

CONICET



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



Hospital de Niños
Dr. Luis Calvo Mackenna
Hospital Autogestionado de Alta Complejidad Fundado en 1942



fcfm
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

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

^{1,2}Lázaro A. Gibert-García, ³Rodrigo Guerra, ⁴Jocelyn Dunstan, ⁵Julia Palma, ^{1,2}Axel J. Soto, ^{1,2}Ana Maguitman, ^{1,2}Carlos Chesñear.

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

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

³FCFM, Universidad de Chile, Chile.

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

⁵Hospital Luis Calvo Mackenna, Santiago, Chile.

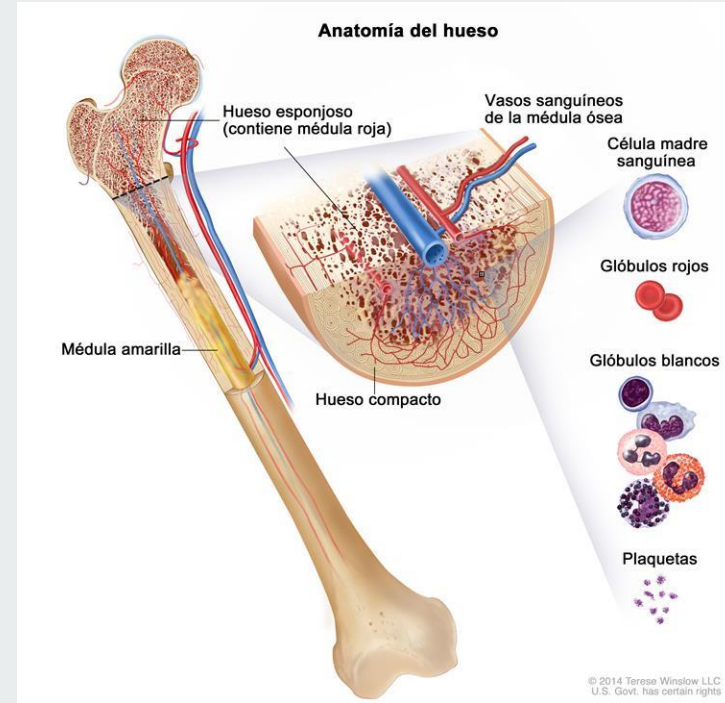
¿Qué es el trasplante de células madre (trasplante de médula ósea)?

- Es un tratamiento médico que reemplaza la médula ósea por células sanas. Las células de reemplazo pueden proceder del propio cuerpo del paciente (autólogo) o de un donante (allogénico).
- Se puede utilizar para tratar ciertos tipos de cáncer, como leucemia, mieloma y linfoma, y otras enfermedades del sistema inmunitario y de la sangre que afectan la médula ósea.



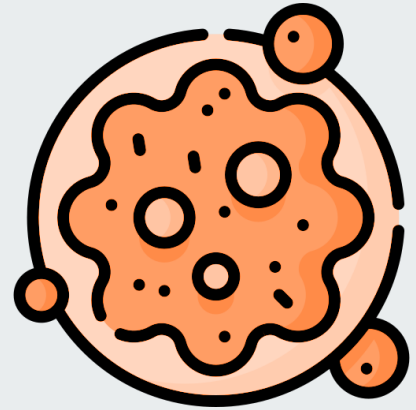
¿Qué es la médula ósea?

- La médula ósea es tejido blando y esponjoso en el cuerpo que contiene células madre hematopoyéticas.
- Se encuentra en el centro de la mayoría de los huesos.
- También se encuentran células madre hematopoyéticas en la sangre que circula por todo el cuerpo.



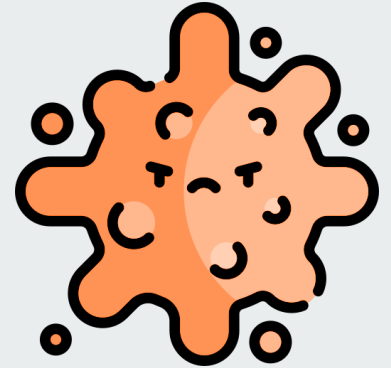
Enfermedad de Injerto contra huésped

- 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 tejidos del huésped, y constituye una complicación frecuente luego de un trasplante de médula ósea.



