

مدل سازی و تحلیل گرافی برای کشف روابط پنهان میان بیماری ها و علائم: رویکردی داده محور برای تشخیص افتراقی و هموقوعی علائم

سپیده غفاری^۱، رقیه خشا^۲

مقاله پژوهشی

چکیده

مقدمه: درک روابط پنهان میان بیماری ها و علائم یکی از چالش های اساسی در تشخیص افتراقی و تحلیل هموقوعی علائم در حوزه پزشکی است. با توجه به پیچیدگی تعاملات بالینی، شناسایی الگوهای مشترک میان بیماری ها می تواند در بهبود فرایند تشخیص و تصمیم گیری بالینی نقش مهمی ایفا کند. هدف این پژوهش، ارائه رویکردی داده محور برای تحلیل گرافی به منظور کشف و مدل سازی روابط پنهان میان بیماری ها و علائم است.

روش بررسی: در این پژوهش از مجموعه داده عمومی «مجموعه داده پیش بینی بیماری-علامت» در پلتفرم کگل شامل ۴۹۲۰ بیماری و مجموعه ای از علائم، استفاده شد که شامل ۴۱ بیماری و ۱۳۱ علامت منحصر به فرد بود. با بهره گیری از ترکیب خطی شاخص های فرکانس وقوع و ضریب کاپا، یک گراف دوبخشی وزن دار بیماری-علامت ایجاد گردید. سپس با استخراج گراف های تک بخشی بیماری-بیماری و علامت-علامت و به کارگیری الگوریتم های خوشه بندی لووین، ماژولاریتی حریصانه و گبروان-نیومن، ساختار شبکه تحلیل شد. **یافته ها:** یافته ها نشان می دهد که تحلیل گرافی می تواند ابزاری موثر برای کشف روابط پنهان میان بیماری ها و علائم باشد و در بهبود تشخیص افتراقی و طراحی سامانه های هوشمند پشتیبان تصمیم گیری نقش به سزایی ایفا کند.

نتیجه گیری: بلاک چین با ارتقای امنیت و کارایی عملیاتی، تحولی بزرگ در صنعت بهداشت ایجاد می کند. برای موفقیت در پیاده سازی، سیاست گذاران باید بر توسعه زیرساخت های خدمات پزشکی و لجستیک آنلاین تمرکز کنند. این فناوری چارچوبی مستحکم برای پیشبرد سیستم های سلامت هوشمند و بیمار محور فراهم می آورد.

واژه های کلیدی: تشخیص افتراقی؛ بیماری؛ هموقوعی علائم، تحلیل شبکه های پزشکی، گراف دوبخشی، خوشه بندی شبکه ای.

پیام کلیدی: این پژوهش ساختار شبکه ای بیماری ها و علائم را بررسی کرده و روابط پنهان میان آن ها را مشخص می کند. از نتایج پژوهش می توان برای تشخیص دقیق تر بیماری ها، شناسایی علائم کلیدی و بهبود کارایی سامانه های پشتیبان تصمیم گیری پزشکی استفاده کرد.

تاریخ انتشار: ۱۴۰۴/۱۰/۱۰

پذیرش مقاله: ۱۴۰۴/۹/۲۷

دریافت مقاله: ۱۴۰۴/۷/۲۰

ارجاع: غفاری سپیده، خشا رقیه. مدل سازی و تحلیل گرافی برای کشف روابط پنهان میان بیماری ها و علائم: رویکردی داده محور برای تشخیص افتراقی و هموقوعی علائم. مدیریت اطلاعات سلامت ۲۲: ۱۳۴-۱۳۹ (۴) ۱۹۸-۱۸۹.

مقدمه

در حوزه سلامت، شناسایی و استخراج ارتباط میان بیماری ها و علائم نه تنها به درک عمیق تر از روابط متقابل بیماری های مختلف کمک می کند، بلکه امکان شناسایی علائم مشترک و ارائه مبنای دقیق تر برای پیشگیری و درمان را فراهم می سازد (۱). به دلیل همپوشانی علائم در بیماری های مختلف و مشاهده هموقوعی علائم، تشخیص دقیق دشوار است و یافتن ارتباط های پنهان میان علائم به کادر درمان در تشخیص دقیق تر و تعیین مؤثرتر برنامه های مراقبتی کمک می کند (۲).

گزارش سازمان بهداشت جهانی در سال ۲۰۱۶ تأکید می کند که مهم ترین وظیفه پزشکان، دستیابی به تشخیص صحیح است (۳). در تاریخچه انفورماتیک پزشکی، ابزارهای یادگیری ماشین و سیستم های پشتیبان تصمیم گیری برای کمک به فرایند تشخیص بیماری ها استفاده شده اند. در میان این ابزارها و سیستم ها، سیستم های خبره، به ویژه الگوریتم های استدلال مبتنی بر مورد، بخش مهمی را تشکیل داده اند، اما فرایند کسب دانش از خبرگان و به روز رسانی پایگاه دانش در این سیستم ها، دشوار و زمان بر بوده است (۳).

تشخیص افتراقی فرایندی است که احتمال ابتلای یک فرد به یک بیماری خاص در مقایسه با سایر بیماری های دارای علائم مشابه، بررسی می شود (۳). خوشه علائم به مجموعه ای از چند علامت اشاره دارد که همزمان رخ می دهند و به آن هموقوعی علائم می گویند (۴).

تحلیل شبکه، امکان شناسایی علائم کلیدی و درک ماهیت پیچیده خوشه های علامتی را فراهم می کند و روابط میان علائم را به صورت بصری نمایش می دهد (۵). تحلیل شبکه علائم برخلاف روش های سنتی خوشه بندی مانند الگوریتم K-میانیگین (K-means)، تعامل متقابل علائم درون و میان خوشه ها را نشان می دهد (۴).

۱. دانشجوی کارشناسی ارشد، مهندسی سیستم های سلامت، دانشکده مهندسی صنایع و سیستم های مدیریت، دانشگاه صنعتی امیرکبیر، تهران، ایران.
۲. استادیار، دانشکده مهندسی صنایع و سیستم های مدیریت، دانشگاه صنعتی امیرکبیر، تهران، ایران

نویسنده طرف مکاتبه: رقیه خشا، استادیار، دانشکده مهندسی صنایع و سیستم های مدیریت، دانشگاه صنعتی امیرکبیر، تهران، ایران

Email: r.khasha@aut.ac.ir

روش بررسی

پژوهش حاضر از نوع تحلیلی-توصیفی است و با بهره‌گیری از تحلیل گراف دوبخشی بیماری-علائم انجام شده است. در این پژوهش، گراف بیماری-بیماری نگاشتی از گراف بیماری-علائم در نظر گرفته شده است به‌طوری که میان دو بیماری زمانی یال برقرار می‌شود که حداقل یک علامت مشترک داشته باشند. به‌طور مشابه، در گراف علامت-علامت در نظر گرفته شده است که دو علامت زمانی یال مشترک دارند که حداقل در یک بیماری مشترک باشند پژوهش حاضر، در سه مرحله انجام می‌گیرد:

۱. تحلیل شبکه‌ای ارتباط میان بیماری‌ها و علائم: یک مدل گراف دوبخشی بیماری-علائم ایجاد می‌گردد که مجموعه گره‌های این گراف، بیماری‌ها و علائم را نمایان می‌سازد.

