

Artículo Original

Evaluación de la replicabilidad y consistencia del triaje de urgencias asistido por inteligencia artificial: Análisis del software comercial Mediktor según perfiles clínicos y demográficos

Eva M. Ferrer Beltran ^{1,*} y Ane Araico Bedia ¹

¹ Departamento Médico de Teckel Medical S.L., Barcelona, España; eferrer@mediktor.com; aaraico@mediktor.com

* Autor correspondencia: eferrer@mediktor.com; ORCID id: 0009-0001-2942-219X; Tel.: +34-644-223-582

DOI: <https://doi.org/10.37536/RIECS.2025.10.2.481>

Resumen: Dado que un triaje eficaz en urgencias es crucial para la seguridad del paciente y la estandarización es clave, este estudio tuvo como objetivo evaluar la replicabilidad y consistencia del software de triaje asistido por inteligencia artificial Mediktor. La metodología se desarrolló en dos fases: una primera para verificar la replicabilidad, utilizando perfiles de paciente y motivos de consulta idénticos; y una segunda para analizar la consistencia, introduciendo perfiles clínicos y demográficos diversos para una misma queja. Los resultados demostraron una replicabilidad del 100 % en la primera fase, generando siempre el mismo flujo de preguntas, prediagnóstico y nivel de urgencia. En la segunda fase, el software adaptó de forma coherente el interrogatorio a las variaciones del perfil del paciente, manteniendo la consistencia en los resultados finales. En conclusión, Mediktor es una herramienta altamente fiable, replicable y consistente, que se adapta al perfil del paciente sin comprometer la exactitud, validando su uso como un soporte seguro para estandarizar el proceso de triaje en entornos clínicos.

Palabras Clave: Inteligencia Artificial, Triage, Diagnóstico Preliminar, Reproducibilidad de los Resultados.

Abstract: Given that effective triage in emergency departments is crucial for patient safety and standardization is key, this study aimed to evaluate the replicability and consistency of the artificial intelligence-assisted triage software, Mediktor. The methodology was developed in two phases: a first phase to verify replicability using identical patient profiles and chief complaints, and a second phase to analyze consistency by introducing diverse clinical and demographic profiles for the same complaint. The results demonstrated 100 % replicability in the first phase, always generating the same question flow, pre-diagnosis, and urgency level. In the second phase, the software coherently adapted the interview based on the patient's profile variations while maintaining consistency in the final results. In conclusion, Mediktor is a highly reliable, replicable, and consistent tool that adapts to the patient's profile without compromising accuracy, thus validating its use as a safe support for standardizing the triage process in clinical settings.

Key words: Artificial Intelligence, Triage, Preliminary Diagnosis, Reproducibility of Results.

1. Introducción

En los servicios de urgencia, el proceso de triaje es esencial para manejar de manera eficiente a los pacientes, en particular cuando la necesidad excede los recursos existentes. Su meta es clasificar a los pacientes en función de la severidad de su estado, dando prioridad a los casos más urgentes [1].

Es fundamental mantener una uniformidad en el triaje para asegurar una valoración equitativa y prevenir fallos en la priorización que puedan impactar la calidad del cuidado y la seguridad del paciente [1]. Para lograrlo, se necesitan sistemas organizados y normalizados, respaldados por protocolos y herramientas tecnológicas que garanticen una evaluación imparcial y repetible.

Además, el uso de programas especializados en triaje puede mejorar la precisión y rapidez en la clasificación, siempre que estos proporcionen resultados homogéneos entre diferentes usuarios y perfiles de pacientes [2].

Conforman algunos de los sistemas de triaje más famosos:

- **Modelo de Triage de Andorra (MAT):** fundamentado en cinco grados de severidad, emplea algoritmos clínicos digitales para realizar una valoración exacta [3].
- **Sistema de Triage Manchester (MTS):** creado en Reino Unido, utiliza categorías de exposición y esquemas de flujo para organizar la evaluación [4].
- **Escala de Triage y Acuidad de Canadá (CTAS):** define cinco grados de urgencia basados en los síntomas principales y los tiempos de respuesta sugeridos [5].

El objetivo de normalizar estos sistemas es minimizar la variabilidad en la valoración inicial de los pacientes. No obstante, la implementación de tecnologías como programas de triaje exige validaciones estrictas para asegurar resultados uniformes entre diferentes grupos de pacientes.

