

Un Primer Abordaje para la Predicción de la Enfermedad de Injerto contra Huésped en Trasplante de Médula Ósea en Pacientes Pediátricos*

Lázaro A. Gibert^{1,2}, Rodrigo Guerra³, Jocelyn Dunstan⁴, Julia Palma⁵,
Axel J. Soto^{1,2}, Ana Maguitman^{1,2}, Carlos Chesñever^{1,2}

¹ Instituto de Ciencias e Ingeniería de la Computación, CONICET-UNS, Argentina

² Dept. Cs. e Ing. de la Computación, Universidad Nacional del Sur, Argentina

³ FCFM, Universidad de Chile, Chile

⁴ DCC & IMC, Pontificia Universidad Católica de Chile, Chile

⁵ Hospital Luis Calvo Mackenna, Santiago, Chile

lazarogarcia@cs.uns.edu.ar, rodrigoguerra@uchile.cl,
jdunstan@uc.cl, jpalma@calvomackenna.cl,

axel.soto@cs.uns.edu.ar, agm@cs.uns.edu.ar, cic@cs.uns.edu.ar

Resumen: La enfermedad de injerto contra huésped (EICH) es una complicación común en un paciente, luego de ser sometido a un trasplante de médula ósea y puede ser mortal. En la actualidad los especialistas no tienen manera de predecir de forma efectiva si el paciente puede desarrollar la enfermedad. A partir de esta problemática, en este trabajo se propone obtener distintos modelos de Aprendizaje Automático con el objetivo de predecir, a partir de los datos a los que tiene acceso el especialista, si el paciente va a desarrollar EICH. Como punto de partida se utilizarán dos datasets representativos: por un lado, un conjunto de 187 ejemplos con 37 atributos correspondiente a datos públicos, y por otro, un conjunto de mayor magnitud proporcionado por el Hospital Luis Calvo Mackenna consistente de 536 ejemplos con 227 atributos (seleccionándose un subconjunto de éstos últimos). Nuestra investigación preliminar a partir de los datasets anteriores muestra que se pueden generar distintos modelos con buena capacidad predictiva. En particular utilizamos Árbol de Decisión, Bosque Aleatorio, XG-Boost, Regresión Logística, Perceptrón Multicapa, AdaBoost y Potenciación de Gradiente, comparando su desempeño utilizando distintas métricas.

Palabras claves: Aprendizaje Automático · Trasplante de médula ósea · Enfermedad de injerto contra huésped.

1. Introducción

La enfermedad de injerto contra huésped (EICH, o también GVHD por las siglas en inglés de *Graft-Versus-Host Disease*) es un padecimiento autoinmune en el cual los linfocitos T del donador desempeñan un papel central en el ataque inmunológico a los

* El presente trabajo es financiado por CONICET (PUE 22920160100056CO, PIBAA 2872021010 1236CO), MinCyT (PICT PRH-2017-0007, PICT 2019-3944), UNS (PGI 24/N051), FONDECYT 11201250, ANID-Chile Basal Funds for Center of Excellence FB210005 (CMM), Millennium Science Initiative Program ICN2021.004 (iHealth) e ICN17.002 (IMFD) y STIC-AmSud.

tejidos del huésped, y constituye una complicación frecuente luego de un trasplante de médula ósea. Las formas aguda y crónica son diferentes desde el punto de vista clínico e inmunológico. La EICH representa una alteración inmunológica con un patrón de producción de citocinas, principalmente de tipo TH1, es de corta duración y afecta únicamente a los epitelios. Su frecuencia oscila desde un 6 % a un 90 % entre los receptores de injertos. Su grado de afectación se relaciona con los factores de riesgo propios de cada paciente, y ocasiona gran morbilidad y mortalidad si no se diagnostica y trata a tiempo. Los pacientes que padecen EICH refieren frecuentemente a un prurito o sensación de dolor cutáneo, en cuyo caso se observa un exantema maculopapular que a veces puede progresar a lesiones ampollosas e incluso a necrosis epidérmica [6].

