APJ. 4(2):21-39, Summer 2023

Arman Process Journal (APJ)

Homepage: armanprocessjournal.ir

A Review of Smart Methods in Diagnosing and Predicting Liver Diseases Using Machine Learning Techniques and Meta Heuristic Algorithms

F. WahabAminiazar[1*], RasoulFarahi[2], Eqbal Khancheh Sepehrardin[3]

[1] Assistant Professor, Department of Computer Engineering and Information Technology, Faculty of Electrical and Computer Engineering , Islamic Azad University, Mahabad Branch, Mahabad, Iran

[2] Instructor, Department of Computer Engineering and Information Technology, Faculty of Electrical and Computer Engineering, Islamic Azad University, Mahabad Branch, Mahabad, Iran.

[3]Instructor, Department of Computer Engineering and Information Technology, Faculty of Electrical and Computer Engineering, Islamic Azad University, Mahabad Branch, Mahabad, Iran

Received: 29 June 2023
Accepted: 10 August 2023

KEYWORDS:
Liver Diseases,
Intelligent Diagnosis and Prediction,
Data Mining,
Machine Learning Techniques,
Metaheuristic Algorithms,

[1] Corresponding author
Aminiazar@iau-mahabad.ac.ir

ABSTRACT

One of the most important problems in the world is the increase in costs in the field of health. One of the important research in recent years to reduce these costs is the prediction of diseases. Liver diseases are one of the most serious diseases in the world, because the liver plays a vital role in the human body, and any disorder in the liver causes serious and sometimes irreparable problems in the body. The symptoms of liver diseases are not known in the early stages, and when the symptoms become clear enough, the situation is usually quite critical. Based on this, early diagnosis of liver diseases is necessary. The purpose of this article is to review methods based on machine learning techniques and metaheuristic algorithms to predict liver diseases. In this article, it is intended to collect and review a comprehensive review of liver patients and diagnosis methods in the field of medicine. For this reason, liver diseases were examined first, and following this examination, we will discuss and analyze the field of early diagnosis of these diseases. In the following, we will introduce various machine learning techniques and metaheuristic algorithms in the diagnosis of liver diseases. After that, we will discuss the methods of diagnosing and predicting liver diseases and we will review these methods in terms of their goals, limitations and capabilities. In this article, for the prediction of liver diseases, different points of view such as: feature selection, classification method, used tools are examined, which are shown by the analysis of classification algorithms. It should be noted that meta-innovative algorithms find an approximate solution faster compared to traditional data mining algorithms and exact methods and usually provide better results compared to deterministic algorithms. . Also, the results show that the ILPD database and MATLAB tools are the most useful in diagnosing and predicting liver diseases.

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
69	3	2

نشریه تخصصی آرمان پردازش، دوره ۴، شماره ۲، تابستان ۱۴۰۲

فصلنامه تخصصی آرمان پردازش
(APJ)
Homepage: www.armanprocessjournal.ir

مروری بر روش‌های هوشمند در تشخیص و پیش‌بینی بیمارهای کبد با استفاده از تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری

وهاب امینی‌آذر[۱*]، رسول فرحی[۲] و اقبال خوانچه سپهرالدین[۳]

[۱*] استادیار ، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد مهاباد، مهاباد، ایران
۲ مربی، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد مهاباد، مهاباد، ایران
۳ مربی، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی واحد مهاباد، مهاباد، ایران

چکیده

یکی از مهم‌ترین مشکلات درجهان، افزایش هزینه‌ها در حوزه سلامت است؛ از پژوهش‌های مهم سال های اخیر برای کاهش این هزینه‌ها، پیش‌بینی بیماری‌ها می‌باشد. بیماری‌های کبد یکی از بیماری های جدی در جهان است، زیرا کبد نقش حیاتی در بدن انسان دارد و هرگونه اختلال در کبد باعث مشکلات جدی و بعضاًجبران ناپذیزی در بدن می‌شود. علائم بیماری‌های کبدی در مراحل اولیه ماهر نمی‌شوند و زمانیکه علائم به اندازه کافی واضح شوند، معمولا وضعیت کاملاً بحرانی است. بر این اساس تشخیص زودهنگام بیماری‌های کبد لازم و ضروری است. هدف از این مقاله مروری بر روش‌های مبتنی بر تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری به منظور پیش‌بینی بیماری‌های کبد می‌باشد. در این مقاله قصد بر آن بوده تا یک بررسی جامعی از بیمارهای کبد، روش‌های تشخیص در حوزه پزشکی را گردآوری و بررسی نمود. بدین جهت در ابتدا بیمارهای کبدی بررسی شده و پیرو این بررسی به بحث و تحلیل حوزه تشخیص زود هنگام این بیماری‌ها خواهیم پرداخت. در ادامه انواع تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در تشخیص بیماری‌های کبد را معرفی خواهیم کرد. پس از آن به بحث در ارتباط با روش‌های تشخیص و پیش‌بینی بیمارهای کبد خواهیم پرداخت و این روش‌ها را از نظر اهداف، محدودیت ها و قابلیت‌ها نقد و بررسی خواهیم کرد. در این مقاله برای پیش‌بینی بیماری‌های کبد دیدگاه‌های مختلف که عبارتند از :انتخاب ویژگی، روش طبقه‌بندی، ابزارهای استفاده شده، مورد بررسی قرار می‌گیرند، که با تحلیل الگوریتم‌های طبقه-بندی نشان می‌دهیم که الگوریتم‌های فراابتکاری در مقایسه با الگوریتم‌های سنتی داده‌کاوی و روش-های دقیق، راه‌حل تقریبی را سریع‌تر پیدا می‌کنند و در مقایسه با الگوریتم‌های قطعی معمولا نتایج بهتری را ارائه می‌دهند. همچنین نتایج گویا این است که که پایگاه داده ILPD و ابزار MATLAB بیشترین کاربرد را در تشخیص و پیش‌بینی بیماری‌های کبد دارند.

واژگان کلیدی:
بیماری‌های کبد،
تشخیص و پیش‌بینی هوشمند،
داده‌کاوی،
تکنیک‌های یادگیری ماشین،
الگوریتم‌های فرابتکاری.

تعداد جداول	تعداد شکل ها	تعداد مراجع
۲	۳	۶۹

۱- مقدمه

در دو دهه گذشته، رشد سریع داده‌ها در پزشکی از چالش‌های اساسی می‌باشد. با توجه به پیشرفت سریع روش‌های بیوتکنولوژی و روش های تجزیه وتحلیل داده‌های زیستی، روش‌های داده‌کاوی در حال رشد می‌باشد. امروزه برای تجزیه و تحلیل و پیش بینی داده‌های بزر پزشکی از الگوریتم‌های داده‌کاوی و یادگیری ماشین استفاده می‌شود. هدف اصلی از تجزیه و تحلیل این داده‌ها، توسعه مدل‌های مختلف پیش بینی در حوزه پزشکی می‌باشد. تشخیص بیماری موضوع داغ در تحقیقات فعلی است. ویژگی اصلی مدل تشخیص بیماری، تصمیم گیری سریع پزشکان و کاهش اشتباهات در تشخیص می-باشد و داده‌های پزشکی را خیلی سریع بررسی می‌کند. همچنین از آنجایی که داده‌های پزشکی شامل تعداد زیادی ویژگی است و محققان تنها زیر مجموعه‌ای از این ویژگی‌ها را در نظر می-گیرند و این ویژگی‌ها برای تشخیص بیماری حیاتی است، بنابراین بررسی این ویژگی‌های مهم باید بسیار دقیق باشد. همچنین باید در نظر داشت که هر کدام از این ویژگی‌ها هزینه‌ای برای بیمار در بر خواهد داشت [۱]. دوران محاسبات مدرن، دستیابی به روش‌های هوشمند در بیوانفورماتی را گسترش داده است. با توجه به عدم اطمینان در داده‌های پزشکی، دریافت اطلاعات قابل درک تبدیل به چالشی بزرگ برای پزشکان شده است. این چالش می‌تواند به تشخیص اشتباه‌ی بیماری منجر شود که بیشتر باعث درمان نامناسب بیماری می‌شود. از این رو اگر متخصصین بتوانند با استفاده از تکنیک‌های هوشمند که قادر به تصمیم‌گیری هستند، با بررسی دقیق داده‌های پیچیده و مبهم ارزیابی خود را انجام دهند، برای بیماران مفید خواهد بود. سیستم‌های هوشمند احتمال وقوع خطاهای پزشکی، هزینه و زمان را کاهش می‌دهد [۲]. به کارگیری دانش داده‌کاوی در حوزه تشخیص بیماری کبد از جمله شیوه-های هوشمندی است که می‌تواند با تحلیل عوامل موثر، مشکلات کبدی را زودتر و دقیق‌تر تشخیص داده و به توسعه تحقیقات پزشکی و تصمیم-گیری‌های علمی در زمینه تشخیص و درمان این بیماری کم کند [۳]. از آنجا که کبد در بسیاری از اعمال حیاتی دخالت دارد، بنابراین با بروز اختلال در کبد انتظار وقوع علائم زیادی می‌رود ولی به علت سیر پیشرونده مشکلات کبدی، اغلب بیماری‌های کبدی تا مراحل پیشرفته، علامت خاصی را نشان نمی‌دهند. فقدان علایم در مراحل اولیه ممکن است موجب تشخیص نادرست بیماری توسط بسیاری از پزشکان گردد که این تشخیص نادرست می‌تواند منجر به درمان اشتباه و تجویز داروی نامناسب و در نتیجه ایجاد عوارض حاد و بلند مدت این بیماری و یا مشکلات دیگر گردد. بنابراین تشخیص زودتر و دقیق‌تر مشکلات کبدی به کمک تجزیه و تحلیل دقیق ویژگی‌های مؤثر یک سیستم تشخیص پزشکی اتوماتیک، جهت درمان صحیح و پیشگیری از آسیب‌های جدی به این عضو حیاتی، ضروری به نظر می‌رسد. به همین منظور استفاده از تکنیک های داده‌کاوی و یادگیری ماشین و بهره‌گیری از الگوریتم‌های فراابتکاری جهت ارائه مدلی هوشمند برای تشخیص زودهنگام این بیماری لازم و ضروری می‌باشد.

پژوهش حاضر با هدف بررسی کارآیی، عملکرد و تاثیرات تکنیک‌های یادگیری ماشین و الگوریتم‌های فرابتکاری در تشخیص و پیش‌بینی هوشمند بیماری‌های کبد انجام شده است با توجه به گستردگی تکنیک-های یادگیری ماشین و الگوریتم‌های فرابتکاری پیش‌بینی می‌شود این تکنیک‌های و الگوریتم‌های بهینه راه‌حل‌های بهینه‌ای برای چالش‌های مختلف پزشکی در زمینه تشخیص و پیش‌بینی هوشمند بیماری‌های کبد ارائه دهند. در این مقاله، ابتدا، تعاریف مربوط به بیمارهای کبد، داده‌کاوی، تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در حوزه سلامت مرور شده است، سپس، کاربرد این تکنیک‌ها و الگوریتم‌ها در بحث تشخیص و پیش‌بینی بیمارهای کبد که در چندین سال اخیر توسط محققین انجام شده است بررسی می‌شوند.

۲- کبد و بیمارهای کبد

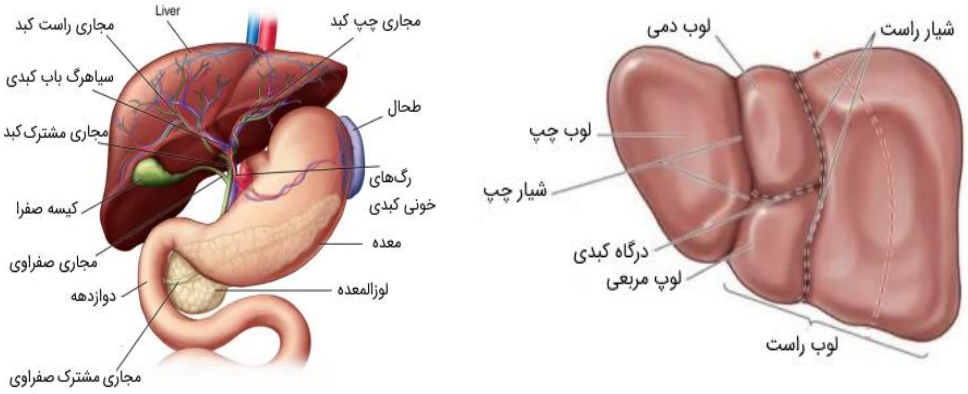
کبد یکی از مهم‌ترین اعضای بدن است. کبد (جگر سیاه) در پشت دنده ها و در بالا و سمت راست معده (شکم) واقع شده است. کبد مخروطی شکل بوده و سطحی صاف، نرم و لاستی مانند دارد. رنگ آن قهوه‌ای متمایل به قرمز تیره است. این اندام که افراد به ندرت درباره‌اش فکر می‌کنند، به قدری مهم است که اگر وظایفش را انجام ندهد در عرض یک روز باعث مرگ فرد می‌شود. کبد بیش از ۵۰۰ وظیفه مختلف را در بدن انجام می‌دهد که از جمله آنها خنثی‌سازی سموم، فراهم کردن قند برای مغز، مقابله با عفونت، ذخیره مواد مغیی و ویتامین‌ها، ساخت پروتئین‌ها و هورمون‌ها، کنترل قند خون و کم به انعقاد خون است.