Enfermedad de Injerto contra huésped

- 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.



Motivación

- Desde la ciencia de la salud no se tiene una forma efectiva de predecir si el paciente puede desarrollar EICH.
- Existen pocos datos clínicos para estudiar patrones de la enfermedad.
- El descubrimiento temprano de los pacientes propensos a contraer EICH ayuda a la toma las medidas necesarias para que evitar que se contraiga la enfermedad y en el peor de los casos la muerte del paciente.



Propuesta



- Se propone modelar utilizando modelos de aprendizaje para predecir si el paciente va a desarrollar la enfermedad o no, utilizando dos bases de datos.
- Se realizó una comparación del desempeño de predicción de cada modelo utilizando distintas métricas.

Antecedentes



- Entre los estudios anteriormente realizados 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 ([Sikora et al. 2019](#)).
- Se han utilizado métodos basados en árboles de decisión que se han comportado de manera satisfactoria ([Kretowska et al. 2018](#)).

Marek Sikora. GuideR: A guided separate-and-conquer rule learning in classification, regression, and survival settings. *Knowledge-Based Systems*, 173:1–14, 2019.

Malgorzata Kretowska. Tree-based models for survival data with competing risks. *Computer Methods and Programs in Biomedicine*, 159:185–198, 2018.

Antecedentes



- Existen también estudios para predecir complicaciones específicas al trasplante de médula ósea como lo es la EICH ([Liu et al. 2022](#)).
- Para la predicción de la EICH se han utilizado varias fuentes de datos, desde registros electrónicos de salud ([Tang et al. 2020](#)) hasta resultados de laboratorio almacenados en forma tabular ([Iwasaki et al. 2022](#)).

Xueou Liu. Dynamic forecasting of severe acute graft-versus-host disease after transplantation. *Natural computational science*, 2:153–159, 2022.

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.

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.

Datasets

- **BMT-CH** es un dataset publicado en Kaggle suministrado por [Sikora et al. \(2019\)](#).
- **HLCM** es un dataset proporcionado por el Hospital Luis Calvo Mackenna, un hospital pediátrico ubicado en **Santiago de Chile**.

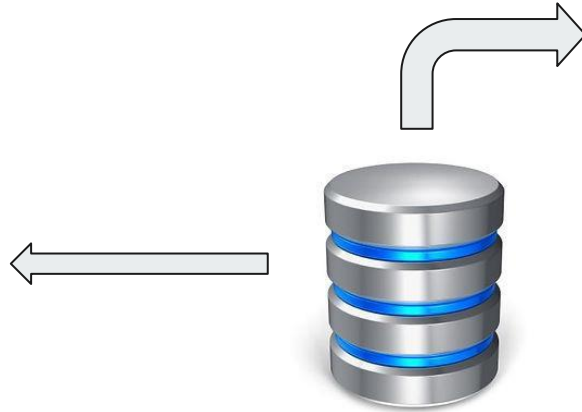


Marek Sikora. *GuidER: A guided separate-and-conquer rule learning in classification, regression, and survival settings*. *Knowledge-Based Systems*, 173:1–14, 2019.

BMT-CH



Describe pacientes pediátricos con varias enfermedades hematológicas.



Casos malignos:

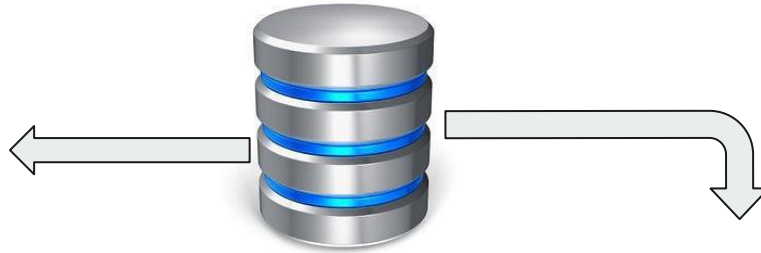
- Leucemia linfoblástica aguda.
- Leucemia mielógena aguda.
- Síndrome mielodisplásico.

Casos no malignos:

- Anemia aplásica grave.
- Anemia de Fanconi.

