روشی نوین برای تشخیص بیماری مبتنی بر زیرساخت اینترنت اشیاء با استفاده از الگوریتم ژنتیک و طبقه‌بندی‌کننده‌های مختلف

* سیدابراهیم دشتی ** مریم نیک‌پور *** مهدی نیک‌پور **** محبوبه جوهری
* استادیار، دانشکده مهندسی برق و کامپیوتر، واحد جهرم، دانشگاه آزاد اسلامی، جهرم، فارس، ایران sayed.dashty@gmail.com
** دانشجوی دکتری، دانشکده مهندسی برق و کامپیوتر، دانشگاه آزاد اسلامی، فارس، ایران nikpoor.maryam@yahoo.com
*** مربی، دانشگاه علوم پزشکی بندرعباس، هرمزگان، ایران nikpoor.maryam@gmail.com
**** کارشناسی ارشد، دانشگاه آزاد اسلامی فسا، فارس، ایران mahbobeh.johari@gmail.com

تاریخ دریافت: ۱۴۰۱/۰۶/۰۶ تاریخ پذیرش: ۱۴۰۲/۰۸/۲۷
صص: ۱۸۲- ۱۶۹

چکیده

فناوری اطلاعات پزشکی و خدمات بهداشتی و درمانی با رفاه ملی و معیشت مردم ارتباط دارد. ادغام پردازش ابری و اینترنت اشیا یک پیشرفت بزرگ در کاربرد پزشکی مدرن خواهد بود. در این تحقیق تمرکز بر روی بیماری مزمن دیابت می‌باشد که یکی از عوامل اصلی مرگ و میر در سراسر جهان محسوب می‌شود. این تحقیق تکنولوژی اطلاعات پزشکی را در زمینه اینترنت اشیا، به ویژه در زمینه کاربرد نظارت و مدیریت پزشکی بکار گرفته است. این مطالعه یک روش مبتکرانه مبتنی بر اینترنت اشیا را برای تشخیص دیابت معرفی می‌کند. یک معماری برای نظارت از راه دور و مدیریت پلت فرم ابر اطلاعات بهداشتی پیشنهاد و تحلیل می‌شود ، اطلاعات بیماران از طریق ابزارهای اینترنت اشیاء پوشیدنی و تعبیه شده بر حسب نیاز استفاده و جمع آوری می‌شود و در نهایت از طریق اینترنت شخص ارسال می‌شود. در این مقاله الگوریتم مبتنی بر الگوریتم ژنتیک و طبقه بندی ترکیبی برای تشخیص دیابت در راستای کمک به نظارت پزشکی ارائه شده است. از الگوریتم‌های ژنتیک برای انتخاب ویژگی‌های مرتبط بر اساس همبستگی آنها با وضعیت دیابت و وابستگی‌های بین ویژگی‌ها استفاده می‌کند. متعاقباً، یک مدل یادگیری مجموعه‌ای انباشته، با ادغام طبقه‌بندی‌کننده‌های SVM، KNN، ANN، درختان و GNB برای دقت بیشتر استفاده می‌شود. نتایج نشان‌دهنده عملکرد برتر رویکرد ما است و پتانسیل آن را برای بهبود مدیریت دیابت و نتایج مراقبت‌های بهداشتی برجسته می‌کند. روش پیشنهادی از سه روش ارزیابی شده است و نتایج حاصل نشان می‌دهد روش پیشنهادی از عملکرد بالاتری به میزان ۹ تا ۵۷ درصد نسبت بـه روشهای پایه برخوردار بوده و به دقت ۹۳ درصد رسیده است.

واژه‌های کلیدی: سلامت هوشمند، یادگیری ماشین، اینترنت اشیا، یادگیری گروهی، بیماری دیابت.

نوع مقاله: علمی

۱- مقدمه

سیاست‌های درگیر با حوزه سلامت، زیرساخت‌های مجهز به اینترنت اشیا را برای پردازش سوابق پزشکی الکترونیکی که نیازمند محاسبات قابل توجه و منابع ارتباطی ایمن است، اتخاذ کرده اند. در این زیرساخت، داده‌ها در ابرهایی که معمولاً دور از مرکز مراقبت‌های بهداشتی قرار دارند ذخیره می‌شود. علاوه بر بعد حقوقی و اخلاقی داده‌های پزشکی [۱]، سرویس‌های ابری سنتی ممکن است مقررات مراقبت‌های بهداشتی را نقض کنند،

نویسنده عهده‌دار مکاتبات: سیدابراهیم دشتی Sayed.dashty@gmail.com

اعتماد بیمار- بیمارستان را نقض کنند، امنیت سایبری را به خطر بیاندازند، رویدادهای غیرقابل پاسخگویی شکست شبکه و موارد دیگر را ایجاد کند. بیماری دیابت یکی از عوامل حیاتی مرگ در سراسر جهان است که در دسترس بودن مقادیر عظیمی از اطلاعات پزشکی نیاز به ابزار یادگیری قدرتمندی را برای کمک به متخصصان پزشکی برای تشخیص بیماری دیابت نشان می‌دهد. روشهای یادگیری ماشین در تشخیص بیماری دیابت بسیار مفید است و سطح کارایی آن را افزایش می‌دهد. هم‌چنین طبق این مرجع، بیماری دیابت یک مسئله حیاتی و درعین‌حال رایج‌ترین بیماری در جهان محسوب می‌شود [۱] که بیماری دیابت را به‌عنوان یک بیماری مزمن و همه‌گیر در نظر گرفته‌اند و هم‌چنین یکی از مهم‌ترین مشکلات سلامتی در ایالات‌متحده محسوب می‌شود و پنجمین علت مرگ‌ومیر، ناشی از این بیماری است[۲]، بنابراین بیماری دیابت یک اختلال مزمن غددی درون‌ریز است که بر متابولیسم بدن اثر می‌گذارد و موجب تغییرات ساختاری آن می‌شود بنابراین بیماری دیابت جزو رایج‌ترین بیماری‌هاست که در بدن انسان عارضه ایجاد می‌کند که از سال ۲۰۱۴ به بعد، گسترش این بیماری از۱۰۰ میلیون بیمار به ۴۲۲ میلیون بیمار از افزایش‌یافته است[۳]. این بیماری در حالت معمولی به بیماری دیابت نوع یک و نوع ۲ تقسیم می‌شود که بیماری دیابت نوع۲ یک بیماری بانفوذ بالا در سراسر جهان است که روند آن هنوز هم در حال افزایش بوده و یکی از علل اصلی مرگ‌ومیر به شمار می‌رود که به پیشرفت روند بیماری قلبی نیز کمک می‌کند [۴]. هم چنین از عوارض این بیماری می‌توان به آب‌مروارید، گلوکز و آسیب عروقی خونی داخل چشم اشاره کرد. بنابراین مراقبت‌های بهداشتی در ساده‌ترین شکل یعنی تشخیص و پیشگیری از بیماری یا درمان هرگونه آسیب‌های پزشکی، نقش مهمی در ارائه یک زندگی مفید برای جامعه دارد [۵]. بنابراین تلاش‌هایی برای کاهش تعداد تست‌های تشخیص بیماری مزمن انجام‌گرفته تا هزینه کلی کاهش یابد. یکی از راه‌حل‌های احتمالی استفاده از تکنیک‌های یادگیری ماشین در داده‌های مراقبت بهداشتی است که این تکنیک‌ها برای پیدا کردن الگوهای مکرر در یک پایگاه داده بزرگ استفاده می‌شود تا از آن اطلاعات مفیدی به دست آید.

علاوه براین، شیوع دیابت نوع ۲ در نوجوانان و جوانان به‌طور چشمگیری افزایش می‌یابد. در افراد مبتلابه دیابت نوع ۲، سابقه ابتلا به دیابت نوع ۲، عوامل اصلی خطر ابتلا به چاقی، سابقه خانوادگی و شیوه زندگی بررسی می‌شود.

ادامه مقاله به شرح زیر است: بخش دوم کارهای مرتبط را که در این حوزه انجام‌شده بررسی‌شده در بخش سوم به بررسی سازمان‌بندی اینترنت اشیا در حوزه سلامت هوشمند و در بخش چهار روش پیشنهادی و درنهایت در بخش آخر به نتیجه‌گیری و کارهای آتی می‌پردازیم.

