

طراحی یک سیستم هوشمند تشخیص هویت بر اساس پردازش سیگنال ECG

مصطفی لنگری زاده^۱، نوید مشتاقی یزدانی^۲، محمد حسن علیائی طرقله^۳

^۱ استادیار، گروه مدیریت اطلاعات سلامت، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی ایران، تهران، ایران

^۲ دانشجوی دکتری، گروه مهندسی برق، دانشگاه آزاد اسلامی، مشهد، ایران

^۳ کارشناسی ارشد مهندسی برق کنترل دانشگاه صنعتی سجاد، مشهد، ایران

چکیده

هدف اصلی این مقاله ارائه روشی برای بررسی سیگنالهای قلبی برای شناسایی افراد می باشد. برای این کار مجموعه داده‌های ۴۸ ثبت ECG دوکاناله از بانک اطلاعاتی MIT-BIT دریافت و بعنوان نمونه مورد استفاده قرار گرفت. هریک از این نمونه‌ها برای زمان یک ساعت و نرخ نمونه برداری ۱۲۸ هرتز برای هر فرد استخراج شده است. هر فایل در قطعات زمانی ۱۰ ثانیه تقسیم بندی و با استفاده از تبدیل موجک تا سطح ۱۰ تجزیه گشت. با انتخاب ویژگی حد آستانه الگوریتم IDE برابر با ۰/۸، از بین ۶۰ ویژگی ارائه شده تنها ۷ ویژگی (انحراف معیار سیگنال، انرژی سیگنال، واریانس سیگنال، توان سیگنال، انحراف معیار سیگنال، نوان سیگنال و انحراف معیار سیگنال) از حد آستانه، مقدار بیشتری را دارا بودند. سپس این ۷ ویژگی نیز به انواع روش های طبقه بندی برای کسب درصد پاسخگویی و هزینه محاسباتی وارد شد. هزینه محاسباتی روش KNN نسبت به سایر روشها در هر دو حالت انتخاب ویژگی کمتر می باشد در صورتی که از لحاظ درصد درستی روش XCSLA در هر دو حالت انتخاب ویژگی بهترین روش بوده و حتی در انتخاب ویژگی به کمک الگوریتم IDE تمامی طبقه بندی‌ها را به درستی تشخیص داده است. با توجه به نتایج حاصل از بکارگیری سیستم های طبقه بندی گسترش یافته به منظور تشخیص هویت افراد از روی سیگنال قلبی بر اساس نشانگان بالینی می توان اینگونه نتیجه گیری کرد که این سیستم می تواند جایگزین مناسبی جهت شناسایی افراد بعنوان یک روش بیومتریک باشد هرچند لازم است تحقیق بیشتری روی این سیستم انجام شود. محدودیت اصلی این سیستم ممکن است در حالات مختلف انسان از جمله عصبانیت، شادی، غم و سایر وضعیت ها باشد که موجب تغییر در ریتم تپش قلب می گردد.

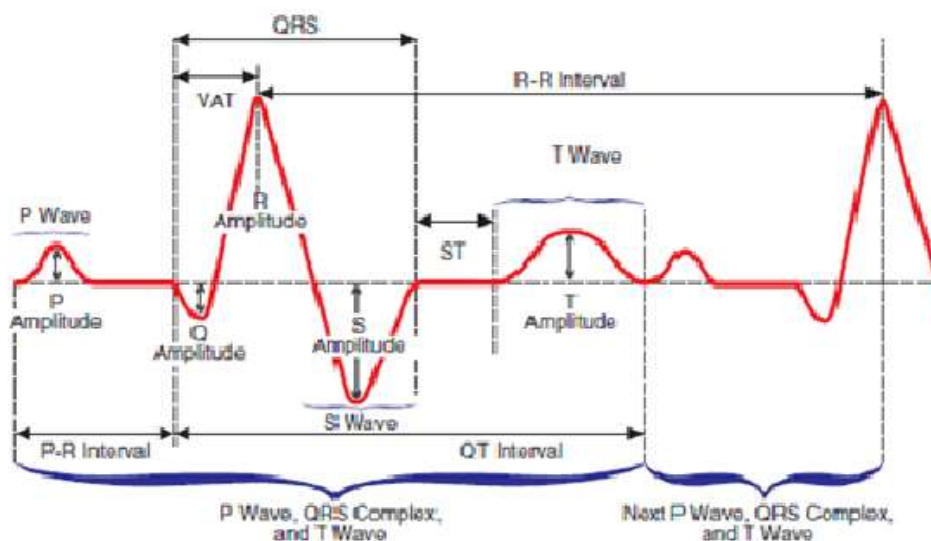
واژه‌های کلیدی: شناسایی افراد، تشخیص هویت، سیگنال قلب، نمونه برداری، تبدیل موجک، الگوریتم، طبقه بندی