En este marco, se examina Mediktor, un programa de triaje asistido por inteligencia artificial (IA) cuya tecnología central evalúa los síntomas de los pacientes y los guía hacia el nivel de atención apropiado de acuerdo a su urgencia. Esta investigación analiza su reproducibilidad y uniformidad en la clasificación de urgencias.

1.1. Objetivo del estudio

El objetivo de este estudio es evaluar la capacidad de replicabilidad y consistencia del software de triaje y prediagnóstico asistido por IA Mediktor. El objetivo es establecer si el software produce un flujo de preguntas y resultados (prediagnóstico y nivel de urgencia) equivalentes al interactuar con pacientes de iguales características, y si conserva la coherencia al incluir perfiles clínicos y demográficos diferentes para una misma razón de consulta.

Objetivos específicos:

- **Analizar la capacidad de reproducción y consistencia en perfiles idénticos:** Es crucial verificar que Mediktor genera un conjunto de preguntas y resultados totalmente reproducibles cuando se presenta el mismo motivo de consulta (escrito de manera exacta) para un paciente con las mismas características demográficas (sexo, edad) y se reproducen las respuestas en la evaluación.
- **Examinar la capacidad del software para adaptarse a diversos perfiles:** Analizar si Mediktor modifica el flujo de preguntas de forma consistente de acuerdo al historial clínico y demográfico del paciente.
- **Establecer la solidez del prediagnóstico y el grado de urgencia:** Analizar si, a pesar de las fluctuaciones en el flujo de preguntas al aplicar diferentes perfiles, el software conserva un nivel de predicción homogéneo en la clasificación de urgencias.

Es fundamental asegurar la fiabilidad y uniformidad de tecnologías como Mediktor para garantizar un cuidado justo y exacto, reduciendo la variabilidad en la interpretación del triaje por los expertos en salud.

2. Material y Métodos

El presente estudio se ha llevado a cabo empleando el software comercial Mediktor, creado por Teckel Medical S.L. [6]. El procedimiento de evaluación y análisis de los datos tuvo lugar entre febrero y marzo de 2025.

Para examinar la seguridad y coherencia del software Mediktor en la clasificación de usuarios en un contexto de triaje de urgencias, los investigadores han elaborado un estudio experimental dividido en dos pasos:

- **Fase 1:** Centrada en examinar el primer objetivo, evaluando seis razones de consulta, cada una examinada 30 veces con un paciente idéntico por razón de consulta.
- **Fase 2:** Dirigida a valorar los objetivos segundo y tercero, examinando cuatro razones de consulta con 30 pacientes distintos (con diferentes perfiles clínicos y demográficos) por cada razón.

Para asegurar la fiabilidad del análisis, se escogieron los motivos de consulta que forman parte del estudio basándose en su frecuencia en los servicios de urgencias hospitalarias y su prevalencia en los registros históricos de Mediktor. Para reducir prejuicios y garantizar la repetibilidad de los hallazgos, se registraron los motivos de consulta de forma estandarizada, manteniendo una escritura uniforme en ambas etapas del estudio y respetando el uso de mayúsculas, argot y terminología particular.

2.1. Fase 1: Evaluación de la consistencia en pacientes con perfiles demográficos idénticos

En esta fase se evaluó si el software Mediktor genera un flujo de preguntas, prediagnósticos y niveles de urgencia consistentes y reproducibles cuando pacientes con características demográficas idénticas ingresan con el mismo motivo de consulta (Ver Tabla I).

Tabla I Perfil demográficamente idéntico (edad, género y ubicación) testeado por cada uno de los 6 motivos de consulta evaluados

	Motivo de consulta	Edad (años)	Género	Ubicación
Caso 1	Tengo dolor de cabeza	24	Femenino	Lisboa (Portugal)
Caso 2	Tengo fiebre de 38’7	73	Masculino	Luanda (Angola)
Caso 3	Me sale moco del oído	35	Mujer trans	Bilbao (España)
Caso 4	Tengo dificultad para miccionar	52	Femenino	Granada (España)
Caso 5	Me ha salido una mancha roja en la cara	40	Masculino	India
Caso 6	No tengo ganas de hacer nada, estoy todo el día llorando	64	Hombre trans	Barcelona (España)

Para ello, se utilizó el software sin proporcionar una historia clínica detallada de los pacientes. Se definieron seis motivos de consulta y, para cada uno, se realizaron 10 repeticiones de tres evaluaciones simultáneas con el mismo perfil de paciente, resultando en un total de 30 evaluaciones por motivo de consulta.

