[REV. MED. CLIN. CONDES - 2022; 33(6) 576-582]

REVISTA MÉDICA CLÍNICA LAS CONDES
https://www.journals.elsevier.com/revista-medica-clinica-las-condes



ARTÍCULO ESPECIAL

Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile

Natural language processing for clinical text in Spanish: The case of waiting lists in Chile

Pablo Báez[a], Antonia Paz Arancibia[b], Matías Ignacio Chaparro[b], Tomás Bucarey[b], Fredy Núñez[c,d], Jocelyn Dunstan[d,e,f]✉.

[a] Centro de Informática Médica y Telemedicina, Facultad de Medicina, Universidad de Chile. Santiago, Chile.
[b] Escuela de Medicina, Facultad de Medicina, Universidad de Chile. Santiago, Chile.
[c] Departamento de Ciencias del Lenguaje, Facultad de Letras, Pontificia Universidad Católica de Chile. Santiago, Chile.
[d] Centro de Modelamiento Matemático, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile. Santiago, Chile.
[e] Iniciativa de Datos & Inteligencia Artificial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile. Santiago, Chile.
[f] Instituto Milenio Engineering for Healthcare, ANID, Chile. Santiago, Chile.

INFORMACIÓN DEL ARTÍCULO

Historia del Artículo:
Recibido: 30 07 2022
Aceptado: 18 10 2022

Key words:
Natural Language Processing; Medical Informatics; Artificial Intelligence, Waiting Lists.

Palabras clave:
Procesamiento del Lenguaje Natural; Informática Médica; Inteligencia Artificial; Listas de Espera.

RESUMEN

Las listas de espera no cubiertas por el Plan de Garantías Explícitas en Salud para nueva consulta de especialidad en Chile se han visto incrementadas por los efectos de la pandemia del coronavirus SARS-CoV-2 (COVID-19). Esto representa un problema debido a la demora en la resolución y priorización de cada caso de derivación al nivel secundario de atención en salud. El objetivo de este artículo es exponer el problema de la lista de espera en el sistema de salud de Chile, y abordarlo como ejemplo de la aplicación de técnicas de Procesamiento del Lenguaje Natural (PLN). Específicamente, se describe una metodología para el reconocimiento de información clave en narrativas médicas. Actualmente, contamos con un conjunto de interconsultas médicas manualmente anotadas en el desarrollo del Corpus de Lista de Espera Chilena, y con una fracción de 2.000 interconsultas en las que las entidades médicas anotadas fueron normalizadas de forma automatizada a los conceptos del Sistema de Lenguaje Médico Unificado empleando el léxico MedLexSp. Este y otros recursos lingüísticos y herramientas de PLN están siendo desarrollados por el grupo de PLN en Medicina del Centro de Modelamiento Matemático de la Universidad de Chile y otros grupos a nivel nacional, los cuales constituyen aportes relevantes que pueden ser transferidos al sistema de salud chileno, con el objetivo de apoyar la gestión del texto clínico en español.

ABSTRACT

The waiting lists not covered by the Explicit Health Guarantee Plan for new specialty consultation in Chile increased due to the effects of the SARS-CoV-2 coronavirus (COVID-19) pandemic. This represents a problem derived from the delay in the resolution and prioritization of each case. This paper aims to describe the issue of the waiting lists in the Chilean health system and present an example of the application of Natural Language Processing (NLP). Specifically, a methodology for recognizing key information in medical narratives

✉ Autor para correspondencia
Correo electrónico: jdunstan@uchile.cl

https://doi.org/10.1016/j.rmclc.2022.10.002
e-ISSN: 2531-0186/ ISSN: 0716-8640/© 2021 Revista Médica Clínica Las Condes.
Este es un artículo Open Access bajo la licencia CC BY-NC-ND
(http://creativecommons.org/licenses/by-nc-nd/4.0/).

is described. Currently, we have a set of manually annotated medical referrals in the development of the Chilean Waiting List Corpus, with a fraction of 2,000 referrals in which the annotated medical entities were automatically normalized to the Unified Medical Language System concepts using the lexicon MedLexSp. The clinical NLP Group of the Center for Mathematical Modeling of the University of Chile, and other national NLP groups, are developing several tools and resources in medicine that can be transferred to the Chilean health system to support managing clinical text in Spanish.

INTRODUCCIÓN

Tal como ocurre en muchos países del continente y el mundo, el sistema de salud en Chile enfrenta problemáticas derivadas del envejecimiento poblacional, la mayor expectativa de vida, la multimorbilidad y el incremento en los servicios de salud, entre otras. Actualmente, diversos aspectos se han agravado debido a la pandemia del coronavirus SARS-CoV-2 (COVID-19), cuyas consecuencias no sólo han impactado negativamente en el ámbito económico, sino también en la salud pública del país. A la par, estamos experimentando una transformación digital sin precedentes, cuyo núcleo es la aplicación de la inteligencia artificial (en adelante IA) para generar herramientas que contribuyan a desarrollar un sistema de salud más eficiente, que cubra las necesidades de los pacientes con soluciones costo-efectivas.