۱. مقدمه

امروزه استفاده از علائم حیاتی و خصوصیات فردی اشخاص برای تشخیص هویت، در بسیاری از مراکز و سیستمهای امنیتی و اطلاعاتی امری ضروری به نظر می رسد. این روش ها بدلیل مزیت هایی که نسبت به شیوه های دیگر تشخیص هویت همانند رمزعبور دارند، پیشرفت قابل توجهی داشته و استفاده از آنها افزایش یافته است. روش های بیومتریک بدلیل یکتایی پارامترهای شناسایی افراد و عدم وجود مشکلاتی که در سیستم های دیگر وجود دارد، از لحاظ امنیتی برتری دارند. در حال حاضر اثر انگشت، عنبیه، صدا و تشخیص چهره از جمله رایج ترین روش های بیومتریک در سیستمهای امنیتی هستند. اما با توجه به این موضوع که در منحصر بفرد بودن ویژگی های صدا و چهره جای شک می باشد [۱۲ و ۱۳] و اثر انگشت و الگوی عنبیه چشم نیاز به تصاویری با رزولوشن بالا (بیشتر از ۴۰۰ dpi) دارند. در این مقاله از سیگنال ECG استفاده شده که یکی دیگر از سیستمهای بیومتریک بوده و مختص شخص است می باشد و قابل جعل و سرقت نمی باشد استفاده میکنیم سیگنال الکتریکی که از گره سینوسی - دهلیزی در قلب آغاز می شود، از طریق سیستم تحریکی - هدایتی به تمام قلب منتشر گشته و بخش کوچکی از آن با عبور از بافت های بدن که هادی نسبتا خوبی هستند به سطح بدن می رسد. با قرار دادن الکترودهایی بر سطح پوست در مقابل قلب می توان پتانسیل الکتریکی تولید شده توسط جریان الکتریکی قلب را ثبت نمود منحنی الکترو کاردیو گرام وقایع الکتریکی سه بعدی را در قلب نشان می دهد. به همین دلیل از اشتقاق های مختلف برای بررسی عملکرد قلب استفاده می گردد و با قرار دادن الکترودها در موقعیت های مختلف، می توان اطلاعات جامعی از فعالیت های الکتریکی قلب بدست آورد. در کارهای بالینی معمولی، از ۱۲ اشتقاق قلبی استفاده می شود که شامل سه اشتقاق دو قطبی، سه اشتقاق تک قطبی اندام ها و شش اشتقاق سینه ای یا جلوی قلبی است. محور یک اشتقاق الکترودیوگرافی که از دو الکتروود مثبت و منفی تشکیل شده است، خطی است فرضی که این دو الکتروود را به هم متصل می کند و جهت آن نیز از الکتروود منفی به سمت الکتروود مثبت می باشد

یک منحنی الکترو کاردیوگرام از امواج زیر تشکیل شده است

- ۱- موج P : نشان دهنده انتشار موج دپلاریزاسیون در دهلیز ها است. موج P نرمال یک موج قرینه و یک کوهانه است.
 - ۲- کمپلکس QRS : نشان دهنده انتشار موج دپلاریزاسیون در بطن ها است
 - ۳- موج T : نشان دهنده انتشار موج رپلاریزاسیون در بطن ها است
- از فاصله بین دو کمپلکس QRS متوالی یا فاصله R-R معمولا برای بدست آوردن تعداد ضربان قلب در دقیقه استفاده می کنند. برای این کار عدد ۶۰ را بر متوسط فاصله R-R تقسیم می کنند. (شکل ۱)



شکل ۱- منحنی الکتروکاردیوگرام

علاوه بر تعداد ضربان قلب، با استفاده از الکتروکاردیوگرام می توانیم اطلاعاتی در مورد ریتم قلب، سرعت هدایت سیگنال و حتی شناسایی انسان [۱] بدست آوریم. در مقالات [۲ و ۳ و ۴ و ۵ و ۶] از ویژگی های حوزه زمان سیگنال ECG در جهت شناسایی افراد استفاده شده و در بعضی تحقیقات دیگر از ویژگی هایی مانند خود همبستگی [۱۰ و ۱۱]، تبدیل فوریه [۷ و ۸] و تبدیل موجک گسسته استفاده شده است [۹]. شناسایی مبتنی بر سیگنال های قلبی شامل سه فرآیند اصلی ثبت داده و سیگنالها، استخراج ویژگی و انطباق ویژگی ها می باشد که قبل از استخراج ویژگی بایستی مرحله پیش پردازش بر روی سیگنالهای ثبت شده انجام گرفته سپس به کمک سیستم هوشمند ویژگی های مهم استخراج شده از بیو سیگنالها انطباق و هویت افراد تشخیص داده می شود. در این مقاله روش جدیدی به منظور تشخیص هویت افراد با استفاده از یک سیستم طبقه بندی توسعه یافته با متغیرهای پیوسته (XCSR) بهبود یافته ارائه می گردد. روش بهبود یافته مبتنی بر استفاده چندگانه از مثال های محدود تربیتی موجود است، بدین معنا که پس از ارائه داده های تربیتی موجود به سیستم و به روز آوری مشخصه های قوانین آن، داده های مزبور به کمک عملگر همگذاری (crossover) به داده های جدید و معناداری تبدیل می گردند که با ارائه اطلاعات جدیدی از محیط و الگو (pattern) های موجود در آن، قدرت تعمیم قوانین تشکیل دهنده XCSR را بهبود می بخشند.

