

ارائه روشی ترکیبی جهت تشخیص بیماری قلبی با بهره‌گیری از رویکردهای انتخاب ویژگی تلفیقی و طبقه‌بندی بهینه

مرجان مطیعی‌زاده و مارال کلاه‌کج

تشخیص را افزایش دهد و مراجعات مکرر بیمار را به مراکز درمانی به حداقل رساند [۷]. با این حال، بسیاری از روش‌های موجود تنها از یک الگوریتم انتخاب ویژگی بهره می‌برند [۸] که ممکن است باعث انتخاب ویژگی‌های زائد و کاهش دقت طبقه‌بندی شود [۹]. همچنین روش‌های طبقه‌بندی متداول دارای محدودیت‌هایی در سرعت و دقت تشخیص هستند [۱۰] و [۱۱]؛ بنابراین اگر بتوان به ساختاری خودکار به‌عنوان ابزاری جدید در علم پزشکی جهت شناسایی غیرتهاجمی آریتمی قلبی مبتنی بر تحلیل سیگنال با دقت و سرعت مناسب دست یافت و نتایج را در اختیار پزشک قرار داد، می‌توان درمانی زود هنگام را همگام با آن آغاز کرد [۱۲] و [۱۳]. این از موضوعات حائز اهمیت در حوزه تشخیص زود هنگام و پیشگیری از حادث شدن بیماری شناخته می‌شود [۱۴].

هدف اصلی این تحقیق، ارائه روشی ترکیبی است که با استفاده از تلفیق الگوریتم‌های انتخاب ویژگی، ویژگی‌های مرتبط‌تر و مؤثرتری را استخراج کند و با بهره‌گیری از طبقه‌بندی ماشین یادگیری حداکثری با کرنل موجک و بهینه‌سازی پارامترهای آن توسط الگوریتم قورباغه بهینه‌شده، دقت و سرعت تشخیص بیماری قلبی را افزایش دهد. در این روش، خروجی‌های الگوریتم‌های مختلف انتخاب ویژگی با یکدیگر تلفیق شده و با استفاده از رأی‌گیری نرم، ویژگی‌هایی که به طور مکرر اهمیت بالایی دارند انتخاب می‌شوند. این رویکرد باعث افزایش دقت و کاهش تأثیر ویژگی‌های زائد می‌گردد. همچنین در این تحقیق، الگوریتم فرابانتکاری قورباغه به منظور بهینه‌سازی پارامترهای طبقه‌بند ماشین یادگیری حداکثری به کار رفته است. به منظور افزایش توانایی کاوش و بهبود سرعت همگرایی، نسخه اصلاح‌شده‌ای از الگوریتم قورباغه ارائه شده که شامل دو تغییر کلیدی می‌باشد: (۱) افزودن مکانیزم وزن‌دهی پویا برای ویژگی‌ها در هر نسل، به گونه‌ای که اهمیت هر ویژگی بر اساس تأثیر آن روی خطا به‌روزرسانی می‌شود. این مکانیزم باعث شناسایی سریع‌تر ویژگی‌های مؤثر و حذف ویژگی‌های زائد می‌شود. (۲) ترکیب الگوریتم قورباغه با الگوریتم ژنتیک در قالب یک فرایند ترکیبی که پس از هر چند نسل قورباغه، عملیات جهش و ترکیب راه‌حل‌ها انجام می‌شود تا از گیرافتادن در بهینه‌های محلی جلوگیری گردد. این اصلاحات منجر به بهبود دقت و سرعت طبقه‌بندی در تشخیص بیماری قلبی شده است. نتایج نشان می‌دهند که این رویکرد نسبت به روش‌های مشابه، عملکرد بهتری در تشخیص بیماری قلبی دارد.

به‌صورت خلاصه، جهت دستیابی به نتایج مطلوب، کار اصلی در این مقاله به شرح ذیل می‌باشد:

- معرفی روش ترکیبی انتخاب ویژگی با رأی‌گیری نرم
- به‌کارگیری طبقه‌بندی ماشین یادگیری حداکثری با کرنل موجک
- بهینه‌سازی پارامترهای طبقه‌بند توسط الگوریتم قورباغه بهینه‌شده

چکیده: بیماری‌های قلبی یکی از مهم‌ترین عوامل مرگ‌ومیر در جهان محسوب می‌شوند و تشخیص زود هنگام آنها از اهمیت بالایی برخوردار است. روش‌های موجود در انتخاب ویژگی برای تشخیص بیماری قلبی معمولاً محدود به استفاده از یک الگوریتم واحد بوده و ممکن است منجر به انتخاب ویژگی‌های زائد یا حذف ویژگی‌های مهم شوند که این امر دقت طبقه‌بندی را کاهش می‌دهد. در این مقاله، روش ترکیبی جدیدی برای انتخاب ویژگی ارائه شده که با بهره‌گیری از تلفیق نرم نتایج چند الگوریتم انتخاب ویژگی، ویژگی‌های کارآمدتر و مرتبط‌تر شناسایی می‌شوند. همچنین برای افزایش دقت و سرعت تشخیص، از طبقه‌بندی ماشین یادگیری حداکثری با کرنل موجک استفاده شده که پارامترهای آن توسط نسخه اصلاح‌شده الگوریتم فرابانتکاری قورباغه بهینه می‌شوند. این الگوریتم اصلاح‌شده شامل مکانیزم وزن‌دهی پویا و ترکیب با الگوریتم ژنتیک می‌باشد که به بهبود دقت و سرعت طبقه‌بندی کمک می‌کند. برای اثبات توانمندی و تعمیم‌پذیری روش، آن بر روی سه مجموعه داده معتبر UCI آزمایش شده که نتایج ارزیابی نشان می‌دهند دقت مدل پیشنهادی به ۹۳٫۳٪ رسیده است. نتایج بیانگر توانمندی و تعمیم‌پذیری بالای روش پیشنهادی در تشخیص بیماری قلبی می‌باشد.

کلیدواژه: انتخاب ویژگی، الگوریتم قورباغه، تشخیص بیماری قلبی، طبقه‌بندی ELM، کرنل موجک.

۱- مقدمه

بیماری‌های قلبی و عروقی در سراسر جهان یکی از اصلی‌ترین دلایل مرگ‌ومیر به شمار می‌روند. بر اساس گزارش سازمان بهداشت جهانی (WHO)، سالیانه بیش از ۱۷ میلیون نفر بر اثر این بیماری‌ها جان خود را از دست می‌دهند که این رقم حدود ۳۰٪ کل مرگ‌ومیرهای جهانی را تشکیل می‌دهد [۱] و [۲]. پیش‌بینی‌ها نشان می‌دهد که این تعداد تا سال ۲۰۳۰ به بیش از ۲۳ میلیون نفر افزایش یابد [۳]. تشخیص دقیق و زود هنگام بیماری‌های قلبی به علت پیچیدگی عملکرد قلب و تفاوت‌های فردی، کاری تخصصی و زمان‌بر است که معمولاً نیاز به تجربه بالینی فراوان دارد [۴].

در سال‌های اخیر، استفاده از سیستم‌های مبتنی بر یادگیری ماشین [۵] و انتخاب ویژگی [۶] برای تحلیل داده‌های پزشکی و به‌ویژه تشخیص بیماری قلبی افزایش یافته که می‌تواند تا حد چشمگیری دقت و صحت

این مقاله در تاریخ ۶ دی ماه ۱۴۰۰ دریافت و در تاریخ ۴ خرداد ماه ۱۴۰۴ بازنگری شد.

مرجان مطیعی‌زاده، گروه کامپیوتر، واحد اهواز، دانشگاه آزاد اسلامی، اهواز، ایران، (email: miss.motiee@gmail.com)

مارال کلاه‌کج (نویسنده مسئول)، گروه مهندسی کامپیوتر، واحد سوسنگرد، دانشگاه آزاد اسلامی، سوسنگرد، ایران، (email: maral.kolahkaj@iau.ac.ir).

حداقل مربعات ماشین بردار پشتیبانی دوقلو، یک روش مبتنی بر ماشین بردار پشتیبان دوقلوی مبتنی بر مدل خطی وزنی پیشنهاد شده است. آزمایش‌های انجام‌شده روی مجموعه داده‌های قلبی منجر به دستیابی به دقت ۸۵ درصد در حالت دوکلاسه با ۳۰۳ نمونه شده است.

در [۲۵]، یک روش کارآمد برای یادگیری مدل‌های رگرسیون لجستیک در درخت پیشنهاد شده است. آنها حداقل رگرسیون زاویه‌ای را برای به‌روزرسانی مدل رگرسیون در LogitBoost به کار می‌گیرند تا الگوریتم به‌طور مؤثری مدل‌های رگرسیون لجستیک پراکنده را که از متغیرهای ورودی مربوطه یاد می‌گیرد، بیاموزد. برای داده‌های قلبی در حالت دوکلاسی به دقت ۸۳٪ درصد دست یافتند.

Wang و همکاران [۲۶] مدلی جدید معرفی کرده‌اند که ترکیبی از یادگیری تنک، تابع پین‌بال و مفهوم Universum در طبقه‌بندی‌های غیرموازی است. این مدل با هدف بهبود دقت و کاهش پیچیدگی محاسباتی در مسائل طبقه‌بندی توسعه یافته و نیز یک قانون پالایش ایمن برای کاهش ابعاد داده و تسریع فرایند آموزش ارائه شده است. نتایج تجربی نشان داده‌اند که این مدل نسبت به مدل‌های مشابه، عملکرد بهتری در دقت و کارایی دارد.

در [۱۳]، یک روش انتخاب زیرمجموعه ویژگی با استفاده از ترکیب یک الگوریتم ژنتیک با یک الگوریتم فرااکتشافی شبیه‌سازی شده و همراه با طبقه‌بندی ماشین بردار پشتیبان ارائه گردیده است. در تحقیق آنها به‌واسطه طبقه‌بندی مدل برای داده‌های قلبی با ۳۰۳ نمونه و در حالت دوکلاسی دقت ۸۴٪ درصد محاسبه شده است.

تحقیق [۲۷] با یکپارچه‌سازی انتساب و انتخاب ویژگی‌های مبتنی بر الگوریتم ژنتیک، پیشرفت در رویکرد گروهی را پیشنهاد می‌دهد. انتخاب ویژگی، تعداد الگوهای ازدست‌رفته را کاهش می‌دهد و باعث افزایش سرعت طبقه‌بندی خواهد شد. همچنین کسری از نمونه‌های جدید را که می‌توان توسط گروه طبقه‌بندی کرد تا حد زیادی افزایش داد. در این تحقیق برای داده‌های قلبی با حالت پنج‌کلاسه و ۳۰۳ نمونه به دقت ۵۸٪ درصد دست پیدا کردند.

