مجله مدیریت اطلاعات سلامت
سال هجدهم/ شماره چهارم/ مهر و آبان ۱۴۰۰
Published by *Vesnu Publications*
DOI: 10.22122/him.v18i4.4384

تحلیل وضعیت پژوهش‌های رشته غدد درون‌ریز و متابولیسم در ایران با استفاده از روش‌های متن‌کاوی

ام‌البنین اسدی قادیکلایی[۱]، نجلا حریری[۲]، مریم خادمی[۳]، فهیمه باب‌الحوائجی[۴]

مقاله پژوهشی

چکیده

مقدمه: با توجه به اهمیت و جایگاهی که حوزه غدد درون‌ریز در بخش سلامت دارد، تسهیل بازیابی اطلاعات در این حوزه می‌تواند بسیار مهم باشد. مطالعه حاضر با هدف مدل‌سازی موضوعی مقالات منتشر شده پژوهشگران ایرانی در حوزه غدد درون‌ریز و متابولیسم در پایگاه استنادی علوم انجام گرفت.

روش بررسی: این تحقیق از نوع توصیفی بود و با روش متن‌کاوی انجام شد. چکیده مقالات با استفاده از کلید واژه‌های منتخب سرعنوان موضوعی پزشکی MeSH (Medical Subject Headings) از پایگاه استنادی علوم استخراج گردید. ۵۵۵۲ مقاله از سال ۱۹۷۷ تا سال ۲۰۱۹ بازیابی شد. سپس متن چکیده‌ها در نرم‌افزار MATLAB مورد تحلیل و بررسی قرار گرفت و دسته‌بندی شد.

یافته‌ها: دسته‌های موضوعی متشکل از ۲۰ واژه و در ۴۸ دسته استخراج گردید. بیماری دیابت با ۷۱۴۵ بار تکرار، بیشتر از سایر موضوعات مورد توجه پژوهشگران ایرانی قرار گرفته است. دسته موضوعی مربوط به بیماری‌های سندرم متابولیک با بیشترین تعداد مقالات (۳۰۴ مقاله) و دسته موضوعی شماره ۴۷ که مربوط به بیماری‌های نرمی استخوان بود، کمترین تعداد مقالات (۵۱ مقاله) را به خود اختصاص داد.

نتیجه‌گیری: پژوهشگران ایرانی به تحقیقاتی با موضوع سندرم متابولیک بیشتر و به موضوعاتی مانند نرمی استخوان کمتر پرداخته بودند. مباحثی شامل Dwarfism، Parathyroid Diseases، Pituitary Diseases، Gonadal Disorders، Polyendocrinopathies و Autoimmune که در موضوعات حاصل از مدل‌سازی موضوعی وجود نداشت، بیانگر خلأ موجود در پژوهش‌های محققان ایرانی می‌باشد که بر لزوم توجه بیشتر بر این حوزه‌ها تأکید می‌شود.

واژه‌های کلیدی: مدل‌سازی موضوعی؛ بیماری‌های غدد درون‌ریز؛ متابولیسم؛ متن‌کاوی؛ تخصیص پنهان دریکله

پیام کلیدی: نتایج پژوهش حاضر نشان داد که حوزه‌های موضوعی تخصصی شامل Dwarfism، Parathyroid Diseases، Pituitary Diseases، Gonadal Disorders، Polyendocrinopathies و Autoimmune با وجود اهمیت، به جهت شیوع در جامعه مورد غفلت قرار گرفته‌اند.

دریافت مقاله: ۱۴۰۰/۶/۸ پذیرش مقاله: ۱۴۰۰/۷/۱۴ تاریخ انتشار: ۱۴۰۰/۷/۱۵

ارجاع: اسدی قادیکلایی ام‌البنین، حریری نجلا، خادمی مریم، باب‌الحوائجی فهیمه. **تحلیل وضعیت پژوهش‌های رشته غدد درون‌ریز و متابولیسم در ایران با استفاده از روش‌های متن‌کاوی.** مدیریت اطلاعات سلامت ۱۴۰۰؛ ۱۸ (۴): ۱۶۵-۱۶۰

مقدمه

افزایش تعداد اسناد و حجم انبوه داده‌های متنی پزشکی، جستجو در این حجم از اطلاعات که اغلب ساختار نیافته هستند را با چالش‌های بسیاری مواجه کرده است. تحلیل مجموعه اسناد متنی و پردازش زبان طبیعی و همچنین، شناسایی الگو و ساختار در نمونه‌های داده در حجم انبوه، نیازمند توسعه ابزارهای قدرتمند و روش‌های آماری می‌باشد. تکنیک‌های بسیاری جهت سهولت این روند و ارایه سریع‌تر این امر توسعه یافته‌اند (۱).

