

La voz del paciente como herramienta diagnóstica: análisis de texto libre en enfermedades reumáticas

Inés Pérez – Sancristóbal¹, Nils Steinz², Ling Qin², Tjardo Maarseveen², Floor Zegers^{2,3}, Barbara Bislawska Axnäs⁴, Esther Toledano – Martínez⁵, Luis Rodríguez – Rodríguez⁶ and Rachel Knevel^{2,7}

1 Universidad Complutense de Madrid, Facultad de Medicina, Madrid, España 2 Leiden University Medical Center, Departamento de Reumatología, Leiden, Países Bajos, 3 Leiden University Medical Center, Departamento de Ciencia de Datos Biomédicos, Leiden, Países Bajos, 4 Elsa Science, 5 Hospital Clínico San Carlos, Departamento de Reumatología, Madrid, España, 6 Grupo de Patología Musculoesquelética, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria (IdISSC), Madrid, España, 7 Translational and Clinical Research Institute, Newcastle University, Newcastle Upon, Reino Unido

Introducción:

Los síntomas musculoesqueléticos suponen un reto diagnóstico en reumatología debido a la superposición de manifestaciones clínicas. Aunque herramientas como la anamnesis tradicional y encuestas estructuradas son fundamentales, pueden omitir información crucial y retrasar el diagnóstico temprano en condiciones como osteoartritis (OA), fibromialgia (FM) y enfermedades reumáticas inmunomediadas (imRD). Además, las discrepancias entre la descripción de los pacientes y la interpretación clínica pueden afectar la precisión diagnóstica.

Este estudio evaluó si los síntomas reportados pueden predecir enfermedades reumáticas mediadas inmunológicamente (imRD), osteoartritis (OA) y fibromialgia (FM), y si la predicción varía entre pacientes diagnosticados y no diagnosticados.

Métodos:

Usando técnicas de procesamiento del lenguaje natural (NLP) y aprendizaje automático (ML), analizamos las respuestas en texto libre de pacientes obtenidas de la encuesta "Rheumatic@" entre junio de 2021 y febrero de 2023. Los datos se dividieron en conjuntos de entrenamiento (66%) y validación (33%). Se testaron 8 modelos de ML utilizando vectorización *Term Frequency-Inverse Document Frequency* (TF-IDF) y validación cruzada (2x5), seleccionando el óptimo según el área bajo la curva (AUC-ROC). Este fue optimizado y aplicado a diagnósticos existentes y nuevos. Los puntos de corte óptimos se identificaron mediante gráficos de dispersión *Swarm plots*, maximizando el valor predictivo positivo (VPP) y la especificidad en OA y FM, y la sensibilidad y el valor predictivo negativo (VPN) en imRD.

Tabla 1. Características sociodemográficas básicas de la población estudiada. **Verdaderos negativos = pacientes que no reportaron ninguno de los diagnósticos de los 3 grupos de enfermedad.*

Características	Todos	OA	FM	imRD	Verdaderos negativos*
N	8454	3930 (46%)	1457 (17%)	1747 (21%)	2393 (28%)
Mujeres (%)	6578 (78%)	3135 (80%)	1355 (93%)	1228 (70%)	1881 (79%)
Grupo de edad					
< 40 (%)	767 (9%)	70 (2%)	155 (11%)	124 (7%)	359 (15%)
40-50 (%)	1409 (17%)	400 (10%)	314 (22%)	233 (13%)	538 (22%)
50-60 (%)	3094 (37%)	1461 (37%)	602 (41%)	535 (31%)	901 (38%)
60 + (%)	3184 (38%)	1999 (51%)	386 (26%)	855 (49%)	595 (25%)

