مقاله پژوهشی

بهبود تشخیص بیماری‌های قلبی- عروقی با استفاده از سیستم تصمیم‌یار هوشمند با رویکرد رایانش مه

Doi: 10.30508/kdip.2023.368618.1061

حجت آزادروش[۱] | حمید طباطبایی[۲]
۱-گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد مشهد، ایران، مشهد

تاریخ دریافت: ۱۴۰۱/۱۱/۰۹
تاریخ پذیرش: ۱۴۰۲/۰۳/۱۲
صفحه: ۰۰ - ۰۰

چکیده

بیشترین نوع بیماری‌های قلبی شامل؛ بیماری مادرزادی قلبی، نارسایی قلبی، کاردیومیوپاتی، بیماری روماتیسمی قلب، تنگی ریوی و بیماری عروق کرونر است. تشخیص بیماری‌های قلبی - عروقی از طریق علائم، یک چالش بزرگ در شرایط جهانی فعلی است. اگر به موقع تشخیص داده نشود، ممکن است عامل مرگ و میر شود. به دلیل دسترسی محدود پزشکان متخصص قلب به مناطق دورافتاده، یک سیستم پشتیبان تصمیم گیری هوشمند با رویکرد رایانش مه می‌تواند به عنوان یک راهکار موثر در بهبود تشخیص بیماری‌های قلبی- عروقی استفاده شود. هدف این مقاله ارائه یک سیستم تصمیم‌یار هوشمند به منظور بهبود تشخیص بیماری‌های قلبی- عروقی براساس رایانش مه است. در ابتدا، اطلاعات و پرونده‌های پزشکی ۱۰۰ بیمار جمع آوری می‌شوند. سپس داده‌های ورودی، باید پاک‌سازی و نرمال‌سازی شوند، سپس بر اساس الگوریتم‌های استخراج ویژگی، ویژگی‌ها استخراج، انتخاب و وزن‌دهی می‌شوند. بعد از آن با استفاده از الگوریتم ماشین بردار پشتیبان، داده‌ها طبقه‌بندی می‌شوند. نتایج نشان می‌دهد که معیار دقت در روش‌های پیشنهادی بالاتر از مقاله پایه است. در روش پیشنهادی ممتیک- ژنتیک دقت ۹۳ درصد و در روش پیشنهادی گراف کاوی- ژنتیک دقت ۹۱ درصد و در مقاله پایه دقت ۸۶ درصد می‌باشد. روش‌های پیشنهادی توانسته سطح بیشتری را نسبت به مقاله پایه در منحنی ROC ارائه شده، پوشش دهد. در روش ممتیک- ژنتیک در نقطه صفر مقدار TPR=۰٫۳۸ بوده و در روش گراف‌کاوی- ژنتیک در نقطه صفر مقدار TPR=۰٫۲ است. درحالی که در مقاله پایه،TPR مقدار آن صفر در نظرگرفته شده است. همچنین روش ممتیک-ژنتیک سریع‌تر می تواند به مقدار یک در TPR برسد. نقطه‌ای که به مقدار ۱ رسیده است، ۰٫۴ می‌باشد، خطای میانگین مربعات روش‌های پیشنهادی عملکرد بهتری نسبت به مقاله پایه داشته و برای روش ترکیبی ممتیک- ژنتیک و گراف کاوی- ژنتیک مقادیر ۱۱و ۱۵درصد را به خود اختصاص داده است.

کلمات کلیدی: سیستم تصمیم یار هوشمند، رایانش مه، سیستم‌های اطلاعاتی، تشخیص بیماری قلبی - عروقی.

۱- مقدمه

بیماری‌های قلبی-عروقی اولین مشکلی است که برای دنیا وجود دارد. بیماری‌های قلبی- عروقی بیش از مرگ افراد در اولین حمله قلبی رخ می‌دهند. اما نه تنها برای حمله قلبی، مشکلاتی برای سرطان سینه، سرطان ریه و بطن وجود دارد. مصرف شیر، و غیره ضروری است که می‌تواند میزان شیوع بیماری‌های قلبی-عروقی را در هزاران نمونه از بین ببرد. در حال حاضر بیماری‌های قلبی- عروقی دلیل اصلی مرگ و میر در جهان است. در طول چند سال گذشته، انواع بیماری از قلب اتفاق می‌افتد. دلایل مختلفی وجود دارد که عامل خطر ابتلا به بیماری‌های قلبی را افزایش می‌دهد. این بیماری بیشتر در مردان به دلیل عادت به سیگار کشیدن دیده می‌شود (علی، کریم و محمد[1]، ۲۰۲۲). شواهد موجود در تغییر سبک زندگی مردم نشان می‌دهد که شیوع بیماری قلبی-عروقی در ایران رو به افزایش است. بیشترین نوع معمول از بیماری‌های قلبی؛ بیماری مادرزادی قلبی، نارسایی قلبی، فشارخون بالا، کاردیومیوپاتی، بیماری روماتیسمی قلب، تنگی ریوی و عروق کرونر بیماری عروق است. اگر عملکرد قلب به درستی انجام نشود و هر کدام از ناراحتی‌های قلبی بروز داده شود به بخش‌های دیگر بدن نیز تأثیر می‌گذارد و بنابراین لازم است که این بیماری سریع تشخیص داده شود (لیمام، زوهاری، و اوسالتی[2]، ۲۰۲۲).

امروزه بسیاری از بیمارستان‌ها از سیستم مدیریت اطلاعات بیمارستان برای مدیریت داده‌های بیماران و یا بهداشت و درمان استفاده می‌کنند. این سیستم‌ها مقدار زیادی داده شامل اعداد، متون، نمودارها و عکس‌ها ایجاد می‌کنند. متاسفانه این داده‌ها به ندرت در تصمیم‌گیری‌های کلینیکی مورد استفاده قرار می‌گیرند. سوالی که ذهن اکثر متخصصان این حوزه را به خود مشغول کرده این است که چگونه می‌توان داده‌ها را تبدیل به اطلاعات مفید نمود که کادر پزشکی را قادر به تصمیمات کلینیکی هوشمندانه نماید. از سوی دیگر، روش‌های داده‌کاوی روی داده‌ها به دنبال تایید آنچه از قبل وجود دارد، نیستند بلکه به دنبال مشخص کردن الگوهای از قبل شناخته نشده هستند. همچنین هزینه و عوارض جانبی روش‌های تشخیصی بیماری قلب سبب شد که محققان به دنبال روش‌های ارزان و با دقت بالا برای تشخیص این بیماری باشند (لینگ، وانگ، تان و لی[3]، ۲۰۱۷). سیستم‌های تصمیم یار هوشمند، سیستم‌های اطلاعاتی هستند که با بکارگیری مدل‌ها، داده، اطلاعات و دانش جمع آوری شده، جهت یاری رساندن به مدیران جهت حل مسائل غیر ساخته و شبه ساخت یافته توسعه می‌یابند. برای انجام هر پروژه، تعریف متدولوژی خاص لازم است. برای تعریف متدولوژی خاص نیز باید به سراغ الگوهای فرآیند رفت تا مجموعه مفیدی از قطعات متدولوژی مختلف حاصل آید. این سیستم‌ها دارای قابلیت‌های تحلیلی پیشرفته‌ای هستند که به کاربران اجازه داده شده است تا از مدل‌های تصمیم‌گیری مختلفی برای تحلیل اطلاعات استفاده کنند. سیستم‌های هوشمند) سیستم خبره و شبکه‌ی عصبی (دارای ساختار، اجزا و قابلیت‌هایی هستند که در مجموع قابلیت تصمیم‌گیری را ارتقا می‌دهند. به همین دلیل، از آنها در موارد بسیاری در پزشکی استفاده شده است (پوگازندی و همکاران[4]، ۲۰۱۷؛