مدل‌های موضوعی، از مهم‌ترین و عمده‌ترین تکنیک‌های پیشرفته یادگیری ماشینی (بدون نظارت) هستند که به صورت گسترده در متن‌کاوی مورد استفاده قرار می‌گیرند و برای پیدا کردن الگوهای پنهان معنایی در حجم انبوه داده به کار می‌روند (۳، ۲). از سوی دیگر، یکی از بهترین راه‌های بررسی تولیدات علمی، بررسی مقالات منتشر شده در هر حوزه علمی است که مسیر علمی آینده یک رشته را مشخص می‌کند و می‌توان سیاست‌گذاری‌ها و سرمایه‌گذاری‌ها را در رشته‌های علمی مشخص نمود (۴).

در این بین، رشته غدد درون‌ریز و متابولیسم به جهت شیوع زیاد بیماری‌های مرتبط، هزینه‌های مالی ناشی از درمان و میزان مرگ و میر، دارای اهمیت فراوانی است و لزوم انجام پژوهش‌های علمی در این زمینه را نشان می‌دهد. مطالعاتی که

مقاله حاصل پایان‌نامه مقطع دکتری تخصصی به شماره ۹۷۲/۲۰۹۶۱/۱۰۰۰۹۰/۱۲۳ می‌باشد که با حمایت دانشگاه آزاد اسلامی واحد علوم و تحقیقات انجام شده است.

۱- دانشجوی دکتری تخصصی، علم اطلاعات و دانش‌شناسی، گروه علوم ارتباطات و دانش‌شناسی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۲- استاد، علم اطلاعات و دانش‌شناسی، گروه علوم ارتباطات و دانش‌شناسی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۳- دانشیار، ریاضی کاربردی، گروه ریاضی کاربردی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران
۴- دانشیار، علم اطلاعات و دانش‌شناسی، گروه علوم ارتباطات ودانش شناسی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
نویسنده طرف مکاتبه: مریم خادمی؛ دانشیار، ریاضی کاربردی، گروه ریاضی کاربردی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران
Email: khademi@azad.ac.ir

سبب کاهش بار مالی و میزان مرگ و میر این بیماران خواهد شد (۵). از جمله شایع‌ترین این بیماری‌ها می‌توان به دیابت، چاقی، افزایش چربی خون، پوکی استخوان و اختلالات ناشی از کمبود ید اشاره کرد. سالانه افراد زیادی به بیماری دیابت مبتلا می‌شوند و تا سال ۲۰۱۰ شیوع آن به بیش از ۲۰۰ میلیون نفر در سراسر جهان رسیده است که این تعداد تا سال ۲۰۲۵ به ۳۰۰ میلیون نفر خواهد رسید (۶). انتظار می‌رود با انجام مدل‌سازی موضوعی، بتوان وضعیت تحقیقات و پژوهش‌ها را در این رشته منجسم‌تر نمود و مسیر آینده برای پژوهشگران روشن‌تر گردد.

نتایج مطالعه محمدی و همکاران که با هدف بررسی برون‌دادهای علمی رشته غدد درون‌ریز و متابولیسم به انجام رسید، نشان داد که اگرچه رشد تولیدات علمی ایران در حوزه غدد درون‌ریز و متابولیسم قابل توجه است، اما ارتقای رتبه علمی آن از نظر کمیت و کیفیت، نیاز به برنامه‌ریزی دقیق‌تری دارد (۷). عزیزی در تحقیق خود به بررسی تولیدات علمی ایران در جهان و آسیای جنوب غربی پرداخت و به این نتیجه رسید که تولیدات علمی ایران در دو دهه اخیر افزایش فوق‌العاده‌ای داشته و این افزایش از سال ۱۳۸۵ به بعد شدت بیشتری یافته و ایران در محدوده این سال حایز رتبه ۱۹ و ۲۰ جهانی شده است. این روند به همین صورت ادامه یافته و رتبه ایران در سال ۱۳۹۱، ۱۶ جهانی و اول منطقه بوده است (۸). امامی و همکاران در پژوهشی، برون‌دادهای علمی رشته غدد درون‌ریز و متابولیسم را در ایران و خاورمیانه مورد بررسی قرار دادند و دریافتند که کشور ترکیه از لحاظ تعداد انتشارات و ارجاعات متنی در خاورمیانه پیشرو می‌باشد و ایران رتبه دوم از نظر تعداد انتشارات و استنادات محلی و رتبه سوم از لحاظ تعداد استناد جهانی را به خود اختصاص داده‌است (۹). شاهمرادی و همکاران مطالعه‌ای را با هدف بررسی تولیدات علمی کشور ایران در حوزه موضوعی غدد درون‌ریز ومتابولیسم از لحاظ همکاری‌های علمی انجام دادند و نتیجه‌گیری کردند که همکاری‌های بین‌المللی، تعامل با کشورهای پیشرو و همکاری‌های بین رشته‌ای، روند رو به افزایشی دارد (۱۰).

