Arch Argent Pediatr. 2026;124(4):e202510882

Reducción de la entropía y puntaje F: aplicaciones de la teoría de la información en la evaluación de pruebas diagnósticas

Eduardo Cuestas[1,2,3], María E. Cieri[2], L. Johana Escobar Zuluaga[2], María M. Ruiz Brünner[2]

RESUMEN

Este trabajo propone un enfoque desde la teoría de la información y el aprendizaje automático, aplicando el puntaje F y la reducción de la entropía para analizar el flujo informativo entre el paciente, la prueba diagnóstica utilizada para evaluar la enfermedad y los evaluadores que interpretan los resultados, para enriquecer la valoración crítica de las pruebas diagnósticas.

El puntaje F, al integrar el valor predictivo positivo y la sensibilidad, ofrece una medida sintética menos dependiente de la prevalencia. La reducción de la entropía constituye una métrica integral que cuantifica la disminución de incertidumbre y ganancia informacional de una prueba diagnóstica, permitiendo comparaciones precisas entre diferentes exámenes mediante una única medida de precisión.

Para su aplicación práctica, diseñamos una calculadora específica que integra estos indicadores. Su implementación facilitaría la interpretación de estudios diagnósticos y contribuiría a mejorar la toma de decisiones clínicas.

Palabras clave: *valor predictivo de las pruebas; incertidumbre; teoría de la información; entropía; pruebas diagnósticas de rutina.*

doi (español): http://dx.doi.org/10.5546/aap.2025-10882
doi (inglés): http://dx.doi.org/10.5546/aap.2025-10882.eng

Cómo citar: Cuestas E, Cieri ME, Escobar Zuluaga J, Ruiz Brünner MM. Reducción de la entropía y puntaje F: aplicaciones de la teoría de la información en la evaluación de pruebas diagnósticas. *Arch Argent Pediatr.* 2026;124(4):e202510882.

[1] *Servicio de Pediatría y Neonatología, Hospital Privado Universitario de Córdoba, Cátedra de Pediatría, Córdoba, Argentina;*
[2] *Instituto Universitario de Ciencias Biomédicas de Córdoba (IUCBC), Centro de Investigación en Medicina Traslacional Severo R. Amuchástegui (CIMETSA); Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Córdoba, Argentina;*
[3] *2.da Cátedra de Clínica Pediátrica, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Argentina.*

Correspondencia para *Eduardo Cuestas: eduardo.cuestas@iucbc.edu.ar*

Financiamiento*: Ninguno.*

Conflicto de intereses*: Ninguno que declarar.*

Recibido*: 2-9-2025*
Aceptado*: 18-12-2025*

Esta obra está bajo una licencia de Creative Commons Atribución-No Comercial-Sin Obra Derivada 4.0 Internacional.
Atribución — Permite copiar, distribuir y comunicar públicamente la obra. A cambio se debe reconocer y citar al autor original.
No Comercial — Esta obra no puede ser utilizada con finalidades comerciales, a menos que se obtenga el permiso.
Sin Obra Derivada — Si remezcla, transforma o crea a partir del material, no puede difundir el material modificado.

INTRODUCCIÓN

El uso adecuado de la literatura médica en el marco de la medicina basada en la evidencia disminuye los errores médicos.[1] Evaluar la precisión diagnóstica de los exámenes complementarios es un paso esencial para mejorar la seguridad del paciente y la calidad de la atención. Las métricas tradicionales —sensibilidad, especificidad y valores predictivos positivo (VPP) y negativo (VPN)— son fundamentales en la evaluación diagnóstica. Sin embargo, presentan limitaciones al comparar pruebas en distintos contextos clínicos.[2] La sensibilidad y la especificidad pierden utilidad una vez obtenido el resultado, momento en el que el pediatra debe considerar la probabilidad posprueba, representada por el VPP y el VPN. Estas últimas, a su vez, dependen de la prevalencia, lo que dificulta su comparación entre poblaciones con perfiles epidemiológicos distintos.[3]

Aunque la teoría de la información de Claude Shannon (1948) ha tenido aplicaciones exitosas en múltiples campos,[4] su empleo en evaluación diagnóstica médica sigue siendo limitado.[5,6] Los conceptos de reducción de incertidumbre (entropía) y precisión-exhaustividad[3,7] ofrecen herramientas sólidas para hacer más predecible una fuente de información.[8] La reducción de la incertidumbre mide cuánto una prueba aclara la presencia o ausencia de enfermedad, mientras que el puntaje F sintetiza el equilibrio entre sensibilidad (precisión) y VPP (exhaustividad), con relativa independencia de la prevalencia.