۲- کارهای مرتبط

در این بخش، مراقبت‌های بهداشتی در ساده‌ترین شکل یعنی تشخیص و پیشگیری از بیماری یا درمان هرگونه آسیب‌های پزشکی است که یک نقش مهمی در ارائه زندگی برای جامعه بر عهده دارد که یکی از نگرانی‌ها چگونگی ارائه خدمات بهتر با هزینه کمتر هست که تکنیک‌های یادگیری ماشین در دستیابی به این هدف کمک می‌کند.که پیشرفت‌های قابل‌توجه در بیوتکنولوژی و علوم پزشکی موجب تولید قابل‌توجهی از داده‌ها شده است، مانند داده‌های ژنتیکی و اطلاعات بالینی که از پرونده‌های الکترونیکی سلامت الکترونیکی (EHRs) تولیدشده است. برای این منظور، استفاده از روشهای یادگیری ماشین و داده‌ها در علوم زیستی در حال حاضر، بیشتر از قبل، حیاتی و ضروری در تلاش برای تبدیل هوشمندانه تمام اطلاعات موجود به دانش ارزشمند است که محققان در مرجع [۶] از تکنیک‌های یادگیری ماشین به‌منظور ارائه یک روش هوشمند ترکیبی با استفاده از تجزیه‌وتحلیل مؤلفه‌های اصلی و مدل مخلوط Gaussian، طبقه‌بندی و رگرسیون درختان Cart به‌منظور پیش‌بینی بیماری‌های مزمن استفاده کرده‌اند هم‌چنین در مرجع [۷] یک سیستم هوشمند ترکیبی جدید برای طبقه‌بندی بیماری دیابت با استفاده از تکنیک‌های یادگیری ماشین ارائه‌شده که از الگوریتم‌های خوشه-بندی برای خوشه‌بندی مجموعه داده‌های بیماران و از ماشین

بردار پشتیبانی برای طبقه‌بندی انواع بیماری و از رویکرد PCA برای کاهش ابعاد استفاده‌شده که نتایج نشان داده این روش متشکل از خوشه‌بندی، PCA، SVM دقت طبقه‌بندی خوبی به دست آورده است. علاوه براین در مرجع [۸] محققان به بررسی و مقایسه دقت طبقه‌بندی داده‌های مختلف با استفاده از تکنیک‌های یادگیری ماشین برای پیش‌بینی بیماری مزمن قلب پرداخته‌اند که در این بررسی از الگوریتم‌هایی از قبیلSVM, RF,DT, KNN,NB و درخت تصمیم استفاده‌شده هم‌چنین از الگوریتم‌های تقویتی نظیر Bagging, Boosting, Satcking نیز استفاده شده است که نتایج نشان داده که استفاده از روش Bagging با روش SVM به‌دقت بالای ۸۱،۸۴ درصد رسیده است. هم‌چنین در مرجع [۹] محققان به توسعه یک چارچوب نیمه‌خودکار مبتنی بر یادگیری ماشین برای شناسایی دیابت نوع ۲ از طریق پرونده الکترونیک سلامت ارائه کرده اند که در این چارچوب به ارزیابی و مقایسه کارایی تکنیک‌ها جهت شناسایی مدلهای فراوانی از مدلهای یادگیری ماشین از قبیل:SVM, RF,DT, KNN,NB و رگرسیون منطقی استفاده‌شده که نتایج تجربی نشان می‌دهد که چارچوب پیشنهادی می‌تواند افراد مبتلابه دیابت نوع ۲ را با میانگین ۹۸ درصد تشخیص دهد. درنهایت در مرجع [۱۰] محققان، مجموعه انواع مختلف تکنیک‌ها و الگوریتم‌های یادگیری ماشین و ابزارهایی که برای تحلیل بیماری و فرآیند تصمیم‌گیری مورد استفاده قرار می‌گیرد را مورد بررسی قرار داده اند که به‌منظور تجزیه‌وتحلیل الگوریتم‌های یادگیری ماشین برای تشخیص بیماری‌های مختلفی از قبیل بیماری قلبی، دیابت، کبد و بیماری هپاتیت مورداستفاده قرارگرفته است. درنتیجه یادگیری ماشین و داده‌کاوی نقش مهمی برای کشف دانش در پایگاه داده دارد که یک فرایند تکراری از تمیز کردن داده‌ها؛ ادغام داده‌ها، انتخاب داده و تشخیص الگو و شناخت دانش داده دارد که در حوزه پزشکی نیز حضور قابل‌توجهی دارد که برای تشخیص انواع بیماری‌های مزمن مانند: دیابت، سرطان ریه، بیماری قلبی، نارسایی کلیه، سنگ کلیه و اختلالات کبدی استفاده می‌شود.

۳- سازمان‌بندی اینترنت اشیا در سلامت هوشمند

اینترنت اشیا به‌عنوان یک زیرساخت جهانی برای جامعه اطلاعاتی موردتوجه قرارگرفته و خدمات پیشرفته‌ای به‌وسیله اتصالات (فیزیکی و مجازی) بر اساس اطلاعات موجود و درحال‌توسعه، اطلاعات و ارتباطات در اختیار قرار داده و ازآنجایی‌که بسیاری از فن‌آوری‌ها، خدمات و استانداردهای مختلف را در برمی‌گیرد، که در ده سال آینده به‌عنوان سنگ گوشه‌ای از بازار فناوری اطلاعات در نظر گرفته می‌شود. از دیدگاه منطقی، یک سیستم IoT را می‌توان به‌عنوان مجموعه‌ای از دستگاه‌های هوشمند که بر پایه همکاری تعامل دارند، برای نشان دادن یک هدف مشترک نشان داد که اینترنت اشیا جهان مادی را با جهان اطلاعات ادغام کرده و سرویس‌های معمولی و برنامه‌های کاربردی را برای بهبود زندگی انسانها فراهم می‌کند بی‌تردید، ارتباطات موجودات (اشیاء یا چیزها) در زمینه اینترنت اشیا (IoT)، نقش فعالی در فعالیت‌ها، دستگاه‌ها و پروسه‌های انسانی بازی می‌کند [۱۳،۱۲،۱۱].

بنابراین اینترنت اشیا برای وسایل و حس‌گرهای پزشکی متصل شده نقش مهمی در صنعت مراقبت بهداشتی نسل بعدی برای مراقبت از بیماران باکیفیت دارد. با توجه به افزایش تعداد سالمندان و افراد معلول، یک نیاز فوری به زیرساخت‌های مراقبت‌های بهداشتی در زمان واقعی برای تحلیل داده‌های مراقبت‌های بهداشتی بیماران وجود دارد تا از مرگ‌ومیر قابل‌پیشگیری جلوگیری شود [۱۴] زیرا ارائه‌دهندگان مراقبت‌های بهداشتی به‌طور مداوم رفتار و سلامت سالمندان را نظارت می‌کنند که سیستم مورد انتظار باید وظایفی مانند شناسایی و پیشگیری از تصادفات و انتقال پارامترهای بدن از قبیل: ضربان قلب، دمای بدن، فشارخون، سطح قند خون و...به محل کار (بیمارستان) را انجام دهد که در حوزه سلامت هوشمند استفاده از دستگاه‌های پوشیدنی به‌تدریج قابلیت‌های خود را در دهه‌های گذشته افزایش داده است. [۱۵] که از مزایای مراقبت‌های بهداشتی مبتنی بر IoT در مراقبت از بیماران می‌توان به:

• امکان مشاهده وضعیت بیمار سلامتی بیمار

(بـه‌صـورت جزئی و دقیق) در هرلحظه
• نظارت بر وضعیت بیماران در بیمارستان و خانه سالمندان
• جلـوگیری از هزینـه‌هـای درمـانی غیرضـروری و ارائـه حمایت‌های پزشکی مناسب در زمان مناسب
• ارتقا کیفیت زندگی برای افرادی که نیـاز بـه پشـتیبانی یـا نظارت دائم دارند.

در نتیجه تکنولوژی اینترنت اشـیا در صـنعت مراقبـت‌هـای شخصی دارای مزایای بالقوهای بوده، ولی با چالش‌هـای فراوانـی روبه‌رو است [۱۶] پیشرفت‌های بهداشتی اینترنت اشیاء و یا سلامت هوشمند، بدون چالش نیست کـه یکـی از چـالش‌هـای اصـلی IOT در سـلامت هوشـمند تشـخیص و پیش‌بینی دقیق نتایج بیماری مزمن می‌باشد. که با کمک آن می‌توان به پیشرفت این بیماری و جلوگیری از اثرات منفی آن رسید. برای این منظور داده‌ها از ابزارهاو وسایل اینترنت اشیاء مثل ساعت و سنسورهای هوشمند که به بدن بیماران متصل می‌باشد استفاده می‌شود و اطلاعات آنها پایش می‌شود.