مدل‌سازی موضوعی، یکی از تکنیک‌های خوشه‌بندی در داده‌کاوی و متن‌کاوی است. این روش، ساختارهای موضوعی پنهان را در مجموعه‌ای از اسناد متنی نشان می‌دهد و موجب تولید روش‌هایی برای دستیابی به اسناد به وسیله مرور، جستجو و ایجاد خلاصه آرشیوهای بزرگ متنی می‌شود. موضوعات در این روش، دسته‌ای از واژگان هستند که با هم رخ می‌دهند (۱۱). مدل تخصیص پنهان دیریکله LDA (Latent Dirichlet Allocation)، یکی از پرکاربردترین و اساسی‌ترین الگوریتم‌های مدل‌سازی موضوعی احتمالاتی است (۱۲) و در مطالعه حاضر مورد استفاده قرار گرفت.

تحقیق حاضر با توجه به ضرورت انجام پژوهش در حوزه غدد، با هدف تعیین این رشته از لحاظ موضوعی، شناسایی نقاط ضعف و نقاط قوت آن صورت گرفت و با توجه به دقت و کارایی مدل‌سازی موضوعی و الگوریتم LDA، این روش‌ها در اجرای هدف مورد نظر استفاده گردید.

روش بررسی

این مطالعه از نوع توصیفی بود و با روش متن‌کاوی انجام شد. برون‌دادهای علمی با استفاده از واژگان مستند شده سرعنوان موضوعی پزشکی MeSH (Medical Subject Headings) در پایگاه استنادی علوم که شامل ۱۱ توصیف‌گر اصلی «بیماری‌های غدد درون‌ریز، بیماری‌های غده آدرنال، بیماری‌های استخوان- غدد، دیابت شیرین، کوتاهی قد، سرطان‌های غدد درون‌ریز، اختلالات غدد جنسی، بیماری‌های پاراتیروئید، بیماری‌های هیپوفیز، پلی‌اندوکرینوپاتی‌های خودایمن، بیماری‌های تیروئید، سل- غدد» و توصیف‌گرها و واژگان مدخل بیماری‌های متابولیک و بیماری‌های استخوان بود، بر اساس عنوان مقالات و محدود کردن نویسندگان به کشور ایران و در گروه موضوعی غدد درون‌ریز و متابولیسم، مورد جستجو قرار گرفت.

در مجموع، ۷۸۹۰ مقاله طی سال‌های ۱۹۷۷ تا ۲۰۱۹ بازیابی شد که از این تعداد، ۲۳۳۸ مقاله تکراری یا فاقد چکیده بود و حذف گردید. ۵۵۵۲ مقاله باقی ماند. مقالات به نرم‌افزار مدیریت منابع EndNote نسخه ۹ و چکیده آن‌ها به صورت متن ساده به نرم‌افزار MATLAB وارد شد و پیش‌پردازش داده‌ها صورت گرفت.

پیش‌پردازش داده‌ها از مرحله Tokenize (تبدیل متن به توکن) شروع و با ریشه‌یابی کلمات با استفاده از نرمالیزه کردن ادامه یافت. سپس نشانه‌های ویراستاری از متن حذف شد وکلمات ایست (شامل واژه‌ها و لغاتی که با وجود تکرار مکرر در متن مقالات، از نظر معنایی دارای اهمیت کمی هستند؛ مانند «اما»، «ولی»، «که»، «یا» و...، بسیاری از افعال، اسامی، قیود، صفات و کلمات ربط و تعریف نیز ایست واژه شناخته شده‌اند) حذف گردید. حذف این کلمات، موجب بهبود نتایج و همچنین، کاهش بار محاسبات و افزایش سرعت پردازش خواهد گردید. بر این اساس، کلمات ایست در مرحله پیش‌پردازش حذف می‌شوند (۱۳). سپس از کلیه واژه‌ها خروجی تهیه گردید و جهت حذف واژه‌های نامرتبط و عام، در اختیار ۴ فوق تخصص غدد درون‌ریز و متابولیسم قرار گرفت. حذف واژه‌های عام و نامرتبط در سه مرحله صورت گرفت و در هر مرحله واژه‌های نامرتبط حذف گردید و تا حذف کلیه واژه‌های عام و نامرتبط ادامه یافت.