BMT-CH

Contiene 187 ejemplos caracterizados por 37 atributos.



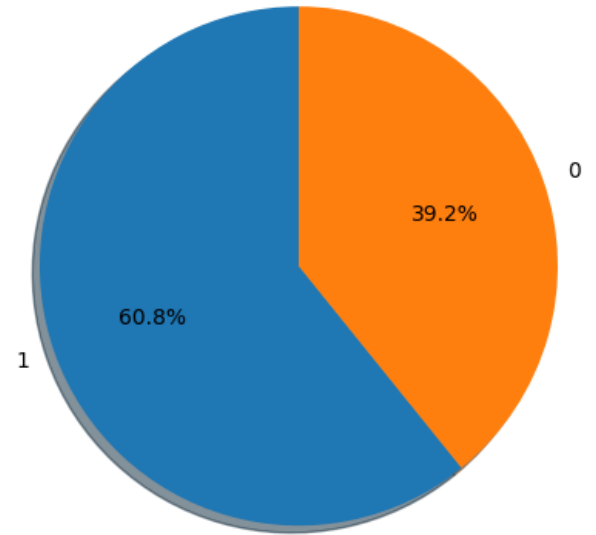
Solo se utilizaron aquellos atributos que son pre-trasplante quedando 31 en la selección.

BMT-CH



Distribución de las clases:

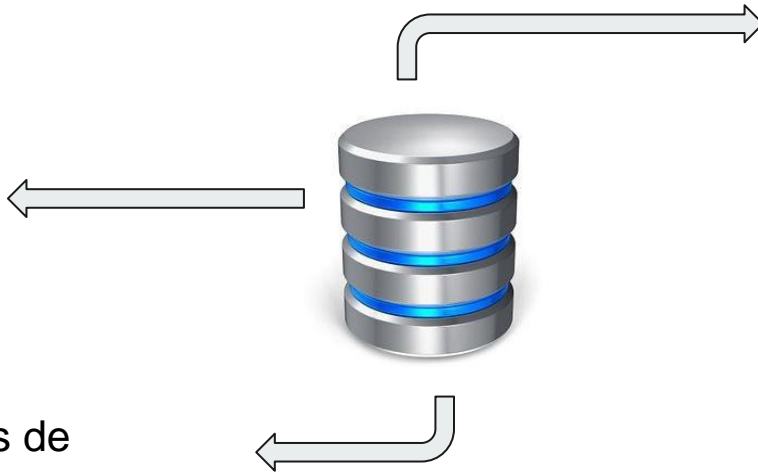
- 104 pacientes padecen de EICH (clase 1).
- 67 pacientes no padecen la enfermedad (clase 0).



HLCM



Describe pacientes pediátricos con varias enfermedades hematológicas.



Contiene tanto trastornos malignos como no malignos.

Presenta diferentes tipos de donantes:

- Autólogos.
- Alogénicos.

HLCM



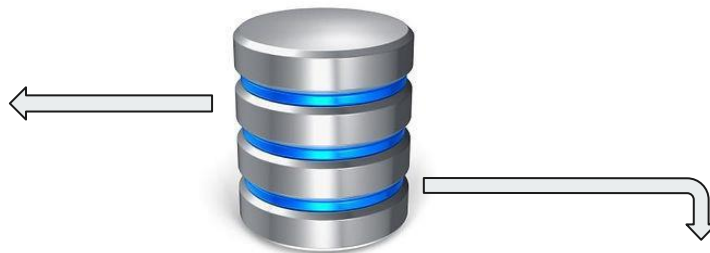
Tipos de donantes Alogénicos:

- Emparentado.
- No emparentado compatible.
- Emparentado parcialmente compatible.

HLCM



Cuenta con 536 ejemplos
caracterizados por 227 atributos.



Se seleccionaron solo los
atributos pre-trasplante.

