

Predicción de Rendimiento en Aféresis Plaquetaria mediante Machine Learning

Integración de Variables Históricas del Donante con Validación Temporal Rigurosa

Mario Miguel Barbosa Rengifo, MD, MSc
Luisa Fernanda Ospina Hincapie. Bsc MSc

Congreso · Abril 2026

Agenda

01

Introducción

Aféresis plaquetaria · Problema clínico · Estado del arte

02

Métodos

Población · Ingeniería de características · Algoritmos ML

03

Resultados

Características donantes · Performance modelos · Variables predictoras

04

Discusión

Hallazgos principales · Implicaciones clínicas · Limitaciones

05

Conclusiones

Aportes · Aplicación clínica · Direcciones futuras

INTRODUCCIÓN

Aféresis Plaquetaria y la Necesidad de Predicción

Contexto clínico · Estado del arte · Hipótesis de trabajo

¿Qué es la Aféresis Plaquetaria?

Introducción · Relevancia clínica

► Procedimiento esencial en medicina transfusional

- Separación celular automatizada del donante
- Permite obtener 2 ó 3 unidades terapéuticas
- Un procedimiento = múltiples unidades (vs. 1 en sangre total)

► Factores que determinan el rendimiento:

- Recuento plaquetario inicial del donante
- Peso corporal, sexo, hemoglobina
- Historial individual de donaciones previas

Impacto Operativo

Inventario

Planificación proactiva de unidades disponibles

Citaciones

Programación estratégica de donantes de alto rendimiento

Eficiencia

Optimización del uso de equipos y personal técnico

El Problema Clínico

Estado del arte · Brecha en la literatura

¿Por qué predecir el rendimiento antes del procedimiento?

Rendimiento variable: 2 ó 3 unidades por procedimiento (22.6% producen 3u)

Impacto directo en gestión de inventarios y planificación hospitalaria

Modelos previos reportan solo 75%–82% de accuracy

Limitación crítica: usaban solo variables contemporáneas del donante

Limitación metodológica: modelos previos usaban variables del procedimiento actual → resultados inflados

 **Hipótesis: Incorporar historial del donante + solo variables pre-procedimiento → superar el estado del arte**

MÉTODOS

Diseño del Estudio, Población y Algoritmos

5,712 procedimientos · 42 variables · Solo información pre-procedimiento

Diseño del Estudio

Retrospectivo observacional · Cali, Colombia · 2012–2025

5,712

Procedimientos
analizados

2,230

Donantes
únicos

13 años

Seguimiento
2012–2025

42

Variables
predictoras

Criterios de Inclusión / Exclusión

- ✓ Procedimientos con resultado 2 ó 3 unidades
- ✓ Datos completos de variables pre-procedimiento
- ✗ Datos faltantes o valores plaquetarios implausibles
- ✗ Rendimientos atípicos (1u o >3u, <1% del total)

Fuente de Datos

- Registros electrónicos del banco de sangre
- Fundación Valle del Lili, Cali, Colombia
- Variables demográficas, clínicas y del procedimiento
- Historial completo de donaciones previas por donante
- Aprobado por Comité de Ética · Exención de consentimiento

Validación Temporal 80/20

Entrenamiento: primeros 80% cronológicamente (n=4,569) | Prueba: últimos 20% (n=1,143)
Sin cross-validation → preserva la dependencia temporal natural de los datos clínicos

Algoritmos de Machine Learning Evaluados

6 algoritmos · class_weight='balanced' · 300 árboles

Extra Trees

★ MEJOR

XGBoost

2°

Random Forest

3°

Gradient Boosting

4°

Regresión Logística

5°

Árbol de Decisión

6°

Configuración común para ensemble:

- 300 árboles · Profundidad máxima 15 · class_weight='balanced' (manejo desbalance 77:23)
- StandardScaler solo para Regresión Logística · Ajustado exclusivamente en entrenamiento
- random_state=42 para reproducibilidad · Métricas: Accuracy, AUC-ROC, Sens., Espec., F1

RESULTADOS

Características de Donantes, Modelos Predictivos y Variables Clave

88.5% accuracy · AUC-ROC 0.934 · Validación temporal estricta

Características de los Donantes

N=2,230 donantes únicos · 5,712 procedimientos · 2012–2025

38.6

Edad media
± 39.5 años (DE)

