemento de la base de datos de egresos nombrada anteriormente, que contaba con la información de todos los procedimientos que se realizaron a un número determinado de pacientes, pertenecientes al llamado CMBD.

Inicialmente se realiza una descripción de los esquemas de codificación que utiliza el Hospital para identificar los diagnósticos y procedimientos que asigna a los pacientes hospitalizados. Luego se hace una descripción del CMBD y las características seleccionadas para la evaluación de los clasificadores. Posteriormente, se hace una descripción de las codificaciones empleadas para el manejo del conjunto de datos como

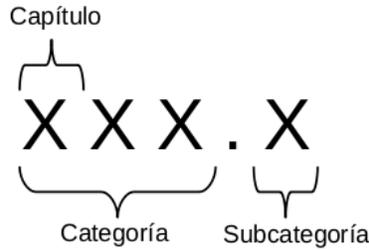
entrada a los clasificadores y finalmente se hace una descripción de la cohorte de estudio.

## 4.2. Esquemas de codificación del CMBD.

### 4.2.1. Codificación CIE-10.

CIE-10 es la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud – 10ª Revisión, en inglés *ICD-10 - International Statistical Classification of Diseases and Related Health Problems - 10th revision*. El objetivo de la codificación CIE es poder hacer un registro sistemático de la información, que permita hacer análisis, interpretaciones y comparaciones entre los datos de mortalidad y morbilidad de pacientes reportados en diferentes países y en diferentes épocas. Se utiliza para convertir la terminología médica de diagnósticos y otros problemas de la salud en códigos alfanuméricos que facilitan su almacenamiento y recuperación para hacer análisis posteriores.[44]

La CIE está dividida en 22 capítulos, donde cada capítulo hace referencia a las Categorías de Diagnóstico Principal (*Major Diagnostic Categories – MDC*) ordenados en grupos. Cada capítulo corresponde a la patologías o motivos de consulta por cada sistema corporal o bien enfermedad y causa de la misma. Cada capítulo contiene una lista de categorías de tres caracteres, cada una de las cuales puede ser dividida en diez (máximo) subcategorías de cuatro caracteres. Se utiliza un código alfanumérico que contiene una letra en la primera posición y números en las posiciones dos, tres y cuatro. El cuarto carácter está precedido por un punto decimal. Por lo tanto, los códigos van de A00.0 a Z99.9.[45]. Ver figura 4-1.



**Figura 4-1:** Estructura del código CIE-10.

En la tabla 4-1 se listan los capítulos en los cuales está dividido el CIE-10, cada capítulo contiene la información de un sistema de órganos o etiología que son mutuamente excluyentes y donde se evidencia que en la mayoría está representado por una letra única del alfabeto romano, es decir, en el atributo denominado *Diagnósticos CIE10 - MDC* cada letra corresponde a una enfermedad relacionada con un sistema corporal específico, excepto la letra **D**, que se usa tanto en el *Capítulo II - Tumores (Neoplasias)*, como en el *Capítulo III - Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan el mecanismo de la inmunidad*; y la letra **H**, que se usa tanto en el *Capítulo VII - Enfermedades del ojo y sus anexos* como en el *Capítulo VIII - Enfermedades del oído y de la apófisis mastoideas*. Cuatro capítulos (capítulos I, II, XIX y XX) usan más de una letra en la primera posición de sus códigos. [44] Esta división en sistemas que expone el MDC es relevante para este trabajo de investigación, porque como se describe en la sección 4.3.2 *Información de diagnóstico*, uno de los atributos seleccionados para las matrices de entrada a los modelos propuestos es este primer elemento alfanumérico del código CIE-10.

Capítulo	Códigos	Título
I	A00-B99	Ciertas enfermedades infecciosas y parasitarias
II	C00-D48	Tumores [neoplasias]
III	D50-D89	Enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan el mecanismo de la inmunidad
IV	E00-E90	Enfermedades endocrinas, nutricionales y metabólicas
V	F00-F99	Trastornos mentales y del comportamiento
VI	G00-G99	Enfermedades del sistema nervioso

VII	H00-H59	Enfermedades del ojo y sus anexos
VIII	H60-H95	Enfermedades del oído y de la apófisis mastoides
IX	I00-I99	Enfermedades del sistema circulatorio
X	J00-J99	Enfermedades del sistema respiratorio
XI	K00-K93	Enfermedades del sistema digestivo
XII	L00-L99	Enfermedades de la piel y del tejido subcutáneo
XIII	M00-M99	Enfermedades del sistema osteomuscular y del tejido conjuntivo
XIV	N00-N99	Enfermedades del sistema genitourinario
XV	O00-O99	Embarazo, parto y puerperio
XVI	P00-P96	Ciertas afecciones originadas en el período perinatal
XVII	Q00-Q99	Malformaciones congénitas, deformidades y anomalías cromosómicas
XVIII	R00-R99	Síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte
XIX	S00-S98	Traumatismos, envenenamientos y algunas otras consecuencias de causas externas
XX	V00-Y98	Causas externas de morbilidad y de mortalidad
XXI	Z00-Z99	Factores que influyen en el estado de salud y contacto con los servicios de salud
XXII	U00-U99	Códigos para propósitos especiales

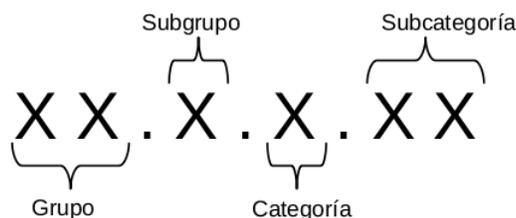
Fuente: [45] World Health Organization. (2004). International statistical classification of diseases and related health problems (Vol. 1). World Health Organization.

**Tabla 4-1:** Capítulos en que se divide el CIE-10.

#### 4.2.2. Codificación CUPS.

CUPS es la Clasificación Única de Procedimientos en Salud, los cuales corresponden a un ordenamiento lógico y detallado de los procedimientos y servicios en salud que se realizan en Colombia, según los describe la Resolución número 0005851 de 2018 del Ministerio de Salud y Protección Social. El código, de seis caracteres, está conformado por niveles jerárquicos que constituyen su estructura que representan cada

procedimiento o servicio de salud. La conformación del código permite ubicarlo con exactitud según el nivel jerárquico, tanto en forma general como detallada de manera sistemática y concatenada. En la figura 4-2 se ilustra la estructura del código.



**Figura 4-2:** Estructura del código CUPS

- **Grupo:** representa el capítulo en el cual se encuentra ubicado, señala: La zona anatómica específica para los Capítulos 01 al 14; la unidad de producción específica para los Capítulos 15 al 24; el proceso en la colectividad, para el Capítulo 25; las condiciones, factores y tipos de proceso en el trabajo para los Capítulos 26 y 27; y Los servicios de salud para el Capítulo 28.
- **Subgrupo:** identificado con el tercer elemento alfanumérico del código que representa el tipo de procedimiento, el tipo de imagen, el tipo de área técnica, el tipo de acción, el tipo de estrategia, el tipo de fase en la atención, el tipo de nivel institucional o territorial y el tipo de servicios.
- **Categoría:** está representada por el cuarto elemento alfanumérico e indica en forma genérica o global la nomenclatura del procedimiento.
- **Subcategoría:** corresponde a los dos últimos caracteres que definen con mayor precisión y detalle el procedimiento genérico de acuerdo a variables como: especificidad en la zona anatómica de la región operatoria o diagnóstica, técnica, tecnología, método, extensión, disciplina del conocimiento, agente etiopatogénico, tipo de muestra, entre otras.

La Lista Tabular de los códigos CUPS consta de cinco secciones: *Sección 00 - Procedimientos quirúrgicos*, *01 - Procedimientos no quirúrgicos*, *02 - Procedimientos e intervenciones sobre la comunidad, su entorno y salud*, *03 - Procedimientos e intervenciones hacia la protección de la salud de los trabajadores* y *04 - Servicios de Salud*. En la tabla 4-2 se ilustran algunos ejemplos de esta codificación CUPS.

No.	Sección	Código	Descripción
1	00	47.1.1.10	Apendicectomía vía laparoscópica
2	01	87.0.0.03	Radiografía de base de cráneo
3	02	A4.1.2.02	Uso racional de insecticidas
4	03	T1.0.0.03	Estudio de iluminación
5	04	S0.1.1.00	Consulta externa de complejidad baja

**Tabla 4-2:** Ejemplos de códigos CUPS por sección.

### 4.3. Descripción del Conjunto Mínimo Básico de Datos (CMBD).

La información es originada en la IPS Universitaria sede Clínica León XIII, un Hospital de Tercer Nivel de complejidad en Medellín Colombia, el cual según los indicadores de gestión de 2018, contaba con 626 camas distribuidas en: Hospitalización adultos 531, Hospitalización pediátrica 21 y UCI-UCE 74. Este volumen proporciona inicialmente 110.779 registros de egresos de pacientes de los años 2015, 2016, 2017 de todo el año y 2018 hasta octubre. Son cuatro bases de datos separadas por año, la información es llamada “Conjunto Mínimo Básico de Datos - CMBD”, donde cada registro representa la información de egreso de pacientes que se divide en tres grandes grupos:

1. Información demográfica como edad, género, fechas de ingreso y egreso, EPS, entre otras.
2. Información de diagnósticos, diagnóstico principal que hacen referencia a la causa primaria de consulta del paciente y/o diagnósticos secundarios que hacen

referencia a las posibles complicaciones o comorbilidades que puede presentar el paciente, estos diagnósticos están codificados con el estándar CIE-10.

3. Información de procedimientos, procedimiento principal y procedimientos secundarios, esta información está disponible en pacientes que debido a su condición requirieron consumir recursos hospitalarios como quirófanos o pruebas diagnósticas como imagenología entre otras.

Esta información del egreso es necesaria para que un software de clasificación tradicional de Grupos Relacionados por el Diagnóstico conocido como *DRG-grouper*, pueda etiquetar cada egreso de paciente con un GRD.

Año	2015	2016	2017	2018
Egresos	27.869	28.204	30.169	24.537*
Características	89	67	165	168

\* Estos registros corresponden a la información de egresos hasta octubre de 2018

**Tabla 4-3:** Tamaño de los archivos de CMBD.

Como se puede observar en la tabla 4-3, las cuatro bases de datos que contienen la información de año a año son diferentes entre sí. La información del 2015 cuenta con 89 atributos o variables, la del 2016 con 67, la del 2017 con 165 y finalmente el archivo de 2018 cuenta con 168 características. Cabe destacar que el archivo CMBD del 2015 fue etiquetado con uno de los tantos softwares llamados *DRG-grouper*, mientras que en los otros años se usó un clasificador diferente; es decir el método para etiquetar la base de datos cambió del 2015 respecto a los años restantes. Por esta razón, se decidió descartar la información del año 2015 del proceso de selección de la cohorte de estudio. Luego de esta decisión, la base de datos seleccionada hasta el momento estaba representada por 82.910 registros de egresos y 168 variables.

Es claro que, así se elimine la información de 2015, seguirá existiendo disparidad entre las características de las bases de datos de los años 2016 al 2018, esto sin importar que fueron procesadas y etiquetadas con el mismo software *DRG-grouper*. Por esta razón y

teniendo en cuenta tanto los artículos relacionados en el estado del arte en la sección 3.4, donde se evidencia que para el proceso de predicción o clasificación se pueden usar características del CMBD con la mayor importancia según su análisis, como la experiencia de expertos locales, se decide seleccionar solo algunas características que están presentes en el CMBD. Los atributos seleccionados se relacionan en la tabla 4-4. En estas nueve características seleccionadas se evidencian los tres grupos de información descritos anteriormente.

No	Atributo	Tipo de dato	Posibles valores del atributo
1	Mes de la admisión	nominal	12 (e.g. Diciembre)
2	Edad en años documentada en el momento de la admisión	continua	108 (1 año, . . . ,99 años, 102 años)
3	Género	nominal	2 (Masculino, Femenino)
4	Tipo de admisión	nominal	2 (Urgencia, Programada)
5	Grupo etario	nominal	5 (De_28_dias_a_1_año, De_2_año_a_14_años, De_15_años_a_44_años, De_45_años_a_74_años y Mayor_de_75_años)
6	Diagnósticos CIE10 - MDC	nominal	20 (A; C; D; Z...)
7	Diagnósticos CIE10 - (La letra y dos dígitos)	nominal	1057 ('A01_', 'A02_', 'A03_', 'A04_', ... , 'Z94_', 'Z98_')
8	Diagnósticos CIE10 - Código completo	nominal	1692 ('A020', 'A030', 'A031', ... , 'Z940', 'Z944')
9	Procedimientos CUPS	nominal	1592 (10111, 10115, 1102 ,...,P00993, S22223)

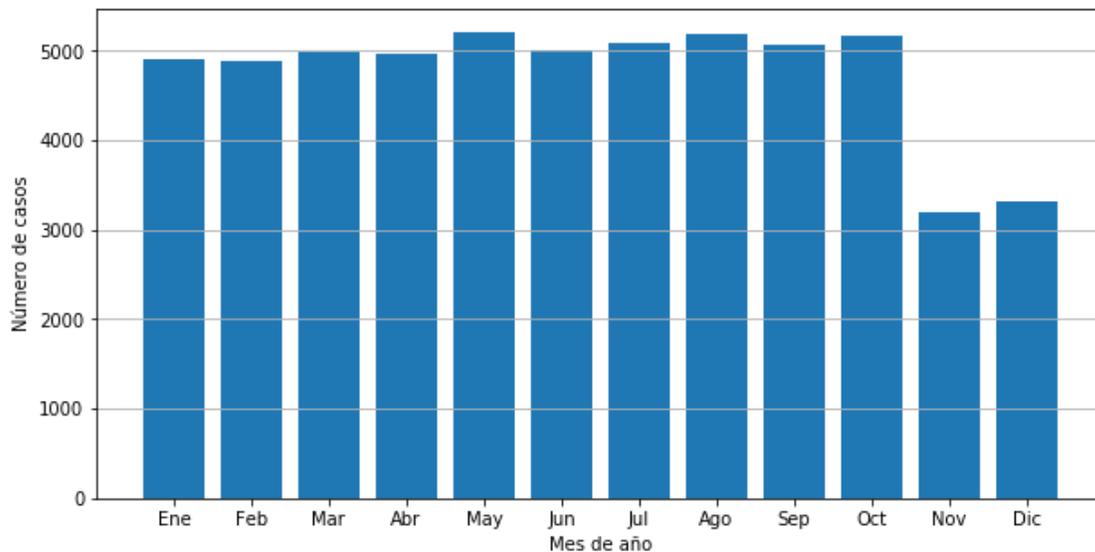
**Tabla 4-4:** Descripción de atributos resultantes.

El CMBD es un conjunto de datos completo y confiable, esto debido a; 1) como se describe en la sección 1.3.2, la información de diagnósticos y procedimientos se almacena en la historia clínica del paciente, desde su ingreso, es decir, la apertura del número de atención, y es evolucionada por el médico tratante cada 24 horas para pacientes en piso y cada 12 horas para pacientes en UCI y UCE, hasta que el paciente es dado de alta; y 2) debido a que esta información es enviada al clasificador *DRG-groupes* que maneja el hospital para la clasificación tradicional de GRD, esta información

es revisada y validada por un personal asistencial que se encarga de comparar la información narrativa de la historia clínica con la codificación hecha para los diagnósticos y procedimientos.

### 4.3.1. Información demográfica.

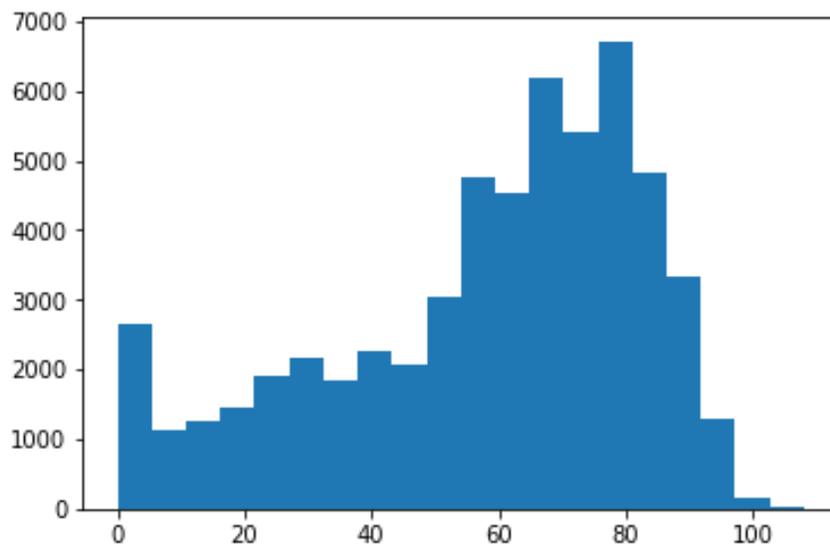
- **Mes de admisión:** relaciona el mes en que el paciente fue admitido al hospital. Es una variable nominal y toma los valores de 1 a 12, de Enero a Diciembre. En la gráfica 4-1 se relacionan el número de ingresos de pacientes por cada mes del año, en este caso la cantidad es la sumatoria de los ingresos de cada mes por los tres años de evaluación. Los meses con más ingresos fueron mayo y agosto con 5209 y 5195 respectivamente y los dos meses con menos ingresos fueron noviembre y diciembre con 3.202 y 3.319 ingresos de pacientes respectivamente.



**Gráfica 4-1:** Relación de número de ingresos por cada mes del año.

- **Edad en años documentada en el momento de la admisión:** hace referencia a la edad del paciente en el momento de la admisión, donde la menor edad pertenece a los pacientes con cero (0) años, los cuales corresponden a infantes entre los 28 y 364 días de nacido y la edad más avanzada es de 108 años. En la

gráfica 4-2 se relacionan las frecuencias de las edades de los registros de egresos de pacientes. Aquí se evidencia que la mayoría de los pacientes son adultos mayores, con una media en 58,02 y una desviación estándar de 24,79.



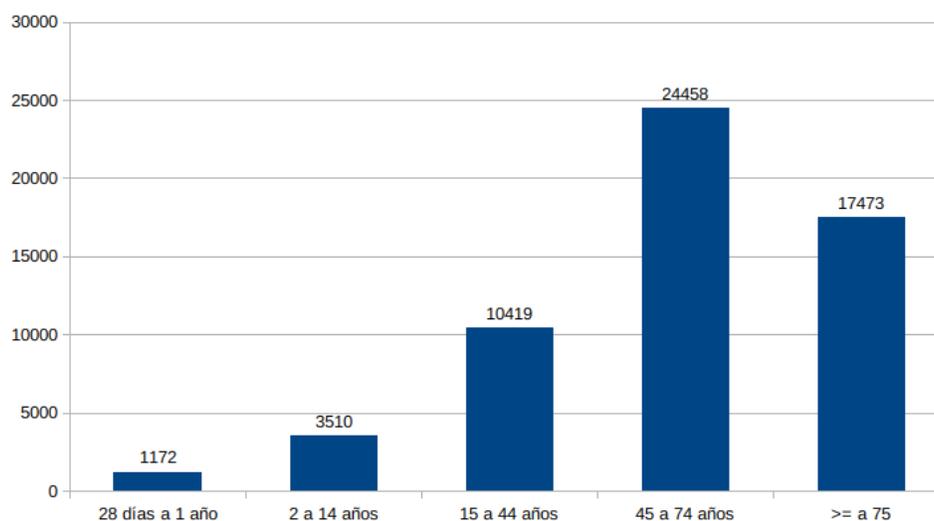
**Gráfica 4-2:** Frecuencia de edad de pacientes.

Esta distribución de edades está acorde con las patologías atendidas por el Hospital caso de estudio, el cual no es un centro neonatal y es por esta razón que no hay registros con pacientes neonatales, es decir, con 28 días o menos de nacido.

- **Género:** se encuentran 27.118 hombres que representan 47,55% y 29.914 mujeres correspondientes al 52,45% del total del conjunto de datos.
- **Tipo de admisión:** hace referencia a la forma en que ingresa el paciente a hospitalización. El hospital maneja dos tipos de ingresos, Urgencia y Programado, el primero se refiere a pacientes que ingresaron por el área de urgencias hospitalarias debido a algún problema que pone en riesgo la integridad o incluso la vida del paciente y este ingreso está definido por el Triage que se hace cuando el paciente ingresa al área. Los casos Programados hacen referencia a dos tipos de pacientes; 1) paciente que contacta al hospital antes de ingresar, este es el

caso de pacientes que van a ingresar para cirugía y 2) paciente de traslado, es el paciente que es trasladado a uno de los servicios del Hospital desde otra IPS. Los pacientes que ingresaron por urgencia durante estos tres años de evaluación fueron 50.469 (88,49%) y los programados fueron 6.563 (11,51%).

- **Grupo etario:** Es una variable que utiliza el Hospital para agrupar a los pacientes según las edades, se manejan cinco grupos; 1) Edad entre 28 días y 1 año, 2) Edad entre 2 y 14 años, 3) Edad entre 15 y 44 años, 4) Edad entre 45 y 74 años y 5) Edad mayores o iguales a 75. En la gráfica 4-3 se evidencia que la gran mayoría de los pacientes que se atienden en el Hospital son adultos y adultos mayores ya que el 73,5% son pacientes mayores de 45 años.



**Gráfica 4-3:** Número de pacientes por rango de edades.

### 4.3.2. Información de diagnóstico.

Para efectos de este trabajo de investigación, solo se utiliza el diagnóstico principal en el proceso de clasificación, se descartan los diagnósticos secundarios porque según estudios previos y la disponibilidad de la información en el Hospital caso de estudio, este parámetro no tiene gran aporte en la predicción del paciente en un GRD. Cómo se

relaciona en la sección 4.2.1, los diagnósticos están codificados con el estándar CIE-10. Teniendo en cuenta la literatura y con el objetivo de mejorar el proceso de clasificación, desde el código de diagnóstico principal (Diagnósticos CIE10 - Código completo, por ejemplo: *I110 - Enfermedad cardíaca hipertensiva con insuficiencia cardíaca (congestiva)*) se construyen dos características adicionales; el Diagnósticos CIE10 - MDC (por ejemplo: *I - Capítulo IX: Enfermedades del sistema circulatorio*) y el Diagnósticos CIE10 - (La letra y dos dígitos, por ejemplo: *I10 - Hipertensión esencial (primaria)*).

- **Diagnósticos CIE10 - Código completo:** hace referencia al código CIE-10 que determina el diagnóstico principal registrado en el CMBD. La tabla 4-5 relaciona el top 30 de los diagnósticos principales asignados a los pacientes hospitalizados durante los años 2016 a 2018. Se evidencia que los diagnósticos atendidos están acordes con la complejidad del Hospital caso de estudio.

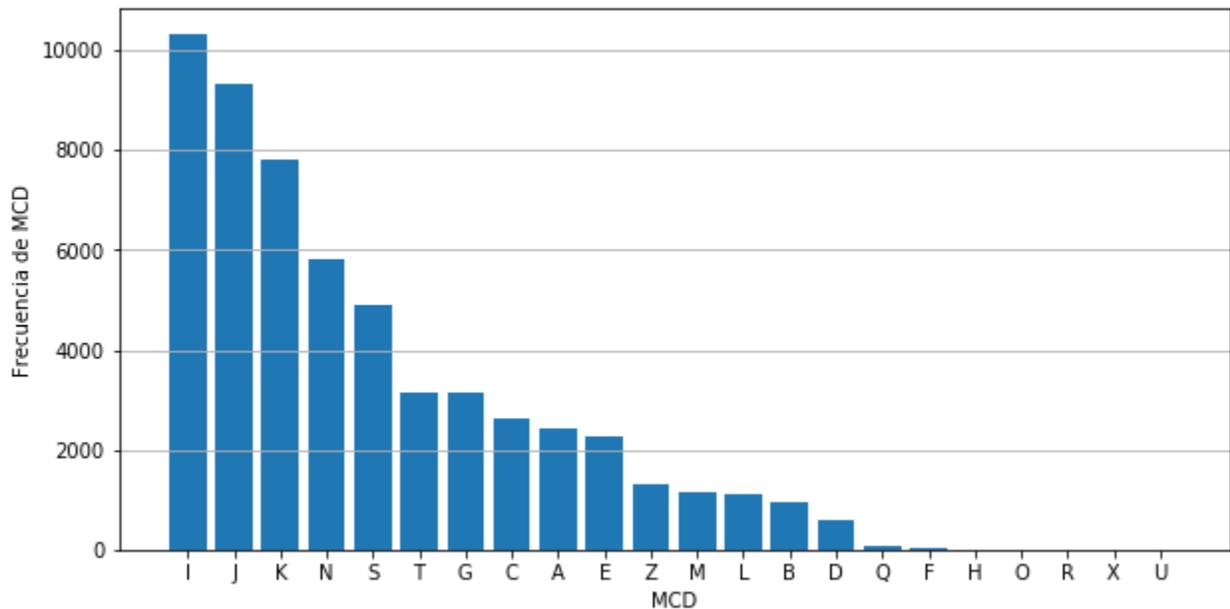
No	Código	Descripción del Diagnóstico	Frec.
1	N390	Infección de vías urinarias, sitio no especificado	3609
2	J441	Enfermedad pulmonar obstructiva crónica con exacerbación aguda, no especificada	2762
3	J188	Otras neumonías, de microorganismo no especificado	2391
4	I678	Otras enfermedades cerebrovasculares especificadas	1598
5	I110	Enfermedad cardíaca hipertensiva con insuficiencia cardíaca (congestiva)	1563
6	Z511	Sesión de quimioterapia por tumor	1244
7	I132	Enfermedad cardiorenal hipertensiva con insuficiencias cardíaca y renal (congestiva)	1069
8	J440	Enfermedad pulmonar obstructiva crónica c/infecc. Ag.de vías respiratorias inferiores	999
9	A09	Otras gastroenteritis y colitis no especificadas de origen infeccioso	980
10	J181	Neumonía lobar, no especificada	980

11	I120	Enfermedad renal hipertensiva con insuficiencia renal	908
12	K800	Cálculo de la vesícula biliar con colecistitis aguda	856
13	I219	Infarto agudo de miocardio, sin otra especificación	717
14	G401	Epilepsia y sindr. Epil.sintomat.relac.c/localizac.(FOCALES) (parc.) Y c/ataques parc.simples	694
15	G408	Otras epilepsias	681
16	I200	Angina inestable	656
17	N201	Cálculo del uréter	610
18	T824	Complicación mecánica de catéter para diálisis vascular	579
19	E100	Diabetes mellitus insulino dependiente con coma	561
20	T814	Infección consecutiva a procedimiento, no clasificada en otra parte	561
21	E107	Diabetes mellitus insulino dependiente con complicaciones múltiples	543
22	A90	Fiebre del dengue (dengue clásico)	456
23	C349	Tumor maligno de los bronquios o del pulmón, parte no especificada	446
24	N200	Cálculo del riñón	433
25	J157	Neumonía debida a mycoplasma pneumoniae	430
26	K290	Gastritis aguda hemorrágica	429
27	K746	Otras cirrosis del hígado y las no especificadas	424
28	K573	Enfermedad diverticular del intestino grueso sin perforación ni absceso	410
29	E105	Diabetes mellitus insulino dependiente con complicaciones circulatorias periféricas	402
30	J46	Estado asmático	387

**Tabla 4-5:** Top 30 de los diagnósticos más frecuentes.

- **Diagnósticos CIE10 - MDC:** hace referencia al primer elemento del código alfanumérico CIE-10 del diagnóstico principal reportado en el CMBD para cada

paciente. En la gráfica 4-4 se ilustra la frecuencia de los capítulos de los MCD (sección 4.2.1), donde los mayores valores de ocurrencia son para I, J y K que son enfermedades relacionadas con los sistemas circulatorio, respiratorio y digestivo respectivamente.

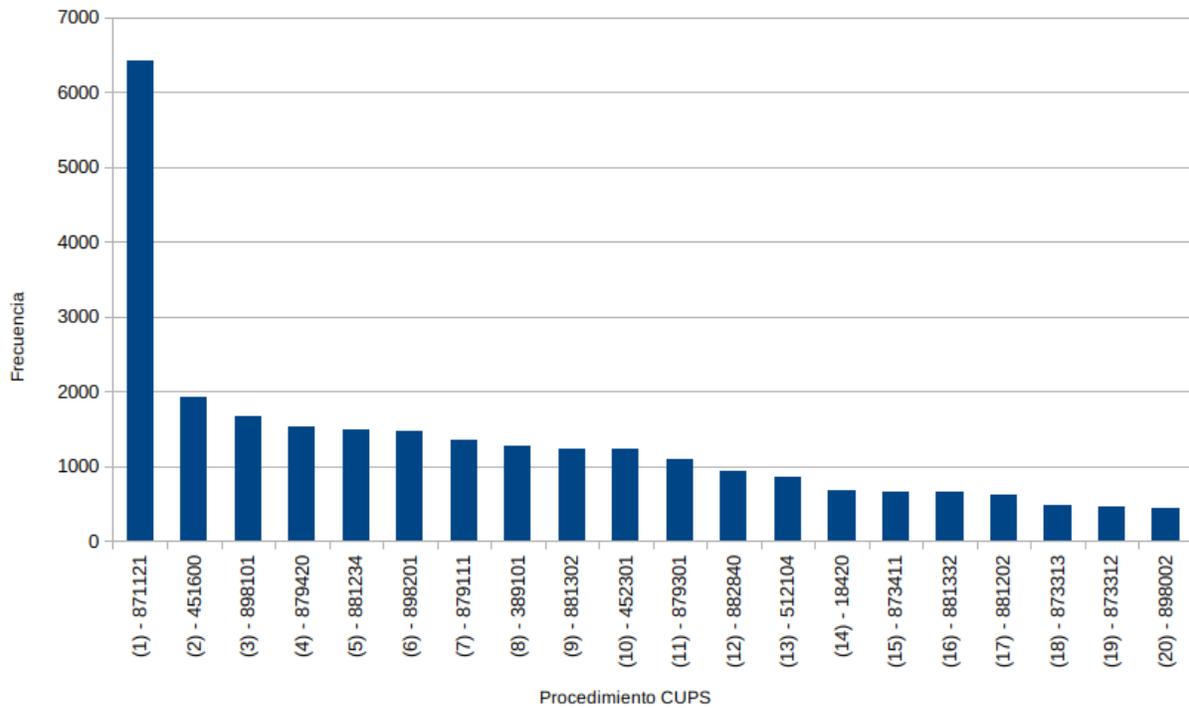


**Gráfica 4-4:** Frecuencia de los MCD.

- **Diagnósticos CIE10 - (La letra y dos dígitos):** describe la categoría de código CIE-10 del diagnóstico principal reportado en el CMBD.

#### 4.3.3. Información de procedimientos.

- **Procedimientos CUPS:** relaciona el código CUPS asignado a cada paciente que requirió como mínimo un procedimiento durante su estancia hospitalaria. Es posible que un registro de egreso de un paciente tenga cero (0) o varios procedimientos asociados.



**Gráfica 4-5:** Frecuencia de los procedimientos CUPS

La gráfica 4-5 relaciona el top 20 de la frecuencia de los procedimientos CUPS registrados en la base de datos durante los tres años evaluados. El procedimiento que más realizó fue el 871121 - *Radiografía de tórax (p.a. o a.p., lateral, decúbito lateral, oblicuas o lateral con bario)*.

#### 4.4. Codificación de características.

Con el objetivo de no afectar los resultados de los clasificadores debido a valores numéricos significativamente diferentes, como es el caso de la edad que tiene un rango de 0 a 108 y teniendo en cuenta también que los clasificadores no reciben datos categóricos, se hace normalización de atributos numéricas y codificación de los categóricos como se describe a continuación.

Para el caso de las variables “Mes de la admisión” y “Edad en años documentada en el momento de la admisión”, se hizo una normalización de las variables, para cada caso se tomaron todos los valores y se dividieron en el valor mayor de este conjunto, como se representa en la ecuación 4-1, de esta forma el rango de estas variables queda entre 0 y 1.

$$Valor = \frac{x_i}{x_{max}} \quad (4-1)$$

Para el caso de las variables “Género” y “Tipo de ingreso”, se realizó una codificación binaria, así; “Género” = {0: Femenino, 1: Masculino} y “Tipo de ingreso” = {0: Programado, 1: Urgencia}.

En cuanto a las variables categóricas “Grupo etario”, “Diagnósticos CIE10 - MDC”, “Diagnósticos CIE10 - (La letra y dos dígitos)” y “Diagnósticos CIE10 - Código completo” se utilizó una codificación “*One-Hot-Encoding*”, la cual es una estrategia que consiste en agregar una columna para cada valor distinto que exista en la característica a codificar y, para cada registro, se marca con un uno (1) la columna a la que pertenezca dicho registro y se marcan las demás con ceros (0).

Finalmente, para la variable “Procedimientos CUPS” se desarrolló una tabla de frecuencia, que a diferencia de la codificación “*One-Hot-Encoding*”, la tabla de frecuencia pone cero (0) si no se realizó el procedimiento o un número diferente de cero, el cual representa el número de veces que se le realizó un procedimiento a un paciente durante su estancia hospitalaria.

#### **4.5. Cohorte de estudio.**

Con el objetivo de trabajar con la mayor cantidad de registros y una representación significativa de Grupos Relacionados por el Diagnóstico, teniendo en cuenta la

epidemiología del Hospital caso de estudio y conservando la estabilidad de los clasificadores por problemas de desbalanceo, se opta por trabajar con el 70% (aproximadamente) de la población registrada en el CMBD entre los años 2016 y 2018. Finalmente, se obtiene que el número de etiquetas únicas, es decir, el número de Grupos Relacionados por el Diagnóstico diferentes para este subconjunto es de 47, las cuales se relacionan en la tabla 4-6.