بیماری کبد یکی از بیماری‌های شایع و خطرناک است که امروزه به یکی از عوامل اصلی مرگ و میر در جهان تبدیل شده است. تشخیص به موقع این بیماری می‌تواند در کنترل، درمان و پیشگیری از عوارض آن بسیار مؤثر باشد [۴]. تشخیص و درمان کبد مانند هر بیماری دیگر، معمولا با گرفتن شرح حال از بیمار و معاینه وی شروع می-گردد. برخی از این معاینات بالینی اولیه شامل بررسی ریتم تپش قلب، عملکرد ریه‌ها در هنگام تنفس و وجود توده‌های غیرطبیعی در ناحیه شکم می‌باشد. از آنجا که علائم بالینی دیده شده در بیماری‌های کبدی در سایر بیماری‌ها نیز عمومیت دارد، در خیلی از موارد، این علائم به تنهایی قابل استناد نیستند؛ بنابراین شناسایی و تشخیص بیماری کبد مستلزم به کارگیری روش‌های آزمایشگاهی مانند الترا سونوگرافی و یا MRI است [۴]. از آنجا که درصد قابل توجهی از بیماران کبدی از افزایش وزن و چاقی به ویژه چاقی شکمی رنج می‌برند، روش‌های آزمایشگاهی جهت تشخیص بیماری مطمئن نیستند. همچنین این روش‌ها در همه جا در دسترس نمی‌باشند [۵]. مشکلات کبدی شامل طیف وسیعی از بیماری‌ها و نارسایی‌ها هستند که به بافت کبد یا عملکرد آن آسیب می‌رسانند. مهم‌ترین آنها عبارتند از [۴]:

- **کبد چرب:** این بیماری التهاب کبدی است ،در این بیماری، تجمع بیش از حد چربی در کبد، موجب اختلال در فعالیت طبیعی بافت کبد می‌گردد که می‌تواند سیر پیشرونده پیدا کرده و باعث نارسایی کبد و یا سیروز کبدی گردد.
- **هپاتیت حاد ویروسی:** این بیماری التهاب شایع کبدی است که به وسیله ویروس‌های هپاتیت A، B، C،Dو E ایجاد می‌گردد.
- **سیروز کبدی:** سیروز کبدی یکی از جدیدترین بیماری‌های کبدی می‌باشد که در نتیجه بیماری‌های مزمن کبدی مانند هپاتیت B ،هپاتیت C و کبد چرب پدید می‌آید و می‌توان آن را به عنوان یک وضعیت پیش‌سرطانی در نظر گرفت. در این بیماری سلول‌های کبدی دیگر به وظایفشان عمل نمی‌کنند و کبد دچار التهاب مداوم، کاهش تدریجی عملکرد و فرسودگی می‌گردد.
- **سرطان کبد:** سرطان کبد می‌تواند تومور بدخیمی از خود کبد باشد که در بیشتر موارد بعد از ابتلا به سیروز کبدی پدید می‌آید و یا به دلیل سرطان دیگر اعضای بدن مانند دهان، ریه، معده، پانکراس، روده بزرگ و حتی مثانه باشد که به کبد منتقل شده و نوعی سرطان ثانویه ایجاد کرده باشد .این بیماری علاوه بر این که موجب نارسایی کبد می‌شود، امکان سرایت آن به سایر اعضای بدن به خصوص ریه‌ها، غدد فوق کلیوی و استخوان‌ها هم وجود دارد. در شکل (۱) قسمت‌های مختلف کبد نشان داده شده است.

۳- داده‌کاوی در حوزه سلامت

امروزه در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف از اهمیت فراوانی برخوردار است. مراکز پزشکی با مقاصد گوناگونی به جمع‌آوری این داده‌ها می‌پردازند. تحقیق روی این داده‌ها و به دست آوردن نتایج و الگوهای مفید در رابطه با بیماری‌ها، یکی از اهداف استفاده از این داده‌ها است. حجم زیاد این داده‌ها و سردرگمی حاصل از آن مشکلی است که مانع رسیدن به نتایج قابل توجه می‌شود.



شکل ۱: کبد و قسمت های مختلف آن

بنابراین از داده کاوی برای غلبه بر این مشکل و به دست آوردن روابط مفید بین عوامل خطرزا در بیماری‌ها استفاده می‌شود. داده کاوی به عنوان فرآیندی برای یافتن الگوها و ارتباطات در پایگاه داده به همراه استفاده از اطلاعات برای ساختن مدل‌های پیش‌بینی تعریف شده است [۶]. و همچنین از آن به عنوان فرآیندی برای انتخاب، اکتشاف و ساختن مدل‌ها با استفاده از انبوه داده‌های ذخیره شده برای کشف الگوهای از پیش موجود نیز یاد می‌شود [۷]. از داده‌کاوی برای شناسایی روابط و الگوهای نو، صحیح، قابل فهم و به صورت بالقوه مفید در درون داده‌ها با استفاده از ترکیب مجموعه داده‌ها و استخراج الگوهای پیچیده برای انسان استفاده می‌شود [۸]. داده‌کاوی، گامی مهم در کشف و استخراج دانش می‌باشد؛ این اصطلاح به معنای کاوش مجموعه داده‌های بزرگ برای استخراج الگوهای ناشناخته بین داده‌ها است [۹]. از آنجایی که کشف روابط بین داده‌ها در روش‌های سنتی آماری بسیار مشکل است [۱۰]، لذا کاربرد داده‌کاوی به سرعت در بخش‌های وسیعی از قبیل سازمان‌های ارائه خدمات بهداشتی، پیش‌بینی مالی و پیش‌بینی هوا گسترش یافته است [۱۱].

داده‌کاوی در مراقبت سلامت، شاخه بسیار مهمی در تشخیص و فهم عمیق‌تر داده‌های پزشکی می‌باشد. داده‌کاوی بهداشتی درمانی، درصدد حل مسائل دنیای واقعی در تشخیص و درمان بیماری هاست [۱۲]. همچنین در مراقبت بهداشتی درمانی، زمینه تحقیقاتی مهمی برای پیش‌بینی بیماری‌ها و درکی عمیق‌تر از داده‌های بهداشتی درمانی محسوب می‌شود. داده‌کاوی بهداشتی درمانی قصد دارد تا مسائل دنیای واقعی را در تشخیص و درمان بیماری‌ها حل کند [۱۳]. پژوهشگران از این روش برای تشخیص بیماری‌های مختلف استفاده می‌کنند و برای اینکار از روش‌ها و الگوریتم‌های مختلفی بهره می‌برند که میزان صحت و دقت آن‌ها با یکدیگر متفاوت است [۱۴]. کاربردهای داده‌کاوی می‌توانند در تمام بخش‌های صنعت مراقبت بهداشتی درمانی به صورت جدی مثمرثمر باشند. برای مثال سازمان‌های ارائه‌دهنده خدمات بیمه

بهداشتی درمانی می‌توانند کلاهبرداری و سوء استفاده‌ها را شناسایی نمایند؛ سازمانهای بهداشتی درمانی تصمیمات مربوط به مدیریت ارتباط با مراجعه‌کنندگان را اتخاذ نمایند، پزشکان، درمانهای موثر و بهترین شیوه‌های درمانی را شناسایی نمایند و بیماران،خدمات مراقبتی بهتر و رضایت بخش‌تر را دریافت نمایند [۱۵و۱۶]. شناخت روش‌های پیشگیرانه و درمانی در بیماری‌های مزمن از قبیل آسم از طریق داده-کاوی میسر می‌شود و این روش نقش بارزی در شناخت الگوها دارد [۱۷]. ابزارها و روش‌های بسیاری برای داده‌کاوی و تحلیل آن‌ها وجود دارد که الگوریتم‌های مشابهی را پیاده‌سازی کرده‌اند. این پژوهش با استفاده از روش‌های تحلیل و داده‌کاوی قصد دارد تا با تحلیل داده‌های گردآوری شده برای بیماری کبد، بخشی از دانش پنهان درآنرا کشف نموده و گزارش کند. فرآیند کشف دانش دارای چندین مرحله می‌باشد که دقت در انجام هر کدام از این مراحل، روی کیفیت کل فرآیند تاثیرگذار است. این مراحل عبارتند از:

۱) **فهم و تعریف مسئله**: مرحله اول فهم حوزه کاری و مسئله-ای است که سعی در پیدا کردن راه حل برای آن داریم.درک کامل مسئله پیش‌نیاز ضروری برای انتخاب روش مناسب داده‌کاوی و کشف دانش مفید از میان داده‌ها می‌باشد [۱۸].

۲) **جمع آوری و پیش پردازش داده ها**: پیش‌پردازش داده‌ها خود شامل مراحلی است. این مراحل عبارتند از: یکپارچه-سازی، حذف نویزها، مقابله با مقادیر مفقوده و تغییر شکل داده‌ها، کاهش تعداد داده ها و یا تعداد ویژگی‌ها و غیره [۲۵].

۳) **داده‌کاوی**: مرحله سوم همان مرحله داده‌کاوی می‌باشدکه با انجام آن الگوها و مدل‌های پنهان درمیان داده‌ها استخراج می‌گردد. در این مرحله ما باید نخست وظیفه داده‌کاوی و سپس روش داده‌کاوی را انتخاب نماییم [۱۸].

۴) **تفسیر و ارزیابی نتایج**: مرحله چهارم شامل تفسیر نتایج به دست آمده از مرحله سوم می‌باشد. لزوما نتایج به دست آمده از مرحله سوم قابل اطمینان نیستند و ممکن است بیانگر دانش مفید و قابل استفاده نباشند. به همین خاطر باید این نتایج را به گونه‌ای ارزیابی نمود. برای ارزیابی مدل به دست آمده میتوان در ابتدای امر داده‌ها را به دو دسته آموزش[۳] و آزمایش[۴] تقسیم نمود، مدل را روی داده‌های آموزش ساخت و سپس نتایج به دست آمده را روی داده‌های تست آزمایش کرد و دقت مدل را محاسبه نمود [۱۸].

۵) **استفاده از دانش کشف شده**: مرحله آخر استفاده از دانش استخراج شده از داده‌ها به صورت عملی می‌باشد. در حقیقت هدف از انجام مراحل مختلف کشف دانش، دستیابی به نتایجی است که بتوان از آنها در دنیای واقعی و برای بهبود کارایی سازمان‌ها استفاده کرد. این دانش استخراج شده می-تواند به عنوان یک سیستم کمک تصمیم در دنیای واقعی مورداستفاده قرار گیرد. داده‌کاوی روی داده‌های پزشکی از اهمیت بالایی برخوردار است و طراحی سیستم‌های کمک تصمیم جهت یاری رساندن به پزشکان در زمینه تشخیص نوع بیماری و یا انتخاب نوع درمان مناسب، با کمک داده کاوی می‌تواند کمک شایانی در زمینه نجات جان انسان‌ها انجام دهد [۱۹].

۴- یادگیری ماشین در حوزه سلامت

یادگیری ماشین یک رشته علوم کامپیوتر است که از الگوریتم‌های کامپیوتری برای شناسایی الگوها در داده‌های بزرگ استفاده می‌کند و به پیش‌بینی نتایج مختلف کمک می‌کند. در یادگیری ماشین، ماشین می‌تواند یک مجموعه داده را تجزیه و تحلیل کند و بر اساس آنچه آموخته، تصمیم گیری کند. یادگیری ماشین نقش مهمی را در تصمیم‌گیری‌های پزشکی، شناسایی به موقع بیماران ایفا می‌کند. روش‌های یادگیری ماشین به سه گروه نظارتی(Supervised learning algorithm) و بدون نظارت (Unsupervised learning algorithm) و تقویتی(Reinforcement learning algorithm) تقسیم می‌شوند [۲۰].