76.8

Peso medio
± 13.3 kg (DE)

57.1%

Sexo masculino
(1,274 de 2,230)

61.8%

Grupo O
(más frecuente)

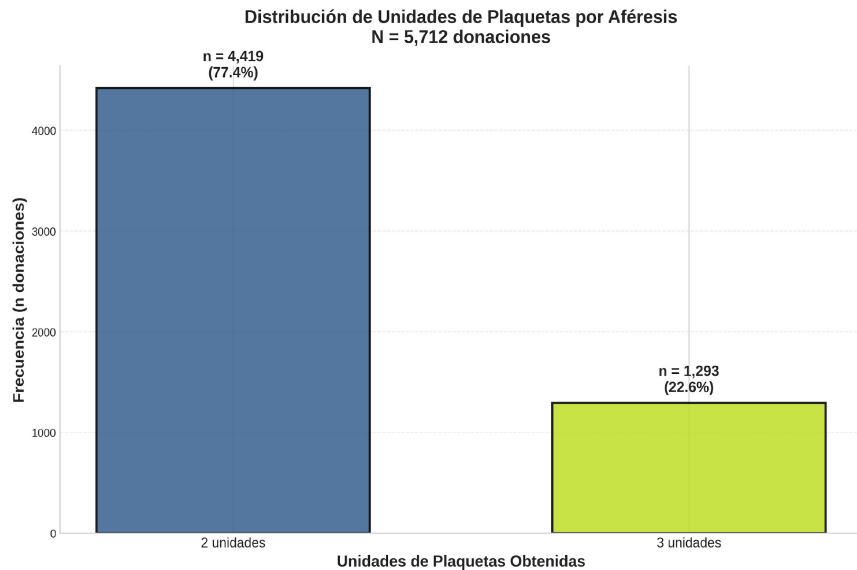
Variable	Media ± DE	Mediana [Q25–Q75]
Edad (años)	38.6 ± 39.5	37 [28–46]
Peso (kg)	76.8 ± 13.3	76 [67–85]
Estatura (cm)	167 ± 14	168 [160–174]
Hb (g/dL)	14.8 ± 1.6	14.8 [13.9–15.7]
Plaquetas ($\times 10^3/\mu\text{L}$)	317 ± 118	306 [272–347]

64.4% de donantes realizaron una sola donación (n=1,436)

4.5% con >10 donaciones (donantes altamente recurrentes)

Distribución del Rendimiento y Comparación de Grupos

2 unidades (77.4%) vs. 3 unidades (22.6%) · p<0.001



Diferencias clave entre grupos

Variable	2 unidades	3 unidades	Diferencia	P
Plaquetas iniciales	297 ± 96	387 ± 155	+90 ×10 ³ /μL	p<.001
Peso corporal	76.3 ± 12.7 kg	81.4 ± 13.3 kg	+5.2 kg	p<.001
Diferencial plaq.	135 ± 94	198 ± 210	+62 ×10 ³ /μL	p<.001
Hemoglobina	14.9 ± 1.7 g/dL	14.5 ± 1.1 g/dL	-0.4 g/dL	p<.001
Edad	40.6 ± 29.3 a	39.8 ± 11.6 a	-0.9 a	p=.31 NS

Hallazgo clave: El recuento plaquetario inicial es el predictor individual más importante — donantes 3u tienen un 30% más de plaquetas en promedio

Comparación de Performance: 6 Modelos ML

Validación temporal 80/20 · n=1,143 prueba

Modelo	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Extra Trees ★	0.885	0.770	0.736	0.753	0.934
XGBoost	0.880	0.745	0.758	0.751	0.934
Random Forest	0.878	0.748	0.740	0.744	0.943
Gradient Boosting	0.877	0.741	0.744	0.742	0.939
Reg. Logística	0.868	0.680	0.842	0.753	0.924
Árbol de Decisión	0.835	0.645	0.685	0.664	0.782

Todos los modelos ensemble superan ampliamente el estado del arte reportado en literatura (75%–82%)

Extra Trees seleccionado como modelo final por mejor balance accuracy/interpretabilidad.
AUC-ROC 0.934 → clasificación «excelente» (umbral clínico >0.90, Hosmer & Lemeshow 2013)