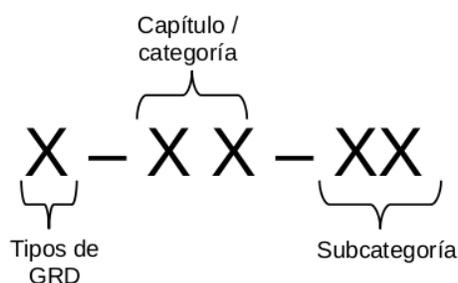
IT	Código GRD	Descripción de GRD	Egresos
1	C-09-07	Neumonía y neumonitis	4656
2	C-09-08	Enfermedad pulmonar obstructiva crónica	3806
3	C-13-04	Enfermedad infecciosa del riñón y vía urinaria	3709
4	C-08-04	Insuficiencia cardíaca	3086
5	C-05-18	Otros trastornos del sistema nervioso	2342
6	C-10-08	Infecciones gastrointestinales y de cavidad abdominal	2191
7	C-04-04	Diabetes	2188
8	C-05-14	Epilepsia y convulsiones	2077
9	C-10-10	Otros trastornos de aparato gastrointestinal	1738
10	C-13-05	Cálculo renal y de vía urinaria	1406
11	C-08-16	Otros trastornos y lesiones cardiovasculares	1335
12	C-10-15	Otros trastornos de vesícula y vía biliar	1334
13	P-10-12	Colecistectomía laparoscópica	1324
14	C-02-02	Admisión para sesión de quimioterapia	1246
15	P-12-13	Otros procedimientos en cadera y fémur excepto artroplastia	1180
16	C-08-01	Síndrome coronario agudo	1122
17	C-13-03	Insuficiencia y fallo renal	1065
18	C-10-02	Gastritis y úlcera gastroduodenal	1023
19	P-08-12	Cateterismo cardíaco y coronariografía con o sin angioplastia sin inserción de stent coronario	999

20	C-05-16	Traumatismo de cráneo	921
21	C-10-01	Enfermedad neoplásica de aparato gastrointestinal & retroperitoneo	898
22	C-09-05	Enfermedad neoplásica de aparato respiratorio y mediastino	894
23	C-10-12	Neoplasia maligna de hígado, vía biliar y páncreas	877
24	P-11-01	Injerto de piel	867
25	C-10-11	Cirrosis, hepatitis y fallo hepático	850
26	C-17-02	Envenenamiento y efectos tóxicos causado por fármacos, drogas y alimentos	793
27	C-12-04	Otra fractura traumática, excepto de pelvis, fémur y columna vertebral	787
28	C-12-07	Traumatismos múltiples, heridas superficiales y lesiones traumáticas articulares, miotendinosas y ligamentarias	782
29	C-11-02	Infecciones de piel, tejido subcutáneo, músculo y tegumentos	777
30	C-12-14	Otros trastornos osteomioarticulares	761
31	C-09-06	Otras infecciones respiratorias excepto neumonía	713
32	C-01-04	Infecciones virales excepto virus de inmunodeficiencia humana	695
33	C-01-06	Infección por virus de inmunodeficiencia humana	652
34	C-08-15	Complicaciones ocasionadas por dispositivos y procedimientos cardiovasculares	639
35	C-05-15	Migraña y otros dolores de cabeza	637
36	C-10-04	Diverticulosis y diverticulitis del intestino delgado y grueso	629
37	P-12-15	Otros procedimientos en rodilla y pierna excepto artroplastia	619
38	P-12-11	Procedimientos en miembros superiores excepto mano	596
39	C-09-09	Asma	570
40	P-10-02	Procedimientos de mayor complejidad sobre tracto gastrointestinal	567
41	C-08-12	Arritmia cardiaca y trastornos de conducción	552
42	C-17-03	Complicación de tratamientos clínicos y/o quirúrgicos	547
43	C-13-09	Otros trastornos de riñón y aparato urinario	546

44	C-05-06	Hemorragia intracraneal no traumática	515
45	C-01-02	Infección asociada a procedimientos clínicos y/o quirúrgicos	515
46	C-11-01	Enfermedades no infecciosas de piel, tejido subcutáneo y tegumentos	510
47	C-10-14	Otros trastornos hepáticos	496

**Tabla 4-6:** 47 etiquetas de GRD más frecuentes en el 70% de los datos de 2016 a 2018.

En la tabla 4-6 donde se describen los GRD seleccionados, la columna del código tiene un código alfanumérico que consta de tres partes como se ilustra en la figura 4-3.



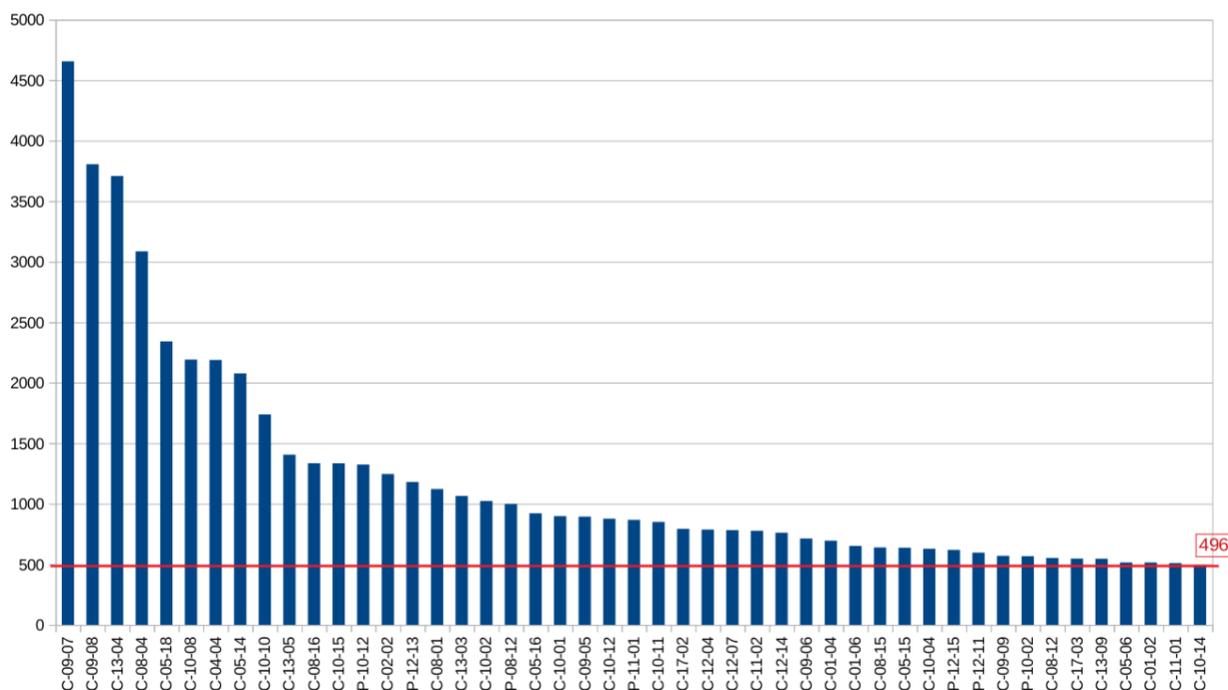
**Figura 4-3:** Estructura del código GRD

Como se describió en la sección 2.2.1 *Clasificación de pacientes en GRD - El método tradicional*, existen dos tipos de GRD; médicos y quirúrgicos, en la codificación del conjunto de datos, este parámetro está dado por el primer elemento alfanumérico del código, llamado *Tipos de GRD* (figura 4-3). Para este caso en particular y como se lista en la tabla 4-6 está representado con dos letras, la **C** para indicar que es un GRD médico y la **P** para indicar que es un GRD quirúrgico, en nuestra cohorte de estudio representan el 74.46% y 25.54% respectivamente. El segundo elemento del código, llamado *Capítulo/categoría* hace referencia a los capítulos de MDC (ver tabla 4-1). Finalmente, la sección del código llamado *Subcategoría*, hace referencia a los códigos en base a la casuística y realidad asistencial.

La gráfica 4-6 ilustra el desbalance que hay entre la clase (GRD) con más registros, conocida como la clase mayoritaria y la que menos tiene (clase minoritaria) en todo el conjunto de datos seleccionado. En el capítulo 5: *Investigación Experimental*, nos

ocuparemos de explicar cuál fue el tratamiento que se le dio a este problema de desbalance de clases.

Finalmente, se obtiene que la cohorte de estudio cuenta con 57.032 registros de egresos correspondientes a los años 2016 hasta octubre de 2018, de los cuales 14.566 son pacientes que durante su estancia hospitalaria le realizaron por lo menos un procedimiento y que estarán clasificados como GRD quirúrgicos y el resto, es decir, 42.466 registros de egresos están catalogados como GRD médicos. Como se describe en la sección 2.2.



Gráfica 4-6: Gráfica de frecuencia de GRD

#### 4.6. Conclusiones.

El producto final obtenido en este capítulo es una base de datos filtrada, depurada y codificada, la cual fue obtenida luego de hacer un análisis de la epidemiología del

Hospital caso de estudio, con la identificación de los GRD que tenían un mayor impacto en el quehacer hospitalario, tanto el costos como en frecuencia de casos y de llevar a cabo la identificación de las características más relevantes que podrían aportar valor al proceso de predicción de pacientes en GRD y dando cumplimiento así a los dos primeros objetivos de este trabajo de investigación.

Una de las características relevantes de la base de datos principal utilizada en este trabajo de investigación, el CMBD, es que los datos están estandarizados y claramente validados por el personal asistencial. Lo primero debido a que este CMBD no contiene información de texto libre, todos los atributos que esta contiene son etiquetas que identifican un valor asociado al paciente, como es el caso de los diagnósticos principales y secundarios que se describen en este capítulo. Lo segundo se debe a que un grupo de personal asistencial está a cargo de la validación de la información que ingresa al *DRG-grouper*, hacen la comparación de la información consignada en la historia clínica de los pacientes con los datos estandarizados en el CMBD y verifican así la confiabilidad de la información.

Uno de los objetivos importantes en el análisis de la información de las bases de datos y la construcción de la cohorte de estudio usados en inteligencia computacional, es determinar la calidad de la información, la disponibilidad de la misma y eliminar la mayor cantidad de ruido que pueden generar algunos datos, que bien están disponibles para usar, no aportan al sistema o lo que puede ser más grave, confunden. Por tal razón, teniendo la información del año 2015 y sabiendo esta está clasificada con el *DRG-grouper* diferente, el cual implementa el algoritmo tradicional con una variante de GRD europea, mientras que los otros años se clasifican con un *DRG-grouper*, que según el fabricante [45] usa una variante desarrollada con base en la epidemiología latinoamericana, que entre otras cosas, la comparación de los CMBD tenía atributos diferentes. Por esta razón se asume que los 27.869 egresos de 2015 en vez de aportar a la mejora en el sistema, puede generar ruido que disminuya el rendimiento de las técnicas implementadas.

# Capítulo 5: Investigación Experimental

## 5.1. Introducción.

Existen varias técnicas que cumplen con los parámetros requeridos para el desarrollo del problema que se plantea en este trabajo de investigación, técnicas de Machine Learning que permiten trabajar con sistemas multiclase, el sistema de clasificación temprana que se plantea aquí tiene como salida 47 clases diferentes, donde uno de los grandes retos es el manejo del desbalance de las clases, la clase con más registros es aproximadamente diez veces más grande que la clase que menos pacientes tiene.

En este capítulo se describen las métricas utilizadas para evaluar el desempeño de las técnicas, se hace una descripción de la metodología propuesta para la clasificación temprana (predicción) de pacientes en un GRD, se da la justificación de por qué se usaron estas técnicas y no otras diferentes o adicionales, una descripción de estas técnicas y los resultados obtenidos con cada una de ellas. Finalmente se encontrará un comparativo entre los resultados obtenidos y las respectivas conclusiones.

## 5.2. Proceso clínico de admisión y evolución de pacientes.

Para poder plantear una metodología para hacer la predicción de pacientes en GRD durante la estancia hospitalaria del paciente, se debe conocer el proceso de admisión y

evolución de los pacientes hospitalizados, para tal efecto en la figura 5-1 se ilustra cómo el Hospital caso de estudio clasifica los pacientes que recibe casos espontáneos con seguridad social o particular, espontáneos sin seguridad social, remitidos y XPRESS. Todos estos ingresan por el área de urgencias, aunque con condiciones diferentes, por ejemplo, el paciente XPRESS, es aquel que llega con una urgencia vital e ingresa para ser atendido sin pasar por el triaje, al igual que el paciente remitido, llega directamente a habitación ya que está siendo remitido desde otra IPS y ya está diagnosticado.

Es importante aclarar en este punto que los egresos casos de estudio de este proyecto de investigación, son los pacientes que están clasificados como hospitalizados según el criterio definido por el Hospital y estos son los que ingresan según la clasificación anterior y cumplen más de 24 horas internados, a partir de este momento es que se consideran como hospitalizado y son los que hacen parte de la base de datos de egresos hospitalarios.

Como parte de los hallazgos de esta etapa de revisión, se encontró que en los procesos estandarizados del Hospital, está que los médicos hacen ronda de pacientes hospitalizados cada 24 horas, la frecuencia de rondas médicas se duplican en pacientes en UCI y UCE, pero en términos generales, se puede concluir que cada 24 horas los pacientes hospitalizados van a tener una actualización en la evolución de su historia clínica, la cual en todos los casos consiste en revisión de pruebas diagnósticas, seguimiento a la respuesta del paciente a los tratamientos o procedimientos médicos realizados y la anamnesis actual del paciente. La diferencia principal radica en que a los pacientes en UCI y UCE se les puede deteriorar su estado de salud rápidamente y por esta razón el seguimiento debe ser más frecuente.

Bajo este contexto se toma la decisión de trabajar con pacientes hospitalizados, los cuales son los relacionados en la base de datos de egresos y de acuerdo con lo investigado en los procesos de atención hospitalización, de estos pacientes se conoce la evolución clínica (información) y frecuencia de actualización de esta evolución (tiempo), lo cual es la entrada para los modelos de predicción. También se decide

emplear modelos para reevaluar el GRD del paciente hospitalizado cada 24 horas, ya que es en estos periodos de tiempo que se alimenta la base de datos con información nueva y puede cambiar o no la predicción del GRD.

La figura 5-1 ilustra todo lo descrito anteriormente, se especifica el tipo de paciente, la clasificación en el triaje, los tiempos de actualización de la evolución de la historia clínica y el responsable de cada evolución.

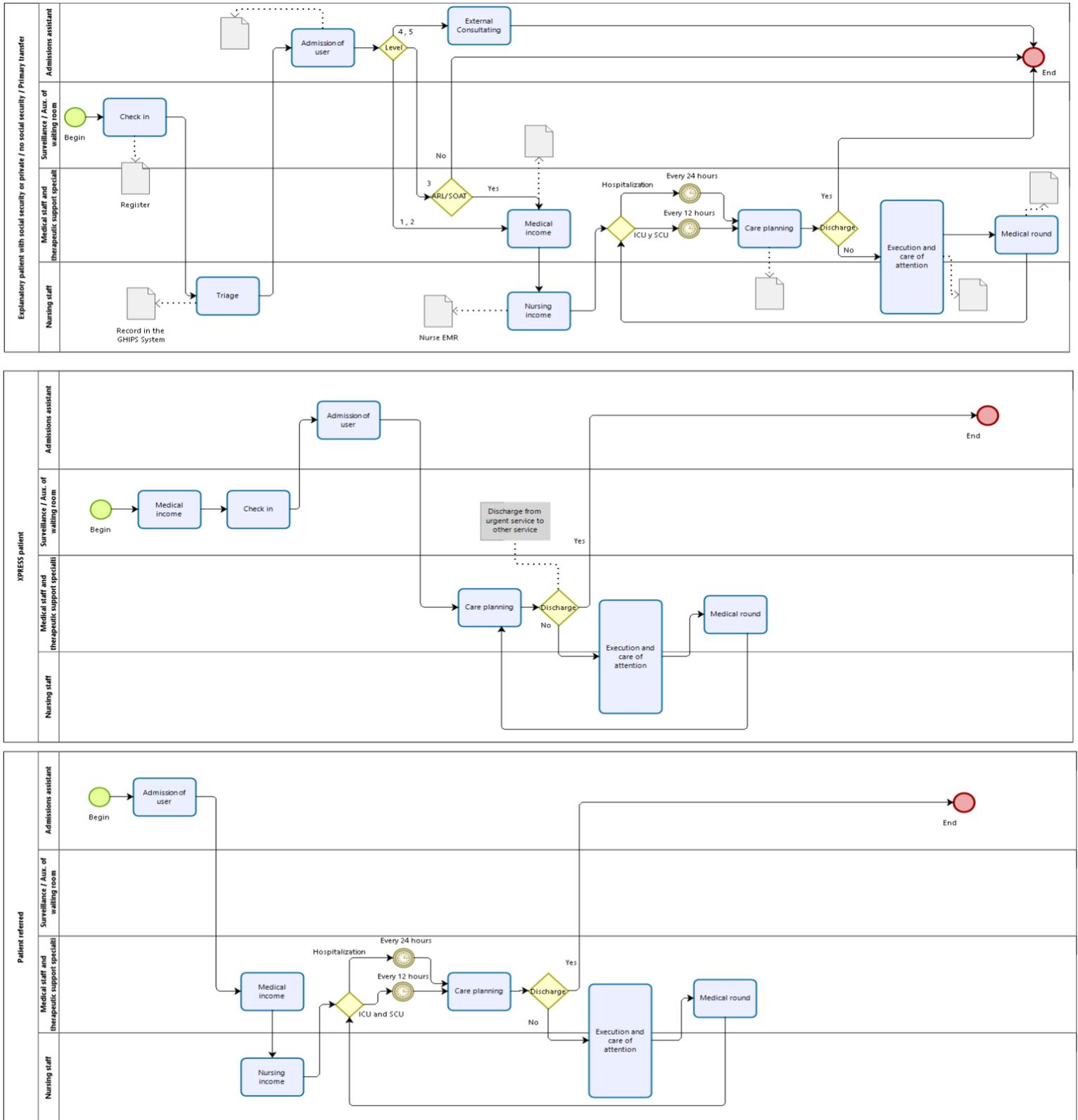


Figura 5-1: Flujo de ingreso y hospitalización de paciente.

### 5.3. Estrategia de evaluación temprana.

El objetivo principal de este trabajo de investigación es determinar la viabilidad de la clasificación temprana (predicción) en GRD de pacientes hospitalizados. Para cumplir el objetivo se adopta como estrategia construir modelos de clasificación en diferentes momentos de la estancia del paciente apoyados en la trazabilidad clínica que se tiene de este desde que es ingresado a hospitalización hasta que se le da el alta.

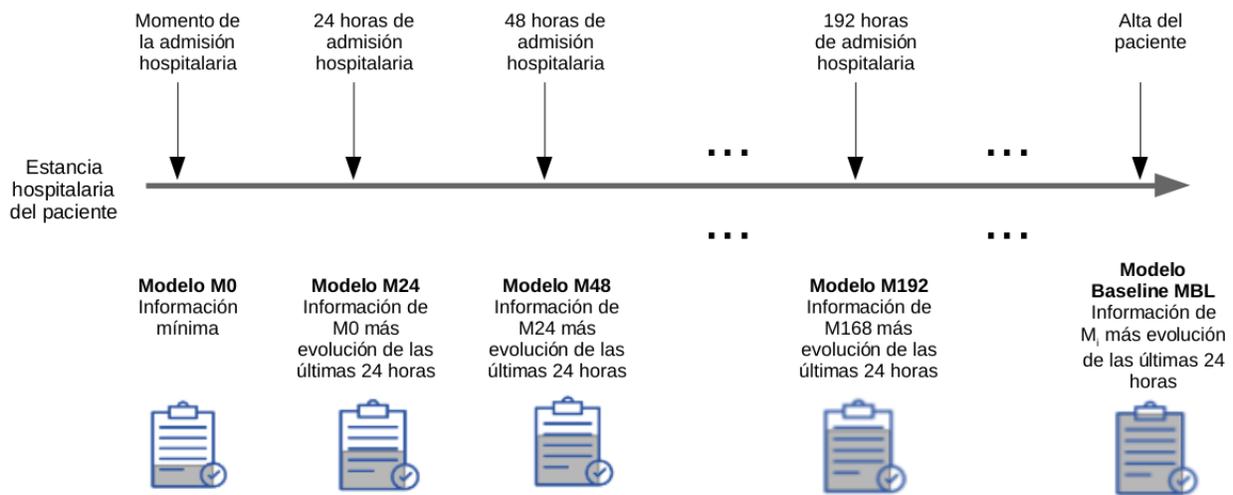
La estrategia adoptada fue la de evaluar la clasificación de GRD de un paciente cada 24 horas desde que es admitido en hospitalización hasta que se le da la alta hospitalaria; es decir, clasificar el paciente en un GRD en el momento de la admisión hospitalaria, cuando la información clínica es escasa ya que solo se contaría con el diagnóstico principal evaluado por el médico que hace el ingreso, luego a las 24 horas de la admisión hospitalaria, donde según la operación del hospital, el médico tratante debió haber evolucionado el paciente y la historia clínica con la actualización del diagnóstico principal, si aplica; diagnósticos secundarios, si aplican; procedimientos, si aplica y pruebas diagnósticas realizadas entre otros, luego a las siguientes 24 horas, es decir, a las 48 horas de la admisión y después de la segunda evolución de la historia clínica del paciente hacer una nueva clasificación de este en GRD y así sucesivamente hasta que sea dado de alta.

Clínicamente se espera que; si la información es actualizada correctamente, la predicción del GRD se mantenga durante toda la estancia hospitalaria; si el médico tratante hace cambios en el diagnóstico principal o en los procedimientos hechos durante las evoluciones anteriores, la predicción de GRD del paciente también pueda cambiar.

Para efectos de materializar esta metodología se desarrollaron nueve modelos llamados M0, M24, ... , M192, donde el modelo M0 se utiliza para predecir el GRD del paciente cuando es admitido, el modelo M24 se utiliza para la predicción de GRD del paciente cuando este lleva 24 horas hospitalizado, M48 para predecir el GRD a las 48 horas de la admisión hospitalaria y así sucesivamente hasta el día ocho de la estancia hospitalaria

(192 horas desde la hospitalización), que sería el modelos M192, para este último caso se espera que si el paciente aún no ha sido dado de alta, tenga documentado al menos las ocho evoluciones. Para generalizar el nombre dado a los modelos, se usó la **M** de modelo y el número corresponde al número de horas desde la admisión del paciente hasta el momento en que es evaluado por el modelo. Por ejemplo, M168, predice el GRD de pacientes que tienen información desde la admisión y llevan 168 horas hospitalizados (7 días).

Finalmente, con el objetivo de tener un punto de comparación entre un modelo alimentado con la información completa de pacientes y modelos en el tiempo con información parcial, se genera un modelo *Baseline - MBL* que clasifica los pacientes en GRD usando el 100% de la información; es decir, al igual que el *DRG-grouper* clasifica a los pacientes cuando ya han sido dados de alta. De la misma forma se resalta que se utilizaron los resultados del *DRG-grouper* para refinar este modelo MBL. En la figura 5-2 se ilustra la metodología planteada.



**Figura 5-2:** Metodología propuesta para predicción de GRD.

Como se describió anteriormente, a medida que los días de estancia del paciente aumentan, se incrementa también la información a evaluar en el clasificador; por esta razón, por cada modelo que se plantea se genera un matriz codificada (4.4. Codificación

de características) de entrada al clasificador. La matriz de entrada del modelo M0 es la de menor tamaño ya que en este punto los registros de los pacientes no cuentan con información de procedimientos, mientras que la matriz de entrada al modelo M24 aumenta su tamaño en 697 características, lo que indica que durante las primeras 24 horas de hospitalización de los pacientes se realizaron 697 procedimientos diferentes. La entrada del modelo M48 tiene 934 características más que M0, implica que se realizaron 237 procedimientos diferentes adicionales en los segundos días de estancia hospitalaria a los pacientes. La tabla 5-1 relaciona los tamaños de las matrices de entrada a los nueve modelos de predicción y el modelo BL. Es de resaltar que en el CMBD se pudieron identificar 1.592 procedimientos diferentes, los cuales son atributos de alto impacto económico porque son recursos físicos como quirófanos, cuartos estériles, área de diagnósticos, entre otros que son usados para la recuperación apropiada de los pacientes.

No	Modelo	Tamaño de conjunto de datos	Etiqueta	Información demográfica	Diagnósticos			Procedimientos
					CIE-10	CIE-10 3 caracteres	CIE-10 MDC	
1	<b>M0</b>	(57032,2778)	1	9	1692	1057	20	0
2	<b>M24</b>	(57032,3475)	1	9	1692	1057	20	697
3	<b>M48</b>	(57032,3712)	1	9	1692	1057	20	934
4	<b>M72</b>	(57032,3843)	1	9	1692	1057	20	1065
5	<b>M96</b>	(57032,3921)	1	9	1692	1057	20	1143
6	<b>M120</b>	(57032,3990)	1	9	1692	1057	20	1212
7	<b>M144</b>	(57032,4041)	1	9	1692	1057	20	1263
8	<b>M168</b>	(57032,4093)	1	9	1692	1057	20	1315
9	<b>M192</b>	(57032,4135)	1	9	1692	1057	20	1357
10	<b>MBL</b>	(57032,4370)	1	9	1692	1057	20	1592

**Tabla 5-1:** tamaño de entradas por modelos evaluado.

#### 5.4. Métricas empleadas para el análisis.

Para evaluar el desempeño de los modelos de ML, se utilizó la medida estadística *F1-score*. la cual es:

$$F1 = 2 * \textit{precisión} * \textit{recall} / (\textit{precision} + \textit{recall})$$

*F1-score* tiene valores entre 0 y 1 y se utiliza para combinar las medidas de precisión y *recall* en un sólo valor. Esta métrica es útil cuando en los problemas de clasificación es igual de importante la precisión y el *recall*, es práctico para hacer comparaciones simultáneas de ambas medidas. Ambas medidas, precisión y *recall*, son directamente proporcionales a la medida *F1-score*, si alguna de las dos medidas baja o sube, igual lo hará la métrica *F1-score*.

Donde la precisión es:

$$\text{precisión} = \frac{TP}{TP + FP}$$

Y el *recall* se halla:

$$\text{recall} = \frac{TP}{TP + FN}$$

TP: True Positive (Verdadero Positivo)

FN: False Negative (Falso Negativo)

## 5.5. Entrenamiento y prueba.

Para el entrenamiento se cuenta con una matriz de entrada a cada uno de los modelos de predicción propuestos y el modelo BL; es decir, diez matrices de entrada. Adicionalmente, cada una de ellas con tamaños diferentes como se relaciona en la tabla 5-1. Es por esto que cada modelo propuesto (M0, ... M192) debe ser entrenado y validado con su propia matriz de entrada.

Con el objetivo de evitar el sobre ajuste, durante la etapa de entrenamiento de cada modelo se utilizó el método de validación cruzada de *k-fold*, el cual consiste en dividir aleatoriamente el conjunto de muestra en una serie de grupos (*folds*) del mismo tamaño, donde *k* indica el número de particiones o grupos (*folds*), en los que se divide el conjunto

de datos. Luego de tener estas  $k$  particiones ( $k$ -folds) aleatorias, se usan las  $k-1$  particiones para la etapa de entrenamientos y la partición restante para la etapa de prueba. [54] Para esta investigación en particular se usó validación cruzada con cinco folds, es decir, el proceso de entrenamiento se repite cinco veces y utiliza muestras diferentes en cada ciclo del entrenamiento, el sistema mezcla las muestras y las parte en 80% para entrenamiento y 20% para prueba. El promedio de las precisiones halladas en cada ciclo es la precisión general de la etapa de entrenamiento y prueba.

Durante el proceso de separación de las muestras en la etapa de entrenamiento, dicha separación se hace de forma estratificada, es con el objetivo de usar proporciones iguales de clases y evitar así favorecer las clases con más muestras, lo que podría terminar en un sobre ajuste del modelo o no contar con las suficientes muestras de comparación en las clases pequeñas y producir así que los modelos se equivoquen más con las clases minoritarias. La estratificación se hace por porcentaje, seleccionando el 80% de cada clase para el entrenamiento y el restante para la prueba, garantizando así la igualdad en la proporción por número de muestras por cada clase. En la tabla 5-2 se relacionan los GRD y el número de muestras que seleccionaron aproximadamente para las etapas de entrenamiento y prueba la proporción de 80% y 20%.

GRD	Clase	Muestras para entrenamiento	Muestras para prueba	Total de muestras
C-09-07	0	3725	931	4656
C-09-08	1	3045	761	3806
C-13-04	2	2968	741	3709
C-08-04	3	2469	617	3086
C-05-18	4	1874	468	2342
C-10-08	5	1753	438	2191
C-04-04	6	1751	437	2188
C-05-14	7	1662	415	2077
C-10-10	8	1391	347	1738

C-13-05	9	1125	281	1406
C-08-16	10	1068	267	1335
C-10-15	11	1068	266	1334
P-10-12	12	1060	264	1324
C-02-02	13	997	249	1246
P-12-13	14	944	236	1180
C-08-01	15	898	224	1122
C-13-03	16	852	213	1065
C-10-02	17	819	204	1023
P-08-12	18	800	199	999
C-05-16	19	737	184	921
C-10-01	20	719	179	898
C-09-05	21	716	178	894
C-10-12	22	702	175	877
P-11-01	23	694	173	867
C-10-11	24	680	170	850
C-17-02	25	635	158	793
C-12-04	26	630	157	787
C-12-07	27	626	156	782
C-11-02	28	622	155	777
C-12-14	29	609	152	761
C-09-06	30	571	142	713
C-01-04	31	556	139	695
C-01-06	32	522	130	652
C-08-15	33	512	127	639
C-05-15	34	510	127	637
C-10-04	35	504	125	629
P-12-15	36	496	123	619

P-12-11	37	477	119	596
C-09-09	38	456	114	570
P-10-02	39	454	113	567
C-08-12	40	442	110	552
C-17-03	41	438	109	547
C-13-09	42	437	109	546
C-05-06	43	412	103	515
C-01-02	44	412	103	515
C-11-01	45	408	102	510
C-10-14	46	397	99	496

**Tabla 5-2:** Relación GRD estratificados para entrenamiento y prueba.

## 5.6. Técnicas usadas en la investigación experimental.

Existen diferentes técnicas en el Machine Learning para la creación de modelos de clasificación o regresión. Por la naturaleza de este proyecto, es necesario el uso de modelos de clasificación multiclase. En esta investigación se optó por usar dos técnicas de inteligencia artificial, se usaron la Red Neuronal Artificial (RNA) y el *eXtrem Gradient Boosting* conocido como XGBoost.

### 5.6.1. Red Neuronal Artificial – RNA.

Para la implementación de la técnica de ML, las Redes Neuronales Artificiales, se probaron diferentes arquitecturas, redes neuronales artificiales densas *Feed-forward* y redes neuronales convolucionales, pero rápidamente se observó que no se necesitaban arquitecturas muy complejas para obtener resultados aceptables. Se optó por trabajar con arquitecturas de redes neuronales artificiales densas, se probó con varias capas ocultas y estas a su vez con diferentes números de neuronas.

Para definir la arquitectura de la red con la que se trabajaron estos experimentos, se tomó como punto de partida la estructura usada en Gentimis et al. (2017) [55], trabajo que manejó datos similares a los descritos en esta investigación y el objetivo fue la predicción de días de estancia de pacientes hospitalizados, lo cual es uno de los parámetros obtenidos con la predicción de GRD. Con base en lo anterior y como se describe en las siguientes secciones, se determinó que la mejor arquitectura de red para esta implementación es una red que cuenta con dos capas ocultas densas (*full connected*), la primera con 256 neuronas y la segunda con 128 neuronas, ambas capas usan una función de activación *relu*. Finalmente, a la salida de la red, se cuenta con una señal de salida que va desde  $Y_0$  hasta  $Y_{46}$ , lo que corresponde a las 47 clases para clasificación con las que cuenta el sistema y que serán representadas en forma binario, tendrá el valor de uno (1) la salida donde el sistema considera que es la clase a la que pertenece la muestra, el resto de las salidas serán cero (0) y la función de activación utilizada en esta capa de salida es una *softmax*. Cabe destacar también que la función de pérdida utilizada en esta RNA fue la *sparse\_categorical\_crossentropy*.

En la figura 5-3, se puede observar que, a la entrada a la red, la señal de entrada se ilustra de 1 hasta  $p$ , donde para el modelo definido como *Baseline* el valor de  $p$  es 4370, que corresponde al número máximo de características que este sistema va a tener, es decir, al alta del paciente, cuanto se tiene el 100% de la información almacenada durante la estancia hospitalaria del paciente.

