مجله مدیریت اطلاعات سلامت
سال شانزدهم/ شماره سوم/ مرداد و شهریور ۱۳۹۸
Published by *Vesnu Publications*
DOI: 10.22122/him.v16i3.3886

تشخیص اختلالات کبد با استفاده از ترکیب سیستم استنتاج عصبی– فازی تطبیقی و الگوریتم بهینه‌سازی ازدحام ذرات

محمد حسن احمدی[۱]، محمدرضا رمضان‌پور[۲]، ریحانه خورسند[۳]

مقاله پژوهشی

چکیده

مقدمه: بروز بیماری‌های کبد می‌تواند فرد را در طولانی مدت مستعد سرطان کبد نماید که از مرگبارترین نوع سرطان‌ها در جهان به شمار می‌رود و در عین حال، قابل پیشگیری است. تشخیص زودهنگام بیماری‌های کبدی، امری ضروری جهت درمان آن‌ها می‌باشد. هدف از انجام پژوهش حاضر، دسته‌بندی وضعیت بیماران کبدی بر اساس شاخص‌های آزمایشگاهی با استفاده از رویکرد داده‌کاوی بود.

روش بررسی: در این مطالعه توصیفی، از الگوریتم‌های بهینه‌سازی ازدحام ذرات PSO (Particle Swarm Optimization) و سیستم استنتاج عصبی- فازی تطبیقی ANFIS (Adaptive Neuro-Fuzzy Inference System) برای تشخیص اختلالات کبد در افراد سالم و بیمار استفاده گردید. بدین ترتیب، داده‌ها از پایگاه داده معتبر دانشگاه کالیفرنیا- ارواین UCI (University of California-Irvine) دریافت شد. برای ارزیابی روش پیشنهادی، معیارهای دقت، حساسیت و صحت مورد استفاده قرار گرفت.

یافته‌ها: ترکیب ANFIS و الگوریتم PSO با متوسط صحت ۹۹/۱۴ درصد برای مجموعه داده مبتلایان به بیماری‌های کبدی در هند ILPD (Indian Liver Patient Dataset) قادر به تشخیص اختلالات کبدی بود.

نتیجه‌گیری: نتایج به دست آمده حاکی از توانمندی بالای مدل ANFIS در تشخیص اختلالات کبد می‌باشد. مدل پیشنهادی نسبت به سایر مدل‌های مورد مقایسه دارای حداقل میزان خطا و بیشترین صحت و دقت است. بنابراین، به کارگیری این مدل در زمینه تشخیص بیماری‌های کبد پیشنهاد می‌شود.

واژه‌های کلیدی: داده‌کاوی؛ تشخیص؛ بیماری‌های کبد؛ بیماران؛ اکتشافی

پیام کلیدی: بیماری‌های کبد در مراحل اولیه هیچ علایمی ندارند. بنابراین، کوچک‌ترین نشانه‌ای باید جدی گرفته شود. تست‌های عملکرد کبدی، آزمایش‌های خون هستند که جهت کمک به تشخیص و نظارت بر بیماری یا آسیب کبدی استفاده می‌شود. با این تست‌ها می‌توان عملکرد کبد را بررسی نمود. داده‌کاوی روش آسان، سریع و دقیقی برای تشخیص و پیش‌بینی انواع بیماری‌ها می‌باشد و می‌تواند موجب کاهش زمان تشخیص بیماری و بهبود دقت و صحت تشخیص شود. در مطالعه حاضر با ارایه مدل پیشنهادی مبتنی بر داده‌کاوی، میزان صحت ۹۹/۱۴ درصد در تشخیص اختلالات کبد به دست آمد.

دریافت مقاله: ۱۳۹۸/۲/۲۸ پذیرش مقاله: ۱۳۹۸/۵/۱۰ تاریخ انتشار: ۱۳۹۸/۵/۱۵

ارجاع: احمدی محمد حسن، رمضان‌پور محمدرضا، خورسند ریحانه. **تشخیص اختلالات کبد با استفاده از ترکیب سیستم استنتاج عصبی- فازی تطبیقی و الگوریتم بهینه‌سازی ازدحام ذرات.** مدیریت اطلاعات سلامت ۱۳۹۸؛ ۱۶ (۳): ۱۲۱-۱۱۵

مقدمه

کبد بزرگ‌ترین عضو داخلی بدن به شمار می‌رود (۱) که نقش مهمی در فرایند گوارش، دستگاه هاضمه و متابولیسم مواد غذایی به عهده دارد و کمک می‌کند که ریزمغذی‌های حیاتی به اشکال و انواع قابل استفاده برای بدن تبدیل شوند (۲). اصطلاح بیماری کبدی برای بسیاری از بیماری‌ها و اختلالاتی به کار می‌رود که موجب می‌شود کبد به درستی نتواند به فعالیت طبیعی خود ادامه دهد و دچار نارسایی گردد (۳). سرطان کبد پنجمین سرطان شایع جهان و از کشنده‌ترین سرطان‌های گوارشی به شمار می‌رود و به دلیل شناسایی در مراحل پیشرفته، سومین سرطان با میزان مرگ و میر بالا می‌باشد (۵، ۴). تشخیص و پیش‌بینی انواع بیماری‌ها با استفاده از تکنیک‌های داده‌کاوی امکان‌پذیر است (۶). کشف الگوهای مفید بین بیماری و علایم بالینی و آزمایشگاهی بیمار، از جمله کاربردهای داده‌کاوی در پزشکی می‌باشد. منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیرمجموعه از داده‌های بیمار و تشخیص بیماری را بیان می‌کند (۷). در مورد کاربرد داده‌کاوی در تشخیص و پیش‌بینی بیماری‌های مختلف، می‌توان به تحقیق Mohapatra و همکاران اشاره کرد که

مقاله حاصل پایان‌نامه کارشناسی ارشد به شماره ۱۹۰۴۱۰۰۹۹۶۲۰۰۷ می‌باشد که با حمایت دانشگاه آزاد اسلامی واحد مبارکه انجام شده است.