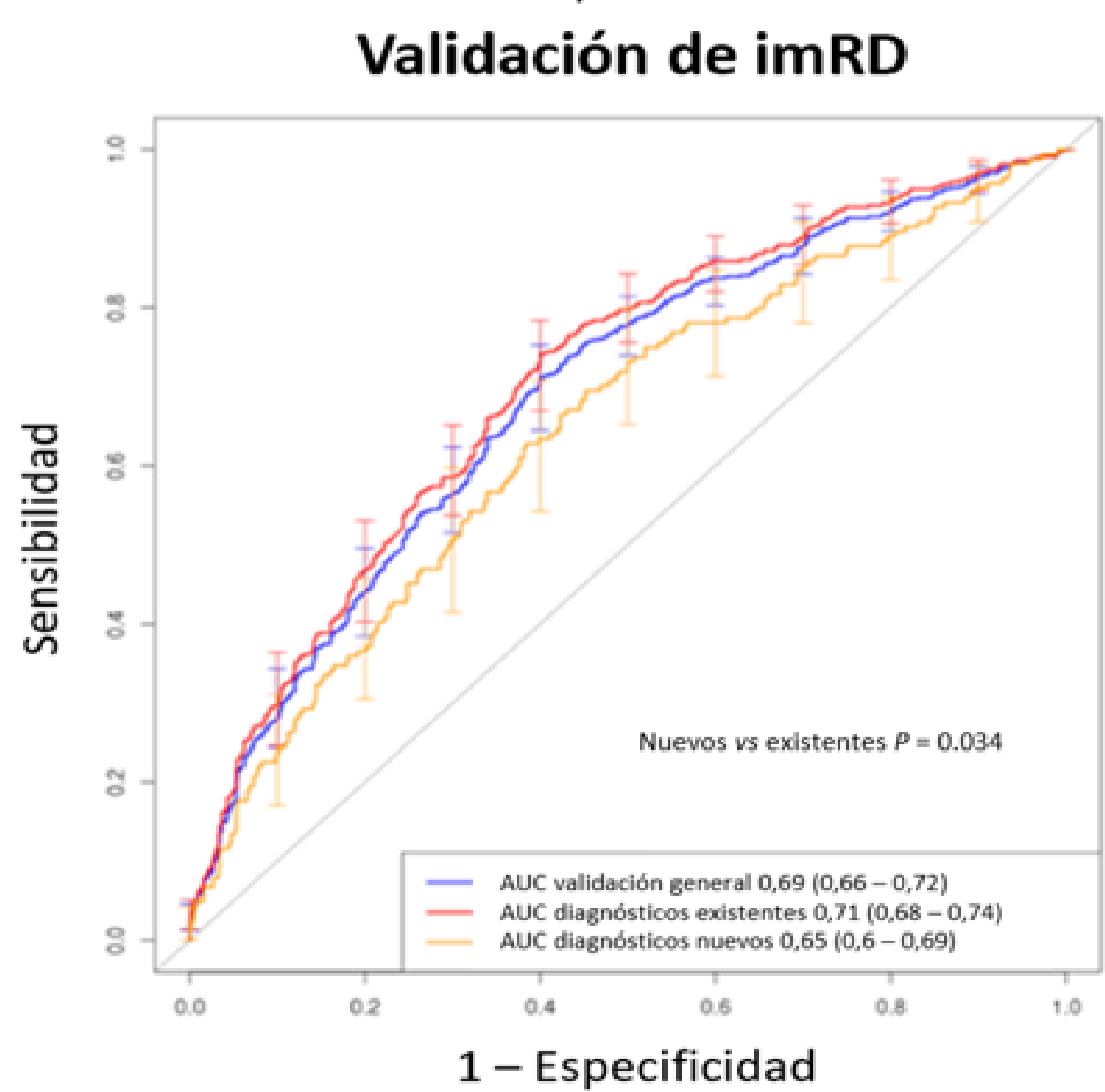
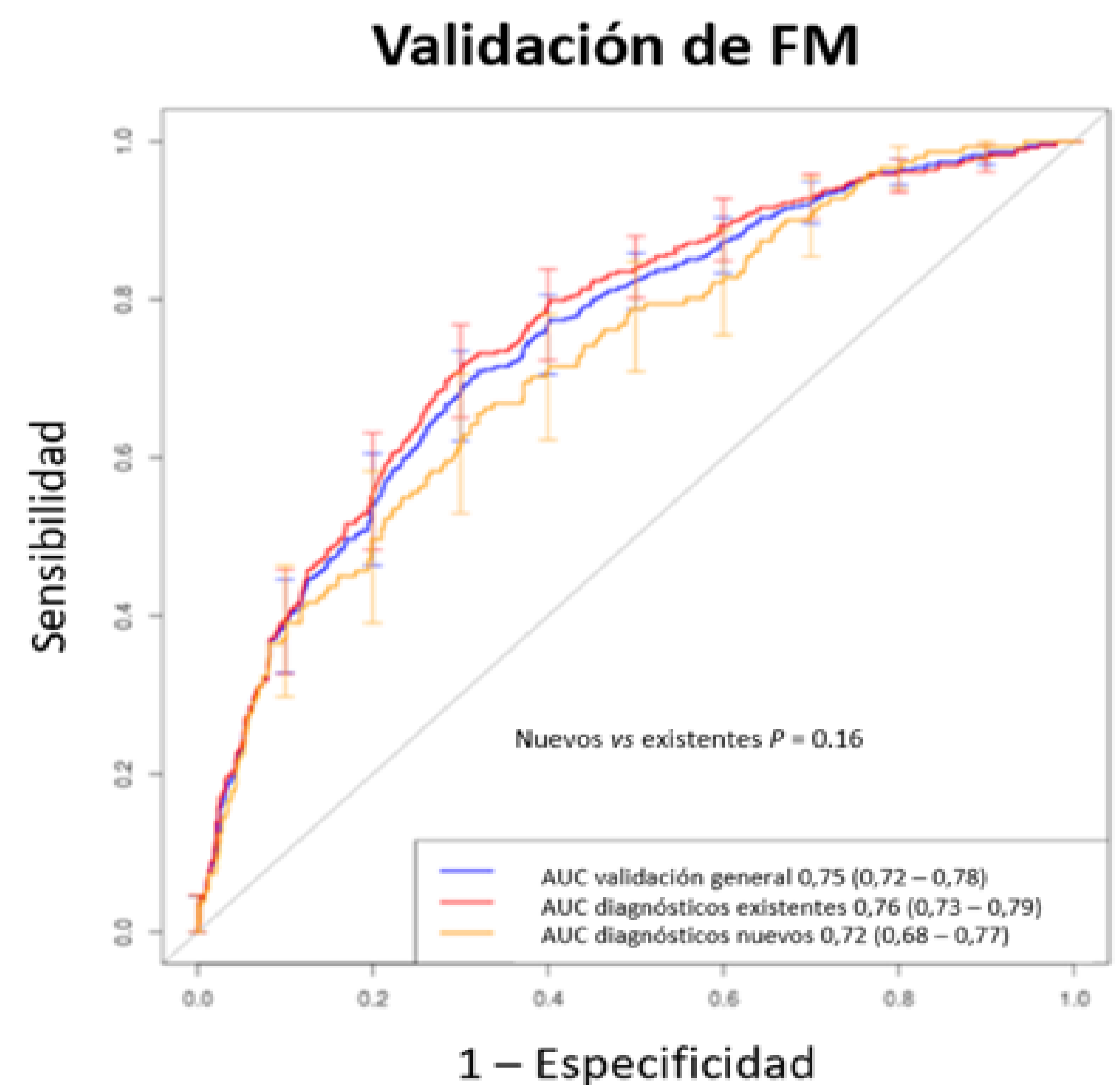
