

XXXI CONGRESO  
INTERNACIONAL



# XXXI CONGRESO INTERNACIONAL SCAI 2025

SOCIEDAD CHILENA DE ALERGIA E INMUNOLOGIA

[www.scai.cl](http://www.scai.cl)

[www.congresoscai.cl](http://www.congresoscai.cl)



# Inteligencia artificial en inmunología: Uso en predicción de alergias

**Dr. Francisco Sidgman Galleguillos**

**Inmunólogo**

**Hospital del Salvador/ Clínica Bupa Santiago**

**Puerto Varas, 3 de octubre de 2025**



## Hoja de ruta

- Introducción
- ¿Qué es la inteligencia artificial?
- ¿Por qué utilizar IA?
- Potenciales usos
  - Bioseguridad
  - Predecir resultados de TPO en alergia alimentaria y fármacos
- Futuro de la IA

## Introducción

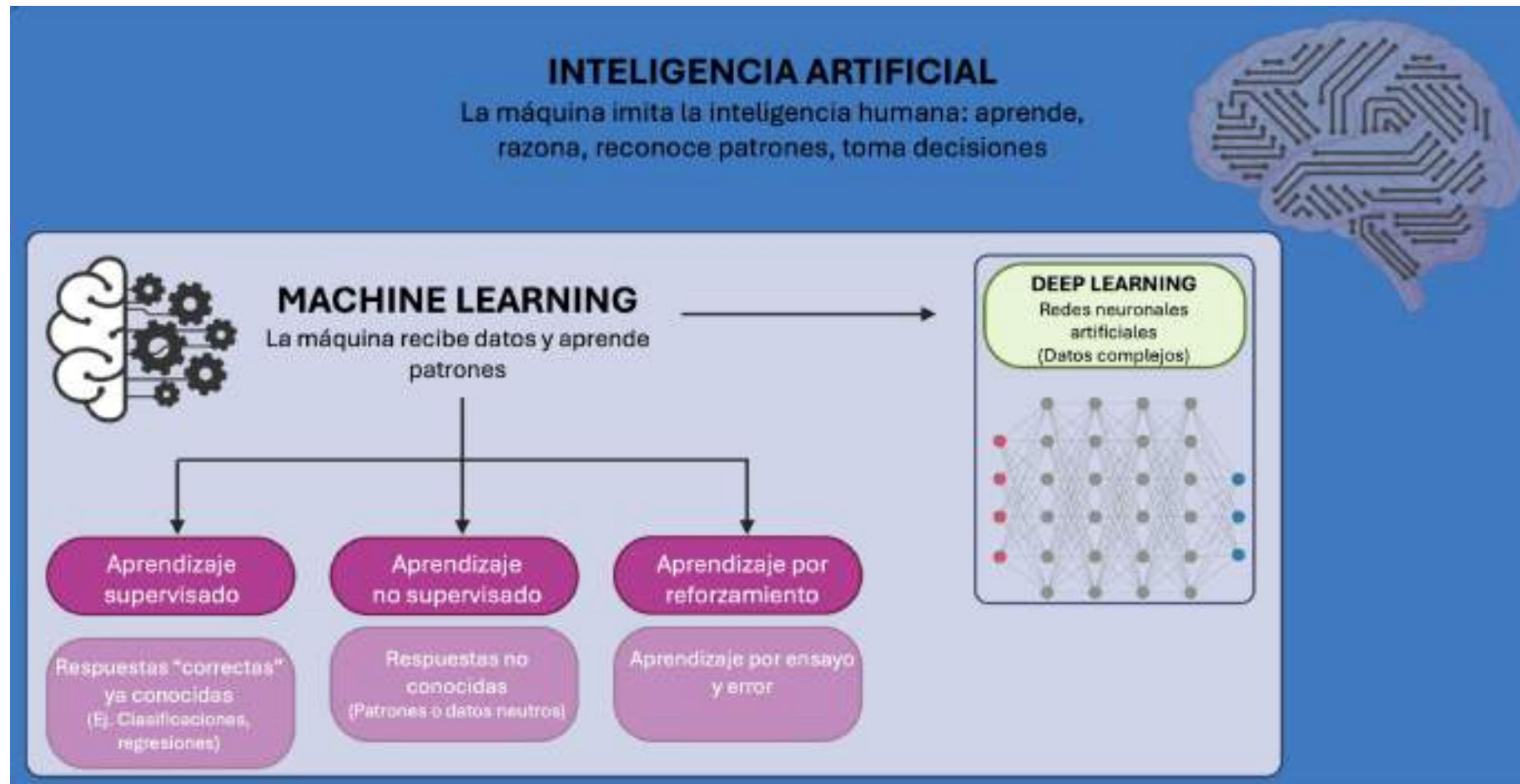
- Aumento progresivo de la prevalencia de alergia a nivel mundial
- Desafíos actuales en alergología:
  - Métodos: IgE, prick test, historia clínica
  - Limitaciones en sensibilidad/especificidad
  - Riesgo en pruebas de provocación oral



**INTELIGENCIA ARTIFICIAL**



# ¿Qué es la inteligencia artificial?



Allergy. 2023;78:2085–2088.



## ¿Cómo aprende la IA?

Ingreso

- Síntomas, IgE, historia clínica, genética

Proceso

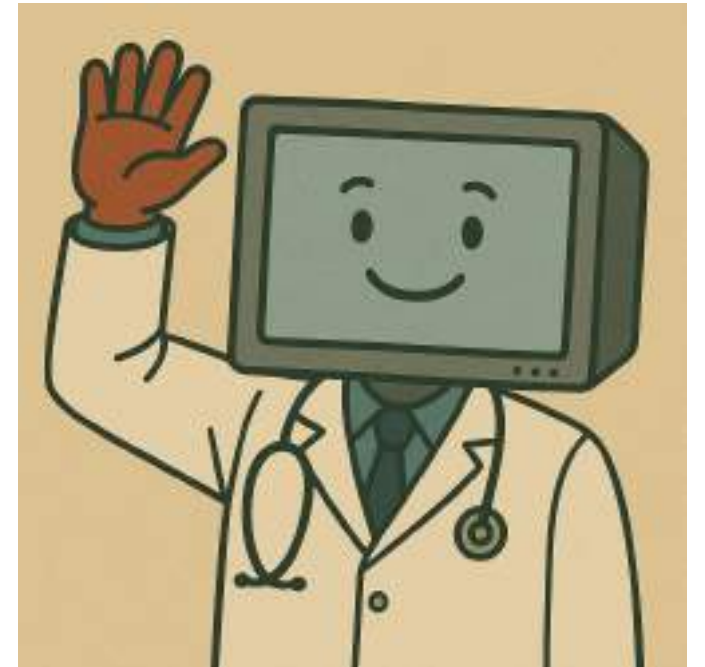
- Patrones aprendidos mediante algoritmos.