۱- دانشجوی کارشناسی ارشد، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد مبارکه، دانشگاه آزاد اسلامی، اصفهان، ایران

۲- استادیار، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد مبارکه، دانشگاه آزاد اسلامی، اصفهان، ایران (نویسنده طرف مکاتبه)

Email: mr.ramezanpoor@gmail.com

۳- استادیار، مهندسی کامپیوتر، گروه مهندسی کامپیوتر، واحد دولت‌آباد، دانشگاه آزاد اسلامی، اصفهان، ایران

میزان صحت ۷۵/۸۳ درصدی را برای مجموعه داده اختلالات کبدی گزارش کردند (۸).

در پژوهش Baitharu و Pani، از شبکه عصبی پرسپترون برای پیش‌بینی بیماری‌های کبد استفاده گردید. نتایج آزمایش‌ها نشان داد که شبکه عصبی پرسپترون چند لایه MLP (Multilayer Perceptron) نسبت به دیگر روش‌ها دارای دقت ۷۱/۵۹ درصدی است. در مطالعه مذکور از تعدادی از کاربردی‌ترین دسته‌بندی‌ها برای پیش‌بینی بیماری‌های کبد استفاده شد و نتایج حاکی از آن بود که این الگوریتم‌ها نتوانستند به میزان صحت بالایی جهت پیش‌بینی بیماری‌های کبد دست یابند، اما نشان دهنده عملکرد قابل قبول MLP که از الگوریتم‌های پرکابرد شبکه‌های عصبی مصنوعی در مباحث دسته‌بندی داده‌های پزشکی است، نسبت به سایر الگوریتم‌ها می‌باشد (۹).

در مطالعه آبدار و همکاران، از الگوریتم‌های دسته‌بند در تشخیص زودهنگام بیماری‌های کبد استفاده شد. آن‌ها از الگوریتم‌های درخت تصمیم C5.0 و آشکارساز تعاملی خودکار CHAID Chi (Chi Automatic Interaction Detector) برای پیش‌بینی بیماری کبد بر روی مجموعه داده‌های پایگاه داده بیماران کبدی هندی ILPD (Indian Liver Patient Dataset) استفاده کردند. نتایج آزمایش‌ها دقت ۹۷/۳۵ درصدی را نشان داد. میزان صحت بالا در تشخیص بیماری و عملکرد مناسب الگوریتم تقویت شده C5.0 و CHAID در شناسایی بیماری کبد و تولید قوانین بیشتر، از جمله مزایای این روش بود. همچنین، بر اهمیت عامل جنسیت در تشخیص بیماری کبد که در مطالعات مشابه کمتر مورد بررسی قرار گرفته است، تأکید شد (۱۰).

تحقیق Islam و همکاران به مقایسه چهار الگوریتم دسته‌بندی یادگیری ماشین جنگل تصادفی RF (Random Forest)، ماشین بردار پشتیبان SVM (Support Vector Machine)، شبکه‌های عصبی مصنوعی ANN (Artificial Neural Network) و رگرسیون لجستیک LR (Logistic Regression) در پیش‌بینی بیماری کبد چرب پرداخت. بر اساس نتایج، الگوریتم‌های RF، SVM، ANN و LR به ترتیب دقت‌های ۶۵، ۶۹، ۶۹ و ۷۰ درصد را نشان دادند. از مزایای این مدل، تشخیص مؤثر بیماری کبد چرب افراد با استفاده از غربالگری اولیه بدون کمک گرفتن از سونوگرافی شکم می‌باشد (۱۱).

در مطالعه Kadu و همکاران، تشخیص ناهنجاری‌های کبد با استفاده از SVM مطرح شد. آن‌ها از SVM به عنوان یک دسته‌بندی‌کننده برای جداسازی کبد طبیعی از کبد بیمار استفاده کردند. از آن‌جا که داده‌ها به صورت خطی قابل جدا شدن نیست، از دسته‌بندی غیر خطی و تکنیکی به نام ترفند Kernel استفاده کردند که SVM با Kernel تابع پایه شعاعی RBF (Radial Basis Function)، میزان دقت ۸۲/۵۳ درصد را نشان داد. انتخاب شاخص‌های مناسب برای Kernelهای غیر خطی دشوار می‌باشد و در صورتی که امکان انتخاب بهترین شاخص‌ها با استفاده از الگوریتم فرابتکاری فراهم گردد، ممکن است میزان نتایج بهبود یابد (۱۲).

در پژوهش ابراهیمی خامنه و همکاران، شناسایی بیماران مبتلا به سیروز در معرض ابتلا به سرطان کبد به کمک رویکردهای داده‌کاوی مطرح شد. میزان تأثیرگذاری مشخصه‌های مورد بررسی بیماران در پیش‌بینی صحیح سرطان در بیماران سیروتیک، از مزایای تحقیق آنان بود که درخت تصمیم توانست میزان دقت ۸۸ درصدی را برای بیماران با اتیولوژی ویروسی و دقت ۹۲ درصدی را برای بیماران با اتیولوژی غیر ویروسی ارایه نماید (۱۳). با توجه به افزایش میزان مبتلایان به اختلالات کبد در ایران (که زمینه‌ساز ابتلا به سرطان کبد و به وجود آمدن بار سنگین هزینه‌های پزشکی بر جامعه می‌باشد) و جهت تشخیص دقیق و سریع بیماری‌های کبدی، در مطالعه حاضر روش هوشمندی برای تشخیص سلامت یا بیماری افراد ارایه گردید.