۳- روش پیشنهادی

مدل پیشنهادی ترکیبی از روش‌های انتخاب ویژگی و تلفیق پاسخ‌ها و نیز طبقه‌بندی پاسخ‌هاست. همچنین برای بهینه‌نمودن جواب‌ها از بهینه‌سازی طبقه‌بندی استفاده شده است. روند کار به این صورت است که ابتدا پیش‌پردازش‌هایی به‌منظور بهبود کیفیت ویژگی‌ها انجام می‌گیرد. در ادامه ویژگی‌ها نرمال شده و در مرحله انتخاب ویژگی، تعدادی از ویژگی‌ها به‌عنوان زیرمجموعه ویژگی کلینیکی انتخاب می‌شوند. مرحله یادگیری و طبقه‌بندی در دو گام انجام می‌گیرد؛ به طوری که ابتدا توسط اعمال روش ماشین با امکان یادگیری حداکثری و بهبود آن توسط نسخه بهبودیافته الگوریتم قورباغه، نمونه‌های سالم و دارای بیماری نارسایی قلبی از هم تفکیک می‌شوند. در مرحله دوم محک‌های ارزیابی خطا و دقت برای هر دو بخش داده آزمایشی و آموزشی پیاده‌سازی می‌شوند. در شکل ۱ فلوچارت کلی روش پیشنهادی نشان داده شده است.

۳-۱ پیش‌پردازش

ابتدا پیش‌پردازش‌هایی به‌منظور بهبود کیفیت ویژگی‌ها بر روی آنها انجام می‌گیرد. در این مرحله با استفاده از نرمال‌سازی و یافتن مقادیر ازدست‌رفته از طریق مشابه‌ترین همسایگی پیش‌پردازش انجام می‌شود.

• بهبود دقت و سرعت تشخیص بیماری قلبی نسبت به روش‌های موجود

۲- پیشینه تحقیق

در جهت تشخیص بیماری، روش‌های انتخاب ویژگی مورد توجه بوده که از آن جمله می‌توان به [۱۵] اشاره کرد که از جستجوی هارمونی (HSA) و جستجوی محلی تصادفی (SLS) جهت انتخاب ویژگی‌ها استفاده نموده و نهایتاً برای یافتن پارامترهای طبقه‌بندی ماشین بردار پشتیبان از الگوریتم جستجوی هارمونی- جستجوی محلی تصادفی ترکیبی نیز استفاده کرده‌اند. این تحقیق مشابه با [۱۶] بود، با این تفاوت که تکنیک‌های جستجو تغییر کرده است.

استفاده از الگوریتم‌های مبتنی بر انتخاب ویژگی شامل الگوریتم کلونی مورچگان و الگوریتم زنبور عسل از جمله تکنیک‌های فرااکتشافی است که می‌توانند در تفکیک ویژگی‌های مناسب و به‌تبع آن، طبقه‌بندی مناسب به کار گرفته شوند [۱۷]. در تحقیق Tapkan و همکاران [۱۸] روش طبقه‌بندی حساس به هزینه با عنوان الگوریتم BEE-Miner با استفاده از الگوریتم زنبور عسل بهبودیافته، پیشنهاد شده است. در این تحقیق برای داده قلبی دوکلاسه با ۲۷۰ نمونه، بر اساس روش پیشنهادیشان، دقتی معادل ۷۵٪/۶۷ به دست آمده است.

مرجع [۱۹] یک شبکه عصبی جدید معرفی می‌کند که با بهره‌گیری از اطلاعات گراف برای طبقه‌بندی داده‌های ناقص طراحی شده است. رویکرد پیشنهادی بدون نیاز به تکمیل‌سازی داده‌های ناقص، ویژگی‌های مؤثر را از طریق ساختار گرافی استخراج کرده و عملکرد مدل را در مواجهه با داده‌های ناقص به‌طور قابل توجهی بهبود می‌دهد.

در [۲۰] برای بهیگی مدل SPODES از الگوریتم ایمنی مصنوعی بهره گرفته شده است. در این تحقیق، آنها داده‌های نارسایی قلبی پنج‌کلاسه را به دوکلاسه تبدیل کردند و دقت ۸۳٪/۸۰ را به دست آوردند. ایراد کار آنها در آن بود که برای حالت چندکلاسی دقت مناسبی به دست نیاورده بودند.

در [۲۱]، یک روش شناسایی مجدد افراد بین مدالیته به‌صورت نیمه‌نظارتی و مبتنی بر اعتماد پیشنهاد شده است. روش معرفی‌شده از برچسب‌گذاری خودکار نمونه‌ها با سطح اعتماد بالا بهره می‌برد تا یادگیری بین تصاویر رنگی و مادون قرمز را بهینه‌سازی کرده و دقت بازشناسایی را در سناریوهای چندحسی بهبود دهد.

مرجع [۲۲] مدل جدیدی مبتنی بر معیار نابارگی جرمی ارائه داده است. این مدل برای بهبود دقت طبقه‌بندی در داده‌های نویزی و غیرقطعی طراحی شده و از مفاهیم فازی برای مقابله با عدم قطعیت و از فاصله‌های جرمی برای سنجش شباهت مؤثرتر میان نمونه‌ها استفاده می‌کند. نتایج تجربی نشان می‌دهند که FTSVM پیشنهادی عملکرد بالاتری نسبت به سایر روش‌های فازی دارد.

در [۲۳]، انواع مختلف کرنل‌هایی که در ماشین بردار پشتیبان استفاده شده، با استفاده از داده‌های زیست پزشکی مورد بررسی و مقایسه قرار گرفته است. نتایج تجربی نشان می‌دهد که استفاده از نوع کرنل مناسب در طبقه‌بندی‌کننده ماشین بردار پشتیبان بر روی داده‌های زیست پزشکی، نتایج طبقه‌بندی موفقیت‌آمیز و قابل اطمینان‌تری را ارائه می‌دهد.

در [۲۴] بر اساس روش ماشین بردار پشتیبانی دوقلو پیش‌بینی و

ج) روش Relief

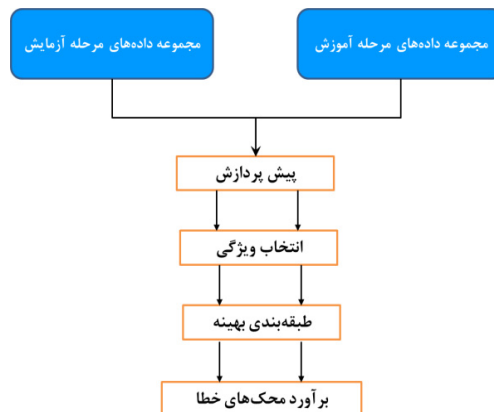
ایده اصلی این الگوریتم آن است که هرچه اختلاف بین یک ویژگی در نمونه انتخاب‌شده و نزدیک‌ترین برخورد کمتر باشد، این ویژگی بهتر است و به‌علاوه یک ویژگی خوب آن است که اختلاف بین اندازه آن ویژگی و نزدیک‌ترین شکست وی بیشتر باشد. الگوریتم پس از تعیین نزدیک‌ترین برخورد و نزدیک‌ترین شکست، وزن‌های ویژگی‌ها را به‌روزرسانی می‌کند. بعد از تعیین فاصله برای تمام نمونه‌های موجود در مجموعه نمونه‌ها، الگوریتم، ویژگی‌هایی را که وزن آنها کمتر یا مساوی با یک حد آستانه است را حذف می‌کند و سایر ویژگی‌ها را به‌عنوان زیرمجموعه ویژگی جواب بازمی‌گرداند.

۳-۳ تلفیق نرم با بهره‌گیری از روش ویکور

یکی از روش‌های کارآمد در زمینه انتخاب ویژگی، شیوه ویکور^۳ یا ارزیابی رتبه‌بندی به‌عنوان یک روش مبتنی بر تحلیل چندمعیاره تصمیم‌گیری^۴ (MCDM) برای حل چالش‌های داده استفاده می‌شود و در آن معیارها تناسب‌پذیر و ناسازگار هستند و تصمیم‌گیری مستقیم در مورد آنها قدری پیچیده است. در تهیه چنین راه‌حل سازمان‌دهی‌شده‌ای، روش ویکور مفاهیم «مزیت قابل قبول» و «پایداری قابل قبول» را برای تعیین حداکثر «گروهی از اکثریت» و حداقل «ناکارآمدی یک ویژگی به‌خصوص» به کار می‌برد [۲۹]. بر اساس مفاهیم ذکرشده، الگوریتم ویکور دارای محک‌هایی است [۳۰]. با ورود ویژگی‌ها به‌عنوان داده‌های آموزشی، هر یک از شیوه‌های انتخاب ویژگی به‌عنوان مدلی از انتخاب ویژگی، مجموعه‌ای از ویژگی‌های کارآمد را استخراج می‌کنند. به عبارت دیگر، اولین گام اساسی مبتنی بر یادگیری ماشینی پیش از طبقه‌بندی بیماری، برگزینش زیرمجموعه‌ای از ویژگی‌های مناسب است. سپس زیرمجموعه ویژگی‌ها را به انتخابگر ویژگی می‌دهیم تا ویژگی‌هایی که کمترین همبستگی را با یکدیگر دارند، به‌عنوان ویژگی‌های مؤثر به کار ببندیم. به عبارت بهتر با این کار جهت پس‌زدن مشخصه‌هایی اقدام می‌کنیم که با فرایند تشخیص بیماری ارتباط کمتری دارند و در صورت حذف آنها، نه‌تنها حجم اطلاعاتی که قرار است پردازش شوند، کاهش می‌یابد، بلکه سرعت تحلیل داده‌ها نیز تا حد چشمگیری افزایش پیدا می‌کند.

۳-۴ طبقه‌بندی با ماشین یادگیری حداکثری

ماشین‌های یادگیرنده حداکثری دارای مکانیسمی درخور توجه هستند و می‌توانند اتصال‌دهنده‌ای میان یادگیری ماشین و یادگیری بیولوژیکی باشند. این شبکه، ساختاری مشابه با شبکه‌های عصبی مصنوعی با تغذیه پیش‌رونده دارد که دارای یک لایه از گره‌های مخفی هستند و پارامترهای گره‌های لایه مخفی آن نیازی به تنظیم ندارند. گره‌های لایه مخفی می‌توانند تصادفی مقدار بگیرند و هیچ‌گاه به‌روزرسانی نشوند یا می‌توانند بدون هیچ تغییری، پارامترهای قبلی خود را داشته باشند. در اغلب موارد، یادگیری مبتنی بر یک مرحله می‌باشد که برای یک مدل خطی کفایت می‌کند؛ اما زمانی که مسئله غیرخطی و پیچیده باشد، آن‌گاه نیازمند به وجود هسته (کرنل) خواهیم بود که از عمده ماشین‌های یادگیرنده حداکثری همراه با کرنل می‌توان به کرنل‌های خطی، چندجمله‌ای، تابع پایه شعاعی و کرنل موجک اشاره نمود. در میان کرنل‌ها به طور معمول



شکل ۱: مراحل الگوریتم پیشنهادی.

۳-۲ انتخاب ویژگی

در این مرحله، ابتدا از سه روش انتخاب ویژگی فیلتری عام، Wrapper و Relief جهت انتخاب ویژگی استفاده می‌شود. سپس از سه بردار ویژگی حاصل از سه روش مذکور به روش رأی‌گیری نرم، بهترین بردار ویژگی انتخاب خواهد شد. در ادامه هر یک از معیارها به‌صورت مجزا تشریح می‌شوند.