Salida

- Riesgo o diagnóstico predicho

## ¿Por qué utilizar la inteligencia artificial?

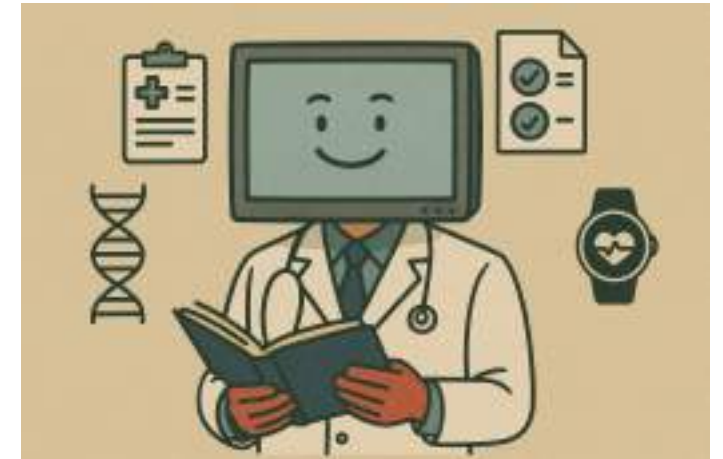
- Creación progresiva de grandes bancos de datos
- Revisión manual -> lenta y con riesgos de error
- IA → permite integrar múltiples biomarcadores y datos heterogéneos.

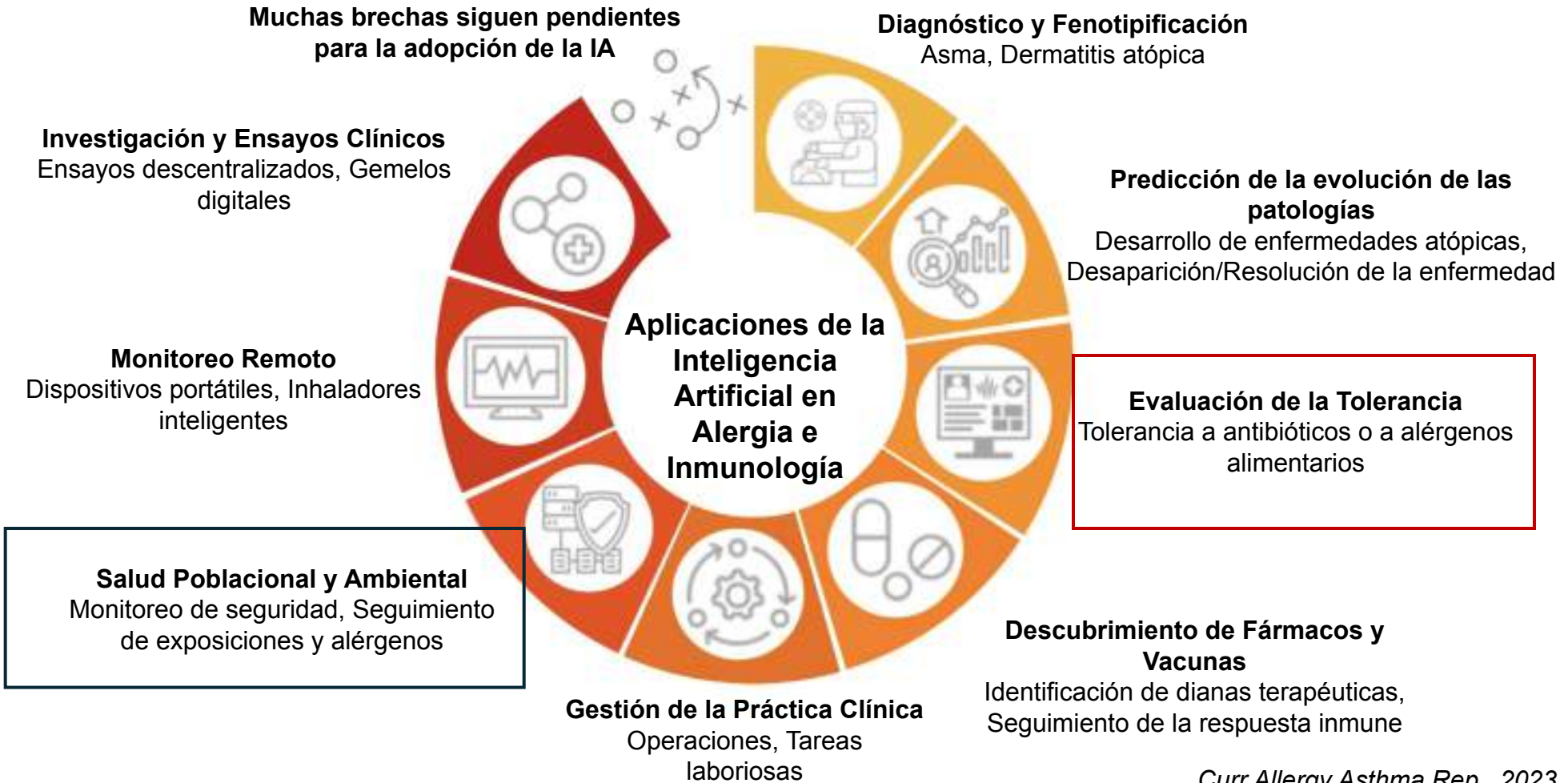




# “Ingreso”: fuentes de datos para modelos predictivos

- **Clínicos:** historia clínica, cuestionarios de calidad de vida, antecedentes familiares.
- **Pruebas de laboratorio:** IgE total, IgE específica, IgG4, triptasa, test de activación de basófilos.
- **Omics:** transcriptómica, metabolómica, epigenética (CpG methylation).
- **Imágenes / “salud digital”:** espirometría digital, fotoanálisis de lesiones cutáneas, sensores portátiles.