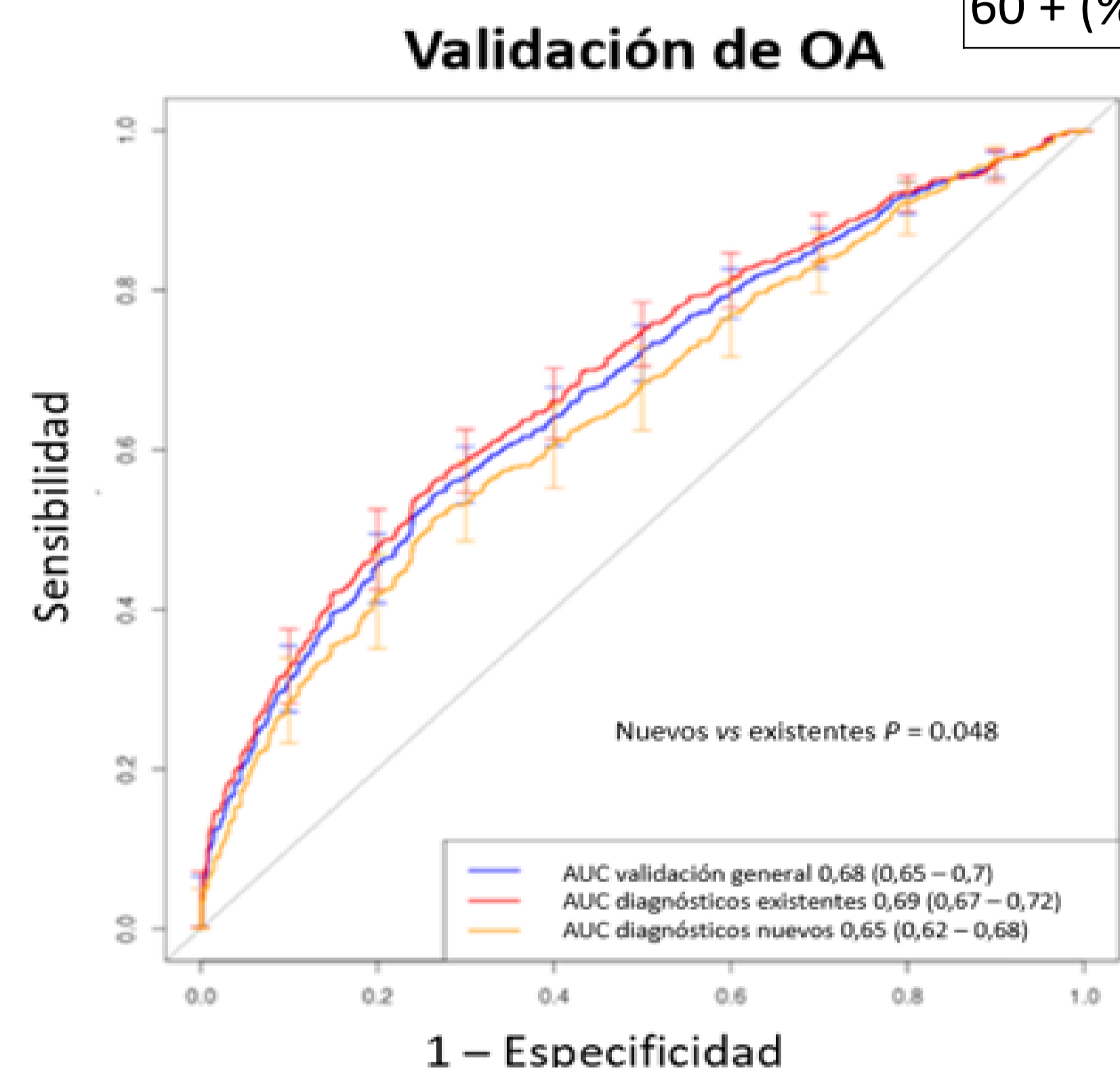


