



Original

La era del big data: análisis del lenguaje natural mediante la aplicación de folksonomía

Laia Sans^a, Ismael Vallvé^b, Joan Teixidó^b, Josep Manel Picas^b, Jordi Martínez-Roldán^c y Julio Pascual^{a,*}

^a Servicio de Nefrología, Hospital del Mar, Barcelona, España

^b Bismart, Barcelona, España

^c Dirección de Innovación y Transformación Digital, Hospital del Mar, Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 28 de mayo de 2020

Aceptado el 15 de septiembre de 2021

On-line el 30 de octubre de 2021

Palabras clave:

Big data

Folksonomía

Procesamiento del lenguaje natural

Nefrología

R E S U M E N

Antecedentes y objetivo: Gran parte de la información médica que se deriva de la práctica clínica habitual queda recogida en forma de lenguaje natural en los informes médicos. Clásicamente, la extracción de información clínica para su posterior análisis a partir de los informes médicos requiere de la lectura y revisión manual de cada uno de ellos con la consiguiente inversión de tiempo. El objetivo de este proyecto piloto ha sido evaluar la utilidad de la folksonomía para la extracción y análisis rápido de los datos que contienen los informes médicos.

Material y métodos: En este proyecto piloto hemos utilizado la folksonomía para el análisis y la rápida extracción de datos de 1.631 informes médicos de alta de hospitalización del Servicio de Nefrología del Hospital del Mar sin necesidad de crear una base de datos estructurada previamente.

Resultados: A partir de determinadas preguntas sobre la práctica médica habitual (tratamiento hipoglicemiante de los pacientes diabéticos, tratamiento antihipertensivo y manejo de los inhibidores del sistema renina angiotensina durante el ingreso en nefrología y análisis de datos relacionados con la esfera emocional de los pacientes renales) la herramienta ha permitido estructurar y analizar la información contenida en texto libre en los informes de alta.

Conclusiones: La aplicación de folksonomía a los informes médicos nos permite transformar la información contenida en lenguaje natural en una serie de datos estructurados y analizables de manera automática sin necesidad de proceder a la revisión manual de los mismos.

© 2021 Sociedad Española de Nefrología. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Autor para correspondencia.

Correo electrónico: julpascual@gmail.com (J. Pascual).

<https://doi.org/10.1016/j.nefro.2021.09.006>

0211-6995/© 2021 Sociedad Española de Nefrología. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The big data era: The usefulness of folksonomy for natural language processing

ABSTRACT

Keywords:

Big data
Folksonomy
Natural language processing
Nephrology

Background: A huge amount of clinical data is daily generated and it is usually filed in clinical reports as natural language. Data extraction and further analysis requires reading and manual review of each report, which is a time consuming process. With the aim to test folksonomy to quickly obtain and analyze the information contained in medial reports we set up this study.

Methods and objectives: We have used folksonomy to quickly obtain and analyse data from 1631 discharge clinical reports from Nephrology Department of Hospital del Mar, without the need to create an structured database.

Results: After posing some questions related to daily clinical practice (hypoglycaemic drugs used in diabetic patients, antihypertensive drugs and the use of renin angiotensin blockers during hospitalisation in the nephrology department and data related to emotional environment of patients with chronic kidney disease) this tool has allowed the conversion of unstructured information in natural language into a structured pool of data for its further analysis.

Conclusions: Folksonomy allows the conversion of the information contained in clinical reports as natural language into a pool of structured data which can be further easily analysed without the need of the classical manual review of the reports.

© 2021 Sociedad Española de Nefrología. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

El término *big data* hace referencia a una cantidad de datos cuyo volumen, variabilidad y velocidad necesaria de procesamiento hace muy complejo su análisis utilizando los sistemas manuales o software estándar para su manejo^{1,2}.

En el ámbito sanitario se generan cada día millones de datos derivados de la asistencia a los pacientes. Hoy en día y en nuestro entorno, la utilización de la historia clínica electrónica es generalizada y los avances tecnológicos pueden facilitarnos el análisis de los datos recogidos. El análisis de los datos de las historias clínicas permite realizar un control de calidad de las actuaciones médicas, así como obtener datos observacionales de cohortes de pacientes para generar evidencia científica y seleccionar individuos con determinadas características para proponer su participación en ensayos clínicos.

Si bien algunos datos de los obtenidos son numéricos (datos de laboratorio o de la toma de constantes) la mayor parte de ellos quedan recogidos en la historia clínica de los pacientes en forma de lenguaje natural (por ejemplo, los datos obtenidos de la anamnesis del paciente, de la exploración física, del tratamiento, de las diversas exploraciones complementarias o los propios diagnósticos). Transformar todos estos valiosos datos recogidos en lenguaje natural en una serie de datos estructurados implica una importante inversión de tiempo, ya que requiere de un trabajo manual consistente en la lectura de la historia clínica, la identificación y obtención de los datos que previamente se hayan considerado de interés, la creación y alimentación de bases de datos que deben estructurarse en datos cuantitativos (se requiere la transformación de la información recogida en lenguaje natural en variables numéricas)

y, finalmente, el análisis de estos datos. Este proceso, además de suponer un consumo importante de tiempo, no permite el reanálisis de nuevos datos una vez que han quedado recogidos los parámetros considerados de interés en el proyecto inicial, a menos que se inicie de nuevo el proceso de revisión y obtención manual de información. Este hecho hace también imposible el reanálisis en tiempo real de las nuevas historias clínicas que se vayan generando y que sean de interés para un proyecto particular; cualquier reconsideración conlleva rehacer todo el proceso manual.

