مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره ششم شماره اول. ص ۳۲-۴۵، ۱۳۹۸

سیستم تصمیم‌یار پزشکی برای تشخیص بیماری دیابت و هپاتیت مبتنی بر ترکیب بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب

فرهاد سلیمانیان قره چپق[1،*]، سید کیوان موسوی[2]

• دریافت مقاله: ۱۳۹۷/۴/۲۷ • پذیرش مقاله: ۱۳۹۷/۷/۱۹

مقدمه: سیستم‌های تصمیم یار پزشکی در قالب یک برنامه کامپیوتری طراحی می‌شوند و به متخصصان پزشکی در اتخاذ تصمیمات تشخیص بیماری، کمک می‌کنند. هدف اصلی این گونه سیستم‌ها در واقع یاری رساندن به پزشکان در زمینه تشخیص بیماری می‌باشد، بدین‌معنی که یک پزشک می‌تواند با سیستم تعامل داشته باشد و در تحلیل داده‌های بیمار، تشخیص‌دهی و سایر فعالیت‌های پزشکی از سیستم کمک بگیرد.

روش: این مطالعه از نوع توصیفی–تحلیلی بود. مجموعه داده‌ها شامل ۷۶۸ رکورد دیابت با ۸ ویژگی و ۱۵۵ رکورد هپاتیت با ۱۹ ویژگی می‌باشند که از سایت جهانی UCI تهیه شده‌اند و از الگوریتم بهینه‌سازی اجتماع ذرات برای انتخاب ویژگی و از الگوریتم کرم شب‌تاب برای طبقه‌بندی بیماری دیابت و هپاتیت به دو کلاس سالم و ناسالم استفاده شد. از ۸۰ درصد داده‌ها جهت آموزش و از ۲۰ درصد باقی‌مانده جهت آزمون استفاده شد.

نتایج: بررسی اولیه نشان داد صحت الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای مجموعه داده دیابت به‌ترتیب برابر با ۸۴/۴۱ و ۸۲/۰۸ درصد و برای مجموعه داده هپاتیت به ترتیب برابر با ۸۱/۸۴ و ۸۰/۳۴ درصد به دست آمد. همچنین صحت مدل پیشنهادی برای مجموعه داده دیابت و هپاتیت به ترتیب برابر ۹۵/۳۸ و ۹۴/۰۹ درصد بود.

نتیجه‌گیری: بر اساس یافته‌های این مطالعه، مدل پیشنهادی در مقایسه با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب از نرخ خطای کمتری در تشخیص بیماری برخوردار بود. یافته‌های این پژوهش می‌تواند به پزشکان در تشخیص به موقع بیماری دیابت و هپاتیت کمک نماید.

کلید واژه‌ها: سیستم تصمیم یار پزشکی، تشخیص بیماری، بیماری دیابت، بیماری هپاتیت، الگوریتم بهینه‌سازی اجتماع ذرات، الگوریتم کرم شب‌تاب

• **ارجاع:** سلیمانیان قره چپق فرهاد، موسوی سید کیوان. سیستم تصمیم‌یار پزشکی برای تشخیص بیماری دیابت و هپاتیت مبتنی بر ترکیب بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۸؛ ۶(۱): ۴۵-۳۲.

۱. دکترای مهندسی کامپیوتر، استادیار، گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران

۲. دانشجوی دکترای مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران

* **نویسنده مسئول:** ارومیه، کیلومتر دو جاده فرودگاه، دانشگاه آزاد اسلامی ارومیه، دانشکده فنی و مهندسی

• **شماره تماس:** ۰۹۱۴۱۷۶۴۴۲۷

• **Email:** bonab.farhad@gmail.com

مقدمه

سیستم‌های تصمیم‌یار، با تحلیل اطلاعات و دانش جمع‌آوری شده طی سال‌ها در انبوهی از داده‌های مختلف بیماران، پزشکان را جهت انجام تصمیم‌گیری‌های دقیق و برنامه‌ریزی برای تشخیص بیماری‌ها یاری می‌دهند [۱]. سیستم‌های تصمیم‌یار خانواده‌ای از سیستم‌های اطلاعاتی شامل ابزارها و تکنیک‌ها جهت حمایت از پزشکان برای تصمیم‌گیری‌های دقیق درباره وضعیت بیماران هستند [۲]. مزایای سیستم‌های تصمیم‌یار پزشکی شامل فراهم کردن دسترسی آسان و سریع به داده‌های بیماران، توانایی ارائه گزارش‌های پیچیده و نگهداری حجم زیادی از داده‌ها می‌باشند [۳]. فراهم نمودن یک مراقبت جامع، مناسب و با کیفیت از بیماران، نیازمند گردآوری، تجزیه‌وتحلیل و ارائه داده‌های کلینیکی بسیار پیچیده است که با استفاده از داده‌های بیماران انجام می‌شود.

این سیستم‌ها با استفاده از اطلاعات و دانش پزشکی به تشخیص عوارض گوناگون و تجویز توصیه‌های پزشکی برای بیماران اقدام می‌نمایند [۴]. این سیستم‌ها برای جایگزینی پزشکان طراحی نشده‌اند و تنها جهت کمک‌رسانی به متخصصان علوم پزشکی در امر تشخیص بیماری‌ها بر اساس یکسری از قواعد تجربی ارائه شده‌اند. این سیستم‌ها که برای تشخیص بیماری‌ها و تجویز داروهای لازم به کار می‌روند، دارای قابلیت‌های متعددی نظیر دریافت اطلاعات، یادگیری، تفسیر داده‌ها و آموزش‌دهی می‌باشند [۴].

حجم بالایی از داده‌ها در پرونده‌های بیماران گردآوری و ذخیره شده است و تنها زمانی این داده‌ها مفید واقع می‌گردند که بررسی و آنالیز شوند و اطلاعات مفید در میان داده‌ها کشف شود و برای تشخیص و نوع بیماری استفاده شوند. لزوم سیستم تصمیم‌یار در پزشکی می‌تواند در کاهش انواع خطاهای پزشکی مؤثر باشد [۵]. اغلب مسائل در پزشکی مدرن بسیار پیچیده می‌باشند و دلایل منطقی برای تصمیم‌گیری‌های دقیق وجود ندارد. به همین دلیل تصمیم‌گیری‌های پزشکان معمولاً دلخواه و متغیر می‌باشد. از سوی دیگر حجم مفید اطلاعات پزشکی که حتی با یک محدوده کوچک تشخیصی در ارتباط است، به قدری زیاد می‌باشد که تصمیم‌گیری سریع و دقیق را دشوار خواهد ساخت. با توجه به این مسائل به آسانی قابل درک است که پزشکان با چالش عظیمی در فرآیند تشخیص پزشکی مواجه هستند که نیاز به پیدایش سیستم‌های تصمیم‌یار در پزشکی را آشکار خواهد نمود.

امکان وقوع خطاهای پزشکی در هر زمان و مکان وجود دارد و پیامد آن از آسیب‌های کوچک تا حوادث کشنده متفاوت است. مؤسسه پزشکی آمریکا در گزارش سال ۱۹۹۷ خود اعلام کرد، سالیانه بین ۴۴ تا ۹۸ هزار آمریکایی جان خود را به علت خطاهای پزشکی از دست می‌دهند، این در حالی است که از این رقم، ۷۰۰۰ مورد به اشتباهات دارویی تعلق می‌گیرد که این میزان در مقایسه با آمار مرگ‌ومیر ناشی از حوادث موتوری، سرطان سینه و ایدز هم بیشتر است. خطاهای پزشکی به عنوان یک مشکل جهانی می‌تواند منجر به افزایش مرگ‌ومیر بیماران، افزایش طول مدت بستری و افزایش هزینه‌های درمانی و همچنین سبب سلب اعتماد و نارضایتی بیماران از سیستم‌های ارائه خدمات درمانی و ایجاد استرس و تعارض‌های اخلاقی برای پرسنل درمانی و کاهش کیفیت مراقبت گردد [۶]؛ بنابراین برای بهبود بیمار، پزشکان مستلزم سیستم‌های تصمیم‌یار برای شناسایی وقایع تهدیدکننده و شناخت کامل فرآیند بیماری هستند.