در آخرین مرحله، دسته‌های موضوعی و ابر واژه‌ها با استفاده از الگوریتم LDA استخراج شد و مدل‌سازی صورت گرفت. دسته‌های موضوعی شامل گروه‌های ۴۸ تایی و در هر گروه ۱۰۰ واژه بود و توسط ۴ فوق تخصص غدد درون‌ریز و متابولیسم با توجه به ارتباط آن‌ها به حوزه‌های موضوعی خاص نام‌گذاری گردید. در ابرهای واژه‌ای، کلمه‌ای که کلمات هر دسته بیشترین ارتباط را به آن واژه داشتند، با استفاده از نرم‌افزار مشخص شد.

بیماری‌های با بیشترین میزان تکرار و مقالات با احتمال بیش از ۲۰ درصد ارتباط استخراج شدند. روش مورد استفاده در پژوهش حاضر، میزان ارتباط مقاله- موضوع را به صورت عددی بین صفر تا ۱ به عنوان احتمال وابستگی گزارش می‌کند. وقتی احتمال وابستگی مقاله‌ای با موضوعی صفر باشد، یعنی این مقاله راجع به آن موضوع نیست. همچنین، وقتی احتمال وابستگی مقاله‌ای با موضوعی ۱ باشد، بدین معنی است که مقاله راجع به موضوع بحث می‌کند.

در نرم‌افزار MATLAB از چهار روش avb(approximate variational Bayes)، cgs(collapsed Gibbs sampling)، savb(stochastic approximate variational Bayes) و cvb0(variational Bayes, zeroth) به منظور پیاده‌سازی الگوریتم LDA استفاده می‌شود (۱۴). رایج‌ترین راه جهت ارزیابی مدل‌های احتمالی محاسبه لگاریتم، احتمال وقوع است. لگاریتم احتمال وقوع یا معیار سرگشتگی (Perplexity) در مدل‌های موضوعی با استفاده از رابطه ۱ محاسبه می‌شود (۱۵).

$$Perplexity\ (D_{test}) = exp\left\{-\frac{\sum_{d=1}^{M} Log\ p\ (W_d)}{\sum_{d=1}^{M} N_d}\right\}$$ رابطه ۱

با توجه به داده‌های جدول ۱، روش cgs بهترین عملکرد را نسبت به بقیه روش‌ها داشت. بنابراین، از این روش جهت محاسبه سرگشتگی استفاده شد.

جدول ۱: میزان معیار سرگشتگی بر اساس مدت زمان محاسبه در مرحله آموزش و ارزیابی (تست)

روش	معیار سرگشتگی		زمان
	ارزیابی	آموزش	
cgs	۳۳۳/۳	۶۵۷/۳	۷۱
avb	۴۵۴/۹	۱۰۰۲/۸	۵۳
cvb0	۳۳۱/۹	۵۳۹/۹	۶۲۹
savb	۳۵۱	۷۹۰/۷	۲۱۵

جهت ارزیابی تناسب میزان سرگشتگی مطالعه حاضر، نتایج با مقالات مشابه مورد مقایسه قرار گرفت. Ray و همکاران در مطالعه خود بر روی متون هندی، به میزان سرگشتگی بین ۴۲۰ تا ۱۲۶۰ رسیدند (۱۶). میزان سرگشتگی در تحقیق حاضر، عددی بین ۳۳۳/۳ تا ۶۵۷/۳ محاسبه گردید که نسبت به پژوهش‌های قبل، تناسب بسیار قابل قبولی دارد و نتایج از لحاظ سرگشتگی مورد تأیید قرار گرفت.

یافته‌ها

نتایج حاصل از مدل‌سازی موضوعی نشان داد که ۶ بیماری دارای بیشترین میزان تکرار در مقالات پژوهشگران ایرانی بوده‌اند. بر این اساس، بیماری دیابت بیشترین میزان تکرار در میان سایر بیماری‌ها را به خود اختصاص داد (جدول ۲).

جدول ۲: بیماری‌های با بیشترین میزان تکرار در کل بیماری‌ها

بیماری	تعداد تکرار	
Diabetes	Diabetes	۷۱۴۵
Diabetic		۴۳۵۹
Mellitus		۲۲۷۴
t2dm		۹۰۴
Mets	۹۲۲	
polymorphism	۸۹۹	
Cardiovascular	۸۳۸	
Cancer	۸۳۷	
Obesity	۸۳۱	

سپس دسته‌های موضوعی دارای بیشترین تعداد مقالات با احتمال وابستگی بیش از ۲۰ درصد استخراج گردید. ۱۰ حوزه موضوعی دارای بیشترین تعداد مقاله با احتمال ارتباط بالای ۲۰ درصد در جدول ۳ ارایه شده است.