HLCM-Limpieza



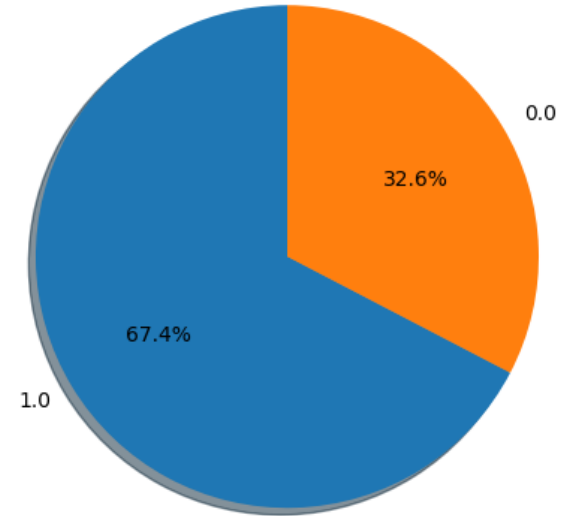
- Descartados atributos con más del 50% de valores faltantes.
- Atributos con menos del 50% de valores faltantes (**Imputación múltiple por ecuaciones encadenadas**).
- Quedando una selección de 51 atributos.
- De los 536 ejemplos se tomaron solo 438 ya que el resto fueron sometidos a trasplantes autólogos, en cuyos casos no se desarrolla la EICH.

HLCM-Limpieza

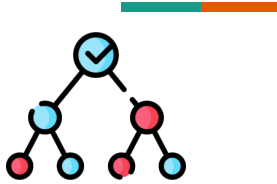


Distribución de las clases:

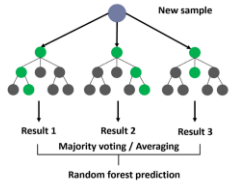
- 295 pacientes padecen de EICH (clase 1).
- 143 pacientes no padecen la enfermedad (clase 0).



Modelos



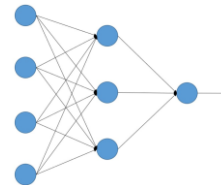
Árbol de Decisión



Bosque Aleatorio



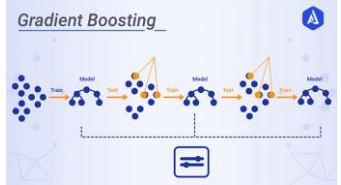
XGBoost



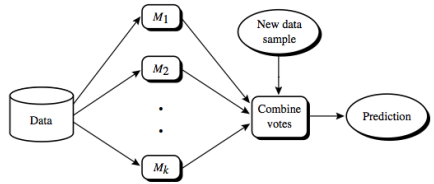
Perceptrón Multicapa



Regresión Logística



Potenciación del Gradiente



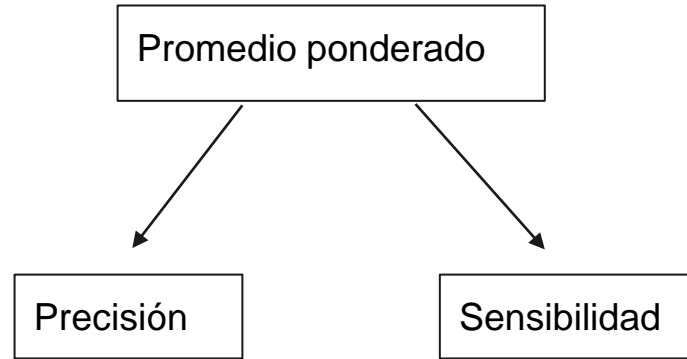
AdaBoost

Métricas



- Exactitud (accuracy).
- Especificidad (recall para la clase 0).
- Sensibilidad (recall para la clase 1).
- Precisión (precisión para la clase 1).
- Medida-F (F_β -score).

Medida F (F_{β} -score)



- Si $\beta = 1$ media armónica.
- Si $\beta < 1$ + precisión.
- Si $\beta > 1$ + sensibilidad.
- $\beta = 0.5, 1$ y 2 .

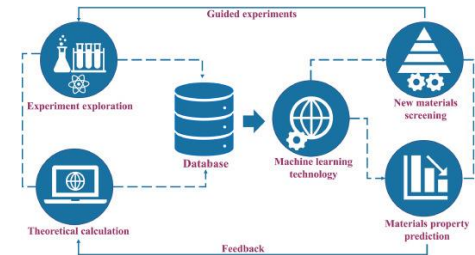
Medida F (F_β -score)



- En este caso es más importante la métrica de sensibilidad, ya que los falsos positivos tienen menos impacto en la salud del paciente que los falsos negativos.
- Priorizar la métrica F_2 es importante para minimizar el riesgo de perder casos positivos.