Este trabajo tuvo por objetivo aplicar y adaptar herramientas derivadas de la teoría de la información y el aprendizaje automático —el puntaje F y la reducción de la entropía— para la evaluación y comparación de pruebas diagnósticas. Asimismo, se presenta una calculadora específica para su implementación con el fin de enriquecer el análisis crítico y fortalecer la toma de decisiones basadas en evidencia.

MÉTODOS

En primer lugar, empleamos el puntaje F —tradicionalmente utilizado en ciencias de la computación— para la evaluación de pruebas diagnósticas médicas.[3] Este indicador fue adaptado para estimar el desempeño de la prueba en la detección de pacientes enfermos, reduciendo la influencia de la prevalencia al combinar la sensibilidad con el VPP. Matemáticamente, se expresa como la media armónica de ambas métricas:

$$Puntaje\ F = 2 \cdot \frac{VPP \cdot Sensibilidad}{VPP + Sensibilidad}$$

El puntaje F refleja el equilibrio entre sensibilidad, es decir, la capacidad de detectar correctamente a los enfermos, y el valor predictivo positivo, que indica la probabilidad de que un resultado positivo corresponda a un caso verdadero. Valores altos señalan buen rendimiento diagnóstico global en una sola métrica.

La entropía de Shannon *(H)* mide la incertidumbre asociada con una distribución de probabilidad. En medicina, esto se traduce en la incertidumbre diagnóstica antes y después de una prueba. Una prueba útil debería reducir la entropía, es decir, reducir la incertidumbre sobre si un paciente tiene o no una enfermedad.

Fórmula general de entropía de Shannon:[8]

$$H(x) = -\Sigma p_i \cdot log2(p_i)$$

Para eventos binarios (enfermo/sano), el máximo ocurre cuando $p = 0{,}5$ (máxima incertidumbre).

Cálculo de la reducción de entropía diagnóstica:

Se parte de una tabla 2 × 2 con los datos de la prueba (VP = verdaderos positivos, FP = falsos positivos, FN = falos negativos, VN = verdaderos negativos y N = total de casos) y se calculan:

1. Entropía preprueba:

H preprueba=[FP+VN]VN.(log2(N)-log2(FP+VN))
+[VP+FN]VN.(log2(N)-log2(VP+FN))

2. Entropía posprueba para resultados positivos (n positivo):

H posprueba (+)
=[VP]Vn positivo.(log2(n positivo)-log2(VP))
+[FP]Vn positivo.(log2(n positivo)-log2(FP))

3. Entropía posprueba para resultados negativos (n negativo):

H posprueba (-)
=[FN]Vn negativo.(log2(n negativo)-log2(FN))
+[VN]Vn negativo.(log2(n negativo)-log2(VN))

4. Reducción de entropía:

Δ H= H preprueba - [n positivoVN .H posprueba (+)
+ n negativoVN . H posprueba (-)

5. Interpretación:

A mayor *Δ*, mayor capacidad de disminuir la incertidumbre diagnóstica. Si la prueba fuera

perfecta (100 % de sensibilidad y 100 % de especificidad), la incertidumbre se reduciría 100 %.

Para ilustrar la aplicación práctica, analizamos un ejemplo con datos tomados del artículo "Tira reactiva de orina para el diagnóstico de infecciones del tracto urinario en lactantes febriles de 2 a 6 meses".[9] Estos datos se emplearon para calcular sensibilidades, especificidades, VPP y VPN, junto con verdaderos y falsos positivos/negativos, razones de verosimilitud, reducción de la entropía, puntaje F y precisión diagnóstica.

Este estudio está exento de revisión ética institucional por involucrar únicamente datos preexistentes de fuentes públicas sin posibilidad de identificación directa o indirecta de los sujetos.