۲. تبدیل موجک

تبدیل موجک ابزاری است که داده ها، توابع و یا عملگرها را به اجزای فرکانسی متفاوت تقسیم می نماید و سپس هر جزء را با دقتی متناسب با مقیاس آن بررسی می نماید. همانند تبدیل فوریه، تبدیل پیوسته موجک یک تابع با جمع حاصل ضرب تابع مزبور در تابع موجک مقیاس شده و شیفته یافته در کل بازه زمانی تعریف می شود. نکته قابل توجه آنست که مقیاس بزرگ موجک معادل فرکانس های پایین و مقیاس کوچک معادل فرکانس های بالاست. هر تابعی که به عنوان موجک به کار گرفته می شود دارای میانگین صفر و انرژی واحد است برای آسان تر شدن کار با تبدیل موجک معمولاً گسسته سازی آن به صورت دودویی انجام میگردد با این تعریف موجک گسسته به صورت زیر نوشته می شود.

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}} \Psi(2^{-j}t - k) \quad (1)$$

با توجه به رابطه (۱) و تعریف تبدیل پیوسته موجک، تبدیل گسسته موجک سری زمانی $f(n)$ مطابق رابطه زیر نوشته می‌شود.

$$C(j,k) = \sum_{n \in \mathbb{Z}} f(n) \Psi_{j,k}(n) \quad (2)$$

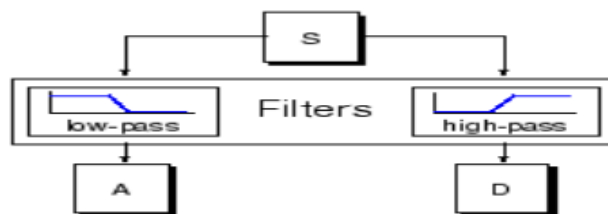
که n در این رابطه (گسسته) معادل t است. بازسازی سیگنال هم با استناد به مطالب یاد شده به سادگی از رابطه (۳) قابل محاسبه است.

$$f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C(j,k) \Psi_{j,k}(t) \quad (3)$$

شایان ذکر است که در فرآیند بازسازی سیگنال این امکان وجود دارد که برحسب نیاز تنها قسمت‌های معینی از آن بازسازی شود بنابراین می‌توان به دلخواه بخشی از سیگنال را حذف کرد. این قابلیت در کاربردهایی نظیر کاهش نویز سیگنال به کار می‌رود. آنچه که گفته شد در فرآیند بازسازی مطابق رابطه (۴) قابل اجرا است.

$$f_j(n) = \sum_{k \in \mathbb{Z}} C(j,k) \Psi_{j,k}(n) \quad (4)$$

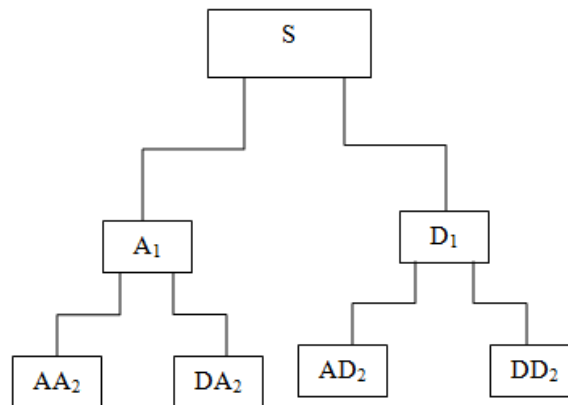
بنابراین می‌توان سیگنال را به مؤلفه‌هایی مرتبط با هر مقیاس تجزیه نمود. در تبدیل موجک یک سیگنال به دو قسمت تجزیه می‌شود که در آن تجزیه با فیلتر پایین‌گذر را تخمین و تجزیه با فیلتر بالا‌گذر را جزئیات می‌نامند. (شکل ۲)



شکل ۲- فیلترهای تبدیل موجک گسسته

۲.۲. ویولت پکت

این تبدیل یک تجزیه ویولت تعمیم یافته است که در آن جزئیات نیز به مانند تقریبها می توانند تجزیه شوند. در شکل ۳ تجزیه یک سیگنال را با استفاده از ویولت پکت تا سطح دوم مشاهده می شود.