روش بررسی

در این مطالعه توصیفی، از مجموعه داده ILPD موجود در مخزن یادگیری ماشین پایگاه داده معتبر دانشگاه کالیفرنیا- ارواین UCI (University of California-Irvine) استفاده گردید. این مخزن مجموعه‌ای از پایگاه داده و مجموعه داده‌هایی است که به منظور آنالیز الگوریتم‌های مختلف در حوزه یادگیری ماشین، مورد استفاده قرار می‌گیرند. مجموعه داده ILPD مربوط به بیماران کبد هندوستان و شامل ۴۱۶ بیمار مبتلا به اختلالات کبد و ۱۶۷ فرد سالم است که از این تعداد، ۴۴۱ مرد و ۱۴۲ نفر زن می‌باشند. از این مجموعه داده برای بررسی و تشخیص اختلالات ایجاد شده در کبد انسان استفاده می‌گردد که با بررسی تعدادی از خواص خون شخص حاصل می‌شود. این مجموعه داده شامل ۵۸۳ رکورد می‌باشد که هر رکورد از ۱۰ ویژگی تشکیل شده است. ویژگی‌های هر رکورد در جدول ۱ ارایه شده است.

جدول ۱: ویژگی مشخصه‌های موجود در مجموعه داده (ILPD) Indian Liver Patient Dataset

نام مجموعه داده	نوع مشخصه	نام مشخصه در مدل
ILPD	عددی	Age (سن)
	اسمی	Gender (جنسیت)
	عددی	TB
	عددی	DB
	عددی	AlkPhos
	عددی	SGPT
	عددی	SGOT
	عددی	TP
	عددی	ALB
	عددی	A/G

ILPD: Indian Liver Patient Dataset; TB: Total Bilirubin; DB: Direct Bilirubin; ALP: Alkaline Phosphatase; SGPT: Serum Glutamic Pyruvic Transaminase; SGOT: Serum Glutamate-Oxaloacetate Transaminase; TP: Total Proteins; ALB: Albumin; A/G: Albumin and Globulin Ratio

مقدار متغیرهای میانگین، چارک اول، چارک دوم، چارک سوم و انحراف معیار تمام ۱۰ ویژگی در جدول ۲ نشان داده شده است.

شکل ۱ چارچوب روش پیشنهادی را نشان می‌دهد. این چارچوب از پنج بخش اصلی تشکیل شده که جزییات هر بخش در ادامه بیان شده است.

پیش‌پردازش: اولین مرحله از روش پیشنهادی پیش‌پردازش است. قبل از استفاده از داده‌ها برای پردازش، عملیات پیش‌پردازش داده‌ها شامل برطرف

جدول ۲: میانگین، چارک‌ها و انحراف معیار ویژگی‌های موجود در مجموعه داده

مشخصه متغیر	Age (سن) (سال)	Gender (جنسیت) (درصد)	TB	DB	AlkPhos	SGPT	SGOT	TP	ALB	A/G
میانگین	۴۴/۷۵	مردان ۷۵/۶ زنان ۲۴/۴	۳/۳۰	۱/۴۹	۲۹۰/۵۸	۸۰/۷۱	۱۰۹/۹۱	۶/۴۸	۳/۱۴	۰/۹۵
Q۱	۳۳	–	۰/۸۰	۰/۲۰	۱۷۵/۰۰	۲۳/۰۰	۲۵/۰۰	۵/۸۰	۲/۶۰	۰/۷۰
Q۲	۴۵	–	۱/۰۰	۰/۳۰	۲۰۸/۰۰	۳۵/۰۰	۴۲/۰۰	۶/۶۰	۳/۱۰	۰/۹۳
Q۳	۵۸	–	۲/۶۰	۱/۳۰	۲۹۸/۰۰	۶۱/۰۰	۸۷/۰۰	۷/۲۰	۳/۸۰	۱/۱۰
انحراف معیار	۴/۰۲	–	۲/۴۹	۱/۶۷	۱۵/۵۸	۱۳/۵۱	۱۶/۹۹	۱/۰۹	۰/۸۰	۰/۳۲

TB: Total Bilirubin; DB: Direct Bilirubin; ALP: Alkaline Phosphatase; SGPT: Serum Glutamic Pyruvic Transaminase; SGOT: Serum Glutamate-Oxaloacetate Transaminase; TP: Total Proteins; ALB: Albumin; A/G: Albumin and Globulin Ratio

کردن مشکل داده‌های پرت با استفاده از الگوریتم خوشه‌بندی K-means که از جمله محبوب‌ترین الگوریتم‌های یادگیری بدون نظارت است، بر روی مجموعه داده ILPD صورت گرفت و در آن مجموعه داده‌ها به چهار خوشه تقسیم شد و داده‌های پرت حذف گردید.

