



Título: Evaluación de una herramienta de atención remota de triaje asistida por Inteligencia Artificial (IA) (INFERMEDICA® SYMPTOM CHECKER) y elaboración de un modelo teórico de adopción de un sistema de IA en el acceso a la atención urgente u ordinaria en una Comunidad Autónoma

Autores:

Damián Gallegos Lemos, Javier Iturria Soler, Arancha Real Díaz

Tutor:

Pablo Serrano Balazote

Escuela Nacional de Sanidad
Instituto de Salud Carlos III
Sociedad Española de Informática de la Salud

Master en Dirección de Sistemas y TIC de la Salud y en
Digitalización Sanitaria

Octubre, 2021

Tabla de contenido

Resumen	4
1. Introducción	5
1.1. La atención remota como alternativa de uso creciente	5
1.2. La atención remota en las Unidades de Urgencia Hospitalaria	5
1.2.1. Estándares y Recomendaciones en las Unidades de Urgencia Hospitalaria	5
1.2.2. Modelos de Triage	6
1.2.3. Saturación de Servicios de Urgencia Hospitalaria	7
1.2.4. Justificación de uso de los Symptom checkers	7
1.2.5. Oportunidades para los Symptom Checkers.....	9
2. Referencias al contenido del temario del Máster	9
3. Objetivos	10
4. Metodología	10
5. Resultados	11
5.1. Descripción de Symptomate, Infermédica®	11
5.1.1. Características técnicas	12
5.1.2. Metodología de validación clínica de la base de datos de información médica	13
5.1.3. Metodología para el desarrollo del algoritmo de Infermedica: la red bayesiana	14
5.1.4. Procesamiento de Lenguaje Natural.....	17
5.1.5. Descripción del proceso diagnóstico	17
5.1.6. Descripción del proceso de triaje.....	18
5.1.7. Arquitectura del Sistema	22
5.1.8. Validación externa del algoritmo de Infermédica®.....	23
5.2. Casos de uso del algoritmo de Infermédica®.....	25
5.3. Evaluación de Symptomate, Infermédica®.....	27
5.3.1. Aplicación del método de valoración de aplicaciones móviles de salud en español iSYScore de la fundación iSYS a Symptomate de infermédica®	28
5.3.2. Aplicación de las recomendaciones de Yun Liu. et al. sobre cómo evaluar herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos médicos.....	29
5.4. Propuesta de un Modelo Teórico para para incorporar un symptom checker en el flujo de atención urgente u ordinaria para el acceso al sistema de salud de una Comunidad Autónoma	36
5.4.1. Matriz de análisis DAFO.....	36
5.4.2. Encuesta sobre conocimiento sobre soluciones de Triage o Symtom checkers con Inteligencia artificial en SNS	40
5.4.3. Modelo de triaje virtual asistido por IA y acceso a servicios médicos	43
6. Discusión	46
7. Bibliografía citada.....	49

8. Webgrafía	51
--------------------	----

Resumen

La pandemia de COVID-19 ha supuesto, no solo una sobrecarga y en ocasiones colapso de los sistemas sanitarios alrededor del mundo, sino que también ha motivado un cambio en el modo de buscar atención sanitaria por parte de los ciudadanos. En este sentido, se ha evidenciado que las consultas remotas por teléfono y/o a través de plataformas de telemedicina o de gestión remota de pacientes (GRP) se postulan como una alternativa sencilla, eficaz y económica para solucionar los requerimientos de atención por parte de los pacientes, evidenciando la potencialidad de las atenciones remotas en modalidades síncronas. Al mismo tiempo han favorecido también poder plantear la alternativa para modalidades de atención asíncronas (por ejemplo, anamnesis estructuradas y asistidas por inteligencia artificial (IA), o los verificadores de síntomas o symptom checkers en inglés).

El presente trabajo evalúa a Symptomate un symptom checker asistido por Inteligencia artificial de la empresa Infermedica, a través de dos metodologías de evaluación: un método de valoración de aplicaciones móviles de salud en español, el índice iSYScore de la fundación iSYS, que valora la aceptabilidad y la experiencia de usuario y la segunda descrita en JAMA por Yun Liu; et. al. que utiliza un proceso de 3 pasos para derivar, validar y establecer la eficacia clínica de herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos. En ambas la herramienta obtuvo una buena valoración.

Adicionalmente se realiza una propuesta de modelo teórico para el establecimiento de un flujo de atención urgente u ordinaria para el acceso al sistema de salud en una Comunidad Autónoma a través de un triaje gestionado por un symptom checker asistido por IA. Para comprender mejor el contexto de posible implementación se utilizaron dos instrumentos: una matriz de análisis DAFO y una encuesta a responsables TIC de las CCAA sobre su conocimiento acerca la implementación de soluciones de symptom checkers.

En la gestión de los servicios de urgencia hospitalaria, los symptom checkers tendrían la potencialidad de aliviar los servicios de urgencia hospitalaria saturados, ofrecer una valoración estructurada de los síntomas de un paciente facilitando el diagnóstico al médico y gestionar el agendamiento de citas según su prioridad. Además de la optimización del uso de los recursos y consecuente reducción de costes en el sistema sanitario.

No obstante, una de las principales limitaciones que tendría la implantación de un flujo de atención urgente u ordinaria de acceso al sistema de salud de una Comunidad Autónoma a través de un sistema de triaje virtual es la poca evidencia disponible para recomendar el uso de symptom checkers. Es necesario por tanto un mayor esfuerzo para la generación de estudios de alto nivel de evidencia científica, incluyendo estudios experimentales en poblaciones representativas donde se vaya a utilizar el algoritmo de triaje, para obtener una base científica con mayor solidez que nos permita realizar una recomendación más firme sobre el uso de esta herramienta científica y de gestión. La propuesta de modelo recogida en este estudio podría trasladarse a una Investigación-Acción que podría contribuir a la generación de evidencia.

Esta propuesta trata de motivar un cambio estructural por parte de los sistemas de salud, de los profesionales y de los usuarios que contribuya a la transformación del sector de la salud hacia un modelo más sostenible y conectado, a través de fomentar una medicina proactiva, con un ciudadano más activo e informado y centrado en la prevención y el uso eficiente de los recursos.

1. Introducción

1.1. La atención remota como alternativa de uso creciente

La pandemia de COVID-19 ha supuesto, no solo una sobrecarga y en ocasiones colapso de los sistemas sanitarios alrededor del mundo, sino que también ha motivado un cambio en el modo de buscar atención sanitaria por parte de los ciudadanos, ante el miedo a la transmisión y contagio en los centros sanitarios. Una necesidad imperante es brindar una respuesta alternativa a la ciudadanía cuya demanda de atención es creciente.

En este sentido, se ha evidenciado que las consultas remotas por teléfono y/o a través de plataformas de telemedicina o de gestión remota de pacientes (GRP) se postulan como una solución sencilla, eficaz y económica para muchas dudas de pacientes, además de saltarse ese riesgo de potencial transmisión en las consultas in-situ y como alternativa a suplir estas demandas de atención en los servicios de salud, tanto de primaria como especializada.

Se ha evidenciado que los pacientes buscan cada vez más participar más en su asistencia sanitaria. Como resultado, las herramientas digitales para el cuidado de la salud (herramientas de salud en línea y móviles) han proliferado y el uso por parte de los pacientes ha aumentado drásticamente. Se estima que 1 de cada 3 adultos estadounidenses se conectaron a Internet para intentar auto diagnosticarse un problema médico ya en 2013. El hecho de que los pacientes en la actualidad hayan incrementado el uso de herramientas digitales de atención médica es una oportunidad que facilitaría la introducción de nuevos procesos en los flujos de atención urgente u ordinaria para determinar el acceso al sistema de salud.

Las atenciones por vía telefónica, en donde el caso lo permite, han solucionado muchos requerimientos por parte de los pacientes; evidenciando la potencialidad de las atenciones remotas en modalidades síncronas, pero al mismo tiempo han favorecido también poder plantear la alternativa para modalidades de atención asíncronas (por ejemplo, anamnesis estructuradas y asistidas por inteligencia artificial (IA), o los verificadores de síntomas o symptom checkers en inglés).

El reto de los servicios de salud es cómo ampliar el acceso a la atención remota al paciente satisfecho manteniendo los estándares de atención basada en la evidencia con alta calidad, en el momento justo y en el lugar correcto, pudiendo ser este en muchos casos el propio domicilio del ciudadano con atención telemática o presencial, y en los casos necesarios según el nivel de triaje, en los servicios de urgencias hospitalarios.

1.2. La atención remota en las Unidades de Urgencia Hospitalaria

1.2.1. Estándares y Recomendaciones en las Unidades de Urgencia Hospitalaria

Una unidad de urgencia hospitalaria (UUH) debe disponer de un sistema de triaje, validado, integrado en el sistema informático del hospital, que permita la rápida clasificación del paciente en función de su gravedad, asignándole un tiempo máximo de espera para ser atendido en el recurso apropiado. El triaje es un método de trabajo estructurado que permite una evaluación rápida de la gravedad del paciente, de una manera reglada, válida y reproducible, con el propósito de priorizar el orden de atención médica y determinar el lugar y los medios que precise. La cartera de servicios de apoyo

a la UUH, debe estar en relación con el volumen de urgencias y con las características del hospital donde se ubique la unidad.¹

1.2.2. Modelos de Triage

Antes de la pandemia de COVID-19, el flujo de acceso a los servicios de salud ofrecía como única alternativa la manera presencial. Todavía ahora, una considerable cantidad de pacientes acceden al sistema de salud a través de los servicios de urgencia hospitalaria, donde se realiza un triaje. El concepto de triaje hace referencia a la disponibilidad de una escala de clasificación válida, útil y reproducible que permita realizar la clasificación de los pacientes en base a su grado de urgencia basándose en el grado de urgencia/gravedad. El sistema de triaje ha de ser dinámico, fácil de entender y rápido de aplicar, ha de estar estandarizado o normalizado y debe tener un alto nivel de concordancia inter-observador.²

En la actualidad, existen varios modelos de clasificación normalizados, estructurados, estandarizados y universalizados a 5 niveles, adaptados al área de urgencias: la Australian Triage Scale (ATS/NTS); la Canadian Emergency Department Triage and Acuity Scale (CTAS); el Manchester Triage System (MTS); el Emergency Severity Index (ESI); y el Model Andorrà de Triage (MAT), que se adoptó como modelo estándar para España con el nombre de Sistema Español de Triage (SET) según las recomendaciones de la Sociedad Española de Medicina de Urgencias y Emergencias (SEMES).^{3, 4} En España, los sistemas de triaje que han alcanzado mayor difusión son el MAT-SET y el MTS.

Un abordaje alternativo ha sido el desarrollo de algoritmos mediante tecnologías de Machine Learning que permiten a los modelos aprender directamente de los datos. Se han generado muchos modelos para la ayuda a la decisión en triaje basados en árboles de decisión, regresión logística, redes neuronales y redes Bayesianas.^{5, 6}

¹ Ministerio de Sanidad y Política Social. Unidad de urgencias hospitalaria Estándares y recomendaciones. 2010.

² Gómez Jiménez J. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias, Emergencias. 2003; 15:165-74.

³ Gómez Jiménez J, Burgues I, Pàmies S. Clinical management of a hospital emergency department: quality indicators, benchmarking and case-mix analysis., Gestión Hospitalaria., 2004;15,(1),:3:12

⁴ FitzGerald G, Jelinek GA, Scott D, Gerdts MF. Emergency department triage revisited. Emerg Med J. 2010 Feb;27(2):86-92.

⁵ Abad-Grau MM, Ierache J, Cervino C, Sebastiani P. Evolution and challenges in the design of computational systems for triage assistance. Journal of Biomedical Informatics. junio de 2008;41(3):432-41.

⁶ Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. Diagn Progn Res. diciembre de 2020;4(1):16.

1.2.3. Saturación de Servicios de Urgencia Hospitalaria

Uno de los retos que viene arrastrando el Sistema Nacional de Salud (SNS) desde hace ya unos años, es la no adecuación de las visitas a los servicios de urgencia hospitalaria por parte de la población, resultando en un mayor coste de la atención sanitaria, peor eficiencia de los servicios de urgencia, descenso de los indicadores de calidad y la posibilidad de poner en riesgo la seguridad del paciente.^{7, 8}

Los Servicios de Urgencias Extrahospitalarios (SUEs) deberían ser la puerta de entrada a la atención urgente, de la misma manera en que la Atención Primaria (AP) es la entrada al Sistema Nacional de Salud. A pesar de estar diseñado con este fin, un gran porcentaje de pacientes acude directamente a los Servicios de Urgencia Hospitalarios (SUH), en una práctica conocida como bypass, utilizando los recursos para patologías de segundo nivel en personas que no lo necesitan.⁹

Aunado a otras causas, la consecuencia de este proceso es la saturación de las urgencias; situación en la que se han descrito aumento de la mortalidad hospitalaria a los 30 días, mayor mortalidad del trauma grave en urgencias, retrasos en el tratamiento antibiótico y del dolor, demoras en las cirugías de urgencias, aumento del tiempo “puerta-aguja” en el infarto agudo de miocardio, prolongación de la estancia hospitalaria, demoras para tomografía axial computarizada (TAC), aumento del número de errores del personal, más readmisiones, mayor mortalidad en la unidad de cuidados intensivos (UCI) por retrasos en el ingreso, entre otros.¹⁰

1.2.4. Justificación de uso de los Symptom checkers

En la actualidad, muchos pacientes buscan cada vez más opciones para consultar en la web sobre sus síntomas para obtener diagnósticos, lo cual se ha visto incrementado debido a la pandemia. La Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los hogares de 2019 del INE revela que el 72% de los internautas busca información sobre salud. En el contexto actual y con el impacto de la pandemia esta realidad se ha acentuado y los ciudadanos reclaman nuevas formas de contactar con los sistemas de salud y su personal sanitario.

Los sistemas asistidos por IA de verificación de síntomas (symptom checkers), se postulan como alternativa para los pacientes que utilizan estos recursos, ya que el resultado entregado por el motor de inferencia de estos sistemas se constituye en una parte importante del triaje. A día de hoy instituciones de salud o compañías de seguros,

⁷ Ochoa Gómez J, Villar Arias A, Ignacio Ruiz Azpiazu J, Bragado Blas L, Gimeno Villa C, Ramalle-Gómara E. Visitas inapropiadas al servicio de urgencias de un hospital general. *Medicina Clínica*. enero de 2000;115(10):377-8.

⁸ Sempere-Selva T, Peiró S, Sendra-Pina P, Martínez-Espín C, López-Aguilera I. Inappropriate use of an accident and emergency department: Magnitude, associated factors, and reasons—An approach with explicit criteria. *Annals of Emergency Medicine*. junio de 2001;37(6):568-79.

⁹ Reinoso Hermida S, Díaz Grávalos GJ, Robles Castiñeiras A, Pereiro Sánchez E, Fernández Cardama E, López González A. Forma de acceso a los Servicios de Urgencia hospitalarios y pertinencia de la consulta. *Rev Clin Med Fam*. octubre de 2011;4(3):205-10.

¹⁰ Una propuesta de modelo fisiológico de servicio de urgencias hospitalario. Principios de funcionamiento, tipificación de la saturación y pautas para el rediseño. *An Sist Sanit Navar* [Internet]. abril de 2017 [citado 23 de agosto de 2021];40(1). Disponible en: <https://recyt.fecyt.es/index.php/ASSN/article/view/52722/35615>

entre otras, utilizan la tecnología de pre-diagnóstico para recomendar un nivel adecuado de ayuda (desde el autocuidado hasta la sala de emergencias) para re-dirigir a los pacientes a los centros médicos adecuados. Son también una alternativa para su uso en los sistemas públicos de salud en circunstancias precisas, como sistema complementario para reforzar la atención no presencial, y lo más crucial, para disminuir la presión de atención de urgencias, en especial para condiciones ambulatorias de baja complejidad, pero también para activar recursos del sistema de emergencias médicas 112.

Recientemente han surgido iniciativas de parte de instituciones de gran reconocimiento, como el National Health Service (NHS) y la Clínica Mayo, que han puesto en marcha su propio sistema de symptom checker en línea¹¹ vislumbrando el potencial que estos sistemas tienen para influenciar al paciente a la hora de buscar atención médica y como apoyo a los sistemas de salud.¹²

La existencia de una población de pacientes en crecimiento y la escasez de médicos, requiere de alternativas para lograr la expansión de servicios de salud manteniendo la asequibilidad, la calidad, y la comodidad que los pacientes esperan. De hecho, miles de personas se comunican con chatbots o centros de llamadas asociados con su proveedor de atención médica donde se está observando un número creciente de instituciones de salud que utilizan aplicaciones de pre-diagnóstico para la clasificación/triage y mejora del manejo de los flujos de pacientes. De esta manera, las visitas físicas pueden ser complementadas por interacciones más eficientes y cómodas con el personal sanitario.