الگوریتم‌های فرا ابتکاری، ابزارهایی بسیار مناسب برای تشخیص و پیش‌بینی می‌باشند که امروزه در زمینه‌های مختلفی کاربرد پیدا کرده‌اند [۷]. تشخیص و شناسایی بیماری‌های مختلف در علم پزشکی، یکی از زمینه‌های پرکاربرد الگوریتم‌های فرا ابتکاری محسوب می‌شود که در سال‌های اخیر تحقیقات و مطالعات زیادی پیرامون آن انجام شده است [۸،۹]. هدف این پژوهش این است که مدل ترکیبی بر مبنای الگوریتم بهینه‌سازی اجتماع ذرات [۱۰] و کرم شب‌تاب [۱۱] جهت تشخیص بیماری دیابت (۷۶۸ نمونه با ۸ ویژگی) [۱۲] و هپاتیت (۱۵۵ نمونه با ۱۹ ویژگی) [۱۳] با استفاده از الگوریتم‌های فرا ابتکاری پیشنهاد دهد. در این پژوهش یک سیستم تصمیم‌یار بر مبنای ترکیب الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای تشخیص بیماری دیابت [۱۴] و هپاتیت [۱۵،۱۶] پیشنهاد شد. در مدل پیشنهادی از الگوریتم بهینه‌سازی اجتماع ذرات برای انتخاب ویژگی و از الگوریتم کرم شب‌تاب برای طبقه‌بندی استفاده شد. انتخاب ویژگی‌های مؤثر و مهم به روند دقت تشخیص و همچنین به افزایش دقت طبقه‌بندی کمک می‌کنند. هدف مدل پیشنهادی افزایش دقت تشخیص بیماری دیابت و هپاتیت است.

در ادامه به برخی از الگوریتم‌های پیشنهاد شده برای تشخیص بیماری دیابت و هپاتیت اشاره شد. هر مدل از الگوریتم‌های متنوعی مانند داده‌کاوی یا هوش مصنوعی استفاده کرده‌اند. همچنین هر مدل مزایا و معایبی دارد. در مطالعه‌ای مدل

ترکیبی بر مبنای K-Means و رگرسیون لجستیک برای پیش‌بینی بیماری دیابت پیشنهاد شده است [۱۷]. نتایج بر روی ۷۶۸ نمونه بیمار نشان داده که مقدار صحت مدل ترکیبی برابر با ۹۵/۴۲ درصد است [۱۷]. محققین مدلی بر مبنای شبکه عصبی مصنوعی تابع شعاعی پایه برای پیش‌بینی بیمار دیابت پیش‌بینی کرده‌اند [۱۸]. نتایج بر روی ۷۶۸ نمونه بیمار نشان داده که صحت مدل ترکیبی برابر ۷۳/۹۱ است [۱۸]. جهانی و همکاران [۱۹] از شبکه عصبی مصنوعی به منظور پیش‌بینی ابتلا به دیابت در میان افراد مستعد دیابت استفاده کرده‌اند. این پژوهش از نوع کاربردی و جامعه هدف آن متشکل از ۵۴۵ فرد بیمار و سالم از مرکز دیابت دانشگاه علوم پزشکی همدان جمع‌آوری گردیده است. بررسی اولیه نشان داد که دقت شبکه عصبی، ۸۸ درصد، می‌باشد. بعد از به‌روزرسانی وزن‌ها با الگوریتم ممتیک دقت آن به ۹۳/۲ درصد افزایش یافت.

خسروانیان [۲۰] مدلی بر مبنای ترکیب روش رگرسیون بردار پشتیبان و الگوریتم کرم شب‌تاب جهت تشخیص بیماری دیابت پیشنهاد داده است. به منظور شبیه‌سازی و ارزیابی سیستم طراحی شده، از مجموعه داده مربوط به بیماران دیابتی هندی موجود در مجموعه داده یادگیری ماشین دانشگاه ایروین کالیفرنیا که شامل ۷۶۸ نمونه است، استفاده شد. درنهایت مقدار شاخص‌های حساسیت، اختصاصیت، صحت و دقت به ترتیب معادل ۸۴/۰۴، ۹۵/۲۶، ۹۱/۲۵، ۹۰/۸۰ درصد است. رافع و اربابی پژوهشی تحت عنوان «استفاده از تکنیک‌های داده‌کاوی جهت تشخیص دیابت با استفاده از چربی خون» برای تشخیص بیماری دیابت پیشنهاد داده‌اند؛ داده‌های موردنیاز این تحقیق از پایگاه داده یکی از آزمایشگاه‌های شهرستان نهاوند استخراج شده است که شامل داده‌های ۵۷۰۶ نمونه است. درصد صحت درخت تصمیم C4.5 برابر ۹۰/۰۲ درصد می‌باشد [۲۱].

عامری و همکاران مطالعه‌ای تحت عنوان «استخراج دانش از داده‌های بیماران دیابتی با استفاده از روش درخت تصمیم C5.0» برای تشخیص بیماری دیابت پیشنهاد کرده‌اند؛ تعداد نمونه‌های اولیه بیماران ۸۵۶ رکورد بوده است که پس از پالایش و حذف رکوردهایی کـه اطلاعات اصلی آن‌ها وجود نداشت به ۲۵۴ رکورد نهایی رسیده است. درصد صحت در درخت تصمیم C5.0 برابر ۸۹/۷۴ درصد و در شبکه عصبی مصنوعی برابر ۵۱/۲۸ درصد می‌باشد [۲۲]. عاملی و منتظر [۲۳]، مدلی برمبنای نظریه گواه برای ترکیب نتایج حاصل از شبکه عصبی مصنوعی و ماشین بردار پشتیبان پیشنهاد داده‌اند. داده‌های بیماری نیز بر اساس اطلاعات مربوط به ۲۱۲۵ بیمار مشکوک به دیابت است که از بیمارستان فیروزگر تهران گردآوری شده است. خروجی نهایی حاصل از ترکیب شبکه عصبی مصنوعی و ماشین بردار پشتیبان برابر با ۸۱/۷۵ درصد است. در پژوهشی دیگر با عنوان «آنالیز عملکرد تکنیک‌های تلفیقی داده‌کاوی جهت پیش‌بینی بیماری دیابت» برای تشخیص بیماری دیابت پیشنهاد کرده‌اند [۲۴]. روش پیشنهادی بر روی داده‌های دریافت شده از سیستم پزشکی دانشگاه کالیفرنیا که شامل ۷۶۸ نمونه می‌باشند اعمال شده است. نتایج ارزیابی‌ها نشان می‌دهد که به صورت کلی روش ترکیبی Bagging، با هر دو الگوریتم پایه J48 و Logistic دارای عملکرد بالاتری نسبت به الگوریتم آدابوست و الگوریتم‌های مستقل J48 و Logistic در تشخیص بیماری می‌باشد.

مدل RE-RX [۲۵] بر مبنای ترکیب شبکه عصبی مصنوعی چندلایه و درخت تصمیم C4.5 برای تشخیص بیماری هپاتیت پیشنهاد شده است. مقدار صحت در مدل ترکیبی برابر با ۸۳/۲۴ درصد است. تشخیص بیماری هپاتیت با استفاده از تئوری مجموعه راف و k نزدیک‌ترین همسایه انجام شده است [۲۶]. مقدار صحت مدل ترکیبی برابر با ۸۴/۵۲ است. بهروش و صادق زاده [۲۷] مطالعه‌ای تحت عنوان «تشخیص بیماری هپاتیت با رویکرد ترکیبی جستجوی فاخته و حداقل مربعات ماشین بردار پشتیبان» پیشنهاد کرده‌اند. در این پژوهش، از الگوریتم فاخته برای انتخاب ویژگی و طبقه‌بندی با استفاده از حداقل مربعات ماشین بردار پشتیبان انجام شده است. دقت به دست آمده از نتایج پیاده‌سازی بر روی ۱۵۵ نمونه برابر ۹۹/۱۸ درصد می‌باشد. خدارحمی و روحانی [۲۸] هم مطالعه‌ای با عنوان «تشخیص بیماری هپاتیت با استفاده از ترکیب الگوریتم K-Means و الگوریتم بهینه‌سازی علف‌های هرز» پیشنهاد داده‌اند. نتایج استفاده از خوشه‌بندی K-Means و الگوریتم بهینه‌سازی علف‌های هرز حاکی از دقت ۱۰۰ درصد در خوشه‌بندی داده‌های آزمایشی و ۹۸/۹ درصد در خوشه‌بندی داده‌های آموزشی است. محققین در مطالعه‌ای تحت عنوان «تشخیص بیماری هپاتیت با استفاده از تحلیل مؤلفه‌های اساسی و LS-SVM» برای تشخیص بیماری هپاتیت پیشنهاد کرده‌اند [۲۹]. در ابتدا با استفاده از تحلیل مؤلفه اساسی، فضای ویژگی کاهش داده شده و سپس از طبقه‌بند ماشین بردار پشتیبان حداقل مربعات برای طبقه‌بندی استفاده شده است که دقت مدل ترکیبی برابر با ۹۹/۷۵ است.