La tecnología del *natural language understanding* (NLU) o *natural language processing* (NLP) es de gran ayuda para poder convertir de manera rápida y automática toda la información recogida en texto libre en una estructura ordenada, y así poder proceder a un análisis mucho más rápido de toda la información.

La mayoría de los sistemas que realizan NLU o NLP requieren de una ontología o entidad maestra para, posteriormente, analizar los documentos³. Es decir, es necesario decidir con anticipación cuáles son los términos o etiquetas de interés (antes de iniciar la obtención de datos) y, por lo tanto, no permiten descubrir ningún término que no esté ya dispuesto en la ontología (distribución *top down*).

La utilización de la **folksonomía** (procede de los términos «folk» y «taxonomía») permite la obtención de información contenida en texto libre sin la necesidad previa de generar una entidad maestra de términos de interés, lo que aporta una evidente ventaja sobre los sistemas habituales de NLP. Esta solución de analítica avanzada transforma los documentos de texto no estructurado en documentos de texto estructurado pudiendo permitir el descubrimiento de información sin requerir de un proyecto inicial cerrado de términos de



Figura 1 – La solución propuesta por Bismart se basa en un flujo de datos que comienza con la incorporación en la base de datos de conocimiento de los documentos PDF aportados por Hospital del Mar. Sobre estos documentos se aplican procesos de OCR y la definición de los campos que queremos importar de cada documento. Una vez almacenados en la base de datos e identificados los campos, el sistema inicia el proceso de folksonomía, detectando palabras o grupos de palabras importantes en el conjunto de documentos. Una vez que la herramienta Folksonomy ha extraído la información que se necesita para trabajar, ésta se presenta en una Web para que pueda ser consultada, modificada o para ejecutar de nuevo el proceso a petición.

búsqueda antes del inicio de la obtención de información. Por tanto, la folksonomía permitiría resaltar automáticamente las etiquetas de conceptos en lenguaje natural para revelar el contenido interno. La **folksonomía** es un sistema automático de clasificación en tiempo real, basado en las etiquetas y la frecuencia con la que aparecen y es la única forma viable de poder trabajar con cantidades enormes de documentos. La forma de trabajar de este sistema se conoce como *bottom up* y la solución Bismart Folksonomy es el primer software que puede gestionar este tipo de clasificación (<https://bismart.com/es/inicio/>).

La utilización de algoritmos de NLP junto con la folksonomía en el ámbito médico permitiría no invertir más tiempo en la generación de bases de datos que el que requiere la actividad asistencial habitual y el análisis en tiempo real de los nuevos datos recogidos. Así, el *big data*, aportaría beneficios significativos al sector médico⁴.

Si bien existen publicaciones sobre la utilización de NLP en el ámbito médico, no existen en nuestro conocimiento experiencias previas de la aplicación de la folksonomía para la obtención de datos en el ámbito de la medicina en general, ni en el de la Nefrología en particular. En este artículo reportamos la primera experiencia piloto en la utilización de folksonomía junto con inteligencia artificial en NLP para, a partir de algunas preguntas eminentemente relacionadas con la práctica médica habitual, analizar los datos clínicos de los informes de altas hospitalarias del Servicio de Nefrología del Hospital del Mar de Barcelona en un determinado periodo y comprobar el rendimiento de este sistema para el análisis automático de datos.

Metodología

Datos

Se recogieron 1.631 informes de altas de hospitalización del Servicio de Nefrología del Hospital del Mar entre los años 2016 y 2018. Se procedió a la anonimización de los documentos en formato PDF desde un equipo situado en el Hospital del