El objetivo de este trabajo es aplicar una serie de modelos de Aprendizaje Automático (en inglés *Machine Learning*) a dos datasets, uno disponible públicamente y otro proporcionado por el Hospital Luis Calvo Mackenna ubicado en Santiago de Chile, que contienen datos sobre trasplantes de médula ósea en niños para lograr predecir si un paciente desarrollará EICH o no.

2. Antecedentes

El desarrollo de técnicas de Aprendizaje Automático para apoyar a los especialistas médicos en su labor ha sido objeto de atención entre varios investigadores en esta área de la inteligencia artificial. Entre los estudios se propone la estrategia de inducción de reglas para predecir si un paciente sobrevivirá al trasplante de médula ósea y los factores que hacen posibles estos resultados [16]. Además de los métodos de inducción de reglas para la predicción de supervivencia, también se han utilizado métodos basados en árboles de decisión que se han comportado de manera satisfactoria [9]. El área de interés no se limita a la predicción de supervivencia de los pacientes, existen también estudios para predecir complicaciones específicas al trasplante de médula ósea como lo es la EICH [10]. Para la predicción de la EICH se han utilizado varias fuentes de datos, desde Registros electrónicos de salud [17] hasta resultados de laboratorio almacenados en forma tabular [8]. Además se ha trabajado en formas más específicas de esta enfermedad prediciendo la EICH aguda de grado alto (4 ó 5) [15], y la EICH crónica [5].

3. Descripción de los datasets

El análisis reportado en este trabajo utiliza un conjunto de datos públicos y un conjunto de datos proporcionado por el Hospital Luis Calvo Mackenna, un hospital pediátrico chileno, ubicado en la comuna de Providencia, Santiago de Chile. El conjunto de datos públicos, el cual denominaremos como **BMT-CH**, describe pacientes pediátricos con varias enfermedades hematológicas: trastornos malignos (e.g., leucemia linfoblástica aguda, leucemia mielógena aguda, leucemia mielógena crónica, síndrome mielodisplásico) y casos no malignos (e.g., anemia aplásica grave, anemia de Fanconi con adrenoleucodistrofia ligada al cromosoma X). Todos los pacientes se sometieron al trasplante de células madre hematopoyéticas de un donante alogénico no emparentado. Este dataset se puede encontrar publicado en la web de Kaggle [1]. El conjunto contiene 187 ejemplos caracterizados por 37 atributos. El dataset cuenta con atributos post-trasplante por lo que seleccionamos solo aquellos que son pre-trasplante quedando en la selección 31 atributos, además tiene valores faltantes por lo que para implementar los modelos se eliminaron dichos registros de pacientes eliminando un total de 45

pacientes. Esto resultó en una disminución en el número de observaciones del dataset de 187 a 142. Considerando que, en cada atributo con datos faltantes, los mismos representan menos del 10 % de los datos, podemos eliminar estos registros [11].

El conjunto de datos proporcionado por el Hospital Luis Calvo Mackenna, al cual llamaremos **HLCM**, describe también pacientes pediátricos con varias enfermedades hematológicas, tanto trastornos malignos como no malignos. A diferencia del conjunto mencionado anteriormente, este presenta diferentes tipos de donantes (e.g., alogénico emparentado y no emparentado compatible, alogénico emparentado parcialmente compatible). Este conjunto de datos cuenta con 536 ejemplos caracterizados por 227 atributos. Por correctitud, utilizamos solo los atributos pretrasplante, por lo tanto, de estos 227 atributos seleccionamos 51 para realizar la predicción. Además, se descartaron aquellos atributos que tuviesen más del 50 % de valores faltantes, mientras que a las columnas que tienen menos del 50 % de valores faltantes se les aplicó el imputador denominado Imputación múltiple por ecuaciones encadenadas [2]. En adición a esto, de los 536 ejemplos se tomaron solo 438 ya que el resto fueron sometidos a trasplantes autólogos (i.e., las células madre del donante provienen del receptor, es decir, el paciente es su propio donante), en cuyos casos no se desarrolla la EICH. En ambos datasets tomamos la decisión de separar los datos en conjuntos de testeo y entrenamiento, asignando el 20 % de los datos al conjunto de testeo y el resto al conjunto de entrenamiento.