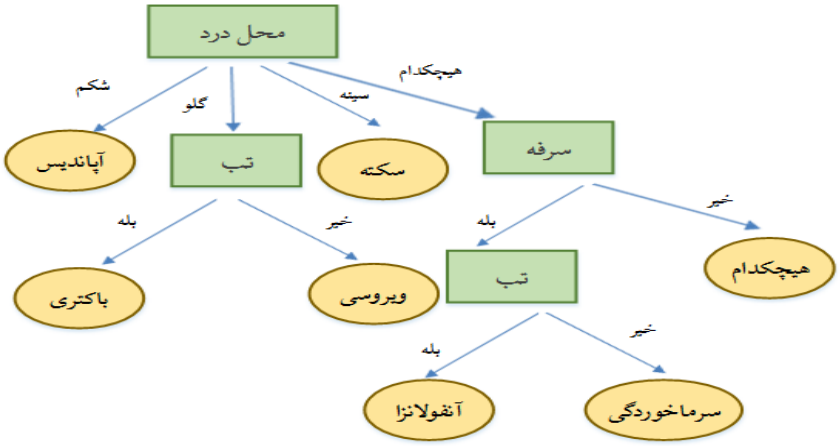
۴-۱- الگوریتم‌های نظارتی

این نوع از الگوریتم‌ها دارای دو نوع متغیرمی‌باشد؛ نوع اول که متغیرهای مستقل نامیده می‌شوند، یا چند متغیر هستند که قرار است بر اساس مقادیر آن‌ها، یک متغیر دیگر پیش‌بینی شود. نوع دوم متغیرهای وابسته یا هدف یا خروجی هستند و قرار است مقادیر آن‌ها کم این الگوریتم‌ها پیش‌بینی شود. برای این منظور باید تابعی ایجادشود که ورودی‌ها (متغیرهای مستقل) را گرفته و خروجی موردنظر (متغیر وابسته یا هدف) را تولید کند. فرآیند یافتن این تابع که در حقیقت کشف رابطه‌ای بین متغیرهای مستقل و متغیرهای وابسته است، فرآیند آموزش(Training Process) گفته می‌شود که روی داده‌های موجود اعمال می-شود و تا رسیدن به دقت لازم ادامه می‌یابد. نمونه‌هایی از این الگوریتم‌ها عبارتند از درخت‌های تصمیم (Desiccation Tree)، ماشین بردار پشتیبان (Support Vector Machine)، جنگل تصادفی (Random Forest Algorithm)، و رگرسیون (Regression Algorithm) و کلاسه‌بندی‌کننده بیزین ساده (Naïve Bayes) [۲۰].

۴-۱-۱- الگوریتم درخت تصمیم

درخت تصمیم ، یکی از ابزارهای قوی و متداول برای رده‌بندی و پیش-بینی می‌باشد. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یکسری قواعد توضیح داده می‌شود. ساختار درخت

تصمیم، یک ساختار درختی، شبیه فلوچارت است. بالاترین گره در درخت، گره ریشه است و گره‌های برگ، رده‌ها را مشخص می‌کنند [۲۰].



شکل ۲: نمونه‌ای از یک درخت تصمیم

پیدایش درخت تصمیم، شامل دو مرحله است: مرحله رشد و ایجاد درخت و مرحله هرس درخت با هدف حداقل کردن خطای پیش‌بینی [۲۰و۲۱]. تمام الگوریتم‌های ایجاد درخت، با نگرش بالا به پائین ایجاد می‌شوند. روش‌های متفاوتی برای ایجاد درخت وجود دارد. یکی از روش‌های معمول برای ایجاد درخت، انتخاب معیاری برای انشعاب گره-های بالایی به تعدادی زیرگره می‌باشد. انتخاب نقطه شکست و ایجاد انشعاب در درخت از اهمیت خاصی برخوردار است. مهمترین معیارها برای انتخاب نقطه شکست، دو معیار جینی (Gini) و انتروپی (Entropy) هستند [۲۱]. این معیارها، مبتنی بر ناخالصی هستند. همواره متغیری برای انشعاب، انتخاب می‌شود که باعث کاهش ناخالصی شود. در انتخاب نقطه شکست، متغیری که زیرگروهش به یکی از رده‌ها (برگ) تبدیل شود، اولویت دارد.

۴-۱-۲- الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان نوع خاصی از شبکه‌های عصبی هستند که بر خلاف سایر انواع شبکه عصبی مانند شبکه عصبی چندلایه (Multi . Layer Perceptrons)، شبکه عصبی پایه شعاعی(Radial basis Function) به جای کمینه کردن خطا، اقدام به کمینه کردن ریسک عملیاتی کلاسه‌بندی یا مدل‌سازی می‌کند. این ابزار، بسیار قدرتمند است و در زمینه‌های مختلفی چون کلاسه‌بندی، خوشه‌بندی و مدلسازی (رگرسیون) می‌تواند مورد استفاده قرار بگیرد یکی از الگوریتم‌های مهم از میان ماشینهای بردار پشتیبان، رگرسیون بردار پشتیبان (Support Vector Regression) می‌باشد. در آمارکلاسیک روشهای کلاسه‌بندی و رگرسیون بر پایه فرضیات محدودکننده‌ای بنا شده که در آن مدلهای توزیع احتمال و یا توابع چگالی احتمال، معلوم هستند. متأسفانه بسیاری اوقات، در عمل اطلاعات کافی درباره توزیع احتمال متغیرهای مورد مطالعه در دسترس نیست. در چنین مواقعی به روشهایی نیاز داریم که بدون دانستن توزیع احتمال، به خوبی عمل کنند. همچنین در بیشتر مطالعات با اطلاعاتی در فضاهایی با ابعاد بالا مواجه هستیم [۲۲]. برای استفاده از روش‌های کلاسیک آماری در چنین شرایطی، نیازمند نمونه-هایی با حجم بالا می‌باشیم که ممکن است در عمل فراهم کردن آن میسر نباشد. یکی از روش‌هایی که برای حل چنین مشکلاتی مورد استفاده قرار می‌گیرد، استفاده از الگوریتم‌های یادگیری ماشین است. ماشین بردار پشتیبان یک تکنیک جدید از روش‌های یادگیری ماشین است و می‌توان محبوبیت کنونی آنرا با محبوبیت شبکه‌های عصبی دردهه‌های گذشته مقایسه کرد. ماشین بردار پشتیبان در پیش‌بینی‌ها نسبت به شبکه‌های عصبی از دقت بالاتری برخوردار است. از طرفی تعیین وضعیت شبکه و قابلیت تعمیم شبکه عصبی برای وظایف مدلسازی / یادگیری هنوز به خوبی حل نشده است، در حالی که مدل ماشین بردار پشتیبان به خوبی تعمیم‌پذیر است. ویژگی مهم ماشین بردار پشتیبان این است که بر خلاف الگوریتم‌های کلاسیک و رگرسیون-های خطی که به وسیله مینیمم کردن قدر مطلق خطا یا توان دوم خطا عمل می‌کنند، آن‌ها ریسک عملیاتی را مینیمم می‌کنند. ماشین بردار پشتیبان برخلاف شبکه‌های عصبی با مشکل گیر افتادن در مینیمم‌های محلی تابع خطا مواجه نمی‌باشد . همچنین با استفاده از هسته‌های غیرخطی قادر به تصمیم‌گیری غیرخطی نیز می‌باشد. انتخاب هسته‌های مناسب برای ماشین بردار پشتیبان، منجر به برتری آن نسبت به سایر

رویکردهای مبتنی بر تصمیم‌گیری خطی شده است. مدل‌های ماشین‌ها بردار پشتیبان به دو گروه عمده مدل کلاسه‌بندی ماشین بردار پشتیبان و مدل رگرسیون ماشین بردار پشتیبان تقسیم‌بندی می‌شوند. از مدل کلاسه‌بندی ماشین بردار پشتیبان جهت حل مسائل کلاسه‌بندی داده-هایی که در کلاس‌های مختلف قرار می‌گیرند استفاده می‌گردد و مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی کاربرد دارد [۲۰و۲۱و۲۲].

۴-۱-۳- الگوریتم جنگل تصادفی

روش جنگل تصادفی یک روش غیر پارامتری و متعلق به خانواده روش-های دسته جمعی که در اواخر قرن نوزدهم از روش‌های ماشین یادگیری به دست آمد. این الگوریتم را که شامل مجموعه‌ای از درخت‌های دسته-بندی و رگرسیونی است اولین بار بریمن توسعه داد [۲۳]. روش جنگلی تصادفی از مجموعه‌ای از درخت‌ها که با بازسازی داده‌های آموزشی به کار گرفته شده است ساخته می‌شود در حالت عادی مجموعه‌ای از نمونه‌ها هستند که به شکل تصادفی با جایگزین داده‌های آموزشی اصلی شکلی می‌گیرند. ترکیب سه پارامتر در الگوریتم جنگل تصادفی ضروری است اول اینکه چه تعداد درخت باید ساخته شود دیگر اینکه چه تعداد از متغیرها در ایجاد یک گره برای شبکه شرکت می‌کنند و پارامتر سوم به اندازه گره بر می‌گردد که عمق درخت ساخته شده را نشان می‌دهد در زمان ساخته شدن یک درخت مجموعه آموزشی جدید با جایگزینی مجموعه داده‌های آموزشی اصلی ساخته می‌شود.

۴-۱-۴- الگوریتم رگرسیون

پیشگویی مقادیر پیوسته می تواند توسط تکنیک‌های آماری که رگرسیون نامیده می شود مدل سازی شوند. هدف تحلیل رگرسیون تعیین بهترین مدلی است که بتواند متغیر خروجی با متغیرهای ورودی متعدد را تعیین کند. بیشتر حالات تحلیل رگرسیون حالتی است که تعیین کننده چگونگی ارتباط متغیر Y با یک یا چند متغیر X1 , X2 Xn , ... , باشد. Y معمولا خروجی پاسخ یا متغیر وابسته نامیده می شود و X1-X ورودی‌ها، برگشت کننده ها، متغیرهای توضیحی یا متغیرهای مستقل نامیده می‌شوند. [۲۴].

۴-۱-۵- کلاسه‌بندی‌کننده بیزین ساده

کلاسه‌بندی‌کننده بیزین ساده بر اساس قانون احتمال بیز، با فرض استقلال مقادیر ویژگی‌ها بر اساس ویژگی برچسب، طراحی شده است.لغت ساده درعنوان این کلاسه‌بندی کننده به همین موضوع اشاره می‌کند. با توجه به الگوریتم ساده این کلاسه‌بندی کننده، سرعت محاسبه آن بالا است. این سادگی باعث مدیریت ساده یک پایگاه داده با تعداد زیادی ویژگی می‌شود. این کلاسه‌بندی کننده برای پیش‌بینی بر چسب یک رکورد در مقایسه با سایر کلاسه‌بندی‌کننده‌ها به داده‌های آموزشی کمتری نیاز دارد [۲۵].

۴-۲- الگوریتم‌های بدون نظارت

در این نوع از الگوریتم ها ، متغیر هدف وجود ندارد و خروجی الگوریتم، نامشخص است .الگوریتم‌های یادگیری ماشین بدون نظارت زمانی استفاده میشود که اطلاعات مورد استفاده برای آموزش، طبقه-بندی و برچسب‌گذاری نشده‌اند. الگوریتم K – نزدیکترین همسایه (K-Nearest Neighbor) و تجزیه و تحلیل مولفه اصلی (Principal component analysis) و از این دسته هستند [۲۶].

۴-۲-۱- الگوریتم K –نزدیکترین همسایه

این روش با استفاده از الگویی که از قبل یاد گرفته داده‌های آزمون را دریافت می‌نماید و داده‌های موجود و اضافه شده را با هم مقایسه می‌کند. این روش اطلاعات آموزش دارای برچسب‌گذاری صحیح به فراگیر منتقل می‌شود. که اینجا منظور از K بیشترین داده همبسته گرفته شده از مجموعه آموزش است. معمولاً از K به عنوان نماینده‌ای برای داده‌های آزمون استفاده می‌کنند [۲۶ و ۲۷].

۴-۲-۲- الگوریتم تجزیه و تحلیل مولفه اصلی

یک روش چندبخشی است که جدولی از داده‌های موجود را بررسی و تجزیه وتحلیل می‌کند و مشاهدات در آن توسط چندین متغیر وابسته کمی همبسته تعریف می‌شوند. این روش یکی از محبوب‌ترین و قدیم-یترین روش‌های تجزیه تحلیل به حساب می‌آید ایده اصلی این روش برداشت اطلاعات مهم از جدول،نشان دادن آن به عنوان مجموعه‌ای از متغیرهای متعامد است [۲۷].

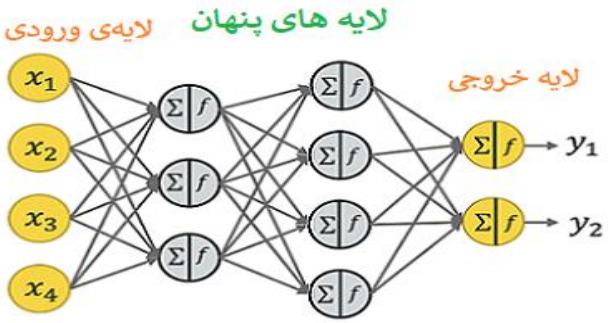
۴-۳- ا لگوریتم‌های تقویتی

نوع سوم از الگوریتم‌ها که شاید بتوان آن‌ها را در زمره الگوریتم‌های بدون نظارتی هم دسته‌بندی کرد، الگوریتم‌های تقویتی و یا یادگیری تقویتی هستند. در این نوع از الگوریتم‌ها، ی ماشین برای گرفتن یک تصمیم خاص، آموزش داده می‌شود و ماشین بر اساس موقعیت فعلی (مجموعه متغیرهای موجود) و رفتارهای مجاز (مرلا حرکت به جلو ، حرکت به عقب و ...) یک تصمیم را می‌گیرد که در دفعات اول، این تصمیم می‌تواند کاملاً تصادفی باشد و به ازای هر رفتاری که بروز می‌دهد، سیستم یک امتیاز به او می‌دهد و از روی این امتیاز، ماشین متوجه می‌شود که تصمیم درست را اتخاذ کرده است یا نه، که در دفعات بعد در آن موقعیت ، همان رفتار را تکرار کند یا رفتار دیگری را امتحان کند. با توجه به وابسته بودن حالت و رفتار فعلی به حالات و رفتارهای قبلی، فرآیند تصمیم‌گیری مارکوف، یکی از مثال‌های این گروه از الگوریتم‌ها می‌تواند باشد . الگوریتم‌های شبکه‌های عصبی چندلایه

(multilayer perceptron) هم می‌توانند ازاین دسته به حساب آیند [۲۷].