1- Ali, Kareem, & Mohammed
2- Limam, Zouhair, & Oueslati
3- Ling, Wong, Tan, & Lee
4- Pughazendi etal

سومیا و سامیترا[1]، ۲۰۱۷).

یکی از مزایای این سیستم‌ها، در نظر گرفتن راه‌حل‌های متنوع‌تر است. هوش مصنوعی به پزشک کمک می‌کند تا متغیرهای بیشتر و متنوع‌تری را در زمان تشخیص بیماری یا انتخاب درمان در نظر بگیرد. به عبارتی، با توجه به محدودیت یادآوری ذهن، پزشک ممکن است تمام متغیرهای لازم برای تصمیم‌گیری برای نمونه علایم یا نتایج آزمایش‌ها را در آن واحد در نظر نگیرد یا آن‌ها را فراموش کند یا در پی کسب اطلاعات در خصوص آن نباشد. اما از آنجا که روابط بین این متغیرها در زمان طراحی سیستم در آن لحاظ می‌گردد، بنابراین احتمال نادیده گرفتن برخی از این عوامل یا در نظر گرفتن تأثیر آنها کمتر یا بیشتر از حد معقول، کاهش می‌یابد. بنابراین با توجه به کیفیت تعریف این روابط، می‌توان انتظار داشت تا تصمیمات پزشکان دقیق‌تر شود (خمامی، رضوانیان، میبدی و باقری[2]، ۲۰۲۱؛ چن و همکاران[3]، ۲۰۱۴). هدف مقاله حاضر این است که یک سیستم تصمیم یار هوشمند به منظور بهبود تشخیص بیماری‌های قلبی-عروقی مبتنی بر رایانش مه ارائه شود.

۲- مبانی نظری

کایا و همکارانش یک سیستم پشتیبانی تصمیم‌گیری برای پیش‌بینی خطر ابتلا به بیماری‌های قلبی-عروقی در بیماران هندی با استفاده از روش یادگیری ماشین پیشنهاد کردند. آنها از الگوریتم ژنتیکی برای تصمیم‌گیری در مورد الگوی اثر بالا و مقدار بهینه آنها استفاده کرده‌اند. همچنین از روش‌های نظری برای اجرای الگوریتم یادگیری ماشین استفاده شد (کایا، اوران و ارسلان[4]، ۲۰۱۱). فلورنس و همکارانش برای تشخیص بیماری مادرزادی قلب، از یک طبقه‌بندی کننده مبتنی بر قانون فازی استفاده کردند که بیماری ساختاری و عملکردی قلب را تعیین می‌کند. آنها از روش رای‌گیری وزنی و روش تک برنده استفاده کرده‌اند. نتیجه نشان داده است که روش رای‌گیری وزنی به طور کلی باعث افزایش دقت طبقه‌بندی بیماری مادرزادی قلب شده است (فلورنس، اما، آناپورانی، و ملاتهی[5]، ۲۰۱۴).

آنوجی یک روش نورونی- فازی لایه بردار برای پیش‌بینی در مورد وقوع بیماری قلبی عروق کرونر توسط نرم‌افزار متلب پیشنهاد کردند. اجرای رویکرد یکپارچه عصبی-فازی دارای خطای کم و عملکرد بالا در تجزیه و تحلیل نتیجه برای وقایع بیماری عروق کرونر ایجاد کرده است (آنوجی[6]، ۲۰۱۲). امیر ساراج و همکارش، یک سیستم متخصص فازی برای تشخیص بیماری قلبی پیشنهاد کردند. نویسندگان در این مقاله، چندین متغیر از جمله درد قفسه سینه، فشار خون، کلسترول، قند خون، حداکثر ضربان قلب، جنس الکتروکاردیوگرافی[7]، ورزش به عنوان ورودی وضعیت تشخیص بیمار استفاده شده است (امیرساراج و راجسواری[8]، ۲۰۱۸). مالایم و همکارش، در کار خود از الگوریتم های مختلف برای پیش بینی بیماری از چندین ویژگی هدف استفاده کردند. آنها روش‌های پیش‌بینی حمله قلبی هوشمند و مؤثر را با استفاده از داده‌کاوی ارائه داده‌اند. برای پیش‌بینی حمله قلبی به طور قابل توجهی ۱۵ ویژگی ذکر شده است. علاوه بر این ۱۵ مورد ذکر شده در ادبیات پزشکی و سایر تکنیک‌های داده‌کاوی، به عنوان مثال، سری زمانی، خوشه‌بندی و قوانین انجمن را در بر می‌گیرد (مالایم و آلیبوغلو[9]، ۲۰۱۶). مانوگرن و همکارانش از الگوریتم مبتنی بر یادگیری چند هسته‌ای با سیستم استنباطی عصبی-فازی برای تشخیص بیماری استفاده کردند. تجزیه و تحلیل مجموعه داده‌های مربوط به بیماری‌های قلبی حاکی از آن است که افراد در مجموعه سنی ۴۰-۶۰ سال باید از علائم بیماری قلبی آگاه باشند زیرا این مجموعه سن بیشتر خطرات بیماری قلبی را به همراه دارد. مردان نسبت به جنسیت زن خطرات بیشتری از بیماری قلبی دارند (مانوگرن، واراساراجان، و پریان[10]، ۲۰۱۸). آما، یک سیستم پیش‌بینی حمله قلبی هوشمند و مؤثر را با

1- Sowmiya, & Sumitra
2- Khomami, Rezvanian, Meybodi, & Bagheri
3- Chen etal
4- Kaya, Oran, & Arslan
5- Florence, Amma, Annapoorani, & Malathi
6- Anooj
7- ECG
8- Amirtharaj, & Rajeswari
9- Mülayim, & Alaybeyolu
10- Manogaran,Varatharajan, & Priyan

استفاده از داده‌کاوی و شبکه عصبی مصنوعی ارائه دادند. آنها از خوشه‌بندی K به معنی استخراج داده‌های مناسب برای حمله قلبی استفاده کردند. همچنین از الگوریتم MAFIA برای استخراج الگوهای مکرر استفاده کردند (آما[1]، ۲۰۱۲).