RESULTADOS

Según los datos de la *Tabla 1*, la esterasa leucocitaria se confirma como la prueba más robusta para el diagnóstico de infección urinaria, al alcanzar la mayor reducción de entropía (48,3 %), reflejo de la ganancia informativa más significativa y de la disminución más marcada de la incertidumbre entre la situación pre- y posprueba. Este indicador integra todos los componentes de la tabla 2 × 2 ponderados por su frecuencia relativa, mostrando de manera simultánea la capacidad tanto de confirmar como de descartar enfermedad.

En este sentido, la reducción de la entropía constituye el parámetro que mejor expresa la ganancia informativa real en la comparación relativa de pruebas diagnósticas en la práctica clínica, al traducir de forma tangible la disminución de la incertidumbre. Asimismo, la esterasa leucocitaria alcanza el puntaje F más elevado (68,9 %), resultado del óptimo equilibrio entre sensibilidad (91,2 %) y valor predictivo positivo (55,3 %), lo que la torna menos dependiente de la prevalencia y, por ende, más confiable.

En contraste, los nitritos, si bien exhiben mayor especificidad (98,9 %) y valor predictivo positivo (80,8 %), presentan una sensibilidad claramente inferior (36,7 %), mientras que el recuento de leucocitos muestra un rendimiento intermedio.

En síntesis, la esterasa leucocitaria, al conjugar el mejor equilibrio entre puntaje F y reducción de la entropía o incertidumbre, se posiciona como la prueba más informativa y consistente para la práctica clínica.

La *Figura 1* muestra la variación del VPP (líneas continuas) y del puntaje F (líneas discontinuas) según la prevalencia, para sensibilidades y especificidades de 0,90, 0,80, 0,70 y 0,60. A medida que la prevalencia aumenta, el VPP se incrementa de forma marcada, siguiendo una relación racional no lineal que evidencia su fuerte dependencia de la prevalencia en todos los niveles de desempeño diagnóstico.

Tabla 1. Comparación del valor diagnóstico de las pruebas de esterasa leucocitaria, nitritos y sedimento de orina vs. urocultivo[1]

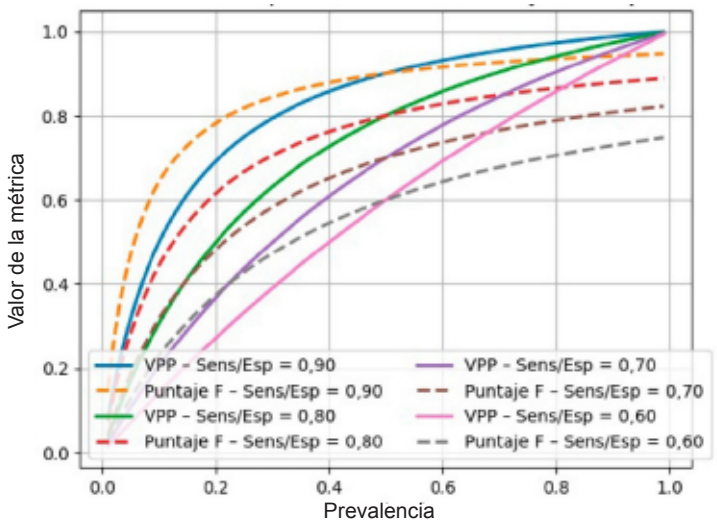
Métrica	Esterasa leucocitaria	Nitritos	Recuento de leucocitos ≥5 × CGA
Verdaderos positivos, n	810	373	759
Falsos negativos, n	78	645	121
Verdaderos negativos, n	6217	7950	3489
Falsos positivos, n	655	89	742
Sensibilidad, %	91,2	36,7	86,2
Especificidad, %	90,5	98,9	82,5
Valor predictivo positivo, %	55,3	80,8	50,6
Valor predictivo negativo, %	98,8	92,5	96,6
Razón de verosimilitud positiva	9,57	33,15	4,92
Razón de verosimilitud negativa	0,10	0,64	0,17
Reducción de la entropía, %	48,3	20,9	33,1
Puntaje F, %	68,9	13,8	63,7
Precisión, %	90,5	91,9	83,1

CGA: campo de gran aumento.