Mar, donde un programador de Bismart tuvo acceso físico sin conexión a internet para proceder con el eliminando de los encabezado de los mismos dónde quedan recogidos todos los datos de filiación de los pacientes mediante un proceso automatizado desarrollado por Bismart (Barcelona) utilizando lenguaje Python, pudiendo únicamente identificar el género, necesario para la posterior aplicación de fórmulas de estimación de filtrado glomerular. Una vez cortadas las cabeceras de los documentos de altas médicas asegurando su anonimización, se cargaron en el portal de Bismart Folksonomy, procediendo a su conversión a texto mediante un sistema OCR de Microsoft Cognitive Systems, y mediante algoritmos de detección de patrones se separaron los distintos campos con base en los apartados de los informes (diagnósticos, motivo de consulta, antecedentes personales, tratamiento habitual, exploraciones complementarias, evolución y tratamiento al alta), para posteriormente almacenar la información ya anonimizada en la nube del Hospital del Mar. A continuación, se realizaron procesos de normalización de datos, lematización, se aplicaron algoritmos, se folksonomizó y se instaló el portal web Bismart Folksonomy. Bismart dispone de un proceso automático que despliega en la nube escogida una máquina virtual, la base de datos y todos los servicios necesarios para el portal Bismart Folksonomy, de manera que el despliegue técnico es un proceso relativamente sencillo (fig. 1). Dicho portal cumple todas las normativas GDPR (*general data protection regulation*) existiendo un registro de accesos y de modificaciones o consultas que se hacen de los datos, quedando la hora, el usuario y la dirección IP registrados en un log de la herramienta.

En los documentos aparecían términos médicos y acrónimos propios de la especialidad. Además, los informes estaban indistintamente escritos en dos idiomas (catalán y castellano) e incluso en un mismo informe podían aparecer términos en ambos idiomas. Esto añadió mayor complejidad para la extracción de datos, pero dado que la folksonomía trabaja con términos y no con idiomas, la creación de sinónimos de palabras entre ambos idiomas o de palabras y acrónimos permite la identificación del término de búsqueda independientemente del idioma.

Grado de enfermedad renal crónica

En el campo de la Nefrología es de gran importancia la clasificación del grado de enfermedad renal crónica (ERC) de los pacientes⁵, dado que tiene implicaciones pronósticas y terapéuticas. La revisión de la clasificación de la enfermedad o situación renal según la información recogida en el apartado «diagnósticos» de los informes de alta, permitió únicamente identificar el grado de ERC en unos 300 informes. Gracias a que la herramienta permite añadir sinónimos, se asignaron las palabras «grau», «estadió» y «estadi» a la palabra «grado», lo que permitió encontrar el grado de ERC en 768 informes. Para conseguir clasificar el grado de ERC del resto de los informes, así como para identificar los informes con diagnóstico de insuficiencia renal aguda, se generaron algoritmos con reglas heurísticas para la correcta identificación de la situación de enfermedad renal con base en: a) la presencia de las palabras «fracaso renal agudo» y sinónimos en el apartado diagnóstico, que implicaba la etiqueta de insuficiencia renal aguda, b) la presencia de las palabras «ingreso para receptor de trasplante renal» y sinónimos en motivo de

consulta implicaba la etiqueta ERC grado 5, c) la identificación de las palabras «enfermedad renal crónica grado X» y sinónimos entre los antecedentes personales permitía el etiquetado de los informes como ERC grados 1 al 5, d) la utilización de la creatinina en la analítica de ingreso con la edad y género (datos recogidos entre las variables antropométricas) permitió el cálculo del filtrado glomerular estimado introduciendo la fórmula de CKD-EPI⁶ en el software (*chronic kidney disease epidemiology collaboration*). A pesar de ello, los algoritmos no permitieron clasificar 79 documentos en cuanto a la situación renal, por lo que se procedió a la revisión y asignación manual de la situación renal en aquellos que quedaron sin clasificar. Así, la aplicación de las herramientas del software (creación de etiquetas, sinónimos y algoritmos con reglas heurísticas) permitieron la clasificación automática del grado de enfermedad renal a pesar de no estar directamente recogido en el apartado «diagnósticos» de las historias clínicas en el 95% del total de informes, requiriendo la revisión manual de únicamente un 5% de ellos. De esta manera, todos los informes quedaron clasificados como: fracaso renal agudo, ERC grados 1- 5 o sin enfermedad renal (fig. 2).