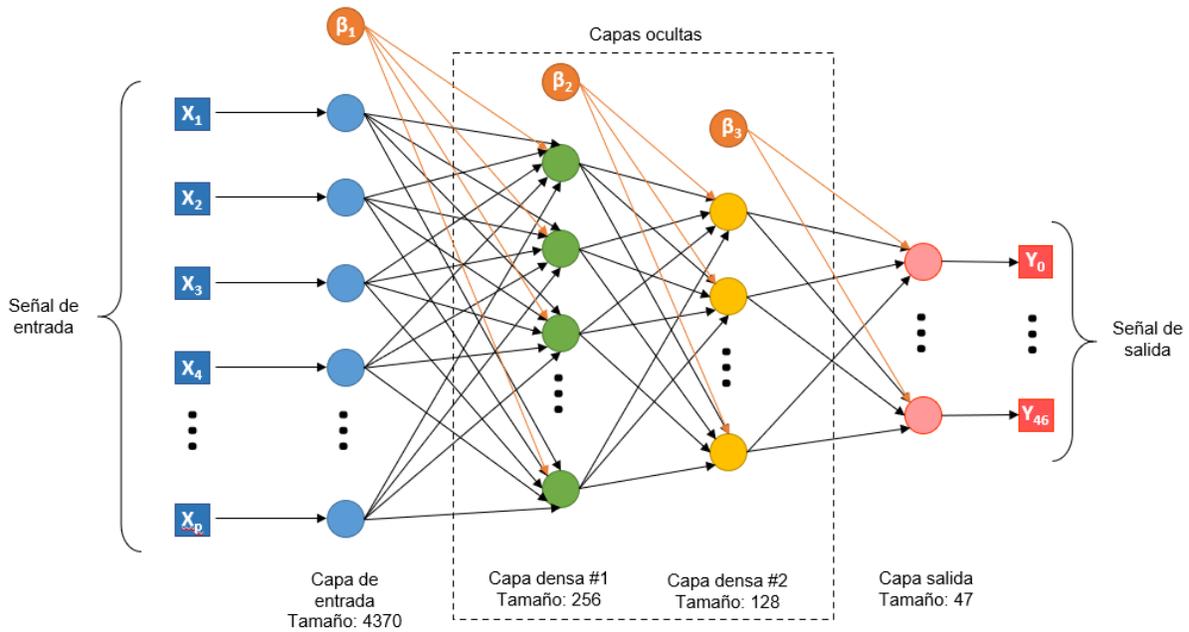


Figura 5-3: Arquitectura de Red Neuronal Artificial usada en el experimento.

### 5.6.1.1. Manejo de desbalance de las muestras usando la RNA.

Uno de los posibles problemas a los que se enfrenta este tipo de proyecto es el desbalance de las clases. En este particular, como se describe en la tabla 4-6, la clase con más muestras tiene 4.656 registros y la clase más pequeña cuenta con 496 registros. Para manejar este problema con la RNA, se utilizó el método `compute_class_weight(class_weight, classes, y)` de la implementación de la RNA en *Scikit-Learn*, donde los parámetros usados fueron `class_weight = 'balanced'` y los pesos de las clases tienen la forma:

$$w = \frac{n_{sample}}{n_{classes} * np.bincoun(y)}$$

El método `bincount()` de NumPy se utiliza para obtener la frecuencia de cada elemento contenidos dentro de una matriz numpy. Para este caso en particular, retorna la frecuencia de las clases del conjunto de datos.

En las tablas 5-3 y 5-4 se muestran los pesos  $w_i$  que este método le asigna a cada GRD (clase) en las etapas de entrenamiento y prueba respectivamente, donde el primer dato corresponde al peso  $w_0$  que se le da a la clase 0 (GRD C-09-07) que es la que más muestras tiene y el último dato corresponde a  $w_{46}$ , peso que está asociado a la clase 46 (GRD C-10-14), la cual es la clase minoritaria. Se puede observar en ambas tablas que los pesos  $w$  asignados por la RNA en durante las dos etapas son prácticamente iguales, por ejemplo, el  $w_0$  es 0,26070541 y 0,26027836 para entrenamiento y prueba respectivamente, de igual forma pasa con los valores del  $w_{46}$ , son 2,44616539 y 2,44766817 en las mismas etapas respectivamente. Lo anterior se debe a que, para la separación de las muestras desde el conjunto de datos original, se usó el parámetro *stratified* con valor *True* el en método *cv* de la API *xgboost* y esto permite que la separación de las muestras se haga por porcentaje aplicado a cada clase, de esta forma se garantiza que las proporciones de las clases, sin importar el desbalance, sean iguales entre sí.

GRD	Peso								
C-09-07	0,26070541	C-08-16	0,31892534	C-10-01	0,32719935	C-12-14	0,39332834	C-09-09	0,51821113
C-09-08	0,55398041	C-10-15	0,55461317	C-09-05	0,58431267	C-09-06	0,69815073	P-10-02	0,86322459
C-13-04	0,90929556	P-10-12	0,90929556	C-10-12	0,91615817	C-01-04	0,97404981	C-08-12	1,02873693
C-08-04	1,13982120	C-02-02	1,08143392	P-11-01	1,18574806	C-01-06	1,54147248	C-17-03	1,21390957
C-05-18	1,31767661	P-12-13	1,35066434	C-10-11	1,38337273	C-08-15	1,35632355	C-13-09	1,42812891
C-10-08	1,39931939	C-08-01	1,52933490	C-17-02	1,55132214	C-05-15	1,56129849	C-05-06	1,59462670
C-04-04	1,70074897	C-13-03	1,74663248	C-12-04	1,86039782	C-10-04	1,90417188	C-01-02	1,89673371
C-05-14	2,03590704	C-10-02	1,92684059	C-12-07	1,95791867	P-12-15	2,12966592	C-11-01	2,13904771
C-10-10	2,19712140	P-08-12	2,22226009	C-11-02	2,21718644	P-12-11	2,35710597	C-10-14	2,35710597
C-13-05	2,38021485	C-05-16	2,44616539						

**Tabla 5-3:** pesos  $w$  asignados a las 47 clases en la etapa de entrenamiento (*train* en la RNA).

GRD	Peso								
C-09-07	0,26027836	C-08-16	0,31842201	C-10-01	0,3270164	C-12-14	0,39273768	C-09-09	0,51777596
C-09-08	0,55324007	C-10-15	0,55450606	C-09-05	0,58390156	C-09-06	0,69832608	P-10-02	0,86234573
C-13-04	0,90756236	P-10-12	0,91097424	C-10-12	0,91787556	C-01-04	0,97316927	C-08-12	1,02677605
C-08-04	1,13764859	C-02-02	1,08178191	P-11-01	1,18783897	C-01-06	1,54343407	C-17-03	1,21768417
C-05-18	1,31695190	P-12-13	1,35373826	C-10-11	1,38468085	C-08-15	1,36134353	C-13-09	1,42540676
C-10-08	1,40068872	C-08-01	1,53366550	C-17-02	1,55332788	C-05-15	1,56334935	C-05-06	1,59420493
C-04-04	1,70647288	C-13-03	1,74330323	C-12-04	1,86399345	C-10-04	1,90802479	C-01-02	1,90802479
C-05-14	2,03629537	C-10-02	1,93855319	C-12-07	1,97007438	P-12-15	2,12560657	C-11-01	2,14441725
C-10-10	2,20290135	P-08-12	2,22311146	C-11-02	2,22311146	P-12-11	2,35261310	C-10-14	2,35261310
C-13-05	2,37567793	C-05-16	2,44766817						

**Tabla 5-4:** pesos  $w$  asignados a las 47 clases en la etapa de prueba (*test* en la RNA).

### 5.6.1.2. Ajuste de parámetros e hiperparámetros de la RNA.

Una vez se define la arquitectura, se hallan los parámetros e hiperparámetros de la red neuronal evaluando una malla de parámetros, que es descrita en la tabla 5-5. Es de aclarar que para determinar estos ajustes, se hizo con el modelo *Baseline* – MBL, con el 100% de la información de la base de datos de entrada y teniendo como objetivo que se debía alcanzar la mejor clasificación posible en comparación con el *DRG-grouper*.

No.	Parámetro	Valores de la malla
1	Tasa de aprendizaje	[ 0.001 , 0.005, 0.01 , 0.05 , 0.1 , 0.5 , 1 ]
2	Épocas	[ 20, 50, 100, 500, 1000 ]
3	Tamaño del <i>batch</i>	[ 256, 512, 1024, 2048 ]
4	Optimizador	[ adam , SGD , RMSprop ]
5	<i>Dropout</i>	[ 0, 0.1, 0.2, 0.5, 0.8 ]

**Tabla 5-5:** Malla de evaluación de parámetros e hiperparámetros de la RNA.

Utilizando el método *GridSearchCV* de *Scikit-Learn*, se hace el proceso de ajuste de los parámetros e hiperparámetros de la red neuronal artificial determinada, los valores que mejor resultado de precisión arrojaron fueron los que se relacionan en la tabla 5-6. El *Dropout* de 0.1 determinado en este ajuste, se hace entre las dos capas densas ocultas.

No.	Parámetro	Valores
1	Tasa de aprendizaje	0.005
2	Épocas	20
3	Tamaño del <i>batch</i>	1024
4	Optimizador	adam
5	<i>Dropout</i>	0.1
6	Semilla	[ 7, 14, 21, 1, 31 ]

**Tabla 5-6:** Parámetros e hiperparámetros determinados para la RNA.

Luego de completar la optimización y seleccionar los parámetros e hiperparámetros, se hicieron los experimentos usando semillas diferentes, con el objetivo de generar muestras aleatorias diferentes para los entrenamientos y probar el desempeño del modelo en condiciones diferentes. Los valores usados para la semilla se relacionan en el ítem 5 de la tabla 5-6.

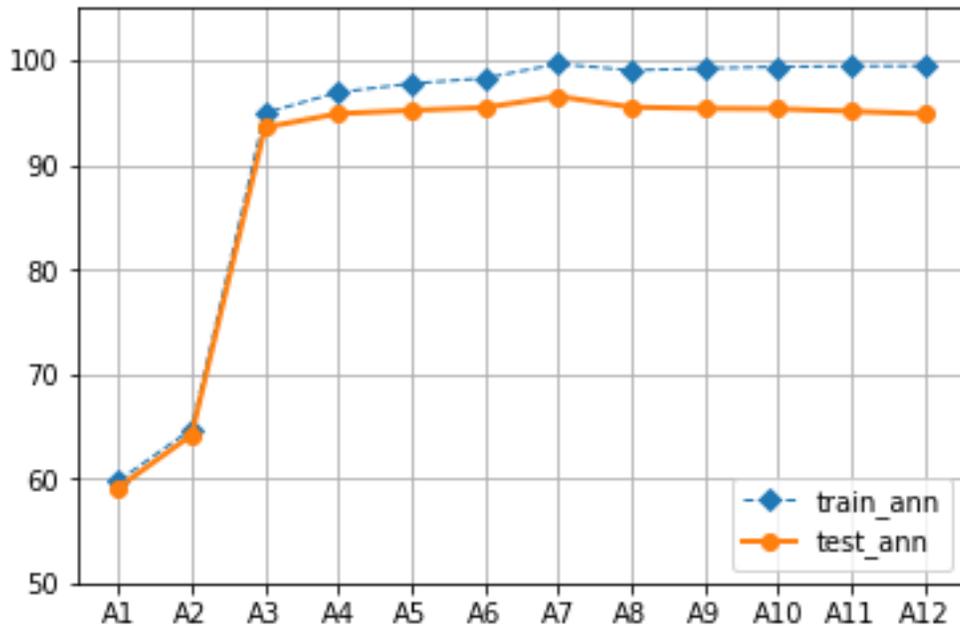
### 5.6.1.3. Análisis de sensibilidad del modelo de la RNA.

El objetivo de un análisis de sensibilidad es determinar qué tanto puede variar el desempeño del modelo haciendo cambios en los hiperparámetros. Después de tener el mejor modelo MBL entrenado y para determinar este comportamiento, se hacen variaciones a la estructura de la RNA como se muestra en la tabla 5-7.

Arquitectura	Capas de la red	Neuronas por capa
A1	2	5 , 3
A2	1	4
A3	1	8
A4	1	16
A5	1	32
A6	1	64
<b>A7</b>	<b>2</b>	<b>256 , 128</b>
A8	1	512
A9	2	1024 , 512
A10	5	1024 , 512 , 256 , 128 , 64
A11	5	1024 , 1024 , 1024 , 1024 , 1024
A12	5	2048 , 2048 , 2048 , 2048 , 2048

**Tabla 5-7:** Arquitecturas para medición de sensibilidad de la RNA.

En la gráfica 5-1 se ilustra el comportamiento de la red cuando se varían los hiperparámetros que se relacionan en la tabla 5-7. Se observa que con las dos primeras arquitecturas la precisión es muy baja, pero con aumentar unas cuantas neuronas al modelo la precisión mejora rápidamente, así como se ve en el paso de la A2 a la A3; solo con agregar cuatro neuronas el modelo mejora considerablemente, pasa de un 65% de precisión a un 94%. De ahí en adelante, con los cambios en las arquitecturas los cambios de precisión no son muy sustanciales. Incluso en las tres últimas arquitecturas (A10, A11 y A12) se observa que empieza a haber una tendencia al sobreajuste. Finalmente se determina que la arquitectura con el mejor desempeño es la A7, con valores de precisión en entrenamiento de 99,00% +/- 0,02 y prueba de 96,56% +/- 0,04.



**Gráfica 5-1:** Análisis de sensibilidad del modelo MBL usando la RNA.

#### 5.6.1.4. Ejecución de experimentos RNA.

El entrenamiento se ejecutó en un clúster de 12 núcleos y 354 Gb de RAM con sistema operativo Centos 7. Aquí se ejecutaron las redes neuronales planteadas para este modelo de ML. En la tabla 5-8 se relacionan los tiempos que tomaron en entrenarse los modelos de RNAs planteados; se observa una leve tendencia al aumento desde el modelo M0 con un tiempo de 5,483 minutos hasta el modelo MBL que tomó 7,350 minutos en entrenarse. Esto se debe a que la matriz de entrada del modelo M0 tiene un tamaño de (57.032 , 2.777), mientras que la dimensión del modelo MBL es (57.032 , 4.369). La tendencia no es completamente lineal porque este clúster se compartió con otros proyectos de investigación, es decir, el recurso no era dedicado.

No.	Modelo	Tiempo de ejecución (minutos)	Tamaño matriz de entrada
1	M0	5,483	(57.032 , 2.777)
2	M24	6,067	(57.032 , 3.474)
3	M48	6,467	(57.032 , 3.712)
4	M72	6,917	(57.032 , 3.842)
5	M96	7,067	(57.032 , 3.920)
6	M120	6,733	(57.032 , 3.989)
7	M144	7,300	(57.032 , 4.040)
8	M168	6,967	(57.032 , 4.092)
9	M192	7,133	(57.032 , 4.134)
10	MBL	7,350	(57.032 , 4.369)

**Tabla 5-8:** Tiempo entrenamiento por modelo – RNA.

### 5.6.2. Extreme Gradient Boosting – XGBoost.

El XGBoost (eXtreme Gradient Boosting) es un algoritmo de Machine Learning de ensamblaje basado en árboles de decisión. Utiliza todas las características del Gradient Boosting, incluida el algoritmo de Gradiente Descendente para minimizar los errores en modelos secuenciales, con la ventaja que el XGBoost optimiza el Gradient Boosting a través del procesamiento paralelo, poda de árboles, manejo de valores faltantes y regularización para evitar sobreajustes y/o sesgos. Este es un algoritmo de ML que se ha popularizado por los altos desempeños que ha obtenido en concursos como *Kaggle* y *KDD Cup 2015*, donde dio la pelea a algoritmos como redes neuronales artificiales.[56]

Para la implementación de la técnica de ML, se usó la librería XGBoost y la implementación que esta tiene en *Scikit-Learn*. Se recurrió al método *XGBClassifier* y se manejó el desbalance de las clases asignando los pesos  $w$  de forma inversamente proporcional al porcentaje de representación de cada clase problema. Se usó el

parámetro  $n\_job$  para mejorar los tiempos de procesamiento del modelo en la etapa de entrenamiento, ya que este parámetro permite seleccionar el número de núcleos (cores) del clúster que se necesitan emplear para el procesamiento.

### 5.6.2.1. Manejo de desbalance de las muestras usando el XGBoost.

Al igual que otras técnicas de ML, el desbalance de las clases es uno de los problemas que hay que manejar para poder obtener resultados aceptables. Para este caso en particular, de la información relacionada en la tabla 4-6, la clase minoritaria (496 registros) es el 10,65% de la clase mayoritaria (4.656 registros). Para manejar este problema, se trabajó con el método *fit* también de la API XGBoost implementada en *Scikit-Learn*, se utilizó el parámetro *sample\_weight( array\_like )*, donde el *array\_like* es un arreglo de *Numpy* con los pesos asignados a cada una de las muestras de entrada a los modelos. Para el cálculo de dichos pesos se usa una función que calcula el porcentaje que tienen cada clase en el total del conjunto de datos, así;

$$\%class = \frac{100 * n\_samples\_per\_class}{n\_samples}$$

Estos valores son normalizados con respecto a la clase que menos muestras tiene, de tal forma que a la clase que tiene menos muestras se le asigna un peso  $w$  de 1 y a la clase que más muestras tiene se le asigna un peso  $w$  según la relación correspondientes, es decir;

$$w0 = \frac{\%_{min\_class}}{\%_{max\_class}}$$

De esta forma se garantiza que se les dará un mayor peso a las muestras de la clase minoritaria y por el contrario a la clase mayoritaria se penalizará con el peso más bajo obtenido con esta función.

En las tablas 5-9 y 5-10 se muestran los pesos  $w_i$  que este método le asigna a las clases en las etapas de entrenamiento y prueba respectivamente, donde el primer dato corresponde al peso  $w_0$  que se le da a la clase 0 que es la que más muestras tiene y el último dato corresponde al  $w_{46}$ , peso que corresponde a la clase 46, la cual es la clase con menos registros. Se observa en ambas tablas que los pesos asignados en las dos etapas son prácticamente iguales, por ejemplo, el  $w_0$  es 0,106577 y 0,106337 para entrenamiento y prueba respectivamente y para la clase 46, el peso  $w_{46}$  es 1,000000 en ambas etapas del proceso. Al igual que el análisis hecho en la sección de RNA, lo anterior se debe a que, para la separación de las muestras desde el conjunto de datos original, se usó el parámetro *stratified* con valor *True* el en método *cv* de la API *xgboost* y esto permite que la separación de las muestras se haga por porcentaje aplicado a cada clase, de esta forma se garantiza que las proporciones de las clases, sin importar el desbalance, sean iguales entre sí.

GRD	Peso								
C-09-07	0,106577	C-08-16	0,130378	C-10-01	0,133805	C-12-14	0,160794	C-09-09	0,211959
C-09-08	0,226469	C-10-15	0,226857	C-09-05	0,239013	C-09-06	0,285612	P-10-02	0,352889
C-13-04	0,371723	P-10-12	0,372071	C-10-12	0,374882	C-01-04	0,398195	C-08-12	0,420551
C-08-04	0,442586	C-02-02	0,465962	P-11-01	0,485330	C-01-06	0,496871	C-17-03	0,538670
C-05-18	0,552925	P-12-13	0,555245	C-10-11	0,566334	C-08-15	0,572046	C-13-09	0,583824
C-10-08	0,626183	C-08-01	0,630159	C-17-02	0,634185	C-05-15	0,638264	C-05-06	0,651888
C-04-04	0,696491	C-13-03	0,714029	C-12-04	0,760536	C-10-04	0,776908	C-01-02	0,778431
C-05-14	0,789264	C-10-02	0,802002	C-12-07	0,832285	P-12-15	0,870614	C-11-01	0,874449
C-10-10	0,898109	P-08-12	0,906393	C-11-02	0,908467	P-12-11	0,963592	C-10-14	0,963592
C-13-05	0,973039	C-05-16	1,000000						

**Tabla 5-9:** pesos  $w$  asignados a las 47 clases en la etapa de entrenamiento (*train* en el XGBoost)

GRD	Peso								
C-09-07	0,106337	C-08-16	0,130092	C-10-01	0,133423	C-12-14	0,160454	C-09-09	0,211087
C-09-08	0,226027	C-10-15	0,226027	C-09-05	0,237981	C-09-06	0,284483	P-10-02	0,352313
C-13-04	0,370787	P-10-12	0,370787	C-10-12	0,373585	C-01-04	0,397590	C-08-12	0,419492
C-08-04	0,440000	C-02-02	0,464789	P-11-01	0,482927	C-01-06	0,495000	C-17-03	0,538043
C-05-18	0,550000	P-12-13	0,553073	C-10-11	0,562500	C-08-15	0,572254	C-13-09	0,582353
C-10-08	0,622642	C-08-01	0,630573	C-17-02	0,634615	C-05-15	0,638710	C-05-06	0,651316
C-04-04	0,692308	C-13-03	0,71223	C-12-04	0,761538	C-10-04	0,773438	C-01-02	0,779528
C-05-14	0,785714	C-10-02	0,798387	C-12-07	0,831933	P-12-15	0,868421	C-11-01	0,876106
C-10-10	0,900000	P-08-12	0,908257	C-11-02	0,908257	P-12-11	0,961165	C-10-14	0,961165
C-13-05	0,970588	C-05-16	1,000000						

**Tabla 5-10:** pesos  $w$  asignados a las 47 clases en la etapa de prueba (*test* en el XGBoost).

Si bien los métodos de ponderación de los pesos  $w$  en ambas técnicas, RNA y XGBoost, son diferentes, hay que destacar que la proporción de estos se conserva, es decir, la proporción entre el peso dado a la clase 0 y el peso dado a la clase 46 son 9.38287160976061 y 9.38288748979611 para la RNA y el XGBoost respectivamente. En otras palabras, para ambos casos, el peso  $w_{46}$  que correspondiente a la clase 46 (clase minoritaria), es aproximadamente 9.3828 veces más grande que el peso  $w_0$  de la clase 0 (clase mayoritaria). Como parte de estas experimentaciones, se corrieron la RNA con los pesos del XGBoost y viceversa, y los resultados fueron consistentes. Para el caso de la RNA con los pesos del XGBoost, los resultados comparativos de *accuracy* se relacionan en la tabla 5-11.

	Media de accuracy en entrenamiento	Media de accuracy en prueba
<b>RNA con pesos de RNA</b>	99.68% (+/- 0.04%)	95.09% (+/- 0.75%)
<b>RNA con pesos de XGBoost</b>	99.68% (+/- 0.02%)	95.04% (+/- 0.64%)

**Tabla 5-11:** Comparativo de accuracy la RNA en entrenamiento y prueba con pesos de la RNA y el XGBoost.

	Media de accuracy en entrenamiento	Media de accuracy en prueba
XGBoost con pesos de XGBoost	97.07%	93.53%
XGBoost con pesos de RNA	98.65%	94.33%

**Tabla 5-12:** Comparativo de accuracy del XGBoost en entrenamiento y prueba con pesos del XGBoost y la RNA.

Como se observa en las tablas 5-11 y 5-12, no hay diferencias significativas entre las medidas de precisión entre las dos técnicas cuando se corren con los pesos propios y los pesos de la técnica contraria. Esto implica que los resultados obtenidos con ambas ponderaciones son completamente consistentes y comparativos.

### 5.6.2.2. Ajuste de parámetros e hiperparámetros del XGBoost.

Para la evaluación del desempeño de los diferentes modelos en el XGBoost se usó una validación cruzada con cinco *folds*; y para determinar los parámetros e hiperparámetros que se debían usar en la configuración final de este método de ensamble de árboles se usa la malla que se muestra en la tabla 5-13.

No.	Parámetro	Valores
1	<i>learning_rate</i>	[ 0.1 , 0.2 , 0.2 , 0.8 ]
2	<i>n_estimators</i>	[ 100 , 200 , 300 , 1000 ]
3	<i>max_depth</i>	[4,5,6]
4	<i>min_child_weight</i>	[ 1 , 3 , 5 ]
5	<i>Reg. gamma</i>	[ 0.0 , 0.1 , 0.2 , 0.3 , 0.4 ]
6	<i>subsample</i>	[ 0.6 , 0.7 , 0.75 , 0.8 , 0.85 , 0.9]
7	<i>colsample_bytree</i>	[ 0.6 , 0.7 , 0.75 , 0.8 , 0.85 , 0.9 ]
8	<i>reg_alpha</i>	[ 0 , 1e-5 , 0.001 , 0.005 , 0.01 , 0.05 , 0.1 , 1 , 100 ]
9	<i>booster</i>	gbtree , dart
10	<i>stratified</i>	True

**Tabla 5-13:** Malla de evaluación de parámetros e hiperparámetros del XGBoost.

Luego de hacer el proceso de ajuste de los parámetros e hiperparámetros del XGBoost, el mejor desempeño se obtuvo con los valores que se relacionan en la tabla 5-14.

No.	Parámetro	Valores
1	<i>learning_rate</i>	0.1
2	<i>n_estimators</i>	300
3	<i>max_depth</i>	5
4	<i>min_child_weight</i>	1
5	<i>gamma</i>	0.2
6	<i>subsample</i>	0.8
7	<i>colsample_bytree</i>	0.8
8	<i>reg_alpha</i>	0.1
9	<i>booster</i>	gbtree
10	<i>stratified</i>	True

**Tabla 5-14.** Parámetros e hiperparámetros determinados para el XGBoost.

Posterior al ajuste de los parámetros y durante las etapas de entrenamiento y prueba se usaron diferentes semillas para la selección de las muestras e introducir de esta manera un nivel de aleatorización. Estas diferentes semillas permiten verificar el desempeño de los modelos en condiciones de muestras diferentes; para esto se usó el parámetro *seed* del método *XGBClassifier* de la API *xgboost* implementada en *Scikit-Learn* con los valores [27, 14, 100, 70, 35] y el parámetro *random\_state* = [42, 21, 133, 100, 35] perteneciente al método *train\_test\_split* implementado en *Scikit-Learn*.

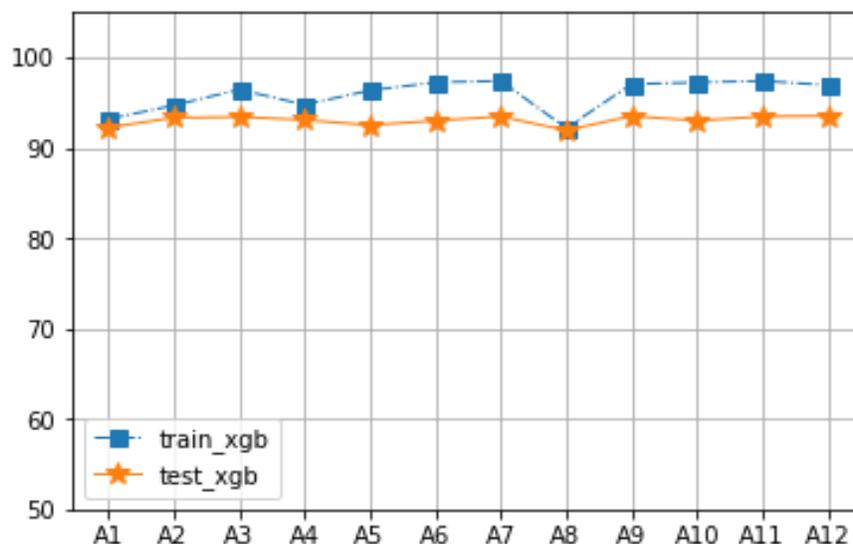
### 5.6.2.3. Análisis de sensibilidad del modelo del XGBoost.

En esta sección se evalúa qué tanto puede variar el desempeño del modelo en el XGBoost haciendo cambios en los hiperparámetros. Luego de tener el mejor modelo MBL entrenado, para determinar este comportamiento, se hacen variaciones a la estructura del XGBoost como se muestra en la tabla 5-15.

Arquitectura	Estimadores	Profundidad máxima
A1	100	3
A2	100	5
A3	100	10
A4	200	3
A5	200	5
A6	200	10
A7	200	15
A8	300	3
<b>A9</b>	<b>300</b>	<b>5</b>
A10	300	10
A11	300	15
A12	1000	5

**Tabla 5-15:** Arquitecturas para medición de sensibilidad del XGBoost.

En la gráfica 5-2 se ilustra el comportamiento de este algoritmo de ensamblaje cuando se varían los hiperparámetros. A diferencia de la sensibilidad observada en la implementación de la RNA para este MBL, la precisión con el XGBoost es menos variable entre las arquitecturas planteadas en la tabla 5-13, lo que implica una mayor estabilidad y una menor sensibilidad al cambio de hiperparámetros en el XGBoost, las arquitecturas A1 y A2 son las más sencilla y las que tiene la precisión más baja, 92% y 93% respectivamente para la fase de prueba, pero a pesar de esto, son valores aceptables en comparación con el comportamiento de las dos primeras arquitectas (A1 y A2) de la RNA. De ahí en adelante, el cambio en la precisión de las arquitecturas no es muy sustancial, la mejor precisión para la fase de prueba es de 93,55%. Finalmente, se determina que la arquitectura con el mejor desempeño es la A9, con valores de precisión en entrenamiento de 97,04% +/- 0,01 y prueba de 93,55% +/- 0,02.



Gráfica 5-2: Análisis de sensibilidad del modelo MBL usando el XGBoost.

#### 5.6.2.4. Ejecución de experimentos con el XGBoost.

Los experimentos realizados con el método XGBoost, al igual que la evaluación de la RNA descrita en la sección 5.3.1.4, se ejecutaron en un clúster de 12 núcleos y 354 Gb de RAM con sistema operativo Centos 7. Aquí se ejecutaron los modelos de ensamble de árboles planteados para este modelo de ML. En los tiempos que tomaron en entrenarse los modelos del XGBoost, se observa un aumento significativo en la etapa de entrenamiento desde el modelo M0 hasta el modelo MBL. Esto se debe a que la matriz de entrada del modelo M0 tienen un tamaño de (57.032 , 2.777), mientras que el modelo MBL tiene una dimensión de (57.032 , 4.369). La tendencia no es completamente lineal porque este clúster no era un recurso dedicado. En la tabla 5-16 se relacionan los promedios de los tiempos en minutos. Por ejemplo, el modelo M0 tardó 374,95 minutos en entrenarse, lo que equivale a 6,249167 horas. Y para el modelo MBL, el tiempo en horas fue de 14,443056.

No.	Modelo	Tiempo de ejecución (minutos)	Tamaño matriz de entrada
1	M0	374,950000	(57.032 , 2.777)
2	M24	404,416667	(57.032 , 3.474)
3	M48	480,883333	(57.032 , 3.712)
4	M72	549,250000	(57.032 , 3.842)
5	M96	555,866667	(57.032 , 3.920)
6	M120	562,633333	(57.032 , 3.989)
7	M144	528,616667	(57.032 , 4.040)
8	M168	545,650000	(57.032 , 4.092)
9	M192	545,316667	(57.032 , 4.134)
10	MBL	866,583333	(57.032 , 4.369)

**Tabla 5-16:** Tiempo entrenamiento por modelo - XGBoost.

## 5.7. Evaluación de los modelos propuestos en la metodología de predicción.

En esta sección se hace un análisis de los desempeños en las dos técnicas de ML empleadas, de los diferentes modelos propuestos en la metodología de clasificación temprana de pacientes en GRD. Primero se abordará el análisis de los resultados del modelo M0, el cual evalúa la predicción del GRD del paciente en el momento de la admisión hospitalaria, luego se hará un análisis conjunto del desempeño de los modelos de predicción siguientes, desde M24 hasta M192.

### 5.7.1. Evaluación de modelo M0.

Como se ha indicado, en términos de capacidad de cómputos, el modelo M0 es el más sencillo de los nueve propuesto, ya que su matriz de entrada es la más pequeña, su tamaño es de (egresos , características) = (57.032 , 2.777), representa la admisión del

paciente y donde el número de procedimientos realizados es cero (0). Los resultados obtenidos de precisión se relacionan en la tabla 5-17, donde se observa un mejor resultado con la RNA. En términos numéricos, de cada 100 pacientes que son admitidos en hospitalización, 93 son clasificados en los GRD que le corresponden. Esto es un logro importante porque a diferencia del *DRG-grouped*, estos 93 pacientes fueron clasificados utilizando únicamente el diagnóstico principal. En esta misma tabla 5-17 se pueden observar las desviaciones estándar de las métricas, las cuales se obtuvieron con las mediciones de *accuracy* en cada uno de los cinco *folds* ejecutados, se puede inferir que durante todo el proceso las mediciones son muy consistentes ya que las desviaciones son muy pequeñas, 0.02% y 0.38% en la etapa de prueba para la RNA y el XGBoost respectivamente.

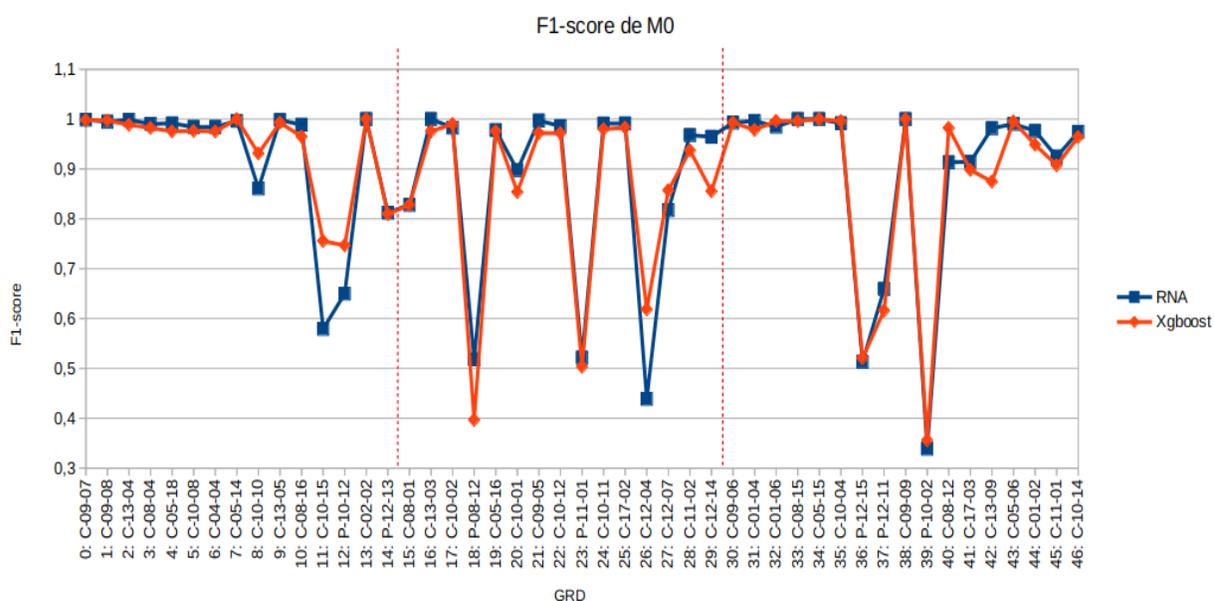
	Precisión	
Técnica	Entrenamiento	Prueba
RNA	0,9415 ± 0,0003	0,9306 ± 0,0002
XGBoost	0,9228 ± 0,0015	0,8911 ± 0,0038

**Tabla 5-17:** Precisión de la RNA y el XGBoost en el modelo M0.

En la gráfica 5-3 se ilustran la métrica *F1-score* para las dos técnicas (RNA y XGBoost) en los resultados de la fase de prueba del modelo M0 por cada uno de los GRD. Se observa una tendencia similar en ambas técnicas, se equivocaron o acertaron en las mismas proporciones, con algunas diferencias significativas en los GRD C-10-15, P-10-12, P-08-12 y C-12-04, de lo cual nos ocuparemos más adelante.