En EEUU el 25% de las consultas presenciales son “consultas informativas” que fácilmente podrían ser atendidas a distancia.¹³ De esta forma, modalidades de atención asíncronas como las anamnesis estructuradas y asistidas por IA, o los verificadores de síntomas o symptom checkers permiten emitir y responder las consultas cuando paciente y profesional, respectivamente, tengan el momento propicio para ello.

Por otra parte, la facilidad de acceso a servicios de información médicos (de varios tipos y calidad) a través de Internet ha dado lugar a que los pacientes traten de diagnosticarse a sí mismos al encontrar información en línea, lo que generalmente conduce a la falsa impresión que están tan bien informados sobre el problema de salud como si se tratara de una consulta a un médico. Sin embargo, para obtener diagnósticos o auto-triage los verificadores de síntomas asistidos por inteligencia artificial (IA) en línea o Symptom Checkers en inglés, pueden ser una mejor opción frente a una en la que los usuarios, a través de utilizar un motor de búsqueda, encuentran información en línea y optan por la primera opción plausible, que típicamente es la opción con la trampa del sesgo de confirmación.¹⁴

¹¹ Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self-diagnosis and triage: audit study. BMJ. 8 de julio de 2015; h3480.

¹² Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S-C, Leung L-P. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. Hong Kong Journal of Emergency Medicine. julio de 2020;27(4):217-22.

¹³ Ashley N D Meyer, et.al. ; Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study; J Med Internet Res 2020 | vol. 22 | iss. 1 | e14679 | p. 1

¹⁴ Adam Zagorecki, Piotr Orzechowskia, Katarzyna Hołownia; A System for Automated General Medical Diagnosis using Bayesian Networks; MEDINFO 2013 C.U. Lehmann et al. (Eds.) © 2013 IMIA and IOS Press. This article is published online with Open Access by IOS Press and

1.2.5. Oportunidades para los Symptom Checkers

Los symptom checkers pueden ser una opción para aliviar los servicios de urgencia hospitalaria saturados, permitiendo a los pacientes indecisos sobre la urgencia de su condición, realizar un auto triaje e instruir a los pacientes con condiciones no urgentes abstenerse de acudir a la urgencia, y en los casos que así lo requieran ser derivados con la prioridad necesaria a Atención Primaria o a los SUEs.

En estudios observacionales se ha descrito que los pacientes más jóvenes son los que más acuden de forma espontánea a los SUH sin consultar previamente a su MAP ni a un SUE,^{6, 7} lo que supone un nicho para los sistemas informatizados debido a la mayor penetración en este sector de la población.

Ante la situación de colapso asistencial constante en el SNS, sobre todo en los SUH, donde los engranajes de la estructura del SNS no están suficientemente lubricados y ante un desequilibrio tan grande entre demanda y oferta, urge buscar herramientas y mecanismos necesarios para poner fin a este problema enquistado.¹⁵ Esto constituye una oportunidad única para los sistemas telemáticos basados en algoritmo tipo symptom checkers para integrar y coordinar la asistencia, juntos a una buena práctica de gestión del cambio, ofrecen una oportunidad para mejorar la calidad asistencial y bajar los costes en el SNS.

La puesta en marcha de un sistema de symptom checker con una integración adecuada a los sistemas de salud regionales tienen la capacidad de apoyar a los SUEs, evitar la saturación de los SUH, gestionar el agendamiento de citas según su prioridad en Atención Primaria, ofrecer una valoración estructurada de los síntomas de un paciente facilitando el diagnóstico al médico, ofrecer datos explotables para los gerentes sanitarios y organismo de Salud Pública, y además ser la entrada a los servicios de Telemedicina del SNS.

2. Referencias al contenido del temario del Máster

En este trabajo de fin de máster hemos tratado de aplicar los conceptos aprendidos a lo largo del curso, de los siguientes temas:

- ✓ **Tema 1.2** El Sector de las TIC para la Salud
- ✓ **Tema 2.1** La planificación TIC. Aspectos generales. La planificación estratégica. Articulación con la planificación operativa. La gobernanza TI. Efectos de la transformación digital en la planificación estratégica. Vinculación con el entorno estadístico
- ✓ **Tema 2.9.** Entornos / Metodologías / Plataformas de desarrollo. Desarrollos en el ámbito científico. Entornos bibliográficos. Gestión de contenidos, documentales. Repositorios. Control de versiones. Desarrollos para movilidad. Herramientas de backend

distributed under the terms of the Creative Commons Attribution Non-Commercial License.
doi:10.3233/978-1-61499-289-9-461

¹⁵ Miró O. El usuario y su uso de las urgencias. Anales Sis San Navarra [Internet]. diciembre de 2009 [citado 23 de agosto de 2021];32(3). Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272009000500001&lng=en&nrm=iso&tlng=en

- ✓ **Tema 3.2** La Protección de Datos de Salud. Marco Legislativo, aspectos sectoriales y nuevo marco europeo
- ✓ **Tema 3.5:** “Análisis de Entorno y habilidades del CIO para un Sistema Sanitario en cambio continuo”.
- ✓ **Tema 4.3** Sistemas, servicios y aplicaciones personales. “Personal Health Record”, Apps. Salud2.0, Servicios en la red para público en general, Redes de usuarios
- ✓ **Tema 4.7.** Analítica y modelos predictivos en Salud
- ✓ **Tema 4.8** Seguridad del paciente. Control de acceso, errores medicación. Ayuda a la decisión. Herramientas de ayuda al diagnóstico

3. Objetivos

1. Describir y evaluar una herramienta de atención remota de verificación de síntomas asistida por Inteligencia Artificial (symptom checker) que intenta orientar al paciente en la toma de decisiones respecto a la búsqueda de asistencia sanitaria (Symptomate de la empresa Infermédica®)
2. Proponer un modelo teórico para incorporar un symptom checker en el flujo de atención urgente u ordinaria para el acceso al sistema de salud de una Comunidad Autónoma

4. Metodología

Para la realización de este estudio se llevó a cabo una revisión de la literatura sobre las características técnicas y los casos de uso de la herramienta Symptomate de la empresa Infermédica®. Se consultó el dossier técnico de la herramienta y se llevó a cabo una búsqueda bibliográfica en la base de datos electrónica Medline.

Posteriormente se realizó una evaluación de la herramienta Symptomate. En el ámbito de la evaluación de aplicaciones móviles hay diversas iniciativas para la valoración y/o acreditación; así, podemos mencionar por ejemplo a MyHealthApps , donde los evaluadores son las asociaciones de pacientes, AppSaludable de la Junta de Andalucía o Health Apps Library del National Health Service (NHS) británico, o el iSYScore de la fundación iSYS española, entre otros.

Se analizó la calidad de la herramienta así como la potencialidad de la misma para ser incorporada en el flujo de acceso al sistema de salud, buscando ofrecer un servicio alternativo de gestión remota de pacientes enfocado en distribuir la demanda de atención sanitaria espacial y temporalmente, discriminando la necesidad de atención en servicios de urgencia hospitalaria de manera inmediata o en atención primaria de manera preferente o diferida. Se utilizaron las siguientes metodologías de evaluación:

iSYScore ¹⁶ es un proyecto que evalúa apps de salud en español con el objetivo de ofrecer una aproximación crítica a la confiabilidad de las apps en salud orientado al público general, así como ayudando a los médicos de atención primaria en un primer cribado sobre la idoneidad de la app a recomendar a sus pacientes. Esta herramienta fue construida con un enfoque sistemático basado en la evidencia y gracias al consenso

¹⁶ I. Grau, B. Kostov, J.A. Gallego, F. Grajales III, L. Fernández-Luque, y Sisó-Almirall A. Método de valoración de aplicaciones móviles de salud en español: el índice iSYScore. : Semergen. 2016;42(8):575---583

de expertos mediante un proceso Delphi, a través del cual se desarrolló un baremo y un catálogo de aplicaciones que ayuda a recomendar las mejores aplicaciones para usuarios o sanitarios a través de 3 diferentes dimensiones: 1) popularidad e interés (puntuación máxima de hasta 11 puntos); 2) confianza y calidad (puntuación máxima de hasta 18 puntos), y 3) utilidad (puntuación máxima de hasta 18 puntos). Se seleccionó esta metodología de evaluación para Symptomate de Infirmédica® puesto que es una herramienta desarrollada en nuestro medio para aplicaciones en español que incluye criterios de excelencia y que permite realizar una aproximación crítica a la confiabilidad, además de ayudar a los desarrolladores a mejorar la calidad de sus aplicaciones. La evaluación se realizó a partir de la revisión de las publicaciones existentes sobre Symptomate y por acuerdo entre los tres evaluadores.

Adicionalmente se aplicó la evaluación descrita en JAMA por Yun Liu; et. al.¹⁷ que utiliza un proceso de 3 pasos para derivar, validar y establecer la eficacia clínica de herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos. Esta metodología establece un conjunto de elementos a evaluar a través de unas preguntas guía, lo cual es de especial utilidad para recomendar la implementación de este tipo de soluciones en la práctica clínica o en la gestión de los servicios de salud. Mediante una búsqueda de la literatura relacionada con symptom checkers en enfermedades de carácter general, y por acuerdo entre los tres evaluadores, realizamos una interpretación metodológica, técnica y clínica, de Symptomate de Infirmédica® según la guía de Yun Liu. et al.

Finalmente, y en base al resultado de la evaluación, se propuso un modelo teórico que incluye un verificador de síntomas (Symptom Checker) asistido por inteligencia artificial, como parte del flujo de atención urgente u ordinaria para el acceso al sistema de salud en una comunidad autónoma. Para elaborar la propuesta de modelo teórico y comprender mejor el contexto de posible implementación se utilizaron dos instrumentos: una matriz de análisis DAFO (debilidades, amenazas, fortalezas y oportunidades) y una encuesta a responsables TIC de las 17 CCAA sobre su conocimiento acerca la implementación de soluciones de symptom checkers. El DAFO se realizó por parte de los tres evaluadores. La encuesta se realizó a través de la plataforma survey monkey y con llamadas telefónicas, en caso de no respuesta a la encuesta on-line.

5. Resultados

5.1. Descripción de Symptomate, Infirmédica®

La herramienta de atención remota de triaje asistida por Inteligencia Artificial (IA) (symptom checker) Symptomate de la empresa Infirmédica® es una Interfaz de programación de aplicaciones (API) para el triaje de pacientes y el diagnóstico médico preliminar, pensado para implementar un verificador de síntomas inteligente o un formulario de admisión de pacientes adaptado a cualquier aplicación de salud de un proveedor de atención sanitaria.

La solución consiste en un motor/base de conocimientos médicos de alto rendimiento con miles de síntomas, afecciones, factores de riesgo y pruebas de laboratorio que se cruzan entre sí de manera similar a una red en la que, dados los síntomas y los factores

¹⁷ Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *JAMA*. 2019;322(18):1806–1816. doi:10.1001/jama.2019.16489

de riesgo de un paciente, permite procesar la información asistida por IA y calcular las probabilidades de condiciones diagnósticas. Gracias a esta recopilación y su capacidad de procesar datos instantáneos, los pacientes reciben un pre-diagnóstico confiable, además de acceso a información médicamente certificada que facilita el proceso a los pacientes y los guía hacia el tratamiento adecuado a través de una consulta remota o presencial según sea el caso. Esto es posible gracias a los sofisticados algoritmos estadísticos que utilizan para realizar el razonamiento diagnóstico.

El papel de la herramienta de inferencia es calcular las probabilidades de condiciones factibles y elegir las preguntas más relevantes para hacer a continuación. La herramienta de inferencia es un algoritmo estadístico patentado para diagnóstico diferencial que utiliza el conocimiento subyacente almacenado en la base de conocimiento médico.

El motor de inferencia proporciona una conversación dinámica y personalizada con el paciente, mucho más eficaz que una regla fija o un sistema basado en aplicar un árbol de decisiones. Además del conocimiento basado en expertos utiliza el aprendizaje automático para reunir información de conjuntos de datos médicos anónimos y utiliza bucles de retroalimentación para mejoras continuas y afinamiento a través del tiempo.

5.1.1. Características técnicas

De la documentación técnica de Symptomate de la empresa Infermédica®¹⁸ presentamos a continuación sus características más destacadas:

La base de conocimientos médicos (Metabase) comprende más de 1260 síntomas, 680 condiciones médicas y 133 factores de riesgo que se interconectan entre sí que permiten un pre-diagnóstico preciso de los pacientes y el triaje pre-hospitalario; ésta se revisa y amplía continuamente de acuerdo con las demandas específicas de pacientes y médicos.

Infermédica® cuenta con un equipo médico que trabaja en los diferentes conceptos médicos, utilizando fuentes reconocidas conocimiento médico y desarrollamos conceptos junto con expertos certificados para cubrir diversas especializaciones médicas; al igual que para validar los conceptos en varios idiomas.

La base de datos de información médica lleva alrededor de 10 años en constante desarrollo y continuamente se agregan nuevas condiciones, síntomas, factores de riesgo y pruebas de laboratorio mejorando así su colección enciclopédica. Esta Metabase y la calidad de los datos es el fundamento de la herramienta y donde se enfoca el diseño de este verificador de síntomas, emulando el ejercicio que hacen los médicos en el diagnóstico de sus pacientes.

Las claves para hacer de esta base de datos de información médica confiable es consecuencia del trabajo del equipo médico que se estima en algo más de 34.000 horas construyendo y validando el más alto nivel de contenido médico. El trabajo del equipo médico sigue los principios de la medicina basada en la evidencia para cada contenido de la Metabase y se validan contrastándolas con fuentes confiables, como revistas médicas bien establecidas (p.e. Best Practice by British Medical Journal, New England Journal of Medicine, Lancet, Clinical Key), guías y publicaciones de agencias especializadas y organizaciones que operan a nivel mundial (p.e. OMS, CDC), así como numerosos artículos médicos con un alto factor de impacto (p.e. NEJM, Lancet).

¹⁸ <https://developer.infermedica.com/docs/introduction>

Infermédica® ha establecido un proceso de desarrollo de contenido que evita errores y mantiene la más alta calidad de información médica publicable. Las etapas clave incluyen definir un alcance de mejora del contenido y la estructura existentes, obtener conocimiento experto, especificar criterios de prueba, realizar pruebas y desplegar procedimientos. Este proceso es dinámico en tanto que se verifica y actualiza permanentemente lo que exige también validar la magnitud y capacidad del modelo probabilístico de la herramienta que vincula todos los conceptos médicos de la base de con síntomas asociados, afecciones, estudios de casos, El modelo de inferencia matemática como la base de datos médica son elementos complementarios de la solución y no pueden existir el uno sin el otro.

Por otra parte, a medida que se construye la base de datos de información médica, se toma en consideración el cómo los pacientes la utilizarán y cómo comunican sus síntomas. Junto con el contenido médico profesional, se incluyen nombres comunes y frases de síntomas que suelen utilizar los pacientes. Ésta es la clave de la eficacia de la base de datos; términos comunes vinculados al vocabulario médico. Con cada nuevo concepto, se conectan el lenguaje del paciente con la terminología médica de una manera comprensible e inequívoca.

5.1.2. Metodología de validación clínica de la base de datos de información médica

La validación clínica de las condiciones y síntomas (conceptos médicos) y demás datos que se incluyen en la Metabase, garantiza un nivel constante de rendimiento de todo el sistema de inteligencia artificial por lo que las aplicaciones se ejecutan a través de una batería de pruebas, que incluyen todos los aspectos médicos y técnicos:

1. Revisión de pares: Este paso se lleva a cabo una vez que un nuevo concepto médico está listo. Durante la revisión por pares, los demás editores de contenido médico analizan los nuevos conceptos. Juntos proponen, discuten e introducen posibles mejoras. Al mismo tiempo, validan todos los datos y fuentes relacionados con el nuevo concepto médico.
2. Prueba de aceptación: Una vez que se aprueba el nuevo concepto, el equipo médico comienza a trabajar en casos de pruebas clínicas para probar su rendimiento general y verificar el desempeño del concepto médico recién agregado en el conjunto de la base de datos. Para ello se utilizan miles de casos clínicos bien documentados de un síntoma o afección determinados de pacientes reales descritos en la literatura científica; esto ha permitido automatizar el proceso.

Una vez que se ha agregado el nuevo caso de prueba, la Metabase lo valida con el motor de diagnóstico de Infermédica® y verifica su corrección con todos los demás síntomas y condiciones en la base de datos. A día de hoy, los casos de prueba tienen una tasa de éxito del 93% para afecciones comunes. Después de la verificación positiva, el nuevo caso de prueba pasa a formar parte de Metabase y se utiliza en pruebas de regresión.

3. Revisión de expertos; A intervalos regulares, todo el contenido médico se somete a una verificación adicional. Los médicos del panel de expertos de Infermédica® realizan una revisión adicional. Se trata de especialistas con experiencia en diversos campos. Durante la revisión de expertos, inspeccionan cuidadosamente los conceptos nuevos y existentes para identificar posibles incompatibilidades.