پینتو و همکارانش برای توصیف آنتی بیوتیک‌هایی با مصرف محدود، یک آی‌دی‌اس[2] مبتنی بر عامل تعریف کردند، این سیستم با داشتن انواع مختلفی از عامل‌ها کار می‌کند. این عامل‌ها، دستیاران آزمایشگاهی هستند که نتایج آزمایشگاه و متخصصان داروسازی را به عهده می‌گیرند و با استفاده از دانش خود در مورد داروهای پزشکی و وضعیت فعلی، راهکارهای درمانی جایگزین را ارائه داده اند (پینتو، کابرال و گومز[3]، ۲۰۱۷).

آروی و همکارانش یک سیستم عامل تلفن همراه برای جمع‌آوری داده‌های پزشکی از بیمارستان‌های مختلف و سایر منابع داده استفاده کردند. یک سیستم تصمیم‌یار برای تشخیص پژشکی بیمار نشان داده شده است که بیمار داده‌های پزشکی خود را با پر کردن پرسش‌نامه بر روی گوشی خود انجام می‌دهد. سپس سیستم تصمیم‌یار بر اساس داده‌های بدست آمده و آنالیز آن بیماری وی را تشخیص می‌دهد. همچنین این سیستم گزینه‌های مختلفی را در اختیار بیمار قرار می‌دهد که بیمار با توجه به نوع بیماری یا علائم آیکون‌های موجود را انتخاب می‌کند و سپس به پرسش‌های آن بخش پاسخ می‌دهد (آروی، پرینی، و واس[4]، ۲۰۱۸).

رانی و همکارانش یک سیستم تصمیم‌یار هوشمند برای پیش‌بینی بیماری قلبی بر اساس یادگیری ماشین ارائه دادند. در این مقاله، نویسندگان یک سیستم پشتیبانی تصمیم‌یار هوشمند ترکیبی را پیشنهاد کرده‌اند که می‌تواند به تشخیص زودهنگام بیماری قلبی-عروقی بر اساس پارامترهای بالینی بیمار کمک کند. نویسندگان از روش‌های چند متغیره مبتنی بر الگوریتم معادلات زنجیره‌ای برای رسیدگی به مقادیر گمشده استفاده کرده‌اند. یک الگوریتم انتخاب ویژگی ترکیبی از الگوریتم ژنتیک و حذف ویژگی بازگشتی برای انتخاب ویژگی‌های مناسب از مجموعه داده‌های موجود استفاده شده است. علاوه بر این، برای پیش پردازش داده‌ها از تکنیک اقلیت مصنوعی و روش‌های اسکالر استاندارد استفاده شده است. در آخرین مرحله از توسعه سیستم ترکیبی پیشنهادی، نویسندگان از ماشین بردار پشتیبان، رگرسیون لجستیک، جنگل تصادفی و طبقه‌بندی‌کننده آدابوست استفاده کرده‌اند. مشخص شده است که این سیستم دقیق‌ترین نتایج را با طبقه‌بندی جنگل تصادفی ارائه کرده است. نتایج کار آنها دارای دقت ۸۶٫۶٪ می‌باشد که نسبت به برخی از سیستم‌های پیش‌بینی بیماری قلبی موجود در ادبیات پیشین بهتر است (رانی، کومار، احمد، و جاین[5]، ۲۰۲۱).

روش گراف کاوی

یکی از روش‌های گراف کاوی، روش حریصانه Kempe است. رویکرد حریصانه‌ی استفاده شده توسط Kempe با یافتن بهترین رأس برای فعالسازی با استفاده از روش جستجوی فراگیر شروع می‌شود. انجمن در گراف به‌عنوان مجموعه‌ای متراکم از گره‌های متصل بهم در نظر گرفته می‌شود که به‌صورت خلوت به سایر بخش‌های گراف متصل شده‌اند. به عبارت دیگر، یک انجمن در یک گراف متشکل از گروه‌های از گره‌ها می‌باشد که روابط بیشتری در داخل انجمن با هم دارند و روابط بسیار کم و خلوتی با سایر گروه‌ها در گراف دارند. هر گره را به عنوان یک انجمن مقداردهی اولیه می‌کند و دو انجمن را با درنظر گرفتن بیشترین افزایش ماژولاریتی یا کمترین کاهش ماژولاریتی ادغام می‌کند. این فرایند تا زمانیکه تمام گره‌های شبکه متعلق به یک انجمن باشند، تکرار می‌شوند. این الگوریتم از معنای فیزیکی ماژولاریتی استفاده می‌کند، هرچه ماژولاریتی بیشتر باشد ساختار انجمن مربوطه بهتر است. در زیر شبه کد الگوریتم Kempe نشان داده شده است. یک رأس فعال می‌شود و سپس مدل انتشار به دفعات دلخواه به کار گرفته می‌شود. پس از آزمایش همه‌ی رئوس، رأسی که تعداد بیشتری از رئوس را فعال کرده باشد انتخاب می‌شود.

1- Amma
2- IDSS
3- Pinto, Cabral, & Gomes
4- Árvai, Perényi, & Vass
5- Rani, Kumar, Ahmed, & Jain

Algorithm 1 GeneralGreedy(G, k)

intialize $S = \emptyset$ and $N = 20000$
for $i = 1$ to k **do**
  for each vertex $v \in V \setminus S$ **do**
    $s_v = 0$.
    for $i = 1$ to N **do**
      $s_v \mathrel{+}= |RanCas(S \cup \{v\})|$
    end for
    $s_v = s_v / N$
  end for
  $S = S \cup \{\arg\max_{v \in V \setminus S}\{s_v\}\}$
end for
output S.

شکل(۲): شبه کد الگوریتم Kempe (کایا و همکارانش،۲۰۱۱)

برمبنای شکل شماره (۲)، هر یک ازرئوس باقیمانده به این رأس اولیه اضافه می‌شود و دوباره شبیه‌سازی‌ها برای انتخاب بهترین رأس برای اضافه کردن به مجموعه رئوس اولیه اجرا می‌شود. این فرآیند آنقدر تکرار می‌شود تا k رأس انتخاب شوند. در این روش می‌توان تعداد رئوس اولیه‌ی دلخواه فعال را مشخص کرد. این روش از روش‌های دیگر ساده‌تر است زیرا تنها به گراف شبکه و تعداد دلخواه رئوس تأثیرگذار به عنوان ورودی نیاز دارد. اما در این روش اجتماعات در نظر گرفته نشده‌اند و همچنین، این روش اگرچه کارایی مناسبی دارد، اما سرعت بسیار کمی به ویژه در سیستم‌های بزرگ دارد.