En cada repetición, tres usuarios con características idénticas en cuanto a edad, género y ubicación ingresaron simultáneamente el mismo motivo de consulta y siguieron el mismo flujo de preguntas. El objetivo fue determinar si la herramienta genera un patrón replicable en la formulación de preguntas, en la asignación de prediagnósticos y en la determinación del nivel de urgencia. Por cada motivo de consulta se utilizó un único perfil de paciente idéntico durante las 30 repeticiones.

2.2. Fase 2: Evaluación de la adaptación del software a perfiles clínicos diferenciados

En esta segunda fase, el estudio se centró en analizar la capacidad del software Mediktor para adaptar su flujo de preguntas y respuestas en función de las características clínicas y demográficas del paciente, asegurando que este ajuste se realice de manera lógica y coherente, tal como lo haría el personal de enfermería en un triaje hospitalario.

Para ello, se diseñaron 30 perfiles médicos distintos, con variaciones en su historial clínico y demográfico, asegurando una distribución equitativa en términos de edad, género y condiciones médicas preexistentes (Ver Materiales Suplementarios). Cada uno de estos perfiles fue evaluado en el software Mediktor mediante cuatro motivos de consulta, estableciendo una duración fija de los síntomas para garantizar la comparabilidad de los resultados (Ver Tabla II).

Tabla II Motivos de consulta y duración de los síntomas de cada uno de los casos de estudio para cada uno de los 30 perfiles evaluados

	Motivo de consulta	Duración de los síntomas
Caso 1	Tengo dolor de cabeza	Desde hace 2 días
Caso 2	Tengo fiebre de 38'7	Desde hace 10 horas
Caso 3	Me sale moco del oído	Desde hace 3 días
Caso 4	Tengo dificultad para miccionar	Desde hace 8 horas

El objetivo de esta fase fue determinar si el software ajusta su proceso de evaluación en función de las características individuales de los pacientes, manteniendo la coherencia en la formulación de preguntas, en la clasificación del nivel de urgencia y en la emisión de prediagnósticos y recomendaciones clínicas.

Asimismo, se realizó una comprobación de repetibilidad para verificar si el comportamiento del software Mediktor es consistente al utilizar la característica de perfil en comparación con su uso sin perfil, como se evaluó en la Fase 1 del estudio. Para ello, se evaluaron los cuatro casos en tres repeticiones consecutivas con un único perfil (Perfil 1), con el fin de comprobar si el flujo de preguntas y los resultados obtenidos son replicables al emplear un perfil demográfica y clínicamente idéntico.

2.3. Procesamiento y análisis de los datos

La recolección de datos fue realizada por tres colaboradoras formadas específicamente en la aplicación del protocolo del estudio. Para asegurar la imparcialidad y objetividad del proceso, se estableció como criterio fundamental que estas colaboradoras no tuvieran ninguna vinculación profesional o comercial con Teckel Medical S.L., ni experiencia previa con la herramienta evaluada. Su función se limitó exclusivamente a la aplicación del protocolo y registro de datos según las indicaciones establecidas por los investigadores.

Durante el desarrollo de las evaluaciones, y con el objetivo de garantizar un mayor grado de repetibilidad, se seleccionó en cada caso una evaluación de referencia, definida como aquella en la que el flujo de preguntas y los resultados se repitieron con mayor frecuencia en todas las repeticiones. Esta evaluación representativa se tomó como caso control, de modo que cualquier desviación en el flujo de preguntas o en los resultados respecto a esta fue considerada como falta de repetibilidad. Para mantener la consistencia en las respuestas, todas las preguntas adicionales introducidas por el software fueron respondidas con "No lo sé" o "No".

Todos los datos recopilados en ambas fases fueron documentados y sistematizados, y se encuentran disponibles en los Materiales Suplementarios del estudio. Se llevó a cabo un análisis comparativo de los resultados con el propósito de identificar patrones de consistencia en la generación de preguntas, la asignación de prediagnósticos y la determinación del nivel de urgencia.

