



Diseño de una herramienta para la extracción semiautomática de información en procesos de evaluación de tecnología en salud, enfocada en dispositivos médicos para Instituciones Prestadoras de Servicios de salud acreditadas en Colombia

Mabel Catalina Zapata Álvarez

Trabajo de investigación presentado para optar al título de Magíster en Ingeniería

Asesores

Jenny Kateryne Aristizábal, Magíster (MSc) en Ingeniería Biomédica

Juan Guillermo Barreneche, Magíster en Administración de Negocios

Universidad de Antioquia

Facultad de Ingeniería

Maestría en Ingeniería

Medellín, Antioquia, Colombia

2022

Cita	Zapata Álvarez [1]
Referencia	[1] M.C. Zapata Álvarez, “Diseño de una herramienta para la extracción semiautomática de información en procesos de evaluación de tecnología en salud, enfocada en dispositivos médicos para Instituciones Prestadoras de Servicios de salud acreditadas en Colombia”, Tesis de maestría, Maestría en Ingeniería, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2022.



Maestría en Ingeniería

Grupo de Investigación Bioinstrumentación e Ingeniería Clínica (GIBIC).



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes

Decano/Director: Jesús Francisco Vargas Bonilla

Jefe departamento: Sara Cristina Vieira Agudelo.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Agradecimientos

Quiero agradecer en primera instancia a mis asesores, Jenny Katerlyne Aristizábal y Juan Guillermo Barreneche, quienes me guiaron durante este proceso con su conocimiento y dedicación.

A mi madre Maribel Álvarez Rendón y abuela Luz Rendón Muñoz quienes con su amor y entrega me han apoyado durante todo mi proceso académico hasta este momento.

A mi equipo de trabajo Javier García Ramos, Lucía Uribe y Óscar Darío Saldarriaga que siempre confiaron y me alentaron, además de brindarme su apoyo incondicional y sus conocimientos. A mi amiga incondicional Paula Andrea Pérez quien estuvo siempre que necesité una voz de aliento.

A la IPS Universitaria Clínica León XIII, Hospital Pablo Tobón Uribe, Hospital Universitario San Vicente Fundación y Clínica Universitaria Bolivariana donde logré caracterizar sus procesos de Evaluación de Tecnología en Salud y realizar las pruebas de usabilidad de la herramienta propuesta en este trabajo de investigación.

Al Grupo de Investigación en Bioinstrumentación e Ingeniería Clínica – GIBIC.

Contenido

Resumen	10
Abstract.....	12
1. Introducción.....	14
1.1. Contexto	14
1.2. Objetivos.....	17
1.2.1. General.....	17
1.2.2. Específicos.....	17
2. Estado del arte	18
3. Marco teórico	21
3.1. Evaluación de Tecnología en Salud	21
3.1.1. Implementación de mini ETS	28
3.1.2. Contexto colombiano para la evaluación de tecnología en salud basada en hospitales	28
3.2. Procesamiento de Lenguaje Natural	30
3.3. Métodos de análisis de lenguaje	31
3.3.1. Métodos clásicos – conteo de palabras y análisis estadísticos	31
3.3.2. Métodos clásicos – análisis sintáctico	33
3.3.3. Métodos de word embeddings (incrustación de palabras).....	33
3.4. Aprendizaje automático - algoritmos de clasificación	34
3.4.1. Máquinas de Soporte Vectorial – SVM por sus siglas en inglés	34
3.4.2. Árboles de decisión - DT por sus siglas en inglés.....	35
3.5. Aprendizaje profundo.....	36
3.6. Métricas de desempeño.....	39
3.6.1. Matriz de confusión	39
3.6.2. Exactitud	40
3.6.3. Precisión.....	40
3.6.4. Recall (Sensitividad)	40
3.6.5. Especificidad	41
3.6.6. Área bajo la curva ROC	41
3.7. Análisis <i>RobotReviewer</i>	41

3.7.1.	Robot reviewer	41
3.7.1.1.	Método de aprendizaje, corpus, vocabulario y espacio de características 41	
3.7.1.2.	Método de clasificación	43
3.7.1.3.	Principales resultados	44
3.7.1.4.	Actualizaciones.....	45
3.7.2.	BERT	45
3.7.2.1.	Bert embeddings.	46
3.7.2.2.	Modelo de atención	46
3.7.2.3.	Tareas de preentrenamiento y entrenamiento.	49
3.7.2.4.	Principales resultados	50
3.7.3.	SciBERT	51
3.7.3.1.	Tareas de entrenamiento y ajuste	51
3.7.3.2.	Principales resultados	52
4.	Metodología y Resultados	53
4.1.	Caracterización del proceso de evaluación de tecnología en salud en hospitales colombianos.....	54
4.1.1.	Resultados caracterización	58
4.1.2.	Resultados individuales por hospital	62
4.2.	Base de datos.....	69
4.2.1.	Elección de dispositivos médicos de enfoque	70
4.2.2.	Levantamiento de criterios de inclusión.....	71
4.2.3.	Búsqueda, filtro y consolidación.....	72
4.3.	Clasificación de estudios a incluir en síntesis de evidencia y extracción de información	79
4.3.1.	Análisis general de la herramienta RobotReviewer	82
4.3.2.	Identificación de algoritmos de clasificación y extracción de información	82
4.3.3.	Entrenamiento y validación de algoritmos.....	85
4.4.	Herramienta de integración de búsqueda, extracción de información y síntesis de evidencia clínica	92
4.4.1.	Levantamiento de árbol de objetivos y priorización de tareas	94
4.4.2.	Integración de actividades.....	96

4.4.3. Validación.....	114
5. Conclusiones	121
5.4. Contribuciones.....	122
5.5. Limitaciones.....	123
5.6. Trabajo futuro.....	123
6. Referencias.....	124
7. Anexos	134
7.1. Instructivo de uso	134
7.2. Prueba de usabilidad	134
7.3. Ejecución pruebas de usabilidad	134
7.4. Congresos y publicaciones	134

Índice de figuras

Figura 1. Ciclo de vida de la tecnología en salud.	22
Figura 2. Intensidad de uso y evaluación de tecnología en salud Fuente: [53], [54]	23
Figura 3. Estrategias de ETS con enfoque hospitalario, adaptado según [30]	25
Figura 4. Representación de palabras a través de TF-IDF.....	32
Figura 5. Representación de SVM para un problema de clasificación de 2 clases.....	35
Figura 6. Representación árboles de decisión	36
Figura 7. Esquema redes neuronales.....	36
Figura 8. Nodo o neurona en redes LSTM	38
Figura 9. Matriz de confusión	40
Figura 10. Área bajo la curva ROC.....	41
Figura 11. Representación de aprendizaje multi-tarea en robot reviewer	42
Figura 12. Ejemplo de BERT embeddings	46
Figura 13. Ejemplo gráfico mecanismo de atención.....	48
Figura 14. Capa de transformador, adaptado de [39]	49
Figura 15. Metodología general de trabajo	54
Figura 16. Metodología implementada para y caracterización de ETS	55
Figura 17. Metodología para el levantamiento de base de datos de reportes de ETS ..	69
Figura 18. Información reportes de evaluación de tecnología.....	75
Figura 19. Pirámide de evidencia clínica [135]	76
Figura 20. Metodología para la adaptación de algoritmos RobotReviewer.....	80
Figura 21. Metodología clasificación de tipo de estudios	81
Figura 22. Etiquetado de resúmenes, reconocimiento de términos PICO.	84
Figura 23. Priorización de estudios.....	89
Figura 24. Distancia coseno entre artículos recuperados con estrategia de búsqueda.	90
Figura 25. Resultados de priorización de evidencia clínica	91
Figura 26. Metodología para el diseño de herramienta de síntesis semiautomática	92
Figura 27. Proceso de pruebas de usabilidad, extraído de [144]	94
Figura 28. Tareas que se integran en la herramienta de síntesis semiautomática	95
Figura 29. Indicador PYPL de popularidad de lenguajes de programación	97
Figura 30. Indicador TIOBE de popularidad de lenguajes de programación	97

Figura 31. Ayudante de búsqueda avanzada PubMed	98
Figura 32. Campo donde se buscarán las palabras de la estrategia sintetizada	99
Figura 33. Ítem para incluir, excluir o unir las palabras en la estrategia de búsqueda ..	99
Figura 34. Ventana para síntesis de estrategia de búsqueda	100
Figura 35. Campo para ingresar términos de búsqueda	100
Figura 36. Recuadros para ingresar términos PICO a estrategia de búsqueda	101
Figura 37. Visualización de la cantidad de resultados obtenidos en la búsqueda	104
Figura 38. Barra de progreso de análisis de resultados.....	105
Figura 39. Visualización de artículos con texto resaltado	106
Figura 40. Tooltip botones "Anterior" y "siguiente"	107
Figura 41. Etiqueta de posición de navegación y priorización de resultados.....	107
Figura 42. Visualización de parámetros de ranking de artículos	108
Figura 43. Clasificación de tipo de estudio según modelo entrenado	109
Figura 44. Casilla de verificación para incluir artículos en el reporte de síntesis.....	110
Figura 45. Botones que permiten generación automática de reporte de síntesis.....	110
Figura 46. Sección uno de reporte: Parámetros y estrategia de búsqueda	111
Figura 47. Sección dos: resumen de artículos presentados en reporte	112
Figura 48. Sección tres: resultados (artículos).....	113
Figura 49. Sección tres: resultados. Ejemplo de artículo con resumen no disponible.	113
Figura 50. Resultados prueba de usabilidad.....	116

Índice de tablas

Tabla 1. Síntesis de la herramienta AdHopHTA mini ETS	26
Tabla 2. Resultados para la predicción de sesgo de un documento completo	44
Tabla 3. Proporción de estudios para los cuales se identificaron sentencias de texto relevantes para cada dominio de sesgo.	45
Tabla 4. Tareas para la evaluación de sistemas de PLN contempladas GLUE [76].....	50
Tabla 5. Desempeño de BERT en tareas de GLUE.....	51
Tabla 6. Principales resultados SciBERT, adaptado de [40]	52
Tabla 7. Estructura de encuesta diseñada para guiar la caracterización de ETS.....	56
Tabla 8. Capacidad instalada instituciones incluidas en la caracterización.....	58
Tabla 9. Resultados caracterización por preguntas e institución	59
Tabla 10. Principales semejanzas y diferencias identificadas en los procesos de ETS	68
Tabla 11. Intervenciones y tecnología empleada en el diagnóstico y tratamiento de enfermedades cardiovasculares. Principales resultados analizados	71
Tabla 12. Criterios de inclusión reportes de evaluación.....	72
Tabla 13. Información derivada del reporte de evaluación de tecnología: "Use of Biventricular Pacing in Atrioventricular Heart Block"	77
5. Tabla 14. Desempeño modelo de reconocimiento de términos PICO	85
Tabla 15. Valores de hiperparámetros optimizados SVM	86
Tabla 16. Valores de hiperparámetros optimizados árboles de decisión	86
Tabla 17. Resultados clasificación ensayos clínicos aleatorizados	87
Tabla 18. Resultados clasificación revisiones sistemáticas	87
Tabla 19. Reporte de evaluación de tecnología replicado	89
Tabla 20. Parámetros usados para priorización de evidencia	92
Tabla 21. Uso de bases de datos en hospitales para hacer revisión de literatura	102
Tabla 22. Descripción de los usuarios que ejecutaron la prueba de usabilidad	115
Tabla 23. Respuestas a la pregunta 6 de la prueba de usabilidad.....	117
Tabla 24. Respuestas a la pregunta 7 de la prueba de usabilidad.....	117
Tabla 25. Respuestas a la pregunta 8 de la prueba de usabilidad	118
Tabla 26. Respuestas a la pregunta 9 de la prueba de usabilidad.....	118
Tabla 27. Respuestas a la pregunta 10 de la prueba de usabilidad.....	119

Resumen

La evaluación de la seguridad y efectividad derivada del uso de la tecnología se conoce como Evaluación de Tecnologías en Salud (ETS), este proceso debe ser sistemático y está enfocado en los efectos previstos y no previstos del uso de la tecnología, y en el largo y corto plazo. Dentro del ciclo de vida de la tecnología la evaluación hace parte de las primeras etapas y tiene una gran relevancia, dado que, desarrollarla adecuadamente potencia el cumplimiento de las demás. Estas etapas son: *evaluación*, presupuesto, adquisición, educación, cumplimiento, mantenimiento, seguridad, reposición. La ETS se implementa como una buena herramienta para liderar la correcta incorporación de tecnología en cualquier sistema de salud, particularmente a nivel hospitalario es realizada por profesionales de la salud y contribuye a optimizar los recursos de las instituciones. En el mundo existen cuatro grandes metodologías para desarrollar la ETS en hospitales, en Colombia, la más común es la mini-ETS, debido a la complejidad de los equipos de trabajo y su enfoque en apoyar la toma de decisiones. Aunque la contribución de la ETS a la toma de decisiones podría mejorar las inversiones y la seguridad del paciente, en el país no es obligatorio bajo la normativa nacional, pero sí por la acreditación en salud, que es un proceso voluntario enfocado en garantizar y mejorar la calidad de la atención en salud. La ETS en los hospitales, como cualquier otro proceso relacionado con síntesis de evidencia clínica, puede ser complejo y lento debido a la dificultad para identificar la evidencia necesaria para responder una pregunta de investigación y al análisis de la información disponible. Dentro de los asuntos a apoyar y fortalecer en los procesos de ETS se encuentran la generación de adecuadas preguntas de investigación, estrategias de búsqueda y análisis de la evidencia, así como la necesidad de mejorar la trazabilidad del proceso por medio de una adecuada documentación del mismo.

El procesamiento del lenguaje natural (PLN) y el aprendizaje automático (y profundo) son herramientas poderosas que se utilizan para automatizar tareas en muchos campos de la medicina. En particular, la ETS a nivel hospitalario puede usar estas herramientas para potenciar y automatizar la recopilación de evidencia y reducir la carga de trabajo de los investigadores.

Para aportar significativamente en la solución del problema planteado, este trabajo tuvo como objetivo principal desarrollar una herramienta para automatizar la recuperación y síntesis de evidencia clínica en hospitales colombianos acreditados. El primer paso para alcanzar este objetivo fue caracterizar el proceso de ETS en cuatro hospitales acreditados del país, aplicando una encuesta basada en el modelo mini ETS, el Manual AdhopHTA y estudios recientes relacionados con el desarrollo de la ETS en Latinoamérica y Colombia. El resultado más relevante muestra que en todos los hospitales la ETS es realizada por un grupo interdisciplinario de profesionales, asimismo, la evaluación clínica de la tecnología se basa en datos generados a través de búsquedas bibliográficas y la experiencia clínica. En la mayoría de los casos, la incorporación de tecnología en salud es precedida por un proceso de ETS, esta una de las mejores prácticas identificadas, así como la definición de los resultados en salud y los beneficios clínicos de la tecnología antes de evaluarla. Los hospitales que fueron consultados usan PubMed como la base de datos para sus búsquedas,

por lo cual es la base de datos que se consulta a través de la herramienta de síntesis diseñada. En segundo lugar, a través de una búsqueda manual se recopilaron 49 informes de ETS y a partir de la descripción de los estudios clínicos incluidos en cada uno de estos, se levantó una base de datos de artículos. Se recopilaron 523 estudios clínicos, que evaluaban dispositivos médicos destinados a diagnosticar o tratar afecciones cardiovasculares. A partir de este conjunto de datos, se compararon diferentes técnicas para clasificar Ensayos Clínicos Controlados Aleatorios (ECA) y Revisiones Sistemáticas (SR). Adicionalmente, se propuso y aplicó en un informe de ETS un método para priorizar títulos y resúmenes en el proceso de ETS. Finalmente, se diseñó una herramienta de acceso abierto con la integración de la búsqueda, filtro y síntesis de información para el proceso de ETS, la cual fue evaluada a través de una prueba de usabilidad por profesionales de los hospitales acreditados mencionados anteriormente. De los resultados obtenidos a través de la prueba de usabilidad se destaca que se diseñó una herramienta muy fácil de usar, intuitiva para el usuario y muy útil. También se identificó que todos los usuarios que usaron la herramienta convinieron en que facilita el proceso de síntesis de estrategias y búsquedas, así como la visualización de resultados. Además, se resaltó la importancia de la trazabilidad del proceso que se genera a través de la síntesis de reportes e indicaron tendencia por seguirla usando.

Palabras clave: Evaluación de Tecnología en Salud, Procesamiento de Lenguaje Natural, acreditación en salud, usabilidad, automatización.

Abstract

To assess the safety and effectiveness arising from the use of technology is known as Health Technology Assessment (HTA), this process must be systematic and is focused on the intended and unintended effects, and in the long and short terms. The HTA is implemented as a great tool to lead the right choices in any health system. The Hospital-Based HTA is conducted by hospitals staff and contributes to optimizing the institutions' resources. There are four big approaches to developing HTA in hospitals, in Colombia, the most common is the mini-HTA, due to the organizational complexity and its focus on managerial decision making. Although its contribution to decision-making could improve budget usage and patient safety, it is not compelled by national regulation, but it is by accreditation in health, which is a voluntary process focused on guaranteeing and improving the quality in healthcare. The HTA in hospitals and any process related to evidence synthesis can be complex and time-consuming. Among the issues to be supported and strengthened are the formulation of the right research questions, search strategies evidence analysis, and finally, the documentation of the HTA complete process.

Natural Language Processing (NLP) and machine (and deep) learning are powerful tools used to automate tasks in many fields of medicine. Particularly, HB-HTA can use it to leverage and automate the evidence gathering to reduce the researchers' workload.

As mentioned earlier, the HTA is a structured and systematic process, this work aimed to develop a tool to automate the retrieval and synthesis of clinical evidence in accredited Colombian hospitals. The first step to reach this goal was to characterize the HTA process in four accredited hospitals in the country, applying a survey based on the mini HTA model, the AdhopHTA Handbook, and recent studies related to the development of HTA in Latin America and Colombia. The most relevant result shows that in all hospitals the HTA is performed by an interdisciplinary group of professionals, also, the clinical evaluation of health technology is based on data generated through literature searching and clinical experience. Most often health technology incorporation is led up by the HTA process, it is one of the best practices identified, as well as the definition of health outcomes and clinical benefits of technology before assessing it. The database consulted by all hospitals is PubMed, therefore, the automated search in the developed tool is accessed through this virtual library. In the second place through a manual search, 49 reports of HTA were gathered and from the description of the clinical studies included in each of these, a database of HTA reports was raised. 523 clinical studies were collected from those reports, which evaluated medical devices intended to diagnose or treat some cardiovascular conditions. From this dataset, different techniques to classify Random Controlled Clinical Trials (RCT) and Systematic Reviews (SR) were compared. Additionally, a method for screening titles and abstracts in the HTA process was proposed and applied in an HTA report. Finally, an open-access tool was designed with the integration of the search, screening, and synthesis of information for the HTA process, it was evaluated through a usability test by professionals of the accredited hospitals mentioned above. The users rated the tool as easy to use, very useful, and intuitive. This tool facilitates the search strategy synthesis, the searching, and the screening process. Also, the users highlighted the utility of

the reports generated by the tool in the traceability of the clinical evidence synthesis and they stated their preference for using the application for future processes.

Keywords: Health Technology Assessment, Natural Language Processing, accreditation in health, usability, automate.

1. Introducción

1.1. Contexto

La Organización Mundial de la Salud define la tecnología médica como la aplicación de conocimientos y habilidades organizados en forma de dispositivos, medicinas, vacunas, procedimientos y sistemas desarrollados para resolver un problema de salud y mejorar la calidad de vida [1]. La evaluación de la tecnología médica es una forma de investigación que analiza las consecuencias clínicas, económicas y sociales derivadas de su uso, en el corto y en el mediano plazo, y sus efectos directos e indirectos, deseados y no deseados [2].

La Evaluación de Tecnología en Salud (ETS) implica la consideración de varios criterios dentro de los cuales se encuentran los clínicos, regulatorios, organizacionales, técnicos, económicos y de operación [3] [4]. Clínicamente es importante determinar la seguridad, eficacia y efectividad de la tecnología sometida a evaluación [5]. En términos económicos se deben evaluar los costos y efectos sobre la salud del uso de la tecnología, para esto existen varias estrategias como lo son las evaluaciones de costo-beneficio, costo-utilidad, costo-eficacia, costo-efectividad, entre otros. El tipo de estrategia implementada depende de las unidades de medida elegidas para evaluar las consecuencias o efectos del uso de la tecnología [6]. Así mismo, hay otros aspectos éticos, organizacionales y sociales que son de interés dentro de un proceso de ETS [4]. El principal objetivo de una evaluación robusta de los criterios mencionados es guiar adecuadamente la toma de decisiones a diferentes niveles de un sistema de salud, ya sean nacionales (nivel macro), hospitales (nivel meso) o profesionales (nivel micro) [4], [7], [8].

La ETS en Colombia a nivel macro es desarrollada principalmente por el Instituto Nacional de Evaluación de Tecnología en Salud (IETS), a nivel meso es abordada por clínicas, hospitales y aseguradoras y del nivel micro son responsables los profesionales de la salud y pacientes. El IETS tiene como objetivo la producción de información basada en evidencia científica en cuanto a la evaluación de la tecnología en el país para guiar la toma de decisiones relacionadas con la misma. Sin embargo, el mismo ha declarado su incapacidad de evaluar toda la tecnología que ingresa al país debido a su alta difusión [9]. Aun cuando en el país existe una institución encargada de la ETS a nivel macro, es necesario que la evaluación en los niveles meso y micro se desarrolle profundamente también, de tal manera que puedan complementarse y apoyar la toma de decisiones en el sistema de salud [4], [9]. Por ejemplo, evaluar adecuadamente los dispositivos médicos destinados al diagnóstico y tratamiento de enfermedades cardiovasculares puede tener un alto impacto en el sistema de salud colombiano porque la atención en salud de estas enfermedades representa aproximadamente el 7% de todo el presupuesto de la salud del país [10]. Adicionalmente, en los procesos de acreditación en alta calidad en salud, específicamente en los estándares 132 y 133 (GT2, GT3) de gestión de tecnología se exige a las instituciones prestadoras de servicios de salud (IPS) contar con un proceso para la planeación, la gestión y la evaluación de la tecnología, así como tener una política organizacional para adquisición, incorporación, monitorización, control y reposición de la tecnología [11].

En cualquier nivel de ETS, para llevar a cabo principalmente las evaluaciones de seguridad, eficacia y efectividad de la tecnología existen dos métodos epidemiológicos diferentes, los cuales son los métodos de datos primarios y los métodos de datos secundarios. Los métodos de datos primarios consisten en la generación de nueva información por medio de estudios diseñados y ejecutados usando protocolos prospectivos o retrospectivos. Los métodos de datos secundarios se basan en la recolección, análisis y síntesis de los resultados de estudios previos [12] [13]. Entendiéndose la ETS como una extensión de la Medicina Basada en Evidencia, es posible afirmar que es la ETS una forma de investigación que analiza la evidencia disponible para tomar una decisión (métodos de datos secundarios). [14] Dentro de éstos últimos, pueden encontrarse revisiones sistemáticas de la literatura, metaanálisis, modelamiento matemático, entre otros. Sin embargo, tanto nacional como internacionalmente, existe una problemática en cuanto a la poca diseminación de la evidencia específicamente para la evaluación de dispositivos médicos, por lo cual se dificulta su evaluación [3]. Así mismo, se tiene una ETS para dispositivos médicos menos ejecutada y poco usada para guiar la toma de decisiones en salud en los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) [15]. Otros factores que se han identificado como barreras para la implementación de la ETS en Colombia es la falta de adecuados y claros formatos de presentación, falta de redes de políticas y apoyo normativo insuficiente [16]. En su estudio sobre barreras y facilitadores para el uso de ETS en Colombia, Florian Dams y colaboradores destacan la necesidad de formar técnicamente el recurso humano, así como mantener informes más concisos y estandarizados con el fin de reducir cargas de trabajo. También se destaca la relevancia de mejorar la comunicación y cooperación entre investigadores y gestores de políticas [16].

Las revisiones sistemáticas (RS) de la literatura consisten en una búsqueda estructurada de información por medio de criterios de elegibilidad previamente establecidos con el objeto de identificar, analizar y sintetizar toda la evidencia existente para dar respuesta a una pregunta de investigación específica [17]. Generalmente en la ETS de dispositivos médicos, estas preguntas están enfocadas a responder sobre la pertinencia de incorporación de estos dispositivos y cómo debería hacerse [1]. Las RS constituyen una parte fundamental dentro de un procesos de evaluación de tecnología en salud [18]. Sin embargo, el proceso de búsqueda y filtro de información para su posterior análisis y síntesis en las RS toma mucho tiempo y representa altas cargas de trabajo para el personal de salud que las desempeña [19]. Esto se debe a que en cualquier proceso de RS se debe hacer una exhaustiva búsqueda en las fuentes de información pertinentes y además existe una evidencia muy limitada y de difícil identificación para evaluar seguridad, eficacia, efectividad y costo beneficio particularmente en dispositivos médicos [20]. Una revisión sistemática rápida enfocada en proveer en los hospitales una referencia para la toma de decisiones sobre equipos, dispositivos y procedimientos médicos, tiene un tiempo de producción entre 6 y 12 semanas según AdHopHTA (siendo esta una de las más cortas entre los tipos de reportes de ETS enfocado en hospitales) [21]. En los últimos años, la comunidad científica se ha enfocado en automatizar las revisiones sistemáticas, de manera que se pueda reducir el tiempo de ejecución de estas, así como reducir el riesgo de sesgo en su

ejecución. Los sesgos de las revisiones sistemáticas van desde sesgos de selección (no identificar toda la evidencia sobre un tema en específico), hasta la inclusión errónea de los estudios primarios a analizar [22].

Los avances en el procesamiento de lenguaje natural, minería de texto y el aprendizaje de máquina han permitido generar algoritmos que pueden reproducir el esfuerzo humano para realizar las RS [18], [23], [24]. La automatización de procesos como la identificación de estudios para RS por medio del uso de minería de texto han mostrado reducciones de carga de trabajo de entre un 30 y 70 % [25]. La evaluación de riesgo de sesgo para los ensayos clínicos aleatorizados incluidos en una revisión sistemática es una tarea importante dentro de la misma, y por esto, varios investigadores han mostrado que la evaluación de riesgo de sesgo en ensayos clínicos aleatorizados puede ser automatizada con una exactitud razonable (71.0% para el análisis automático vs 78.3% para el análisis manual) y la extracción automática de información para soportar las predicciones de sesgo tuvieron una calidad tan alta como aquella información que se extrajo de forma manual [19]. En Colombia, se tiene que el IETS ha implementado técnicas de aprendizaje automático y procesamiento de grandes volúmenes de datos para hacer ETS, no obstante, manifiesta que al ser la ETS una labor específica de tecnología particulares, no se tiene un proceso estandarizado y reproducible en el cual desarrollen ETS de manera automatizada [26].

A partir de lo expuesto surge la necesidad de implementar herramientas y procesos que permitan apoyar a los profesionales en la ejecución de revisiones sistemáticas de la ETS enfocadas en dispositivos médicos, con el objetivo de guiar la toma de decisiones para la incorporación y selección de tecnología en las instituciones prestadoras de salud en Colombia. Esta necesidad puede ser abordada por medio de la automatización de subprocesos de las revisiones sistemáticas, tal que se reduzcan los tiempos globales de ejecución del proceso de ETS.

Este documento está estructurado como sigue: la primera sección contiene el contexto general en el que se enmarca, el objetivo general y objetivos específicos planteados. Después se presenta el estado del arte y marco teórico tanto de procesamiento de lenguaje natural y revisiones sistemáticas de la literatura como el panorama general de la evaluación de tecnología en salud. A continuación, se sintetizan la metodología y resultados para cada uno de los objetivos planteados y finalmente se esbozan las principales conclusiones, contribuciones del trabajo de investigación, limitaciones y trabajo futuro.

1.2. Objetivos

1.2.1. *General*

Desarrollar una herramienta semiautomática para la extracción de información en revisiones sistemáticas a partir de evidencia clínica en procesos de evaluación de tecnología en salud, enfocada en dispositivos médicos para IPS acreditadas o en proceso de acreditación en Colombia.

1.2.2. *Específicos*

- ✓ Seleccionar y caracterizar una metodología de evaluación de tecnología en salud de dispositivos médicos, enfocada y aplicada en instituciones prestadoras de servicios de salud acreditadas o en proceso de acreditación en Colombia.
- ✓ Consolidar una base de datos de evaluaciones de tecnología en salud en dispositivos médicos a partir de reportes de evaluación de instituciones reconocidas por entidades gubernamentales y no gubernamentales.
- ✓ Adaptar los algoritmos de la herramienta *RobotReviewer*, con el fin de hacer transferencia de aprendizaje y automatizar la extracción información de fuentes de evidencia clínica.
- ✓ Diseñar una herramienta que integre la automatización de algoritmos de búsqueda, clasificación y extracción de información para revisiones sistemáticas, enfocada en fuentes de evidencia clínica ajustada a la metodología de ETS elegida en IPS colombianas.

2. Estado del arte

La Evaluación de Tecnología en Salud (ETS) es una herramienta que facilita la toma de decisiones en cualquier nivel del sistema de salud que sea aplicada, requiere del análisis de múltiples criterios y la integración de profesionales de diferentes áreas. Asegurar inicialmente que la tecnología sea segura y efectiva es una de las primeras etapas de la ETS y es fundamental en cualquier proceso de ETS [27]. Lo anterior se puede evidenciar dentro de los modelos más recientes para llevar a cabo las ETS, uno de ellos es el modelo Core desarrollado por la agencia europea para la evaluación de tecnología en salud (EUnetHTA). Este modelo se basa en el análisis de nueve dominios, de los cuales el tercero y cuarto corresponden a la seguridad y efectividad clínica de la tecnología respectivamente [28]. La iniciativa, AdhopHTA también integra estos dos criterios, haciendo una subdivisión del dominio de efectividad clínica en dos aspectos, los cuales son tamaño del efecto del resultado clínico y calidad de evidencia [21]. Así mismo, AdhopHTA propone una herramienta denominada “*AdHopHTA mini-HTA template*” la cual es una evolución del “*mini-HTA template*” desarrollado por el centro Danés para evaluación de tecnología en salud [29]. Siendo este último uno de los cuatro grandes modelos para hacer ETS en hospitales a nivel mundial [30] y principalmente el que más se desarrolla en el contexto latinoamericano y colombiano [31], [32].

Dentro de los diferentes organismos conformados para guiar la aplicación de evaluación de tecnología se identifica uniformidad en la prioridad de evaluar la misma con base en análisis sistemático de la mejor evidencia disponible con el fin de tomar las decisiones más convenientes para los pacientes [29], [33]. Tal como se puede evidenciar por medio de *SuRe Info*, página web reciente que resume información para la recuperación de información en ETS, basada en investigaciones sobre la identificación de evidencia efectiva y eficiente para los diferentes aspectos de la evaluación de tecnologías de la salud y síntesis de evidencia

En los últimos años, la automatización de tareas ha tomado cada vez más acogida en diversos campos del conocimiento y el dominio biomédico no es la excepción. A través de esta automatización es posible facilitar los procesos de revisión de la literatura para tomar decisiones relacionadas principalmente con aspectos de seguridad y efectividad de la evaluación de tecnología en salud. Esto se ve reflejado en el interés creciente de automatizar procesos como lo son las revisiones sistemáticas (RS) de literatura, fundamentales para la toma de decisiones en salud. Tanto así, que se generó una colaboración internacional para la automatización de revisiones sistemáticas (ICASR por sus siglas en inglés) [34]. Dentro de esta colaboración se encuentran organizaciones como el Instituto Nacional de Ciencias de la Salud Ambiental (NIEHS) y el instituto Cochrane, que está enfocado en sintetizar evidencia clínica de la mejor calidad a través de revisiones sistemáticas para apoyar las decisiones relacionadas con salud en el mundo [35]. En la página web de la colaboración actualmente se encuentran 112 herramientas (de acceso libre y bajo licencias) para apoyar la ejecución de las etapas de una revisión sistemática, algunas de ellas son la síntesis de estrategias de búsqueda, filtrado de estudios, extracción de información y listas de chequeo para análisis crítico de revisiones [36]. La automatización de actividades como éstas se logra a través de la implementación del Procesamiento de

Lenguaje Natural (PLN), el cual está enfocado en hacer posible comunicación entre humanos y máquinas por medio del uso de lenguaje humano, en el caso de este trabajo de investigación se trabaja con el lenguaje expresado por medio de la escritura [37].

Los avances que se han realizado en la automatización de tareas de las revisiones sistemáticas por medio del PLN están enfocados en tareas tales como extracción y análisis de información, análisis de sesgo y filtrado de artículos. Desde su implementación, los métodos de PLN han variado mucho, van desde el conteo de palabras hasta la implementación de redes neuronales profundas. Actualmente, el estado del arte con relación al PLN se basa en redes neuronales por medio del uso de Transformadores Bidireccionales para el Entendimiento del Lenguaje (*BERT* por sus siglas en inglés) [38]. Su éxito radica en la implementación del mecanismo de atención, en el cual se analiza la interacción de cada una de las palabras de una sentencia con las demás [39]. *BERT* fue desarrollado en 2019 y a partir de su publicación y con base en este modelo múltiples investigadores de todos los campos han entrenado otros para ser utilizados en dominios específicos. *SciBERT* [40] y *BioBERT* [41] son dos adaptaciones de *BERT* para el entendimiento del lenguaje en el campo médico. *SciBERT* actualmente es implementado en la herramienta *RobotReviewer* como una mejora para extraer texto que soporta el análisis de sesgo en ensayos clínicos aleatorizados (con una exactitud del 71.0% en su primera versión) [19]. En un estudio realizado en 2021 se identificó que el uso de esta y otras herramientas para la automatización de SR reduce el tiempo de ejecución de las revisiones según los investigadores que las ejecutan, entre los cuales se encuentran asesores para la ETS [42]. Otra herramienta desarrollada por los mismos desarrolladores de *RobotReviewer* es *RobotSearch*, que con base en algoritmos de aprendizaje de máquina (máquinas de soporte vectorial) fue entrenado para la clasificación de ensayos clínicos con un desempeño para el área bajo la curva ROC de 0.987 [43]. Otro software útil al analizar información en revisiones de literatura es *PICO Parser* [44], un algoritmo entrenado para identificar los términos PICO (Paciente, Intervención, Comparador, Resultados – Outcomes en inglés) en ensayos clínicos. Este último es adaptado de *BLUEBERT*, un modelo basado en *BERT* que demostró los mejores resultados al ser evaluado bajo una métrica de Evaluación para el Entendimiento del Lenguaje Biomédico (*BLUE* por sus siglas en inglés) [45]. Es importante mencionar que estos desarrollos han sido producto de investigaciones del Centro Nacional para la Información Biotecnológica (NCBI) apoyados por la Biblioteca Nacional de Medicina (NLM) de Estados Unidos.

Todas estas herramientas están basadas en el análisis de Ensayos Clínicos Aleatorizados, los cuales son los diseños de estudios preferidos para evaluar tecnología, en conjunto con otras revisiones sistemáticas y reportes de evaluación de tecnología previos. Alineado con el avance de los algoritmos de PLN y aprendizaje de máquina y profundo, en el campo médico y de investigación se han realizado avances que pueden contribuir a facilitar la automatización de las revisiones sistemáticas. Dentro de dichos avances se encuentran los esfuerzos por estandarizar el reporte de los estudios como *PRISMA* [46] (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) y *CONSORT* (CONSolidated Standards of Reporting Trials) [47]. Así mismo, iniciativas como la de la FDA sobre la

unificación de definiciones y términos en ensayos clínicos relacionados con afecciones cardiovasculares [48].

A través de la revisión e integración del estado del arte mencionado con relación al PLN y los modelos más recientes para evaluar tecnología en salud, en este trabajo de investigación se desarrolló una herramienta de síntesis semiautomática de evidencia clínica. Dicha herramienta fue diseñada con el objetivo de ser implementada y apoyar y eventualmente mejorar los procesos de ETS en los hospitales acreditados del país [49].

3. Marco teórico

El marco teórico de este trabajo de investigación se aborda desde dos brazos, el primero hace referencia a la Evaluación de Tecnología en Salud, las metodologías existentes para su implementación en hospitales, el contexto de la mini-ETS en el mundo y en Colombia. El segundo está enfocado en los métodos de automatización como el Procesamiento de Lenguaje Natural y aprendizaje de máquina. En esta segunda parte se exponen las bases y principales modelos que se tuvieron en cuenta para el cumplimiento de los objetivos planteados.

3.1. Evaluación de Tecnología en Salud

La evaluación de tecnología en salud se define como el análisis de las implicaciones y resultados de la implementación de una tecnología en salud. Teniendo en cuenta los efectos deseados y no deseados tanto en el corto como en el largo plazo. El análisis de la intervención o implementación de la tecnología comprende varios criterios, dentro de los cuales se encuentran criterios económicos, de seguridad, efectividad, criterios éticos, sociales y en algunos casos organizacionales [50]. Se entiende tecnología en salud como todas las actividades, dispositivos e instrumentos usados con el objeto de promover la salud y prevenir, tratar, diagnosticar, monitorear, rehabilitar una enfermedad, así como reemplazar y/o dar soporte a distintas partes del cuerpo, es decir, resolver problemas sanitarios y mejorar la calidad de vida [1].

La evaluación de la tecnología en salud debe realizarse durante todo el ciclo de vida de la tecnología médica. Este ciclo de vida se define por el Ministerio de salud y protección social (MSPS) de Colombia en dos etapas, las cuales son premercado y posmercado como se puede visualizar en la Figura 1.

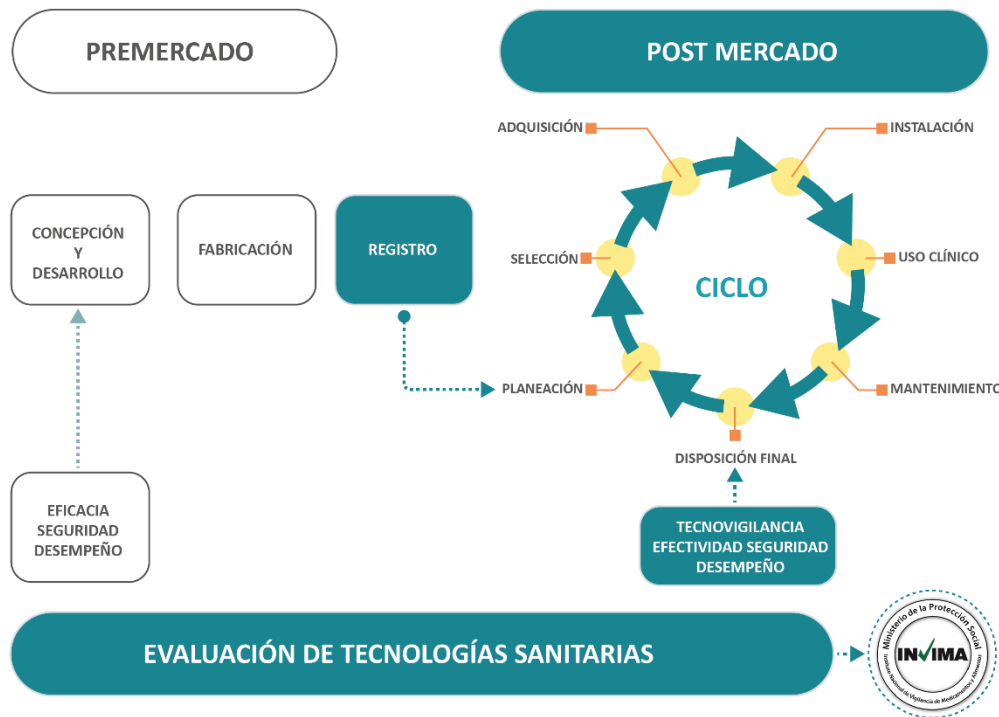


Figura 1. Ciclo de vida de la tecnología en salud.

A nivel meso u hospitalario generalmente se lleva a cabo toda la etapa de poscomercialización, donde se realizan actividades de planeación, selección, adquisición, instalación, mantenimiento, bajas, reposición, traslados y disposición final [51]. Si bien estas actividades se realizan mayormente en los hospitales, la normativa colombiana insta a los proveedores de tecnología (ya sea fabricantes o importadores) a trabajar en conjunto con las instituciones para garantizar que la tecnología se use en condiciones óptimas. Lo anterior está regulado específicamente en el artículo 38 *Posventa de los dispositivos médicos considerados equipos biomédicos y su mantenimiento* del decreto 4725 de 2005 [52]. La intensidad de uso y la evaluación de la tecnología dependen de la etapa del ciclo de vida en el que se encuentre y varían en función del tiempo como se puede observar en la Figura 2.

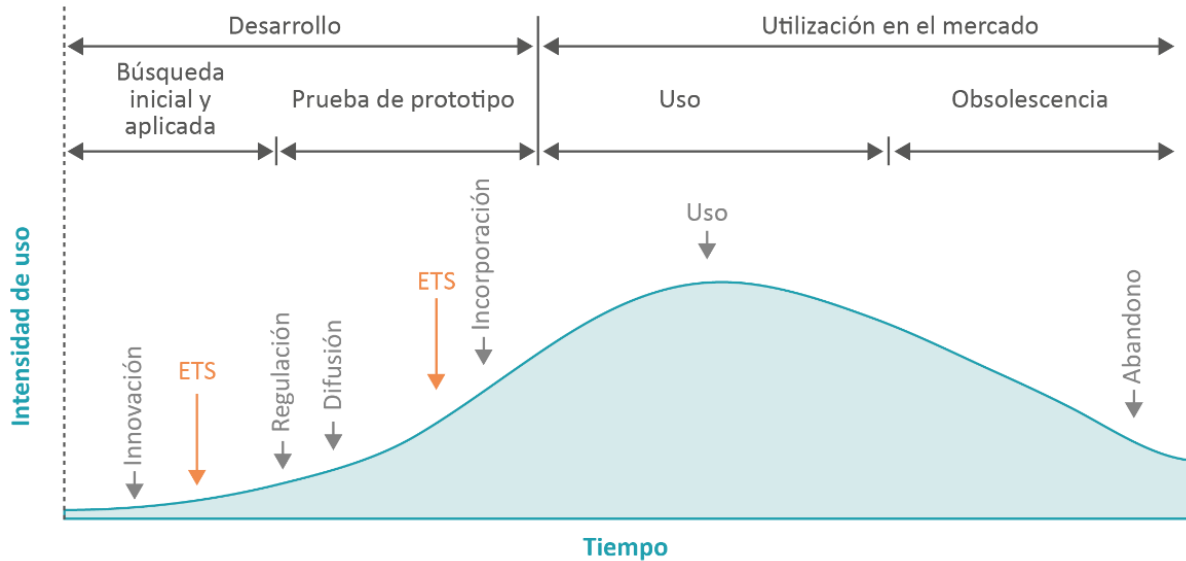


Figura 2. Intensidad de uso y evaluación de tecnología en salud en función del tiempo Fuente: [53], [54]

Los principales aspectos para evaluar en un proceso de ETS son varios, a continuación de describen los más abordados en la literatura:

- ✓ **Seguridad:** Se relaciona con los posibles daños y eventos adversos causados derivados por la implementación de la tecnología [4].
- ✓ **Eficacia:** Evalúa que la probabilidad de que el uso de la tecnología dé los resultados esperados en condiciones ideales para cierto paciente [4].
- ✓ **Efectividad:** Evalúa que la probabilidad de que el uso de la tecnología de los resultados esperados en condiciones reales para cierto paciente [4].
- ✓ **Costo-beneficio:** Evalúa la relación entre los recursos consumidos (unidades monetarias) y los resultados producidos medidos en unidades monetarias [6].
- ✓ **Costo-utilidad:** Evalúa la relación entre los recursos consumidos (unidades monetarias) y la supervivencia ajustada por calidad de vida [6].
- ✓ **Ética:** Analiza factores como la justicia, el derecho a la salud, la imparcialidad, obligación moral, entre otros aspectos éticos que se relacionan con la adopción de la tecnología [5].
- ✓ **Aspectos organizacionales:** Evalúa como la adopción de la tecnología puede modificar la estructura organizacional de una institución en relación a los profesionales y su rol dentro de la misma [4].
- ✓ **Social:** Tiene relación con el impacto que generaría la implementación de la tecnología en la población en cuanto a aspectos sociales, culturales, políticos y organizacionales [5].