۴- مدل ساختار نظارت بر سلامت و مراقبت پزشکی

با توسعه سریع فن آوری اطلاعات، حجم داده نیز با سرعت شگفت انگیزی افزایش می‌یابد. به تازگی، پردازش ابری و اینترنت اشیا داغ‌ترین موضوع در صنعت فناوری اطلاعات هستند. پردازش ابری مزایایی در مقیاس پذیری عالی و قیمت پایین دارد، در حالی که تکنیک اصلی اینترنت اشیا مانند حسگر و RFID قبلا در مقیاس وسیع مورد استفاده قرار گرفته است. بسیاری از شرکت‌های فناوری معروف مانند مایکروسافت، آمازون، آی بی ام و گوگل قبلا ابر خود را به صورت موفقیت آمیز ساخته اند و خدمات ابری را در مدیریت اطلاعات، ذخیره‌سازی داده‌ها، جستجوی اطلاعات و ... ارائه می‌دهند. با این حال، آشنایی مردم نسبت به ابر رایانه و اینترنت اشیا در برنامه پزشکی مدرن کافی نیست. اینترنت اشیا تولید فناوری اطلاعات است. اینترنت اشیا رشته اطلاعات در فرصت‌های توسعه و تغییر است. کمیته اتحادیه اروپا بر این باور بود که برنامه توسعه اینترنت از طریق حل مسائل مدرن جامعه در آینده حل و فصل خواهد شد و سهم بسیار بزرگی را به همراه خواهد داشت. راهبردهای مدرن با استفاده از تکنولوژی اطلاعات مدرن تحت رفتار ساختار مدیریت مدرن قرار می‌گیرد. در این راستا سه درخواست اساسی شامل سرویس بهتر ؛ هزینه کمتر و سرعت سریعتر می‌باشد. فناوری اطلاعات پزشکی کاربرد گسترده ای در برنامه‌های مدرن دارد.

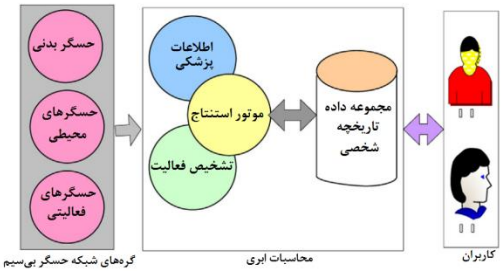
فناوری اطلاعات پزشکی و خدمات بهداشتی و درمانی با رفاه ملی و معیشت مردم ارتباط دارد. ادغام پردازش ابری و اینترنت اشیا یک پیشرفت بزرگ در کاربرد پزشکی مدرن خواهد بود. از آنجا که محاسبات ابری دارای مزایایی در مقیاس وسیع است، قابلیت اطمینان بالا، مجازی‌سازی، راندمان بالا و امکان پذیر ساختن ابر عمومی در بیمارستان‌ها و بیماران می‌توانند به اشتراک‌گذاری منابع، صرفه جویی در هزینه و ایجاد نظارت پزشکی و مدیریت سیستم با راندمان بالا کمک کنند. اینترنت اشیا، به عنوان یک پشتیبان مهم در انجام نظارت و مدیریت ایمن، با کارایی وکیفیت بالاست، تکنیک‌های اصلی آن مانند حسگرهای RFID و حسگرهای الکترومغناطیسی صوتی عکس می‌تواند موفقیت‌های زیادی در انتقال اطلاعات پزشکی، نظارت پزشکی هوشمند و مکان دقیق داشته باشد. اینترنت اشیا راحتی زیادی به ویژه در نظارت و مدیریت ردیابی بیمار را به بیمارستان اختصاص می‌دهد. با توسعه سریع اینترنت و ادغام پردازش ابری و اینترنت اشیا، پلت فرم مراقبت و نظارت پزشکی فرصت جدیدی را برای بیمارستان و حتی تمام جامعه فراهم می‌کند.

این مقاله تکنولوژی اطلاعات پزشکی را در زمینه پردازش ابری و اینترنت اشیا، به ویژه در زمینه کاربرد نظارت و مدیریت پزشکی را بکار گرفته است. یک معماری مدل برای نظارت از راه دور و مدیریت پلت فرم ابر اطلاعات بهداشتی پیشنهاد و تحلیل می‌شود و سپس یک الگوریتم مبتنی بر الگوریتم ژنتیک کارآمد برای نظارت پزشکی و مدیریت کاربرد پردازش ابری ارائه شده است که این مقاله از طریق ایجاد یک زمان‌بندی کارآمد و با الگوریتم ژنتیک این مطالعه را بهبود می‌بخشد.

پلت فرم ابر کنترل از راه دور در اطلاعات بهداشتی شـامل حسگرهای بدن، شـبکه حسـگر، مـاژول ارتبـاطی، درگـاه[۱] مربـوط بـه خانـه، تجزیـه و تحلیـل اطلاعـات پزشـکی و پردازش پلت فرم، وسایل پزشکی و غیـره اسـت. شـکل ۱ یک معماری کلـی پلـت فـرم نظـارت از راه دور اطلاعـات

[۱] gateway

مراقبت‌های بهداشتی را نشان می‌دهد که شامل سه ماژول اصلی یعنی حسگرهای شبکه حسگر بی‌سیم، مرکز محاسبات ابری و کاربران است و همچنین می‌تواند به عنوان یک نمونه اولیه از اینترنت اشیا در نظر گرفته شود.



شکل ۱. معماری کلی پلت فرم نظارت از راه دور اطلاعات مراقبت‌های بهداشتی

این سیستم دارای اجزایی به شرح زیر است:

- مرکز محاسباتی در شبکه حسگر بی‌سیم، ماژول محاسبات ابری، موتور استنتاج و غیره
- مرکز برنامه‌ریزی منابع بین کاربر و پایگاه داده.
- مرکز ارتباطی میان محاسبات ابری، شبکه حسگر بی‌سیم و ماژول‌های کاربر.

در این محیط حسگرهای بدن قادر به ثبت و حذف شبکه ناحیه پزشکی بدن می‌باشند. حسگر پزشکی بدن را می‌توان به حسگر تعبیه شده بر روی بدن و حسگر پوشیدنی تقسیم کرد. این حسگر باید مصرف انرژی کم، کوچک و تهاجم کمی به بدن انسان داشته باشد. مزیت حسگر پوشیدنی راحتی استفاده از آن است؛ اما از زمانی که حالت سنجش محدود شده است، کاربرد آن محدود شده است. حسگر تعبیه شده بر روی بدن می‌تواند محدوده آن را گسترش دهد، اما با این حال باید از طریق عمل در بدن کاشت شود و بنابراین دست‌کاری آن دشوار است. علاوه بر این حسگرهای کنترلی سلامت بدن بر روی انسان تمرکز می‌کنند، بنابراین باید احتمال جابجایی و تحرک آن‌ها را در نظر گرفت. حسگرهای معمول در سیستم کنترلی سلامت بدن عبارت از فشارخون، درجه حرارت بدن، موقعیت و غیره می‌باشد. ماژول ارتباطی، اطلاعات پزشکی بدن را به gateway خانه یا تلفن همراه انتقال می‌دهد. این اطلاعات پزشکی به مرکز ذخیره‌سازی داده‌ها و پردازش داده ارسال می‌شود. سپس پس از پردازش سیستم خبره، راهنمایی‌های بهداشتی به بیمار یا بیمارستان ارسال می‌شود.

اطلاعات پزشکی جمع‌آوری‌شده توسط شبکه حسگر متنوع است. داده‌های پرونده سلامت شخصی بسیار زیاد است و به سرعت در حال افزایش است؛ بنابراین مقدار زیادی از اطلاعات نیاز به طبقه‌بندی خودکار، تجزیه و تحلیل و پردازش دارند. اطلاعات برای همه ارائه‌دهندگان خدمات در دسترس است. به عنوان مثال، مرکز خدمات ایمنی ارائه خدمات امنیتی؛ مرکز کنترل اضطراب بیماری در پیشگیری از وضعیت بیماری گروهی را کنترل و مدیریت می‌کند. مرکز خدمات اضطراری با توجه به اطلاعات استخراج شده کمک‌های اولیه ارائه می‌دهد؛ بیمارستان تشخیص از راه دور تحت اطلاعات پزشکی را فراهم می‌آورد؛ مرکز توان‌بخشی راهنمایی‌های توان‌بخشی از راه دور با توجه به انواع اطلاعات بازیابی فراهم می‌کند. مرکز خدمات بهداشتی، از طریق استخراج اطلاعات بهداشتی خود، به کاربر راهنمایی می‌کند. به این ترتیب باید ابزار به اشتراک‌گذاری منابع شبکه مورد نیاز باشد. علاوه بر این، یک روش مؤثر برای استخراج اطلاعات ضروری است. شبکه حسگر بدن باید قابلیت انتقال قابل اعتماد را دارا باشد. مطالعات امروزی نشان می‌دهد که حسگر تعبیه شده در بدن تمایل به استفاده از ۴۰۰ مگاهرتز دارد و حسگر پوشیدنی مایل به استفاده از فرکانس ۲/۴ گیگاهرتز است. بنابراین هر کدام دارای قابلیتها و محدودیتهای خاص خود می‌باشند، با توجه به محیط و موارد دیگر می‌توان از نوع مناسب استفاده نمود.