۲. کشف بیماری‌های مشابه: بر اساس گراف دوبخشی، یک گراف بر پایه مجموعه بیماری‌ها ایجاد می‌شود تا هدف خوشه‌بندی بیماری‌های مشابه را محقق سازد.

۳. بررسی شبکه هم‌وقوعی علائم: بر اساس گراف دوبخشی، یک گراف بر پایه مجموعه علائم ایجاد می‌شود تا روابط کلیدی میان علائم را شناسایی نماید.

• داده‌های مورد استفاده

داده‌های مورد استفاده برای پیاده‌سازی روش پیشنهادی، از مجموعه داده عمومی «مجموعه داده پیش‌بینی بیماری-علامت» در پلتفرم کگل استفاده شده است. کگل، پلتفرم آنلاین برای برگزاری رقابت‌های علم داده در حوزه‌های تجاری و پژوهشی است (۱۶). هر سطر نشان‌دهنده یک بیماری خاص و مجموعه‌ای از علائم آن بیماری است که در ستون‌های مربوطه ثبت شده است. ستون اول بیانگر نام بیماری‌ها و سایر ستون‌ها بیانگر علائم مرتبط با هر بیماری است. این مجموعه داده شامل ۴۹۲۰ ردیف، ۱۸ ستون و حداکثر ۱۷ علامت برای هر بیماری است. پس از تبدیل داده‌های جدولی به گراف دوبخشی، ۴۱ بیماری و ۱۳۱ علامت منحصر به فرد شناسایی شده است.

• وزن‌دهی

برای وزن‌دهی یال‌های موجود در گراف بیماری-علائم، از دو شاخص فرکانس وقوع و ضریب کاپا استفاده شده است. فرکانس وقوع بر اساس تعداد بیماری‌های سنجیده می‌شود که هم‌زمان، یک بیماری و آن علامت خاص را دارند. در این حالت، وزن یال میزان شیوع هر علامت در ارتباط با آن بیماری را بیان می‌کند. ضریب کاپا معیاری برای سنجش میزان توافق و ارزیابی هم‌پوشانی میان بیماری‌ها و علائم است و نشان می‌دهد که آیا یک علامت خاص معمولاً با یک بیماری مشخص همراه است یا خیر.

در این پژوهش، ترکیب‌های خطی میان آن دو شاخص نیز به‌عنوان وزن یال‌ها در نظر گرفته شده است. برای مقایسه موثر میان این دو ضریب، ضریب فرکانس وقوع به بازه $[0, 1]$ ، و ضریب کاپا از بازه $[-1, 1]$ به بازه $[0, 1]$ انتقال یافته است و ترکیب‌های خطی بر اساس مقدار نرمال شده ساخته شده است. ترکیب خطی از رابطه ۱ به‌دست می‌آید:

رابطه ۱

$$\text{combined_weighted} = (\alpha)kappa_scaled + (1-\alpha)frequency_scaled$$

$$\alpha \in \{0.1, 0.2, \dots, 0.9\}$$

شناسایی خوشه‌های علائم با روش‌های تحلیلی به گروه‌بندی علائم مشابه، کمک می‌کند. مداخلات فردی مبتنی بر درمان جداگانه علائم، بدون تمرکز بر روابط میان آن‌ها، به بهبود ناچیز علائم منجر شده و با اتلاف منابع درمانی و افزایش هزینه‌های ارایه خدمات سلامت همراه است (۴). در مقابل، شناسایی خوشه‌های بیماران با الگوهای علائم مشابه، امکان طراحی مداخلات هدفمند متناسب با نیازهای هر گروه از بیماران را فراهم کرده و به ارایه مراقبت بالینی شخصی‌سازی شده کمک می‌کند (۶).

در سال‌های اخیر، از روش‌های گراف‌کاوی و تحلیل شبکه در حوزه‌های مختلفی از جمله حوزه سلامت استفاده شده است. در میان این روش‌ها، تشخیص اجتماع به‌عنوان یکی از رویکردهای محبوب شناخته می‌شود (۷). این فرایند با استفاده از معیارهای شباهت یا فاصله انجام می‌شود (۸). هدف از تشخیص اجتماع، تقسیم شبکه به بخش‌هایی است که درون هر بخش، گره‌ها به‌طور قوی‌تری به یکدیگر متصل باشند. روش‌های خوشه‌بندی، ابزار مهمی برای درک ساختار و رفتار سیستم‌ها محسوب می‌شوند (۹).

مطالعه نظام‌مند بیماری‌های انسانی از طریق تحلیل روابط میان موجودیت‌های زیست‌پزشکی پیشرفت چشمگیری داشته است و استفاده از گراف‌ها برای نمایش بیماری‌ها، علائم و روابط میان آن‌ها رواج یافته است (۱). از مهم‌ترین کاربرد گراف‌های دوبخشی در زیست‌پزشکی می‌توان به بازکاربرد یابی دارو بر اساس شبکه‌های دارو-بیماری (۱۰)، پیش‌بینی ارتباط بالقوه میان ریز RNAها و بیماری‌ها در شبکه‌های ریز-RNA بیماری (۱۱) و کشف نقش‌های عملکردی RNAهای غیر کدکننده طولانی از طریق شبکه‌های تعامل-IncrRNA پروتئین (۱۲) اشاره کرد. مشاهده می‌شود که بیشتر مطالعات بر جنبه‌هایی نظیر ژنتیک متمرکز بوده‌اند (۱).

توسعه سیستم‌های تشخیص پزشکی از مهم‌ترین کاربردهای گراف بیماری-علائم است که با بهره‌گیری از علائم، تشخیص‌های ممکن را ارایه می‌دهند (۱۳). برای ساخت این گراف‌ها، می‌توان از ترکیب واژگان MeSH و پایگاه داده PubMed استفاده کرد. واژگان MeSH دقت بالایی دارند، اما به‌روزرسانی سالانه و محدودیت پوشش اصطلاحات بیماری، از نقاط ضعف آن است (۱۴). بهره‌گیری از آنتولوژی‌های استاندارد بیماری‌ها و علائم امکان استخراج و ترکیب خودکار مفاهیم را فراهم می‌کند (۱۳). پرونده الکترونیک سلامت و نتایج آزمایشگاهی، منابع ارزشمندی محسوب می‌شوند، اما محدودیت‌هایی مانند خطاهای مستندسازی، پیچیدگی تحلیل متون روایی و چالش‌های حریم خصوصی، استفاده از آن‌ها را دشوار می‌سازد (۱۴).