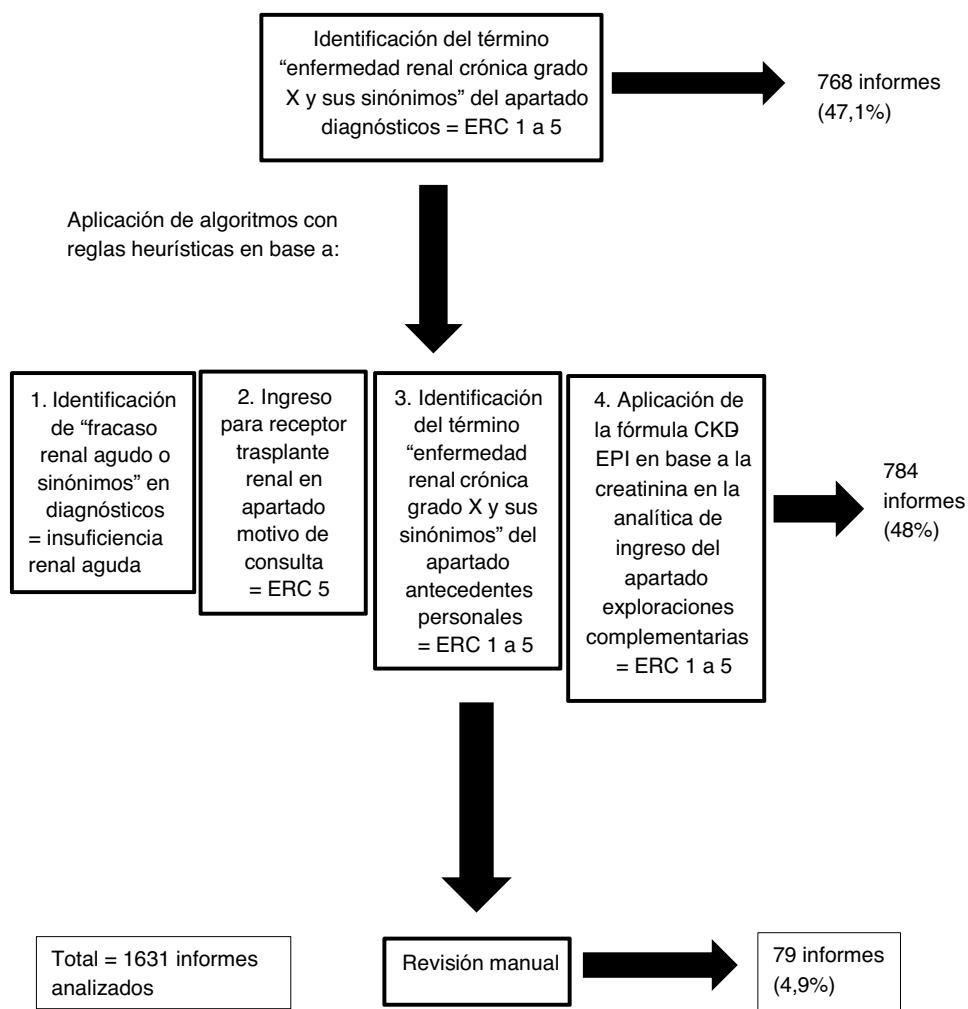


Figura 2 – Algoritmo aplicado para la clasificación de la situación de enfermedad renal de los informes incluidos en el análisis.

Preguntas planteadas como prueba piloto

Como prueba piloto se plantearon tres cuestiones:

1.^a ¿Cómo tratamos a los pacientes diabéticos con enfermedad renal? Teniendo en cuenta que la metformina es el hipoglucemiente oral principalmente usado en población diabética, ¿cuál es nuestra actitud en cuanto a la prescripción de este fármaco en situación de enfermedad renal? ¿Cuántos y que características tienen los pacientes con diagnóstico de acidosis láctica por metformina?

2.^a ¿Cómo tratamos la hipertensión arterial de los pacientes con enfermedad renal crónica? ¿Son los inhibidores del sistema renina angiotensina la familia de fármacos más utilizada teniendo en cuenta sus beneficios en cuanto a nefroprotección? ¿Cuál es la actitud de los nefrólogos del Servicio de Nefrología del Hospital del Mar en relación con la retirada o mantenimiento de los **inhibidores del sistema renina angiotensina** de los pacientes que ingresan en el Servicio de Nefrología?

3.^a ¿Cuál es el porcentaje de ingresos en nefrología que reciben algún **tratamiento hipnótico/sedante/antidepresivo** a pesar de no incluirse diagnósticos relacionados con esta patología entre los antecedentes del paciente?

Resultados

Tratamiento de la diabetes mellitus

A pesar de que la metformina sigue siendo el hipoglucemiente oral más usado y con mayor evidencia de eficacia en el tratamiento de la diabetes tipo 2 por sus beneficios en cuanto a morbi-mortalidad, su uso en el paciente con enfermedad renal está restringido⁷. En situación de ERC se debe ajustar la dosis y su utilización está contraindicada en situación de ERC avanzada, pudiéndose asociar a la presencia de acidosis láctica, especialmente cuando se administra a pacientes con filtrado glomerular inferior a 30 mL/min/1,73m²⁸. Así, en situación de ERC avanzada únicamente pueden utilizarse hipoglucemiantes orales de otras familias o se debe plantear la insulinización de los pacientes diabéticos. Se analizó el tratamiento que recibían los pacientes diabéticos que ingresaban en el servicio de Nefrología del Hospital del Mar y las características de los pacientes con diagnóstico de acidosis láctica por metformina.