روش ممتیک

الگوریتم ممتیک، یک الگوریتم مبتنی بر جمعیت است که برای مسایل بهینه‌سازی پیچیده و بزرگ مورد استفاده قرار می‌گیرد. ایده اصلی این الگوریتم، به کارگیری یک روش جستجوی محلی در درون ساختار الگوریتم ژنتیک برای بهبود کارآیی فرایند تشدید هنگام جستجو است. در شکل شماره (۳)، شبه کد الگوریتم ممتیک نشان داده شده است.

```
Memetic Algorithm Template
Begin
    INITIALIZE population;
    EVALUATE each candidate;
    Repeat Until (TERMINIATION CONDITION) Do
        SELECT parents;
        RECOMBINE to produce offspring;
        MUTATE offspring;
        IMPROVE offspring via Local Search;
        EVALUATE offspring;
        SELECT individuals for next generation;
    endDo
End
```

شکل (۳): شبه کد الگوریتم ممتیک (کایا و همکارانش،۲۰۱۱)

الگوریتم پیشنهادی شامل دو استراتژی است که شامل زمان‌بندی داده‌ها بر پایه الگوریتم ممتیک و طرح بیداری است. استراتژی اول داده‌های اضافی را بر پایه زمان‌بندی با الگوریتم ممتیک غیر فعال می‌کند. سپس از طرح بیداری برای مدیریت بهینه‌سازی برای استخراج و آموزش داده‌ها در هر مرحله استفاده می‌شود. الگوریتم ممتیکی از دو بخش عملگرهای الگوریتم ژنتیک و جستجوی محلی ساخته شده است. جمعیت اولیه معمولا بصورت تصادفی ساخته می‌شود و پروسه تکاملی بوسیله عمل‌گرهای الگوریتم ژنتیک مانند انتخاب، تقاطع و جهش انجام می‌گیرد. هر مشخصه شامل راه‌حلی با مقدار برازش متناظر است که از تابع برازش به دست می‌آید. از تابع برازش برای

مشخص کردن بهینگی راه حل استفاده می‌شود. بعد از تکمیل عملیات از جستجوی محلی برای افزایش مقدار برازش استفاده می‌شود. با تکرار این عملیات جمعیت جدید از افراد برتر تولید می‌شود و نتایج حاصله به نقطه بهینه نزدیک خواهند شد.

الگوریتم ممتیک جهت تعیین مقدار بهینگی و نزدیک شدن به جواب نهایی در کروموزوم موجود در یک مسئله از تابع برازندگی استفاده می‌کند. ساختار برای تعیین بردار پوشش در کروموزوم k ام از رابطه زیر استفاده می‌شود:

فرمول (۱)

$$\bar{w}(k) = (lk, 1.\pi 1) \vee (lk, 2.\pi 2) \ldots \vee (lk, N.\pi N)$$

می‌توان درصد پوشش یک کروموزوم را به صورت زیر محاسبه نمود:

فرمول (۲)

$$\varepsilon^k = \frac{\|\bar{w}(k)\|^2}{M}$$

که در آن صورت کسر نشان دهنده تعداد داده‌های فعال موجود در پایگاه داده است. نسبت ابزار داده برای داده k ام به صورت زیر محاسبه می‌شود:

فرمول (۳)

$$N_1^k = \frac{\sum_{g=1}^{N} lk, g}{N}$$

الگوریتم ممتیک در ابتدا مجموعه‌ای از جواب‌های اولیه را رمزگذاری می‌کند. آنگاه این الگوریتم میزان مطلوبیت هر یک از جواب‌ها را بر اساس یک تابع برازندگی محاسبه کرده و با استفاده از عملگرهایی چون تقاطع و جهش، جواب‌های جدیدی را تولید می‌کند. در پایان، هر نسل روی مجموعه‌های از جواب‌های آن نسل، یک جستجوی محلی با هدف تشدید فرایند جستجو انجام می‌شود تا کیفیت جواب‌های بهینه محلی افزایش یابد. در ادامه، شکل شماره (۴) شبه کد جستجوی محلی در الگوریتم ممتیک نشان داده شده است.

```
Step 1: Input a population Pop_k = {u_1, u_2, ..., u_ρ}_k to the local search unit.
Step 2: Let u_i be the i-th chromosome in Pop_k, and N be the length of u_i. According to the definition
        of allele ℓ_{i,j}, we know that u_i is a binary string composed of ℓ_{i,j}, where ∀i, 1≤ j ≤ N.
          For i = 1 to ρ do
           For   j = 1 to N do
              If the value of ℓ_{i,j} is equal to one then
                   Record the allele ℓ_{i,j} in the array S_i.
              End
            End
          End
Step 3: For y = 1 to ρ do
            Let h be the length of S_y.
              For l = 1 to h do
                 fitness1 = fit(u_y); // fit() is the fitness function.
                 In u_y, let the allele c_tmp, corresponding to the l_th element in S_y, be zero.
                 fitness2 = fit(u_y); // Re-evaluate the fitness
                 If fitness1 > finess2 then
                    Let c_tmp equal to one.
                 End
              End
          End
Step 4: Output the improved Pop_k.
```

شکل (۴): شبه کد جستجوی محلی در الگوریتم ممتیک (فلورنس و همکاران، ۲۰۱۴)

جستجوی محلی به کارگرفته شده در الگوریتم ممتیک، می‌تواند ضعف روش‌های تکاملی همچون الگوریتم ژنتیک را در تشدید فرایند جستجو رفع کند. شرط توقف الگوریتم پیشنهادی این است که مقدار کروموزوم پس تعداد تکرای که بوسیله حد آستانه مشخص می‌شود تغییر نکند.

ماشین بردار پشتیبان

گفتیم که بعد از استخراج داده‌ها نیاز است که داده‌ها طبقه‌بندی شوند، برای این کار از یکی از روش‌های مطرح داده‌کاوی یعنی ماشین بردار پشتیبان استفاده شده است. قبل از تقسیم خطی، برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها به وسیله تابع ϕ به فضای با ابعاد خیلی بالاتر برده می‌شود.

برای اینکه بتوان مساله ابعاد خیلی بالا با استفاده از این روش‌ها حل کرد از قضیه دوگانی لاگرانژ برای تبدیل مساله مینیمم‌سازی مورد نظر به فرم دوگانی آن که در آن به جای تابع پیچیده ϕ که ما را به فضایی با ابعاد بالا می‌برد، تابع ساده‌تری به نام تابع هسته که ضرب برداری تابع ϕ است، استفاده می‌شود. از توابع هسته مختلفی از جمله هسته‌های نمایی، چند جمله‌ای و سیگموید می‌توان استفاده نمود. فرض شود مجوعه داده آموزشی D شامل n عضو در اختیار است که به صورت رابطه زیر تعریف می‌شود.