Un problema de salud pública que se ha acrecentado significativamente por la pandemia y que puede ser abordado con el apoyo de la IA es el de las listas de espera. En particular, y tal como se explicará en detalle en las siguientes secciones, la razón de interconsulta en la lista de espera para nueva consulta de especialidad en hospitales públicos chilenos está escrita en texto no estructurado, un problema que ha motivado por varios años nuestro trabajo como grupo de Procesamiento de Lenguaje Natural en Medicina del Centro de Modelamiento Matemático de la Universidad de Chile (http://pln.cmm.uchile.cl).

En este artículo describimos el problema de las listas de espera en Chile y lo presentamos como ejemplo de la aplicación de procesamiento de lenguaje natural (en adelante PLN). Nos centraremos en cómo esta rama de la inteligencia artificial puede proporcionar soluciones a problemas de salud pública, y mostraremos otras tareas de PLN en el contexto clínico chileno.

EL SISTEMA DE SALUD CHILENO Y LAS LISTAS DE ESPERA

El sistema de salud en Chile se compone de un sistema mixto de atención integrado por el seguro público, el Fondo Nacional de Salud (en adelante FONASA), y uno privado formado por Instituciones de Salud Previsional (ISAPRES)[1]. Es de interés el sistema público, ya que al año 2019 concentraba el 78% de la población[2].

En el sistema público no se puede acceder directamente a una atención con un especialista. Cuando, a juicio del equipo de salud, se requiere la opinión de éste, se emite una solicitud de interconsulta junto con una derivación al nivel secundario de atención en salud. Es en estos centros donde se solicitan exámenes, se definen tratamientos y seguimiento. Ante la imposibilidad de atender esta consulta inmediatamente, surge el concepto de lista de espera[3]. En términos generales, las listas de espera pueden ser definidas desde dos perspectivas:

• Genéricamente, se definen como un conjunto de personas que, en un momento dado, se encuentran en espera de ser atendidas para una consulta de especialidad médica u odontológica, para un procedimiento o prueba diagnóstica o para una intervención quirúrgica programada, solicitada por un profesional médico u odontológico autorizado en la red de salud y teniendo documentada tal petición.

• Operacionalmente, se entienden como el universo de registros que no poseen una causal de salida y aquellos identificados por la causal de salida Nº3, es decir, aquellos pacientes que se encuentran en re-evaluación para el diagnóstico y por ende, no salen de las listas de espera[4].

Las Garantías Explícitas en Salud (en adelante, GES) corresponden a los beneficios, garantizados por ley, para la población afiliada tanto a las ISAPRES como a FONASA[5], correspondiente a un conjunto de 87 patologías. Algunos ejemplos de las patologías cubiertas por el GES son: enfermedad renal crónica etapa 4 y 5, diabetes mellitus tipo I y II, e hipertensión arterial[6].

Las garantías exigibles para las patologías GES son:
Acceso: es el derecho a recibir las atenciones correspondientes, según la Ley, para cada patología listada.
Oportunidad: tiempos máximos de espera para el otorgamiento de las prestaciones.
Protección financiera: la persona beneficiaria cancelará un porcentaje de la afiliación.
Calidad: otorgamiento de las prestaciones por un prestador acreditado o certificado[5].

El problema que presenta la lista de espera radica en la demora de resolución y priorización de cada caso, más que en el número de personas que conforman dicha lista[7]. Específicamente, durante el año 2016 fallecieron 16.625 personas esperando ser atendidas por un médico especialista para patologías no cubiertas por el GES, mientras que, en el mismo año, 993 personas fallecieron espe-

rando en una lista de espera GES. Lo anterior se estableció combinando tanto las listas de espera para consultas de especialidad como la lista quirúrgica[8]. Así, al 31 de marzo del 2021, el número de personas en lista de espera para patologías no GES ascendía a 1.932.422[9]. En consecuencia, tanto la alta mortalidad de la lista de espera no GES, como el tiempo que los pacientes esperan para ser atendidos, nos ha motivado como grupo a estudiar esta lista de espera en particular.

Esta situación es preocupante para la población chilena y las autoridades, por lo que deben buscarse medidas para revertir este tiempo prolongado de espera, junto con sus desfavorables consecuencias. Para abordar este problema, es importante identificar la dinámica en la que se registra la información. El médico tratante primero debe ingresar la información clínica del paciente a una plataforma de ficha clínica electrónica, la que contiene campos tanto de información estructurada como no estructurada, en donde se justifica la solicitud. La información estructurada puede ser categórica o numérica, mientras que la no estructurada corresponde a información de tipo imágenes, señales o texto[10]. Del total de la información, un 40% corresponde a las denominadas 'narrativas clínicas' o texto libre[11].