4. Revisión técnica: Se comprueban si el nuevo concepto se ajusta a las directrices internas. Cada nuevo concepto debe estar perfectamente alineado con la estructura del contenido médico, por lo que buscan posibles problemas que involucren la estructuración del contenido médico. Estos pueden ser desde síntomas duplicados hasta jerarquía alterada de síntomas o problemas con parámetros numéricos recién introducidos.
5. Pruebas de regresión: Una vez se confirman los nuevos conceptos médicos, se construye un modelo de diagnóstico actualizado. En este punto, se hacen pruebas de regresión para validar cómo el contenido recién introducido afecta el rendimiento del modelo anterior. Comparar versiones de modelos de diagnóstico con pruebas de regresión ayuda a garantizar la estabilidad del sistema y medir continuamente su comportamiento. Señalar que el rendimiento real en las pruebas varía del 93% para los casos de afecciones comunes al 85% para todos los casos.
6. Pruebas manuales: El paso final del ciclo de creación de contenido consiste en que los médicos prueban las afecciones y los síntomas recién introducidos de una manera que refleja cómo éstos o los pacientes los usarían en la experiencia de la vida real. Aunque este método es subjetivo, ayuda a detectar elementos que podrían interferir con la experiencia del usuario y permite corregirlos antes de que ser incluidos.

5.1.3. Metodología para el desarrollo del algoritmo de Infermedica: la red bayesiana

El sistema está destinado al público en general, considerando que no hay la posibilidad de proporcionar formación sobre su uso, por lo que es imprescindible que la información se introduzca de forma intuitiva y eficiente en el tiempo. Un segundo aspecto a tener en cuenta es el sistema tiempo de respuesta, asumiendo que el proceso de diagnóstico normalmente incluye de 10 a 20 preguntas únicas para llegar a un diagnóstico presuntivo, por lo que cada interacción debería tomar menos de un segundo para proporcionar una experiencia de usuario adecuada. Para lograr eso, el proceso consiste en dividir el modelo de dominio (Metabase de información médica) en varias redes bayesianas que incluían variables superpuestas. A continuación, se detalla la metodología aplicada.

La red bayesiana (RB) es un modelo especialmente útil para dominios que implican incertidumbre y consiste en un gráfico dirigido y acíclico en el que los nodos representan las variables aleatorias y los arcos indican dependencias probabilísticas directas. Así, una o más distribuciones de probabilidad están asociadas con cada nodo en un modelo RB. Las distribuciones de probabilidad única (probabilidades previas) se asocian con nodos que no tienen arcos entrantes (sin padres) en la parte gráfica, mientras que los nodos restantes tienen múltiples distribuciones de probabilidad condicionales asociados con ellos, normalmente almacenados en tablas de probabilidad condicional (TPC).

Una RB es de hecho una representación compacta de la distribución de probabilidad conjunta (DPC) sobre las variables incluidas en el modelo; explotando aquellas independientes entre el resto de variables mediante representación gráfica, reduciendo en gran medida el número de distribuciones de probabilidad necesarias para especificar el DPC. La RB permite una respuesta eficiente de consultas relacionadas con probabilidades condicionales arbitrarias que involucran las variables en el modelo, como cuál es la probabilidad de una determinada enfermedad respecto del conjunto de síntomas observados.

Una de las debilidades particulares de la RB es el tamaño de los TPC; en general, el número de probabilidades necesarias para especificar un TPC crece exponencialmente en el número de nodos padres, lo que hace la especificación del TPC no sea práctica. Para enfrentar el problema de grandes TPC, un enfoque particularmente útil para aplicaciones de diagnóstico es el modelo noisy-OR que es una versión probabilística del OR lógico. Asumiendo que un nodo de efecto E (síntoma) tiene varias posibles causas $C = C_1, \dots, C_n$ (enfermedades). El noisy-OR asume que cada una de las causas es capaz de producir el efecto con alguna probabilidad p_i , pero mientras está ausente es incapaz de producir el efecto con certeza. Además, la probabilidad de fuga p_l se introduce y corresponde a la probabilidad del efecto estar presente dado que ninguna de las causas está presente. La distribución arbitraria del TCP se expresaría de la siguiente manera:

$$P(E = true | C) = 1 - \frac{\prod_{i \in C^+} (1 - p_i)}{1 - p_l}$$

Donde C^+ representa un subconjunto del conjunto de causas C que están en el estado actual. El noisy-OR reduce el número de probables distribuciones requeridas para especificar un TCP de exponencial a lineal respecto del número de padres. La obtención de las probabilidades requeridas para el noisy-OR es muy intuitiva: el parámetro p_i es la probabilidad de que la causa i produzca el efecto, asumiendo que todas las demás causas están ausentes; con el beneficio adicional que ofrece el noisy-OR de reducir la complejidad del razonamiento de los algoritmos.

A partir de este modelo de RB se asume para la obtención los datos médicos a partir de un modelo gráfico bipartito: una red de dos capas con los nodos de enfermedades como capa superior, y las observaciones (síntomas) ubicadas en la capa inferior, donde y los arcos procedentes de los nodos de la enfermedad van a las observaciones. Todas las observaciones se supone que son modelos de tipo noisy-OR, detallado en el párrafo precedente. Este tipo de modelo no es un concepto nuevo y se conoce en la literatura por sus siglas en inglés como BN2O.

El trabajo del equipo de expertos médicos facilita la creación del modelo de tipo BN2O a partir de tres tipos distintos de información:

- Conexión entre enfermedades y síntomas (observaciones) relacionados; donde el papel de los expertos es identificar los síntomas relevantes para un grupo de enfermedades dadas, pero limitadas a las adecuadas para un autodiagnóstico por parte de un paciente.
- Probabilidades previas sobre enfermedades: estos datos pueden obtenerse de la literatura médica y de bases de datos, pero a veces es necesario ajustarlos para reflejar con mayor precisión la población de usuarios esperada (por ejemplo, considerando dinámicas como en brotes de gripe)
- Probabilidad condicional de que ocurra un síntoma en presencia de una enfermedad dada.

Adicionalmente se incluyen algunas restricciones adicionales: información del tipo de que el embarazo es imposible en hombres e inclusión de factores de riesgo como antecedentes de cáncer, edad, etc.

Este enfoque tiene debilidades particulares, como el supuesto que las enfermedades son estadísticamente independientes entre ellas, lo cual es obviamente incorrecto, ya que algunas enfermedades a menudo coexisten; así como también las limitaciones del modelado en lo referente a factores de riesgo que influyen en las probabilidades de pre-existencia para muchas enfermedades.

La solución posee una herramienta separada para crear automáticamente el modelo de RB a partir de los datos obtenidos. La versión de la base de datos médica está en continua evolución y revisión y debido a la densa conectividad de la red, que es uno de los factores clave de rendimiento para la inferencia en la RB, el uso de algoritmos exactos y aproximados puede llegar a ser inmanejable o tardar demasiado en proporcionar resultados fiables (el objetivo es responder en menos de 1 segundo) para lo cual el sistema divide el modelo original en un conjunto de modelos más pequeños (sub-modelos), para los cuales el tiempo de consulta acumulativo es consecuente con la meta establecida con el menor número de sub-modelos posibles.

El problema de dividir el modelo en sub-modelos no conectados es que elimina información sobre dependencias entre nodos que se encuentran en diferentes modelos; por lo que para reducir este efecto el sistema hace lo siguiente:

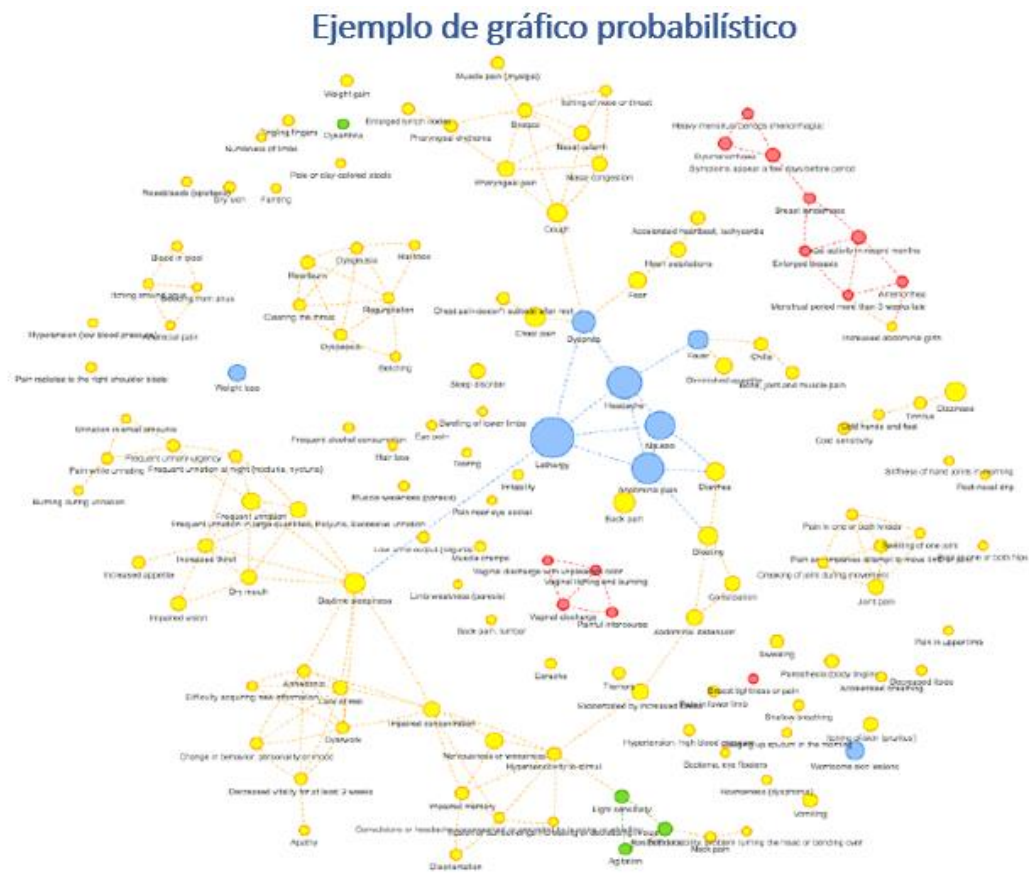
- Da preferencia a un número menor de sub-modelos
- Permite nodos de observación repetitivos en los diferentes sub-modelos
- Ubica los nodos de enfermedades que comparten los mismos síntomas en el mismo sub-modelo, a través de un criterio específico desarrollado para medir la interdependencia entre dos nodos de enfermedades según el número de síntomas comunes y los parámetros noisy-OR de éstos.

Esto gracias al desarrollo de un algoritmo especial de agregación jerárquica que sirve para identificar una división óptima de la modelo original en un subconjunto de modelos. El algoritmo comienza con un conjunto de modelos de RB y luego fusiona iterativamente algunos de los modelos en modelos más grandes. Este paso se realiza en el procedimiento de fusión de modelos (***fuse_models***). El procedimiento toma como entrada un conjunto de modelos V_{in} y para las salidas un conjunto de modelos V_{out} con menor cardinalidad (número de elementos).

Al comienzo del proceso se empieza con V_{in} con n modelos de RB, donde n es el número de enfermedades en la base de datos. Cada uno de los modelos iniciales de RB consta de un nodo de enfermedad única con sus nodos hijos que corresponden a todas las observaciones/síntomas para esta enfermedad. El algoritmo llama iterativamente al procedimiento ***fuse_models*** hasta que los modelos resultantes no superan los requisitos computacionales establecidos como criterios de eficiencia.

La operación ***fuse_models*** se define de la siguiente manera: La función distancia (u, v) es básicamente una medida de similitud entre dos modelos de RB u y v . En este contexto, la medida de similitud es el número de nodos de observación compartidos entre los dos modelos u y v . También se aplican medidas más elaboradas de similitud, como, por ejemplo, las basadas en fortalezas de las observaciones compartidas, que requerirían tomar en cuenta los parámetros noisy-OR. El procedimiento de combinar (u, v) fusiona dos modelos de RB en uno. Dado que ambos modelos son redes del tipo BN2O, la fusión equivale a combinar estructuralmente dos modelos, que es una tarea sencilla, y actualizar los parámetros de fuga O “noisy”. El algoritmo termina cuando el criterio de rendimiento basado en el tiempo de diagnóstico combinado se consigue. ^{ibídem}

Grafico 1. Ejemplo de gráfico probabilístico de Symptomate



5.1.4. Procesamiento de Lenguaje Natural

La API de Infermedica usa tecnología de procesamiento de lenguaje natural, lo que permite que Symptomate comprenda los conceptos clínicos (síntomas y factores de riesgo) mencionados por los usuarios como texto en lenguaje natural. Para el usuario el servicio es fácil de usar, pudiendo éste enviar el mensaje original incluso con errores gramaticales u ortográficos y el sistema (end-point) lo procesará y hará todo lo posible para detectar menciones de síntomas o factores de riesgo.

5.1.5. Descripción del proceso diagnóstico

El proceso de diagnóstico comienza realizando varias preguntas al usuario como su edad, sexo, peso y altura (para determinar el índice de IMC), después de lo cual se le pide especificar su síntoma principal y, opcionalmente, seleccionar en un modelo de las partes del cuerpo la localización de su motivo preocupación o malestar. Los datos recopilados ingresan como evidencia en todos los sub-modelos. En este punto se inicia el procedimiento iterativo y se genera una lista de prelación con las sospechas de posibles enfermedades. Para determinar la próxima pregunta a formular, se usa una puntuación basada en el valor de la información y la entropía cruzada, siendo la observación más valorada la determinante. Debido a que la misma observación puede aparecer en varios sub-modelos, donde además cada uno de estos sub-modelos puede tener una puntuación diferente, asumimos que debe tomarse el valor más alto de la puntuación. De esta manera, se le presenta al usuario la pregunta relacionada con la observación con la puntuación; teniendo éste siempre la opción de omitir la pregunta

(cuando no está seguro de la respuesta), en cuyo caso se le presenta la pregunta con la segunda puntuación más alta, y así cada respuesta a una pregunta se agrega al conjunto de pruebas y el proceso iterativamente.

El proceso se detiene cuando la probabilidad de diagnosticar la enfermedad es mayor del 70% y se han hecho al menos 9 preguntas, o en su defecto, un número fijo de pasos se ha alcanzado. De hecho, la condición de finalización (parada positiva) se establece empíricamente para mejorar la experiencia del usuario: Se ha determinado que los casos de diagnósticos exitosos (70% alcanzado) tienen una distribución normal en la campana de Gauss, y el percentil 95 de esta distribución se utiliza como un umbral para informar al usuario que: *“el sistema no es capaz de hacer un diagnóstico en ese momento y que proseguir en la investigación podría conducir potencialmente a un diagnóstico erróneo”*.

Si el proceso de análisis determina que el diagnóstico más probable cumple la condición de finalización (parada positiva), se presenta al usuario un ranking de enfermedades sospechosas. En la mayoría de los casos, la lista incluye de una a tres enfermedades.

5.1.6. Descripción del proceso de triaje

Además del criterio de valoración / **diagnóstico**, la API de Infermedica proporciona un criterio de valoración complementario / **triaje** que categoriza los casos de pacientes en función de la gravedad de los síntomas/observaciones informados y la gravedad de las enfermedades/condiciones probables.

El end point / **triaje** utiliza el mismo motor de diagnóstico que alimenta el end point /**diagnóstico** para calcular la clasificación de las posibles enfermedades/condiciones. El algoritmo de clasificación de triaje tiene en cuenta la gravedad de las condiciones más probables identificadas por el motor de diagnóstico, así como la aparición de cualquier síntoma alarmante, factor de riesgo o bandera roja.

La respuesta entregada por la solución contiene los siguientes elementos:

- Un nivel de triaje (5 niveles) del caso,
- una lista de síntomas/observaciones graves,
- una causa raíz que explica la lógica interna del algoritmo de triaje subyacente,
- una bandera que indica si se puede aplicar una teleconsulta con un médico en lugar de una visita presencial

El nivel de triaje se devuelve como un atributo **triage_level** con uno de los siguientes valores:

- Nivel I: Emergencia_ambulancia: los síntomas informados son muy graves y el paciente puede requerir atención de emergencia. El paciente debe llamar a una ambulancia ahora mismo,
- Nivel II: Emergencia: la evidencia reportada parece grave y el paciente debe acudir a un departamento de emergencias. Si el paciente no puede llegar al servicio de urgencias más cercano, debe llamar a una ambulancia,
- Nivel III: Consulta en 24h: el paciente debe consultar a un médico en un plazo de 24 horas. Si los síntomas empeoran repentinamente, el paciente debe acudir al servicio de urgencias más cercano,
- Consulta: el paciente puede requerir una evaluación médica y puede necesitar programar una cita presencial con un médico. Si los síntomas empeoran, el paciente debe consultar a un médico de inmediato.