La ETS implica como se definió anteriormente el análisis de las consecuencias que se derivan del uso de esta, lo cual es una actividad amplia. Existen diferentes procesos que componen la ETS, los cuales permiten tener un análisis completo y suficiente de la tecnología. Dentro de estos procesos se encuentran una evaluación regulatoria, la cual tiene el objetivo de dar un aval o registro nacional para la utilización de la tecnología. La

evaluación para políticas en salud en la cual se establece normativamente su utilización o implementación dentro del sistema nacional de salud. La evaluación en hospitales (nivel meso) la cual define la adopción de tecnologías de acuerdo con la situación particular de una institución y finalmente una evaluación para adquisición donde se da una selección de la mejor oferta técnica, clínica y económica posible de acuerdo con las necesidades de una institución y la población que atiende [4].

La ETS se ha convertido en una herramienta para guiar la toma de decisiones en los sistemas de salud en todos sus niveles de aplicación, sin embargo, en una encuesta realizada entre los años 2014 y 2015 se identificó que en América Latina solamente en Brasil se obliga a implementar la ETS la toma de decisiones en salud [55]. Así mismo, se resalta que una de las barreras existentes para ejecutar ETS es la falta de personal idóneo y de recursos financieros para realizar estas actividades y se menciona la estrecha relación que existe entre personal capacitado y las decisiones basadas en la evidencia.

La baja diseminación de literatura que demuestre el impacto económico de la implementación de ETS es una brecha para su análisis principalmente en países como Colombia. Sin embargo, algunos estudios publicados ilustran el impacto que puede tener la implementación de un proceso de ETS [56]. Ejemplos de esto se tienen en Canadá, donde se indica que cerca del 71% de las recomendaciones realizadas en los reportes de evaluación de la unidad de ETS en el Hospital universitario de McGill son aceptadas y tenidas en cuenta en sus decisiones. Adicionalmente, se reporta un ahorro de \$CAN 1,140,958 debido a la implementación de las recomendaciones de 15 reportes de ETS entre los años 2004 y 2011 [57]. Así mismo la unidad de ETS en Calgary estimó un ahorro de \$ CAN 1'000.000 anualmente debido a la evaluación y estandarización de suministros ortopédicos [58]. En Austria, el 48% de las recomendaciones realizadas por el Instituto Ludwig Boltzmann fueron consistentes con las decisiones de los hospitales y 40 % de las tecnologías no recomendadas fueron incorporadas con restricciones. El ahorro estimado fue de 3 millones de euros por cada reporte de ETS (con una alta variación) [59].

La Evaluación de Tecnología en Salud en Hospitales

Las actividades de ETS son desarrolladas en diferentes ámbitos y de acuerdo con estos se puede clasificar en tres niveles, los cuales son el macro, meso y micro. El nivel macro es abarcado por países e instituciones que analizan un impacto nacional de la adopción de una tecnología. El nivel meso es abordado en instituciones prestadoras de servicios de salud y analiza el impacto sobre la población que atiende. Finalmente, el nivel micro es abordado por profesionales de la salud para guiar la toma de decisiones en la práctica clínica diaria con sus pacientes. Criterios como seguridad y efectividad son transversales a todos los niveles de evaluación de tecnología en salud y dan cuenta de posibles daños o efectos adversos que se pueden presentar como consecuencia de la evaluación de tecnología, así como de la probabilidad que un individuo pueda obtener beneficios de la implementación de la tecnología en condiciones reales o normales de uso [4].

En particular, a nivel hospitalario se caracteriza por analizar detalladamente los impactos organizacionales que puede tener la adopción de una tecnología en salud. Se

enfoca en el grupo de pacientes que interviene y su perfil epidemiológico y también hace un análisis de la capacidad de adopción de la tecnología [4]. A lo largo del tiempo se ha desarrollado la ETS a nivel hospitalario con diferentes modelos o metodologías, principalmente desde el año 2005, globalmente se han hecho esfuerzos por estandarizar el proceso con la colaboración de diferentes organizaciones.

En el año 2006 la sociedad internacional de evaluación de tecnología en salud HTAi, clasificó las metodologías para realizar ETS en hospitales en cuatro grandes grupos como se ilustra en la Figura 3. Esta clasificación se estableció con base en dos criterios, el objeto de la ETS (toma de decisiones o guiar la práctica clínica) y el nivel de complejidad del equipo de trabajo que desarrolla la evaluación (profesionales independientes o equipos de trabajo interdisciplinarios) [30]. Los dos modelos enfocados en guiar la toma de decisiones son la unidad de evaluación de tecnología en salud y la mini-Evaluación de Tecnología en Salud (mini-ETS). Según la HTAi, las unidades de ETS son el nivel más complejo de evaluación e implica profesionales dedicados completamente a realizar ETS, en la mini-ETS profesionales independientes recogen y analizan información del contexto organizacional para informar a tomadores de decisiones, este modelo fue tomado del Centro Danés de Evaluación de Tecnología en Salud (DACEHTA).

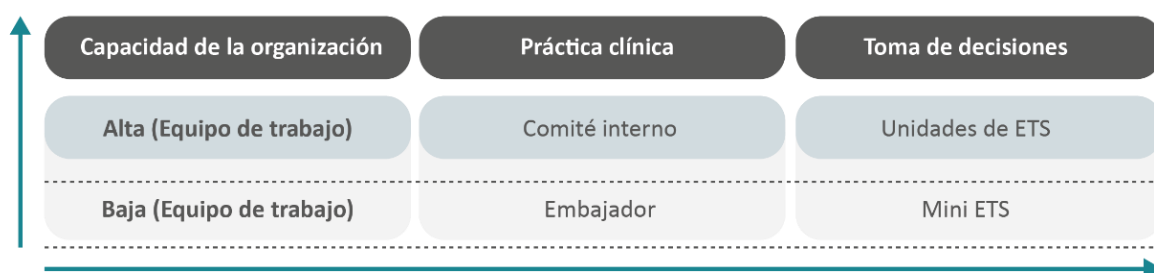


Figura 3. Estrategias de ETS con enfoque hospitalario, adaptado según [30]

El Centro Danés desarrolló la mini-ETS en 2000 como una herramienta que pudiera usarse fácilmente en los hospitales y en organizaciones regionales que necesitaran evaluar la implementación de una nueva tecnología. Esta herramienta guía la evaluación de tecnología a través de una lista de chequeo que abarca los aspectos relacionados con la ETS y se enfoca en la tecnología, los pacientes a impactar, los aspectos organizacionales y económicos [29].

Para el año 2015, se desarrolló el proyecto AdhopHTA en Europa, el cual pretendía adoptar la evaluación de tecnología en salud basada en hospitales (ETS-BH) en dicho continente. En este proyecto participaron más de 380 personas de 20 países, dentro de sus resultados se encuentran 15 principios de buenas prácticas de ETS-BH, pautas para guiar la evaluación de tecnología de acuerdo con cuatro dimensiones (proceso de evaluación, liderazgo, estrategias y colaboración, recursos e impacto), la platilla de mini-ETS de AdhopHTA, entre otros. AdhopHTA considera la mini-ETS como un tipo de reporte, corto y estructurado sobre el uso de una tecnología en una población en específico, generalmente sintetizado en una lista de preguntas relacionadas con aspectos clínicos, económicos,

organizacionales y de seguridad [33]. De este proyecto surgió una nueva plantilla de mini-ETS, la cual es una ‘actualización’ o evolución de la herramienta desarrollada por DACEHTA que integra las experiencias y resultados de la investigación desarrollada por AdhopHTA.

Herramienta AdHopHTA mini ETS

Esta herramienta está constituida por 28 preguntas enfocadas en sintetizar un completo reporte de evaluación de tecnología que cubra todas las dimensiones relevantes. En la Tabla 1 se sintetiza el enfoque de cada una de las preguntas:

Tabla 1. Síntesis de la herramienta AdHopHTA mini ETS

Preguntas	Enfoque	Características adicionales
1	Resumen	Enfocado en sintetizar un resumen de la evaluación, motivaciones, aspectos de seguridad, efectividad, comparador y recomendaciones de la evaluación
2-7	Información básica	Estas preguntas sintetizan quien propone la evaluación, los autores del reporte, actores involucrados, conflictos de intereses, y una definición muy clara de la tecnología a evaluar por medio del uso de la estrategia TICO. Tecnología, Indicación, Comparador y Resultados (Outcomes)
8-12	Aspectos metodológicos generales y reporte *	Esta sección está enfocada en describir metodológicamente cómo se lleva a cabo la síntesis de evidencia con relación a los siguientes aspectos: <ul style="list-style-type: none"> • Revisión de literatura (Fechas de búsqueda, términos clave utilizados, bases de datos consultadas, criterios de selección, diagramas de flujo, etc) • Fuentes adicionales de información • Calidad de la información disponible • Referencias • Estudios en progreso
13-23	Resultados por dominio	Enfocado en describir los resultados de acuerdo con los dominios de: <ul style="list-style-type: none"> • Efectividad clínica • Seguridad del paciente • Aspectos económicos • Aspectos organizacionales • Percepción del paciente • Aspectos estratégicos • Otros aspectos potenciales
24-28	Discusión, conclusión y recomendaciones	Sección enfocada en describir la discusión con relación a los resultados de la evidencia identificada y dar las conclusiones y recomendación pertinente con relación a la adopción de la tecnología en evaluación. Esta última de acuerdo con los resultados y análisis ejecutado

*Este trabajo de investigación está enfocado en apoyar la síntesis de evidencia, por lo cual principalmente ayudará a responder por esta sección de la evaluación de tecnología en salud en los hospitales.

Algunos aspectos importantes para dar cumplimiento a las preguntas 8 a la 12 de la Tabla 1 están relacionados con las estrategias para sintetizar las preguntas de investigación adecuadas, ejecutar la búsqueda y analizar la calidad de la evidencia clínica disponible. Para el planteamiento de la pregunta de investigación existen estrategias tales como el uso del acrónimo PICO o TICO [4], los cuales están enfocados en que la pregunta contenga la información más relevante y se pueda llevar a resultados que den una adecuada respuesta. A continuación, se describe que significa cada una de ellas:

- ✓ **P:** pacientes, donde se describen las principales características de la población objeto, dentro de estas se encuentran, por ejemplo: edad, género, educación y nivel de riesgo.
- ✓ **I:** intervención, este ítem requiere la descripción de la intervención en salud a evaluar, ya sea un medicamento, procedimiento o dispositivo médico.
- ✓ **C:** comparador, en este ítem se requiere describir la intervención con la cual se va a comparar el medicamento, procedimiento o dispositivo de interés.
- ✓ **O:** *Outcomes* o resultados, aquí se deben desagregar todos los resultados de interés en la evaluación, por ejemplo: cambios en la mortalidad, enfermedad, efectos adversos, calidad de vida de los pacientes o estancia hospitalaria.

En el caso de la estrategia TICO se tiene lo siguiente:

- ✓ **T:** tecnología, es la variación de la estrategia anterior, dado que está enfocado en que específicamente se describa la tecnología (generalmente dispositivos médicos) a evaluar.
- ✓ **I:** indicación, en este ítem se describen las características de la población sobre la cual se aplicará la tecnología, son las mismas que se sintetizan en el término Paciente de la estrategia PICO.
- ✓ **C:** comparador, este ítem se conserva de la estrategia PICO y hace referencia a describir la intervención con la cual se va a comparar la tecnología de interés.
- ✓ **O:** Outcomes o resultados, este ítem se conserva también de la estrategia PICO y precisa de la descripción de los principales resultados clínicos de interés a medir debido al uso de la tecnología objeto de evaluación.

Con relación a la calidad de la información disponible a la que se hace referencia en las preguntas 8 a la 12 de la Tabla 1, existen diferentes estrategias para hacer este análisis y varían de acuerdo al tipo de estudio que se esté analizando. A continuación, se describirán algunas de las estrategias más relevantes y la calidad que se evalúa:

- ✓ Para ensayos clínicos la calidad del estudio se analiza en función del riesgo de sesgo, este riesgo se clasifica en seis grandes categorías: aleatorización de participantes, asignación, ocultamiento de intervención para participantes y personal, ocultamiento para análisis de resultados, reporte incompleto de resultados, reporte selectivo [60].
- ✓ En las revisiones sistemáticas se analizan varios criterios, dentro de los más relevantes están: Un protocolo establecido antes de la revisión, adecuada búsqueda de literatura, justificación de estudios excluidos, riesgo de sesgo de los estudios incluidos, métodos de análisis, evaluación de la presencia e impacto probable de sesgo en la publicación. Una estrategia para analizar una revisión sistemática es el checklist AMSTAR (*Assessment of Multiple Systematic Reviews*), que es una herramienta de evaluación crítica de revisiones sistemáticas de intervenciones en salud [61]. Analizar las revisiones bajo la herramienta PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) también es una manera de verificar la calidad de estas [46].

- ✓ Para estudios observacionales se sigue una dinámica similar a la anterior, dentro de las estrategias para analizarlos se encuentra la declaración de la iniciativa STROBE (*Strengthening the Reporting of Observational studies in Epidemiology*) [62].
- ✓ Con relación a estudios de exactitud diagnóstica también se cuenta con una herramienta para analizar la calidad de la evidencia, la cual se denomina QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*) [63].

Una vez es analizado cada estudio en particular, existen estrategias para analizar globalmente la calidad de evidencia recuperada para guiar la revisión. Dentro de estas se destaca la escala GRADE (Grading of Recommendations, Assessment, Development and Evaluation), que está enfocado en dar una calificación total de calidad. Se basa tanto en el tipo de estudios incluidos como en el análisis de sesgo que se realiza a cada uno de ellos de manera individual [64].

3.1.1. Implementación de mini ETS

Desde su desarrollo, la mini-ETS ha sido implementada y gradualmente se ha convertido en un modelo de gran interés para el desarrollo de la ETS en diferentes países y también en el contexto colombiano.

En Noruega, tecnologías específicas financiadas por el servicios de salud noruego como medicamentos son evaluadas desde una perspectiva macro o nacional y el resto de tecnologías es posible evaluarlas a un nivel hospitalario y generalmente esta evaluación es guiada bajo el modelo de mini-ETS [65]. En 2017, fue llevado a cabo un estudio para identificar la idoneidad de la implementación de mini-ETS en hospitales universitarios franceses, los investigadores concluyeron que es una herramienta que puede potenciar la implementación de la ETS, sin embargo, previamente es necesaria una adaptación del modelo al contexto local [66]. También en 2017, se publicó la experiencia inicial de la implementación de ETS-BH en Brasil, de acuerdo con las entrevistas realizadas en hospitales del núcleo de ETS-BH entre 2011 y 2012 se identificó que la mayoría de las evaluaciones fueron reportadas en formato de mini-ETS [31]. El primer programa de ETS-BH implementado en un hospital público de Argentina también hizo uso de la herramienta mini-ETS, adaptándola al contexto local con el objeto de identificar claramente la tecnología a evaluar según los requerimientos de los profesionales de la institución [32].

3.1.2. Contexto colombiano para la evaluación de tecnología en salud basada en hospitales

En un estudio realizado en el año 2018, se identificaron las principales barreras y facilitadores para la implementación de la evaluación de tecnología en salud en Colombia. En este estudio participaron diferentes actores del sistema de salud, incluyendo algunos profesionales del Instituto de Evaluación de Tecnología en Salud (IETS). Dentro de los principales facilitadores se identificó que existe recurso humano calificado para realizar ETS, el tiempo para realizar la evaluación es considerado adecuado y existe disponibilidad para hacer investigación. Como barreras se identificaron el insuficiente soporte legal y la ausencia de políticas para desarrollar ETS, así como poco involucramiento de tomadores de decisiones o de los gerentes de los hospitales en el proceso de evaluación. Llama la atención que otra de las barreras identificadas sea la necesidad de adecuados formatos para la

presentación de informes de ETS [16], teniendo en cuenta las herramientas existentes posiblemente adaptables al contexto colombiano.

Una encuesta realizada en 2017 en hospitales de la ciudad de Medellín con el objeto de caracterizar las mejores prácticas del proceso de adquisición de tecnología ilustró que dentro de los criterios que se evalúan en los hospitales principalmente se encuentran los criterios clínicos y de seguridad del paciente. También se identificó que dentro de las principales restricciones para adquirir tecnología es la evidencia médico-científica de no utilización de esta y los reportes de eventos adversos registrados en las bases de datos [67].

En general, la ETS en el país no es obligatoria en todas las instituciones de salud debido a la ausencia de su exigencia en la resolución 3100 de 2019, que consolida los procedimientos y condiciones de inscripción de los prestadores de servicios de salud y de habilitación de los servicios de salud [68]. Sin embargo, la acreditación en salud si lo plantea dentro de sus estándares de gestión de la tecnología, donde se requiere de un proceso para la planeación, la gestión y la evaluación de la tecnología [11]. De acuerdo con lo anterior, las instituciones acreditadas en salud en el país cuentan con procesos estructurados y verificados para la ETS. Particularmente la ETS es requerida bajo los estándares 132 (proceso para la planeación, la gestión y la evaluación de la tecnología) y 133 (política organizacional definida, implementada y evaluada para adquisición, incorporación, monitorización, control y reposición de la tecnología) del Manual de Acreditación en Salud Ambulatorio y Hospitalario (Resolución 5095 de 2018) [69]. Dentro de los criterios que se deben evaluar se encuentran y que se pueden apoyar desde el planteamiento de este trabajo de investigación se encuentran:

- ✓ El análisis para la incorporación de nueva tecnología, incluyendo: **evidencias de Seguridad**, disponibilidad de información sobre fabricación, confiabilidad, precios, mantenimiento y soporte, inversiones adicionales requeridas, comparaciones con tecnología similar, tiempo de vida útil, garantías, manuales de uso, representación y demás factores que contribuyan a una incorporación eficiente y efectiva.
- ✓ Las facilidades, las comodidades, la privacidad, el respeto y los demás elementos para la humanización de la atención con la tecnología disponible y **la información sobre beneficios y riesgos para los usuarios**.
- ✓ La evaluación de eficiencia, **costo-efectividad, seguridad**, impacto ambiental y demás factores de evaluación de la tecnología.
- ✓ La **evidencia de seguridad**.
- ✓ Evaluación de la confiabilidad, incluyendo el análisis de las fallas y **eventos adversos reportados** por otros compradores.
- ✓ **Las condiciones de seguridad para su uso**.
- ✓ Las alternativas disponibles.
- ✓ Evaluación de **costo beneficio, utilidad y costo efectividad de la tecnología**

Otro estudio realizado en el país, que buscó plantear un método de incorporación estratégica de tecnología biomédica para instituciones de salud, planteó el uso de la mini-ETS como metodología de evaluación cuando se trata de la incorporación de una tecnología nueva en la institución. Esta evaluación es realizada por un comité de evaluación, parte de

una solicitud y es requisito previo para el proceso de incorporación realizado por un comité de adquisición en el hospital [70].

Teniendo en cuenta la definición planteada anteriormente para las unidades de evaluación de tecnología en salud y a la poca cantidad de instituciones acreditadas en el país (47 en todo el país), podría intuirse que las unidades de ETS en los hospitales en el país son mucho menores que estas.

3.2. Procesamiento de Lenguaje Natural

El lenguaje natural se define como la capacidad del ser humano para expresarse y comunicarse a través del sonido articulado o de otros sistemas de signos [71]. El Procesamiento de Lenguaje Natural (PLN) es un campo de la computación que se enfoca en el desarrollo de modelos o algoritmos eficaces para hacer posible la comunicación entre humanos y máquinas por medio del lenguaje natural. Así, el PLN si bien tiene muchas tareas transversales a todos lenguajes, en cada uno de estos se precisa del entendimiento particular de cada uno para lograr el desarrollo de adecuados modelos que permitan la comunicación planteada entre máquinas y humanos. El PLN tiene una gran cantidad de aplicaciones tales como la traducción automática, el análisis de sentimientos, síntesis de discurso, recuperación y extracción de información, respuesta a preguntas y reconocimiento del habla [72].

El PLN analiza aspectos lingüísticos como la gramática, sintáctica y semántica ya sea en el lenguaje de manera oral o escrita. En particular para automatizar la síntesis de evidencia clínica, se debe procesar el lenguaje natural de manera escrita, dado la forma en la que se presenta la misma. En determinadas aplicaciones es útil hacer uso de algunos métodos de preprocesamiento del lenguaje para mejorar el rendimiento de los modelos generados y optimizar los recursos computacionales para tales efectos. Los métodos más comunes en relación con el preprocesamiento de texto son las siguientes:

Tokenización

División en pequeñas partes de una secuencia de texto de interés. Estas partes pueden ser oraciones, palabras, letras. Existen diferentes técnicas y librerías de tokenización, más adelante se describirá aquellas de interés con relación a *RobotReviewer* y el estado del arte de PLN [37].

Remoción de palabras vacías

Eliminación de palabras que no tienen significado como artículos, pronombres y preposiciones. Estas palabras pueden variar dependiendo de las aplicaciones y el vocabulario del corpus sobre el cual se esté trabajando [37].

Desambiguación de significado

Esta tarea se basa en identificar cuál definición o que significado de una palabra se ajusta más al contexto en el que se encuentra (cuando la palabra tiene varios significados). Por ejemplo: 'artery (arteria)', palabra que de acuerdo con la Real Academia de la lengua

Española (RAE) y con el diccionario de Cambridge (para el idioma inglés), tiene 2 significados diferentes. Este significado puede ser inferido de acuerdo con el uso que se le dé dentro de una oración [37].

Etiquetado gramatical

Consiste en clasificar las palabras de una oración de acuerdo con su respectiva denominación (sustantivo, verbo, pronombre, adjetivo, adverbio, preposición, conjunción). Este etiquetado puede influir de manera relevante en el significado de una oración [37].

Lematización y stemming

Proceso mediante el cual se genera la raíz de una palabra, en la lematización la palabra generada es el lema correspondiente, mientras que en el stemming la palabra generada no necesariamente es un lema [37].

3.3. Métodos de análisis de lenguaje

3.3.1. Métodos clásicos – conteo de palabras y análisis estadísticos

Para llevar a cabo el desarrollo de algoritmos para el entendimiento del lenguaje a través del uso de máquinas, es pertinente hacer una interpretación de este por medio de una representación numérica. A lo largo de los años, esta representación ha tenido variaciones y en los últimos años se ha incrementado el interés en que esta representación contenga la mayor información posible en relación con la sintáctica y semántica del lenguaje. A continuación, se presentan algunas de las técnicas existentes para hacer la representación mencionada.

Bag of Words – BoW (bolsa de palabras)

Este método es el más simple de todos para representar numéricamente un texto, sin embargo, es la base de muchos otros. Por medio de este no se analiza el significado o la relación entre las palabras, sino que solamente se identifica su ocurrencia dentro de un texto. A partir del vocabulario que compone el corpus de interés, se define un orden aleatorio y a partir de este orden, cada uno de los elementos del corpus es dividido en palabras (tokenizado) y a cada una de ellas se le asigna un número que corresponde con su ocurrencia en el elemento analizado, obteniéndose un vector para cada elemento con un tamaño igual a las dimensiones del vocabulario. Este método puede generar vectores dispersos y a medida que el vocabulario incrementa, el tamaño de los mismos también lo hace, haciendo que el procesamiento de estos vectores dispersos genere un costo computacional alto cuando se trata de vocabularios muy amplios.

Term Frequency – Inverse Document Frequency (frecuencia de palabras – frecuencia inversa de documento)

En este método, a cada palabra de un texto se le asigna un valor $W_{x,y}$ que se obtiene a través de la Ecuación 1.

$$W_{x,y} = tf_{x,y} * \log \left(\frac{N}{df_x} \right)$$

Ecuación 1. Frecuencia de términos – Frecuencia Inversa de Documento

Donde,

$tf_{x,y}$ = frecuencia de x en y

df_x = número de documentos que contienen x

N = número total de documentos

De esta manera N puede estar representado por cantidad de documentos, cantidad de frases, párrafos o ítems que compone el corpus de análisis. Algunas características de este método son que para palabras raras o de baja ocurrencia se obtiene valores de $W_{x,y}$ altos y para las stopwords o palabras de parada en el texto, se obtiene valores pequeños, como se ilustra en la Figura 4 Este método es usado para identificar palabras importantes en un texto [73].



Figura 4. Representación de palabras a través de TF-IDF

Distancia coseno

Este método es utilizado para comparar documentos (textos, oraciones). En la recuperación de información esta técnica puede ranquear documentos con relación a un vector de palabras clave. Los documentos son comparados con vectores de frecuencia (los términos pueden ser n-gramas de una, dos o más palabras). La distancia coseno está dada por la Ecuación 2:

$$sim(x, y) = \frac{x \cdot y}{||x|| ||y||}$$

Ecuación 2. Distancia coseno

Donde, x, y son los vectores a comparar,

$\|x\|$ es la norma euclidiana del vector $x = (x_1, x_2, \dots, x_p)$, definida como:

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2}$$

$\text{sim}(x, y)$ es el coseno del ángulo entre los vectores x y y .

Un valor de 0 indica un ángulo 90° entre los vectores, lo que significa disimilitud, y un valor de 1 indica un ángulo de 0° (o vectores paralelos) y la mayor similitud posible entre los vectores.

3.3.2. Métodos clásicos – análisis sintáctico

Colocaciones y n-gramas

El término *colocaciones* hace referencia a secuencias de palabras que ocurren juntas de manera frecuente, una característica particular es que no es posible sustituir una palabra por otra de similar significado porque la expresión pierde algo de sentido. Ejemplos de esto en el dominio médico son: dispositivo médico y presión arterial. Este tipo de análisis es tan relevante que en algunos casos aportan mayor información palabras que ocurren de manera conjunta en un texto que las palabras analizadas de manera aislada. Los N-gramas son la partición de un texto en frases con una cantidad N de palabras conjuntas, así, un bigrama está representado por dos palabras conjuntas, un trigramas por tres palabras, etc. Las colocaciones son bigramas frecuentes, que generalmente aparecen más frecuentemente que otras palabras individuales [37].

Extracción de entidades

También conocido como Reconocimiento de Entidades Nombradas, el cual busca localizarlas y clasificarlas en categorías predefinidas, como personas, organizaciones, lugares, expresiones de tiempo y cantidades.

3.3.3. Métodos de word embeddings (incrustación de palabras)

Estos métodos buscan obtener una representación de las palabras en la cual aquellas que tengan un significado similar puedan tener una representación numérica semejante, así mismo, buscan que esta representación contenga información acerca de la relación que existe entre las palabras. Actualmente existen diversos métodos de *word embeddings* y variaciones de estos, dentro de los más comunes se encuentran Word2Vec y GloVe (Global Vectors for Word representation). En el año 2019, *BERT*, un modelo pre entrenado basado en transformadores bidireccionales para el entendimiento del lenguaje, se convirtió en el estado del arte principal del PLN. Este modelo es de particular interés porque una de sus derivaciones (*SciBERT*) es usada por *RobotReviewer*.

Word2Vec

Es un modelo de representación distribuida de palabras, el objetivo de las representaciones distribuidas del lenguaje es encontrar una función f que sea capaz de

convertir cada palabra en un vector de forma que la proximidad entre vectores sea equivalente a la proximidad semántica de las palabras. Está basado en redes neuronales y tiene dos tipos de arquitecturas: *continuous skip-gram* y *continuous bag of words* (CBOW). En la arquitectura skip-gram, el modelo predice las palabras contexto a partir de la palabra central. En la arquitectura CBOW, el modelo predice la palabra central a partir de una ventana de palabras contexto (la predicción es independiente del orden de las palabras contexto) [74].

Glove (global vectors for word representation)

Es un modelo de representación distribuida de palabras, basado en aprendizaje no supervisado para obtener vectores que representan las palabras. Es entrenado con base en la matriz de co-ocurrencia de las palabras en un corpus determinado [75].

BERT

Este modelo pre entrenado utiliza un método de *Word embeddings* denominado *wordPiece embeddings*, el cual se basa en una transformación de las palabras en vectores predeterminados de dimensión fija 768. El vocabulario de este método está compuesto por 30,522 tokens, dentro de los cuales se encuentran palabras, números y símbolos [38].

3.4. Aprendizaje automático - algoritmos de clasificación

3.4.1. Máquinas de Soporte Vectorial – SVM por sus siglas en inglés

Para tareas de clasificación, las SVM buscan un hiperplano que separa diferentes clases, este hiperplano es elegido con el objetivo de maximizar la distancia entre la frontera de decisión y todas las muestras. El hiperplano es un espacio de dimensiones $n - 1$, donde n es el número de características que representan cada clase. En la mayoría de los casos, las clases no son totalmente separables, por lo cual se realiza una penalización a las muestras mal clasificadas. El principal objetivo de la SVM es manejar el balance entre dicha penalización y el margen de ajuste de cada clase a los datos. Un ejemplo de la clasificación a través de las SVM es ilustrado en la Figura 5, para un espacio de características de dos dimensiones y asumiendo un problema linealmente separable.

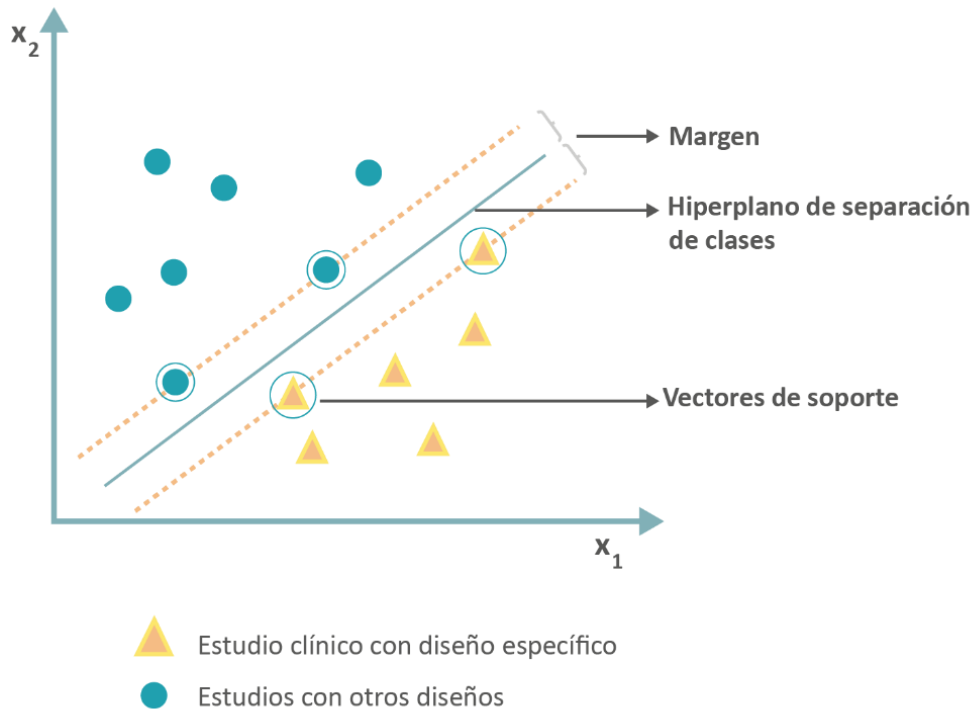


Figura 5. Representación de SVM para un problema de clasificación de 2 clases

3.4.2. Árboles de decisión - DT por sus siglas en inglés

Los árboles de decisión pertenecen al aprendizaje supervisado y son usados en tareas de regresión y clasificación, siendo esta última donde más se aplican. Como su nombre lo indica usa un modelo de decisiones en forma de árbol. En este tipo de algoritmos existen dos nodos, nodo de decisión y nodo hoja, en el primero se toma una decisión y existen múltiples posibilidades o *ramas* y en el segundo se tiene información sobre la salida o la decisión tomada y no existen ramas adicionales. Las reglas para decidir en los nodos de decisión son obtenidas a partir de las características de los datos a procesar. Esta arquitectura se puede observar en la Figura 6. El número de observaciones mínimas para efectuar una división, el número de observaciones que deben tener todos los nodos hoja y la profundidad del árbol son hiperparámetros a optimizar en este tipo de algoritmos para obtener un adecuado modelo de predicción o clasificación.

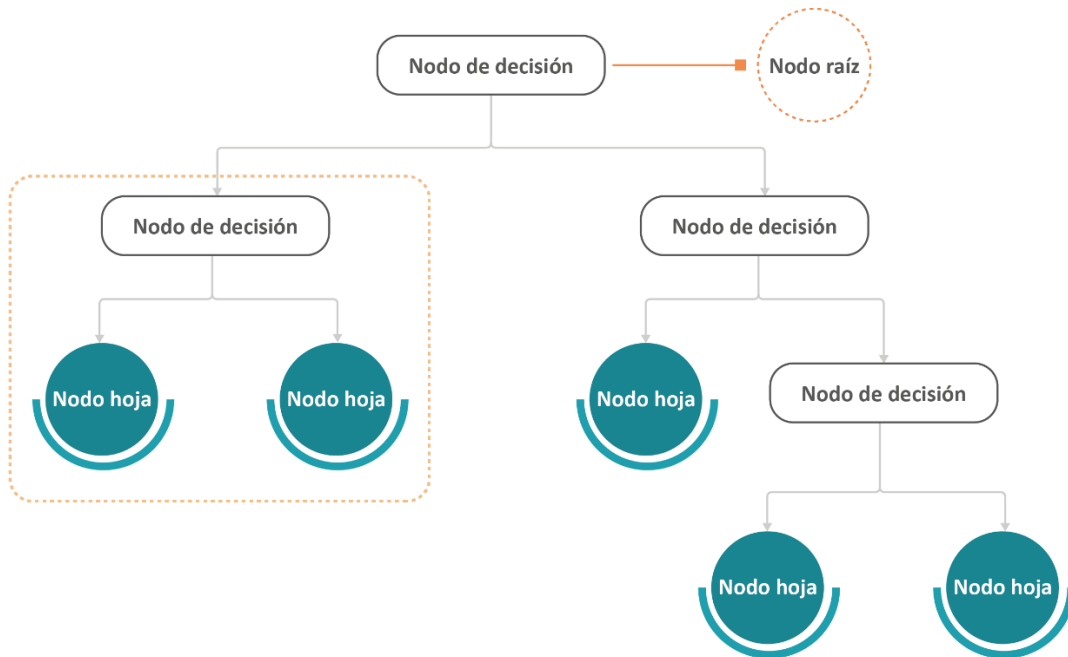


Figura 6. Representación árboles de decisión

3.5. Aprendizaje profundo

Otras técnicas implementadas tanto para modelos de clasificación y regresión automáticos son las redes neuronales. Una red neuronal está compuesta por capas y estas a su vez están integradas por nodos o neuronas, como se ilustra en la Figura 7 (simulando las conexiones cerebrales humanas). Cada neurona está asociada con las demás y posee un peso y un sesgo (o umbral), si la salida de un nodo en particular supera el valor de dicho umbral el nodo es activado y la información se envía a dicha capa, de otra manera la información es transmitida a la siguiente capa.

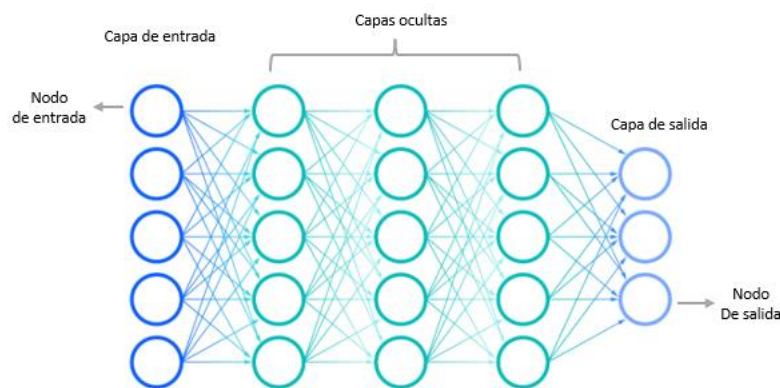


Figura 7. Esquema redes neuronales

Como en cualquiera de estos algoritmos, las redes neuronales se basan en información de la cuál deben “aprender” e incrementar su exactitud o la métrica de desempeño que se quiera optimizar. Tienen una amplia aplicación en tareas de reconocimiento de imágenes y PLN [76].

En una red neuronal simple cada nodo o neurona tiene su propia función de activación, para un modelo de regresión lineal con una función de activación umbral escalón, cada neurona tendrá una información de entrada, un peso, un sesgo y una salida, dicho modelo se puede sintetizar como se ilustra en la Ecuación 3 y la Ecuación 4:

$$\sum_{i=1}^m w_i x_i + b = w_1 x_1 + w_2 x_2 + \dots + w_m x_m + b$$

Donde,

$w_i =$ peso de cada neurona

$x_i =$ información de entrada o vector de datos

$b =$ sesgo o bias

Ecuación 3. Modelo de regresión por nodo en redes neuronales simples

Cada una de estas neuronas tendrá salida de acuerdo a su función de activación, para este ejemplo en particular se tiene que:

La salida es 1 cuando:

$$\sum w_1 x_1 + b \geq 0$$

Y 0, cuando:

$$\sum w_1 x_1 + b < 0$$

Ecuación 4. Salida función de activación escalón

Existe un tipo de redes neuronales, denominadas redes recurrentes, las cuales poseen dos tipos de dato de entrada: el dato actual y el dato de la activación anterior y sus salidas son la predicción actual y la salida actual. Generalmente, son implementadas en tareas en las cuales la información está planteada como una secuencia de datos, es decir, el dato siguiente depende del o de los anteriores. En estas redes son capaces de “memorizar” los parámetros de entrada inmediatos y los anteriores. Una aplicación particular son las denominadas LSTM (Long-Short Term Memory por sus siglas en inglés), estas redes tienen como característica que son capaces de “recordar” datos inmediatamente anteriores o con una ocurrencia más antigua. En la Figura 8 se ilustra una neurona del tipo red neuronal LSTM, donde se puede observar que la salida depende de: el estado de larga memoria (estado de neurona), la salida anterior (estado oculto) y el dato de entrada actual. Las LSTM

usan una serie de compuertas que controlan el flujo de la información que para a través de ella es almacenada u olvidada por la red. Generalmente existen tres tipos de compuertas: de olvido, de entrada y de salida. Cada una puede interpretarse como un filtro y tiene su propia red neuronal.

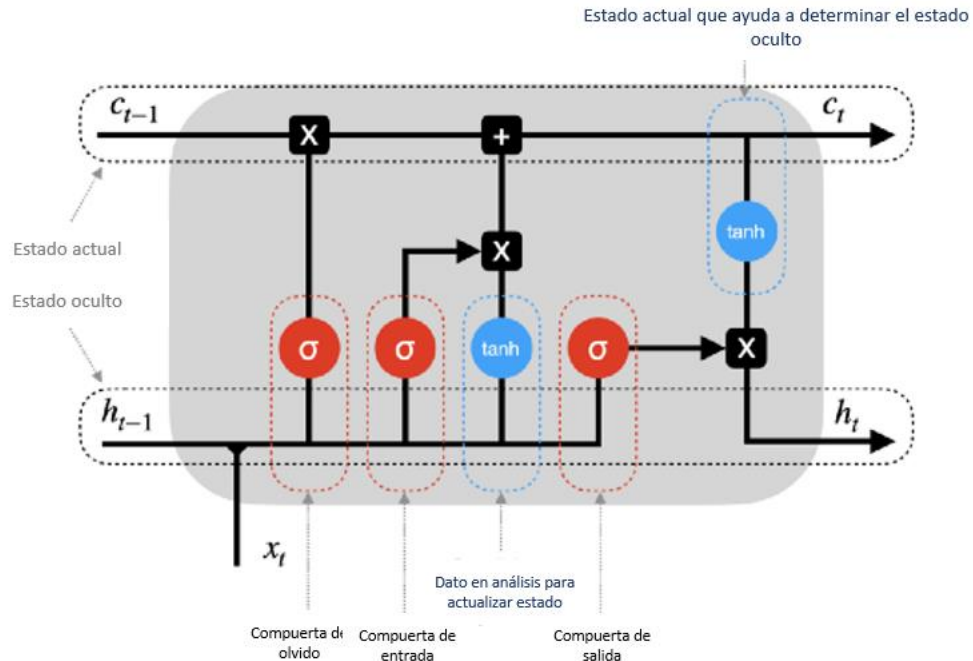


Figura 8. Nodo o neurona en redes LSTM

Cada una de las compuertas anteriores se comporta de la siguiente manera:

- ✓ **Compuerta de olvido:** Usada en el primer paso del proceso, en esta etapa se decide qué información del estado de la neurona es útil dado el nuevo dato de entrada y el estado oculto anterior. Aquellos datos que se consideran irrelevantes por la compuerta de olvido son multiplicados por un número muy cercano a cero y tendrá menos influencia en los pasos siguientes. Esta compuerta decide qué partes de la memoria a largo plazo deben ser olvidados de acuerdo con el estado oculto anterior y el nuevo dato de la secuencia.
- ✓ **Compuerta de entrada:** En el siguiente paso se involucran la red de memoria nueva y la compuerta de entrada, en donde se determina qué de la nueva información debería adicionarse a la red de memoria a largo plazo (estado de la neurona), teniendo en cuenta el estado oculto previo y el dato de entrada. La red de memoria nueva y la compuerta de entrada son redes en sí mismas y ambas toman la misma entrada: el estado oculto y el nuevo dato. Es decir, las mismas entradas de la compuerta de olvido.
La red de memoria nueva es una red neuronal activada a través de la función tangente hiperbólica (\tanh) que ha aprendido como combinar el estado oculto anterior y el nuevo dato para general un nuevo vector de memoria actualizado.

Este vector contiene información del nuevo dato de estrada teniendo el contexto de del estado oculto anterior, este vector contiene la información para actualizar cada componente del estado de la neurona dado el nuevo dato. En este caso se ejemplifica la función \tanh porque sus valores van en el rango $[-1, 1]$, es decir, pueden ser negativos, esto es implementado si se desea reducir el impacto de un componente en el estado de la célula.