La información estructurada es de fácil análisis y comprensión, lo que posibilita generar documentos equivalentes y acotados, pero puede omitir información relevante respecto al proceso diagnóstico. Por otro lado, el texto libre podría contener faltas de ortografía, uso de abreviaciones, jerga local, entre otros, lo que dificulta su análisis. Su contenido, sin embargo, es de gran relevancia, ya que las narrativas contienen información detallada sobre enfermedades, su sintomatología y el fundamento clínico escrito por profesionales de la salud.

Al analizar las interconsultas que forman parte de la lista de espera no GES, una contribución a la resolución del problema podría ser la implementación de criterios de prioridad, la realización de estadísticas de la derivación, la determinación de los casos que pueden ser resueltos por medio de telemedicina, la identificación de factores de riesgo y, finalmente, la búsqueda de relaciones con los antecedentes familiares, entre otros. Todas estas tareas pueden ser agilizadas al utilizar técnicas de PLN[12,13].

PROCESAMIENTO DE LENGUAJE NATURAL (PLN)

El PLN es una rama de la inteligencia artificial que puede ser utilizada para el análisis del texto y discurso producido por humanos[14]. Específicamente, el PLN tiene como objetivos estudiar, diseñar y aplicar sistemas informáticos que faciliten la comunicación entre personas y, entre personas y máquinas[15,16]. En otras palabras, se persigue imitar artificialmente algunos de los aspectos de la capacidad humana para el lenguaje, lo que se traduce en procesos de producción y comprensión.

En definitiva, el PLN busca entregar soluciones a los problemas concretos derivados del intento por reproducir artificialmente los patrones con los que funcionan la mente y el lenguaje humanos, para transferirlos a la relación entre humanos y máquinas. Algunas de las tareas tradicionales que aborda el PLN son las siguientes:

1. Desarrollo de sistemas de diálogo: sistemas conversacionales que tienen por objetivo optimizar la precisión del proceso comunicativo, ya sea por medio de un chat o del reconocimiento acústico-fonético por parte de la máquina[17].

2. Extracción y recuperación de información: ámbito de investigación dedicado al desarrollo de procedimientos informáticos para encontrar documentos que contienen datos textuales no estructurados (es decir, que carecen de una estructura semánticamente abierta y fácil de usar por una máquina)[18].

3. Traducción automática o semiautomática: conjunto de técnicas para el desarrollo de programas especializados que buscan traducir datos textuales desde una lengua origen hacia una lengua meta, a partir de implementaciones estadísticas[19] o neuronales[20].

En cuanto a la aplicación del PLN en el ámbito de la medicina, algunas de las herramientas de mayor utilidad sirven para la extracción automática de información en documentos electrónicos, como lo son las fichas clínicas[21]. Una gran proporción de la información contenida en las fichas clínicas es no estructurada, lo que hace que su análisis sea de mayor dificultad pues no puede ser fácilmente resumida[22]. El procesamiento de las fichas clínicas permite asistir a los profesionales de la salud en estudios retrospectivos, en la toma de decisiones clínicas, obtención y comparación de datos sobre enfermedades poco frecuentes, creación de estadísticas e identificación de factores de riesgo[23], así como también en la evaluación de eficiencia y costos del cuidado de la salud[24].

Otras aplicaciones de PLN en medicina son: detección de reacciones adversas a medicamentos[25,26], detección de eventos y síntomas que preceden al diagnóstico de cáncer[27,28], evaluación del riesgo de suicidio en personas con depresión[29,30], asociación de comorbilidades y otras aplicaciones que pueden ayudar a elaborar estudios epidemiológicos[11]. Asimismo, cabe nombrar la utilidad en la asignación automática de códigos CIE-10 (u otras nomenclaturas)[31,32] al texto libre, normalizando diversas expresiones empleadas para describir la misma patología.

Uno de los enfoques más comunes en PLN en el ámbito médico es emplear el texto clínico (fichas clínicas o parte de ellas como reportes de imagenología, anatomía patológica o interconsultas, entre otros) para construir un *corpus*. Un *corpus* es un conjunto ordenado de datos o textos que sirve de base a una investigación, y que es representativo de un ámbito o variedad lingüística particular[33]. Estos *corpus* pueden enriquecerse con información interpretativa

mediante un proceso denominado anotación[34]. La información contenida en los textos es analizada por una persona con experiencia (anotador), quien identifica y señala la información clave (por ejemplo, las enfermedades que presentan los pacientes). Estos textos anotados pueden ser utilizados para entrenar sistemas de reconocimiento automatizado de dicha información clave[11].

Precisamente, la detección de entidades nombradas (NER por su sigla en inglés), consiste en el reconocimiento de clases semánticas en texto libre no estructurado[35]. La detección de entidades como enfermedades, signos y síntomas o procedimientos, son ejemplos de NER en el contexto médico. Sin embargo, gran parte del trabajo previo está desarrollado en la lengua inglesa, lo que junto a la carencia de recursos lingüísticos para la lengua española en el dominio clínico, supone una dificultad para el análisis de textos en esta lengua[22].