۵- روش پیشنهادی برای تشخیص بیماری مبتنی بر نظارت پزشکی با اینترنت اشیا

معرفی تشخیص بیماری بر اساس شبکه اینترنت اشیا (IoT) نشان دهنده یک پیشرفت پیشگامانه در زمینه مراقبت‌های بهداشتی و سلامت عمومی است. اینترنت اشیا با شبکه به هم پیوسته دستگاه‌ها و حسگرها، نقشی اساسی در ایجاد انقلابی در نظارت، پیشگیری و مدیریت بیماری ایفا می‌کند. در این سیستم نوآورانه، حسگرها و دستگاه‌های IoT به صورت استراتژیک در مکان‌های مختلف، مانند امکانات مراقبت‌های بهداشتی، فضاهای عمومی و حتی داخل خانه‌ها مستقر می‌شوند. این حسگرها برای جمع‌آوری طیف متنوعی از داده‌ها، از جمله علائم حیاتی، عوامل محیطی و اطلاعات سلامتی تولید

شده توسط کاربر، مجهز هستند. سپس این داده‌ها به صورت بلادرنگ به یک پلتفرم مرکزی منتقل می‌شوند، جایی که می‌توان آن‌ها را با الگوریتم‌های پیشرفته تجزیه و تحلیل و پردازش کرد.

یکی از مزایای کلیدی استفاده از اینترنت اشیا در تشخیص بیماری، توانایی آن در ارائه اطلاعات به موقع و دقیق است. با پایش مداوم علائم حیاتی و پارامترهای محیطی، دستگاه‌های اینترنت اشیا می‌توانند علائم هشدار دهنده اولیه بیماری‌های عفونی، شرایط مزمن یا حتی عوامل محیطی را که ممکن است مشکلات سلامتی را تشدید کنند، شناسایی کنند. این رویکرد پیشگیرانه، متخصصان مراقبت‌های بهداشتی را قادر می‌سازد تا به سرعت مداخله کنند، از گسترش بیماری‌ها جلوگیری کرده و شدت بیماری را کاهش دهند. اینترنت اشیا همچنین نظارت از راه دور بیمار را تسهیل می‌کند، که به ویژه در مدیریت بیماری‌های مزمن ارزشمند است. بیماران می‌توانند از دستگاه‌های پوشیدنی متصل به شبکه اینترنت اشیا برای ردیابی وضعیت سلامتی و پایبندی خود به برنامه‌های درمانی استفاده کنند. ارائه دهندگان مراقبت‌های بهداشتی می‌توانند به این داده‌ها در زمان واقعی دسترسی داشته باشند و آنها را قادر می‌سازد تا تصمیمات آگاهانه بگیرند و مراقبت شخصی را بدون نیاز به بازدیدهای مکرر حضوری ارائه دهند. علاوه بر این، اینترنت اشیا می‌تواند در ردیابی شیوع بیماری‌های عفونی در مقیاس بزرگ بسیار مفید باشد. با تجزیه و تحلیل داده‌های حسگرهای مختلف، مقامات بهداشتی می‌توانند نقاط داغ بیماری را شناسایی کنند، روندها را پایش کنند و منابع را به طور مؤثر تخصیص دهند. این رویکرد مبتنی بر داده‌ها می‌تواند به کاهش تأثیر شیوع و همه‌گیری کمک کند و پاسخی چابک‌تر و کارآمدتر را ممکن می‌سازد. اینترنت اشیا ابتکارات بهداشت عمومی را با ترویج اقدامات پیشگیرانه افزایش می‌دهد. دستگاه‌های هوشمند می‌توانند اطلاعات و هشدارهای بهداشتی را در زمان واقعی منتشر کنند، به مردم در مورد خطرات بیماری‌ها آموزش دهند و رفتارهای سالم را تشویق کنند. به عنوان مثال، دماسنج‌های هوشمند می‌توانند دمای بالا را در افراد تشخیص دهند و به آن‌ها اطلاع دهند که به دنبال مراقبت‌های پزشکی باشند یا خود را ایزوله کنند و به مهار زودهنگام بیماری‌ها کمک کنند.

یکی از روش‌های کلیدی برای جمع‌آوری اطلاعات مربوط به بیماری دیابت از بیماران، استفاده از سنسورهای نظارت مداوم اطلاعات حیاتی بیمار است. این حسگرهای مینیاتوری در زیر پوست بیمار، معمولاً روی شکم، کاشته می‌شوند و به طور مداوم سطح اطلاعات حیاتی بیمار را اندازه گیری می‌کنند. این داده‌ها جمع آوری شده و به صورت بی سیم به گیرنده یا گوشی هوشمند منتقل می‌شوند، جایی که می‌توان آن‌ها را بیشتر تجزیه و تحلیل و تجسم کرد. حسگرهای اطلاعات حیاتی بیمار داده‌های بی‌درنگ را ارائه می‌دهند و بیماران و ارائه‌دهندگان مراقبت‌های بهداشتی را قادر می‌سازند تا بر روند گلوکز را نظارت کنند و درباره دوز انسولین، انتخاب‌های غذایی و رژیم‌های ورزشی تصمیم‌گیری آگاهانه بگیرند. علاوه بر این، ماهیت مداوم داده‌های اطلاعات حیاتی بیمار به تشخیص زودهنگام خطرات مرتبط با دیابت کمک می‌کند و امکان مداخلات به موقع را برای جلوگیری از نوسانات خطرناک در سطح قند خون فراهم می‌کند. این روش‌ها برای جمع‌آوری اطلاعات مرتبط با دیابت، از فناوری‌های اینترنت اشیا استفاده می‌کنند تا دیدی جامع از مدیریت دیابت به بیماران و ارائه‌دهندگان مراقبت‌های بهداشتی ارائه کنند. این راه‌حل‌ها با نظارت مستمر سطوح گلوکز و تجویز انسولین، افراد مبتلا به دیابت را قادر می‌سازد تا تصمیمات آگاهانه بگیرند و به تیم‌های مراقبت‌های بهداشتی اجازه می‌دهد تا مداخلات مراقبتی شخصی‌تر و به موقع‌تری را ارائه دهند.

استفاده از فناوری یادگیری ماشین در سالهای اخیر محبوبیت زیادی پیداکرده است بطوریکه یادگیری ماشین علمی است که کامپیوتر بدون اینکه یکپارچه برنامه‌ریزی شوند با استفاده از داده‌های آموزشی که از قبل به سیستم Train می-شوند یاد می‌گیرند که در این مقاله از رویکرد یادگیری گروهی که یک الگوریتم تقویتی به‌منظور بهبود دقت پیش‌بینی بیماری دیابت ارائه‌شده است که قبل از استفاده از این روش، ما از رویکرد انتخاب ویژگی به‌منظور یافتن ویژگی هایی که درروند تشخیص سریع دیابت واریانس بیشتری دارند استفاده می‌کنیم تا نتایج دقت پیش‌بینی را با استفاده از نتایج انتخاب ویژگی به دست بیاوریم. انتخاب ویژگی، فرآیند یافتن متغیرهای مرتبط برای یک مدل پیش‌بینی است. این تکنیک‌ها می‌توانند برای

شناسایی و حذف ویژگی‌های غیرضروری و بی‌معنی که باعث کاهش صحت مدل پیش‌بینی شوند. انتخاب ویژگی و انتخاب نمونه دو مرحله مهم پیش‌پردازش داده‌ها درداده کاوی است، که در آن سابق باهدف حذف برخی از ویژگی‌های نامناسب و یا ازکارافتاده از یک مجموعه داده اختصاص داده‌شده است و دومی در حذف داده‌های معیوب است. الگوریتم‌های ژنتیکی برای این وظایف در مطالعات مرتبط به‌طور گسترده‌ای مورداستفاده قرارگرفته‌اند. بااین‌حال، این دو وظیفه قبل از پردازش اطلاعات به‌طورکلی در ادبیات جداگانه در نظر گرفته‌شده است. بنابراین هدف این مطالعه، انتخاب و انتخاب نمونه بر اساس الگوریتم‌های ژنتیکی با استفاده از اولویت‌های مختلف برای بررسی عملکرد طبقه‌بندی بر روی مجموعه داده‌های مختلف است.

که در این مقاله ما از الگوریتم ژنتیک (GA) برای انتخاب ویژگی‌های بهینه استفاده می‌کنیم. این روش فضای زیرمجموعه‌های احتمالی را برای به دست آوردن مجموعه‌ای از ویژگی‌هایی که حداکثر دقت پیش‌بینی‌شده را به حداقل می‌رساند و ویژگی‌های نامناسب را به حداقل می‌رساند، بررسی می‌کند. ما یک همبستگی چندگانه را در یک تابع تناسب استفاده‌شده توسط GA برای ارزیابی تناسب هر زیرمجموعه ویژگی در رابطه با دامنه آن معرفی می‌کنیم.