Para identificar si es necesario recordar el vector de memoria nueva es necesaria la interacción con la compuerta de entrada. Esta compuerta es una red activada con una función de activación sigmoide la cual actúa como filtro e identifica qué componentes del vector de memoria nuevo deben ser recordados. Esta red tendrá una salida de valores entre 0 y 1 (debido a la función de activación sigmoide). Similar a la compuerta de olvido, una salida cercana a cero indica que no es necesario recordar o actualizar el elemento del estado de la neurona.

Las salidas de las compuertas anteriores son multiplicadas, el vector resultante es agregado al estado actual de la neurona, convirtiéndose en la memoria a largo plazo de la red.

- ✓ *Compuerta de salida:* En esta compuerta se decide el nuevo estado oculto, para esto se requieren tres cosas: el nuevo estado de la neurona, el estado oculto previo y el dato de entrada. La función de activación es sigmoide, de tal manera que se puede interpretar como la anterior, es decir, un filtro, el cual es aplicado al nuevo estado de la neurona. Esto asegura que solo la información necesaria es almacenada en el nuevo estado oculto. Antes de aplicar este filtro, la información atraviesa una función \tanh llevando los valores al rango entre -1 y 1.

El resumen el paso a paso del proceso en la LSTM es el siguiente: aplicar la función \tanh al estado actual de la neurona para obtener el estado suavizado de la neurona; pasar el estado oculto previo y el dato de entrada a través de la función de sigmoide para obtener el vector filtro; aplicar este vector filtro al estado suavizado de la neurona y finalmente obtener el nuevo estado oculto [77].

3.6. Métricas de desempeño

Las métricas de desempeño comúnmente usadas para problemas de clasificación en el aprendizaje automático se basan en los valores predichos comparados con los valores reales de cada una de las clases, de esta comparación nace la matriz de confusión y a partir de ella se derivan diferentes métricas de desempeño.

3.6.1. Matriz de confusión

La matriz de confusión se construye a partir de comparar los valores predichos y reales de cada una de las clases que componen la tarea de clasificación a realizar. De esta manera,

se identifican cuantas de las clases fueron clasificadas correctamente y cuantas incorrectamente. En un caso ideal para una clasificación de dos clases, los falsos positivos y falsos negativos deben ser cero. En la Figura 9 se ilustra como se construye una matriz de confusión bi-clase.



Figura 9. Matriz de confusión

3.6.2. Exactitud

Esta métrica pretende indicar que tan cerca se está del resultado verdadero. Se expresa por la cantidad de valores predichos correctamente (verdaderos positivos más verdaderos negativos) sobre el total de los casos clasificados, como se ilustra en la Ecuación 5.

$$\text{Exactitud} = \frac{VP + VN}{VP + FP + FN + VN}$$

Ecuación 5. Exactitud

Cuando existe un desbalance de clases esta métrica no es útil, dado que podría estar indicando un valor muy alto, pero no reflejar realmente un buen desempeño del algoritmo clasificador.

3.6.3. Precisión

Esta métrica identifica los elementos clasificados correctamente como positivos del total de elementos clasificados como positivos.

$$\text{Precisión} = \frac{VP}{VP + FP}$$

Ecuación 6. Precisión

3.6.4. Recall (Sensitividad)

Esta métrica identifica los elementos clasificados correctamente de todos los valores positivos de la clase. Es comúnmente usada cuando se busca precisamente enfocarse en la correcta clasificación de la clase positiva o de interés. Se define como los valores positivos clasificados correctamente sobre el total de muestras positivas.

$$\text{Recall (Sensitivity)} = \frac{VP}{VP + FN}$$

Ecuación 7. Sensitividad

3.6.5. Especificidad

Esta Métrica identifica los elementos que fueron clasificados correctamente de los valores negativos de la clase. Es comúnmente usada cuando se busca precisamente enfocarse en la correcta clasificación de la clase negativa. Se define como los valores negativos clasificados correctamente sobre el total de muestras negativas.

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

Ecuación 8. Especificidad

3.6.6. Área bajo la curva ROC

La curva ROC, grafica la tasa de verdaderos positivos (Sensitividad) comparada con la tasa de falsos positivos (1- Especificidad). En clasificación de datos, un modelo perfecto tendrá un área bajo la curva de 1, mientras que un valor de 0.5 implica una capacidad de predicción nula o aleatoria. En la Figura 10 se ilustra gráficamente esta métrica.

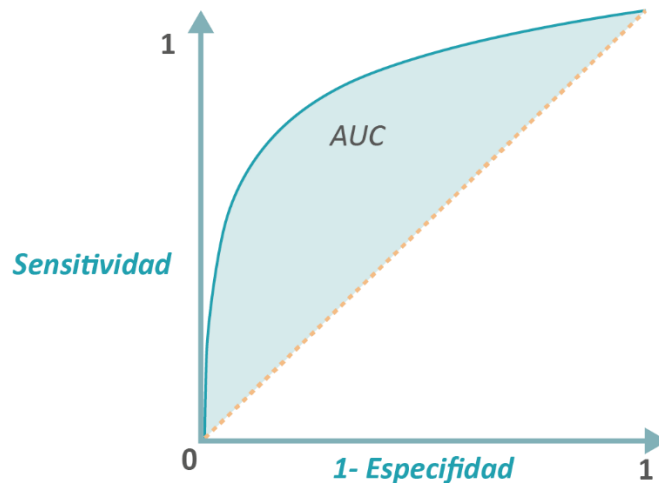


Figura 10. Área bajo la curva ROC

3.7. Análisis RobotReviewer

RobotReviewer es una herramienta entrenada para apoyar la predicción de sesgo en Ensayos Clínicos Aleatorizados (ECA), además de identificar automáticamente este tipo de estudio. Para su entrenamiento se usaron revisiones sistemáticas del instituto Cochrane, el cual está integrado por una red de profesionales (de más de 190 países) dedicados a la síntesis de evidencia clínica para guiar la toma de decisiones en salud, es un referente mundial en este ámbito [35].

3.7.1. Robot reviewer

3.7.1.1. Método de aprendizaje, corpus, vocabulario y espacio de características

Este sistema fue entrenado con 12,808 ensayos clínicos de la base de datos de revisiones sistemáticas de Cochrane (CDSR por sus siglas en inglés). Para el levantamiento y etiquetado

del corpus se usa supervisión a distancia. A partir de las revisiones sistemáticas de Cochrane, se identificaron ensayos clínicos incluidos en las mismas, aquellos que aparecieran en 2 o más búsquedas específicas se incluyeron en el corpus. Los ECA incluidos se etiquetan a partir de los juicios de los autores de las revisiones sistemáticas y las sentencias que soportan esos juicios fueron extraídas de los ensayos si en las revisiones se usaban exactamente las palabras del ECA para predecir el riesgo de sesgo [19].

El sistema se basa en dos niveles de representación numérica del texto de acuerdo con las dos tareas específicas que tiene como objeto. La primera es la clasificación de los ECA en dos clases: riesgo de sesgo ‘bajo’ y riesgo de sesgo ‘alto/incierto’. La segunda, es la clasificación de sentencias como ‘relevante’ y ‘no relevante’ en la justificación para la clasificación del riesgo de sesgo. Los dos niveles para la representación de los ECA en el espacio de características son el *nivel documento* y el *nivel sentencias u oraciones* [19].

Inicialmente, el texto se pre-procesa haciendo remoción de *stopwords* y convirtiendo a minúsculas. El vocabulario del nivel documento son los uni y bigramas que aparecen en al menos dos ECA y el vocabulario del nivel sentencias son los uni y bigramas que aparecen en al menos dos sentencias que justifican el riesgo de sesgo. Cada instancia es representada con el método de *BagOfWords* [19].

En la Figura 11 se representa gráficamente el espacio de características y el enfoque de aprendizaje multi-tarea que implementa *Robot reviewer*.

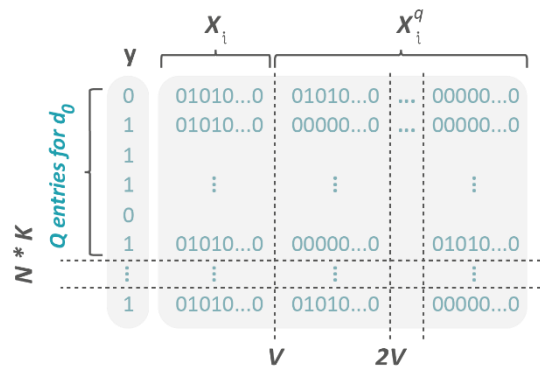


Figura 11. Representación de aprendizaje multi-tarea implementado en robot reviewer

Donde:

y = vector objetivo con etiquetas de riesgo de sesgo para cada documento

x_i = vector de características nivel documento y nivel sentencia

k = Dominios de sesgo

d = documento o artículo

V = longitud de vocabulario

N = cantidad total de documentos o artículos

3.7.1.2. Método de clasificación

El sistema usa la variante *multi-taks o multi-tarea* del método de clasificación *soft-margin support vector machine*. Para los modelos de documento y sentencia, se usa un modelo de clasificación definido por un vector w , tal que $y = L(w * x)$. L mapea el puntaje en una categoría de 0 cuando $w * x < 0$ y 1 cuando $w * x \geq 0$. Se emplea la función de pérdida Hinge y una regularización Ridge o L2 para ajustar el modelo, el cual es una máquina de soporte vectorial (SVM) de margen suave con kernel lineal [19]. La función de pérdida Hinge parte de que el puntaje de la clase correcta es mayor que la suma de los puntajes de las clases incorrectas por un margen de seguridad generalmente de uno. Dicha función de pérdida se ilustra en la Ecuación 9.

$$hinge = \sum_{j \neq y_i} \max(0, S_j - S_{y_i} + 1)$$

Ecuación 9. Función de pérdida Hinge

Para evitar el sobre entrenamiento del modelo, minimizar varianza y ajuste de la función de pérdida a los datos, se implementan varias técnicas de regularización, en este caso se implementó la regularización *Ridge* o *L2*. Esta regularización consiste en reducir los coeficientes del modelo sin anular ninguno de ellos, minimizando la suma de los cuadrados del error (SSE por sus siglas en inglés), como se ilustra en la Ecuación 10 [78].

$$SSE = \sum_{i=1}^k \left(\sum_{j=i}^n (x_{ij} * B_j + B_0) - y_i \right)^2 + a \sum_{j=1}^n B_j^2$$

Donde,

$B_j =$ coeficientes del modelo

$x_{ij} =$ datos de entrada

$y_i =$ salida del modelo

Ecuación 10. Regularización L2

Teniendo en cuenta lo expuesto anteriormente, el modelo de clasificación objetivo se ilustra en la Ecuación 11.

$$w = \arg \min_w \sum_i \{hinge - loss(w * x_i, y_i)\} + \alpha ||w||^2$$

Ecuación 11. vector w de clasificación

El modelo fue entrenado (la estimación de w) minimizando el objetivo a través del algoritmo gradiente descendente estocástico. El modelo incluye un hiper parámetro α , el cual balancea entre la regularización y la pérdida empírica. Este hiper parámetro se ajustó por medio de validación cruzada en un rango de valores 10^{-4} y 10^{-1} , equidistantes en el espacio logarítmico [19].

Debido a que usar el texto completo y la interacción de características produce un espacio de características muy grande, se usó la técnica *hashing* para mantener el manejo del modelo. Este método mapea cadenas de texto en vectores a través de una función *hashing*, lo cual ha mostrado, funcionar bien para múltiples tareas de clasificación de texto de gran tamaño [19]. Específicamente la técnica Hashing es relevante para elevar el alcance de aplicación del aprendizaje de máquina, particularmente, es usada para reducir la dimensión de datos, conservando la distancia Euclidiana. Lo anterior se logra a través de una proyección de una matriz dispersa aleatoria $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, con m menor que n , [79].

3.7.1.3. Principales resultados

En la Tabla 2 se presentan los resultados obtenidos para predecir el riesgo de sesgo en un documento para cada dominio específico. Donde se tiene que, Baseline es la exactitud obtenida al etiquetar los documentos con mayoría de clase en un dominio. *RobotReviewer* tiene entrenados tres modelos diferentes, el primero se entrenó con BoW separada para cada dominio, el segundo es el modelo multi-tarea explicado anteriormente, en que se comparte información de todos los dominios para la representación del BoW y se adiciona la relevancia de las sentencias como característica. El tercer modelo se entrena igual que el segundo, excluyendo los dominios de sesgo 5 y 6. El ítem Cochrane, estima la exactitud humana al evaluar el riesgo de sesgo al analizar el mismo ensayo clínico usado en otra revisión sistemática.

Tabla 2. Resultados para la predicción de sesgo de un documento completo

Dominio	Ensayos (n)	Referencia	Modelo 1	Modelo 2	Modelo 3	Cochrane	Valor P (Modelo 2 vs Cochrane)
Riesgo General	6610	56.4%	69.3%	71.0%	-	78.3%	P<0.001
1. Generación de secuencia aleatoria	1225	59.3%	72.5%	73.9%	75.8%	84.8%	P<0.001
2. Ocultamiento de asignación	2089	53.7%	72.4%	74.0%	73.3%	80.0%	P<0.001
3. Cegamiento de participantes y personal	1051	50.4%	72.6%	73.0%	73.7%	78.1%	P=0.003
4. Cegamiento para evaluación de resultados	250	57.7%	64.0%	61.5%	67.4%	83.2%	P<0.001
5. Reporte incompleto de resultados	1149	60.9%	63.9%	65.1%	-	71.3%	P<0.001
6. Reporte selectivo	846	59.9%	61.8%	67.6%	-	73.0%	P=0.010

Se puede observar que, aunque las predicciones de otras revisiones sistemáticas de Cochrane tuvieron una exactitud mayor en la predicción de sesgo para cada dominio, esta exactitud no es significativa estadísticamente, pues el valor P obtenido del modelo 2 en comparación con Cochrane para cada dominio es inferior a 0.05. Sin embargo, los autores declaran que esta diferencia estadística debe ser evaluada con una base de datos más amplia.

En la Tabla 3, se ilustra la proporción de estudios para los cuales se identificaron sentencias relevantes para la evaluación de sesgo de cada dominio usando cuatro métodos diferentes. Baseline hace referencia a elegir una sentencia aleatoria del documento e identificar su relevancia para predecir sesgo, top1 indica la elección de la frase más relevante, top3 indica la elección de las tres sentencias más relevantes y Cochrane indica las sentencias identificadas como relevantes para un dominio en revisiones sistemáticas de Cochrane.

Tabla 3. Proporción de estudios para los cuales se identificaron sentencias de texto relevantes para cada dominio de sesgo.

Domain	Trials (n)	baseline	top1	top3	cochrane	top1 v cochrane	top3 v cochrane
Overall	378	0.5%	45.0%	60.4%	56.5%	-11.6% (-18.5% to -4.4%); P<0.001	+3.9% (-3.2% to +10.9%); P=0.141
1. Random sequence generation	81	0.0%	55.6%	65.4%	60.5%		
2. Allocation concealment	75	0.0%	44.0%	60.0%	60.0%		
3. Blinding of participants and personnel	76	0.0%	55.3%	72.4%	68.4%		
4. Blinding of outcome assessment	56	0.0%	39.3%	62.5%	57.1%		
5. Incomplete reporting of outcomes	67	3.0%	40.9%	57.6%	50.8%		
6. Selective reporting	23	0.0%	0.0%	4.6%	4.6%		

De la Tabla 3 se puede observar que el mayor porcentaje de sentencias relevantes identificadas para predecir el riesgo de sesgo en un dominio específico se obtuvo al extraer las tres sentencias más importantes del documento. Inclusive se logró un mayor porcentaje de identificación de sentencias que con las extraídas por humanos (Cochrane).

3.7.1.4. Actualizaciones

Teniendo en cuenta los rápidos avances y las nuevas investigaciones en el modelado del lenguaje, la herramienta *RobotReviewer* actualmente hace uso del modelo pre entrenado *SciBERT* [40], el cual es un modelo pre entrenado para el entendimiento de lenguaje específico en el dominio biomédico, que se basa en el modelo pre entrenado *BERT*. Ambos modelos se describen a continuación.

3.7.2. BERT

BERT es un modelo pre entrenado para el entendimiento del lenguaje basado en transformadores bidireccionales profundos. Este modelo pre entrenado utiliza un método de *Word embeddings* denominado *wordPiece embeddings*, el cual se basa en una transformación de las palabras en vectores predeterminados de dimensión fija de 768. El

vocabulario de este método está compuesto por 30,522 tokens, dentro de los cuales se encuentran palabras, números, símbolos, entre otros [38].

Una de las características que posicionaron a *BERT* como el estado del arte en tareas del PNL es que todas las oraciones que ingresan al sistema tienen tres representaciones que se concatenan para formar matrices de dimensión 768xN, donde N es el número de tokens de una sentencia. Estas tres representaciones se dan en tres niveles y se describen a continuación.

3.7.2.1. Bert embeddings.

BERT maneja una representación de las sentencias que se basa en tres tipos de ‘embeddings’ en los niveles de token, sentencia y posición. Para cada token existe una representación en un arreglo que mapea cada token en un vector fijo de dimensión 768, este nivel se basa en *WordPiece Embeddings*. El ‘embedding’ de las sentencias se realiza con el fin de identificar primera y segunda posición, específicamente para lograr las tareas de parafraseo, respuesta a preguntas, similitud, etc. En este nivel cada token de la primera sentencia se representa con un ‘0’ y cada token de la segunda sentencia se representa con un ‘1’. Finalmente, en el embedding de posición a cada token se asigna un número correspondiente a la posición en la que aparece en las sentencias. Esta descripción se puede observar en la Figura 12, la cual es un ejemplo de la representación del ‘embedding’ para la sentencia: “A total of 25 women affected by clinical stress urinary incontinence (SUI) were enrolled. After undergoing urodynamics assessment, pelvic organ prolapse quantification, urine culture, Q-tip test ...”.

A | total | of | 25 | women | affected | by | clinical | stress | urinary | incontinence | (SUI) | were | enrolled. | After | undergoing | (...) | Q-tip | test

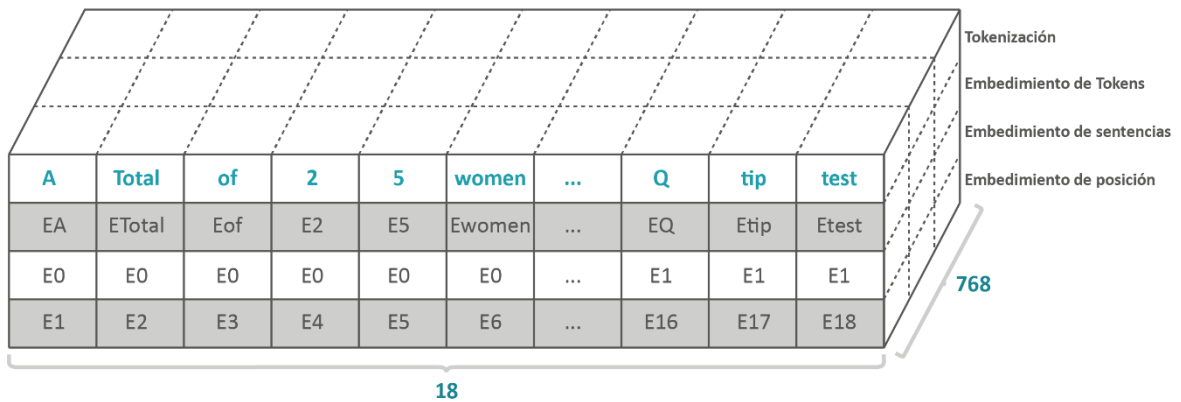


Figura 12. Ejemplo de BERT embeddings

3.7.2.2. Modelo de atención

Adicionalmente a su método de embeddings, *BERT* transforma las entradas por medio de un modelo de atención denominado *multi-head self attention*. Para describir este modelo de atención es necesario definir inicialmente el modelo *self attention* o

autoatención para finalmente llegar a la definición de *multi-head self attention* o autoatención multi-cabeza.

Self attention se denomina al mecanismo por medio del cual, en una sentencia dada, es posible que cada token de la sentencia 'interactúe' con los demás e identifique a cuál debería prestar más atención. La salida de este mecanismo son los resultados de esta interacción y los puntajes de atención que se asignan a cada token en relación con el interés. Para una sentencia de entrada de vectores que representan N tokens, se obtendrá una salida de N vectores.

Cada entrada (en este caso, vector de representación de un token), es representada por tres vectores denominados, *key*, *value* y *query* (clave, valor y requerimiento). Esta representación se obtiene de multiplicar cada entrada con un conjunto de 'pesos' para *query*, *key* y *value*, estos conjuntos de pesos son generalmente números pequeños y son inicializados de manera aleatoria antes de entrenar un modelo. Posterior a obtener las tres representaciones para cada entrada, la interacción entre estas hasta obtener una salida para cada entrada, de la siguiente manera:

- Obtención de los puntajes de atención: producto punto entre los vectores *query* de la entrada a analizar y cada vector *key*.
- Obtención de *softmax* para cada puntaje del paso anterior
- Multiplicación de puntajes de atención con los vectores *value* para obtener valores ponderados
- Suma de valores ponderados para obtener la salida correspondiente a la entrada
- Se repite el proceso para cada entrada

Este proceso se ilustra gráficamente en la Figura 13, donde para un vector de entrada: {1,0,1,0} se obtiene una salida: {2,7,1.5}.

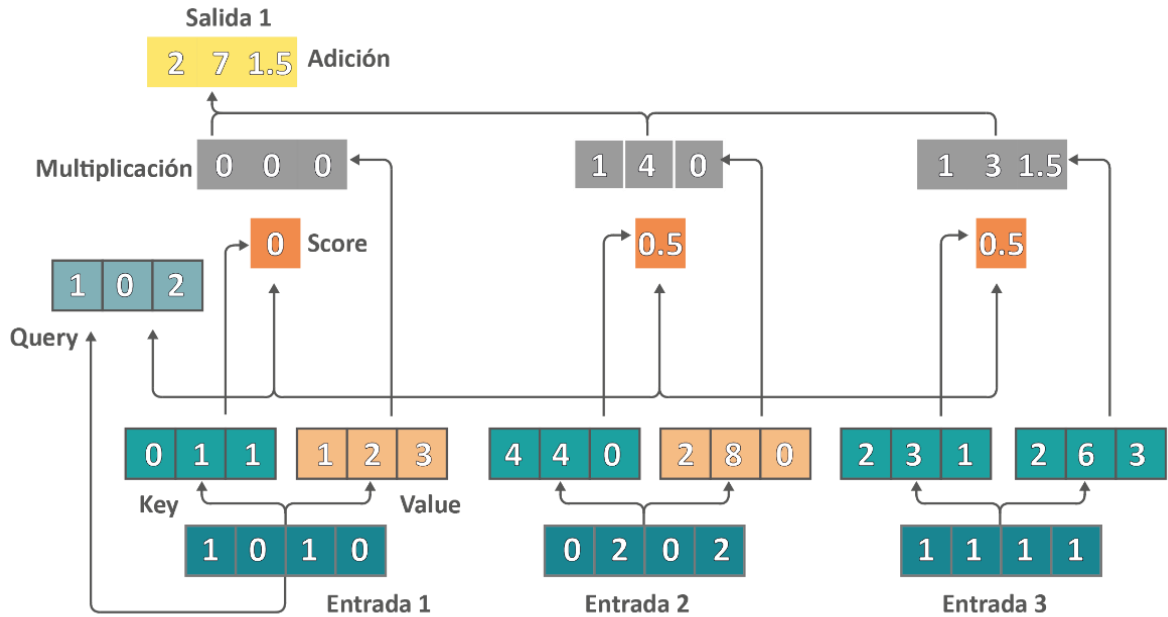


Figura 13. Ejemplo gráfico mecanismo de atención

El multi-head self attention hace referencia a cada conjunto de valores *key*, *query* y *value*, se realiza una proyección lineal en las dimensiones D_k , D_q , D_v , respectivamente. En cada proyección se ejecuta el mecanismo *self attention* en paralelo, donde las salidas resultantes son concatenadas y proyectadas nuevamente para obtener los valores finales correspondientes a cada par entrada-salida.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o$$

Ecuación 12. Mecanismo de atención

$$\text{donde, } head = Attention(Q * W^q, K * W^k, V * W^v)$$

Donde las proyecciones son matrices de parámetros $w_i^Q \in \mathbb{R}^{d_{model} * dk}$, $w_i^K \in \mathbb{R}^{d_{model} * dk}$, $w_i^V \in \mathbb{R}^{d_{model} * dv}$ y $w_i^O \in \mathbb{R}^{hdv * d_{model}}$

En este modelo se usó $h = 8$, y $dk = dv = \frac{d_{model}}{h} = 64$.

Finalmente, *BERT* posee una arquitectura multi-capa (N=6) de transformadores bidireccionales, la arquitectura de cada capa se ilustra en la Figura 14.

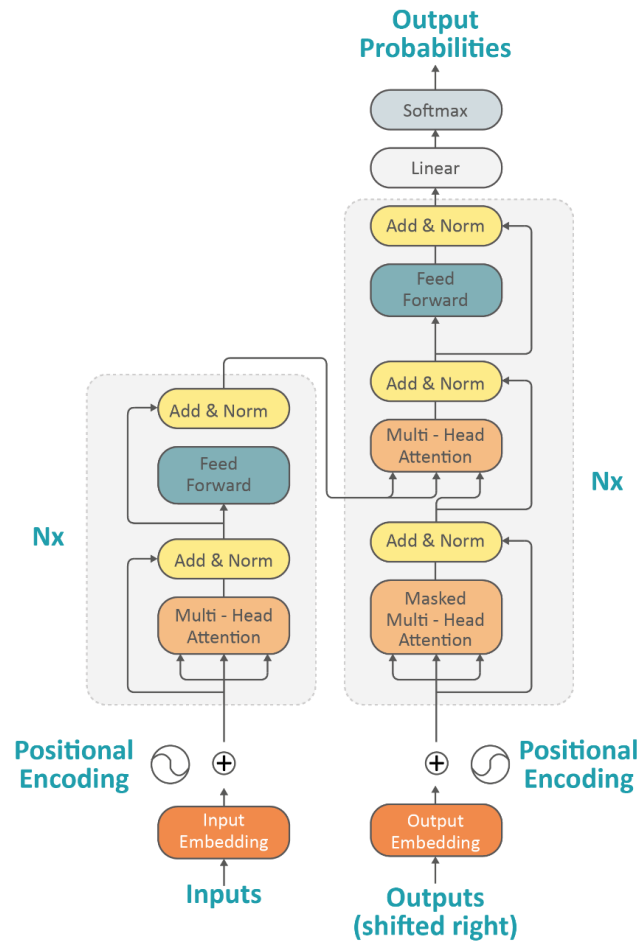


Figura 14. Capa de transformador, adaptado de [39]

El entrenamiento de *BERT* se dividió en dos partes, un preentrenamiento basado en dos tareas específicas: predicción de la siguiente frase y la predicción de palabras enmascaradas, y un ajuste que se realizó por medio de varias tareas contempladas en *GLUE*, *SQuAD* v1.1 y 2.0 y *SWAG* [38]. El preentrenamiento se realizó con el corpus de *BookCorpus* (con 800M palabras) y con artículos de wikipedia (con 2,500M palabras), ignorando listas, tablas y encabezados.

3.7.2.3. Tareas de preentrenamiento y entrenamiento.

GLUE (General Language understanding evaluation) es una colección de recursos para el entrenamiento, evaluación y análisis de los sistemas de PLN. Para la evaluación de un sistema de PLN, *GLUE* tiene nueve tareas de entendimiento del lenguaje basadas en datasets existentes y elegidos con el objeto de cubrir un amplio rango de actividades, niveles de dificultad y tipos de textos [80]. La métrica de evaluación de las tareas puede variar entre ellas, el dataset usado, su descripción, la métrica de evaluación y algunos ejemplos se ilustran en la Tabla 4.

Tabla 4. Tareas para la evaluación de sistemas de PLN contempladas GLUE [80]

Base de datos	Descripción	Ejemplo	Métrica
CoLA	Is the sentence grammatical or ungrammatical?	"This building is than that one." = Ungrammatical	Matthews
SST-2	Is the movie review positive, negative, or neutral?	"The movie is funny , smart, visually inventive , and most of all, alive ." = .93056 (Very Positive)	Exactitud
MRPC	Is the sentence B a paraphrase of sentence A?	A) "Yesterday , Taiwan reported 35 new infections , bringing the total number of cases to 418 ." B) "The island reported another 35 probable cases yesterday , taking its total to 418 ." = A Paraphrase	Exactitud/Puntaje F1
STS-B	How similar are sentences A and B?	A) "Elephants are walking down a trail." B) "A herd of elephants are walking along a trail." = 4.6 (Very Similar)	Pearson/Spearman
QQP	Are the two questions similar?	A) "How can I increase the speed of my internet connection while using a VPN?" B) "How can Internet speed be increased by hacking through DNS?" = Not Similar	Exactitud/Puntaje F1
MNLI-mm	Does sentence A entail or contradict sentence B?	A) "Tourist Information offices can be very helpful." B) "Tourist Information offices are never of any help." = Contradiction	Exactitud
QNLI	Does sentence B contain the answer to the question in sentence A?	A) "What is essential for the mating of the elements that create radio waves?" B) "Antennas are required by any radio receiver or transmitter to couple its electrical connection to the electromagnetic field." = Answerable	Exactitud
RTE	Does sentence A entail sentence B?	A) "In 2003, Yunus brought the microcredit revolution to the streets of Bangladesh to support more than 50,000 beggars, whom the Grameen Bank respectfully calls Struggling Members." B) "Yunus supported more than 50,000 Struggling Members." = Entailed	Exactitud
WNLI	Sentence B replaces sentence A's ambiguous pronoun with one of the nouns - is this the correct noun?	A) "Lily spoke to Donna, breaking her concentration." B) "Lily spoke to Donna, breaking Lily's concentration." = Incorrect Referent	Exactitud

SQuAD (Stanford Question Answering Dataset) v1.1 y 2.0 son bases de datos de preguntas de la universidad de Stanford. *SQuAD V1.1* contiene preguntas que se responden a partir de un segmento de texto del correspondiente pasaje o la pregunta podría ser no contestable. *SQuAD V2.0* agrega preguntas que no tienen respuesta y evalúa la capacidad del sistema para identificarlo y abstenerse de dar una respuesta [81].

SWAG (Situations With Adversarial Generations) es una base de datos amplia que se basa en la inferencia o el sentido común, contiene 113.000 preguntas con múltiple respuesta acerca de situaciones comunes. Cada pregunta es la descripción de una escena o situación y la tarea del modelo que se entrene es debe identificar cuál de las cuatro posibles opciones es la siguiente escena o situación [82].

3.7.2.4. Principales resultados

En la Tabla 5, se puede observar que *BERT* supera el desempeño de otros modelos como *OpenAI GPT* y *BiLSTM+ELMO+Atm*.

Tabla 5. Desempeño de BERT en tareas de GLUE

Base de datos	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Promedio
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.2	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
<hr/>									
BERT _{BASE}	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT _{LARGE}	86.7/85.9	72.1	92.7	94.9	60.5	86.5	89.3	70.1	82.1

Para las tareas *QQP* y *MRPC* se reporta el puntaje F1, el coeficiente de correlación de Spearman se reporta para *STS-B* y la exactitud para las demás tareas.

Con relación a *SQuAD V1.1* y *V2* también superó el desempeño de modelos anteriores a él, con puntajes F1 de 93.2 (1.5 puntos por encima) para *V 1.1* y de 83.1 (5.1 puntos por encima) para *V2.0*. Para *SWAG*, *BERT* también superó los modelos previos como *ESIM+ELMO* con un 27.1% y *OpenAI GPT* con 8.3%. e incluso superó el desempeño de un humano experto con 1.3% de exactitud. La exactitud de *BERT* para esta tarea fue de 86.3%.

A partir de *BERT* se han diseñado diferentes modelos especializados en diferentes ámbitos y tareas del procesamiento de lenguaje natural. *SciBERT* es uno de ellos y como se mencionó anteriormente, la última versión de *RobotReviewer* hace uso de este lenguaje. A continuación, se describen sus tareas de entrenamiento, el corpus con el que fue entrenado y sus principales resultados.

3.7.3. SciBERT

Es un modelo pre entrenado para el entendimiento del lenguaje natural específico del dominio científico. Fue entrenado con un corpus de 1.14M de artículos de *Semantic Scholar*, el corpus consistió en un 18% de artículos relacionados con ciencias de la computación y 82% de artículos del dominio biomédico. Este modelo también contó con su propio vocabulario (*SCIVOCAB*), creando un nuevo *WordPiece* en el vocabulario específico, a través del uso de la librería *SentencePiece*. El solapamiento entre el vocabulario de *BERT* y *SciBERT* es del 42%, lo que ilustra una gran diferencia entre las palabras usadas en el dominio científico y los dominios de texto general [40].

3.7.3.1. Tareas de entrenamiento y ajuste

Este modelo se evaluó por medio de las siguientes tareas del PLN:

- Reconocimiento de entidades nombradas (*NER* por sus siglas en inglés)
- Extracción del término PICO (paciente, intervención, comparador y resultados)
- Clasificación de relación (*REL* por sus siglas en inglés)
- Análisis de dependencia (*DEP* por sus siglas en inglés)

Para estas tareas se implementaron bases de datos de texto relacionadas con cada una de ellas, dentro de las que se encontraron: *EBM-NLP*, la cual contiene términos PICO de resúmenes de ensayos clínicos. *SciERC* que contiene entidades y relaciones de resúmenes de ciencias de computación. También se usó *ACL-ARC* y *SciCite* que asigna etiquetas a sentencias de artículos científicos que citan otros artículos. Finalmente, se usó la base de datos *The Paper Field* la cual contiene la clasificación de los títulos de los artículos en siete campos de estudio [40].

3.7.3.2. Principales resultados

De los principales resultados de este modelo (Tabla 6) particularmente en el dominio biomédico, se puede observar que *SciBERT* superó los modelos existentes hasta su desarrollo para tres bases de datos relacionadas con tareas de extracción de término PICO, Reconocimiento de entidades nombradas y clasificación de relación. (El término frozen indica un ‘congelamiento’ de los pesos de la red, el cual se compara con el desempeño obtenido luego del ajuste del modelo). *BC5CDR* es una base de datos de 1500 artículos de PubMed etiquetado con 4409 nombres de sustancias químicas, 5818 enfermedades y 3116 interacciones entre químicos y enfermedades. *ChemProt* es una base de datos que contiene texto etiquetado con sustancias químicas/medicamentos y genes/proteínas [40].

Tabla 6. Principales resultados SciBERT, adaptado de [40]

Campo de Aplicación	Tarea	Base de datos	Estado del arte	BERT-Base		SCIBERT	
				Pesos Congelados	Ajuste	Pesos Congelados	Ajuste
Bio	NER	BCSCDR (Li et al., 2016)	88.85 ⁷	85.08	86.72	88.73	90.01
		JNLPBA (Collier and Kim, 2004)	78.58	74.05	76.09	75.77	77.28
		NCBI disease (Dogan et al., 2014)	89.36	84.06	86.88	86.39	88.57
	PICO	EBM-NLP (Nye et al., 2018)	66.30	61.44	71.53	68.30	72.28
	DEP	GENIA (Kim et al., 2003) - LAS	91.92	90.22	90.33	90.36	90.43
		GENIA (Kim et al., 2003) - UAS	92.84	91.84	91.89	92.00	91.99
REL	ChemProt (Kringelum et al., 2016)	76.68	68.21	79.14	75.03	83.64	
CS	NER	SciERC (Luan et al., 2018)	64.20	63.58	65.24	65.77	67.57
	REL	SciERC (Luan et al., 2018)	n/a	72.74	78.71	75.25	79.97
	CLS	ACL-ARC (Jurgens et al., 2018)	67.9	62.04	63.91	60.74	70.98
Multi	CLS	Paper Field	n/a	63.64	65.37	64.38	65.71
		SciCite (Cohan et al., 2019)	84.0	84.31	84.85	85.42	85.49
Promedio				73.58	77.16	76.01	79.27

4. Metodología y Resultados

Para dar cumplimiento a los objetivos planteados se siguió la metodología que se describe en cada sección de resultados, los cuales se presentan con relación a cada uno de los objetivos específicos del trabajo de investigación. En términos generales el flujo de trabajo de cada objetivo contribuyó al desarrollo del objetivo general como se sintetiza en la Figura 15. Por medio del cumplimiento del primer objetivo se identificó la principal base de datos usada en las instituciones para realizar búsquedas de evidencia, así mismo se comprendió que la búsqueda se realiza de manera exigente pero no exhaustiva como lo exige una revisión sistemática. El segundo objetivo permitió levantar una base de datos de reportes de ETS por medio de las cuales se identificaron los estudios clínicos incluidos para hacer las evaluaciones. Por medio de esta base de datos se identificó el título, resumen y el diseño de cada estudio, así mismo se identificaron las estrategias de búsqueda implementadas. Dando paso así al entrenamiento de un algoritmo de identificación de ensayos clínicos y revisiones sistemáticas y se presenta una estrategia para priorizar los estudios que fueron resultado de la búsqueda en la base de datos consultada.

Posteriormente, con el desarrollo del objetivo cuatro se pudo integrar la creación de estrategias de búsqueda en PubMed, búsqueda remota, clasificación de diseño de estudios y extracción de información (términos PICO) en los procesos de síntesis de evidencia clínica. De esta manera se logró obtener una herramienta semiautomática para ser usada en los procesos de evaluación de tecnología en salud (ETS) en las instituciones acreditadas de la ciudad que facilite la síntesis de evidencia clínica.

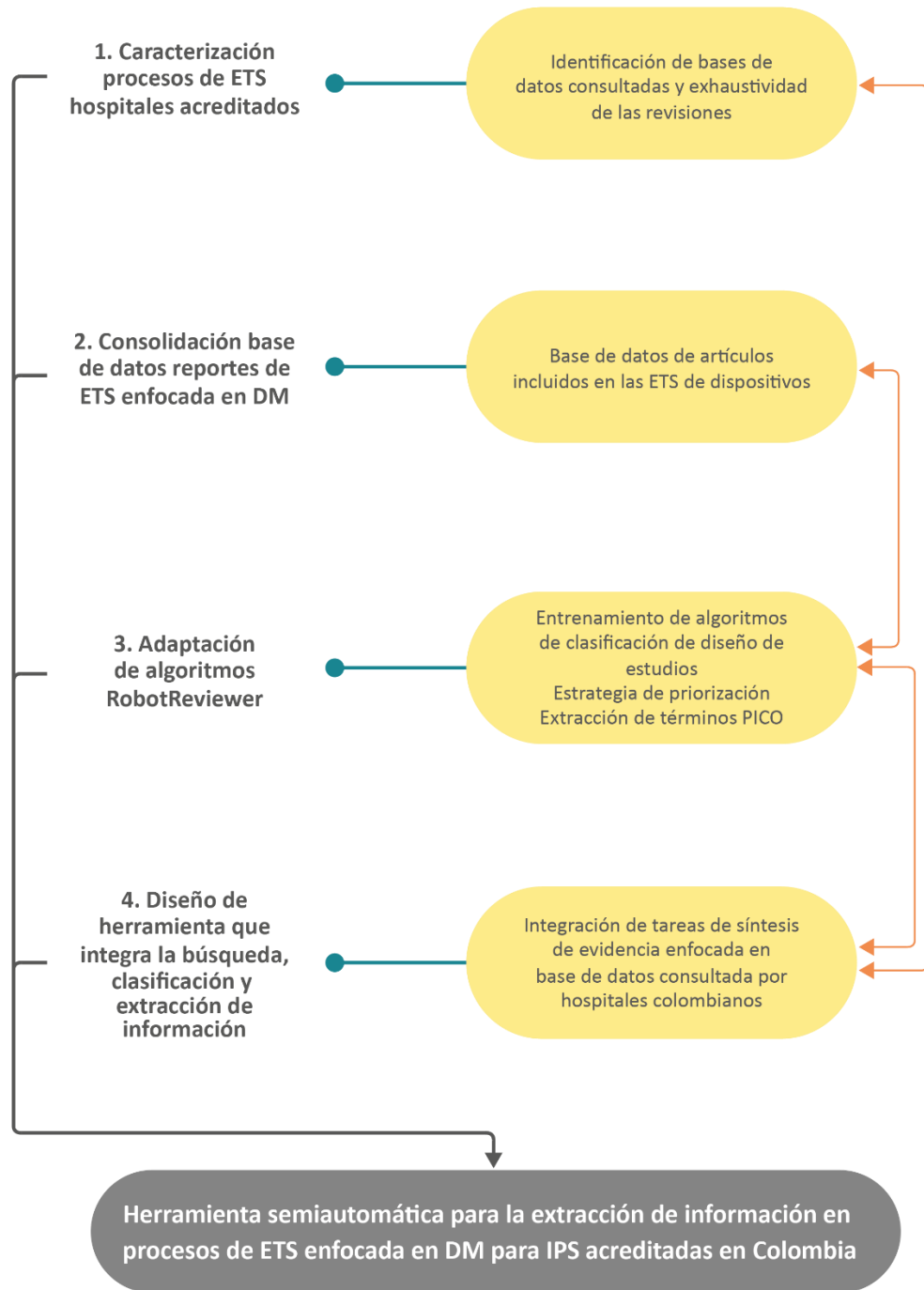


Figura 15. Metodología general de trabajo

4.1. Caracterización del proceso de evaluación de tecnología en salud en hospitales colombianos

Para llevar a cabo este objetivo se siguió la metodología que se sintetiza en la Figura 16.

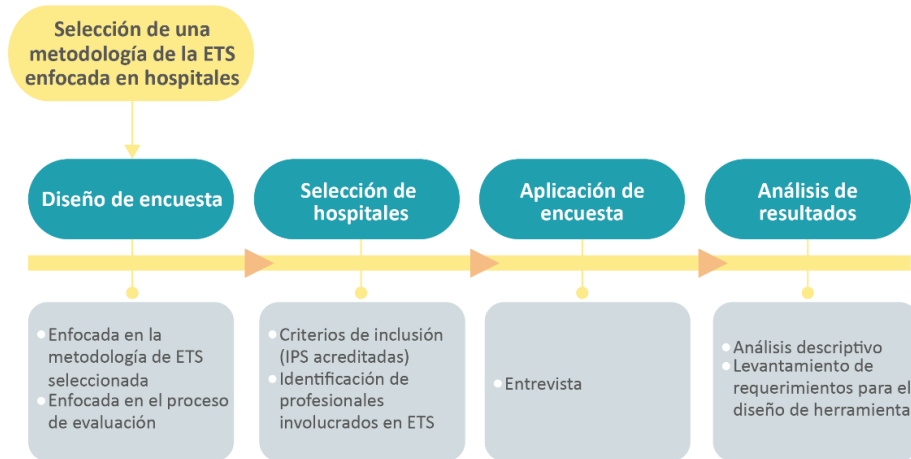


Figura 16. Metodología implementada para la selección y caracterización de una metodología de ETS enfocada y aplicada en IPS acreditadas

Inicialmente se realizó la selección de una metodología de ETS, los criterios tenidos en cuenta para esta selección fueron los siguientes:

- ✓ Que fuera una metodología enfocada en hospitales
- ✓ Que se utilizara para guiar la toma de decisiones sobre selección e incorporación de tecnología a los hospitales
- ✓ Que fuera una metodología ampliamente utilizada en Colombia y en el mundo
- ✓ Que contara con suficiente información metodológica disponible para su aplicación

Esta selección se basó en una revisión bibliográfica tanto de las diferentes metodologías enfocadas en el ámbito hospitalario como de su implementación en el mundo, Latinoamérica y Colombia (en el nivel meso).