فرمول (۴)

$$D=\{(x_i,y_i)|\, x_i \in R^P, y_i \in x_i, y_i)|\, x_i \in R^P, y_i \in \{-1\ ,1\}\}_{i=1}^{n}\}_{i=1}^{n}$$

جایی که مقدار y برابر یک یا -۱ و هر Xi یک بردار حقیقی P بعدی است. هدف پیدا کردن ابر صفحه جداکننده با بیشترین فاصله از نقاط حاشیه‌ای است که نقاط $yi=1$ را از نقاط $yi=-1$ جدا کند. هر ابر صفحه می‌تواند به صورت مجموعه‌ای از نقاط X که شرط در رابطه زیر را ارضا کند، تعریف شود:

فرمول (۵)

$$w.x-b=0$$

جایی که W بردار نرمال است، که به ابر صفحه عمود است.

قرار است W,b را طوری انتخاب شود که بیشترین فاصله بین ابر صفحه‌های موازی که داده‌ها را از هم جدا می‌کنند، ایجاد شود. این ابر صفحه با استفاده از رابطه زیر تعریف می‌شود:

فرمول (۶)

$$w.x-b=1$$
$$w.x-b=-1$$

اگر داده‌های آموزشی جدایی پذیر خطی باشند، می‌توان دو ابر صفحه در حاشیه نقاط به طوری که هیچ نقطه مشترکی نداشته باشند، در نظر گرفت و سپس سعی شود، فاصله آنها را، ماکسیمم کرد. برای اینکه از ورود نقاط به حاشیه جلوگیری شود، شرایطی که در ادامه است، اضافه می‌شود:

فرمول (۷)

$$w.x_i-b\geq -1 \quad for\ x_i \qquad w.x_i-b\geq 1 \quad for\ x_i$$

این می‌تواند به صورت رابطه زیر نوشت:

فرمول (۸)

$$y_i\,(w.x_i-b)\geq 1 \qquad for\ all \quad 1\leq i\leq n$$

با کنار هم قرار دادن این دو یک مسئله بهینه‌سازی به دست می‌آید، که در رابطه زیر نشان داده شده است:

فرمول (۹)

$$Minimize\ (in^{w,b}):||W|| \qquad Subject\ to\ (for\ any\ i=1,\dots\dots n)$$
$$y_i\,(w.x_i-b)\geq 1$$

می‌توان رابطه بالا را با استفاده از ضرایب منفی لاگرانژ به صورت رابطه زیر بدست آورد.

فرمول (۱۰)

$$\underset{W,b,a}{Min}\{\underset{W,b,a}{Min}\{1/2\ ||\ W\}||^2-\sum_{i=1}^{n}a_i[y_i(W.X_i-b)-1]|^2-\sum_{i=1}^{n}a_i[y_i(W.X_i-b)-1]|$$

۳- روش تحقیق

در ابتدا داده‌ها از طریق پرسش‌نامه جمع‌آوری می‌شود. این داده‌ها شامل اطلاعات و پرونده پزشکی ۱۰۰ نفر از بیماران می‌باشد. سپس داده‌های ورودی، ابتدا باید پاکسازی و نرمالسازی شوند و سپس بر اساس الگوریتم‌ها استخراج ویژگی، ویژگی‌ها انتخاب می‌شوند و به آن‌ها وزن داده می‌شود. در مرحله پیش پردازش داده‌ها سه گام اساسی شامل شناخت داده‌ها، پاک‌سازی داده‌ها و تبدیل داده‌ها انجام می‌شود. برای شناخت داده‌ها و پاک‌سازی داده‌ها از ماتریس همبستگی و برای تبدیل داده‌ها از الگوریتم Min-Max استفاده شده است. بعد از پیش‌پردازش داده‌ها و آماده‌سازی آنها نیاز است که داده‌ها استخراج شده و آموزش ببینند تا یک مدل درستی برای تشخیص و تعیین ارتباط بین بیماری‌ها ایجاد شود. برای این منظور از دو روش گراف کاوی و روش ممتیک به صورت جدا استفاده شده است. یعنی یکبار داده‌ها به وسیله روش‌های گراف کاوی استخراج و آموزش می‌بینند، یکبار هم با استفاده از الگوریتم ممتیک. هر دو الگوریتم از روش‌های مطرح در این زمینه هستند و برای اینکه بتوانیم در این پژوهش بهترین مدل را ارائه کنیم لازم است از دو روش برای مقایسه باهم استفاده کنیم. و در قسمت ارزیابی می‌توان این دو روش را از لحاظ دقت و پیچیدگی‌های زمانی مقایسه کرد. در ادامه شکل شماره (۵)، روش تحقیق نشان داده شده است.

شکل (۵): نمودار کلی روش تحقیق

همانطور که گفته شد برای استخراج و آموزش داده‌ها از دو روش گراف کاوی و ممتیک به صورت جداگانه استفاده خواهیم کرد. درگراف‌کاوی مبحث تشخیص انجمن نقش اساسی و مهم در بسیاری از حوزه‌ها مانند جامعه‌شناسی، زیست‌شناسی و علوم کامپیوتر ایفا می‌کند. مبحث تشخیص انجمن در شبکه و گراف‌ها از لحاظ تئوری و رفتاری برای تجزیه و تحلیل توپولوژی شبکه، همچنین برای تجزیه و تحلیل عملکرد و پیش‌بینی رفتار در شبکه‌های پیچیده اهمیت بسیاری دارد. از این رو تشخیص انجمن در گراف و همچنین بررسی ساختار انجمن در گراف بیش از پیش مورد توجه قرار گرفته‌است. روش دوم، استفاه از الگوریتم ممتیک است. این روش یک الگوریتم مبتنی بر جمعیت است که برای مسایل بهینه‌سازی پیچیده و بزرگ مورد استفاده قرار می‌گیرد.