[1] *Tomados de: Hunt KM, et al.* Urine Dipstick for the Diagnosis of Urinary Tract Infection in Febrile Infants Aged 2 to 6 Months. Pediatrics. *2025;155:e2024068671.*

Los cálculos fueron realizados utilizando la calculadora gratuita en línea Diagnostic Test Metrics *desarrollada por los autores, disponible en:* https://ejcuestas.github.io/diagnostic-calculator/

Figura 1. Influencia de la prevalencia sobre el valor predictivo positivo (VPP) y el puntaje F en pruebas diagnósticas con sensibilidad y especificidad equivalentes*



**Las simulaciones se realizaron en Python (3.14.2) utilizando NumPy para generar poblaciones simuladas y explorar un rango continuo de prevalencias entre 0,01 y 0,99, manteniendo sensibilidad y especificidad constantes en cada escenario (0,90; 0,80; 0,70; 0,60). Se utilizó el método de Monte Carlo con 100 000 iteraciones por condición para asegurar estabilidad numérica. El VPP se calculó mediante la fórmula bayesiana estándar y el puntaje F a partir del VPP y la sensibilidad según la fórmula canónica. Las curvas resultantes se representaron con Matplotlib. No se aplicaron métodos de ajuste ni inferencia, dado que el objetivo fue ilustrar de manera determinística el comportamiento comparativo de ambas métricas.*

Incluso con una sensibilidad y especificidad de 0,80, el VPP disminuye sustancialmente cuando la prevalencia baja, pasando de valores muy altos en prevalencias elevadas a valores mucho menores en prevalencias intermedias y bajas. En este escenario, el VPP desciende de aproximadamente 0,97 a una prevalencia de 0,90 a alrededor de 0,73 cuando la prevalencia cae a 0,40, lo que implica una disminución absoluta de 0,24 puntos y una reducción relativa del 24,7 %.

En contraste, con el mismo desempeño diagnóstico, el puntaje F muestra una variación mucho más atenuada en el mismo rango de prevalencias. Sus curvas (líneas punteadas) adoptan una forma suavemente polinomial, más estable y menos susceptible a los cambios en la prevalencia. Con una sensibilidad y especificidad de 0,80, el puntaje F desciende de aproximadamente 0,88 a una prevalencia de 0,90 a 0,76 cuando la prevalencia es de 0,40, lo que representa una disminución absoluta de 0,12 puntos y una reducción relativa de 13,6 %, aproximadamente la mitad de la observada en el VPP. Esta diferencia cuantitativa muestra que, aunque el puntaje F no es completamente independiente de la prevalencia, su variación ante los cambios en la prevalencia es considerablemente menor que la del VPP, especialmente en pruebas con desempeño diagnóstico moderado a alto.

Para facilitar su aplicación en la práctica clínica, hemos desarrollado específicamente una calculadora en línea que permite a los lectores calcular directa y simultáneamente el puntaje F y la reducción de la entropía introduciendo los datos básicos (VP, FP, VN, FN) en la tabla de contingencia 2 × 2, disponible en:

https://ejcuestas.github.io/diagnostic-calculator/

DISCUSIÓN

El análisis de los datos de tablas diagnósticas 2 × 2 mediante métricas derivadas de la teoría de la información y el aprendizaje automático, tales como la reducción de la entropía y el puntaje F, muestra el potencial de estas herramientas para fortalecer la evaluación crítica de la evidencia.