Posteriormente se realizó el diseño de una encuesta de caracterización enfocada en la metodología de evaluación seleccionada en el paso anterior. Esta encuesta contó con un total de 28 preguntas, 8 de las cuales fueron de respuesta abierta, 10 de respuesta cerrada y 10 de múltiples respuestas como se ilustra en la Tabla 7. Esta encuesta contó con tres secciones, una parte inicial que pretendía conocer la estructura organizacional o los profesionales involucrados en el proceso de ETS y el tipo de tecnologías que se evalúan en los hospitales. Una etapa enfocada en la ejecución de la evaluación con énfasis principalmente en los aspectos más relevantes sobre la síntesis de evidencia clínica. Finalmente, una sección de cierre en la cual se pretendía conocer el impacto y la implementación de las recomendaciones de los reportes de ETS.

Tabla 7. Estructura de encuesta diseñada para guiar la caracterización de ETS

Sección	Descripción	Tipo de pregunta	Cantidad de preguntas
Introducción	El objeto de estas preguntas fue conocer los profesionales involucrados en el proceso de ETS y las tecnologías evaluadas en la institución.	Respuesta abierta	2
		Respuesta cerrada	2
		Múltiples respuestas	2
Proceso y documentación	Esta es la sección principal de la encuesta y tuvo el objetivo de caracterizar la revisión de evidencia clínica, iniciando con la pregunta de investigación hasta la síntesis del reporte de ETS.	Respuesta abierta	5
		Respuesta cerrada	7
		Múltiples respuestas	8
Información complementaria	En esta sección se abordó el análisis del impacto y la implementación de las recomendaciones obtenidas del proceso de ETS.	Respuesta abierta	1
		Respuesta cerrada	1

Una vez se tuvo el diseño de la encuesta se seleccionaron los hospitales para caracterizar sus procesos de evaluación, el criterio de selección fue que contaran con acreditación en salud. Es importante resaltar que se enfocó en estas instituciones dado que, según lo expuesto anteriormente son aquellas que garantizan procesos de evaluación de tecnología en salud porque son exigidos por la normativa de acreditación. Lo anterior no indica que algunos hospitales no acreditados no tengan procesos de evaluación, pero son más escasos y el enfoque de este trabajo de investigación no fue evaluar la ETS de los hospitales.

Finalmente, la aplicación de las encuestas de caracterización se realizó por medio de entrevistas dirigidas con uno de los profesionales que hacen parte de los procesos de ETS de los hospitales seleccionados. Esta caracterización se realizó durante el segundo trimestre del año 2020, dada la emergencia sanitaria las entrevistas se realizaron de manera virtual. Luego, se realizó un análisis descriptivo de los resultados consolidados, con base en este análisis se extrajeron aspectos relevantes que servirían como insumo para el planteamiento de la herramienta de síntesis de evidencia semiautomática que se propone en este trabajo de investigación.

Como resultados de este objetivo y de acuerdo a la metodología planteada, inicialmente se eligió una metodología de evaluación en salud enfocada en hospitales, a partir de las cuatro grandes metodologías identificadas a nivel mundial por la HTAi [30]. Esta elección se basó en el análisis de los criterios planteados inicialmente y se describe a continuación:

- **El uso de la ETS para apoyar la toma de decisiones para la incorporación de tecnología en el país.** Bajo esta premisa y con base en lo expuesto sobre las metodologías de ETS y la Figura 3, se reducen a dos las metodologías enfocadas en este objetivo: la mini-ETS y las unidades de evaluación.

- **El contexto colombiano en relación con la ETS.** Donde se tienen recursos limitados en el sistema de salud y no es común la vinculación de profesionales en los hospitales para dedicarse exclusivamente a realizar actividades de ETS. Sino por el contrario asignar diferentes responsabilidades, dentro de ellas apoyar la toma de decisiones a nivel gerencial o de direccionamiento con base en el conocimiento técnico de las diferentes áreas.
- **La implementación de la mini-ETS a nivel global.** Como se expuso anteriormente la mini-ETS es un modelo ampliamente utilizado en el mundo y en Latinoamérica con la respectiva adaptación al contexto local en el que se lleva a cabo. Por este motivo se cuenta con amplia información sobre casos de éxito y formas de implementación como lo expone el HandBook de la iniciativa AdHopHTA [33].
- **Los diferentes abordajes de los modelos y formatos de reportes de evaluación de tecnología en salud.** Es decir, el enfoque en la toma de decisiones, el nivel de complejidad de las metodologías y la ‘facilidad’ que representa el abordaje de la ETS a través de la mini-ETS.
- **La concordancia entre los formatos de mini-ETS y el formato de verificación de criterios de los reportes de evaluación desarrollado por la red internacional de agencias para la evaluación (INAHTA por sus siglas en inglés).** Con este último criterio se garantiza que desarrollar evaluación a través de metodología mini-ETS cumple con los criterios establecidos por la comunidad mundial de agencias de evaluación de tecnología.

Teniendo en cuenta lo expuesto, se seleccionó la mini-ETS como modelo o metodología de evaluación de tecnología en salud para guiar la caracterización de los procesos de evaluación en los hospitales de la ciudad de Medellín.

Es importante resaltar que en países como Colombia la ETS se basa principalmente en la revisión de la evidencia clínica disponible generada en países productores de tecnología en salud. Esto se puede constatar al contrastar la correlación entre los países que más aportan en el reporte de ensayos clínicos en el mundo (Estados Unidos, China, Japón, Alemania, Reino Unido) [83] y los mayores productores de dispositivos médicos (Estados Unidos, Alemania, Francia, Irlanda) [84]. Según AdhopHTA, los reportes de mini-ETS se basan en un análisis retrospectivo y revisiones de la literatura disponible, así como de opiniones de expertos, aunque este análisis es sistemático no es tan exhaustivo como en los informes completos de ETS. Esta característica de sistematicidad no exhaustiva es coherente con el tiempo y la disponibilidad de los profesionales para realizar ETS en los hospitales en el contexto colombiano.

Como resultado de una amplia revisión de literatura se identificaron los formatos de mini-ETS de AdhopHTA, DACEHTA e INAHTA además de información complementaria y con base en esta información se consolidó la encuesta que se resume en la Tabla 9. En la cual también se ilustran los resultados obtenidos después de realizar las entrevistas guiadas a los profesionales que participan de los procesos de ETS.

De acuerdo con el criterio de selección de los hospitales a caracterizar, en las entrevistas participaron profesionales involucrados en el proceso de ETS de las cuatro instituciones

prestadoras de servicios de salud acreditadas en alta calidad de la ciudad de Medellín, la capacidad instalada de estas instituciones se ilustra en la Tabla 8.

Tabla 8. Capacidad instalada instituciones incluidas en la caracterización (Fuente: Registro especial de prestadores de servicios de salud)

Institución/Ítem	Ciclo de acreditación	ICONTEC	JCI	Camas adultos	Camas pediátricas	UCI adultos	Quirófanos
Hospital Pablo Tobón Uribe (HPTU)	4° Acreditado con excelencia	X	X	298	68	96	14
Hospital San Vicente	3°	X		237	120	93	16
Clínica Universitaria Bolivariana	1°	X		104	14	30	5
Hospital General de Medellín	4°	X		253	50	58	8
IPS universitaria	Autoevaluación	X		486	21	83	17

Todas estas instituciones se encuentran acreditadas por el Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC), adicionalmente el HPTU cuenta con acreditación otorgada por la Comisión Conjunta Internacional (JCI por sus siglas en inglés), institución de certificación estadounidense. Con relación a esta última se exige bajo el estándar GLD.7 el liderazgo para la toma de decisiones relacionadas con la compra y el uso de los recursos (humanos y técnicos) con base en las implicaciones de calidad y seguridad de las mismas. Dentro de los criterios contemplados en este estándar se encuentran:

- Búsqueda de datos para informar **decisiones relacionadas** con la compra y **uso de nuevas tecnologías**
- Uso de información de las implicaciones de **calidad y seguridad** de la selección de **tecnología médica**
- Uso de información de las implicaciones de calidad y seguridad de la selección y decisiones relacionadas con el talento humano
- Monitoreo de resultado de las decisiones y uso de información para evaluar y mejorar la calidad de plan de adquisición y decisiones de asignación de recursos
- Uso de **recomendaciones** de organizaciones profesionales y otras autoridades en el proceso de **toma de decisiones**

Por lo tanto, es de esperarse un proceso de ETS mucho más robusto para esta institución.

La encuesta diseñada estuvo distribuida en tres grandes secciones: introducción, aspectos metodológicos del proceso e impacto de la ETS en las instituciones. La cantidad, el tipo de preguntas, las secciones y el objeto de estas se ilustran en la Tabla 7.

4.1.1. Resultados caracterización

Los resultados de la caracterización se presentan en la Tabla 9 y posteriormente se hace una descripción de los principales resultados por cada institución.

Tabla 9. Resultados caracterización por preguntas e institución

Institución Pregunta	San Vicente Fundación	Hospital Pablo Tobón Uribe	Clínica Universitaria- Universidad Pontificia Bolivariana	IPS Universitaria - Clínica León XIII
¿Quién realiza una revisión de la literatura para evaluar la nueva tecnología?	Médicos, epidemiólogos.	Médicos, epidemiólogos.	Médicos, Epidemiólogos	Médicos
¿Los reportes de ETS generalmente son revisados por otros profesionales no relacionados con el proceso?	Si	No	No	Si
¿A qué tipo de tecnología en salud se le hace ETS?	Medicamentos, Equipos médicos, Dispositivos médicos, Procedimientos	Medicamentos, Equipos médicos, Dispositivos médicos, Procedimientos, Tecnologías de información	Equipos médicos, Dispositivos médicos	Medicamentos, dispositivos médicos, equipos biomédicos
¿Ha existido algún conflicto de intereses al realizar ETS?	No	No	No	No
¿Qué estrategia se implementa para definir claramente el objeto de la ETS?	PICO (Paciente, Intervención, Comparador, Resultados)	PICOT (Paciente, Intervención, Comparador, Resultados, Tiempo)	TICO (Tecnología, Intervención, Comparador, Resultados) y PICO (Paciente, Intervención, Comparador, Resultados)	Tecnología a incorporar, pacientes objetivo, tecnología que se usa actualmente y principales desenlaces que justifican la incorporación de la nueva tecnología
¿Cómo se identifica la necesidad de una revisión sistemática de la literatura para dar respuesta a la ETS?	Para todas las evaluaciones se hace una revisión de la literatura.	Para todas las evaluaciones se hace una revisión de la literatura, independiente de si existen o no reportes de ETS y revisiones adicionales.	Consulta inicial de fuentes de ETS	Durante la evaluación no se hace revisión sistemática o no hay evidencia de ello. Por lo tanto, no hay evidencia de identificar revisiones previas antes de ejecutar búsqueda de evidencia de otros estudios
¿Se hace un planteamiento de la pregunta de investigación?	Si	Si	Si	No específicamente
¿Se determinan criterios de inclusión y exclusión de la evidencia a incluir en la revisión de la literatura?	Si	Si	No	Se describen los criterios PICO sin ser una estrategia estandarizada, para los demás criterios no se determinan criterios de manera exhaustiva

Institución Pregunta	San Vicente Fundación	Hospital Pablo Tobón Uribe	Clínica Universitaria- Universidad Pontificia Bolivariana	IPS Universitaria - Clínica León XIII
¿Se documentan las estrategias de búsqueda implementadas para recuperar la evidencia?	Si	Si	Si	No
¿Quién o quiénes seleccionan la literatura a incluir o excluir?	Médicos que conforman un listado que el comité tiene predefinido	Médicos - epidemiólogos	Médicos	Médicos
¿La selección de evidencia y la extracción de información se hace de manera aislada entre los profesionales que la realizan?	No	Si	No	No
¿Qué tipo de evidencia clínica se usa para llevar a cabo la ETS?	Información generada a partir de búsqueda de la literatura, Información generada a partir de la experiencia clínica	Información generada a partir de búsqueda de la literatura	Información generada a partir de búsqueda de la literatura	Información generada a partir de búsqueda de la literatura
¿Qué tipos de estudios se incluyen en las revisiones de literatura?, principalmente en la evaluación de dispositivos médicos.	Ensayos clínicos aleatorizados, Estudios de cohorte, Revisiones sistemáticas, Metaanálisis	Ensayos clínicos aleatorizados, Estudios de cohorte, Estudios de casos y controles, Estudios transversales, Revisiones sistemáticas, Metaanálisis, Reportes de evaluación de tecnología en salud	Ensayos clínicos aleatorizados, Revisiones sistemáticas, Metaanálisis, Reportes de evaluación de tecnología en salud	Ensayos clínicos aleatorizados, Estudios de cohorte, Estudios de casos y controles, Estudios transversales, Revisiones sistemáticas, Metaanálisis, Reportes de evaluación de tecnología en salud*
¿Qué fuentes de información se consultan en las búsquedas de literatura?	No identificado	GRADE	Lista de chequeo de la iniciativa para el fortalecimiento de la presentación de informes de estudios observacionales en epidemiología (STROBE), Herramienta Cochrane para análisis de sesgos, Lista de chequeo de la red de colegios escocesa (SIGN), Lista de chequeo del programa de habilidades de evaluación crítica (CASP)	No indica

Institución	San Vicente Fundación	Hospital Pablo Tobón Uribe	Clínica Universitaria-Universidad Pontificia Bolivariana	IPS Universitaria - Clínica León XIII
Pregunta				
¿Qué fuentes de información se consultan para identificar la evidencia basada en la experiencia clínica?	Base de datos recalls FDA, GESTIÓN DE INFORMES DE SEGURIDAD - ALERTAS - RECALLS E HURTOS (RISARH) INVIMA, Base de datos de fabricantes y usuarios de dispositivos médicos (MAUDE) FDA, Alertas de seguridad de ECRI	Base de datos recalls FDA, GESTIÓN DE INFORMES DE SEGURIDAD - ALERTAS - RECALLS E HURTOS (RISARH) INVIMA	Base de datos recalls FDA, GESTIÓN DE INFORMES DE SEGURIDAD - ALERTAS - RECALLS E HURTOS (RISARH) INVIMA, Base de datos de fabricantes y usuarios de dispositivos médicos (MAUDE) FDA, Alertas de seguridad de ECRI	Base de datos recalls FDA, GESTIÓN DE INFORMES DE SEGURIDAD - ALERTAS - RECALLS E HURTOS (RISARH) INVIMA
¿Qué metodología se implementa para evaluar la calidad de la evidencia consultada?	No identificado	Calificación de las recomendaciones de evaluación (GRADE)	Diseño propio	No indica
Definición de tecnología a evaluar, el comparador, la población, los resultados a medir	2 semanas	2 semanas	2 horas	No se lleva registro
Planteamiento de pregunta de investigación y planificación de búsqueda de literatura	2 semanas	1-2 semanas	2 horas (en conjunto con actividad anterior)	No se lleva registro
Ejecución de la revisión (búsqueda, recuperación, inclusión y exclusión, extracción de información, análisis de la calidad de información)	2 semanas	1-2 semanas	10 horas	No se lleva registro
Síntesis de recomendaciones	2 semanas	1 semana	2 horas	1 semana después de que se recibe la solicitud de evaluación y se verifican los estudios presentados.

Institución Pregunta	San Vicente Fundación	Hospital Pablo Tobón Uribe	Clínica Universitaria-Universidad Pontificia Bolivariana	IPS Universitaria - Clínica León XIII
¿Se hacen referenciaciones con otras instituciones para identificar la experiencia en la implementación de la tecnología evaluada?	Si	No	Si	Si
¿Cuáles de los siguientes aspectos organizacionales se tienen en cuenta para la evaluación de una tecnología?	Personal y cargas de trabajo, Adecuaciones de espacio, Impacto en los usuarios relacionados con información, entrenamiento, capacitación, Ambiente y organización del trabajo, En qué momento sería adecuado introducir la tecnología en la institución, Interdependencia entre servicios del hospital y colaboración con otras instituciones	Personal y cargas de trabajo, Adecuaciones de espacio, Impacto en los usuarios relacionados con información, entrenamiento, capacitación, Ambiente y organización del trabajo, En qué momento sería adecuado introducir la tecnología en la institución, Interdependencia entre servicios del hospital y colaboración con otras instituciones	Personal y cargas de trabajo, Adecuaciones de espacio, Impacto en los usuarios relacionados con información, entrenamiento, capacitación, Ambiente y organización del trabajo, Interdependencia entre servicios del hospital y colaboración con otras instituciones	Personal y cargas de trabajo, Adecuaciones de espacio, Impacto en los usuarios relacionados con información, entrenamiento, capacitación, En qué momento sería adecuado introducir la tecnología en la institución, Interdependencia entre servicios del hospital y colaboración con otras instituciones
¿Las recomendaciones realizadas han sido aceptadas e implementadas en las decisiones para la incorporación de la tecnología en la institución?	Sí	Sí	Sí	Sí
¿Se ha estimado el ahorro monetario obtenido a causa de la implementación de las recomendaciones de las ETS en la institución?	Solamente en un caso específico	No se ha estimado	No se ha estimado	No se ha estimado

4.1.2. Resultados individuales por hospital

Hospital Pablo Tobón Uribe

La ETS en el Hospital Pablo Tobón Uribe parte de una solicitud hecha por los profesionales de la institución, la cual se realiza en un formato determinado y ajustado según las necesidades del hospital. Con base en esta solicitud, el comité de evaluación se

realiza la respectiva revisión de literatura, previo a un filtro realizado por el área de gestión de calidad donde se identifica cual tecnología debe ir a comité de evaluación o se puede adquirir sin evaluarse. La solicitud de evaluación de la tecnología es muy importante, dado que, a partir de esta se identifica por qué se está solicitando la tecnología, cuales resultados se espera obtener con su implementación y cuál es la tecnología (comparador) que se usa actualmente para suplir la necesidad planteada. El profesional que solicita y propone la implementación de la tecnología es una fuente importante de información que se tiene en cuenta para guiar algunas etapas del proceso de medición.

El comité de evaluación de tecnología de la institución está conformado por el director general del hospital, personal administrativo, médicos, y un ingeniero biomédico con formación en epidemiología.

La revisión de la literatura es realizada generalmente por dos médicos con formación en epidemiología, quienes hacen al final de la revisión una comparación de la información recuperada por cada uno para hacer la síntesis de evidencia. En el comité del HPTU tanto el análisis de sesgo de los estudios epidemiológicos revisados como la calidad de la evidencia recuperada se analiza por medio de la escala *GRADE* (Grading of Recommendations Assessment, Development and Evaluation). Una vez se finaliza el proceso de revisión, se escriben dos informes con las recomendaciones basadas en dicha revisión, se hace una descripción narrativa de los principales hallazgos en la revisión y se presenta en el comité donde las recomendaciones y conclusiones son discutidas y aprobadas. De cada evaluación surgen dos tipos de informes, uno completo donde se hace una descripción amplia de todos los aspectos de la revisión y uno corto que es publicado y es de acceso libre en la página del hospital. Los informes no son revisados por pares externos, sin embargo, siempre se analiza su contenido por medio de una lista de chequeo de los mínimos que debe cumplir. Es importante resaltar que en el hospital no se basan en resultados de ETS externas, sino que siempre se realizan sus propias evaluaciones y en los casos en los que se identifican informes de evaluación a través de la revisión, estos son descritos dentro de los hallazgos como parte de la evidencia consultada.

En el Hospital Pablo Tobón Uribe la recomendación de la ETS para la toma de decisiones es muy importante en este proceso, sin embargo, al existir otros factores (por ejemplo: económicos, organizacionales, éticos) que pueden influir en las mismas, no es posible afirmar que las decisiones tomadas siempre han estado alineadas con las recomendaciones dadas. También es importante resaltar que las evaluaciones tienen un carácter no vinculante, es decir, los profesionales de la salud no están obligados a adherirse a las recomendaciones, dado que, sobre estas últimas prima el criterio médico.

En cuanto a la evaluación de dispositivos médicos, como para cualquier evaluación, en el hospital se basan en la evidencia de mejor calidad y en particular no se describió ninguna brecha específicamente para hacer la ETS de este tipo de tecnologías. Sin embargo, si se expresó que la poca cantidad de evidencia disponible los ha llevado a dejar evaluaciones inconclusas en dos o tres ocasiones, principalmente cuando la tecnología está relacionada con enfermedades huérfanas. Cuando no se logra realizar una revisión de la literatura para hacer la ETS dada la poca disponibilidad de información o ausencia de la misma, se realizan

referenciaciones con otras instituciones, sin embargo, estas no se consideran parte de la ETS y no son ámbito de trabajo del comité de evaluación del hospital. También se identificó no se ha evaluado en la institución el impacto económico de la implementación o no de las recomendaciones. Programas o estudios para evaluar la experiencia del paciente con relación al uso de la tecnología tampoco son parte de las actividades ejecutadas por el comité de evaluación, sin embargo, si se realizan en el hospital como parte del mejoramiento continuo y el sostenimiento de los estándares de acreditación en salud. Algunos aspectos organizacionales tampoco son considerados por el comité.

Clínica Universitaria Bolivariana

En la clínica se cuenta con un formato de solicitud de inclusión de nuevas tecnologías, el cual es recibido por el comité de gestión de tecnología donde se analiza y se asigna al responsable de la revisión de literatura. El director científico de la clínica es el responsable de la evaluación de nuevas tecnologías, quién también apoya el grupo de evaluación de tecnología biomédica en la comparabilidad tecnológica. La búsqueda y síntesis de literatura es realizada por médicos y en necesidades particulares se solicita apoyo al grupo de epidemiólogos. Los informes de evaluación de la clínica no cuentan con evaluadores externos. El comité de gestión de la tecnología está conformado por médicos, profesionales de ingeniería biomédica y coordinación de farmacia.

La revisión de la literatura se inicia con una búsqueda de previos reportes de evaluaciones de tecnología de otras instituciones, y si la evaluación existe se analiza el cumplimiento de criterios comparados con la tecnología que se está evaluando en la institución, y cuando no existen otras ETS se realiza una revisión y síntesis de evidencia propia en la institución. La revisión y síntesis de evidencia no es necesariamente una revisión sistemática, sino que se centra en identificar metaanálisis o revisiones sistemáticas que cumplan con criterios de calidad y esta revisión es llevada a cabo por dos médicos de manera conjunta. Una vez finalizado el análisis de evidencia se elabora un informe de ETS que no son públicos.

La metodología para la evaluación de tecnología, específicamente del criterio organizacional, está en proceso de construcción, y considera un levantamiento de información institucional a partir de investigación de campo o encuestas realizadas al interior de la clínica para identificar aspectos relevantes en la ETS. También se hace referenciación y levantamiento de información en otras instituciones para estructurar aspectos organizacionales de la implementación de la tecnología como las curvas de aprendizaje, el personal necesario para la operación, entre otros. En el aspecto organizacional de seguridad se analizan los eventos adversos derivados del uso de la tecnología.

Todas las recomendaciones realizadas se tienen en cuenta en la toma de decisiones, a la fecha de realización de esta entrevista, se han evaluado pocas tecnologías y la recomendación ha sido la no adquisición de estas, a raíz de esto tampoco se han realizado análisis del impacto de la implementación o no de una tecnología con base en las recomendaciones realizadas.

Hospital San Vicente fundación

En el hospital San Vicente también se cuenta con un formato de solicitud de la tecnología, el cual es recibido en el área de ingeniería biomédica y es analizado y filtrado por medio de una matriz de clasificación o árbol de decisión. La tecnología con un nivel de riesgo III siempre pasa a ser objeto de evaluación en el comité de ETS del hospital, esto debido a que es el nivel más alto de riesgo en la clasificación nacional regulada por el decreto 4725 de 2005 [52]. En el país existen realmente pocos equipos biomédicos con esta clasificación, esta particularidad solo la expresó esta institución. La de riesgo IIB se analiza y dependiente del volumen de adquisición y costo se evalúa en comité o continúa el proceso de adquisición en el área de ingeniería biomédica. Toda la tecnología que se analiza en el comité no necesariamente se evalúa, sino que a partir de un consenso se identifica si existe un interés en la adopción de la tecnología o no y dependiendo de ello se hace la ETS. Tecnología de riesgo I y IIA no se evalúa en comité. El comité de evaluación de tecnología de la institución está conformado por los directores del hospital (sede Medellín y Rionegro), el director de la unidad de investigación, el gerente de compras, la coordinadora de farmacia, la ingeniera biomédica, el coordinador TI y jefes de gestión clínica.

Una vez la tecnología se asigna para evaluación en el comité, se selecciona de una lista de 35 integrantes los profesionales que realizarán la revisión de la literatura, que generalmente son dos. Cuando estos profesionales tienen formación en epidemiología, realizan de manera individual la revisión, en caso contrario se asigna un asesor metodológico que guía la revisión. La evaluación también se basa en la mejor evidencia disponible y se baja de calidad en los estudios epidemiológicos de acuerdo con los resultados que se van obteniendo en el proceso. La falta de evidencia para realizar evaluación se cubre con referenciones a otras instituciones y la identificación de experiencias. Al finalizar el proceso de revisión se estructura un informe de evaluación que contiene toda la evidencia analizada y las recomendaciones que surgieron a partir de la misma. Estos informes no son públicos.

En la institución todas las recomendaciones que surgen de la ETS del comité son la base para la toma de decisiones de adquisición de tecnología. Tampoco se han realizado estudios para analizar el impacto de las recomendaciones del comité de evaluación sobre la implementación o no de la tecnología. Solamente en un caso particular se ha realizado un seguimiento y análisis de la implementación de la tecnología para evaluar el impacto económico de la misma.

IPS Universitaria

La IPS Universitaria es una de las instituciones con mayor capacidad del departamento de Antioquia (ver Tabla 8). Dentro de su plan estratégico se busca el reconocimiento como Hospital Universitario de la Universidad de Antioquia, meta para la cual requiere la acreditación en salud como paso previo [85]. Como parte de su proceso para llegar a la acreditación se encuentra el fortalecimiento de la evaluación de tecnología en salud que se lleva a cabo, motivo por el cual la IPS Universitaria se incluye dentro de esta caracterización pese a no ser un hospital acreditado aún en el país.

En la institución se cuenta con dos formatos de solicitud de evaluación de tecnología. Uno es destinado a las solicitudes de medicamentos y dispositivos médicos y el otro a las solicitudes de equipos biomédicos. Aunque en ambos formatos se solicita información sobre la evidencia clínica disponible que soporta la incorporación de la tecnología, solamente en aquellos relacionados con medicamentos y dispositivos médicos son en los que se diligencia completamente en este ítem. Por este motivo, la caracterización del proceso estará enfocada principalmente en estos dos tipos de tecnología.

Actualmente no existe un comité consolidado de evaluación de tecnología, pero en el proceso están involucrados en primera instancia, el personal asistencial que realiza la solicitud de inclusión de la tecnología y quienes reciben la solicitud, la analizan y toman una decisión de incorporación. Los cargos que hacen esta recepción son la coordinación de farmacia, coordinación de toxicología y farmacovigilancia e ingeniería biomédica líder en tecnovigilancia. Dentro de la solicitud de evaluación se debe diligenciar la tecnología a incluir, principales desenlaces o motivos por los cuales se requiere incluir, los pacientes en los cuales tendrá impacto y la tecnología que se usa actualmente en la institución (comparador). Aunque no se implementa la estrategia PICO específicamente bajo ese nombre, en los reportes se relacionan cada uno de los criterios necesarios o base que plantea esta estrategia. Así mismo, el solicitante debe adjuntar mínimo dos estudios clínicos que soporten las afirmaciones sobre seguridad y efectividad de la nueva tecnología. Cuando no se cuenta con evidencia disponible se realiza una referenciación con otras instituciones para identificar experiencias exitosas con relación al uso de la tecnología. Así mismo, se describen el principal servicio clínico donde será utilizado y los servicios alternos que se verían beneficiados indirectamente con la incorporación de la tecnología. También es importante y se describen los beneficios para el personal asistencial que lo usará.

Una vez la solicitud es recibida, se hace un proceso de verificación de la evidencia clínica adjunta. Desde la coordinación de farmacia se realiza el estudio de costos asociado al uso de la tecnología. Finalmente, conjuntamente las áreas responsables hacen una recomendación en conjunto con base en la información analizada hacen una recomendación. Todas las decisiones que se toman en este proceso son adoptadas por la IPS para la incorporación o no de la tecnología.

En el reporte de evaluación de tecnología con el que cuenta la clínica se consigna todo lo relacionado con los criterios PICO, los estudios incluidos en su análisis, el estudio de costos y la decisión de adopción o no. No se evidencia registro de límite temporal de la búsqueda, restricciones de idioma, estrategias de búsqueda implementadas, herramientas para análisis crítico de la evidencia, estudios excluidos o tipo de diseño de estudios a incluir. Es decir, no se evidencia una metodología clara para realizar la revisión de literatura. Así mismo, se evidencia que esta revisión no es exhaustiva, sino que se basa principalmente en identificar rápidamente estudios que validen la inclusión de la tecnología.

Análisis general

De manera general, se identificó que, en todas las instituciones abordadas, la ETS parte de una necesidad o solicitud de incorporación de la tecnología. Esta solicitud generalmente

es elaborada con base en formatos estandarizados en cada institución según su contexto y tienen en común que pretenden identificar tanto la tecnología a evaluar como los beneficios y los impactos que se pretenden tener con la incorporación de la tecnología. Para las instituciones es muy importante cuantificar e identificar los indicadores institucionales que incrementarán a raíz de la implementación de la tecnología en cuestión. Se logró identificar que los formatos implementados para generar la solicitud de evaluación de ETS tienen las características que se sintetizan en la etapa inicial de los formatos de mini-ETS.

En particular llama la atención la diferencia de tiempos de ejecución de los procesos de evaluación entre las instituciones (Tabla 9). Sin embargo, existe una correlación alta entre la estructura del proceso, el comité evaluador y los tiempos de ejecución en cada institución, es decir, entre menos participantes y/o sesiones de trabajo se realizan en función de una evaluación de tecnología, los tiempos de ejecución son menores y la cantidad de informes también es reducida. En hospitales como el HPTU y el hospital San Vicente Fundación se identificaron tiempos globales de ejecución de ETS similares (dos meses y medio); estas dos instituciones tienen una capacidad instalada grande y procesos de evaluación de tecnología estructurados y comités de evaluación integrados por profesionales de diferentes áreas. La Clínica Universitaria Bolivariana, tiene la menor capacidad instalada entre todas las instituciones analizadas y en donde el proceso de evaluación aún se encuentra en fase de implementación y los profesionales involucrados son pocos, aquí el tiempo de ejecución es menor, además de que la cantidad de tecnologías evaluadas es mucho menor en comparación a la cantidad de tecnologías evaluadas en el HPTU y el San Vicente. En esta institución se describe un tiempo de ejecución en horas (14 horas en total), el cual es el tiempo más corto de todos los analizados. Esto también puede deberse no solo a la 'madurez' del comité de evaluación y a la cantidad de profesionales que lo integran sino también a la estrategia de trabajo que se aborda en los mismos, dado que, el establecimiento de reuniones periódicas con los profesionales genera gran dispersión en el tiempo.

Se identificó que la metodología de mini-ETS es la implementada en todas las instituciones, debido al enfoque de evaluar para apoyar la toma de decisiones, que los profesionales que la realizan tienen funciones adicionales a las de evaluar tecnología y que la etapa inicial para solicitar una ETS cumple con los lineamientos sintetizados en los formatos de mini-ETS. Se identificó que en todas las instituciones se evalúan medicamentos y dispositivos médicos, estos últimos en menor medida.

Teniendo en cuenta la caracterización realizada se identificó PubMed como la base de datos que consultan todos los hospitales en sus procesos de evaluación de tecnología, esto es importante para el cumplimiento del objetivo específico 4, acerca del desarrollo de la herramienta semiautomática para síntesis de evidencia. Dado que, es necesario consultar de manera remota las bases de datos para luego identificar la información que corresponde con las estrategias de búsqueda sintetizadas. También se identificó que no siempre se lleva un registro de la trazabilidad de las estrategias de búsqueda implementadas y resultados obtenidos en las bases de datos, la herramienta que se propone más adelante apoya la síntesis de reportes estructurados con esta información. Así mismo, facilita la extracción de información relevante de los estudios clínicos, extrayendo los términos relacionados con

paciente, intervención y resultados de forma automática. Por medio del reporte que se genera a través de la herramienta de síntesis es posible identificar la tecnología objeto de evaluación, comparador, resultados a medir e información relacionada con los pacientes. Contar con este registro incrementa la confiabilidad de las recomendaciones realizadas y la posibilidad de reproducir la revisión en otro momento que se pueda llegar a necesitar.

Por medio de la implementación de los ítems sintetizados anteriormente es posible contribuir a tener procesos de ETS que garanticen decisiones de incorporación de tecnología con base en las evidencias de seguridad disponibles para garantizar la efectividad y seguridad de la misma. Esto potencia la gestión de tecnología en las instituciones y permite alcanzar el cumplimiento de los estándares de la acreditación en salud relacionados.

En la Tabla 10 se ilustran las principales semejanzas y diferencias identificadas en los procesos de evaluación de tecnología de los hospitales analizados.

Tabla 10. Principales semejanzas y diferencias identificadas en los procesos de ETS

Semejanzas	Diferencias
Procesos de evaluación realizados por equipos de trabajo interdisciplinarios	Tiempos de ejecución heterogéneos y en un caso particular no se lleva registro del tiempo empleado para la evaluación
Cada ETS parte de una solicitud detallada	Estructura completa del proceso de evaluación
Se identifican los pacientes, tecnología, resultados clínicos a evaluar y comparador como etapa inicial del proceso **	No en todas las instituciones se lleva registro de la trazabilidad de la revisión de literatura **
Se describen los principales indicadores o procesos a mejorar con la incorporación de la nueva tecnología en el hospital	
ETS ampliamente implementada para la evaluación de medicamentos y en menor medida en dispositivos médicos	
No se realiza evaluación del impacto de las recomendaciones realizadas en casa informe de ETS	
Consulta de bases de datos como PubMed para la síntesis de evidencia **	
Consulta de alertas en INVIMA y FDA	

**** Por medio de la herramienta propuesta se pretende impactar en estos aspectos del proceso de síntesis de evidencia**

En la Tabla 10 se ilustran los principales insumos que se obtuvieron de esta caracterización para implementar en la herramienta de síntesis semiautomática de evidencia clínica. Dentro de las cuales se destacan: la identificación de Pacientes, Intervención o Tecnología, Comparador y Resultados a analizar en la ETS, también la consulta de la base de datos PubMed como principal fuente de información en todas las instituciones y el mejoramiento de la trazabilidad del proceso de ETS.

4.2. Base de datos

Una vez caracterizado el proceso de evaluación con base en la metodología mini-ETS, se procedió a levantar una base de datos de reportes de evaluación de tecnología. El objetivo principal de esta base de datos es tener un punto de partida para replicar los procesos de síntesis de evidencia por medio de la herramienta que se diseñe. La información que se consolida a través de estos reportes permite el entrenamiento de algoritmos de clasificación y priorización de estudios clínicos usados para hacer recomendaciones en dichos reportes. Finalmente, dichos algoritmos se implementarán en la herramienta con el fin de que el usuario pueda obtener una clasificación, priorización y extracción semiautomática de información. La metodología implementada para hacer el levantamiento de esta base de datos se sintetiza en la Figura 17.

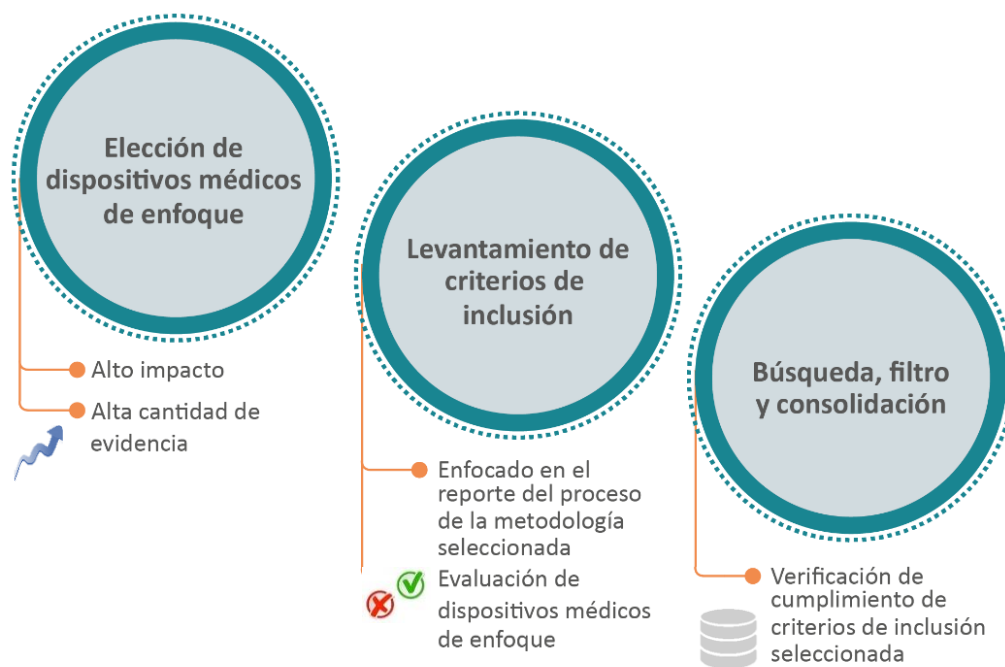


Figura 17. Metodología para el levantamiento de base de datos de reportes de ETS

Como primer paso con el objetivo de realmente lograr levantar una base de datos de un tamaño que permitiera el entrenamiento de los algoritmos de clasificación y acotar el alcance de este trabajo se buscó focalizar la base de datos en reportes de ETS de dispositivos médicos específicos. Las características de estos fueron:

- ✓ Dispositivos médicos de alto impacto en el país y el mundo.
- ✓ Dispositivos que tuvieran una amplia diseminación de evidencia clínica.
- ✓ Que correspondieran con las intervenciones y diagnósticos de las principales causas de morbilidad y mortalidad

Posteriormente se realizó un levantamiento de criterios de inclusión de los reportes de ETS en la base de datos, dentro de los cuales se tuvieron en cuenta los siguientes:

- ✓ Reportes de enfocados en evaluar los dispositivos médicos de enfoque seleccionados en el paso anterior.
- ✓ Reportes que concordaran con la metodología de mini-ETS, seleccionada de acuerdo con el análisis y caracterización de los procesos de evaluación de hospitales realizada para el cumplimiento del primer objetivo de este trabajo de investigación. Para esto se identificaron las características de un reporte de mini-ETS de acuerdo con las recomendaciones realizadas por el *Handbook* de AdHopHTA, DACEHTA y la lista de chequeo de reportes de ETS de INAHTA.

Finalmente, para el proceso de búsqueda, filtro y consolidación se verificó el cumplimiento de los criterios de inclusión.

Luego de aplicar la metodología descrita se obtuvieron los siguientes resultados:

4.2.1. Elección de dispositivos médicos de enfoque

De acuerdo con los criterios de enfoque mencionados anteriormente, se identificó que en Colombia la principal causa de muerte en la población general durante los años 2005 a 2018 fueron las enfermedades del sistema circulatorio. Además, de que se prevee una estabilización de esta causa y un posterior incremento. Esto de acuerdo con el Análisis de la Situación de Salud del país para el año 2020 [86]. Adicionalmente, se identificó que el 17 % de la población en el país se ve afectada por enfermedades coronarias y que en el país se invierte aproximadamente el 7 % del gasto en salud para el tratamiento y diagnóstico de este tipo de enfermedades[10]. Teniendo en cuenta lo anterior, los dispositivos médicos en los cuales se enfocó el desarrollo de este trabajo de investigación fueron aquellos destinados al diagnóstico y tratamiento de las enfermedades del sistema circulatorio, dado que cumplen con los criterios definidos en la metodología descrita.

Después de elegir el enfoque de los dispositivos médicos se realizó una revisión de la literatura por medio de la cual se identificaron los principales padecimientos del sistema cardiovascular y procedimientos llevados a cabo para su diagnóstico y tratamiento. Así mismo, se identificaron algunos de los desenlaces comúnmente evaluados al analizar los resultados del diagnóstico y tratamiento de las diferentes intervenciones. Dentro de la revisión de literatura, lo más relevante fue el reporte de consenso sobre las definiciones de desenlaces en ensayos clínicos sobre el sistema cardiovascular, consolidado por la iniciativa para la recolección de datos estandarizada en ensayos cardiovasculares y la Agencia de medicamentos y Alimentos estadounidense (FDA por sus siglas en inglés) [48]. A partir de esta revisión se identificaron términos clave para guiar la búsqueda de reportes de evaluación de tecnología relacionados con los mismos, los cuales se resumen en la Tabla 11. Estos términos se clasificaron en tres componentes: Intervención, tecnología y resultados, se resumieron en inglés debido a que la mayoría de información se encuentra reportada en este idioma y la síntesis de evidencia a realizar por medio de la herramienta objeto de este trabajo se planteó en inglés.

Tabla 11. Intervenciones y tecnología empleada en el diagnóstico y tratamiento de enfermedades cardiovasculares. Principales resultados analizados

Intervención	Resultados	Tecnologías
Coronary angiography	surgical mortality rates	Radiographic Fluoroscopic Systems
Peripheral angiography	rate of surgical complications	Scanning Systems
Balloon angioplasty	hospital acquired infections	Ultrasonic Cardiac Intravascular
Stent implantation	Clinical Cardiovascular Outcomes	Magnetic Resonance imaging
Coronary artery bypass grafting (CABG)	Quality adjusted life years	computed tomography scanners
Cardiothoracic Surgery	sensitivity	Scanning Systems Gamma Camera
Cardiovascular surgery	specificity	positron emission tomography scan
Cardiac ablation	positive and negative predicative values (PPV, NPV)	Surgical robotic systems
Open heart surgery	likelihood ratio	Electrocardiographs
Heart valve repair	the area under the ROC curve (AUC)	Electrophysiology monitoring system
Heart valve repair replacement	Youden's index	Angiography Interventional
	diagnostic odds ratio (DOR)	

4.2.2. Levantamiento de criterios de inclusión

Adicional al enfoque de los reportes de evaluación de tecnología en los dispositivos médicos seleccionados, se hizo un levantamiento de criterios de inclusión que debían cumplir estos reportes, principalmente en el desarrollo metodológico de la búsqueda y recuperación de evidencia para hacer recomendaciones. Estos criterios están basados en los formatos de mini-ETS de AdHopHTA, de DACEHTA y lista de chequeo de INAHTA. Estos criterios de inclusión se resumen en la Tabla 12. Estos formatos son los correspondientes a la mini-ETS, dado que es la metodología de evaluación seleccionada en el primer objetivo de este trabajo.