Este análisis permite evaluar la fiabilidad y replicabilidad del software en escenarios clínicos controlados, así como su capacidad de adaptación a diferentes perfiles de pacientes, garantizando que los principios de equidad y precisión sean preservados en el proceso de triaje automatizado.

2.4. Consideraciones éticas

El estudio se realizó íntegramente con pacientes ficticios, sin la participación de sujetos humanos ni animales, y sin el uso de datos clínicos reales. Dado que se trató exclusivamente de simulaciones en el software Mediktor, no fue necesaria la aprobación de un comité de ética.

3. Resultados

3.1. Fase 1: Evaluación de la consistencia en pacientes con perfiles demográficos idénticos

En esta fase, se analizaron los seis casos expuestos en la Tabla I, cada uno correspondiente a un motivo de consulta específico, evaluado 30 veces con un mismo perfil de paciente también descrito en la Tabla I.

Para determinar si el software comercial Mediktor produce evaluaciones replicables y consistentes, se han establecido los siguientes criterios de aceptación, donde se definen los resultados 1 y 0 como evaluación repetible y no repetible, respectivamente:

1. Evaluación repetible (1) si se cumplen todas las siguientes condiciones:
 - a. Todas las preguntas del flujo de evaluación son las mismas en cada repetición (no se omiten, agregan ni varían).
 - b. El orden de las preguntas es idéntico en todas las evaluaciones.
 - c. El nivel de urgencia y el prediagnóstico asignado son los mismos en todas las repeticiones.
2. Evaluación no replicable (0) si alguna de las condiciones anteriores no se cumple.

Los resultados detallados de cada evaluación se encuentran disponibles en los Materiales Suplementarios.

Los resultados presentados en la Figura 1 muestran que todos los casos analizados, correspondientes a los motivos de consulta "Tengo dolor de cabeza", "Tengo fiebre de 38,7°C", "Me ha salido una mancha roja en la cara", "No tengo ganas de hacer nada, estoy todo el día llorando", "Me sale moco del oído" y "Tengo dificultad para miccionar", fueron 100 % replicables en todas las evaluaciones. Esto incluye tanto la repetibilidad del flujo de preguntas como la replicabilidad de los resultados de urgencia y prediagnóstico, los cuales se mantuvieron consistentes en todas las iteraciones.

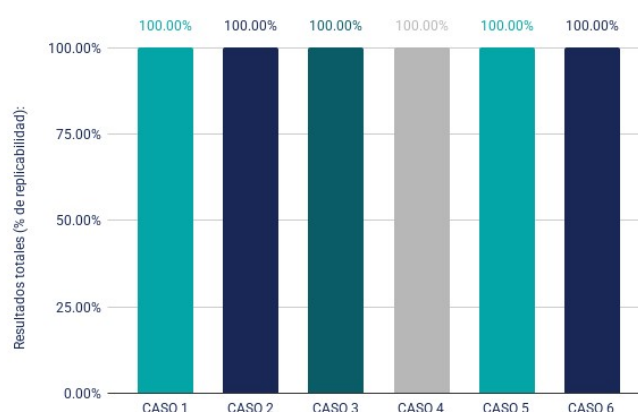


Figura 1 Resultados del porcentaje de replicabilidad (eje vertical) de cada caso definido en la Tabla I (eje horizontal). Cada caso contiene 30 evaluaciones por cada tipo de perfil.

3.2. Fase 2: Evaluación de la adaptación del software a perfiles clínicos diferenciados

En esta fase, se analizaron los cuatro casos expuestos en la Tabla II, cada uno correspondiente a un motivo de consulta específico, evaluado una vez por cada uno de los 30 perfiles definidos en los

Materiales Suplementarios. Adicionalmente, se analizaron los cuatro casos expuestos en la Tabla II tres veces consecutivas con un mismo perfil (Perfil 1) para estudiar la repetibilidad del software al usar la característica de perfil.