Diseño de los experimentos

- En cada caso se dividió el dataset en un subconjunto de entrenamiento (80%) y otro de pruebas (20%).
- Se utilizó la misma semilla al realizar la división para asegurar que se escogieran los mismos datos en cada caso.
- Se realizó la selección de hiperparámetros en cada modelo utilizando validación cruzada con 10 folds.
- Se repitió cada validación cruzada 5 veces.
- Se escogió el modelo de mejor comportamiento en la validación cruzada utilizando como métrica principal, la medida-F.





Resultados

BMT-CH



Métricas	Árbol de Decisión	Bosque Aleatorio	XGBoost	Regresión Logística	Potenciación del 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

HLCM



Métricas	Árbol de Decisión	Bosque Aleatorio	XGBoost	Regresión Logística	Potenciación del 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

Interpretabilidad de los modelos

¿Por qué es importante la explicabilidad o interpretabilidad de los modelos?

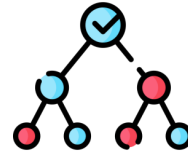
- Permite obtener una intuición sobre el funcionamiento general del modelo.
- Permite conocer los atributos a los que el modelo considera más importantes.
- Permite llegar a conclusiones sobre las razones por las cuales el modelo está haciendo las predicciones de la manera en que las hace.
- Genera una mayor confianza en el modelo.



Interpretabilidad de los modelos



Regresión Logística



Árboles de Decisión

Interpretables

Interpretabilidad de los modelos



Regresión Logística:

- Cada atributo tiene asociado a un coeficiente.
- Permite intuir a qué atributos el modelo le está dando mayor importancia durante las predicciones.

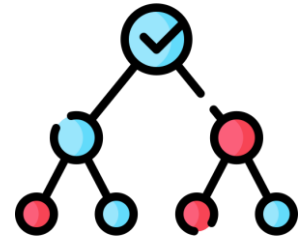


Interpretabilidad de los modelos



Árboles de decisión:

- Permite analizar los caminos que tomó el modelo para llegar a la predicción.
- Permite calcular la importancia de los atributos analizando cuánto ha disminuido la varianza de un nodo u hoja en comparación con el nodo principal.

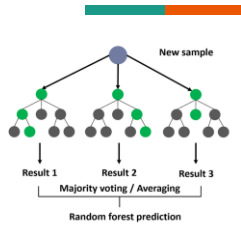


Importancia de atributos.



Datasets	Regresión Logística	Árbol de Decisión
BMT-CH	Edad del paciente entre 5 y 10 .	CMV status (0)
	Peso corporal del paciente.	alelos
	Factor rh +	CD34x1 ⁶ por kg
	Coincidencia grupo sanguíneo	antigen.
HLCM	Profilaxis(choice Metilprednisolona (MP,MTP))	Edad de TPH
	Grupo sanguíneo del paciente	Quimioterapia (Antitimoglobulina)
	Tipo de diagnóstico (Oncológico)	Quimioterapia (Citarabina)
	Factor rh +	Profilaxis