Tabla 12. Criterios de inclusión reportes de evaluación

Criterios
Incluir un resumen de la evaluación/estudio planteado
Incluir la pregunta de investigación planteada
Delimitar claramente paciente, intervención, comparador y resultados (PICO) o tecnología, propósito de uso, comparador y resultados (TICO)
Especificar detalles de la búsqueda de literatura realizada como: <ul style="list-style-type: none">• Bases de datos consultadas• Indicar las restricciones idiomáticas• Indicar términos clave o estrategias de búsqueda implementada• Listado de estudios incluidos• Criterios de inclusión y exclusión explícitos
Deben estar enfocados en la evaluación de los dispositivos médicos de interés

4.2.3. Búsqueda, filtro y consolidación

A partir de una búsqueda de agencias de ETS alrededor del mundo se consolidó un banco de reportes de evaluación de tecnología en salud relacionados con los términos sintetizados en la Tabla 11. Los reportes incluidos fueron 49 y se listan a continuación:

- ✓ Use of Biventricular Pacing in Atrioventricular Heart Block [87]
- ✓ Left Ventricular Assist Devices for Destination Therapy [88]
- ✓ Intravascular Ultrasound [89]
- ✓ Magnetic Resonance Imaging for Patients with Implantable Cardiac Devices: A Review of Safety and Guidelines [90]
- ✓ Magnetic Resonance Imaging (MRI) for the Assessment of Myocardial Viability [91]
- ✓ Positron Emission Tomography for the Assessment of Myocardial Viability [92]
- ✓ Positron emission tomography for the assessment of myocardial viability [93]
- ✓ Transcatheter Aortic Valve Implantation (TAVI) for Treatment of Aortic Valve Stenosis: An Evidence-Based Analysis [94]
- ✓ Percutaneous Ventricular Assist Devices: A Health Technology Assessment [95]
- ✓ Multidetector Computed Tomography for Coronary Artery Disease Screening in Asymptomatic Populations [96]

- ✓ Transcatheter Aortic Valve Implantation in Patients With Severe, Symptomatic Aortic Valve Stenosis at Intermediate Surgical Risk: A Health Technology Assessment [97]
- ✓ Comparative Effectiveness of Radiofrequency Catheter Ablation for Atrial Fibrillation [98]
- ✓ Computed Tomography Angiography for Diagnosis of Stroke or Transient Ischemic Attack: A Review of the Clinical Effectiveness [99]
- ✓ Use of Cardiac Resynchronization Therapy [100]
- ✓ Cardiac Resynchronization Therapy in Heart Failure [101]
- ✓ Magnetic resonance imaging of patients with suspected (non-ischaemic) dilated cardiomyopathies [102]
- ✓ Coronary Computed Tomographic Angiography for Detection of Coronary Artery Disease [103]
- ✓ Ablation for Atrial Fibrillation [104]
- ✓ Real time measurement of cardiac output and other cardiac flow parameters (without concurrent cardiac imaging) using continuous wave Doppler techniques [105]
- ✓ The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation [106]
- ✓ Transcatheter aortic valve implantation (TAVI) for the treatment of patients with severe symptomatic aortic stenosis who are at intermediate surgical risk [107]
- ✓ Transthoracic Echocardiography in Adult Patients with Ischaemic Stroke: A Review of the Diagnostic Yield and Cost Effectiveness [108]
- ✓ Computed Tomography Angiography versus Computed Tomography for the Diagnosis and Management of Hyperacute Stroke: A Review of Comparative Clinical Evidence and Guidelines [109]
- ✓ What is the clinical effectiveness, safety and cost effectiveness of endovascular therapy using mechanical thrombectomy devices for patients with acute ischaemic stroke? [110]
- ✓ Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction [111]
- ✓ Cardiac magnetic resonance imaging for the diagnosis of coronary artery disease [112]
- ✓ Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary syndrome [113]
- ✓ Extracorporeal membrane oxygenation for cardiac indications in adults: A health technology assessment [114]
- ✓ Esophageal Doppler Ultrasound-Based Cardiac Output Monitoring for Real-Time Therapeutic Management of Hospitalized Patients [115]
- ✓ Remote monitoring of implantable cardioverter-defibrillators, cardiac resynchronization therapy and permanent pacemakers: A health technology assessment [116]

- ✓ Intravascular Diagnostic Procedures and Imaging Techniques Versus Angiography Alone in Coronary Artery Stenting: Comparative Effectiveness Review [117]
- ✓ Cardiac Nuclear Imaging [118]
- ✓ Single photon emission computed tomography for the diagnosis of coronary artery disease: An evidence-based analysis [119]
- ✓ Intracoronary radiation therapy [120]
- ✓ Clinical effectiveness and cost- effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation [121]
- ✓ Stenting for Peripheral Artery Disease of the Lower Extremities [122]
- ✓ Positron Emission Tomography for Cardiovascular Disease: A Review of the Clinical Effectiveness [123]
- ✓ Endovascular Repair of Descending Thoracic Aortic Aneurysm: Review of Literature [124]
- ✓ Surgical aortic valve replacement with the ATS Enable[®] sutureless aortic valve for aortic stenosis [125]
- ✓ Post-Discharge Cardiac Monitoring of Stroke Patients: Clinical Effectiveness and Guidelines [126]
- ✓ Transcatheter Aortic Valve Replacement in Severe Aortic Stenosis: A Review of Comparative Durability and Clinical Effectiveness Beyond 12 Months [127]
- ✓ Screening for Cardiovascular Disease Risk With Electrocardiography: An Evidence Review for the U.S. Preventive Services Task Force [128]
- ✓ Stress echocardiography for the diagnosis of coronary artery disease [129]
- ✓ Medical Imaging Prior to Endovascular Therapy for Ischemic Stroke : Clinical Effectiveness and Guidelines [130]
- ✓ Cardiac MRI for myocardial stress perfusion and viability imaging in patients with known or suspected coronary artery disease [131]
- ✓ Insertion, replacement or removal of a cardiac resynchronization therapy device capable of defibrillation (CRT-D) for mild, moderate or severe chronic heart failure [132]
- ✓ Catheter ablation of atrial fibrillation [133]
- ✓ Evaluación de la cirugía guiada por imagen: eficacia, seguridad e impacto económico de la resonancia magnética intraoperatoria [134]

Posteriormente, se hizo una lectura y extracción de información de los reportes de ETS obtenidos, a partir del cual se consolidó una base de datos con la siguiente estructura:

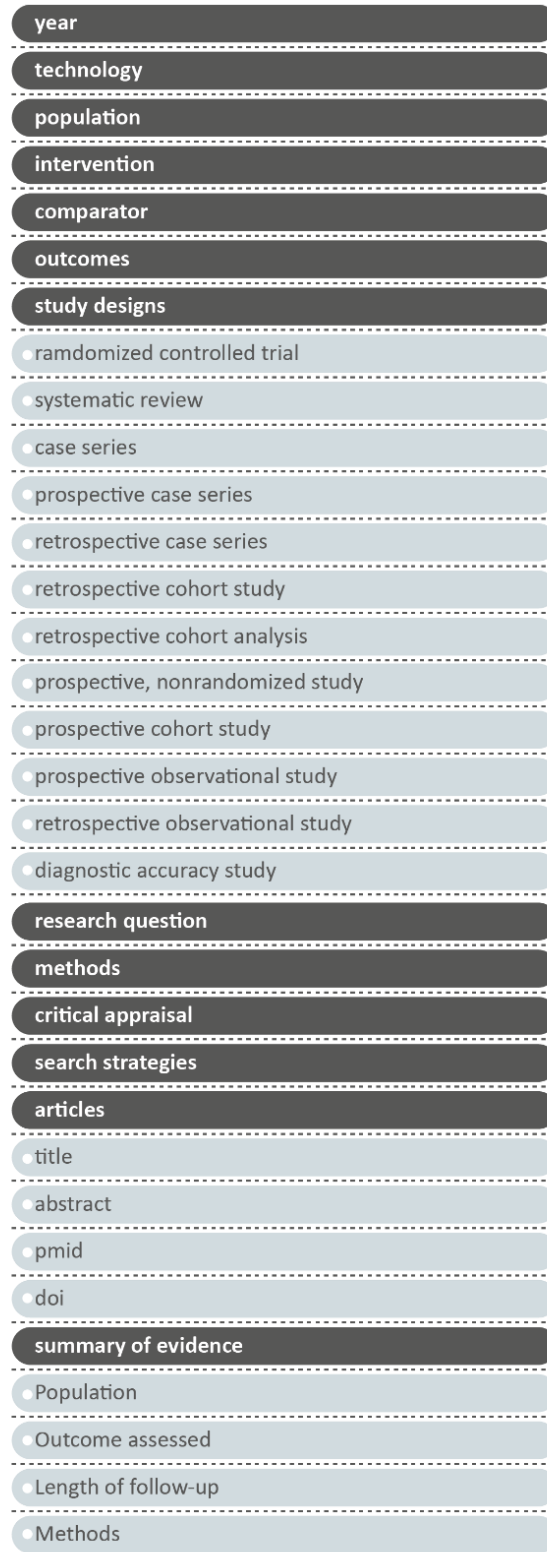


Figura 18. Información reportes de evaluación de tecnología

Para cada reporte de ETS se extrajeron cada uno de los ítems relacionados anteriormente. El ítem de diseño de estudio que contiene 12 tipos de estudios se construyó a partir de la asignación dada a los estudios en cada reporte. Teniendo en cuenta la pirámide de calidad de evidencia, este trabajo se enfoca en los dos primeros tipos de estudios, los cuales son: ensayos clínicos aleatorizados y revisiones sistemáticas, como se ilustra en la Figura 19. Para el ítem de resumen de evidencia, aunque para cada estudio variaba su presentación de un reporte a otro, se conservó la estructura más común reportada. En esta estructura se listaban las características de la población objeto del estudio, tiempo de seguimiento, metodología y los principales resultados obtenidos. En la Tabla 13, se ilustra un ejemplo de uno de los ensayos clínicos (*adverse effect of right ventricular pacing prevented by biventricular pacing during long-term follow-up: a randomized comparison*) incluidos en un reporte de ETS (*Use of Biventricular Pacing in Atrioventricular Heart Block*) y la información que se extrajo para la base de datos objeto de este trabajo de investigación.

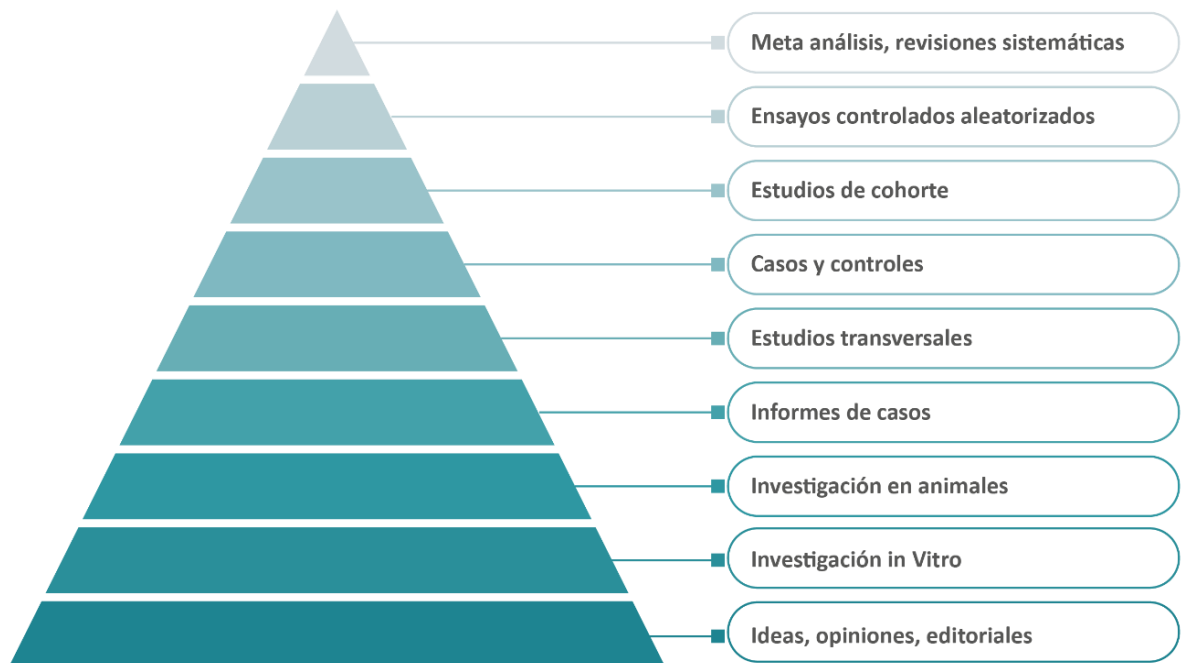


Figura 19. Pirámide de evidencia clínica [135]

Tabla 13. Información derivada del reporte de evaluación de tecnología: "Use of Biventricular Pacing in Atrioventricular Heart Block"

Item	Descripción
Año	2016
Título	Use of Biventricular Pacing in Atrioventricular Heart Block
Tecnología	biventricular pacing
Población	patients with a third-degree heart block
Intervención	use of biventricular pacing in atrioventricular heart block
Comparador	right ventricular pacing
Resultados	primary outcomes were measures of lv function (lvef, lvesv, lvedv) and ventricular dyssynchrony (qrs width, ivmd, lv dyssynchrony index). secondary outcomes included quality of life score and 6-minute walk test. mortality or hospitalization
Diseño de estudios a incluir	randomized controlled trials (rcts), observational studies, systematic reviews and meta-analyses
Pregunta de investigación	not specified
Información general	<p>we carried out a search for relevant randomized controlled trials (rcts), observational studies, systematic reviews and meta-analyses reporting on efficacy, safety and cost-effectiveness for bvp in heart block patients. we also made a search for relevant clinical guidelines. the search for randomized controlled trials and observational studies was limited to the databases maintained by the cochrane library and pubmed. we also searched for randomized controlled trials in progress from clinicaltrials.gov. we carried out a search for health technology assessment (hta) reports in the databases maintained by york university (http://www.york.ac.uk/inst/crd/), by the national institute for health research (nih) (http://www.hta.ac.uk/), and the us agency for healthcare research and quality (ahrq) (http://www.ahrq.gov/research/findings/ta/index.html). the search was conducted by two of the authors (es and ls). the last search was conducted on july 26, 2015. we retained only studies published in english and in adult subjects.</p>
	<p>the quality of the rcts in terms of risk of bias was assessed on the basis of random sequence generation, allocation concealment, blinding of participants/personnel, blinding of outcome assessment, and incomplete outcome data using the cochrane collaboration's tool for assessing risk of bias.¹² each entry for these categories assesses the risk of bias as 'low', 'high', or 'unclear'. we also evaluated if there is potential conflict of interest attributable to sources of funding. the risk of bias in the observational studies was assessed using the newcastle-ottawa scale.¹³ the risks of bias ratings were carried out by two co-authors (ls and na).</p>
	<p>• (biventricular pacing [title/abstract] or cardiac resynchronization therapy [title/abstract] or biventricular pacemaker [title/abstract]) and (heart block [title/abstract] or av lock [title/abstract] or atrioventricular block [title/abstract] or av-block [title/abstract] or bradycardia [title/abstract])</p> <p>• (biventricular pacing [title/abstract] or biventricular pacemaker [title/abstract] or cardiac resynchronization therapy [title/abstract] or biv stimulation [title/abstract]) and (upgrade [title/abstract] or upgrading [title/abstract])</p>
Estudios incluidos*	<p>Título</p> <p>adverse effect of right ventricular pacing prevented by biventricular pacing during long-term follow-up: a randomized comparison</p>

Item	Descripción	
Estudios incluidos*	<p>aims: to investigate whether biventricular (biv) pacing preserves left ventricular ejection fraction (lvef) and reduces lv dyssynchrony when compared with standard dual-chamber right ventricular (rv) pacing in consecutive patients with high-grade atrioventricular block during 3 years of pacing.</p> <p>methods and results: fifty patients were randomized to rv pacing or biv pacing. lvef was measured using three-dimensional echocardiography. tissue doppler imaging was used to quantify lv dyssynchrony in terms of the standard deviation of the time-to-peak velocity (ts-sd). lvef differed significantly between the two groups during 3 years of pacing (anova: p=0.003). lvef in the rv group decreased from 59±5% at baseline to 53±11% (p=0.01), while lvef remained unchanged in the biv group (57±7% at baseline vs. 58±10% (p=0.40). after 3 years of follow-up, we observed no difference in lv dyssynchrony, lv remodelling or measurements of clinical heart failure (n-terminal pro-brain natriuretic peptide, walking test, and new york heart association functional class) between the two groups. however, in the rv group, but not in the biv group, dyssynchrony increased significantly (p=0.005) during follow-up. furthermore, adverse lv remodelling was observed in the rv group with increased systolic volume and thinning of the lv septum.</p> <p>conclusion: biv pacing preserves lvef and minimizes lv dyssynchrony during long-term follow-up. adverse remodelling observed during 3 years of rv pacing was prevented by biv pacing. however, the adverse impact of rv pacing on lv function was not reflected in measures of clinical heart failure.</p>	
	Resumen	
	PMID	21857020
	DOI	10.1093/ejehocard/jer136
Diseño del estudio	randomized controlled trial	1
	systematic review	0
	case series	0
	prospective case series	0
	retrospective case series	0
	retrospective cohort study	0
	retrospective cohort analysis	0
	prospective, nonrandomized study	0
	prospective cohort study	0
	prospective observational study	0
	retrospective observational study	0
	diagnostic accuracy study	0

Item		Descripción
Síntesis de evidencia	population	patients with permanent or paroxysmal high grade av block. n=25. age (years), mean (sd) or median (min, max) 76 (71, 81). sex male, % 68.bvp. n=25. age (years), mean (sd) or median (min, max) 76 (71, 81). sex male, % 68.rvp
	outcome assessed	not specified
	length of follow-up	3 years
	methods	randomized controlled trial

* Para efectos de ejemplo se describe solo un artículo, sin embargo, este reporte de ETS cuenta con 12 artículos asociados. Entre estos se encuentran 11 ensayos clínicos aleatorizados y 1 estudio de corte prospectivo

En total la base de datos cuenta con:

- 49 reportes de evaluación de tecnología en salud
- 556 artículos (523 únicos, se debe a que algunos reportes de evaluación incluyeron los mismos estudios)
 - ✓ 97 revisiones sistemáticas
 - ✓ 143 ensayos clínicos aleatorizados
 - ✓ 283 estudios observacionales y de exactitud diagnóstica

No todos los reportes de evaluación cumplen con todas las características de los criterios de inclusión resumidos en la Tabla 12. Algunos de ellos no describen explícitamente la pregunta de investigación, la estrategia de búsqueda implementada o la metodología para hacer análisis crítico de la evidencia. Sin embargo, para todos los reportes se le logró relacionar los estudios incluidos en cada uno de ellos, así como su respectivo título, PMI, resumen y tipo de diseño del estudio. Esto último es fundamental para la tarea de clasificación que se presenta en los resultados del siguiente objetivo específico.

4.3. Clasificación de estudios a incluir en síntesis de evidencia y extracción de información

En esta sección se describen la metodología planteada y los resultados obtenidos para el cumplimiento del tercer objetivo específico de este trabajo de investigación: *Adaptar los algoritmos de la herramienta RobotReviewer, con el fin de hacer transferencia de aprendizaje y automatizar la extracción información de fuentes de evidencia clínica*. De acuerdo con lo anterior se planteó la metodología que se ilustra en la Figura 20.

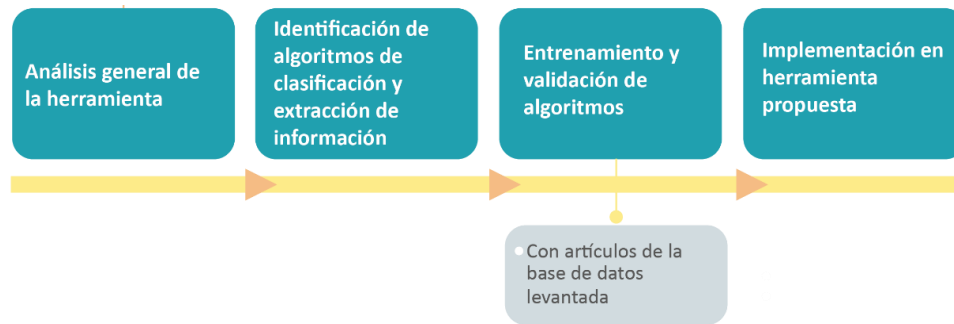


Figura 20. Metodología para la adaptación de algoritmos RobotReviewer

Esta metodología partió de un análisis de la herramienta *RobotReviewer*, que tuvo como fin identificar los algoritmos de aprendizaje de máquina implementados para la clasificación y extracción de información. Este análisis se realizó por medio de la lectura de los artículos producto de estas investigación, análisis de los repositorios de acceso libre que se relacionan en los artículos y referenciación adicional para profundizar en el entendimiento de los resultados que describen los investigadores. Los artículos se listan a continuación:

- ✓ RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials[19].
- ✓ Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide [43].
- ✓ BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [38].
- ✓ SCIBERT: A Pretrained Language Model for Scientific Text [40].
- ✓ Attention Is All You Need[39]
- ✓ Pretraining to Recognize PICO Elements from Randomized Controlled Trial Literature [44].

Posteriormente, se realizó el entrenamiento de algoritmos de clasificación supervisada para ensayos clínicos y revisiones sistemáticas, se propuso una metodología para priorizar los artículos a incluir en la etapa de filtro y se implementó un modelo existente para la de extracción de los términos PICO. Esta clasificación se entrenó a partir los títulos y resumen que hacen parte de la base de datos levantada para el cumplimiento del objetivo específico dos.

Para esta clasificación de tipo de estudios se siguió la metodología que se sintetiza en la Figura 21.

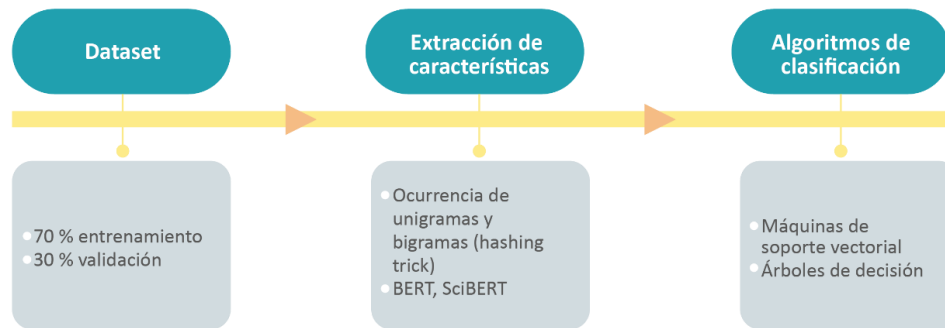


Figura 21. Metodología clasificación de tipo de estudios

De acuerdo con la Figura 21, inicialmente la base de datos se dividió en un 70% de datos para entrenamiento y un 30% para validación, esto para cada tipo de estudio, ensayos clínicos aleatorizados y revisiones sistemáticas.

Posteriormente se planteó una extracción de características de los textos de título y resumen de los estudios por medio de tres técnicas, entre ellas: vectores de ocurrencia de palabras con unigramas y bigramas, *BERT* y *SciBERT embeddings*. Normalmente, una tarea de clasificación basada en métodos de conteo de palabras no es escalable, dado que, el algoritmo dependerá del vocabulario de entrenamiento y cada vez que una nueva palabra sea presentada al sistema, no será reconocida. Por esta razón, se usó una técnica *hashing* para la extracción de características (matrices de ocurrencia de palabras). Este método mapea el token o palabra en un espacio de características de una dimensión determinada, en este caso $n = 2^{20}$. Este método se aplicó por medio de la librería *Scikit-learn* en *Python*, a través del módulo de extracción de características de texto con *HashingVectorizer* y se hizo remoción de palabras vacías. En todos los casos los vectores fueron normalizados utilizando la norma '*L2*', la función de pérdida cuadrática de *Hinge*.

Con relación a la extracción de características de *BERT* y *SciBERT*, esta se aplicó por medio de la herramienta *WEBERT* [136] diseñada para *Python*, con el objetivo de extraer características dinámicas y estáticas de *Bidirectional Encoder Representations from Transformers (BERT)*. Para efectos de esta clasificación se usaron las características estáticas, las cuales son el promedio, desviación estándar, kurtosis, mínimo y máximo de cada neurona (768 en total), es decir, que el espacio de características está determinado por un total de $n = 4608$.

Una vez se extraen las características, se utilizaron dos algoritmos de clasificación los cuales fueron máquinas de soporte vectorial y árboles de decisión. Para el entrenamiento de estos algoritmos se usó un método de rejilla para identificar los mejores los hiperparámetros correspondientes a cada uno de los algoritmos. Este método de ajuste de

parámetros fue *GridSearch* de la librería *Sklearn*, que se enfoca en identificar la mejor combinación de parámetros de acuerdo a las métricas de evaluación determinadas.

Para la clasificación bajo **máquinas de soporte vectorial** optimizaron los hiperparámetros *C* y *gamma* (para kernel no lineal). En los **árboles de decisión** se optimizaron la profundidad, número de muestras requeridas para dividirse y mínimo número de muestras requerido para ser un nodo

Las métricas de desempeño analizadas fueron la sensibilidad, precisión, especificidad y área bajo la curva ROC. Los algoritmos fueron implementados en Python.

Una vez se finalizó esta etapa se implementaron dichos algoritmos en la herramienta propuesta para la síntesis semiautomática de evidencia clínica. Los resultados que se obtuvieron de la implementación de esta metodología se ilustran a continuación, en función de cada una de las cuatro grandes etapas sintetizadas en la Figura 20.

4.3.1. *Análisis general de la herramienta RobotReviewer*

Como resultado de esta etapa de la metodología se sintetizó el análisis se despliega en el marco teórico descrito anteriormente, que principalmente se enfocó en identificar los algoritmos usados por la herramienta *RobotReviewer* en su procesamiento de ensayos clínicos (identificación de tipo de estudio y predicción de riesgo de sesgo). Estos resultados se describen en la sección 3.7 *Análisis RobotReviewer* de este documento. Este análisis abarcó los algoritmos de aprendizaje y procesos de PLN implementados, así como una descripción de BERT y SciBERT como modelos que marcaron el estado del arte en el PLN en todos los ámbitos y también en el dominio biomédico.

4.3.2. *Identificación de algoritmos de clasificación y extracción de información*

Como se ilustró anteriormente, en todas las instituciones la revisión de evidencia clínica se guía por la mejor calidad de evidencia disponible. La escala de calidad de evidencia en revisiones sistemáticas se resume en la Figura 19. De acuerdo con esto, la recuperación de información en la herramienta que se propone en este trabajo se rige por el mismo principio, priorizando revisiones sistemáticas y ensayos clínicos aleatorizados. Es importante mencionar que, identificar ETS de otras instituciones o países que se ajusten a criterios similares de la tecnología y población a evaluar es una buena práctica, ejecutada principalmente por la Clínica Bolivariana. Sin embargo, los reportes de ETS que no se logren identificar en fuentes de información de búsqueda susceptible a automatización no se lograrán recuperar de manera automática.

Con relación a las herramientas para el análisis de sesgo con relación a la evidencia clínica recuperada en las revisiones de la literatura, no se identificó homogeneidad en las instituciones en el uso de alguna en específico. Únicamente, se identificó que el HPTU usa la escala GRADE tanto para valorar la calidad de la evidencia recuperada como para analizar el sesgo de la misma, las otras instituciones no tienen una herramienta determinada.

En general, se identificó que el enfoque de mini-ETS se puede ajustar al diseño de la herramienta de síntesis semiautomática de evidencia clínica para apoyar la ETS en las instituciones de salud de la ciudad. Si bien los procesos caracterizados, las metodologías de

trabajo y las experiencias son diferentes en cada institución, pueden ajustarse a la metodología de trabajo y formato de reportes de mini-ETS reportados en la literatura con adaptaciones al contexto local. Esto se expuso en la sección 4.1. *Caracterización del proceso de evaluación de tecnología en salud en hospitales colombianos.*

Siguiendo la metodología para la ejecución de revisiones sistemáticas, en este caso para la síntesis de evidencia semiautomática, una vez recuperada la información de la base de datos consultada, es pertinente realizar un filtro según título y resumen para identificar los potenciales estudios a incluir en la síntesis. Teniendo en cuenta lo anterior, se expondrá la metodología y resultados obtenidos tanto para clasificar el tipo de estudios de los artículos recuperados como la propuesta de una estrategia de priorización de los mismos.

4.3.2.1. *Clasificación de estudios clínicos*

A partir del análisis anterior se identificó que la herramienta ha usado dos tipos procesamiento de lenguaje natural para hacer su clasificación, dentro de las cuales se encuentran el uso de *SciBERT* como última actualización y el uso de una función hashing. Los algoritmos de clasificación que mejores resultados demostraron fueron las máquinas de soporte vectorial. A partir de esta información se realizó en entrenamiento para clasificar tanto ensayos clínicos como revisiones sistemáticas, es decir, se entrenaron algoritmos donde la extracción de características y clasificación se realizó por medio de los métodos identificados en el análisis de *RobotReviewer*.

4.3.2.2. *Extracción de información*

Con relación a la extracción de información específica de los artículos, por medio de la revisión de literatura transversal a este trabajo de investigación se identificó un modelo de extracción de términos PICO (Paciente/Problema, Intervención, Comparador y Resultados-Outcomes) que se basa en BERT. Este modelo se implementó sin modificaciones en la herramienta de síntesis que se plantea en este trabajo y su análisis se despliega a continuación.

El modelo fue entrenado por Tian Kang y colaboradores [44], a partir de la premisa de que la estrategia PICO es muy importante en la formulación de preguntas de investigación para recuperar evidencia de la literatura y en la práctica de Medicina Basada en Evidencia. Inicialmente los investigadores generaron un corpus a partir de 170 resúmenes de ensayos clínicos de la base de datos Medline y se etiquetaron las palabras o entidades de cada uno de estos a partir de una guía diseñada por un profesional médico y un investigador informático. Cada resumen fue etiquetado al menos dos veces por profesionales diferentes a través de BRAT, una herramienta virtual para el etiquetado de textos [137]. Las etiquetas están divididas en seis grupos: Palabras relacionadas con Pacientes, Intervención (Comparador contenido dentro del grupo de intervenciones), Resultados, Medidas y Calificadores. Un ejemplo de este etiquetado se ilustra en la Figura 22.

Biventricular pacing preserves **left ventricular performance** in patients with **high-grade atrio-ventricular block**; a randomized comparison with **DDD(R) pacing** in 50 consecutive patients.

Aims: We aimed to investigate whether **biventricular (BIV) pacing** minimizes **left ventricular (LV) dyssynchrony** and preserves **LV ejection fraction (LVEF)** as compared with **standard dual-chamber DDD(R) pacing** in consecutive patients with **high-grade atrio-ventricular (AV) block**.

Methods and results: Fifty patients were randomized to **DDD(R) pacing or BIV pacing**. **LVEF** was measured using three-dimensional echocardiography.

Tissue-Doppler imaging was used to quantify **LV dyssynchrony** in terms of **number of segments with delayed longitudinal contraction (DLC)**.

LVEF was not different between groups after 12 months ($P = 0.18$).

In the **DDD(R) group** **LVEF** decreased significantly from 59.7(57.4-61.4)% at baseline to 57.2(52.1-60.6)% at 12 months of follow-up ($P = 0.03$), whereas **LVEF** remained unchanged in the **BIV group** [58.9(47.1-61.7)% at baseline vs. 60.1(55.2-63.3)% after 12 months ($P = 0.15$)].

Figura 22. Etiquetado de resúmenes, reconocimiento de términos PICO. Adaptado de [44]

Este modelo se basa en una red neuronal recurrente (bidireccional Long Short Term Memory - LSTM) como base y una capa de salida Condition Random Field CFR (biLSTM-CRF). Se preentrenó en un corpus de 5000 resúmenes publicado por un conjunto de investigadores en EBM [138], y posteriormente se hizo transferencia de aprendizaje y se ajustó el modelo de reconocimiento PICO con los 170 resúmenes etiquetados. Los resultados obtenidos se ilustran en la Tabla 14. El desempeño se midió comparando el modelo con y sin preentrenamiento y con los métodos de etiquetado Raw y BIO. La principal métrica de desempeño que se tuvo en cuenta para elegir el mejor modelo fue el puntaje F1.

5. Tabla 14. Desempeño modelo de reconocimiento de términos PICO, adaptado de [44]

		No Pre + Raw		No Pre + BIO		Pre + Raw		Pre + BIO	
		Mejor Desempeño	Promedio	Mejor Desempeño	Promedio	Mejor Desempeño	Promedio	Mejor Desempeño	Promedio
Datos para evaluación (Palabras)	Precisión	0.78	0.76	0.84	0.85	0.93	0.87	0.86	0.83
	Sensitividad	0.62	0.63	0.68	0.66	0.80	0.70	0.71	0.7
	Puntaje F1	0.69	0.66	0.75	0.74	0.89	0.78	0.78	0.73
Datos para evaluación (Toquens)	Precisión	0.54	0.52	0.58	0.53	0.74	0.61	0.63	0.63
	Sensitividad	0.53	0.51	0.56	0.52	0.74	0.56	0.64	0.61
	Puntaje F1	0.53	0.52	0.57	0.54	0.74	0.58	0.64	0.62

Dentro de los principales motivos para implementar este modelo de reconocimiento PICO en la herramienta de síntesis de este trabajo son los mismos que se relacionan como contribuciones innovadoras en el trabajo de Tian King y colaboradores que son:

- ✓ Es el primer sistema de PLN disponible públicamente y de acceso libre para el reconocimiento de elementos PICO y sus atributos o medidas en resúmenes de ensayos clínicos.
- ✓ Este modelo logra resultados comparables e incluso mayores para algunas categorías con relación a los demás modelos que se han publicado anteriormente.

Actualmente el sistema cuenta con dos modelos, uno que se explicó anteriormente y el segundo es un modelo basado en *BlueBERT*, este último es el que se utiliza en la herramienta de síntesis, pues como se mencionó anteriormente *BERT* es el estado del arte para las tareas de entendimiento en el PLN. *BlueBERT* es un modelo preentrenado en resúmenes de PubMed y notas clínicas. *BLUE* (Biomedical Language Understanding Evaluation) es un referente para la Evaluación de modelos entrenados para el entendimiento de lenguaje biomédico y se inspiró en *GLUE*. Al evaluar con *BLUE* los modelos basados en *BERT* y *ELMo* se identificó que *BERT* presentó los mejores resultados, las bases de datos, modelos pre entrenados y códigos están disponibles al público [45]. Es importante mencionar que estas herramientas son el resultado de investigaciones realizadas en la rama de la biología computacional del *NLM/NCBI*.

4.3.3. Entrenamiento y validación de algoritmos

4.3.3.1. Clasificación de tipo de estudio

Una vez se realizó la extracción de características por medio de los métodos de PLN expuestos en la metodología, se obtuvieron los valores óptimos de los hiperparámetros de los algoritmos de clasificación por medio de una búsqueda en rejilla o *grid search*. Para esta búsqueda se optimizaron los parámetros haciendo una combinación de las técnicas de

extracción de características planteadas con cada uno de los algoritmos de clasificación. Es decir, se entrenaron máquinas de soporte vectorial con los datos procesados a través de las matrices de coocurrencia de palabras (técnica *hashing*), *BERT* y *SciBert*. Así mismo se realizó para los árboles de decisión, con lo cual fueron en total seis tipos de clasificación.

Los valores en los cuales se variaron los parámetros de las máquinas de soporte vectorial se sintetizan en la Tabla 15, así como los mejores parámetros obtenidos.

Tabla 15. Valores de hiperparámetros optimizados SVM

Máquina de Soporte Vectorial	
Kernel	Linear, radial basis function*, polynomial
C	1, 10, 100*, 1000
Gamma	1e-2*, 1e-3, 1e-4

* Valores con mejor desempeño en la clasificación

Tabla 16. Valores de hiperparámetros optimizados árboles de decisión

Los valores en los cuales se variaron los parámetros de los árboles de decisión se sintetizan en la Tabla 16, así como los mejores parámetros obtenidos.

Decision Tree	
Depth	8*, 10, 12, 16
Samples required to split	2*, 3, 4, 5
Minimum number to be a leaf node	1*, 2, 5, 10, 15

* Valores con mejor desempeño en la clasificación

Teniendo en cuenta lo anterior, se obtuvieron los resultados de clasificación para ensayos clínicos aleatorizados y revisiones sistemáticas como se sintetiza en las Tabla 17 y Tabla 18 respectivamente.

Tabla 17. Resultados clasificación ensayos clínicos aleatorizados

	SVM - BERT	SVM - SciBERT	SVM - Hashing v	DT - BERT	DT - SciBERT	DT - Hashing v
Exactitud (prom ± des)	0.922 ± 0.027	0.861 ± 0.046	0.926 ± 0.023	0.776 ± 0.040	0.897 ± 0.047	0.856 ± 0.036
Precisión (prom ± des)	0.866 ± 0.038	0.841 ± 0.054	0.906 ± 0.041	0.776 ± 0.055	0.881 ± 0.058	0.882 ± 0.064
Sensitividad (prom ± des)	0.951 ± 0.021	0.891 ± 0.044	0.951 ± 0.013	0.774 ± 0.051	0.912 ± 0.049	0.912 ± 0.049
Especificidad (prom ± des)	0.866 ± 0.038	0.841 ± 0.054	0.906 ± 0.041	0.776 ± 0.055	0.881 ± 0.047	0.881 ± 0.047
AUC (prom ± des)	0.972 ± 0.014	0.923 ± 0.013	0.974 ± 0.014	0.776 ± 0.040	0.897 ± 0.047	0.897 ± 0.047

Tabla 18. Resultados clasificación revisiones sistemáticas

	SVM - BERT	SVM - SciBERT	SVM - Hashing v	DT - BERT	DT - SciBERT	DT - Hashing v
Exactitud (prom ± des)	0.934 ± 0.034	0.852 ± 0.026	0.911 ± 0.009	0.757 ± 0.052	0.794 ± 0.049	0.856 ± 0.036
Precisión (prom ± des)	0.905 ± 0.060	0.881 ± 0.043	0.988 ± 0.018	0.751 ± 0.070	0.823 ± 0.076	0.882 ± 0.064
Sensitividad (prom ± des)	0.976 ± 0.019	0.812 ± 0.035	0.829 ± 0.008	0.767 ± 0.077	0.750 ± 0.047	0.822 ± 0.040
Especificidad (prom ± des)	0.905 ± 0.060	0.881 ± 0.043	0.988 ± 0.018	0.751 ± 0.070	0.823 ± 0.076	0.881 ± 0.064
AUC (prom ± des)	0.983 ± 0.011	0.930 ± 0.023	0.977 ± 0.013	0.757 ± 0.052	0.793 ± 0.048	0.855 ± 0.036

Analizando los resultados expuestos en las Tabla 17 y Tabla 18, la combinación de *SVM* y *hashing vectorizer* para la clasificación de ensayos clínicos, fue aquella que mejor resultado dio en cuanto a sensibilidad, especificidad y el valor del área bajo la curva ROC, todas tres por encima de 0.900.

Con relación a la clasificación de revisiones sistemáticas se tiene que las máquinas de soporte vectorial también son el mejor algoritmo de clasificación para las métricas evaluadas. Sin embargo, la extracción de características que mostró mejores resultados para la sensibilidad y AUC fue BERT, ambas por encima de 0.975, mientras que la sensibilidad tuvo mejores resultados con *hashing vectorizer* con un puntaje de 0.988.

Con relación al análisis anterior y en comparación con lo publicado por I. J. Marchall [43] y colaboradores, en su trabajo sobre aprendizaje de máquina para identificar Ensayos clínicos aleatorizados, se puede decir que el desempeño obtenido por el modelo entrenado es realmente bueno. En su estudio, utilizan el dataset Clinical Hedges (49025 artículos publicados entre los años 2000 y 2001), y obtienen un valor AUC de 0.987. Los investigadores afirman que los artículos reportados bajo lineamientos estandarizados y explícitos como lo son las pautas CONSORT (actualizadas en 2010) [47] podrían mejorar el rendimiento de las técnicas de aprendizaje de máquina y los filtros tradicionales. Los resultados obtenidos para el modelo que se presenta en la Tabla 17 y que fue entrenado con 523 artículos, publicados entre 1995 y 2015, indican un valor de AUC de 0.974. Este modelo de clasificación reafirma la hipótesis de I. J. Marchall et. al.[43], dado que se han reproducido resultados similares con significativamente menos muestras.

Es importante mencionar que no solamente se propone un modelo de clasificación de ensayos clínicos aleatorizados sino también de revisiones sistemáticas, estos dos tipos de estudios epidemiológicos son estándar de oro para hacer recomendaciones. Dentro de las revisiones de literatura sobre clasificación automática de estudios no se identificaron estudios que apuntaran a la clasificación automática de revisiones sistemáticas. Además es relevante, dado que un paso importante en una revisión sistemática es identificar si existen revisiones previas sobre el tema en cuestión, como lo plantea el CDR's en su guía para llevar a cabo revisiones en el ámbito de la salud [139].

4.3.3.2. *Priorización de estudios a incluir en síntesis de evidencia*

Generar un modelo de predicción para revisiones sistemáticas que determine los estudios que se deben incluir o excluir de la misma es una labor compleja, teniendo en cuenta que en muchas ocasiones los expertos no filtran exactamente igual un conjunto de artículos en función de una pregunta de investigación específica. Por este motivo además de clasificar los estudios, se plantea establecer un ranking o priorización de los estudios que se recuperen de una búsqueda específica de manera tal que dentro de los n primeros artículos se encuentre la mayor cantidad de información relevante para emitir recomendaciones o simplemente parar la revisión, como se ilustra en la Figura 23.

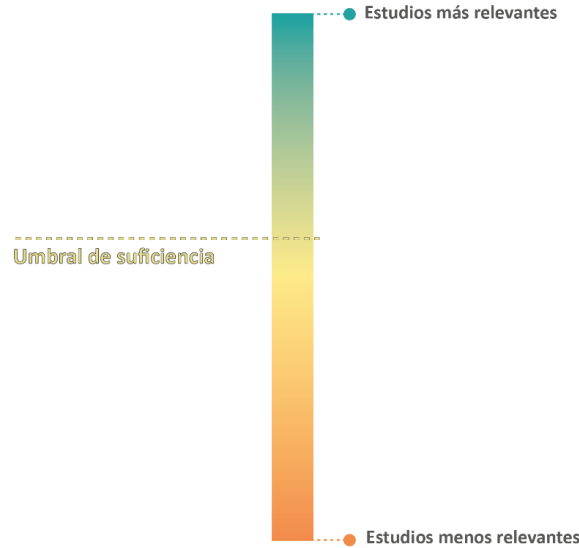


Figura 23. Priorización de estudios

Para este objetivo se reprodujo la estrategia de búsqueda del reporte de evaluación de tecnología que se sintetiza en la Tabla 19.