APORTES AL PROBLEMA DE LA LISTA DE ESPERA CON PLN

Nuestro grupo construyó el primer *corpus* de texto clínico chileno de libre acceso, a partir de datos de la lista de espera no cubierta por el Plan GES para nueva consulta de especialidad. Esta muestra está compuesta por 2.592.925 interconsultas integradas por las 50 especialidades médicas y dentales[22]. Adicionalmente, empleamos 10.000 interconsultas para construir un *corpus* manualmente anotado con hallazgos clínicos, resultados de test o de laboratorio, signos y síntomas, enfermedades, partes del cuerpo, medicamentos, abreviaciones, miembros de la familia y procedimientos de laboratorio, diagnósticos y de tratamiento[12].

El *corpus* anotado se empleó para entrenar un modelo de NER, el cual permite identificar de forma automática enfermedades, medicamentos y partes del cuerpo, con un F1-*score* de 82,9%, 84,3% y 86,5%, respectivamente. El F1-*score* es una métrica ampliamente utilizada en la estimación del rendimiento de sistemas de PLN o de recuperación de la información, la cual combina la precisión y el *recall* o exhaustividad[36] y permite hacer comparaciones entre diversas aproximaciones para resolver una tarea en cuestión. Nuestro objetivo es que dicho *corpus*, que actualmente se encuentra disponible gratuitamente para uso no comercial (https://zenodo.org/record/5591011#.Yt3-K-zMI-R), permita reducir la lista de espera mediante la identificación de casos que, a su vez, pueden ser solucionados mediante otras estrategias tales como telemedicina y la priorización de pacientes mediante análisis de comorbilidades. Otro aspecto importante que queremos abordar en futuras investigaciones y colaboraciones con el Ministerio de Salud es la identificación de procedimientos pendientes.

CODIFICACIÓN AUTOMÁTICA

Una de las dificultades en el análisis de texto no estructurado radica en la variabilidad del lenguaje que puede ser empleado para referirse a una misma entidad; específicamente, casos de ambigüedad en la asignación de diferentes etiquetas lingüísticas para el mismo sentido. Por ejemplo, las expresiones «cáncer de mama», «tumor mamario», «ca. mama», «tu. mamario», «carcinoma mamario», pueden ser empleadas, indistintamente, para hacer alusión a la misma enfermedad. Una de las soluciones para reducir esta variabilidad es la normalización de los términos médicos, mediante la asignación de, por ejemplo, su código correspondiente desde los conceptos del Sistema de Lenguaje Médico Unificado (UMLS por su sigla en inglés).

El UMLS *Metathesaurus*, desarrollado por la Biblioteca Nacional de Medicina de los Estados Unidos[37], es una herramienta creada principalmente para resolver dos barreras importantes frente a la capacidad de las máquinas para extraer información: la variedad de nombres para referirse al mismo concepto, como ya lo hemos mencionado, y la ausencia de un formato establecido para distribuir terminologías. Esta herramienta contiene un compilado de nombres, relaciones e información asociada de una variedad de sistemas biomédicos que integra más de dos millones de nombres de aproximadamente 900.000 conceptos de vocabulario médico, y no sólo eso, sino que también posee más de 12 millones de relaciones entre todos estos conceptos.

Actualmente, contamos con una fracción de 2.000 interconsultas médicas anotadas en el *Corpus* de Lista de Espera Chilena, cuyas entidades médicas fueron normalizadas de forma automatizada empleando el léxico MedLexSp[38,39] asignándole uno o múltiples códigos únicos de identificación a cada entidad. Este recurso estará disponible para su libre uso próximamente.

OTROS SISTEMAS DE PLN CLÍNICO DESARROLLADOS EN CHILE

El Plan Nacional del Cáncer 2018-2028 en Chile propone una línea estratégica para fortalecer los Sistemas de Registro, Información y Vigilancia del Cáncer. A partir de lo anterior, el PLN puede apoyar la extracción automática de información y la sistematización de bases de datos para este fin. Recientemente, hemos desarrollado un sistema de apoyo que facilita la codificación CIE-O de la morfología y topografía de los tumores en los informes de patología, una tarea esencial para los registros de cáncer[40]. Este sistema puede probarse en el siguiente enlace: https://topomorfo.oncodata.org.

Otro de nuestros desarrollos en esta área permite detectar automáticamente menciones de metástasis a distancia en reportes de imagenología. Específicamente, la clasificación TNM de tumores aporta información relevante en la definición de un estadio, definiendo las características del tumor primario (T), la posible propagación a ganglios linfáticos cercanos (N) y la presencia o ausencia de metástasis a otras partes del cuerpo (M). Como antecedente, es relevante considerar que la detección manual de los parámetros TNM consume mucho tiempo y horas de personal, pues implica

el análisis individual de cada reporte. Tanto este desarrollo como el anterior se enmarcan en una colaboración con la Fundación Arturo López Pérez, considerado como el mayor centro oncológico de Chile, que atiende a más de 50 mil pacientes al año, la mayoría pertenecientes al sistema público de salud[41].