۱-۵- انتخاب ویژگی توسط الگوریتم ژنتیک

به‌صورت ویژه، یک‌راه‌حل بالقوه برای یک مسئله بهینه‌سازی، به‌عنوان یک دنباله ژن که به‌عنوان کروموزوم هم شناخته می‌شود، در الگوریتم ژنتیک رمزگذاری شده است. جمعیتی از راه‌حل‌های احتمالی از کروموزوم‌ها، به‌سوی راه‌حل‌های بهتر، با داشتن تعداد کمتری از ویژگی‌ها تکامل پیداکرده و در نهایت به راه‌حل بهینه دست پیدا می‌کند. معمولاً چنین راه‌حل‌هایی، در یک رشته دو دویی ۰ و ۱ رمزگذاری می‌شوند. در زنجیره دودویی ۰ نشان می‌دهد که ویژگی‌های مرتبط انتخاب نشده و از مجموعه ویژگی حذف شده‌اند، در حالی که ۱ بر این دلالت دارد که ویژگی موردنظر، انتخاب‌شده است.

کروموزوم‌های اولیه به صورت تصادفی به عنوان والدین انتخاب می‌گردند که می‌توانند یک نسل[1]جدید را تولید کنند که این امر از طریق اعمال عملگرهای ژنتیک، یعنی انتخاب، جهش و تقاطع در فرآیند تکامل حاصل می‌گردد. عملکرد هر کروموزوم توسط یک تابع برازش[2]و شایستگی مورد بررسی قرار گرفته که نزدیکی هر کروموزوم به راه‌حل را اندازه‌گیری می‌کند. این تابع شایستگی، تفکیک‌پذیری[3]نمونه‌ها با دسته‌های مختلف را مدنظر قرار داده و معیار تکراری بودن[4]در بین ویژگی‌ها را در برمی‌گیرد. در این حال، می‌توان راه‌حل‌هایی را پیدا کرد که دربردارنده تعداد کوچک‌تری از ویژگی‌ها باشند. پس از این مرحله ارزیابی، یک گروه از والدین[5](منشاء ها) از جمعیت حاضر با استفاده از انتخاب چرخ رولت[6]و روش تقاطع تک نقطه‌ای[7] انتخاب می‌شوند. تعدادی از بهترین اعضا مستقیماً به عنوان بخشی از جمعیت جدید بر مبنای یک راهبرد نخبه گرایانه[8] کپی می‌شوند. پس از این، عملیات تقاطع و جهش برای شکل دادن به یک نسل جدید، مورد استفاده قرار می‌گیرند. در هر نسلی، تعداد کروموزوم‌ها در جمعیت با حذف کروموزوم‌هایی که دارنده کمترین مقادیر شایستگی هستند، بدون اصلاح باقی می‌ماند. این فرآیند تا زمانی تکرار می‌شود که یک کیفیت از پیش تعیین‌شده از شایستگی یا تعداد حداکثر تکرارها به دست آید. بهترین کروموزوم در جمعیت نهایی، راه‌حل بهینه برای مسئله انتخاب ویژگی می‌باشد.

حال مراحل این الگوریتم را با جزئیات بیشتری مورد بررسی قرار می‌دهیم :

در مرحله‌ی اول یک جمعیت تصادفی از ۱۰۰ کروموزوم متفاوت داریم که این کروموزوم‌ها شامل مقادیر صفر و یک هستند. هر کروموزوم به صورت یک بردار ۵۶بعدی (۵۶×۱) می‌باشد که در اینجا ستون‌های صفر بیانگر عدم انتخاب ویژگی و ستون‌های یک بیانگر انتخاب ویژگی هستند. عدد ۵۶ بیانگر تعداد ویژگی‌های مجموعه داده مورد استفاده است. حال در مرحله‌ی دوم بایستی

1. Offspring
2. Fitness Function
3. Separability
4. Redundant Measurement
5. Parent
6. Roulette-Wheel
7. Single-point Crossover Method
8. Elitist Strategy

ستون‌های صفر حذف شوند و فقط ستون‌های یک باقی بمانند و عمل طبقه بندی بر رو ویژگی‌های موجود انجام گردد.

۲-۵- تابع ارزیابی الگوریتم ژنتیک

یک راه‌حل بالقوه برای یک مسئله بهینه‌سازی، به‌عنوان یک دنباله از اعداد باینری که به‌عنوان کلونی هم شناخته می‌شود در الگوریتم ژنتیک رمزگذاری شده است. جمعیتی از راه‌حل‌های احتمالی از کروموزوم‌ها و کشورهای تحت سلطه آن‌ها، به سوی راه‌حل‌های بهتر، با داشتن تعداد کمتری از ویژگی‌ها تکامل پیدا کرده و درنهایت به راه‌حل بهینه دست پیدا می‌کند. معمولاً چنین راه‌حل‌هایی، در یک رشته دو دویی ۰ و ۱ رمزگذاری می‌شوند. در زنجیره دودویی ۰ نشان می‌دهد که ویژگی‌های مرتبط انتخاب نشده و از مجموعه ویژگی حذف شده‌اند، در حالی که ۱ بر این دلالت دارد که ویژگی موردنظر، انتخاب‌شده است. راه‌حل‌های اولیه به صورت تصادفی به‌عنوان جمعیت کروموزوم‌ها انتخاب می‌گردند که می‌توانند یک نسل[1]جدید را تولید کنند که این امر از طریق ترکیب وتقاطع کروموزوم‌ها و سپس اعمال فرآیند جهش و به روز رسانی آن‌ها حاصل می‌گردد. عملکرد هر کروموزوم توسط یک تابع برازش[2]و شایستگی موردبررسی قرارگرفته که نزدیکی هر راه‌حل به راه‌حل را اندازه‌گیری می‌کند. در بسیاری از رویکردهای انتخاب ویژگی قبلی تنها از یک تابع هدف بر اساس بررسی توزیع داده‌ها پشتیبانی می‌گردد که این امر منجر به حذف بسیاری از راه‌حل‌های کارآمدی می‌گردد که ممکن است بر اساس اهداف کارایی دیگری دارای قدرت باشند. در واقع ارزیابی این زیرمجموعه‌ها با یک روش انجام می‌شود که منجر به پس زدن زیرمجموعه‌هایی با خواص خوب با توجه به معیار دیگری می‌شود. در این زمینه، راه‌حل با افزونگی بسیار کم یا همبستگی بسیار بالا که دو هدف کارایی بسیار مهم می‌باشند، می‌تواند توسط فرآیند انتخاب رد شود. برای افزایش تنوع زیرمجموعه‌های انتخاب شده ، در این مقاله فضای جستجوی بسیار بزرگ مسئله با یک روش چند منظوره مورد بررسی قرار می‌گیرد. ارزیابی کیفیت اهداف کارایی بیان شده براساس اطلاعات متقابل[3]صورت می‌پذیرد تا به‌طور جداگانه میزان همبستگی داده‌ها و افزونگی زیر مجموعه ویژگی‌های انتخابی را اندازه‌گیری کند. این دو معیار هر دو کیفیت اختصاصی ویژگی‌های انتخاب شده و کیفیت زیر مجموعه را اندازه‌گیری می‌کند.

- فرموله‌سازی مسئله

در این بخش ما فرموله سازی معیارهای انتخاب ویژگی به عنوان تابع هدف الگوریتم ژنتیک و نحوه محاسبه آن‌ها را ارائه می‌دهیم. اطلاعات متقابل یک شاخص خوب برای مطالعه وابستگی بین یک ویژگی و طبقه‌بندی و افزونگی بین ویژگی‌های تصادفی است. X و Y را به عنوان دو متغیر تصادفی با قوانین احتمال گسسته در نظر بگیرید. اطلاعات متقابل دو متغیر X و Y با $I(X,Y)$ نشان داده شده و از طریق $P(X)$ و $P(Y)$ و $P(X,Y)$ به صورت زیر تعریف می‌شود:

$$I(X;Y)=\sum_{y\in\Omega Y}\sum_{x\in\Omega X}P(x,y).\log\frac{P(x,y)}{P(x).P(y)} \qquad (۱)$$

که در آن Ω_X و Ω_Y به ترتیب فضاهای نمونه X و Y به ترتیب هستند. هنگامی که دو متغیر X و Y به یکدیگر وابسته هستند. $I(X,Y)$ زیاد است و در نقطه مقابل هنگامی که X و Y به‌طور کامل از هم مستقل هستند، $I(X,Y)$ برابر با صفر است. بعد از محاسبه اطلاعات متقابل، ما دو هدف کارایی مختلف را برای انتخاب بهترین زیرمجموعه از بین راه‌حل‌های تولیدی توسط الگوریتم ژنتیک را دنبال می‌کنیم. اولین هدف کیفیت مجموعه ویژگی‌های انتخابی در راستای پیش‌بینی هدف داده‌ها که تحت عنوان کلاس بیان می‌گردد، را ارزیابی می‌کند و دومین هدف به بررسی افزونگی و تکراری بودن مجموعه ویژگی‌ها می‌پردازد. این دو هدف در ادامه بیان شده‌اند.