Tabla 19. Reporte de evaluación de tecnología replicado

Reporte: Use of Biventricular Pacing in Atrioventricular Heart Block	
Pacientes	✓ Patients with a third-degree heart block
Intervención o tecnología	✓ Biventricular pacing
Comparador	✓ Right ventricular pacing
Resultados	<ul style="list-style-type: none"> ✓ Primary outcomes were measures of LV function (LVEF, LVESv, LVEDv) and ventricular dyssynchrony (QRS width, IVMD, LV dyssynchrony index). ✓ Secondary outcomes included quality of life score and 6-minute walk test, mortality or hospitalization.
Estrategia de búsqueda	(Biventricular pacing [Title/Abstract] OR cardiac re-synchronization therapy [Title/Abstract] OR biventricular pacemaker [Title/Abstract]) AND (heart block[Title/Abstract] OR AV block [Title/Abstract] OR atrioventricular block[Title/Abstract] OR AV-block [Title/Abstract] OR bradycardia [Title/Abstract]).
Artículos obtenidos	367 e incrementando
Artículos incluidos	8

Inicialmente se planteó establecer este ranking con base en la medición de la distancia coseno entre la estrategia de búsqueda y cada uno de los artículos recuperados, sin embargo, una vez computada, en la Figura 24 se pudo evidenciar que los estudios incluidos (Clase 1) en la revisión están totalmente embebidos en los no incluidos (Clase 0). Como se observa en la Figura 24, al obtener esta información se reafirmó la hipótesis de que no tiene mucho sentido generar un modelo de clasificación exacta de estudios incluidos y excluidos y se potencia la idea de generar una priorización.

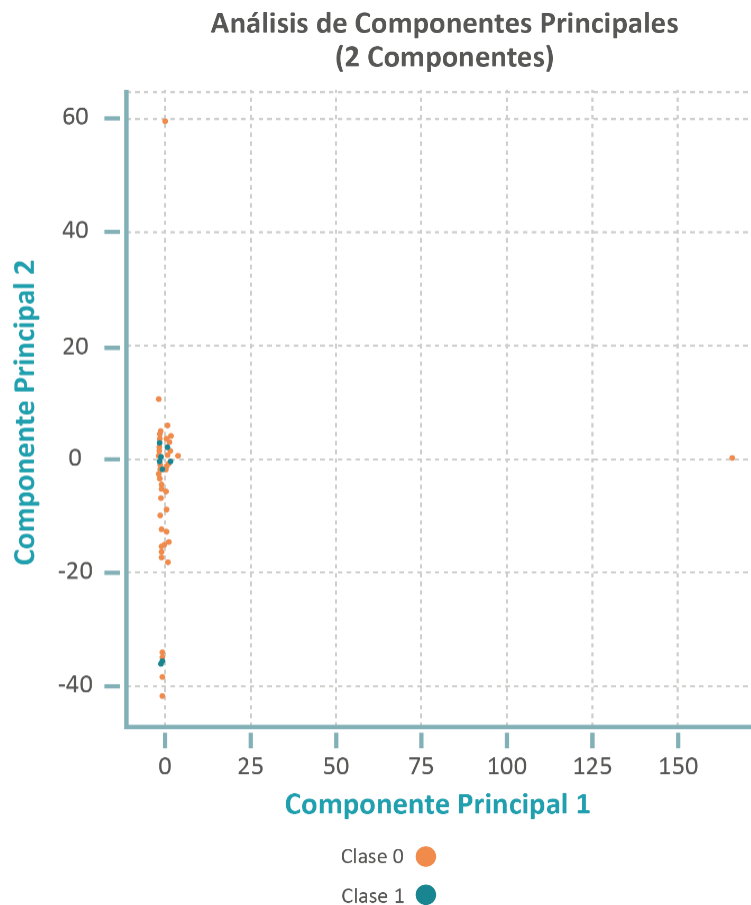


Figura 24. Distancia coseno entre artículos recuperados con estrategia de búsqueda reproducida (Análisis de componentes principales)

Para realizar esta priorización se calcula inicialmente el valor de TF-IDF (frecuencia de término -frecuencia inversa de documento) de acuerdo con la Ecuación 1. Este valor se computó tanto para unigramas como para bigramas. Posteriormente se realiza una suma ponderada de los términos más relevantes que aparecen en cada artículo, dándole a cada término un peso dentro de la suma y ese vector de pesos para la suma es determinado por el valor de TF-IDF para cada término. Así cada artículo es organizado en orden de "importancia" de acuerdo a la aparición de los términos más relevantes en cada uno de ellos, que se refleja en el resultado de dicha suma ponderada. En la Figura 25 se ilustra la "cantidad" de información relevante en un conjunto de artículos tanto para un orden

aleatorio como para un orden establecido de acuerdo con la suma ponderada planteada en tres escenarios diferentes. Estos escenarios son la suma ponderada del índice TF-IDF de los términos y pesos que se sintetizan en la Tabla 20. Se puede observar que en los tres escenarios planteados es posible reducir hasta en un 50% la cantidad de estudios a revisar porque dentro de estos se encuentra el total de artículos incluidos en el reporte de evaluación de tecnología. Se obtuvieron mejores resultados al computar la suma ponderada con unigramas que con bigramas, motivo por el cual, en la herramienta propuesta la priorización de los estudios se realiza con base en bigramas.

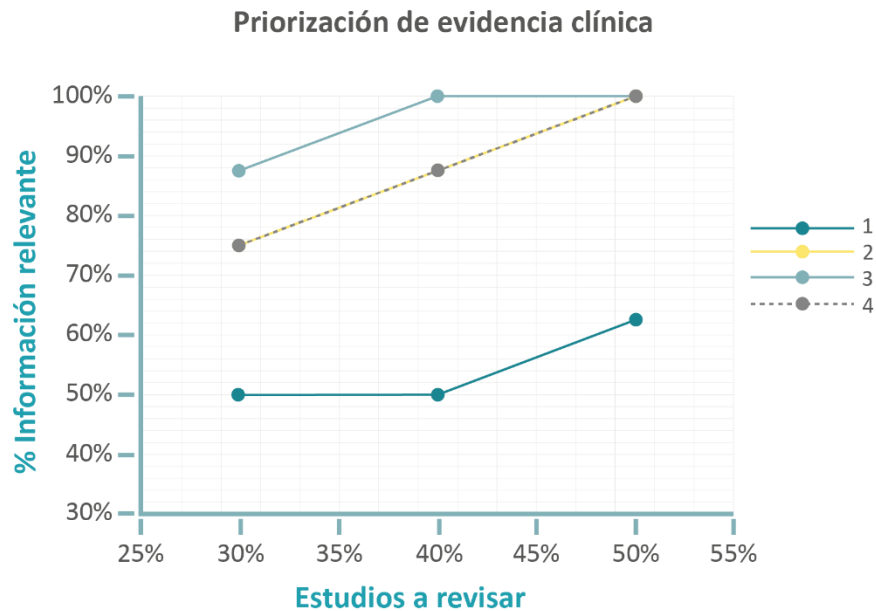


Figura 25. Resultados de priorización de evidencia clínica con base en TF-IDF y términos clave

Cada uno una de las líneas 1,2,3 y 4 que se ilustran en la Figura 25, corresponden con la cantidad de evidencia a identificar al realizar la priorización con base en los términos correspondientes sintetizados en la Tabla 20.

Tabla 20. Parámetros usados para priorización de evidencia

	Key terms	Vector w
1	Random	NA
2	'Biventricular pacing', 'left ventricular', 'ventricular ejection', 'ejection fraction'	[0.25, 0.25, 0.25, 0.25]
3	'biventricular', 'left', 'ejection', 'fraction'	[0.7, 0.05, 0.2, 0.05]
4	'Biventricular pacing', 'left ventricular', 'ventricular ejection', 'ejection fraction'	[0.5, 0.125, 0.25, 0.125]

Si bien las revisiones de literatura tienden a ser un proceso estandarizado y exhaustivo, para la ETS no siempre se desarrolla con el mismo rigor y disponibilidad de evidencia a analizar. Puesto que en cualquier proceso de ETS la evidencia relevante depende de parámetros como pacientes, intervención (tecnología), comparación y resultados, la evidencia para tomar una decisión dependerá de las preguntas de investigación específicas y del juicio de los investigadores. Por lo tanto, como propone G. Kontonatsios [24], la revisión puede detenerse cuando los investigadores así lo decidan, según la baja probabilidad de que los estudios restantes representen información relevante o se haya terminado el tiempo para hacer una recomendación. Adicionalmente, durante la caracterización de los procesos de ETS en los hospitales de la ciudad, se identificó que las revisiones si bien generalmente son sistemáticas, no son exhaustivas en la recopilación de toda la evidencia relacionada con la tecnología de interés. Esto último se corresponde con la sentencia que expone que la revisión de literatura puede parar una vez los investigadores lo decidan, dada la evidencia suficiente para tomar una decisión y/o la baja probabilidad de identificar información útil en los demás estudios.

4.4. Herramienta de integración de búsqueda, extracción de información y síntesis de evidencia clínica

Finalmente, para el diseño de la herramienta de síntesis de evidencia se planteó la metodología que se ilustra en la Figura 26.

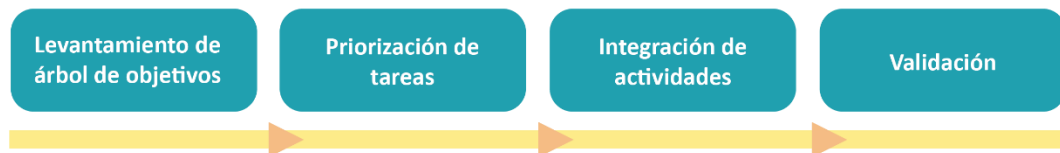


Figura 26. Metodología para el diseño de herramienta de síntesis semiautomática de evidencia clínica

Inicialmente se levantó un árbol de objetivos con el cual se plantearon las etapas de las revisiones sistemáticas de literatura que se impactarían con el diseño de la herramienta. Este planteamiento se realizó con base en los pasos reportados en la literatura para hacer revisiones sistemáticas. Los cuales son:

- ✓ Planteamiento de una pregunta de investigación
- ✓ Diseño de un protocolo para ejecutar la revisión
- ✓ Búsqueda de información
- ✓ Selección de estudios
- ✓ Extracción y síntesis de información
- ✓ Recomendaciones para guiar la toma de decisiones

Todo este proceso debe ir acompañado de una adecuada documentación que permita conservar la trazabilidad de la misma.

Posteriormente se realizó una priorización de las etapas de las revisiones sistemáticas a automatizar, esta priorización se realizó con base en la caracterización del proceso de evaluación de tecnología en salud que se llevó a cabo en la primera parte del desarrollo de este trabajo de investigación. Además de tener en cuenta los lineamientos para realizar un reporte de ETS de la metodología seleccionada que fue la mini-ETS.

Dentro de la metodología planteada también se consideró realizar una validación de la herramienta por medio de una prueba de usabilidad. La usabilidad es un factor muy importante dentro de la calidad de un software [140]. Existen múltiples definiciones para la misma, la ISO/IEC 25010 (Modelos de calidad de sistemas y software), la define como *el grado en el que un software puede ser utilizado por usuarios específicos para lograr objetivos específicos con eficacia, eficiencia y satisfacción en un contexto de uso específico* [141]. Esta norma hace parte de la familia de normas ISO 25000, Evaluación y Requerimientos para Calidad de Sistemas y Software (SQuaRE por sus siglas en inglés) [142]. Una reciente revisión sistemática sobre estudios de usabilidad identificó que en general la mayoría de los estudios evalúa los atributos mencionados anteriormente además de la *facilidad o capacidad de aprendizaje* de uso. Adicionalmente, resalta que dentro de los métodos más usados para evaluar la usabilidad se encuentran los Test de Usabilidad, Evaluación Heurística y los cuestionarios [143].

Con base en la información anterior y siguiendo la propuesta para realizar pruebas de usabilidad descritas por G. Sánchez y colaboradores [144], se diseñó, ejecutó y analizó una prueba de usabilidad de la herramienta propuesta. La prueba se aplicó a 6 usuarios, de acuerdo con [145] y [146], esta cantidad de personas permite identificar tantos problemas de usabilidad como se encontrarían usando muchos más participantes. La caracterización de los mismos se presentará en los resultados de la aplicación de la prueba.

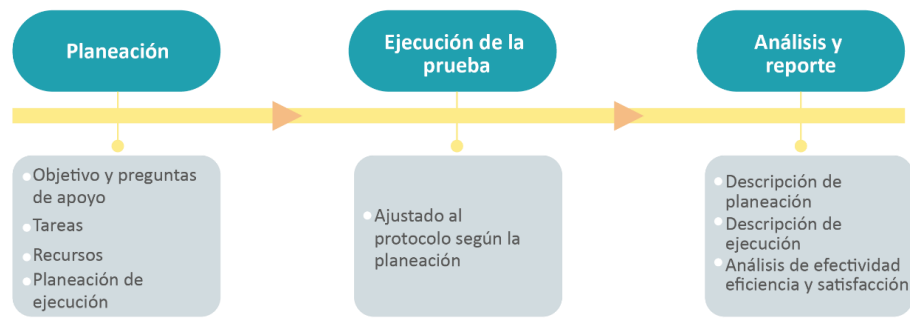


Figura 27. Proceso de pruebas de usabilidad, extraído de [144]

Siguiendo el proceso anterior (Figura 27), se diseñó la prueba de usabilidad que buscó identificar el nivel de usabilidad de la aplicación por medio de preguntas enfocadas en la facilidad de uso, intuitividad, utilidad, satisfacción e información de ayuda (íconos desplegable, videos e instructivo de uso). El cuestionario completo con el cual se llevaron a cabo las pruebas se encuentra anexo a este documento: *7.2 Prueba de usabilidad*.

Una vez se aplicó la metodología descrita para el diseño de la herramienta de síntesis de evidencia clínica se obtuvieron los resultados que se describen a continuación:

4.4.1. Levantamiento de árbol de objetivos y priorización de tareas

Una vez caracterizados los procesos de ETS en las instituciones acreditadas que se describen en la sección *4.1 Caracterización del proceso de evaluación de tecnología en salud en hospitales colombianos*, se definió que las actividades a automatizar para apoyar dichos procesos son la búsqueda, selección o filtro de estudios y una parte de la extracción y síntesis de información de las revisiones sistemáticas. Estas etapas a apoyar se ilustran en la Figura 28. También se realiza una documentación de la estrategia de búsqueda, resultados obtenidos, clasificación, ranking o priorización y filtro de estudios.

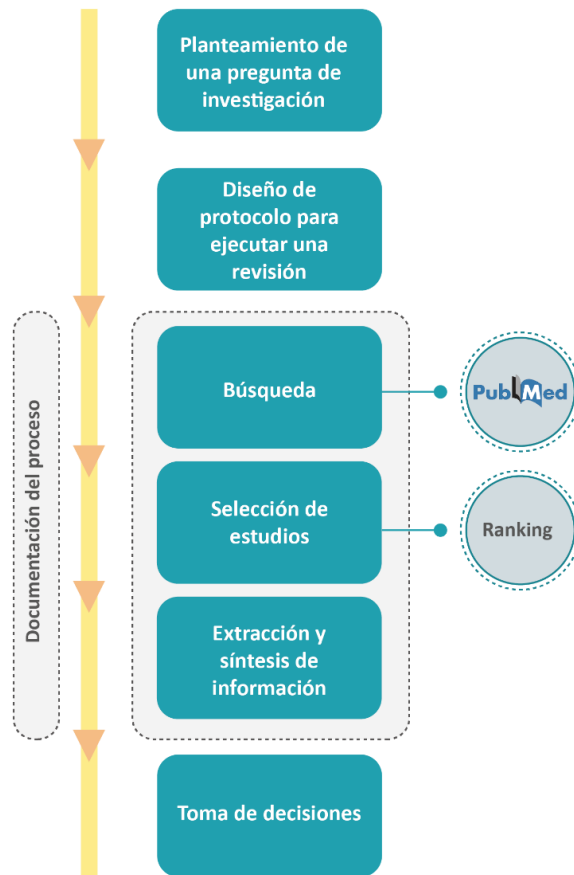


Figura 28. Tareas que se integran en la herramienta de síntesis semiautomática de información

Específicamente las tareas que se desarrollan en la herramienta hasta llegar a la síntesis de información son las siguientes:

- ✓ Generación de estrategia de búsqueda y documentación de la misma
- ✓ Conexión con PubMed y búsqueda de artículos por medio de la estrategia de búsqueda generada
- ✓ Síntesis de resultados de la búsqueda en Excel y archivos de texto por cada artículo
- ✓ Extracción de palabras P(paciente) I(intervención y comparador) O(Resultados – Outcomes) en cada uno de los resúmenes recuperados
- ✓ Síntesis de resultados en pantalla que permite:
 - Visualizar cada artículo con sus términos PICO resaltados
 - Navegar a través de los resultados
 - Priorizar los artículos de acuerdo con las expresiones más relevantes y vector de pesos
 - Clasificación según tipo de evidencia (Ensayos clínicos, revisiones sistemáticas y otros), según etiqueta de la base de datos y cantidad de resultados
 - Seleccionar estudios a incluir en la síntesis

- Generar reportes en pdf con los parámetros de ranking, clasificación de tipo de estudio y artículos asignados y documentación del proceso de búsqueda.

Es importante mencionar que no es del interés de esta herramienta la optimización de estrategias de búsqueda. Esta actividad está dada de acuerdo a la experiencia del investigador y ajuste según los resultados obtenidos con cada una de las estrategias generadas.

4.4.2. Integración de actividades

La herramienta que se desarrolló integró las tareas de:

- ✓ Síntesis de estrategias de búsqueda y búsqueda en PubMed de manera remota
- ✓ Clasificación de tipo de estudios por medio de los algoritmos entrenados durante el desarrollo del tercer objetivo específico de este trabajo
- ✓ Extracción de términos PICO por medio de la implementación del modelo PICO Parser expuesto.
- ✓ Visualización y análisis de información con base en las dos actividades anteriores
- ✓ Documentación que apoya la trazabilidad del proceso de síntesis de evidencia clínica.

Estas tareas fueron elegidas de acuerdo los resultados de la priorización planteada en la actividad anterior de la metodología.

A continuación, se detallan los lenguajes de programación implementados para su desarrollo y cada una de las funcionalidades con las que cuenta la herramienta de síntesis.

4.4.2.1. Lenguajes de programación implementados en el desarrollo de la herramienta de síntesis semiautomática de evidencia clínica.

La implementación de la herramienta propuesta se desarrolló en el lenguaje de programación *Python*, este es un lenguaje multiplataforma y de código abierto que puede utilizarse en diferentes ámbitos como el procesamiento de datos, desarrollo de aplicaciones tanto web como de escritorio, la seguridad informática, entre otros. Dentro de las principales características por las que se eligió es la gran cantidad de documentación que tiene, lo que se puede ver reflejado en que es el lenguaje más popular del mundo como lo demuestran los índices PYPL y TIOBE. PYPL es un indicador creado a partir del análisis de la frecuencia de búsqueda de tutoriales del lenguaje que se hacen en Google [147]. Como se puede observar en la Figura 29, Python ha sido el lenguaje más popular desde el 2018 y viene en una tendencia de incremento desde el año 2000.

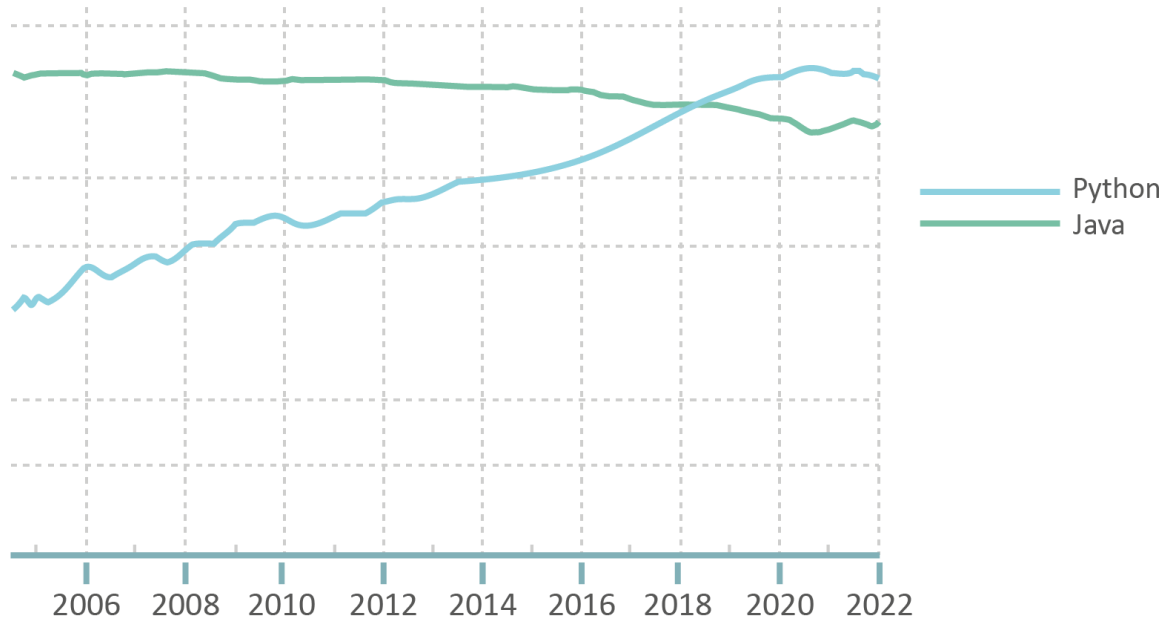


Figura 29. Indicador PYPL de popularidad de lenguajes de programación a nivel mundial, tomado de [147]

El índice TIOBE es denominado el índice de la comunidad de la programación, mide la popularidad de un lenguaje con base en su aparición en las búsquedas en los motores de búsqueda Google, Google Blogs, Yahoo, Baidu, YouTube, MSN y Wikipedia. Este indicador fue creado por la compañía TIOBE, especializada en analizar la calidad de los software [148]. En la Figura 30 se observa que Python también es el más popular en la actualidad bajo este indicador.

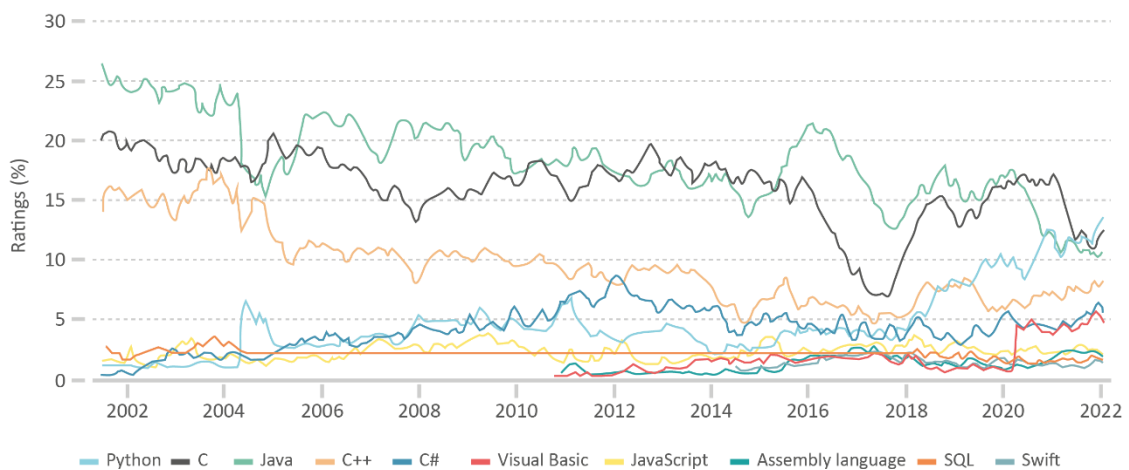


Figura 30. Indicador TIOBE de popularidad de lenguajes de programación a nivel mundial, adaptado de [148]

La interfaz gráfica de la herramienta se implementó por medio de PyQt 5, el cual es un binding o adaptación de Qt Designer al lenguaje de Python [149]. Qt Designer es una herramienta para diseñar interfaces gráficas de usuario (GUIs por sus siglas en inglés) con

Qt Widgets. La posibilidad de crear una interfaz gráfica en la forma what-you-see-is-what-you-get es atractivo, dado que, permite visualizar una interfaz muy similar al resultado que se obtendrá en la ejecución de la misma. Adicionalmente es un entorno de trabajo multiplataforma de código abierto compatible con una amplia variedad de sistemas operativos y hardware. Qt tiene un impacto en más de 70 industrias y es usada por un millón de desarrolladores a nivel mundial [150].

4.4.2.2. *Funcionalidades de la herramienta*

A continuación, se describirá a detalle cada una de estas funcionalidades o tareas y características adicionales que se tuvieron en cuenta para el diseño y la facilidad de uso de la herramienta:

✓ ***Generación de estrategia de búsqueda y documentación de la misma***

Teniendo en cuenta que la fuente de información es la base de datos PubMed, debido a los resultados obtenidos en la caracterización de los procesos de ETS, se diseñó una ventana de búsqueda (Figura 34, Figura 35) que reprodujera el entorno de búsqueda de la base de datos (Figura 31, Figura 32 y Figura 33).

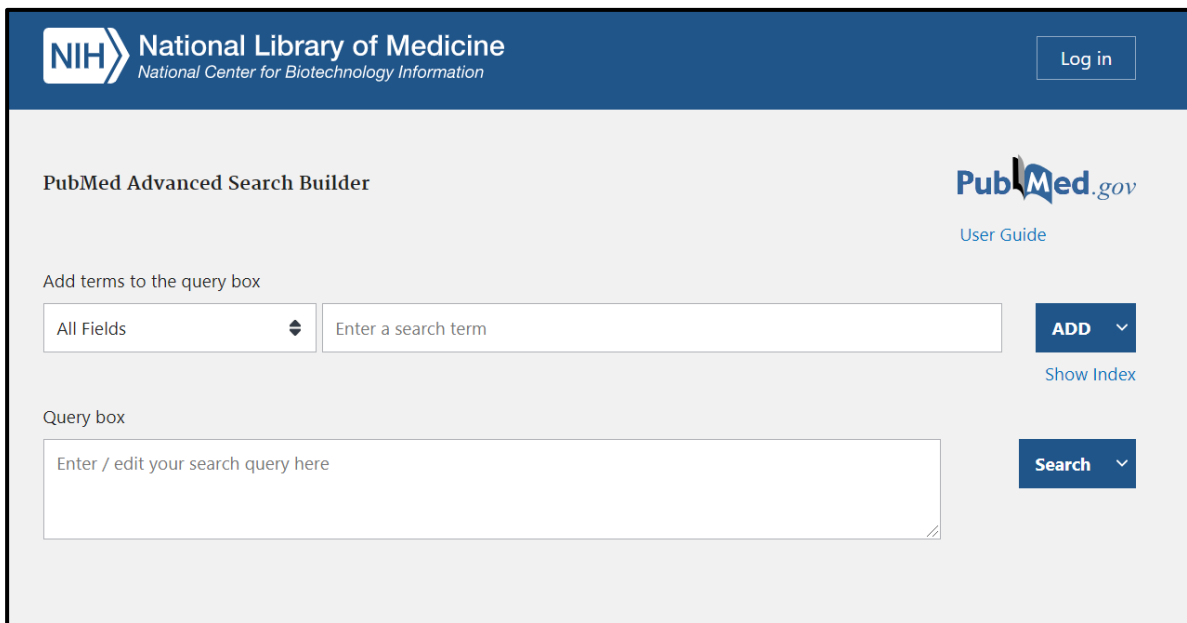


Figura 31. Ayudante de búsqueda avanzada PubMed

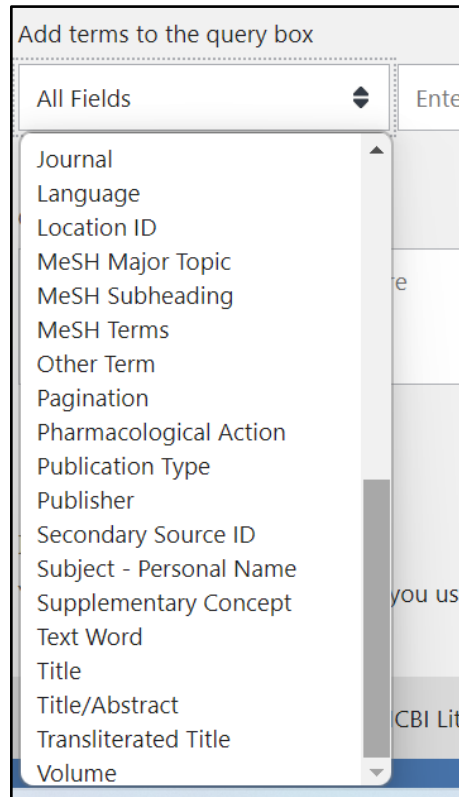


Figura 32. Campo donde se buscarán las palabras de la estrategia sintetizada PubMed

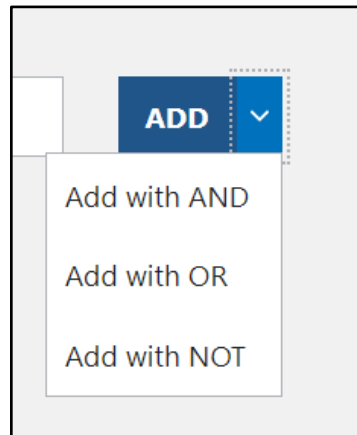


Figura 33. Ítem para incluir, excluir o unir las palabras en la estrategia de búsqueda PubMed

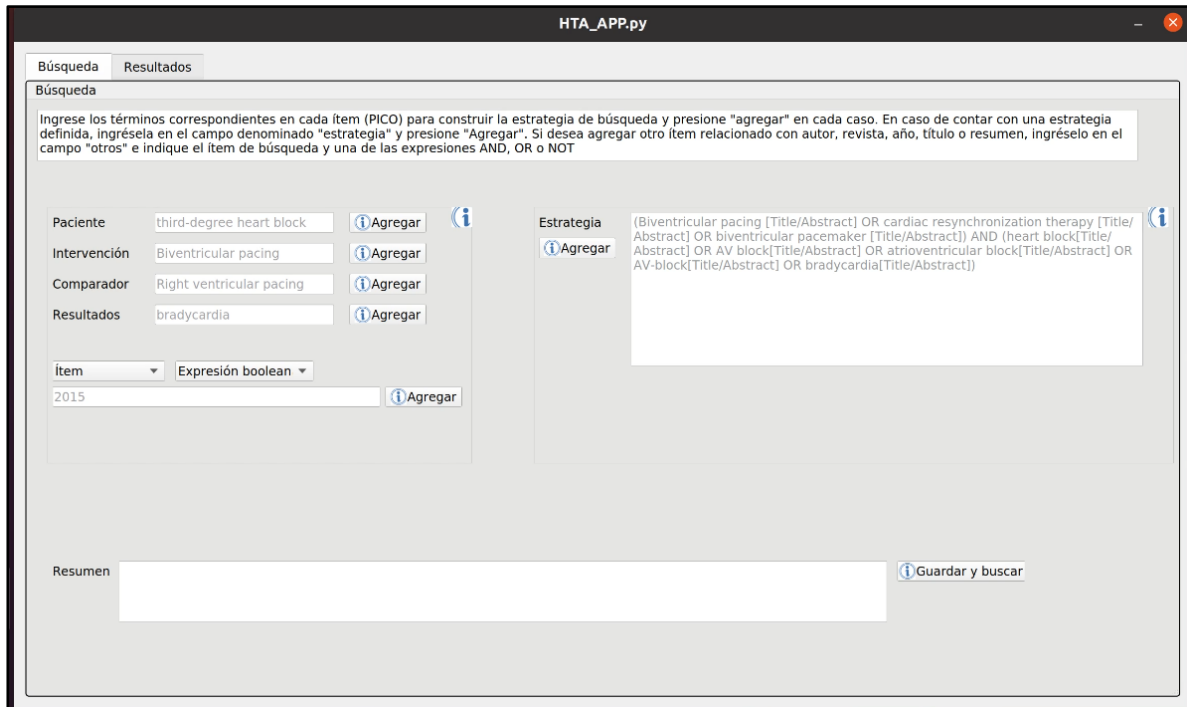


Figura 34. Ventana para síntesis de estrategia de búsqueda

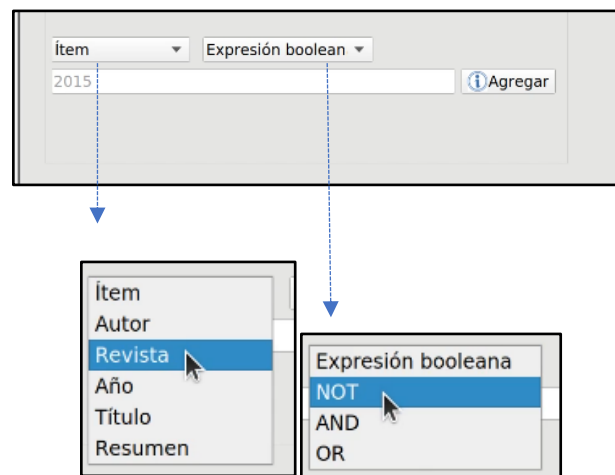
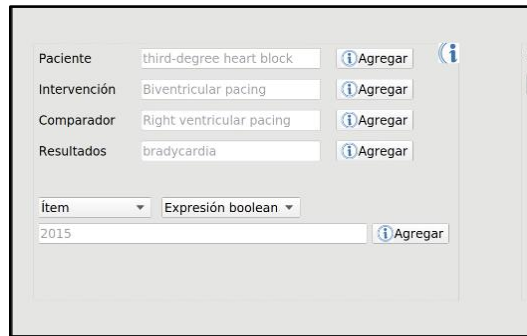


Figura 35. Campo para ingresar términos de búsqueda como autores, revista y año de publicación

Adicionalmente se agregó una ventana para que en caso tal que el usuario tenga una estrategia de búsqueda predeterminada pueda agregarla, guardarla y buscar. Como se puede observar en la Figura 34, ambos procesos se presentan en paralelo con el objetivo de dar al usuario dos opciones para sintetizar la estrategia de búsqueda, cada una de ellas enmarcada en recuadros grises para ubicar mejor la vista del usuario. Cada uno de los recuadros para ingresar palabras a la estrategia de búsqueda se usan de la misma manera (escribir palabra objetivo y presionar botón agregar para ingresarla a la estrategia de búsqueda), la utilidad de la subdivisión en cuatro recuadros es la de guiar al usuario a ingresar los términos PICO de su pregunta de investigación y que no se pase por alto algún ítem importante (Figura 36).



The image shows a search strategy builder interface with the following fields and values:

Field	Value	Action
Paciente	third-degree heart block	Agregar
Intervención	Biventricular pacing	Agregar
Comparador	Right ventricular pacing	Agregar
Resultados	bradycardia	Agregar
Item	2015	Agregar

Figura 36. Recuadros para ingresar términos PICO a estrategia de búsqueda

De la Figura 34 se puede observar que cada uno de los campos para construir estrategias de búsqueda tiene información ejemplo predeterminada, esta información está sintetizada con el fin de ejemplificar al usuario el uso de la herramienta. Siempre que se inicializa la aplicación se podrá visualizar, el usuario puede hacer uso de la misma o modificarla para sintetizar estrategias. Una vez el usuario modifica estas palabras el color del texto cambia a negro, lo cual indica el cambio de información de ejemplo a información de uso real.

✓ **Conexión con PubMed y búsqueda de artículos por medio de la estrategia de búsqueda generada**

Una tarea importante de la revisión de la literatura que demanda tiempo de ejecución y que es susceptible de automatización es la búsqueda automática de información. Como se puede observar en la Tabla 21, durante la caracterización del proceso de ETS se identificó que PubMed y FDA recalls son bases de datos usadas en todos los hospitales encuestados. Cada una de estas bases de datos tienen un enfoque diferente, PubMed y EMBASE son implementadas para identificar estudios clínicos que demuestren la seguridad y efectividad de la tecnología (a nivel general), FDA recalls e INVIMA brindan información acerca de eventos adversos durante el uso de la misma (a nivel específico: marca y modelo de los dispositivos). Por su parte, ECRI brinda información técnica y comparativa de varias tecnologías a nivel mundial, realiza evaluaciones de tecnología, entre otros servicios, sin embargo, es necesario el pago de cada uno de estos, por este motivo no es ampliamente utilizada por las instituciones en el país. De acuerdo con estos resultados, la fuente de información a consultar de manera remota es PubMed. Esta es una base de datos de acceso libre enfocada en la búsqueda y recuperación de información biomédica y ciencias de la salud, cuyo objetivo es mejorar la salud a nivel global. Contiene más de 33 millones de citas y resúmenes de literatura biomédica. Esta base de datos no contiene artículos completos, sin embargo, se puede acceder a los mismos cuando estos están disponibles desde otras fuentes como los sitios de las revistas y *PubMed Central (PMC)* [151]. Esta última es una base de datos de artículos completos la Biblioteca Nacional de Medicina del Instituto Nacional de Salud (NLM/NIH por sus siglas en inglés). Actualmente PMC cuenta con más de 7 millones de textos completos [152].

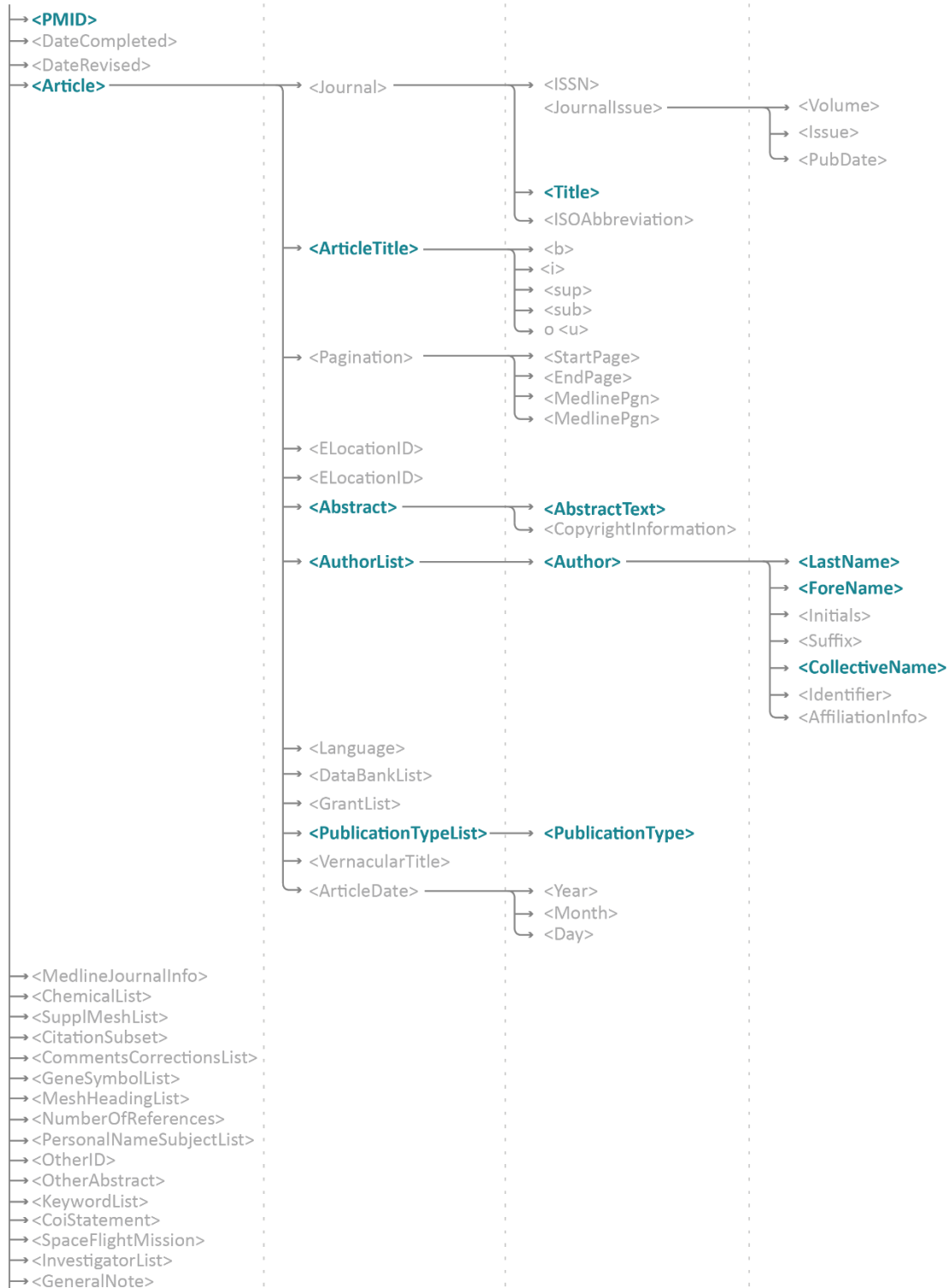
Tabla 21. Uso de bases de datos en hospitales para hacer revisión de literatura en ETS

Base de datos / Institución	HPTU	Clínica Bolivariana	San Vicente Fundación	IPS Universitaria
PUBMED	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
EMBASE	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
FDA	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
ECRI	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
INVIMA	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Sí
 No

Esta automatización depende de características propias de cada fuente de información. Bases de datos como PubMed, Embase y la FDA cuenta con APIs (Interfaz de Programación de Aplicaciones). Para efectos de esta herramienta se utilizó *Entrez Programming Utilities o E-utilities*, la cual es una API pública con acceso al sistema *Entrez* (Sistema de Búsqueda de Base de Datos Cruzada de Consulta Global) del NCBI (Centro Nacional de Información Biotecnológica) que permite acceder a todas las bases de datos de *Entrez* incluyendo PubMed, PMC y otras.

La estructura de los resultados que se obtienen a partir de la consulta es la siguiente:



Los datos que se extraen de cada artículo son los *resultados* en la estructura de datos presentada anteriormente, es decir, PMID, título, resumen, autor(es) y tipo de publicación.

El PMID es un número único de identificación que se le asigna a cada registro en la base de datos de PubMed. Si el usuario de la herramienta desea no buscar por medio de una

estrategia de búsqueda, sino que cuenta con un listado de artículos que desea consultar, puede ingresar los PMID en la ventana de búsqueda y se consultarán dichos artículos.

✓ **Síntesis de resultados de la búsqueda en Excel y archivos de texto por cada artículo**

A partir de los datos obtenidos de la consulta en PubMed, para cada uno de los artículos se escribe un archivo de texto que contiene la información descrita y se genera un archivo de Excel con la misma información, esto con el objetivo de guardar los resultados obtenidos y posteriormente consultarlos desde la interfaz de visualización de la aplicación. Cada vez que se ejecuta una estrategia de búsqueda estos archivos se reescriben con la intención de no llenar espacios de memoria innecesarios en los computadores donde se corra el programa, en caso tal de que el usuario quiera guardarlos, tiene acceso a través de las carpetas con información transitoria.

✓ **Extracción de palabras P(paciente) I(intervención y comparador) O(Resultados – Outcomes) en cada uno de los resúmenes recuperados**

Esto se logra por medio de la conexión con el sistema de PLN para extracción de términos PICO, producto de investigaciones de NLM/NCBI, como se expuso en la sección en 4.3.2.2 Extracción de información. Previo a la extracción de términos, en pantalla se muestra al usuario la cantidad de artículos que se obtuvieron de la búsqueda y si desea proceder con la extracción de información, priorización y clasificación. También existe la opción de no analizar y regresar a ajustar la estrategia de búsqueda en función de obtener más o menos resultados según sea el caso. Esto se puede observar en la Figura 37.