Un algoritmo de clasificación que emplea técnicas de PLN e historias clínicas anonimizadas de un hospital de Chile fue desarrollado por Ramos et al.[42] para clasificar los diagnósticos de los pacientes discriminando entre las clases 'cáncer' frente a 'no cáncer' y 'cáncer de mama' frente a 'otro cáncer'. Este algoritmo podría utilizarse como herramienta de apoyo y recomendación del diagnóstico de los pacientes, principalmente para médicos que inician sus labores en sectores alejados con poco personal especializado como, por ejemplo, hospitales rurales.

Existen varios trabajos enfocados en el uso secundario de datos médicos y clasificación de textos desarrollados por investigadores de la Universidad de Concepción[43-45]. En esta propuesta se emplearon textos clínicos en español, provenientes del Hospital Clínico Regional Dr. Guillermo Grant Benavente de Concepción, para identificar y extraer información sobre el estado de tabaquismo de los pacientes, mediante técnicas de PLN y minería textual[46], junto con información sobre medidas de peso corporal y comorbilidades[47]. En ambos casos, los datos extraídos fueron posteriormente utilizados como *corpus* para algoritmos de clasificación de textos. A partir de esta implementación, por ejemplo, fue posible determinar si un paciente era fumador, no fumador, fumador actual o fumador pasado.

Lecaros et al.[48], por su parte, examinaron derivaciones contenidas en el *corpus* de la lista de espera para detectar casos de pacientes con psoriasis y, así, determinar la incidencia de dicha patología en Chile. Recientemente, Figueroa-Barra et al.[49] emplearon técnicas de PLN dentro de un sistema de análisis automático del lenguaje para identificar y predecir esquizofrenia en el primer episodio de psicosis. Este sistema se empleó en entrevistas clínicas en español y se basa en el análisis de 30 rasgos lingüísticos que permiten distinguir los controles sanos de los pacientes con esquizofrenia crónica, y predecir el diagnóstico de esquizofrenia en pacientes con primer episodio de psicosis.

CONCLUSIONES

El texto clínico representa una proporción importante de la información registrada de los y las pacientes. Su procesamiento masivo y utilización en la toma de decisiones requiere herramientas del estado del arte en PLN; correspondiente a una de las tres ramas de la inteligencia artificial, junto con la visión por computador y la robótica.

Un área que hemos comenzado a desarrollar es el estudio de las propiedades lingüísticas de distintos textos clínicos, lo que es relevante para desarrollar sistemas en diversas especialidades médicas. En efecto, el reconocimiento de patrones lingüísticos presenta implicaciones significativas en términos de la generalización, ya que las herramientas de PLN suelen adaptarse de manera eficiente a dominios estrechos, con lenguajes bien definidos y comprendidos[50]. Además, nuestro equipo colabora con grupos en España y Argentina, con el fin de aunar esfuerzos para la construcción de recursos lingüísticos que consideren la variedad de estructuras presentes en la lengua española, y de esta manera evaluar sus similitudes o diferencias.

Para avanzar en el uso del PLN en español es necesario obtener narrativas clínicas, frecuentemente anotadas por humanos, así como modelos computacionales que permitan realizar la tarea en cuestión. Ejemplos de tareas relevantes de PLN clínico son: la detección de información clave, la clasificación de textos o la asignación de códigos internacionales. En el presente artículo, mostramos ejemplos de los avances en estas tareas, lideradas por nuestro grupo de investigación y otros grupos nacionales.

Según lo anterior, recibir apoyo político y gubernamental es esencial para que trabajos como estos puedan ser transferidos al sistema de salud chileno. En particular, desde octubre de 2022 hemos comenzado a colaborar con el Departamento de Estadísticas e Información de Salud (DEIS) del Ministerio de Salud. El objetivo de esta colaboración es transferir nuestro conocimiento adquirido en el trabajo con texto libre de la lista de espera no GES para apoyar la gestión de ésta. En particular, nos hemos propuesto codificar todas las enfermedades mencionadas en las razones de interconsultas, y con ello apoyar estrategias tales como la detección de pacientes GES en la lista no GES, el Registro Nacional de Cáncer o la articulación con el Hospital Digital para el uso de la telemedicina.

Un desafío actual en la investigación en PLN clínico es lograr acceso a narrativas médicas preservando la privacidad de los pacientes y las regulaciones cada vez más estrictas en algunos países. Una solución es la creación de un *corpus* sintético que preserve las propiedades lingüísticas del *corpus* original, pero cuya información no guarde relación con los pacientes reales[51]. Este enfoque es muy promisorio, dado que podría corregir las graves brechas de representación de grupos identitarios[52].