La identificación de los pacientes diabéticos se realizó con base en la presencia de este diagnóstico en el apartado «diagnósticos» en los informes de alta. Así, se identificó un porcentaje de informes con el diagnóstico diabetes inferior al esperado, por lo que se amplió la búsqueda con nuevas reglas heurísticas, asignando el diagnóstico de diabetes a aquellos informes en los que se encontrara entre el tratamiento habitual algún fármaco hipoglucemiente. Dado el elevado número de hipoglucemiantes disponibles en el mercado, la inclusión de cada uno de ellos de forma individual en las búsquedas generaba una mayor complejidad al proyecto. Así, para etiquetar cada medicamento se decidió utilizar la clasificación ATC (*Anatomical Therapeutic Chemical classification system*) de la Agencia Española del Medicamento (AEMPS), que agrupa estos compuestos por principios activos. La clasificación ATC nos proporciona información tanto del nombre comercial como

del principio activo y en el proceso de análisis de datos se recorrió todo el texto buscando cualquiera de ambos. En el caso de identificar un nombre comercial, gracias a la clasificación ATC se puede inferir el principio activo y el grupo ATC al que corresponde mediante la aplicación de algoritmos de análisis de grafos. Hay que tener en cuenta que hay grupos ATC que pueden contener dos o más principios activos, por lo que se añadió esta lógica al algoritmo de detección.

En el caso de los fármacos para el tratamiento de la diabetes, estos corresponden al grupo A, subgrupo A10 de la clasificación ATC.

Se identificaron finalmente 637 de los 1.631 informes con el diagnóstico de diabetes (39,05% de los informes). En la **tabla 1** se muestran los tratamientos hipoglucemiantes recibidos por estos pacientes en función de su grado de enfermedad renal (subgrupos A10A para la insulina, A10BA para la metformina, A10BH para los inhibidores de la DPP4, A10BX para repaglinida e inhibidores del transportador SGLT2 –requiriendo búsqueda por nombre comercial y principio activo-, A10BB para los derivados de las sulfonilureas y A10BG para las tiazolidinonas). Así, el tratamiento más utilizado en estos pacientes renales con diabetes es la insulina (337 informes contenían insulina en el tratamiento habitual al ingreso), seguida de la metformina (85 informes contenían metformina en la medicación al ingreso). En cinco de estos 85 informes, se identificó en el apartado diagnóstico el término «láctica» y sinónimos, objetivándose así cinco casos de acidosis láctica por metformina (tres episodios en el contexto de un fracaso renal agudo, un paciente con ERC estadio 3 y un paciente con ERC estadio 4). En todos los casos, el motivo de consulta resultó ser una gastroenteritis aguda (cuatro con diarreas y uno con síndrome emético). Además, en todos los casos, excepto en el paciente con ERC estadio 3, se requirió la realización de hemodiálisis en el contexto del deterioro de función renal y acidosis láctica por metformina.

En 102 informes no se detectó la presencia de ningún tratamiento hipoglucemiente, por lo que se concluyó que un 16% de los pacientes, seguían únicamente tratamiento dietético de su diabetes.

Tratamiento de la hipertensión arterial

La prevalencia de hipertensión arterial en la enfermedad renal es elevada, coexistiendo ambas patologías en el 80-85% de los pacientes con enfermedad renal⁹.

Los beneficios renales y cardiovasculares de los inhibidores del sistema renina angiotensina (inhibidores SRA) en pacientes con ERC han sido ampliamente demostrados¹⁰ y deberían formar parte del tratamiento de la hipertensión arterial de los pacientes renales. Sin embargo, ante situaciones de descompensaciones agudas de la función renal, suelen retirarse. El retraso en la reintroducción de los mismos una vez resuelto el episodio descompensador, podría implicar un empeoramiento en el pronóstico de nuestros pacientes¹¹.

Se analizó el tratamiento antihipertensivo que reciben los pacientes con diagnóstico de hipertensión arterial y que ingresan en el Servicio de Nefrología del Hospital del Mar. El diagnóstico de hipertensión arterial se identificó en 1.520 de los 1.631 informes disponibles (93,19% de los informes). Para ello, bajo el paraguas de la etiqueta «hipertensión arterial»

Tabla 1 – Número de informes que contienen las distintas opciones terapéuticas para el tratamiento de la diabetes en los pacientes que ingresan en el Servicio de Nefrología

	INS	MTF	-DPP4	RGL	SLN	-SGLT2	GLP1	TZN
FRA	78	16	5	0	3	0	0	0
ERC 1	12	10	5	0	1	0	0	0
ERC 2	15	21	2	2	0	0	0	0
ERC 3	46	32	8	8	1	2	2	0
ERC 4	74	6	8	13	0	0	0	1
ERC 5	182	0	27	24	1	0	0	0
Total	337	85	55	47	6	2	2	1

FRA: fracaso renal agudo; ERC: enfermedad renal crónica; INS: insulina; MTF: metformina; -DPP4: Inhibidores de la dipeptil peptidasa 4; RGL: repaglinida; SLN: derivados sulfonilureas; -SGLT2: inhibidores del transportador sodio glucosa; GLP1: agonistas del péptido similar al glucagón tipo 1; TZN: tiazolidinadonas.

Tabla 2 – Prescripción de fármacos antihipertensivos por familias en la medicación al ingreso y medicación al alta

	Medicación al ingreso	Medicación al alta
Diuréticos de asa	549	585
Diuréticos tiacídicos y afines	126	112
Diuréticos ahorradores de potasio	79	94
Betabloqueantes	611	724
Calcioantagonistas dihidropiridínicos	532	639
Calcioantagonistas no dihidropiridínicos	18	10
Inhibidores del SRA	437	317
Alfabloqueantes	153	177
Hidralazina	78	97
Antiadrenérgicos de acción central	5	3
Total	2.588	2.758

SRA: sistema renina angiotensina.