Modelo Extra Trees: Performance Detallada

Matriz de confusión · Métricas clínicas

88.5%

Accuracy
(IC95% 86.6%–90.3%)

0.934

AUC-ROC
(excelente >0.90)

93.1%

Especificidad
para 2 unidades

73.6%

Sensibilidad
para 3 unidades

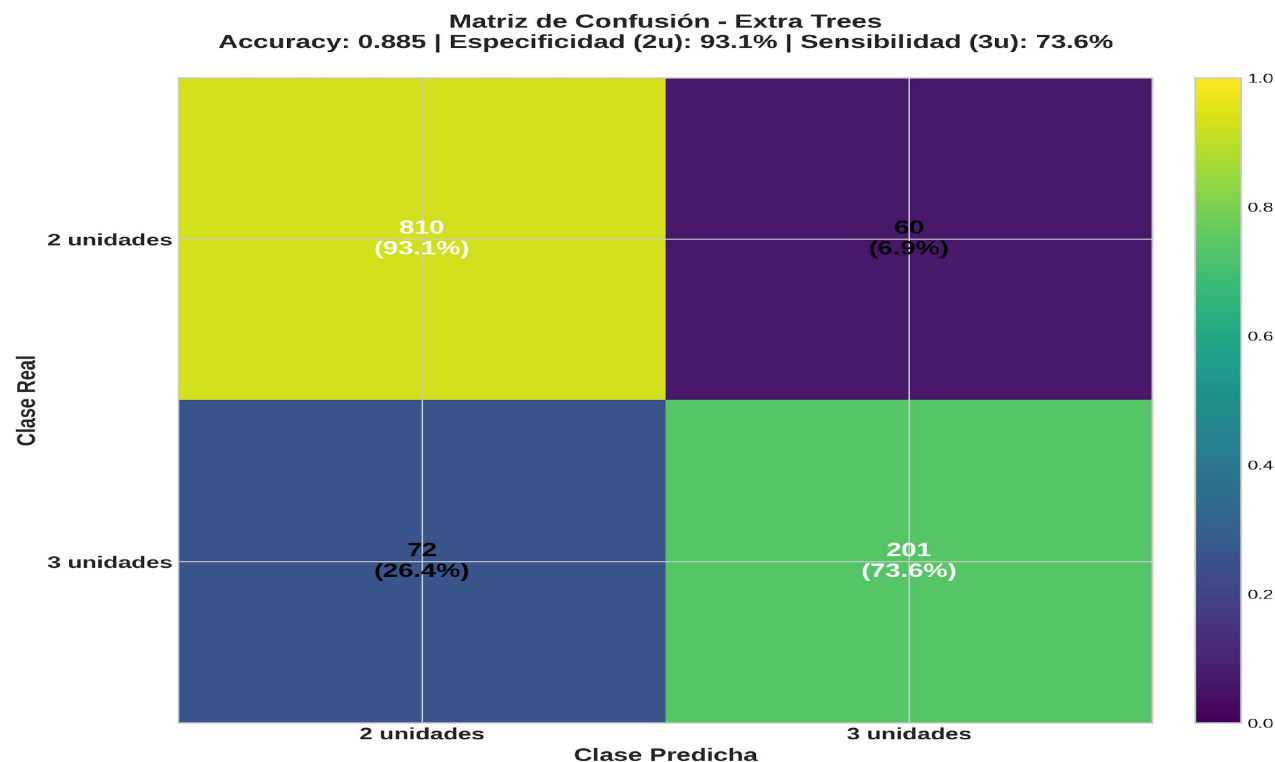
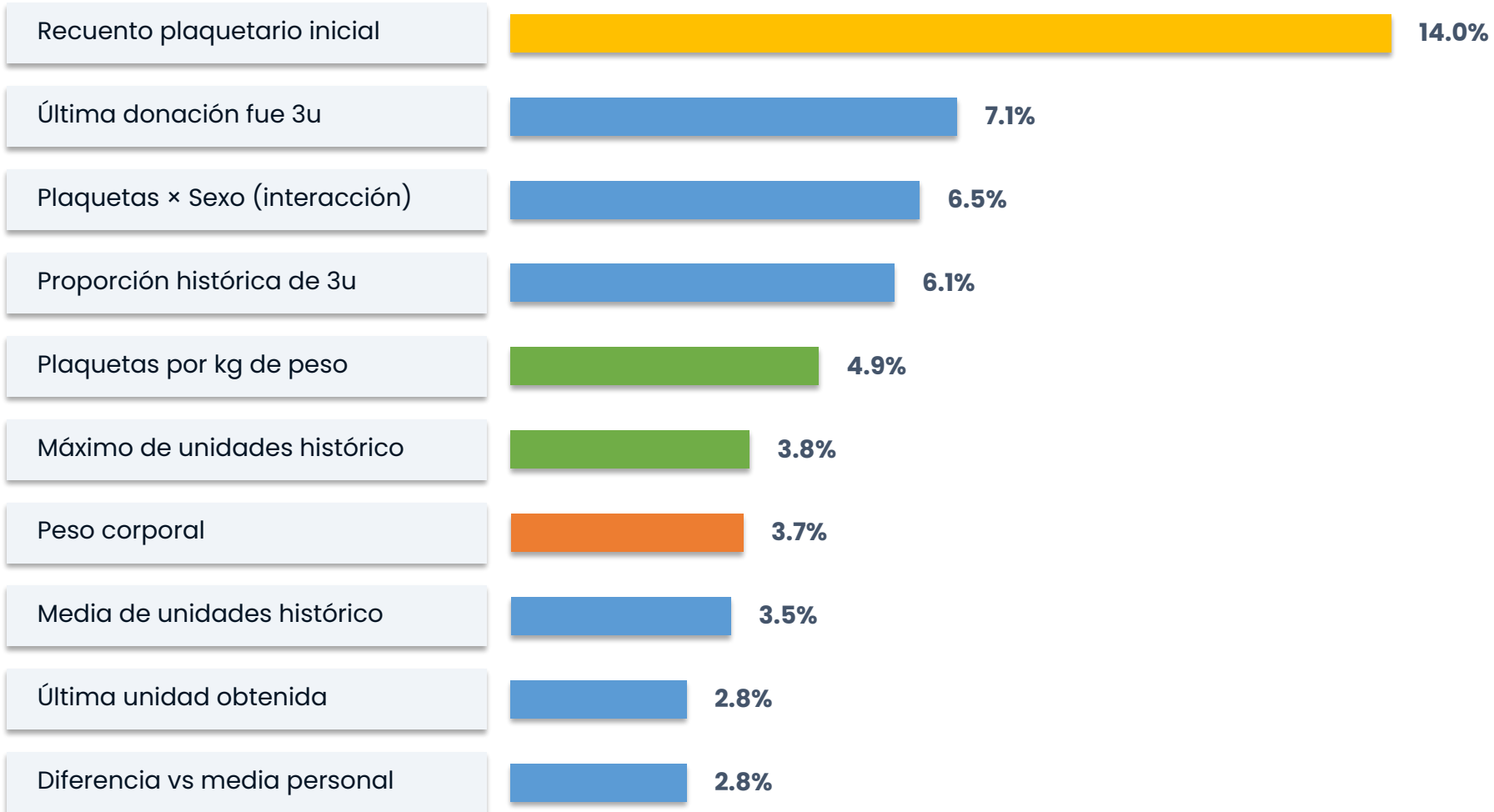