- همبستگی ویژگی‌ها با کلاس

برای هر زیرمجموعه از ویژگی‌ها، ما مفهوم ارتباط را که از طریق وابستگی بیان می‌گردد را تعریف می‌کنیم که در واقع از طریق محاسبه میانگین اطلاعات متقابل بین هر یک از ویژگی‌های موجود در مجموعه ویژگی‌های انتخابی با مجموعه هدف داده‌ها یعنی کلاس داده‌ها که از طریق

1. Offspring
2. Fitness Function
3. Nutual Information (MI)

متغیر C نشان داده می‌شود، به دست می‌آید. نحوه محاسبه این رابطه در زیر نشان داده شده است:

$$D_S = \frac{1}{|S|}\sum_{X_i \in S} I(X_{i;c}) \quad (۲)$$

که در آن I(xi, c) بیان‌کننده مقدار اطلاعات متقابل بین ویژگی iام در زیرمجموعه ویژگی انتخابی کاندید و کلاس داده‌ها می‌باشد و بیان‌کننده این واقعیت است که این ویژگی با چه کیفیتی کلاس داده‌ها را بیان می‌کند.

- کاهش افزونگی بین ویژگی‌های انتخابی

در یک راه‌حل انتخابی ممکن است، دو یا بیش از دو ویژگی در هدف اول یعنی ارتباط و همبستگی با کلاس بسیار خوب باشند، اما این ویژگی‌ها به نحوی دارای افزونگی بوده و بتوان با داشتن یکی از آن‌ها به مقادیر دیگر ویژگی‌ها دست یافت. در این شرایط می‌بایست ویژگی‌های انتخابی در راستای کاهش افزونگی و حذف ویژگی‌های تکراری بررسی گردند. در اینجا از اطلاعات متقابل برای ارزیابی افزونگی بین ویژگی‌ها استفاده می‌گردد. در اینجا دو متغیر i و j از یک مجموعه ویژگی‌های کاندید توسط الگوریتم ژنتیک به‌صورت $(X_i; X_j)_{i,j=1,\dots,m_{i\in j}}$ نشان داده شده و از طریق رابطه زیر محاسبه می‌گردد:

$$R_S = \frac{1}{|S|^2}\sum_{X_i, X_j \in S} I(X_i; X_j) \quad (۳)$$

بعد از محاسبه دو هدف مورد استفاده و تحقیق در این مقاله یعنی همبستگی داده‌ها با کلاس و همچنین افزونگی ویژگی، تابع هدف و نهایی این تحقیق برای ایجاد توازن مناسب در بین این دو هدف بیان می‌گردد. مدل ما دیدگاهی منحصر به فرد و نوآورانه را برای استفاده از الگوریتم‌های ژنتیک در تشخیص بیماری دیابت معرفی می‌کند. در حالی که الگوریتم‌های ژنتیک واقعاً در مطالعات قبلی مورد استفاده قرار گرفته‌اند، نوآوری ما در استفاده از یک رویکرد مبتنی بر فیلتر برای محاسبه تابع هدف در طول انتخاب ویژگی نهفته است. به طور سنتی، الگوریتم‌های ژنتیک در انتخاب ویژگی عمدتاً بر به حداقل رساندن همبستگی بین ویژگی‌های انتخاب شده برای کاهش افزونگی تمرکز می‌کنند. در مقابل، رویکرد ما فراتر از این پارادایم مرسوم است. هدف ما نه تنها به حداقل رساندن همبستگی بین ویژگی‌های انتخاب‌شده است، بلکه افزایش همبستگی بین این ویژگی‌های انتخاب‌شده و متغیر هدف، که وضعیت دیابت در مورد ما است، در اولویت قرار دارد. این تابع دوهدفه نشان دهنده یک رویکرد جدید و ظریف تر برای انتخاب ویژگی است، رویکردی که به طور خاص برای نیازهای تشخیص دیابت طراحی شده است. هدف ما افزایش معیار اول یعنی همبستگی ویژگی‌ها با کلاس و کاهش هدف دوم یعنی افزونگی مابین ویژگی‌هاست. در اینجا از یک تابع هدف به‌صورت رابطه زیر استفاده می‌گردد:

$$Fitness = \frac{D_S}{R_S} \quad (۴)$$

در این تابع هدف همبستگی داده‌ها در صورت کسر و هدف دوم یعنی افزونگی در مخرج کسر قرار می‌گیرد. با بزرگ شدن صورت و کوچک شدن مخرج حاصل کسر بزرگ شده و کیفیت بالاتری را نشان می‌دهد؛ بنابراین یک تعادل و توازن ما بین این دو هدف کارایی برقرار می‌گردد.

۶- طبقه بندی با یادگیری گروهی

یادگیری گروهی یکی از روش‌های یادگیری ماشین بوده که از الگوریتم‌های فرا یادگیری برای دستیابی به نتایج پیش‌بینی کننده بهتر استفاده می‌کنند تا بتوان از هر الگوریتم یادگیری دقت بهتری به دست آورد. بسیاری از الگوریتم‌های محبوب یادگیری ماشین درواقع Ensemble هستند. در این مقاله، از الگوریتم‌های Stack Ensemble که متشکل از الگوریتم‌های مختلفی که در ادامه شرح داده خواهد شد استفاده می‌شود که درواقع این روش یک کلاس از الگوریتم‌های است که شامل آموزش "یادگیرنده‌های قوی" برای پیدا کردن ترکیبی مطلوب از فراگیران پایه استفاده می‌کند که جزئیات این الگوریتم‌ها در زیر شرح داده‌شده است :

Algorithm Stacking
Input: training data D = {Xi, yi}^{m} i=1
Ouput: ensemble classifier H
Step 1: learn base-level classifiers
for t= ۱:to T do
learn ht based on D
end for
Step2: construct new data set of predictions

for i= to m do
D_h = {Xi, yi}, where Xi } = {h1 (Xi),…,h_T (Xi)}
end for
Step3: learn a meta -classifier
learn H based on D_h
return H

شکل ۲. الگوریتم Stacking

یادگیری گروهی فرایندی است که از طریق آن چندین مدل، مانند طبقه‌بندی‌ها و یا متخصصان، به‌طور استراتژیک تولید می‌شوند و برای حل یک مسئله اطلاعات محاسباتی خاص ترکیب می‌شوند. یادگیری گروهی‌عمدتاً برای بهبود (طبقه‌بندی، پیش‌بینی، تقریب عملکرد، و غیره) عملکرد یک مدل استفاده می‌شود به همین دلیل است که روشهای گروهی در بسیاری از مسابقات معتبر آموزشی، مانند مسابقه Netflix، KDD ۲۰۰۹ و Kaggle قرار می‌گیرد.

۷-ارزیابی نتایج

ما به‌منظور ارزیابی روشهای مورداستفاده و تعیین اینکه چگونه می‌توان از بین مدلهای موجود، مدلی را در نظر گرفت که نسبت به بقیه روشها، دارای بیشترین دقت پیش‌بینی داشته باشد که به مقایسه روشهای دسته‌بندی و یافتن روش مناسب و کارا در این مقاله از تجزیه‌وتحلیل هزینه سود(ماتریس درهم‌ریختگی)، منحنی ROC و سایر مسائل مربوط به انتخاب مدل از قبیل دقت و غیره استفاده کردهایم.

اندازه‌گیری عملکرد برای تعیین اثربخشی الگوریتم طبقه-بندی مورداستفاده قرار می‌گیرد به این صورت که در مسائل دسته‌بندی با دودسته، می‌توان هزینه دسته‌بندی را با یک ماتریس هزینه نشان داد به‌این‌ترتیب که برای دو نوع خطای مثبت اشتباه (FP) و منفی اشتباه (FN) و دو نوع دسته‌بندی به سمت مثبت درست (TN) و منفی درست (TN) که هزینه و سودهای متفاوتی قائل می‌شود که در جدول ۱ نمایش داده‌شده است.