Figura 1. Resultados de los modelos optimizados en los distintos escenarios sobre el conjunto de validación.

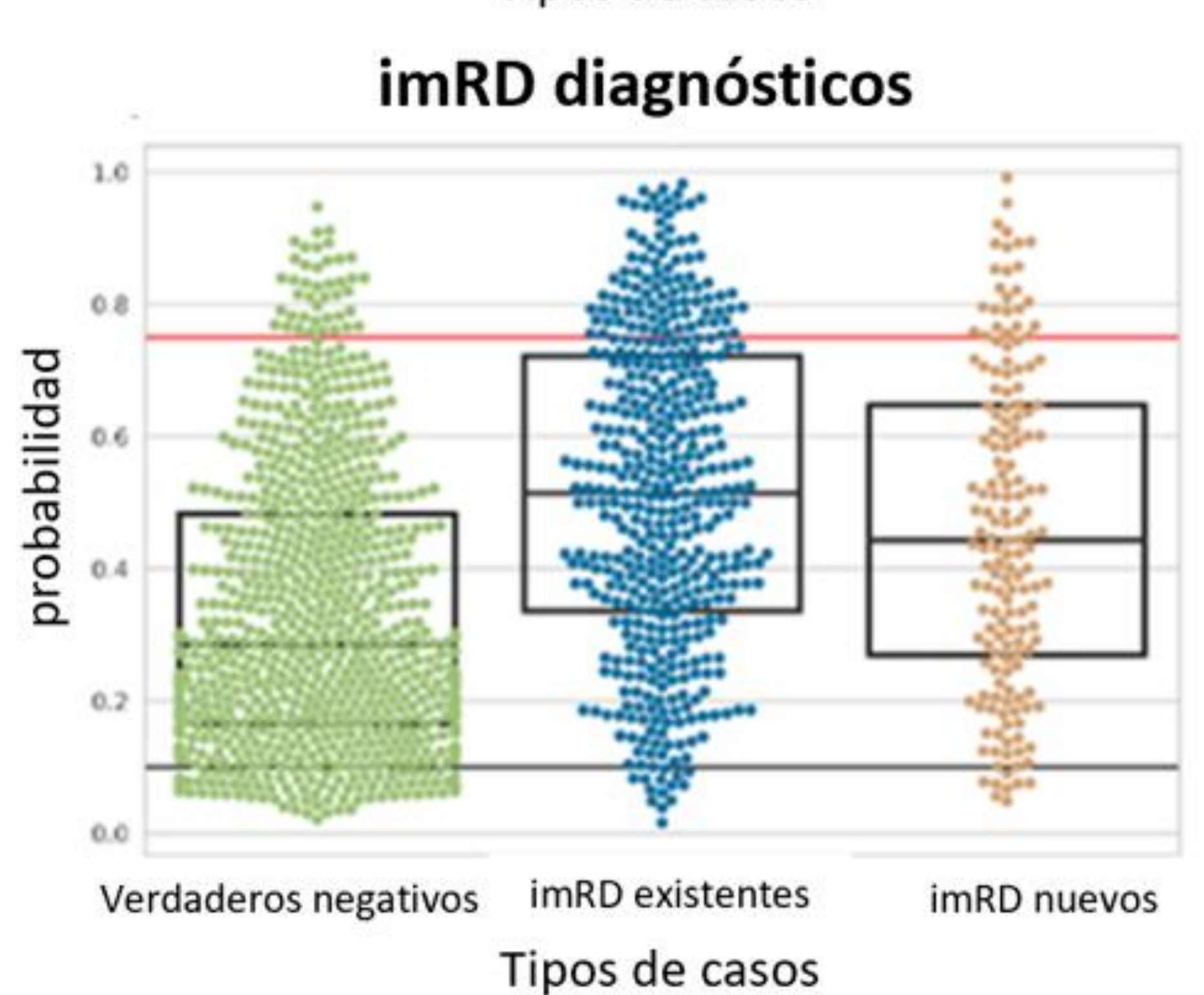
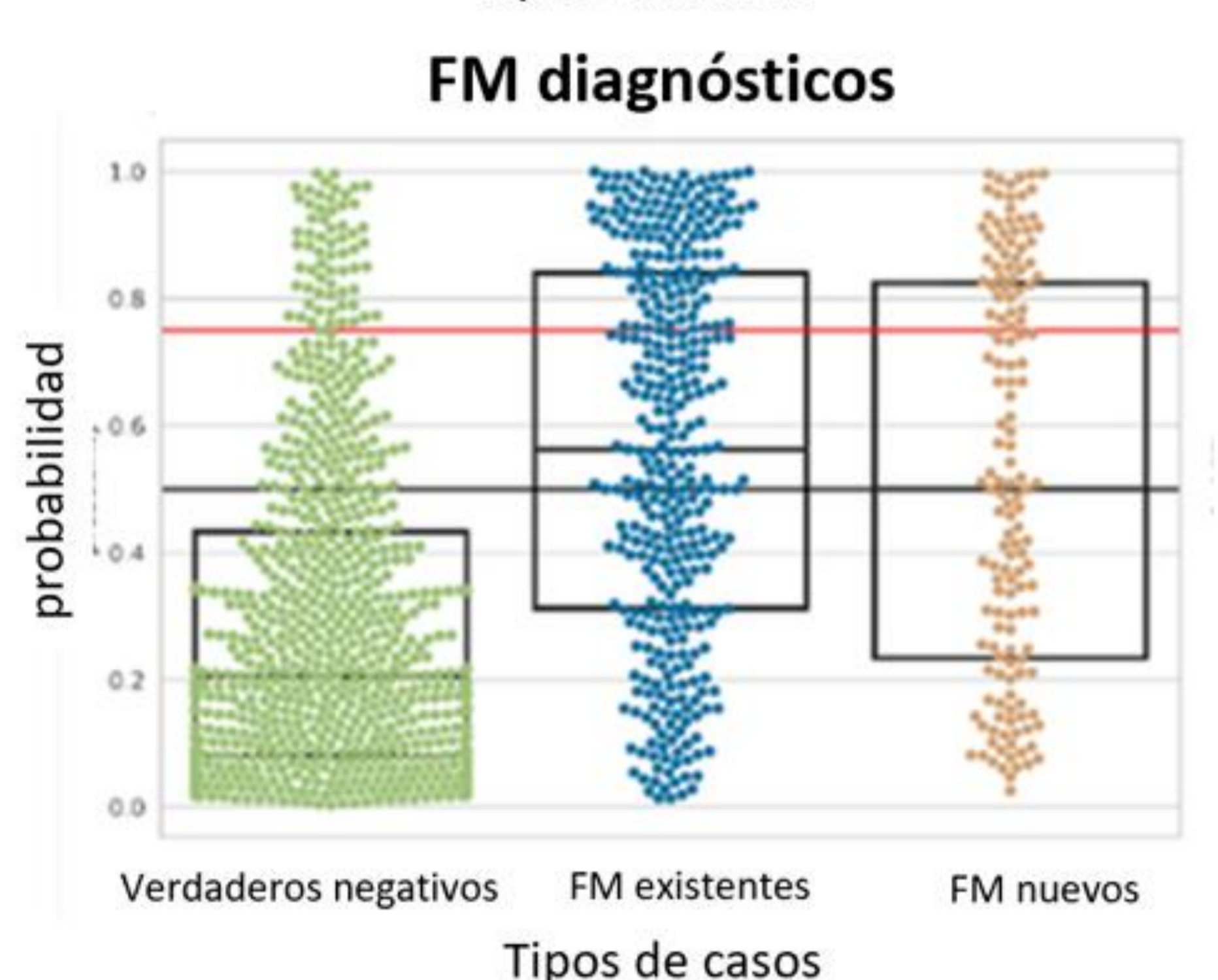
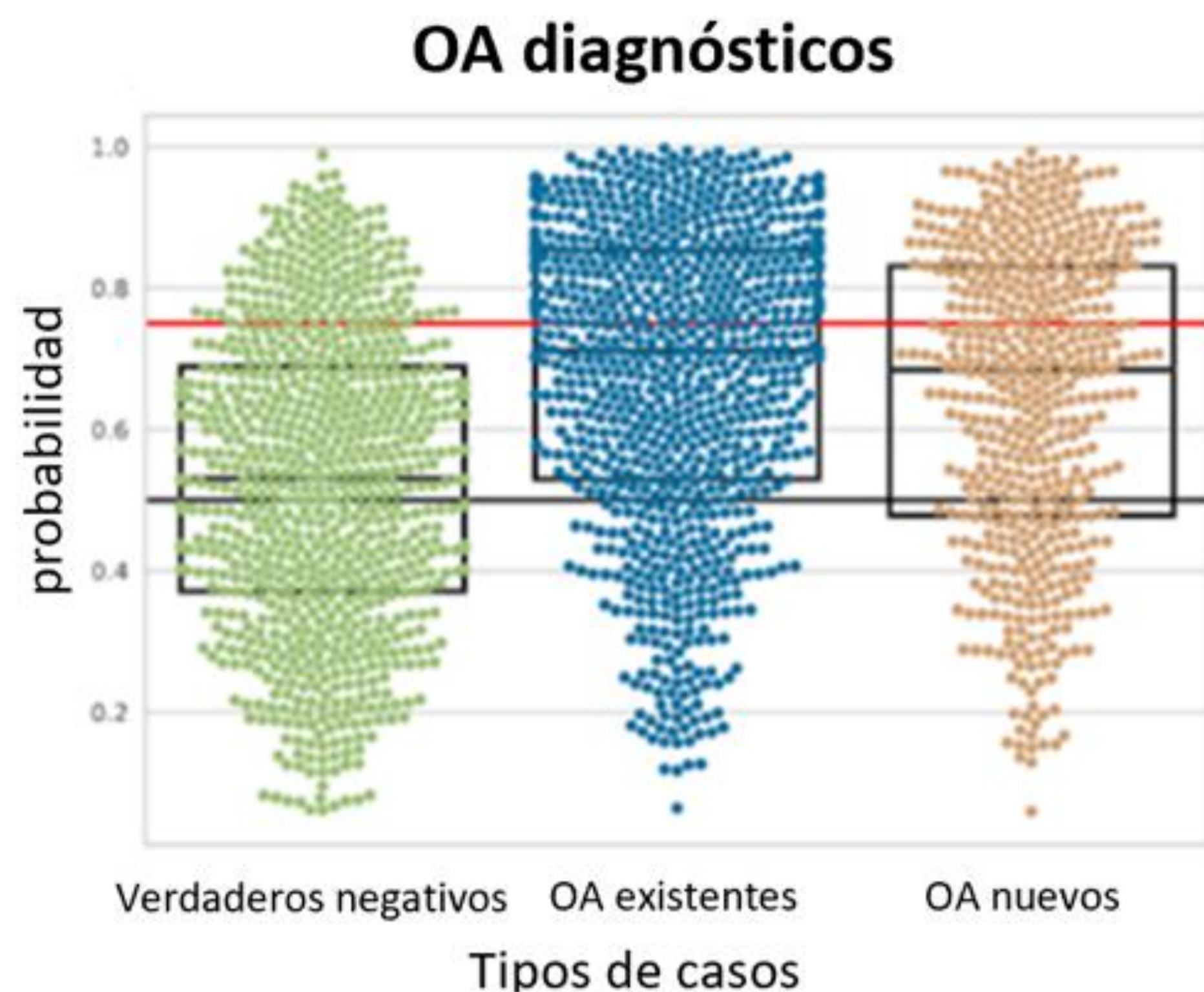