Figura 4. Matriz de confusión: 810 VP (2u) + 201 VP (3u) de 1,143 casos de prueba

Variables Predictoras Más Importantes

Gini importance · Extra Trees · Top 10 de 42 variables



DISCUSIÓN

Hallazgos, Implicaciones Clínicas y Limitaciones

Comparación con literatura · Valor clínico · Limitaciones del estudio

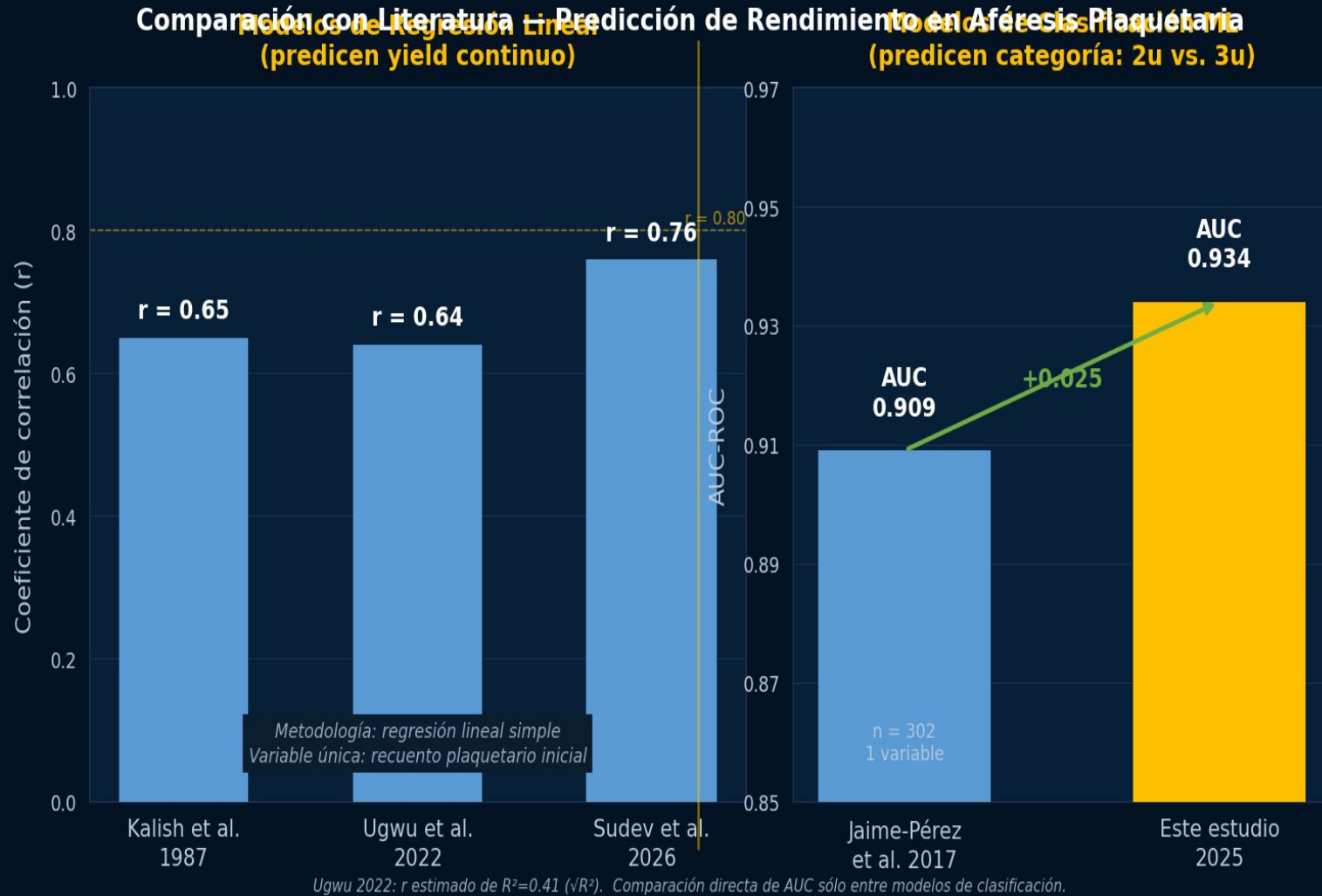
Estado del Arte — Predicción de Rendimiento en Aféresis Plaquetaria