*Curr Allergy Asthma Rep . 2023 Jun;23(6)/351-362.*



# Modelos de IA en clínica

- **Árboles de decisión**
  - Random forests/gradient boosting
  - Ensembles (métodos combinados).
- **Métricas:**
  - AUC (área bajo la curva) -> capacidad global de un modelo de distinguir entre clases ( ej. alérgico vs no alérgico)
  - Calibración
  - Validez externa
- **Interpretabilidad:**
  - SHAP, LIME -> evaluación de las variables que influyen más
- **Problemas:**
  - Desbalance: datos representados no equitativamente
  - Sobreajuste: memorización de errores y desbalance

*Eur J Pediatr.* 2024;183(7):2757-65



# Comparativa métodos

## Métodos clásicos

Regresiones, exámenes diagnósticos (SPT/IgE con umbrales), scores predictivos (PENFAST)

## Métodos de IA

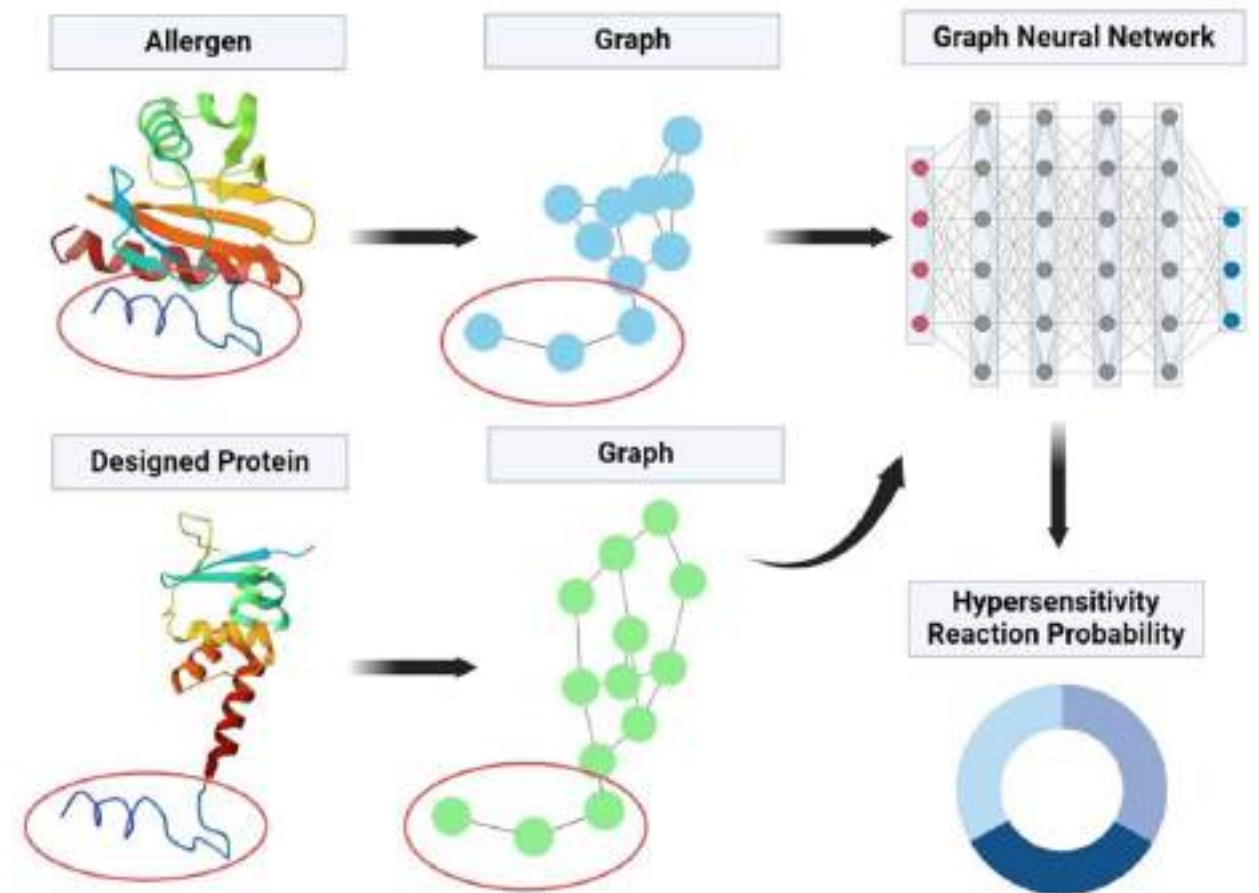
Random forest, gradient boosting, redes neuronales.

**Ventajas IA:** mayor precisión, integración de múltiples variables, detección de interacciones complejas.

**Limitaciones IA:** riesgo de sobreajuste, necesidad de **validación externa**, interpretabilidad -> difícil generalización

# Bioseguridad y vigilancia: prediciendo alergenicidad

- Bases de datos de alérgenos: uso de secuencias conocidas (AllergenOnline, Allergome).
- IA aplicada a secuencias de proteínas: identifica patrones estructurales comunes a alérgenos.