Varias condiciones deben darse para que nuevas tecnologías como el PLN sean utilizadas tanto en el sistema público como el privado. Obviamente, es necesaria la voluntad política, por un lado, y el acceso a poder de cómputo, por otro. Además, se debe promover una estrategia que permita asegurar un uso ético de la IA. En efecto, la Organización Mundial de la Salud (https://www.who.int/publications/i/item/9789240029200) recomienda seis principios que deben tenerse en cuenta, los que incluyen: no perder la autonomía humana en la toma de decisiones, incluir medidas de mejoramiento continuo, explicabilidad de los modelos, responsabilidad

sobre las tareas que está realizando la máquina, garantizar un uso no-discriminatorio y sustentabilidad. En definitiva, el cumplimiento de todas estas condiciones no es una tarea sencilla. No obstante, contar con ellas como meta constituye una ayuda relevante para crear consciencia y promover el trabajo interdisciplinario y el mejoramiento continuo de los modelos.

PLN en medicina es un área transdisciplinaria por definición, en la que confluyen las ciencias de la computación, la medicina y la lingüística. Para lograr un mayor impacto, se requiere del trabajo conjunto de profesionales de la salud y tomadores de decisión que conozcan y valoren los alcances de la inteligencia artificial aplicada al ámbito de la salud.

Declaración de conflicto de interés
Los autores declaran no tener conflictos de intereses.

Financiamiento
Este trabajo ha sido financiado por la ANID a través de los Fondos Basales para Centros de Excelencia FB210005 (Centro de Modelamiento Matemático), Fondecyt de Iniciación 11201250 (J. Dunstan) y Fondecyt de Postdoctorado 3210395 (P. Báez). Además, la investigación conducida por J. Dunstan es apoyada por los Institutos Milenio ICN2021_004 (iHealth) e ICN17_002 (IMFD).

REFERENCIAS BIBLIOGRÁFICAS

1. *Cómo funciona el Sistema de Salud en Chile - Cómo funciona el Sistema de Salud en Chile [Internet]. Orientación en Salud. Superintendencia de Salud, Gobierno de Chile. [citado el 23 de julio de 2022]. Disponible en: http://www.supersalud.gob.cl/difusion/665/w3-article-17328.html*
2. *Informe CDD: Caracterización sociodemográfica y socioeconómica en la población asegurada inscrita. Gobierno de Chile. [Internet]. División de Desarrollo Institucional, Departamento de Estudios y Estadísticas; 2020 [citado el 23 de julio de 2022]. Disponible en: https://www.fonasa.cl/sites/fonasa/adjuntos/Informe_caracterizacion_poblacion_asegurada*
3. *Estay R, Cuadrado C, Crispi F, González F, Alvarado F, Cabrera N. Desde el conflicto de listas de espera, hacia el fortalecimiento de los prestadores públicos de salud: Una propuesta para Chile. Cuad Med Soc. 2017;57(1):49-63.*
4. *Cordero A, Valderrama S, Duran L, Santelices P, Olave M, Jiron P, et al. Norma técnica para el registro de las listas de espera. 2016; Disponible en: https://www.minsal.cl/sites/default/files/files/Nueva%20Norma%20de%20Listas%20de%20Espera%202011.pdf*
5. *¿Qué es el AUGE o GES? - ¿Qué es el AUGE o GES? [Internet]. Preguntas Frecuentes. Superintendencia de Salud, Gobierno de Chile. [citado el 18 de julio de 2022]. Disponible en: http://www.supersalud.gob.cl/consultas/667/w3-article-4605.html*
6. *Garantías Explícitas en Salud (GES) [Internet]. Orientación en Salud. Superintendencia de Salud, Gobierno de Chile. [citado el 18 de julio de 2022]. Disponible en: http://www.supersalud.gob.cl/difusion/665/w3-propertyvalue-1962.html*
7. *Martinez DA, Zhang H, Bastias M, Feijoo F, Hinson J, Martinez R, et al. Prolonged wait time is associated with increased mortality for Chilean waiting list patients with non-prioritized conditions. BMC Public Health. 2019;19(1):233. doi: 10.1186/s12889-019-6526-6*
8. *MINSAL. Estado de Situacion Personas Fallecidas en Listas de Espera No-GES y Garantias Retrasadas GES [Internet]. 2017 [citado el 25 de julio de 2022]. Disponible en: https://www.minsal.cl/wp-content/uploads/2018/01/Informe-Final-Comision-Asesora-LE-y-Garantias-Retrasadas-GES-17082017_.pdf*
9. *MINSAL. Glosa 06 - Lista de espera no GES y garantías de oportunidad GES retrasadas [Internet]. 2021 [citado el 23 de julio de 2022]. Disponible en: https://www.minsal.cl/wp-content/uploads/2021/05/ORD-1204-DIGERA-Glosa-06.pdf*
10. *Villena F, Dunstan J. Obtención automática de palabras clave en textos clínicos: una aplicación de procesamiento del lenguaje natural a datos masivos de sospecha diagnóstica en Chile. [Automatic keyword retrieval from clinical texts: an application of natural language processing to massive data of Chilean suspected diagnosis]. Rev Med Chile. 2019;147(10):1229-1238.*
11. *Dalianis H. Clinical text mining: Secondary use of electronic patient records. Springer Nature; 2018.*
12. *Báez P, Villena F, Rojas M, Durán M, Dunstan J. The Chilean Waiting List Corpus: a new resource for clinical named entity recognition in Spanish. In: Proceedings of the 3rd Clinical Natural Language Processing Workshop. Association for Computational Linguistics. 2020. p.291-300. doi 10.18653/v1/2020.clinicalnlp-1.32*
13. *Báez P, Bravo-Marquez F, Dunstan J, Rojas M, Villena F. Automatic Extraction of Nested Entities in Clinical Referrals in Spanish. ACM Trans Comput Healthc Health. 2022;3(3):1-22. doi: 10.1145/3498324*
14. *Jurafsky D, Martin JH. Speech and Language Processing [Internet]. Third Ed. draft. 2022 [citado el 18 de julio de 2022]. Disponible en: https://web.stanford.edu/~jurafsky/slp3/ed3book_jan122022.pdf*
15. *Periñán Pascual JC, Mairal Usón R. Bringing Role and Reference Grammar to natural language understanding. Proces Leng Nat. 2009;(43):265-273.*
16. *Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol. 2020;9(2):14. doi: 10.1167/tvst.9.2.14*
17. *Gerbino E, Baggia P, Giachin E, Rullent C. Analysis and Evaluation of Spontaneous Speech Utterances in Focused Dialogue Contexts. Proc. of ESCA Workshop. 1995.*
18. *Savoy J, Gaussier É. Information Retrieval. In: Indurkhya N, Damerau FJ. (eds). Handbook of Natural Language Processing. Second Ed. Chapman and Hall/CRC; 2010.*
19. *Sinhal RA, Gupta KO. Machine translation approaches and design aspects. IOSR J Comput Eng. 2014;16(1):22-25.*
20. *Dabre R, Chu C, Kunchukuttan A. A Survey of Multilingual Neural Machine Translation. ACM Comput Surv. 2020;53(5):99:1-99:38. doi: 10.1145/3406095*