La reducción entrópica constituye una métrica integral que cuantifica la disminución de incertidumbre y ganancia informacional de una prueba diagnóstica, permitiendo comparaciones precisas entre diferentes herramientas diagnósticas mediante una medida única de precisión.[6]

Paralelamente, el puntaje F facilita una comparación más matizada entre pruebas

diagnósticas, al trascender las limitaciones del VPP y ofrecer una medida del rendimiento predictivo global, con relativa independencia de la prevalencia, reflejando el equilibrio entre la identificación correcta de casos verdaderos y la minimización de clasificaciones erróneas, superando el valor del VPP y la sensibilidad considerados aisladamente.[5]

Nuestros resultados revelan que la esterasa leucocitaria proporcionó la mayor reducción de la entropía (48,3 %) y el mejor puntaje F (68,9 %). Este hallazgo subraya la superioridad diagnóstica de este método cuando se evalúa a través de métricas que integran la precisión (sensibilidad) y la exhaustividad (VPP).[4]

Esta perspectiva representa un avance hacia la personalización de las decisiones clínicas. La comprensión de la reducción de la incertidumbre asociada con cada paso diagnóstico permite una toma de decisiones más informada. Esta aproximación podría facilitar la toma de decisiones personalizadas en tiempo real en diferentes entornos, lo que representa una mejora en comparación con la práctica habitual, que a menudo se basa en la evaluación del riesgo poblacional a partir de la historia clínica y la demografía del paciente.[4,8]

Nuestros hallazgos constituyen un modelado estadístico-matemático que requiere validación clínica posterior. Aunque la implementación de la entropía de Shannon y el puntaje F constituye una aplicación muy prometedora de la teoría de la información al análisis diagnóstico, sus ventajas clínicas deben ser confirmadas mediante estudios prospectivos que evalúen su desempeño en la práctica real.

La principal limitación radica en que estas métricas, al igual que los indicadores tradicionales, no siempre reflejan las consecuencias clínicas de falsos positivos y negativos. Esta consideración queda supeditada al criterio clínico y la experiencia pediátrica, donde el costo de un falso negativo puede superar al de un falso positivo, o viceversa, según el escenario clínico específico.[10]

En conclusión, los resultados indican que tanto el puntaje F como la reducción de la entropía pueden constituir herramientas valiosas para el análisis comparativo de pruebas diagnósticas y para el diseño de futuros estudios y revisiones sistemáticas. Su aplicación en diferentes poblaciones permitirá corroborar la estabilidad y la validez de estos indicadores en diversos contextos clínicos. ■

REFERENCIAS

1. Cosby K, Yang D, Fineberg HV. Assessing Diagnostic Performance. *NEJM Evid*. 2024;3(2):EVIDra2300232. doi:10.1056/EVIDra2300232.
2. Van Rijsbergen CJ. Information Retrieval. 2nd ed. Michigan: Butterworths; 1979.
3. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12(1):5979. doi: 10.1038/s41598-022-09954-8.
4. Casagrande A, Fabris F, Girometti R. Fifty years of Shannon information theory in assessing the accuracy and agreement of diagnostic tests. *Med Biol Eng Comput*. 2022;60(4):941-55. doi:10.1007/s11517-021-02494-9.
5. Krause P. Information Theory and Medical Decision Making. *Stud Health Technol Inform*. 2019;263:23-34. doi: 10.3233/SHTI190108.
6. Benish WA. A Review of the Application of Information Theory to Clinical Diagnostic Testing. *Entropy (Basel)*. 2020;22(1):97. doi:10.3390/e22010097.
7. Mosquera C, Ferrer L, Milone DH, Luna D, Ferrante E. Class imbalance on medical image classification: towards better evaluation practices for discrimination and calibration performance. *Eur Radiol*. 2024;34(12):7895-903. doi: 10.1007/s00330-024-10834-0.
8. He S, Chong P, Yoon BJ, Chung PH, Chen D, Marzouk S, et al. Entropy removal of medical diagnostics. *Sci Rep*. 2024;14(1):1181. doi: 10.1038/s41598-024-51268-4.
9. Hunt KM, Green RS, Sartori LF, Aronson PL, Chamberlain JM, Florin TA, et al. Urine Dipstick for the Diagnosis of Urinary Tract Infection in Febrile Infants Aged 2 to 6 Months. *Pediatrics*. 2025;155(4):e2024068671. doi: 10.1542/peds.2024-068671.
10. Taylor RA, Sangal RB, Smith ME, Haimovich AD, Rodman A, Iscoe MS, et al. Leveraging artificial intelligence to reduce diagnostic errors in emergency medicine: Challenges, opportunities, and future directions. *Acad Emerg Med*. 2025;32(3):327-39. doi: 10.1111/acem.15066.