بررسی‌ها نشان می‌دهد تا کنون پژوهشی به‌طور هم‌زمان به تشخیص افتراقی بیماری‌ها و هم‌وقوعی علائم با روش‌های مبتنی بر گراف نپرداخته است. با توجه به پیچیدگی روابط میان بیماری‌ها و علائم، بهره‌گیری از گراف‌های دوبخشی وزن‌دار می‌تواند شبکه‌ای میان بیماری‌ها و علائم ایجاد کند که علاوه بر نشان دادن ارتباط مستقیم میان آن‌ها، هم‌وقوعی علائم و شناسایی بیماری‌های مشابه را فراهم می‌کند. از این‌رو، در این پژوهش روش‌های داده‌محور مبتنی بر گراف برای ارایه تشخیص دقیق‌تر، بررسی خواهند شد.

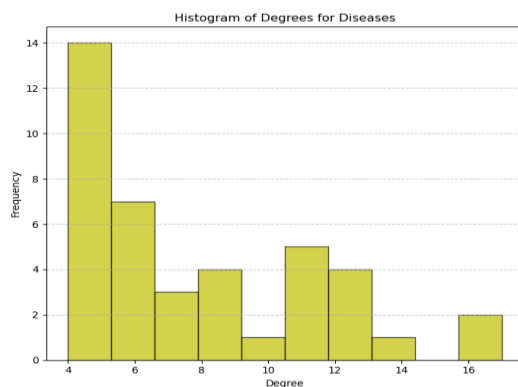
Network می‌سجد (۱۷). الگوریتم لووین (Louvain) ماهیت حریصانه دارد و برای گراف‌های وزن‌دار مناسب است. این الگوریتم به دلیل همگرایی سریع و توانایی در تقسیم‌بندی سلسله‌مراتبی، کاربرد گسترده‌ای دارد (۷). الگوریتم ماژولاریتی حریصانه (Greedy Modularity) با رویکردی ساده و کارا، ماژولاریتی شبکه را افزایش می‌دهد و گره‌ها را به گروه‌های متمایز تقسیم می‌کند (۱۸). الگوریتم گیروان-نیومن (Girvan-Newman) با حذف تدریجی یال‌های دارای مرکزیت بینابینی بالا، اجتماعات را به صورت سلسله‌مراتبی شناسایی می‌کند (۱۹).

• پیاده‌سازی

برای پیاده‌سازی این پژوهش از زبان برنامه‌نویسی پایتون و کتابخانه NetworkX 3.2.1 استفاده شده است. برای مصورسازی گراف‌ها از نرم‌افزار Gephi 0.10 بهره‌گیری شده است.

یافته‌ها

تعداد گره‌ها و یال‌ها در گراف دویخشی به ترتیب برابر با ۱۷۲ و ۳۲۱ است. چگالی هر دو بخش، ۰/۰۶ است. درجه هر بیماری نشان‌دهنده تعداد علائم مرتبط با آن بیماری است. بیماری‌های سرماخوردگی، سل، تب دنگی، هیپاتیت E، کم‌کاری تیروئید، هیپوگلیسمی، هیپاتیت B، آبله‌مرغان، هیپاتیت A، پرکاری تیروئید، حصبه، ذات‌الریه و دیابت به ترتیب بیشترین درجه را دارند؛ سرماخوردگی با درجه ۱۷ و سل با درجه ۱۶ بیشترین ارتباط را با علائم دارند، در مقابل دیابت با درجه ۱۰ ارتباط کمتری با علائم دارد. درجه هر علامت نشان‌دهنده تعداد بیماری‌های مرتبط با آن علامت است. علائم خستگی، استفراغ، تب بالا، بی‌اشتهایی، سردرد و حالت تهوع درجه بالایی دارند..



شکل ۱: هیستوگرام درجه بیماری‌ها و علائم

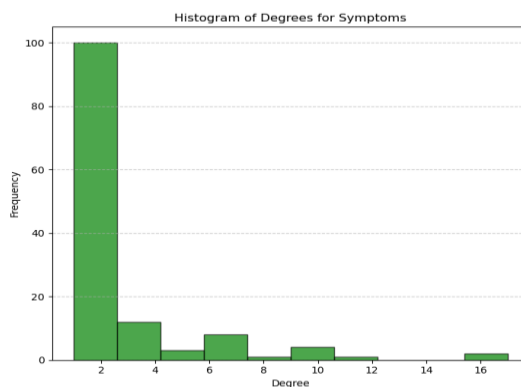
است؛ بنابراین مقدار بیشتر این شاخص بیانگر فراگیرتر بودن آن گره در شبکه است. بیماری‌های سرماخوردگی، سل و تب دنگی بیشترین ارتباط را با علائم مختلف دارند و علائم خستگی، استفراغ و تب بالا در بیشترین تعداد بیماری‌ها مشاهده می‌شوند. مرکزیت نزدیکی بیانگر میزان دسترسی سریع یک گره به سایر گره‌های شبکه است. گره‌هایی با مقدار بالاتر این شاخص، سریع‌تر با سایر بخش‌های شبکه ارتباط برقرار می‌کنند. تب دنگی، آبله‌مرغان و حصبه در میان بیماری‌ها و خستگی، استفراغ و تب بالا در میان علائم با بیشترین نزدیکی، نقش محوری در ساختار کلی شبکه دارند.

در فرایند وزن‌دهی گراف‌های بیماری-بیماری و علامت-علامت، در صورتی که دو گره دارای یک موجودیت مشترک باشند، وزن یال میان آن‌ها با میانگین وزن‌های متناظر تعیین می‌شود. در مواردی که بیش از یک موجودیت مشترک وجود دارد، دو رویکرد لحاظ می‌گردد؛ رویکرد جمع اوزان که بر فراوانی تاکید دارد و با افزایش تعداد آن، وزن یال افزایش می‌یابد و رویکرد میانگین اوزان که شدت را لحاظ می‌کند و وزن یال متأثر از میانگین آن‌ها خواهد بود. هر دو رویکرد برای تمامی مجموعه وزن‌های از پیش تعیین شده، اعمال شده است.

• خوشه‌بندی شبکه‌ای

برای سنجش میزان شباهت میان بیماری‌ها از سه معیار شبکه‌ای شباهت جاکارد (Jaccard Similarity)، همبستگی پیرسون (Pearson Correlation) و شباهت کسینوسی (Cosine Similarity) استفاده شده است. شاخص شباهت جاکارد نسبت تعداد همسایه‌های مشترک دو بیماری را به کل همسایگان آن‌ها محاسبه می‌کند. شاخص همبستگی پیرسون به جای تمرکز صرف بر شمار همسایه‌های مشترک، الگوی روابط هر بیماری با سایر بیماری‌ها را در کل شبکه با یکدیگر مقایسه می‌کند. شاخص شباهت کسینوسی نسبت همسایگان مشترک دو بیماری را به حاصل ضرب تعداد همسایگان آن‌ها محاسبه می‌کند که تأثیر تفاوت زیاد در درجات گره‌ها را کاهش داده و امکان مقایسه عادلانه‌تری را فراهم می‌سازد.