نرمال‌سازی: در این مرحله، داده‌ها با استفاده از روش نرمال‌سازی مینیمم-ماکزیمم در بازه صفر و ۱ قرار می‌گیرند (رابطه ۱) تا تأثیر شاخص‌ها در یک سطح قرار گیرد. در رابطه مذکور، x عددی است که می‌خواهیم نرمال شود. قبل از ورود داده‌ها به شبکه باید نرمال شوند؛ چرا که وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت شبکه می‌شود.

$$z = \frac{x-\min(x)}{[\max(x)-\min(x)]}$$ رابطه ۱

انتخاب ویژگی: انتخاب ویژگی را می‌توان به عنوان فرایند شناسایی ویژگی‌های مرتبط و حذف ویژگی‌های غیر مرتبط و تکراری با هدف مشاهده زیرمجموعه‌ای از ویژگی‌ها که مسأله را به خوبی و با حداقل کاهش درجه کارایی تشریح می‌کند، تعریف نمود. در پژوهش حاضر تصمیم به انتخاب ویژگی‌های مناسب جهت بهبود کارایی و سادگی مدل گرفته شد و برای رسیدن به نتایج بهینه‌تر، از الگوریتم PSO استفاده گردید. آغاز کار الگوریتم مذکور در روش پیشنهادی به این صورت است که گروهی از ذرات یا راه‌حل‌ها به صورت تصادفی ساخته و مقداردهی اولیه می‌شود و هر ذره به عنوان یک نقطه در فضای ۱۰ بعدی (به دلیل وجود شاخص) برای مجموعه داده ILPD در نظر گرفته می‌شود و با حرکت دادن این ذرات در طی تکرارهای متوالی و با به‌روز کردن موقعیت ذرات، سعی در یافتن راه‌حل بهینه برای مسأله می‌نمایند.

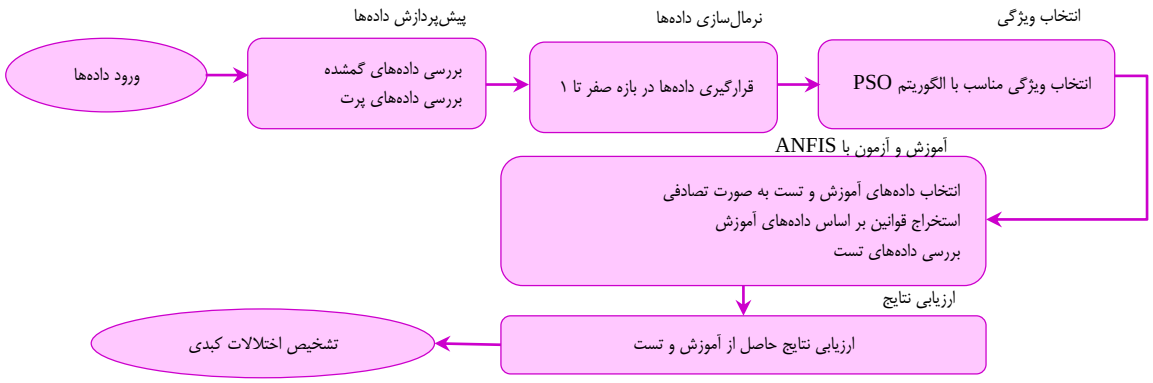
گام‌های اصلی الگوریتم PSO برای روش پیشنهادی در ادامه آمده است (شکل ۲).

الف. الگوریتم PSO با یک گروه از جواب‌های تصادفی به عنوان ذرات شروع به کار می‌کند. هر جواب مسأله معادل با یک ذره می‌باشد.

ب. برای هر جواب، مقدار برازندگی هر ذره ارزیابی می‌شود و در صورتی که مقدار برازندگی بهتری حاصل شود، موقعیت ذره به‌روز می‌گردد.

ج. بهترین موقعیت جدید کل گروه پیدا می‌شود. اگر برازندگی بهترین موقعیت جدید بهتر از گروه قبل باشد، بهترین موقعیت کل جمعیت با استفاده از تابع هدف (رابطه ۲) سنجیده و به‌روز می‌شود که در رابطه مذکور، n تعداد کل ویژگی‌ها، S تعداد زیرمجموعه ویژگی منتخب، جمله اول ضریبی از دقت شناسایی و جمله دوم ضریبی از نرخ کاهش ویژگی‌ها می‌باشد. مجموع ضرایب α و β برابر با ۱۰۰ در نظر گرفته می‌شود.

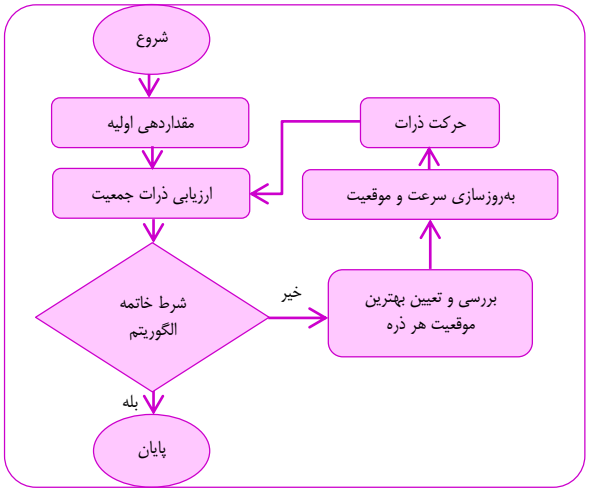
$$\text{Fitness function} = \alpha \times \text{Accuracy} + \beta \times \frac{n+S}{n}$$ رابطه ۲



