

پیشنهاد راهکار جدیدی جهت تشخیص بیماری آلزایمر با استفاده از درخت تصمیم

سامیه خسروی

دانشکده برق و کامپیوتر دانشگاه بیرجند، skhosravi@birjand.ac.ir

چکیده - امروزه بیماری آلزایمر بسیار شیوع پیدا کرده است، و تشخیص سریع آن یکی از موارد بسیار مهمی می باشد که می تواند در پیشگیری و درمان افراد بسیار موثر واقع شود. از اینرو در این مقاله روشی برای پیش بینی بیماری آلزایمر پیشنهاد شده است. در اینجا با استفاده از الگوریتم ژنتیک و درخت تصمیم گیری سعی شده است تا راه حلی بیان شود که با استفاده از پارامترهایی که از طریق داده کاوی بدست آمده است بیماری آلزایمر در افراد پیش بینی شود. سیستم ارائه شده در اینجا قادر است تا تصمیماتی را اتخاذ کند که این تصمیمات می تواند در جهت رسیدن به سیستمی توصیه گر خودمختار بسیار سودمند باشد. روش پیشنهادی در اینجا یک روش خودآموز می باشد، که قادر است تا با استفاده از الگوریتم ژنتیک در ابتدا یک طبقه بندی را انجام دهد که باعث افزایش سرعت الگوریتم نیز می شود و در ادامه آن ما از درخت تصمیم گیری خودمان استفاده نمودیم، که قادر است با دقت بالایی تصمیمات را اتخاذ نماید. کلید واژه- درخت تصمیم گیری، داده کاوی، تشخیص آلزایمر.

تنها راه قطعی تشخیص بیماری آلزایمر در یک فرد، معاینه

مغز وی بعد از مرگ اوست. اما می توان آزمایش های دیگری را

برای اندازه گیری قوه ذهنی شناختی فرد، تشخیص زوال عقل و رد کردن بیماری های دیگر انجام داد. می توان با گرفتن تاریخچه پزشکی از فرد و همچنین سوال هایی مانند:

- نشانه های فرد

- تاریخچه پزشکی خانواده فرد

- بیماری هایی که اکنون دارد یا قبلا داشته است

- داروهایی که اکنون مصرف می کند یا قبلا مصرف کرده

است

- رژیم غذایی، مشروبات الکلی یا عادات دیگر زندگی فرد [4].

در این بیماری، ابتدا توانایی های شناختی به تدریج کاهش می یابد که مشکل برگشتناپذیری است و بیشتر بخش های مربوط به حافظه را درگیر می کند. متأسفانه از زمان شروع این بیماری تا هنگام بروز نشانه در فرد، فاصله زمانی زیادی وجود دارد؛ بنابراین پزشک زمانی از بیماری اطلاع پیدا می کند که در مراحل پیشرفته قرار دارد. بدین دلیل، پیش بینی و تشخیص زودهنگام این بیماری در بهبود عملکرد فرد موثر خواهد بود [4][5].

یکی از عوامل مهمی که معمولا در ایجاد بیماری آلزایمر دخیل و موثر می باشد، عامل ژنتیک است. اگر فردی در خانواده مبتلا به این بیماری باشد، احتمال ابتلای فردی دیگر در خانواده

۱- مقدمه

تشخیص بیماری آلزایمر در اوایل شکل گیری آن، بسیار با اهمیت است. در حال حاضر از نشانه های زیستی مختلفی مانند: مقدار نرون های ازدست رفته در نواحی مختلف مغز و اتصالات تخریب شده بین آن ها که از تصاویر ساختاری MRI مغز قابل استخراج است، برای تشخیص اختلالات به وجود آمده در مغز به دلیل ابتلا به بیماری های شناختی مانند آلزایمر و اختلالات خفیف شناختی استفاده می شود. مسئله مهم در علم پزشکی، چگونگی استفاده از این نشانه ها جهت تشخیص هرچه دقیق تر بیماری های شناختی، مانند آلزایمر و اختلالات خفیف شناختی است [1].

بیماری آلزایمر نوعی زوال عقل پیشرفته است. زوال عقل اصطلاح وسیع تری برای شرایطی است که توسط آسیب ها و بیماری های مغزی به وجود می آیند، که تأثیر منفی روی حافظه و تفکر و رفتار دارند. این تغییرات زندگی روزمره را مختل می کنند [2].

طبق گفته انجمن آلزایمر، این بیماری ۶۰ تا ۸۰ درصد نمونه های زوال عقل را شامل می شود. در اکثر موارد، بیماری بعد از ۶۵ سالگی تشخیص داده می شود. اگر این بیماری زودتر از ۶۵ سالگی تشخیص داده شود، معمولا نشان دهنده بیماری آلزایمر زودرس است [3].

تشخیص دهد، و به بیماران اقدامات درمانی پیشنهاد دهد [9,10].

در پژوهشی تحت عنوان "تشخیص بیماری قلبی با استفاده از مکانیزم حل فازی" که در سال ۲۰۱۳ توسط دکتر ای وی سنتیل کومار انجام شد، از یک مکانیزم پیشرفته فازی برای تشخیص بیماری قلبی با پنج لایه که هر لایه گره‌های مربوط به خود را داشت، استفاده شد. در این پژوهش برای پیاده‌سازی مکانیزم فازی از نرم افزار متلب استفاده گردید. این روش در مقایسه با روش قلبی کارایی بهتری در تشخیص بیماری قلبی داشت [9,11].