Para evaluar los resultados de esta fase, se han establecido los siguientes criterios de aceptación, que evalúan por un lado la repetibilidad del flujo de preguntas de la evaluación, y por otro lado, evalúa la repetibilidad del resultado obtenido. En ambos casos, se definen los valores 1 y 0 como resultado repetible y no repetible, respectivamente:

1. Repetibilidad del flujo de preguntas:
 - a. Flujo de preguntas repetible (1) si se cumplen todas las siguientes condiciones:
 - i. Mismas preguntas.
 - ii. Mismo número de preguntas totales.
 - b. Flujo de preguntas no replicable (0) si alguna de las condiciones anteriores no se cumple.
2. Repetibilidad del resultado obtenido:
 - a. Resultado repetible (1) si se cumplen todas las siguientes condiciones:
 - i. Mismo nivel de urgencia.
 - ii. Misma primera enfermedad del prediagnóstico o Mismas 3 enfermedades en el prediagnóstico en cualquier orden.
 - b. Resultado no replicable (0) si alguna de las condiciones anteriores no se cumple.

Los resultados detallados de cada evaluación se encuentran disponibles en los Materiales Suplementarios.

Acerca del estudio de consistencia del flujo de preguntas y resultados al usar un único perfil definido demográficamente y clínicamente (Perfil 1), los resultados presentados en la Figura 2 muestran que todos los casos analizados, correspondientes a los motivos de consulta "Tengo dolor de cabeza", "Tengo fiebre de 38,7°C", "Me sale moco del oído" y "Tengo dificultad para miccionar", fueron 100 % replicables, tanto en el flujo de preguntas como en el resultado de urgencia y prediagnóstico. En todos los casos, no se observaron variaciones en el flujo de preguntas, y los resultados se mantuvieron consistentes en todas las iteraciones.

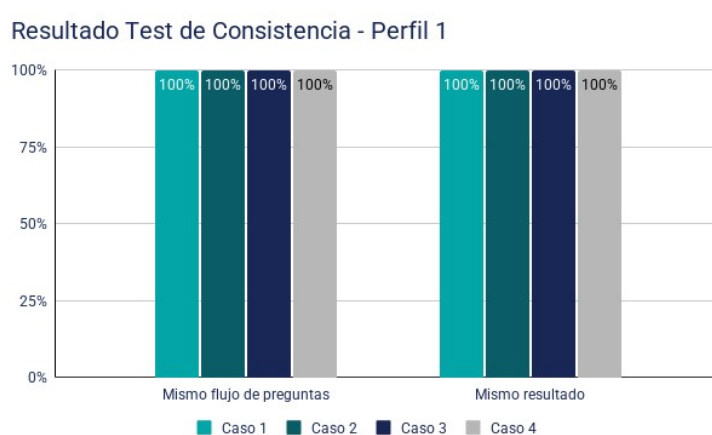


Figura 2 Porcentaje de replicabilidad (eje vertical) del flujo de preguntas y de los resultados (eje horizontal) para cada caso definido en la Tabla II (leyenda), tras tres iteraciones consecutivas con el Perfil 1, descrito en los Materiales Suplementarios.

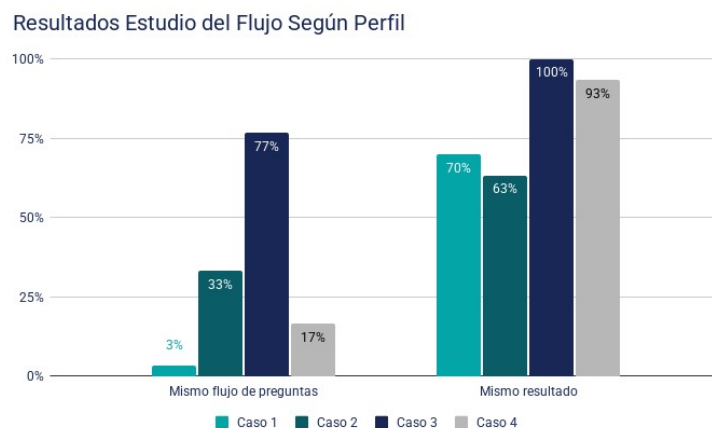


Figura 3 Porcentaje de replicabilidad (eje vertical) del flujo de preguntas y de los resultados (eje horizontal) para cada caso definido en la Tabla II (leyenda), tras una iteración para cada uno de los perfiles (desde el Perfil 1 al Perfil 30), descritos en los Materiales Suplementarios.