۴-۳-۱- شبکه عصبی چند لایه

شبکه عصبی مصنوعی،یک مدل محاسباتی الهام گرفته شده از ساختار و رفتار شبکه عصبی بیولوژیکی است. شبکه عصبی از یک گروه به هم پیوسته از نرون‌ها مصنوعی تشکیل شده که پردازش اطلاعات در آن با استفاده از ارتباط بین نرون‌ها انجام می‌شود. شبکه عصبی دارای معماری‌های پیاده‌سازی مختلفی است. برای پیش-بینی و کلاسه‌بندی داده‌ها، معمولا از شبکه عصبی پیش‌خور چند لایه استفاده می‌شود. این نوع شبکه عصبی، از یک لایه ورودی، و یک لایه خروجی و چندین لایه پنهان ایجاد می‌شود. اطلاعات اولیه شبکه در لایه ورودی دریافت شده و پس از محاسبات و وزن‌دهی در لایه پنهان به لایه خروجی ارسال می شود. الگوریتم‌های مختلفی برای وزن‌دهی ارتباطات بین لایه‌ها بر اساس مقادیر لایه ورودی تعریف شده است. الگوریتم انتشار، یک الگوریتم یادگیری با ناظر بوده ، که دارای دو مرحله انتشار و تغییر وزن‌هاست. این دو مرحله، آنقدر تکرار شده تا کارایی شبکه به اندازه کافی برسد. الگوریتم‌های لونبرگ_مارکوارت، گرادیانت مزدوج پل_ریبیر ، سایر الگوریتم‌های یادگیری هستند. مقداردهی اولیه وزن‌ها بصورت تصادفی است. زمان آموزش در شبکه عصبی طولانی است و بیشتر در مواردی استفاده می‌شود که زمان یادگیری، در آنها مهم نباشد. در شکل (۲) زیر یک شبکه عصبی با سه لایه نمایش داده شده است [۲۸].



شکل ۳: ساختار شبکه عصبی چند لایه

۵- الگوریتم‌های فراابتکاری در حوزه سلامت

نظام سلامت بدون شک بر اقتصاد کشور و زندگی روزمره مردم تأثیر زیادی خواهد گذاشت. داده‌های دقیق در حوزه سلامت توسط سختافزارهای جدید و مدرن جمع‌آوری می‌شوند. اما ارزیابی وضعیت داده‌های جمع‌آوری شده نیاز به ابزارهای پیش‌بینی و تشخیص قوی دارد. از این رو الگوریتم‌های داده‌کاوی همراه با فراابتکاری می-تواند یک راه‌حل مفید در ساخت مدل هوشمند سلامت باشد. از آنجا که الگوریتم‌های فراابتکاری می‌تواند برای حل مسائل داده‌کاوی و یا افزایش عملکرد و دقت مدل داده‌کاوی و کم به تجزیه و تحلیل داده‌های ناشناس در حوزه سلامت استفاده شود، در نتیجه آنها بخش مهمی از تحقیقات مدرن در حوزه سلامت می‌باشند. اگرچه الگوریتم‌های داده‌کاوی سنتی می‌توانند نتایج مفیدی در حوزه سلامت ارائه دهند اما استفاده از الگوریتم‌های فراابتکاری می‌تواند در بهبود نتایج و پیش بینی دقیق‌تر بسیار مؤثر باشد. به‌طور معمول الگوریتم‌های فرابتکاری را می‌توان به عنوان بخشی از یادگیری ماشین و محاسبات نرم در نظر گرفت. خصوصیات اصلی الگوریتم‌های فراابتکاری این است که علاوه بر ورودی و خروجی دارای اپراتورهای انتقال، ارزیابی و تعیین می‌باشند که به دفعات تا زمانیکه مطابق شرایط توقف از پیش تعیین شده و یا اینکه جستجو همگرا شود، تکرار می‌شوند. معمولا این الگوریتم‌ها در حوزه سلامت برای حل مشکلات داده‌کاوی مانند خوشه‌بندی و طبقه‌بندی اولیه داده‌های ناشناخته و قوانین مربوط به الگوها استفاده می‌شوند. الگوریتم‌های فراابتکاری در حوزه سلامت، یک راه‌حل ممکن که نیاز به روش‌های هوشند دارند را فراهم می-کنند. در مقایسه با شبکه عصبی و الگوریتم‌های دقیق، الگوریتم-های فراابتکاری معمولاً یک راه‌حل تقریبی را سریعتر پیدا می-کنند و در مقایسه با الگوریتم‌های قطعی و مبتنی بر قانون معمولا نتایج بهتری را ارائه می‌دهند. از آنجا که الگوریتم‌های فراابتکاری در مقایسه با دیگر الگوریتم‌ها قادر به ارائه‌ی راه‌حل تقریبی در زمان کمتر می‌باشند و همچنین برای ارتقاء عملکرد الگوریتم‌های داده‌کاوی استفاده می‌شوند، در سالهای اخیر برای حل مشکل درحوزه سلامت بسیار مورد توجه محققین قرارگرفته است [۲۹].

۶-مروری بر کارهای پیشین در زمینه تشخیص بیماری‌های کبد

از آنجایکه نرخ مرگ و میر ناشی از بیماری‌های کبدی بالا می‌باشد و با توجه به اینکه این دسته از بیماری‌ها در مراحل اولیه بیماری با

روشهای ساده قابل درمان هستند و با گیشت زمان و تأخیر در تشخیص بیماری امکان درمان بسیار کم می‌شود. تشخیص زودهنگام بیماری کبد از اهمیت به سزایی برخوردار است. تحقیقات متعددی برای تشخیص کبد بیمار انجام شده که هر کدام از این پژوهش‌ها از روش‌ها و الگوریتم‌های متفاوتی استفاده نموده‌اند. در این بخش به طبقه‌بندی رویکردهای جاری در تشخیص بیماری کبد که شامل روش‌های هوشمند دادهکاوی و یادگیری ماشین، الگوریتم‌ها فراابتکاری و معیارهای آن پرداخته می‌شود.

۶-۵-۱- تشخیص بیماری کبد با استفاده از الگوریتم‌های داده کاوی و یادگیری ماشین

در این روش‌های از انواع الگوریتم‌های یادگیری ماشین برای طبقه بندی بیماری استفاده می‌کنند. یادگیری ماشین زیر مجموعه‌ای از هوش-مصنوعی است. با استفاده از تکنیک‌های یادگیری ماشین، کامپیوتر، الگوهای موجود در داده‌ها را یادگرفته و می‌تواند از آن استفاده کند. در فرآیندهای یادگیری‌ماشین، داده‌ها بسیار اهمیت دارند. در واقع این داده‌ها هستند که به الگوریتم داده می‌شوند و الگوریتم از روی آنها یادگیری را انجام می‌دهد. بنابراین هرچقدر هم از الگوریتم‌های قوی استفاده شود تا زمانیکه داده‌های ناب به الگوریتم داده نشود، طبقه‌بندی پاسخی نادقیق و غیرصحیح خواهد داشت. معمولا روش‌های داده‌کاوی به منظوریافتن و استخراج اطلاعات پنهان، الگوها و روابط مشخص در حجم انبوهی از داده‌ها با هدف پیش‌بینی رویدادها و نتایج استفاده می‌شود. با داده‌کاوی می‌توان تمام داده‌های نامرتبط، بی‌فایده و تکراری را از مجموعه داده‌ها حذف کرد، موارد را شناسایی کرده و از آن‌ها برای رسیدن به اهداف از پیش تعیین شده بهره برد و به فرایند تصمیم‌گیری آگاهانه سرعت ببخشید [۳۰].

رامانا و همکاران [۳۱] در سال ۲۰۱۹ بر روی روش‌های طبقه‌بندی یادگیری ماشین از جمله J48, IBK , Bagging, Rip, Naive Bayes و شبکه پرسپترون چند لایه (multilayer perceptron) با استفاده از مجموعه داده ILPD تحقیقی صورت دادند و از نتایج بدست آمده روش Bagging دارای دقت ۶۹,۳ درصد و عملکرد بهتری نسبت به بقیه الگوریتم‌ها بود.

کبیر و همکاران [۳۲] در سال ۲۰۱۹ روشی ترکیبی با استفاده از روش تقویتی boosting (Gradient boosted machine)، جنگل تصادفی،شبکه عصبی عمیق بر روی مجموعه داده‌ها ILPD ارائه نمودند که نتایج شبیه‌سازی نشان‌دهنده این است که روش پیشنهادی نسبت به سایر روش‌ها دارای دقت ۷۱/۸۰ درصد می‌باشد.

وی و همکاران [۳۳] با ترکیب روشی برای انتخاب ویژگی‌ها و سپس طبقه‌بندی داده‌ها با استفاده از الگوریتم‌های یادگیری ماشین، مدلی مؤثر برای تشخیص بیماری کبد چرب معرفی کردند. این مدل از چهار مرحله تشکیل شده است؛ ۱- پیش پردازش داده‌ها شامل تمیز کردن داده ها، رفع داده‌های از دست رفته و تبدیل داده‌ها می‌باشد. برای این مرحله از آزمون تکنیک نمونه گیری استفاده شده است. ۲- انتخاب ویژگی‌ها: فرایند انتخاب بهترین زیر مجموعه از ویژگی‌ها برای استفاده در ساخت مدل. در این تحقیق برای انتخاب ویژگی‌ها از فرایند رتبه‌بندی اطلاعات یا همان آنتروپی استفاده شده است و ویژگی‌هایی که دارای وزن بیشتر از صفر بودند در مدل نهایی به کار برده شده است. ۳- ساختمان مدل: انتخاب یک طبقه‌بند مناسب برای پیش‌بینی بهتر. مدل‌های طبقه‌بندی برای پیش‌بینی و شناسایی بیماران کبد چرب از قبیل جنگل تصادفی، شبکه عصبی مصنوعی، نانوبیز و رگرسیون لجستیک می‌باشد. ۴- اعتبار سنجی: مجموعه داده به دو گروه آموزشی و تست (Kn-1:1) تقسیم می‌شوند ۷۰۰ بیمار دارای کبد چرب وجود داشت. ۲۲ بیمار که دارای سن کمتر از ۳۱ سال بودند، حذف شدند؛ ۱۲۳ بیمار بدلیل اطلاعات ناقص از تحقیق حذف شدند و در مجموع ۵۷۷ بیمار که با تست اولترا سونوگرافی تمام معیارها را داشتند ، مورد مطالعه قرار گرفتند. سن بیماران مبتلا به کبد چرب ۱۲/۶ ± ۵۴/۱ سن بیمارانی که کبد چرب نداشتند ۱۵/۲ ± ۴۹/۴بود. ۲۰۷ نفر از مردان دارای کبد چرب و ۶۶ نفر بدون کبد چرب بودند. نتایج این تحقیق نشان می‌دهد که مدل‌های یادگیری ماشین برای تشخیص بیماری کبد چرب مناسب می‌باشد و به ویژه مدل جنگل تصادفی برای پیش‌بینی با استفاده از حداقل متغیرهای بالینی عملکرد بهتری نسبت به مدل‌های دیگر دارد.

کومار و همکاران [۳۴] روشی را با استفاده از الگوریتم‌های K – نزدیکترین همسایه، نانوبیز، درخت تصمیم و جنگل تصادفی برای طبقه بندی و تشخیص بیماری کبد ارائه داده‌اند. نتایج آزمایشات گویا این است که الگوریتم C5.0 دارای عملکرد بهتری نسبت به دیگر الگوریتم‌ها بوده است.

پتوری و همکاران [۳۵] از روش یادگیری نیمه نظارتی مبتنی بر شبکه‌های فازی ـ عصبی برای تشخیص بیماری کبد استفاده نمودند. در این مقاله نشان داده شده که اگر نمونه‌های کمی به مجموعه داده اصلی اضافه شود مدل دوباره آموزش می‌بیند و داده‌های پرت را با مدل فازی تشخیص می‌دهد . دقت مدل ارائه شده در این تحقیق بر روی مجموعه داده ILPD دقت بین ۸۰-۷۰ درصد می‌باشد.