## AllergyPred: a web server for allergen prediction

Emanuel Kemmler<sup>1</sup>, Emma Katherine Fath<sup>2,3,4</sup>, Robert Preissner<sup>1</sup>, Priyanka Banerjee<sup>1,4,\*</sup> Nucleic Acids Research, 2025, 53, W4–W10

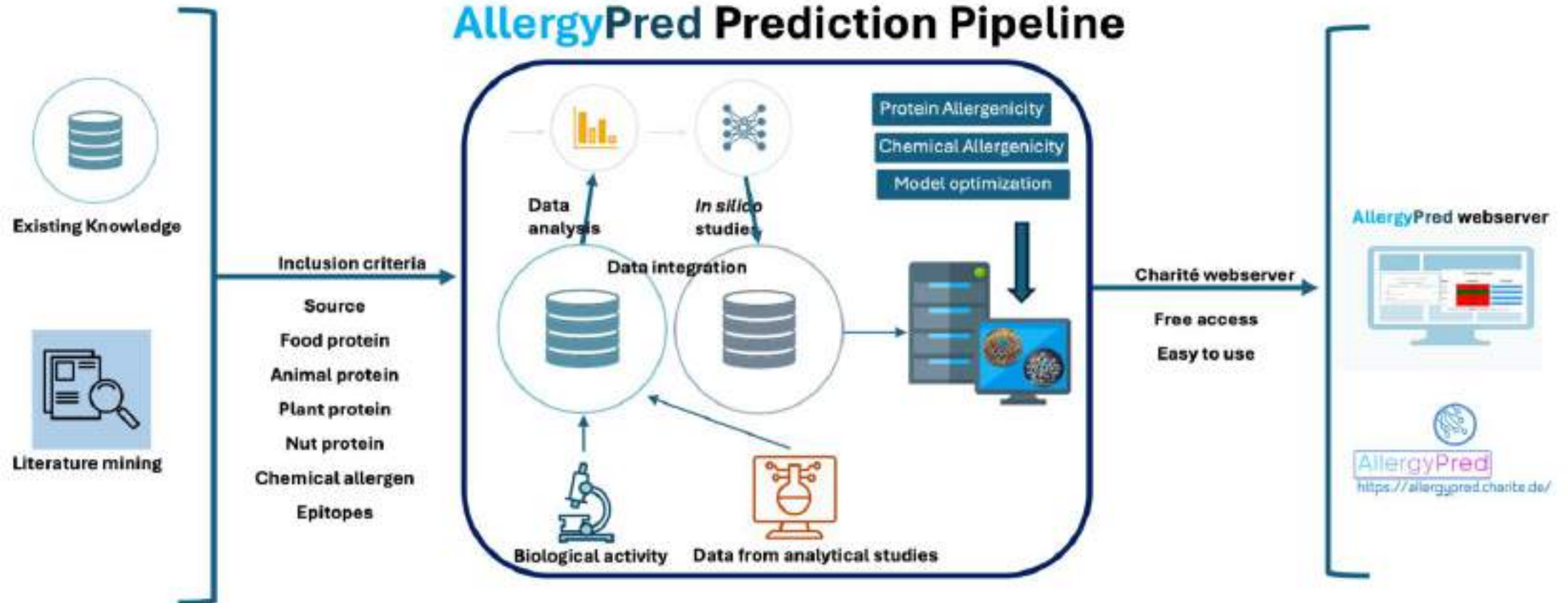
- Objetivo: Crear una plataforma abierta para predecir alergenicidad de moléculas biológicas (proteínas, péptidos) y químicas (fármacos, cosméticos, contaminantes).
- Modelos incluidos (5 en total):
  - Proteínas de alimentos
  - Proteínas animales
  - Proteínas vegetales
  - Proteínas de frutos secos
  - Moléculas químicas (haptenos, drogas, cosméticos)

Modelos de deep learning:

- Redes neuronales que reconocen motivos estructurales relacionados con la unión a IgE.



## Graphical abstract





**Table 1.** Performance evaluation on the CV set, validation set, and test set for the protein-induced allergenicity models reported in the AllergyPred platform

Protein Allergenicity model	Training set (CV)		Validation set		Test set		
	Accuracy	F1	Accuracy	F1	Accuracy	F1	MCC
Food-protein	0.99	0.99	0.98	0.93	0.97	0.88	0.87
Animal-derived	1.00	1.00	0.98	0.93	0.96	0.85	0.84
Plant-derived	0.99	0.99	0.97	0.94	0.96	0.89	0.87
Nut- derived	1.00	1.00	0.99	0.98	0.98	0.91	0.91

**Table 2.** Performance evaluation on the CV set and validation test set for the chemical-induced allergenicity models reported in the AllergyPred platform

Chemical Allergenicity model	5-fold CV set	Test set
Accuracy	0.84	0.82
Sensitivity	0.85	0.81
Specificity	0.83	0.83
F1	0.74	0.75

## Aplicación clínica/farmacéutica:

Evaluación de seguridad de nuevos alimentos procesados, transgénicos o fármacos biológicos.



Los resultados de la predicción de AllergyPred deben interpretarse como hipótesis, y no como conclusiones definitivas; se recomiendan validaciones adicionales en in vitro o verificación cruzada con otros modelos.



# Predicción de tolerancia: ¿el paciente es o no alérgico?

- Predicción de resultados de test de provocación oral (TPO).
- Severidad de reacciones.
- Delabeling de fármacos.
- Anafilaxia y SCARs.

*Front Allergy. 2024;5:1456585.  
Allergy Asthma Proc. 2025;46(1):3-8.*



# Alergia alimentaria: test de provocación oral

- TPO= *gold standart*, pero riesgoso.
- IA busca reducir TPO innecesarios.
- Variables: historia clínica, SPT, slgE.



*J Allergy Clin Immunol.* 2024;153(3):885-7  
*Diagnostics (Basel).* 2024;14(18):2069.



# Prediction of pediatric peanut oral food challenge outcomes using machine learning

Jonathan Gryak, PhD,<sup>a</sup> Aleksandra Georgievska, BS,<sup>a</sup> Justin Zhang, BS,<sup>b</sup> Kayvan Najarian, Rajan Ravikumar, MD,<sup>c</sup> Georgiana Sanders, MD MS,<sup>a,h</sup> and Charles F. Schuler IV, MD<sup>a,h</sup>  
*J Allergy Clin Immunol Global.* 2024;3:100252.

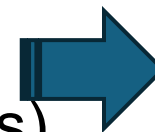
Objetivo: Mejorar la predicción de resultados de TPO en niños con sospecha de alergia a maní.

Cohortes:

LEAP (n=463, 52 reacciones, 11.2%).

IMPACT (n=140, 100% reacciones)

Universidad de Michigan (n=46, 17% reacciones)



Validación

Variables clínicas (15): edad, sexo, raza, SPT (pápula/eritema), IgE total, IgE específica a maní y sus componentes (Ara h 1–9).