یکی از روش‌های مؤثر برای کمک به تشخیص بیماری‌ها، استفاده از شبکه عصبی مصنوعی و روش‌های داده‌کاوی است

[۳۰]. مدل‌های پیاده‌سازی شده با شبکه عصبی مصنوعی و ماشین بردار پشتیبان و شبکه عصبی مصنوعی خودسازمان‌ده به ترتیب دارای دقت تشخیص ۸۶/۹۸ درصد و ۱۰۰ درصد می‌باشند. ابراهیمی و همکاران [۳۱] از درخت تصمیم‌گیری، شبکه عصبی مصنوعی و ماشین بردار پشتیبان به‌منظور تشخیص بیماری هپاتیت استفاده کرده‌اند. نتایج به دست آمده نشان می‌دهد که شبکه عصبی مصنوعی دقت بالاتری نسبت به سایر الگوریتم‌ها دارد. با استفاده از شبکه عصبی مصنوعی دقت ۸۹/۷۴ به دست آمده است. برای تشخیص بیماری هپاتیت از ترکیب مجموعه راف و یادگیری ماشین، استفاده شده است [۳۲]. مدل ترکیبی شامل دو مرحله است. در مرحله اول، ویژگی‌های زائد از طریق مجموعه راف حذف می‌شوند و مرحله دوم، فرآیند دسته‌بندی با ویژگی‌های باقی‌مانده به وسیله یادگیری ماشین انجام می‌شود. نتایج ارزیابی بر روی ۱۵۵ نمونه نشان داده است که مدل ترکیبی با کاهش ۷۸/۹۵ درصد ویژگی‌ها، بالاترین دقت دسته‌بندی یعنی مقدار ۱۰۰ درصد را به دست آورده است.

با توجه به توسعه سیستم‌های تصمیم‌یار در حوزه پزشکی، در این مطالعه، هدف ارائه یک سیستم تصمیم‌یار برای تشخیص بیماری دیابت و هپاتیت برای یاری رساندن به پزشک برای تشخیص اقدامات لازم بعدی است. در مدل پیشنهادی از الگوریتم بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب استفاده شد. در ابتدا ویژگی‌ها با استفاده از الگوریتم بهینه‌سازی اجتماع ذرات انتخاب شدند و سپس از الگوریتم کرم شب‌تاب برای طبقه‌بندی افراد بیمار و سالم استفاده شد.

روش

این مطالعه از نوع توصیفی-تحلیلی است که بر اساس ویژگی‌های ورودی به پیشگویی وضعیت بیماران دیابت و هپاتیت از نظر سالم یا ناسالم بودن می‌پردازد. داده‌های مورد استفاده در این پژوهش از مجموعه داده مربوط به بیماران مبتلا به دیابت و هپاتیت، موجود در مجموعه داده یادگیری ماشین دانشگاه ایروین، کالیفرنیا تأمین شده است [۱۲،۱۳]. مجموعه‌داده‌ها شامل ۷۶۸ رکورد دیابت با ۸ ویژگی و ۱۵۵ رکورد هپاتیت با ۱۹ ویژگی می‌باشند. در نهایت مجموعه داده دیابت را می‌توان یک ماتریس ۷۶۸*۸ در نظر گرفت و مجموعه داده هپاتیت را می‌توان یک ماتریس ۱۵۵*۱۹ تلقی کرد. از مجموعه داده هپاتیت، ۲۰ درصد یعنی ۳۲ نفر ناسالم و ۸۰ درصد یعنی ۱۲۳ نفر سالم می‌باشند.

سیستم تصمیم‌یار تشخیص بیماری، شامل مراحل پیش پردازش داده‌ها، نرمال‌سازی داده‌ها و طبقه‌بندی می‌باشد. در مرحله طبقه‌بندی، نمونه‌ها در دو مرحله آموزش و آزمایش با استفاده از الگوریتم کرم شب‌تاب، ارزیابی شدند. در ابتدا توسط نمونه‌های آموزش مدل ساخته شد؛ یعنی نوع کلاس‌های سالم و بیمار جدا شدند و سپس در مرحله آزمایش با استفاده از نمونه‌های جدید، مدل ارزیابی شد و نمونه‌های آزمایش به کلاس‌هایی که مشخص شده‌اند، اختصاص داده شد. در شکل ۱، فلوچارت مدل پیشنهادی آورده شد.

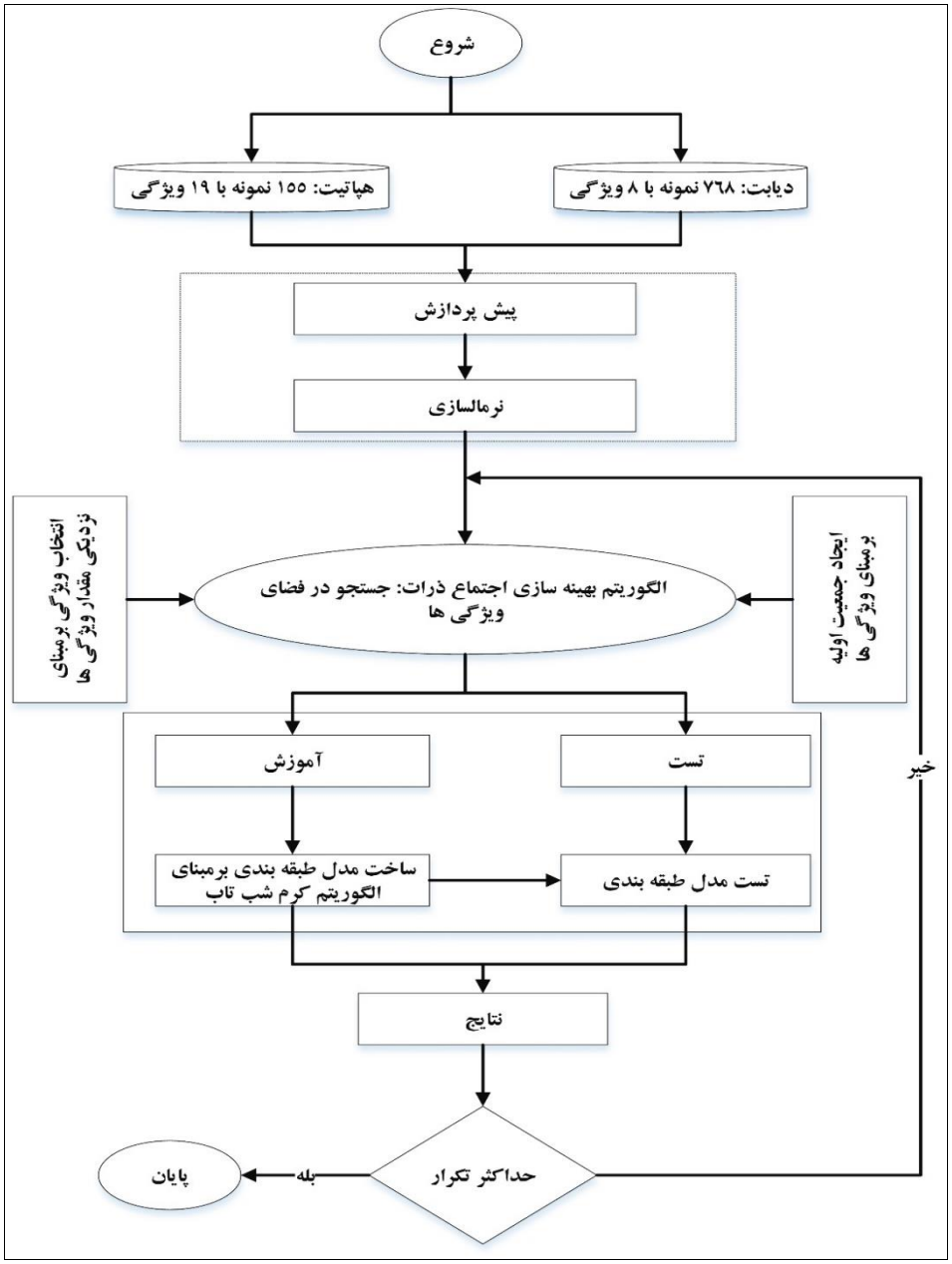
الف) پیش پردازش داده‌ها

ازآنجایی‌که که در مجموعه داده‌ها، ویژگی‌هایی با مقادیر مفقود وجود دارند. در این بخش، به عنوان فاز پیش‌پردازش، مقادیر این ویژگی‌های مفقودی با استفاده از روش مبتنی بر شباهت پیش‌بینی شده‌اند. در این روش، به ازای هر نمونه که دارای مقادیر مفقودی است با استفاده از فاصله اقلیدسی نزدیک‌ترین نمونه به آن پیدا شده و مقادیر نامعلوم این نمونه، با استفاده از نزدیک‌ترین نمونه پیدا شده مشخص گردید.