شکل ۱: چارچوب روش پیشنهادی

ANFIS: Adaptive Neuro-Fuzzy Inference System; PSO: Particle Swarm Optimization

http://him.mui.ac.ir



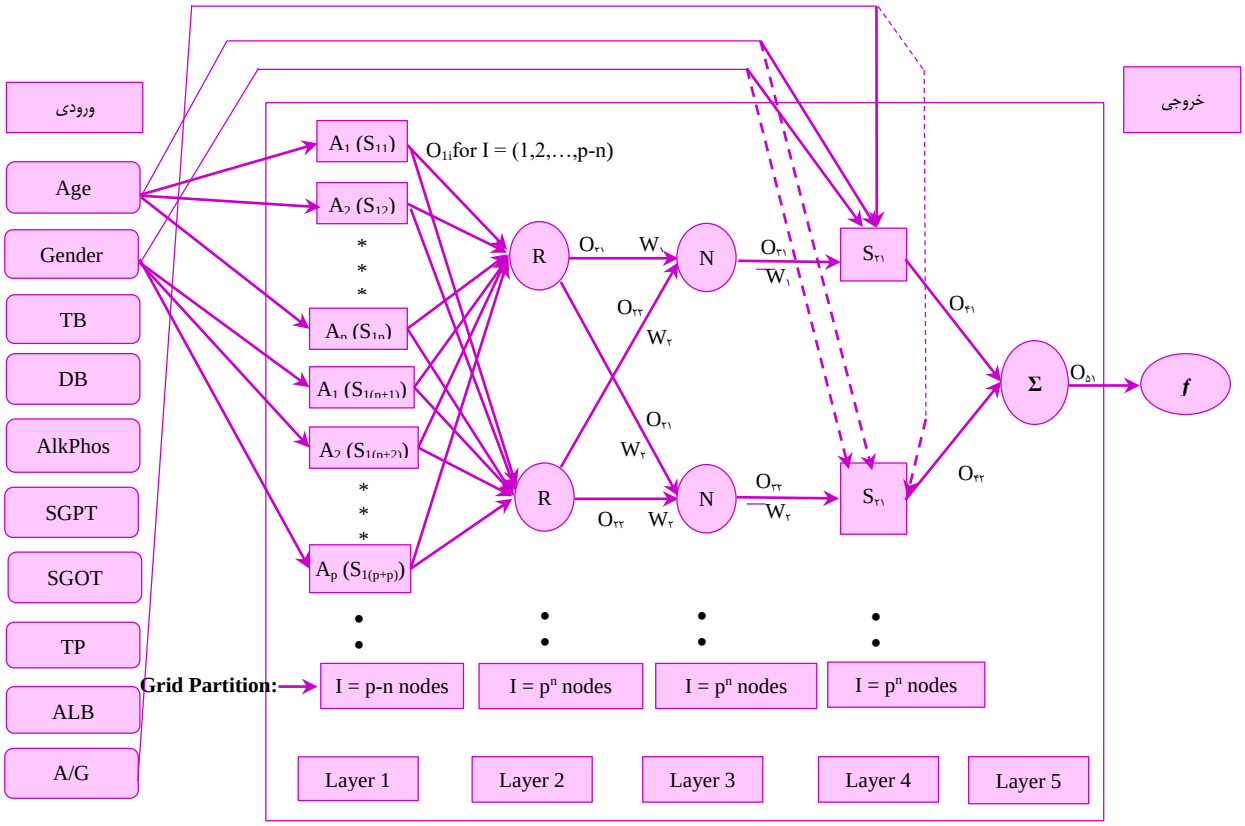
شکل ۲: مراحل الگوریتم Particle Swarm Optimization (PSO)

د. اگر معیار توقف برآورده شود، الگوریتم نیز متوقف می‌شود. معیار توقف در مطالعه حاضر، حداکثر تعداد ۱۰۰۰ تکرار یا خطای کمتر از ۰/۱ درصد در نظر گرفته شد.

ه. برای هر مشخصه، سرعت و موقعیت به‌روز می‌شود و به گام ۲ برمی‌گردد.

پس از اعمال الگوریتم PSO بر روی ویژگی‌ها، مؤثرترین ویژگی‌های تشخیص اختلالات کبدی انتخاب و وارد ANFIS می‌گردد. در این الگوریتم، از شاخص‌های مختلفی استفاده شده که مقادیر آن در بازه‌های متفاوتی مورد ارزیابی قرار گرفته است. برای بررسی دقیق‌تر، این شاخص‌ها با توجه به بازه خواسته شده درون حلقه قرار گرفته شد. برای بررسی دقیق‌تر شاخص‌هایی همچون تعداد ذرات اولیه و تعداد تکرار، با هر بار اجرای برنامه، داده‌ها مجدد به صورت تصادفی انتخاب می‌شود تا شرایط مختلفی مورد بررسی قرار گیرد.

آموزش و آزمون: مرحله آموزش و آزمون با ANFIS صورت می‌گیرد که قابلیت خوبی در آموزش، ساخت و طبقه‌بندی دارد. ساختار ANFIS از ۵ لایه تشکیل شده است. لایه اول فرایند فازی‌سازی را انجام می‌دهد. در این لایه، نوع تابع عضویت و تعداد آن مشخص و تمام قوانین موجود تشکیل می‌شود. در لایه دوم، میزان اثرگذاری هر قانون محاسبه می‌گردد که می‌توان آن را لایه استنتاج هم نامید. در این لایه، قوانین تعریف می‌شود. در لایه سوم، میزان اثرگذاری هر قانون با توجه به تأثیر قوانین دیگر نرمال می‌شود. در لایه چهارم، خروجی هر یک از قوانین به دست می‌آید که محاسبه خروجی‌های وزن‌دار انجام می‌گیرد. در لایه پنجم، خروجی‌های لایه‌های چهارم با هم جمع می‌شود و خروجی سیستم فازی را تشکیل می‌دهد (شکل ۳).



شکل ۳: Adaptive Neuro-Fuzzy Inference System (ANFIS) مورد استفاده در روش پیشنهادی

TB: Total Bilirubin; DB: Direct Bilirubin; ALP: Alkaline Phosphatase; SGPT: Serum Glutamic Pyruvic Transaminase; SGOT: Serum Glutamate-Oxaloacetate Transaminase; TP: Total Proteins; ALB: Albumin; A/G: Albumin and Globulin Ratio

http://him.mui.ac.ir

ANFIS ویژگی‌های انتخاب شده از مرحله استخراج ویژگی توسط الگوریتم PSO را به عنوان ورودی دریافت می‌کند و ۷۰ درصد تعداد نمونه‌ها را به عنوان داده آموزش، ۲۰ درصد را به عنوان داده تست و ۱۰ درصد باقی‌مانده را برای ارزیابی در نظر می‌گیرد؛ البته داده‌های آموزش، تست و ارزیابی به صورت تصادفی انتخاب می‌شوند.