شکل ۳- تجزیه سیگنال با استفاده از ویولت پکت تا سطح دوم

به عنوان می توان با استفاده از ویولت پکت سیگنال S را به صورت $AA_2 + DA_2 + D_1$ نشان داد که ویولت معمولی قادر به نحوه بیان سیگنال به این صورت نمی باشد

۳. روش IDE

مسئله انتخاب ویژگی، یکی از مسائلی است که در مبحث یادگیری ماشین و در بررسی مسائل شناخت الگو از دیدگاه آمار مطرح است. این مسئله در بسیاری از کاربردها به مانند طبقه بندی اهمیت به سزایی دارد، زیرا در این کاربردها تعداد زیادی ویژگی وجود دارد، که بسیاری از آنها یا غیر قابل استفاده هستند و یا اینکه بار اطلاعاتی چندانی ندارند. حذف کردن این ویژگی ها مشکلی از لحاظ اطلاعاتی ایجاد نمی کند، ولی حذف ویژگی های نامرتبط و اضافی بار محاسباتی را کاهش می دهد. علاوه بر این باعث می شود که اطلاعات غیر مفید زیادی را به همراه داده های مفید ذخیره کنیم

این روش از تابع ارزیابی مبتنی بر فاصله استفاده می نماید که از شش مرحله تشکیل شده است. در مرحله اول به محاسبه فاصله میانگین نمونه های هر وضعیت پرداخته سپس به محاسبه فاصله میانگین وضعیت ها می پردازیم. در مرحله بعد به محاسبه فاکتور واریانس در مرحله اول می پردازیم. پس از آن در مرحله سوم به محاسبه مقدار ویژه متوسط همه نمونه ها در هر وضعیت پرداخته می شود و پس از آن فاصله میانگین بین نمونه های وضعیت را محاسبه می کنیم در مرحله پنجم فاکتور جبران محاسبه می گردد و در نهایت در مرحله آخر نسبت فاصله میانگین نمونه های هر وضعیت را به فاصله میانگین بین نمونه های وضعیت های متفاوت محاسبه و سپس آن را نرمال سازی می کنیم. بنابراین مؤثرترین ویژگی ها می توانند از مجموعه ویژگی ها بر اساس معیار فاصله از بزرگترین مقدار تا کوچکترین مقدار انتخاب شوند

۴. روش های مختلف طبقه بندی جهت تشخیص عیوب

اتوماتیک کردن کارها و نقش تشخیص الگو در تحقق آن، یکی از مهمترین دلایلی است که موجب شده است شناسایی الگو جایگاه ویژه ای در پژوهش های جدید برای تشخیص انواع بیماری ها داشته باشد. در طراحی سیستمهای تشخیص الگو نیاز به داده هایی از سیستم هایی داریم که به خوبی آن ها را توصیف کنند. برای این منظور مشخصه هایی از آن بیماری ها را در نظر میگیریم که با آنها بتوانیم بین بیماری های مربوط به کلاسهای مختلف تمایز قایل شویم.

۴.۱. الگوریتم KNN

الگوریتم K-NN یا نزدیکترین همسایه یکی از الگوریتم های یادگیری مبتنی بر نمونه است. این الگوریتم در فاز یادگیری تنها نمونه های آموزشی را ذخیره می کند. برای مشخص کردن طبقه (class) یک نمونه داده، الگوریتم مزبور فاصله این نمونه را با سایر نمونه های آموزشی محاسبه میکند. پس از محاسبه فاصله، بین k نزدیک ترین نمونه آموزشی به نمونه تست کنونی یک رای گیری اکثریتی برگزار و برچسب اکثریت این نمونه به نمونه تست اختصاص داده می شود.

۴.۲. الگوریتم SVM

SVM یک روش برای کلاس بندی داده های خطی و غیر خطی می باشد. این روش، ابتدا از یک نگاشت غیر خطی برای تبدیل داده ی اولیه به ابعاد بالا استفاده کرده و سپس در بعد جدید به دنبال بهترین فرایضحه ی جدا ساز می گردد. SVM این فرا صفحه را با استفاده از بردارهای پشتیبان و حاشیه ها که توسط بردار های پشتیبان تعریف می شوند، می یابد. برای جدا کردن داده های متعلق به کلاس های گوناگون، مرزهای تصمیم متعددی وجود دارند اما همواره به دنبال مرز تصمیم یا فرایضحه ی جداساز با حاشیه ی بیشینه یا MMH می باشیم که داده را با دقت بیشتر و خطای کمتر جداسازی می کند.

۴.۳. شبکه های عصبی

شبکه های عصبی، با قابلیت قابل توجه آنها در استنتاج معانی از داده های پیچیده یا مبهم، میتواند برای استخراج الگوها و شناسایی روشهایی که آگاهی از آنها برای انسان و دیگر تکنیک های کامپیوتری بسیار پیچیده و دشوار است به کار گرفته شود. یک شبکه عصبی تربیت یافته می تواند به عنوان یک متخصص در مقوله اطلاعاتی ای که برای تجزیه تحلیل به آن داده شده به حساب آید.