برای شناسایی گروه‌های بیماری مشابه، از سه الگوریتم خوشه‌بندی و معیار ماژولاریتی برای مقایسه استفاده شده است. ماژولاریتی، میزان خوشه‌بندی گره‌ها در جوامع را از طریق مقایسه تراکم ارتباطات درون گروهی با یک شبکه مدل تهی (Null Model)



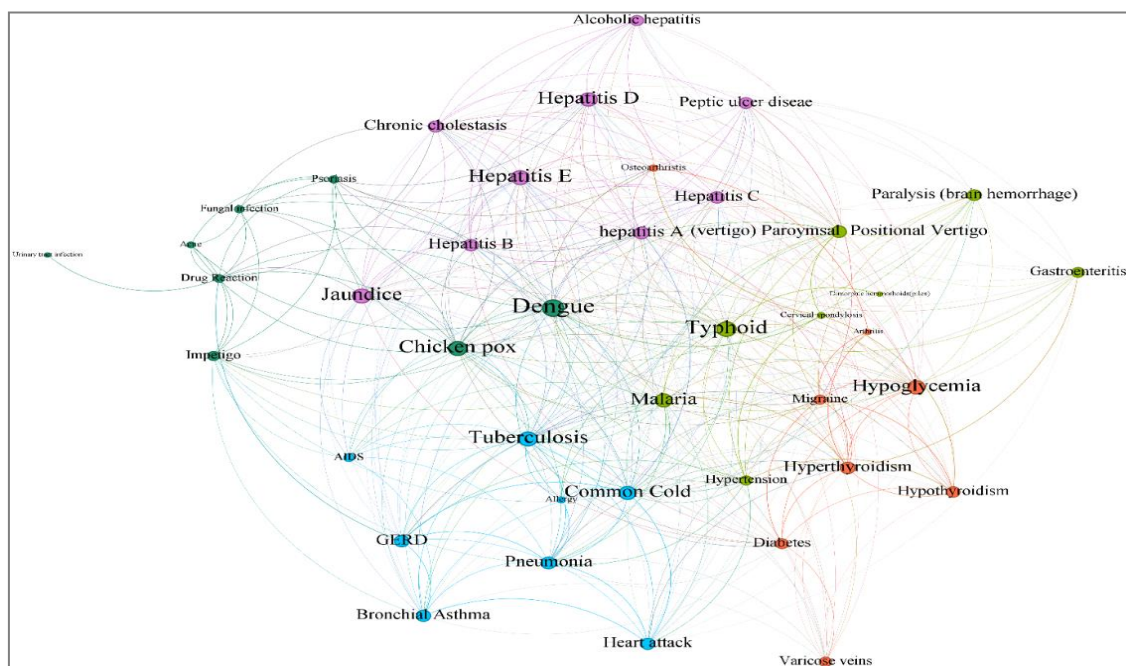
هیستوگرام سمت چپ، توزیع درجه گره‌های بیماری را نشان می‌دهد که بیشترین فراوانی در درجه ۴ تا ۵ متمرکز شده است. هیستوگرام سمت راست، توزیع درجه علائم را نمایش می‌دهد که اکثریت علائم بسیار اختصاصی و تنها به یک یا دو بیماری مرتبط هستند و تعداد اندکی از علائم دارای درجه بالا هستند که نقش علامت‌های کلیدی را ایفا می‌کنند. محاسبه مرکزیت‌های درجه‌ای، نزدیکی و بینابینی برای گره‌های بیماری‌ها و علائم میزان اهمیت و نقش هر گره در ساختار شبکه را مشخص می‌کند. مرکزیت درجه‌ای نشان می‌دهد هر بیماری با چند علامت و هر علامت با چند بیماری در ارتباط

مرکزیت بینایی نشان می‌دهد یک گره تا چه اندازه نقش پل ارتباطی میان سایر گره‌ها را ایفا می‌کند. گره‌هایی با مقدار بالاتر این شاخص در حفظ پیوستگی شبکه و ارتباط بین خوشه‌ها نقش کلیدی دارند. تب دنگی، حصه و سرماخوردگی در بیماری‌ها و خستگی، استفراغ و بثورات پوستی در علایم، بیشترین نقش واسطه‌ای را ایفا می‌کنند.

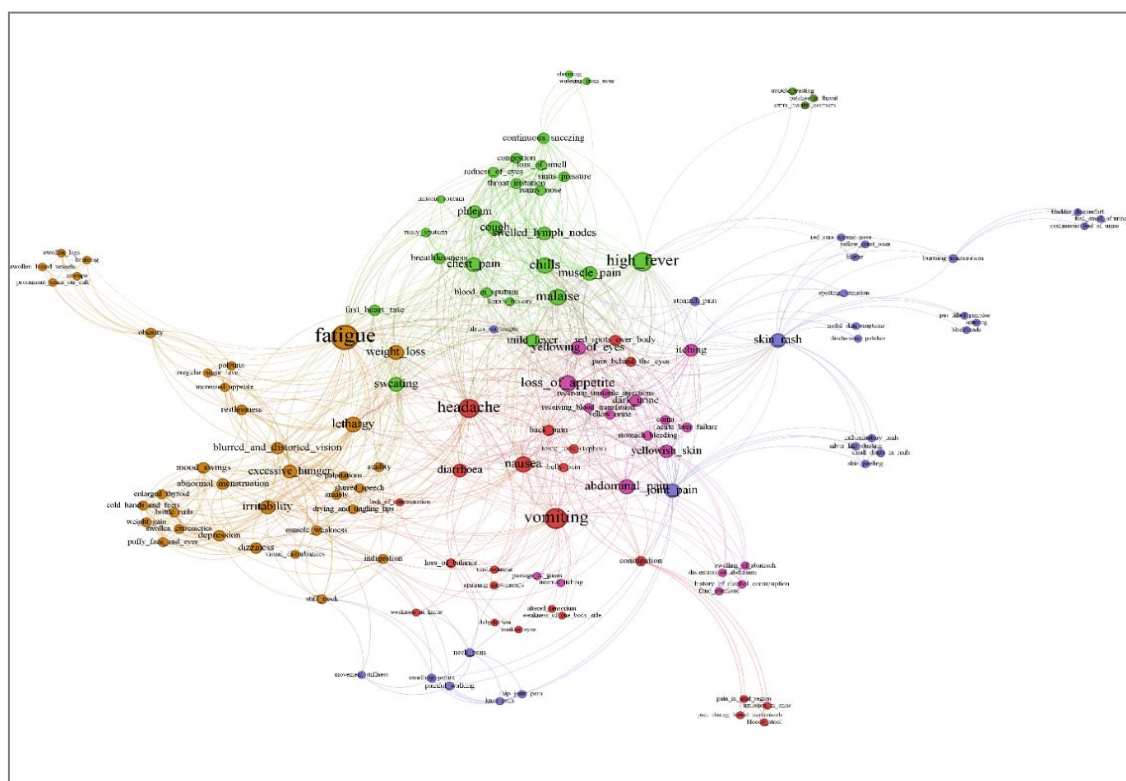
تعداد گره‌ها و تعداد یال‌ها در گراف بیماری-بیماری به ترتیب برابر با ۴۱ و ۳۸۰ است. چگالی این گراف، ۰/۴۶ است. میانگین درجه‌ها ۱۹ است؛ یعنی هر بیماری به‌طور متوسط با ۱۹ بیماری دیگر علایم مشترک دارد. بیماری تب دنگی بیشترین تعداد علایم را دارد و پس از آن بیماری‌های حصه، یرقان، آبله‌مرغان و هپاتیت E قرار دارند. بیماری‌های تب دنگی و سل با بالاترین مرکزیت نزدیکی، کمترین فاصله را با سایر گره‌ها دارند. تب دنگی با بالاترین مرکزیت بینایی به‌عنوان پل میان خوشه‌های مختلف بیماری عمل می‌کند.