- Auto-cuidado: se recomienda una consulta médica, pero no es estrictamente necesaria; el paciente debe observar sus síntomas y consultar a un médico si los síntomas empeoran dentro de las 24 horas.

Gráfico 2. Interfaz gráfico del triaje de Symptomate/Introducción

Introducción


Paciente

Síntomas

Regiones

Entrevista

Resultados



¡Hola!

Está a punto de realizar un chequeo médico breve (3 min), seguro y anónimo. Sus respuestas serán analizadas cuidadosamente y se le dirá las posibles causas de sus síntomas.

Introducción

Paciente


Síntomas

Regiones


Entrevista

Resultados

Por favor, seleccione su género



Mujer



Hombre

Gráfico 3. Interfaz gráfico del triaje de Symptomate/Paciente

Introducción

Paciente

Síntomas

Regiones

Entrevista

Resultados

Por favor, seleccione su edad

45

−

+

Por favor, seleccione todas las afirmaciones que se apliquen a usted

Escoja una respuesta en cada fila.

Tengo sobrepeso o estoy obeso

☐ Sí

☐ No

☒ No lo sé

Fumo cigarrillos

☐ Sí

☒ No

☐ No lo sé

He sufrido una lesión recientemente

☐ Sí

☒ No

☐ No lo sé

Tengo el colesterol alto

☐ Sí

☐ No

☒ No lo sé

Tengo hipertensión

☐ Sí

☒ No

☐ No lo sé

Gráfico 4. Interfaz gráfico del triaje de Symptomate /Síntomas y regiones

Introducción

Paciente

Síntomas

Regiones

Entrevista

Resultados

Añada sus síntomas

Por favor, utilice la búsqueda o haga click en el cuerpo.

Buscar, p. ej. dolor de cabeza

Q

Añada sus síntomas

Por favor, utilice la búsqueda o haga click en el cuerpo.

Buscar, p. ej. dolor de cabeza

Q

Dolor de cabeza

Fotofobia

Rotar modelo

Seleccione las regiones

Por favor, seleccione la región en la que vive y las regiones a las que ha viajado en los últimos 12 meses.

20

Gráfico 5. Interfaz gráfico del triaje de Symptomate /Entrevista Red Flags

Introducción
Paciente
Síntomas
Regiones
Entrevista
Resultados

¿Tiene alguno de los síntomas siguientes?

☐ Ilusión de lo que le rodea tiene un tamaño diferente (más grande o más pequeño) del que realmente tienen

☐ Cuello rígido

☐ Dolor alrededor de los ojos

☐ Agitación

☐ Dolores de cabeza en serie

☐ Dificultad para hablar

Gráfico 6. Interfaz gráfico del triaje de Symptomate /Síntomas

Introducción
Paciente
Síntomas
Regiones
Entrevista
Resultados

¿Cuál es la intensidad de su dolor de cabeza?

Leve 1 2 3 4 5 6 7 8 9 10 Insoportable

¿Ha tenido dolores de cabeza similares de manera frecuente en los últimos tres meses?

Tiene que escoger una opción.

☒ Sí

☐ No

Habitualmente, ¿cuánto tiempo suelen durar sus dolores de cabeza?

Tiene que escoger una opción.

☐ De segundos a minutos

☒ De pocos minutos a 4 horas

☐ De 4 horas a 3 días

☐ Más de 3 días

No entiendo esta pregunta

Introducción
Paciente
Síntomas
Regiones
Entrevista
Resultados

¿Dónde está situado su dolor de cabeza?

Tiene que escoger una opción.

☐ Ocupa toda mi cabeza

☒ En un lado de mi cabeza

☐ En la parte frontal de mi cabeza

☐ En la parte de atrás

☐ A los costados, en la sien

No entiendo esta pregunta

¿Tiene alguno de los síntomas descritos abajo?

Escoja una respuesta en cada fila.

Náuseas ☐ Sí ☒ No ☐ No sé

Vómitos ☐ Sí ☒ No ☐ No sé

Diarrea o heces al día ☒ No ☐ No sé

Evaluando sus respuestas...

Introducción
Paciente
Síntomas
Regiones
Entrevista
Resultados

¿Hay algo que empeore su dolor de cabeza?

Escoja una respuesta en cada fila.

Empeora en la mañana ☐ Sí ☐ No ☒ No sé

El estrés ☒ Sí ☐ No ☐ No sé

Inclinar la cabeza hacia abajo ☐ Sí ☒ No ☐ No sé

No entiendo esta pregunta

Introducción
Paciente
Síntomas
Regiones
Entrevista
Resultados

¿Cómo describiría su dolor de cabeza?

Escoja una respuesta en cada fila.

Pulsátil o palpitante ☒ Sí ☐ No ☐ No sé

Se siente «punzante» o como un «taladro» ☐ Sí ☒ No ☐ No sé

Se siente como presión alrededor de mi cabeza ☐ Sí ☐ No ☒ No sé

No entiendo esta pregunta

Gráfico 7. Interfaz gráfico del triaje de Symptomate /Resultados del triaje



5.1.7. Arquitectura del Sistema

El sistema de software responsable de proporcionar el diagnóstico es implementado como un servicio web. En el corazón de la herramienta se encuentra un sistema distribuido en paralelo a múltiples motores del tipo RB que son responsables de realizar consultas a RB individuales. Cada uno de estos motores se basa en el software de RB de propósito general conocido como SMILE¹⁹.

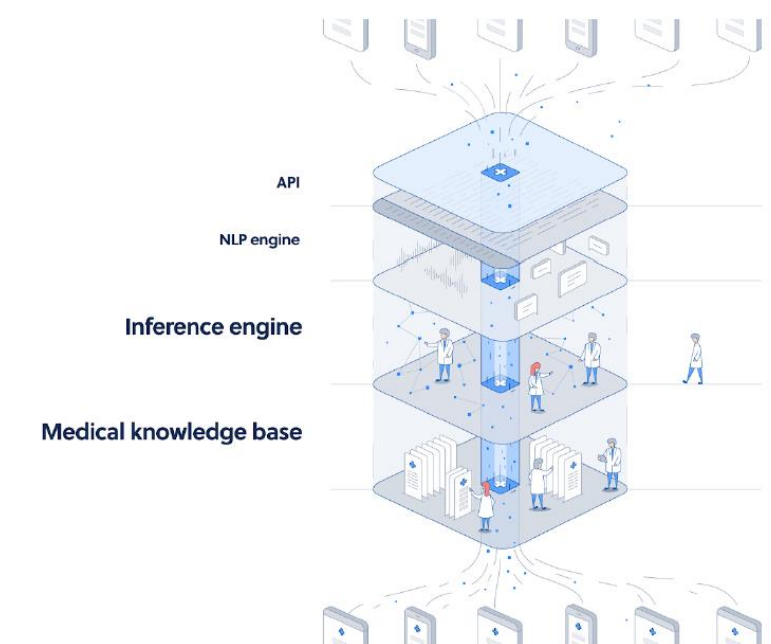
Para lograr escalabilidad y confiabilidad, los motores de RB son apátridas: toda la información necesaria está codificada en la consulta (query) y en los datos de resultado. El sistema distribuido está diseñado como un sistema de dos capas con la primera capa responsable de distribuir las consultas entrantes entre n servidores a nivel de hardware, que se implementa como un round-robin²⁰. La segunda capa divide cada consulta entre varios motores de RB. Esto lo hace el *Model Manager*, una pieza específica de software que se encarga de distribuir la carga de consultas de manera eficiente e integrando los resultados y mantiene en la memoria una serie de *Modelos de RB precargados* en forma de instancias de motores de tipo SMILE. La funcionalidad de diagnóstico se expone a través de una API web que se pone a disposición de los desarrolladores. Los sitios web

¹⁹ SMILE es el motor de inferencia desarrollado en el Laboratorio de Sistemas de Decisiones de la Universidad de Pittsburgh, y es utilizado para realizar inferencias de diagnóstico (<http://genie.sis.pitt.edu>).

²⁰ **Planificación Round-robin:** uno de los algoritmos de planificación de procesos más simples dentro de un sistema operativo que asigna a cada proceso una porción de tiempo equitativa y ordenada.

que utilizan Symptomate, Infermédica® son entidades independientes y se comunican con el sistema de diagnóstico a través de llamadas a una API web.

Gráfico 8. Arquitectura del sistema Symptomate



5.1.8. Validación externa del algoritmo de Infermédica®

En un estudio observacional reciente realizado en Australia, se evaluaron 36 symptom checkers que aportan un diagnóstico clínico o recomendación de triaje. El estudio consistió en ponerlos a prueba utilizando apuntes clínicos de condiciones o enfermedades a través de 1.170 pruebas dirigidas a evaluar su capacidad diagnóstica y 688 para evaluar su capacidad de triaje. El estudio comparó symptom checkers basados en IA con otros, obteniendo mejores resultados en los que usan IA (los symptom checkers que utilizan algoritmos de inteligencia artificial presentaron el diagnóstico correcto en primer lugar en el 46% de las pruebas (IC del 95%, 40-57%), en comparación con el 32% (IC del 95%, 26-38%) para otros symptom checkers).

El algoritmo de Infermédica® logró muy buenos resultados en precisión diagnóstica. Siendo el algoritmo que más acertó en ofrecer el diagnóstico correcto. En 4 de cada 31 casos obtuvo el diagnóstico correcto entre los 3 primeros diagnósticos. En 25 de cada 31 casos obtuvo el diagnóstico correcto entre los 10 primeros diagnósticos. Y en 19 de cada 31 casos la aplicación aportó su primer posible diagnóstico siendo este el correcto.²¹

En las conclusiones resaltan el hecho de que el diagnóstico no consiste en una única evaluación, sino es proceso que requiere conocimiento, experiencia, examen y pruebas clínicas tomando en consideración además el paso del tiempo, situación imposible de replicar en una sola interacción en línea. Por otro lado, destaca que los consejos de triaje pueden ser la función más importante de las herramientas de symptom checkers en línea. Es importante destacar que todos los symptom checkers incluidos en el estudio advierten que su servicio no sustituye a la consulta de un médico.

²¹ Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Medical Journal of Australia. Junio de 2020;212(11):514-9.

Pese a esta apreciación, en el mismo estudio, al comparar la precisión en su recomendación de triaje, Symptomate, así como la mayoría de los symptom checkers estudiados tiende a huir del riesgo, sobre todo en los casos más leves. Mientras en los casos que requieren cuidados de emergencia y de urgencia obtuvo resultados por encima de la media, en los casos no urgentes y aptos para el autocuidado obtuvo resultados por debajo de la media. De 11 casos que requerían cuidado con carácter emergente, clasificó correctamente a 9. Y de 11 casos que requerían cuidados urgentes, clasificó correctamente a 6.

A partir de esta información se podría inferir que Symptomate como la mayoría de symptom checkers tiende a sobre triar (siendo el paciente es categorizado en un nivel de mayor urgencia al que le corresponde), produciendo mayor carga al sistema de salud, pero ganando seguridad a cambio. Es importante destacar que el sobre-triaje no comporta ningún riesgo para el paciente a diferencia del sub-triaje (cuando un paciente es clasificado en un nivel de menor urgencia al que le corresponde), situación que se busca evitar más en un sistema de triaje por sus evidentes consecuencias: condicionar un alargamiento del tiempo de inicio de la visita médica y un riesgo para el paciente, tanto de deterioro clínico como de efectos adversos. Es muy importante tener en cuenta que el contexto remoto (GRP) es muy diferente del de una Sala de Urgencias Hospitalaria donde sí se afecta la atención sobre otros pacientes y se podría condicionar un inadecuado consumo de recursos.

En una SHU para evitar el sub-triaje, se ha introducido el concepto de reevaluación periódica. Se ha establecido que los pacientes de nivel I requieren reevaluación continua, los de nivel II cada 15 minutos, los de nivel III cada 30 minutos, los de nivel IV cada 60 minutos y los de nivel V cada 120 minutos.²²

Para el contexto de GRP, al momento de implementar un sistema de triaje pre-hospitalario asistido por IA con una herramienta de verificación de síntomas, se podría considerar incluir en el flujo del proceso de atención una acción de outbound²³ desde el centro de llamadas que cumpla con la función de reevaluación continua remota.

El estudio citado concluye que, a pesar de sus limitaciones y la falta de regulación de los programas de salud en línea, los symptom checkers pueden ser útiles. Los datos epidemiológicos se pueden rastrear en sitios como WebMD, proporcionando datos oportunos a los profesionales de la salud. Los usuarios pueden informarse sobre su propia salud, mejorando potencialmente las relaciones médico-paciente o incluso mejorar los flujos asistenciales cuando éstos están directamente implantados en los servicios de salud. A medida que el público se vuelve más educado, activo e informado sobre la salud, las herramientas en línea serán un recurso cada vez más popular y un recurso importante para reducir la carga/presión sobre los sistemas de atención de la salud, en particular en la atención de emergencia, a través de lograr que los consejos de triaje de las herramientas en línea orienten con la mayor precisión posible y rápidamente a las personas hacia la atención adecuada.

²² Gómez Jiménez J. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias. Emergencias. 2003; 15: 165-174

²³ **Outbound:** o llamadas salientes realizadas por agentes para seguimientos; estos contactos con los usuarios se realizan desde múltiples canales, no solo llamadas telefónicas.

5.2. Casos de uso del algoritmo de Infermédica®

A raíz de la pandemia por COVID-19, la estrategia de muchos sistemas de salud a nivel mundial se dirigió hacia la implantación de sistemas desarrollados en remoto para ayudar a la gente a sobrellevar sus preocupaciones y dudas frente al apareamiento de sintomatología compatible con la infección respiratoria descrita para este coronavirus. Autoridades como la OMS, CDC, Ministerios de Salud de diferentes países, Cruz Roja, hospitales y compañías de seguros brindan chatbots gratuitos que asisten a la población para informar sobre el coronavirus, para coordinación de recursos y flujos de asistencia sanitaria y urgencias. Dentro de este grupo se encuentran bots para verificación de síntomas o symptom checkers cuyo número se ha incrementado considerablemente.

Varios estudios se han puesto en marcha para analizar el impacto de este tipo de soluciones en el sistema sanitario al igual que la aceptación de los mismos entre la población (p.e. la experiencia del usuario UX), donde queda clara la potencialidad de su implementación tanto como auxiliares en la atención remota, como elemento a tener en cuenta al momento de plantear la transformación digital de un sistema de salud.²⁴

De ahí que, se hayan considerado también otros usos de los verificadores de síntomas o symptom checkers como por ejemplo el papel de los mismos en la distribución de flujos de usuarios con miras a mejorar el funcionamiento de los centros médicos en el ámbito de la atención primaria de salud.²⁵

En el estudio de Munsch et.al.,²⁶ donde comparan 10 verificadores de síntomas de COVID-19 basados en la web a través de la evaluación de 50 informes de casos de COVID-19 y 410 casos de control que no son de COVID-19; el algoritmo de infermédica® demostró un buen equilibrio entre sensibilidad y especificidad siendo uno de los dos algoritmos con mejores resultados. Los resultados se reportaron como sensibilidad, especificidad, puntuación F1 y coeficiente de correlación de Matthews (MCC).

El objetivo del estudio fue el de evaluar la precisión diagnóstica de los verificadores de síntomas de COVID-19 basados en la web con los siguientes resultados:

La tarea de clasificación de los diferentes symptom checkers para clasificar entre COVID-19-positivo y COVID-19-negativo para casos de “alto riesgo” de entre los 460 casos de prueba arrojó (ordenados por puntaje F1): Symptoma (F1 = 0.92, MCC = 0.85), **Infermedica (F1 = 0.80 , MCC = 0.61)**, Centros para el Control y la Prevención de Enfermedades (CDC) de EE. UU. (F1 = 0.71, MCC = 0.30), Babylon (F1 = 0.70, MCC = 0.29), Cleveland Clinic (F1 = 0.40, MCC = 0.07), Providence (F1 = 0.40, MCC = 0.05), Apple (F1 = 0.29, MCC = -0.10), Docyet (F1 = 0.27, MCC = 0.29), Ada (F1 = 0.24, MCC

²⁴ Hoehn and Bongard-Blanchy; Heuristic Evaluation of COVID-19 Chatbots; Luxembourg National Research Fund (FNR), SLANT, 13320890; https://orbilu.uni.lu/bitstream/10993/44913/1/2020-12-02Conversational_UX_analysis_of_COVID_19_chatbots.pdf

²⁵ Nan-Guang Tan et al., Virtual care to increase military medical centre capacity in the primary health care setting: A prospective self-controlled pilot study of symptoms collection and telemedicine; J Telemed Telecare 2020 Oct 4; 1357633X20959579. doi: 10.1177/1357633X20959579. <https://pubmed.ncbi.nlm.nih.gov/33016187/>

²⁶ Munsch N, Martin A, Gruarin S, Nateqi J, Abdarrahmane I, Weingartner-Ortner R, Knapp B; Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study, J Med Internet Res; 2020;22(10):e21299; URL: <https://www.jmir.org/2020/10/e21299>; DOI: 10.2196/21299

= 0.27) y Your.MD (F1 = 0,24, MCC = 0,27). Para "alto riesgo" y "riesgo medio" combinados, el rendimiento fue: Symptoma (F1 = 0.91, MCC = 0.83) **Infermedica (F1 = 0.80, MCC = 0.61)**, Cleveland Clinic (F1 = 0.76, MCC = 0.47), Providence (F1 = 0,75, MCC = 0,45), Your.MD (F1 = 0,72, MCC = 0,33), CDC (F1 = 0,71, MCC = 0,30), Babylon (F1 = 0,70, MCC = 0,29), Apple (F1 = 0,70, MCC = 0,25), Ada (F1 = 0,42, MCC = 0,03) y Docyet (F1 = 0,27, MCC = 0,29).