Con respecto al estudio de adaptación del software Mediktora a perfiles clínicos y demográficos diferenciados, los resultados presentados en la Figura 3 muestran que los casos 1, 2 y 4, correspondientes a los motivos de consulta "Tengo dolor de cabeza", "Tengo fiebre de 38,7°C" y "Tengo dificultad para miccionar", presentaron una menor repetibilidad en el flujo de preguntas (3 %, 33 % y 17 %, respectivamente), mientras que la replicabilidad del resultado fue mayor en comparación (70 %, 63 % y 93 %, respectivamente).

Por otro lado, el caso 3, correspondiente al motivo de consulta "Me sale moco del oído", fue 100 % replicable en el resultado, con un 77 % de repetibilidad en el flujo de preguntas, lo que indica una mayor consistencia en comparación con los demás casos.

En todos los casos, la variabilidad en el flujo de preguntas se debe a la inclusión o exclusión de determinadas preguntas, ya que el evaluador adapta el interrogatorio en función de las características clínicas y demográficas del paciente.

Cabe destacar que todos los casos en los que el resultado (nivel de urgencia y prediagnóstico) no fue replicable también presentaron un flujo de preguntas no replicable. Esto sugiere una correlación entre la variabilidad en el flujo de preguntas y la replicabilidad del resultado: cuando el flujo de preguntas es replicable, el resultado también lo es; mientras que si el flujo de preguntas varía, el resultado puede o no mantenerse constante.

4. Discusión

4.1. Fase 1: Evaluación de la consistencia en pacientes con perfiles demográficos idénticos

Los resultados obtenidos en la Fase 1 indican que el software evaluado logra un alto grado de replicabilidad en todos los casos analizados. Los motivos de consulta correspondientes a los casos 1, 2, 3, 4, 5 y 6 fueron 100 % replicables, lo que implica que tanto el flujo de preguntas, la clasificación del nivel de urgencia, como el prediagnóstico fueron completamente consistentes en las 30 evaluaciones realizadas para cada uno de estos casos.

Aunque todos los casos fueron 100 % replicables en términos de los resultados obtenidos, es importante destacar que el sistema de sinónimos y procesamiento de lenguaje natural (NLP) empleado por el software puede influir en la replicabilidad de la evaluación, especialmente en lo que respecta al flujo de preguntas. Si el motivo de consulta está redactado de manera diferente (por ejemplo, con variaciones en la redacción), esto puede generar pequeñas discrepancias en la forma en que el software interpreta la consulta, lo que podría dar lugar a ligeras variaciones en el flujo de preguntas. No obstante, el diseño del sistema garantiza que, a pesar de estas diferencias en la redacción del motivo de consulta, el flujo de preguntas siga siendo similar y coherente, y que los resultados de prediagnóstico y urgencia sigan siendo replicables de manera consistente.

4.2. Fase 2: Evaluación de la adaptación del software a perfiles clínicos diferenciados

Los resultados de esta fase del estudio muestran que, al evaluar los cuatro motivos de consulta con un mismo perfil definido tanto demográfica como clínicamente (Perfil 1), el flujo de preguntas y los resultados fueron 100 % replicables. Esto confirma la alta consistencia del software, independientemente del uso de la función de determinación de perfil. La replicabilidad total en el triaje, la clasificación del nivel de urgencia y el prediagnóstico refuerza la capacidad del sistema para generar evaluaciones estandarizadas y precisas.

El estudio sobre la adaptación de Mediktor a perfiles clínicos y demográficos diferenciados, realizado con 30 perfiles evaluados en cuatro motivos de consulta, muestra que la replicabilidad del flujo de preguntas y los resultados varía según el caso y las características del paciente.

El motivo de consulta "Me sale moco del oído" presentó la mayor consistencia, con un 77 % de repetibilidad en el flujo de preguntas y un 100 % en los resultados. Esto sugiere que las patologías relacionadas con este síntoma tienen menor variabilidad diagnóstica en función de factores como la edad, el sexo o los antecedentes médicos.