الف) روش فیلتر

اگر انتخاب ویژگی مستقل از هر گونه الگوریتم یادگیری انجام شود، آن روش را شیوه فیلتر یا حلقه باز گویند. در این مورد، ویژگی‌های نامطلوب پیش از استنتاج کنار زده خواهند شد. در روش مورد نظر، ابتدا ویژگی‌های مربوط به بیماری در چند گروه مشخص تقسیم‌بندی می‌شوند و برای هر گروه (G_i) یک امتیاز به نام Sc_i به هر گروه مطابق (۱) تعلق می‌گیرد. گروه‌ها بر اساس وزنی که دریافت کرده‌اند، مرتب می‌شوند و گروهی که بر اساس بهترین ویژگی‌ها در رأس این امتیازبندی قرار می‌گیرد، به‌عنوان خروجی اولیه شناخته خواهد شد

$$Sc_i = \frac{\delta_{i+} - \varepsilon_{i+}}{\delta_{i-} - \varepsilon_{i-}} \quad (1)$$

که در آن δ_{i+} و ε_{i+} به ترتیب نمایانگر انحراف متوسط و انحراف استاندارد ویژگی‌های بیماری قلبی و δ_{i-} و ε_{i-} نیز به ترتیب نمایانگر انحراف متوسط و انحراف استاندارد نمونه‌های سالم هستند.

ب) روش Wrapper

اگر روند ارزیابی انتخاب ویژگی با یک الگوریتم طبقه‌بندی در ارتباط باشد، روش انتخاب ویژگی را پیچشی یا حلقه بسته می‌نامند. روش پیچشی، جستجو در فضای زیرمجموعه‌ها را بر اساس تخمین دقت ناشی از انتخاب یک زیرمجموعه خاص تحت شرایط الگوریتم طبقه‌بندی مورد استفاده به‌عنوان معیاری از بهینگی آن زیرمجموعه انجام می‌دهد [۲۸]. این روش از استراتژی‌های جستجو مانند انتخاب متوالی روبه‌جلو^۱ و انتخاب متوالی روبه‌عقب^۲ استفاده می‌کند. در این رویکرد، تمامی زیرمجموعه‌های ممکن از ویژگی‌ها در نظر گرفته می‌شود و با ارزیابی همه حالت‌ها، بهترین آنها که کمترین خطای عمومی را به همراه دارد، انتخاب می‌شود.

3. VIKOR

4. The Field of Multiple Criteria Decision Making

1. Sequential Forward Selection

2. Sequential Backward Selection

اگر این تغییر موقعیت، قورباغه‌ای با شایستگی بهتر تولید کرد، این قورباغه جایگزین قورباغه بدتر می‌شود و در غیر این صورت، قورباغه با بهترین شایستگی در کل جمعیت (بهینه فرامحلی) X_g جایگزین X_b در (۳) شده و قورباغه جدیدی تولید می‌شود. اگر قورباغه‌ای با شایستگی بهتر تولید شود، این قورباغه جایگزین قورباغه بدتر می‌شود و در غیر این صورت، قورباغه جدید به صورت تصادفی تولید و جایگزین بدترین قورباغه خواهد شد.

(۴) به هم آمیختن جمعیت: پس از تکامل درونی چندین نسل، تمام مجموعه‌ها به هم آمیخته و بر اساس ارزش شایستگی آنها به صورت نزولی مرتب می‌شوند. سپس دوباره به چند زیرمجموعه تقسیم می‌گردند و روند تکامل در هر مجموعه تا زمانی که به معیار توقف برسد ادامه می‌یابد.

۳-۶ ترکیب

در الگوریتم پیشنهادی با استفاده از جستجوی سرتاسری الگوریتم قورباغه در فضای جستجو، به وزن‌های بهینه‌ای می‌رسیم که این وزن‌ها به شبکه عصبی یادگیری حداکثری سوئیچ می‌گردد. در الگوریتم قورباغه با توجه به ساختار شبکه عصبی پیشنهادی هر قورباغه، برداری x خانه‌ای مطابق با شکل ۲ است که شامل وزن‌ها و بایاس شبکه عصبی می‌باشد. در واقع بردار (w, b) یک قورباغه به گونه‌ای تعریف می‌شود که در آن w_i برابر با وزن‌ها و b_i برابر بایاس شبکه عصبی مصنوعی می‌باشند.

مراحل تعیین وزن‌های بهینه توسط الگوریتم قورباغه به این صورت است که قورباغه‌های جستجوگر در فضای جستجو (وزن‌های اولیه) و به صورت تصادفی جستجو را انجام می‌دهند. در هر تکرار میزان شایستگی راه‌حل‌ها با شایستگی به دست آمده توسط بهترین قورباغه‌ها مقایسه شده و بهترین راه‌حل به دست می‌آید (هرچه خطای به دست آمده از هر قورباغه کمتر باشد، میزان شایستگی آن راه‌حل بیشتر است). سپس جستجو تا رسیدن به تکرار آخر ادامه دارد و نهایتاً بهترین وزن برای شبکه به دست می‌آید. در هر مرحله از جستجو اگر راه‌حل یافت شده بهتر از راه‌حل قبلی باشد با آن جایگزین شده و در غیر این صورت راه‌حل قبلی باقی می‌ماند. بنابراین الگوریتم قورباغه پس از X بار تکرار، بهترین وزن‌های بهینه را با بالاترین دقت و کمترین هزینه پیدا می‌کند و شبکه عصبی پیشنهادی با وزن‌های بهینه‌ای که توسط الگوریتم قورباغه پیدا شده است، مورد ارزیابی قرار می‌گیرد.

۳-۷ بهینه‌سازی الگوریتم قورباغه

در نسخه اصلی، الگوریتم قورباغه پارامترهای طبقه‌بندی را بهینه می‌کند. در ادامه به انجام اصلاحاتی جهت بهینه‌سازی این الگوریتم که شامل دو مرحله است، می‌پردازیم.

الف) افزودن وزن‌دهی پویا در انتخاب ویژگی:

- در نسخه اصلاح شده، یک مرحله وزن‌دهی پویا روی ویژگی‌ها اضافه می‌کنیم که در هر نسل الگوریتم قورباغه، اهمیت ویژگی‌ها بر اساس میزان تأثیر آنها روی خطا به روزرسانی شود.
- این کار باعث می‌شود ویژگی‌های مؤثرتر زودتر شناسایی شده و الگوریتم سریع‌تر همگرا شود.

ب) ترکیب الگوریتم قورباغه با الگوریتم ژنتیک:

- بعد از چند نسل الگوریتم قورباغه، از عملیات جهش و ترکیب الگوریتم ژنتیک برای تنوع‌دادن به جمعیت راه‌حل‌ها استفاده می‌شود.

$$(w, b) \begin{bmatrix} w_1 & w_2 & \dots & w_i & b_1 & b_2 & \dots & b_i \end{bmatrix}$$

شکل ۲: ساختار عملکردی هر قورباغه در برآورد پارامتر کرنل شبکه عصبی با امکان یادگیری حداکثری.

مدل مویک پاسخ بهینه‌تری به همراه دارد و سریع‌تر عمل می‌کند [۳۱]. در میان همه روش‌های طبقه‌بندی مبتنی بر الگوریتم ماشین یادگیرنده حداکثری، ایرادی که همواره خودنمایی می‌کند، دقت کم است که به سبب بیش‌برازش رخ می‌دهد. بیش‌برازش به معنی طبقه‌بندی یک کلاس در کلاس یا کلاس‌های دیگر است که به‌طور معمول زمانی که الگوریتم یک کلاس را بیشتر یاد بگیرد، احتمال قرارگیری در کلاس دیگر کمتر خواهد بود. برای غلبه بر این مشکلات نیاز است تا برخی از پارامترهای الگوریتم ماشین یادگیرنده حداکثری به‌واسطه وجود کرنل تصحیح و تنظیم شوند. کرنل انتخابی، کرنل مویک است که خود دارای سه پارامتر تنظیمی است. در این تحقیق این پارامترها را توسط الگوریتم قورباغه تنظیم می‌کنیم.

۳-۵ تعیین پارامترهای شبکه عصبی با الگوریتم تکاملی قورباغه

الگوریتم جهش قورباغه^۱ (SFL) یکی از الگوریتم‌های الهام‌گرفته از طبیعت است که گروهی از قورباغه‌ها به چندین زیرمجموعه^۲ تقسیم می‌شوند که هر قورباغه فرهنگ مختص به خود را دارد و می‌تواند از فرهنگ‌ها^۳ یا ایده‌های قورباغه‌های دیگر در طول روند تکامل استفاده کند [۳۲]. مراحل اجرای این الگوریتم در زیر شرح داده شده است:

(۱) تولید جمعیت اولیه: همانند تمامی الگوریتم‌های اکتشافی^۴، جمعیت اولیه (قورباغه‌ها) به صورت تصادفی^۵ از میان بازه مسئله، تولید می‌شود. قورباغه‌ها بر اساس شایستگی‌شان به صورت نزولی مرتب می‌شوند و بر اساس روندی خاص به زیرمجموعه‌های مختلف تقسیم می‌شوند.

(۲) دسته‌بندی قورباغه‌ها: فرض بر آن است که جمعیت اولیه با p قورباغه تولید شده و p قورباغه به m مجموع تقسیم می‌شوند. روند تقسیم‌بندی قورباغه‌ها بدین صورت است که قورباغه اول به مجموعه اول، قورباغه دوم به مجموعه دوم و قورباغه m ام به مجموعه m ام و قورباغه $m+1$ ام به مجموعه اول تعلق دارند. این روند به صورت مشابه تا قورباغه آخر تکرار می‌شود. هر مجموعه m شامل n قورباغه است به طوری که (۲) برآورده شود

$$p = m \times n \quad (2)$$

(۳) مراحل جستجوی محلی در الگوریتم SFL: در هر مجموعه موقعیت قورباغه i بر اساس اختلاف بین قورباغه بهتر با بهترین شایستگی X_b و قورباغه بدتر با بدترین شایستگی X_w با استفاده از (۳) به دست می‌آید

$$D(i) = rand() \times (X_b - X_w) \quad (3)$$

که $rand()$ یک عدد تصادفی یکنواخت بین صفر تا یک است. موقعیت جدید قورباغه توسط (۴) به دست می‌آید که D_{max} ماکسیمم تغییراتی است که در موقعیت قورباغه می‌توان اعمال کرد

$$X_w = currentpositionX_w + D_i \quad (4)$$

1. Shuffled Frog Leaping
2. Memplex
3. Culture
4. Heuristic
5. Random

این مجموعه داده شامل اطلاعات ۲۷۰ بیمار است که هدف در آن، پیش‌بینی احتمال ابتلا به بیماری قلبی است. هر رکورد دارای ۱۳ ویژگی عددی شامل عواملی نظیر فشار خون، میزان کلسترول، سن، وضعیت قند خون، ضربان قلب ماکسیمم و سایر عوامل بالینی مرتبط با بیماری قلبی می‌باشد. متغیر هدف این مجموعه داده یک متغیر دودویی است که وجود یا عدم وجود بیماری قلبی را نشان می‌دهد (۰ برای فرد سالم و ۱ برای فرد مبتلا). داده‌ها به صورت کامل بوده و فاقد مقادیر گمشده می‌باشند. این مجموعه داده از نظر حجم متوسط، بسیار متعادل و استاندارد بوده و برای مقایسه الگوریتم‌های یادگیری ماشین در حوزه تشخیص بیماری قلبی کاربرد گسترده‌ای دارد.