پس از آموزش داده‌های ورودی با استفاده از توابع یادگیری مختلف، داده‌ها مورد ارزیابی قرار می‌گیرند.

طبق چارچوب پیشنهادی، عملیات انتخاب ویژگی با استفاده از الگوریتم PSO انجام و سپس با استفاده از دسته‌بند ANFIS دسته‌بندی می‌گردد. می‌توان از توابع یادگیری مختلفی برای مدل‌سازی با ANFIS استفاده کرد که در تحقیق حاضر از ۸ تابع که در جدول ۳ آمده است، استفاده و نتایج هر یک با هم مقایسه گردید. با توجه به این که داده‌ها از پایگاه داده استاندارد UCI گرفته شده است، تمامی مجوزهای لازم جهت انجام مطالعه کسب و ملاحظات اخلاقی در کلیه مراحل اجرایی رعایت گردید.

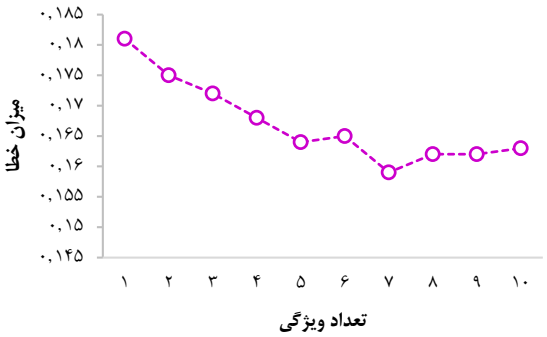
یافته‌ها

شرایط مطالعه: پس از ورود داده‌های خام و انجام عملیات پیش‌پردازش و نرمال‌سازی داده‌ها، طبق الگوریتم پیشنهادی بر روی مجموعه داده، فرایند انتخاب ویژگی با الگوریتم PSO انجام شد که ۱۰ ویژگی از مجموعه داده ILPD انتخاب گردید و سپس با استفاده از دسته‌بند ANFIS به دو دسته سالم و بیمار دسته‌بندی شد. برای نشان دادن اهمیت انتخاب ویژگی، دو سناریو طراحی گردید. در سناریوی اول، دسته‌بندی بدون الگوریتم PSO با تمام ویژگی‌های مجموعه داده انجام می‌شود و در سناریوی دوم، ابتدا تعداد ویژگی‌ها با استفاده از الگوریتم PSO کاهش داده می‌شود و سپس به ANFIS وارد می‌گردد. برای پیش‌بینی و تشخیص سرطان کبد، از الگوریتم پیشنهادی و جهت ارزیابی روش پیشنهادی از روش K-لایه (K-Fold) که در آن K = ۱۰ می‌باشد، استفاده شد. در این روش، داده‌ها به K زیرمجموعه تقسیم می‌گردد و از این زیرمجموعه‌ها هر بار یکی برای اعتبارسنجی و K-۱ تای دیگر برای آموزش سیستم مورد استفاده قرار می‌گیرد. در پایان، میانگین نتایج K بار اعتبارسنجی به عنوان تخمین نهایی اعلام می‌شود.

شاخص‌های مورد ارزیابی: به منظور ارزیابی روش پیشنهادی، از متغیرهای صحت، دقت و حساسیت استفاده گردید.

الف. سناریوی اول (دسته‌بندی با تمام ویژگی‌های مجموعه داده): عملیات دسته‌بندی با ANFIS بر روی مجموعه داده ILPD با ۱۰ ویژگی صورت می‌گیرد. نتایج کامل ANFIS بر روی مجموعه داده ILPD در جدول ۳ ارایه شده است. در این مجموعه داده، ۴۱۶ نمونه بیمار و ۱۶۷ نمونه سالم مورد بررسی قرار گرفت و نتایج صحت، دقت و حساسیت برای هر اجرا با ۸ تابع یادگیری نشان داده شد. با توجه به جدول ۳، نتایج حاصل از صحت بر روی داده‌های تست مجموعه داده ILPD در بازه ۶۴/۳۲ تا ۸۲/۶۸ درصد می‌باشد.

ب. سناریوی دوم (دسته‌بندی با ویژگی‌های کاهش یافته): در این سناریو، عملیات انتخاب ویژگی با الگوریتم PSO بر روی مجموعه داده ILPD انجام پذیرفت و مشخص شد که با انتخاب تعداد ۷ ویژگی، الگوریتم PSO کمترین خطا به میزان ۰/۱۵۹ درصد را دارد (شکل ۴).



شکل ۴: عملیات انتخاب ویژگی‌های مؤثر بر روی مجموعه داده (ILPD) Indian Liver Patient Dataset

سپس خروجی الگوریتم PSO که شامل ۷ ویژگی بود، وارد ANFIS شد. نتایج عملیات بر روی مجموعه داده ILPD در جدول ۳ ارایه شده است. در این مجموعه داده، ۴۱۶ نمونه بیمار و ۱۶۷ نمونه سالم مورد بررسی قرار گرفت و نتایج صحت، دقت و حساسیت برای هر اجرا با ۸ توابع یادگیری به دست آمد. بر این اساس، می‌توان استنباط نمود که میزان صحت به دست آمده با توابع یادگیری مختلف بر روی داده‌های تست مجموعه داده ILPD انتخاب ویژگی شد و در بازه ۹۱/۷۷ تا ۹۹/۱۴ درصد می‌باشد. بیشترین مقدار صحت، حساسیت و دقت به ترتیب ۹۹/۱۴، ۹۹/۷۶ و ۹۹/۰۴ درصد بود که نشان دهنده عملکرد بالای ANFIS است.