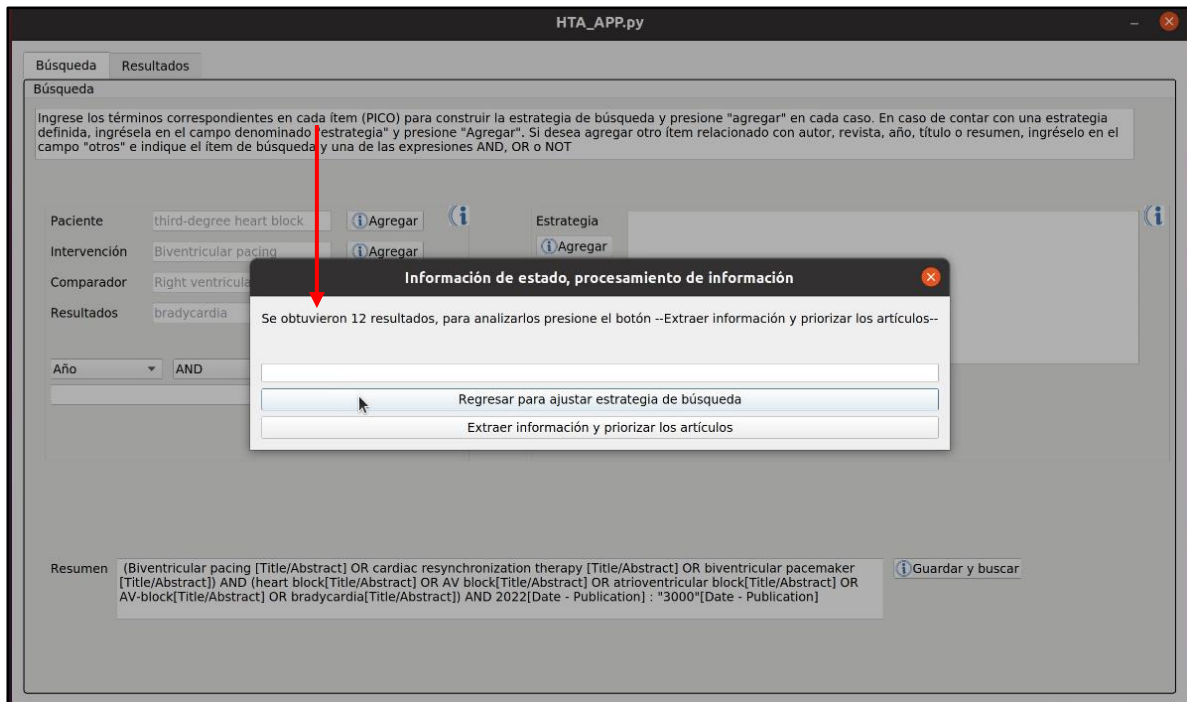


Figura 37. Visualización de la cantidad de resultados obtenidos en la búsqueda

Una vez el usuario procede con el análisis de la información se inicializa una barra de progreso con el objetivo de ilustrar el avance del análisis y que el usuario en todo momento de la ejecución del programa pueda identificar qué está sucediendo. Esto se puede observar en la Figura 38.

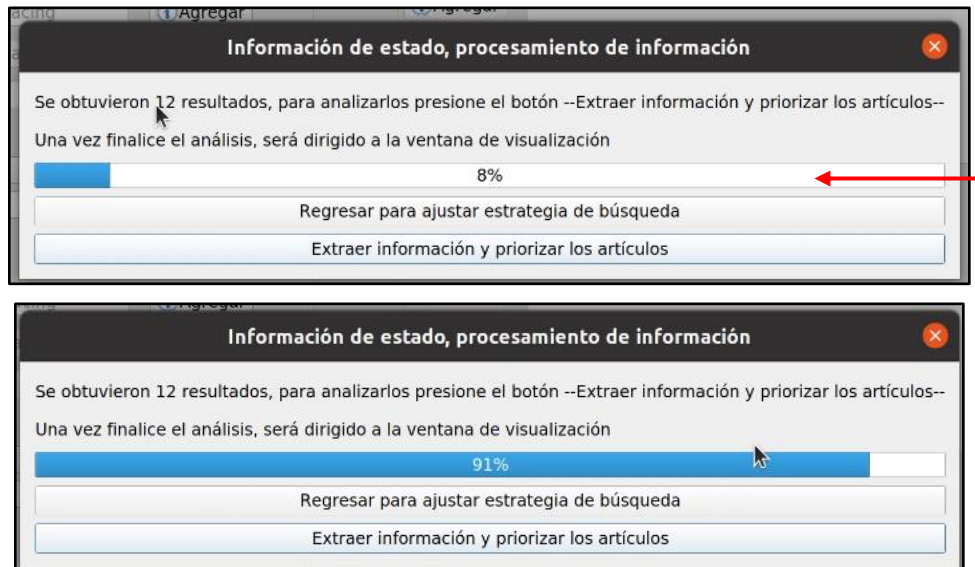


Figura 38. Barra de progreso de análisis de resultados

✓ **Síntesis de resultados en pantalla que permite:**

- **Visualizar cada artículo con sus términos PICO resaltados**

Para cada uno de los artículos encontrados se obtiene la siguiente información: PMID, título, resumen, autor(es), año de publicación, tipo de publicación según PubMed y tipo de publicación de acuerdo con modelo de clasificación entrenado (Ensayos clínicos, revisiones sistemáticas). Para título y resumen se resaltan las palabras que están relacionadas con Paciente, Tecnología y Comparador y Resultados, en la ventana de visualización cada artículo se despliega como el ejemplo que se ilustra a continuación:

Cada color representa un grupo de palabras así:

Paciente

Tecnología (Intervención)

Resultados

Biventricular pacing preserves **left ventricular performance** in patients with **high-grade atrio-ventricular block**: a randomized comparison with **DDD(R) pacing** in 50 consecutive patients.

Aims: We aimed to investigate whether **biventricular (BiV) pacing** minimizes left ventricular (LV) dyssynchrony and preserves **LV ejection fraction (LVEF)** as compared with **standard dual-chamber DDD(R)** pacing in consecutive patients with **high-grade atrio-ventricular (AV) block**. Methods and results: Fifty patients were randomized to DDD(R)

pacing or BiV pacing. LVEF was measured using three-dimensional echocardiography. Tissue-Doppler imaging was used to quantify LV dyssynchrony in terms of number of segments with delayed longitudinal contraction (DLC). LVEF was not different between groups after 12 months (P = 0.18). In the DDD(R) group LVEF decreased significantly from 59.7(57.4-61.4)% at baseline to 57.2(52.1-60.6)% at 12 months of follow-up (P = 0.03), whereas LVEF remained unchanged in the BiV group [58.9(47.1-61.7)% at baseline vs. 60.1(55.2-63.3)% after 12 months (P = 0.15)]. Dyssynchrony was more prominent in the DDD(R) group than in the BiV group at baseline (2.2 +/- 2.2 vs. 1.4 +/- 1.3 segments with DLC per patient, P = 0.10); and at 12 month follow-up (1.8 +/- 1.9 vs. 0.8 +/- 0.9 segments with DLC per patient, P = 0.02). NT-proBNP was unchanged in the DDD(R) group during follow-up (122 +/- 178 pmol/L vs. 91 +/- 166 pmol/L, NS) but decreased significantly in the BiV-group (from 198 +/- 505 pmol/L to 86 +/- 95 pmol/L after 12 months, P = 0.02). Conclusion: BiV pacing minimizes LV dyssynchrony, preserves LV function, and reduces NT-proBNP in contrast to DDD(R) pacing in patients with high-grade AV block.

En la Ventana de visualización se puede observar cómo se ilustra en la Figura 39.

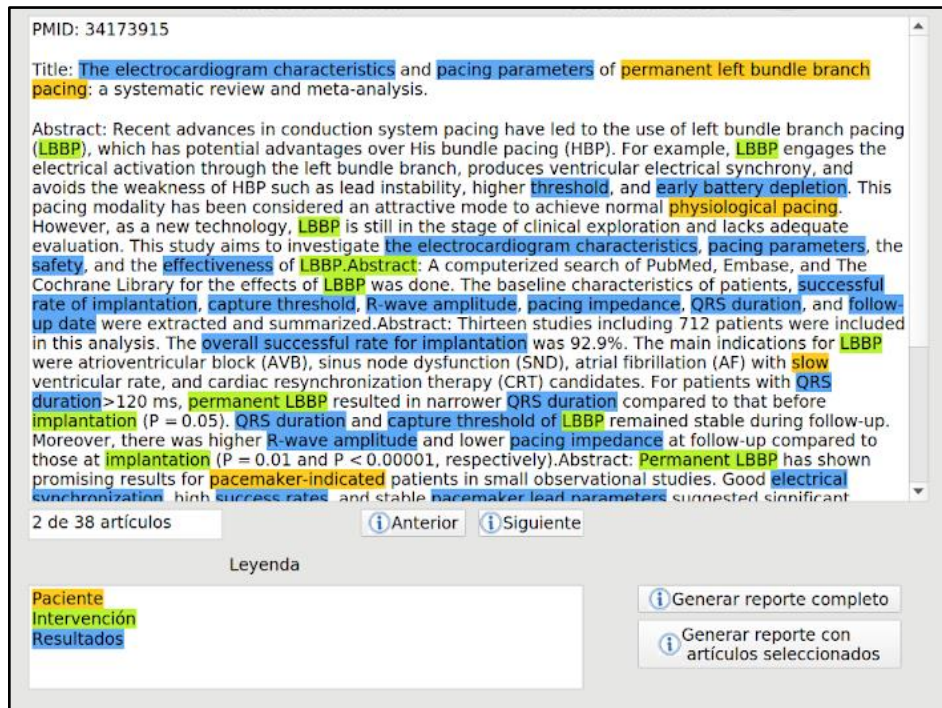


Figura 39. Visualización de artículos con texto resaltado

✓ Navegar a través de los resultados

Cuando se obtienen varios resultados, en la ventana de visualización se ubican dos botones que permiten ir hacia el siguiente y el anterior artículo, cada uno de estos botones tiene un ícono de información en el cual se indica su funcionalidad (Figura 40. Tooltip botones "Anterior" y "siguiente"). Así mismo se cuenta con una etiqueta de estado en la cual se indica la posición del artículo actual dentro del total de resultados. Esta posición no

es aleatoria, sino que corresponde con la priorización que se realiza de los artículos de acuerdo con las palabras relevantes como se expuso en la sección 4.3.3.2. de Priorización.

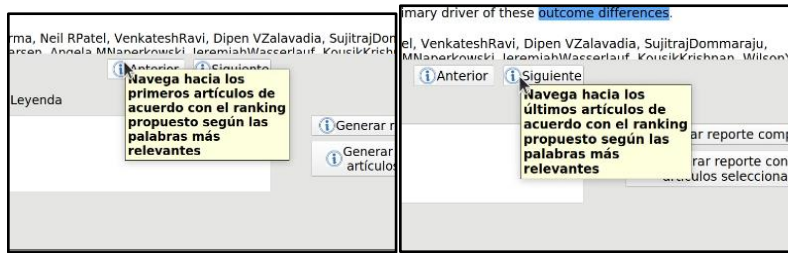


Figura 40. Tooltip botones "Anterior" y "siguiente"

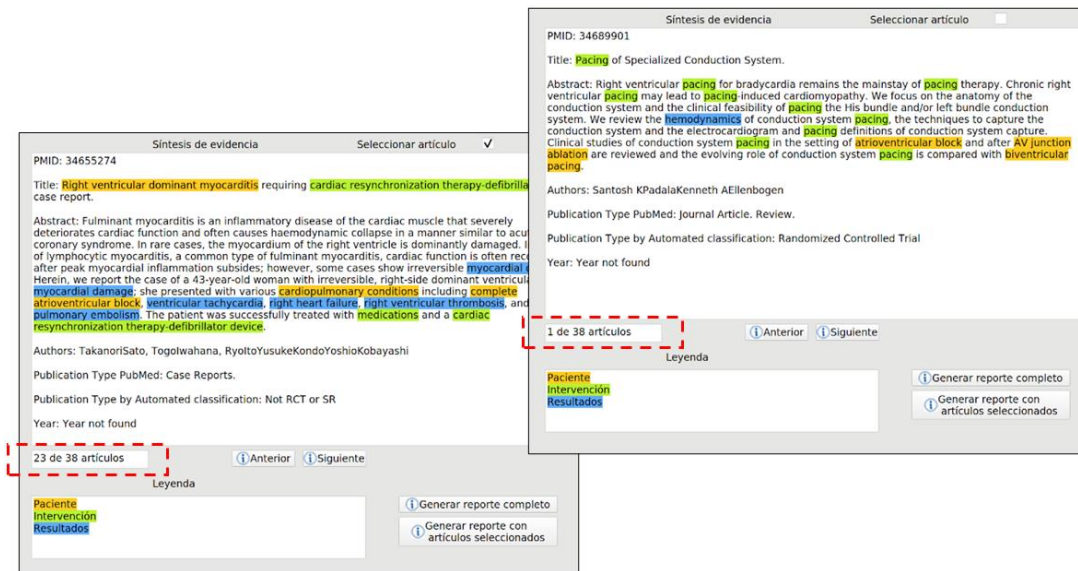


Figura 41. Etiqueta de posición de navegación y priorización de resultados

✓ **Priorizar los artículos de acuerdo con expresiones más relevantes y vector de pesos**

La priorización de los artículos obtenidos se puede observar al navegar a través de los resultados como se explicó en el ítem anterior. Adicionalmente, si el usuario genera el reporte completo, la sección dos de este reporte consiste en el listado de los títulos de los artículos en el orden establecido por la priorización. Así mismo en el lado izquierdo de la ventana de visualización se ubican las palabras y el vector de pesos con el cual se realizó dicho ranking como se observa en la Figura 42.

Así mismo, se tiene un ícono de ayuda que orienta al usuario sobre la interpretación de los parámetros que se sintetizan.

The screenshot shows the HTA_APP.py interface with search results. On the left, there is a search strategy and classification summary. The main area displays a search result for PMID: 34481985, including the title, abstract, and authors. At the bottom, a table shows the ranking parameters for different expressions.

Expresión	Priorización	Importancia
crt		0.31
pacing		0.27
lbbap		0.22
rvp		0.2

Figura 42. Visualización de parámetros de ranking de artículos

✓ **Clasificación según tipo de evidencia (Ensayos clínicos, revisiones sistemáticas y otros), según etiqueta de la base de datos y cantidad de resultados**

Para el total de resultados obtenidos se analizan aquellos que fue posible recuperar título y resumen, en el lado izquierdo de la ventana de visualización (Figura 43) se indican la cantidad de ensayos clínicos y revisiones sistemáticas que se identificaron con el modelo de clasificación entrenado y expuesto en la sección 4.3.3.1 Clasificación de tipo de estudio. Así mismo cada artículo que se ilustra en pantalla cuenta con la información del tipo de estudio según el modelo presentado y la etiqueta de la base de datos, como se señala también en la Figura 43.

The screenshot shows the HTA_APP.py interface with the following sections:

- Búsqueda** / **Resultados** tabs.
- Resumen** section:
 - Estrategia de búsqueda:** (Biventricular pacing [Title/Abstract] OR cardiac resynchronization therapy [Title/Abstract] OR biventricular pacemaker [Title/Abstract] AND (heart block[Title/Abstract] OR AV block[Title/Abstract] OR atrioventricular block[Title/Abstract] OR AV-block[Title/Abstract] OR bradycardia[Title/Abstract]) AND 2022[Date - Publication] : "3000"[Date - Publication])
 - Resultados:** 12
 - Disponibles (título - resumen):** 11
- Clasificación** section (highlighted with a red dashed box):
 - Ensayos clínicos:** 8
 - Revisiones sistemáticas:** 0
 - Otros:** 4
- Priorización** table:

Expresión	Priorización	Importancia
crt		0.31
pacing		0.27
lbbap		0.22
rvp		0.2
- Síntesis de evidencia** section:
 - PMID:** 34481985
 - Title:** Clinical outcomes of left bundle branch area pacing compared to right ventricular pacing: Results from the Geisinger-Rush Conduction System Pacing Registry.
 - Abstract:** Left bundle branch area pacing (LBBAP) has been shown to be a feasible option for patients requiring ventricular pacing. Abstract: The purpose of this study was to compare clinical outcomes between LBBAP and RVP among patients undergoing pacemaker implantation METHODS: This observational registry included patients who underwent pacemaker implantations with LBBAP or RVP for bradycardia indications between April 2018 and October 2020. The primary composite outcome included all-cause mortality, heart failure hospitalization (HFH), or upgrade to biventricular pacing. Secondary outcomes included the composite endpoint among patients with a prespecified burden of ventricular pacing and individual outcomes. Abstract: A total of 703 patients met inclusion criteria (321 LBBAP and 382 RVP). QRS duration during LBBAP was similar to baseline (121 ± 23 ms vs 117 ± 30 ms; P = .302) and was narrower compared to RVP (121 ± 23 ms vs 156 ± 27 ms; P < .001). The primary composite outcome was significantly lower with LBBAP (10.0%) compared to RVP (23.3%) (hazard ratio [HR] 0.46; 95% confidence interval [CI] 0.306-0.695; P < .001). Among patients with ventricular pacing burden >20%, LBBAP was associated with significant reduction in the primary outcome compared to RVP (8.4% vs 26.1%; HR 0.32; 95% CI 0.187-0.540; P < .001). LBBAP was also associated with significant reduction in mortality (7.8% vs 15%; HR 0.59; P = .03) and HFH (3.7% vs 10.5%; HR 0.38; P = .004). Abstract: LBBAP resulted in improved clinical outcomes compared to RVP. Higher burden of ventricular pacing (>20%) was the primary driver of these outcome differences.
 - Authors:** Parikshit SSharma, Neil RPatel, VenkateshRavi, Dipen VZalavadia, SujitrajDommaraju, VarunGarg, Timothy Blarcan, Angela MManarkowski, JeremiahMaccardauf, KovicKrichnan, WilsonYoung
 - 1 de 12 artículos**
 - Leyenda:** Paciente, Intervención, Resultados
 - Botones:** Generar reporte completo, Generar reporte con artículos seleccionados

Figura 43. Clasificación de tipo de estudio según modelo entrenado etiqueta de PubMed

✓ **Seleccionar estudios a incluir en la síntesis**

A medida que el usuario navega a través de los resultados obtenidos puede hacer el filtro de los artículos que desea incluir en una revisión completa de los artículos, esto lo puede hacer haciendo clic en la casilla de verificación que se ubica en la esquina superior derecha de la ventana de visualización de resultados (Figura 44). También es posible deseleccionar el artículo volviendo a hacer clic sobre la casilla, en el reporte de artículos seleccionados solo se incluirán aquellos sobre los que al oprimir el botón de *Generar reporte con artículos seleccionados* tengan en la casilla. Al pasar el cursor por esta casilla también se despliega un ícono de ayuda que indica al usuario la funcionalidad de esta.

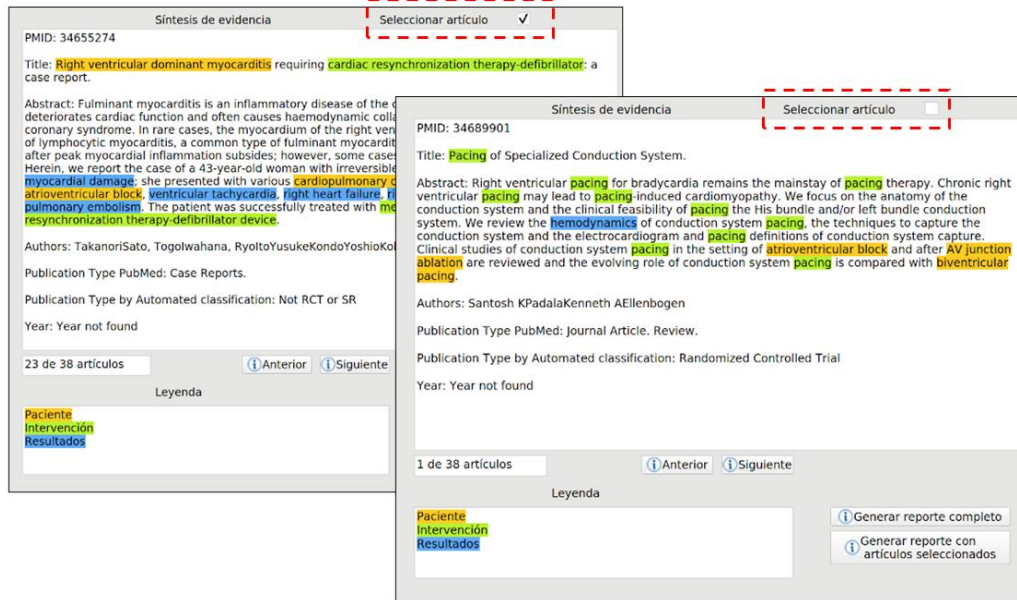


Figura 44. Casilla de verificación para incluir artículos en el reporte de síntesis de evidencia

✓ **Generar reportes en pdf con los parámetros de ranking, clasificación de tipo de estudio y artículos asignados y documentación del proceso de búsqueda.**

Una vez el usuario finaliza el análisis de los resultados puede generar un reporte en el cual se documenta toda la información relevante sobre el mismo. Este reporte se genera de manera automática al presionar alguno de los dos botones de *Generar reporte*. Como se puede observar en la Figura 45 cada uno de los botones también cuenta con ícono de ayuda para el usuario.

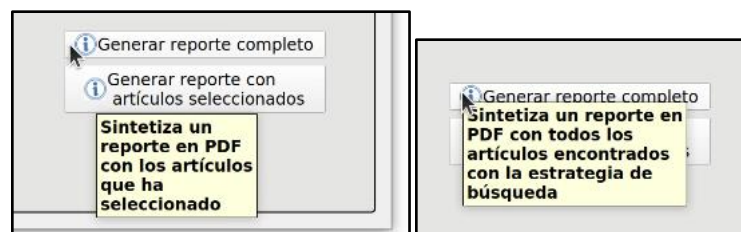


Figura 45. Botones que permiten generación automática de reporte de síntesis

Se pueden generar dos tipos de reporte, uno que cuenta con todos los artículos obtenidos en la búsqueda y otro que cuenta solamente con los artículos que el usuario selecciona dentro del filtro que realiza por medio de la casilla de verificación expuesta anteriormente. Este último se diseñó para agilizar el proceso de filtro desde la navegación de los resultados, y con el objetivo de tener un reporte con los artículos incluidos por el investigador, sin embargo, también puede ser útil para subdividir los resultados y analizarlos por etapas o por diferentes investigadores.

Cada reporte cuenta con tres grandes secciones las cuales se dividieron así: parámetros y estrategia de búsqueda, resumen de resultados obtenidos y resultados. A continuación, se describirá el contenido de cada uno de estos capítulos.

Parámetros y estrategia de búsqueda:

Inicialmente se sintetizan las palabras usadas en la estrategia de búsqueda que están relacionadas con los parámetros de paciente, tecnología, comparador y resultados. Se indica el idioma en el cual se ejecuta la búsqueda y como se mencionó anteriormente esta herramienta está diseñada para el PLN en inglés, por lo tanto, todos los reportes tendrán este lenguaje predeterminado. Se sintetiza la estrategia de búsqueda implementada y el total de artículos encontrados, también se indican la cantidad de artículos para los cuales no fue posible recuperar el resumen. Esto último se debe a que esos artículos en particular requieren un pago para la visualización. Esto último debido a las características de privacidad particulares de las revistas donde están publicados dichos artículos. Finalmente se indican los tipos de estudios identificados según el modelo de clasificación entrenado. Un ejemplo de este capítulo del reporte se ilustra en la Figura 46.

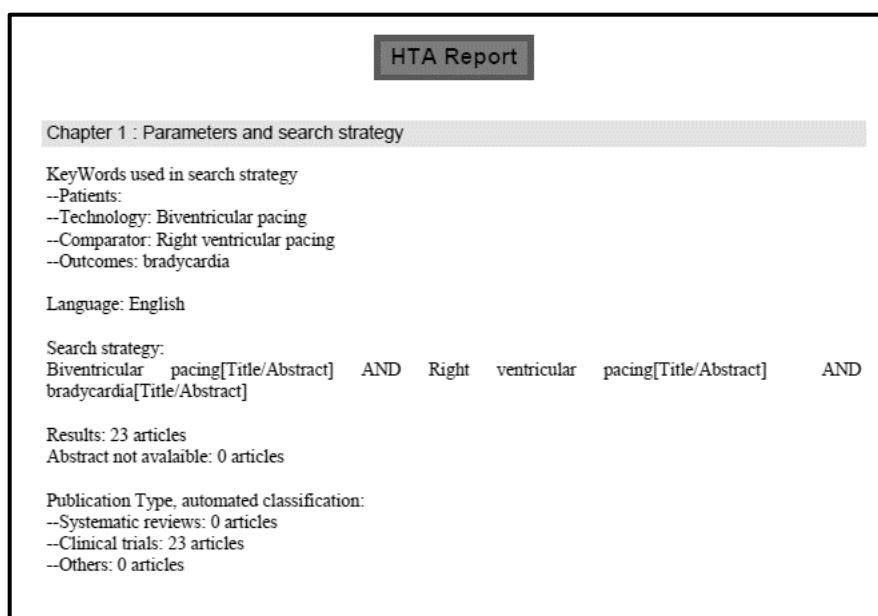


Figura 46. Sección uno de reporte: Parámetros y estrategia de búsqueda

Esta sección busca dar respuesta a los lineamientos de los reportes de Evaluación de Tecnología en Salud, en los cuales es muy importante documentar los detalles de la revisión sistemática implementada con el fin de que sea reproducible por otros y minimizar el sesgo en sus resultados.

Resumen:

En este apartado se listan los títulos de los artículos que se encontrarán a lo largo del reporte. Aquí se diferencian los reportes completos de los reportes con los artículos, dado que, el listado dependerá de la cantidad de artículos que el usuario haya seleccionado para el reporte o si en caso contrario requiere la síntesis con todos los artículos encontrados. Este capítulo luce como se muestra en la Figura 47.

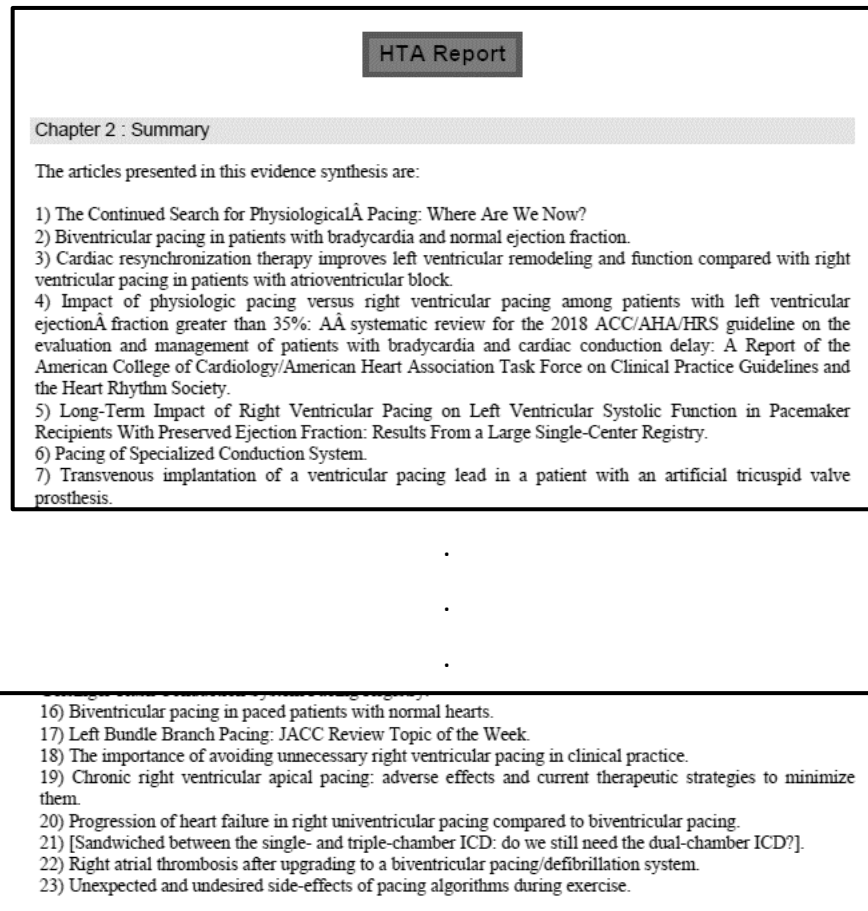


Figura 47. Sección dos: resumen de artículos presentados en reporte

Resultados:

Para cualquier tipo de reporte que se elija los resultados se presenta con la misma estructura, para cada artículo se indica la información que se presenta en la ventana de visualización de resultados. Esta información es: PMID, título, resumen, autores, tipo de publicación según PubMed y según modelo de clasificación y año de publicación. Un ejemplo de esta síntesis se ilustra en la Figura 48. Cuando no se logra recuperar uno de los ítems anteriores se indica como se ilustra en la Figura 49.

Chapter 3 : Articles

PMID: 19915220

Title: Biventricular pacing in patients with bradycardia and normal ejection fraction.

Abstract: Observational studies suggest that conventional right ventricular apical pacing may have a deleterious effect on left ventricular function. In this study, we examined whether biventricular pacing is superior to right ventricular apical pacing in preventing deterioration of left ventricular systolic function and cardiac remodeling in patients with bradycardia and a normal ejection fraction. Abstract: In this prospective, double-blind, multicenter study, we randomly assigned 177 patients in whom a biventricular pacemaker had been successfully implanted to receive biventricular pacing (89 patients) or right ventricular apical pacing (88 patients). The primary end points were the left ventricular ejection fraction and left ventricular end-systolic volume at 12 months. Abstract: At 12 months, the mean left ventricular ejection fraction was significantly lower in the right-ventricular-pacing group than in the biventricular-pacing group (54.8±9.1% vs. 62.2±7.0%, P<0.001), with an absolute difference of 7.4 percentage points, whereas the left ventricular end-systolic volume was significantly higher in the right-ventricular-pacing group than in the biventricular-pacing group (35.7±16.3 ml vs. 27.6±10.4 ml, P<0.001), with a relative difference between the groups in the change from baseline of 25% (P<0.001). The deleterious effect of right ventricular apical pacing occurred in prespecified subgroups, including patients with and patients without preexisting left ventricular diastolic dysfunction. Eight patients in the right-ventricular-pacing group (9%) and one in the biventricular-pacing group (1%) had ejection fractions of less than 45% (P=0.02). There was one death in the right-ventricular-pacing group, and six patients in the right-ventricular-pacing group and five in the biventricular-pacing group were hospitalized for heart failure (P=0.74). Abstract: In patients with normal systolic function, conventional right ventricular apical pacing resulted in adverse left ventricular remodeling and in a reduction in the left ventricular ejection fraction; these effects were prevented by biventricular pacing. (Centre for Clinical Trials number, CUHK_CCT00037.)

Authors: Cheuk-Man Yu, Joseph Yat-Sum Chan, Qing Zhang, Razali Omar, Gabriel Wai-Kwok Yip, Azlan Hussin, Fang Fang, Kai Huat Lam, Hamish Chi-Kin Chan, Jeffrey Wing-Hong Fung

Publication Type PubMed: Comparative Study. Journal Article. Multicenter Study. Randomized Controlled Trial. Research Support, Non-U.S. Gov't.

Publication Type by Automated classification: Randomized Controlled Trial

Year: 2009

Figura 48. Sección tres: resultados (artículos)

Evidence synthesis

Chapter 3 : Articles

PMID: 18083954

Title: Upgrading from right ventricular pacing to biventricular pacing in pacemaker patients with chronic heart failure: Heart failure.

Abstract not found

Authors: Christophe Leclercq

Publication Type PubMed: Journal Article. Review.

Publication Type by Automated classification: Randomized Controlled Trial

Year: 2008

Figura 49. Sección tres: resultados. Ejemplo de artículo con resumen no disponible

Los resultados aparecen en el mismo orden que se presentan en el resumen, en la visualización de resultados de la herramienta y corresponden con el ranking o la priorización realizada.

En el anexo *Instructivo de uso*, se puede observar todo el proceso de uso de la herramienta, desde la síntesis de la estrategia de búsqueda, recuperación de información, análisis, visualización, síntesis de reporte y visualización de reporte.

4.4.3. Validación

De acuerdo con la metodología planteada, la validación de la herramienta propuesta se realizó por medio de la aplicación de la prueba de usabilidad que se diseñó para tal fin. Todas las pruebas se realizaron con éxito, durante su ejecución se conservó la homogeneidad de su aplicación y se contó con diferentes momentos, los cuales fueron:

- ✓ **Inicio:** se realizó una presentación de la aplicación y sus principales funcionalidades. Durante esta presentación se indicó el marco del trabajo de investigación en la cual se desarrolló y el objetivo tanto de la prueba como de la aplicación.
- ✓ **Capacitación:** se realizó una explicación sobre el uso de la herramienta haciendo uso de los ejemplos con los que se inicializa por defecto la aplicación. Durante la explicación se ilustró cómo llevar a cabo las tareas que indicaba la prueba de usabilidad, tales como: sintetizar estrategias de búsqueda, búsqueda, navegar sobre los resultados, filtrar artículos de interés, sintetizar reportes y visualizarlos.
- ✓ **Uso:** una vez se finalizaba la explicación, se procedió a que el usuario ejecutara la prueba por medio de la encuesta diseñada para tal fin. Para esta etapa se desarrolló un instructivo de uso, para que en caso tal de que el usuario tuviera dudas que no lograra resolver por medio de la información de ayuda presentada en la aplicación, pudiera consultarlo. Este instructivo se encuentra como anexo a este trabajo de investigación: 7.1 Instructivo de uso.
- ✓ **Calificación:** una vez el usuario finalizaba la prueba, se procedió a dar respuesta a las preguntas planteadas.

Los resultados obtenidos se describirán siguiendo la estructura de la prueba diseñada. Inicialmente se realizó una caracterización básica de cada usuario donde se identificó nombre, edad, perfil profesional y nivel de escolaridad. Esta caracterización se sintetiza en la Tabla 22.

Tabla 22. Descripción de los usuarios que ejecutaron la prueba de usabilidad

	Perfil	Institución	Nombre	Edad
Usuario 1	Bioingeniero	IPS Universitaria	Santiago Suárez	25
Usuario 2	Médico, MSc epidemiología	Clínica UPB	Juan Guillermo Barrientos *	50
Usuario 3	Médico, MSc epidemiología	Clínica UPB	Esteban Villegas Arbeláez	29
Usuario 4	Bioingeniero, especialista en evaluación económica en salud	Fundación Instituto neurológico de Colombia	Santiago Marín	35
Usuario 5	Química Farmacéutica	IPS Universitaria	Natalia Restrepo *	33
Usuario 6	Médico, MSc epidemiología	Hospital Pablo Tobón Uribe	Elkin Arango *	53

*Profesionales que hacen parte de los comités de evaluación de tecnología de sus instituciones y que participaron en la caracterización de los procesos de ETS en la etapa inicial de este trabajo de investigación

Las preguntas de la prueba se dividieron en dos grupos, el primero estuvo compuesto de preguntas de opción múltiple, en la cual se pidió al usuario calificar en una escala de uno a cinco su percepción sobre cinco ítems, los cuales son:

- ✓ **Facilidad de uso.** Siendo 1 muy difícil y 5 muy fácil.
- ✓ **Intuitividad.** Siendo 1 poco intuitivo y 5 muy intuitivo.
- ✓ **Utilidad.** Siendo 1 poco útil y 5 muy útil.
- ✓ **Información.** Siendo 1 insuficiente y 5 suficiente.
- ✓ **Satisfacción.** Siendo 1 poco satisfecho y 5 muy satisfecho.

Los tres primeros parámetros se evaluaron para cada una de las tareas a ejecutar, las dos últimas se evaluaron de manera general.

En la Figura 50 se presentan los resultados obtenidos para cada uno de los parámetros evaluados.

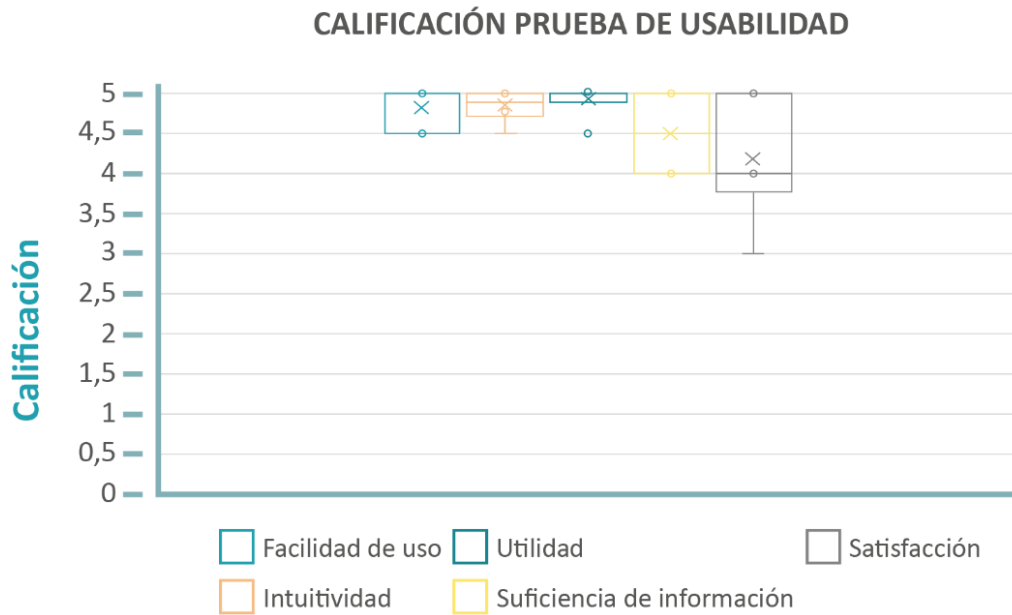


Figura 50. Resultados prueba de usabilidad

Se obtuvo una calificación favorable en todos los parámetros evaluados. En particular con relación a la facilidad de uso, el promedio de la calificación fue de 4.83 con una desviación estándar de 0.24, lo que indica que la aplicación es muy fácil de usar. Así mismo se identificó que la interfaz es intuitiva con una calificación de 4.83 y una desviación estándar de 0.19 para este parámetro. La calificación más alta se obtuvo para el parámetro de utilidad con un 4.92, es decir que los usuarios encontraron muy útil la aplicación y cada una de sus funcionalidades. Con relación a la información de ayuda se obtuvo una calificación de 4.50 (con desviación de 0.50), la cual es una calificación que indica que la información presentada es suficiente para apoyar al usuario en el uso adecuado de las funcionalidades de la aplicación.

Con relación a la pregunta de satisfacción se obtuvo un promedio de calificación de 4.17, lo que indica que los usuarios están satisfechos con la aplicación versus el objetivo de la misma. **Objetivo:** Esta aplicación está diseñada para apoyar y mejorar la trazabilidad de la etapa de filtrado de estudios según título y resumen en las revisiones sistemáticas de una evaluación de tecnología en salud. Este parámetro tiene una desviación estándar de 0.69, la cual es más alta que los demás, lo que indica una mayor dispersión de las respuestas obtenidas. Esto puede deberse a que, aunque la pregunta se enfocó en evaluar la aplicación con relación al objetivo descrito inicialmente, el usuario tiene una **expectativa** con base en su experiencia profesional, es decir, que el usuario ya ha experimentado con herramienta existentes, lo cual sesga su criterio de calificación y eventualmente su percepción puede estar inclinada a tener una interfaz que enfatice en sus necesidades particulares y no las colectivas.

El segundo grupo estuvo compuesto por preguntas de respuesta abierta, donde se buscó obtener mayor información sobre la percepción del usuario con relación a la herramienta presentada. Los resultados se presentan a continuación:

Tabla 23. Respuestas a la pregunta 6 de la prueba de usabilidad

¿Qué parte de la aplicación más le gusta?, ¿por qué?	
Usuario 1	La generación automática de los reportes con base en la selección que el usuario tenga, considero que posibilita una muy buena estrategia para una búsqueda bibliografía corta y que al estar priorizada genera alta confiabilidad sobre los resultados obtenidos
Usuario 2	La identificación de los componentes de la pregunta PICO
Usuario 3	Colores en los resultados, fácil de identificar
Usuario 4	La forma de ingresar la pregunta PICO porque facilita el proceso de búsqueda
Usuario 5	Es fácil de usar y muy intuitiva
Usuario 6	El poder realizar la búsqueda con la estrategia PICO. Facilita la búsqueda para personas no expertas

Tabla 24. Respuestas a la pregunta 7 de la prueba de usabilidad

¿Qué parte de la aplicación usó menos?, ¿por qué?	
Usuario 1	Todas las partes se usaron en conjunto dentro de la aplicación, puesto que son parámetros sustanciales y relevantes para sintetizar la búsqueda bibliográfica
Usuario 2	Sería importante en mostrar información estadística de forma independiente
Usuario 3	Creo que hace falta adicionar opción de filtrar
Usuario 4	Ninguna
Usuario 5	Usamos todo
Usuario 6	Usé todo por igual

De las Tabla 23 y Tabla 24 se evidencia que todos los elementos de la interfaz son necesarios para su uso, dado que, todos se usan para llevar a cabo el proceso de síntesis de evidencia. Como principales fortalezas los usuarios resaltaron como positivo la manera de ingresar la estrategia de **búsqueda por medio de términos PICO**, la **extracción de información** de texto y **resaltarlo** en los resultados y finalmente la **síntesis de los reportes**.

Tabla 25. Respuestas a la pregunta 8 de la prueba de usabilidad

¿Qué piensa acerca de la interfaz?	
Usuario 1	Es una interfaz limpia, que es agradable a la vista y que permite diferenciar fácilmente las secciones que tiene incorporadas
Usuario 2	Es práctica y útil, importante continuar con línea de procesamiento de datos estadísticos
Usuario 3	Amigable
Usuario 4	Que es una interfaz intuitiva, permite hacer una revisión apropiada de los artículos encontrados
Usuario 5	Es amigable y fácil de usar
Usuario 6	Me parece un desarrollo útil... el resultado final, intuyo, será una maravilla

Con relación a la interfaz en términos generales se tuvo una aceptación absoluta acerca de la misma, se encontró **amigable, práctica, útil y fácil de usar**. Se resalta la recomendación de implementar un diseño más llamativo conservando las características antes mencionadas, las cuales la hacen simple y eficiente en términos del cumplimiento de objetivos. Esto se puede evidenciar en la Tabla 27.