ایده اصلی این الگوریتم، به کارگیری یک روش جستجوی محلی در درون ساختار الگوریتم ژنتیک برای

بهبود کارآیی فرایند تشدید هنگام جستجو است. بعد از استخراج و آموزش داده‌ها نوبت به طبقه‌بندی داده‌ها می‌رسد. برای این منظور از الگوریتم ماشین بردار پشتیبان استفاده شده است. علت انتخاب ماشین بردار پشتیبان دقت بالای آن در طبقه‌بندی داده‌ها و دور بودن از پیچیدگی محاسباتی است. ماشین بردار پشتیبان، یک طبقه‌بندی کننده است. با توجه به داده‌های آموزشی دارای برچسب (یادگیری نظارت شده)، الگوریتم یک صفحه نمایش بهینه را ارائه می‌دهد که نمونه‌های جدیدی را طبقه‌بندی می‌کند. هدف از الگوریتم ماشین بردار پشتیبان، یافتن یک داده‌ها و دسته‌بندی آن‌ها در فضای Nبعدی از ویژگی‌ها است. مبنای کاری دسته‌بندی کننده ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است. و در تقسیم خطی داده‌ها سعی می‌شود خطی انتخاب گردد که حاشیه اطمینان بیشتری داشته باشد. بعد از طبقه‌بندی داده‌های نیاز از مدل‌های بدست آمده مورد ارزیابی و سنجش قرار بگیرند، برای این منظور از ماتریس درهم ریختگی استفاده می‌شود. این ماتریس ابزار مفیدی برای تحلیل چگونگی عملکرد روش در طراحی مدل در تشخیص داده‌ها یا مشاهدات دسته‌های مختلف است. حالت ایده آل این است که بیشتر داده‌های مرتبط با مشاهدات روی قطر اصلی ماتریس قرار گرفته باشند و مابقی مقادیر ماتریس صفر یا نزدیک صفر باشند.

شرایط آزمایش

در ابتدا داده‌ها از طریق پرسش‌نامه جمع‌آوری می‌شود. این داده‌ها شامل اطلاعات و پرونده پزشکی ۱۰۰ نفر از بیماران می‌باشد. در ادامه شکل شماره (۶) برخی از رکوردهای پایگاه‌داده افراد نشان داده شده است.

	A	B	C	D	E	F	G	H
1	Age	Marital_St	Gender	Weight_C	Cholester	Stress_Ma	Trait_Anx	2nd_Heart_Attack
2	60	2	0	1	150	1	50	Yes
3	69	2	1	1	170	0	60	Yes
4	52	1	0	0	174	1	35	No
5	66	2	1	1	169	0	60	Yes
6	70	3	0	1	237	0	65	Yes
7	52	1	0	0	174	1	35	No
8	58	2	1	0	140	0	45	No
9	59	2	1	0	143	0	45	Yes
10	60	2	0	0	139	0	45	No
11	51	1	1	0	174	1	40	No
12	52	1	0	0	189	1	65	No
13	70	2	1	1	147	1	50	Yes
14	52	2	1	2	160	0	40	Yes
15	74	3	1	2	178	0	75	Yes
16	64	2	1	2	236	1	80	Yes
17	69	2	0	1	146	1	50	Yes
18	58	2	0	0	141	0	45	No
19	68	1	0	0	172	0	60	No
20	66	1	0	0	172	0	60	No
21	63	0	1	1	138	1	50	No
22	50	1	1	0	174	1	40	No
23	60	2	0	1	146	1	50	Yes
24	70	2	1	1	238	0	60	Yes
25	54	1	0	0	172	1	35	No
26	75	1	0	2	178	0	75	Yes
27	72	3	0	1	236	0	65	Yes
28	59	1	1	2	202	0	70	Yes
29	60	2	1	0	140	0	45	No

شکل (۶): تعدادی از رکوردهای پایگاه داده تهیه شده

۴- یافته‌های تحقیق

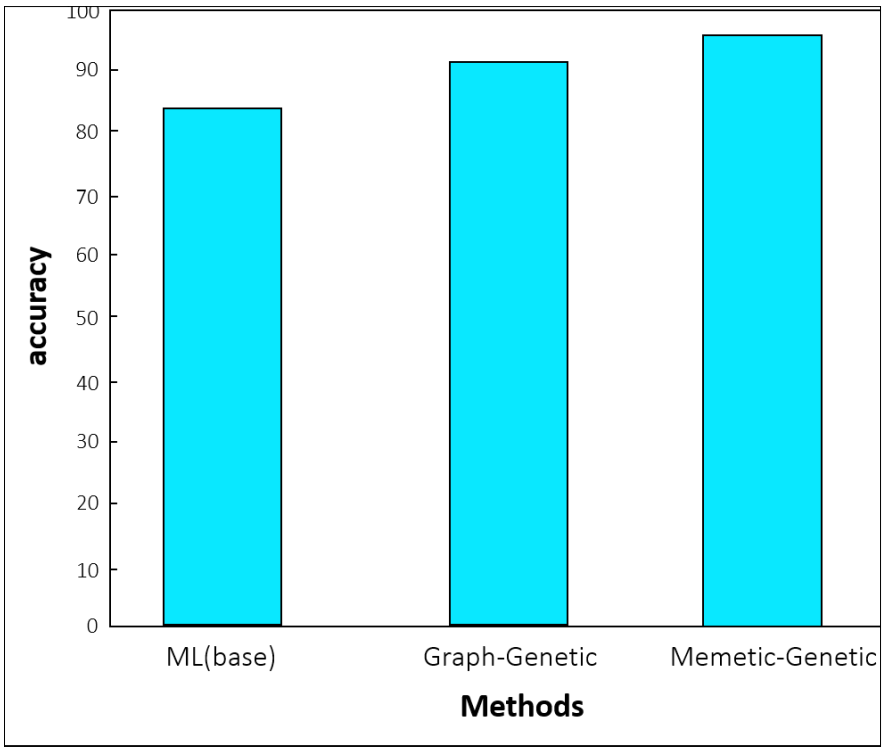
بدست آوردن دقت تشخیص

دقت طبقه‌بندی کننده به معنی درصد تشخیص صحیح تصاویر اثرانگشت واقعی از جعلی می‌باشد که حاصل تقسیم تعداد تشخیص درست بر کل نمونه‌ها می‌باشد که با ضرب این عدد در ۱۰۰، درصد دقت بدست می‌آید. در ادامه در رابطه۱۱ دقت بدست آمده است.

فرمول (۱۱)

$$accuracy = \frac{TP + TN}{(TP + FN + FP + TN)}$$

در رابطه بالا، TP[1] مقادیر مثبت حقیقی، TN[2] مقادیر منفی حقیقی، FN[3] مقادیر منفی کاذب و FP[4] مقادیر مثبت کاذب را نشان می‌دهند.



نمودار (۱): مقایسه نتایج حاصل از معیار دقت الگوریتم‌های پیشنهادی و مقاله پایه

همان‌طور که در نمودار شماره (۱) نشان داده شده است، معیار دقت در روش‌های پیشنهادی بالاتر از مقاله پایه است. در روش پیشنهادی ممتیک-ژنتیک دارای دقت ۹۳ درصد و در روش پیشنهادی گراف کاوی-ژنتیک دارای دقت ۹۱ درصد و در مقاله پایه دارای دقت ۸۶ درصد می‌باشد.