جدول ۳: مقایسه نتایج Adaptive Neuro-Fuzzy Inference System (ANFIS) با انتخاب ویژگی و بدون انتخاب ویژگی

توابع عضویت ANFIS	صحت (درصد)		حساسیت (درصد)		دقت (درصد)	
	ANFIS + PSO	ANFIS	ANFIS + PSO	ANFIS	ANFIS + PSO	ANFIS
Gbellmf	۹۴/۵۱	۷۲/۹۰	۹۸/۰۰	۸۵/۴۴	۹۴/۲۳	۷۴/۷۶
Trimf	۹۶/۵۷	۷۵/۶۴	۹۸/۷۷	۸۵/۶۸	۹۶/۳۹	۷۹/۰۹
Trapmf	۹۹/۱۴	۸۲/۶۸	۹۹/۷۶	۸۹/۰۸	۹۹/۰۴	۸۶/۳۰
Gaussmf	۹۴/۱۷	۶۹/۱۳	۹۷/۵۱	۷۹/۳۵	۹۴/۲۳	۷۶/۶۸
gauss2mf	۹۱/۷۷	۷۱/۵۳	۹۴/۸۸	۸۱/۲۵	۹۳/۵۱	۷۸/۱۳
Pimf	۹۱/۷۷	۶۴/۳۲	۹۶/۲۳	۷۶/۰۰	۹۲/۰۷	۷۳/۰۸
Dsigmf	۹۳/۸۳	۷۳/۹۳	۹۷/۵۰	۸۳/۸۵	۹۳/۷۵	۷۸/۶۱
Psigmf	۹۴/۱۷	۷۰/۵۰	۹۷/۷۵	۷۹/۹۰	۹۳/۹۹	۷۸/۳۷

ANFIS: Adaptive Neuro-Fuzzy Inference System; PSO: Particle Swarm Optimization

http://him.mui.ac.ir

بحث

نتایج مطالعه حاضر نشان داد که انتخاب ویژگی‌های مؤثر با استفاده از الگوریتم PSO در دسته‌بندی داده‌های اختلالات کبد اثربخش و مفید می‌باشد. به طور کلی، در بهترین حالت، صحت پاسخگویی ANFIS با الگوریتم PSO برای مجموعه داده ILPD برابر با ۹۹/۱۴ است. در روش پیشنهادی، پس از ورود داده‌های خام و انجام عملیات پیش‌پردازش و نرمال‌سازی داده‌ها، انتخاب ویژگی با PSO انجام گرفت و ۷ ویژگی مؤثرتر از بین ۱۰ ویژگی انتخاب گردید و سپس با استفاده از دسته‌بند ANFIS، به دو دسته سالم و بیمار دسته‌بندی شد. روش مذکور نسبت به تحقیقات آبدار و همکاران (۱۰) و Kadu و همکاران (۱۲) که بر روی مجموعه داده یکسان انجام شده است، عملکرد بهتری را نشان داد. با توجه به بررسی‌ها و نتایج حاصل از ارزیابی‌های صورت گرفته در بررسی حاضر، اهمیت استفاده از مدل‌سازی در مطالعه بیماری‌ها، بیشتر مشخص شد.

از جمله محدودیت‌های مطالعه حاضر می‌توان به عدم دسترسی به داده‌های واقعی بیماران در ایران اشاره نمود. از طرف دیگر، عدم ثبت اطلاعات بیماران به صورت دیجیتال، جمع‌آوری آن‌ها را با مشکل مواجه کرد که در صورت رفع این محدودیت‌ها، به طور قطع می‌توان به نتایج بهتری برای بیماران ایرانی دست یافت.

نتیجه‌گیری

در پژوهش حاضر، یک تکنیک دسته‌بندی شامل ترکیب ANFIS با الگوریتم PSO ارایه شد و کارایی آن روی مجموعه داده استاندارد مربوط به بیماران کبدی مورد بررسی و ارزیابی قرار گرفت. با توجه با نتایج به دست آمده، می‌توان نتیجه گرفت که ترکیب ANFIS با الگوریتم PSO عملکرد مناسبی دارد و روش مناسبی در جهت تشخیص اختلالات کبد است. با توجه به کمبود و محدودیت منابع مالی و نیروی انسانی متخصص در حوزه بهداشت و درمان و همچنین، مدیریت بهینه منابع، می‌توان گفت که روش پیشنهادی کمک شایانی در بهبود ارایه خدمات می‌کند. در این روش با شناسایی شاخص‌های ضروری و مؤثر در تشخیص بیماری توسط پزشکان، هزینه‌های تحمیل شده بر بیمار، بیمارستان و صنعت بیمه کاهش چشمگیری می‌یابد و به تشخیص دقیق و سریع بیماری در مراحل اولیه کمک می‌شود تا بیمار عمر بیشتر وکیفیت زندگی بهتری داشته باشد.

پیشنهادها

استفاده از ویژگی‌های مؤثر در تشخیص اختلالات کبد، می‌تواند در افزایش دقت و کارایی روش پیشنهادی مؤثر باشد. از این‌رو، پیشنهاد می‌گردد این موضوع در مطالعات آینده مد نظر قرار گیرد. همچنین، پیشنهاد می‌شود از سایر تکنیک‌های داده‌کاوی برای طراحی مدل‌های پیش‌بینی استفاده گردد و کارایی آن‌ها مورد بررسی قرار گیرد. با توجه به دقت بالای روش پیشنهادی و نیاز در حوزه پزشکی، توسعه نرم‌افزاری نتایج تحقیق حاضر به صورت برنامه‌های کاربردی توصیه می‌شود.