گراف	ردیف	نام الگوریتم	مقدار ماژولاریتی	تعداد خوشه‌ها	وزن مربوطه
بیماری-بیماری	۱	لووین	-/۲۷۸۹	۴	kappa_sum
	۲	ماژولاریتی حریصانه	-/۲۷۴۴	۵	kappa_sum
	۳	گیروان-نیومن	-/۰۴۳۴	۹	alpha_0.4_mean
علامت-علامت	۱	لووین	-/۴۳۹۸	۶	kappa_mean
	۲	ماژولاریتی حریصانه	-/۳۹۰۳	۱۱	kappa_sum
	۳	گیروان-نیومن	-/۱۵۱۶	۸	alpha_0.6_mean





(ب)



(ج)

شکل ۲: نتایج خوشه‌بندی الگوریتم‌ها: (الف) الگوریتم لووین در گراف بیماری-بیماری، (ب) الگوریتم ماژولاریتی حریصانه در گراف بیماری-بیماری، (ج) الگوریتم لووین در گراف علامت-علامت

نتایج الگوریتم‌های خوشه‌بندی برای گراف‌های بیماری-بیماری و علامت-

علامت در جدول ۲ آمده است.

جدول ۲: نتایج الگوریتم خوشه‌بندی

گراف	الگوریتم	شماره	اندازه	درصد	اجزای هر خوشه
بیماری-بیماری	لووین	۱	۸	۱۹/۵۱	آکنه، آبله‌مرغان، تب دنگی، واکنش دارویی، عفونت قارچی، زرد زخم، پسوریازیس و عفونت ادراری
		۲	۹	۲۱/۹۵	هیپاتیت الکلی، کلستاز مزمن، هیپاتیت B، هیپاتیت C، هیپاتیت D، هیپاتیت E، یرقان، زخم گوارشی و هیپاتیت A
		۳	۱۲	۳۹/۲۷	ایدز، آلرژی، آسم، برونش، سرماخوردگی، هموروئید، رفلاکس معده، گاستروانتریت، سکنه قلبی، مالاریا، ذات‌الریه، سل و حصه
		۴	۱۲	۳۹/۳۷	سرگیجه موضعی حمله‌ای، آرتریت، اسپوندیلوز گردن، دیابت، فشار خون بالا، پرکاری تیروئید، هیپوگلیسمی، کم‌کاری تیروئید، میگرن، استئوآرتریت، فلج ناشی از خون‌ریزی مغزی و واریس
		۱	۹	۲۱/۹۵	هیپاتیت الکلی، کلستاز مزمن، هیپاتیت B، هیپاتیت C، هیپاتیت D، هیپاتیت E، یرقان، زخم گوارشی و هیپاتیت A
	ماژولاریتی حریصانه	۲	۸	۱۹/۵۱	سرگیجه موضعی حمله‌ای، اسپوندیلوز گردن، هموروئید، گاستروانتریت، فشار خون بالا، مالاریا، فلج ناشی از خون‌ریزی مغزی و حصه
		۳	۸	۱۹/۵۱	ایدز، آلرژی، آسم، برونش، سرماخوردگی، رفلاکس معده، سکنه قلبی، ذات‌الریه و سل
		۴	۸	۱۹/۵۱	آرتریت، دیابت، پرکاری تیروئید، هیپوگلیسمی، کم‌کاری تیروئید، میگرن، استئوآرتریت و واریس
		۵	۸	۱۹/۵۱	آکنه، آبله‌مرغان، تب دنگی، واکنش دارویی، عفونت قارچی، زرد زخم، پسوریازیس و عفونت ادراری
		۱	۳	۲/۲۹	روابط خارج از ازدواج، تحلیل عضلانی و ضایعات گلو
علامت-علامت	لووین	۲	۲۳	۱۷/۵۶	اختلال سطح هوشیاری، کمر درد، شکم درد، اسهال خونی، یبوست، کم‌آبی بدن، اسهال، سردرد، سوزش در مقعد، عدم تمرکز، بی‌ثباتی، حالت تهوع، درد پشت چشم، درد حین دفع مدفوع، درد در ناحیه مقعد، دانه‌های قرمز روی بدن، سرگیجه، گودی زیر چشم، چهره بیمار، عدم تعادل، استفراغ، ضعف و ضعف یک طرف بدن
		۳	۱۸	۱۳/۷۴	دل درد، نارسای حاد کبد، کما، تیرگی ادرار، اتساع شکم، اضافه بار مایعات، سابقه مصرف الکل، خارش، خارش داخلی، بی‌اشتهایی، دفع گاز، دریافت خون، دریافت تزریقات غیر استریل، خون‌ریزی معده، ورم معده، رنگ ادرار، زردی چشم و زردی پوست
		۴	۲۵	۱۹/۰۸	خلط خونی، تنگی نفس، درد قفسه سینه، لرز، گرفتگی، عطسه مداوم، سرفه، سابقه خانوادگی، ضربان قلب بالا، تب بالا، از دست دادن بویایی، کسالت، تب خفیف، خلط شفاف، درد عضلانی، خلط، قرمزی چشم، آبریزش بینی، خلط زنگاری رنگ، لرزیدن، فشار سینوس، تعریق، تورم غدد لنفاوی، گلو درد و آبریزش چشم
	لووین	۵	۳۵	۲۶/۷۲	قاعدگی نامنظم، ترشح اسید معده، اضطراب، تاری و انحراف دید، ناخن شکننده، کبودی، سردی دست و پا، گرفتگی عضلانی، افسردگی، گیجی، خشکی و گزگز کردن لب‌ها، بزرگی غده تیروئید، احساس گرسنگی شدید، خستگی، پر اشتهایی، سوءهاضمه، نوسان قند خون، تحریک‌پذیری، کاهش سطح هوشیاری، نوسانات خلقی، ضعف عضلانی، چاقی، تیش قلب، تکرر ادرار، واریس ساق پا، یف صورت و چشم‌ها، بی‌قراری، اختلال گفتاری، سفتی گردن، تورم رگ‌های خونی، ورم اندام‌ها، ورم پاها، اختلالات بینایی، افزایش وزن و کاهش وزن
		۶	۲۷	۲۰/۶۱	جوش سر سیاه، درد مثانه، تاول، سوزش ادرار، احساس مداوم ادرار، لکه‌های پوستی با رنگ غیرطبیعی، بوی بد ادرار، درد مفصل ران، قارچ ناخن، درد مفاصل، زانو درد، خشکی مفاصل، گردن درد، ضایعات پوستی، پا درد حین راه رفتن، جوش‌های چرکی، زخم قرمز دور بینی، جای زخم، پوسته‌ریزی نقره‌ای رنگ، پوسته شدن پوست، بثورات پوستی، فرورفتگی ناخن، ادرار خونی، معده درد، ورم مفاصل، آفت زبان و ترشحات دلمه‌بسته زرد رنگ