۵. سیستم یادگیرنده XCS

یاد گیری ماشین به طیف گسترده ای از الگوریتم های یادگیری با مربی و بدون مربی اطلاق می شود که هدف از آنها در حوزه داده کاوی، پرهیز از جستجوی سراسری (exhaustive search) داده ها و جایگزین کردن این گونه جستجوی زمان بر با روش های هوشمندانه ای است که از طریق یافتن الگو (ها) ی موجود در میان داده ها، دسته بندی و یا مدل سازی رفتار آنها را به سادگی امکان پذیر می سازد. در دو دهه اخیر روش های زیادی در حوزه داده کاوی ارائه شده است که در آنها از انواع الگوریتم های یادگیری با مربی (supervised)، بدون مربی (unsupervised) و یا تقویتی (reinforcement) برای اهدافی همچون تشخیص و تخصیص الگو استفاده می شود. از جمله موفق ترین این روش ها می توان به سیستم های طبقه بند (classifier systems) اشاره کرد.

در حالت کلی، سیستم های طبقه بند شامل مجموعه ای از قوانین با فرمت «اگر- آنگاه» هستند که هر قانون راه حل بالقوه ای برای مسئله هدف ارائه می نماید. این مجموعه قانون به تدریج با بکار گیری یک سازوکار یادگیری تقویتی ارزیابی شده و در فواصل زمانی مشخصی به کمک یک الگوریتم ژنتیک بروز آوری می شود. در جریان این تکامل تدریجی، سیستم رفتار محیط را می آموزد و سپس در فاز کاربرد، پاسخ های مناسبی را به پرسش (query) های مطرح شده از سوی کاربر ارائه می کند.

اولین سیستم طبقه بند در سال ۱۹۷۶ توسط هالند با عنوان (Learning Classifier System) LCS پیشنهاد شد. در این سیستم، ارزش هر قانون با شاخصی به نام «قوت» (strength) ارزیابی می‌شد. قوت یک قانون متناسب با میزان پاسخگویی صحیح آن به مثال‌های تربیتی در چارچوب ضوابط یادگیری تقویتی افزایش می‌یافت و در فواصل زمانی مشخصی، یک الگوریتم جستجوی تکاملی (عموماً الگوریتم ژنتیک) مسئولیت تولید قوانین جدید و حذف قوانین ناکارآمد را برعهده می‌گرفت. در پایان مرحله تربیت، این مجموعه قوانین توانایی نسبی آن را داشت که در مواجهه با پرسش‌های جدید، راه حل‌های قابل قبولی را ارائه نماید. در عین حال عملکرد موفق LCS منوط به انتخاب مقادیر مناسب برای پارامترهای کنترلی سیستم بود که عمدتاً وابسته به تجربه طراح این سیستم بود.

از زمان به وجود آمدن LCS، انواع دیگری از سیستم‌های طبقه بند پیشنهاد شده است که از میان آن‌ها می‌توان به سیستم‌های طبقه بند توسعه یافته (Extended Classifier Systems: XCS) اشاره کرد. تا قبل از معرفی XCS در سال ۱۹۹۵، توانایی این سیستم‌ها در بدست آوردن پاسخ‌های مناسب بسیار محدود بود. اما از آن زمان این سیستم‌ها به تدریج به عواملی هوشمندتر و دقیق‌تر تبدیل شدند و اکنون اعتقاد بر این است که XCS و نسخه‌های بهبود یافته آن قادرند مسائل پیچیده‌ای را بدون نیاز به تنظیم پارامترها حل کنند. با معرفی سیستم طبقه بندی با متغیرهای پیوسته (XCSR)، برخی ضعف‌های ذاتی سیستم‌های طبقه بندی دودویی (binary) نظیر ناتوانی در معرفی بازه‌های مشخص مقادیر متغیرها تا حد زیادی برطرف شد و امروزه این سیستم‌ها به عنوان یکی از موفق‌ترین عامل‌های یادگیرنده (Learning Agents) برای حل مسایل داده کاوی در محیط‌های نیمه مشاهده پذیر شناخته شده‌اند. [۱۴]

بر اساس رهیافت متداول برای تربیت XCSR، تنها برازندگی قوانینی افزایش می‌یابد که به داده‌های تربیتی پاسخ صحیح بدهند. این بدان معناست که شانس هر قانون برای حذف نشدن و مشارکت در فرایند تولید قوانین جدید، بستگی مستقیم به نحوه پاسخگویی آن به داده‌های تربیتی دارد و تعیین واقع‌گرایانه این شانس نیازمند بکارگیری تعداد زیادی از داده‌های تربیتی است. از آنجا که در مسایل واقعی، تعداد داده‌های تربیتی معمولاً محدود بوده و افزایش تعداد داده‌ها به سادگی میسر نیست، معمولاً استفاده از XCSR در اینگونه کاربردها از نظر زمان و هزینه محاسباتی توجیه پذیر نمی‌باشد. در ادامه این مقاله روش جدیدی برای بهبود عملکرد و افزایش نرخ همگرایی XCSR با استفاده از داده‌های محدود تربیتی ارائه می‌گردد.