Figura 2. Swarm plots que representan la probabilidad de ser un caso negativo o positivo en el conjunto de datos de validación.

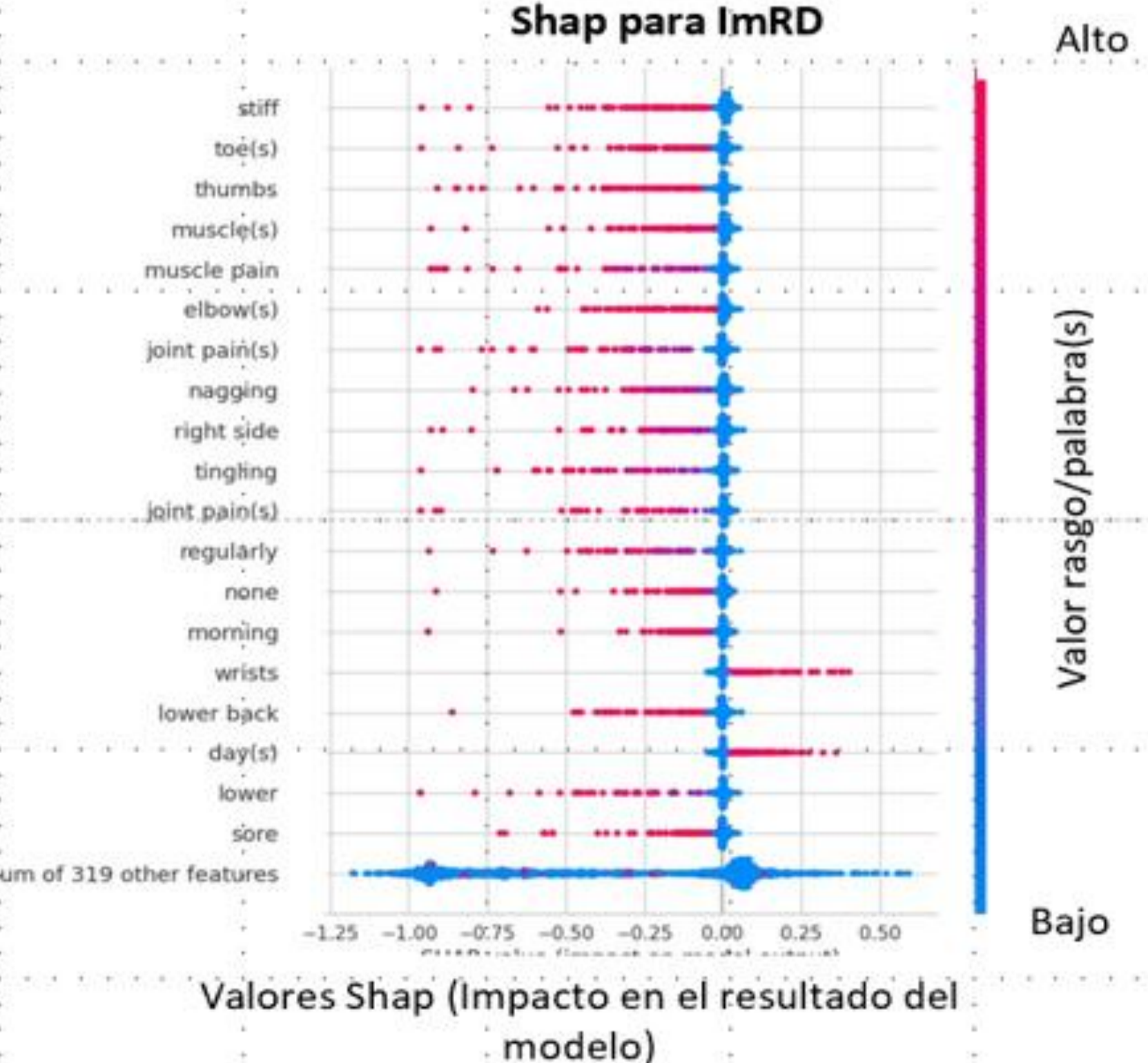
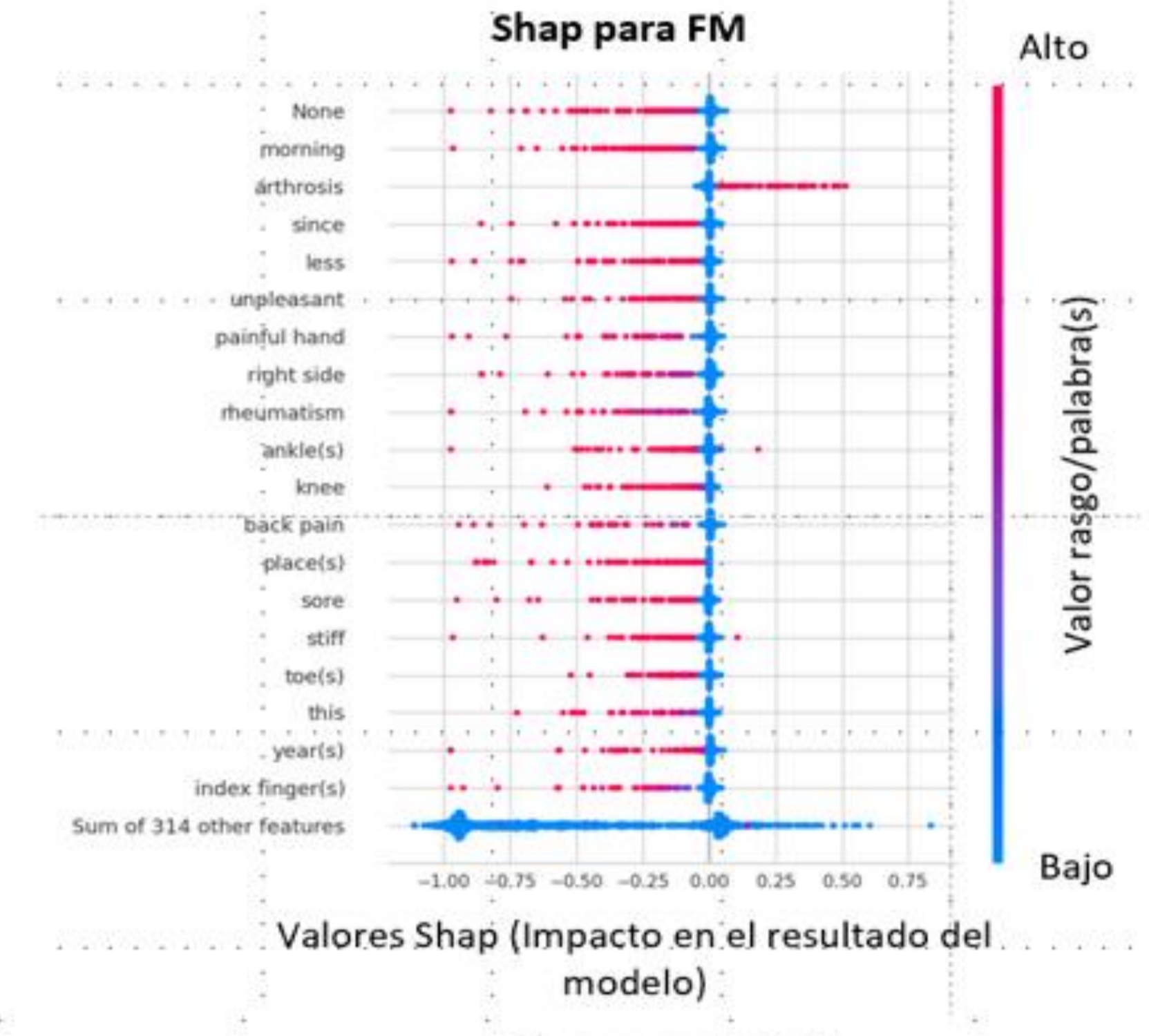
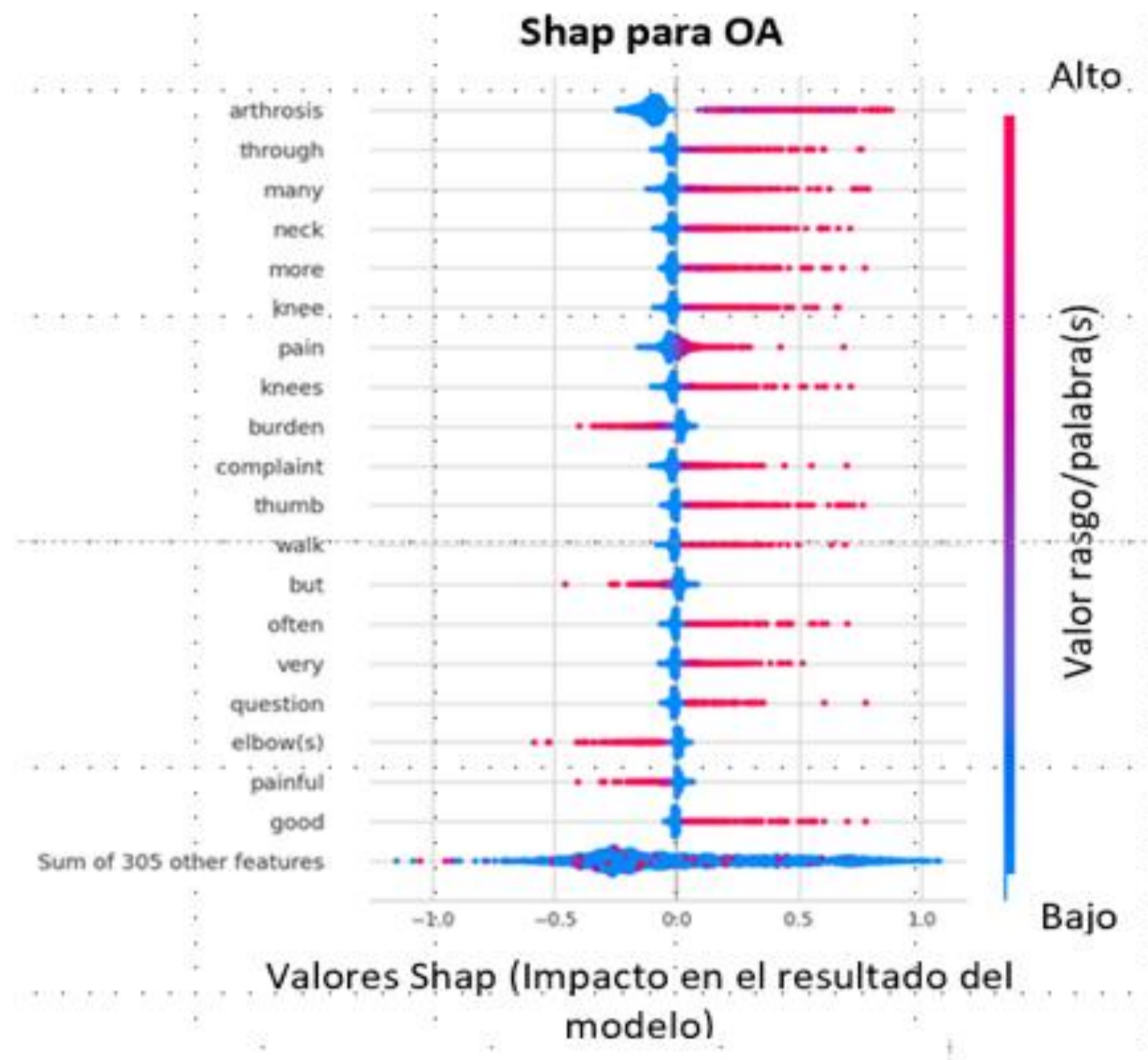