Siguiendo con el análisis de la gráfica, si dividiéramos la gráfica en tres segmentos (líneas punteadas rojas en la gráfica); 1) los GRD con más clases, 2) los GRD de frecuencia intermedia y 3) los GRD con las clases minoritarias. En promedio, los primeros 15 GRD con mayor frecuencia tienen una media del *F1-score* de 0,9219 y 0,9396 para la RNA y el XGBoost respectivamente; para el segundo grupo que serían del 16 al 30, el promedio de la métrica *F1-score* es el más bajo de los tres grupos, con 0,8587 para la

RNA y 0,8465 para el XGBoost, en este grupo de GRD es donde se observan mediciones de *F1-score* con más tendencia a la baja, es decir, es la sección de GRD donde más se equivoca el modelo M0. Para el caso del tercer grupo de GRD ilustrados en la gráfica, se observa la medición más baja en ambas técnicas en el GRD P-10-02, pero se compensa con otras mediciones que son superiores a 0,90, para este grupo los promedios de las medidas de *F1-score* son 0,8913 para la RNA y 0,8836 para el XGBoost.

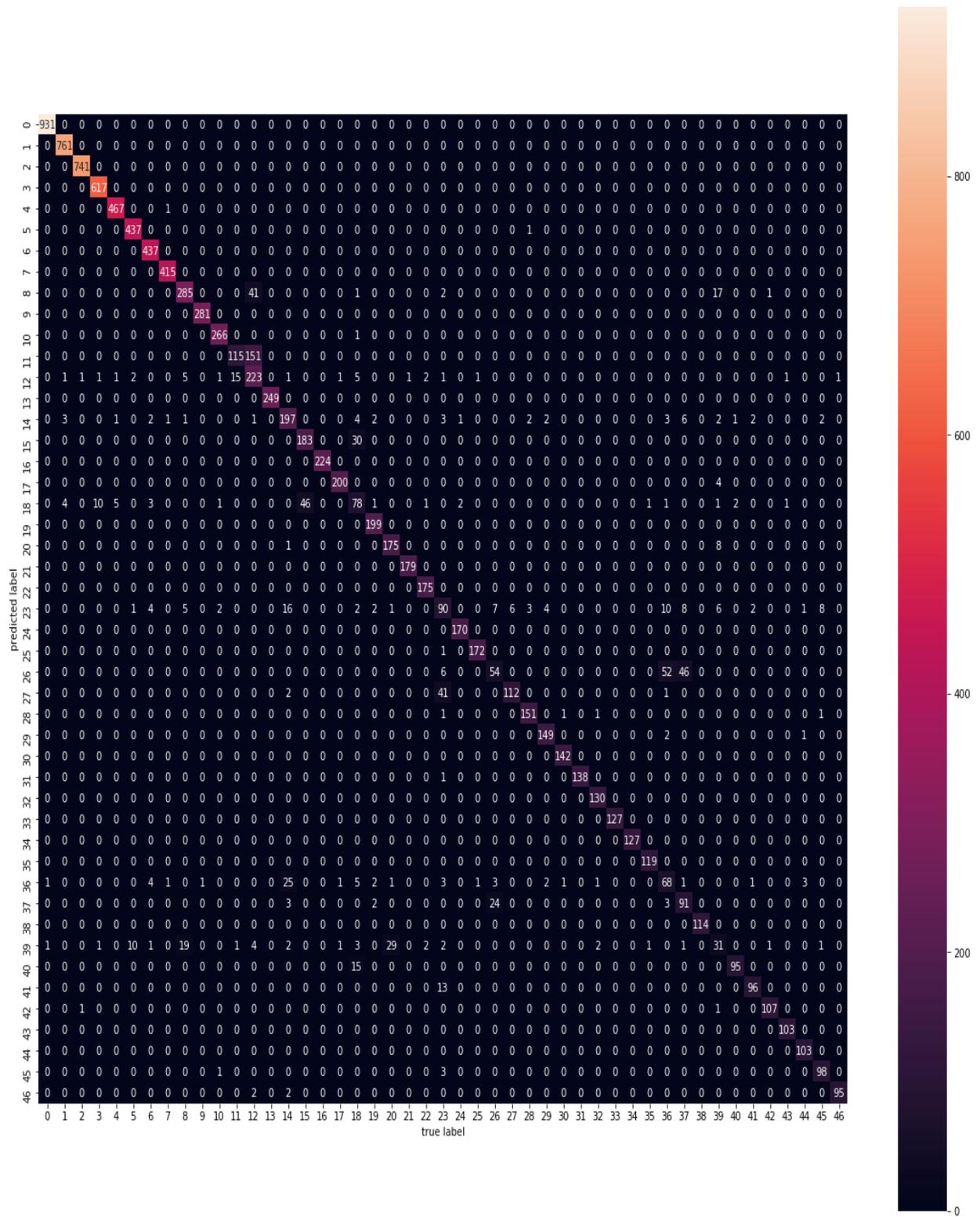


**Gráfica 5-3:** Métrica *F1-score* de la RNA y el XGBoost para todos los GRD.

De este gráfico podemos inferir que si bien el *F1-score* más bajo está en uno de los GRD de las clases minoritarias, no quiere decir que por tener pocas muestras es que clasifica mal, ya que hay varios GRD con menos muestras y con *F1-score* entre 0,90 y 0,99. Bajo este mismo análisis, pero pasando al grupo de GRD mayoritarios, existen dos GRD con medidas *F1-score* por debajo de 0,65 en la RNA y por debajo de 0,75 en el XGBoost.

La gráfica 5-4 es la matriz de confusión resultante de la evaluación del modelo M0 implementado en la RNA. Se observa que la clase 11 (266 registros) y 12 (264 registros) correspondiente a los GRD *C-10-15: Otros trastornos de vesícula y vía biliar* y *P-10-12:*

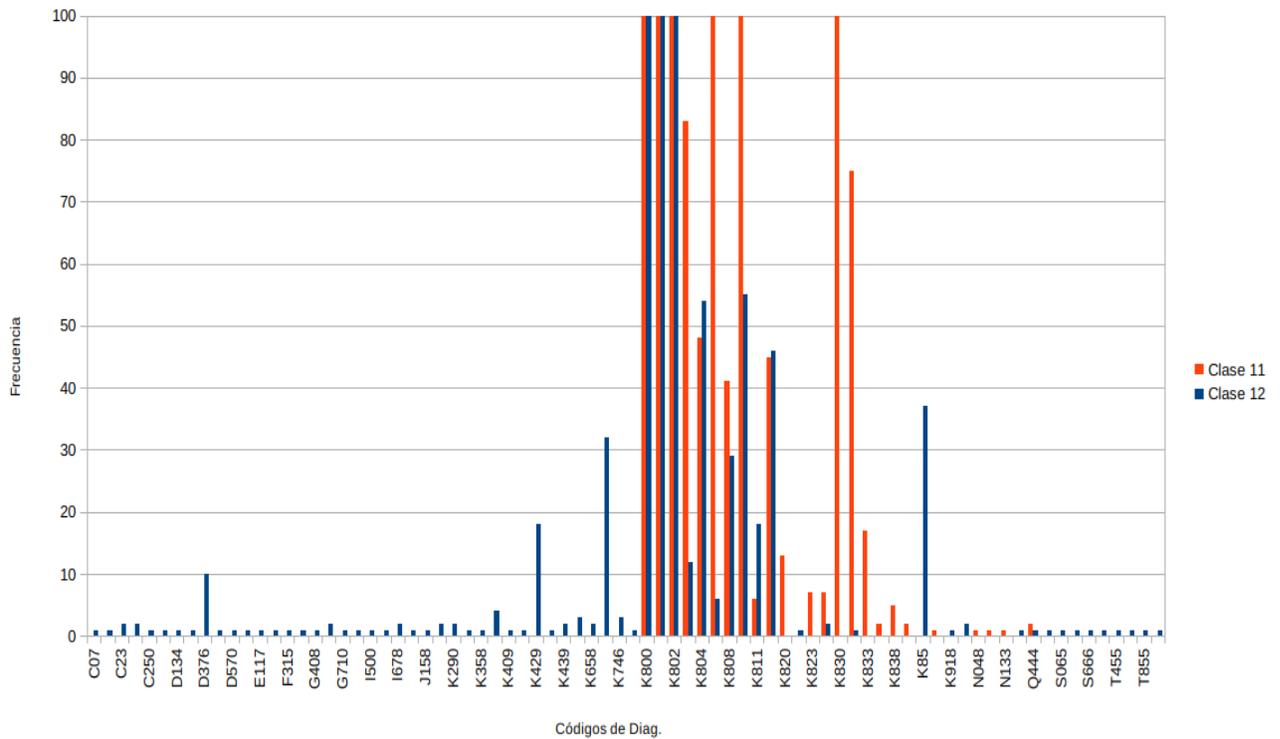
*Colecistectomía laparoscópica respectivamente*, presentan una gran cantidad de errores en la clasificación. Por ejemplo, se muestra que 151 casos de 264 de la clase 12 fueron clasificados en la clase 11, esto se atribuye a que el GRD quirúrgico *P-10-12* es la operación que se realiza para extirpar la vesícula biliar debido a la presencia de cálculos biliares que causan dolor o una infección, usando instrumental de laparoscopia y los diagnósticos principales asociados a estos dos GRD, en su mayoría son los mismo, como se ve en la gráfica 5-5 que, aunque el espectro de diagnósticos asociados a las clase 12 es grande, hay un gran número de estos que se comparten con todos los diagnósticos asociados para la clase 11, en otras palabras, el 87,5% de los diagnósticos asociados a la clase 12 (*P-10-12*) son los mismos que representan el 82,4% de los diagnósticos asociados a la clase 11 (*C-10-15*). Los diagnósticos compartidos se listan en la tabla 5-18.



Gráfica 5-4: Matriz de confusión del M0 con la RNA.

No.	Código CIE-10	Descripción CIE-10
1	K800	Cálculo de la vesícula biliar con colecistitis aguda
2	K801	Cálculo de la vesícula biliar con otra colecistitis
3	K802	Cálculo de la vesícula biliar sin colecistitis
4	K803	Cálculo de conducto biliar con colangitis
5	K804	Cálculo de conducto biliar con colecistitis
6	K805	Cálculo de conducto biliar sin colangitis ni colecistitis
7	K808	Otras colelitiasis
8	K810	Colecistitis aguda
9	K811	Colecistitis crónica
10	K818	Otras colecistitis

**Tabla 5-18:** Códigos de diagnóstico comunes para los GRD C-10-15 (clase 11) y P-10-12 (clase 12)



**Gráfica 5-5:** Frecuencia de Diagnósticos asociados a los GRD C-10-15 y P-10-12.

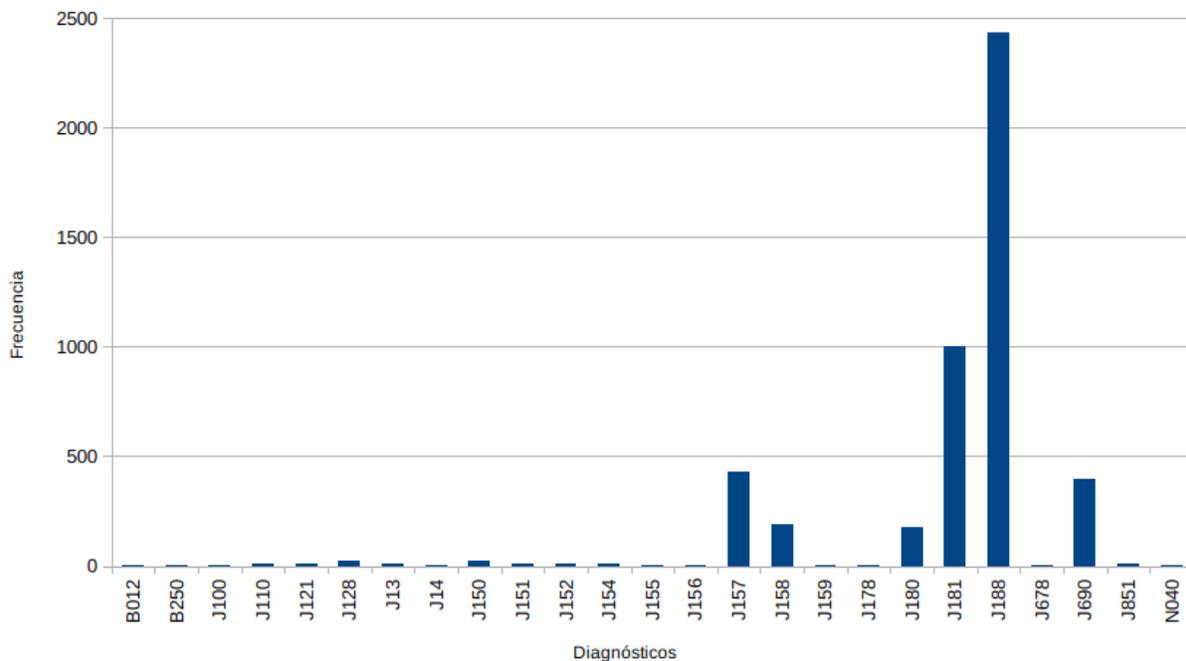
Siguiendo con el análisis de los GRD mal clasificado, pero esta vez en el grupo de GRD con cantidad de clases medio, resalta la medida *F1-score* para el GRD quirúrgico *P-08-12: Cateterismo cardiaco y coronariografía con o sin angioplastia sin inserción de stent coronario*, corresponde a la clase 18 y tiene 199 registros en la etapa de prueba del modelo. Tiene un *F1-score* de 0,5183 para la implementación del modelo M0 en la RNA y de 0,3967 para el XGBoost, el resultado de esta métrica se atribuye a la gran variedad de diagnósticos principales asociados a este GRD, en total tiene 157 diagnósticos diferentes en los 999 registros que tiene esta clase en todo el conjunto de datos, los diagnósticos para este GRD son tan diversos que para este caso en particular provienen de 17 de las 20 MDC existentes en los datos relacionadas en la tabla 4-1.

Finalmente, en el análisis de los GRD mal clasificados por el modelo M0, existe en el grupo de los GRD con clases minoritarias, el GRD quirúrgico *P-10-02: Procedimientos de mayor complejidad sobre tracto gastrointestinal*, clase 39 y 113 registros para la etapa de prueba del modelo con un *F1-score* de 0,3388 y 0,3558 para la RNA y el XGBoost respectivamente. Al igual que el caso anterior, se deben a la gran variedad de diagnósticos principales asociados a este GRD, existen 170 diagnósticos diferentes en los 567 registros de este GRD y provienen de 16 de las 20 MDC existentes en los datos relacionadas en la tabla 4-1.

Haciendo el mismo análisis, pero ahora con las clases que quedaron correctamente clasificadas en un 100%, es el caso del GRD *C-09-07: Neumonía y neumonitis*, que como se ilustra en gráfica 5-4 (matriz de confusión) muestra que todos los 931 registros de la clase 0 quedan correctamente clasificados, esta es la clase con más registros (4.656) y obtuvo un *F1-score* de 0,9989 en la RNA y 0,9979 en el XGBoost, estos resultados se deben a que los registros cuentan con 28 diagnósticos diferentes en los 4.656 registros de la clase 0 (C-09-07) y que son provenientes de tres MDC; de estos tres, uno de ellos (*Capítulo X: Enfermedades del sistema respiratorio*. Ver tabla 4-1) representa el 99% de los registros para esta clase. La gráfica 5-6 ilustra que la frecuencia de los diagnósticos asignados a los pacientes clasificados en este GRD, aproximadamente el 76% de los diagnósticos, están representados en tres códigos; el *J180: Bronconeumonía, no*

especificada, el J181: Neumonía lobar, no especificada y el J188: Otras neumonías, de microorganismo no especificado.

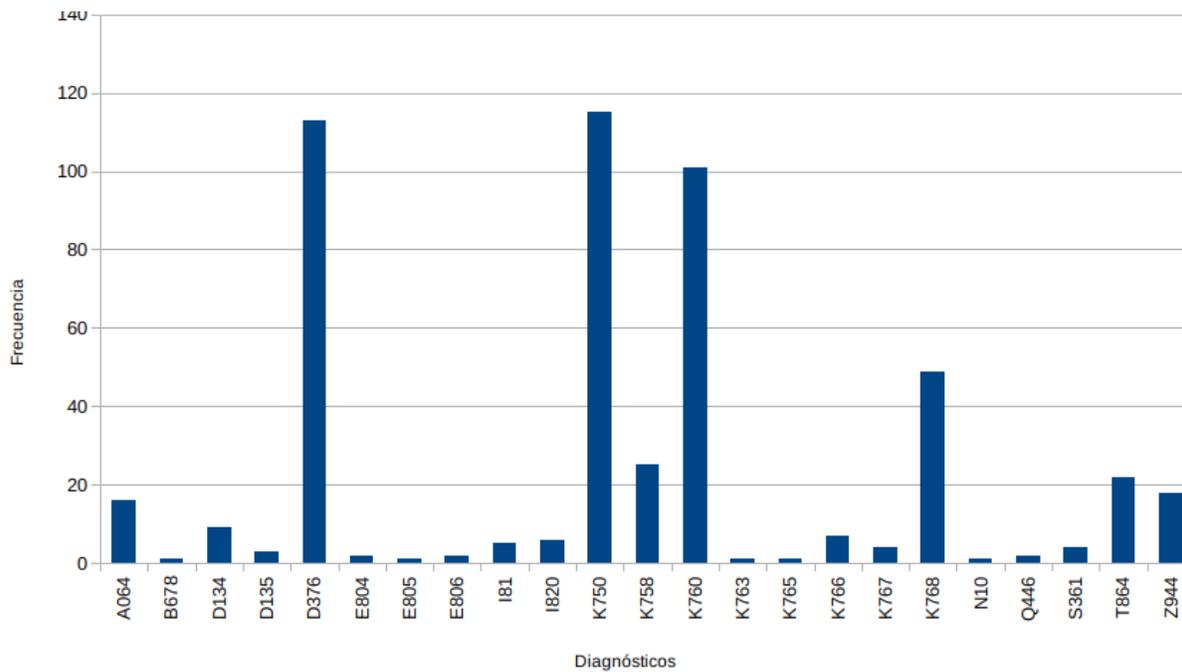
Si se hace el mismo análisis, esta vez para el GRD C-10-14: Otros trastornos hepáticos, que corresponde a la clase más pequeña (clase 46 con 496 registros) podemos observar lo siguiente: para la RNA la métrica *F1-score* es de 0,9744 y para el XGBoost es de 0,9641. Analizando la matriz de confusión para estos mismos GRD C10-14 (gráfica 5-4), se ilustra que 95 de los 99 registros quedan correctamente clasificados. Este buen resultado se debe a que cada egreso cuenta con 23 diagnósticos diferentes, que a su vez están representados en 11 de 22 MDC y a que el 60% de los registros pertenecen al capítulo XI: *Enfermedades del sistema digestivo* del MDC y que están representados en los diagnósticos listados en la tabla 5-19. En la gráfica 5-7 se ilustran los 23 diagnósticos diferentes y su frecuencia para la clasificación de pacientes en el GRD C-10-14 (clase 46)



**Gráfica 5-6:** Gráfica de frecuencia de diagnósticos la clase 0 (GRD C-09-07).

No.	Código CIE-10	Descripción CIE-10
1	K750	Absceso del hígado
2	K758	Otras enfermedades inflamatorias del hígado, especificadas
3	K760	Degeneración grasa del hígado, no clasificada en otra parte
4	K763	Infarto del hígado
5	K765	Enfermedad veno-oclusiva del hígado
6	K766	Hipertensión portal
7	K767	Síndrome hepatorenal
8	K768	Otras enfermedades especificadas del hígado

**Tabla 5-19:** Códigos de diagnóstico asociados a la clase 46 (DRG C-10-14) que representan el 60%



**Gráfica 5-7:** Gráfica de frecuencia de diagnósticos la clase 46 (GRD C-10-14).

### 5.7.2. Evaluación de los modelos M24 al M192.

En la tabla 5-20 se relacionan los valores de *F1-score* de las 47 etiquetas (GRD) por cada uno de los nueve modelos de predicción definidos y el modelo MBL, implementados en la RNA. Para el caso del XGBoost, esta misma métrica se relaciona en la tabla 5-21.

ETIQUETA	número de muestras	M0	M24	M48	M72	M96	M120	M144	M168	M192	BL
C-09-07	4.656	0,9989	0,9989	0,9989	0,9989	0,9989	0,9989	0,9984	0,9984	0,9984	0,9979
C-09-08	3.806	0,9948	0,9948	0,9948	0,9948	0,9948	0,9948	0,9948	0,9941	0,9948	0,9935
C-13-04	3.709	0,9987	0,9993	0,9993	0,9993	0,9973	0,9973	0,9980	0,9987	0,9980	0,9973
C-08-04	3.086	0,9904	0,9896	0,9896	0,9896	0,9879	0,9871	0,9887	0,9845	0,9895	0,9887
C-05-18	2.342	0,9915	0,9915	0,9925	0,9894	0,9915	0,9915	0,9947	0,9915	0,9915	0,9915
C-10-08	2.191	0,9842	0,9809	0,9831	0,9831	0,9831	0,9820	0,9798	0,9820	0,9809	0,9853
C-04-04	2.188	0,9842	0,9842	0,9842	0,9831	0,9831	0,9853	0,9842	0,9864	0,9887	0,9842
C-05-14	2.077	0,9964	0,9964	0,9964	0,9952	0,9952	0,9964	0,9940	0,9952	0,9976	0,9940
C-10-10	1.738	0,8610	0,8722	0,8823	0,8968	0,9143	0,8925	0,9161	0,9086	0,9158	0,9405
C-13-05	1.406	0,9982	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9982	0,9982
C-08-16	1.335	0,9888	0,9888	0,9907	0,9870	0,9889	0,9907	0,9907	0,9907	0,9888	0,9888
C-10-15	1.334	0,5793	0,6466	0,6501	0,7340	0,7786	0,7325	0,8362	0,8321	0,8142	0,8516
P-10-12	1.324	0,6501	0,6655	0,6828	0,6929	0,7089	0,7411	0,7868	0,7636	0,7860	0,8098
C-02-02	1.246	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
P-12-13	1.180	0,8124	0,8294	0,8352	0,8412	0,8397	0,8386	0,8423	0,8520	0,8519	0,8322
C-08-01	1.122	0,8281	0,8151	0,8412	0,8597	0,8753	0,8796	0,8756	0,8858	0,8899	0,8879
C-13-03	1.065	1,0000	0,9978	0,9965	1,0000	0,9933	0,9956	0,9933	0,9978	0,9933	0,9978
C-10-02	1.023	0,9828	0,9828	0,9878	0,9878	0,9828	0,9828	0,9829	0,9854	0,9854	0,9878
P-08-12	999	0,5183	0,3291	0,4672	0,5503	0,6154	0,6346	0,6599	0,6462	0,6730	0,6485
C-05-16	921	0,9779	0,9677	0,9778	0,9778	0,9779	0,9754	0,9779	0,9803	0,9827	0,9777
C-10-01	898	0,8974	0,8564	0,8872	0,8595	0,8922	0,8762	0,8872	0,9049	0,9040	0,9282
C-09-05	894	0,9972	0,9972	0,9972	0,9972	0,9972	0,9972	0,9972	0,9972	0,9944	0,9972
C-10-12	877	0,9859	0,9859	0,9859	0,9859	0,9859	0,9859	0,9887	0,9887	0,9859	0,9859

P-11-01	867	<b>0,5217</b>	<b>0,5249</b>	<b>0,5346</b>	<b>0,5817</b>	<b>0,6254</b>	<b>0,6073</b>	<b>0,6316</b>	<b>0,6195</b>	<b>0,6667</b>	<b>0,7294</b>
C-10-11	850	0,9913	0,9913	0,9913	0,9913	0,9913	0,9913	0,9913	0,9913	0,9913	0,9913
C-17-02	793	0,9914	0,9886	0,9885	0,9886	0,9885	0,9886	0,9886	0,9886	0,9857	0,9914
C-12-04	787	<b>0,4390</b>	<b>0,5775</b>	<b>0,6549</b>	<b>0,6634</b>	<b>0,6704</b>	<b>0,6772</b>	<b>0,6766</b>	<b>0,6822</b>	<b>0,7075</b>	<b>0,7268</b>
C-12-07	782	0,8175	0,8667	0,8704	0,8889	0,9073	0,9177	0,9172	0,9125	0,9191	0,9205
C-11-02	777	0,9679	0,9679	0,9679	0,9711	0,9679	0,9646	0,9646	0,9711	0,9742	0,9615
C-12-14	761	0,9644	0,9587	0,9379	0,9524	0,9401	0,9375	0,9321	0,9235	0,9235	0,8856
C-09-06	713	0,9930	0,9965	0,9965	0,9965	0,9965	0,9895	0,9930	0,9965	0,9965	0,9930
C-01-04	695	0,9964	0,9929	0,9928	0,9929	0,9892	0,9892	0,9856	0,9892	0,9892	0,9856
C-01-06	652	0,9848	0,9848	0,9848	0,9848	0,9811	0,9848	0,9811	0,9738	0,9848	0,9811
C-08-15	639	<b>1,0000</b>									
C-05-15	637	<b>1,0000</b>	0,9961	<b>1,0000</b>							
C-10-04	629	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917
P-12-15	619	<b>0,5132</b>	<b>0,4000</b>	<b>0,5100</b>	<b>0,4176</b>	<b>0,4255</b>	<b>0,3832</b>	<b>0,3522</b>	<b>0,3636</b>	<b>0,4162</b>	<b>0,4545</b>
P-12-11	596	<b>0,6594</b>	<b>0,5234</b>	<b>0,5837</b>	<b>0,5413</b>	<b>0,6215</b>	<b>0,5944</b>	<b>0,5690</b>	<b>0,5200</b>	<b>0,6693</b>	<b>0,6957</b>
C-09-09	570	<b>1,0000</b>									
P-10-02	567	<b>0,3388</b>	<b>0,2870</b>	<b>0,3118</b>	<b>0,3581</b>	<b>0,3133</b>	<b>0,2618</b>	<b>0,2619</b>	<b>0,3676</b>	<b>0,3439</b>	<b>0,4606</b>
C-08-12	552	0,9135	0,9340	0,9727	0,8800	0,9772	0,9725	0,9725	0,9725	0,9725	0,9630
C-17-03	547	0,9143	0,9143	0,9292	0,9464	0,9488	0,9455	0,9502	0,9327	0,9378	0,9450
C-13-09	546	0,9817	0,9863	0,9864	0,9864	0,9863	0,9727	0,9774	0,9818	0,9774	0,9686
C-05-06	515	0,9904	0,9904	0,9904	0,9904	0,9904	0,9855	0,9855	0,9904	0,9855	0,9904
C-01-02	515	0,9763	0,9763	0,9763	0,9763	0,9763	0,9668	0,9761	0,9810	0,9668	0,9714
C-11-01	510	0,9245	0,9395	0,9302	0,9352	0,9346	0,9346	0,9439	0,9484	0,9390	0,9528
C-10-14	496	0,9744	0,9848	0,9898	0,9898	0,9898	0,9898	0,9898	0,9898	0,9898	0,9898

Tabla 5-20: F1-score por modelo por clase en la RNA.

ETIQUETA	número de muestras	M0	M24	M48	M72	M96	M120	M144	M168	M192	BL
C-09-07	4.656	0,9979	0,9973	0,9946	0,9941	0,9941	0,9941	0,9957	0,9941	0,9952	0,9952
C-09-08	3.806	0,9974	0,9967	0,9941	0,9934	0,9934	0,9934	0,9941	0,9941	0,9948	0,9941
C-13-04	3.709	0,9886	0,9886	0,9912	0,9892	0,9892	0,9886	0,9919	0,9892	0,9899	0,9899



C-05-15	637	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
C-10-04	629	0,9958	0,9958	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917	0,9917
P-12-15	619	0,5203	0,6025	0,5667	0,6068	0,6667	0,6946	0,7377	0,7311	0,7426	0,7083
P-12-11	596	0,6160	0,6432	0,7012	0,7323	0,7668	0,7769	0,8049	0,7871	0,8048	0,7273
C-09-09	570	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
P-10-02	567	0,3558	0,3626	0,4318	0,4615	0,5083	0,5109	0,5000	0,5217	0,5109	0,6154
C-08-12	552	0,9821	0,9730	0,9908	0,9862	0,9862	0,9862	0,9862	0,9863	0,9908	0,9818
C-17-03	547	0,8981	0,8930	0,9304	0,9264	0,9345	0,9345	0,9427	0,9386	0,9511	0,9427
C-13-09	546	0,8745	0,8426	0,8841	0,8646	0,8559	0,8485	0,8621	0,8609	0,8621	0,8458
C-05-06	515	0,9952	0,9952	0,9952	0,9952	0,9952	0,9952	0,9952	0,9952	0,9952	0,9952
C-01-02	515	0,9488	0,9528	0,9528	0,9573	0,9573	0,9474	0,9519	0,9519	0,9524	0,9479
C-11-01	510	0,9074	0,9116	0,9245	0,9202	0,9202	0,9333	0,9378	0,9327	0,9289	0,9202
C-10-14	496	0,9641	0,9588	0,9744	0,9694	0,9694	0,9694	0,9645	0,9694	0,9645	0,9645

**Tabla 5-21:** *F1-score* por modelo por clase en XGBoost

En la sección 5.5.1 se analizaron algunos casos, los mejores evaluados y los que presentaron desempeños más pobres, como fueron los casos de C-10-15, P-10-12, P-08-12 y P-10-02 (desempeño más pobre); y C-09-07 y C-10-14 que son GRD entre los mejores desempeños y tienen la particularidad de que son las clases con más registros (4.656) y la clase con menos registros (496) respectivamente, en todo el conjunto de datos. Lo que implica que el manejo del desbalance de clases en este proyecto de investigación es el apropiado.

Ahora bien, con base en la información relacionada en las tablas 5-20 y 5-21, podemos observar unos GRD con clasificación perfecta en todos los modelos propuestos, es decir, la predicción de los modelos desde M0 hasta M192 para los grupos C-02-02, C-08-15 y C-09-09 en la RNA; y C-05-14, C-05-15 y C-09-09 en el XGBoost fueron clasificados correctamente el 100% de las muestras. En término clínicos el paciente en la admisión fue clasificado en un GRD, solo usando el diagnóstico principal el modelo M0 predijo el GRD del paciente y durante toda la estancia, sin importar los procedimientos realizados o las complicaciones o comorbilidades presentadas, el paciente no cambió de GRD

cuando fue evaluado por otros modelos hasta llegar al MBL. Este comportamiento se debe a que existen unos pocos diagnósticos que conllevan a la clasificación de estos GRD.

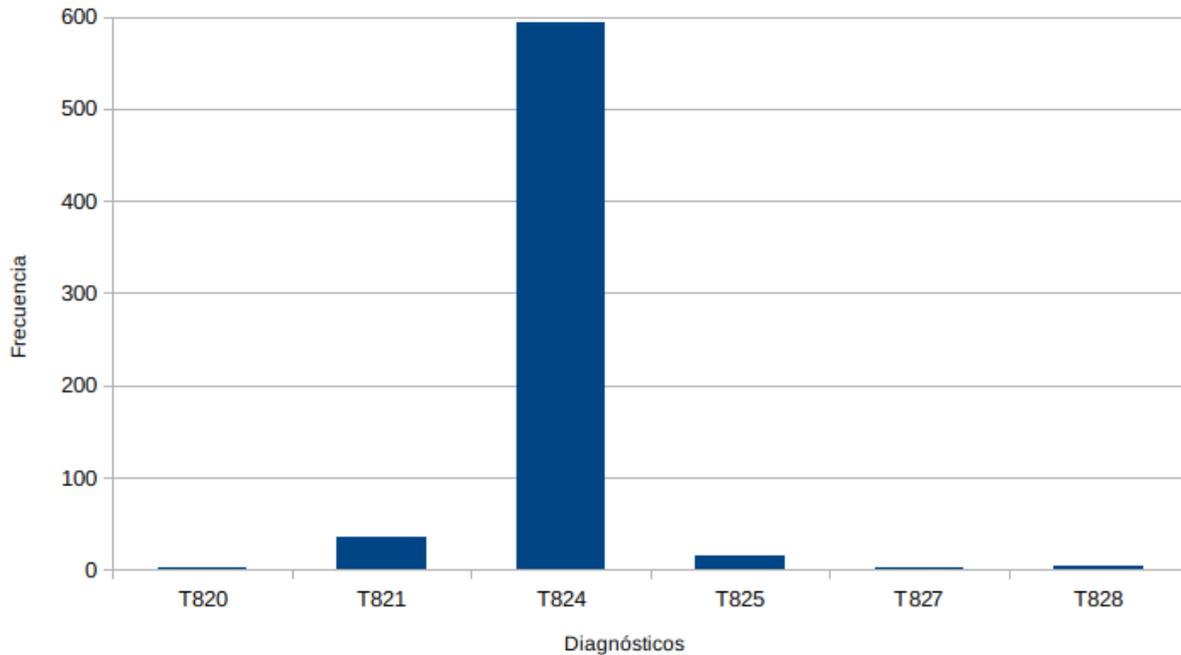
El *C-02-02: Admisión para sesión de quimioterapia* es uno de los GRD que tiene un desempeño del 100% en todos los modelos de predicción propuestos con la técnica RNA, este resultado se debe a que los diagnósticos están concentrados sólo en dos códigos CIE-10 que son el *Z511: Sesión de quimioterapia por tumor* y *Z512: Otra quimioterapia*. De los 1.246 registros que tiene este GRD (clase 13) el 99.0% están representados en el diagnóstico *Z511*. Clínicamente, todos los pacientes que son admitidos en hospitalización por motivo de estos dos diagnósticos son pacientes que ingresan para sesión de quimioterapia.