بحث

این مطالعه با هدف تحلیل روابط شبکه‌ای میان بیماری‌ها و علائم، شناسایی خوشه‌های معنادار و استخراج گره‌های کلیدی انجام شد. نتایج نشان داد که هپاتیت D و هپاتیت E بیشترین شباهت را به یکدیگر دارند که ناشی از هم‌پوشانی بالای الگوهای علامتی و ویژگی‌های پاتوفیزیولوژیک مشترک آن‌ها است. آسم برونشی و ذات‌الریه به دلیل اشتراک علائم تنفسی در یک خوشه قرار گرفتند که با درک بالینی موجود از بیماری‌های ریوی هم‌خوانی دارد. نزدیکی شبکه‌ای دیابت و واریس بیانگر اشتراک برخی علائم متابولیک است که اغلب به‌عنوان بیماری‌های همراه مشاهده می‌شوند. در بیماری‌های عصبی، ارتباط میان فلج ناشی از خون‌ریزی مغزی و سرگیجه موضعی حمله‌ای نشان‌دهنده هم‌پوشانی علائم عصبی و مسیره‌ای تشخیصی مشترک است.

اگرچه بیماری‌های حصبه و سل در شاخص‌های جاکارد و کسینوسی شباهت بالایی دارند، همبستگی کمتر پیرسون نشان می‌دهد که این دو بیماری الگوی ارتباطی متفاوتی با سایر بیماری‌ها در شبکه دارند. از نظر بالینی، این یافته تأکید می‌کند که شباهت ظاهری علائم نباید به‌تنهایی مبنای تشخیص قرار گیرد و توجه به الگوی کلی علائم ضروری است.

تحلیل معیارهای مرکزیت نشان داد که بیماری تب دنگی با بالاترین میزان مرکزیت در شبکه بیماری‌ها از نظر الگوی علائم، با طیف وسیعی از بیماری‌ها در ارتباط است. بیماری سل بر اساس معیار نزدیکی، نقش محوری در شبکه بیماری‌ها دارد. از دیدگاه بالینی، چنین بیماری‌هایی در مواجهه با بیماران دارای علائم غیراختصاصی، زودتر در تشخیص افتراقی مدنظر قرار می‌گیرند. در میان علائم، خستگی به‌عنوان یک علامت مرکزی شناسایی شد که مطابق با مطالعات، به‌عنوان یک نشانه بسیار شایع در طیف وسیعی از بیماری‌ها شناخته می‌شود (۲۰،۱۹).

Lu و همکاران (۲۰۲۳)، خوشه‌های نسبتاً پایدار علائم از جمله خوشه‌های روان‌شناختی و گوارشی و علائم محوری در طول درمان را گزارش داده‌اند که با یافته‌های این پژوهش در خصوص پایداری خوشه‌ها و نقش کلیدی علائمی مانند خستگی هم‌راستا است (۲۱).

Kim و همکاران (۲۰۰۵)، تأکید می‌کنند که خوشه‌های علائم باید دارای همبستگی درون‌خوشه‌ای قوی‌تر نسبت به بین‌خوشه‌ای و پایداری نسبی در زمان باشند (۲۲).

Qi و همکاران (۲۰۲۴)، نشان داده‌اند که علائمی مانند خستگی، درد و اختلال خواب به‌صورت درهم‌تنیده و متقابل بروز می‌کنند و مداخلات باید به ساختار شبکه‌ای توجه کنند (۲۳). هم‌سویی این مطالعات با یافته‌های مطالعه حاضر از جمله شناسایی خوشه‌های مشابه و نقش علائم مرکزی و رابط، اعتبار چارچوب شبکه‌ای ارایه شده را تقویت می‌کند.

میانگین بالای ضریب خوشه‌بندی در گراف‌ها نشان‌دهنده تمایل قوی بیماری‌ها و علائم به تشکیل خوشه‌های مشخص است. در گراف بیماری - بیماری، الگوریتم

لووین چهار خوشه اصلی را شناسایی کرد. خوشه نخست شامل بیماری‌های پوستی و عفونی است که اغلب با نشانه‌های بالینی قابل‌مشاهده همراه هستند. خوشه دوم، بیماری‌های مرتبط با دستگاه گوارش، کبد و کیسه صفرا را شامل می‌شود که معمولاً علائم گوارشی مشترک دارند. خوشه سوم شامل بیماری‌های عفونی، سیستمیک و عمومی است که از طریق علائم غیراختصاصی مانند تب و ضعف به یکدیگر مرتبط شده‌اند. خوشه چهارم شامل بیماری‌های غدد و متابولیک، اختلالات عصبی، بیماری‌های اسکلتی - عضلانی و بیماری‌های مزمن قلبی - عروقی است که اغلب در بیماران با بیماری‌های مزمن مشاهده می‌شوند. این خوشه‌بندی می‌تواند به پزشک کمک کند تا هنگام مواجهه با یک بیمار، به‌جای بررسی تک‌تک بیماری‌ها به‌صورت مجزا، مجموعه‌ای از بیماری‌های مرتبط را هم‌زمان در تشخیص افتراقی مدنظر قرار دهد. این موضوع در بیماران مسن با بیماری‌های مزمن چندگانه اهمیت دارد.