ب) نرمال‌سازی داده‌ها

اگر مقادیر ویژگی‌های مجموعه داده در دامنه متفاوتی قرار داشته باشند، احتمال بروز خطا در یافته‌ها افزایش می‌یابد. به قرار دادن داده‌های یک جامعه آماری در دامنه مشابه، نرمال‌سازی گفته می‌شود. ازآنجایی‌که ویژگی‌های مجموعه داده‌ها دارای مقادیر مختلفی هستند، به منظور یکسان‌سازی، مقدار ویژگی‌ها به فاصله بین ۰ و ۱ تبدیل می‌گردند. در فرآیند نرمال‌سازی، کوچک‌ترین عدد مربوط به هر ویژگی با min و بزرگ‌ترین عدد با max نشان داده می‌شود و x مقدار هر ویژگی می‌باشد. عمل نرمال‌سازی طبق معادله ۱، انجام شد [۳۳]. مقدار new_max و new_min به ترتیب برابر ۱ و ۰ می‌باشند.

$$y = \frac{x - \min}{\max - \min} * [new\_\max - new\_\min] + new\_\min \quad (۱)$$



شکل ۱: فلوچارت مدل پیشنهادی

ج) انتخاب ویژگی

در مدل پیشنهادی در مرحله اول، با استفاده از الگوریتم بهینه‌سازی اجتماع ذرات، جمعیت اولیه که همان مقدار ویژگی‌ها هستند را تشکیل داد. روش کار برای مجموعه داده دیابت بدین‌صورت است که الگوریتم بهینه‌سازی اجتماع ذرات از ۳۰ ذره و هر ذره نیز از ۸ ویژگی (۸ بُعد) تشکیل شده است. موقعیت هر ذره توسط این ۸ ویژگی تعریف شد. همچنین روش کار برای مجموعه داده هپاتیت بدین‌صورت است که الگوریتم بهینه‌سازی اجتماع ذرات از ۳۰ ذره و هر ذره نیز از ۱۹ ویژگی (۱۹ بعُد) تشکیل شده است. موقعیت هر ذره توسط این ۱۹ ویژگی تعریف شد.

در مرحله دوم، انتخاب ویژگی با استفاده از ذرات انجام گرفت. مجموعه داده‌هایی که دارای ابعاد (تعداد ویژگی) زیادی هستند، علیرغم مزایایی که دارند، چالش‌های محاسباتی زیادی را ایجاد می‌کنند. یکی از مشکلات داده‌های با ابعاد زیاد این است که در بیشتر مواقع تمام ویژگی‌های داده‌ها برای کشف قوانین لازم و ضروری نیستند. به همین دلیل در بسیاری از زمینه‌ها کاهش ابعاد داده یکی از مباحث مهم است. در اغلب موارد بسیاری از

ویژگی‌های کاندید برای کار یادگیری، نامربوط یا زائد هستند و کارایی به‌کارگیری الگوریتم یادگیری را بدتر خواهند کرد و زمان محاسبات را افزایش می‌دهند؛ به عبارتی دقت یادگیری و سرعت آموزش ممکن است به میزان چشمگیری با این ویژگی‌های زائد بدتر شود؛ بنابراین انتخاب ویژگی‌های مرتبط و ضروری در مرحله پیش‌پردازش از اهمیتی بنیادین برخوردار است. یک انتخاب ویژگی مناسب کارایی یک مدل یادگیری را افزایش می‌دهد. هدف پژوهش حاضر، انتخاب زیرمجموعه‌ای از ویژگی‌ها با استفاده از الگوریتم بهینه‌سازی اجتماع ذرات جهت استفاده از آن‌ها در طبقه‌بندی است؛ به‌عبارت‌دیگر هدف، پیدا کردن یک زیرمجموعه‌ای از ویژگی‌ها با حداقل تعداد ممکن است که برای تشخیص افراد دارای بیماری دقت بالایی داشته باشند؛ بنابراین از روش‌های مبتنی بر انتخاب ویژگی استفاده شد. در این پژوهش از روش انتخاب ویژگی بر مبنای الگوریتم بهینه‌سازی اجتماع ذرات استفاده شد.

در الگوریتم بهینه‌سازی اجتماع ذرات، ذرات در محیط پخش می‌شوند و ویژگی‌هایی که بهترین مقدار را در کل مجموعه داده‌ها دارند و طبقه‌بندی به آن‌ها وابسته است را پیدا می‌کنند. یک مسئله مهم در مدل پیشنهادی این است که الگوریتم بهینه‌سازی اجتماع ذرات برای انتخاب ویژگی باید از حالت پیوسته به حالت گسسته تبدیل شود. در حالت گسسته، حرکت ذرات به صفر و یک محدود می‌شود و مفهوم سرعت به مفهوم احتمال، تغییر یافته و v احتمال یک بودن x را بیان می‌کند. بدین‌معنی که مقدار v به یک مقدار بین [۰،۱] نگاشت شده که این مقدار بیانگر احتمال یک بودن x است. برای پیاده‌سازی رویکرد الگوریتم بهینه‌سازی اجتماع ذرات گسسته، ابتدا سرعت ذره در هر بعد محاسبه می‌شود. سپس، این مقدار با استفاده از تابع محدود کننده سیگموید طبق معادله ۲ به مقداری بین صفر و یک نگاشت می‌شود و در نهایت موقعیت ذره iام در بعد dام طبق معادله ۳ به‌روزرسانی شد.

$S(v_i) = sigmoid(v_i) = \frac{1}{1+e^{-v_i}}$	(۲)
$x_{i+1} = \begin{cases} 0 & if\ rand() \geq S(v_i) \\ 1 & if\ rand() < S(v_i) \end{cases}$	(۳)

هر ذره در الگوریتم بهینه‌سازی اجتماع ذرات از سه بردار d بُعدی تشکیل شده است که d بُعد فضای جستجو می‌باشد. برای ذره iام، این سه بردار عبارت‌اند از: موقعیت فعلی ذره (x_i)، سرعت حرکت ذره (v_i) و بهترین موقعیتی ($p_{best,i}$) که ذره تا به حال تجربه کرده است (x_i). به عنوان یک جواب برای مسئله محاسبه گردید. اگر این موقعیت بهتر از جواب‌های قبلی باشد در ($p_{best,i}$) ذخیره می‌شود. در هر تکرار (x_i) و (v_i) جدیدی به دست می‌آیند و منظور از اجرای الگوریتم، بهتر کردن ($p_{best,i}$)، برای رسیدن به بهترین پاسخ است. درنتیجه بهترین موقعیتی که به وسیله همه ذرات پیدا شده است به صورت (g_{best}) نشان داده شد. در مرحله ابتدایی الگوریتم، ذرات با موقعیت‌ها و سرعت‌های تصادفی ایجاد شدند. در طی اجرای الگوریتم، موقعیت و سرعت هر ذره در مرحله $t+1$ام از الگوریتم، از روی اطلاعات مرحله قبلی ساخته شدند. بهترین میزان تناسب هر ذره، pbest و بهترین میزان تناسب در میان گروه ذرات، gbest نامیده شد. این فرآیند تا زمانی که معیار حداکثر تکرار برآورده نشود، ادامه خواهد داشت.

د) طبقه‌بندی

نزدیکی دو نمونه، با به دست آوردن تشابه و یا فاصله میان این دو نمونه محاسبه شد. هر نمونه می‌تواند از انواع داده‌ها تشکیل شده باشد که باید تشابه میان آن‌ها بررسی شود. به منظور طبقه‌بندی نمونه‌ها از معیار فاصله طبق الگوریتم کرم شب‌تاب استفاده شد. برای این که طبقه‌بندی نمونه‌ها با دقت بالاتری انجام گیرد، معیار فاصله بین کرم‌های شب‌تاب تغییر داده شد. بدین‌صورت که از بهترین ذره الگوریتم بهینه‌سازی اجتماع ذرات برای به دست آوردن فاصله بین ویژگی‌ها استفاده شد. در معادله ۴، gbest بهترین ذره، d ابعاد ویژگی‌ها و x موقعیت ذرات در kام بعد(ابعاد) می‌باشد. مقدار هر ذره با مقدار ویژگی‌های همسایه مقایسه شد و نزدیک‌ترین ذرات به عنوان نمونه جدید به کلاس‌ها تخصیص داده شد.

$$r_{ij} = \left\| gbest_i - x_j \right\| = \sqrt{\sum_{k=1}^{d} (gbest_{i,k} - x_{j,k})^2} \quad (۴)$$

بعد از این که تشخیص نمونه‌ها برای دو کلاس سالم و بیمار توسط الگوریتم بهینه‌سازی اجتماع ذرات انجام شد بر مبنای بهترین ذره، بهترین ویژگی‌های نزدیک به هم انتخاب شدند؛ الگوریتم کرم شب‌تاب عمل طبقه‌بندی بین نمونه‌ها را بر مبنای فاصله انجام داد و در واقع نمونه‌ها را بدون احتمال به کلاس‌های درست واگذار نمود.