۶. روش پیشنهادی XCSLA

در روش پیشنهادی، ابتدا با اضافه کردن یک فیلد LA به انتهای تمامی قوانین که به صورت رندوم تولید می‌گردند می‌توان از آنها به عنوان حافظه‌هایی که وضعیت را ذخیره می‌کنند در نظر گرفت. همانطور که می‌دانیم در اتوماتای یادگیر سلولی بسته سیگنال‌های تقویتی تنها تابعی از عمل انتخابی هر سلول و عمل‌های انتخابی سلول‌های همسایه می‌باشد در روش پیشنهادی مجموعه داده‌های تربیتی برای اصلاح مشخصه‌های قوانین (شامل «پیش‌بینی»، «خطای پیش‌بینی» و «برازندگی» و «مقدار» LA)، بکار گرفته می‌شود. این کار با استفاده از روابط زیر صورت می‌گیرد:

بروزآوری پیش‌بینی و خطای پیش‌بینی

$$\text{If } \text{expi} < 1/\beta \text{ then } P_i = P_i + (R - P_i) / \text{expi}, \quad \varepsilon_i = \varepsilon_i + (|R - P_i| - \varepsilon_i) / \text{expi}, \quad \text{la} = \text{la} + (R - P_i) / 1 - \text{expi}$$

$$\text{If } \text{expi} \geq 1/\beta \text{ then } P_i = P_i + \beta (R - P_i), \varepsilon_i = \varepsilon_i + \beta (|R - P_i| - \varepsilon_i), \quad \text{La} = \text{La} + (R - P_i) / 1 - \text{expi}$$

بروزآوری برازندگی

$$\text{If } \varepsilon_i < \varepsilon_0 \text{ then } k_i = 1$$

$$\text{If } \varepsilon_i \geq \varepsilon_0 \text{ then } k_i = \beta (\varepsilon_i / \varepsilon_0) - \gamma + \beta (\text{Lai} - \text{fi})$$

$$\text{Fi} = \text{fi} + \beta [\text{Lai} / \sum k_j] - \text{fi}$$

که در این روابط، β نرخ یادگیری، γ توان دقت قانون، ε خطای پیش بینی، \exp تجربه قانون، P پیش بینی قانون، R پاداش دریافتی از محیط، k دقت قانون، La مقدار فیلد اتوماتای اضافه شده و f برازندگی آن است. اندیس i نیز شماره قانون را در مجموعه قوانین نشان می‌دهد. پس از آن از $action\ set$ های تمامی مثال های تربیتی به صورت رندوم و با استفاده از چرخ رولت والدین را انتخاب کرده و عملیات ژنتیک را انجام می‌دهیم. سپس قوانینی که در $match\ set$ قرار گرفتند و دارای $fitness$ بسیار پایین می باشند، مقدار $action$ فیلدی که دارای بزرگترین $fitness$ را در آن $match\ set$ می باشد را به جای مقدار $action$ خود انتخاب می کنند و مقدار فیلد La خود را بروز رسانی می کنند. بدین ترتیب امکان کم شدن $min\ action\ size$ وجود دارد و در صورت بروز این عمل، قانون جدیدی که تولید می شود با یکی از قوانینی که دارای کوچکترین مقدار $fitness$ در کل مجموعه قوانین بود جایگزین می شود تا بدین ترتیب قوانین قوی تر تولید شده جایگزین قوانین ضعیفی گردیده که امکان عدم استفاده از آن ها تا انتهای فرآیند تربیت وجود دارد. نحوه انتخاب قوانین جدید بدین صورت می باشد که با استفاده از روش «انتخاب تصادفی با باقی مانده» از میان فیلد هایی که بخش شرط داده ها را به وجود می آورند در یک جدول به نام $mating\ pool$ ذخیره می گردند نحوه انتخاب زوج های گوناگونی انتخاب شده با استفاده از فیلد اتوماتای یادگیر اضافه شده و روش همگذاری انجام می شود. بدین ترتیب که رشته هایی که نمایشگر بخش شرط داده می باشند به ترتیب مقدار فیلد La موجود در آنها مرتب گردیده سپس جفت رشته ابتدایی با استفاده از روش همگذاری میانی یک جفت فرزند جدید (داده جدید) جایگزین میکنند که بخش عملکرد داده های جدید نیز با استفاده از داده های موجود تولید می شود این عمل تا آنجا ادامه می یابد که شرط توقف یادگیری (در این مثال اجرای بیست باره مثال های تربیتی بر روی قوانین ایجاد شده) به کمک داده های کامل شده ارضاء گردد [۱۵].