Figura 3. Valores de SHAP para las palabras discutidas en los tres grupos de enfermedades.

Resultados:

- Los algoritmos tipo *Support Vector Machine (SVM)* lograron el mejor rendimiento.
- **AUC diagnósticos existentes: OA 0.69, FM 0.76, imRD 0.71; nuevos diagnósticos: OA 0.65, FM 0.72, imRD 0.65.**
- En imRD, el corte: **0.75** priorizó **especificidad** (19% derivaciones directas, 5% falsos positivos) y uno de **0.15** priorizó **sensibilidad** (92% derivaciones generales, 80% falsos positivos).
- Los valores explicativos (*Shapley additive explanations, SHAP*) mostraron que palabras específicas favorecen la predicción de OA, y la ausencia de algunas palabras fue clave en FM e imRD.

Conclusiones:

- El análisis de texto libre mediante NLP y ML muestra gran potencial para distinguir OA, FM e imRD y optimizar el proceso diagnóstico.
- Aunque no son totalmente confiables para un diagnóstico autónomo por ahora, su integración en herramientas de IA es prometedora.
- La colaboración entre ingenieros, informáticos y clínicos es clave para desarrollar herramientas diagnósticas precisas y transparentes.



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101059522. Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor the granting authority can be held responsible for them.

SPIDERR

Detalles de contacto: ineper10@ucm.es