بر اساس داده‌های جدول ۳، بیشترین تعداد مقالات با احتمال وابستگی بیشتر از ۲۰ درصد، در دسته موضوعی سندرم متابولیک وجود داشت که بیان‌کننده توجه بیشتر پژوهشگران ایرانی به این موضوع است. در دسته موضوعی Rickets (نرمی استخوان)، کمترین تعداد مقالات قرار گرفت که نشان دهنده پرداختن کمتر محققان ایرانی به این حوزه موضوعی می‌باشد.

بررسی و مقایسه مدل موضوعی حاضر با MeSH نشان داد که موضوعاتی مانند Prediabetes، DM and life style management، Vitamin D deficiency، DM and infertility، Diabetic foot ulcer، DM and infection، Stem cell transplant، Pancreas transplantation in DM، Statin therapy، Healthy diet و Genetic in T2DM در MeSH جایگاه مستقلی ندارند؛ در صورتی که در نتایج حاصل از تحلیل‌های پژوهش حاضر دارای جایگاه مستقلی می‌باشند که نشان از فعالیت محققان ایرانی در این حوزه‌ها است.

جدول ۳: حوزه‌های موضوعی با بیشترین تعداد مقاله

حوزه موضوعی	تعداد مقاله
Diabetes	۳۶۶۱
Metabolic syndrome	۳۰۴
genetic in T2DM	۲۴۵
Stem cell transplant	۲۴۳
DM and CVD	۲۳۶
Treatment of T1DM	۲۱۹
Osteoporosis	۲۰۵
Osteoarthritis	۲۰۲
Vitamin D deficiency	۱۹۷
Dyslipidemia	۱۹۱

بحث

از لحاظ میزان توزیع بیماری‌ها در تحقیقات پژوهشگران ایرانی، بیماری دیابت بیشترین میزان تکرار را به خود اختصاص داد. همان‌طور که در چارچوب ملی ارایه خدمات در بیماری دیابت اشاره شده است، این بیماری شایع‌ترین بیماری متابولیک در جهان می‌باشد و سازمان جهانی بهداشت از آن به عنوان همه‌گیری نهفته یاد کرده است به‌طور قطع، باید این حوزه مورد توجه بیشتری قرار گیرد. شیوع دیابت در جهان طی سال‌های ۱۹۸۰ تا ۲۰۱۴ حدود دو برابر شده است. میزان شیوع این بیماری در ایران نیز طی ۳ دهه گذشته دو برابر شده است (۱۷) و از این لحاظ، نتایج مطالعه حاضر مورد تأیید قرار می‌گیرد.

پس از دیابت، بیماری‌های سندرم متابولیک قرار داشت. در پژوهش عظیمی‌نژاد و همکاران، میزان شیوع سندرم متابولیک در ۵۵ درصد زنان و ۳۰ درصد مردان گزارش گردید (۱۸). ضابطیان و همکاران در پژوهش خود، میزان شیوع سندرم متابولیک در ایران را بر اساس تعریف سازمان جهانی بهداشت، ۱۸/۴ درصد اعلام کردند (۱۹).

بیماری Polymorphism با ۸۹۹ بار تکرار، پس از بیماری‌های سندرم متابولیک در رده بعدی بیماری‌های پرتکرار قرار گرفت. این بیماری که چندشکلی ژنتیکی نامیده می‌شود، تحت عنوان وقوع هم‌زمان دو یا چند ژنوتیپ یا آلل ناپیوسته در یک جمعیت تعریف شده است (۲۰). با توجه به نقش ژنتیک در بسیاری از بیماری‌های حوزه غدد مانند چاقی، دیابت، سندرم متابولیک و...، پرتکرار بودن این بیماری بسیار طبیعی است.

بیماری‌های قلبی- عروقی CVD (Cardiovascular disease) با ۸۳۸ بار تکرار پس از آن، بیشترین میزان توجه پژوهشگران ایرانی را به خود جلب کرده بود. CVD بیشترین دلیل مرگ و میر را در ایران به خود اختصاص داده است (۲۱). خسروی بروجنی و همکاران در مطالعه خود، ۳۲ درصد میزان مرگ و میر

ایرانیان را در اثر CVD و سکته‌های مغزی گزارش نمودند (۲۲).

سرطان‌های حوزه غدد در جایگاه بعدی قرار گرفت. سرطان‌ها سومین عامل مرگ و میر پس از CVD در ایران هستند و شایع‌ترین سرطان در حوزه غدد، سرطان تیروئید می‌باشد که ۲/۳ درصد کل سرطان‌ها را شامل می‌شود. گزارش ارایه شده در تحقیق حق‌پناه و همکاران، اهمیت سرطان‌های حوزه غدد را مورد تأیید و تأکید قرار می‌دهد (۲۳).