Modelos evaluados: ensembles de Naïve Bayes, SVM, Random Forest, y LUCCK.



- Resultados:
  - LEAP: AUC 0.993, sensibilidad 98,6%, especificidad 100%.
  - Validación UM–IMPACT: AUC 0.871, sensibilidad 76.3%, especificidad 98%.
- Interpretabilidad: SHAP mostró que los predictores más importantes fueron tamaño del eritema, tamaño de la pápula y nivel de sIgE a maní (Ara h 2).
- Conclusión: ML pueden complementar los métodos clásicos, reduce la necesidad de TPO y mejora la confianza diagnóstica



## Pero...

- Tamaño limitado de cohortes
- Variabilidad de cohortes, diseño de estudio y prevalencia
- Desbalance de clases
  - LEAP solo 11% tuvo TPO (+)
- Generalización -> datos solo de EEUU
- Requiere validación externa prospectiva

# Alergia a fármacos

- Problema: sobrediagnóstico
- Inteligencia artificial y “*Large Language Model*” en sistemas de apoyo a la decisión clínica.
- Ejemplo: HELIOT.

## Objetivos:

- Reducir etiquetas falsas
- Generar recomendaciones clínicas
- Desetiquetado más seguro.



arXiv [Preprint]. 2024. arXiv:2409.16395.  
Allergy Asthma Proc. 2025;46(1):3-8.



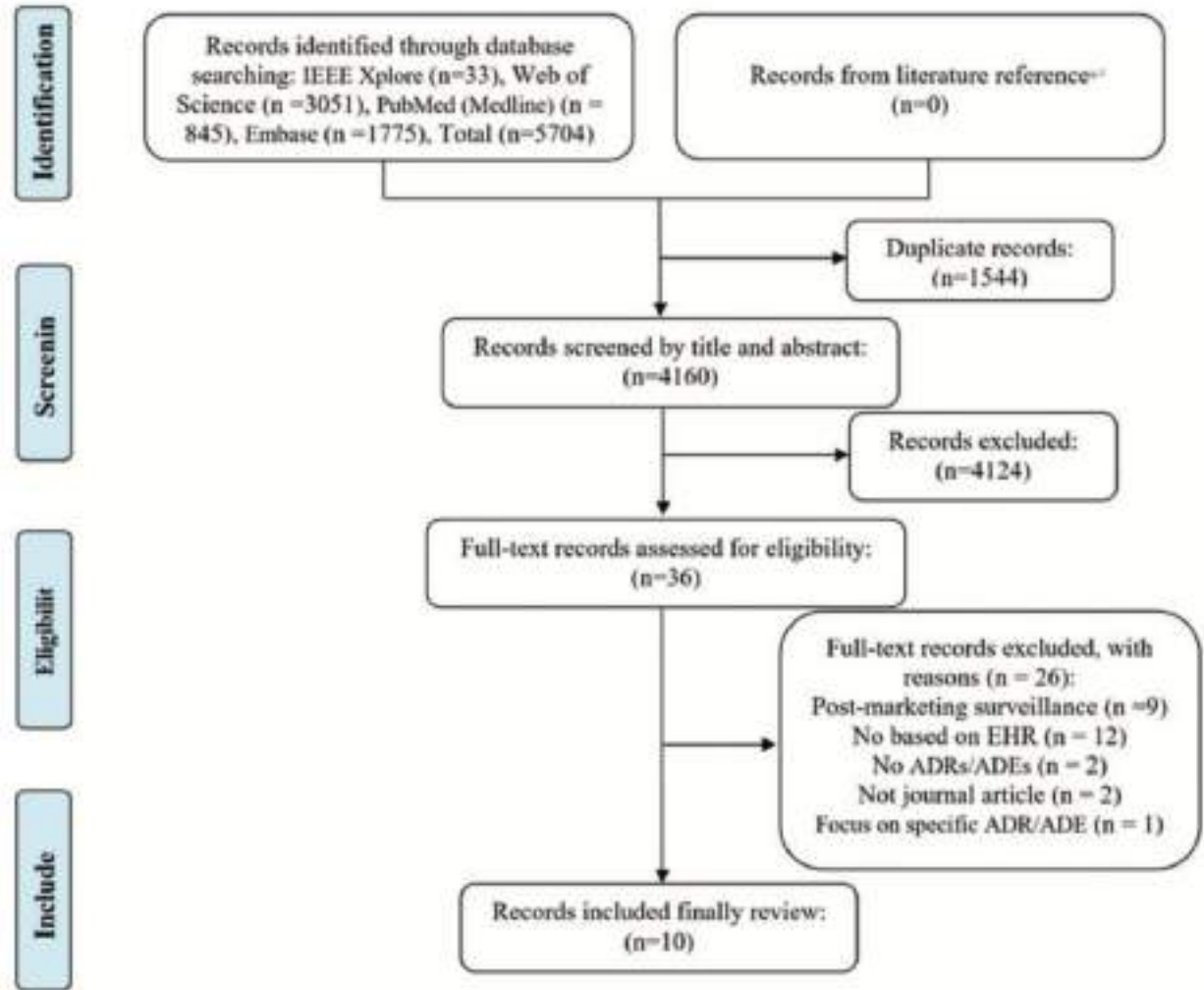
# Machine learning to predict adverse drug events based on electronic health records: a systematic review and meta-analysis

Qiaozhi Hu<sup>1,2,\*</sup>, Jiafeng Li<sup>3,\*</sup>, Xiaoqi Li<sup>2</sup>, Dan Zou<sup>1</sup>, Ting Xu<sup>1</sup> and Zhiyao He<sup>1,4</sup>

*Journal of International Medical Research* 2024, Vol. 52(12)  
1–19

Objetivo: Evaluar desempeño de modelos ML en predicción de reacciones adversas a fármacos (RAM) basados en historias clínicas electrónicas.

Algoritmos evaluados: Random Forest, Gradient Boosting, Redes Neuronales, SVM, XGBoost.