ه) ارزیابی مدل

در این پژوهش، جهت ارزیابی کارایی طبقه‌بندی از پنج معیار دقت، بازخوانی، اندازه‌گیری، صحت و نرخ خطا استفاده شد. دقت نشان دهنده تعداد نمونه‌های درست به تعداد کل نمونه‌ها است. معیار بازخوانی نشان داد که اگر نمونه‌ای، نوع A تشخیص داده شد با چه احتمالی نوع A است. دقت و بازخوانی طبق معادله ۵ و ۶ محاسبه شد [۱۷،۱۸].

شماره	توضیحات	معادلات
(۵)	دقت	$\text{Precision} = \frac{TP}{TP + FP}$
(۶)	بازخوانی	$\text{Recall} = \frac{TP}{TP + FN}$
(۷)	F-Measure	$F - \text{Measure} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})}$
(۸)	صحت	$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$
(۹)	نرخ خطا	$\text{ErrorRate} = 1 - \text{Accuracy}$

پارامترهای درست مثبت (TP)، درست منفی (TN)، کاذب مثبت (FP)، کاذب منفی (FN) از پارمترهای اصلی برای معیارهای ارزیابی هستند. پارامتر درست مثبت، بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی مثبت تشخیص داده است. پارامتر درست منفی، بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است. پارامتر کاذب مثبت، بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و الگوریتم دسته‌بندی دسته آن‌ها را به اشتباه مثبت تشخیص داده است. پارامتر کاذب منفی، بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و الگوریتم دسته‌بندی دسته آن‌ها را به‌اشتباه منفی تشخیص داده است.

نتایج

در این بخش ارزیابی و نتایج مدل پیشنهادی بر روی دو مجموعه داده دیابت و هپاتیت انجام شد. مجموعه داده‌های دیابت و هپاتیت برگرفته از داده‌های دانشگاه کالیفرنیا در ارواین (University of California, Irvine) مشهور به UCI می‌باشند. مجموعه داده دیابت شامل ۷۶۸ نمونه و ۸ ویژگی است. مجموعه داده هپاتیت شامل ۱۵۵ نمونه ۱۹ ویژگی است. پیاده‌سازی داده‌ها در محیط Matlab نسخه ۲۰۱۷ انجام شد. تعداد تکرارها، مهم‌ترین فاکتور در ارزیابی مدل پیشنهادی بود. جدول ۱، نتایج الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب را با ۵۰۰ بار تکرار نشان داد. نتایج نشان داد که مقدار صحت در الگوریتم بهینه‌سازی اجتماع ذرات بیشتر است. الگوریتم بهینه‌سازی اجتماع ذرات از روش یادگیری و استفاده از دانش بهترین ذره بهره می‌گیرد. هماهنگی و تنظیم ضرایب وزنی الگوریتم بهینه‌سازی اجتماع ذرات یکی از موضوعاتی است که نیازمند مقدار دقیق و مناسب برای آن‌ها است. در مورد تمامی مسائل بررسی شده با استفاده از الگوریتم بهینه‌سازی اجتماع ذرات، ضرایب وزنی از مقادیر ۰/۵، ۱/۵، ۱/۵ برای w و c1 و c2 استفاده می‌کنند.

این جدول نتایج الگوریتم‌ها را بر روی مجموعه داده دیابت و هپاتیت نشان داد. صحت بر روی مجموعه داده دیابت در

الگوریتم بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب به ترتیب برابر با ۸۴/۴۱ و ۸۲/۰۸ درصد است. همچنین صحت بر روی مجموعه داده هپاتیت در الگوریتم بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب به ترتیب برابر با ۸۱/۸۴ و ۸۰/۳۴ درصد است. نتایج بر روی دو مجموعه داده نشان داد که نرخ خطا در الگوریتم بهینه‌سازی اجتماع ذرات کمتر است.

جدول ۱: نتایج الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب

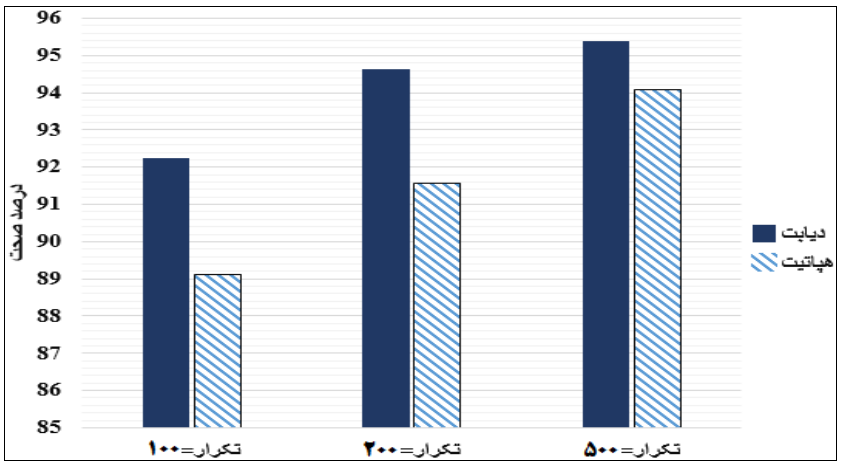
مجموعه داده	دیابت		هپاتیت	
معیارها	بهینه‌سازی اجتماع ذرات	الگوریتم کرم شب‌تاب	بهینه‌سازی اجتماع ذرات	الگوریتم کرم شب‌تاب
دقت	۸۵/۳۴	۸۴/۵۲	۸۲/۱۵	۸۱/۴۹
بازخوانی	۸۶/۹۱	۸۴/۷۹	۸۳/۰۶	۸۲/۵۷
F-Measure	۸۶/۱۲	۸۴/۶۵	۸۲/۶۰	۸۲/۰۳
صحت	۸۴/۴۱	۸۲/۰۸	۸۱/۸۴	۸۰/۳۴
نرخ خطا	۱۵/۵۹	۱۷/۹۲	۱۸/۱۶	۱۹/۶۶

در جدول ۲ نتایج مدل پیشنهادی بر مبنای تکرارهای مختلف نشان داده شد. بیشترین درصد صحت متعلق به ۵۰۰ تکرار است. درصد صحت در ۵۰۰ تکرار برای بیماری دیابت و هپاتیت به ترتیب برابر با ۹۵/۳۸ و ۹۴/۰۹ درصد است. همچنین درصد دقت و بازخوانی برای مجموعه داده دیابت به ترتیب برابر با ۹۴/۱۹ و ۹۵/۲۲ درصد است. و برای مجموعه داده هپاتیت به ترتیب برابر با ۹۳/۸۶ و ۹۴/۰۵ درصد است.

جدول ۲: نتایج مدل پیشنهادی بر مبنای تکرارهای مختلف

مجموعه داده	تعداد تکرار	دقت	بازخوانی	F-Measure	صحت	نرخ خطا
دیابت	۱۰۰	۹۱/۶۲	۹۲/۰۱	۹۱/۸۱	۹۲/۲۳	۷/۷۷
	۲۰۰	۹۳/۲۷	۹۳/۹۲	۹۳/۵۹	۹۴/۶۲	۵/۳۸
	۵۰۰	۹۴/۱۹	۹۵/۲۲	۹۴/۷۰	۹۵/۳۸	۴/۶۲
هپاتیت	۱۰۰	۸۹/۰۳	۸۹/۴۶	۸۹/۲۴	۸۹/۱۲	۱۰/۸۸
	۲۰۰	۹۱/۱۴	۹۱/۳۵	۹۱/۲۴	۹۱/۵۶	۸/۴۴
	۵۰۰	۹۳/۸۶	۹۴/۰۵	۹۳/۹۵	۹۴/۰۹	۵/۹۱

در شکل ۲ نمودار مقایسه درصد صحت مدل پیشنهادی بر مبنای تعداد تکرار برای مجموعه داده‌های دیابت و هپاتیت نشان داده شد. همان‌طور که مشاهده شد تعداد تکرار با درصد صحت رابطه مستقیم دارد. یعنی اگر تعداد تکرار بیشتر باشد درصد صحت بیشتر خواهد بود. در شکل ۲، محور افقی نمودار، بیانگر تعداد تکرار و محور عمودی بیانگر درصد صحت است. نمودارها برای حالت‌های ۱۰۰، ۲۰۰ و ۵۰۰ بار تکرار ترسیم شده‌اند. بیشترین درصد صحت در حالت ۵۰۰ بار تکرار است.