پورپانش و همکاران [۳۶] ترکیب سیستم فازی ـ عصبی Min-Max الگوریتم فراابتکاری طوفان مغزی (Brain storm optimization) روشی برای تشخیص بیماری کبد ارائه نمودند. از الگوریتم BSO برای انتخاب ویژگی‌ها و از سیستم فازی عصبی برای طبقه‌بندی استفاده کردند. دقت مدل ارائه شد (FMM- BSO) در این مقاله بر روی مجموعه داده BUPA دقت بین ۷۵-۶۰ درصد می‌باشد.

تارن و همکاران [۳۷] در سال ۲۰۱۹ از روش یادگیری نیمه‌نظارتی همراه با سیستم فازی ـ عصبی Min- Max روش ترکیبی را برای طبقه‌بندی و تشخیص بیماری کبد ارائه کردند. عملکرد مدل ارائه شده نسبت به بقیه مدل‌های فازی ـ عصبی بهتر بوده است.

لی و همکاران [۳۸] از ترکیب الگوریتم خوشه بندی فازی (c-Min) و سیستم فازی ـ عصبی، مجموعه داده ILPD را در دو فاز طبقه‌بندی نمودند و نشان دادند که دقت این مدل طبقه‌بند ۷۴/۸۳ درصد می‌باشد. حاقگو و همکاران [۳۹] با ترکیب شبکه عصبی پرسپترون چند لایه و مدل جنگل تصادفی به تشخیص بیماری کبد بر روی مجموعه داده BUPA پرداختند و دقت مدل ارائه شده ۸۵/۲۹ درصد می‌باشد.

خو و همکاران [۴۰] با استفاده از فرایند پردازش تصویر و مدل طبقه بندی ماشین بردار پشتیبان برای تشخیص بیماری کارسینوم هپاتوسلولا (Hepatocellular carcinoma) که از شایعترین نوع بیماری بدخیم کبد در سراسر جهان می‌باشد بر اساس تصویر التراسونوگرافی، روشی با دقت بالاتر از ۸۹/۲۵ درصد ارائه نمودند.

دانگتین و همکاران [۴۱] روش ترکیبی را با استفاده از شبکه عصبی کانولوشن و تابع SVD و مبدل گسسته موج (DWT-SVD) ارائه نمودند هدف از این روش کاهش زمان اجرا و کاهش فضای اشغال شده مربوط به تصاویر کبدی موجود در روی هارد دیس و همچنین عملکرد بهتر طبقه بند است. این مقاله از تصاویر سی تی اسکن برای طبقه‌بندی توده‌های خوشخیم و بدخیم استفاده کرده است. نتایج خروجی دقت ۹۷/۳ درصدی را نشان می‌دهد.

سابرامینیا و همکاران [۴۲] برای تشخیص کبد چرب و کبد سیروتوز (cirrhotic live) با استفاده از تصاویر رادیولوژی مدلی ارائه نمودند که در این مدل برای انتخاب ویژگی‌ها از ترکیب دیفرانسیل تکاملی با KNN (KNN-DEFS) و همین طور ترکیب دیفرانسیل تکاملی با NB-DEFS (Naïve-Bayes) و نهایتا مدل ماشین بردار پشتیبان به منظور طبقه‌بندی استفاده نمودند. با مقایسه دو روش پیشنهادی، روش ترکیبی NB-DEFS دارای عملکرد بهتری می‌باشد.

داس و همکاران [۴۳] روشی برای تشخیص ضایعات سرطانی در تصاویر ستی‌اسکن کبد ارائه نمودند. در این روش پیشنهادی ابتدا کبد با استفاده از فرایند تقسیم بندی با مارکرهایی جدا شد و در نهایت با استفاده از الگویتم گوسین (Gaussian mixture model) تجزیه گردید. پس از تقسیم‌بندی تومو، ویژگی‌های مختلف بافت استخراج شد. این ویژگی ها در سه دسته به شبکه عصبی عمیق (Deep Neural Network) داده شدند. در طی ۲۰۰ مرحله دقت طبقه‌بندی ۹۸/۳۸درصد بود. میرشا و همکاران [۴۴] یک الگوریتم ترکیبی را به منظور پایین آوردن میزان خطرات اختلات کبد ارائه نمودند. در این پژوهش از مجموعه داده مربوط به بیماران مبتلا به کبد استفاده شده است. این مجموعه داده دارای ۵ ویژگی با محدوده مقدار متفاوت است. معیار ارزیابی در این پژوهش نرخ خطا و دقت است، میزان دقت در این پژوهش برابر با ۸۶ درصد است.

مولود آبدار و همکارانش [۴۵] ازدو الگوریتم درخت تصمیم یعنی الگوریتم‌های C5.0 و CHAID برای پیش‌بینی بیماری کبد استفاده کردند. در این پژوهش، ۵۸۳ نمونه از مجموعه داده بیماری کبدی از پایگاه داده UCI در نظر گرفته شد. این مجموعه داده شامل ۴۱۶ پرونده بیماری کبد و ۱۹۷ پرونده از کبد سالم است. این مجموعه داده توسط الگوریتم های C5.0 تقویت شده و CHAID مورد تجزیه وتحلیل قرار گرفت. نتایج نشان داد که هر دو الگوریتم تاثیر قابل توجهی در پیش بینی بیماری کبد با توجه به قوانین تولید شده دارند. مقایسه ی کارایی این دو الگوریتم مشخص کرد الگوریتم C5.0 تقویت شده عملکرد بهتری نسبت به CHAID دارد و با بررسی عوامل موثر در فواصل کوچک صحت الگوریتم افزایش می‌یابد و در واقع الگوریتم C5.0 همه‌ی فاکتورها را در نظر می‌گیرد که این نشان دهنده‌ی این است که این الگوریتم از فاکتورها با احتیاط بیشتری استفاده می‌کند، در حالی که الگوریتم CHAID تنها از پنج فاکتور فوق استفاده می‌کند. بنابراین الگوریتم C5.0 با صحت ۹۳,۷۵ درصد نسبت به الگوریتمCHAID صحت ۶۵ درصد بالاتر و همچنین قوانین بیشتری دارد. یکی از ویژگی-های مهم در مورد الگوریتم C5.0، امکان اعمال روش‌های تقویتی در آن است. روش‌های تقویتی در الگوریتم C5.0 منجر به افزایش صحت و سرعت در ایجاد قوانین می‌شود. سیندوج و همکاران[۴۶] به پیش‌بینی اختلات کبد با استفاده از الگوریتم‌های درخت تصمیم J48, CARTپرداخته‌اند. در این پژوهش نویسندگان از مجموعه داده استاندارد مربوط به کبدکه شامل ۱۲ ویژگی اصلی می‌باشد، برای آموزش داده‌ها استفاده کرده‌اند. نتایج ازمایشات نشان‌دهنده این است که میزان دقت در این مدل در بهترین حالت برابر ۷۰ درصد می‌باشد. ساکینا و همکاران [۴۷] با استفاده از الگوریتم ترکیبی ژنتیک و رویکرد فازی روشی برای تشخیص اختلات کبدی ارائه نمودند. در این پژوهش قوانین ایجاد شده از نظر معیار دقت مورد بررسی قرار گرفته اند و از میان قوانین ایجاد شده ۱۲ قانون نرخ بالاتری نسبت به قوانین دیگر بدست آورده‌اند. سیندوج و همکاران [۴۸] با استفاده از تحلیل رگرسیون به بررسی اختلات کبدی پرداخته‌اند. در این مقاله، از مجموعه داده‌های اختلالات بیماران کبدی پایگاه داده UCI استفاده شده است، که شامل ۷ ویژگی می‌باشد. این مجموعه داده شامل ۳۴۵ رکورد است که از این تعداد ۲۴۲ رکورد برای آموزش استفاده شده است. نتایج آزمایشات بر روی این داده‌ها نشان‌دهنده این است میزان خطا در بهترین حالات ممکن برای رویکرد پیشنهادی برابر با ۰,۳ می‌باشد. خادیجیا و همکاران [۴۹] به پیش‌بینی بیمارهای کبد با استفاده از الگوریتم‌های یادگیری ماشین پرداخته‌اند. این پژوهش با چهار طبقه‌بندی با استفاده از جنگل‌تصادفی، نانو بینز، الگوریتم K- نزدیکترین همسایه و ماشین بردار پشتیبان بررسی شده است، مجموعه داده‌ها مورد استفاده در این پژوهش LIPD می‌باشد. نتایج شبیه‌سازی‌ها نشان‌دهنده این است الگوریتم جنگل تصادفی بهترین دقت را در میان این طبقه‌بندها با دقت ۷۷,۰۶ درصد بدست اورده است. در [۵۰] از الگوریتم ماشین بردار پشتیبان برای دسته بندی بیماران کبدی بر روی دو مجموعه داده کبد استفاده شده است که با ترکیب ویژگی های مختلف مانند SGOT،SGPT و Alkaline Phosphatesدر مجموعه داده‌ها، دقت، نرخ خطا، حساسیت، شیوع و

مشخصه را اندازه‌گیری کرده‌اند که نتایج در ۶ ویژگی انتخاب شده در مجموعه داده ILPD بهتر از مجموعه داده BUPA نشان داده شده است که دلیل این تفاوت، وجود تعداد ویژگی‌های مفید بیشتر در مجموعه داده ILPD دانسته‌اند.

در [۵۱] یک مطالعه تطبیقی بین مجموعه داده ILPD و مجموعه داده BUPA از مخزن داده UCI انجام داده‌اند که در تمام الگوریتم‌های انتخابی، مجموعه داده ILPD دارای عملکرد بهتری نسب به مجموعه داده BUPA نشان داده است که در بین دسته‌بندهای انتخابی، دسته بند KNN بهترین نتیجه را بر روی مجموعه داده ILPD با ترکیب تمام ویژگی‌ها نشان داده است که برای تجزیه و تحلیل تفاوت‌ها در دو مجموعه داده از روش‌های تحلیل واریانس یا ANOVA و تحلیل واریانس چند متغیره یا MANOVA استفاده کرده‌اند. در آمار ANOVAمجموعه ای از مدل‌های آماری است که به منظور تجزیه و تحلیل تفاوت بین گروه ها استفاده می‌شود و MANOVA یک روش برای مقایسه میانگین نمونه‌های چند متغیره می‌باشد و زمانی از روش چند متغیره استفاده می‌شود که دو یا تعداد بیشتری متغیرهای وابسته وجود داشته باشد.

در [۵۲] از الگوریتم‌های دسته‌بندی چون دسته بند بیز ساده، ، C4.5شبکه عصبی پس‌انتشار، ماشین بردار پشتیبان برای دسته‌بندی بیماری کبد استفاده کرده‌اند که نتایج را بر اساس چهار معیار دقت، صحت، حساسیت و مشخصه ارزیابی کرده‌اند که در مجموعه داده AP نتایج بهتری نسبت به مجموعه داده BUPA نشان داده است که دلیل بر وجود ویژگی‌های مفید موجود در این مجموعه داده دانسته‌اند و الگوریتم های KNNو پس انتشار و ماشین بردار پشتیبان نتایج بهتری با تمام ترکیبات ویژگی‌ها نشان داده است.

در [۵۳] یک رویکرد شبکه عصبی احتمالی برای دسته‌بندی داده‌های جمع‌آوری شده از مناطق ساحل شمالی هند با استفاده از نرافزار متلب ارائه شده است. در این مقاله، روش‌هایی مانند ماتریس اغتشاش، مساحت زیر منحنی، بهترین عملکرد ارزیابی، مقدار R توپولوژی، SOM موفقیت، اتصالات همسایه ،SOM فاصله وزن همسایه، موقعیت وزن SOM با استفاده از نرم افزار متلب بر روی مجموعه داده مناطق ساحل شمالی تجزیه و تحلیل شده‌اند. روش شبکه عصبی احتمالی، اعتبار سنجی داده ها را با دسته بندی ۱۰۰ درصد درست ارائه داد. این رویکرد احتمالی نتایج خوبی برای مجموعه داده های سرطان نیز ارائه داد. در [۵۴] یک مدل ترکیبی هوشمند برای تعیین انواع بیماری کبد بر اساس ادغام شبکه‌های عصبی مصنوعی، فرآیند سلسله مراتبی تجزیه و تحلیل و روش‌های استدلال مبتنی بر حالت ارائه شده است. داده‌های استفاده شده در این تحقیق، از ۵۱۰ بیمار سرپایی در مرکز پزشکی تایوان در طول یکسال در سال های ۲۰۰۵ تا ۲۰۰۹ جمع آوری شده که ۲۱۰ نفر سالم و ۳۰۰ نفر بیمار کبدی (هپاتیت مزمن، هپاتیت الکلی، سیروز کبد، هپاتی B و غیره) تشخیص داده شده است.