4. Resultados

Al llevar a cabo el experimento se entrenaron varios modelos de Aprendizaje Automático tales como Árbol de Decisión [18], Bosque Aleatorio [3], XGBoost [4], Perceptrón Multicapa [13], Regresión Logística [7], Potenciación del Gradiente [12] y AdaBoost [14], adaptándolos según el caso para que modelen problemas de clasificación, seleccionando los mejores modelos usando la validación cruzada. Una vez aplicados todos estos modelos se realizó una comparación teniendo en cuenta distintas métricas de clasificación.

Para la comparación de resultados utilizamos las siguientes métricas: exactitud (accuracy), precisión (precision), sensibilidad (recall), especificidad (recall para la clase 0) y medida-F (F_β -score). También realizamos la selección de hiperparámetros de cada modelo utilizando validación cruzada con 10 folds y repitiendo cada validación cruzada 5 veces, buscando encontrar el modelo que mejor se comportara en la validación cruzada utilizando como métrica principal la **medida-F**. Una vez seleccionados los hiperparámetros de los modelos se midió el error en el conjunto de testeo con la métrica F_β , que es un promedio ponderado de precisión y sensibilidad con la particularidad de que si el β es menor que 1 le da mayor peso a la precisión y en caso de que β sea mayor a 1 presta mayor atención a la sensibilidad. En nuestro análisis se utilizaron los valores de β 0.5, 1 y 2. Cabe mencionar que en este caso es más importante la métrica de sensibilidad, considerando que predecir de manera incorrecta que un paciente desarrollará EICH y luego no la desarrolle (falso positivo) tiene menos impacto en la salud del paciente que predecir que no la desarrollará y luego sufra de la enfermedad (falso negativo). En tal sentido, priorizar la métrica F_2 es importante para minimizar el riesgo de perder casos positivos. En el cuadro 1 se pueden observar los resultados de aplicar los métodos de Aprendizaje Automático al dataset BMT-CH, mientras que en el cuadro 2 se muestran los resultados sobre el dataset HLCM.

Cuadro 1. Métricas de los modelos en el dataset BMT-CH

Métricas	Árbol de Decisión	Bosque aleatorio	XGBoost	Regresión Logística	Potenciación de Gradiente	AdaBoost	Perceptrón multicapa
Exactitud	0,52	0,66	0,59	0,62	0,55	0,62	0,62
$F_{0,5}$	0,61	0,70	0,67	0,69	0,64	0,67	0,67
F_2	0,69	0,87	0,63	0,75	0,77	0,89	0,89
F_1	0,65	0,77	0,65	0,72	0,70	0,77	0,77
Especificidad	0,18	0,18	0,55	0,36	0,09	0,00	0,00
Sensibilidad	0,72	0,94	0,61	0,78	0,83	1,00	1,00
Precisión	0,59	0,65	0,69	0,67	0,60	0,62	0,62

Cuadro 2. Métricas de los modelos en el dataset HLCM

Métricas	Árbol de Decisión	Bosque aleatorio	XGBoost	Regresión Logística	Potenciación de Gradiente	AdaBoost	Perceptrón multicapa
Exactitud	0,67	0,66	0,60	0,61	0,65	0,66	0,67
$F_{0,5}$	0,72	0,72	0,69	0,70	0,71	0,72	0,72
F_2	0,91	0,87	0,77	0,76	0,86	0,86	0,91
F_1	0,80	0,79	0,73	0,73	0,78	0,78	0,80
Especificidad	0,00	0,10	0,21	0,28	0,07	0,14	0,00
Sensibilidad	1,00	0,93	0,80	0,78	0,93	0,92	1,00
Precisión	0,67	0,68	0,67	0,69	0,67	0,68	0,67

A partir de nuestra experimentación observamos que en el dataset BMT-CH el modelo de Bosque Aleatorio resultó ser el de mayor exactitud, mientras que el Perceptrón Multicapa y AdaBoost comparten el mayor F_2 y mayor sensibilidad, luego XGBoost tiene la mayor especificidad (sensibilidad para la clase negativa) y precisión. En el caso del dataset privado HLCM los modelos de Árbol de Decisión y Perceptrón Multicapa comparten la mejor exactitud, F_2 y sensibilidad, mientras que Regresión Logística muestra la mejor especificidad y precisión.