Tabla 26. Respuestas a la pregunta 9 de la prueba de usabilidad

¿Qué le parece la manera en que las características e información están presentadas?	
Usuario 1	La presentación de la información, especialmente en los resultados al seleccionar las palabras relacionadas con cada una de las 3 clasificaciones, es un aspecto para resaltar, debido a que con una rápida inspección se puede tener interés o no en cierto artículo que fue encontrado por la aplicación
Usuario 2	Revisar contraste de colores en el texto de la pregunta PICO para no fatigar al lector cuando sean muchos artículos
Usuario 3	Muy buena, añadiría el tema de: a) filtros, b) selección o visualización de resultados críticos
Usuario 4	Es muy útil cuando los resultados son pocos, pero podría complicarse con búsquedas que presentan muchos artículos, podría tener algunos elementos que permitan filtrar los resultados
Usuario 5	Me parece que están muy bien presentados, los colores son acordes y fáciles de identificar. Solo recomendaría que en los resultados también se ordene por tipo de publicación, para hacer más fácil la selección de los artículos
Usuario 6	Me parece adecuada. Sugiero que la interfaz tenga un diseño atractivo para el usuario final

De manera general todos los usuarios convinieron en que las características e información se presentan adecuadamente, sin embargo, se recibió la retroalimentación de permitir o agregar una manera de **usar filtros en la visualización de resultados** para hacer más eficiente su revisión. Esta recomendación es totalmente válida y al ser homogénea su necesidad entre los usuarios es viable y pertinente su implementación. Con relación al contraste del texto con el fondo en la visualización de resultados solo se objetó por parte de un usuario, los demás estuvieron cómodos. Por este motivo se propone no cambiar permanentemente los colores de visualización sino dejar al usuario la elección de contraste de colores para analizar los resultados.

Tabla 27. Respuestas a la pregunta 10 de la prueba de usabilidad

¿Seguiría usando esta aplicación? Si no, ¿por qué no?	
Usuario 1	Si, respecto al ámbito investigativo es un avance sustancial en la extracción y consolidación de la información, además de que permite poder abordar mayor cantidad de información en un menor tiempo, lo que al final repercute en un mayor tiempo dedicado al análisis de los resultados y un menos tiempo dedicado a la parte operativa de la búsqueda de la información
Usuario 2	Si
Usuario 3	Si
Usuario 4	Si, es fundamental para iniciar cualquier proceso de evaluación de tecnología
Usuario 5	Si, me parece muy útil y amigable, muestra resultados fácilmente y es práctico la forma como selecciona palabras clave
Usuario 6	Si, porque facilita las búsquedas y la organización de la información

De la Tabla 27 se puede evidenciar que el **100%** de los usuarios que participaron en la prueba convienen en que **seguirían usando la aplicación**, resaltando su utilidad en el proceso de síntesis de evidencia, la facilidad de uso, la presentación de la información y la trazabilidad que se genera del proceso. Este resultado es beneficioso para este trabajo de investigación, dado que, propicia la implementación de las mejoras sugeridas por los usuarios en aplicación y el despliegue de la misma en los ámbitos para los cuales fue diseñada.

Todos los resultados aquí presentados son una transcripción de los resultados obtenidos durante la ejecución de las pruebas, estas se encuentran como anexo: 7.3 Ejecución pruebas de usabilidad.

A partir de los resultados y principalmente las respuestas relacionadas con la utilidad de la herramienta propuesta y a la voluntad unificada de usarla en procesos de ETS en los hospitales incluidos en la prueba, es posible concluir que realmente puede apoyar los procesos para guiar la toma de decisiones en salud. De esta manera se demuestra que la implementación de la automatización de procesos en el dominio de la evaluación de

tecnología en salud puede contribuir a reducir la carga de trabajo de los profesionales y mejorar continuamente el proceso.

5. Conclusiones

La evaluación de tecnología en salud es una herramienta muy valiosa en cualquier nivel de implementación en un sistema de salud para apoyar la toma de decisiones con relación a la incorporación de tecnología. Aunque su importancia para incorporar tecnología pertinente y optimización de recursos es un hecho, en Colombia se realiza en muy pocos hospitales del país. Sin embargo, se logró evidenciar que en aquellas instituciones en que se realiza, se llevan a cabo buenas prácticas y generalmente existe cierta homogeneidad en la metodología para su aplicación. Se identificó que los principales anfitriones para desarrollar estos procesos de revisión de literatura son profesionales de la medicina, epidemiología y química farmacéutica y en pocas ocasiones se ven involucrados los profesionales de la ingeniería biomédica. También se evidenció que, si bien se realiza ETS, aún no existe el hábito de evaluar o cuantificar los beneficios o consecuencias de guiar la incorporación de tecnología teniendo como base los resultados de la ETS. Sin embargo, en la mayoría de las instituciones realizan esta práctica, las recomendaciones que surgen del proceso de evaluación son tenidas en cuenta al momento de adquirir tecnología en salud. Teniendo en cuenta el enfoque para la toma de decisiones y la estructura organizacional destinada a realizar ETS, es posible afirmar que la metodología que mejor se ajusta a las instituciones de salud en Colombia es la mini-ETS. A través de esta última sentencia y la caracterización realizada, se da cumplimiento al primer objetivo específico de este trabajo.

Con relación al segundo objetivo específico: *Consolidar una base de datos de evaluaciones de tecnología en salud en dispositivos médicos a partir de reportes de evaluación de instituciones reconocidas por entidades gubernamentales y no gubernamentales*, se tiene un cumplimiento completo. Dado que, se logró consolidar una base de datos de reportes de evaluación de tecnología en salud enfocados en dispositivos médicos para el diagnóstico y tratamiento de las enfermedades cardiovasculares. Esta base de datos permitió el entrenamiento de los algoritmos para la clasificación del tipo de publicación (ensayos clínicos y revisiones sistemáticas) que se incluyen en dichos reportes. El mejor algoritmo de clasificación fue la máquina de soporte vectorial, para la clasificación de ensayos clínicos se obtuvo un AUC de 0.974 y para revisiones sistemáticas un AUC de 0.983.

Además, esta base de datos permitió validar un algoritmo para priorizar los estudios clínicos que se obtienen de una búsqueda de tal manera que el investigador obtenga en los primeros artículos la cantidad suficiente de información para incluir en la revisión de literatura deseada y poder hacer recomendaciones. Esto se logró por medio de la aplicación del Procesamiento de Lenguaje Natural, específicamente el análisis de matrices de coocurrencia de palabras, cálculo de frecuencia de términos (TF-IDF), métodos de *word embeddings* de *BERT* y los algoritmos de aprendizaje de máquina como máquinas de soporte vectorial y árboles de decisión. Con el entrenamiento de dichos algoritmos se dió cumplimiento al tercer objetivo específico planteado: *Adaptar los algoritmos de la herramienta RobotReviewer, con el fin de hacer transferencia de aprendizaje y automatizar la extracción información de fuentes de evidencia clínica.*

Finalmente, se logró el diseño de una herramienta que integra todo el desarrollo de la automatización de algoritmos de búsqueda, búsqueda, clasificación y extracción de información para las revisiones sistemáticas. Por medio de esta herramienta es posible visualizar, analizar y filtrar los estudios obtenidos de la búsqueda de información en la base de datos PubMed. Esta última base de datos se eligió por medio de la caracterización que se realizó a Instituciones Prestadoras de Servicios de Salud acreditadas de la ciudad de Medellín. Por medio de esta caracterización también se contribuye a tener un mejor panorama sobre la implementación de la evaluación de tecnología en salud en los hospitales de la ciudad y la participación de la ingeniería clínica en dichos procesos.

De acuerdo con los resultados obtenidos de durante la ejecución de las pruebas de usabilidad de la herramienta propuesta se puede concluir que se implementó una interfaz amigable, intuitiva y fácil de usar. Todos los usuarios indicaron que es una aplicación útil y expresaron su interés en continuar usando la aplicación debido principalmente a la facilidad para ejecutar búsquedas, visualizar resultados y generar trazabilidad del proceso. Con esto último se confirma la oportunidad de continuar en esta línea de trabajo con el fin de mejorar y ampliar la automatización de estos procesos en las Instituciones Prestadoras de Servicios de Salud en Colombia.

Este trabajo de investigación partió de la necesidad de implementar herramientas y procesos que permitan apoyar la ejecución de revisiones sistemáticas de la ETS enfocadas en dispositivos médicos, con el objetivo de guiar la toma de decisiones para la incorporación de tecnología en las IPS acreditadas en Colombia. Con los resultados ilustrados es posible concluir que se logró automatizar algunas de las tareas necesarias para la síntesis de evidencia clínica, la cual es un criterio relevante dentro de la acreditación en salud, que exige toma de decisiones con base en las evidencias de seguridad disponibles. Adicionalmente, como en cualquier sistema que esté enfocado en mejorar la calidad de sus procesos, la trazabilidad que permite la herramienta es un insumo que aporta en gran medida a la gestión de tecnología en dichas instituciones.

5.4. Contribuciones

Este trabajo de investigación permitió generar los aportes que se indican a continuación:

- ✓ Una caracterización detallada del proceso de evaluación de tecnología en salud en instituciones acreditadas de la ciudad, enfocada principalmente en la búsqueda de evidencia clínica para analizar la seguridad, efectividad y eficacia de la tecnología.
- ✓ Un consolidado de reportes de evaluación de tecnología en salud enfocados en dispositivos médicos para tratar afecciones cardiovasculares, el cuál puede servir como base de para futuras evaluaciones de tecnología en las instituciones de salud
- ✓ Se presenta una metodología para minimizar el tiempo de filtro y análisis de artículos en la etapa de selección de estudios por medio de título y resúmenes en las revisiones de literatura. Así mismo se presentan los algoritmos para clasificar ensayos clínicos y revisiones sistemáticas de otro tipo de diseños.

- ✓ Finalmente, se entrega una herramienta de acceso libre que permite apoyar los procesos de evaluación de tecnología en salud y facilita la síntesis de evidencia. Generando mayor trazabilidad y la posibilidad de estandarizar los reportes de evaluación y en general los procesos de ETS.
- ✓ Por medio de este trabajo de investigación se aporta una herramienta útil para mejorar la implementación de ETS en las instituciones acreditadas (o en proceso de acreditación). Dado que, facilita la trazabilidad de la evaluación de tecnologías y apoya la incorporación de la misma con base en las evidencias de seguridad disponibles como lo es la síntesis de evidencia clínica que se puede realizar por medio de esta herramienta.

5.5. Limitaciones

Este trabajo estuvo enfocado en dispositivos médicos con una alta diseminación de evidencia, por lo tanto, es pertinente ampliar su alcance y estudiar su aplicación y resultados exitosos en otro tipo de dispositivos médicos e inclusive otro tipo de tecnologías en salud.

No fue posible caracterizar el proceso de evaluación de tecnología en salud del hospital General de Medellín, institución que también ha obtenido la acreditación en salud y es relevante dentro de la red de prestadores de servicios de salud de la ciudad. Este trabajo se desarrolló durante la pandemia por COVID – 19, por lo que la dinámica hospitalaria restringió en cierta medida disponibilidad del personal de salud e interacción entre personas.

5.6. Trabajo futuro

Con base en el presente trabajo es posible implementar la herramienta de síntesis de evidencia en las instituciones hospitalarias y rectificar su utilidad durante procesos rutinarios de ETS. Asimismo, al ser de código abierto, es posible hacer integraciones de otras actividades relacionadas con la revisión de literatura que surjan a raíz de su uso y las necesidades clínicas. Como, por ejemplo, se resalta la pertinencia de profundizar en la extracción y procesamiento de datos estadísticos de los artículos obtenidos en las búsquedas, para finalmente presentarlos y hacer más robusto el apoyo que se dé a los profesionales que realizan el análisis y toman decisiones con base en esta información

6. Referencias

- [1] OMS, "Evaluación de tecnologías sanitarias aplicadas a los dispositivos médicos." 2012.
- [2] OTA, "Assessing the Efficacy and Safety of Medical Technologies," no. September. pp. 1–133, 1978, doi: PB-286929.
- [3] B. M. da S. S. de Ciência., "Directriz Metodológica: elaboración de estudios para evaluación de equipos médicos." Brasilia, 2013.
- [4] J. G. Barrientos Gómez, A. E. Marín Castro, L. Becerra Ruiz, and M. A. Tobón Arango, "La evaluación de nuevas tecnologías en salud en hospitales: revisión narrativa," *Med. UPB*, vol. 35, no. 2, pp. 120–134, 2017, doi: 10.18566/medupb.v35n2.a06.
- [5] CENETEC-Salud, "Evaluación de tecnologías para la salud. Documento metodológico," *Secretaría de Salud. Subsecretaría de Integración y Desarrollo del Sector Salud*. México D.F., pp. 1–77, 2010.
- [6] Ministerio de la Protección Social, "Guía para la evaluación de tecnologías de salud (ETS) en instituciones de servicios de salud (IPS)," p. 118, 2011.
- [7] A. Pichon-Riviere, N. C. Soto, F. A. Augustovski, and S. G. Martí, "Evaluación de tecnologías sanitarias para la toma de decisiones en Latinoamérica: principios de buenas prácticas," *Rev Panam Salud Publica Pan Am. J. Public Heal.*, vol. 41, pp. 1–8, 2017, doi: 10.26633/RPSP.2017.138.
- [8] CAHTA, "EUnetHTA Work Package 8. Handbook on Hta Capacity Building," *Health Technol. Assess. (Rockv)*., no. October, 2008.
- [9] J. H. R. M. M. . M. . DPH, "Evaluación de tecnologías en salud a nivel Meso del Sistema de Salud."
- [10] World Heart Federation, "Cost of heart disease in Latin America," 2016, pp. 30–32.
- [11] M. I. RIACHI GONZÁLEZ, "Manual de Acreditación en Salud Ambulatorio y Hospitalario," vol. 1. 2018.
- [12] J. O. B. T. S. Toma, T. Pereira, T. Vanni, "Síntese de evidências para políticas de saúde," *Avaliação de Tecnologias de Saúde & Políticas Informadas por Evidências Avaliação de Tecnologias Informadas por Evidências*. pp. 183–208, 2017.
- [13] M. C. Kapoor, "Types of studies and research design," *Indian J. Anaesth.*, vol. 60, no. 9, pp. 626–630, Sep. 2016, doi: 10.4103/0019-5049.190616.
- [14] U. of Liverpool, "Evidence Synthesis and Health Technology Assessment," *Institute of Population Health*, 2022. [Online]. Available: <https://www.liverpool.ac.uk/population-health/research/groups/evidencesynthandhta/>. [Accessed: 11-Mar-2022].
- [15] A. Auraen, R. Fujisawa, G. de Lagasnerie, and V. Paris, "How OECD health systems define the range of good and services to be financed collectively.," *OECD Heal. Work. Pap.*, no. 90, p. 112 many ref, 2016, doi: 10.1787/5jlnb59ll80x-en.
- [16] F. Dams, J. L. Gonzalez Rodriguez, K. L. Cheung, B. F. M. Wijnen, and M. Hiligsmann, "Relevance of barriers and facilitators in the use of health technology assessment in Colombia," *J. Med. Econ.*, vol. 21, no. 5, pp. 510–517, 2018, doi: 10.1080/13696998.2018.1449751.
- [17] J. W. & S. L. Cochrane Collaboration, "Cochrane handbook for systematic reviews of

- interventions,” *IEEE International Symposium on Information Theory - Proceedings*. 2008, doi: 10.1109/ISIT.2017.8006970.
- [18] E. Beller *et al.*, “Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR),” pp. 1–7, 2018.
- [19] I. J. Marshall and B. C. Wallace, “RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials,” vol. 44, no. 0, pp. 193–201, 2016, doi: 10.1093/jamia/ocv044.
- [20] R. Tarricone, G. Callea, M. Ogorevc, and V. Prevolnik Rupel, “Improving the Methods for the Economic Evaluation of Medical Devices,” *Heal. Econ. (United Kingdom)*, vol. 26, pp. 70–92, 2017, doi: 10.1002/hec.3471.
- [21] L. K. Sampietro-Colom L, K. K. Cicchetti A, F. B. Pasternack I, W. C. Rosenmöller M, W. J. Kahveci R, and et al. Kiivet RA, “THE AdHopHTA HANDBOOK,” p. 222, 2015.
- [22] A. M. Drucker, P. Fleming, and A. W. Chan, “Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews,” *J. Invest. Dermatol.*, vol. 136, no. 11, pp. e109–e114, 2016, doi: 10.1016/j.jid.2016.08.021.
- [23] R. Borah, A. W. Brown, P. L. Capers, and K. A. Kaiser, “Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry,” *BMJ Open*, vol. 7, no. 2, pp. 1–8, 2017, doi: 10.1136/bmjopen-2016-012545.
- [24] G. Kontonatsios *et al.*, “A semi-supervised approach using label propagation to support citation screening,” *J. Biomed. Inform.*, vol. 72, pp. 67–76, 2017, doi: 10.1016/j.jbi.2017.06.018.
- [25] A. O’Mara-Eves, J. M. , James Thomas, and M. M. and S. Ananiadou2, “Using text mining for study identification in systematic reviews: a systematic review of current approaches,” *Syst. Rev. J.*, pp. 1–22, 2016, doi: 10.1145/2945078.2945085.
- [26] J. L. G. Noreña, “RESPUESTA DERECHO DE PETICIÓN,” Bogotá, 2019.
- [27] F. M. Martinez-licona, M. Ortiz-posadas, and R. Ortiz-pedroza, “State of the Art of the Evaluation of Health Technologies in Latin America,” no. December, 2019, doi: 10.17488/rmib.40.3.12.
- [28] EUnetHTA, *EUnetHTA Joint Action 2, Work Package 8. HTA Core Model ® version 3.0 (Pdf)*, vol. 3. 2016.
- [29] DACEHTA, *Introduction to mini-hta –. Danish*, 2005.
- [30] A. Cicchetti, M. Marchetti, R. Dibidino, and M. Corio, “Hospital based health technology assessment. Worldwide survey. HTAi interest sub-group Hospital Based HTA 2008.” 2008.
- [31] C. C. de A. Pereira, R. D. S. Rabello, and F. T. S. Elias, “HOSPITAL-BASED HEALTH TECHNOLOGY ASSESSMENT IN BRAZIL: AN OVERVIEW OF THE INITIAL EXPERIENCES,” *Int. J. Technol. Assess. Health Care*, vol. 33, no. 2, pp. 227–231, Jan. 2017, doi: 10.1017/S0266462317000459.
- [32] G. Demirdjian, “A 10-YEAR HOSPITAL-BASED HEALTH TECHNOLOGY ASSESSMENT PROGRAM in A PUBLIC HOSPITAL in ARGENTINA,” *Int. J. Technol. Assess. Health Care*, vol. 31, no. 1–2, pp. 103–110, 2015, doi: 10.1017/S0266462315000124.
- [33] L. Sampietro-Colom *et al.*, “A Handbook and a Toolkit for Hospital-Based Health

- Technology Assessment,” *Value Heal.*, vol. 18, no. 7, p. A556, 2015, doi: 10.1016/j.jval.2015.09.1796.
- [34] A. M. O’Connor, G. Tsafnat, S. B. Gilbert, K. A. Thayer, and M. S. Wolfe, “Moving toward the automation of the systematic review process: A summary of discussions at the second meeting of International Collaboration for the Automation of Systematic Reviews (ICASR),” *Syst. Rev.*, vol. 7, no. 1, pp. 3–7, 2018, doi: 10.1186/s13643-017-0667-4.
- [35] Cochrane, “¿Quiénes somos?,” 2022. [Online]. Available: <https://www.cochrane.org/es/about-us>. [Accessed: 19-May-2022].
- [36] S. TOOLBOX, “THE SYSTEMATIC REVIEW TOOLBOX,” ICASR, 2022. [Online]. Available: <http://systematicreviewtools.com/guidance.php>. [Accessed: 19-May-2022].
- [37] E. L. S. Bird, E. Klein, *Natural Language Proccesing with python*, First. O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472., 2009.
- [38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” no. Mlm, 2018.
- [39] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [40] I. Beltagy, K. Lo, and A. Cohan, “SciBERT: A Pretrained Language Model for Scientific Text,” pp. 3613–3618, 2019, doi: 10.18653/v1/d19-1371.
- [41] J. Lee *et al.*, “BioBERT: A pre-trained biomedical language representation model for biomedical text mining,” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020, doi: 10.1093/bioinformatics/btz682.
- [42] A. M. Scott, C. Forbes, J. Clark, M. Carter, P. Glasziou, and Z. Munn, “Systematic review automation tool use by systematic reviewers, health technology assessors and clinical guideline developers: tools used, abandoned, and desired,” *medRxiv*, 2021.
- [43] I. J. Marshall, A. Noel, S. Joël, K. James, and T. Byron, “Machine learning for identifying Randomized Controlled Trials : An evaluation and practitioner ’ s guide,” no. October 2017, pp. 602–614, 2018, doi: 10.1002/jrsm.1287.
- [44] C. W. Tian Kanga, Shirui Zoub, “Pretraining to Recognize PICO Elements from Randomized Controlled Trial Literature,” *Stud. Health Technol. Inform.*, vol. 0, 2019, doi: 10.3233/SHTI190209.Pretraining.
- [45] Z. L. Yifan Peng, Shankai Yan, “Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets,” *Natl. Cent. Biotechnol. Inf.*, no. iv, 2019.
- [46] D. Moher *et al.*, “Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement,” *Rev. Esp. Nutr. Humana y Diet.*, vol. 20, no. 2, pp. 148–160, 2015, doi: 10.1186/2046-4053-4-1.
- [47] D. Moher *et al.*, “CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.,” *BMJ*, vol. 340, 2010, doi: 10.1136/bmj.c869.
- [48] K. A. Hicks *et al.*, “2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials,” *Circulation*, vol. 137, no. 9, pp. 961–972, 2018, doi: 10.1161/CIRCULATIONAHA.117.033502.

- [49] J. Isojärvi and J. Glanville, "Evidence-based searching for health technology assessment: Keeping up to date with SuRe Info," *Int. J. Technol. Assess. Health Care*, no. May, 2021, doi: 10.1017/S026646232100009X.
- [50] Hta. INAHTA, "Health Technology Assessment," *HTA Glossary*, 2020. [Online]. Available: <http://htaglossary.net/health+technology+assessment+%28HTA%29>. [Accessed: 10-Apr-2020].
- [51] J. Villanueva Padilla and F. M. Licona, "Análisis del Ciclo de Vida de la Tecnología Médica desde una Aproximación Integral Life Cycle Analysis of Medical Technology from an Integrative Approach," vol. 8, no. 2, pp. 7–12, 2010.
- [52] Ministerio de la Protección Social, "DECRETO 4725 DE 2005." Bogotá, p. 34, 2005.
- [53] F. S. Sônego, "Estudo de Métodos de Avaliação de Tecnologias em Saúde Aplicada a Equipamentos Eletromédicos," pp. 1–92, 2007.
- [54] M. Contó, "INTERACCIONES ENTRE ETS Y GTS: AÑADIENDO VALOR A LOS SISTEMAS Y SERVICIOS DE SALUD." Medellín, 2019.
- [55] A. L. Fernanda Lessa, Francisco Caccavo, Stephanie Curtis, Stéphanie Ouimet-Rathé, "Fortalecimiento y ejecución de la Evaluación de Tecnología en Salud y del proceso toma de decisiones en la región de las Américas," *Rev. Panam. Salud Pública*, vol. 41, 2017, doi: 10.26633/RPSP.2017.165.
- [56] M. P. Gagnon, M. Desmartis, T. Poder, and W. Witteman, "Effects and repercussions of local/hospital-based health technology assessment (HTA): A systematic review," *Syst. Rev.*, vol. 3, no. 1, pp. 1–14, 2015, doi: 10.1186/2046-4053-3-129.
- [57] M. M., "The Impact of Reports of the Technology Assessment Unit of the McGill University Health Centre," 2012.
- [58] R. C. Lee, D. Marshall, C. Waddell, D. Hailey, and D. Juzwishin, "Health technology assessment, research, and implementation within a health region in Alberta, Canada," *Int. J. Technol. Assess. Health Care*, vol. 19, no. 3, pp. 513–520, 2003, doi: 10.1017/S0266462303000448.
- [59] I. Zechmeister and I. Schumacher, "The impact of health technology assessment reports on decision making in Austria," *Int. J. Technol. Assess. Health Care*, vol. 28, no. 1, pp. 77–84, 2012, doi: 10.1017/S0266462311000729.
- [60] C. Collaboration, *Manual Cochrane de revisiones sistemáticas de intervenciones*, 5.1., no. March. Barcelona: https://es.cochrane.org/sites/es.cochrane.org/files/uploads/Manual_Cochrane_510_reduit.pdf, 2011.
- [61] B. J. Shea *et al.*, "AMSTAR 2 : a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions , or both," pp. 1–9, 2017, doi: 10.1136/bmj.j4008.
- [62] C. Centre and J. P. Vandembroucke, "Strengthening the reporting of observational studies in epidemiology (STROBE) statement : guidelines for reporting observational studies," vol. 335, no. october, pp. 20–22, 2007.
- [63] P. F. A. W. S. M. E. W. S. M. J. J. D. J. B. R. M. M. G. L. J. A. C. S. P. M. M. B. Whiting, "QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies," *Ann. Intern. Med.*, no. 4, 2011.
- [64] G. H. Guyatt *et al.*, "GRADE guidelines: 9. Rating up the quality of evidence," *J. Clin.*

- Epidemiol.*, vol. 64, no. 12, pp. 1311–1316, 2011, doi: 10.1016/j.jclinepi.2011.06.004.
- [65] H. Arentz-hansen *et al.*, “Safer introduction of new health technologies,” 2018. [Online]. Available: <https://tidsskriftet.no/en/2018/05/kronikk/safer-introduction-new-health-technologies>. [Accessed: 23-May-2020].
- [66] N. Martelli *et al.*, “Harmonizing Health Technology Assessment Practices in University Hospitals: To What Extent Is the Mini-Hta Model Suitable in the French Context?,” *Int. J. Technol. Assess. Health Care*, vol. 33, no. 02, pp. 307–314, 2017, doi: 10.1017/s0266462317000393.
- [67] K. J. Salazar-Flórez, S. Botero-Botero, and C. N. Jiménez-Hernández, “Adquisición de tecnología biomédica en IPS colombianas: Comparación y mejores prácticas,” *Rev. Gerenc. y Polit. Salud*, vol. 15, no. 31, pp. 88–118, 2016, doi: 10.11144/Javeriana.rgyps15-31.atbi.
- [68] MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL, “Resolución No. 3100 de 2019.pdf.” Colombia, p. 71, 2019.
- [69] Ministerio de Salud y de la Protección Social de Colombia, “Resolución 5095 de 2018.” p. 81, 2018.
- [70] J. C. M Ardila, A Gomez, “Método de incorporación estratégica de tecnología Biomédica Para Instituciones De Salud,” pp. 35–42, 2016.
- [71] RAE, “Lenguaje,” 2020. [Online]. Available: <https://dle.rae.es/lenguaje?m=form>.
- [72] P. Goyal, S. Pandey, and K. Jain, “Deep Learning for Natural Language Processing,” *Deep Learn. Nat. Lang. Process.*, 2018, doi: 10.1007/978-1-4842-3685-7.
- [73] S. Robertson, “Understanding inverse document frequency: On theoretical arguments for IDF,” *J. Doc.*, vol. 60, no. 5, pp. 503–520, 2004, doi: 10.1108/00220410410560582.
- [74] T. Demeester, T. Rocktäschel, and S. Riedel, “Distributed Representations of Words and Phrases and their Compositionality Tomas,” *EMNLP 2016 - Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 1389–1399, 2016, doi: 10.18653/v1/d16-1146.
- [75] C. M. J. Pennington, R. Socher, “GloVe: Global Vectors for Word Representation Jeffrey,” vol. 31, no. 6, pp. 682–687, 2017, doi: 10.1080/02688697.2017.1354122.
- [76] IBM, “Recurrent Neural Networks,” 2022. [Online]. Available: <https://www.ibm.com/cloud/learn/recurrent-neural-networks>. [Accessed: 28-Sep-2022].
- [77] R. Dolphin, “LSTM Networks | A Detailed Explanation,” 2022. [Online]. Available: <https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9>. [Accessed: 28-Sep-2022].
- [78] R. Roy, “Regularization in Machine Learning,” 2022. [Online]. Available: <https://towardsdatascience.com/regularization-in-machine-learning-6fbc4417b1e5>. [Accessed: 28-Sep-2022].
- [79] C. Freksen, L. Kamma, and K. G. Larsen, “Fully understanding the hashing trick,” *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, pp. 5389–5399, 2018.
- [80] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, “Glue: A multi-task benchmark and analysis platform for natural language understanding,” *7th Int. Conf. Learn. Represent. ICLR 2019*, pp. 1–20, 2019.
- [81] P. Rajpurkar, R. Jia, and P. Liang, “Know what you don’t know: Unanswerable

- questions for SQuAD,” *ACL 2018 - 56th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 2, pp. 784–789, 2018, doi: 10.18653/v1/p18-2124.
- [82] R. Zellers, Y. Bisk, R. Schwartz, and Y. Choi, “SWAG: A large-scale adversarial dataset for grounded commonsense inference,” *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 93–104, 2020, doi: 10.18653/v1/d18-1009.
- [83] WHO (World Health Organization), “Number of clinical trials by year, country, region and income group,” *WHO*, 2021. [Online]. Available: <https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/number-of-clinical-trials-by-year-country-who-region-and-income-group>. [Accessed: 14-May-2022].
- [84] GetReskilled, “¿What are the Top Medical Device Companies in the World in 2020?,” 2020. [Online]. Available: <https://www.getreskilled.com/medical-device-companies/>. [Accessed: 14-May-2022].
- [85] I. UNIVERSITARIA, “Plan Estratégico 2017-2026,” 2017. [Online]. Available: <https://ipsuniversitaria.com.co/plan-estrategico-2017-2026/>. [Accessed: 31-Jan-2022].
- [86] Ministerio de Salud y Protección Social, “Análisis de la Situación de Salud ASIS-Colombia 2020,” *Minist. Salud y Prot. Soc. Colomb.*, p. 98, 2020.
- [87] “Use of Biventricular Pacing in Atrioventricular Heart Block,” 2016.
- [88] D. A. B.-M. and P. J. H. from Ms Linda Mundy, Mr Brent Hodgkinson, Ms Tracy Merlin, “National Horizon Scanning Unit Horizon scanning report Left Ventricular Assist Devices for Destination Therapy,” *Adelaide Heal. Technol. Assess.*, no. March, 2004.
- [89] A. K. Kirsten Howard, Simon Eckermann, “Intravascular Ultrasound April 2001 MSAC application 1031 Assessment report,” *Medicare Serv. Advis. Comm.*, no. April, 2001.
- [90] C. A. for D. and T. in Health, “Magnetic Resonance Imaging for Patients with Implantable Cardiac Devices: A Review of Safety and Guidelines - NCBI Bookshelf,” pp. 1–32, 2019.
- [91] D. C. Isbell and C. M. Kramer, “Magnetic resonance for the assessment of myocardial viability,” 2010.
- [92] Medical Advisory Secretariat, “Positron emission tomography for the assessment of myocardial viability,” *Ontario Health Technology Assessment Series*. 2005.
- [93] Medical Advisory Secretariat, “Positron emission tomography for the assessment of myocardial viability,” 2010.
- [94] S. Sehatzadeh *et al.*, “Transcatheter aortic valve implantation (TAVI) for treatment of aortic valve stenosis: An evidence update,” *Ont. Health Technol. Assess. Ser.*, vol. 13, no. 1, pp. 1–40, 2013.
- [95] H. Q. Ontario, “Ontario health technology assessment series: Percutaneous ventricular assist devices: A health technology assessment,” *Ont. Health Technol. Assess. Ser.*, vol. 17, no. 2, pp. 1–97, 2017.
- [96] Medical Advisory Secretariat, *Multidetector computed tomography for coronary artery disease screening in asymptomatic populations: evidence-based analysis.*, vol. 7, no. 3. 2007.
- [97] H. Q. Ontario, “Transcatheter aortic valve implantation in patients with severe aortic valve stenosis at low surgical risk: A health technology assessment,” *Ont. Health*

- Technol. Assess. Ser.*, vol. 20, no. 14, pp. 1–148, 2020.
- [98] A. for H. R. and Quality, “Comparative Effectiveness of Radiofrequency Catheter Ablation for Atrial Fibrillation,” *Agency Healthc. Res. Qual.*, 2009.
- [99] C. A. for D. and T. in Health, “Computed Tomography Angiography for Diagnosis of Stroke or Transient Isquemic Attack: A Review of the Clinical Effectiveness,” *Can. Agency Drugs Technol. Heal.*, no. November, 2016.
- [100] Agency for Healthcare Research and Quality (AHRQ), “Use of Cardiac Resynchronization Therapy,” 2019.
- [101] and N. D. Eva Suarthana,* Nisha Almeida, “Cardiac Resynchronization Therapy in Heart Failure,” 2016.
- [102] M. C. S. and A. P. T. M. Ms Sharon Kessels, Dr Ruchi Mittal, Dr Judy Morona, Ms Skye Newton, Mr Ben Ellery, Ms Joanne Milverton, Ms Jacqueline Parsons, Ms Arlene Vogan, “Magnetic resonance imaging of patients with suspected (non-ischaemic) dilated cardiomyopathies,” 2016.
- [103] S. D. P. Daniel A, Ollendorf, Alexander Gohler, “CORONARY COMPUTED TOMOGRAPHIC ANGIOGRAPHY FOR DETECTION OF CORONARY ARTERY DISEASE,” 2009.
- [104] M. A. Secretariat, “Ablation for atrial fibrillation,” 2006.
- [105] T. S. M. S. A. Committee, “Real time measurement of cardiac output and other cardiac flow parameters (without concurrent cardiac imaging) using continuous wave Doppler techniques,” 2009.
- [106] N. Calvert, D. Hind, C. Beverley, and A. Davidson, “The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation,” 2003.
- [107] SHTG, “Transcatheter aortic valve implantation (TAVI) for the treatment of patients with severe symptomatic aortic stenosis who are at intermediate surgical risk,” *SHTG Publ.*, no. 91, 2019.
- [108] C. A. for D. and T. in Health, “Transthoracic Echocardiography in Adult Patients with Ischaemic Stroke: A Review of the Diagnostic Yield and Cost Effectiveness,” 2014.
- [109] C. A. for D. and T. in Health, “Computed Tomography Angiography versus Computed Tomography for the Diagnosis and Management of Hyperacute Stroke: A Review of Comparative Clinical Evidence and Guidelines,” *Can. Agency Drugs Technol. Heal.*, 2013.
- [110] H. I. Scotland, “What is the clinical effectiveness, safety and cost effectiveness of endovascular therapy using mechanical thrombectomy devices for patients with acute ischaemic stroke?,” 2016.
- [111] G. Mowatt *et al.*, “Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction,” *Health Technol. Assess. (Rockv.)*, vol. 8, no. 30, 2004, doi: 10.3310/hta8300.
- [112] Medical Advisory Secretariat, “Cardiac magnetic resonance imaging for the diagnosis of coronary artery disease,” 2010.
- [113] S. Goodacre *et al.*, “Systematic review, meta-analysis and economic modelling of diagnostic strategies for suspected acute coronary syndrome,” *Health Technol.*

- Assess. (Rockv).*, vol. 17, no. 1, pp. 1–210, 2013, doi: 10.3310/hta17010.
- [114] W. Is, T. Health, T. Assessment, W. Did, T. Health, and T. Assessment, “Extracorporeal membrane oxygenation for cardiac indications in adults: A health technology assessment,” *Ont. Health Technol. Assess. Ser.*, vol. 20, no. 8, pp. 1–121, 2020.
- [115] AHRQ, “Esophageal Doppler Ultrasound-Based Cardiac Output Monitoring for Real-Time Therapeutic Management of Hospitalized Patients,” 2007.
- [116] M. Wah, “Remote monitoring of implantable cardioverter-defibrillators, cardiac resynchronization therapy and permanent pacemakers: A health technology assessment,” *Ont. Health Technol. Assess. Ser.*, vol. 18, no. 7, pp. 1–199, 2018.
- [117] Agency for Healthcare Research and Quality U.S. Department of Health and Human Services, “Intravascular Diagnostic Procedures and Imaging Techniques Versus Angiography Alone in Coronary Artery Stenting: Comparative Effectiveness Review,” 2013.
- [118] I. for C. and E. Review, “Cardiac Nuclear Imaging,” 2013.
- [119] O. H. T. Assessment, *Single photon emission computed tomography for the diagnosis of coronary artery disease: An evidence-based analysis*, vol. 10, no. 8. 2010.
- [120] Medical Advisory Secretariat, “Intracoronary radiation therapy,” 2001.
- [121] L. V. and L. M. D Hartwell, J Colquitt, E Loveman, AJ Clegg, H Brodin, N Waugh, P Royle, P Davidson, “Clinical effectiveness and cost- effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation,” 2005.
- [122] Medical Advisory Secretariat, “Stenting for Peripheral Artery Disease of the Lower Extremities,” 2010.
- [123] C. A. for D. and T. in Health, “Positron Emission Tomography for Cardiovascular Disease: A Review of the Clinical Effectiveness,” 2010.
- [124] Medical Advisory Secretariat, “Endovascular Repair of Descending Thoracic Aortic Aneurysm: Review of Literature,” 2005.
- [125] A. Sinclair *et al.*, “Surgical aortic valve replacement with the ATS Enable[®] sutureless aortic valve for aortic stenosis,” *Technol. Assess. Unit McGill Univ.*, 2013.
- [126] C. A. for D. and T. in Health, “Post-Discharge Cardiac Monitoring of Stroke Patients: Clinical Effectiveness and Guidelines,” 2015.
- [127] C. A. for D. and T. in Health, “Transcatheter Aortic Valve Replacement in Severe Aortic Stenosis: A Review of Comparative Durability and Clinical Effectiveness Beyond 12 Months,” *Can. Agency Drugs Technol. Heal.*, 2013.
- [128] M. Daniel E. Jonas, MD, MPH Shivani Reddy, MD, MSc Jennifer Cook Middleton, PhD Colleen Barclay, MPH Joshua Green, BA Claire Baker Gary Asher, MD, “Screening for Cardiovascular Disease Risk With Electrocardiography: An Evidence Review for the U.S. Preventive Services Task Force,” 2018.
- [129] O. H. T. Assessment, “Stress echocardiography for the diagnosis of coronary artery disease,” 2010.
- [130] C. F. Charlotte Wells, “Medical Imaging Prior to Endovascular Therapy for Ischemic Stroke : Clinical Effectiveness and Guidelines,” 2017.
- [131] M. S. A. Committee, “Cardiac MRI for myocardial stress perfusion and viability imaging in patients with known or suspected coronary artery disease,” 2016.

- [132] Medical services Advisory Committee, "Insertion, replacement or removal of a cardiac resynchronisation therapy device capable of defibrillation (CRT-D) for mild, moderate or severe chronic heart failure," no. June, 2013.
- [133] C. D. Hans Van Brabandt, Mattias Neyt, "Catheter ablation of atrial fibrillation," *Health Technol. Assess. (Rockv).*, pp. 253–278, 2013, doi: 10.5124/jkma.2006.49.9.837.
- [134] C. R. Rieiro, "Evaluación de la cirugía guiada por imagen: eficacia, seguridad e impacto económico de la resonancia magnética intraoperatoria," 2013.
- [135] H. O. Esmaili, M. Hadialijanvand, H. Doosti, and M. T. Shakeri, "Meta analysis in medical research," *Koomesh*, vol. 14, no. 1, pp. 39–46, 2012.
- [136] P. A. Perez-Toro, "PauPerezT/WEBERT: Word Embeddings using BERT," 2020, doi: 10.5281/ZENODO.3964244.
- [137] P. Stenetorp *et al.*, "brat: a Web-based Tool for NLP-Assisted Text Annotation," *In Proceedings of the Demonstrations Session at EACL 2012*, 2012. [Online]. Available: <https://brat.nlplab.org/about.html>.
- [138] B. Nye, J. J. Li, I. J. Marshall, and B. C. Wallace, "A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature," *arXiv Prepr. arXiv*, 2018.
- [139] C. for R. and Dissemination, "CDR's guidance for undertaking reviews in health care," *Centre for Reviews and Dissemination, University of York*, no. 479. pp. 6–16, 2009, doi: 10.1192/bjp.111.479.1009-a.
- [140] R. V. Prieto, Z. R. Luis, Y. Ofelia, and J. Morell, "Procedimiento para realizar pruebas de usabilidad Procedure for performing usability test."
- [141] International Organization for Standardization, "ISO / IEC 25010 System and software quality models." .
- [142] International Organization for Standardization, "ISO / IEC 2500n – SQuaRE." [Online]. Available: <https://iso25000.com/index.php/normas-iso-25000/iso-25010?id=10&limit=3>. [Accessed: 19-Jan-2022].
- [143] A. S. Kalpna Sagar, "A systematic review of software usability studies," *Int. J. Inf. Technol.*, 2017, doi: 10.1007/s41870-017-0048-1.
- [144] S. M. G, "Proceso de Pruebas de Usabilidad de Software Software Usability Testing Process Introducción," pp. 27–32, 2018.
- [145] T. L. J. Nielsen, "A Mathematical Model of the Finding of Usability Problems," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1993, doi: 10.1145/169059.169166.
- [146] U. S. D. of health and H. Services, "The Research-Based Web Design & Usability Guidelines, Enlarged/Expanded edition.," *Washington: U.S. Government Printing Office*, 2006. [Online]. Available: <https://www.usability.gov/about-us/index.html>. [Accessed: 19-Jan-2022].
- [147] PYPL, "PYPL PopularitY of Programming Language," *Pierre Carbonnelle*, 2022. [Online]. Available: <https://pypl.github.io/PYPL.html>. [Accessed: 13-Jan-2022].
- [148] TIOBE, "TIOBE Index for January 2022," *TIOBE WEB PAGE*, 2022. [Online]. Available: <https://www.tiobe.com/tiobe-index/>. [Accessed: 13-Jan-2022].
- [149] R. C. Limited, "PyQt5 5.15.6," *Riverbank Computing Limited*, 202AD. [Online]. Available: <https://pypi.org/project/PyQt5/>. [Accessed: 13-Jan-2022].

- [150] QT, "QT Company," *QT web page*, 2022. [Online]. Available: <https://www.qt.io/company#why>. [Accessed: 13-Jan-2022].
- [151] National Library of Medicine, "PubMed Overview," *Internet*, 2022. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/about/>.
- [152] N. L. of Medicine, "PMC Central," *Internet*, 2022. .
- [153] M. C. Zapata, J. G. Barreneche, and J. K. Nieto Aristizabal, "Automation of Study Design Classification and Clinical Evidence Ranking for Health Technology Assessment of Medical Devices," in *Applied Computer Sciences in Engineering*, 2021, pp. 190–201.

7. Anexos

7.1. Instructivo de uso

7.2. Prueba de usabilidad

7.3. Ejecución pruebas de usabilidad

7.4. Congresos y publicaciones

Presentación en Workshop on Engineering Applications 2021 como autora de:

Automation of Study Design Classification and Clinical Evidence Ranking for Health Technology Assessment of Medical Devices [153].

Zapata M.C., Barreneche J.G., Nieto Aristizabal J.K. (2021) Automation of Study Design Classification and Clinical Evidence Ranking for Health Technology Assessment of Medical Devices. In: Figueroa-García J.C., Díaz-Gutierrez Y., Gaona-García E.E., Orjuela-Cañón A.D. (eds) Applied Computer Sciences in Engineering. WEA 2021. Communications in Computer and Information Science, vol 1431. Springer, Cham. https://doi.org/10.1007/978-3-030-86702-7_17