جدول ۱. ماتریس درهم‌ریختگی

Confusion Matrix			Classified As:	
			Negative	Positive
Actual Class	Negative		TN	FP
	Positive		FN	TP

۱-۷- روش‌های ارزیابی

روش‌های ارزیابی	معیارهای ارزیابی
دقت طبقه‌بندی ساده ترین اندازه‌گیری عملکرد دقت است که به معنای درصد دسته‌های درست پیش‌بینی‌شده است و از فرمول زیر محاسبه می‌شود.	$\frac{TP+TN}{TP+FP+TN+FN}$
حساسیت:نرخ مثبت واقعی:اگر نتیجه برای شخص مثبت باشد در چند درصد موارد مدل نیز مثبت خواهد بود که از فرمول زیر محاسبه می‌شود.	$\frac{TP}{TP+FN}$
خاصیت:نرخ منفی واقعی: اگر نتیجه برای شخص منفی باشد در چند درصد موارد مدل نیز نتیجه منفی خواهد بود که از فرمول زیر محاسبه می‌شود.	$\frac{TN}{TN+FP}$
خصوصیات گیرنده (ROC): که برای ارزیابی مدلها، یک نموداری می‌کشد که هرکدام از این مدل ها، که سطح زیر نمودار بیشتری داشته باشد. آن مدل، مدل بهتری است که از معیارهای زیر برای مقایسه مدل‌های دسته‌بندی به‌طور گرافیکی استفاده می‌شود.	$TPR=\frac{TP}{TP+FN}$ $FPR=\frac{FP}{FP+TN}$
اگر مدل نتیجه مثبت بدهد احتمال اینکه شخص به دیابت دچار شود چقدر است:PPV	$PPV=\frac{TP}{TP+FP}$
.اگر مدل نتیجه مثبت بدهد احتمال اینکه شخص به دیابت دچار شود چقدر است:NPV	$NVP=\frac{TN}{TN+FN}$

۲-۷- شبیه‌سازی

در این بخش به شبیه‌سازی مقاله خود پرداخته و ارزیابی می‌کنیم. محیط شبیه‌سازی ما در ماژول Jupyter مبتنی بر برنامه‌نویسی پایتون ورژن ۳,۶ انجام شده است. هم چنین عملکرد دسته‌بندهای مختلف را می‌توان با به‌کارگیری آن در مورد یک مثال مقایسه کرد که به

منظور شبیه‌سازی دسته‌بندها از مجموعه داده بیماران پیمای دیابتی از سایت UCI بهره گرفته شده که این مجموعه داده دارای ۹ متغیر و ۷۶۸ رکورد است که این متغیرها و اسامی اختصاری در جدول ۲ آمده است.

جدول ۲. متغیرهای مسئله و اسامی اختصاری

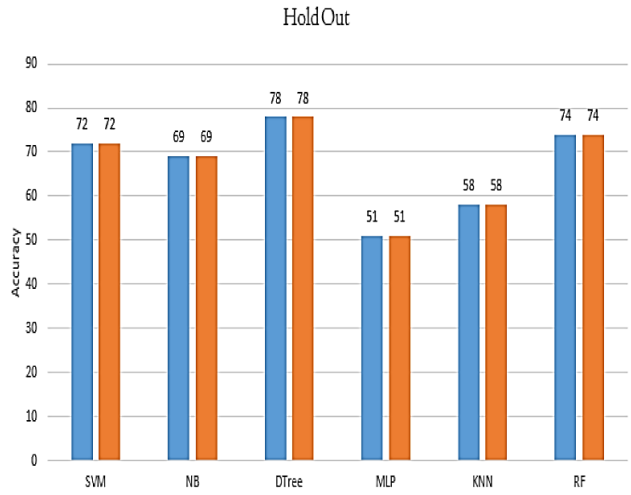
P.NO	Number of times Pregant
PG	Plasma Glucose Concentration
DBP	Diastolic Blood Pressure
TSFT	Triceps Skin Fold Thickness
SI	Two Hour Serum Insulin
BMI	Body Moss Index
DPF	Diabetes Pedigree Function
AGE	Age
C	Class Varible

در نتیجه اشکال زیر، نتایج شبیه‌سازی را برای تک تک الگوریتم‌های مختلف استفاده شده نشان می‌دهد. همان طور که اشاره شد ما از الگوریتم ژنتیک به منظور استخراج ویژگی‌های تأثیرگذار در تشخیص سریع دیابت استفاده کرده ایم که نتایج شبیه‌سازی نشان می‌دهد از بین متغیرهای مجموعه داده فیچرهای ۱- تعداد دفعات بارداری، ۲-غلظت گلوکز دوساعته، ۳-میزان فشارخون، ۴-سطح انسولین و ۵-سن به عنوان فیچرهای مؤثر با دقت ۸۶ درصد انتخاب شدند که نتایج طبقه‌بندی را بر اساس همین ویژگی‌های به دست آمده انجام دادیم که جدول زیر پارامترهای انتخابی برای الگوریتم ژنتیک را نشان می‌دهد.

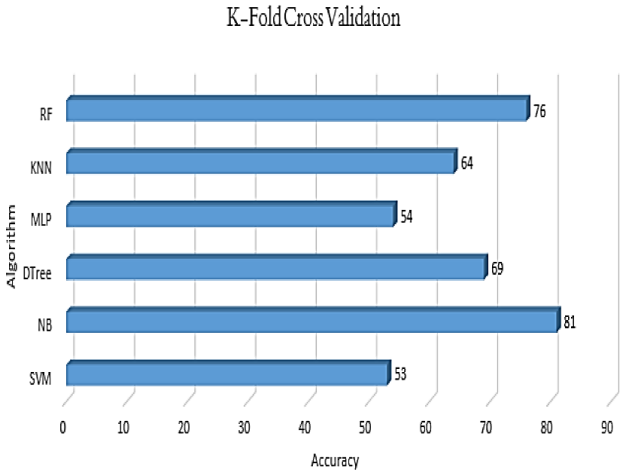
جدول ۳. پارامترهای انتخابی برای الگوریتم ژنتیک

نرخ آمیزش	نرخ جهش	Gen	تعداد جمعیت
۰,۲	۰,۱	۵۶	۴۰

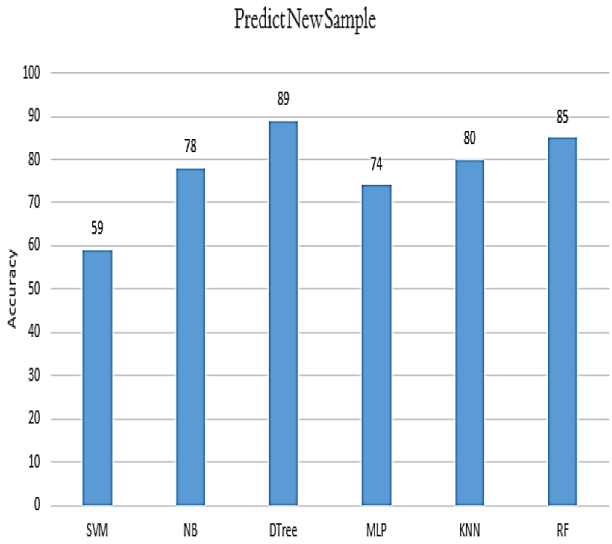
در ادامه به‌مراتب شرح داده خواهد شد که در ابتدا از روش ارزیابی HoldOut برای ارزیابی الگوریتم‌های پایه استفاده کردیم که نتایج شبیه‌سازی در شکل زیر نشان داده شده است.



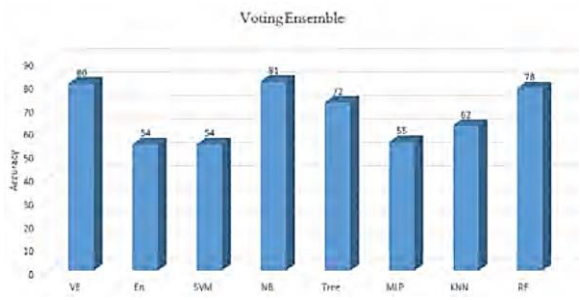
شکل ۳: نتایج به دست آمده از روش HoldOut



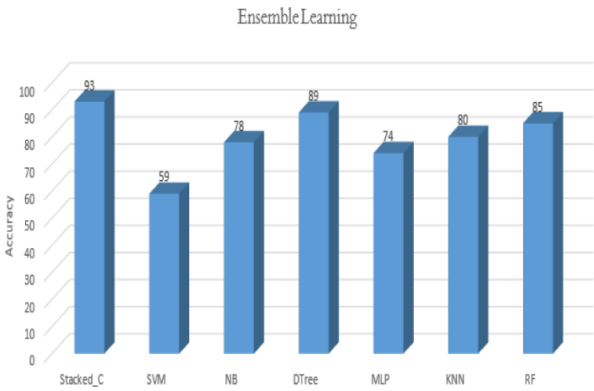
شکل ۴. نتایج به دست آمده از روش K-Fold



شکل ۵. نتایج به دست آمده از دقت پیش‌بینی



شکل ۶. نتایج به دست آمده از روش رأی‌گیری اکثریت



شکل ۷. نتایج به دست آمده از روش پیشنهادی

در نهایت معیارهای عملکرد (ماتریس درهم ریختگی) به ازای روش‌های مختلف روی بیماران دیابتی در جدول ۷- معیارهای عملکرد الگوریتم پیشنهادی را نشان می‌دهد.