۷. روش انجام کار

مجموعه داده هایی که در مورد تشخیص هویت افراد جمع آوری شده است شامل ویژگی های متفاوتی می باشد که دسته ای از آنها فقط بر اساس داده های آزمایشگاهی تشکیل شده است. یکی از معتبرترین پایگاههای داده در زمینه ECG بانک اطلاعاتی MIT-BIT [۱۶] می باشد. این پایگاه داده شامل ۴۸ ثبت ECG دو کاناله می باشد که با زمان یک ساعت و نرخ نمونه برداری ۱۲۸ هرتز برای هر فرد استخراج شده است. هرکدام از فایل ها را در قطعات زمانی ۱۰ ثانیه تقسیم بندی نموده و با استفاده از تبدیل موجک آن را تا سطح ۱۰ تجزیه نموده و برای هر سطح ویژگی های آنروپی، توان، انرژی، واریانس، انحراف معیار و میانگین استخراج شد. بدین ترتیب ۶۰ ویژگی از آن ها استخراج می شود. برای بار نخست تمامی این ۶۰ ویژگی را به روش های طبقه بندی اشاره شده در این مقاله وارد کرده و زمان و درصد پاسخگویی را با ویژگی های انتخاب شده توسط الگوریتم انتخاب ویژگی IDE که دارای اطلاعات مهم و بیشتری می باشند مقایسه می نماییم. به این منظور برای انتخاب ویژگی حد آستانه برای الگوریتم IDE برابر با ۰/۸ در نظر گرفته شد که از بین ۶۰ ویژگی ارائه شده تنها ۷ ویژگی از حد آستانه مقدار بیشتری را دارا بودند که عبارت بودند از:

۱- انحراف معیار سیگنال (AAD^۳)

۲- انرژی سیگنال (AAAAADAY^۷)

۳- واریانس سیگنال (ADAADA^۶)

۴- توان سیگنال (AADA^۴)

۵- انحراف معیار سیگنال (ADDA^۴)

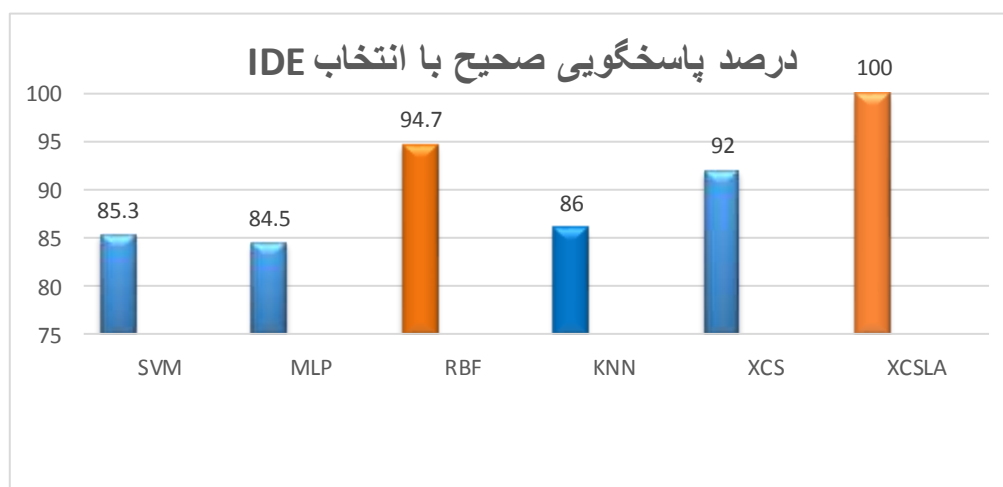
۶- توان سیگنال (ADDDDAAY^۷)

۷- انحراف معیار سیگنال (AAAA^۴)

سپس این ۷ ویژگی نیز به انواع روش های طبقه بندی برای کسب درصد پاسخگویی و هزینه محاسباتی وارد شد. برای اینکه نتایج آزمایشات در زمان اجرای محاسبات در شرایط یکسانی قرار بگیرند در تمامی تست ها از رایانه ای با مشخصات CPU=۲،۴ و ۴GB RAM استفاده شده است. همان طور که در جدول ۱ مشخص است هزینه محاسباتی روش KNN از تمامی روش های فوق در هر دو حالت انتخاب ویژگی کمتر می باشد در صورتی که از لحاظ درصد درستی روش XCSLA در هر دو حالت انتخاب ویژگی بهترین روش بوده و حتی در انتخاب ویژگی به کمک الگوریتم IDE تمامی طبقه بندی ها را به درستی تشخیص داده است.