Por otra parte, el algoritmo de infermedica® fue mencionado por Hoehn and Bongard-Blanchy entre los primeros 3 de una lista de 24 soluciones entre chatbots y symptom checkers; donde las investigadoras evaluaban la experiencia de usuario para el contexto de la interfaz conversacional, tomando en consideración como fortalezas la retroalimentación inmediata, lenguaje familiar, coherencia redacción y diseño visual, así como deficiencias el poco control y libertad del usuario, falta menú permanente y opciones de ayuda, falta de comprensión del contexto y las capacidades de gestión de la interacción).^{ibidem 24} Calificando a la solución de infermedica® como aquella con la interfaz gráfica de usuario más atractiva.

Finalmente, en el contexto de la telemedicina, el symptom checker de infermedica® fue bien valorado en como parte del flujo de asistencia del primer nivel de atención en el estudio piloto prospectivo auto-controlado de recopilación de síntomas y telemedicina conducido en un centro médico militar de Singapur por Nan-Guang Tan et- al.; donde a más de validar la seguridad de la telemedicina en atenciones remotas, se demuestra con éxito el uso del symptom checker como parte del flujo de asistencia donde el 85,7% de los pacientes estuvo de acuerdo en que el verificador de síntomas era fácil de usar y donde además tanto los pacientes como los médicos coincidieron que el verificador de síntomas permitió consultas más claras y rápidas.^{ibidem 25}

Existen también varias experiencias del uso del algoritmo de IA desarrollado por Infermedica que detallamos a continuación, y que se presentan como estudios de caso en la web de la compañía ²⁷ por parte de proveedores de servicios de salud privados con implementaciones exitosas en uso de Symptomate.

PZU Zdrowie, uno de los proveedores privados de atención médica más grandes de Polonia, tiene implementado un sistema de Triage a través de un Call Center habilitado por IA (symptomate para Call Centers) para asistir a su personal médico y no médico a clasificar y guiar a los pacientes hacia la atención más adecuada y ayudar a los pacientes a obtener la atención médica adecuada evitando el tener visitas innecesarias a múltiples especialistas.

El triaje asistido por IA de Symptomate ayuda a los operadores del centro de llamadas a realizar un diagnóstico previo y una clasificación rápida y precisa, detectando casos urgentes rápidamente y previniendo citas innecesarias al recomendar los servicios sanitarios más adecuados o re-dirigiendo las visitas presenciales a teleconsultas a través de un proceso estandarizado de entrevistas de triaje.

Los datos recopilados hasta 2019 demuestran que PZU Zdrowie re-dirigió al 37% de los pacientes con nuevos problemas de salud a consultas de telemedicina, el volumen de llamadas recibidas en 2019 fue de 2.4 millones con un máximo de 15 mil llamadas al día.

Médis, la aseguradora de salud más reconocida de Portugal, también tiene la implementación de la herramienta Call Center Triage de Infermedica en el Clinical

²⁷ <https://infermedica.com/case-studies/pzu-zdrowie> , <https://infermedica.com/case-studies/medis>

Contact Center de Médis. Esto ha permitido a Médis mejorar la experiencia de sus pacientes y de su personal de enfermería, y agilizar los procesos operativos al mejorar las capacidades de triaje mejorando el proceso de entrevistas y evitando visitas y citas de atención urgente innecesarias.

Desde la implementación inicial del sistema hace aproximadamente un año, Médis ha aplicado la solución de Infermedica en aproximadamente 24.000 procesos de triaje, con el 99,4% de los síntomas de los clientes cubiertos por la herramienta. Además es un canal para el acceso a otros servicios proporcionados por Médis como, por ejemplo, conexión al servicio de telemedicina para brindar servicios de atención adaptados a las necesidades del paciente.

La comparación con la intención inicial de los clientes antes de llamar a Médis Clinical Contact Center muestra que la herramienta ha evitado visitas urgentes innecesarias (ha bajado de 17% A 8%) y ha aumentado la frecuencia de la opción de autocuidado (incremento de 17% a 35%). Además, la implementación de esta nueva solución de triaje clínico ha llevado a una mejor experiencia para las el personal de enfermería y los usuarios de Médis.

Dadas las capacidades de aprendizaje automático de symptomate para call centers, el sistema mejora con el tiempo y brinda sugerencias y recomendaciones más precisas. Además, la nueva solución permite una mejor experiencia de usuario (UX) para el personal de enfermería de Médis.

La implementación en Médis ha permitido a Infermedica aumentar la capacidad de su solución de Triage para recomendar las acciones adecuadas para determinados síntomas. El equipo de Infermedica enriqueció la base de datos de factores de riesgo, condiciones y síntomas, con un enfoque especial en pediatría. Durante el proyecto, el equipo médico de Infermedica, con el apoyo de Médis, agregó y modificó aproximadamente 200 factores médicos en la base de datos.

5.3. Evaluación de Symptomate, Infermedica®

Los usuarios de apps de salud demandan cada vez más posibilidades de empoderamiento, por lo que es importante garantizar que la información a la que acceden es adecuada y de calidad. Por otro lado, se requiere validar la pertinencia de incluir herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos en los servicios de salud, así como en la mejora de la gestión de los mismos. Para estos dos objetivos se han empleado dos metodologías de evaluación: iSYScore de la fundación iSYS y las recomendaciones de Yun Liu. et al. sobre cómo evaluar herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos médicos

5.3.1. Aplicación del método de valoración de aplicaciones móviles de salud en español ISYScore de la fundación iSYS a Symptomate de infermédica®

A continuación se presenta el resultado de la valoración según el ISYScore ^{ibidem16}, a partir de la revisión de las publicaciones existentes sobre Symptomate y por acuerdo entre los evaluadores, en las 3 diferentes dimensiones: 1) popularidad e interés (puntuación máxima de hasta 11 puntos); 2) confianza y calidad (puntuación máxima de hasta 18 puntos), y 3) utilidad (puntuación máxima de hasta 18 puntos).

Tabla 1: Valoración ISYScores aplicada a Symptomate

Interés popular	Valoración	Puntuación
Usuarios puntúan positivamente la app	Si	4/4
Disponible en 2 plataformas (iOS y Android)	Si	3/3
Declarado de interés por alguna asociación de afectados	No	0/4
Confianza		
Sellos de calidad reconocidos	Marcado CE, PN-EN ISO 13485, cumple GDPR	18/18*
¿Está vinculada a bibliotecas de administraciones públicas?	No	
Utilidad (validez externa)		
Investigación que demuestre que la app es útil	Poca evidencia para los symptom checkers en general y para Symptomate de Infermédica®. Basado en estudios observacionales	
Investigación sobre una muestra pequeña de usuarios (menos de 30 usuarios)	Sobre estudios observacionales o pacientes no reales	3/3
Declaración de una sociedad científica o asociación de afectados	La NHS y la clínica Mayo han puesto en marcha sistemas de Symptom checkers, aunque no basados en Symptomate de Infermédica®. Experiencias positivas trabajando con aseguradoras privadas y proveedores de salud como Medis (Portugal) y PZU Zdrowie (Polonia) para apoyo online de triaje telefónico. 450 organizaciones y los ministerios de salud de Polonia y Ucrania actualmente utilizan Symptom checker COVID-19 Risk Assesment de Infermédica® con buenos resultados.	2.5/3
Proporciona información	En su web, publicaciones y en la aplicación.	3/3
Proporciona seguimiento útil en salud (trackers). Conecta con un equipo de salud	No proporciona trackers, pero si está preparada para conectar con la historia clínica electrónica o personal de salud de apoyo en triajes telefónicos.	1,5/3
Enlaza con otros afectados o usuarios	No	0/3
Utiliza juegos para promocionar la salud	No	0/3

* Este indicador valora la presencia de sellos de calidad reconocidos o apps vinculadas a bibliotecas de las administraciones públicas, en cuyo caso las app obtienen directamente los 18 puntos disponibles. Symptomate obtuvo por tanto la puntuación completa

El iSYScore proporciona una puntuación orientativa, pero no es un sello de calidad o de seguridad. Es solamente una guía que evalúa apps para el público en general ---no profesional--- de salud y no valora a una app dentro de un proceso asistencial. La puntuación se basa principalmente en la descripción de la aplicación proporcionada por el promotor y la posibilidad de contrastarla.

El resultado de este análisis muestra que Symptomate obtuvo una puntuación de 35 puntos; siendo la puntuación máxima 47 puntos y 34 puntos los que recibió la App mejor valuada con ISYScore en el artículo de Grau et al., por lo que se considera una muy buena valoración.

5.3.2. Aplicación de las recomendaciones de Yun Liu. et al. sobre cómo evaluar herramientas basadas en aprendizaje automático para el establecimiento de diagnósticos médicos

En los últimos años, se han desarrollado muchas herramientas nuevas de diagnóstico clínico utilizando complicados métodos de aprendizaje automático. Independientemente de cómo se obtenga una herramienta de diagnóstico, debería ser evaluada utilizando un proceso de 3 pasos para adquirir, validar y establecer la eficacia clínica de la misma. Las herramientas basadas en aprendizaje automático deben evaluarse según el tipo del modelo de aprendizaje automático utilizado y su idoneidad para el tipo de datos de entrada y el conjunto de datos referencia.

Los modelos de aprendizaje automático suelen tener configuraciones adicionales previamente especificadas llamadas hiperparámetros, que deben ajustarse a un conjunto de datos independientes a su vez de aquel conjunto de datos de validación. Al mismo tiempo, el resultado de este conjunto de validación, con el que se evalúa el modelo, se denomina referencia estándar y debe evaluarse rigurosamente frente a un estándar de oro aceptado o calificación experta.

La Guía propuesta por Yun Liu. et al.^{ibídem17}, pretende facilitar la comprensión de los médicos sobre modelos de aprendizaje automático a través de: (1) enfatizar la importancia de una adecuada validación del modelo de aprendizaje automático y resaltar las diferencias en este proceso relativo a la validación de métodos más tradicionales de desarrollo de modelos estadísticos; (2) revisar los conceptos básicos de los modelos de aprendizaje automático; y (3) revisar cómo deberían implementarse los modelos de aprendizaje automático en medicina clínica. Con este marco de referencia, presentamos a continuación el análisis del sistema de aprendizaje automático Symptomate desarrollado por Infermedica® utilizando reglas similares a las que se aplican a cualquier sistema diseñado para la ayuda en la toma de decisiones para el diagnóstico médico.

Listas de verificación

Tabla 2. Listas de verificación de artículos que analizan Symptomate como herramienta de aprendizaje automático para hacer predicciones clínicamente relevantes para triaje pre-hospitalario

Descripción de la tarea de predicción	Symptomate de la empresa Infermedica® proporciona una API para el triaje de pacientes y el diagnóstico médico preliminar pensado para implementar un verificador de síntomas inteligente o un formulario de admisión de pacientes adaptado a cualquier aplicación de salud de un proveedor de atención sanitaria.
---------------------------------------	---

Datos	Input	Entre 10 y 20 preguntas al usuario relacionadas con su sexo, edad, factores de riesgo, países visitados recientemente y síntomas. Corresponde a datos que se puede responder fácilmente el usuario.
	Output	Ranking de las enfermedades sospechadas, siendo en la mayoría de casos entre una y tres enfermedades, y una recomendación de triaje asociada. La recomendación de triaje evidentemente posee gran relevancia para la gestión clínica.
	Etiquetas	El estándar de referencia tiene alta fiabilidad: la selección de conjunto de síntomas para cada enfermedad, realizada por un experto clínico, de casos publicados en revistas médicas de prestigio.
	Población de Pacientes	Múltiples fuentes provenientes de la literatura. 1.800.000 personas anónimas utilizan Symptomate. Para el mismo algoritmo de Infermedica en la versión de triaje para call center, el volumen de llamadas recibidas en 2019 fue de 2.4 millones con un máximo de 15 mil llamadas al día en Polonia y en para el último año en Portugal 24 mil llamadas.
	Desarrollo / División de validación	Solo un data set para el desarrollo y un data set para validación. No se describe tuning set o cross validation set. En este caso se utilizó un punto operativo, umbral o punto de corte (probabilidad de la enfermedad más posible superior al 70% y al menos 9 preguntas respondidas) ⁹ , a partir del cual, si el output del algoritmo está por encima del punto de corte, la característica que el modelo está tratando de identificar se considera presente. A pesar de las modificaciones posteriores basadas en la aproximación suave de evidencia ²⁵ . Esta metodología trae problemas de calibración, generando falsos positivos cuando se valida en la población general con solo pocos casos de pacientes con la entidad clínica que se está intentando identificar; sin embargo, el objetivo del algoritmo no es un diagnóstico preciso, si no realizar un triaje, que preferiblemente tenga aversión al riesgo.
	Cantidad de datos en el conjunto de desarrollo	La base de conocimientos médicos (Metabase) comprende más de 1260 síntomas, 680 condiciones médicas, 486 pruebas de laboratorio y 133 factores de riesgo que se interconectan entre sí que permiten un pre-diagnóstico preciso de los pacientes y el triaje pre-hospitalario.
	Cantidad de datos en conjunto de validación	Miles de casos clínicos bien documentados son seleccionados de la literatura científica y sirven para probar regularmente el motor de diagnóstico. La precisión se confirma cuando el motor de diagnóstico clasifica la condición dentro de las 3 o 5 primeras posiciones, según la complejidad del caso, y la clasificación de diagnóstico diferencial. A 1/06/2021 la tasa de aprobación fue 95,98%
Machine Learning (ML)	Método	<p>Aprendizaje supervisado, redes bayesianas múltiples, distribuidas y paralelas. Se asume que todas las observaciones se comportan con el modelo de probabilidad noisy-QR.</p> <p><u>Parámetros:</u></p> <ol style="list-style-type: none"> 1. Conexión entre las enfermedades y las observaciones: conjunto de síntomas relevantes para cada enfermedad. 2. Probabilidades previas de las enfermedades: datos obtenidos de la literatura médica y bases de datos. En algunos casos se ajustan para que reflejen más

		<p>precisamente a la población usuaria (por ejemplo, el dinamismo de las epidemias de gripe).</p> <p>3. Probabilidad condicional de que exista un síntoma dada la enfermedad.</p> <p>4. Se incluyeron adicionalmente restricciones utilizando la aproximación suave de evidencia²⁸: información como el hecho de que el embarazo es imposible en hombres, y factores de riesgo tales como antecedentes personales de cáncer, edad, etc. Tal aproximación tiene debilidades como asumir que las enfermedades son independientes estadísticamente entre sí, lo que obviamente es incorrecto ya que muchas enfermedades coexisten.</p>
	Proceso de Entrenamiento	Requiere de expertos que predefinan las características discriminativas y ayudar activamente al algoritmo a identificarlas. Algoritmo antiguo de Machine Learning.
	Aumento/Incremento de datos	La base de conocimientos médicos se revisa y amplía continuamente de acuerdo con las demandas específicas de pacientes y médicos.
	Hiperparámetros que fueron optimizados	<p><u>Sub-modelos:</u></p> <p>. Se prefirió dividir el modelo en pequeños sub-modelos no conectados para manejar la densa red de conexiones de enfermedades y síntomas, disminuyendo el tiempo de consulta.</p> <p>. Los modelos se ajustaron para ser escasos, se permitió la repetición de observación a lo largo de los nodos de diferentes sub-modelos y se trató de colocar los nodos de enfermedad que comparten los mismos síntomas en el mismo sub-modelo.</p> <p>Se desarrolló un criterio especial para medir la interdependencia entre dos nodos de enfermedades basado en el número de síntomas y el parámetro noisy-QR de los síntomas compartidos.</p> <p>Para crear los sub-modelos se desarrolló un algoritmo de agregación.</p> <p><u>Punto operativo, umbral o punto de corte:</u> probabilidad de la enfermedad más posible superior al 70% y al menos 9 preguntas respondidas)⁹, a partir del cual, si el output del algoritmo está por encima del punto de corte, la característica que el modelo está tratando de identificar se considera presente</p>
	Uso de datos para afinación o tuning	Solo un data set para el desarrollo y un data set para validación. No se describe tuning set o cross validation set. No aplica.
	Tiempo necesario para aplicar el modelo por end-point de datos	Menos de 1 segundo.