مجموعه داده MIT-BIH Arrhythmia

پایگاه داده MIT-BIH Arrhythmia در سال ۱۹۸۰ توسط مرکز پزشکی Beth Israel Hospital وابسته به دانشگاه Harvard و با همکاری مؤسسه فناوری ماساچوست (MIT) ایجاد شده است. در این پایگاه داده، نمونه‌ها شامل ۴۸ بیمار هستند که از میان بیش از ۴۰۰۰ نوار ECG انتخاب شده‌اند. در این پایگاه داده، بیش از ۱۰۰۰۰۰ ضربان قلب برچسب‌گذاری شده وجود دارد که توسط متخصصان قلب و بر اساس استانداردهای انجمن قلب آمریکا دسته‌بندی شده‌اند. هر ضربان به یکی از کلاس‌های مختلف مانند ضربان نرمال، ضربان زودرس بطنی، فیبریلاسیون دهلیزی و دیگر انواع آریتمی نسبت داده شده است. در این مرحله، داده‌هایی که در حال حاضر در دسترس هستند و داده‌هایی که برای ساخت مدل مورد نیاز بوده‌اند، تعیین شدند. برای شروع پژوهش بر اساس مطالعات کتابخانه‌ای و استانداردهای موجود در پایگاه داده MIT-BIH، ویژگی‌هایی همچون شکل سیگنال، فاصله RR، نوع ضربان و سایر پارامترهای مرتبط استخراج و مورد استفاده قرار گرفته‌اند.

در ادامه، متغیرهای تعیین‌شده برای ایجاد مدل به دو دسته متغیرهای هدف و متغیرهای پیشگو دسته‌بندی شدند که متغیر هدف ابتلا یا عدم ابتلا و سایر متغیرها به عنوان متغیر پیشگو مورد استفاده قرار گرفتند. جهت انجام آزمایش‌ها، نحوه تقسیم داده‌ها همانند سایر روش‌های شناسایی بیماری قلبی، ۷۰٪ (جهت آموزش) در برابر ۳۰٪ (جهت آزمایش) در نظر گرفته شد.

۴-۲ طبقه‌بندی

شبکه عصبی با امکان یادگیری حداکثری می‌تواند در صورت انتخاب مقادیر مناسب برای پارامترهای آن از قبیل تعداد لایه‌ها و نرون‌ها، نگاهی غیرخطی با دقت دلخواه تشکیل دهد. در واقع شبکه عصبی با امکان یادگیری حداکثر تنها دارای یک لایه مخفی میانی است و به همین دلیل به عنوان یک شبکه عصبی کارآمد محسوب می‌گردد.

۴-۳ بهینه‌سازی پارامترها

برای الگوریتم جهش ترکیبی قورباغه، برخی مقادیرهای اولیه تعریف می‌شود. در مرحله بعد، شبکه با امکان یادگیری حداکثری به وسیله داده‌های آموزشی و پارامترهای محاسبه‌شده، آموزش دیده و مدل عصبی مورد نظر ساخته می‌شوند و البته به سبب بروز ساختارهای مختلف می‌توان میزان خطای حاصل از طبقه‌بندی را پیش‌بینی نمود. پارامترهای مورد استفاده در الگوریتم جهش ترکیبی قورباغه به صورت تجربی و با آزمون و خطا طبق جدول ۱ به دست آمدند.

جدول ۱: پارامترهای الگوریتم جهش ترکیبی قورباغه در تنظیم پارامترهای ماشین با امکان یادگیری حداکثری.

پارامتر	مقدار
سایز ممیتیک اکتشاف	۱۰
تعداد ممیتیک‌های جهش ترکیبی	۵
تعداد والدین	حداکثر ۴ و حداقل ۲
تعداد فرزندان	حداکثر ۴
حداکثر تکرار چرخه جهش	۱۰
اندازه گام	۲

• این کار باعث جلوگیری از گیرافتادن در بهینه‌های محلی و افزایش دقت می‌شود.

۴- پیاده‌سازی روش پیشنهادی

همان گونه که پیش از این عنوان شد، نارسایی قلبی یکی از دلایل اصلی مرگ‌ومیر در سراسر جهان است. در بخش قبل نیز اشاره کردیم که تشخیص نارسایی قلبی یک کار چالش‌برانگیز است؛ به خصوص در کشورهای در حال توسعه و کمتر توسعه‌یافته که کمبود متخصصان و تجهیزات انسانی در آنها وجود دارد. از این رو روش ترکیبی پیشنهاد گردید که بتواند به عنوان یک سیستم هوشمند در تشخیص خودکار نارسایی قلبی مؤثر باشد. مدل‌سازی سیستم پیشنهادی در محیط برنامه‌نویسی متلب انجام پذیرفته و هدف به دست آوردن سطح مطلوبی از دقت بهینه‌سازی در تشخیص این بیماری برای کلاس‌های چندگانه است. سیستم مورد استفاده دارای پردازنده Core i۷ با فرکانس مسیر ۲٫۷ گیگاهرتز و ۴ گیگابایت RAM است. همچنین نسخه متلب مورد استفاده، نسخه ۲۰۱۹ بی می‌باشد که دارای افزونه‌های آنلاین و آخرین نسخه به‌روزرسانی شده از جعبه ابزارهای این نرم‌افزار است.

۴-۱ مجموعه داده

در این پژوهش از سه مجموعه داده تشخیص بیماری قلبی Cleveland Heart Disease dataset، Statlog Heart dataset و MIT-BIH Arrhythmia dataset از مخزن داده‌های سایت مرجع UCI و MIT استفاده شده است [۳۳] و [۳۴]. در ادامه به شرح هر کدام از مجموعه داده‌ها پرداخته می‌شود.

مجموعه داده Cleveland Heart Disease

مرکز پزشکی Long Cleveland Clinic Foundation Beach و VA در سال ۱۹۹۸، این پایگاه داده را ایجاد کرد که در آن نمونه‌ها شامل ۳۰۳ فرد مورد بررسی است که دربرگیرنده ۲۹۷ نمونه کامل و ۶ نمونه با مقادیر از دست‌رفته است. این پایگاه داده شامل ۱۳ علامت بیماری و یک صفت تشخیصی است که به وجود بیماری قلبی بر اساس علائم موجود در بیمار اشاره دارد که یک مقدار عددی صفر (کمتر از ۵۰٪ تنگی عروق) یا ۱ (به معنای تنگی عروق بیشتر از ۵۰٪) است. در این مرحله، داده‌هایی که در حال حاضر در دسترس هستند و داده‌هایی که برای ساخت مدل نیاز بود، تعیین شدند. برای شروع پژوهش بر اساس مطالعات کتابخانه‌ای و مجموعه داده Cleveland، ۱۴ ویژگی در نظر گرفته شده است.

مجموعه داده Statlog Heart

پایگاه داده Statlog Heart بخشی از پروژه Statlog^۱ بوده و توسط

جدول ۲: نتایج حاصل از پیاده‌سازی مدل اول در مرحله آزمایش به روش تقسیم داده‌ها با CV برابر ۱۰ مرکب از اعمال بیشینه ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی دو کلاسی.

فولد آزمون	انتخاب تعداد کم نرون در لایه مخفی (۳ تا ۶)		انتخاب تعداد متوسط نرون در لایه مخفی (۷ تا ۱۱)		انتخاب تعداد نرون زیاد در لایه مخفی (۱۲ تا ۱۸)	
	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت
فولد ۶	۰٫۷۸۵۶	۰٫۲۱۴۳	۰٫۱۷۹۷	۰٫۸۲۰۳	۰٫۲۴۴۵	۰٫۷۵۵۵
	۰٫۷۸۶۷	۰٫۲۱۳۲	۰٫۸۳۰۰	۰٫۱۷۰۰	۰٫۸۲۴۳	۰٫۱۷۵۷
فولد ۸	۰٫۷۷۲۵	۰٫۲۲۹۳	۰٫۱۸۰۰	۰٫۸۲۰۰	۰٫۲۲۴۶	۰٫۷۷۵۴
	۰٫۷۷۲۵	۰٫۲۲۷۵	۰٫۸۲۸۸	۰٫۱۷۱۲	۰٫۷۴۴۴	۰٫۲۵۵۶
فولد ۹	۰٫۸۰۰۷	۰٫۲۱۹۶	۰٫۱۵۵۳	۰٫۸۴۴۷	۰٫۲۳۴۴	۰٫۷۶۵۶
	۰٫۸۰۰۷	۰٫۲۱۹۳	۰٫۸۶۷۴	۰٫۱۳۲۶	۰٫۷۷۰۲	۰٫۲۲۹۸
میانگین	۰٫۷۹۲۸	۰٫۲۱۹۱۷	۰٫۸۴۰۷	۰٫۸۲۴۷	۰٫۷۷۷۱	۰٫۷۸۳۱

جدول ۳: نتایج حاصل از مدل اول در مرحله آزمایش به روش تقسیم داده‌ها (۳۰/۷۰) مرکب از اعمال حداکثر ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی پنج کلاسی.

کلاس ۱	کلاس ۲	کلاس ۳	کلاس ۴	کلاس ۵	Accuracy
کلاس ۱	۰٫۶۱	۰٫۱۲	۰٫۱۱	۰٫۰۹	۰٫۰۷
کلاس ۲	۰٫۰۷	۰٫۵۹	۰٫۱۴	۰٫۱۷	۰٫۰۳
کلاس ۳	۰٫۰۸	۰٫۱۵	۰٫۵۷	۰٫۱۴	۰٫۰۶
کلاس ۴	۰٫۰۶	۰٫۱۱	۰٫۱۷	۰٫۵۵	۰٫۱۱
کلاس ۵	۰٫۰۱	۰٫۰۷	۰٫۱۱	۰٫۱۴	۰٫۵۸
دقت نهایی					
۰٫۵۸					

جدول ۴: نتایج حاصل از مدل دوم در مرحله آزمایش به روش تقسیم داده‌ها با CV برابر ۱۰ مرکب از اعمال تعداد متوسط ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی دو کلاسی.