بدست آوردن منحنیROC

منحنی ROC یک منحنی است که در آن محور Y را TPR و محور X را FPR تشکیل داده است. این منحنی یکی از مهمترین معیارهای ارزیابی عملکرد مدل‌های طبقه‌بندی شده یا چند لایه می‌باشد. این معیار مناسب، می‌تواند به اندازه‌گیری مدل‌ها در آستانه‌های مختلف بپردازد.

1- True positive
2- True Negative
3- false Negative
4- false positive

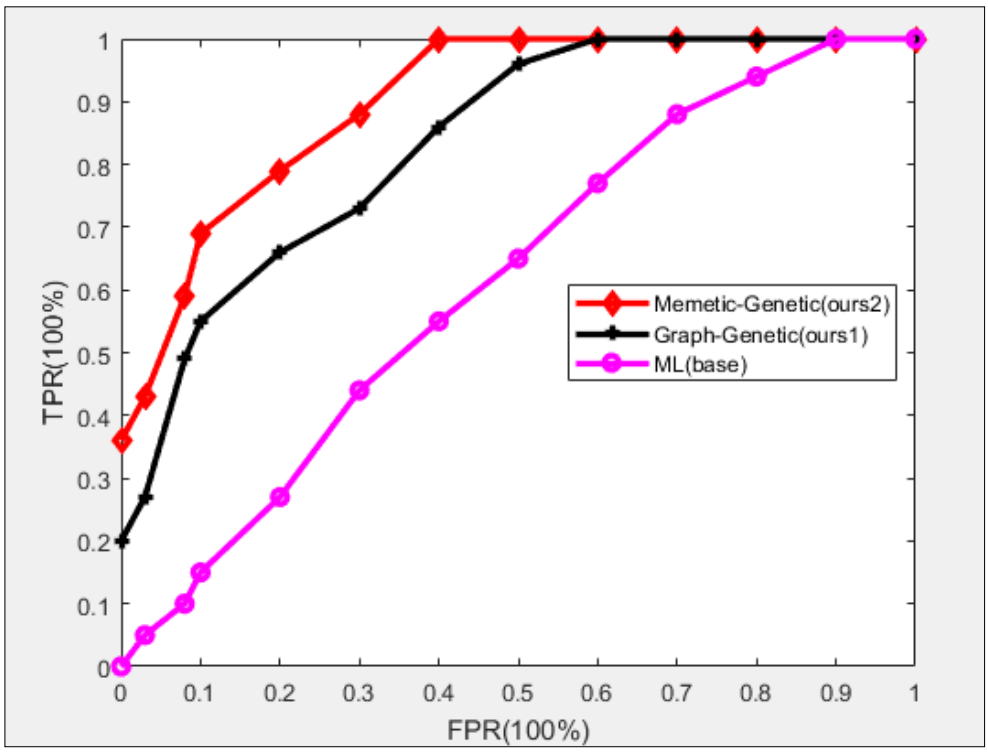
معیار نرخ مثبت کاذب[1]، نشان دهنده نسبت شناسایی نمونه‌های مثبت کاذب(تخصیص منابع) به تمام موارد منفی است. که از رابطه(۱۲) زیر بدست می آید.

فرمول (۱۲)
$$FPR = \frac{FP}{(FP + TN)}$$

FP نشان می‌دهد که کلاس‌های مثبت و منفی، کلاس‌های مثبت پیش‌بینی می‌شوند. در حالیکه TN نشان می‌دهد که کلاس‌های منفی، منفی پیش‌بینی می‌شوند.

معیار نرخ مثبت حقیقی[2] نسبت شناسایی نمونه‌های واقعی به همه موارد مثبت را نشان می‌دهد. که از رابطه (۱۳) زیر بدست می‌آید:

فرمول (۱۳)
$$TPR = \frac{TP}{(TP + FN)}$$



نمودار (۲): مقایسه نتایج حاصل از منحنی ROC الگوریتم‌های پیشنهادی و مقاله پایه

در نمودار شماره (۲)، ROC را مشاهده می دهد. همانطور که از نمودار پیدا است، روش‌های پیشنهادی با دلایل زیر توانسته‌اند سطح بیشتری را نسبت به مقاله پایه فراهم کنند:

۱- ارزیابی در نقطه صفر : در نقطه صفر، مقدار TPR برای روش ممتیک- ژنتیک برابر با ۰٫۳۸ و برای روش گراف کاوی- ژنتیک برابر با ۰٫۲ است. در مقابل، در مقاله پایه، TPR در نقطه صفر برابر با صفر است. این نتایج نشان می‌دهد که روش‌های پیشنهادی در تشخیص صحیح بیماری‌ها در مرحله اولیه بهبود یافته‌اند.

۲- سرعت دستیابی به TPR : روش ممتیک- ژنتیک به سرعت توانسته به مقدار ۱ در TPR برسد. نقطه‌ای که به

1- false positive rate (FPR)
2- true positive rate (TPR)

مقدار ۱ رسیده است، در این روش ۰٫۴ است. این نتیجه نشان می‌دهد که روش ممتیک-ژنتیک بهبود قابل توجهی در تشخیص صحیح بیماری‌ها در مراحل پیشرفته نشان داده است.

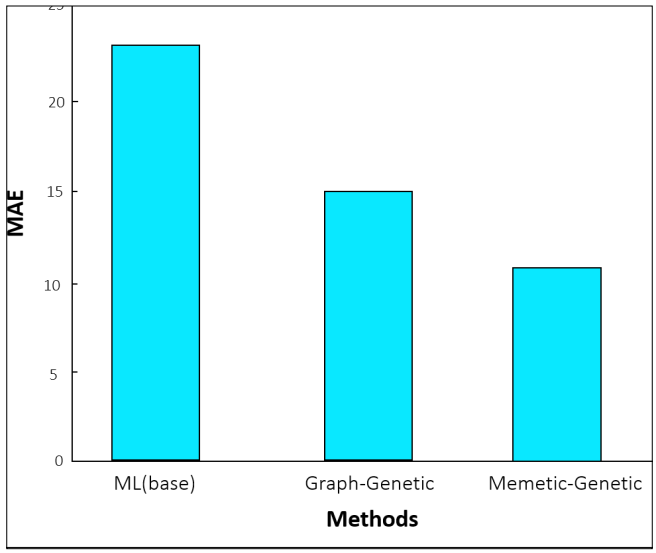
از این رو، با توجه به دلایل فوق، روش‌های پیشنهادی در مقایسه با روش پایه، نتایج بهتری در تشخیص و تعیین ارتباط بین بیماری‌ها به ارمغان آورده‌اند.