²⁸ D'Ambrosio B. Symbolic probabilistic inference in large BN2O networks. In: Proc. 10th Conf. on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann; 1994. pp. 128-135.

Evaluación	Criterios de exclusión/inclusión del conjunto de datos	No aplica
	Métrica de rendimiento	<p>Fue estándar el uso de la media con IC del 95% para describir los eventos de precisión diagnóstica y precisión de triaje en las publicaciones estudiadas.</p> <p><u>Métricas relacionadas con la precisión diagnóstica del algoritmo:</u></p> <p>G Hill et al: métrica expresada como la media con un IC del 95% de la precisión diagnóstica. (Análisis de varianza de una vía (ANOVA) y prueba t de muestras independiente)</p> <p>Semigran et al: métrica expresada como la media con un IC del 95% de la precisión diagnóstica. (prueba de X^2)</p> <p><u>Métricas relacionadas con la precisión de triaje:</u> métrica expresada como la media con un IC del 95% de la precisión diagnóstica. (Análisis de varianza de una vía (ANOVA) y prueba t de muestras independiente)</p> <p>Semigran et al: Métrica expresada como media con IC 95%</p>
	¿Se usó un conjunto de validación independiente?	Si. Miles de casos clínicos bien documentados independientes.
	Métrica de comparación humana	Protocolos de Schmitt-Thompson, McKesson Directrices de clasificación y clasificación de Manchester
	Comparador humano	<p>Gran desviación entre publicaciones y con la práctica real, desde no utilizar comparador, o utilizar el triaje hecho por personal de salud o el diagnóstico realizado por un clínico. Sugerimos el triaje telefónico como comparador, por ser el caso más ajustado a la práctica real.</p> <p>Kellermann et al: diagnósticos hechos por clínicos. Poote et al: diagnóstico hecho por MAP.</p> <p>Hill et al (viñetas): cada viñeta tenía un diagnóstico y triaje; esto fue comparado con el resultado del symptom checker.</p> <p>Semigran et al (viñetas): cada viñeta tenía un diagnóstico y triaje; esto fue comparado con el resultado del symptom checker.</p> <p>Berry et al: ninguno.</p>
	Humanos vs el rendimiento del modelo de ML	La recomendación de triaje del modelo ML es muy sensible y poco específica, en comparación al triaje realizado por personal de salud. El modelo tiende a evitar el riesgo, dando más seguridad clínica.
	Brecha de desempeño en la validación del desarrollo	No aplica.
	Subgrupo / análisis de sensibilidad	Symptomate utiliza subconjuntos de datos estructurados disponibles para analizar y extraer información adicional que podría aumentar la información proporcionada por la

		megabase de datos médicos. Estos datos incluyen, por ejemplo, una muestra de 500.000 pacientes de uno de sus clientes, a más de los 400.000 resúmenes de visitas ambulatorias del conjunto de datos CDC.gov.
--	--	--

Tabla 3. Descripción general y características Symptomate Infermedica®

Tipo de modelo y ejemplos de uso	Modelo de Probabilístico que permite superar algunas de las imitaciones de los sistemas tradicionales basados en reglas y árboles de decisiones.
Descripción	<p>Actualmente, Symptomate de Infermedica aprende principalmente de expertos capacitados en edición de contenido médico que revisan la literatura disponible basada en evidencia (como Best Practice de British Medical Journal, New England Journal of Medicine, directrices de la OMS y los CDC, y numerosos artículos con un factor de impacto alto) para proporcionar información sobre las condiciones, síntomas, factores de riesgo y sus relaciones probabilísticas. Además de la revisión de los protocolos de triaje existentes, como los Protocolos de Schmitt-Thompson, McKesson Directrices de clasificación y clasificación de Manchester para garantizar la coherencia y la calidad de la plataforma. Hasta la fecha, ha dedicado más de 40.000 horas de trabajo de los médicos editores a construir y validar el contenido médico.</p> <p>El sistema es determinista: proporcionará la misma salida con la misma entrada.</p>
Tipo común de entrada de datos	Además del conocimiento basado en expertos, Infermedica utiliza el análisis de datos para recopilar información de conjuntos de datos médicos anonimizados y utiliza bucles de retroalimentación para mejorar tiempo de respuesta y aprendizaje de la máquina.
Número típico de parámetros	$10^2 \sim 10^4$
Número de hiper parámetros	10^1
Cantidad de puntos de datos requerido	$10^2 \sim 10^6$

A continuación se responden las preguntas de la guía de Yun Liu. et. al.,^{ibidem17}.

Análisis de la validez: evaluando los resultados

¿Qué tan serio es el riesgo de sesgo? ¿Existió un comparador ciego, independiente, y ese comparador fue un estándar de referencia?

En la fase de elaboración del algoritmo el comparador fueron casos publicados en la literatura, seleccionados por un editor médico experto, de fuentes reconocidas y basadas en la evidencia. Los casos publicados en la literatura provenían de fuentes de buena calidad e independientes. Sin embargo, no tenemos datos para comprobar si se utilizó el estándar de referencia en todos los casos. Por lo que no podemos conocer el riesgo de sesgo del modelo.

¿La muestra de pacientes incluyó un espectro apropiado de pacientes a quien se le aplicaría la herramienta de triaje en la práctica?

Infermédica® no aporta información específica sobre la población incluída en los casos clínicos tomados para el desarrollo de su modelo.

¿Existió un data set completamente independiente de validación?

Si. Miles de casos clínicos bien documentados independientes. Aunque no se aporta detalle de los miles de casos utilizados para su algoritmo.

¿Los resultados del test siendo evaluado influenciaron la decisión de realizar un estándar de referencia?

En las publicaciones experimentales con comparador enfocadas en la validación de precisión diagnóstica de symptom checkers y de precisión de triaje (entre ellos Symptomate), el resultado del modelo no influenció la realización del comparador; sin embargo, el comparador de referencia no está muy claro. Proponemos en este sentido, tomar como referencia a la recomendación de triaje realizada por un triaje telefónico que es la función que el modelo pretende suplir. En varios estudios se utilizó el triaje realizado por personal de salud, en otros el diagnóstico realizado por un médico y en otros no se utilizó comparador.

¿Fueron descritos con suficiente detalle los métodos para realizar el modelo como para permitir la replicación?

El modelo exigió un gran trabajo. Existen importantes esfuerzos por parte de Infermédica® tanto desde el punto de vista clínico, como al más técnico para describir su metodología; sin embargo el nivel de granularidad no llega a ser máxima para poder replicar su modelo por otros investigadores.

Análisis de la validez: aplicando los resultados

¿Cuáles fueron los resultados?

En el estudio de Hill et al. basado en viñetas clínicas (síntomas de pacientes obtenidos de casos clínicos de la literatura) los resultados para Symptomate en precisión diagnóstica fueron para los 3 primeros diagnósticos 24/231 (77%) y para los 10 primeros diagnósticos 25/31 (81%). La precisión diagnóstica para todos los symptom checkers

evaluado fue para los primeros 3 diagnósticos 52% (47-59%) y para los 10 primeros diagnósticos de un 58% (53-65%).

En relación a la precisión de triaje symptomate obtuvo 16/31 (52%) de precisión para todos los casos, 9/11 (82%) para casos de emergencia, 6/11 (55%) para casos de urgencia, 0/6 para casos de no urgencia y 1/3 (33%) para casos de autocuidado. La precisión diagnóstica para todos los symptom checkers evaluados fue del 49% (44-54%) para todos los casos, 63% (52-71%) para los casos de emergencia, 56% (52-75%) para los casos urgentes, 30% (11-39%) para los casos no urgente y 40% (26-49%) para los casos de autocuidado.

Los resultados para todos los symptom checkers fueron aportados como la media aritmética con IC 95% para precisión diagnóstica y precisión de triaje.

Este estudio observó buenos resultados para el symptom checker estudiado frente al resto, sobre todo en área de precisión diagnóstica. Sin embargo, el estudio tiene como gran debilidad el ser observacional, no basado en pacientes reales; siendo casi nulos los estudios realizados en pacientes reales del algoritmo de Infirmérica®. Tal hallazgo no es una rareza en el área de estudio de las herramientas de symptom checkers, donde existen muy pocos estudios experimentales en pacientes reales.

¿Fueron presentados los radios de probabilidad, sensibilidad y especificidad para los resultados del modelo o los datos para su cálculo aportada?

Existe escasez de estudios experimentales sobre pacientes reales con el symptom checker a estudio, donde se aporten estas métricas. Sin embargo, en algunos estudios donde se utilizan algoritmos diferentes disponibles en el mercado si se ofrecen.

¿Los resultados me van a ayudar a tratar a mis pacientes?

Si. Consideramos que un sistema de triaje telemático, con relativa precisión de triaje, sería de gran ayuda para la gestión de pacientes en la sanidad.

¿Son los resultados aplicables a mis pacientes?

Desconocemos las características de las poblaciones presentes en los casos clínicos tomados para el desarrollo de este modelo. Tampoco encontramos estudios de validez diagnóstica desarrollados en España. Evaluando exclusivamente los atributos basados en la evidencia del modelo, consideramos que no puede ser extrapolable a la población española hasta el desarrollo de estudios experimentales sobre personas reales que puedan esclarecer la precisión diagnóstica y de triaje, así como la sensibilidad, especificidad, valor predictivo positivo y negativo.

¿Los resultados cambiarán mi manejo?

Basándonos exclusivamente en sus atributos científicos, debido a que el algoritmo no ha sido validado en una población similar a la nuestra, no podemos concluir que cambiará el manejo. Sin embargo, creemos en los beneficios a nivel organizativo de proponer un modelo de atención urgente u ordinaria para el acceso al sistema de salud en una comunidad autónoma e intuimos que se generarían a la vez ahorros al incidir en la disminución de la presión de urgencias de consultas nivel IV y V de triaje.

¿Los pacientes estarán mejor por el resultado del test?

En teoría una herramienta de triaje telemático con una alta tasa de precisión de triaje sobre todo en eventos de emergencia y urgencia, podría beneficiar al sistema de salud y a los pacientes. Sin embargo, basándonos exclusivamente en la robustez científica, el modelo estudiado actualmente podría no funcionar para nuestra población. Si se pudiese validar en nuestra población, la implementación de esta herramienta podría beneficiar al sistema de salud y a los pacientes.

5.4. Propuesta de un Modelo Teórico para para incorporar un symptom checker en el flujo de atención urgente u ordinaria para el acceso al sistema de salud de una Comunidad Autónoma

Aprovechando la madurez las telecomunicaciones como eje integrador de las diferentes soluciones para conectar al ciudadano con el sistema sanitario, proponemos establecer una acceso alternativo de los usuarios, a través de un triaje pre-hospitalario gestionado por un symptom checker asistido por IA, siendo éste complementario a otras soluciones digitales ya existentes, como por ejemplo: los portales de salud (en relación directa con la gestión de citas electrónicas); los expedientes electrónicos y la telemedicina en todas sus modalidades. De esta manera, en su conjunto potenciarían el rol de la gestión remota de pacientes (GRP) en la contención de consultas en las Salas de Urgencia Hospitalarias o en Atención Primaria, distribuyendo la demanda de atención presencial y priorizando aquellos casos que requieran una asistencia de mayor complejidad.

Para elaborar la propuesta de modelo teórico y comprender mejor el contexto de posible implementación se utilizaron dos instrumentos: una matriz de análisis DAFO y una encuesta a responsables TIC de las CCAA sobre su conocimiento acerca la implementación de soluciones de symptom checkers.

5.4.1. Matriz de análisis DAFO

Se realizó un análisis DAFO por parte de los tres evaluadores, para poder tomar decisiones, reformular las estrategias y objetivos sobre el uso de los Symptom Checkers asistidos por Inteligencia Artificial como alternativa complementaria para el triaje en el modelo sanitario. Para ello se partió del modelo sanitario actual y su contexto en la pandemia de COVID-19.

Organización (Interno)	Contexto (Externo)
Debilidades (D)	Amenazas (A)
Fortalezas (F)	Oportunidades (O)

Análisis Interno:

- **Fortalezas** están constituidas por los atributos, cualidades y recursos propios de la herramienta, que son positivos y cuya acción favorable. Son fuerzas impulsoras, capacidades propias destacables que contribuyen positivamente a su integración en el modelo sanitario
 - ✓ Disponibilidad actual por parte de los sistemas sanitarios para desarrollar modelos asistenciales de GRP.
 - ✓ Las plataformas de telemedicina o de gestión remota de pacientes se postulan como una solución más sencilla, eficaz y económica para muchas dudas de pacientes, además de saltarse ese riesgo de potencial contagio.
 - ✓ La mayoría de las CCAA cuentan ya con un acceso web a través del PORTAL DEL PACIENTE donde se ofrecen apps y soluciones para atención del usuario.
 - ✓ Herramienta de software de asistencia médica virtual incorporada en aseguradoras privadas con éxito.
 - ✓ Los sistemas asistidos por IA de verificación de síntomas (symptom checkers) bien valorado por los usuarios
 - ✓ Facilidad de atención a los ciudadanos sin necesidad de presencialidad evitando el contagio ante la actual situación de pandemia
 - ✓ Su utilización disminuye la presión de atención de urgencias, en especial para condiciones ambulatorias de baja complejidad; pero también para activar recursos del sistema del tipo SUMA-112 en casos de real emergencia
 - ✓ La tecnología requerida para la mayoría de servicios que se quieren ofertar está madura.
 - ✓ Medios técnicos ya disponibles o mínima inversión: historia clínica electrónica, teléfono/smartphone/tablet para adaptarlos a procesos de GRP.
 - ✓ Facilidad de acceso a servicios de información médicos (de varios tipos y calidad) a través de Internet ha dado lugar a que los pacientes trataran de diagnosticarse a sí mismos.
 - ✓ La mayoría o todos los agentes implicados en el cambio de modelo asistencial considera que la implantación de tecnología debe ser parte de la solución
 - ✓ Los pacientes en la actualidad han incrementado el uso de herramientas digitales de atención médica
- **Debilidades** en cambio son aquellas situaciones, recursos u otros factores que dan cuenta de las dificultades de la Herramienta; problemas que inciden o limitan las perspectivas de su implementación e impedirían su adecuada integración en el proceso asistencia de triaje.
 - ✓ Ausencia de un marco jurídico específico.
 - ✓ Descentralización del sistema sanitario.
 - ✓ Falta de conocimiento y confianza en las soluciones de la salud electrónica entre los pacientes, los ciudadanos y los profesionales de la salud.

- ✓ Poca evidencia para recomendar symptom checkers
- ✓ Heterogeneidad de los estudios analizados (29 estudios)
- ✓ Estudios de coste / efectividad con resultados positivos, sin embargo, son insuficientes para elaborar una conclusión
- ✓ En relación a la seguridad se encontró que no existe evidencia en el riesgo entre un profesional de la salud y un symptom checker, siendo la evidencia muy débil.
- ✓ En cuanto a efectividad clínica y cumplimiento de las recomendaciones la evidencia es muy limitada como para realizar alguna conclusión.
- ✓ Mayor riesgo en el diagnóstico al no disponer de exploración física y datos aportados por el paciente.
- ✓ En triaje y la capacidad de la herramienta para dirigir al paciente al servicio adecuado, el desempeño varía según el sistema, siendo la evidencia inconsistente

Análisis Externo

Para detectar oportunidades y amenazas del escenario en que se integrarían los symptom checkers así como el posible comportamiento con los diferentes beneficiarios, pagadores, profesionales sanitarios, pacientes.

- **Oportunidades** son factores favorables que encontramos en el contexto y que se deben direccionar para darles utilidad
 - ✓ No existe un modelo estándar de referencia en triaje asíncrono
 - ✓ Alta demanda de pacientes que acceden al sistema de salud a través de los servicios de urgencia hospitalaria, donde se realiza un Triage.
 - ✓ Las políticas actuales fomentan el desarrollo de estrategia asíncronas de atención al paciente en todas las CCAA's.
 - ✓ La tecnología puede ayudar a reducir costes
 - ✓ Existe actualmente mucha inversión en I+D de Fondos europeos que puede consolidarse en productos a corto plazo
 - ✓ La Covid-19 que ha colapsado los sistemas sanitarios estimulando la adopción por parte del sistema de modalidades de atención remota (GRP)
 - ✓ Optimización de las vías clínicas con modelos asistenciales mixtos.
 - ✓ Reducción de los costes de traslado (paciente y cuidadores).
 - ✓ Sostenibilidad ambiental por menos traslados.
 - ✓ Evitar ingresos/visitas a urgencias.
 - ✓ Disminución de riesgos (en situaciones de pandemia).
- **Amenazas** son factores adversos al entorno que impactan, afectan y/o hacen peligrar la implementación de la herramienta
 - ✓ Situación económica adversa motivada por la pandemia
 - ✓ Reducción del poder adquisitivo de los ciudadanos
 - ✓ Población envejecida con limitación para la comprensión y la comunicación telefónica de calidad.
 - ✓ Resistencia al cambio en el modelo de cuidado (preferencia de algunos usuarios e incluso profesionales por los servicios presenciales)

- ✓ Los usuarios no perciben un valor añadido suficiente en algunos de estos servicios
- ✓ Salvo el servicio de tele asistencia de 1ª (y quizá 2ª generación), se trata de un mercado incipiente
- ✓ Rechazo al cambio del modelo asistencial.
- ✓ Afectación a la relación médico-paciente
- ✓ Miedo a los problemas legales por problemas de confidencialidad o errores diagnósticos.
- ✓ La COVID-19 que ha colapsado los sistemas sanitarios
- ✓ En últimos años, se han desarrollado muchas nuevas herramientas de diagnóstico clínico utilizando métodos de aprendizaje automático complicados no siendo bien aceptados por el profesional sanitario.