فولد آزمون	انتخاب تعداد کم نرون در لایه مخفی (۳ تا ۶ نرون)		انتخاب تعداد نرون متوسط لایه مخفی (۷ تا ۱۱)		انتخاب تعداد نرون زیاد در لایه مخفی (۱۲ تا ۱۸)	
	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت
فولد ۶	۰٫۸۳۸۴	۰٫۱۶۱۵	۰٫۱۵۳۶	۰٫۸۴۶۴	۰٫۱۸۲۵	۰٫۸۱۷۵
	۰٫۸۴۹۹	۰٫۱۶۱۵	۰٫۸۳۴۴	۰٫۱۶۵۶	۰٫۸۵۰۰	۰٫۱۵۰۰
فولد ۸	۰٫۸۲۸۳	۰٫۱۷۱۶	۰٫۱۳۶۴	۰٫۸۶۳۶	۰٫۱۷۷۳	۰٫۸۲۲۷
	۰٫۸۵۱۸	۰٫۱۷۱۶	۰٫۸۱۷۸	۰٫۱۸۲۲	۰٫۸۰۹۱	۰٫۱۹۰۹
فولد ۱۰	۰٫۸۵۴۲	۰٫۱۴۵۷	۰٫۱۴۷۲	۰٫۸۵۲۸	۰٫۱۵۰۳	۰٫۸۴۹۷
	۰٫۸۳۴۷	۰٫۱۴۵۷	۰٫۸۵۳۷	۰٫۱۴۶۳	۰٫۷۹۹۷	۰٫۲۰۰۳
میانگین	۰٫۸۳۳۸	۰٫۱۶۳۹۱	۰٫۸۲۶۴	۰٫۸۳۷۹	۰٫۸۰۰۴	۰٫۸۰۴۳

$$precision = \frac{TP}{TP + FP} \quad (۵)$$

ب) فراخوانی: به معنی نسبتی از موارد مثبت است که آزمایش آنها را به درستی به عنوان مثبت علامت‌گذاری می‌کند

$$recall = \frac{TP}{TP + FN} \quad (۶)$$

ج) اختصاصیت: به معنی نسبتی از موارد منفی است که آزمایش آنها را به درستی به عنوان منفی علامت‌گذاری می‌کند

$$Specificity = \frac{TN}{TN + FP} \quad (۷)$$

به طور کلی، نتایج از دو بخش تشکیل شده‌اند. در بخش نخست، نتایج کمی در قالب جداول حالات طبقه‌بندی دو کلاسی و چندکلاسی دیده می‌شود. همچنین در بخش دوم به ارائه نتایج کیفی می‌پردازیم. در جداول ۲ تا ۷ به ترتیب نتایج حاصل از سه مدل پیشنهادی به ازای تغییر تعداد نرون‌ها نشان داده شده است. این نتایج برای تشخیص بیماری نارسایی قلبی در حالت دو کلاسی و چندکلاسی که برای محاسبه ماتریس درهم‌ریختگی و به تبع آن دقت، فراخوانی و اختصاصیت استفاده شده، برای همه نمونه‌های مورد نظر برآورد گردیده است. از سویی این دقت‌ها،

اصلی‌ترین دلیل استفاده از الگوریتم جهش ترکیبی قورباغه آن است که می‌تواند در یافتن بهینه سراسری کارآمد واقع گردد و از این رو کمینه خطای حاصل از طبقه‌بندی را در فضای جواب به دست آورد. نهایتاً برای دستیابی به نتایج بهتر، این الگوریتم با بهره‌گیری از مکانیزم وزن‌دهی پویا و استفاده از عملیات جهش و ترکیب با الگوریتم ژنتیک برای تنوع‌دادن به جمعیت راه‌حل‌ها بهینه می‌شود.

۵- نتایج و تفسیر

در این بخش به ارائه نتایج و تفسیر آنها می‌پردازیم و خروجی‌ها به صورت کمی و کیفی تحلیل خواهند شد.

۵-۱ محک‌های ارزیابی

در حالت دو کلاسی، محک‌های ارزیابی نظیر دقت، فراخوانی، $F1$ -score، AUC، اختصاصیت و منحنی ROC برای تحلیل سیستم‌های پیشنهادی، محاسبه خواهد شد. در حالت پنج کلاسی نیز از معیار Accuracy بهره می‌بریم.

الف) دقت: اولین عامل بررسی مدل پیشنهادی، دقت است که بر اساس آن به طراحی بهینه سیستم‌های گوناگون می‌پردازند. این معیار به معنی نزدیکی مقادیر اندازه‌گیری به یکدیگر است

جدول ۵: نتایج حاصل از مدل دوم در مرحله آزمایش به روش تقسیم داده‌ها (۳۰/۷۰) مرکب از اعمال تعداد متوسط ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی پنج‌کلاسی.

Accuracy	کلاس ۵	کلاس ۴	کلاس ۳	کلاس ۲	کلاس ۱	
۰.۶۸	۰	۰.۰۶	۰.۱۴	۰.۱۲	۰.۶۸	کلاس ۱
۰.۶۵	۰	۰.۱۸	۰.۱۲	۰.۶۵	۰.۰۵	کلاس ۲
۰.۶۲	۰.۰۲	۰.۱۳	۰.۶۲	۰.۱۶	۰.۰۷	کلاس ۳
۰.۵۹	۰.۱۲	۰.۵۹	۰.۱۶	۰.۰۹	۰.۰۴	کلاس ۴
۰.۶۸	۰.۶۸	۰.۱۰	۰.۰۹	۰.۰۴	۰.۰۹	کلاس ۵
۰.۶۴۴						دقت نهایی

جدول ۶: نتایج حاصل از پیاده‌سازی مدل سوم در مرحله آزمایش به روش تقسیم داده‌ها با CV برابر ۱۰ مرکب از اعمال تعداد کمیته ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی دوکلاسی.

فولد آزمون	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت	ماتریس درهم‌ریختگی	دقت	انتخاب تعداد کم نورون در لایه مخفی (۳ تا ۶ نورون)	انتخاب تعداد نورون متوسط لایه مخفی (۷ تا ۱۱)	انتخاب تعداد نورون زیاد در لایه مخفی (۱۲ تا ۱۸)
فولد ۲	۰.۱۸۸۷	۰.۸۰۲۸	۰.۱۳۰۵	۰.۸۶۹۵	۰.۱۹۵۱	۰.۸۰۴۹	۰.۱۹۵۱	۰.۸۶۳۲	۰.۸۰۲۲	۰.۱۹۵۱	۰.۸۰۴۹
فولد ۴	۰.۲۱۲۴	۰.۷۹۹۸	۰.۱۵۴۲	۰.۸۴۷۸	۰.۱۸۷۷	۰.۸۱۲۳	۰.۱۸۷۷	۰.۸۵۱۸	۰.۸۱۷۲	۰.۱۸۷۷	۰.۸۱۲۳
فولد ۶	۰.۲۰۳۵	۰.۷۸۴۵	۰.۱۲۳۵	۰.۸۷۶۵	۰.۱۸۰۱	۰.۸۱۹۹	۰.۱۸۰۱	۰.۸۷۴۶	۰.۸۱۷۸	۰.۱۸۰۱	۰.۸۱۹۹
فولد ۸	۰.۲۲۱۱	۰.۸۰۲۸	۰.۱۸۶۰	۰.۸۶۰۳	۰.۱۷۹۲	۰.۸۲۰۸	۰.۱۷۹۲	۰.۸۵۶۷	۰.۸۲۲۱	۰.۱۷۹۲	۰.۸۲۰۸
فولد ۱۰	۰.۲۱۱۰	۰.۷۹۱۰	۰.۱۲۶۲	۰.۸۷۳۸	۰.۲۲۸۸	۰.۷۷۱۲	۰.۲۲۸۸	۰.۸۷۱۰	۰.۷۷۹۸	۰.۲۲۸۸	۰.۷۷۱۲
میانگین	۰.۷۸۷۹	۰.۷۹۰۶	۰.۷۹۱۴	۰.۸۵۹۸	۰.۸۶۶۲	۰.۸۶۲۷	۰.۸۰۹۶	۰.۸۶۲۷	۰.۸۱۱۹	۰.۸۰۹۶	۰.۸۱۲۴

جدول ۷: نتایج حاصل از پیاده‌سازی مدل دوم در مرحله آزمایش به روش تقسیم داده‌ها (۳۰/۷۰) مرکب از اعمال تعداد کمیته ویژگی‌ها و طبقه‌بندی با مدل پیشنهادی طبقه‌بندی پنج‌کلاسی.

Accuracy	کلاس ۵	کلاس ۴	کلاس ۳	کلاس ۲	کلاس ۱	
۰.۶۴	۰.۰۴	۰.۰۶	۰.۱۴	۰.۱۲	۰.۶۴	کلاس ۱
۰.۶۰	۰.۰۳	۰.۱۷	۰.۱۲	۰.۶۰	۰.۰۸	کلاس ۲
۰.۵۸	۰.۰۳	۰.۱۳	۰.۵۸	۰.۱۴	۰.۱۲	کلاس ۳
۰.۵۴	۰.۰۹	۰.۵۴	۰.۱۵	۰.۱۲	۰.۰۱	کلاس ۴
۰.۶۲	۰.۶۲	۰.۱۲	۰.۱۱	۰.۰۷	۰.۰۸	کلاس ۵
۰.۵۹۶						دقت نهایی

مطابق با روش ارزیابی K -fold، مشابه با آنچه در کلاس‌های مرجع پایگاه داده به دست آمده، محاسبه شده است؛ بدین ترتیب که بهترین دقت در طی هر بار تکرار fold برای K برابر با ۱۰ ثبت می‌گردد. طبقه‌بندی در حالت دوکلاسی به روش تقسیم K -fold می‌باشد که مشخص می‌کند آیا بیماری رخ داده است یا خیر. اما در حالت چندکلاسی تقسیم داده‌ها به شیوه‌ای Hold-out در سه وضعیت تقسیم داده می‌باشد.

۵-۲ نتایج کمی

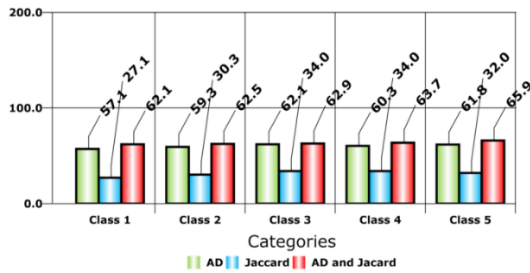
در این بخش، جهت مقایسه عملکرد، حالات زیر در نظر گرفته می‌شود و مشاهده خواهد شد که نتایج با استفاده از تکنیک‌های پیشنهادی چه تفاوتی با یکدیگر خواهند داشت:

- i. مدل اول: در حالت دوکلاسی با بیشینه ویژگی‌ها و تقسیم‌بندی داده به شیوه K -fold ($CV = 10$) و طبقه‌بندی با شبکه پیشنهادی و تقسیم‌بندی داده‌ها به شیوه ۳۰٪ به ۷۰٪ است.
- ii. مدل دوم: در حالت دوکلاسی با متوسط ویژگی‌ها و تقسیم‌بندی داده به شیوه K -fold ($CV = 10$) و طبقه‌بندی با شبکه

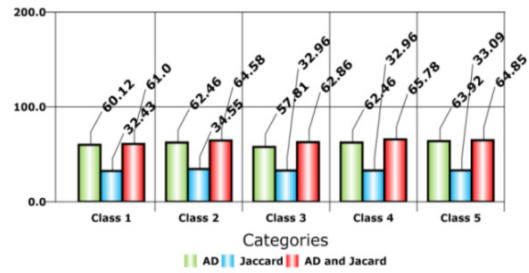
پیشنهادی و تقسیم‌بندی داده‌ها به شیوه ۵۰٪ به ۵۰٪ است.
 iii. مدل سوم: در حالت دوکلاسی با کمیته ویژگی‌ها و تقسیم‌بندی داده به شیوه K -fold ($CV = 10$) و طبقه‌بندی با شبکه پیشنهادی و تقسیم‌بندی داده‌ها به شیوه ۷۰٪ به ۳۰٪ است.