دیابت حوزه موضوعی بود که بیشترین تعداد مقالات را با احتمال بالای ۲۰ درصد شامل می‌شود و نتایج پژوهش‌های متعددی که پیش‌تر ذکر شد، دلیل قرارگیری این حوزه موضوعی در این جایگاه است. پس از آن، بیماری سندرم متابولیک قرار داشت. این بیماری خطر ابتلا به بیماری‌های قلبی، سکته و دیابت را افزایش می‌دهد و می‌تواند دلیل توجه محققان غدد به این حوزه باشد. همچنین، دسته موضوعی دیابت و پیوند سلول‌های بنیادین، دیابت و CVD، ودرمان دیابت نوع ۱ همگی به ترتیب در مطالعات خدائیان و همکاران (۲۴)، کوهی و همکاران (۲۵)، پیردهقان و همکاران (۲۶)، لاریجانی و همکاران (۲۷)، ابوالحسنی و همکاران (۲۸)، حشمت و همکاران (۲۹)، استقامتی و همکاران (۳۰) و ابراهیمی و همکاران (۳۱) از لحاظ اهمیت و میزان شیوع مورد تأیید قرار گرفت که با یافته‌های تحقیق حاضر همسو بود. از محدودیت‌های پژوهش حاضر این بود که ۱۵ مقاله به دلیل عدم دسترسی به چکیده آن‌ها، از روند مطالعه کنار گذاشته شدند.

نتیجه‌گیری

نتایج تحقیق حاضر نشان داد که الگوریتم LDA و روش cgs با توجه به اراه کمترین میزان سرگشتگی، کارایی قابل قبولی در انجام مدل‌سازی موضوعی دارند. بیماری‌هایی که با بیشترین میزان تکرار گزارش شدند، بیماری‌هایی هستند که بیشترین میزان شیوع را به خود اختصاص داده‌اند. همچنین، بررسی که در بخش مقالات با احتمال بیشترین میزان ارتباط به دسته‌های موضوعی صورت گرفت، از نظر تعداد مقالات مرتبط بیان‌کننده میزان شیوع و پرداختن سایر پژوهشگران به این حوزه‌های موضوعی است. دسته‌بندی مقالاتی که دارای احتمال وابستگی بیشتر از ۲۰ درصد هستند، منجر به ایجاد پایگاه داده موضوعی شده است که می‌توان با کد اختصاص یافته به هر مقاله، به آن دسترسی پیدا کرد و برای مطالعه سایر پژوهشگران در این حوزه موضوعی بسیار مفید واقع خواهد شد. نتایج به دست آمده برای سیاست‌گذاران علمی در حوزه غدد درون‌ریز و متابولیسم و پژوهشگران این حوزه در زمینه این که چه حوزه‌هایی مورد غفلت قرار گرفته است و باید به آن‌ها توجه بیشتری شود و در چه مواردی به لحاظ اهمیت شیوع و پیشگیری باید سرمایه‌گذاری شود، راهگشا خواهد بود.

پیشنهادها

پیشنهاد می‌شود پژوهشگران ایرانی توجه بیشتری به حوزه‌های موضوعی Dwarfism، Parathyroid Diseases، Gonadal Disorders، Autoimmune و Polyendocrinopathies که کمتر به آن‌ها پرداخته شده است، نشان دهند.

سیاست‌های تشویقی و حمایتی مؤثر در زمینه تولید اطلاعات علمی توسط سیاست‌گذاران اتخاذ شود. خوشه‌های موضوعی مشخص شده در مطالعه حاضر جهت تهیه منابع علمی پژوهشگران و محققان حوزه غدد مورد توجه قرار گیرد.

تشکر و قدردانی

بدین وسیله از استادان رشته غدد درون‌ریز و متابولیسم، دکتر ناهید هاشمی مدنی، دکتر هدی طاهری، دکتر ملیحه قدیر و دکتر آتوسا نجم‌الدین تشکر و قدردانی به عمل می‌آید.

تضاد منافع

در انجام پژوهش حاضر، نویسندگان هیچ‌گونه تضاد منافعی نداشته‌اند.