Autor (Año)	País · Centro	N	Diseño	Método estadístico	Predictor principal	Métrica clave	Variables	Limitación central
Kalish et al. 1987	EUA	916	Retrospectivo (bivariado)	Correlación de Pearson	Recuento plaquetario pre-aféresis	$r = 0.65$	1	Sin modelo clasificatorio
Jaime-Pérez et al. 2017 ★	México	302	Retrospectivo Regr. + ROC	Regresión lineal + curva ROC	Recuento plaquetario pre-donación	AUC = 0.909 Sens 82% / Esp 93%	1	1 variable; n reducido
Ugwu et al. 2022	Nigeria	153	Retrospectivo (5 años)	Regresión lineal simple	Recuento plaquetario + vol. concentrado	$R^2 = 0.41$ ($r \approx 0.64$)	1-2	Sin AUC; muestra pequeña
Sudev & Vijayalakshmi 2026	India	670	Prospectivo (6 meses)	Regresión múltiple lineal	Recuento plaquetario + IMC	$r = 0.758$ $\beta = 0.657$	2	Sin historial donante; sin AUC
Hussein & Enein 2025	Egipto	900	Prospectivo (3 dispositivos)	Análisis comparativo	Tipo equipo + peso corporal	Descriptivo ($p < 0.05$)	N/A	Sin modelo ML; sin AUC
Este estudio 2025 ♦	Colombia (FVL)	5.712	Temporal 80/20 cronoló.	Extra Trees ML (42 vars)	Recuento + historial completo donante	AUC = 0.934 Acc = 88.5%	42	—

★ Único estudio previo con AUC reportado — comparación directa válida ♦ Este estudio: 19× más casos que ★, 42 variables con historial longitudinal, validación temporal prospectiva N/R = no reportado

Comparación con Literatura — Métricas de Rendimiento

Regresión lineal vs. clasificación ML · Solo Jaime-Pérez 2017 permite comparación directa de AUC



Ventaja de nuestro modelo

AUC
0.934

vs. 0.909 Jaime-Pérez
(+0.025 puntos)

19x

más procedimientos
(5,712 vs. 302)

42

variables predictoras
vs. 1 variable previa

100%

variables pre-procedimiento
validación prospectiva

88.5%

accuracy global
(Sens 73.6% / Esp 93.1%)

Implicaciones Clínicas para el Banco de Sangre

Aplicación directa · Sin modificar flujo de trabajo actual

Optimización de Inventarios

Predicción anticipada del rendimiento esperado
Permite planificación proactiva de citas
Reducción de faltantes y excedentes de plaquetas

Citación Estratégica

Priorización de donantes con alta probabilidad de 3u
Máxima eficiencia en momentos de inventario crítico
Optimización del recurso humano y tecnológico

Eficiencia Operativa

Reducción de procedimientos con bajo rendimiento
Optimización de programación de personal técnico
Mejor utilización de equipos de aféresis

Personalización del Procedimiento

Ratio plaquetas actuales/media personal del donante
Cada recuento se interpreta en contexto individual
Adaptación potencial de parámetros del equipo

✓ El modelo usa SOLO variables disponibles rutinariamente antes del procedimiento — implementación directa sin costos adicionales

Limitaciones y Direcciones Futuras

Fortalezas · Debilidades · Próximos pasos

Limitaciones

- Diseño monocéntrico → generalización limitada
- Sin validación externa en otros centros
- Retrospectivo → potencial sesgo de selección
- No se evaluó el impacto económico directo
- No se incluyeron variables del equipo de aféresis
- (marca, modelo, versión de software)

Fortalezas

- ✓ 5,712 procedimientos → poder estadístico robusto
- ✓ 13 años de seguimiento → variabilidad temporal
- ✓ Validación temporal estricta → escenario real
- ✓ Solo variables disponibles pre-procedimiento
- ✓ Comparación sistemática de 6 algoritmos
- ✓ Historial del donante como innovación metodológica

Direcciones Futuras

- Validación multicéntrica prospectiva (Colombia y región)
- Incorporación de LSTM/Transformers para donantes recurrentes
- Estudio de costo-efectividad e impacto en inventario hospitalario
- Integración con sistema de programación automatizada de donantes

CONCLUSIONES

Aportes, Aplicación Clínica y Relevancia

Conclusiones

1

88.5% accuracy (IC95% 86.6%–90.3%) · AUC-ROC 0.934

El modelo Extra Trees alcanza desempeño «excelente» en predicción de rendimiento de aféresis plaquetaria

2

Validación temporal rigurosa (80/20 cronológico)

Resultados clínicamente aplicables — el modelo predice usando SOLO información disponible antes del procedimiento

3

Las variables históricas del donante importan

Contribución colectiva >20% del poder predictivo — supera lo que cualquier variable individual contemporánea puede aportar

4

Listo para implementación clínica directa

No requiere mediciones adicionales, equipamiento especializado ni modificación del flujo de trabajo actual

Gracias

¿Preguntas?