جدول ۴. معیارهای عملکرد به ازای روش‌های مختلف

	KNN	SVM	GNB	ANN	RFC	TREE
Accuracy	0.73	0.73	0.70	0.70	85	89
Sensitivity	0.65	0.70	0.83	0.76	85	89
Specificity	0.55	0.75	0.56	0.55	85	89
PPV	0.65	0.87	0.78	0.81	85	89
NPV	0.65	0.55	0.64	1.5	85	89
TPR	0.76	0.70	0.83	0.76	0.99	89
FPR	0.44	0.29	0.43	0.44	0.25	0.42

۸- نتیجه‌گیری و کارهای آتی

در این پژوهش ما مسائل باز سلامت هوشمند را با راهکارهای مختلفی تجزیه و تحلیل کرده و در و چند مورد مشاهده شد که سلامت هوشمند می‌تواند به بهبود مراقبت‌های بهداشتی در نظارت مداوم بر سالمندان کمک کرده و مسیر رسیدن به سلامت هوشمند را هموار سازد. در این مقاله ابزارهای مختلف اینترنت اشیاء برای جمع آوری اطلاعات بیماران مورد بررسی قرار گرفت که بر اساس ویژگی و نوع محیط ابزارهای تعبیه شده و پوشیدنی پیشنهاد شدند سپس این اطلاعات از طریق گیت وی آن مکان یا تلفن همراه شخص جهت تحلیل منتقل می‌شوند. در نهایت رویکرد پیشنهادی ما، پس از ارزیابی و مقایسه با الگوریتم‌های پایه نشان داده است که نسبت به روشهای فوق، پیشرفته بوده و دقت به دست آمده با این مدل ۹۳٪ بوده است. بنابراین با توجه به نتایج این شبیه‌سازی می‌توان تصمیم گرفت که رویکرد پیشنهادی ما با توجه به ارزیابی نتایج، عملکرد بهتری نسبت به سایر روش‌ها دارد و با توجه به اینکه الگوریتم‌های پایه به تنهایی در تشخیص و پیش‌بینی بیماری‌ها مؤثر نبوده، در این مقاله از الگوریتم‌های Stack Ensemble با تلفیقی از تکنیک‌های انتخاب ویژگی هیبرید به منظور بهبود دقت پیش‌بینی استفاده و در آخر با الگوریتم‌های پایه مقایسه شده که الگوریتم پیشنهادی ما خطای کمتری نسبت بهروش‌های پایه داشته و به دقت ۹۳ درصد رسیده است.

رویکرد پیشنهادی به تجزیه و تحلیل داده‌های مرتبط با دیابت و تشخیص بیماری، جامع و نوآورانه است. استفاده از الگوریتم‌های ژنتیک برای انتخاب ویژگی یک تکنیک قدرتمند است، زیرا انتخاب مرتبط ترین ویژگی‌ها را با در نظر گرفتن همبستگی آنها با کلاس هدف (وضعیت دیابت) و یکدیگر بهینه می‌کند. این تضمین می‌کند که ویژگی‌های انتخاب شده برای طبقه بندی آموزنده ترین هستند و به طور بالقوه دقت مدل را بهبود می‌بخشد و بیش از حد برازش را کاهش می‌دهد. استفاده از یک مدل یادگیری مجموعه‌ای انباشته که طبقه‌بندی‌کننده‌های مختلفی مانند ماشین‌های بردار پشتیبان (SVM)، k-نزدیک‌ترین همسایه‌ها (KNN)، شبکه‌های عصبی مصنوعی (ANN)، درخت‌های تصمیم‌گیری، و گاوسی بیزی (GNB) را ترکیب می‌کند، یک استراتژی قوی است. هر طبقه‌بندی نقاط قوت و ضعف خود را دارد و با ترکیب آنها در یک مجموعه، مدلی انعطاف‌پذیرتر و دقیق‌تر ایجاد می‌کنید. گروه‌های انباشته اغلب با استفاده از دانش متنوع هر طبقه‌بندی‌کننده، از مدل‌های فردی بهتر عمل می‌کنند. این واقعیت که نتایج برتری روش پیشنهادی را نشان می‌دهد یک دستاورد قابل توجه است. این نشان می‌دهد که رویکرد پیشنهادی نه تنها انتخاب ویژگی‌ها را افزایش می‌دهد، بلکه نقاط قوت طبقه‌بندی‌کننده‌های متعدد را نیز به‌طور بهینه ترکیب می‌کند، که منجر به یک سیستم تشخیص دقیق‌تر و

قوی‌تر بیماری دیابت می‌شود. این می‌تواند پیامدهای قابل توجهی برای بهبود نتایج مراقبت‌های بهداشتی و مدیریت بیمار در زمینه‌های مرتبط با دیابت داشته باشد. با این حال، انجام اعتبارسنجی و آزمایش دقیق برای اطمینان از تعمیم و قابلیت اطمینان روش خود در مجموعه داده‌های مختلف و سناریوهای دنیای واقعی مهم است. برای کارهای آتی از روش پیشنهادی در تشخیص بیماری‌های مختلف مزمن استفاده می‌گردد. با توجه اهمیت تشخیص بیماری قصد داریم در کار آتی سیستم تشخیص بیماری‌های مزمن به صورت زمان واقعی و سیستم‌های توصیه گر و سامانه‌های پیام کوتاه را در کنار روش پیشنهادی پیاده‌سازی کنیم.

منابع

1. Ahmad, M.O., Siddiqui, S.T. (2022). The Internet of Things for Healthcare: Benefits, Applications, Challenges, Use Cases and Future Directions. In: Tiwari, S., Trivedi, M.C., Kolhe, M.L., Mishra, K., Singh, B.K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 318. .

2. Hameed, G., Singh, Y., Haq, S., Rana, B. (2022). Blockchain-Based Model for Secure IoT Communication in Smart Healthcare. In: Singh, P.K., Kolekar, M.H., Tanwar, S., Wierzchoń, S.T., Bhatnagar, R.K. (eds) Emerging Technologies for Computing, Communication and Smart Cities. Lecture Notes in Electrical Engineering, vol 875..

3. Li, W., Chai, Y., Khan, F. et al. A Comprehensive Survey on Machine Learning-Based Big Data Analytics for IoT-Enabled Smart Healthcare System. Mobile Netw Appl 26, 234–252 (2021)..

4. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, 97, 120-127.

5. Somasundaram, S. K., & Alli, P. (2017). A Machine Learning Ensemble Classifier for Early Prediction of Diabetic Retinopathy. Journal of Medical Systems, 41(12), 201.

6. Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An Analytical Method for Diseases Prediction Using Machine Learning Techniques. Computers & Chemical Engineering.

7. Nilashi, M., Bin Ibrahim, O., Mardani, A., Ahani, A., & Jusoh, A. (2016). A soft computing approach for diabetes disease classification. Health Informatics Journal, 1460458216675500.

8. Pouriyeh, S., Vahid, S., Sannino, G., De Pietro, G., Arabnia, H., & Gutierrez, J. (2017, July). A comprehensive investigation and comparison of Machine Learning Techniques in
the domain of heart disease. In Computers and Communications (ISCC), 2017 IEEE Symposium on (pp. 204-207). IEEE.

9. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., ... & Chen, Y. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics, 97, 120-127.

10. Yaqoob I (2021) Blockchain for healthcare data management: opportunities, challenges, and future recommendations. Neural Comput Appl 0123456789.

11. Shaikh, F. K., Zeadally, S., & Exposito, E. (2017). Enabling technologies for green internet of things. IEEE Systems Journal, 11(2), 983-994.

12. Sicari, S., Rizzardi, A., Grieco, L. A., Piro, G., & Coen-Porisini, A. (2017). A Policy Enforcement Framework for Internet of Things Applications in the Smart Health. Smart Health.

13. Velliangiri S, Karthikeyan Karunya P (2020) Blockchain technology: challenges and security issues in consensus algorithm. In: International conference on computer communication and informatics, ICCCI.

14. Hossain, M. S., & Muhammad, G. (2016). Cloud-assisted industrial internet

of things (iiot)–enabled framework for health monitoring. Computer Networks, 101, 192-202.

15. Temko, A. (2017). Accurate wearable heart rate monitoring during physical exercises using PPG. IEEE Transactions on Biomedical Engineering.

16.Philip NY, Rodrigues JJPC, Wang H, Fong SJ, Chen J (2021) Internet of Things for in-home health monitoring systems: current advances, challenges and future directions. IEEE J Sel Areas Commun 39(2):300–310.