## Eventos adversos estudiados:

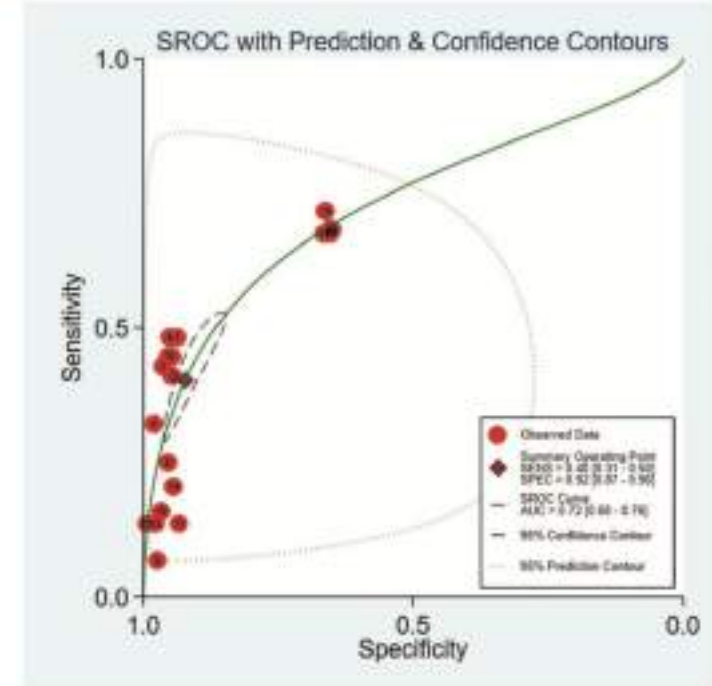
- Reacciones cutáneas severas (SCARs)
- Reacciones a antibióticos, opioides, anticoagulantes
- Efectos adversos cardiovasculares

## Resultados meta-analizados:

- AUC global promedio: 0.82 (IC95% 0.78–0.86)
- ML superó consistentemente a scores clínicos tradicionales.

## Limitaciones:

- Heterogeneidad de definiciones de RAM.
- Pocos estudios con validación externa.
- Necesidad de interpretabilidad (ej. SHAP, LIME).



## Conclusión

ML ofrece alta precisión en predicción de RAM, pero requiere estandarización y validación antes de adopción clínica rutinaria.



## **Predicting Penicillin Allergy: A United States Multicenter Retrospective Study**

*J Allergy Clin Immunol Global. 2024;3:100252.*

- **Objetivo:** Desarrollar un modelo de ML para predecir positividad en pruebas de alergia a penicilina usando datos retrospectivos de 4 hospitales en EE.UU.
- **Métodos**
  - Estudio retrospectivo
  - Cohorte: 4.777 pacientes con etiqueta de alergia a penicilina (PAL). (Clínica Mayo y Hospital General de Massachusetts)
  - Variables: características demográficas (sexo, edad, etnia, edad al momento de la reacción), historia de la reacción (tiempo de inicio, síntomas/signos (anafilaxia diagnosticada por un médico o el paciente), tratamiento y necesidad de atención médica, PENFAST, evaluaciones alérgicas (PT/IDR y/o provocación)
  - Outcome: skin test (+) y/o desafío oral (+).
- Algoritmos: regresión logística y modelos de ML (Gradient Boosted Machine, GBM).
- Interpretabilidad: uso de SHAP para identificar predictores claves.



## Clínica Mayo

- Usó un cuestionario clínico estructurado para caracterizar la reacción índice:
  - **Tipo de reacción** (urticaria, anafilaxia, exantema tardío, síntomas respiratorios, etc.).
  - **Tiempo desde la reacción**
  - **Tiempo de latencia** (inmediata  $<1h$ , tardía  $>1h$ ).
  - **Necesidad de atención médica** (urgencia, hospitalización).
  - **Historia de reacciones múltiples.**



# Hospital General de Massachusetts

- Recolectó información más **simplificada** desde el FCE:
  - Reacción **cutánea/no cutánea**.
  - **Gravedad** (leve, moderada, grave).
  - **Año aproximado** de la reacción.



## Resultados principales

- Positividad confirmada: 11%
- GBM: AUC 0.67, S:2,3%, E:99.2%, VPP: 50% VPN: 75%

TABLE II. Model performance of the gradient-boosted model compared with PEN-FAST in predicting a positive allergy test in all datasets

Dataset/model	No. (%) positives	Performance						Threshold*
		AUC (95% CI)	Accuracy (95% CI) (%)	Sensitivity (95% CI) (%)	Specificity (95% CI) (%)	PPV (95% CI) (%)	NPV (95% CI) (%)	
<b>Mayo enriched test (n = 170)</b>								
Best GBM	43 (25)	0.671 (0.574-0.768)	74.7 (67.5-81.0)	2.3 (0.1-12.3)	99.2 (95.7-100.0)	50.0 (1.3-88.7)	75.0 (67.7-81.3)	0.16
PEN-FAST	43 (25)	0.517 (0.426-0.608)	65.3 (57.6-72.4)	23.3 (11.8-38.6)	79.5 (71.5-86.2)	27.8 (14.2-45.2)	75.4 (67.2-82.4)	0.60
<b>Mayo enriched training (n = 1526)</b>								
Best GBM	388 (25)	0.725 (0.696-0.754)	75.2 (72.9-77.3)	2.8 (1.4-5.0)	99.8 (99.4-100.0)	84.6 (54.6-98.1)	75.1 (72.8-77.2)	0.16
PEN-FAST	388 (25)	0.582 (0.551-0.613)	68.9 (66.5-71.2)	31.4 (26.9-36.3)	81.6 (79.3-83.8)	36.9 (31.6-42.3)	77.7 (75.3-80.1)	0.60
<b>Mayo nonenriched test (n = 2234)</b>								
Best GBM	45 (2)	0.624 (0.541-0.708)	97.8 (97.1-98.4)	2.2 (0.1-11.8)	99.8 (99.3-99.9)	16.7 (0.4-64.1)	98.0 (97.4-98.6)	0.16
PEN-FAST	45 (2)	0.580 (0.490-0.671)	79.0 (77.3-80.7)	35.6 (21.9-51.2)	79.9 (78.2-81.6)	3.5 (2.0-5.6)	98.4 (97.7-98.9)	0.60
<b>MGH nonenriched test (n = 847)</b>								
Best GBM	37 (4)	0.590 (0.500-0.680)	95.6 (94.0-96.9)	0.0 (0.0-9.5)	100.0 (99.5-100.0)	NA	95.6 (94.0-96.9)	0.16
PEN-FAST	37 (4)	0.560 (0.465-0.656)	79.1 (76.2-81.8)	29.7 (15.9-47.0)	81.4 (78.5-84.0)	6.8 (3.4-11.8)	96.2 (94.5-97.5)	0.60