En contraste, los casos "Tengo dolor de cabeza" (3 %), "Tengo fiebre de 38,7°C" (33 %) y "Tengo dificultad para miccionar" (17 %) mostraron mayor variabilidad en el flujo de preguntas. En el caso del dolor de cabeza, la repetibilidad fue baja debido a que el historial clínico influye en la evaluación; por ejemplo, si el paciente tiene hipertensión, se añaden preguntas específicas. Para la fiebre, la estructura del interrogatorio varió según la sospecha diagnóstica del evaluador automatizado, como COVID-19 o mononucleosis. En "Tengo dificultad para miccionar", la variabilidad se debe principalmente al género y a antecedentes médicos, ya que las etiologías pueden diferir significativamente entre hombres y mujeres.

En cuanto a los resultados, la replicabilidad fue mayor que en el flujo de preguntas. "Me sale moco del oído" mostró un 100 % de repetibilidad, mientras que "Tengo dolor de cabeza" alcanzó un 70 %, "Tengo fiebre de 38,7°C" un 63 % y "Tengo dificultad para miccionar" un 93 %. En este último caso, las diferencias en el nivel de urgencia se debieron a factores como la sospecha de pielonefritis en lugar de cistitis o la presencia de fiebre y tratamiento inmunosupresor. En general, los casos de dolor de cabeza y fiebre fueron replicables en prediagnóstico y urgencia, con variaciones mínimas influenciadas por el historial clínico. Estos resultados refuerzan que la variabilidad en los resultados finales depende del flujo de preguntas, reflejando la lógica clínica aplicada según el perfil del paciente.

La capacidad del software para adaptarse a variaciones en los motivos de consulta y aun así generar resultados consistentes refuerza la estabilidad del sistema. Esto subraya la importancia de contar con herramientas de triaje estandarizadas y eficientes para la atención de emergencias. En esta línea, diversos autores destacan la necesidad de optimizar el triaje en urgencias para agilizar la toma de decisiones [7].

Además, la variabilidad en el flujo de preguntas del software es comparable con la observada en el triaje realizado por personal sanitario. Se ha documentado variabilidad intraobservador (un mismo profesional evalúa de manera distinta al mismo paciente en distintos momentos) e interobservador (diferentes profesionales pueden hacer evaluaciones distintas de un mismo caso), lo cual es inherente al juicio clínico [8]. Las diferencias en el software reflejan esta realidad de los entornos clínicos, reforzando su utilidad como herramienta complementaria en la práctica médica.

4.3. Valoración de la herramienta

Los resultados del estudio muestran que Mediktor puede contribuir a optimizar la atención sanitaria, facilitando un triaje más rápido y preciso y respaldando el criterio clínico con orientación basada en evidencia.

Operativamente, la herramienta puede ayudar a agilizar el flujo de trabajo, reducir el tiempo de evaluación inicial y disminuir la carga administrativa, lo que podría mejorar la eficiencia del sistema de salud y la experiencia del paciente.

Asimismo, proporciona información clara sobre acciones recomendadas, contribuyendo a la educación sanitaria y apoyando la toma de decisiones de los profesionales de la salud. A futuro, su

uso podría favorecer una mayor coordinación entre profesionales mediante un sistema de información más estructurado, permitiendo una atención más personalizada.

5. Conclusiones

En la Fase 1 de la investigación, todos los casos estudiados alcanzaron una replicabilidad total en términos de flujo de preguntas, urgencia y prediagnóstico, evaluando seis motivos de consulta con perfiles de paciente idénticos.

En la Fase 2, al utilizar, para diferentes motivos de consulta, un perfil idéntico definido no sólo demográficamente sino también clínicamente, se observó que los resultados seguían siendo totalmente reproducibles, manteniéndose la consistencia en el flujo de preguntas y los resultados de urgencia y prediagnóstico. Demostrando la consistencia del software, independientemente del uso de la función de determinación de perfil.

Sin embargo, cuando se usaron perfiles clínicos y demográficos diferentes, la variabilidad en el flujo de preguntas aumentó, aunque los resultados de urgencia y prediagnóstico siguieron siendo bastante consistentes. El caso "Me sale moco del oído" mostró la mayor reproducibilidad, mientras que "Tengo dolor de cabeza", "Tengo fiebre de 38,7°C" y "Tengo dificultad para miccionar" presentaron más fluctuaciones en las preguntas, lo que refleja cómo las características del paciente impactan en la evaluación. Esto refuerza la capacidad del software para ofrecer resultados estandarizados y confiables, incluso en contextos clínicos con diferentes perfiles.