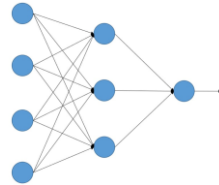
Interpretabilidad de los modelos



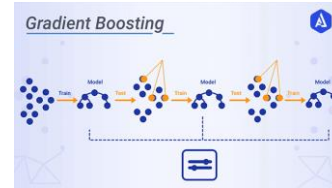
Bosque Aleatorio



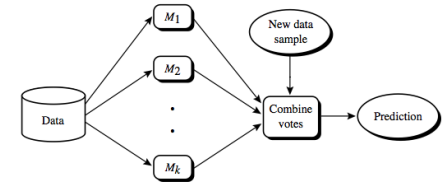
XGBoost



Perceptrón Multicapa



Potenciación del Gradiente



AdaBoost

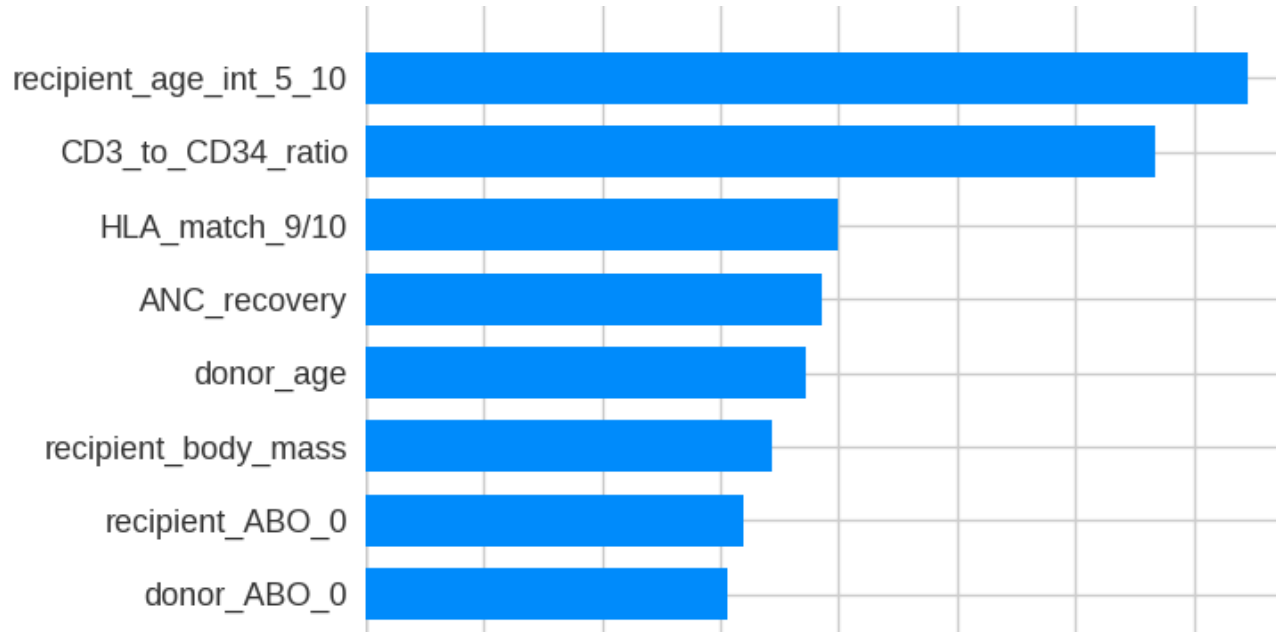
Selección de atributos.
Técnicas de visualización especialmente diseñadas.