Tabla 3 – Situación renal y número de informes de pacientes hipertensos que reciben al ingreso (columna 2) y al alta (columna 3) tratamiento con inhibidores del SRA y el porcentaje de reducción en la prescripción al alta

Total de informes de pacientes con inhibidores del SRA al ingreso (n = 437)	n (al ingreso)	n (al alta)	% de reducción
Fracaso renal agudo	24 (5,5%)	18	25%
Enfermedad renal crónica grado 1	30 (6,9%)	20	33,3%
Enfermedad renal crónica grado 2	22 (5,0%)	19	13,6%
Enfermedad renal crónica grado 3	93 (21,3%)	68	26,9%
Enfermedad renal crónica grado 4	70 (16%)	48	31,4%
Enfermedad renal crónica grado 5	193 (44,2%)	76	60,6%
Sin enfermedad renal crónica*	5 (1,1%)	5	0%

* ingreso para cateterismo suprarrenal como estudio de hiperaldosteronismo primario.

SRA: sistema renina angiotensina.

se incluyeron los sinónimos «hipertensión arterial» y «HTA». Se agruparon los fármacos antihipertensivos que recibían los pacientes por grupos (atc C02A para antihipertensivos de acción central, atc C02C para doxazosina, atc C02DB para hidralazina, atcC03A para diuréticos: 03AA hidroclorotiacida, 03BA clortalidona e indapamida, 03CA diuréticos de asa y 03DA para antialdosterónicos; atc C07 para betabloqueantes, atc C08CA calcioantagonistas dihidropiridínicos y atc C08D no dihidropiridínicos, y atc C09 para los inhibidores del sistema renina angiotensina). Al ingreso los diuréticos fueron los fármacos más utilizados (en 754 informes constaba por lo menos un diurético en la medicación habitual, mayoritariamente [549] diuréticos de asa). Cabe destacar que 34 de los 126 informes en los que constaba administración de diuréticos tiacídicos y afines, correspondían a pacientes con ERC estadio 5, un grado de enfermedad renal en la que se

pierde el efecto diurético de estos fármacos. En segundo lugar, los antihipertensivos más utilizados fueron la familia de los betabloqueantes (611 informes), seguidos de los calcioantagonistas dihidropiridínicos (532 informes) y posteriormente los inhibidores del sistema renina angiotensina (437 informes). El ingreso modificó la pauta de administración de antihipertensivos al alta, tanto por lo que respecta al número total de fármacos antihipertensivos prescritos (al ingreso 2.588 y al alta 2.758) como en el incremento o reducción de la prescripción de determinadas familias de antihipertensivos como se muestra en la [tabla 2](#). En 437 de los 1.520 informes con diagnóstico de hipertensión arterial (28,75%) se identificó entre el tratamiento habitual algún fármaco perteneciente al grupo C09 según la clasificación ATC (inhibidores SRA). La situación renal de los pacientes que recibían tratamiento con este grupo de fármacos al ingreso se muestra en la [tabla 3](#) (columna 2). En la

misma tabla (columna 3) se muestran el número de informes que seguían recibiendo inhibidores del SRA al alta, mostrando la última columna de la tabla, el porcentaje de reducción en la prescripción de inhibidores del sistema renina-angiotensina en el momento del alta de Nefrología. Destacaba una retirada de los inhibidores del SRA en un elevado porcentaje de los informes de alta, mostrando una tendencia creciente en función del grado de enfermedad renal crónica (estadios 3 a 5). En este proyecto no se analizaron las causas de retirada de estos fármacos, si bien conociendo la práctica clínica habitual, probablemente estuviera en relación al deterioro agudo de la función renal, más probable cuanto más avanzada esté la enfermedad renal. Teniendo en cuenta este razonamiento, no parecía plausible con la práctica clínica habitual que el porcentaje de informes que mantenían el tratamiento con inhibidores del SRA al alta entre los catalogados como «fracaso renal agudo» fuese tan elevado (reducción únicamente del 25% en la prescripción al alta). Por ello, se procedió a la revisión manual de estos informes. La revisión manual permitió detectar palabras como «suspender» o «modificar» delante de fármacos inhibidores del SRA, de manera que, en el grupo de fracaso renal agudo, realmente solo siete informes mantenían el tratamiento al alta, habiéndose por lo tanto detectado erróneamente y como falsos positivos 11 informes.

Salud emocional

Hay estudios que muestran que la prevalencia de síntomas depresivos en pacientes con ERC es elevada y que las variables psicosociales juegan un papel importante en la percepción de la calidad de vida de los pacientes renales^{12,13}. Sin embargo, en el trabajo asistencial en los servicios de nefrología, las curas en el ámbito psicológico de nuestros pacientes siguen estando relegadas a un segundo plano.