21. *Oronoz M, Gojenola K, Pérez A, de Ilarraza AD, Casillas A. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. J Biomed Inform. 2015;56:318-32. doi: 10.1016/j.jbi.2015.06.016*
22. *Báez P, Villena F, Zúñiga K, Jones N, Fernández G, Durán M, et al. Construcción de recursos de texto para la identificación automática de información clínica en narrativas no estructuradas. [Construction of text resources for automatic identification of clinical information in unstructured narratives]. Rev Med Chil. 2021;149(7):1014-1022. Spanish. doi: 10.4067/s0034-98872021000701014*
23. *Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical Natural Language Processing in languages other than English: opportunities and challenges. J Biomed Semantics. 2018;9(1):12. doi: 10.1186/s13326-018-0179-8*
24. *Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc. 2019;26(4):364-379. doi: 10.1093/jamia/ocy173*
25. *Jagannatha A, Liu F, Liu W, Yu H. Overview of the First Natural Language Processing Challenge for Extracting Medication, Indication, and Adverse Drug Events from Electronic Health Record Notes (MADE 1.0). Drug Saf. 2019;42(1):99-111. doi: 10.1007/s40264-018-0762-z*
26. *Chen L, Gu Y, Ji X, Sun Z, Li H, Gao Y, Huang Y. Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning. J Am Med Inform Assoc. 2020;27(1):56-64. doi: 10.1093/jamia/ocz141*
27. *Weegar R, Kvist M, Sundström K, Brunak S, Dalianis H. Finding Cervical Cancer Symptoms in Swedish Clinical Text using a Machine Learning Approach and NegEx. AMIA Annu Symp Proc. 2015;2015:1296-1305.*
28. *Weegar R. Mining events preceding a cancer diagnosis. En IEEE Computer Society; 2018:295-296. doi: 10.1109/eScience.2018.00059*
29. *Orooji A, Langarizadeh M. Using of natural language processing techniques in suicide research. Emerg Sci J. 2017;1(2):89-96. doi: 10.28991/esj-2017-01120*
30. *Levis M, Leonard Westgate C, Gui J, Watts BV, Shiner B. Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models. Psychol Med. 2020:1-10. doi: 10.1017/S0033291720000173*
31. *Henry S, Wang Y, Shen F, Uzuner O. The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. J Am Med Inform Assoc. 2020;27(10):1529-1537. doi: 10.1093/jamia/ocaa106. Erratum in: J Am Med Inform Assoc. 2021;28(11):2546.*
32. *Blanco A, Perez-de-Viñaspre O, Pérez A, Casillas A. Boosting ICD multi-label classification of health records with contextual embeddings and label-granularity. Comput Methods Programs Biomed. 2020;188:105264. doi: 10.1016/j.cmpb.2019.105264.*
33. *Dunstan J, Maass A, Tobar F. Una mirada a la era de los datos. 1 ed. Ed. Universitaria SA; 2022. 134 p.*
34. *Fort K. Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects. John Wiley & Sons; 2016.*
35. *Pérez A, Weegar R, Casillas A, Gojenola K, Oronoz M, Dalianis H. Semi-supervised medical entity recognition: A study on Spanish and Swedish clinical corpora. J Biomed Inform. 2017;71:16-30. doi: 10.1016/j.jbi.2017.05.009.*
36. *Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada DE, Fernández-Luna JM (eds). Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-31865-1_25*
37. *Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. Bull Med Libr Assoc. 1993;81(2):217-222.*