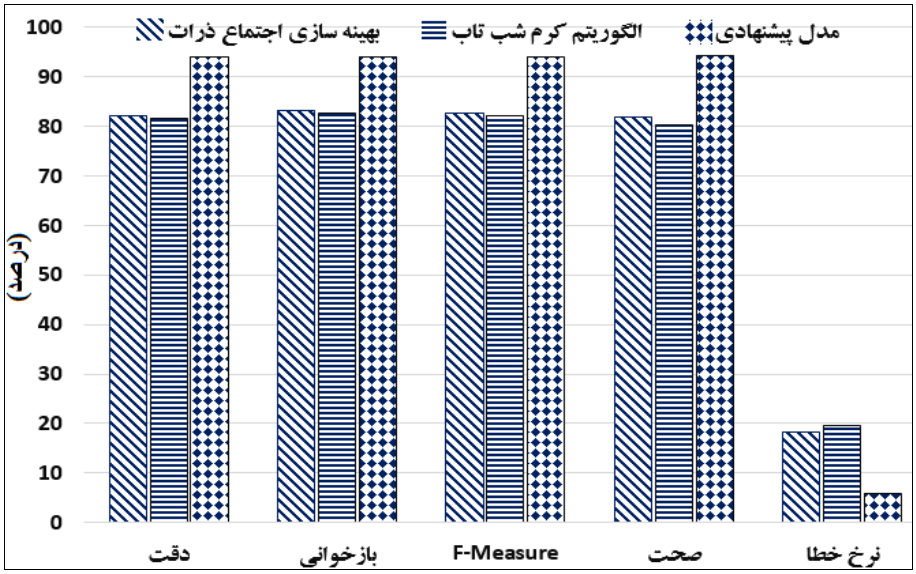
شکل ۲: مقایسه مقدار صحت بر مبنای تعداد تکرار

نتایج نشان داد که مدل پیشنهادی در مقایسه با الگوریتم بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب، مقدار صحت بیشتری دارد. همچنین شکل ۲، نشان داد که درصد صحت مدل پیشنهادی در ۵۰۰ تکرار در مقایسه با حالت‌های ۱۰۰ و ۲۰۰ تکرار، بیشتر است.

در شکل ۳، نمودار مقایسه مدل پیشنهادی با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای مجموعه داده دیابت نشان داده شد. طبق این نمودار روشن است که دقت تشخیص مدل پیشنهادی در مقایسه با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب بیشتر است و همچنین نرخ خطای کمتری دارد. نمودار شکل ۳، طبق نتایج جدول۱ و ۲ بر مبنای ۵۰۰ بار تکرار ترسیم شد. در این نمودار، نرخ همه معیارها نشان داده شد.

شکل ۳: نمودار مقایسه مدل پیشنهادی با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای مجموعه داده دیابت

در شکل ۴، نمودار مقایسه مدل پیشنهادی با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای مجموعه داده هپاتیت نشان داده شد. طبق نمودار شکل ۴، روشن است که دقت تشخیص مدل پیشنهادی در مقایسه با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب بیشتر است و همچنین نرخ خطای کمتری دارد. نمودار شکل ۴، طبق نتایج جدول ۱ و ۲ بر مبنای ۵۰۰ بار تکرار ترسیم شد.



شکل ٤: نمودار مقایسه مدل پیشنهادی با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای مجموعه داده هپاتیت

در جدول ۳، نتایج مدل پیشنهادی بر مبنای تعداد ویژگی‌ها و ۵۰۰ بار تکرار نشان داده شد. برای مجموعه داده دیابت مقدار صحت برای ۴ ، ۵ ، ۶ و ۷ ویژگی به ترتیب برابر با ۹۸/۲۵، ۹۸/۰۶، ۹۷/۴۲ و ۹۶/۱۳ درصد می‌باشد. همچنین برای مجموعه داده هپاتیت مقدار صحت برای ۶، ۸ ، ۱۰، ۱۲ ، ۱۴، ۱۶ و ۱۷ ویژگی به ترتیب برابر با ۹۸/۳۶، ۹۷/۶۵، ۹۷/۱۸، ۹۶/۵۹، ۹۶/۱۴، ۹۵/۳۷ و ۹۵/۱۱ درصد می‌باشد. بیشترین درصد صحت برای مجموعه داده دیابت و هپاتیت به ترتیب برابر با ۹۸/۲۵ و ۹۸/۳۶ درصد است.

جدول ۳، نشان داد که بیشترین درصد معیارهای دقت و بازخوانی متعلق به تعداد ویژگی‌های کمتر است. بیشترین درصد دقت برای مجموعه داده دیابت و هپاتیت به ترتیب برابر با ۹۸/۳۳ و ۹۷/۶۴ درصد است. همچنین بیشترین درصد بازخوانی برای مجموعه داده دیابت و هپاتیت به ترتیب برابر با ۹۸/۶۲ و ۹۸/۱۷ درصد است.

جدول ۳: نتایج مدل پیشنهادی بر مبنای تعداد ویژگی و ۵۰۰ بار تکرار

مجموعه داده	تعداد ویژگی	دقت	بازخوانی	F-Measure	درصد صحت	نرخ خطا
دیابت	۴	۹۸/۳۳	۹۸/۶۲	۹۸/۴۷	۹۸/۲۵	۱/۷۵
	۵	۹۷/۸۶	۹۸/۱۳	۹۷/۹۹	۹۸/۰۶	۱/۹۴
	۶	۹۷/۲۱	۹۷/۸۳	۹۷/۵۲	۹۷/۴۲	۲/۵۸
	۷	۹۶/۹۰	۹۷/۱۵	۹۷/۰۲	۹۶/۱۳	۳/۸۷
هپاتیت	۶	۹۷/۶۴	۹۸/۱۷	۹۷/۹۰	۹۸/۳۶	۱/۶۴
	۸	۹۷/۲۶	۹۷/۳۵	۹۷/۳۰	۹۷/۶۵	۲/۳۵
	۱۰	۹۷/۱۲	۹۷/۷۳	۹۷/۴۲	۹۷/۱۸	۲/۸۲
	۱۲	۹۶/۵۷	۹۶/۸۲	۹۶/۶۹	۹۶/۵۹	۳/۴۱
	۱۴	۹۶/۰۸	۹۶/۳۱	۹۶/۱۹	۹۶/۱۴	۳/۸۶
	۱۶	۹۵/۵۲	۹۵/۷۹	۹۵/۶۵	۹۵/۳۷	۴/۶۳
	۱۷	۹۵/۳۱	۹۶/۱۶	۹۵/۷۳	۹۵/۱۱	۴/۸۹

بحث و نتیجه‌گیری

در این پژوهش یک مدل ترکیبی بر مبنای الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای تشخیص بیماری دیابت و هپاتیت پیشنهاد گردید. در مدل پیشنهادی از الگوریتم بهینه‌سازی اجتماع ذرات برای انتخاب ویژگی و از الگوریتم کرم شب‌تاب برای طبقه‌بندی استفاده شد. انتخاب ویژگی‌های مؤثر و مهم به روند دقت تشخیص و همچنین به افزایش دقت طبقه‌بندی کمک می‌کنند.