TABLE III. Model performance based on "unknown" variables of the index reaction

Dataset/model	Unknown count	Performance						F1 score
		AUC (95% CI)	Accuracy (95% CI) (%)	Sensitivity (95% CI) (%)	Specificity (95% CI) (%)	PPV (95% CI) (%)	NPV (95% CI) (%)	
<b>Mayo enriched test (n = 170)</b>								
	0	0.870 (0.725-1.000)	74.4 (57.9-87.0)	9.1 (0.2-41.3)	100.0 (87.7-100.0)	100.0 (2.5-100.0)	73.7 (56.9-86.6)	16.7000
	1	0.658 (0.534-0.783)	73.5 (63.6-81.9)	0.0 (0.0-13.7)	98.6 (92.6-100.0)	0.0 (0.0-97.3)	74.2 (64.3-82.0)	NA
	2	0.553 (0.287-0.818)	76.0 (54.9-90.6)	0.0 (0.0-45.9)	100.0 (82.4-100.0)	NA	76.0 (54.9-90.6)	NA
	3	NA	87.5 (47.3-99.7)	0.0 (0.0-97.5)	100.0 (59.0-100.0)	NA	87.5 (47.3-99.7)	NA

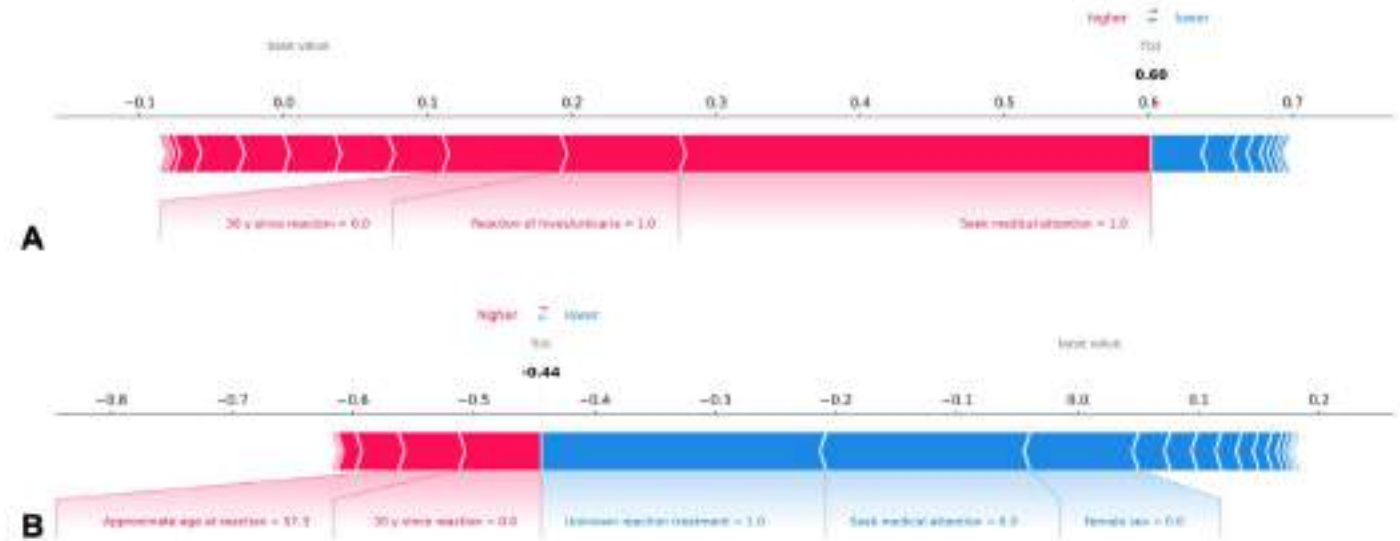
- En casos con **datos completos**, AUC mejoró a 0.87.

### Factores predictivos positivos (SHAP):

- Reacción en el último año.
- Reacción que requirió atención médica.
- Sexo femenino.
- Presencia de urticaria/habones.

### Factores protectores:

- Reacción hace >30 años.
- Inicio retardado (>1 h).



**FIGURE 3.** (A) SHAP force plot of a patient with confirmed penicillin allergy: increasing absolute magnitude of SHAP values (positive or negative) is used to visualize the increasing impact of features on the model direction. Features in red push the model toward the presence of a positive immediate penicillin allergy skin test. (B) SHAP force plot of a patient with disproven penicillin allergy: increasing absolute magnitude of SHAP values (positive or negative) is used to visualize the increasing impact of features on the model direction. Features in blue push the model toward the absence of a positive immediate penicillin allergy skin test. SHAP, Shapley Additive exPlanations.