الگوریتم ماژولاریتی حریصانه با تفکیک دقیق‌تر خوشه‌ها، توانست بیماری‌های مزمن و متابولیک، بیماری‌های تنفسی و ایمنی و بیماری‌های حاد و سیستمیک را به‌صورت مجزا شناسایی کند. این تفکیک دقیق‌تر نشان می‌دهد که ساختار شبکه می‌تواند برای اولویت‌بندی بررسی‌های تشخیصی، انتخاب آزمایش‌های هدفمند و پیش‌بینی بیماری‌های همراه استفاده شود.

در گراف علامت - علامت، الگوریتم لووین شش خوشه مجزا را شناسایی کرد که هر کدام با یک الگوی بالینی قابل‌تفسیر هم‌خوانی دارند. خوشه‌هایی شامل علائم نقص ایمنی، علائم ترکیبی گوارشی و عصبی مرتبط با عفونت، علائم بیماری‌های کبدی، علائم تنفسی و عفونی، علائم متابولیک و غدد و علائم پوستی، مفصلی و ادراری استخراج شدند. از دیدگاه کاربردی، این خوشه‌ها در مراحل اولیه ارزیابی بیمار و به‌ویژه در تریاژ به پزشک کمک می‌کنند تا بدانند کدام علائم احتمالاً بخشی از یک الگوی بالینی واحد هستند و نیاز به بررسی هم‌زمان دارند.

تحلیل کلی نتایج نشان می‌دهد که وجود علائم با ارتباط ضعیف لزوماً به معنای ابتلا به بیماری نادر نیست. در بسیاری از بیماری‌های شایع مانند دیابت یا اختلالات تیروئیدی، مجموعه‌ای از علائم متنوع و ظاهراً نامرتبط مشاهده می‌شود که همگی به یک بیماری متداول بازمی‌گردند. این یافته با شواهد بالینی موجود هم‌خوانی دارد و نشان می‌دهد که پیچیدگی علائم الزاماً نشانگر نادر بودن بیماری نیست. در مقابل، شناسایی خوشه‌های خاص علائم مانند ترکیب علائم عصبی و گوارشی یا علائم پوستی و مفصلی می‌تواند پزشک را به سمت بررسی بیماری‌های مشخص‌تری مانند حصبه یا آرتریت پسوریاتیک هدایت کند. این هم‌خوانی میان خوشه‌های استخراج‌شده و الگوهای شناخته‌شده بالینی، به چارچوب شبکه‌ای پیشنهادی اعتبار می‌بخشد.

پژوهش حاضر با محدودیت‌هایی نیز همراه است. نخست آنکه داده‌های موجود ماهیتی ایستا دارند و شدت و توالی زمانی بروز علائم که در تصمیم‌گیری بالینی اهمیت بالایی دارد، منعکس نمی‌شوند. دوم، برخی بیماری‌ها و علائم نادر در داده‌ها پوشش داده نشده‌اند که تعمیم‌پذیری نتایج را در محیط‌های بالینی محدود می‌سازد.

کاربردهای این مدل در حوزه‌هایی نظیر غربالگری، راه‌اندازی کلینیک‌های تخصصی ترکیبی و تدوین سیستم امتیازبندی خطر مبتنی بر درجه مرکزیت علائم می‌تواند منجر به کاهش خطاهای تشخیصی، حذف مداخلات غیرضروری و در نتیجه افزایش بهره‌وری اقتصادی در نظام سلامت شود. در مجموع، به‌کارگیری این راهکار رویکرد طبابت را از نگاه تک بیماری و واکنشی به‌سوی رویکردی چندبیماری و پیش‌گیرانه سوق خواهد داد.

پیشنهادهای

پیشنهادهای می‌شود مدل ارایه شده روی داده‌های پرونده‌های الکترونیک سلامت پیاده‌سازی شود تا دقت و تعمیم‌پذیری نتایج افزایش یابد. همچنین، برای غلبه بر ماهیت ایستای گراف، بهره‌گیری از مدل‌های شبکه پویا می‌تواند تغییرات علائم و مسیر تکامل بیماری‌ها را در طول زمان مدل‌سازی و پیش‌بینی کند. در نهایت، همکاری مستقیم با متخصصان بالینی برای اعتبارسنجی نتایج و یکپارچه‌سازی دانش تخصصی با یافته‌های داده‌محور ضروری است.

تشکر و قدردانی

این مقاله حاصل تحقیق مستقل بدون حمایت مالی و سازمانی است.

تضاد منافع

در انجام پژوهش حاضر، نویسندگان هیچ‌گونه تضاد منافی نداشتند.

همچنین، اعتبارسنجی خوشه‌ها به‌صورت کیفی و مبتنی بر دانش موجود انجام شده است؛ از این‌رو، اعتبارسنجی و ارزیابی کمی نتایج توسط پزشکان و متخصصان در محیط‌های واقعی درمانی می‌تواند مدنظر قرار گیرد.

نتیجه‌گیری

در این پژوهش با هدف تحلیل ساختاری روابط میان بیماری‌ها و علائم، از روش‌های تحلیل شبکه استفاده شد. با ساخت گراف دوبخشی بیماری-علائم و تبدیل آن به گراف‌های تک‌بخشی بیماری-بیماری و علائم-علائم، امکان بررسی دقیق‌تر ارتباطات میان بیماری‌ها و علائم فراهم شد. نتایج نشان داد که تحلیل شبکه می‌تواند به شناسایی بیماری‌هایی با علائم مشابه و ایجاد خوشه‌های معنادار میان آن‌ها کمک کند. این خوشه‌ها در گراف علائم-علائم، منطبق با گروه‌های بیماری شناخته‌شده بودند که به تشخیص افتراقی و درک بهتر هم‌وقوعی علائم در بیماری‌های مختلف یاری می‌رسانند.

ادغام مدل ارایه شده در سامانه‌های اطلاعات بیمارستانی، زمینه‌ای مناسب برای بهره‌گیری از داده‌های واقعی بیماران در فرایند تشخیص افتراقی فراهم می‌آورد. یافته‌های این پژوهش می‌تواند توسط پزشکان برای پشتیبانی از تصمیم‌گیری در تشخیص، توسط توسعه‌دهندگان سامانه‌های هوشمند سلامت برای طراحی سیستم‌های تصمیم‌یار و توسط سیاست‌گذاران سلامت برای بهبود کیفیت مراقبت و تخصیص بهینه منابع استفاده شود.