38. *Campillos-Llanos L. Medical Lexicon for Spanish (MedLexSp). 2022. doi: 10.20350/digitalCSIC/14656*
39. *Campillos-Llanos L. First steps towards building a medical Lexicon for Spanish with linguistic and semantic information. In Proceedings of the 18th BioNLP Workshop and Shared Task. 2019:152-164. doi:10.18653/v1/W19-5017*
40. *Villena F, Báez P, Peñafiel S, Rojas M, Paredes I, Dunstan J. Automatic Support System for Tumor Coding in Pathology Reports in Spanish. SSRN Electron J [Internet]. 2021. doi:10.2139/ssrn.3982259. Disponible en: https://www.ssrn.com/abstract=3982259*
41. *Memoria de Gestión FALP [Internet]. Fundación Arturo López Pérez. 2017 [citado el 18 de julio de 2022]. Disponible en: https://www.institutoncologicofalp.cl/fundacion/memoria-falp/*
42. *Ramos AA, Allende-Cid H, Taramasco C, Becerra C, Figueroa RL. Application of Machine Learning and Word Embeddings in the Classification of Cancer Diagnosis Using Patient Anamnesis. IEEE Access. 2020;8:106198-106213.*
43. *Flores CA, Figueroa RL, Pezoa JE. Active Learning for Biomedical Text Classification Based on Automatically Generated Regular Expressions. IEEE Access. 2021;9:38767-38777. doi: 10.1109/ACCESS.2021.3064000*
44. *Flores CA, Figueroa RL, Pezoa JE, Zeng-Treitler Q. CREGEX: A Biomedical Text Classifier Based on Automatically Generated Regular Expressions. IEEE Access. 2020;8:29270-29280. doi: 10.1109/ACCESS.2020.2972205*
45. *Flores CA, Figueroa RL, Pezoa JE. FREGEX: A Feature Extraction Method for Biomedical Text Classification using Regular Expressions. Annu Int Conf IEEE Eng Med Biol Soc. 2019;2019:6085-6088. doi: 10.1109/EMBC.2019.8857471*
46. *Figueroa RL, Soto DA, Pino EJ. Identifying and extracting patient smoking status information from clinical narrative texts in Spanish. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:2710-2713. doi: 10.1109/EMBC.2014.6944182*
47. *Figueroa RL, Flores CA. Extracting Information from Electronic Medical Records to Identify the Obesity Status of a Patient Based on Comorbidities and Bodyweight Measures. J Med Syst. 2016;40(8):191. doi: 10.1007/s10916-016-0548-8*
48. *Lecaros C, Dunstan J, Villena F, Ashcroft DM, Parisi R, Griffiths CEM, et al. The incidence of psoriasis in Chile: an analysis of the National Waiting List Repository. Clin Exp Dermatol. 2021;46(7):1262-1269. doi: 10.1111/ced.14713*
49. *Figueroa-Barra A, Del Aguila D, Cerda M, Gaspar PA, Terissi LD, Durán M, et al. Automatic language analysis identifies and predicts schizophrenia in first-episode of psychosis. Schizophrenia (Heidelb). 2022;8(1):53. doi: 10.1038/s41537-022-00259-3*
50. *Zeng QT, Redd D, Divita G, Jarad S, Brandt C, Nebeker JR. Characterizing clinical text and sublanguage: A case study of the VA clinical notes. J Health Med Informat S. 2011;S3. doi:10.4172/2157-7420.S3-001*
51. *Bannour N, Wajsbürt P, Rance B, Tannier X, Névéol A. Privacy-preserving mimic models for clinical named entity recognition in French. J Biomed Inform. 2022;130:104073. doi: 10.1016/j.jbi.2022.104073*
52. *Criado-Perez C. La mujer invisible: descubre cómo los datos configuran un mundo hecho por y para los hombres. España; Seix Barral. 2020.*