Interpretabilidad de los modelos



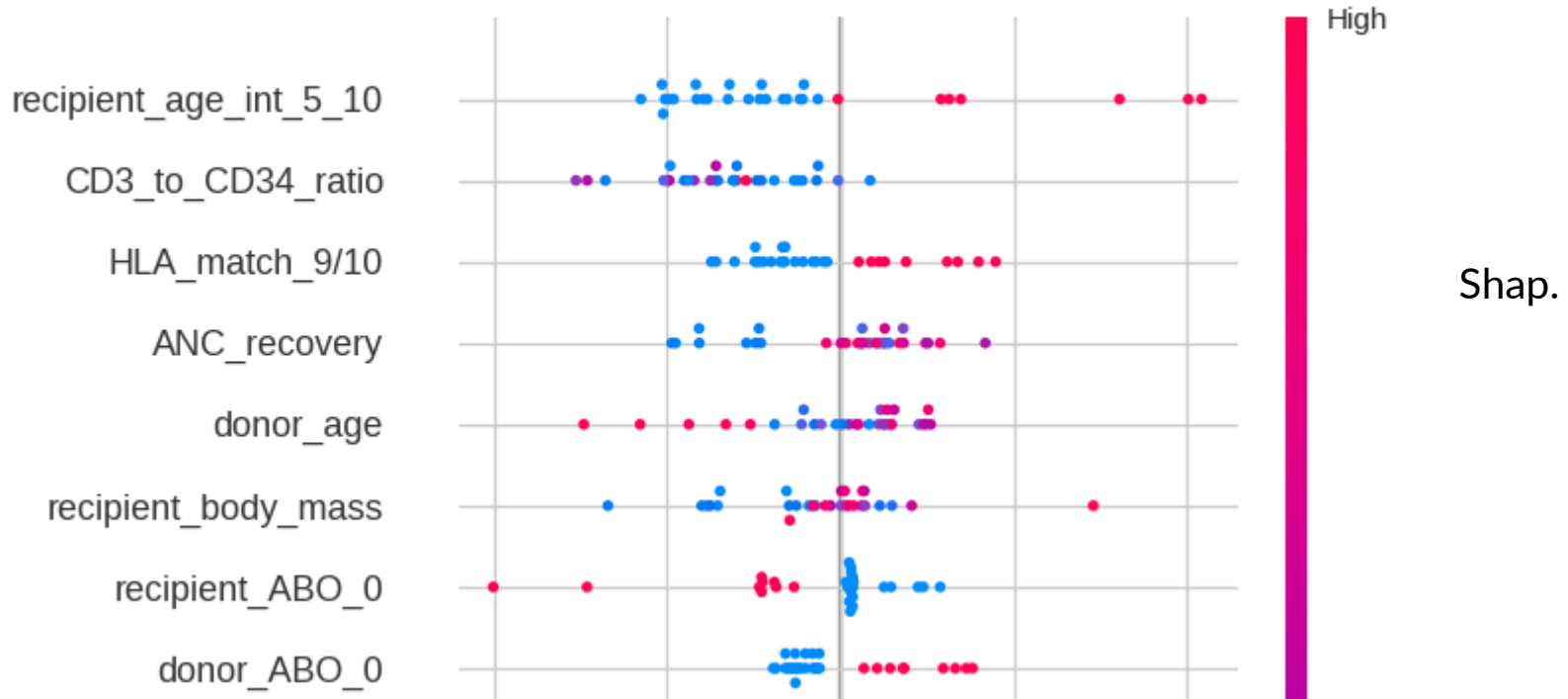
- Eli5.
- Yellowbrick.
- Lime.
- Shap.
- Alibi.

Interpretabilidad de los modelos

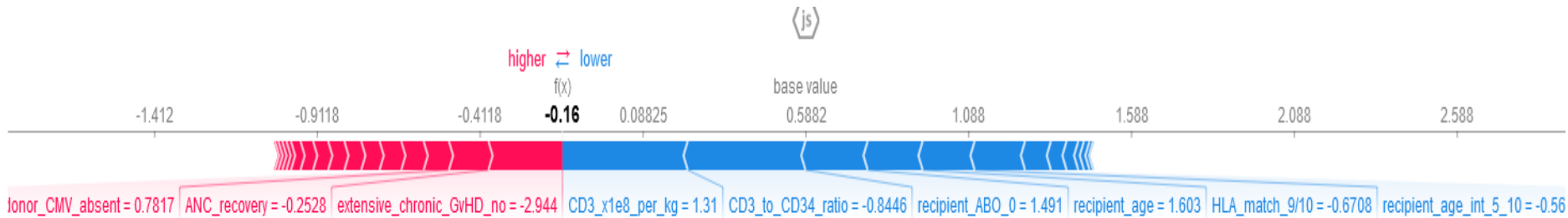


Shap.

Interpretabilidad de los modelos



Interpretabilidad de los modelos



Shap.

Conclusiones



- 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.

Conclusiones



- Mostramos que existen distintos modelos predictivos que pueden formularse, utilizando como punto de referencia dos datasets: HLCCM 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.

Trabajo a futuro



- 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.

Trabajo a futuro



- Hemos comenzado a investigar 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 qué donante es mejor para el paciente).

CONICET



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE



Hospital de Niños
Dr. Luis Calvo Mackenna
Hospital Autogestionado de Alta Complejidad Fundado en 1942



fcfm
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

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

^{1,2}Lázaro A. Gibert-García, ³Rodrigo Guerra, ⁴Jocelyn Dunstan, ⁵Julia Palma, ^{1,2}Axel J. Soto, ^{1,2}Ana Maguitman, ^{1,2}Carlos Chesñear.

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

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

³FCFM, Universidad de Chile, Chile.

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

⁵Hospital Luis Calvo Mackenna, Santiago, Chile.