الف) مدل اول

در حالت دوکلاسی، این مدل مرکب از تعداد ویژگی‌های بیشینه و طبقه‌بندی‌کننده ترکیبی در مرحله نهایی است. در جدول ۲ نتایج حاصل از پیاده‌سازی این مدل به نمایش درآمده است. در این جدول، خروجی‌ها برای حالت دوکلاسی برآورد شده‌اند. در سطر آخر این جدول، اختصاصیت، فراخوانی و دقت محاسبه شده‌اند. همان‌گونه که مشاهده می‌شود، زمانی که ویژگی‌ها زیاد باشد، مدل پیشنهادی در وضعیت تعداد نرون‌های متوسط دارای دقت بهتری است. همچنین در حالت پنج‌کلاسی می‌توان خروجی‌ها را به شیوه تقسیم ۷۰٪ آموزش در برابر ۳۰٪ آزمایش به دست آورد. دقت بهینه به ازای تعداد نرون‌های متوسط در این جدول محاسبه شده است. سایر مقادیر به‌طور متوسط از مقدار متوسط گزارش شده در جدول ۳ کمتر بودند.



شکل ۵: اثر روش Relief در انتخاب ویژگی‌ها با ارزیابی محک‌های جاکارد و درجه متوسط و ترکیب آنها.



شکل ۳: اثر روش Wrapper در انتخاب ویژگی‌ها با ارزیابی محک‌های جاکارد و درجه متوسط و ترکیب آنها.

تا حد زیادی بر دقت دسته‌بندی نهایی این داده و گریز از مشکلاتی چون بیش‌برازش یا زیربرازش مشخص است. در جداول ۶ و ۷ شاهد به نمایش درآمدن نتایج حاصل از مدل سوم روش پیشنهادی هستیم.

به نظر می‌رسد ثبات یا مقاومت بودن الگوریتم در وضعیت تعداد نرون‌های متوسط بالاتر باشد؛ چراکه پراکندگی میان فاکتورهای محک در ارزیابی تغییرات اندکی دارد. در وضعیت انتخاب ویژگی‌های کم (حدود ۳۰٪)، دقت در حالات تعداد نرون‌های کم و متوسط و زیاد به ترتیب ۴٪، ۴٪ افزایش و ۱٪ افزایش داشته است. همچنین اختصاصیت به ترتیب ۵٪، ۳٪ افزایش و ۰.۵٪ افزایش داشته‌اند. فراخوانی هم مانند دو عامل ارزیابی دیگر به ترتیب ۴٪، ۳٪ افزایش و ۲٪ کاهش داشته است. آنچه مشهود است می‌توان برای وضعیت‌های انتخاب متوسط و کم ویژگی برای تعداد نرون‌های کم و متوسط مدل را طراحی کرد، اما در وضعیت تعداد نرون متوسط برای شبکه عصبی با امکان یادگیری حداکثری و تعداد ویژگی کم می‌توان به بالاترین دقت دست یافت؛ هرچند در حالت چندکلاسی، بالاترین سطح دقت مربوط به تعداد ویژگی‌های متوسط بوده است. در مدل سوم، افت دقتی معادل ۵ درصدی نسبت به حالت طبقه‌بندی چندکلاسی و تعداد ویژگی کم گزارش شده است.

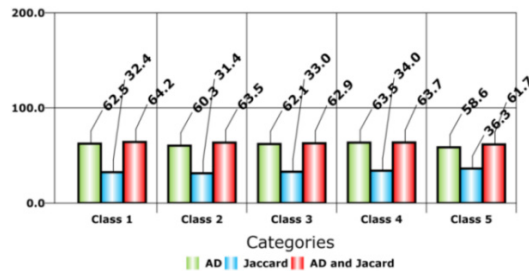
۳-۵ نتایج کیفی

محک‌های جاکارد^۱ و درجه متوسط^۲ (AD) از جمله محک‌های ارزیابی ویژگی‌های انتخاب‌شده، توسط الگوریتم پیشنهادی ترکیبی در نظر گرفته شده‌اند؛ به این معنا که هر آنچه با نام ویژگی در پیش از ورود به طبقه‌بندی به دست می‌آید تا چه اندازه اثر مناسب‌تری به دنبال دارد. بر اساس محاسبه مقادیر ذکر شده می‌توان سطح اهمیت ویژگی‌های انتخاب‌شده را به دست آورد.

ضریب جاکارد شباهت میان مجموعه‌های نمونه محدود را اندازه‌گیری می‌کند و به‌عنوان اندازه اشتراک تقسیم بر اندازه اجتماع مجموعه‌های نمونه تعریف می‌شود

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (۸)$$

که A و B دو مجموعه ویژگی استخراج‌شده هستند و عملاً معیار جاکارد هرچه به ۱ نزدیک‌تر باشد، بهینه‌تر خواهد بود. به نوعی می‌تواند همبستگی میان ویژگی‌ها را نمایش دهد که در شکل‌های ۳ تا ۵ نیز مورد توجه قرار می‌گیرد. همچنین درجه متوسط (AD) به تعداد پیوندهایی که یک گره در یک شبکه به گره‌های دیگر وصل می‌کند، اشاره دارد. در یک شبکه غیرمستقیم تعداد کل پیوندهای L را می‌توان به‌صورت (۹) تعریف کرد



شکل ۴: اثر روش فیلتری عام در انتخاب ویژگی‌ها با ارزیابی محک‌های جاکارد و درجه متوسط و ترکیب آنها.

ب) مدل دوم

مدل دوم مرکب از بردار ویژگی با تعداد ویژگی‌های متوسط است. همچنین از طبقه‌بندی توسط روش شبکه عصبی با امکان یادگیری حداکثری بهینه‌شده با تعداد نرون‌های متغیر استفاده شده است. نتایج حاصل از این پیاده‌سازی در جداول ۴ و ۵ در دو حالت طبقه‌بندی دوکلاسی و طبقه‌بندی چندکلاسی قابل مشاهده است.

در هر سه حالت تعداد نرون، محک فراخوانی نسبت به حالت‌های قبلی افزایش محسوسی داشت. افزایش اختصاصیت نیز با هر سه مدل از تعداد نرون یکسان نبود. در حالت تعداد نرون‌های کم، هر سه فاکتور افزایش پیدا کرده‌اند و به عبارت بهتر در این وضعیت (تعداد ویژگی متوسط و تعداد نرون‌های کم) بالاترین سطح طبقه‌بندی به دست آمده است. در حالت انتخاب ویژگی‌های متوسط، نسبت به حالت انتخاب حداکثر ویژگی‌ها و با وجود تعداد نرون‌های متوسط، اختصاصیت و دقت کاهش ۱ درصدی داشته‌اند، اما فراخوانی افزایش داشته است؛ یعنی نمونه‌های بیمار بهتر شناسایی شده‌اند. به همین ترتیب در وضعیت انتخاب ویژگی‌ها متوسط و تعداد نرون‌های بالا نسبت به انتخاب ویژگی‌ها بالا، سطح دقت و اختصاصیت و نیز فراخوانی افزایش یافته، اما نسبت به دیگر حالات انتخاب تعداد نرون، هنوز دقت آن کمتر است. پرواضح است که اثر الگوریتم و انتخاب تعداد نرون و نیز دخالت ویژگی در افزایش دقت مشهود است. از میان تعداد متوسط از ویژگی‌ها، در فاصله ۵۰ درصدی ویژگی، بیشترین دقت‌ها با تعداد نرون کم و متوسط حاصل آمد و لذا ابعاد ویژگی‌های انتخاب‌شده برای هر داده مربوط به تشخیص بیماری قلبی به تعداد بسیار کمی از ویژگی‌ها تقلیل یافت. به همین ترتیب افزایش چند درصدی دقت (حدود ۶٪) برای حالت طبقه‌بندی چندکلاسه دیده می‌شود که در جدول ۵ به نمایش درآمده است.

ج) مدل سوم

مدل سوم ترکیبی از مراحل پیش‌پردازش داده‌ها، اعمال حداقل ویژگی‌ها (به عبارتی ۳۰٪ کل ویژگی‌ها) و طبقه‌بندی به مدل شبکه عصبی با امکان یادگیری حداکثری در وضعیت تعدادهای مختلف نرون در تک‌لایه این شبکه در حالات دوکلاسه و چندکلاسی است. اثر طبقه‌بندی

1. Jaccard
2. Average Degree

جدول ۸: مقایسه عملکرد مدل‌های یادگیری ماشین و بهینه‌سازی‌شده بر روی مجموعه داده‌های مختلف در تشخیص بیماری قلبی.

مجموعه داده	روش	صحت (Accuracy)	دقت (Precision)	فراخوانی (Recall)	امتیاز F1	سطح زیر منحنی ROC (AUC)
Cleveland Heart Disease dataset	روش ترکیبی GA + SFLA	۰.۹۲۳۳	۰.۹۴۸۷	۰.۹۰۲۴	۰.۹۲۵۰	۰.۹۳۹۸
	بهینه‌سازی GA	۰.۹۲۲۲	۰.۸۶۹۶	۰.۹۷۵۶	۰.۹۱۹۵	۰.۹۴۳۳
	Naive Bayes	۰.۹۰۰۰	۰.۹۲۱۱	۰.۸۵۲۷	۰.۸۸۶۱	۰.۹۶۰۷
	رگرسیون لجستیک	۰.۸۸۸۹	۰.۸۹۷۴	۰.۸۵۳۷	۰.۸۷۵۰	۰.۹۳۹۳
	SVM	۰.۸۸۸۹	۰.۹۴۲۹	۰.۸۰۴۹	۰.۸۶۸۴	۰.۹۳۵۸
	جنگل تصادفی (Random Forest)	۰.۸۵۵۶	۰.۸۵۰۰	۰.۸۲۹۳	۰.۸۳۹۵	۰.۹۳۷۸
	LightGBM	۰.۸۳۳۳	۰.۸۰۹۵	۰.۸۲۹۳	۰.۸۱۹۳	۰.۹۰۰۹
	XGBoost	۰.۸۲۲۲	۰.۸۲۰۵	۰.۷۸۰۵	۰.۸۰۰۰	۰.۹۰۴۹
	K نزدیک‌ترین همسایه (KNN)	۰.۸۴۴۴	۰.۸۸۵۷	۰.۷۵۶۱	۰.۸۱۵۸	۰.۹۲۸۶
	درخت تصمیم‌گیری	۰.۷۰۰۰	۰.۶۴۵۸	۰.۷۵۶۱	۰.۶۹۶۶	۰.۷۰۴۶
Statlog Heart dataset	روش ترکیبی GA + SFLA	۰.۹۱۳۶	۰.۹۶۳۰	۰.۸۱۲۵	۰.۸۸۱۴	۰.۹۱۳۳
	بهینه‌سازی GA	۰.۹۰۱۲	۰.۹۲۸۶	۰.۸۱۲۵	۰.۸۶۶۷	۰.۹۱۳۳
	Naive Bayes	۰.۸۳۹۵	۰.۸۸۰۰	۰.۶۸۷۵	۰.۷۷۱۹	۰.۸۹۴۱
	رگرسیون لجستیک	۰.۸۲۷۲	۰.۸۷۵۰	۰.۶۵۶۲	۰.۷۵۰۰	۰.۸۸۳۹
	SVM	۰.۸۲۷۲	۰.۸۲۱۴	۰.۷۱۸۸	۰.۷۶۶۷	۰.۸۹۷۳
	جنگل تصادفی (Random Forest)	۰.۷۶۵۴	۰.۷۶۰۰	۰.۵۹۳۸	۰.۶۶۶۷	۰.۸۷۷۹
	LightGBM	۰.۸۱۴۸	۰.۸۶۹۶	۰.۶۲۵۰	۰.۷۳۷۳	۰.۹۰۵۰
	XGBoost	۰.۷۷۷۸	۰.۷۹۱۷	۰.۵۹۳۸	۰.۶۷۸۶	۰.۸۸۵۸
	K نزدیک‌ترین همسایه (KNN)	۰.۷۹۰۱	۰.۸۰۰۰	۰.۶۲۵۰	۰.۷۰۱۸	۰.۸۹۳۸
	درخت تصمیم‌گیری	۰.۷۱۶۰	۰.۶۶۶۷	۰.۵۶۲۵	۰.۶۱۰۲	۰.۶۸۹۴
MIT-BIH Arrhythmia dataset	روش ترکیبی GA + SFLA	۰.۸۵۲۵	۰.۸۸۹۸	۰.۸۶۲۷	۰.۸۷۶۰	۰.۹۲۷۳
	Naive Bayes	۰.۷۵۴۱	۰.۷۸۵۷	۰.۷۵۵۹	۰.۷۷۰۵	۰.۸۲۳۲
	رگرسیون لجستیک	۰.۸۳۶۱	۰.۸۷۹۳	۰.۸۲۳۵	۰.۸۵۰۵	۰.۸۹۹۵
	SVM	۰.۷۷۰۵	۰.۸۰۶۱	۰.۷۶۴۷	۰.۷۸۵۰	۰.۸۷۶۰
	جنگل تصادفی (Random Forest)	۰.۷۷۰۵	۰.۷۸۹۵	۰.۷۶۴۷	۰.۷۷۷۰	۰.۸۳۴۹
	LightGBM	۰.۶۱۹۰	۰.۸۸۹۱	۰.۵۸۸۲	۰.۷۱۴۳	۰.۸۳۸۲
	XGBoost	۰.۷۷۰۵	۰.۸۰۶۱	۰.۷۶۴۷	۰.۷۸۵۰	۰.۸۸۴۳
	K نزدیک‌ترین همسایه (KNN)	۰.۷۷۰۵	۰.۸۰۶۱	۰.۷۶۴۷	۰.۷۸۵۰	۰.۸۶۴۷
	درخت تصمیم‌گیری	۰.۷۲۱۳	۰.۷۴۳۶	۰.۷۳۵۳	۰.۷۳۹۴	۰.۷۴۸۶