در [۵۵] مدل استدلال مبتنی بر علت یا CBR استفاده شده است. در این مدل چندین روش دسته‌بندی رایج از تکنیک‌های داده کاوی از جمله شبکه عصبی پس انتشار، درخت دسته بندی و رگرسیون، رگرسیون لجستیک و تجزیه و تحلیل تبعیض آمیز که برای ساخت یک مدل موثر در تشخیص زودهنگام بیماری کبد به منظور افزایش دقت دسته بندی یکپارچه شده‌اند در مقایسه مدل تکی، عملکرد کلی دو نکنیک BPN و CBR بهترین بوده است. برای بالا بردن عملکرد تشخیص، CBR با روش‌های دیگر ترکیب شده و نتایج نشان می‌دهد که دقت و حساسیت هر یک از مدل ترکیبی اضافه شده به CBR بالاتر از دیگر مدل های تکی می‌باشد. از بین تمام مدل‌های ترکیبی اضافه شده به CBRروش BPN – CBR دارای بالاترین دقت و حساسیت و مشخصه می باشد.

در [۵۶] یک مدل ترکیبی با ادغام روش خوشه‌بندی داده مبتنی بر مورد و یک درخت تصمیم فازی برای دسته بندی داده‌های سرطان سینه و اختلالات کبدی پیشنهاد داده‌اند. مجموعه داده اختلالات کبدی مورد استفاده، شامل ۳۴۵ مورد داده و ۵ ویژگی که از آزمایش خون به دست می‌آید، بوده است. میزان صحت روش پیشنهادی، برای اختلالات کبدی ۸۶,۱ درصد بوده است

در [۵۷] یک سیستم با استفاده از دسته‌بندی نظارت شده برای نظارت بر فاکتورهای خطر، برای بیماری قلبی عروقی/عروق مغزی، دیابت، اختلالات کبدی، بیماری کلیوی و فشار خون بالا پیشنهاد نموده‌اند که پس از آن، این فاکتورهای خطر مهم برای پیاده‌سازی دسته‌بندی و ساخت معیارهای هشدار دهنده، استفاده شوند. در این مطالعه رده بند غیر پارامتری K– نزدیکترین همسایگان مورد استفاده قرار گرفته است تعداد ۲۷۲۴ مورد در تایوان برای ارزیابی بیماری‌ها مورد استفاده قرار گرفت. در نهایت، میزان صحت، به دست آمده برای فرد دارای فشار خون و بیماری قلبی برابر با ۷۹,۳۶درصد برای فشار خون و اختلالات کبدی برابر با ۶۵,۳۹ درصدبوده است.

در [۵۸] از الگوریتم‌های طبقه‌بندی برای پیش بینی بیماری‌های کبدی استفاده شده است. در این تحقیق از الگوریتم‌های SVM و Naïve Bayes استفاده گردید. کارایی این دو الگوریتم از لحاظ زمان اجرا و دقت طبقه‌بندی با هم مقایسه شدند. نتایج مطالعات نشان می‌دهد که SVM دارای دقت بیشتری برای پیش بینی بیماری های کبدی است. جین و همکاران [۵۹] یک ارزیابی از نتایج آنالیزهای انجام شده خود با استفاده از الگوریتم‌های مختلف دسته بندی بر روی مجموعه داده‌های جمع آوری شده از بیماران کبدی، ارائه دادند. در این ارزیابی آن‌ها الگوریتم‌های دسته‌بندی (همانند درخت تصمیم، شبکه عصبی چند لایه، نانو بینز، KNN) را بر روی داده های خود اعمال نمودند. در ادمه دو الگوریتم جنگل تصادفی و الگوریتم لجستیک هم بر روی داده‌های انتخابی اعمال شدند و از جهات مختلفی از قبیل حساسیت (Sensitivity)، خاص بودن (recall) و دقت (Accuracy) این الگوریتم‌های اعمال شده مورد بررسی و ارزیابی قرار گرفته شدند. نتایج

بدست آمده از این بررسی‌ها نشان دادند که از جنبه Precisionالگوریتم نانو بینز نتیجه قابل قبول‌تر در مقایسه با بقیه الگوریتم‌های اعمال شده، دارا می‌باشد. ولی از جنبه های دیگر شامل حساسیت، الگوریتم جنگل تصادفی و لجستیک دارای نتایج بهتری در مقایسه با دیگر الگوریتم‌های اعمال شده می‌باشد.

نرجاجی و همکاران [۶۰] از روش‌های مختلف داده کاوی برای رده‌بندی بیماران کبدی و راه‌های تشخیص آن‌ها از افراد سالم استفاده کردند. چهار الگوریتم طبقه بندی Random Forest، Bagging، Naïve Bayesو R SVM بر روی داده‌های موجود، اعمال شدند. همچنین برای ارتقای دقت رده‌بندی انجام شده، یک مدل هیبریدی با نام Neuro SVMبا استفاده از SVM و شبکه‌های عصبی مصنوعی توسط این محققین پیشنهاد گردید. نتایج بررسی‌ها و ارزیابی‌های انجام شده نشان داد که دقت پیشگویی مدل پیشنهادی ۹۸،۸۳ درصد بود.

ویجایارانی و همکاران [۶۱] سعی کردند تا از الگوریتم‌های رده‌بندی برای پیش‌بینی بیماری‌های کبدی استفاده کنند. در این تحقیق، از الگوریتم‌های نانوبیز و ماشین بردار پشتیبان استفاده شد. کارایی این دو الگوریتم از لحاظ زمان اجرا ودقت طبقه‌بندی باهم مقایسه گردید. نتایج ارزیابی نتایج بدست آمده نشان داد که الگوریتم ماشین بردار پشتیبان دارای دقت بیشتری برای پیش بینی بیماری‌های کبدی می باشد.

آلفیشارین و همکاران [۶۲] سعی کردند تا ده خصلت را انتخاب کرده و با استفاده از الگوریتم‌هایی همانند درخت تصمیم، نانو بیز و NBTree اقدام به شناسایی انواع مختلفی از بیماران مبتلا به مشکلات کبدی نمایند. نتایج بررسی‌های این محققین، نشان داد که الگوریتم NBTreeدارای دقت بالاتری بوده و الگوریتم نانو بیز دارای کمترین زمان محاسباتی، در بین سایر الگوریتم‌های استفاده شده می باشد.

راجیسواری و همکاران [۶۳] داده‌های جمع‌آوری شده از انبار داده‌ای UCIرا رده بندی کردند مجموعه آموزشی ایجاد شده از کل مجموعه داده‌های جمع آوری شده از UCI شامل ۳۴۵ نمونه با ۷ ویژگی (خصلت) بود. تمام نمونه‌های انتخابی در دو گروه بر اساس نوع گروه خونی رده‌بندی شد. نتایج مطالعات این محققین نشان دادکه نوع گروه خونی عامل مهمی در ایجاد اختلالات کبدی ناشی از مصرف زیاد الکل می‌تواند محسوب گردد.

هونیتاسوگ و همکاران [۶۴] برای جبران کمبود داده‌های موثر، در کلاس‌های اقلیت یک روش مبتنی بر Over- Sampling را پیشنهاد کردند که در آن به منظور تشخیص بیماری‌های کبدی از تکنیک درخت تصمیم استفاده شد. نتایج بدست آمده از اعمال انواع الگوریتم‌های درخت تصمیم شامل، C4.5 و CART نشان داد که ترکیبی از این الگوریتم‌ها می‌تواند دقت و اعتبار متد پیشنهادی را به افزایش دهد.

بابو و همکاران [۶۵] کارایی الگوریتم سلسله مراتبی خوشه‌بندی را بر روی مجموعه داده بدست آمده از LIPD مورد بررسی و آنالیز قرار دادند. نتایج بدست آمده از این آنالیزها نشان دادند که تکنیک خوشه بندی، اثرات رضایت بخشی در تشخیص‌های پزشکی به خصوص در تشخیص انواع مختلف بیماری‌های کبدی می‌تواند داشته باشد.

۶-۵-۲- تشخیص بیماری کبد با استفاده از الگوریتم‌های فراابتکاری

الگوریتم‌های فراابتکاری طیف وسیعی از الگوریتم‌های مدرن سطح بالا را شامل می‌شود. الگوریتم‌هایی که از طبیعت نشأت گرفته‌اند مانند بهینه سازی ازدحام ذرات((Particle Swarm Optimization)، الگوریتم‌های تکاملی(Evolutionary Algorithms) ازجمله الگوریتم ژنتیک (Genetic Algorithms)، الگوریتم کلونی زنبور عسل (Artificial bee colony algorithm). دو مؤلفه مهم در الگوریتم‌های فرا ابتکاری، افزونگی و تنوع می‌باشد. تعادل بین این دو مؤلفه الگوریتم‌های مفید و کارآمدی را ایجاد می‌کند در الگوریتم‌های فراابتکاری کل فضای راه حل جستجو و راه‌حل‌های متنوع و نزدیک به بهینه ارائه می‌شود و به همین دلیل پیچیدگی زمان در این الگوریتم‌ها کاهش می‌یابد [۴۰].

راجاتی و همکاران [۶۶] برای طبقه‌بندی بیماری‌های کبد با استفاده از تصاویر التراسونوگرافی براساس الگوریتم بهینه‌سازی نهنگ نشان دادند که روش ارائه شده نسبت به مدل‌های طبقه‌بندی یادگیری ماشین بردار پشتیبان، K-نزدیک‌ترین همسایه و جنگل تصادفی دارای عملکرد بهتری می‌باشد.

آناد و [۳] با استفاده از شبکه عصبی مصنوعی رو به جلو برای طبقه-بندی و الگوریتم ازدحام ذرات (Particle Swarm Optimization) برای انتخاب ویژگی‌های بهینه، مدلی ترکیبی ارائه نمودند. در این مدل اتصال بین دونورون دارای وزنی است که در صورت مثبت بودن نشان‌دهنده تقویت نورون و منفی بودن مهار کردن نورون می‌باشد. روش ارائه شده بر روی مجموعه داده ILPD پیاده سازی شده است. نتایج دقت ۹۶/۲۵ درصدی را نشان می‌دهد.

تاشی و همکاران [۱] با استفاده از الگوریتم کلونی مورچه و مدل‌های طبقه بندی یادگیری ماشین مدلی ترکیبی ارائه نمودند. نتایج شبیه‌سازی نشان می‌دهد که مدل ترکیبی کلونی مورچه و ماشین بردار پشتیبان دارای عملکرد بهتری می‌باشد.

در [۶۶] یک سیستم تشخیص خودکار بیماری کبد با رویکرد ترکیبی دو الگوریتم ژنتیک و سیستم ایمنی مصنوعی طراحی کرده‌اند که روش یادگیری سیستم مطابق با الگوریتم ژنتیک می‌باشد که بهترین راه حل را از اطلاعات مشخصه‌های محلی در مسائل پیچیده پیدا می‌کند. این سیستم توسط دو مجموعه داده ILPD و BUPA ارزیابی شده که بالاترین دقت پیش‌بینی با دسته‌بندی با نظارت داده است. این سیستم هوشمند محاسباتی به طور کلی می‌تواند در زمینه های دیگر مانند تشخیص الگو و غیره استفاده شود.

در [۶۷] روشی مبتنی بر الگوریتم بهینه‌سازی گرگ خاکستری برای پیش‌بینی سرطان کبد ارائه شده است. این رویکرد پیشنهادی با استفاده سه مجموعه داده بیماران مزمن کلیوی، سرطان پستان ویسکانسین و بیماری‌های کبد موردآزمایش قرار گرفته است. نتایج ارزیابی‌های برای مجموعه داده‌های بیماران کبدی نشان‌دهنده این است که رویکرد پیشنهادی دارای دقت ۹۱ درصد می‌باشد.