شناخت صحیح و دقیق پزشکان و متخصصین از چگونگی تشخیص و نحوه درمان به موقع بیماری هپاتیت، می‌تواند در جلوگیری از پیشرفت بیماری و کاهش میزان هزینه‌ها، نقش اساسی بازی نماید. سیستم تصمیم‌یار، یک سیستم اطلاعاتی برای کمک به تشخیص‌های پزشکی و راهنمایی متخصصین سلامت در تصمیم‌گیری‌ها جهت مراقبت از بیمار است. با استفاده از سیستم‌های تصمیم‌یار، امکان تبدیل اطلاعات جمع‌آوری شده از افراد به دانش و استفاده از این اطلاعات در زمان مناسب میسر می‌شود. این دانش در قالب مدل تصمیم‌گیری و قوانین می‌تواند به شناسایی و تشخیص درست بیماری کمک کند و شرایط درمان به موقع بیماران را فراهم نماید. سیستم‌های تصمیم‌یار، تعداد زیادی از محاسبات را به سرعت انجام می‌دهند؛ بنابراین این سیستم‌ها می‌توانند برای حل مسائل پیچیده، وضعیت‌های اضطراری که اغلب نیاز به پاسخ‌های سریع دارند مناسب باشند. سیستم‌های تصمیم‌یار می‌توانند با تصمیم‌گیری درست، داده‌ها را پردازش، تجزیه‌وتحلیل کنند. پژوهش‌های گوناگون نقش مثبت این سیستم‌ها را در افزایش دقت و بهبود تصمیم‌گیری پزشکان نشان داده‌اند. ترکیب سیستم‌های تصمیم‌گیری با هوش مصنوعی اثرات مفیدی در سیستم‌های پزشک‌یار ایجاد می‌کند و باعث افزایش دقت در تشخیص نوع بیماری و بهبود درمان بیماران می‌گردد.

با گسترش الگوریتم‌های هوش مصنوعی، سیستم‌های تصمیم‌یار مبتنی بر کامپیوتر اهمیت بسیار زیادی یافته‌اند. سیستم‌های تصمیم‌یار، سیستم‌هایی مبتنی بر دانش می‌باشند و دانش در واقع مهم‌ترین بخش آن‌ها است. در این سیستم‌ها دانش از پزشکان به کامپیوتر منتقل می‌شود. در این پژوهش

یک سیستم تصمیم‌یار با استفاده از ترکیب الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب برای تشخیص بیماری دیابت و هپاتیت پیشنهاد شد. هدف سیستم تصمیم‌یار در این مطالعه، طبقه‌بندی نمونه‌های بیماران به دو کلاس سالم و ناسالم است. ارزیابی بر روی دو مجموعه داده دیابت و هپاتیت انجام گرفت. نتایج نشان داد که مدل پیشنهادی در مقایسه با الگوریتم‌های بهینه‌سازی اجتماع ذرات و کرم شب‌تاب، درصد صحت بیشتری دارد و نرخ خطای آن کمتر است. همچنین درصد صحت مدل پیشنهادی برای مجموعه داده دیابت و هپاتیت به ترتیب با ۵۰۰ بار تکرار برابر با ۹۵/۳۸ درصد و ۹۴/۰۹ درصد بود. مدل پیشنهادی بر مبنای انتخاب ویژگی، ارزیابی شد. نتایج نشان داد که اگر تعداد ویژگی‌ها کمتر باشند درصد صحت نیز بیشتر می‌باشد. درصد صحت مدل پیشنهادی با ۵۰۰ بار تکرار و با ۴ ویژگی برای مجموعه داده دیابت و ۶ ویژگی برای مجموعه داده هپاتیت به ترتیب برابر با ۹۸/۲۵ و ۹۸/۳۶ درصد به دست آمد.

Wu و همکاران [۱۷] با استفاده از K-Means و رگرسیون لجستیک بر روی مجموعه داده دیابت توانستند به صحت ۹۵/۴۲ درصد دست یابند. که دقت تشخیص مدل پیشنهادی برابر با ۹۵/۳۸ درصد می‌باشد. Edla و Cheruku [۱۸] با استفاده از شبکه عصبی مصنوعی تابع شعاعی بر روی مجموعه داده دیابت توانستند به صحت ۷۳/۹۱ درصد برسند. شبکه عصبی مصنوعی در مقایسه با مدل پیشنهادی برای تشخیص بیماری دیابت درصد صحت کمتری دارد. خسروانیان [۲۰] با استفاده از ترکیب روش رگرسیون بردار پشتیبان و الگوریتم کرم شب‌تاب جهت تشخیص بیماری دیابت به شاخص‌های حساسیت، اختصاصیت، صحت و دقت به ترتیب معادل ۸۴/۰۴، ۹۵/۲۶، ۹۱/۲۵، ۹۰/۸۰ درصد دست یافتند. مدل پیشنهادی توانسته است بیماری دیابت را با ۴ ویژگی و ۵۰۰ بار تکرار با دقت ۹۸/۲۵ درصد تشخیص دهد.

Femina و Anto [۲۶] برای تشخیص بیماری هپاتیت از تئوری مجموعه راف و k نزدیک‌ترین همسایه استفاده کرده‌اند که درصد صحت مدل آن‌ها برابر با ۸۴/۵۲ است؛ ولی دقت تشخیص مدل پیشنهادی برای تشخیص بیماری هپاتیت برابر با ۹۴/۰۹ درصد می‌باشد. که مدل پیشنهادی در مقایسه با ترکیب مجموعه راف و k نزدیک‌ترین همسایه، درصد صحت را تا حدودی بالای ۹ درصد افزایش داده است. بهروش و صادق‌زاده [۲۷] از الگوریتم فاخته برای انتخاب ویژگی و از الگوریتم حداقل مربعات ماشین بردار پشتیبان برای طبقه‌بندی استفاده کرده‌اند. و نتایج پیاده‌سازی آن‌ها برابر ۹۹/۱۸ درصد می‌باشد. درصد صحت مدل پیشنهادی در مقایسه با ترکیب الگوریتم فاخته و حداقل مربعات ماشین بردار پشتیبان، کمتر است. مدل‌های پیاده‌سازی شده [۳۰] برای تشخیص بیماری هپاتیت با شبکه عصبی مصنوعی و ترکیب ماشین بردار پشتیبان و شبکه عصبی مصنوعی خودسازمان‌ده دارای ماکزیمم دقت به ترتیب ۸۶/۹۸ و ۱۰۰ درصد می‌باشند. نتایج پژوهش ابراهیمی و همکاران [۳۱] نشان داد که شبکه عصبی مصنوعی دقت بالاتری نسبت به ماشین بردار پشتیبان دارد. صحت به دست آمده از شبکه عصبی مصنوعی برابر با ۸۹/۷۴ درصد است. مدل پیشنهادی نیز توانسته است بیماری هپاتیت را با ۶ ویژگی و ۵۰۰ بار تکرار با دقت ۹۸/۳۶ درصد تشخیص دهد. همچنین درصد صحت مدل پیشنهادی بر روی مجموعه داده هپاتیت برای کل ویژگی‌ها برابر با ۹۴/۰۹ است.

توجه بیشتر به ویژگی‌های بیماران دیابت و هپاتیت و تشخیص دقیق‌تر آن‌ها حتی به اندازه یک درصد باعث خواهد شد که شانس بقاء بیمار افزایش یافته و روند دوره درمان با سرعت بیشتری طی شود؛ لذا در این مطالعه سیستم تصمیم‌یار، با استفاده از ترکیب الگوریتم بهینه‌سازی اجتماع ذرات و الگوریتم کرم شب‌تاب برای تشخیص بیماری دیابت و هپاتیت پیشنهاد شد. مهم‌ترین چالش در تشخیص بیماری، انتخاب ویژگی‌ها است. احتمالاً اساسی‌ترین کاستی در مدل پیشنهادی، عدم انتخاب دقیق ویژگی‌ها باشد. چون مدل پیشنهادی به مقدار صددرصد دست نیافته است. برای مطالعات آینده و بهبود مدل پیشنهادی، موارد زیر پیشنهاد می‌شود:

- استفاده از دیگر روش‌های یادگیری ماشین، از جمله روش‌های مبتنی بر خوشه‌بندی و یا ماشین بردار پشتیبان و الگوریتم k نزدیک‌ترین همسایه.
- استفاده از تکنیک‌های فازی برای انتخاب ویژگی و طبقه‌بندی نمونه‌ها

References

1. Nazari Moghaddam S, Fallah M, Kazemipoor H, Salehipour A, A fuzzy inference- fuzzy analytic hierarchy process-based clinical decision support system for diagnosis of heart diseases. Expert Systems with Applications 2018; 95: 261-71.

2. Malmir B, Amini M, Chang SI. A medical decision support system for disease diagnosis under uncertainty. Expert Systems with Applications 2017;88:95-108.

3. Sudha M. Evolutionary and Neural Computing Based Decision Support System for Disease Diagnosis from Clinical Data Sets in Medical Practice. J Med Syst 2017;41(11):178.

4. Peker M. A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. J Med Syst 2016;40(5):116.

5. Anooj PK. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. Journal of King Saud University - Computer and Information Sciences 2012;24(1):27-40.