جدول ۱- مقایسه نتایج سیستم های طبقه بند جهت تشخیص عیوب

روش انجام کار	هزینه محاسباتی انتخاب اولیه بر حسب دقیقه	هزینه محاسباتی بر IDE انتخاب حسب دقیقه	درصد پاسخ گویی صحیح انتخاب اولیه	درصد پاسخ گویی صحیح انتخاب IDE
SVM	۱۱	۵	۸۲٪	۸۵/۳٪
MLP	۱۳	۴	۸۵٪	۸۴/۵٪
RBF	۱۲	۵	۹۲/۵٪	۹۴/۷٪
KNN	۱۰	۲	۷۷/۵٪	۸۶٪
XCS	۱۶	۶	۸۵٪	۹۲٪
XCSLA	۱۷	۶	۹۱/۵٪	۱۰۰



شکل ۵- درصد پاسخگویی صحیح با انتخاب IDE

۸. نتیجه گیری

در این مقاله، استفاده از سیستم های طبقه بندی گسترش یافته (XCSLA) به منظور تشخیص هویت افراد از روی سیگنال قلبی بر اساس نشانگان بالینی پیشنهاد گردید. بدین منظور داده های ۱۸ فرد نرمال از یکی از معتبرترین پایگاههای داده در زمینه ECG بانک اطلاعاتی استخراج شده، سپس به کمک الگوریتم IDE ۷ ویژگی از ضرایب تبدیل موجک را به انواع روش های طبقه بندی اعمال می کنیم. نتایج این بررسی نشان داد که روش پیشنهادی در مقایسه با دیگر روش های متداول داده کاوی از دقت و نرخ همگرایی بیشتری برخوردار است.

منابع و مراجع

۱. L. Biel, O. Petterson, L. Philipson, and P. Wide. ECG analysis: a new approach in human identification. In Proceedings of the ۱۶th IEEE Instrumentation and Measurement Technology Conference, volume ۱, pages ۵۵۷ – ۵۶۱, Venice, Italy, ۱۹۹۹
۲. R. Palaniappan and S. Krishnan. Identifying individuals using ECG beats. In International Conference on Signal Processing and Communications, pages ۵۶۹ – ۵۷۲, Bangalore, India, ۲۰۰۴
۳. Z. Zhang and D. Wei. A new ECG identification method using Bayes' theorem. In IEEE TENCON, pages ۱ – ۴, Hong Kong, China, ۲۰۰۶.
۴. M. Kyoso and A. Uchiyama. Development of an ECG identification system. In Proceedings of the ۲۳rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume ۴, pages ۳۷۲۱ – ۳۷۲۳, Istanbul, Turkey, ۲۰۰۱.
۵. Y. Singh and P. Gupta. ECG to individual identification. In ۲nd IEEE International Conference on Biometrics: Theory, Applications and Systems, pages ۱ – ۸, Washington, DC, ۲۰۰۸
۶. A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee. Person identification using electrocardiograms. In Canadian Conference on Electrical and Computer Engineering, pages ۱ – ۴, Ottawa, Canada, ۲۰۰۶
۷. S. Saechia, J. Koseeyaporn, and P. Wardkein. Human identification system based ECG signal. In IEEE TENCON, pages ۱ – ۴, Hong Kong, China, ۲۰۰۵.
۸. K.-S. Kim, T.-H. Yoon, J.W. Lee, D.-J. Kim, and H.-S. Koo. A robust human identification by normalized time-domain features of electrocardiogram. In ۲۷th Annual International Conference of the Engineering in Medicine and Biology Society, pages ۱۱۴ – ۱۱۷, Shanghai, China, ۲۰۰۵.
۹. J. Yao and Y. Wan. A wavelet method for biometric identification using wearable ECG sensors. In ۵th International Summer School and Symposium on Medical Devices and Biosensors, pages ۲۹۷ – ۳۰۰, Hong Kong, China, ۲۰۰۸
۱۰. K. Plataniotis, D. Hatzinakos, and J. Lee. ECG biometric recognition without fiducial detection. In Biometrics Consortium Conference, pages ۱ – ۶, Baltimore, MD, ۲۰۰۶.
۱۱. J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz. Activity-aware ECG-based patient authentication for remote health monitoring. In Proceedings of the ۲۰۰۹ International Conference on Multimodal Interfaces, pages ۲۹۷ – ۳۰۴, Cambridge, MA, ۲۰۰۹.
۱۲. D. Zhang, Automated Biometrics—Technologies and Systems, Kluwer Academic, Dordrecht, ۲۰۰۰.
۱۳. A.K. Jain, A. Ross, S. Prabhakar, “An Introduction to Biometric Recognition,” IEEE Trans. on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, January, ۲۰۰۴.
۱۴. M. Shariat Panahi, N. Moshtaghi Yazdani, An Improved XCSR Classifier System for Data Mining with Limited Training Samples, Global Journal of Science, Engineering and Technology, ISSN : ۲۳۲۲- ۲۴۴۱, ۲۰۱۲, Issue ۲, pp. ۵۲-۵۷.

۱۵. نوید مشتاقی، آرزو یزدانی، بهبود الگوریتم XCS با استفاده اتوماتای یادگیرنده، کنفرانس بین المللی کامپیوتر، فناوری اطلاعات و رسانه های دیجیتال، آذر ماه ۱۳۹۲.

۱۶. MIT-BIH rhythmic database directory. Available from :<http://www.physionet.org/cgi-bin/atm/ATM>