Se realizó una búsqueda de los informes que contenían en el tratamiento habitual al ingreso algún fármaco del grupo N05 o N06 según la clasificación (fármacos psicodélicos y psicoanalépticos). Se identificaron 402 informes que contenían alguno de estos fármacos (24,6% de los informes). En cambio, basando la búsqueda en la presencia de algún diagnóstico del ámbito ansioso-depresivo en el apartado «diagnósticos» o en el apartado «antecedentes personales», únicamente se identificaron 45 (2,75%) y 192 (11,77%) informes, respectivamente. Estos datos traducen que la conciencia del médico en cuanto a la prevalencia de trastornos ansioso-depresivos de los pacientes renales es baja a pesar de una elevada prescripción de fármacos para tratar sus síntomas.

Discusión

En este trabajo piloto hemos evaluado la utilidad de la aplicación de folksonomía y técnicas de inteligencia artificial, como NLP, para el análisis de datos de los informes de alta del Servicio de Nefrología a partir de preguntas meramente clínicas. La aplicación de esta tecnología nos ha permitido reducir notoriamente el tiempo empleado para la extracción de información. Únicamente con base en la estructura habitual de los informes de alta y a su redacción en lenguaje natural, se ha podido extraer información relevante que, de no disponer de la

herramienta, hubiera requerido la revisión manual de dichos informes y la generación de bases de datos.

Uno de los aprendizajes tras haber realizado este proyecto piloto es que la redacción clara de la información médica relevante en el ámbito de la nefrología (como es la clasificación de la enfermedad renal) hubiera facilitado y acelerado la obtención de datos. A pesar de una redacción no completamente uniforme y estructurada de las altas hospitalarias, a menudo con falta de información relevante (como la adecuada clasificación de la situación renal de los pacientes), la herramienta ha permitido incluir algoritmos y reglas heurísticas para solventar estas dificultades iniciales.

La instalación del portal Bismart Folksonomy es un proceso relativamente rápido; sin embargo, la aplicación de modificaciones y algoritmos necesarios para un proyecto determinado, requieren una mayor inversión de tiempo inicial. La búsqueda de etiquetas en los documentos es automática, y una vez lanzada, arroja resultados en menos de un minuto. Posteriormente, el ejercicio más laborioso de creación de reglas y sinónimos puede requerir para proporcionar análisis de datos de 2-3 horas de trabajo. Sin embargo, este también será un proceso automático para todos aquellos documentos que se incorporen posteriormente, permitiendo un análisis en tiempo real de todas las altas hospitalarias que se fueran incorporando al sistema y permitiendo, por lo tanto, un análisis en tiempo real de cualquier cuestión a querer explorar, así como crear alarmas que nos permitieran detectar y seleccionar pacientes con determinadas características de interés.

Esta herramienta podría utilizarse también en otros ámbitos asistenciales, como podrían ser las consultas ambulatorias o la actividad en el hospital del día de nefrología, donde se genera un importante volumen de información en lenguaje natural. Además, la obtención y la posibilidad de cruzar datos de informes con resultados de laboratorio u otras exploraciones complementarias no incluidas en informes médicos directamente, incrementaría exponencialmente la información extraída con la aplicación de folksonomía. Este es un aspecto relevante en el análisis que hemos realizado de los tratamientos contenidos en el informe en el momento del alta. Dado que el documento de tratamiento oficial de un paciente es la receta electrónica, sería de gran interés poder contrastar el tratamiento farmacológico contenido en un informe de alta con los datos de la receta electrónica de un mismo paciente. Si bien técnicamente es factible la aplicación de folksonomía a la receta electrónica igual que se ha hecho con los informes de alta, en este proyecto con anonimización completa de la filiación del paciente, este proceso no se podía llevar a cabo.

Hasta donde conocemos, no existen experiencias previas publicadas que hayan trabajado con folksonomía en el ámbito médico. Esta tecnología permite identificar cualquier término de búsqueda sin necesidad de haber definido una ontología o entidad maestra previamente y no existe posibilidad de error en la búsqueda de los términos de interés. Sin embargo, existe la posibilidad de clasificar erróneamente términos de búsqueda por tratarse de falsos positivos o falsos negativos (por ejemplo, términos como «no» o «suspender» delante de nuestros términos de búsqueda representarían falsos positivos). En esta experiencia piloto, ante resultados no congruentes se ha procedido a la revisión y reclasificación manual de dichos informes, pero la herramienta permite la inclusión de reglas

que detectan las negaciones, pudiendo evitar así las tareas manuales de revisión. Finalmente, un aspecto a mejorar es el hecho de que la herramienta de búsqueda del portal Bismart Folksonomy (sección *easy query*) permite la adición de búsqueda de palabras (utilización del término «and») pero en la actualidad no permite la búsqueda de un término u otro (utilización del término «or»), lo que representa cierta limitación en la obtención de información. En esta ocasión, esta limitación en cuanto al término «or» se solventó con la creación de «categorías» (termino que agrupa un conjunto). Un ejemplo de ello sería la clasificación ATC (al término ATC09 se le asociaron todos los fármacos inhibidores del sistema renina angiotensina).