6. Ani R, Jose J, Wilson M, Deepa OS. Modified rotation forest ensemble classifier for medical diagnosis in decision support systems. Advances in Intelligent Systems and Computing 2018; 564: 137-46.

7. Fu YW, Chen HL, Chen SJ, Shen LM, Li QQ. A New Evolutionary Support Vector Machine with Application to Parkinson's disease Diagnosis, International Conference in Swarm Intelligence; 2014; Cham: Advances in Swarm Intelligence; 2014. p. 42-9.

8. Jabbar MA, Deekshatulu BL, Chandra P. An evolutionary algorithm for heart disease prediction. International Conference on Information Processing. ICIP 2012: Wireless Networks and Computational Intelligence; Berlin, Heidelberg: Springer; 2013; p. 378-89.

9. Liang C, Peng L. An automated diagnosis system of liver disease using artificial immune and genetic algorithms. J Med Syst 2013;37(2):9932.

10. Kennedy J, Eberhart R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks; 1995 Nov- Dec 27-1; Perth, WA, Australia, Australia: IEEE; 1995. p. 1942-8.

11. Yang XS. Nature-Inspired Metaheuristic Algorithms. Algorithms. United Kingdom: Luniver Press ; 2008

12. Diabetes data set. [cited 2014 Apr 4]. Available from: https://github.com/renatopp/arff-datasets/blob/master/classification/diabetes.arff

13. Hepatits data set. [cited 2014 Apr 4]. Available from: https://github.com/renatopp/arff-datasets/blob/master/classification/hepatitis.arff

14. Mansourypoor F, Asadi S. Development of a reinforcement learning-based evolutionary fuzzy rule-based system for diabetes diagnosis. Comput Biol Med 2017;91:337-352.

15. El Houby EM. A survey on applying machine learning techniques for management of diseases. Journal of Applied Biomedicine 2018;16(3):165-74.

16. Keerthika U, Sethukkarasi R, Kannan A. A novel rule based classifier for mining temporal medical databases using fuzzy rough set approach. Berlin, Heidelberg: Advances in Computing and Information Technology; 2013; p. 85-95

17. Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. Informatics in Medicine Unlocked 2018;10:100-7.

18. Edla DR, Cheruku R. Diabetes-Finder: a bat optimized classification system for type-2 diabetes. Procedia Computer Science 2017;115:235-42.

19. Jahani M, Rezaeenour J, Mahdavi M, Hadavandi I. Proposing a model for predicting diabetes patients by neural network. Journal of Health Administration 2017; 20(67):24-35. Persian

20. Khosrowanian A. A New Method by Combining the Firefly Algorithm and the Support Vector Regression in the Diagnosis of Diabetes Disease. 2nd Conference Meta-Heuristic Algorithms and Applications in Science and Engineering; 2017 May 11; Najafabad: Payame Noor University Branch Najaf Abad; 2017.

21. Rafeh R, Arbabi M. Data mining techniques to diagnose diabetes using blood lipids. Journal of Ilam University of Medical Sciences 2015;23(4):239-47. Persian

22. Ameri H, Alizadeh S, Barzegari A. Knowledge extraction of diabetics' data by decision tree method. Journal of Health Administration 2013;16(53):58-72. Persian

23. Ameli S, Montazer GA. Improving Diagnosis Diagnosis Diagnosis by Using Artificial Neural Network and Backup Machines. 2nd National Conference on Soft Computing; 2017 Oct 7; Gilan-Rudsar, University of Guilan; 2017.

24. Kashmari S, Elmi Sola Y. Analysis of the Performance of Data Mining Combined Techniques for Diabetes Disease Prevention. 2nd National Conference on Interdisciplinary Research in Computer Engineering Electrical. Mechanical and Mechatronics Engineering; 2017 Sep 20; Boyin Zahra. Boyin Zahra University of Engineering; 2017.

25. Hayashi Y, Fukunaga K. Accuracy of rule extraction using a recursive-rule extraction algorithm with continuous attributes combined with a sampling selection technique for the diagnosis of liver disease. Informatics in Medicine Unlocked 2016;5:26-38.

26. Femina B, Anto S. Disease diagnosis using rough set based feature selection and K-nearest neighbor classifier. International Journal of Multidisciplinary Research and Development 2015; 2(4) 664-68.

27. Behravesh A, Sadeghzadeh M. Diagnosis of Hepatitis Disease Combination of Cuckoo Search and Minimum Squared Support Machine. International Conference on Engineering & IT; 2016 Jun 1; Tehran: University of Tehran; 2016.

28. Khodarahmi R, Rouhani M. Detection of Hepatitis Disease Using the Combination of the K-Means Algorithm and Weed Optimization Algorithm. The International Conference on New Researches in

Engineering Sciences; 2016 Jun 6; Tehran, Shahr University of Technology; 2016.
29. Shahbakhsh P, Rajaei A. Detection of hepatitis disease using basic components analysis and LS-SVM. 2nd International Conference in Research in Electrical Engineering and Computer Science; 2014 May 13; Ramsar: Institute of Higher Education of Kasra, Ramsar; 2016.
30. Kabirirad S, Afshin VR. Prognosis of hepatitis disease by artificial neural network. University National Conference, Development Axis; 2018 Dec 8; Torbat Heydarieh: Payame Noor University Torbat Heydarieh; 2016.
31. Ebrahimi M, Roosta M, Farjami Y. A Comparative Analysis of Data-Based Techniques in the Diagnosis of Hepatitis, International Conference on Nonlinear Systems and Optimization of Electrical and Computer Engineering; 2016 May 26; Dubai: Pandar Andish Rahpo Co.,; 2016.
32. Krashadizadeh Nahid, and Rezaei H. Diagnosis of Hepatitis Disease Using Rave Collection and Machine Learning. National Conference on Computer Engineering and Information Technology Management; 2014 May 29; Tehran: Farzin Tolo and Science Company; 2013.
33. Shilaskar S, Ghatol A. Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. Expert Systems with Applications 2013;40(10):4146-53.

Original Article

Journal of Health and Biomedical Informatics
Medical Informatics Research Center
2019; 6(1): 32-45

A Decision Support System for Diagnosis of Diabetes and Hepatitis, based on the Combination of Particle Swarm Optimization and Firefly Algorithm

Soleimanian Gharehchopogh Farhad [1*], Mousavi Seyyed Keyvan [2]

• **Received:** 18 Jul, 2018 • **Accepted:** 11 Oct, 2018

Introduction: Clinical Decision Support Systems (CDSS) are designed in the form of computer programs that help medical professionals make decisions about disease diagnosis. The main aim of these systems is to assist physicians in diagnosing diseases, in other words, a physician can interact with the system and use them to analyze patient data, diagnose diseases, and other medical activities.

Method: This is a descriptive-analytic study. The datasets include 768 records of diabetes with 8 features and 155 records of hepatitis with 19 features, which were provided by the Global Website of UCI. In this study, the Particle Swarm Optimization (PSO) algorithm was used for Feature Selection (FS) and the Firefly Algorithm (FA) was used to classify diabetes and hepatitis into two healthy and unhealthy classes. 80% of the data was used for training and the remaining (20%) was used for testing.

Results: The experiments showed that the accuracy of the PSO and FA for the diabetes dataset was 84.41% and 82.08%, respectively. Also, the accuracy of the PSO and FA for the hepatitis dataset was 81.84% and 80.34%, respectively. The accuracy of the proposed model for the diabetes and hepatitis datasets was 95.38% and 94.09%, respectively.

Conclusion: According to the results, the proposed model had a lower error rate in diagnosis compared to the PSO and FA. The results of this study can help doctors in timely diagnosis of diabetes and hepatitis.

Keywords: Clinical Decision Support Systems, Diagnosis of disease, Diabetes mellitus, Hepatitis, Particle pool optimization algorithm, Firefly algorithm

• **Citation:** Soleimanian Gharehchopogh F, Mousavi SK. A Decision Support System for Diagnosis of Diabetes and Hepatitis, based on the Combination of Particle Swarm Optimization and Firefly Algorithm**.** Journal of Health and Biomedical Informatics 2019; 6(1): 32-45. [In Persian]

1. Ph.D., in Computer Engineering, Assistant Professor, Computer Engineering Dept., Urmia Branch, Islamic Azad University, Urmia, Iran
2. Ph.D. Student, in Computer Engineering, Computer Engineering Dept., Urmia Branch, Islamic Azad University, Urmia, Iran
*Correspondence: Faculty of Engineering, Urmia Islamic Azad University, 2km Airport, Urmia, Iran.
• **Tel:** 09141764427 • **Email:** bonab.farhad@gmail.com