Conclusiones: Diagnóstico del posicionamiento de la herramienta Symptom checkers en el contexto sanitario del SNS

La situación actual de crisis sanitaria ha motivado un cambio en el modo de buscar atención por parte de los ciudadanos evidenciando la potencialidad de las atenciones remotas y modalidades de atención asíncronas. Los symptom checkers se presentan como instrumentos de triaje válidos, útiles y reproducibles ya que permite realizar la clasificación de los pacientes en base a su grado de urgencia temprana. Esto facilita una mejora en la calidad asistencial, una mayor eficacia y un importante ahorro en costes sanitarios.

De este primer análisis DAFO podemos extraer que en el sector de la salud los Symptom Checkers están en evolución, con ejemplos como iniciativas en el National Health Service (NHS) y la Clínica Mayo, pero existen todavía importantes déficits que hace difícil su implantación en el Sistema Nacional de Salud:

- **De calidad:** Nos hemos encontrado con una escasa evidencia para recomendar symptom checkers, con estudios muy heterogéneos e insuficientes para elaborar una recomendación de adopción en el SNS. Siendo necesario ampliar su dossier científico incorporando ensayos clínicos y estudios de evaluación económico como por ejemplo un análisis de Impacto presupuestario.
- **De Integración:** Los symptom checkers no están concebidos dentro del sistema sanitario de España como parte del flujo de atención de un paciente o como instrumento de triaje, por ello se debería incluir en la cartera de servicios de los hospitales de referencia en las comunidades autónomas, procediendo evaluaciones regionales y generando modificaciones y adaptaciones en base a protocolos y guías de práctica clínica de forma que pacientes y profesionales sanitarios perciban un valor añadido suficiente con estas herramientas de validación de síntomas.
- **De colaboración:** Los symptom checkers han sido desarrolladas por empresas tecnológicas como Infermedica sin contar con el apoyo en todas sus fases de profesionales de la salud o expertos nacionales; lo que da lugar a que no logren satisfacer las necesidades reales y exigencias de la sociedad española.
- Es necesaria una mayor inversión en este campo, pero de forma más organizada y bien dirigida, con procesos que involucren al paciente y que respondan a un trabajo conjunto de los diferentes agentes implicados: administración pública, empresas tecnológicas, proveedores de salud e industria.

- Los symptom checkers se presentan como un factor clave en el desafío de avanzar hacia una sanidad más sostenible, mejorando la efectividad y eficiencia, reduciendo costes y atendiendo a las principales necesidades de nuestra sociedad, caracterizada por el envejecimiento y la prevalencia de las enfermedades crónicas.

5.4.2. Encuesta sobre conocimiento sobre soluciones de Triage o Symptom checkers con Inteligencia artificial en SNS

Se llevó a cabo una encuesta a los responsables de Sistemas de Información y Tecnologías Sanitarias de las Consejerías de Sanidad de las 17 Comunidades Autónomas, que se muestran en la tabla 4.

Tabla 4. Listado de responsables de Sistemas de Información y Tecnología de las Consejerías de Sanidad

CCAA	CARGO	DEPARTAMENTO
Andalucía	Subdirector General	Subdirección de Tecnologías de la Información y Comunicaciones
Aragón	Director	Dirección de Obras, Instalaciones, Equipamientos e Innovación Tecnológica
Asturias	Directora	Dirección de Evaluación y Atención Sanitaria
Baleares	Subdirector General	Subdirección de Tecnología de la Información
Canarias	Jefa de Servicio	Servicio de evaluación de Calidad asistencial y Sistemas de Información
Cantabria	Director General	Dirección General de Transformación digital y atención a los usuarios
Castilla y León	Directora General	Dirección General de Infraestructuras y Tecnologías de la Información
Castilla La Mancha	Director	Dirección de Sistemas de Información
Cataluña	Directora	Agencia de Calidad y Evaluación Sanitarias de Cataluña
Extremadura	Subdirector General	Subdirección de Sistemas de Información
Galicia	Subdirector General	Subdirección General de Sistemas e Tecnologías de la Información
La Rioja	Director	Centro de Información y Atención al Usuario
Madrid	Directora General	Dirección General de Sistemas de Información y equipamientos Sanitarios
Murcia	Subdirector General	Subdirección General de Sistemas de la información
Navarra	Subdirector General	Subdirección General de Sistemas y Tecnologías para la Salud
País Vasco	Subdirector General	Subdirección de Informática y Sistemas de Información
Valencia	Subdirector General	Subdirección General de Sistemas de Información para la Salud

Se escogió como muestra para la encuesta a los responsables de información y tecnología (Chief Information Officer/CIO) de las Consejerías de Salud, ya que el conocer de las cuestiones consultadas recae específicamente en las funciones de identificar todos los procesos mejorables en la organización y que se enumeran a continuación:

- Beneficios para los pacientes y su seguridad
- Incremento de la eficiencia en el proceso asistencial
- Incremento de la eficiencia gestora
- Analizar siempre que sea posible el ROI directo para la organización y en su defecto para la sociedad.
- Promover y proponer todo tipo de innovación
- Colaborar en la aparición de BBDD orientadas al análisis y a la generación de conocimiento
- Colaborar en la inclusión del paciente en la gestión de su enfermedad
- Promover la transición del actual modelo Reactivo, por definición, a un modelo mucho más proactivo mediante el uso intensivo de las TIC

Las preguntas que se incluyeron y las respuestas que se obtuvieron fueron las siguientes:

Gráfico 9. Preguntas 1 y 2

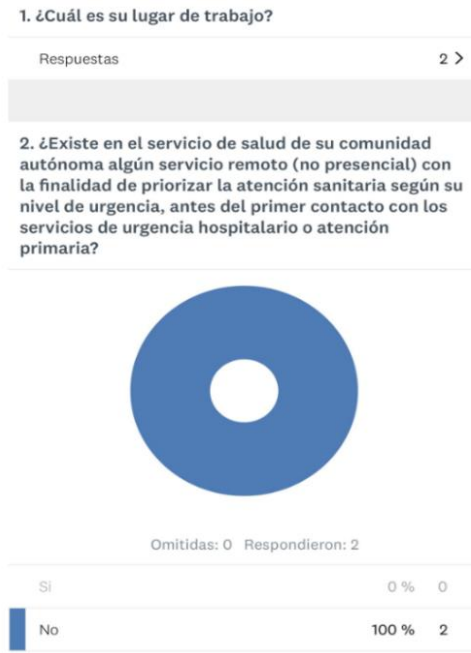
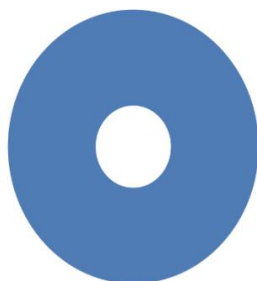


Gráfico 10. Preguntas 3 y 4



Gráfico 11. Preguntas 5 y 6

5. ¿Conoce algún asistente informatizado de triaje o symptom checker basado en inteligencia artificial que se esté utilizando en el sistema de salud de su comunidad autónoma?



Omitidas: 0 Respondieron: 2

Si	0 %	0
No	100 %	2
No se	0 %	0

6. En el caso de que la respuesta a la pregunta anterior sea "Si". ¿Cuál es el nombre de dicho

Respuestas 2 >

Tabla 5. Resultados de la Encuesta

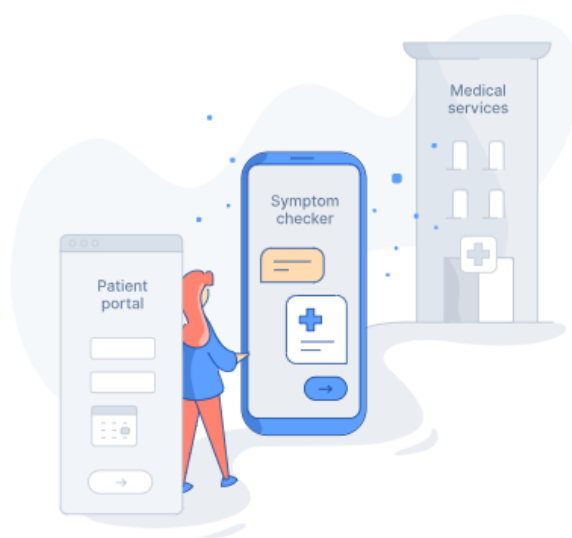
Encuesta	Pregunta 1	Pregunta 2			Pregunta 3		Pregunta 4			Pregunta 5			Pregunta 6		
Survey Monkey	CCAA	Si	No	NC	Omitida	No Omitida	Si	No	NC	Si	No	NC	Si	No	NC
Respuestas	2	0	2	0	2	0	0	2	0	0	2	0	0	2	0
Llamadas Telefónicas	CCAA	Si	No	NC	Omitida	No Omitida	Si	No	NC	Si	No	NC	Si	No	NC
Respuestas	8	0	8	0	8	0	0	8	0	0	8	0	0	8	0
Total	10	0	10	0	10	0	0	10	0	0	10	0	0	10	0
Porcentaje	100%	0%	100%	0%	100%	0%	0%	100%	0%	0%	100%	0%	0%	100%	0%

Diez comunidades autónomas respondieron, concluyendo que no existen servicios remotos cuya finalidad sea la de priorizar la atención sanitaria según su nivel de urgencia anterior al primer contacto con los servicios de urgencia hospitalaria o atención primaria. No conociendo en los servicios de salud contactados ningún asistente informatizado de triaje o symptom checker basado en inteligencia artificial que se este utilizando en el sistema de salud de las comunidades autónomas encuestadas.

5.4.3. Modelo de triaje virtual asistido por IA y acceso a servicios médicos

El siguiente modelo teórico busca incorporar en el flujo de atención urgente u ordinaria de acceso al sistema de salud de una Comunidad Autónoma, un sistema de triaje virtual (clasificación remota /telemática de síntomas), emergencia/urgencia, orientación médica telefónica, seguimiento de pacientes y agendamiento de citas médicas en atención primaria.

Gráfico 10. Modelo de triaje virtual asistido por IA y acceso a servicios médicos



El proceso inicia con la evaluación del requerimiento de atención del ciudadano a través del software de triaje médico que, basado en el resultado de la interacción, puede ofrecer las siguientes alternativas: a) Activar el protocolo emergencia (por ejemplo, SUMMA 112) b) Consulta remota, c) agendar una cita. Los canales de disponibilidad de servicios incluyen teléfono, web y aplicaciones móviles.

El Software de Triaje Médico Remoto, al contar un motor de consulta de alto rendimiento basado en tecnologías de Inteligencia Artificial, permite a partir de la entrada de los síntomas y los factores de riesgo de un paciente, procesar la información (signos y síntomas) de una gran base de conocimiento médico y calcular las probabilidades de condiciones presentes para a continuación generar las preguntas más relevantes para evaluar la gravedad de los síntomas.

Las respuestas del paciente darán como resultado final el triaje de acuerdo a 5 niveles de urgencia: 1) emergencia; 2) urgencia hospitalaria [activación de protocolos de urgencias-pre-hospitalarias y/o unidades móviles medicalizadas; 3) agendamiento de cita médica dentro de las 24 horas; 4) agendamiento de cita médica según disponibilidad 5) asesoramiento médico remoto/telemático (Video, Chat o Teléfono) para evaluar el estado de la salud y emitir recomendaciones de atención en el domicilio según guías clínicas y/o protocolos validados hasta que sea atendido por el personal sanitario (médico o de enfermería) de AP en caso de ser necesario.

Tabla 8. Niveles de triaje y actuaciones correspondientes

1	Inmediato	Emergencia	Activación de protocolos de urgencias-pre-hospitalarias y/o unidades móviles medicalizadas
2	Muy Urgente	Emergencia Hospitalaria	Activación de protocolos de urgencias-pre-hospitalarias y/o unidades móviles medicalizadas
3	Urgente	Derivación Unidades médicas	Agendamiento de cita médica dentro de las 24 horas
4	Estándar/Rutina	Agendamiento de cita médica según disponibilidad	Además: asesoramiento médico remoto por teléfono / chat / videoconferencia a demanda del usuario
5	No Urgente	Asesoramiento médico telefónico según protocolos y guías clínicas SNS	Reevaluación (llamada saliente dentro de las 48 horas) y agendamiento de cita médica según disponibilidad a demanda del usuario

El modelo teórico se propone de la siguiente manera:

El usuario podrá acceder a través del portal del paciente o a través de una plataforma multicanal al triaje remoto del servicio de salud de la comunidad autónoma hipotética. Esto significa que podrá acceder a través de teléfono (convencional o de telefonía IP), acceso vía portal web desde un ordenador o tableta o a través de un aplicativo móvil.

En el primer caso que implica el uso de un teléfono, el usuario accede a un contact center especializado a través de un número corto y gratuito, donde es atendido por un tele-operador capacitado en el uso de la herramienta de triaje con la que formulará las preguntas que el software vaya generando hasta obtener un resultado final que tres escenarios posibles:

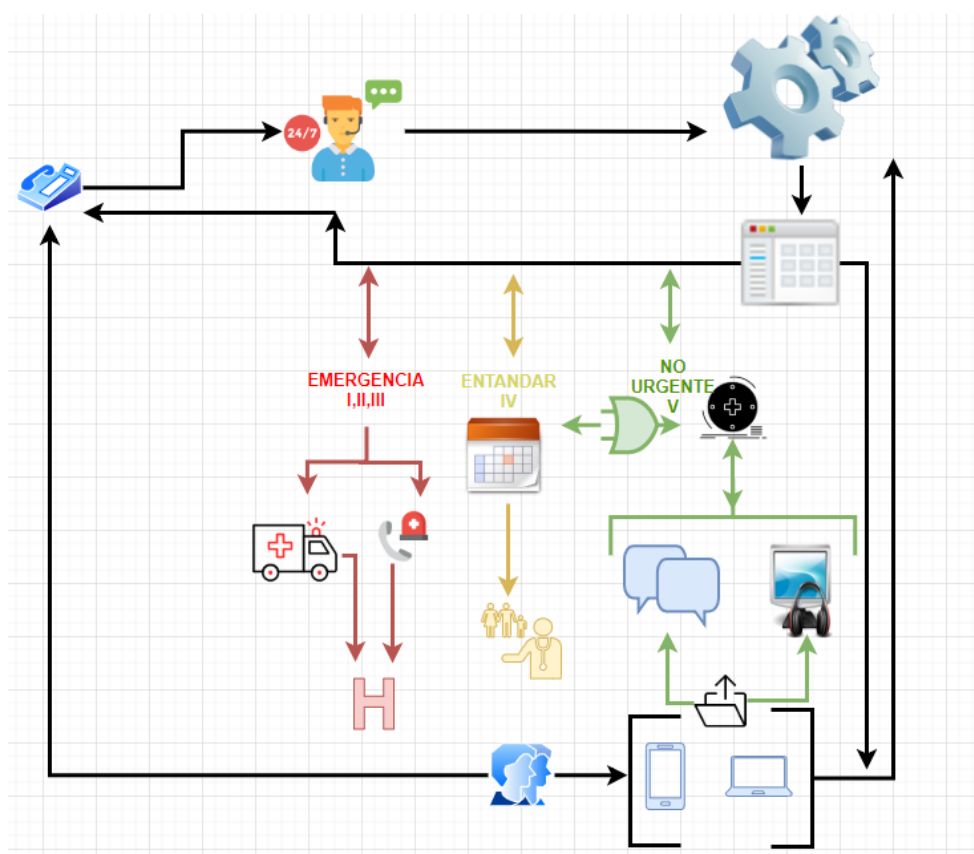
- Que se trate de un resultado ROJO, niveles de triaje I y II de la escala de triaje, en cuyo caso se activarían protocolos de urgencias-pre-hospitalarias y/o unidades móviles medicalizadas conectando con los servicios 112 o SAMUR. Para la categoría III se agendaría una consulta urgente antes de las 24 horas en atención primaria.
- Que se trate de un resultado AMARILLO. Nivel de triaje IV, en cuyo caso el tele-operador ofrecería agendar una cita médica en atención primaria cuando haya disponibilidad y/o la posibilidad de recibir asesoramiento médico remoto a través del teléfono.
- Que se trate de un resultado VERDE. Nivel de triaje V, en cuyo caso el tele-operador ofrecería la posibilidad de recibir asesoramiento médico remoto a través del teléfono y/o agendar una cita médica en atención primaria cuando haya disponibilidad. O en el caso de ser requerido a través de una gestión del contact center en OUTBOUND recibir la asesoría de un profesional sanitario vía Video, Chat o Teléfono.

En el resto de casos, el usuario accede directamente al software de triaje médico desde su portátil, tableta o aplicativo en su Smartphone, donde por autogestión se le formularán preguntas hasta obtener un resultado final idéntico al descrito en el párrafo anterior.

La diferencia radica en que los mensajes devueltos para las emergencias I y II deben igualmente ser auto gestionados por el usuario, ofreciéndole la posibilidad de contactar desde la misma plataforma con los servicios de emergencia 112. Para el resto de escenarios se le ofrecerá auto-agendamiento urgente en casos de nivel III y según disponibilidad para los niveles IV y V; a más de la posibilidad de recibir asesoría médica telemática/remota vía video llamada, chat o teléfono a través de una plataforma de telemedicina.

En el siguiente esquema se describe el flujograma:

Grafico 11. Flujograma del modelo de triaje virtual asistido por un Symptom Checker asistido por IA



Especial atención se presta a las asesorías telemáticas/remotas que requerirán de un servicio de telemedicina que cuente con un plan de atención asistido por guías clínicas y protocolos estandarizados y validados por los profesionales. Una posibilidad sería vincular en el proceso a la Sociedad de Medicina de Familia y Comunitaria (semFYC) y generar guías para atención por tele consulta y la inclusión de las mismas en <https://portal.guiasalud.es/> como parte de los insumos para establecer y fortalecer una estrategia de atención telemática/remota en el marco de un plan nacional de implementación de la telemedicina en la atención de salud en el SNS.

Otro punto a considerar es que deberán estar contempladas medidas técnicas y organizativas para garantizar la protección de la privacidad y protección de datos del paciente de acuerdo con la Ley de Protección de Datos Personales a través del cumplimiento de las siguientes recomendaciones principales en materia de privacidad en aplicación de consulta remota:

- La gestión de la herramienta de symptom checker respetará el principio de minimización de datos personales;
- La implementación del pre-triage y posterior tratamiento de la información, se almacenará de manera segura y se eliminarán los datos personales una vez que no sean necesarios.
- Todo procedimiento contará con el respectivo consentimiento del paciente a utilizar medios telemáticos para la prestación del servicio.

6. Discusión

En este trabajo se ha realizado una descripción y evaluación de un symptom checker, Symptomate de la empresa Infermédica®. La herramienta obtuvo una buena valoración según las metodologías de iSYScore y de Yun Liu. et al. La alta precisión diagnóstica, con la que cuenta Symptomate, es vital y el elemento más importante para el éxito de un servicio de triaje digital. La consecuencia de no tenerla podría afectar la satisfacción del paciente y poca disposición para volver a utilizar el servicio, aumento de la utilización de los servicios de urgencias y posible riesgo en la seguridad del paciente, aunque el abordaje cauteloso de la mayoría de sistemas puede ayudar a mitigar este aspecto.

En la gestión de los servicios de urgencia hospitalaria, los symptom checkers tendrían la potencialidad de aliviar los servicios de urgencia hospitalaria saturados, ofrecer una valoración estructurada de los síntomas de un paciente facilitando el diagnóstico al médico y gestionar el agendamiento de citas según su prioridad.

La implementación de un sistema de triaje pre-hospitalario de gestión remota asistido por IA relacionado con el expediente electrónico y con la gestión de citas electrónicas, buscaría la contención de consultas niveles IV y V en las salas de urgencias hospitalarias y la redistribución del flujo hacia las consultas telemáticas y hacia la atención primaria. Se cumpliría por tanto con un doble propósito; primero: prevenir el autodiagnóstico y segundo: activar los recursos del sistema en caso de necesidad urgente (unidades móviles/ambulancias, salas de urgencias, consultas ambulatorias, hospitales) reduciendo la presión de la demanda a los servicios de salud de mayor complejidad solo a los casos III, II y I que siempre requieren valoración y resolución presencial en una SUH.

Por otro parte, la información generada de la interacción con una herramienta de triaje pre-hospitalario, permitiría establecer reportes en tiempo real sobre la demanda de servicios de salud por parte de la población para un análisis eficiente de los recursos y una óptima toma de decisiones sobre éstos por parte de las autoridades sanitarias. Además de poner a disposición un medio de contacto con la población para el manejo de comunicaciones emergentes directas, focalizadas o masivas a través de mensajería outbound.

Creemos que el triaje pre-hospitalario remoto asistido por IA se podría establecer como el primer filtro efectivo que permitiría destinar los recursos disponibles a quienes realmente los necesitan. Además de la optimización del uso de los recursos y consecuente reducción de costes en el sistema sanitario.

Una ventaja añadida es que una vez abierto el canal de comunicación con el ciudadano gracias a la innovación de utilizar éstas tecnologías existentes, que van demostrando estar bien establecidas y generalmente aceptadas, se abre la posibilidad de la implementación progresiva de un servicio integral de gestión remota de pacientes a través de una plataforma omni-canal de tele-salud.

Se ofrecería por tanto una alternativa para la atención de varios de los principales retos de la salud actual: enfermedades crónicas, demanda creciente de usuarios y déficit de especialistas en determinadas áreas geográficas, a medida que evolucionen las futuras necesidades de salud física y mental de la población en general, tanto durante la pandemia como después de la misma.

No obstante, una de las principales limitaciones que tendría la implantación de un flujo de atención urgente u ordinaria de acceso al sistema de salud de una Comunidad Autónoma a través de un sistema de triaje virtual es la poca evidencia disponible para recomendar el uso de symptom checkers.

En Reino Unido, tras la puesta en marcha de un servicio digital de triaje accesible a la población general por un App o en línea, el Instituto Nacional para la Investigación en Salud (NIHR) inglés realizó una revisión sistemática donde se preguntaban ¿Cuál es el efecto de los servicios digitales diseñados para valorar los síntomas y señalar una recomendación que dirija a los pacientes al servicio adecuado en relación a los resultados clínicos y de los servicios de salud?²⁹

Las conclusiones de este análisis, donde se encontró mucha heterogeneidad entre los estudios incluidos, se enumeran a continuación:

- En general la evidencia fue débil, sin embargo, varía según cada resultado estudiado, siendo ella basada mayoritariamente en estudios observacionales.
- En relación a la seguridad se encontró que no existe evidencia en el riesgo entre un profesional de la salud y un symptom checker, siendo la evidencia muy débil.
- En cuanto a efectividad clínica y cumplimiento de las recomendaciones la evidencia es muy limitada como para realizar alguna conclusión.
- En cuanto a costos y coste / efectividad la mayoría de estudios obtuvo resultados positivos, sin embargo, son insuficientes para elaborar una conclusión.
- Respecto a los efectos en el uso del servicio / diversión los resultados fueron inconsistentes y no se pueden realizar conclusiones.
- En relación a la precisión diagnóstica los symptom checkers parecen ser inferiores a los profesionales de la salud, siendo la evidencia fuerte. Sin embargo, los estudios fueron realizados para condiciones específicas o simulados con casos clínicos de la literatura.

²⁹ Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open*. agosto de 2019;9(8):e027743.

- En triaje y en la capacidad de la herramienta para dirigir al paciente al servicio adecuado, el desempeño varía según el sistema, siendo la evidencia inconsistente.
- La mayoría de los estudios reportan una alta tasa de satisfacción de los pacientes, sin embargo, la evidencia es muy débil debido a que pocos estudios son controlados.

Es necesario por tanto un mayor esfuerzo para la generación de estudios de alto nivel de evidencia científica, incluyendo estudios experimentales en poblaciones representativas donde se vaya a utilizar el algoritmo de triaje, para obtener una base científica con mayor solidez que nos permita realizar una recomendación más firme sobre el uso de esta herramienta científica y de gestión.

Es prioritario continuar recogiendo datos sobre la seguridad de los casos de uso y realizar estudios que hagan una comparación directa del desempeño de múltiples symptom checkers, así como la realización de evaluaciones económicas rigurosas basadas en datos del mundo real. Asimismo, es recomendable realizar investigaciones con pacientes pediátricos, evaluar el posible rol de la teoría del cambio de comportamiento en el desarrollo e implementación de los symptom checkers y realizar investigaciones cualitativas para determinar las percepciones y barreras para su uso por personas que tienen menos familiaridad con las tecnologías digitales.

A pesar de que los symptoms checkers no cuentan con evidencia científica robusta que los apoyen, tienen un hueco en el mundo real. Muchas aseguradoras privadas e incluso algún sistema público como el inglés, ya han puesto en marcha servicios digitales de triaje. Como herramienta de gestión, aunque con una base científicamente no del todo desarrollada, tal y como evidencian algunos casos de uso, traerían beneficios táctiles al sistema de salud y por lo tanto a la población.

En este contexto, nuestra propuesta denominada: “Modelo Teórico para flujo de atención urgente u ordinaria para el acceso al sistema de salud en una Comunidad Autónoma” podría concretarse en una Investigación-Acción (Action Research en inglés) de carácter exploratorio que, confrontada con el servicio de triaje telefónico (comparador estándar), que sería el servicio que se quiere emular, pudiera aportar datos de efectividad real que contribuyeran a la generación de evidencia. Además, podría ser una línea de acción a ser considerada por un servicio de salud en una CA dentro de su estrategia de transformación digital y facilitaría la inclusión de modelos de triaje remoto en los flujos de acceso a los servicios de salud.

Parafraseando a Maturana: realizar *un primer acercamiento científico a un problema de investigación, cuando éste aún no ha sido abordado o no ha sido suficientemente estudiado y las condiciones existentes no son aún determinantes*.³⁰

Finalmente destacar que nuestra propuesta trata de motivar un cambio estructural por parte de los sistemas de salud, de los profesionales y de los usuarios que contribuya a la transformación del sector de la salud hacia un modelo más sostenible y conectado, a través de fomentar una medicina proactiva, con un ciudadano más activo e informado y centrado en la prevención y el uso eficiente de los recursos.

³⁰ MATURANA R., Humberto & y VARELA G., Francisco, El árbol del conocimiento: Las bases biológicas del entendimiento humano. (1era ed.)2003, Ed. Lumen, Buenos Aire, Argentina, p.35

7. Bibliografía citada

1. Ministerio de Sanidad y Política Social. Unidad de urgencias hospitalaria Estándares y recomendaciones. 2010.
2. Gómez Jiménez J. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias, *Emergencias*. 2003; 15:165-74.
3. Gómez Jiménez J, Burgues I, Pàmies S. Clinical management of a hospital emergency department: quality indicators, benchmarking and case-mix analysis., *Gestión Hospitalaria*,. 2004;15,((1),):3:12
4. FitzGerald G, Jelinek GA, Scott D, Gerdts MF. Emergency department triage revisited. *Emerg Med J*. 2010 Feb;27(2):86-92.
5. Abad-Grau MM, Ierache J, Cervino C, Sebastiani P. Evolution and challenges in the design of computational systems for triage assistance. *Journal of Biomedical Informatics*. junio de 2008;41(3):432-41.
6. Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. *Diagn Progn Res*. diciembre de 2020;4(1):16.
7. Ochoa Gómez J, Villar Arias A, Ignacio Ruiz Azpiazu J, Bragado Blas L, Gimeno Villa C, Ramalle-Gómara E. Visitas inapropiadas al servicio de urgencias de un hospital general. *Medicina Clínica*. enero de 2000;115(10):377-8.
8. Sempere-Selva T, Peiró S, Sendra-Pina P, Martínez-Espín C, López-Aguilera I. Inappropriate use of an accident and emergency department: Magnitude, associated factors, and reasons—An approach with explicit criteria. *Annals of Emergency Medicine*. junio de 2001;37(6):568-79.
9. Reinoso Hermida S, Díaz Grávalos GJ, Robles Castiñeiras A, Pereiro Sánchez E, Fernández Cardama E, López González A. Forma de acceso a los Servicios de Urgencia hospitalarios y pertinencia de la consulta. *Rev Clin Med Fam*. octubre de 2011;4(3):205-10.
10. Una propuesta de modelo fisiológico de servicio de urgencias hospitalario. Principios de funcionamiento, tipificación de la saturación y pautas para el rediseño. *An Sist Sanit Navar* [Internet]. abril de 2017 [citado 23 de agosto de 2021];40(1). Disponible en: <https://recyt.fecyt.es/index.php/ASSN/article/view/52722/35615>
11. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self-diagnosis and triage: audit study. *BMJ*. 8 de julio de 2015; h3480.
12. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung S-C, Leung L-P. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. *Hong Kong Journal of Emergency Medicine*. julio de 2020;27(4):217-22.

13. Ashley N D Meyer, et,al. ; Patient Perspectives on the Usefulness of an Artificial Intelligence–Assisted Symptom Checker: Cross-Sectional Survey Study; J Med Internet Res 2020 | vol. 22 | iss. 1 | e14679 | p. 1
14. Adam Zagorecki, Piotr Orzechowska, Katarzyna Hołownia; A System for Automated General Medical Diagnosis using Bayesian Networks; MEDINFO 2013 C.U. Lehmann et al. (Eds.) ©2013 IMIA and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License. doi:10.3233/978-1-61499-289-9-461
15. Miró O. El usuario y su uso de las urgencias. Anales Sis San Navarra [Internet]. diciembre de 2009 [citado 23 de agosto de 2021];32(3). Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272009000500001&lng=en&nrm=iso&tlng=en
16. Grau, B. Kostov, J.A. Gallego, F. Grajales III, L. Fernández-Luque, y Sisó-Almirall A. Método de valoración de aplicaciones móviles de salud en español: el índice iSYScore. : Semergen. 2016;42(8):575---583
17. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. JAMA. 2019;322(18):1806–1816. doi:10.1001/jama.2019.16489
18. SMILE es el motor de inferencia desarrollado en el Laboratorio de Sistemas de Decisiones de la Universidad de Pittsburgh, y es utilizado para realizar inferencias de diagnóstico (<http://genie.sis.pitt.edu>).
19. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Medical Journal of Australia. Junio de 2020;212(11):514-9.
20. Gómez Jiménez J. Clasificación de pacientes en los servicios de urgencias y emergencias: Hacia un modelo de triaje estructurado de urgencias y emergencias. Emergencias. 2003; 15: 165-174
21. Hoehn and Bongard-Blanchy; Heuristic Evaluation of COVID-19 Chatbots; Luxembourg National Research Fund (FNR), SLANT, 13320890;
22. Nan-Guang Tan et al., Virtual care to increase military medical centre capacity in the primary health care setting: A prospective self-controlled pilot study of symptoms collection and telemedicine; J Telemed Telecare 2020 Oct 4; 1357633X20959579. doi: 10.1177/7633X13520959579. <https://pubmed.ncbi.nlm.nih.gov/33016187/>
23. Munsch N, Martin A, Guarín S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, Knapp B; Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study, J Med Internet Res; 2020;22(10):e21299; URL: <https://www.jmir.org/2020/10/e21299>; DOI: 10.2196/21299
24. D'Ambrosio B. Symbolic probabilistic inference in large BN2O networks. In: Proc. 10th Conf. on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kauf- mann; 1994. pp. 128-135.

25. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open. agosto de 2019;9(8):e027743.
26. Maturana R., Humberto & y Varela G., Francisco, El árbol del conocimiento: Las bases biológicas del entendimiento humano. (1era ed.)2003, Ed. Lumen, Buenos Aires, Argentina, p.35

8. Webgrafía

1. <https://orbi.lu.uni/bitstream/10993/44913/1/2020-12-02Conversational UX analysis of COVID 19 chatbots.pdf>
2. <https://developer.infermedica.com/docs/introduction>
3. <https://infermedica.com/case-studies/pzu-zdrowie>
4. <https://infermedica.com/case-studies/medis>
5. <https://infermedica.com/case-studies/covid19>
6. <https://blog.infermedica.com/how-to-improve-emergency-departments-with-ai/>
7. <https://infermedica.com/external-validation>
8. <https://infermedica.com/what-can-we-learn-about-people-using-symptom-checkers>
9. https://blog.infermedica.com/building-a-trusted-quality-digital-medical-knowledge-base/?utm_campaign=what-can-we-learn-about-people-using-symptom-checkers&utm_medium=email&_hsmi=110452169&_hsenc=p2ANqtz-9RzDzCw5HlZg4euDTI7vKpFJ1tiy7dXPd7FLz1ITdowdadjrV9h_IDqIN1b4ibVqCkoBZOiUx6ITLRS�wf85Gc6Doamw&utm_source=report
10. <https://infermedica.com/diagnostic-engine>
11. <https://infermedica.com/test-cases>