معیار MAE[1]

خطای میانگین روشی برای برآورد میزان خطاست که در واقع تفاوت بین مقادیر تخمینی و آنچه تخمین زده شده، است. برای بدست مجموعه یا n داده ه

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)^2$$

$(Y_i - \tilde{Y}_i)^2 (Y_i - \tilde{Y}_i)^2$ مقدار مربع خطای هر داده را محاسبه می‌کند.



نمودار (۳): مقایسه خطای میانگین مربعات الگوریتم‌های پیشنهادی و مقاله پایه

همانطور که از نمودار ۳ مشخص است، خطای میانگین مربعات روش پیشنهادی بهبود قابل توجهی نسبت به مقاله پایه نشان داده است. با استفاده از روش ممتیک-ژنتیک و گراف کاوی-ژنتیک، خطای میانگین مربعات به ترتیب ۱۱ و ۱۵ درصد کاهش یافته است، که این نشان می‌دهد دقت و صحت روش‌های پیشنهادی در تشخیص بیماری های قلبی-عروقی بهبود یافته است.

۵- نتیجه‌گیری

در ابتدا داده‌ها از طریق پرسش‌نامه جمع‌آوری می‌شود. این داده‌ها شامل اطلاعات و پرونده پزشکی ۱۰۰ نفر از بیماران می‌باشد. سپس این داده‌ها با استفاده از دو شبیه‌ساز متلب و ویژال استودیو شبیه‌سازی شدند. روش پیشنهادی با استفاده از معیارهای دقت، منحنی ROC و خطای میانگین مربعات با مقاله پایه مورد مقایسه قرار گرفته است. یافته‌های پژوهش نشان می‌دهد که معیار دقت در روش‌های پیشنهادی بالاتر از مقاله پایه است. در روش پیشنهادی ممتیک- ژنتیک دارای دقت ۹۳ و در روش پیشنهادی گراف کاوی-ژنتیک دارای دقت ۹۱ و در مقاله پایه دارای دقت ۸۶ می‌باشد. روش‌های پیشنهادی توانسته

1- Mean Absolute Error

سطح بیشتری را نسبت به مقاله پایه در منحنی ROC ارائه شده پوشش دهد. در روش ممتیک-ژنتیک در نقطه صفر مقدار TPR=۰٫۳۸ بوده و در روش گراف کاوی-ژنتیک در نقطه صفر مقدار TPR=۰٫۲ است. درحالیکه در مقاله پایه TPR مقدار صفر گرفته است. همچنین در روش ممتیک-ژنتیک سریع‌تر توانسته به مقدار یک در TPR برسد. نقطه‌ای که به مقدار ۱ رسیده است، ۰٫۴ می‌باشد. خطای میانگین مربعات روش‌های پیشنهادی عملکرد بهتری نسبت به مقاله پایه داشته و برای روش ترکیبی ممتیک-ژنتیک و گراف کاوی-ژنتیک مقادیر ۱۱ و ۱۵ درصد را به خود اختصاص داده است. روش‌های ممتیک و ژنتیک هر دو از روش‌های فراابتکاری بودند، لذا توانسته‌اند در ترکیب باهم عملکرد بهتری داشته باشند. به عنوان پیشنهادات آینده می‌توان ترکیبی از سایر روش‌های فراابتکاری چون الگوریتم گرگ خاکستری نیز بهره گرفت.

منابع:

1-Ali, O. M. A., Kareem, S. W., & Mohammed, A. S. (2022, February). Evaluation of electrocardiogram signals classification using CNN, SVM, and LSTM algorithm: A review. In *2022 8th International Engineering Conference on Sustainable Technology and Development (IEC)* (pp. 185-191). IEEE.

2-Amirtharaj, P., & Rajeswari, K. (2018). Prediction of risk score for heart disease in india using machine intelligence. *University Journal of Surgery and Surgical Specialities*, *4*(4).

3-Amma, N. B. (2012, February). Cardiovascular disease prediction system using genetic algorithm and neural network. In *2012 international conference on computing, communication and applications* (pp. 1-5). IEEE.

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

5-Árvai, L., Perényi, D., & Vass, D. (2018, May). Organisational life assistant: How an IT system can help for the ageing society. In *2018 19th International Carpathian Control Conference (ICCC)* (pp. 395-399). IEEE.

6-Chen, C. P., Mukhopadhyay, S. C., Chuang, C. L., Lin, T. S., Liao, M. S., Wang, Y. C., & Jiang, J. A. (2014). A hybrid memetic framework for coverage optimization in wireless sensor networks. *IEEE transactions on cybernetics*, *45*(10), 2309-2322.

7-Florence, S., Amma, N. B., Annapoorani, G., & Malathi, K. (2014). Predicting the risk of heart attacks using neural network and decision tree. *International Journal of Innovative Research in Computer and Communication Engineering*, *2*(11), 7025-7030.

8-Kaya, E., Oran, B., & Arslan, A. (2011). A diagnostic fuzzy rule-based system for congenital heart

disease. *International Journal of Biomedical and Biological Engineering*, *5*(6), 232-235.
9-Khomami, M. M. D., Rezvanian, A., Meybodi, M. R., & Bagheri, A. (2021). CFIN: A community-based algorithm for finding influential nodes in complex social networks. *The Journal of Supercomputing*, *77*, 2207-2236.
10-Limam, H., Zouhair, A., & Oueslati, W. (2022). A New Hybrid Multiclass Approach Based on KNN and SVM. *Journal of Information & Knowledge Management*, *21*(04), 2250061.
11-Ling, T. H. Y., Wong, L. J., Tan, J. E. H., & Lee, C. K. (2015, February). XBee wireless blood pressure monitoring system with microsoft visual studio computer interfacing. In *2015 6th International Conference on Intelligent Systems, Modelling and Simulation* (pp. 5-9). IEEE.
12-Manogaran, G., Varatharajan, R., & Priyan, M. K. (2018). Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. *Multimedia tools and applications*, *77*, 4379-4399.
13-Mülayim, N., & Alaybeyolu, A. (2016, November). Designing of an expert system based on firefly algorithm for diagnosis of Heart Disease. In *2016 20th National Biomedical Engineering Meeting (BIYOMUT)* (pp. 1-4). IEEE.
14-Pinto, S., Cabral, J., & Gomes, T. (2017, March). We-care: An IoT-based health care system for elderly people. In *2017 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1378-1383). IEEE.
15-Pughazendi, N., Sathishkumar, R., Balaji, S., Sathyavenkateshwaren, S., Chander, S. S., & Surendar, V. (2017, August). Heart attack and alcohol detection sensor monitoring in smart transportation system using Internet of Things. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 881-888). IEEE.
16-Rani, P., Kumar, R., Ahmed, N. M. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, *7*(3), 263-275.
17-Sowmiya, C., & Sumitra, P. (2017, March). Analytical study of heart disease diagnosis using classification techniques. In *2017 IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS)* (pp. 1-5). IEEE.