De los métodos analizados, solo Regresión Logística y Árboles de Decisión son considerados modelos interpretables o explicables. Sin embargo, existen diversas técnicas que permiten mejorar la interpretabilidad de los otros modelos, tales como la selección de atributos para identificar los más influyentes en las predicciones del modelo, como así también técnicas de visualización especialmente diseñadas para cada clase de modelo predictivo.

5. Conclusiones

La EICH es una complicación importante que puede aparecer en el contexto de distintos tipos de trasplantes, y en particular en el trasplante pediátrico de médula ósea. La predicción de EICH es un problema abierto para las ciencias médicas, y poder contar con herramientas predictivas que ayuden a la toma de decisiones brinda un gran apoyo para los especialistas. Esto permitiría tomar las medidas necesarias para evitar que se desarrolle o para abordar esta enfermedad de manera temprana.

En este trabajo mostramos que existen distintos modelos predictivos que pueden formularse, utilizando como punto de referencia dos datasets, **HLCM** y **BMT-CH**. De acuerdo a los resultados obtenidos, no es posible identificar un único modelo óptimo en

relación a las diversas métricas de rendimiento presentadas. Actualmente continuamos desarrollando nuestra investigación enfocándonos no solo en la predicción de EICH sino también en la visualización de los datos, brindando herramientas para que los especialistas puedan sacar conclusiones a partir del comportamiento de la población. Además investigamos sobre la explicabilidad de los métodos con el objetivo de aumentar la confianza de los usuarios en los modelos explicando las razones de sus comportamientos y conclusiones. Tanto los modelos predictivos como la explicación de los mismos brindarán al especialista la información necesaria para una mejor toma de decisiones (e.g., decidir que donante es mejor para el paciente).

Referencias

1. Kaggle-BMT-CH:<https://www.kaggle.com/datasets/adamgudys/bone-marrow-transplant-children?resource=download>.
2. Melissa J. Azur. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20:40–49, 2011.
3. Leo Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.
4. Tianqi Chen. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.
5. Jocelyn S. Gandelman. Machine learning reveals chronic graft-versus-host disease phenotypes and stratifies survival after stem cell transplant for hematologic malignancies. *Haematologica*, 104:189–196, 2019.
6. María Antonieta Domínguez Gómez. Enfermedad injerto contra huésped. *Rev Cent Dermatol Pascua*, 21:104–108, 2012.
7. Joseph M. Hilbe. *Text in Statistical Science Logistic Regression Models*. Taylor and Francis Group, ISBN 978-042-914-913-9, 2009.
8. Makoto Iwasaki. Establishment of a predictive model for GVHD-free, relapse-free survival after allogeneic HSCT using ensemble learning. *Blood Advances*, 6:2618–2627, 2022.
9. Malgorzata Kretowska. Tree-based models for survival data with competing risks. *Computer Methods and Programs in Biomedicine*, 159:185–198, 2018.
10. Xueou Liu. Dynamic forecasting of severe acute graft-versus-host disease after transplantation. *Natural computational science*, 2:153–159, 2022.
11. Miguel Marino. Missing data in primary care research: importance, implications and approaches. *Oxford Journals Family Practice*, 38:199–202, 2021.
12. Alexey Natekin. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7:1–21, 2013.
13. Hassan Ramchoun. Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4:26–30, 2016.
14. Robert E. Schapire. Explaining AdaBoost. *Empirical Inference*, pages 37–52, 2013.
15. Meng-Zhu Shen. A comprehensive model to predict severe acute graft-versus-host disease in acute leukemia patients after haploidentical hematopoietic stem cell transplantation. *Experimental Hematology and Oncology*, 11:1–10, 2022.
16. Marek Sikora. GuideR: A guided separate-and-conquer rule learning in classification, regression, and survival settings. *Knowledge-Based Systems*, 173:1–14, 2019.
17. Shengpu Tang. Predicting acute graft-versus-host disease using machine learning and longitudinal vital sign data from electronic health records. *JCO Clinical Cancer Informatics*, 4:128–135, 2020.
18. Yan yan Song. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27:130–135, 2015.