References

- Yang Z, Liu H, Yu H, Hu F. Multi-view disease-symptom association prediction based on disease knowledge graph embedding. In 2023 IEEE International Conference on Medical Artificial Intelligence (MedAI) 2023 Nov 18 (pp. 133-142). IEEE. Beijing, China.
- Pechsiri C, Piriyaikul R. Construction of disease-symptom knowledge graph from web-board documents. Applied Sciences. 2022 Jun 29;12(13):6615.
- Lamba D, Hsu WH, Alsadhan M. Predictive analytics and machine learning for medical informatics: A survey of tasks and techniques. In Machine learning, big data, and IoT for medical informatics 2021 Jan 1 (pp. 1-35). Academic Press.
- Liang M, Zhu S, Zhang W, Knobf MT, Ye Z. Symptom network approach for management in cancer care. Asia-Pacific Journal of Oncology Nursing. 2024 May 1;11(5).
- Kuang Y, Jing F, Sun Y, Zhu Z, Xing W. Symptom networks in older adults with cancer: a network analysis. Journal of Geriatric Oncology. 2024 Apr 1;15(3):101718.
- Finamore P, Spruit MA, Schols JM, Antonelli Incalzi R, Wouters EF, Janssen DJ. Clustering of patients with end-stage chronic diseases by symptoms: a new approach to identify health needs. Aging clinical and experimental research. 2021 Feb;33(2):407-17.
- Shirazi S, Albadvi A, Akhondzadeh E, Farzadfar F, Teimourpour B. A new application of community detection for identifying the real specialty of physicians. International journal of medical informatics. 2020 Aug 1;140:104161.
- Bacciu D, Errica F, Micheli A, Podda M. A gentle introduction to deep learning for graphs. Neural Networks. 2020 Sep 1;129:203-21.
- Hosseini M, Galavi E. Community Detection in Social Networks Using Deep Learning Approach. Business Intelligence Management Studies. 2023 Jun 22;11(44):83-112. [In Persian]
- Gottlieb A, Stein GY, Ruppert E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Molecular systems biology. 2011 Jun 7;7(1):496.
- Li JQ, Rong ZH, Chen X, Yan GY, You ZH. MCMDA: Matrix completion for MiRNA-disease association prediction. Oncotarget. 2017 Feb 3;8(13):21187.
- Zhang W, Yue X, Tang G, Wu W, Huang F, Zhang X. SFPEL-LPI: Sequence-based feature projection ensemble learning for predicting LncRNA-protein interactions. PLoS computational biology. 2018 Dec 11;14(12):e1006616.
- Xia E, Sun W, Mei J, Xu E, Wang K, Qin Y. Mining disease-symptom relation from massive biomedical literature and its application in severe disease diagnosis. In AMIA Annual Symposium Proceedings 2018 Dec 5 (Vol. 2018, p. 1118).
- Zhou X, Menche J, Barabási AL, Sharma A. Human symptoms-disease network. Nature communications. 2014 Jun 26;5(1):4212.
- Bojer CS, Meldgaard JP. Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting. 2021 Apr 1;37(2):587-603.

16. Pavlopoulos GA, Kontou PI, Pavlopoulou A, Bouyioukos C, Markou E, Bagos PG. Bipartite graphs in systems biology and medicine: a survey of methods and applications. *GigaScience*. 2018 Apr;7(4):giy014.
17. Rustamaji HC, Kusuma WA, Nurdianti S, Batubara I. Community detection with greedy modularity disassembly strategy. *Scientific Reports*. 2024 Feb 26;14(1):4694.
18. Zahiri M, Mohammadzadeh J, Harifi S. An improved Girvan-Newman community detection algorithm using trust-based centrality. *Journal of Ambient Intelligence and Humanized Computing*. 2023 Apr;14(4):3755-66.
19. Zhang J, Wang S, Wang Y, Zhuang J, Hang L, Wu Y, Xu D, Huang C. Symptom clusters and network analysis in patients with gynecologic cancer undergoing chemotherapy: A cross-sectional study. *Asia-Pacific Journal of Oncology Nursing*. 2024 Dec 1;11(12):100612.
20. Rha SY, Lee J. Stable symptom clusters and evolving symptom networks in relation to chemotherapy cycles. *Journal of pain and symptom management*. 2021 Mar 1;61(3):544-54.
21. Luo R, Chen H, Liu Y, Sun H, Tang S, Chen Y. Symptom clusters among breast cancer patients in relation to chemotherapy cycles: a longitudinal study. *Supportive Care in Cancer*. 2023 Oct;31(10):573.
22. Kim HJ, McGuire DB, Tulman L, Barsevick AM. Symptom clusters: concept analysis and clinical implications for cancer nursing. *Cancer nursing*. 2005 Jul 1;28(4):270-82.
23. Qi Y, Li H, Chan DN, Ma X, Wong CL. Effects of yoga interventions on the fatigue-pain-sleep disturbance symptom cluster in breast cancer patients: a systematic review. *European Journal of Oncology Nursing*. 2024 Jun 1;70:102594.

Modeling and Graph Analysis for Discovering Hidden Relationships between Diseases and Symptoms: A Data-Driven Approach to Differential Diagnosis and Symptom Co-occurrence

Roghayeh Khasha¹, Sepideh Ghaffari²

Original Article

Abstract

Introduction: Understanding hidden relationships between diseases and symptoms is one of the fundamental challenges in differential diagnosis and comorbidity analysis in medicine. Given the complexity of clinical interactions, identifying communal patterns among diseases can play a crucial role in improving the diagnostic process and clinical decision-making. This study proposes a data-driven, graph-based analytical framework to model and uncover hidden relationships between diseases and symptoms.

Methods: This study utilized the publicly available Disease-Symptom Prediction Dataset from the Kaggle platform, consisting of 4920 records covering 41 diseases and 131 unique symptoms. Using a linear combination of occurrence frequency indices and the kappa coefficient, a weighted bipartite disease-symptom graph was developed. Disease-disease and symptom-symptom unipartite graphs were then extracted, and their network structures were analyzed through the Louvain, Greedy Modularity, and Girvan-Newman clustering algorithms.

Results: The results indicated that fatigue and dengue fever played key roles in the disease-symptom network. Disease clustering using the Louvain and Greedy Modularity algorithms identified meaningful clusters of related diseases. Symptom clustering with the Louvain algorithm revealed six clinically interpretable clusters, which are useful for rapid disease identification. Moreover, seemingly unrelated symptoms can be associated with a common disease, while specific symptom clusters can guide the diagnosis of particular diseases.

Conclusion: The findings indicate that graph-based analysis is an effective tool for uncovering hidden relationships between diseases and symptoms and can play an important role in improving differential diagnosis and designing intelligent decision-support systems.

Keywords: Diagnosis, Differential; Disease; Symptom Co-occurrence; Medical Network Analysis, Bipartite Graph, Community Detection

Received: 12 Oct; 2025

Accepted: 18 Dec; 2025

Published: 31 Dec; 2025

Citation: Khasha R, Ghaffari, S. Modeling and Graph Analysis for Discovering Hidden Relationships between Diseases and Symptoms: A Data-Driven Approach to Differential Diagnosis and Symptom Co-occurrence. Health Inf Manage 2025; 22(4): 189-198.

This article resulted from a Msc thesis with No 23075138.

1. M.Sc. Student, Health Systems Engineering, Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran

2. Assistant Professor, Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran

Corresponding Author: Roghaye Khasha; Assistant Professor, Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran

Email: r.khasha@aut.ac.ir