در ادامه در جدول (۱) مقایسه‌های تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در تشخیص بیماری‌های کبد از نظر سه معیار دقت مشخصه و ویژگی مورد مقایسه و ارزیابی گرفتند و روابط این معیارها ارزیابی به صورت زیر می‌باشد:

دقت: مشهورترین و عمومی‌ترین معیار برای تعیین کارایی یک الگوریتم طبقه‌بندی است و عبارت است از درصدی از موارد که مدل آن‌ها را به درستی دسته‌بندی کرده است.

$$acc = \frac{tp+tn}{tp+tn+fp+fn} \times 100 \qquad (۱)$$

حساسیت: عبارت است از درصدی از موارد مثبت که مدل آن‌ها را به درستی به عنوان مثبت دسته‌بندی کرده است. بعضی مواقع نرخ مثبت واقعی (TPR18) و یا Recall نیز گفته می‌شود.

$$sensitivity = \frac{tp}{tp+fn} \times 100 \qquad (۲)$$

مشخصه: مشخصه یا نرخ منفی واقعی (TNR19) عبارت است از درصدی از موارد مثبت که مدل آن‌ها را به درستی به عنوان منفی دسته‌بندی کرده است.

$$specifcity = \frac{tn}{tn+fp} \times 100 \qquad (۳)$$

جدول (۱): مقایسه‌های تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در تشخیص بیماری‌های کبد از

مرجع	سال انتشار	تکنیک مورد استفاده	مجموعه داده	ابزار	دقت	حساسیت	مشخصه
[۳۱]	۲۰۱۹	Bagging	ILPD	MATLAB	۶۹/۳ درصد	۶۹/۳ درصد	۶۹/۳ درصد
[۳۲]	۲۰۱۹	ترکیب روش تقویتی boosting و جنگل تصادفی و شبکه عصبی عمیق	ILPD	MATLAB	۷۱/۸ درصد	۶۹/۳ درصد	۶۹/۳ درصد
[۳۳]	۲۰۱۹	انتخاب ویژگی‌ها به روش وزن‌دهی آنتروپی و طبقه‌بندی با جنگل تصادفی	داده‌های التراسونوگرافی	MATLAB	۸۶/۳۵درصد	۷۰/۵۹ درصد	۷۲/۲۹ درصد
[۳۴]	۲۰۱۸	C5.0 براساس روش Boosting	ILPD	MATLAB	۷۵/۱۹ درصد	۹۰/۷۲ درصد	۷۸/۵۷ درصد
]h[۳۵	۲۰۱۹	روش یادگیری نیمه نظارتی مبتنی بر شبکه های فازی ـ عصبی	ILPD	MATLAB	۷۰-۸۰ درصد	-	-
[۳۶]	۲۰۱۹	ترکیب سیستم فازی ـ عصبی min- max و الگوریتم فراابتکاری طوفان مغزی	BUPA	MATLAB	۶۰-۷۵ درصد	-	-
[۳۷]	۲۰۱۹	وش یادگیری نیمه نظارتی همراه با سیستم فازی ـ عصبی min- max	ILPD	MATLAB	۷۰/۵۷ درصد	-	-
[۳۸]	۲۰۱۹	ترکیب الگوریتم خوشه بندی فازی (c-mean) و سیستم فازی ـ عصبی	ILPD	Phayton	۷۴/۸۳ درصد	۸۰/۰۶ درصد	۵۴/۳۷ درصد
[۳۹]	۲۰۱۹	ترکیب شبکه عصبی پرسپترون چند لایه و جنگل تصادفی	BUPA	MATLAB	۸۵/۲۹ درصد	۸۰ درصد	۸۹/۴۷ درصد
[۶۶]	۲۰۱۹	الگوریتم بهینه سازی نهنگ	تصاویر التراسونوگرافی	MATLAB	۹۸ درصد	۹۶ درصد	۹۳ درصد
[۱]		ترکیب الگوریتم کلونی مورچه و ماشین بردار پشتیبان	BUPA	-	-	-	-
[۳]	۲۰۱۹	ترکیب شبکه عصبی مصنوعی رو به جلو برای طبقه‌بندی و الگوریتم ازدحام ذرات برای انتخاب ویژگی‌های بهینه	ILPD	MATLAB	۹۵/۴۹ درصد	۹۶/۲۵ درصد	۹۶/۲۵ درصد

[۴۰]	۲۰۱۹	فرایند پردازش تصویر و مدل طبقه بندی ماشین بردار پشتیبان	تصاویر التراسونوگرافی	MATLAB	۸۹/۲۵ درصد	-	-
[۴۱]	۲۰۱۹	شبکه عصبی کانولوشن و تابع SVD و مبدل گسسته موجک	تصاویر التراسونوگرافی	MATLAB	۹۷/۳ درصد	۹۶/۴ درصد	۹۸/۲ درصد
[۴۲]	۲۰۱۹	ترکیب دیفرانسیل تکاملی با نانو بیز و مدل طبقه‌بندی ماشین بردار پشتیبان	تصاویر التراسونوگرافی	MATLAB	۹۷/۹ درصد	-	-
[۴۳]	۲۰۱۹	ترکیب گوسین و شبکه عصبی عمیق	تصاویر التراسونوگرافی	Phayton	۹۸/۳۸ درصد	۱۰۰ درصد	۹۷/۷۲ درصد

نظر معیار دقت، مشخصه و حساسیت

در جدول(۲) مقایسه‌های تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در تشخیص انواع بیماری‌های کبد آمده است.

جدول (۲): مقایسه‌های تکنیک‌های یادگیری ماشین و الگوریتم‌های فراابتکاری در تشخیص انواع بیماری‌های کبد

مرجع	سال انتشار	نوع بیماری	تکنیک مورد استفاده	مزایا	معایب
[۳۹]	۲۰۱۸	سرطان کبد	الگوریتم جنگل تصادفی و شبکه‌های عصبی مصنوعی	استفاده ازالگوریتم شبکه عصبی	پیچیدگی بالا
[۴۵]	۲۰۱۷	اختلالات کبد	الگوریتم‌های C5.0 وCHAID	استفاده از یک مجموعه داده جامع و استاندارد	پیچیدگی بالا
[۴۶]	۲۰۱۶	اختلالات کبد	الگوریتم ترکیبی ژنتیک و رویکرد فازی	آموزش داده‌ها با استفاده از kolfd	پایین بودن میزان دقت
[۴۷]	۲۰۱۹	اختلالات کبد	استفاده از سیستم‌های کامپیوتری	بکارگیری رویکرد فازی	پیچیدگی محاسبات بالا
[۴۸]	۲۰۱۹	اختلالات کبد	تحلیل رگرسیون	سادگی در روش پیشنهادی	-
[۴۹]	۲۰۲۰	سرطان کبد	الگوریتم جنگل تصادفی	استفاده از الگوریتم‌های ساده	عدم استفاده از الگوریتم انتخاب ویژگی و استفاده از ویژگی غیرضروری
[۵۹]	۲۰۲۰	سرطان کبد	الگوریتم بهینه‌سازی گرگ خاکستری	استفاده از الگوریتم تکاملی به منظورانتخاب ویژگی-های برتر	پیچیدگی محاسبات بالا
[۳۱]	۲۰۱۹	بیماری‌های کبد	Bagging	عملکرد بهتری نسبت به الگوریتم‌های J48 IBK , Bagging, Rip, Naive Bayes و شبکه پرسپترون چند لایه (multilayer perceptron)	دقت پایین
[۳۲]	۲۰۱۹	بیماری‌های کبد	ترکیب روش تقویتی boosting و جنگل تصادفی و شبکه عصبی عمیق	دقت بالاتر	زمان حافظه مصرفی بالا و دقت پایین
[۳۳]	۲۰۱۹	بیماری‌های کبد	انتخاب ویژگی‌ها به روش وزن‌دهی آنتروپی و طبقه‌بندی با جنگل تصادفی	عملکرد بهتر جنگل تصادفی نسبت به مدل‌های دیگر	هزینه‌بر و زمان و حافظه مصرفی بالا
[۳۴]	۲۰۱۸	بیماری کبد چرب	C5.0 براساس روش Boosting	عملکرد بهتر C5.0 نسبت به سایر روش‌های پیشین	دقت پایین
[۳۵]	۲۰۱۹	بیماری‌های کبد	روش یادگیری نیمه نظارتی مبتنی بر شبکه های فازی ـ عصبی	تکرار فرایند آموزش برای داده‌های جدید اضافه شده	حافظه مصرفی بالا و دقت پایین
[۳۶]	۲۰۱۹	بیماری‌های کبد	ترکیب سیستم فازی ـ عصبی min- max و الگوریتم فراابتکاری طوفان مغزی	-	دقت پایین
[۳۷]	۲۰۱۹	بیماری‌های کبد	وش یادگیری نیمه نظارتی همراه با سیستم فازی ـ عصبی min- max	-	دقت پایین
[۳۸]	۲۰۱۹	بیماری‌های کبد	ترکیب الگوریتم خوشه بندی فازی (c-mean) و سیستم فازی ـ عصبی	-	دقت پایین
[۳۹]	۲۰۱۹	بیماری‌های کبد	ترکیب شبکه عصبی پرسپترون چند لایه و جنگل تصادفی	-	دقت پایین
[۶۶]	۲۰۱۹	بیماری‌های کبد	الگوریتم بهینه سازی نهنگ	-	حافظه مصرفی بالا و زمان اجرای بالا
[۱]		بیماری‌های کبد	ترکیب الگوریتم کلونی مورچه و ماشین بردار پشتیبان	-	حافظه مصرفی بالا
[۳]	۲۰۱۹	اختلالات کبدی	ترکیب شبکه عصبی مصنوعی رو به جلو برای طبقه‌بندی و الگوریتم ازدحام ذرات برای انتخاب ویژگی‌های بهینه	-	حافظه مصرفی بالا
[۴۰]	۲۰۱۹	بیماری‌های کبد	فرایند پردازش تصویر و مدل طبقه بندی ماشین بردار پشتیبان	استفاده از فرایند پردازش تصویر	پر هزینه
[۴۱]	۲۰۱۹	بیمارهای بدخیم کبد	شبکه عصبی کانولوشن و تابع SVD و مبدل گسسته موجک	کاهش زمان اجراو کاهش فضای اشغال شده	پر هزینه و زمان حافظه مصرفی بالا

[۴۲]	۲۰۱۹	سیتروز کبدی	ترکیب دیفرانسیل تکاملی با نانو بیز و مدل طبقه‌بندی ماشین بردار پشتیبان	-	پر هزینه
[۴۳]	۲۰۱۹	تشخیص ضایعات سرطانی کبد	ترکیب گوسین و شبکه عصبی عمیق	-	پر هزینه

۷- نتیجه‌گیری و کارهای آینده

سلامت هر جامعه مهمترین بحث در پیشرفت و ارتقاء آن جامعه به حساب می‌آید و همه فعالیت‌های انسانی بر این اصل استوار است. با توجه به گسترش تکنولوژی، همه جوامع سعی در بکارگیری فناوری‌های جدید در همه حوزه‌های زندگی انسانی دارند و حوزه پزشکی نیز از این قاعده مستثنی نیست. در سال‌های اخیر پیشرفت‌های قابل توجهی در روش‌ها و ابزارهای هوشمند برای علم پزشکی صورت گرفته است و بکارگیری این ابزارهای هوشمند در تشخیص و کاهش اشتباهات پزشکی، خسارت جانی و مالی کمک شایانی به جامعه پزشکی کرده است. مطالعه روش‌های تشخیص هوشمند نشان می‌دهد که آنها قادر هستند دقت تشخیص را بهبود داده و موارد مشکوک از دست رفته به دلیل خستگی یا بی‌تجربگی فرد خبره را کاهش دهند و در نتیجه اشتباهات تشخیصی ناشی از تنوع زیاد بیماری ها را نیز به حداقل برسانند. از اینرو، استفاده از سیستم‌های هوشمند در پزشکی و ارائه روش‌های جدید در این حوزه مورد توجه پژوهشگران قرار گرفته است و مطالعات و تحقیقات زیادی پیرامون آن انجام می‌گیرد. یکی از مکانیزم-های پرکاربرد در سیستم‌های هوشمند استفاده از مکانیزم داده کاوی است. داده‌کاوی می‌تواند به عنوان ابزار موثری در پیدا کردن رابطه مخفی بین داده بیماران باشد. بدون شک یکی از مهمترین مشکلات در تشخیص بیماری‌های پنهان همچون بیماری مزمن کبدی این است که این بیماری در بسیاری از موارد بدون علائم خاصی در بدن فرد وجود دارد و فرد زمانی از بیماری خود آگاه می‌شود که این بیماری پیشرفت بسیار زیادی را کرده باشد. در این مقاله انواع الگوریتم‌های مبتنی برداد-ه‌کاوی و تأثیر آن‌ها برای پیش‌بینی بیمارهای کبد مورد بررسی قرار گرفت. همچنین چندین روش انتخاب ویژگی برای تشخیص این بیماری موردمطالعه قرار گرفت. نتایج نشان داد که الگوریتم جنگل تصادفی، شبکه‌های فازی ـ عصبی و الگوریتم بردار پشتیبان بیشترین کاربرد را در پیش‌بینی بیمارهای کبد را داشته‌اند و همچنین مجموعه داده‌هایILPD به دلیل دسترسی آسان‌تر و داشتن نویز کمتر به عنوان پرکاربردترین مجموعه داده آموزشی و محیط شبیه‌ساز متلب [۶۸] به دلیل سادگی و قابل فهم بودن به عنوان قوی‌ترین ابزار در زمینه تحقیقاتی بیماری‌های کبدی [۶۹] به شمار می‌روند. با توجه به مطالعات انجام شده و رشد سریع روش‌های دسته‌بندی در کارهای آتی می‌توان با استفاده از روش‌های ترکیبی دقت پیش‌بینی را نسبت به روش‌های موجود افزایش داد.

تعارض منافع

«هیچ‌گونه تعارض منافع توسط نویسندگان بیان نشده است»

منابع و مآخذ

[1] Al-Tashi Q, Rais H, Abdulkadir SJ. Hybrid swarm intelligence algorithms with ensemble machine learning for medical diagnosis. In2018 4th international conference on computer and information sciences (ICCOINS) 2018 Aug 13 (pp. 1-6). IEEE.