El GRD *C-08-15: Complicaciones ocasionadas por dispositivos y procedimientos cardiovasculares* (clase 33), es otro de los GRD que obtuvieron una clasificación temprana del 100% en todos los modelos propuesto para la técnica RNA, lo que se atribuye a que la mayoría de los diagnósticos principales están concentrados en un solo códigos. En la tabla 5-22 se relacionan los códigos y las descripciones de los diagnósticos asociados a este GRD. Como se observa, todos los códigos inician por la letra *T*, lo que corresponde al *Capítulo XIX: Traumatismos, envenenamientos y algunas otras consecuencias de causa externa* del MDC. En la gráfica 5-8 se observan las frecuencias de los diagnósticos y el código *T824* representa el 90.5% de todos los diagnósticos que conllevan a que todos los modelos de predicción propuestos y el MBL tome la decisión de clasificar al paciente en este GRD.

Código CIE-10	Descripción CIE-10
T820	Complicación mecánica de prótesis de válvula cardíaca
T821	Complicación mecánica de dispositivo electrónico cardíaco
T824	Complicación mecánica de catéter para diálisis vascular
T825	Complicación mecánica de otros dispositivos e implantes cardiovasculares

T827	Infección y reacción inflamatoria por otros dispositivos, implantes e injertos cardiovasculares
T828	Otras complicaciones de dispositivos protésicos, implantes e injertos cardiovasculares

**Tabla 5-22:** Diagnósticos relacionados con el GRD C-08-15 (clase 33).

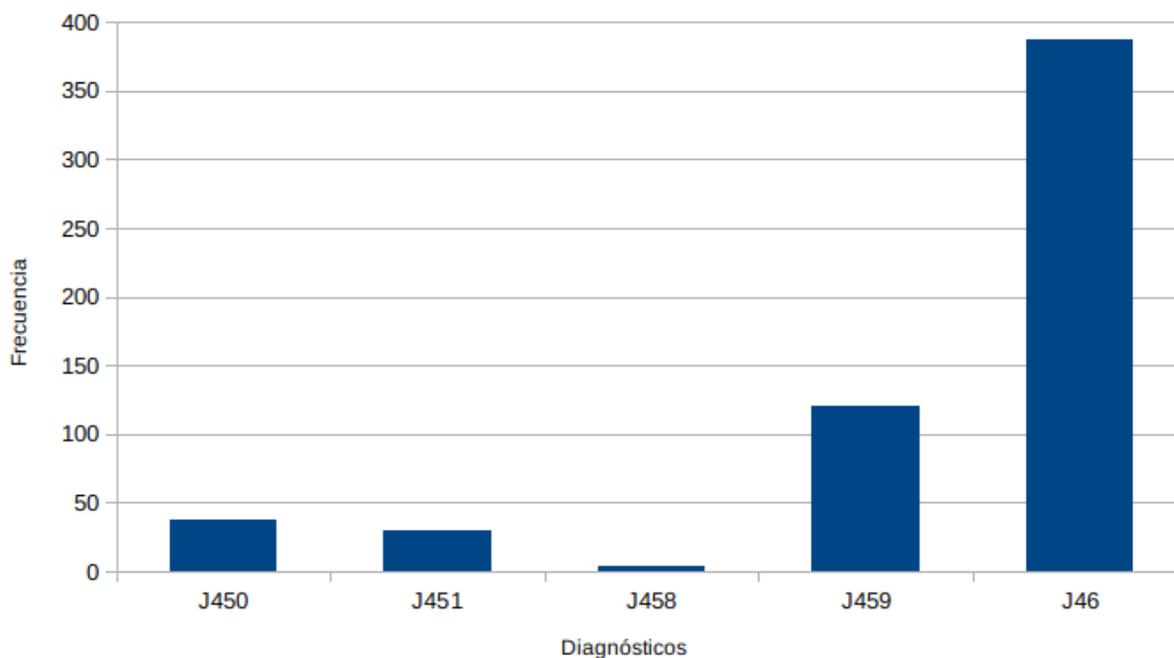


**Gráfica 5-8:** Gráfica de frecuencia de diagnósticos la clase 13 (GRD C-08-15).

La tabla 5-23 y la gráfica 5-9 relacionan los procedimientos que asocian al paciente con el GRD C-09-09: *Asma*. El primer elemento alfanumérico del código del diagnóstico es la letra *J*, corresponde al *Capítulo X: Enfermedades del sistema respiratorio* del MDC. Clínicamente estos tres elementos; Código CIE-10, MDC y GRD están directamente relacionados y explican por qué los modelos sin importar la técnica clasifican el 100% de los pacientes en el GRD correcto.

Código CIE-10	Descripción CIE-10
J450	Asma predominantemente alérgica
J451	Asma no alérgica
J458	Asma mixta
J459	Asma, no especificada
J46	Estado asmático

**Tabla 5-23:** Diagnósticos relacionados con el GRD C-09-09 (clase 38).



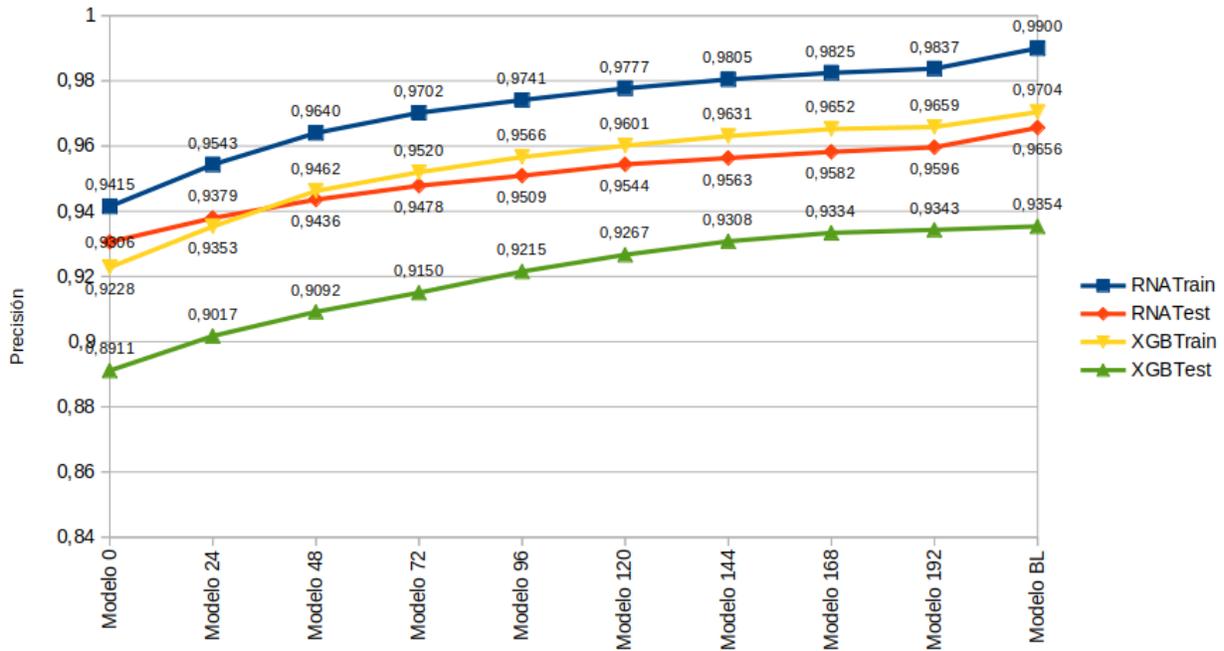
**Gráfica 5-9:** Gráfica de frecuencia de diagnósticos la clase 38 (GRD C-09-09)

Para hacer un análisis más detallado de los GRD en los diferentes modelos de predicción, en Anexo **A** se muestran todas las matrices de confusión, las cuales se pueden contrastar con las tablas 5-18 y 5-19 que tienen las métricas *F1-score* por cada modelo en cada GRD.

### 5.7.3. Comparación de las técnicas RNA y XGBoost.

En términos generales, el desempeño de los modelos en las dos técnicas de ML, la RNA y el XGBoost es muy bueno. En el modelo M0, donde la única información médica disponible es el diagnóstico principal que determinó el médico en el momento de hacer la admisión hospitalaria del paciente, los valores de precisión en la etapa de prueba son 0,9306 para la RNA y 0,8911 para el XGboost. Se observa claramente una tendencia a mejorar el desempeño de las técnicas en los diferentes modelos a lo largo del tiempo. Esto se evidencia con el desempeño del modelo M192, que contiene la información de ocho días de evolución clínica del paciente y donde los valores de precisión fueron de 0,9596 y 0,9343 para las técnicas RNA y XGBoost respectivamente. A medida que avanza la estancia del paciente, se observa una mejora en la precisión de los modelos de la RNA evaluados del 3,50% y del 4,43% en los mismos modelos implementados en el XGBoost. Clínicamente hablando, si se implementa la metodología propuesta con la RNA, durante la admisión (Modelo M0) clasificaría 93 de 100 pacientes correctamente, mientras que el modelo M192 clasificaría 96 de 100 pacientes correctamente; es decir, a medida que aumenta el tiempo y las historias clínicas se van alimentando con la evolución de los pacientes, la precisión en la metodología va mejorando y hay tres pacientes que con la información de los procedimientos realizados cambian su predicción de GRD. Haciendo este mismo análisis con el XGBoost, cuatro pacientes cambiarían su predicción a medida que avance su estancia en el hospital.

Si bien el porcentaje de mejora es mayor en el XGBoost, lo que implica que, con una mayor cantidad de información a la entrada del modelo, el XGBoost incrementa más rápidamente su desempeño que la RNA, pero no es suficiente para superar la precisión que tiene la RNA. En la gráfica 5-10 se ilustran las métricas de precisión para cada uno de los nueve modelos de predicción y el modelo MBL implementados en las dos técnicas de ML (RNA y XGBoost), tanto para las etapas de entrenamiento (*train*) como las de prueba (*test*).



**Gráfica 5-10:** Precisión por modelo en las etapas de entrenamiento y prueba para la RNA y el XGBoost.

En la tabla 5-24 se relaciona la técnica que obtuvo el mejor resultado de *F1-score* en la evaluación de cada uno de los modelos de predicción propuestos y el modelo MBL por cada uno de los 47 GRD evaluados en este experimento. Se asignó la etiqueta (RNA y XGB) y un color a la técnica sobresaliente en cada caso particular. Se observa una mezcla de mejores resultados entre ambas técnicas, pero numéricamente; en 23 GRD la mejor clasificación la dio la RNA, 11 GRD fueron mejor clasificados con el XGBoost y finalmente 13 GRD tuvieron un comportamiento mixto, es decir, para unos GRD en particular, en algunos modelos fue mejor el desempeño de la RNA y en otros el XGBoost.

ETIQUETA	M0	M24	M48	M72	M96	M120	M144	M168	M192	BL
C-09-07	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-09-08	XGB	XGB	RNA	RNA	RNA	RNA	RNA	RNA	RNA	XGB
C-13-04	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-08-04	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-05-18	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-10-08	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-04-04	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-05-14	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-10-10	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-13-05	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-08-16	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-10-15	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
P-10-12	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-02-02	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
P-12-13	RNA	XGB	XGB	RNA	XGB	XGB	XGB	XGB	XGB	XGB
C-08-01	RNA	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-13-03	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-10-02	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
P-08-12	RNA	XGB	XGB	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-05-16	RNA	XGB	RNA	RNA	RNA	RNA	RNA	RNA	RNA	XGB
C-10-01	RNA	RNA	RNA	XGB	XGB	XGB	RNA	RNA	RNA	RNA
C-09-05	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-10-12	RNA	RNA	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
P-11-01	RNA	XGB	XGB	XGB	RNA	XGB	XGB	XGB	RNA	RNA
C-10-11	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-17-02	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-12-04	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-12-07	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	RNA
C-11-02	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-12-14	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-09-06	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-01-04	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-01-06	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-08-15	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-05-15	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	XGB	RNA
C-10-04	XGB	XGB	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
P-12-15	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
P-12-11	RNA	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-09-09	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
P-10-02	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-08-12	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-17-03	RNA	RNA	XGB	RNA	RNA	RNA	RNA	XGB	XGB	RNA
C-13-09	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-05-06	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB	XGB
C-01-02	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-11-01	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
C-10-14	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA

Tabla 5-24: Comparativo de mejores *F1-score* entre la RNA y el XGBoost

Haciendo un análisis en cuanto al consumo de recursos tecnológicos y frente a los tiempos de ejecución de los entrenamientos de las técnicas de ML para cada uno de los nueve modelos de predicción y el modelo MBL establecidos en la metodología propuesta (ver tablas 5-8 y 5-14), claramente se ve que el consumo es mucho más alto en la técnica XGBoost que la RNA. Para el procesamiento de las muestras con la técnica XGBoost se aprovechó la implementación de la paralelización en la tarea de entrenamiento que está implementada en la *API* de *Scikit-Learn*, es decir, se usaban los 12 núcleos de clúster y los tiempos para los diez modelos están entre 6 y 14 horas de ejecución. En comparación con la RNA, donde se usó la *API* implementada en *Keras* y donde no se implementó ningún proceso de paralelización de las tareas y los tiempos de ejecución de los

entrenamientos para los diez modelos estuvieron entre 5 y 7 minutos. A estos tiempos de ejecución, también se le suma el hecho de que el XGBoost guarda en memoria los 300 estimadores por cada modelo, lo que hace que este proceso sea inmanejable en un computador con bajos recursos. Viéndolo desde el punto de vista temporal y de recursos de máquina, se puede decir que es más costosa la implementación de la técnica XGBoost que la RNA.

## 5.8. Importancia de características.

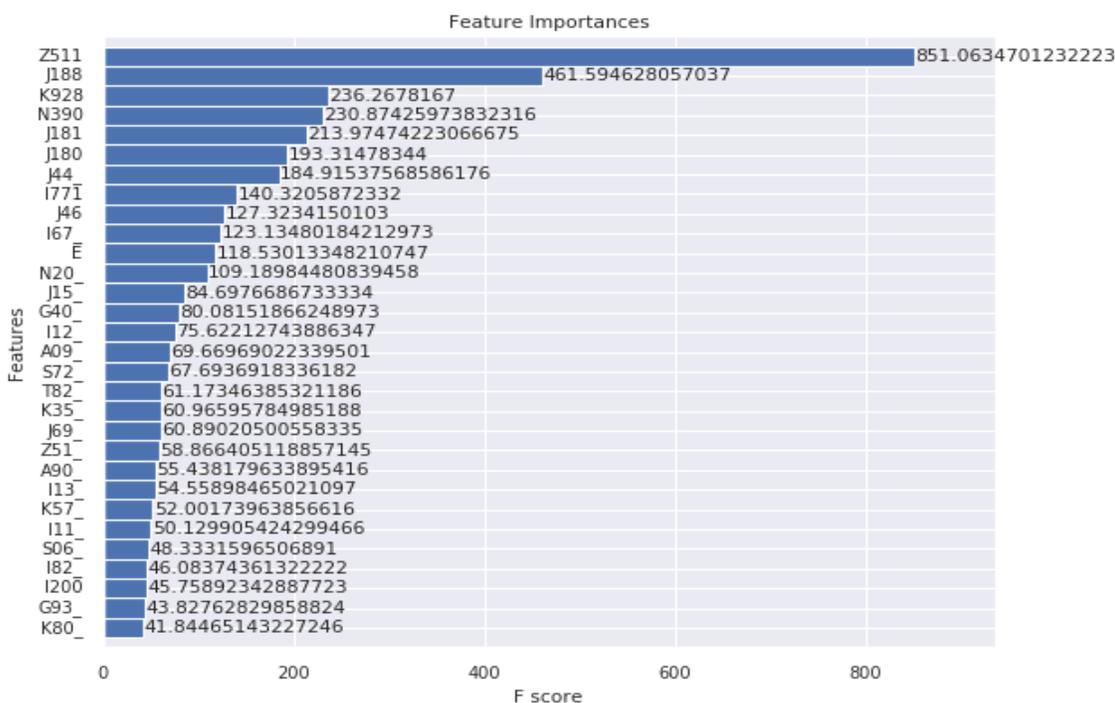
La técnica XGBoost tiene implementados métodos para calcular y graficar la importancia de las características utilizadas en la matriz de entrada a los modelos. La importancia de una característica depende de si el rendimiento de la predicción cambia considerablemente cuando dicha característica se reemplaza con ruido aleatorio [52]. Para analizar la importancia de las características en este trabajo de investigación hay que tener en cuenta los atributos seleccionados y descritos en la sección 4.3 y que se relacionan en tabla 4-4.

Se seleccionaron nueve atributos que luego fueron codificados y las matrices ingresadas a los diferentes modelos cambiaron su estructura, estos nueve atributos se convirtieron en 2.777 características de entrada para el modelo M0 y en 3.474 para el modelo M24 y así sucesivamente como se describe en la tabla 5-1. Así, la importancia de los atributos seleccionados sería la sumatoria de las importancias de las características que las representan.

En la gráfica 5-11 se ilustra el top 30 de la importancia de las características en la ejecución del modelo M0; como es de esperarse, las 30 características hacen referencia a los diagnósticos principales y tiene sentido puesto que, a este modelo solo ingresan la información demográfica y la información del diagnóstico principal, pero como se muestra en la tabla 4-4, el diagnóstico principal está descrito en tres atributos;

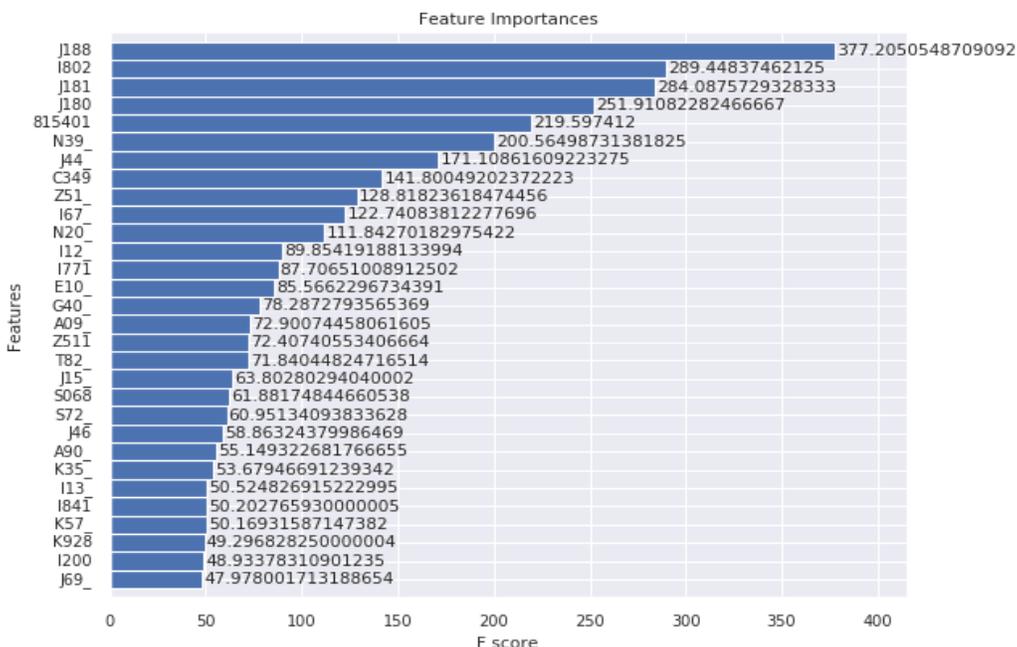
1. Diagnósticos CIE10 – MDC,
2. Diagnósticos CIE10 - (La letra y dos dígitos) y
3. Diagnósticos CIE10 - Código completo.

De este top 30 podemos destacar que nueve características corresponden al *Diagnósticos CIE10 - Código completo*, 20 son del *Diagnósticos CIE10 - (La letra y dos dígitos)* y sólo una representa al *Diagnósticos CIE10 – MDC*; pero sumando las importancias de las características por cada atributo y expresando los resultados en términos de ganancia de importancia, lo encontrado fue: 2.496 para los *Diagnósticos CIE10 - Código completo*, 1.419 para los *Diagnósticos CIE10 - (La letra y dos dígitos)* y finalmente 118 para los *Diagnósticos CIE10 – MDC*. En conclusión, aunque el modelo M0, usó más características que representaban el atributo *Diagnósticos CIE10 - (La letra y dos dígitos)* para hacer la clasificación temprana de los pacientes, fueron más importantes los diagnósticos con el código CIE-10 completo, es decir, lo que representan el atributo *Diagnósticos CIE10 - Código completo*, finalmente para este modelo, no es de mucha importancia el atributo que relaciona los capítulos del MDC.



Gráfica 5-11: top 30 de importancia de características en el modelo M0.

Continuando con el análisis de la importancia de características del modelo M0, el diagnóstico más importante fue el *Z511: Quimioterapia por tumor*, que es el diagnóstico principal que se asignó a pacientes que fueron clasificados en el GRD *C-02-02: Admisión para sesión de quimioterapia* y que fue uno de los pocos que obtuvo una clasificación del 100% en todos los modelos propuestos.



**Gráfica 5-12:** top 30 de importancia de características en el modelo M24.

En la gráfica 5-12, ilustra el top 30 de las características más importantes en la clasificación del modelo M24. A este modelo ya se adiciona a la matriz de entrada los procedimientos realizados a los pacientes durante las primeras 24 horas de la estancia hospitalaria. Por esta razón, en el top 30 de las características más importantes aparece un código CUPS, el código *815401: Reemplazo total de rodilla bicompartimental*, procedimiento que está directamente relacionado con el GRD *C-12-14: Otros trastornos osteomioarticulares (clase 29)*, lo que implica que este procedimiento es una de las características principales para el que el modelo clasifique un paciente en este GRD. Analizando el conjunto de datos para este resultado en particular, se encuentra que se hicieron 16 procedimientos con el código *815401* y todos los egresos que tuvieron este procedimiento fueron clasificados en el GRD *C-12-14*.

Tres de las cuatro características más importantes para este modelo M24 son los procedimientos *J188 - Otras neumonías, de microorganismo no especificado*; *J181 - Neumonía lobar, no especificada* y *J180 - Bronconeumonía, no especificada*. Estos son los procedimientos principales para que el modelo M24 clasifique a los pacientes en el GRD *C-09-07 - Neumonía y neumonitis* (clase 0). Estos mismos procedimientos estuvieron entre las seis características más importantes del modelo M0.

En el Anexo **C** se podrán encontrar todas las gráficas que ilustran la importancia de las características por cada modelo de predicción propuesto y el modelo MBL. En todos, los atributos más importantes son los relacionados con el diagnóstico principal del paciente, en los nueve modelos de predicción, dentro de los top 30 de características, solo siete códigos de procedimientos aparecen como características importantes en la clasificación de los GRD. En la tabla 5-25 se relacionan los procedimientos que fueron los más importantes para la clasificación de los pacientes en los nueve modelos de predicción según la técnica XGBoost, por ejemplo, el procedimiento *815402: Reemplazo total de rodilla tricompartmental*, aparece como característica importante dentro de los top 30 en siete de los nueve modelos evaluados.

No	Código CUPS	Descripción	Frecuencia por Modelo
1	815402	Reemplazo total de rodilla tricompartmental	7
2	3890021	Colocación catéter arterial periférico (línea arterial) ncoc	5
3	13760	Reducción abierta de luxación acromio clavicular	3
4	511102	Colangiografía endoscópica retrograda (transduodenal)	3
5	361601	Anastomosis simple de arteria descendente anterior con injerto venoso vía abierta	2
6	815401	Reemplazo total de rodilla bicompartmental	1
7	881390	Ecografía del abdomen y pelvis como guía de procedimiento quirúrgico o intervencionista	1

**Tabla 5-25:** Procedimientos más importantes para clasificación según el XGBoost.

Continuando con el análisis del Anexo **C**, para el caso del atributo que representa los capítulos del MDC, es decir, que tiene por nombre *Diagnósticos CIE10 - MDC* aporta poco a la clasificación de pacientes, ya que las únicas características de estas que aparecieron en los top 30 fueron *E - Enfermedades endocrinas, nutricionales y metabólicas* y *Z - Factores que influyen en el estado de salud y contacto con los servicios de salud*; la característica *E* apareció en ocho de los nueve modelos de predicción y la *Z* en dos de nueve. Lo que puede entenderse como un aporte bajo del atributo *Diagnósticos CIE10 – MDC* en la clasificación de pacientes.

De los análisis anteriores se puede deducir que las características que más aportan a la clasificación temprana de pacientes en GRD son las que representan el diagnóstico principal, se observa que las características que denotan procedimientos o información demográfica aportan poco o nada a este proceso, esta última no aparece en ninguno de los top 30 de los modelos propuestos.

## **5.9. Conclusiones.**

Con base en los resultados obtenidos podemos concluir que el modelo M0 es un gran acierto en la metodología propuesta para el proceso de predicción porque; 1) es el modelo con la información más escasa para hacer una clasificación; 2) obtiene métricas de precisión del 93% en la RNA y 89% en el XGBoost, que son más altas en comparación con los resultados de otros trabajos reportados en la literatura (ver tabla 3.12); y 3) solo utilizando la información demográfica del paciente y el diagnóstico principal que reporta el médico tratante en el momento de la admisión.

Las características realmente importantes para la predicción de un paciente en un GRD son las relacionadas con el diagnóstico principal, los atributos que denotan el código CIE-10 que reporta el médico tratante en el momento de la admisión hospitalaria. Se llega a esta conclusión mirando en conjunto los modelos de predicción propuestos desde M0 hasta M192, ya que usando la RNA, el modelo M0 clasifica correctamente 93 de 100

pacientes que son admitidos; luego de ocho días de evolución de la historia clínica de los pacientes, solo cambia la clasificación del GRD para tres pacientes, ya que 96 de 100 pacientes son clasificados correctamente por M192; es decir, los procedimientos y las pruebas diagnósticas realizados en los primeros ocho días de la estancia hospitalaria sólo aportan información a la clasificación del 3% de pacientes.

Ambas técnicas procesando los diez modelos planteados presentan en los GRD quirúrgicos; *P-10-12 - Colectomía laparoscópica*, *P-08-12 - Cateterismo cardiaco y coronariografía con o sin angioplastia sin inserción de stent coronario*, *P-11-01 - Injerto de piel*, *P-12-15 - Otros procedimientos en rodilla y pierna excepto artroplastia*, *P-12-11 - Procedimientos en miembros superiores excepto mano* y *P-10-02 - Procedimientos de mayor complejidad sobre tracto gastrointestinal*, los desempeños más pobre. Existe una excepción en este análisis, es el GRD *P-12-13 - Otros procedimientos en cadera y fémur excepto artroplastia*, que si bien no tiene el mejor desempeño del sistema, si tiene las métricas *F1-score* entre el 0,80 y 0,91 corrido en el XGBoost y entre 0,81 y 0,85 usando la RNA. De lo anterior se puede concluir que los GRDs catalogados como quirúrgicos pueden tener la distribución de sus datos de entrada poco consistentes de tal forma que confunden el clasificador y los obliga a cometer errores en la tarea de clasificación, lo que no ocurre con los GRD médicos, que si bien hay algunos GRD de este tipo que no tienen los mejores desempeños, la tendencia general es a que tanto la RNA como el XGBoost tienen una mediciones de *F1-score* altas en la tarea de clasificación.

Haciendo un análisis de la tabla 5-24, se determina que de las 470 métricas *F1-score*, es decir, los 47 GRD por cada uno de los diez modelos establecidos, en 295 la RNA obtuvo un mejor resultado en la clasificación de los GRD, el resto, en las otras 175 mediciones, el mejor desempeño lo obtuvo el XGBoost. Pero si se agrupan los rendimientos por desempeño de modelos de cada GRD, en términos generales la RNA clasificó mejor en 23 de 47 GRD (clases) en los diez modelos, mientras que el XGboost solo alcanzó una clasificación mejor en 11 de 47 GRD en los diez modelos. Los 13 GRD restantes tuvieron un desempeño variado, por ejemplo, para el caso del GRD *C-09-08 - Enfermedad pulmonar obstructiva crónica*, el XGBoost obtuvo mejores resultados de

clasificación en los modelos M0, M24 y MBL, mientras que en el resto de los modelos la RNA fue superior.

Ahora bien, como se ha mencionado, ambas técnicas tuvieron pobres desempeños en los procedimientos catalogados como quirúrgicos; pero sin duda alguna, para estos casos en particular, la técnica que menos se equivocó fue el XGBoost. De los siete GRD quirúrgicos, en tres de ellos el XGBoost obtuvo mejor desempeño en los diez modelos; en otros tres, aunque los mejores resultados de desempeño se repartieron entre ambas técnicas, el número de resultados *F1-score* mejores en los modelos por GRD fue para el XGBoost. Solo en uno de estos siete GRD la RNA fue superior, en el GRD *P-08-12: Cateterismo cardiaco y coronariografía con o sin angioplastia sin inserción de stent coronario*, la RNA clasificó mejor en ocho de los diez modelos, para este caso el XGBoost obtuvo mejor *F1-score* para los modelos M24 y M48.

En términos generales y de gestión administrativos podemos concluir que la metodología tiene la capacidad de predecir pacientes en GRD con una precisión muy alta y que si el diagnóstico principal no cambia desde la admisión del paciente y durante su evolución clínica, se puede saber cuánto tiempo va a durar el paciente hospitalizado y que costo tendrá para el Hospital. Por ejemplo, un paciente que sea admitido con el diagnóstico *J441 - Enfermedad pulmonar obstructiva crónica con exacerbación aguda, no especificada*, el modelo M0 lo clasificará en el GRD *C-09-08 - Enfermedad pulmonar obstructiva crónica*, lo que en términos de gestión hospitalaria significa que el paciente va a estar hospitalizado durante cuatro días [53] y si su diagnóstico principal no cambia, los modelos M24 (día 1) , M48 (día 2), M72 (día 3) y M96 (día 4) clasificarán al paciente en el mismo GRD y el M96 sería el último modelo que clasificaría dicho paciente.

# Capítulo 6:

## Resumen, conclusiones, observaciones y estudios futuros.

### 6.1. Resumen.

Este proyecto de investigación tuvo como objetivo principal usar técnicas de inteligencia computacional para la predicción y clasificación de pacientes en GRD usando como criterio de clasificación correcta del CMBD de la IPS Universitaria Clínica León XIII de la ciudad de Medellín Colombia, el cual fue etiquetado con uno de los tantos softwares *DRG-grouper* existentes en el mercado. Se contó con un segundo conjunto de datos que contenía la información de los procedimientos y pruebas diagnósticas realizadas a los pacientes relacionados en el CMBD.

Con el objetivo de conocer el procedimiento de admisión del Hospital caso de estudio, la primera etapa de la investigación fue modelar el proceso de admisión y de atención hospitalaria para entender los pasos que sigue un paciente y cómo es almacenada la información en la historia clínica que luego alimenta el CMBD. Esta tarea se realizó con ayuda de las personas responsables de las admisiones, cómo llegaban los pacientes, cómo se recibían, clasificaban y admitían una vez se cumplía con el criterio de admisión hospitalaria, que es 24 horas después de que ingresó para los casos de ingreso “Urgente”. Este dato es importante porque el CMBD contiene sólo información de los pacientes que son considerados hospitalizados, es decir, si un paciente dura en

urgencias 12 horas, no es catalogado como hospitalizado. En esta etapa del proceso de hospitalización de los pacientes se determina también que la actualización de la evolución clínica de los pacientes se realiza cada 24 horas y para el caso especial de pacientes en UCI o UCE, dicha evolución se hace cada 12 horas. Esta información nos dio el punto de partida para determinar que se debían diseñar modelos de procesamiento de información cada 24 horas, ya que antes de este periodo no hay información actualizada de la evolución del paciente que pudiera cambiar su predicción en la clasificación del GRD.

La etapa siguiente de la investigación fue seleccionar la población objetivo y los GRD representativos que evidenciaban el mayor impacto en costos y la mayor frecuencia de ocurrencia entre los pacientes hospitalizados. Para esto, se obtuvo la información de los CMBD de los años 2015, 2016, 2017 y 2018, pero luego de un análisis inicial de los atributos y sabiendo que la configuración de los datos es diferente porque se usa para software agrupadores diferentes, se determina sólo trabajar con los registros de egresos de los años 2016 al 2018. Con base en la información de las cantidades de las diferentes clases existentes en el CMBD (clase mayoritaria 4656 registros y clases con solo un registro) y lo que podría representar el problema grande en cuanto al desbalance de las clases se decide seleccionar los GRD con mayor cantidad de registros, los cuales fueron la clase con más registros hasta el GRD que representaba el 10% de esta clase mayoritaria. Lo que dio como resultados la selección de los primeros 47 GRD más frecuentes que representaban el 70% de los datos del CMBD.