است، در حالی که در روش‌های دیگر این اتفاق رخ نداده است. همچنین در الگوریتم‌های کمی تقسیم داده‌ها به هر دو شیوه تحلیل شده‌اند.

۴-۵ مقایسه با روش‌های پایه

در این بخش، روش پیشنهادی با توجه به معیارهای ارزیابی با روش‌های پایه مورد مقایسه قرار می‌گیرد. نتایج مقایسه بر روی سه مجموعه داده مختلف در جدول ۸ و شکل ۶ نشان داده شده است. با توجه به نتایج ارائه‌شده از مجموعه داده Cleveland، مشاهده می‌شود که روش Hybrid GA + SFLA با وجود زمان اجرای پایین‌تر (۰.۱۵) نسبت به GA تنها (۱۹.۲)، بهترین دقت کلی و F1-score را ارائه داده است. مدل Naive Bayes علی‌رغم سادگی، AUC بالایی نشان داده که بیانگر توان بالای تفکیک کلاس‌ها است. SVM و Logistic Regression نیز عملکرد قابل قبولی داشته‌اند. با این حال، بهینه‌سازی پارامترها با GA و SFLA بهبود محسوسی در دقت و فراخوانی ایجاد کرده است. مدل‌های پیچیده مانند LightGBM و XGBoost در این مجموعه داده عملکردی متوسط از خود نشان داده‌اند که ممکن است ناشی از ابعاد و حجم داده باشد. نتایج نسبتاً مشابهی در مجموعه داده‌های Statlog Heart و MIT-BIH Arrhythmia تأییدی

$$L = \frac{1}{N} \sum_{i=1}^n k_i \quad (9)$$

در این شبکه، L تعداد پیوندهای میان ویژگی‌ها، k_i درجه هر گره i ام و n اندازه شبکه است. حال با استفاده از این دو محک و نیز ترکیب هر دو می‌توان میزان اثرگذاری ویژگی‌ها را مطرح کرد. با توجه به اینکه بالاترین میزان دقت در وضعیت تعداد نرون متوسط و به ازای کمینه تعداد ویژگی به دست آمد، در این بخش نیز از این تنظیمات استفاده شده است. در شکل‌های ۳ تا ۵ برای حالت پنج کلاسه و برآورد تک‌تک کلاس‌ها، مقادیر محک جاکارد و درجه متوسط به حالت گرافی مشاهده می‌گردد.

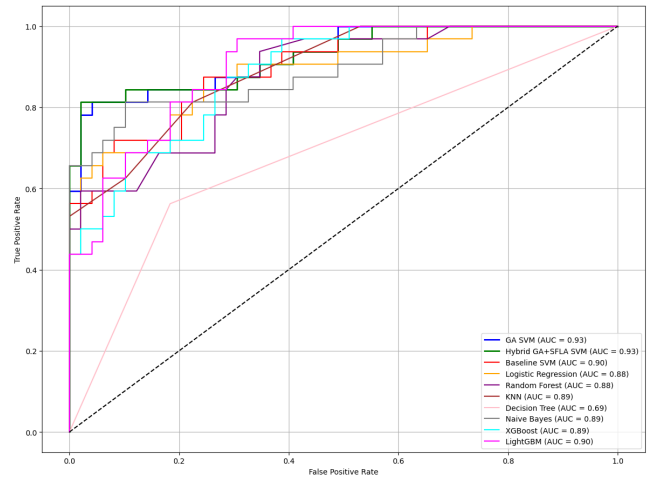
به‌طور خلاصه می‌توان گفت در مقایسه با روش‌های [۱۸] تا [۲۷] و [۳۵] تا [۳۸] که در بخش ۲ مورد بررسی و تحلیل قرار گرفت، مشاهده می‌شود که دقت روش پیشنهادی چه در حالت دوکلاسی و چه در حالت چندکلاسی، سطح تشخیصی مناسب‌تری به دست آورده است. از آنجا که تعداد داده‌های کلینیکی دریافت‌شده از همه پایگاه‌های داده تقریباً برابر است، لیکن تشخیص دقیق بیماری و سطح خطای کمینه تا حدود زیادی وابسته به مدل است. در قیاس با راهکارهای پیشین باید توجه داشت روش پیشنهادی به‌واسطه ساختار متوسطی که دارد، زمان کمتری را در تحلیل و جداسازی داراست. ورودی‌های آن مشخص و دارای ابعاد کمتری

مبتلایان به نارسایی قلبی برآورد نماییم. سیستم یکپارچه معرفی شده در این مقاله، سازوکاری بود که بر اساس آن به بازشناسی بیماری‌های مختلفی چون نارسایی قلبی اقدام نموده و داده پایه ما از مجموعه داده‌های سایت معروف UCI گردآوری شد. الگوریتم بازشناسی در این مقاله، جهت برآورده کردن هدف از محاسبات نرم الهام گرفت. در کمتر تحقیقی، اتکا بر آنالیز وابستگی شدید ویژگی‌ها و به دور از دقت کم صورت پذیرفته است. مراحل کار از مرحله پیش‌پردازش و پالایش ویژگی با مقاردهی مقادیر از دست‌رفته و نیز نرمال‌سازی ویژگی‌ها، انتخاب ویژگی با بهره‌گیری از روش تلفیقی که متشکل از سه روش Wrapper، فیلتری عام و Relief است و طبقه‌بندی به شیوه ماشین یادگیری حداکثری با کرنل موجک بهینه‌شده توسط الگوریتم قورباغه تشکیل شده بود. سپس در چند مرحله، محک‌ها ارزیابی شدند تا اثبات گردد می‌توان از طریق تحلیل متکی بر یادگیری ماشین در داده‌های مربوط به نارسایی قلبی، شناسایی و پیش‌بینی این بیماری را امکان‌پذیر نمود. مشخص شد که استفاده از تکنیک رأی‌گیری نرم در قالب یک راهکار بهینه تلفیقی ویژگی‌ها تا حد زیادی بر افزایش منطقی دقت اثرگذار بوده است. طبقه‌بندی پیشنهادی در این زمینه نیز از طریق یافتن پارامترهای شبکه عصبی با امکان یادگیری حداکثری و کرنل موجک آن و نیز گریز از بیش‌برازش تا حد زیادی دقت را بهینه نموده است.

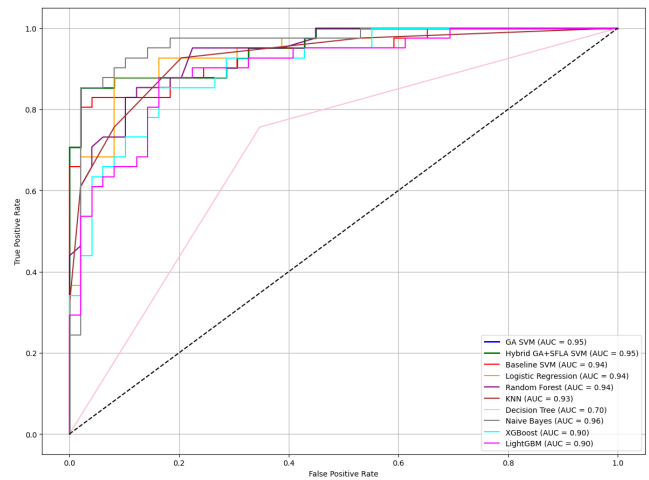
با این حال، داده‌های قلبی و داده‌های مشابه با آن (به‌ویژه برای حالتی که بیش از ۵ کلاس باشد) دارای مقادیر از دست‌رفته بسیار زیادی هستند. بنابراین جهت تحقیقات آینده، پیشنهاد می‌شود راهکار یا تدبیری در مرحله پیش‌پردازش (علاوه بر تکنیک پیشنهادی) ارائه شود که بتوان تا حد قابل ملاحظه‌ای دقت واقعی در تشخیص را بهینه نماید. همچنین پیشنهاد می‌شود که جداسازی سیگنال اصلی از سایر مؤلفه‌ها دقیق‌تر باشد؛ به طوری که ایجاد فضای ویژگی بر پایه روش‌های مختلفی انجام پذیرد. نهایتاً جهت تحقیقات آتی بررسی طبقه‌بندی دیگر جهت تشخیص بیماری پیشنهاد می‌گردد.