En conclusión, la utilización de *big data* en el ámbito médico, en este caso concreto de folksonomía y NLP, puede permitir un importante ahorro de tiempo sin detrimento de la calidad y veracidad de la información obtenida con finalidades de investigación y gestión de calidad de la actividad asistencial realizada.

Conceptos clave

- Cada día se genera un gran volumen de datos clínicos, quedando gran parte de ellos recogidos en forma de lenguaje natural.
- Clásicamente, la extracción y análisis de datos de las historias clínicas se realiza mediante un proceso manual que requiere de una importante inversión de tiempo.
- La utilización de herramientas de *big data*, en concreto de procesamiento de lenguaje natural (*natural language processing* o NLP), permite acelerar de forma evidente este proceso.
- La aplicación de folksonomía como herramienta de NLP no requiere de la creación previa de una entidad maestra que recoja los términos de búsqueda de interés, y este hecho aporta una clara ventaja sobre otras herramientas de NLP.
- Con base en determinadas preguntas clínicas del ámbito de la nefrología y mediante la utilización del software *Bismart Folksonomy*, se ha aplicado la folksonomía para la extracción y análisis de datos de informes de alta de un servicio de nefrología de forma automática.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Este proyecto piloto con la herramienta *Bismart Folksonomy* de la empresa Bismart, se ha llevado a cabo mediante la

financiación de Laboratorios Ferrer, sin que Laboratorios Ferrer haya obtenido ningún dato clínico ni haya guiado el proyecto de investigación. Laia Sans, Jordi Martínez-Roldan y Julio Pascual no tienen ninguna relación profesional con Bismart. Ismael Vallvé y Joan Teixidó trabajan para Bismart y Josep Manel Picas actúa como consultor de dicha empresa.

Los autores agradecen a Laboratorios Ferrer el apoyo para la realización de este estudio.

BIBLIOGRAFÍA

1. Palanisamy V, Thirunavukarasu R. Implications of big data analytics in developing healthcare frameworks – A review. *Journal of King Saud University - Computer and Information Sciences*. 2019;31:415–25.
2. Vigilante K, Escaravage S, McConnell M. Big Data and the Intelligence Community — Lessons for Health Care. *N Engl J Med*. 2019;380:1888–90.
3. Wen A, Fu S, Moon S, El Wazir M, Rosenbaum A, Kaggal VC, et al. Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. *npj Digit Med*. 2019;2:130, 17.
4. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–43.
5. Levey A, de Jong P, Coresh J, El Nahas M, Astor B, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney Int*. 2011;80:17–28.
6. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al. A New Equation to Estimate Glomerular Filtration Rate. *Ann Intern Med*. 2009;150:604–12.
7. Crowley M, Diamantidis C, McDuffie J, Cameron B, Stanifer J, Mock C, et al. Clinical Outcomes of Metformin Use in Populations with Chronic Kidney Disease Congestive Heart Failure, or Chronic Liver Disease: A Systematic Review. *Ann Intern Med*. 2017;166:191–200.
8. Lazarus B, Wu A, Shin JJ, Sang Y, Alexander GC, Secora A, et al. Association of metformin use with Risk of Lactic Acidosis Across the Range of Kidney Function: A community-based Cohort Study. *JAMA Intern Med*. 2018;178:903–10.
9. Kalaitzidis RG, Elisaf MS. Treatment of Hypertension in Chronic Kidney Disease. *Curr Hypertens Rep*. 2018;20:64.
10. Xie X, Liu Y, Perkovic V, Li X, Ninomiya T, Hou W, et al. Renin-Angiotensin System Inhibitors and Kidney and Cardiovascular Outcomes in Patients With CKD: A Bayesian Network Meta-analysis of Randomized Clinical Trials. *Am J Kidney Dis*. 2016;67:728–41.
11. Bhandari S, Ives N, Brettell EA, Valente M, Cockwell P, Topham PS, et al. Multicentre randomized controlled trial of angiotensin-converting enzyme inhibitor/angiotensin receptor blocker withdrawal in advanced renal disease: the STOP-ACEi trial. *Nephrol Dial Transplant*. 2016;31:255–61.
12. Cangini G, Rusolo D, Cappuccilli M, Donati G, La Manna G. Evolution of the concept of quality of life in the population in end stage renal disease. A systematic review of the literature. *Clin Ter*. 2019;170:e301–20.
13. Wang WL, Liang S, Zhu FL, Liu JQ, Wang SY, Chen XM, et al. The prevalence of depression and the association between depression and kidney function and health-related quality of life in elderly patients with chronic kidney disease: a multicenter cross-sectional study. *Clin Interv Aging*. 2019;14:905–13.