Luego de tener la población objetivo, se continuó con la identificación del conjunto de variables que más aportan a la clasificación temprana de GRD desde el conjunto de datos de registros de egresos seleccionados en la etapa anterior. Se hace la selección de los atributos, con base en la información disponible, la revisión de artículos del estado del arte y la experiencia del personal médico que apoya el proceso de investigación. Como resultado de 168 atributos que tienen el CMBD, solo se seleccionan nueve; de los cuales, con base en la información recabada y analizada hasta ese momento, se define la información necesaria para hacer la predicción de los pacientes en GRD. El análisis

llevó a que los atributos necesarios para ingresar a los modelos de clasificación se dividieran en tres grupos; 1) información demográfica, 2) información de diagnóstico y 3) información de procedimientos. El primer grupo cuenta con información con edad, género, fecha de ingreso, etc. del paciente; el segundo grupo contiene información solo del diagnóstico principal, en este punto y con base en los análisis hechos previamente, se descartaron los diagnósticos secundarios o comorbilidades; y el tercer grupo contiene la información de todos los procedimientos y pruebas diagnósticas que se hacen al paciente durante su estancia hospitalaria. Este punto de la investigación fue crucial, porque se determinaron los atributos que iban a ingresar a los modelos de clasificación propuestos.

Luego de contar la población objetivo y haber seleccionado los atributos a ingresar en los modelos, se hizo la codificación de la matriz de ingreso a los modelos de clasificación propuestos, la utilización de la codificación binaria para variables como el género, el tipo de ingreso; la normalización para el caso de la edad y el mes de ingreso; la codificación *One-Hot-Encoding* para el diagnóstico principal y las variaciones propuestas para este atributo; y finalmente la matriz de frecuencia para la codificación de los procedimientos realizados a los pacientes durante toda la estancia hospitalaria.

El siguiente paso fue definir la estrategia de la clasificación temprana de GRD basada en técnicas de inteligencia computacional. Se propuso una metodología en la cual se contempla la clasificación temprana de pacientes en GRD en diferentes momentos de su estancia hospitalaria. Un modelo llamado M0, el cual hace una clasificación temprana del paciente utilizando solo la información demográfica y el diagnóstico principal registrado durante la admisión del paciente. Un segundo modelo llamado M24, el cual hace una clasificación temprana del paciente utilizando la misma información de M0 y adicionando los procedimientos y pruebas diagnósticas realizadas al paciente en las siguientes 24 horas a la admisión. Un tercer modelo M48, que clasifica usando la información de M24 más la información de procedimientos y pruebas diagnósticas que registra el médico entre la hora 24 y 48 de la estancia del paciente. Finalmente hacer

este mismo proceso de crear modelos para evaluar la clasificación temprana de los pacientes cada 24 horas hasta que sea dado de alta.

Con base en la revisión del estado del arte, el uso y la importancia de las técnicas de ML actualmente en el ámbito sanitario, se seleccionaron para este trabajo de investigación implementar la metodología propuesta en una Red Neuronal Artificial (RNA) y el *eXtreme Gradient Boosting* (XGBoost) y evaluar cuál ofrece mejores resultados en la solución al problema planteado.

Finalmente, se procede a realizar los experimentos, a utilizar cada una de las matrices de entrada y evaluar los modelos propuestos, para validar la metodología de clasificación temprana de GRD a partir de criterios de exactitud y desempeño. Se usaron las métricas; precisión y el *F1-score* obteniendo resultados destacados en toda la metodología propuesta. El resultado final fue desempeños por encima del 93% de precisión en el modelo M0, es decir, con la cantidad mínima de información y una precisión del 96% al cabo de ocho días de evolución clínica del paciente. Lo que es un muy buen resultado si se compara con el *DRG-grouper* que requiere el 100% de la información para poder hacer la clasificación al egreso del paciente.

## 6.2. Observaciones finales.

La información usada para crear las matrices de entrada a los modelos planteados, está disponible durante toda la evolución del paciente hospitalizado, ya que es la información demográfica, diagnóstico principal, y los procedimientos y pruebas diagnósticas que se realizan al paciente durante su estancia hospitalaria y viendo los resultados obtenidos, se puede evidenciar el aporte de este trabajo de investigación a la clasificación temprana de pacientes en GRD, ya que a diferencia de los otros trabajos reportados en la literatura y al modo de clasificación del *DRG-grouper*, se demostró que se puede hacer predicción de GRD con la información al momento del ingreso.

Durante el proceso se definió la eliminación de los diagnósticos secundarios debido a que en el CMBD no se contaba con información cronológica de estos; es decir, no había evidencia suficiente que indicara que el médico tratante seleccionó un diagnóstico secundario en el momento de la admisión hospitalaria porque el paciente presentaba alguna complicación o comorbilidad, o si estos diagnósticos secundarios se agregaron a la historia clínica en la evolución del paciente durante su estancia.

Como recomendación para la implementación de esta metodología en el quehacer del Hospital caso de estudio, se deben contemplar dos aspectos; 1) la información confiable y 2) la información oportuna; y para esto se debe hacer un cambio organizacional en el hospital debido a que la responsabilidad de alimentar la información en el sistema es del médico tratante. Bajo esta perspectiva, se evidenció que en muchas ocasiones el diagnóstico principal asignado por el médico tratante no es coherente con lo descrito en la historia clínica, el diagnóstico reportado es un código general y no específico o durante la evolución del paciente se determinó que el diagnóstico principal era diferente al reportado en la historia clínica, pero no se hizo la actualización. Todos estos factores y otros tantos harían que la implementación de la metodología no fuera efectiva.

Para este trabajo de investigación se utilizó el 70% de las muestras del conjunto de datos inicial, lo que implica que los resultados de precisión por encima del 93% desde el modelo M0 hasta el modelo M192 no es sobre el 100% de la población del Hospital caso de estudio. Para hacer la implementación de la metodología con esta limitante se sugiere filtrar los casos de admisión usando el diagnóstico principal que el médico tratante asigne al paciente durante la admisión, ya que como se describe en el desarrollo del trabajo, el código del diagnóstico principal tiene un gran peso en la clasificación del GRD, esto se debe a que el primer elemento alfanumérico del código CIE-10 relaciona el sistema corporal involucrado en la afección del paciente. Entonces, si el paciente es diagnosticado con un código CIE-10 que no existe en el conjunto de datos usado para este trabajo de investigación o su frecuencia es muy baja, por ejemplo, uno, dos o tres registros, no debería ser evaluado con la metodología propuesta ya que habría una alta probabilidad de generar falsos positivos.

Puntualizando los anterior, si un paciente en la admisión es diagnosticado con un código CIE-10 que inicia por la letra *H*, será clasificado en uno de los siguientes grupos de GRD *C-06 y P-06 - Enfermedad clínica y quirúrgica de ojo y anexos, C-07 y P-07 - Enfermedad clínica y quirúrgica de nariz, garganta y oído y P-12-04 - Procedimientos sobre huesos de cráneo y cara*; que no están incluidos en los 47 GRD del conjunto de datos seleccionados para este trabajo de investigación. Similar pasa con los diagnósticos que su código inicia por los literal *F, R y O*, que clasifican a los pacientes en los grupos de GRD *C-18: Enfermedades mentales y trastornos del comportamiento, C-19 - Abuso y/o dependencia de alcohol y drogas, O-15 - Embarazo , parto y puerperio* y en los GRD *C-08-13 - Hipotensión ortostática y síncope, C-09-13 - Otros trastornos respiratorios y de vía aérea y X-00-05 - Error de agrupación*.

### **6.3. Limitaciones y estudios futuros.**

Una de las limitaciones que se tuvo en este proyecto de investigación fue haber eliminado el año 2015 del CMBD, que equivalen a 27.869 egresos de pacientes. Ahora bien, se puede usar esta limitante a favor y plantear un estudio con este número de egresos, hacer una validación de la clasificación de los pacientes en GRD de la información de 2015 versus los otros tres años y probar los modelos con muestras que nunca ha visto y evaluar su desempeño.

Como estudio futuro se propone usar la metodología planteada en este trabajo de investigación con una base de datos diferente, por ejemplo, MIMIC (**M**edical **I**nformation **M**art for **I**ntensive **C**are III). Esta base de datos además de ser muy usada y referenciada en investigación informática y médica es de libre uso y contiene datos sanitarios, completamente anonimizados, asociados con más de cuarenta mil pacientes que se quedaron en las unidades de cuidados críticos del Beth Israel Deaconess Medical Center entre los años 2001 y 2012. Si los resultados de los modelos usando esta base de datos son aceptables, se puede hablar de una generalización del problema, ya que, sin importar la población utilizada, los resultados serán positivos.

En la cohorte de estudio, que fueron las matrices de los datos de entrada a los modelos, se usó el 70% de la población, la justificación; 1) las muestras de los otros GRD era pequeñas y podría ocasionar ruido al sistema, 2) se tomaron los GRD que mayor impacto tenía en la IPS Universitaria Clínica León XIII de la ciudad de Medellín Colombia. Como estudio futuro podría ser evaluar este 30% de las muestras restantes, observar el desempeño de los modelos y determinar la proporción en que disminuye la precisión con más clases, pero con menos muestras que las representen.

---

## Referencias bibliográficas.

- [1] Gaviria, A. Z., & Martín, R. B. (2008). Altas atendidas en medicina interna en los hospitales generales del sistema nacional de salud, año 2005. Jarpyo.
- [2] Rivero, A. (2006). Guía pedagógica para la implementación de los sistemas de clasificación de pacientes (GRD) en instituciones prestadoras de servicios de salud Social MdlP, editor. Bogotá: Ministerio de la Protección Social.
- [3] Cortés Martínez, A. E., Yepes Lujan, F. J., & Agudelo-Londoño, S. M. (2018). El sistema de salud colombiano: grupos relacionados de diagnóstico. Editorial Pontificia Universidad Javeriana.
- [4] Gómez, L., Duque, M. y Carmona, J. (2008). La información contable y de costos e las entidades de salud: una herramienta para Antioquia, 53, 145-172.
- [5] Rivero, A., Sendino, M., Cózar, R., Pascual, M., García, J., & García, C. Análisis y desarrollo de los GDR en el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Madrid. 1999.
- [6] Clínicas de Chile. Grupos relacionados de diagnóstico: Marco teórico y experiencias prácticas. Temas de coyuntura. Chile 2010; 34.
- [7] Quentin, W., Rätto, H., Peltola, M., Busse, R., Häkkinen, U., & EuroDRG group. (2013). Acute myocardial infarction and diagnosis-related groups: patient classification and hospital reimbursement in 11 European countries. European heart journal, 34(26), 1972-1981.
- [8] Quentin, W., Scheller-Kreinsen, D., Geissler, A., Busse, R., y EuroDRG Group. (2012). Apendicectomía y grupos relacionados con el diagnóstico (GRD): clasificación de pacientes y reembolso hospitalario en 11 países europeos. Archivos de cirugía de Langenbeck , 397 (2), 317-326.
- [9] Scheller-Kreinsen, D., Quentin, W., Geissler, A., Busse, R., & EuroDRG Group. (2013). Breast cancer surgery and diagnosis-related groups (DRGs): patient

- classification and hospital reimbursement in 11 European countries. *The Breast*, 22(5), 723-732.
- [10] Bellanger, M. M., Quentin, W., & Tan, S. S. (2013). Childbirth and Diagnosis Related Groups (DRGs): patient classification and hospital reimbursement in 11 European countries. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 168(1), 12-19.
- [11] Paat-Ahi, G., Aaviksoo, A., Świderek, M., & EuroDRG group. (2014). Cholecystectomy and Diagnosis-Related Groups (DRGs): patient classification and hospital reimbursement in 11 European countries. *International journal of health policy and management*, 3(7), 383.
- [12] Gaughan, J., & Kobel, C. (2014). Coronary artery bypass grafts and diagnosis related groups: patient classification and hospital reimbursement in 10 European countries. *Health economics review*, 4(1), 4.
- [13] Tan, S. S., Chiarello, P., & Quentin, W. (2013). Knee replacement and Diagnosis-Related Groups (DRGs): patient classification and hospital reimbursement in 11 European countries. *Knee Surgery, Sports Traumatology, Arthroscopy*, 21(11), 2548-2556.
- [14] Mihailovic, N., Kocic, S., & Jakovljevic, M. (2016). Review of diagnosis-related group-based financing of hospital care. *Health services research and managerial epidemiology*, 3, 2333392816647892.
- [15] Yang, J. J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., ... & Pan, H. (2015). Emerging information technologies for enhanced healthcare. *Computers in Industry*, 69, 3-11.
- [16] Schreyögg J, Tiemann O, Busse R (2006a) Cost accounting to determine prices: How well do prices reflect costs in the German DRG-system? *Health Care Management Sci.* 9:269–279.
- [17] Gartner, D., Kolisch, R., Neill, D. B., & Padman, R. (2015). Machine learning approaches for early DRG classification and resource allocation. *INFORMS Journal on Computing*, 27(4), 718-734.

- 
- [18] Mathauer, I., & Wittenbecher, F. (2013). Hospital payment systems based on diagnosis-related groups: experiences in low-and middle-income countries. *Bulletin of the World Health Organization*, 91, 746-756A.
- [19] Guillain, H. 2008. «Casemix in Switzerland», en J. R. Kimberly, G. de Pouvourville y T. D'Aunno (comps.), *The globalization of managerial innovation in health care*. Cambridge, Cambridge University Press.
- [20] Quentin, W., Geissler, A., Scheller-Kreinsen, D., & Busse, R. (2011). Understanding DRGs and DRG-based hospital payment in Europe. *Diagnosis-Related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. Open University Press and Mc Graw Hill: Berkshire, England, 23-36.
- [21] Gorbanev, I., Cortés, A., Agudelo, S., & Yepes, F. J. (2015). Why are diagnosis-related groups unpopular in Colombia?. *Cadernos de saude publica*, 31(9), 2027-2031.
- [22] Rochell, B.; Roeder, N. 2001. «DRG-basierte Entgeltsysteme in Europa», en M. Arnold, M. Litsch y H. Schellschmidt (comps.), *Krankenhaus-Report 2000: Schwerpunkt: Vergütungsreform mit DRGs*. Stuttgart, Schattauer.
- [23] Street, A., O'Reilly, J., Ward, P., & Mason, A. (2011). DRG-based hospital payment and efficiency: theory, evidence, and challenges. *Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals*, 93-114.
- [24] Geissler, A., Quentin, W., Scheller-Kreinsen, D., & Busse, R. (2011). Introduction to DRGs in Europe: Common objectives across different hospital systems. *Diagnosis related groups in Europe: moving towards transparency, efficiency and quality in hospitals*, 9-21.
- [25] Fetter, R. B. (1984). Diagnosis related groups: the product of the hospital. *Clinical research*, 32(3), 336.
- [26] Mathauer, I., & Wittenbecher, F. (2012). DRG-based payment systems in low-and middle-income countries: implementation experiences and challenges (No. HSS/HSF/DP. E. 10.2). Geneva: World Health Organization.
- [27] Zapata, M. (2018). Importancia del sistema GRD para alcanzar la eficiencia hospitalaria. *Revista Médica Clínica Las Condes*, 29(3), 347-352.

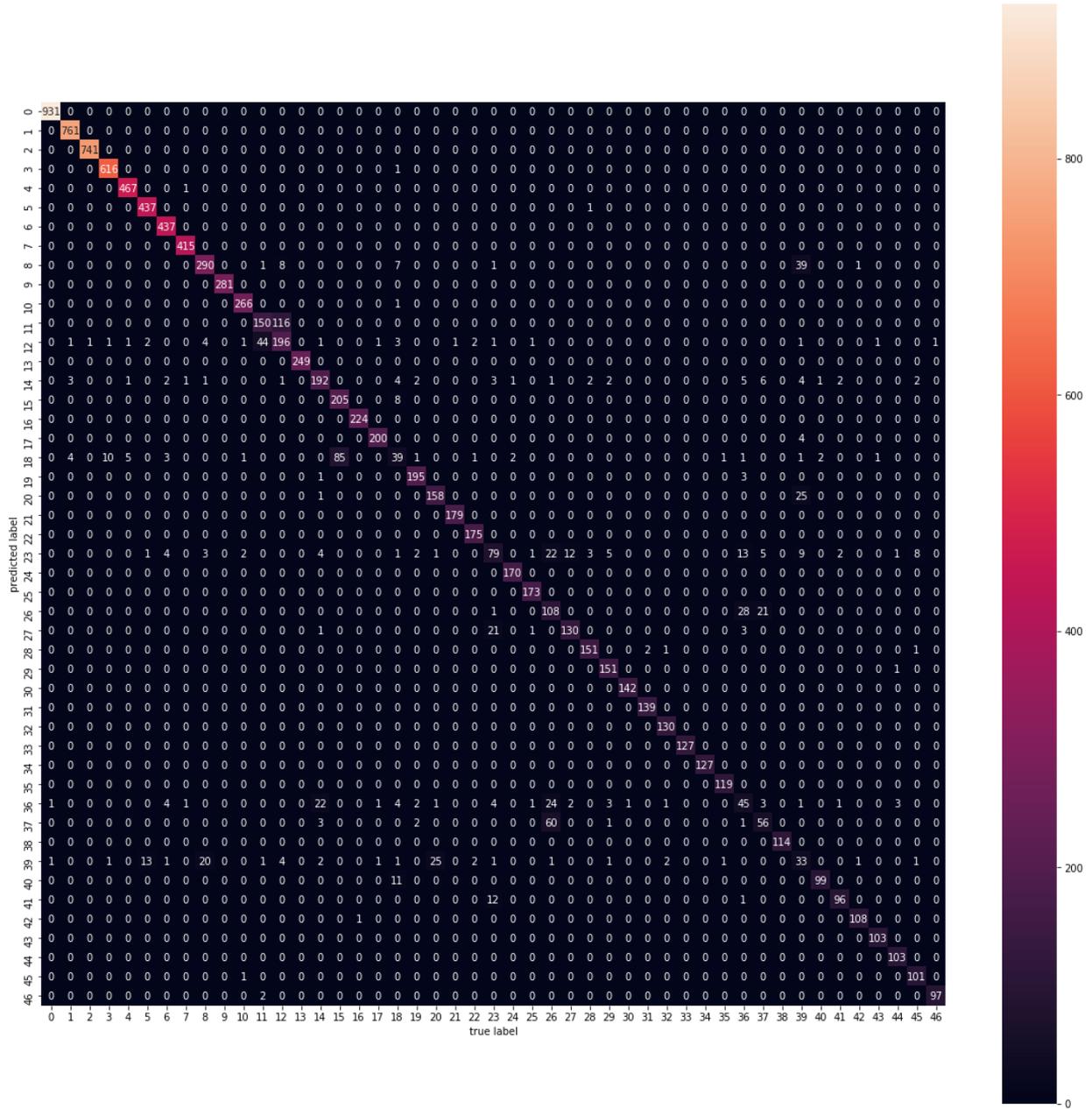
- [28] Calderón, V. G., Huante, I. A. F., Martínez, M. C., Handal, R. E. Y., Antunez, D. P., Soto, M. E., & Koretzky, S. G. (2019). The impact of improving the quality of coding in the utilities of Diagnosis Related Groups system in a private healthcare institution. 14-year experience. *International journal of medical informatics*, 129, 248-252.
- [29] Paolillo, E., Cabrera, D., Martins, L., Scasso, A., Constantin, M., Ferreira, M., ... & Álvarez, J. (2008). Grupos Relacionados por el Diagnóstico (GRD). Experiencia con IR-GRD en el Sanatorio Americano, sistema FEMI. *Revista Médica del Uruguay*, 24(4), 257-265.
- [30] Sierra, P., & Uema, S. (2014). Grupos relacionados por el diagnóstico (GRD). Análisis y desarrollo en un hospital público (Diagnosis Related Groups (DRG). Analysis and development at a public hospital). *Bitácora Digital*, 2(4).
- [31] Ayestarán, J. (4 ago. 2008). Sigesa exporta su software hospitalario a Costa Rica. *Expansión*. Edición impresa. <https://www.expansion.com/>
- [32] Organización para la Cooperación y el Desarrollo Económicos (2017). Estudios de la OCDE sobre los Sistemas de Salud. Costa Rica. Evaluación y recomendaciones. Recuperado de: <https://www.oecd.org/els/health-systems/Estudios-OCDE-sobre-los-Sistemas-de-Salud-Costa-Rica-Evaluaci%C3%B3n-y-Recomendaciones.pdf>
- [33] de Faria, L. B., & Ribeiro, W. C. (2017). Grupos de Diagnósticos Relacionados (DRG): Brasil e Colômbia. *Revista Debate Econômico*, 5(1), 83-97.
- [34] Torres Hinestroza, A., & López Orozco, G. M. (2012). Metodología de costos para instituciones prestadoras de servicios de salud: aplicación de los Grupos Relacionados por el Diagnóstico-GRD.
- [35] Gartner, D., & Padman, R. (2015). Improving Hospital-Wide Early Resource Allocation through Machine Learning. *Studies in health technology and informatics*, 216, 315-319.
- [36] Roeder, N., Rochell, B., Juhra, C., & Mueller, M. (2001). Empirical comparison of DRG variants using cardiovascular surgery data: initial results of a project at 18 German hospitals. *Australian Health Review*, 24(4), 57-80.
- [37] AVEDIAN. [avediangrd.com](http://avediangrd.com). Ciudad de Córdoba, Argentina. ¿Qué es un Agrupador de GRD y cuál es su importancia para los sistemas de salud latinoamericanos?

- <https://avediangrd.com/!/-que-es-un-agrupador-de-grd-y-cual-es-su-importancia-para-los-sistemas-de-salud-latinoamericanos/>
- [38] Tomar D, Agarwal S. A survey on data mining approaches for healthcare. *Int J Biosci Biotechnol* 2013; 5(5):241-66.
- [39] Moon M, Lee SK. Applying of decision tree analysis to risk factors associated with pressure ulcers in long-term care facilities. *Healthc Inform Res* 2017;23(1):43-52.
- [40] Yahya N, Ebert MA, Bulsara M, House MJ, Kennedy A, Joseph DJ, et al. Statistical-learning strategies generate only modestly performing predictive models for urinary symptoms following external beam radiotherapy of the prostate: a comparison of conventional and machine learning methods. *Med Phys* 2016;43(5):2040-52.
- [41] Kuo, C. Y., Yu, L. C., Chen, H. C., & Chan, C. L. (2018). Comparison of models for the prediction of medical costs of spinal fusion in Taiwan Diagnosis-Related Groups by machine learning algorithms. *Healthcare informatics research*, 24(1), 29-37.
- [42] Viangteeravat, T. (2013). Potential identification of pediatric asthma patients within pediatric research database using low rank matrix decomposition. *Journal of clinical bioinformatics*, 3(1), 16.
- [43] Okamoto, K., Uchiyama, T., Takemura, T., Kume, N., Kuroda, T., & Yoshihara, H. (2018). Automatic Selection of Diagnosis Procedure Combination Codes Based on Partial Treatment Data Relative to the Number of Hospitalization Days. *European Journal of Biomedical Informatics*, 14(1).
- [44] World Health Organization. (2004). *International statistical classification of diseases and related health problems (Vol. 2)*. World Health Organization.
- [45] World Health Organization. (2004). *International statistical classification of diseases and related health problems (Vol. 1)*. World Health Organization.
- [46] Rana, A., Rawat, A. S., Bijalwan, A., & Bahuguna, H. (2018, August). Application of multi layer (Perceptron) artificial neural network in the diagnosis system: a systematic review. In *2018 International conference on research in intelligent and computing in engineering (RICE)* (pp. 1-6). IEEE.
- [47] Zhu, X. J. (2005). *Semi-supervised learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences.

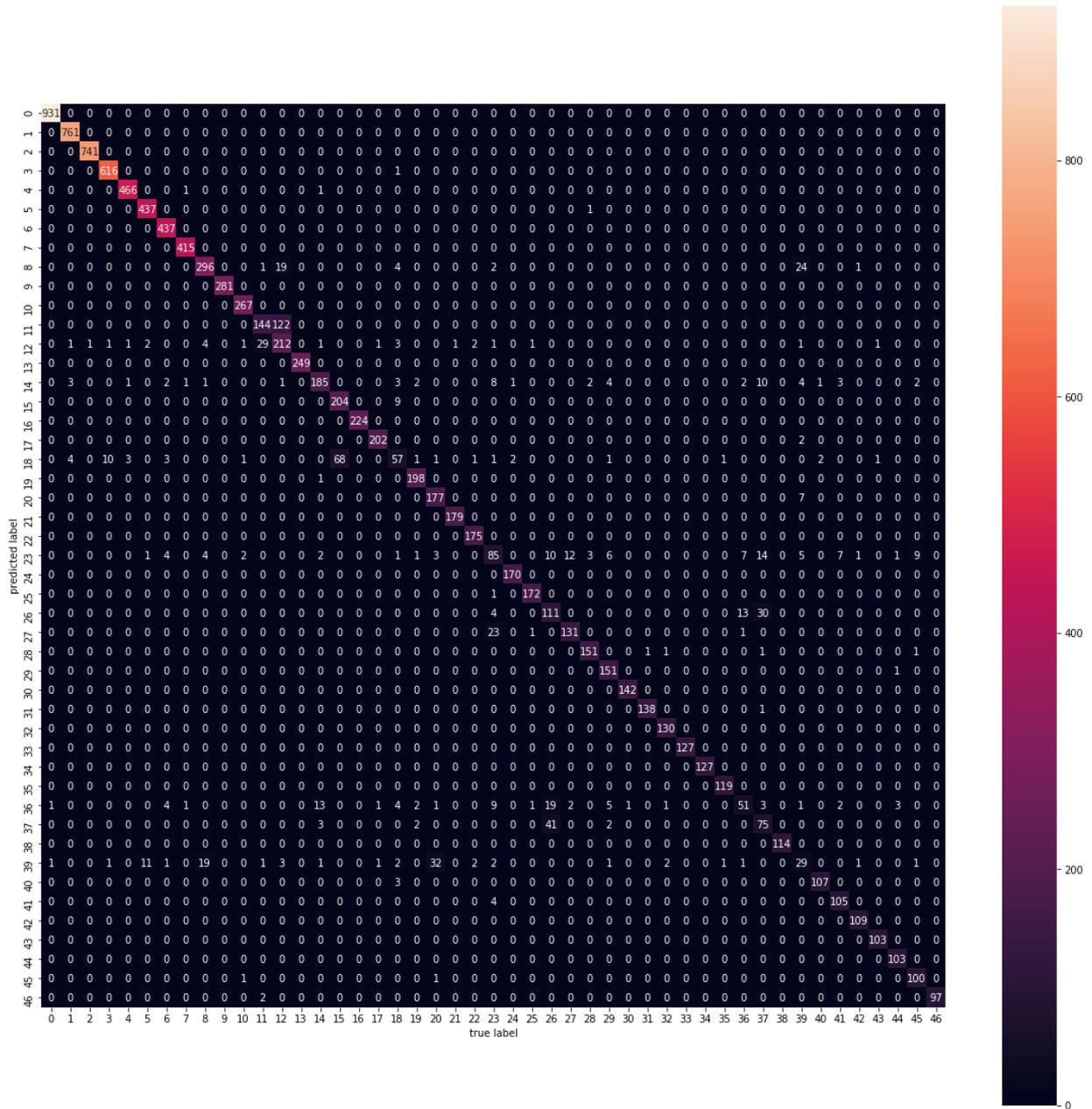
- [48] Khosravanian, A., & Ayat, S. (2016). Diagnosing breast cancer type by using probabilistic neural network in decision support system. *International Journal of Knowledge Engineering*, 2(1).
- [49] Baby, M. S., Saji, A. J., & Kumar, C. S. (2017, April). Parkinsons disease classification using wavelet transform based feature extraction of gait data. In *2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT)* (pp. 1-6). IEEE.
- [50] Garmendia, A., Rios, S. A., Lopez-Guede, J. M., & Graña, M. (2019). Triage prediction in pediatric patients with respiratory problems. *Neurocomputing*, 326, 161-167.
- [51] Ma, B., Meng, F., Yan, G., Yan, H., Chai, B., & Song, F. (2020). Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. *Computers in Biology and Medicine*, 103761.
- [52] Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies*, 10(8), 1168.
- [53] NEW YORK STATE – Department of Health. health.ny.gov. New York, EE.UU. APR-DRG Weights and ALOS for Acute Services. <https://www.health.ny.gov/facilities/hospital/reimbursement/apr-drg/weights/>
- [54] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111-133.
- [55] T. Gentimis, A. Ala’J, A. Durante, K. Cook, and R. Steele, “Predicting hospital length of stay using neural networks on mimic iii data,” in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. IEEE, 2017, pp. 1194–1201
- [56] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016