[2] Singh A, Pandey B. Intelligent techniques and applications in liver disorders: a survey. International Journal of Biomedical Engineering and Technology. 2014 Jan 1;16(1):27-70.

[3] Anand L, Syed Ibrahim SP. HANN: a hybrid model for liver syndrome classification by feature assortment optimization. Journal of medical systems. 2018 Nov;42:1-1.

[4] Sung S, Kim J, Jung Y. Liver-derived exosomes and their implications in liver pathobiology. International journal of molecular sciences. 2018 Nov 22;19(12):3715.

[5] Shaheamlung G, Kaur H, Kaur M. A Survey on machine learning techniques for the diagnosis of liver disease. In2020 International Conference on Intelligent Engineering and Management (ICIEM) 2020 Jun 17 (pp. 337-341). IEEE.

[6] Itani S, Lecron F, Fortemps P. Specifics of medical data mining for diagnosis aid: A survey. Expert systems with applications. 2019 Mar 15;118:300-14.

[7] Arbain AN, Balakrishnan BY. A comparison of data mining algorithms for liver disease prediction on imbalanced data. International Journal of Data Science and Advanced Analytics. 2019 Feb 9;1(1):1-1.

[8] Tanwar N, Rahman KF. Machine Learning in liver disease diagnosis: Current progress and future opportunities. InIOP Conference Series: Materials Science and Engineering 2021 (Vol. 1022, No. 1, p. 012029). IOP Publishing.

[9] ÇAĞLAR MF, ÇETİŞLİ B, Toprak IB. Automatic recognition of Parkinson's disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier. Mühendislik Bilimleri ve Tasarım Dergisi. 2010 Dec 1;1(2):59-64.

[10] Gil D, Johnsson M. Diagnosing Parkinson by using artificial neural networks and support vector machines. Global Journal of Computer Science and Technology. 2009;9(4):63-71.

[11] Kuatbayeva A, Izteleuov N, Kabdoldin A, Abdyzhalilova R. Data mining models for healthcare. Advanced technologies and computer science. 2020 Dec 7(3):11-7.

[12] Ristoski P, Paulheim H. Semantic Web in data mining and knowledge discovery: A comprehensive survey. Journal of Web Semantics. 2016 Jan 1;36:1-22.

[13] Ogundele IO, Popoola OL, Oyesola OO, Orija KT. A review on data mining in healthcare. International Journal of Advanced Research in Computer Engineering and Technology (IJARCET). 2018 Sep;7:698-704.

[14] Ray R. Advances in data mining: Healthcare applications. International Research Journal of Engineering and Technology (IRJET). 2018 Mar;5(03):2395-0056.

[15] Diwani S, Mishol S, Kayange DS, Machuve D, Sam A. Overview applications of data mining in health care: The case study of Arusha region. International journal of computational engineering research. 2013 Aug 8;3(8):73-7.

[16] Amani FA, Fadlalla AM. Data mining applications in accounting: A review of the literature and organizing framework. International Journal of Accounting Information Systems. 2017 Feb 1;24:32-58.

[17] Birjandi SM, Khasteh SH. A survey on data mining techniques used in medicine. Journal of diabetes & metabolic disorders. 2021 Dec;20(2):2055-71.

[18] Gogi VJ, Vijayalakshmi MN. Prognosis of liver disease: Using Machine Learning algorithms. In2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE) 2018 Jul 27 (pp. 875-879). IEEE.

[19] Singh A, Pandey B. Liver disorder diagnosis using linear, nonlinear and decision tree classification algorithms. IJET. 2016;8(5):2059-69.

[20] Buskirk TD, Kirchner A, Eck A, Signorino CS. An introduction to machine learning methods for survey researchers. Survey Practice. 2018 Jan 2;11(1).

[21] Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet]. 2020 Jan;9(1):381-6.

[22] Tahasamadsoltaniheris M, Mahmoodvand Z, Zolnoori M. Intelligent diagnosis of Asthma using machine learning algorithms. International Research Journal of Applied and Basic Sciences. 2013;5(1):140-5.

[23] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal. 2017 Jan 1;15:104-16.

[24] Stephenson N, Shane E, Chase J, Rowland J, Ries D, Justice N, Zhang J, Chan L, Cao R. Survey of machine learning techniques in drug discovery. Current drug metabolism. 2019 Mar 1;20(3):185-93.

[25] Qayyum A, Qadir J, Bilal M, Al-Fuqaha A. Secure and robust machine learning for healthcare: A survey. IEEE Reviews in Biomedical Engineering. 2020 Jul 31;14:156-80.

[26] Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making. 2019 Dec;19(1):1-6.

[27] Scruggs SB, Watson K, Su AI, Hermjakob H, Yates III JR, Lindsey ML, Ping P. Harnessing the heart of big data. Circulation research. 2015 Mar 27;116(7):1115-9.

[28] Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: an overview. InJournal of physics: conference series 2018 Nov (Vol. 1142, p. 012012). IOP Publishing.

[29] Tsai CW, Chiang MC, Ksentini A, Chen M. Metaheuristic algorithms for healthcare: Open issues and challenges. Computers & Electrical Engineering. 2016 Jul 1;53:421-34.

[30] Angelov PP, Gu X. Empirical approach to machine learning. New York: Springer; 2019 Dec 10.

[31] Ramana BV, Boddu RS. Performance comparison of classification algorithms on medical datasets. In2019 IEEE 9th Annual computing and communication workshop and conference (CCWC) 2019 Jan 7 (pp. 0140-0145). IEEE.

[32] Kabir MF, Ludwig SA. Enhancing the performance of classification using super learning. Data-Enabled Discovery and Applications. 2019 Dec;3:1-3.

[33] Wu CC, Yeh WC, Hsu WD, Islam MM, Nguyen PA, Poly TN, Wang YC, Yang HC, Li YC. Prediction of fatty liver disease using machine learning algorithms. Computer methods and programs in biomedicine. 2019 Mar 1;170:23-9.

[34] Kumar S, Katyal S. Effective analysis and diagnosis of liver disorder by data mining. In2018 international conference on inventive

research in computing applications (ICIRCA) 2018 Jul 11 (pp. 1047-1051). IEEE.

[35] Patwary MJ, Wang XZ. Sensitivity analysis on initial classifier accuracy in fuzziness based semi-supervised learning. Information Sciences. 2019 Jul 1;490:93-112.

[36] Pourpanah F, Lim CP, Wang X, Tan CJ, Seera M, Shi Y. A hybrid model of fuzzy min–max and brain storm optimization for feature selection and data classification. Neurocomputing. 2019 Mar 14;333:440-51.

[37] Tran TN, Vu DM, Tran MT, Le BD. The combination of fuzzy min–max neural network and semi-supervised learning in solving liver disease diagnosis support problem. Arabian journal for science and engineering. 2019 Apr 1;44:2933-44.

[38] Le TL. Fuzzy C-means clustering interval type-2 cerebellar model articulation neural network for medical data classification. IEEE Access. 2019 Jan 31;7:20967-73.

[39] Haque MR, Islam MM, Iqbal H, Reza MS, Hasan MK. Performance evaluation of random forests and artificial neural networks for the classification of liver disorder. In2018 international conference on computer, communication, chemical, material and electronic engineering (IC4ME2) 2018 Feb 8 (pp. 1-5). IEEE.

[40] Xu SS, Chang CC, Su CT, Phu PQ. Classification of liver diseases based on ultrasound image texture features. Applied Sciences. 2019 Jan 19;9(2):342.

[41] Doğantekin A, Özyurt F, Avcı E, Koc M. A novel approach for liver image classification: PH-C-ELM. Measurement. 2019 Apr 1;137:332-8.

[42] Subramanya MB, Virmani J. A DEFS Based System for Differential Diagnosis Between Severe Fatty Liver and Cirrhotic Liver Using Ultrasound Images. InMachine Learning in Bio-Signal Analysis and Diagnostic Imaging 2019 Jan 1 (pp. 53-72). Academic Press.

[43] Das A, Acharya UR, Panda SS, Sabut S. Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cognitive Systems Research. 2019 May 1;54:165-75.

[44] Mishra S, Tripathy HK, Mishra B, Sahoo S. Implementation of classification rule mining to minimize liver disorder risks. Int. J. Control Theory Appl. 2017 Nov;10:117-24.

[45] Abdar M, Zomorodi-Moghadam M, Das R, Ting IH. Performance analysis of classification algorithms on early detection of liver disease. Expert Systems with Applications. 2017 Jan 1;67:239-51.

[46] Sindhuja, D., & Priyadarsini, R. J. (2016). Liver Disease Analysis and Accuracy Prediction Using Machine Learning Techniques. *IJCTA*, *9*, 26.

[47] Saxena S, Deep V, Sharma P. Liver Disorder Prediction Due to Excessive Alcohol Consumption Using SLAVE. InEmerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2018, Volume 1 2019 (pp. 193-202). Springer Singapore.

[48] Sindhuja DR, Priyadarsini RJ. A survey on classification techniques in data mining for analyzing liver disease disorder. International Journal of Computer Science and Mobile Computing. 2016 May;5(5):483-8.

[49] Khadija MA, Setiawan NA. Detecting Liver Disease Diagnosis by Combining SMOTE, Information Gain Attribute Evaluation and Ranker. ITSMART: Jurnal Teknologi dan Informasi.;9(1):13-7.

[50] Acharya UR, Fujita H, Bhat S, Raghavendra U, Gudigar A, Molinari F, Vijayananthan A, Ng KH. Decision support system for fatty liver disease using GIST descriptors extracted from ultrasound images. Information Fusion. 2016 May 1;29:32-9.

[51] Ramana BV, Babu MS, Venkateswarlu NB. A critical comparative study of liver patients from USA and INDIA: an exploratory analysis. International Journal of Computer Science Issues (IJCSI). 2012 May 1;9(3):506.

[52] Ramana BV, Babu MS, Venkateswarlu NB. A critical study of selected classification algorithms for liver disease diagnosis. International Journal of Database Management Systems. 2011 May 2;3(2):101-14.

[53] Vital TP, Raju GP, Sreeramamurthy K, Charan VV. A probabilistic neural network approach for classification of datasets collected from north coastal districts of AP, India using MatLab. Procedia Computer Science. 2015 Jan 1;48:715-21.

[54] Lin RH, Chuang CL. A hybrid diagnosis model for determining the types of the liver disease. Computers in Biology and Medicine. 2010 Jul 1;40(7):665-70.

[55] Chuang CL. Case-based reasoning support for liver disease diagnosis. Artificial Intelligence in Medicine. 2011 Sep 1;53(1):15-23.

[56] Fan CY, Chang PC, Lin JJ, Hsieh JC. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing. 2011 Jan 1;11(1):632-44.

[57] Jen CH, Wang CC, Jiang BC, Chu YH, Chen MS. Application of classification techniques on development an early-warning system for

chronic illnesses. Expert Systems with Applications. 2012 Aug 1;39(10):8852-8.

[58] Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR). 2015 Apr 4;4(4):816-20.

[59] Jin H, Kim S, Kim J. Decision factors on effective liver patient data prediction. International journal of Bio-science and Bio-Technology. 2014 Aug;6(4):167-78.

[60] Nagaraj K, Sridhar A. NeuroSVM: a graphical user interface for identification of liver patients. arXiv preprint arXiv:1502.05534. 2015 Feb 19.

[61] Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes algorithms. International Journal of Science, Engineering and Technology Research (IJSETR). 2015 Apr 4;4(4):816-20.

[62] Alfisahrin SN, Mantoro T. Data mining techniques for optimization of liver disease classification. In2013 International Conference on Advanced Computer Science Applications and Technologies 2013 Dec 23 (pp. 379-384). IEEE.

[63] Rajeswari P, Reena GS. Analysis of liver disorder using data mining algorithm. Global journal of computer science and technology. 2010 Nov;10(14):48-52.

[64] Sug H. Improving the prediction accuracy of liver disorder disease with oversampling. InProceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics 2012 Jan 25 (pp. 331-335).

[65] Babu P, Swapna K, Balakrishna T, Venkateswarlu NB. An implementation of hierarchical clustering on Indian Liver Patient Dataset. International Journal of Emerging Technologies in Computational and Applied Sciences. 2014;8(6):543-

[66] Pahareeya J, Vohra R, Makhijani J, Patsariya S. Liver patient classification using intelligence techniques. International journal of advanced research in computer science and software engineering. 2014 Feb;4(2):295-9.

[67] Arasakumar, M., & Sudhakar, P.(2020). AN EFFECTIVE DYNAMIC WEIGHT BASED GREY WOLF OPTIMIZATION ALGORITHM WITH SUPPORT VECTOR MACHINE FOR CLASSIFICATION IN HEALTHCARE INDUSTRY.

[68] https://www.mathworks.com

[69] https://archive.ics.uci.edu/dataset/225/ilpd+indian+liver+patient+dataset

COPYRIGHTS

©2021 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.