مراجع

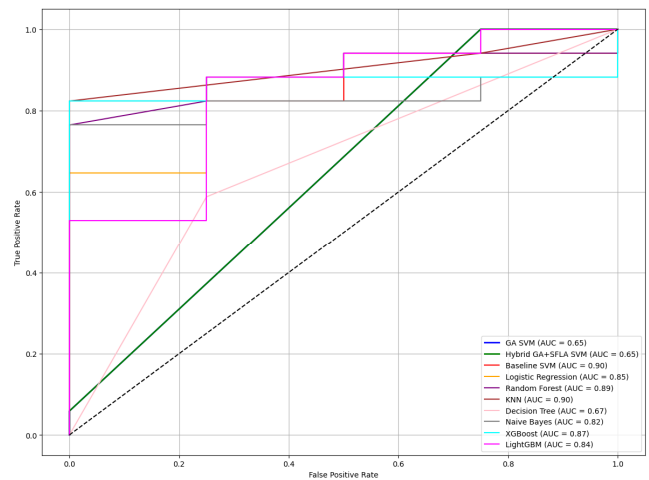
- [1] E. J. Nelwan, E. Widjajanto, S. Andarini, and M. S. Djati, "Modified risk factors for coronary heart disease (CHD) in Minahasa ethnic group from Manado city Indonesia," *J. of Experimental Life Science*, vol. 6, no. 2, pp. 88-94, Apr. 2017.
- [2] C. J. Taylor and J. Moore, "NICE chronic heart failure update guideline 2018," *Primary Care Cardiovascular J.*, vol. 3, no. 9, pp. 1-3, Apr. 2019.
- [3] S. Citlik-Saritas, S. Saritas, R. Cevik-Akyil, and K. Isik, "The effects of Turkish classical music on physiological parameters, pain and analgesic use in patients with myocardial infarction: a non-randomized controlled study," *Eur. J. of Integrative Medicine*, vol. 22, pp. 50-53, Sept. 2018.
- [4] M. Adam, et al., "Automated characterization of cardiovascular diseases using relative wavelet nonlinear features extracted from ECG signals," *Comput. Methods and Programs in Biomedicine*, vol. 161, pp. 133-143, Jul. 2018.
- [5] M. Kolahkaj, A. Harounabadi, and M. Sadeghzade, "A recommender system for web mining using neural network and fuzzy algorithm," *Int. J. of Computer Applications*, vol. 78, no. 8, pp. 20-24, Sept. 2013.
- [6] م. کلاهکج، "ارائه سیستم بازیابی تصاویر مبتنی بر محتوا با بهره‌گیری از یادگیری نیمه‌نظارت‌شده و کاوش الگوهای مکرر انجمنی،" *نشریه مهندسی برق و مهندسی کامپیوتر ایران*، ب- مهندسی کامپیوتر، سال ۲۰، شماره ۳، صص. ۲۴۵-۲۵۲، پاییز ۱۴۰۱.
- [7] K. Oh, Z. Li, B. S. Oh, and K. A. Toh, "Optimizing between data transformation and parametric weighting for stable binary classification," *J. of the Franklin Institute*, vol. 355, no. 4, pp. 1614-1637, Mar. 2018.



(الف)



(ب)



(ج)

شکل ۶: مقایسه منحنی ROC برای روش‌های مختلف بر روی سه مجموعه داده.

بر تعمیم‌پذیری و پایداری مدل پیشنهادی است.

۶- نتیجه‌گیری و کارهای آتی

شناخت دقیق و خودکار برچسب‌های وجود یا عدم وجود بیماری قلبی در مجموعه الگوریتم معرفی شده در این مقاله می‌تواند در ایجاد ارتباط مناسب میان فرد و ماشین و بررسی تغییر میزان اشتباهات پزشک متخصص و متخصص آزمایشگاهی شمر باشد؛ به‌گونه‌ای که قادر خواهیم بود با تفکیک صحیح این دو کلاس، امکان شدت بیماری را در میان

- [26] H. Wang, P. Li, Y. Zheng, K. Jiang, and Y. Xu, "Sparse pinball universum nonparallel support vector machine and its safe screening rule," *Appl. Intelligence*, vol. 55, no. 6, pp. 563-580, Apr. 2025.
- [27] C. T. Tran, M. Zhang, P. Andreae, B. Xue, and L. T. Bui, "An effective and efficient approach to classification with incomplete data," *Knowledge-Based Systems*, vol. 154, pp. 1-16, Aug. 2018.
- [28] A. K. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *Proc. 38th Int. Convention on Information and Communication Technology, Electronics and Microelectronics*, pp. 1200-1205, Opatija, Croatia, 25-29 May 2015.
- [29] J. Hamidzadeh, Z. Mehravaran, and A. Harati, "Feature selection by utilizing kernel-based fuzzy rough set and entropy-based non-dominated sorting genetic algorithm in multi-label data," *Knowledge and Information Systems*, vol. 67, no. 4, pp. 3789-3819, Apr. 2025.
- [30] S. Narayanamoorthy, S. Geetha, R. Rakkiyappan, and Y. H. Joo, "Interval-valued intuitionistic hesitant fuzzy entropy based VIKOR method for industrial robots' selection," *Expert Systems with Applications*, vol. 121, pp. 28-37, May 2019.
- [31] I. Kadhim Ajlan, H. Murad, A. A. Salim, and A. Fadhil Bin Yousif, "Extreme learning machine algorithm for breast cancer diagnosis," *Multimedia Tools and Applications*, vol. 84, pp. 14739-14758, 2024.
- [32] X. Zhang, X. Hu, G. Cui, Y. Wang, and Y. Niu, "An improved shuffled frog leaping algorithm with cognitive behavior," in *Proc. 7th World Congress on Intelligent Control and Automation*, pp. 6197-6202, Chongqing, China, 25-27, Jun. 2008.
- [33] UCI Machine Learning Repository, *Heart Disease*, 1988, available at <https://www.archive.ics.uci.edu/ml/datasets/Heart+Disease>
- [34] G. Moody and R. Mark, *MIT-BIH Arrhythmia Database*, 2025, available at <https://physionet.org/content/mitdb/1.0.0/>
- [35] J. Wu, S. Pan, X. Zhu, Z. Cai, P. Zhang, and C. Zhang, "Self-adaptive attribute weighting for Naive Bayes classification," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1487-1502, Feb. 2015.
- [36] P. Shunmugapriya and S. Kanmani, "A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid)," *Swarm and Evolutionary Computation*, vol. 36, pp. 27-36, Oct. 2017.
- [37] X. J. Shen, Y. Dong, J. P. Gou, Y. Z. Zhan, and J. Fan, "Least squares kernel ensemble regression in reproducing kernel Hilbert space," *Neurocomputing*, vol. 311, pp. 235-244, Oct. 2018.
- [38] C. Yang and X. C. Yin, "Diversity-based random forests with sample weight learning," *Cognitive Computation*, vol. 11, no. 5, pp. 685-696, Oct. 2019.
- [8] M. Kolahhaj, "An image retrieval approach based on feature extraction and self-supervised learning," in *Proc. 2nd Int. Conf. on Distributed Computing and High-Performance Computing*, pp. 46-51, Qom, Iran, 2-3 Mar. 2022.
- [9] C. Berry, D. R. Murdoch, and J. J. McMurray, "Economics of chronic heart failure," *Eur. J. of Heart Failure*, vol. 3, no. 3, pp. 283-291, Jun. 2001.
- [10] J. C. de la Torre, "Hemodynamic instability in heart failure intensifies age-dependent cognitive decline," *J. of Alzheimer's Disease*, vol. 76, no. 1, pp. 63-84, May 2020.
- [11] H. Linusson, U. Johansson, H. Boström, and T. Löfström, "Classification with reject option using conformal prediction," in *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, pp. 94-105, Melbourne, Australia, 3-6 Jun. 2018.
- [12] J. Qiu, J. Xie, D. Zhang, and R. Zhang, "A robust twin support vector machine based on fuzzy systems," *Int. J. of Intelligent Computing and Cybernetics*, vol. 17, no. 1, pp. 101-125, Feb. 2024.
- [13] B. Sahmadi and D. Boughaci, "Hybrid genetic algorithm with SVM for medical data classification," in *Proc. Int. Conf. on Applied Smart Systems*, 6 pp., Medea, Algeria, 24-25 Nov. 2018.
- [14] S. Chen, J. Cao, F. Chen, and B. Liu, "Entropy-based fuzzy least squares twin support vector machine for pattern classification," *Neural Processing Letters*, vol. 51, no. 1, pp. 41-66, Feb. 2020.
- [15] Z. Zainuddin, K. H. Lai, and P. Ong, "An enhanced harmony searches-based algorithm for feature selection: applications in epileptic seizure detection and prediction," *Computers & Electrical Engineering*, vol. 53, pp. 143-162, Jul. 2016.
- [16] M. Nekkaa and D. Boughaci, "Hybrid harmony search combined with stochastic local search for feature selection," *Neural Processing Letters*, vol. 44, no. 1, pp. 199-220, Aug. 2016.
- [17] D. Karaboga and C. Ozturk, "A novel clustering approach: artificial bee colony (ABC) algorithm," *Appl. Soft Computing*, vol. 11, no. 1, pp. 652-657, Jan. 2011.
- [18] P. Tapkan, L. Özbakir, S. Kulluk, and A. Baykasoğlu, "A cost-sensitive classification algorithm: BEE-Miner," *Knowledge-Based Systems*, vol. 95, pp. 99-113, Mar. 2016.
- [19] X. Lai, Z. Zhang, H. Chen, L. Zhang, Z. Li, and W. Lu, "Tracking-removed neural network with graph information for classification of incomplete data," *Appl. Intelligence*, vol. 55, no. 3, pp. 1-20, Feb. 2025.
- [20] J. Wu, S. Pan, X. Zhu, P. Zhang, and C. Zhang, "Sode: self-adaptive one-dependence estimators for classification," *Pattern Recognition*, vol. 51, pp. 358-377, Mar. 2016.
- [21] X. Zhu, et al., "Confidence guided semi-supervised cross-modality person re-identification," *Pattern Recognition*, vol. 165, Article ID: 111669, Sept. 2025.
- [22] X. Wang, G. Wu, G. Hao, and Z. Zhang, "A novel fuzzy twin support vector machine using mass-based dissimilarity measure," *Knowledge and Information Systems*, vol. 55, no. 5, pp. 4233-4300, Jan. 2025.
- [23] B. Aydılek, "Examining effects of the support vector machines kernel types on biomedical data classification," in *Proc. Int. Conf. on Artificial Intelligence and Data Processing*, 4 pp., Maltaya, Turkey, 28-30 Sept. 2018.
- [24] S. Chen, J. Cao, and Z. Huang, "Weighted linear loss projection twin support vector machine for pattern classification," *IEEE Access*, vol. 7, pp. 57349-57360, 2019.
- [25] S. Lee and C. H. Jun, "Fast incremental learning of logistic model tree using least angle regression," *Expert Systems with Applications*, vol. 97, pp. 137-145, May 2018.

مرجان مطبیعی‌زاده دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر در دانشگاه آزاد اسلامی واحد اهواز است. زمینه‌های تحقیقاتی وی شامل یادگیری ماشین و داده‌کاوی می‌باشد.

مارال کلاه‌کج عضو هیأت علمی دانشگاه آزاد اسلامی و پژوهشگر حوزه مهندسی کامپیوتر و هوش مصنوعی است. زمینه‌های پژوهشی وی شامل سیستم‌های توصیه‌گر، الگوریتم‌های بهینه‌سازی و یادگیری ماشین با تمرکز بر تحلیل داده و تصمیم‌گیری هوشمند می‌باشد. از جمله فعالیت‌های پژوهشی او می‌توان به توسعه مدل‌های ترکیبی و گراف‌محور در سیستم‌های توصیه‌گر و به‌کارگیری روش‌های تکاملی در یادگیری عمیق اشاره نمود.