Agradecimientos: Agradecemos a Teckel Medical S.L. por facilitarnos el acceso a su software para la realización de este estudio. Extendemos nuestro reconocimiento a Clara Cambra, Maria Capdevila y Miranda Griñán, estudiantes de Enfermería de las Escuelas Universitarias Gimbernat (EUG), por su apoyo en la recolección de datos siguiendo el protocolo establecido por los investigadores. Asimismo, agradecemos a herramientas de asistencia a la redacción y corrección de texto, como ChatGPT, por su apoyo en la mejora de la presentación del manuscrito.

Contribución de los autores: Los investigadores Eva M. Ferrer Beltran y Ane Araico Bedia concibieron y diseñaron los experimentos; Eva M. Ferrer analizó los datos y escribió el manuscrito; Ane Araico Bedia supervisó los resultados y revisó la versión final del artículo.

Conflictos de Intereses: El proceso de testeo y recopilación de datos fue realizado de manera independiente por Clara Cambra, Maria Capdevila y Miranda Griñán, cuya función se limitó estrictamente a la aplicación del protocolo establecido por los investigadores. Ninguna de ellas tiene vinculación profesional o comercial con Teckel Medical S.L., ni experiencia previa con la herramienta evaluada, lo que asegura la objetividad y neutralidad en la obtención de los datos.

El análisis de los datos y la redacción del estudio fueron llevados a cabo por los investigadores, empleados de Teckel Medical S.L., existiendo una relación contractual y económica que constituye un conflicto de interés potencial. No obstante, los datos fueron obtenidos de manera independiente y el análisis se realizó con rigor metodológico, procurando minimizar cualquier sesgo derivado de esta relación y garantizar la imparcialidad en la evaluación del software.

Abreviaturas

Las siguientes abreviaturas son usadas en este manuscrito:

MAT: Modelo de Triage de Andorra

MTS: Sistema de Triage Manchester

CTAS: Escala de Triage y Acuidad de Canadá

IA: Inteligencia Artificial

Referencias Bibliográficas

1. Soler W, Gómez Muñoz M, Bragulat E, Álvarez A. El triaje: herramienta fundamental en urgencias y emergencias. *An Sist Sanit Navar*. 2010;33(Supl. 1):55-68. <https://doi.org/10.4321/S1137-66272010000200008>

2. Validación clínica de la nueva versión del Programa de Ayuda al Triage (web_e-PAT v3) del Modelo Andorrano de Triage (MAT) y Sistema Español de Triage (SET). Fiabilidad, utilidad y validez en la población pediátrica y adulta [Internet]. Revistaemergencias.org. [cited 2025 Mar 31]. Available from: https://revistaemergencias.org/numeros-anteriores/volumen-18/numero-4/validacion-clinica-de-la-nueva-version-del-programa-de-ayuda-al-triaje-web_e-pat-v3-del-modelo-andorrano-de-triaje-mat-y-sistema-espanol-de-triaje-set-fiabilidad-utilidad-y-validez-en-la-pobla/
3. Gómez Jiménez J, Ferrando Garrigós JB, Vega García JL, Tomás Vecina S, Roqueta Egea F, Chanovas Borràs M. Model Andorrà de Triatge: Bases conceptuals i manual de formació. Servei Andorrà d'Atenció Sanitària; 2003.
4. Mackway-Jones K, Marsden J, Windle J. Emergency Triage. BMJ Publishing Group; 1997.
5. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S. Reliability of the Canadian Emergency Department Triage and Acuity Scale: interrater agreement. Ann Emerg Med. 1999;34(2):155-9.
6. Mediktor - Improving access to healthcare worldwide [Internet]. Mediktor.com. [citado 28 de febrero de 2025]. Disponible en: <https://www.mediktor.com/>
7. Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. Dtsch Arztebl Int. 2010;107(50):892-8. <https://doi.org/10.3238/arztebl.2010.0892>
8. Parenti N, Reggiani ML, Iannone P, Percudani D, Dowding D. A systematic review on the validity and reliability of an emergency department triage scale, the Manchester Triage System. Int J Nurs Stud. 2014;51(7):1062-9. <https://doi.org/10.1016/j.ijnurstu.2014.01.013>



© 2025 por los autores; Esta obra está sujeta a la licencia de Reconocimiento 4.0 Internacional de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/4.0/>.