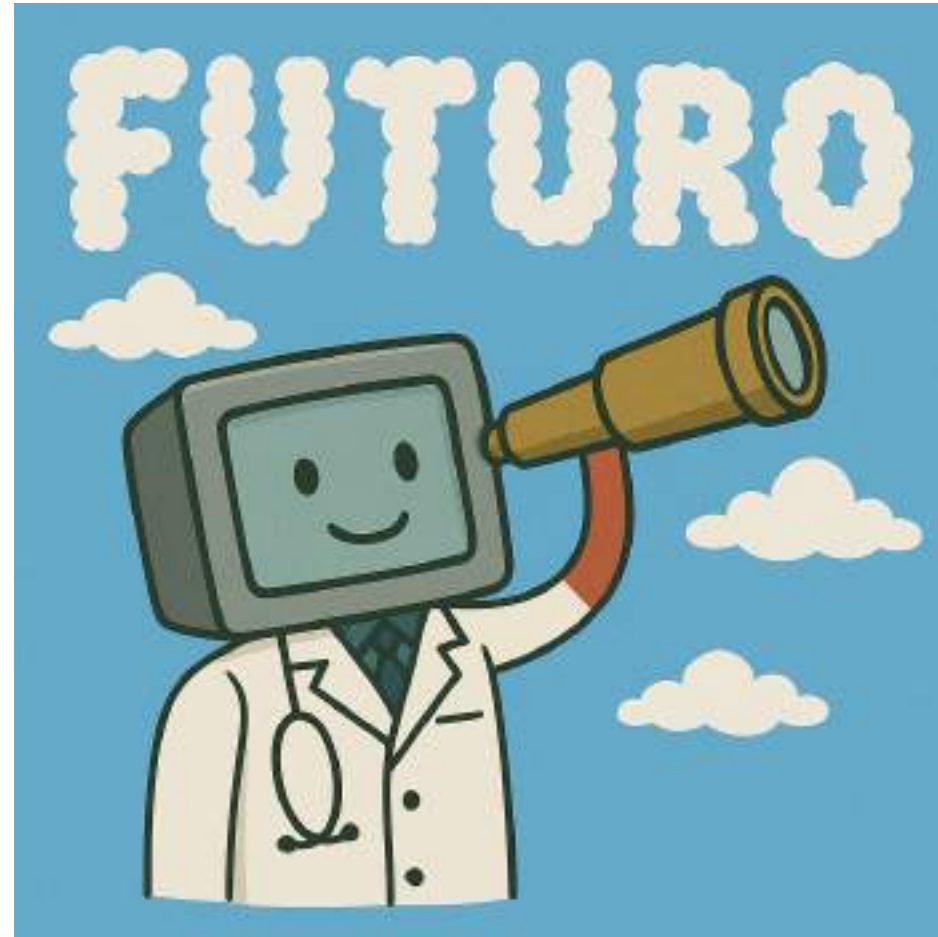
## Conclusión:

- ML aún no logra predicción robusta para uso clínico generalizado.
- Se necesita:
  - Mejor documentación de la reacción índice en registros clínicos.
  - Consenso internacional sobre variables mínimas a recolectar.
  - A pesar de limitaciones, se identifican nuevos predictores clínicos (ej. sexo femenino, atención médica requerida).





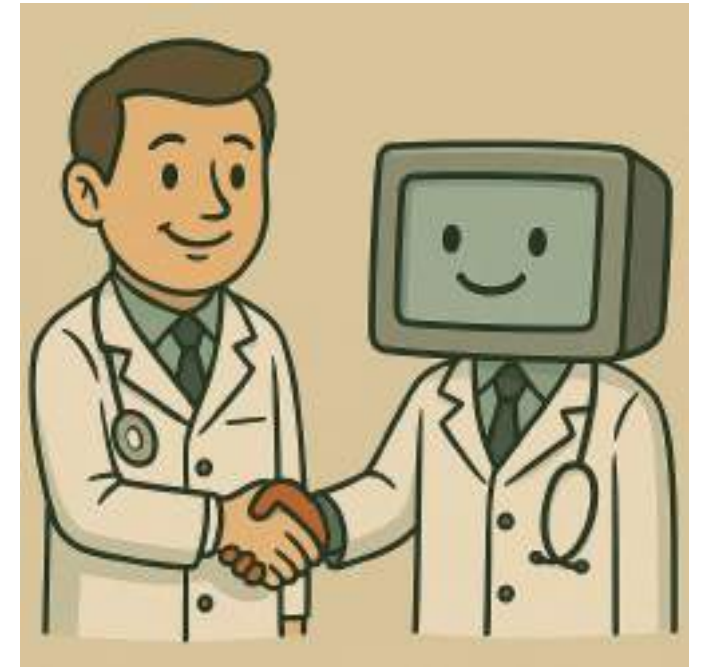
# ¿Qué nos depara el futuro?





# Integración clínica

- Sistema de apoyo en la FCE.
  - Predicción en tiempo real.
  - Variables principales conocidas
- Rol del clínico -> toma de decisión final



*Front Med (Lausanne). 2025;12:1523902.  
Ann Allergy Asthma Immunol. 2025;136(1):3-11.*



# Desafíos actuales

- Modelos dependen de calidad de datos.
  - Registros incompletos o variables
- Desbalance de clases
  - Alérgicos confirmados son minoría, y la mayoría de las etiquetas son falsamente positivas -> tendencia a diagnosticar “no alérgicos”.
- Falta de validación externa
- Explicabilidad -> comprensibles para el clínico y con fundamentos justificables: ¿por qué se hace la predicción?
- Dificultad para generalizar entre poblaciones
  - Validación multicéntrica



*Eur J Pediatr.* 2024;183(7):2757-65  
*Front Med (Lausanne).* 2025;12:1523902.



## ¿Qué debe cumplir un buen modelo?

- Validación externa en otras cohortes.
- Curvas de decisión.
- Métricas más allá de AUC.
- **Evaluación de impacto clínico.**



*Eur J Pediatr.* 2024;183(7):2757-65



## Métricas de impacto

- Reducción de TPO innecesarios.
- Disminución falsos negativos.
- Diagnósticos más rápidos
- Satisfacción paciente/clínico.



*Eur J Pediatr.* 2024;183(7):2757-6



# Riesgos y fallas

- Drift temporal
  - Degradación con el tiempo por sobreajuste a datos locales
  - Mantenimiento insuficiente.
- Falta de confianza del clínico si el modelo no es explicable o falla en casos concretos.



*Eur J Pediatr.* 2024;183(7):2757-65  
*Front Med (Lausanne).* 2025;12:1523902



# Conclusiones

- Puede aumentar la seguridad y eficiencia
  - Necesaria validación rigurosa.
  - Colaboración multidisciplinaria.
  - Implementación responsable
- Perspectivas futuras.
  - Mejora del registro en la ficha clínica
  - Desarrollo de modelos que se puedan generalizar
    - Trabajo multicéntrico y prospectivo
- IA potencia al médico, no lo reemplaza
  - “La decisión final depende del médico y no de la máquina”

*Front Med (Lausanne). 2025;12:1523902  
Ann Allergy Asthma Immunol. 2025;136(1):3-11.*



**Muchas gracias**

[fsidgman@hsalvador.cl](mailto:fsidgman@hsalvador.cl)