References

1. Dieng AB, Ruiz FJR, Blei DM. Topic modeling in embedding spaces. Trans Assoc Comput Linguist 2020; 8: 439-53.
2. Kandula S, Curtis D, Hill B, Zeng-Treitler Q. Use of topic modeling for recommending relevant education material to diabetic patients. AMIA Annu Symp Proc 2011; 2011: 674-82.
3. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. SpringerPlus 2016; 5(1): 1608.
4. Saberi MK, Isfandyari Moghaddam A. Assessment of web citation accessibility and decay of health information and medical librarianship articles indexed in ISI. Health Inf Manage 2011; 8(2): 189-97. [In Persian].
5. Lee KW, Morsi A, Naga O. Endocrine disorders. In: Naga O, editor. pediatric board study guide: A last minute review. Cham, Switzerland: Springer International Publishing; 2015. p. 403-33.
6. Heshmati H, Behnampour N, Khorasani F, Moghadam Z. Prevalence of chronic complications of diabete and its related factors in referred type 2 diabetes patients in Freydonkenar diabetes center. Journal of Neyshabur University of Medical Sciences 2014; 1(1): 36-43. [In Persian].
7. Mohammadi F, Shekofteh M, Kazerani M. Iran's scientific publications in the field of endocrinology and metabolism in the Web of Science: A scientometric analysis. Iran J Endocrinol Metab 2020; 22(2): 127-36. [In Persian].
8. Azizi F. Rank of Iranian endocrinology production in the world and Southwest Asia. Iran J Endocrinol Metab 2014; 16(4): 231-4. [In Persian].
9. Emami Z, Khamseh ME, Madani NH, Hariri N, Alibeyk MR, Ghadiqolaei OA. Trend of scientific productions in the field of Endocrinology and Metabolism in Middle East countries during 2007-2013. null 2018; 12(1): 55-71.
10. Shahmoradi L, Ramezani A, Atlasi R, Namazi N, Larijani B. Visualization of knowledge flow in interpersonal

scientific collaboration network endocrinology and metabolism research institute. J Diabetes Metab Disord 2021; 20(1): 815-23.
11. Wang L, Zhang Y, Zhang Y, Xu X, Cao S. Prescription function prediction using topic model and multilabel classifiers. Evid Based Complement Alternat Med 2017; 2017: 8279109.
12. Hofmann T. Probabilistic latent semantic analysis. arXivc 2021; 1301.6705.
13. Taghva, K, Beckley, R, Sadeh, M. A list of farsistopwords, Technical Report 2003-01. Las Vegas, NV: Information Science Research Institute, University of Nevada; 2003.
14. Brown P, Della Pietra S, Pietra V, Lai J, Mercer R. An estimate of an upper bound for the entropy of English. Comput Linguist 1992; 18(1): 31-40.
15. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinform 2015; 16(13): S8.
16. Ray SK, Ahmad A, Kumar C. Review and implementation of topic modeling in Hindi. Appl Artif Intell 2019; 33(11): 979-1007.
17. Larijani B. Iranian National Service Framework for Diabetes. Tehran, Iran: National Committee for Control and Prevention of Noncommunicable Diseases; 2017. p. 22. [In Persian].
18. Azimi-Nezhad M, Herbeth B, Siest G, Dade S, Ndiaye NC, Esmaily H, et al. High prevalence of metabolic syndrome in Iran in comparison with France: What are the components that explain this? Metab Syndr Relat Disord 2012; 10(3): 181-8.
19. Zabetian A, Hadaegh F, Azizi F. Prevalence of metabolic syndrome in Iranian adult population, concordance between the IDF with the ATPIII and the WHO definitions. Diabetes Res Clin Pract 2007; 77(2): 251-7.
20. Wang Y, He W. Endogenous mitochondrial aldehyde dehydrogenase-2 as an Antioxidant in liver. In: Patel VB, Rajendram R, Preedy VR, editors. The liver. Boston, MA: Academic Press; 2018. p. 247-59.
21. Sarrafzadegan N, Mohammmadifard N. Cardiovascular disease in Iran in the last 40 years: Prevalence, Mortality, morbidity, challenges and strategies for cardiovascular prevention. Arch Iran Med 2019; 22(4): 204-10.
22. Khosravi-Boroujeni H, Mohammadifard N, Sarrafzadegan N, Sajjadi F, Maghroun M, Khosravi A, et al. Potato consumption and cardiovascular disease risk factors among Iranian population. Int J Food Sci Nutr 2012; 63(8): 913-20.
23. Haghpanah V, Soliemanpour B, Heshmat R, Mosavi-Jarrahi AR, Tavangar SM, Malekzadeh R, et al. Endocrine cancer in Iran: Based on cancer registry system. Indian J Cancer 2006; 43(2): 80-5.
24. Khodaeian M, Enayati S, Tabatabaei-Malazy O, Amoli MM. Association between genetic variants and diabetes mellitus in Iranian populations: A systematic review of observational studies. J Diabetes Res 2015; 2015: 585917.
25. Koohi F, Salehiniya H, Mohammadian Hafshejani A. Trends in mortality from cardiovascular disease in Iran from 2006-2010. J Sabzevar Univ Med Sci 2015; 22(4): 630-8. [In Persian].
26. Pirdehghan A, Razavi Z, Rajabi R. Evaluation of the factors influencing diabetic control among adolescents with type 1 diabetes. Avicenna J Clin Med 2020; 26(4): 227-33. [In Persian].
27. Larijani FA, Kalantar Motamedi SM, Keshtkar AA, Khashayar P, Koleini Z, Rahim F, et al. The relation between serum vitamin D levels and blood pressure: A population-based study. Acta Med Iran 2014; 52(4): 290-7.
28. Abolhassani F, Mohammadi M, Soltani A. Burden of osteoporosis in Iran. Iran J Public Health 1970; 33(Suppl 1): 18-28.
29. Heshmat R, Mohammad K, Majdzadeh SR, Forouzanfar MH, Bahrami A, Ranjbar Omrani G, et al. Vitamin D Deficiency in Iran: A Multi-center Study among Different Urban Areas. Iran J Public Health 1970; 37(Suppl 2): 72-8.
30. Esteghamati A, Meysamie A, Khalilzadeh O, Rashidi A, Haghazali M, Asgari F, et al. Third National Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in Iran: Methods and results on prevalence of diabetes, hypertension, obesity, central obesity, and dyslipidemia. BMC Public Health 2009; 9: 167.
31. Ebrahimi H, Emamian MH, Hashemi H, Fotouhi A. Dyslipidemia and its risk factors among urban middle-aged Iranians: A population-based study. Diabetes Metab Syndr 2016; 10(3): 149-56.