تشکر و قدردانی

بدین وسیله از تمام استادانی که در انجام مطالعه حاضر همکاری نمودند، تشکر و قدردانی به عمل می‌آید.

تضاد منافع

در انجام مطالعه حاضر، نویسندگان هیچ‌گونه تضاد منافعی نداشته‌اند.

References

1. Midhila M, Raghesh Krishnan K, Sudhakar R. A study of the phases of classification of liver diseases from ultrasound images and gray level difference weights based segmentation. Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP); 2017 Apr 6-8; Chennai, India.
2. Anisha PR, Reddy CKK, Prasad LVN. A pragmatic approach for detecting liver cancer using image processing and data mining techniques. Proceedings of the 2015 International Conference on Signal Processing and Communication Engineering Systems; 2015 Jan 2-3; Guntur, India.
3. Abdar M. A survey and compare the performance of ibm spss modeler and rapid miner software for predicting liver disease by using various data mining algorithms. Cumhuriyet University Faculty of Science 2015; 36(3): 3230-41.
4. Ramana BV, Babu MSP, Venkateswarlu NB. A critical study of selected classification algorithms for liver disease diagnosis. International Journal of Database Management Systems 2011; 3(2): 101-14.
5. Maimone S, Saffioti F, Oliva G, Di Benedetto A, Alibrandi A, Filomia R, et al. Erectile dysfunction in compensated liver cirrhosis. Dig Liver Dis 2019; 51(6): 843-9.
6. Maynard E. Decompensated cirrhosis and fluid resuscitation. Surg Clin North Am 2017; 97(6): 1419-24.
7. Pazouki M, Sepehri MM, Saberifiroozi M. Discovering hidden cluster structures in patients with cirrhosis based on laboratory data. Govaresh 2014; 19(3): 191-7. [In Persian].
8. Mohapatra P, Chakravarty S, Dash PK. An improved cuckoo search based extreme learning machine for medical data classification. Swarm Evol Comput 2015; 24: 25-49.
9. Baitharu TR, Pani SK. Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. Procedia Comput Sci 2016; 85: 862-70.
10. Abdar M, Zomorodi-Moghadam M, Das R, Ting IH. Performance analysis of classification algorithms on early detection of liver disease. Expert Syst Appl 2017; 67: 239-51.
11. Islam MM, Wu CC, Poly TN, Yang HC, Li YJ. Applications of machine learning in fatty live disease prediction. Stud Health Technol Inform 2018; 247: 166-70.
12. Kadu G, Raut R, Gawande SS. Diagnosis of liver abnormalities using support vector machine. Journal of Science and Engineering Development Research 2018; 3(7): 132-7.
13. Ebrahimi Khameneh M, Sepehri MM, Saberifiroozi M. Using data mining for identify patients at high risk to hepatocellular carcinoma in cirrhosis liver: Preliminary report. Govaresh 2015; 19(4): 265-74. [In Persian].

Health Information Management

Published by *Vesnu Publications*

Vol 16/ No 3/ August & September 2019

DOI: 10.22122/him.v16i3.3886

Diagnosis of Liver Disorders Using a Combination of Adaptive Neuron-Fuzzy Inference System and Particle Swarm Optimization Algorithm

Mohammadhasan Ahmadi[1], Mohammadreza Ramezanpour[2], Reihaneh Khorsand[3]

Original Article

Abstract

Introduction: The incidence of liver diseases in a person can lead to susceptibility to liver cancer in long-term, which is one of the deadliest forms of cancer in the world, which can be prevented. Early diagnosis of liver diseases is essential for their treatment. The purpose of this study was to classify the status of liver patients based on laboratory parameters using the data mining approach.

Methods: In this descriptive study, particle swarm optimization (PSO) algorithms and adaptive neuron-fuzzy inference system (ANFIS) were used to diagnose liver disorders in healthy individuals and patients. The data were taken from University of California-Irvine (UCI) database. Accuracy, sensitivity, and precision criteria were used to evaluate the proposed method.

Results: The combination of ANFIS and PSO algorithm with average accuracy of 14.99 percent was able to detect liver disorders in Indian Liver Patient Dataset (ILPD).

Conclusion: The results of this study indicate the high abilities of ANFIS in liver disorders detection. The proposed model has minimum error, and maximum accuracy and precision compared to other models. The application of this model is proposed in the detection of liver diseases.

Keywords: Data Mining; Diagnosis; Liver Diseases; Patients; Heuristics

Received: 18 May, 2019 Accepted: 01 Aug., 2019 Published: 06 Aug., 2019

Citation: Ahmadi M, Ramezanpour M, Khorsand R. **Diagnosis of Liver Disorders Using a Combination of Adaptive Neuron-Fuzzy Inference System and Particle Swarm Optimization Algorithm.** Health Inf Manage 2019; 16(3): 115-21

Article resulted from MSc thesis No. 19041009962007 funded by Mobarakeh Branch, Islamic Azad University.

1- MSc Student, Computer Engineering, Department of Computer Engineering, Mobarakeh Branch, Islamic Azad University, Isfahan, Iran

2- Assistant Professor, Computer Engineering, Department of Computer Engineering, Mobarakeh Branch, Islamic Azad University, Isfahan, Iran (Corresponding Author) Email: mr.ramezanpoor@gmail.com

3- Assistant Professor, Computer Engineering, Department of Computer Engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran