

Evaluación del rendimiento diagnóstico y la seguridad de la inteligencia artificial en la monitorización cardíaca con dispositivos insertables.

Resumo de: López Loureiro I, Puñal-Rioboo J, Mejuto Martí, T. Evaluación del rendimiento diagnóstico y la seguridad de la inteligencia artificial en la monitorización cardíaca con dispositivos insertables. Madrid: Ministerio de Sanidad. Santiago de Compostela: Agencia Gallega para la Gestión del Conocimiento en Salud, ACIS, Unidad de Asesoramiento Científico-Técnico; Avalia-t; 2024.

Disponible en: <https://extranet.sergas.es/catpb/Docs/cas/Publicaciones/Docs/avalia-t/PDF-3283-es.pdf>

INTRODUCCIÓN

As arritmias son latexos irregulares do corazón que presentan unha sintomatoloxía pouco específica que pode estar asociada a outras doenzas. Así mesmo, estes signos e síntomas poden non presentarse (arritmia asintomática) ou facelo de forma temporal (arritmia paroxística), o que dificulta o seu diagnóstico. Existen diferentes tipos de arritmias, entre as que destaca a fibrilación auricular (FA) por ser a máis frecuente na práctica clínica e responsable dun 20% dos ictus isquémicos. Esta arritmia presenta unha incidencia media de 9,9/1000 persoas/ano en Europa. Para facilitar a detección das arritmias poden empregarse dispositivos de monitorización que realizan un rexistro continuo do electrocardiograma. Existe unha ampla gama de dispositivos con este fin, entre os que destacan os dispositivos subcutáneos implantables por realizaren unha monitorización a longo prazo (2-3 a 5,5 anos). Estes dispositivos dispoñen dun algoritmo que realiza unha análise automática do ECG do doente e grava os rexistros onde reconece unha arritmia (alertas). Co obxectivo de incrementar a súa sensibilidade e detectar todos os eventos que sofre o doente, os algoritmos destes dispositivos comprometen a súa especificidade e xeran unha gran cantidade de alertas que son falsos positivos (FP). Estas alertas deben ser analizadas polo persoal clínico para confirmar o diagnóstico de arritmia, o que supón unha gran carga de traballo. Nos últimos anos incorporáronse algoritmos de intelixencia artificial (IA) para realizar unha análise destes rexistros e así poder cribar os falsos positivos xerados polos dispositivos e reducir a carga de traballo do persoal. Estes algoritmos teñen o reto de reducir os FP sen xerar falsos negativos (FN) que poidan supoñer un impacto na saúde do paciente.

OBXECTIVOS

Avaliar a efectividade en termos de rendemento diagnóstico, a seguridade clínica (FP e FN) e técnica, satisfacción dos clínicos e aspectos económicos, organizativos, éticos, legais e técnicos do uso de algoritmos baseados en IA para a detección de arritmias en doentes con dispositivos implantables.

MÉTODOS

Realizouse unha revisión sistemática da literatura científica con data de 31 de marzo de 2024 en bases de datos especializadas en revisións sistemáticas e informes de avaliación de tecnoloxías sanitarias (Cochrane, Epistemonikos, International Health Technology Assessment Database e Prospero), bases de datos xerais (Medline, Embase, Web of Knowledge) e bases de datos de proxectos de investigación en curso (clinicaltrials.org, Cochrane).



O proceso completouse cunha busca manual específica en metabuscadores e páxinas web de sociedades e organizacións científicas nacionais e internacionais, así como coa revisión manual da bibliografía citada nos estudos seleccionados. Os resultados da busca volcáronse nun xestor de referencias bibliográficas (EndNote 20), co fin de eliminar duplicados e facilitar a xestión documental.

Dous investigadores revisaron e seleccionaron os estudos de forma independente e a cegas empregando a plataforma Covidence. Incluíronse revisións sistemáticas, estudos comparativos, observacionais e de avaliación económica nos que se realizase unha avaliación do uso da intelixencia artificial na análise de electrocardiogramas recompilados por un dispositivo implantable. Excluíronse revisións narrativas, casos clínicos e comunicacións a congresos, así como estudos onde os algoritmos non empregaban IA. Inicialmente, o cribado de resultados realizouse por título e resumo, para posteriormente realizar a revisión a texto completo. En caso de discrepancias, estas resolvéronse mediante consenso. Para avaliar a validez dos estudos e o risco de sesgo, utilizouse a escala QUADAS-2 para as probas diagnósticas e a escala CASPe para os estudos cualitativos. Os resultados extraídos analizáronse de forma cualitativa, e a seguridade avaliouse dende dúas perspectivas: a seguridade clínica e a técnica. Así mesmo, o rendemento diagnóstico valorouse segundo o tipo de arritmia detectada e no seu conxunto.

RESULTADOS

A evidencia analizada procede de catro estudos observacionais de probas diagnósticas e un estudo cualitativo. Nos estudos de probas diagnósticas inclúese un total de 445 pacientes que xeran 6113 rexistros, dos cales se analizan 3761. Respecto ao estudo cualitativo, este inclúe 1136 comentarios de 905 profesionais da saúde. Tres estudos de probas diagnósticas presentaron un deseño prospectivo e un, un deseño retrospectivo. Entre os estudos prospectivos, dous presentan solapamento de doentes. En relación coa calidade da evidencia, os estudos de proba diagnóstica presentan un elevado risco de sesgo, mentres que o estudo cualitativo é de alta calidade.

Dous dos estudos incluídos indican un valor predictivo negativo (VPN) que oscila entre o 98,1% e o 98,6%, mentres que tres estudos describen un rango de FN do 1%-2,9%. Respecto ao valor predictivo positivo (VPP), tres estudos mostran que o rango de valores fluctúa entre o 74,5% e o 89,74%. Así mesmo, tres estudos describen que a IA produce unha redución de FP entre o 60,3% e o 98%. Para a detección da FA, a evidencia de catro estudos indica que a IA mostra unha sensibilidade superior ao 96%, mentres que tres estudos describen unha especificidade entre o 60,3% e o 95,4%. Noutras arritmias, os valores de sensibilidade proveñen de tres estudos e os de especificidade de dous. No caso da taquicardia ventricular (TV), os valores de sensibilidade e especificidade oscilan entre 77,78%-100% e 67,4%-99,73% respectivamente. Na bradicardia, a IA alcanza uns valores de sensibilidade entre o 83,3% e o 100%, mentres que a especificidade oscila entre o 73,3% e o 99,95%. Pola súa banda, na asistolia a sensibilidade sitúase no 93,3%-100%, mentres que a especificidade presenta valores de 85,7%-100%. En relación coa aceptabilidade da tecnoloxía, o estudo cualitativo indica que o persoal clínico presenta unha actitude neutra cara a temas como IA en monitorización cardíaca e IA para a interpretación de ECG. Respecto aos aspectos éticos, sociais, legais, técnicos e económicos, non se puido recuperar suficiente evidencia para responder a estes elementos.

DISCUSIÓN

O corpo da evidencia consiste maioritariamente en estudos de probas diagnósticas con elevado risco de sesgo. Os FN descritos nos estudos son, na súa maioría, FA e taquicardia auricular (TA). Dado que non existe consenso sobre a utilidade clínica do tratamento da FA subclínica na prevención do ictus, o impacto dos FN é difícil de valorar. Respecto á validez diagnóstica dos algoritmos, as diferenzas na sensibilidade e especificidade están influídas polos diferentes dispositivos subcutáneos empregados e o software destes. Por iso, unicamente se poden comparar algoritmos que empreguen os mesmos dispositivos.



Así mesmo, a pouca evidencia localizada e a súa baixa calidade non permiten realizar unha comparación adecuada entre diferentes IA. Dentro da análise da FA, dous dos estudos agrúpanas coa taquicardia auricular, o que tamén podería xerar variabilidade nos resultados. Con respecto á aceptabilidade no uso da IA, o estudo cualitativo incluído foi realizado na súa maior parte entre persoal clínico estadounidense. Polo tanto, a aceptabilidade podería ser diferente no noso contexto sanitario.

CONCLUSIÓNS

Os estudos describen unha taxa de FN entre o 1% e o 2,9%. Porén, non existe suficiente evidencia para valorar as implicacións clínicas destes valores. Dentro da validez diagnóstica, os valores oscilan entre o 77,78% e o 100% para a sensibilidade, mentres que na especificidade fluctúan entre o 60,3% e o 100%. As diferenzas entre os valores veñen determinadas polo tipo de arritmia avaliada, o dispositivo subcutáneo empregado, o seu software e a IA encargada da análise. Os algoritmos de IA mostran un VPP entre o 74,5% e o 89,74%. Así mesmo, producen unha redución de FP entre o 60,3% e o 98%. Non se localizou evidencia que permita valorar o impacto económico, ético, social, legal e técnico derivado da implantación da IA. Así mesmo, non se recuperou información sobre o aforro do tempo de traballo do persoal clínico derivado do uso da IA. Dado o escaso corpo de evidencia localizado e a baixa calidade do mesmo, non se poden realizar conclusións sobre o impacto clínico dos algoritmos de IA na análise de arritmias e redución de FP.