Health Information Management

Published by ***Vesnu Publications***

Vol 18/ No 4/ October & November 2021

DOI: 10.22122/him.v18i4.4384

Analyzing the Status of Endocrinology and Metabolic Research in Iran Using Text Mining Methods

Omolbanin Asadi-Ghadiklaei[1], Nadjla Hariri[2], Maryam Khademi[3], Fahimeh Babalhavaeji[4]

Original Article

Abstract

Introduction: Due to the importance and status of the endocrine field in the health sector, facilitating information retrieval in this field seems to be important. This study endeavored to run topic modeling of articles published by Iranian researchers in the field of endocrinology and metabolism in the science citation database.

Methods: This descriptive study was done by text mining method. In this study, abstracts of articles were extracted from the science citation database using selected keywords of Medical Subject Headings (MeSH). 5552 articles were retrieved from 1977 to 2019, then the text of abstracts was analyzed and categorized in MATLAB software.

Results: Subject categories with 20 items were extracted in 48 categories. Diabetes with 7145 recurrences has been considered by Iranian researchers more than other topics. Subject category related to metabolic syndrome diseases had the highest number of articles (304 articles) and subject category No. 47 which was related to osteoporosis had the lowest number of articles (51 articles).

Conclusion: Iranian researchers have done more research on metabolic syndrome and less research on osteoporosis. Topic categories including Dwarfism, Parathyroid Diseases, Pituitary Diseases, Gonadal Disorders, Polyendocrinopathies, and Autoimmune that did not exist in the topics resulting from topic modeling indicate a gap in the research of Iranian researchers, that emphasizes the need for more attention to these areas.

Keywords: Topic Modeling; Endocrine System Diseases; Metabolism; Text analysis; Latent Dirichlet Allocation

Received: 08 Aug., 2021 Accepted: 06 Oct., 2021 Published: 07 Oct., 2021

Citation: Asadi-Ghadiklaei O, Hariri N, Khademi M, Babalhavaeji F. **Analyzing the Status of Endocrinology and Metabolic Research in Iran Using Text Mining Methods.** Health Inf Manage 2021; 18(4): 160-5.

Article resulted from PhD thesis No. 972/20961/100090/123 funded by Islamic Azad University, Science and Research Branch.

1- PhD Student, Knowledge and Information Science, Department of Communication and Knowledge Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran

2- Professor, Knowledge and Information Science, Department of Communication and Knowledge Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran

3- Associate Professor, Applied Mathematics, Department of Applied Mathematics, South Tehran Branch Islamic Azad University, Tehran, Iran

4- Associate Professor, Knowledge and Information Science, Department of Communication and Knowledge Sciences, Science and Research Branch, Islamic Azad University, Tehran, Iran

Address for correspondence: Maryam Khademi; Associate Professor, Applied Mathematics, Department of Applied Mathematics, South Tehran Branch Islamic Azad University, Tehran, Iran; Email: khademi@azad.ac.ir