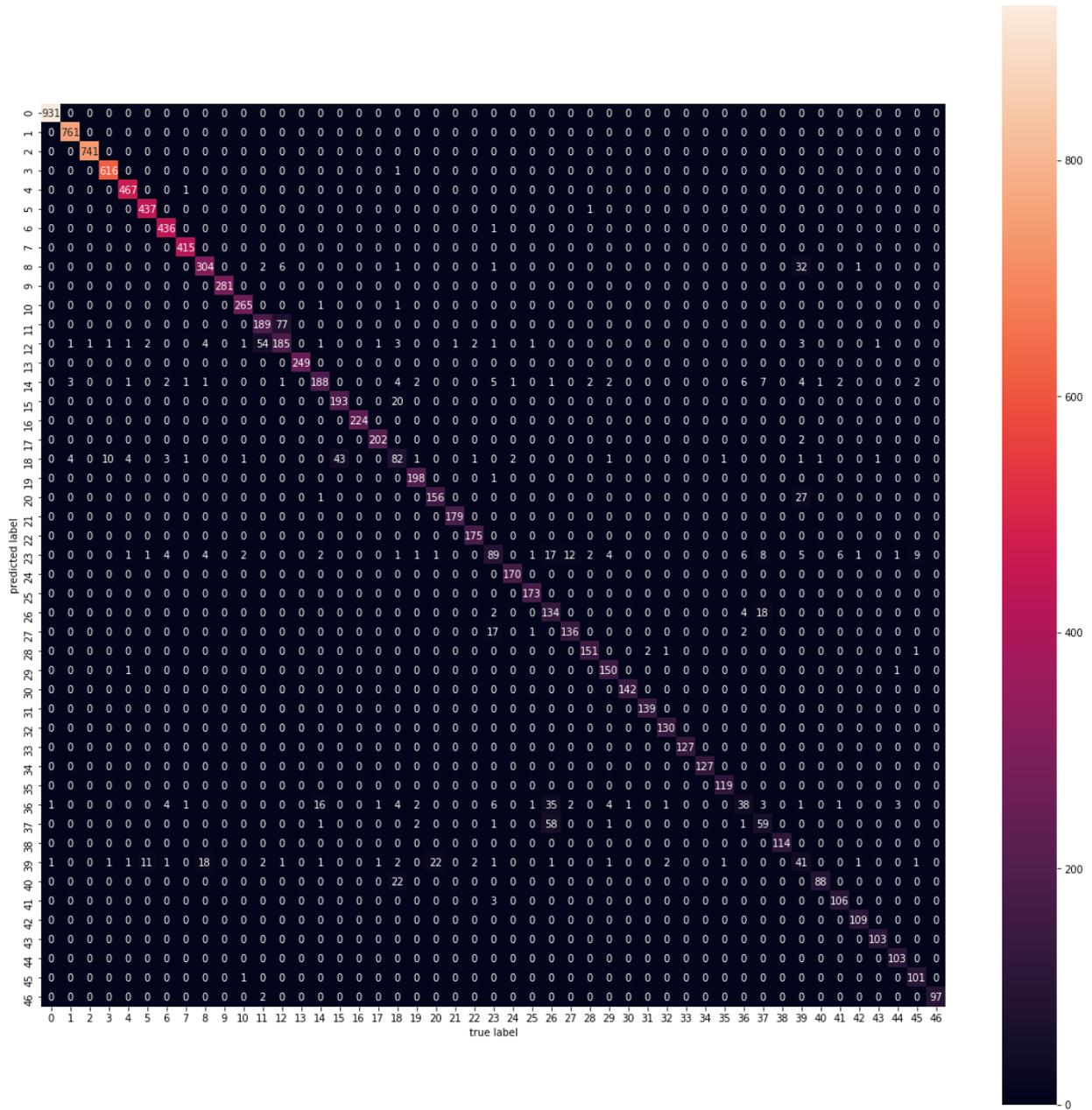
A.2 Modelo M24



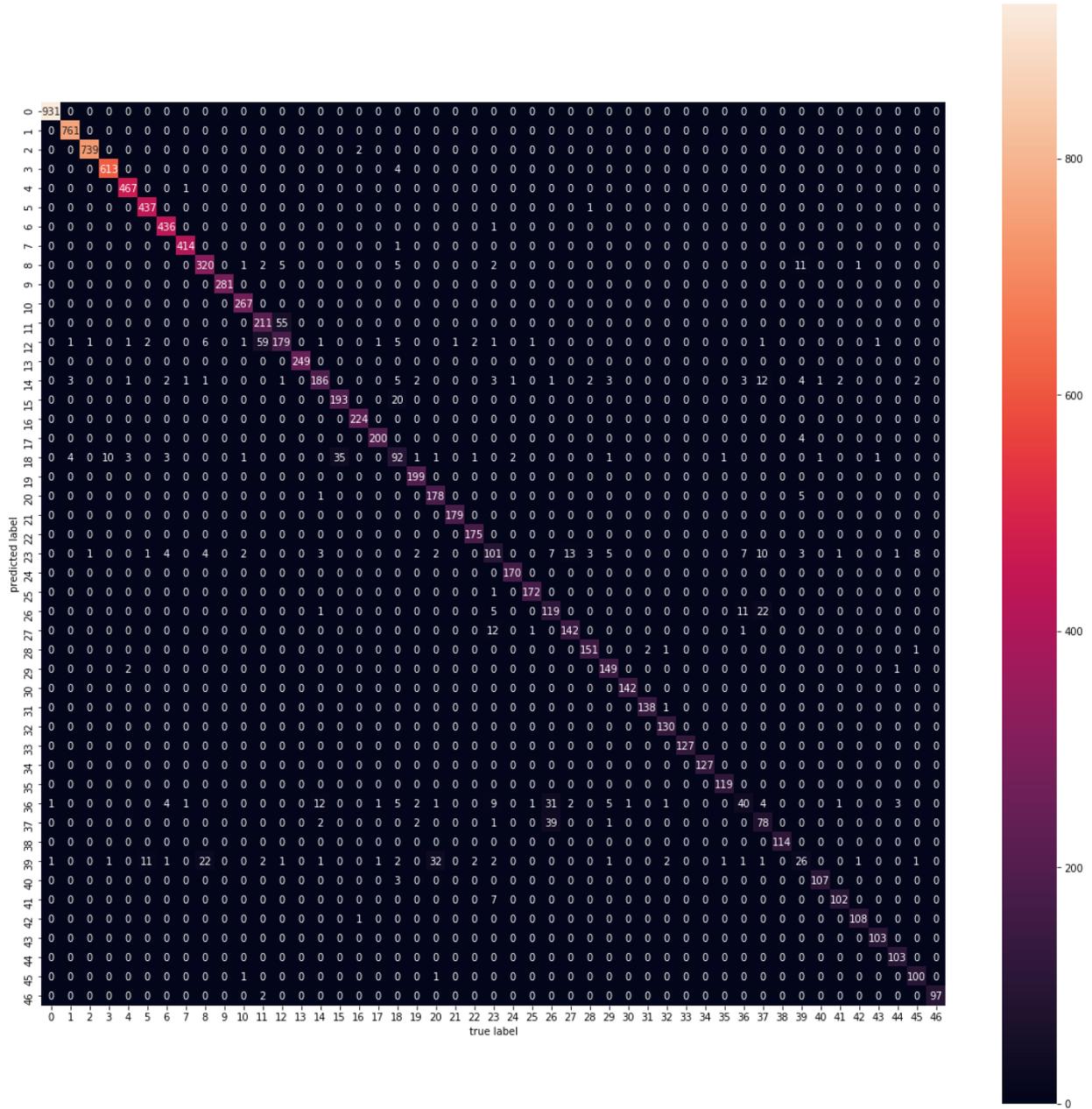
A.3 Modelo M48



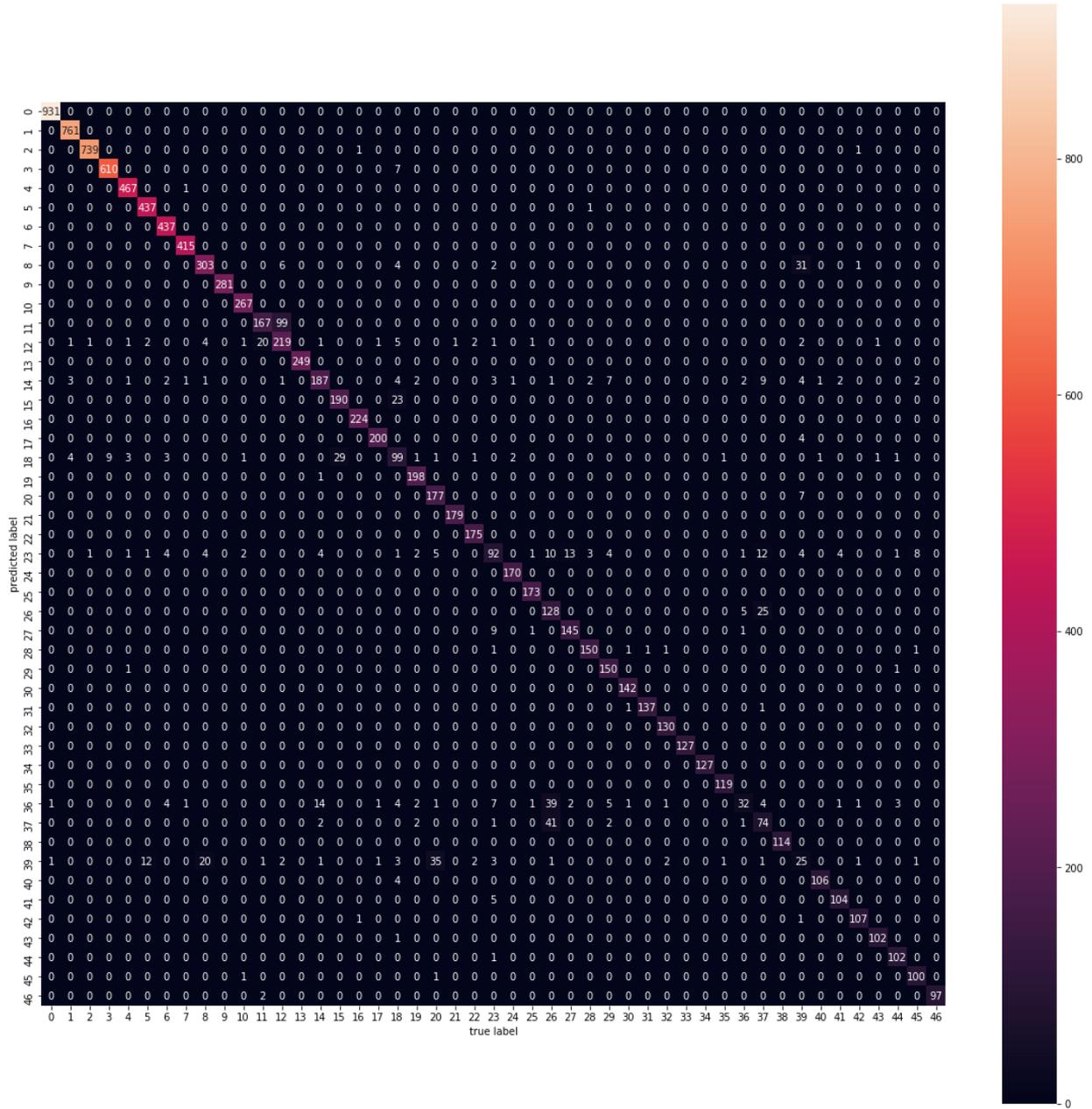
A.4 Modelo M72



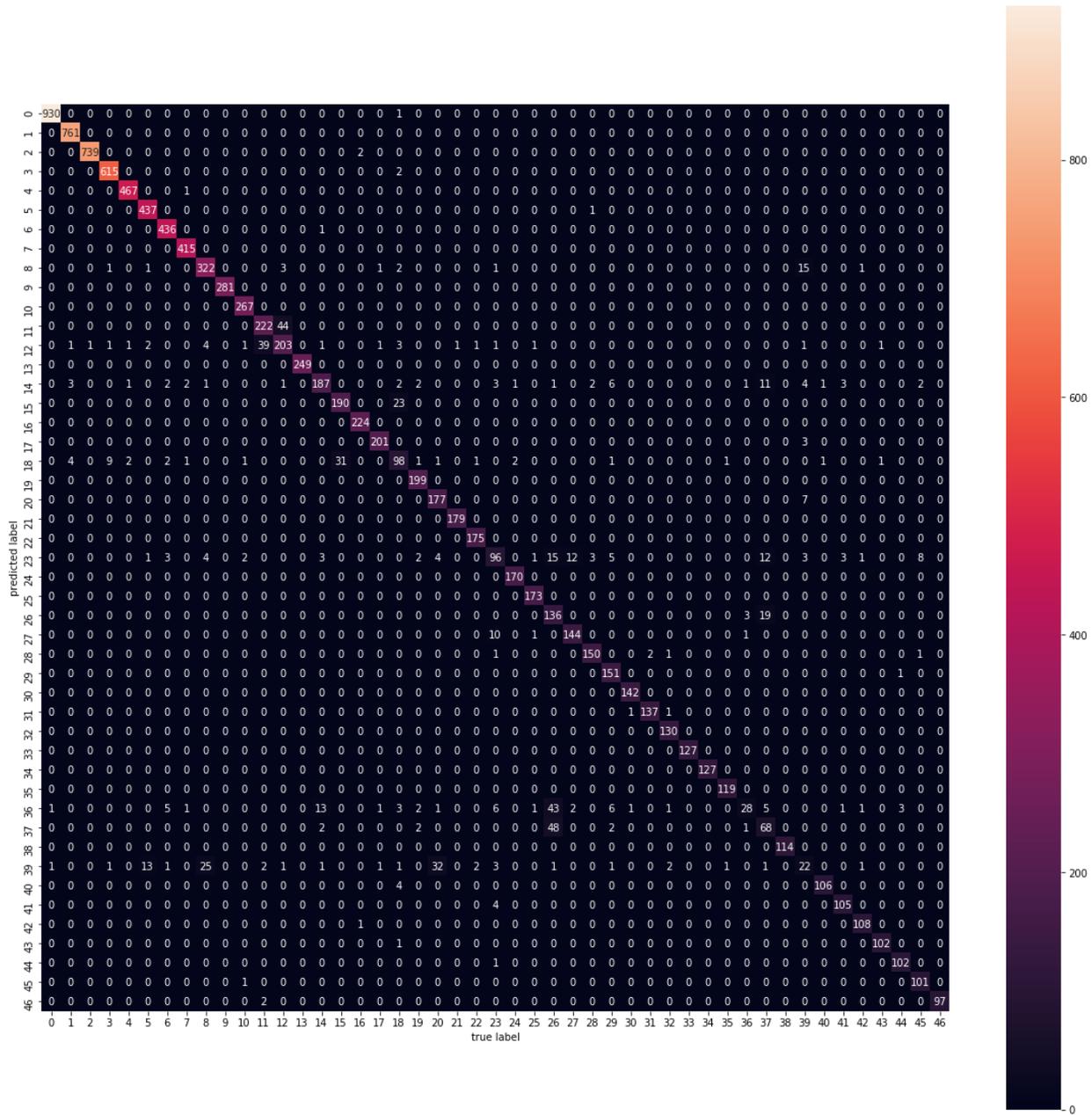
A.5 Modelo M96



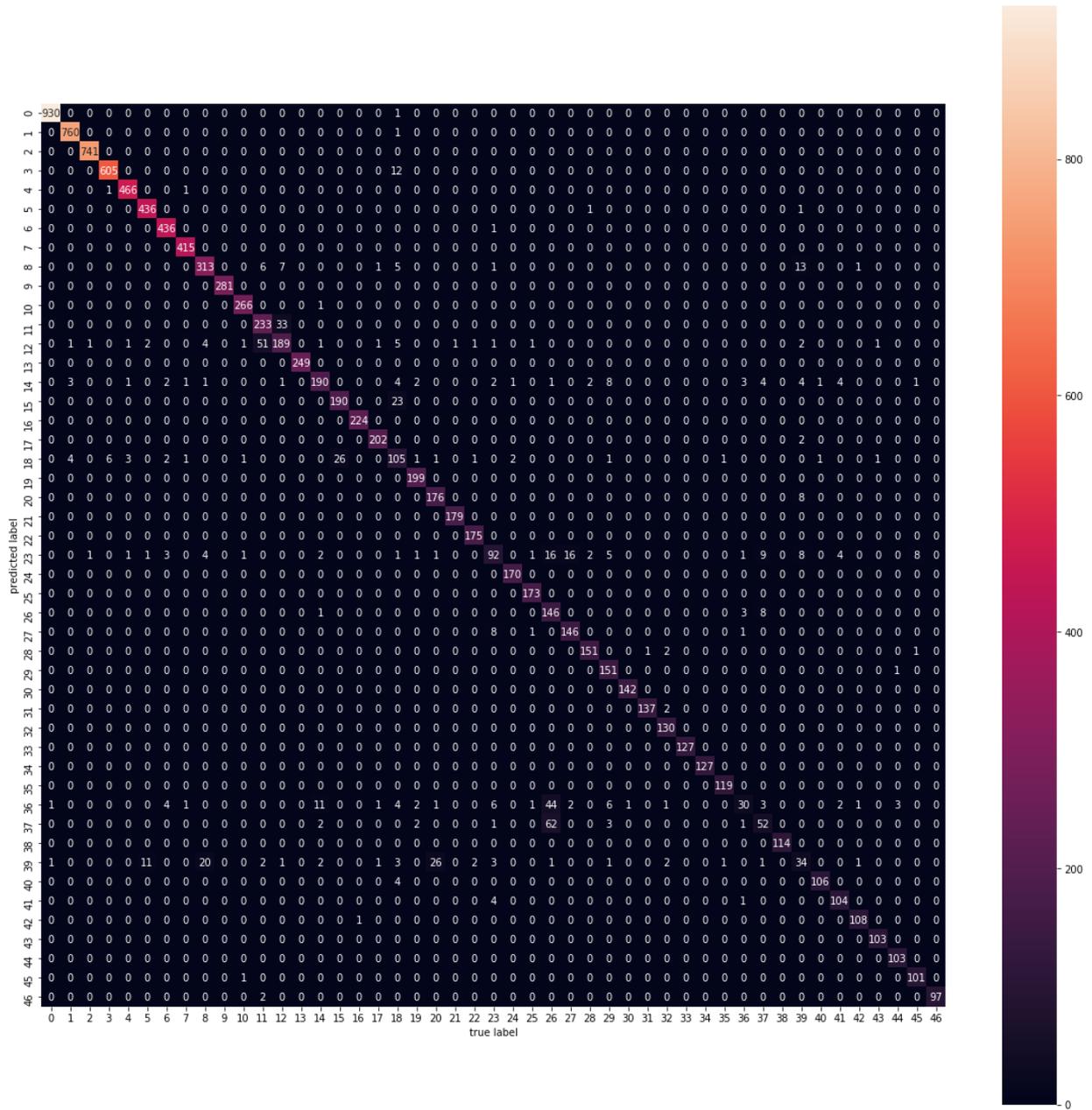
A.6 Modelo M120



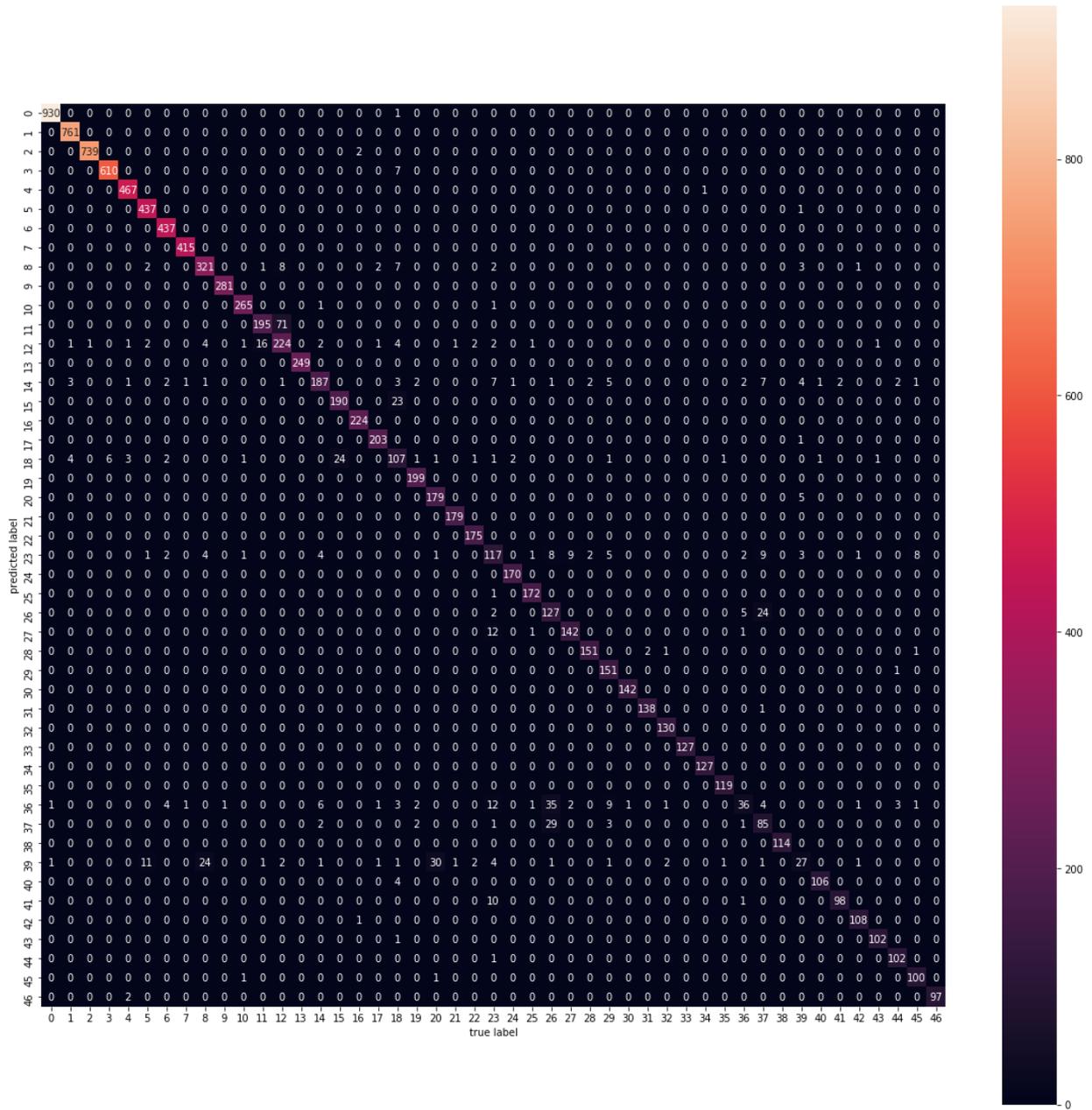
A.7 Modelo M144



A.8 Modelo M168



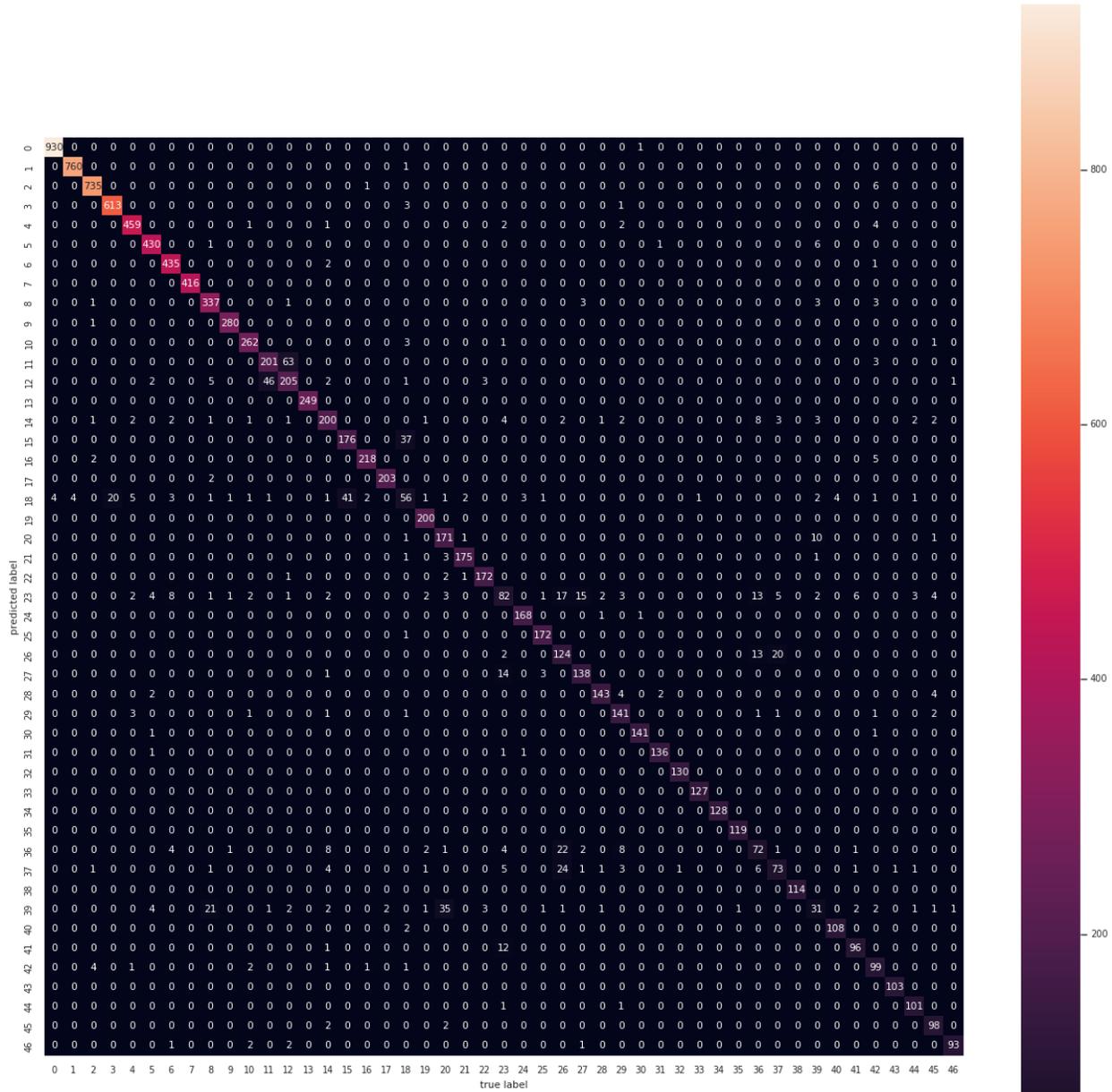
A.9 Modelo M192



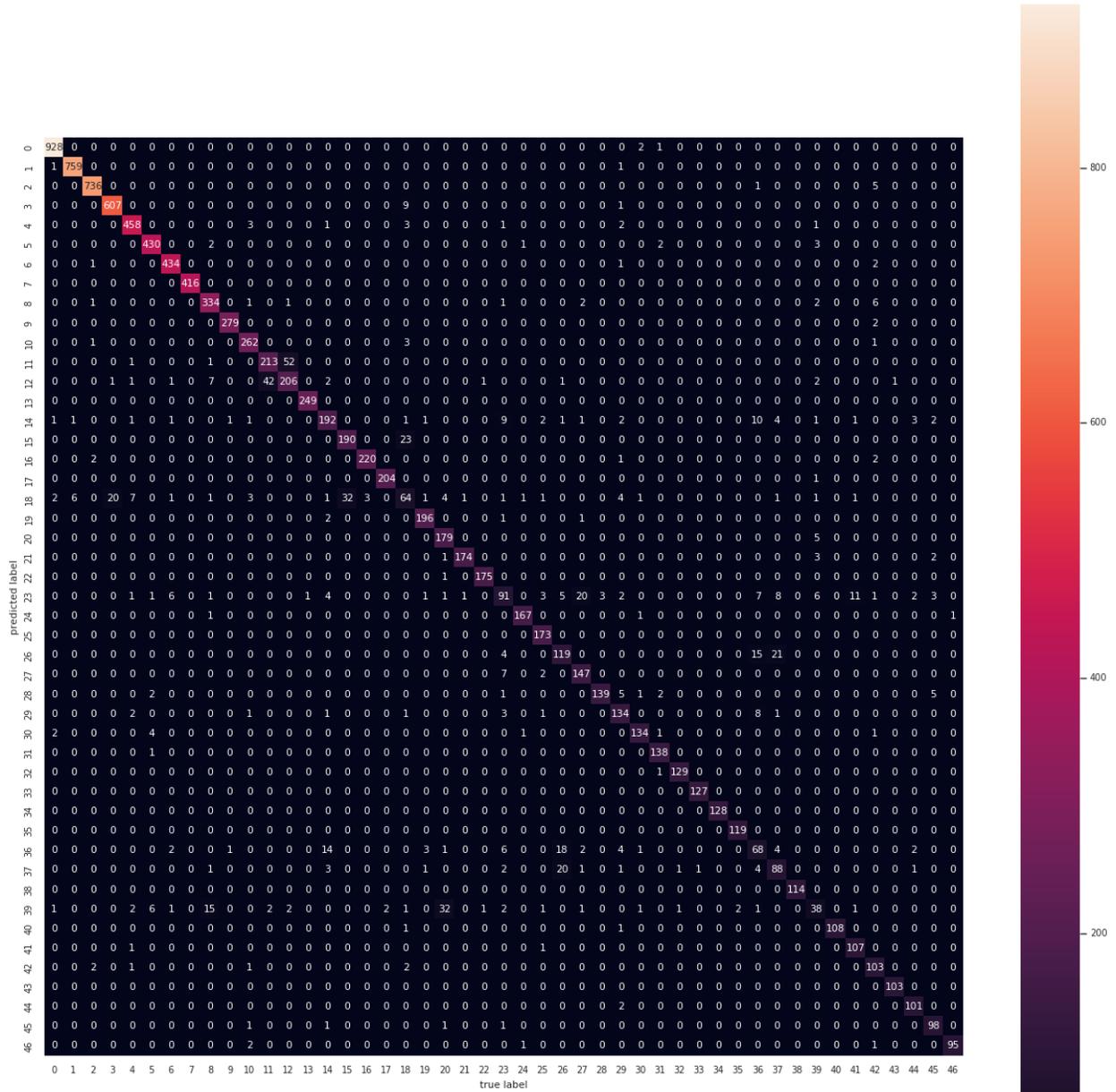




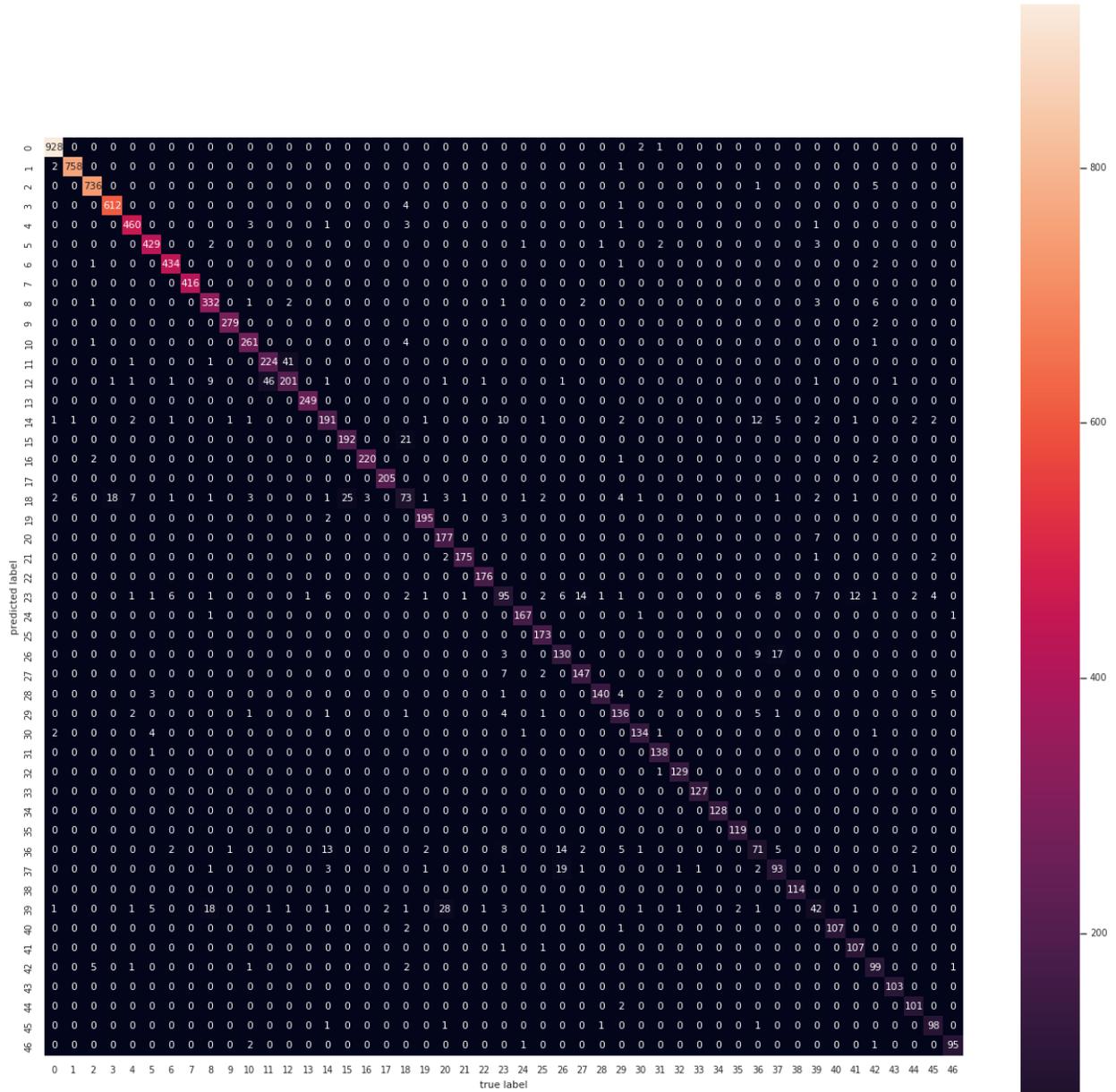
B.2 Modelo M24



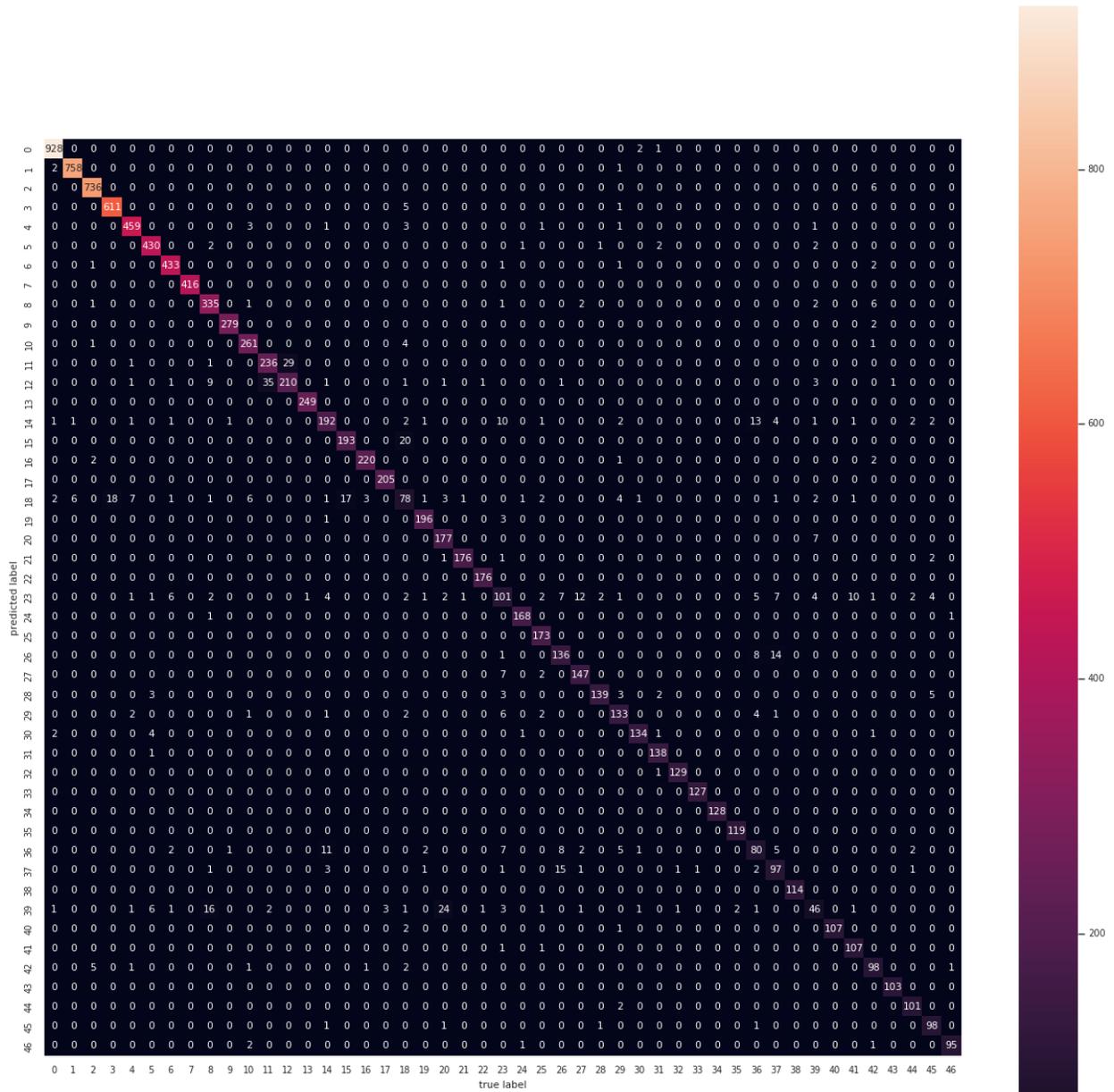
B.3 Modelo M48



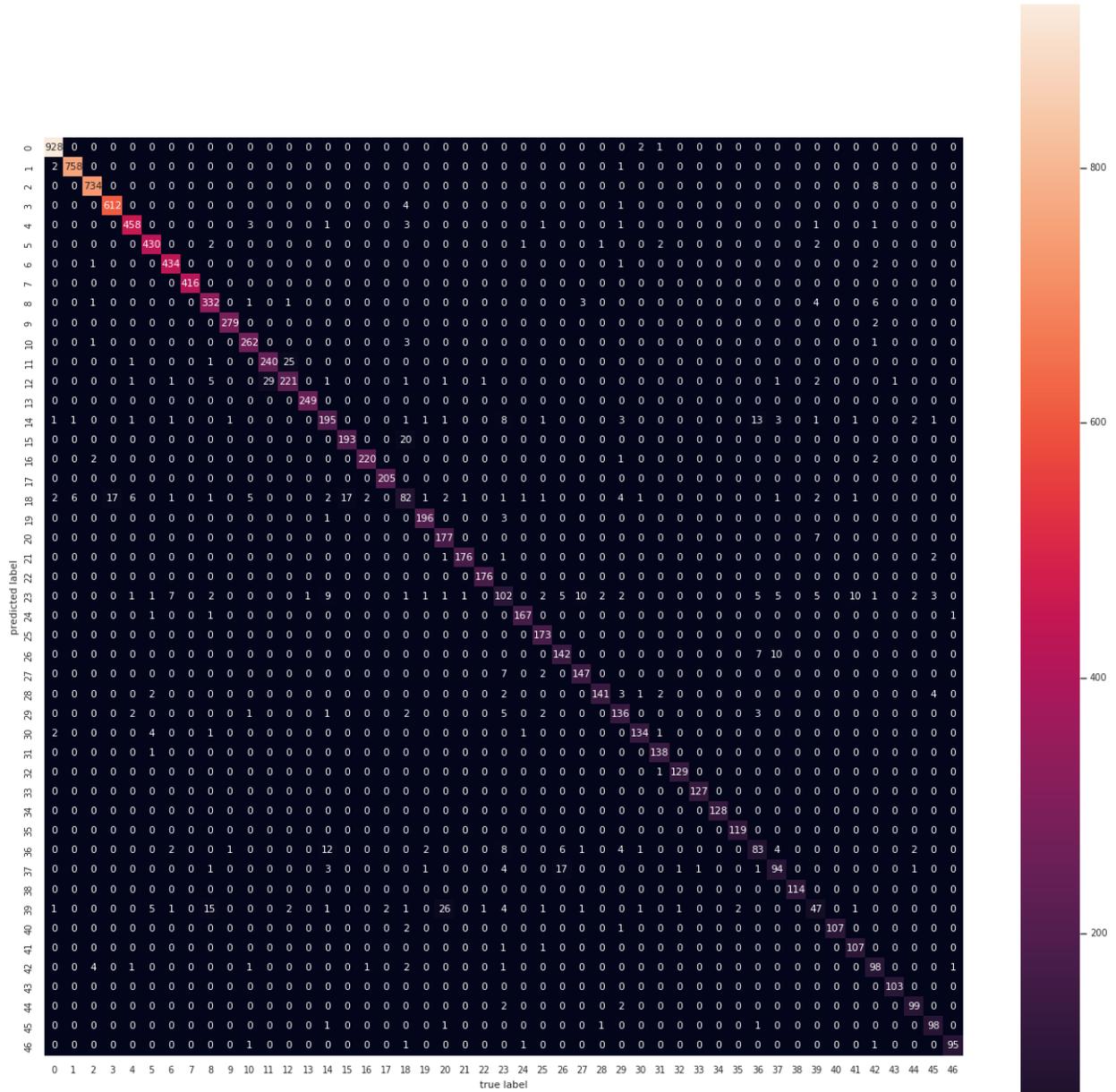
B.4 Modelo M72



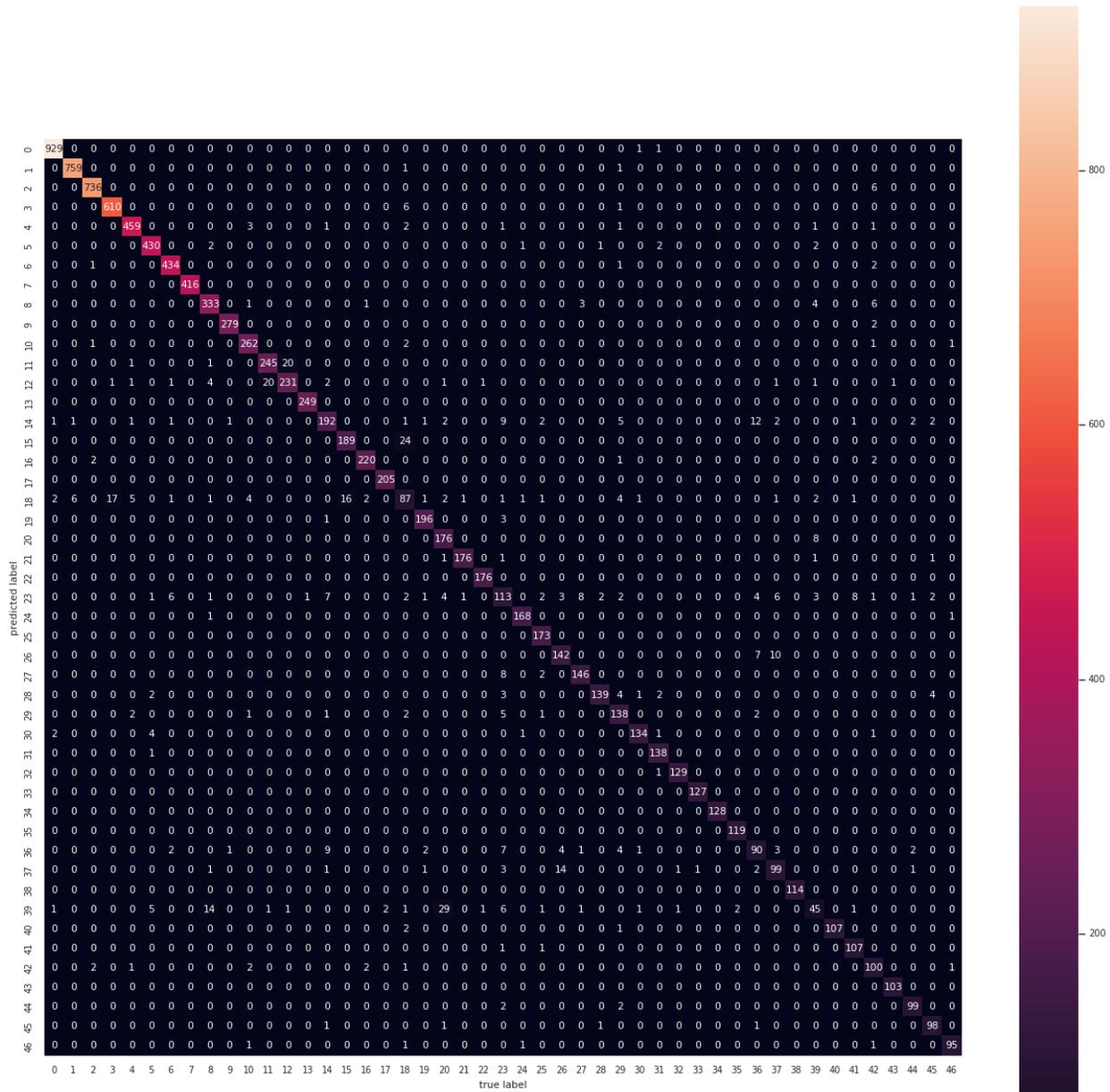
B.5 Modelo M96



B.6 Modelo M120

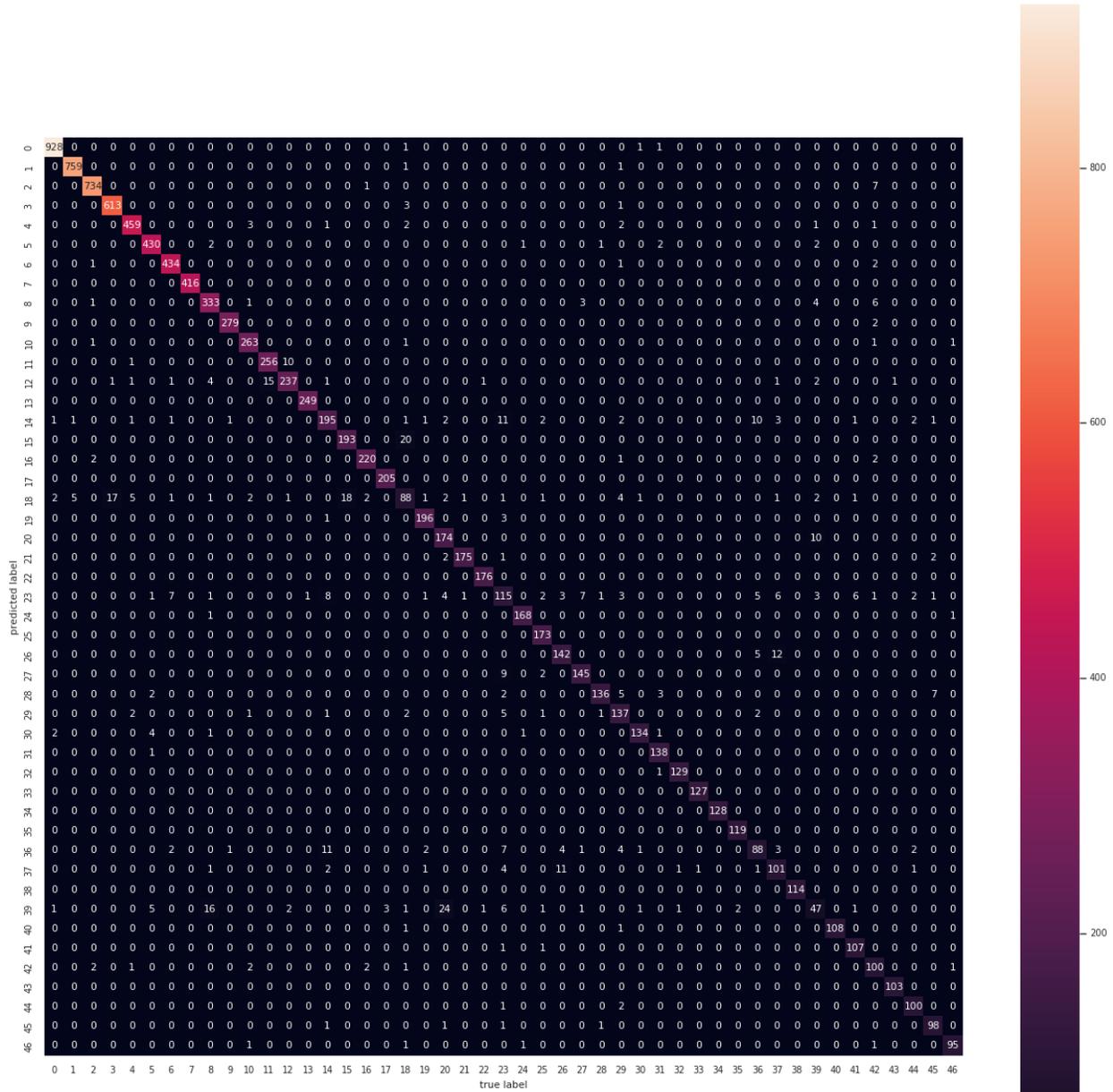


B.7 Modelo M144





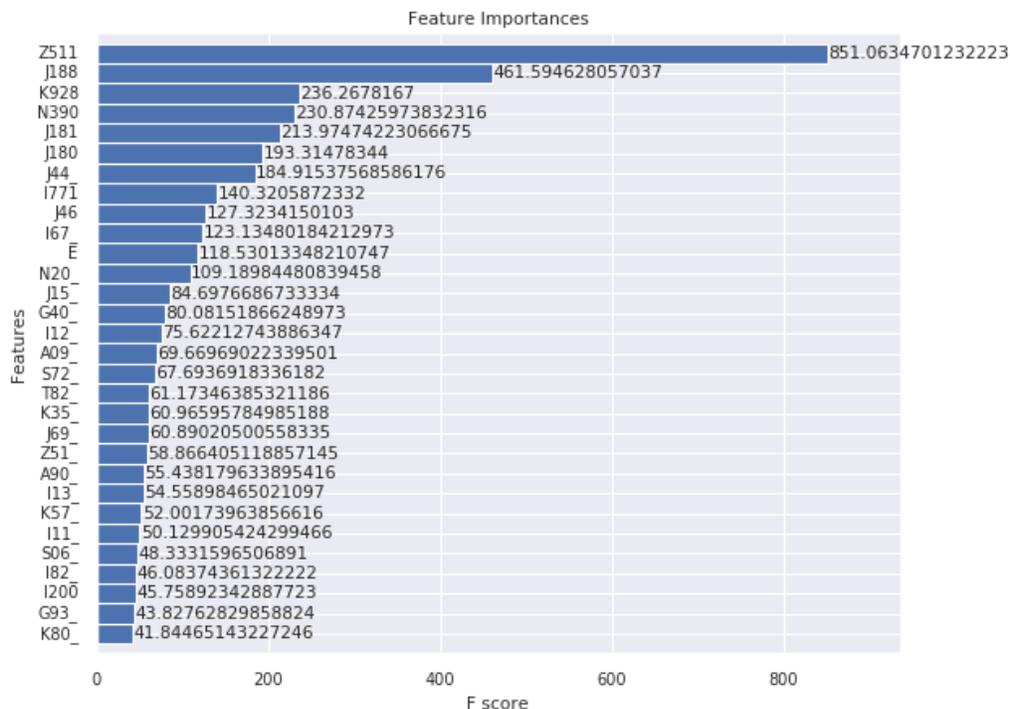
B.9 Modelo M192



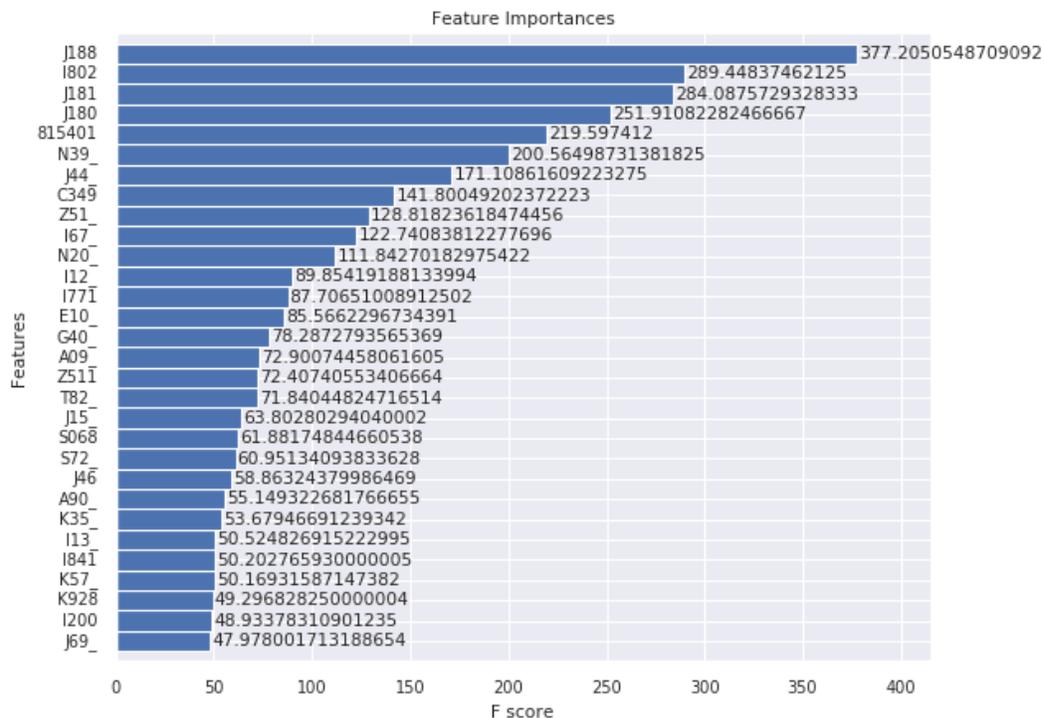


## ANEXO C. Importancia de características en XGBoost

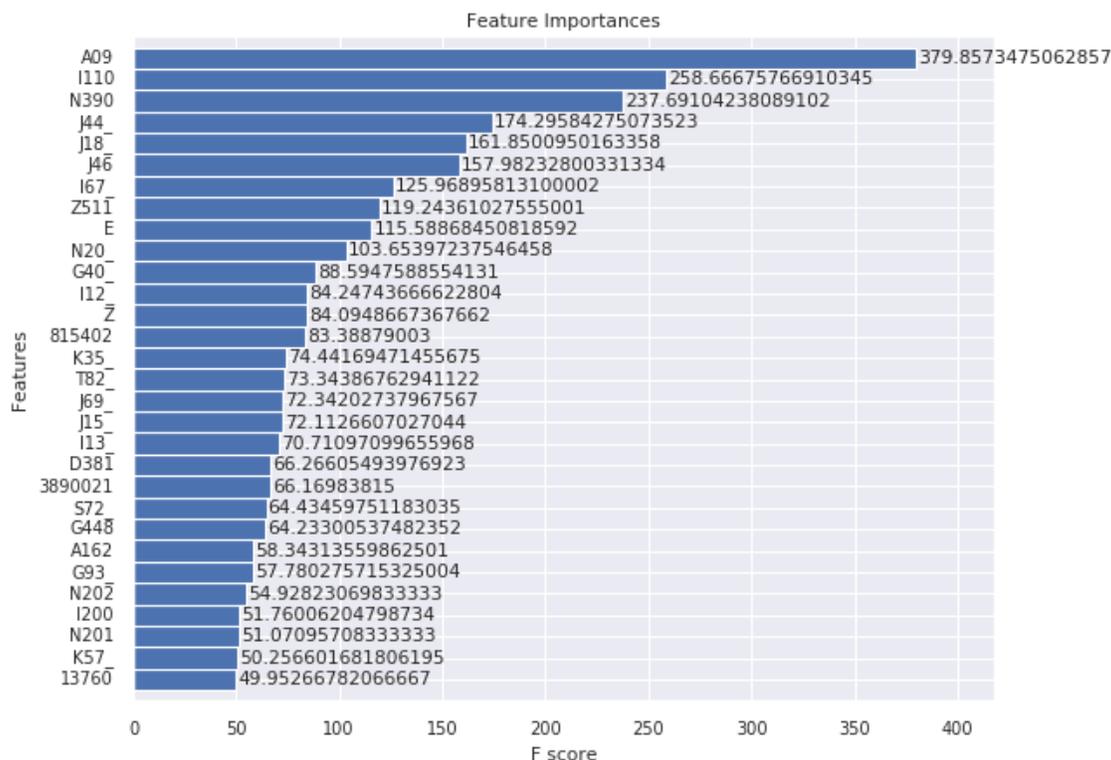
### C.1 Modelo M0



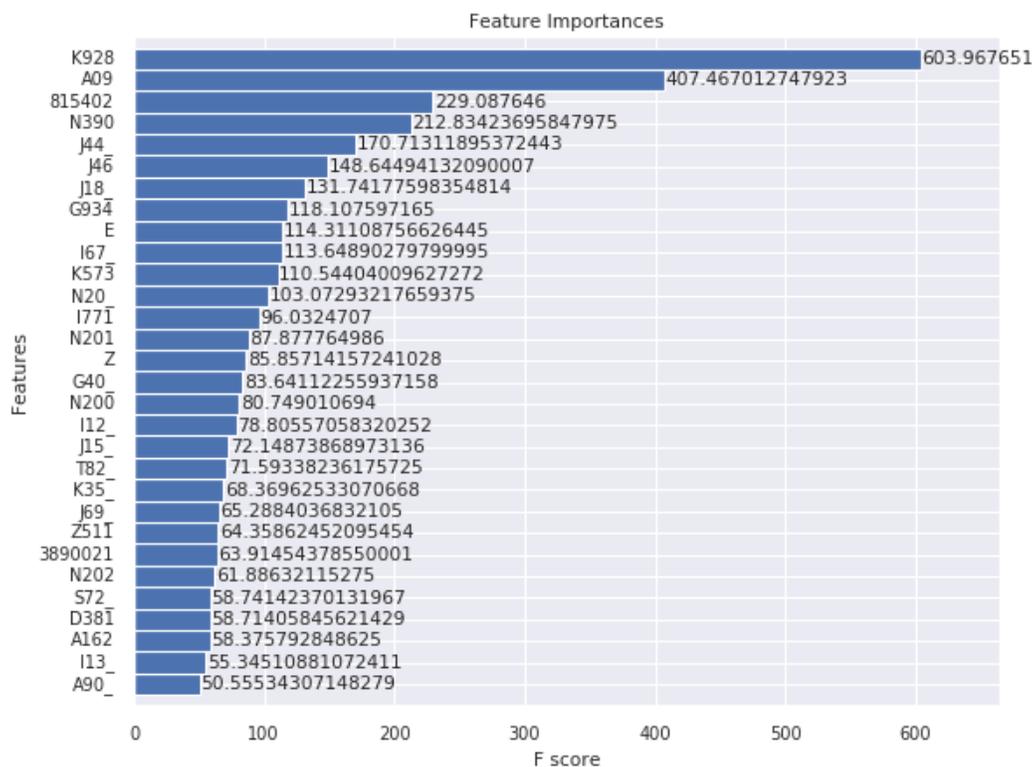
### C.2 Modelo M24



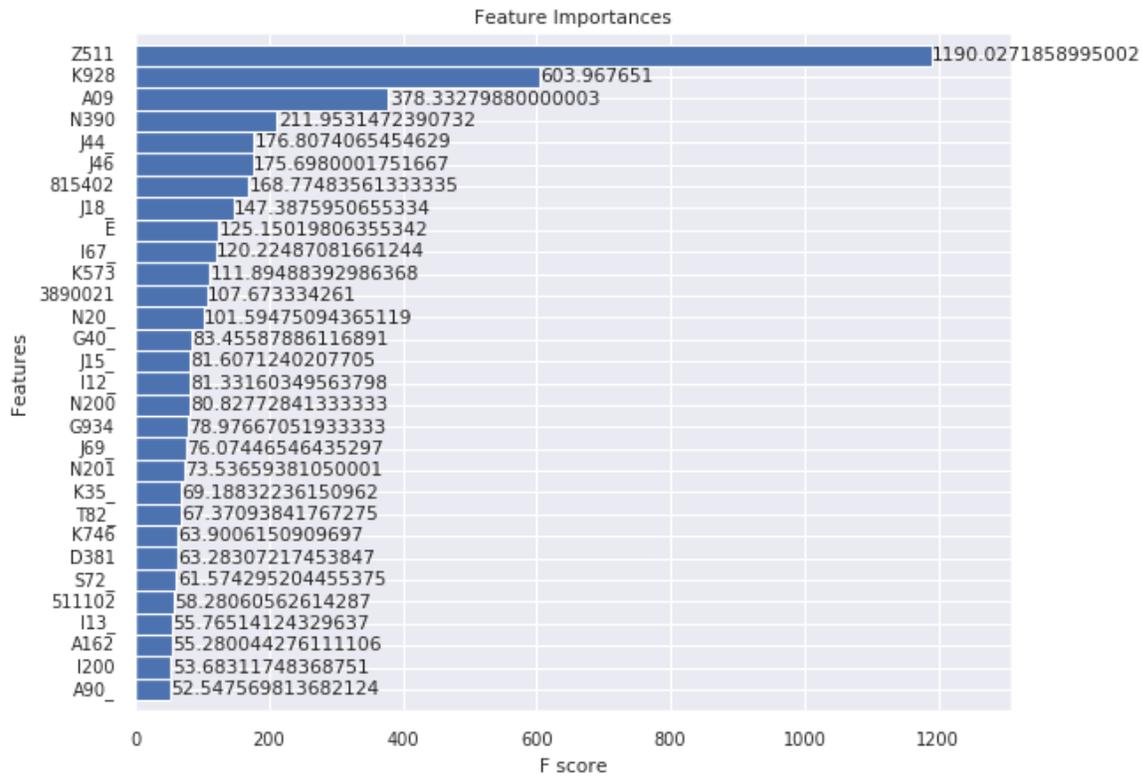
## C.3 Modelo M48



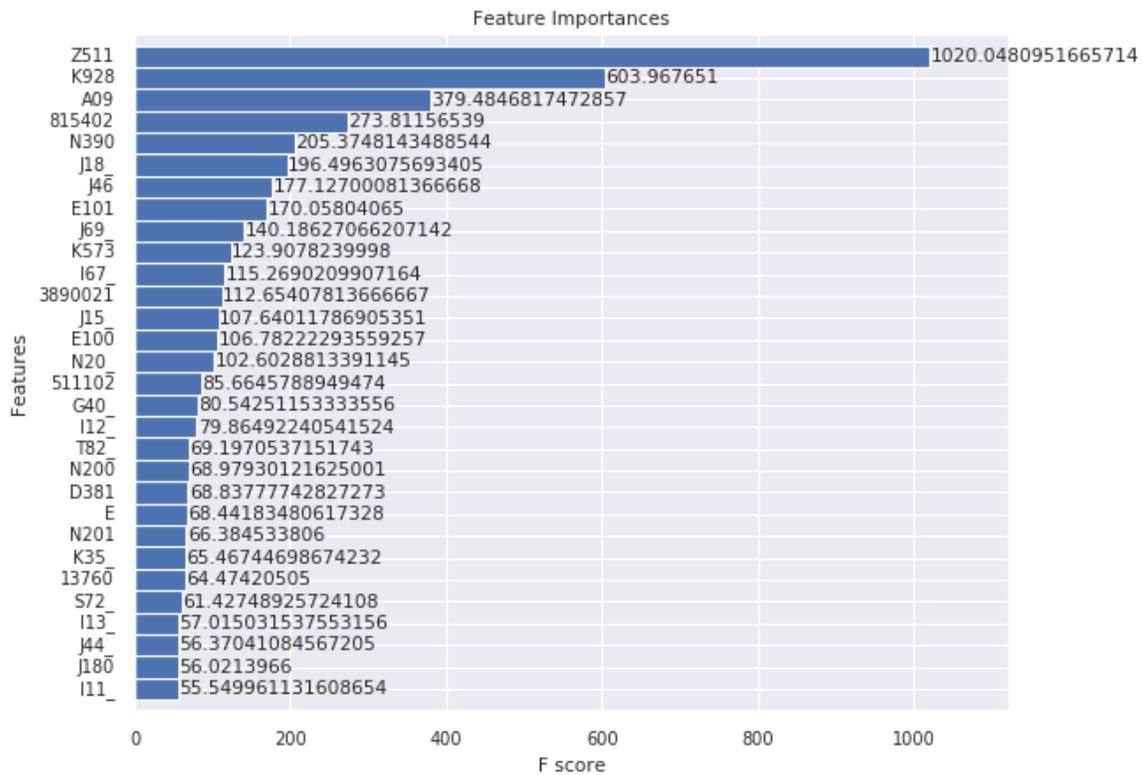
## C.4 Modelo M72



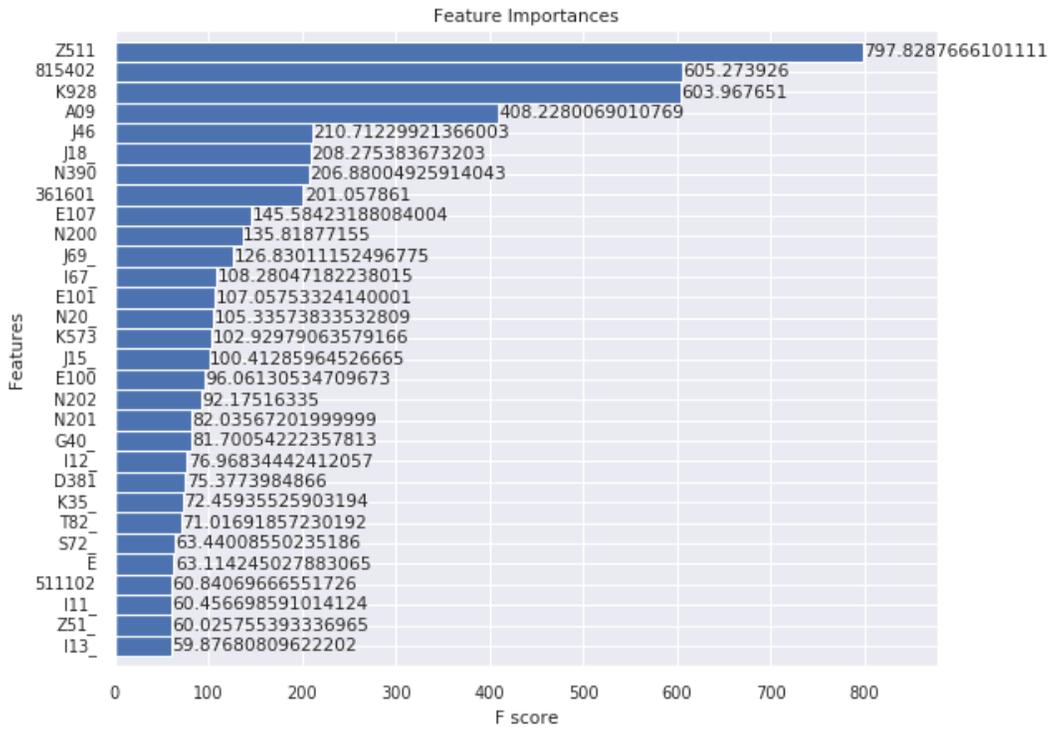
C.5 Modelo M96



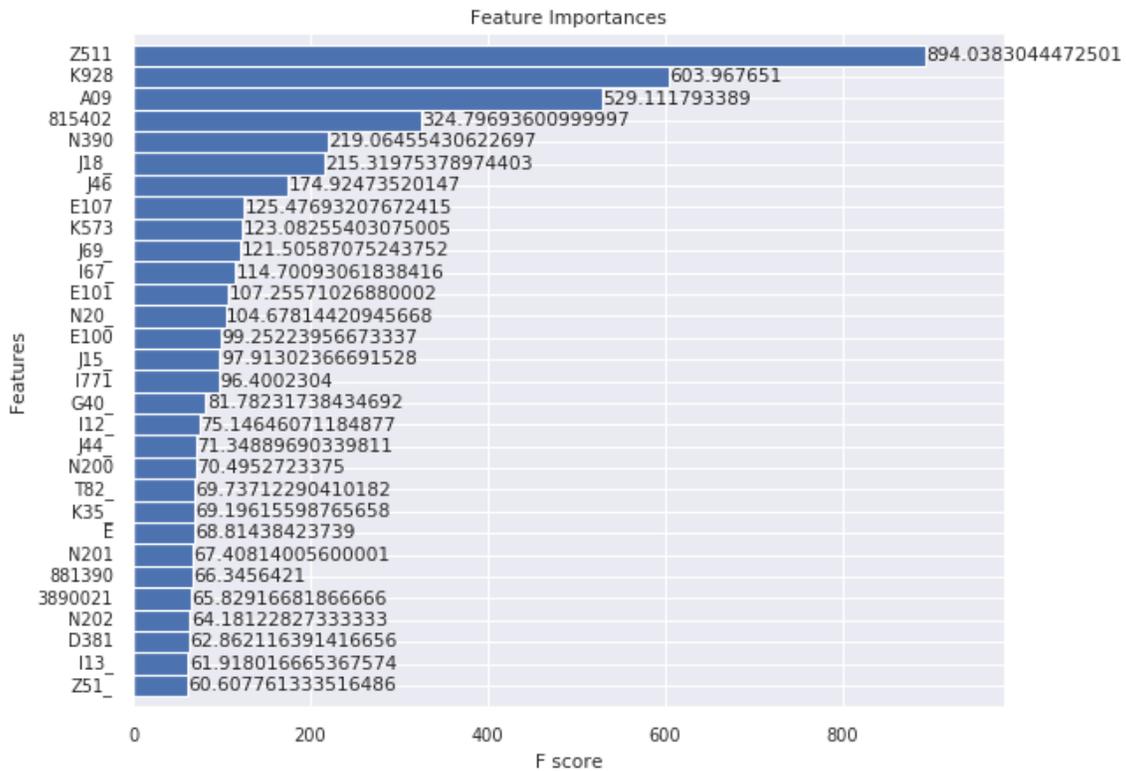
C.6 Modelo M120



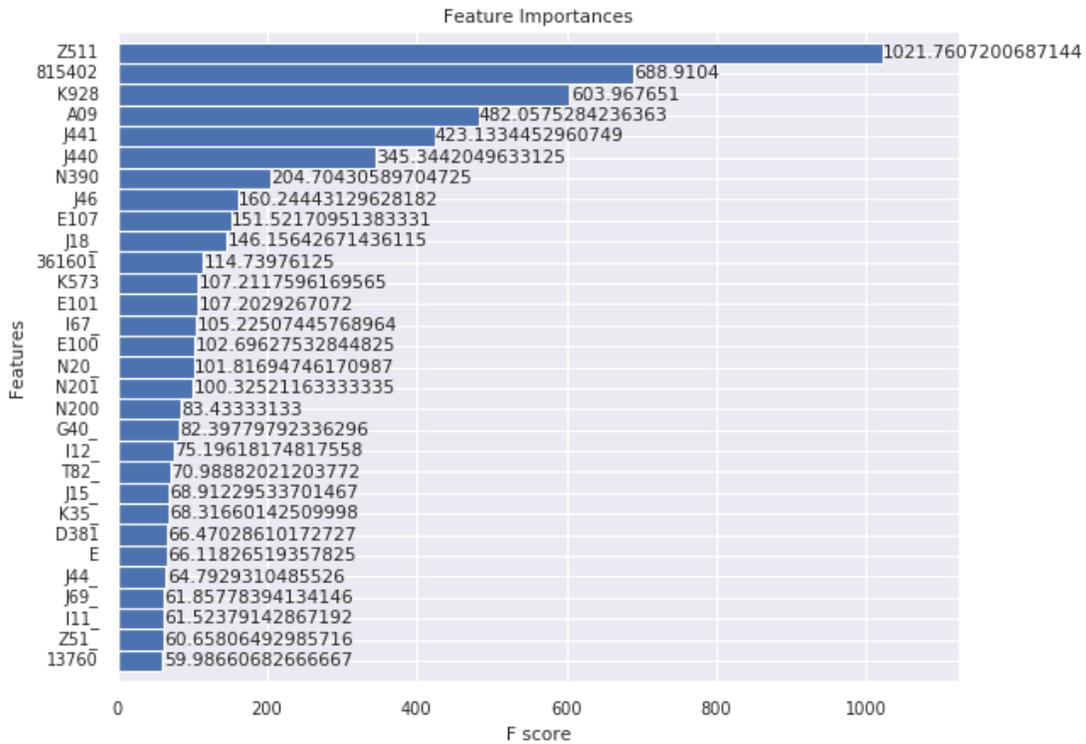
C.7 Modelo M144



C.8 Modelo M168



C.9 Modelo M192



C.10 Modelo Baseline - MBL