در پژوهشی که تحت عنوان "پیش بینی سطح خطر بیماری قلبی با رویکرد مبتنی بر قوانین فازی" که در سال ۲۰۱۴ توسط کانتش کومار آ ای دی، زوژی و پنیال خان بات انجام شد، سیستم خبره‌ای مبتنی بر منطق فازی طراحی گردید و با استفاده از تکنیک داده‌کاوی جمعی از ویژگی‌ها برای طراحی کم شد. این سیستم می‌توانست بیماری قلبی را با توجه به مجموعه داده‌های ماشین یادگیری تشخیص دهد. این سیستم از روش دستی دقیق‌تر بود و ۲ مرحله اصلی داشت. مرحله اول شامل طبقه‌بندی و تشخیص بود و مرحله دوم میزان خطر بیماری تنفسی را تعیین می‌کرد [9,12].

در پژوهشی تحت عنوان "مجازی‌سازی میزان احتمال وجود خطر سرطان پستان با استفاده از منطق فازی در محیط نرم افزاری متلب" که در سال ۲۰۱۳ توسط کی سی لاتا و همکاران انجام شد، سیستم پشتیبان تصمیم‌گیرنده فازی برای کمک در طبقه‌بندی میزان احتمال وجود خطر سرطان پستان طراحی شد. متغیر ورودی این سیستم سن و حالت سرطان بود و متغیر خروجی تعیین میزان احتمال وجود خطر سرطان پستان بود که به سه وضعیت "جدی نیست"، "جدی" و "خیلی جدی" تقسیم بندی می‌شد. این سیستم میزان احتمال وجود خطر سرطان پستان را در فرد ارزیابی می‌کرد. نویسنده در این پژوهش معتقد است هنوز با توسعه تکنولوژی روش و معیاری وجود ندارد که به صورت دقیق میزان احتمال خطر را تخمین بزند [9,13].

در پژوهشی تحت عنوان "طراحی و پیاده‌سازی یک سیستم خبره فازی برای تشخیص و برآورد سطح آسم و بیماری انسداد مزمن ریه" که در سال ۲۰۱۳ توسط اس کریسنا آنند انجام شد، سیستم خبره فازی طراحی گردید که گزارش جزئیات مختلف بیمار را می‌گرفت و با توجه به مشکلات تشخیصی بیمار احتمال مواجهه با بیماری را مشخص می‌کرد. برای کاهش پیچیدگی کل سیستم، زیر سیستم‌هایی با کنترل کننده مستقل هوشمند نیز

وجود دارد. از آنجا که رشد مغزی، تا حدودی به ژنتیک فرد بستگی دارد و این بیماری مربوط به کوچک‌تر شدن حجم مغز به مرور زمان می‌شود، به نظر عامل ژنتیک در پیش‌بینی این بیماری دقیق‌تر از عوامل دیگر عمل می‌کند [6][11].

آلزایمر ششمین علت مرگ در جهان است. هر چه علم پزشکی بیشتر پیشرفت می‌کند، بیشتر این مسأله مطرح می‌شود که یک زمینه ژنتیکی از یک بیماری، احتمال وقوع آن را در سایر اعضای خانواده بیشتر می‌کند، بیماری آلزایمر بیش از اینکه یک بیماری با دلایل محیطی باشد، یک بیماری ژنتیکی است.

با توجه به اهمیت عوامل ژنتیکی در بروز بیماری آلزایمر، این پژوهش به تحلیل مجموعه داده ژن‌هایی که منجر به بیماری آلزایمر می‌شود و با استفاده از الگوریتم‌ها و تکنیک‌های داده‌کاوی می‌توان الگوها و نتایجی را برای محققین فراهم کند، که عوامل ابتلا به بیماری‌های مختلف را نشان دهد و براساس آن نتایج، پزشکان در پیشگیری آن بیماری‌ها اقدام کنند. همچنین در راستای این تحقیق از درخت تصمیم استفاده شده است که موجب می‌شود تا سیستم بتواند پیشنهادهایی را نیز ارائه دهد و براساس درختی که در نتیجه کاویدن داده‌ها حاصل می‌شود می‌توان بیماری را در مراحل ابتدایی شناسایی نمود [7].

۲- مروری بر تاریخچه این پژوهش

در پژوهشی تحت عنوان "یک سیستم خبره فازی برای تشخیص بیماری قلبی" که توسط علی عادل و مهدی نشاط در سال ۲۰۱۰ انجام شد، سیستمی طراحی گردید که بیماری قلبی را تشخیص می‌داد. این سیستم 11 متغیر ورودی و یک متغیر خروجی داشت. متغیرهای ورودی شامل نوع درد قفسه سینه، میزان کلاسترول خون، فشار خون در زمان استراحت، میزان قند خون، نوار قلب فرد در زمان استراحت، حداکثر ضربان قلب، ورزش، آلد پیک، اسکن تالیوم، جنس و سن بیمار بودند و متغیر خروجی تشخیص بیماری قلبی در فرد بود. این سیستم می‌توانست ۹۴٪ بیماری قلبی را تشخیص دهد [8,9]. در پژوهشی تحت عنوان "تشخیص بیماری‌های قلبی به کمک منطق فازی" توسط سنجیو کومار و کورسیمارنجیت کور که در سال ۲۰۱۳ انجام شد، سیستم خبره فازی برای تشخیص بیماری قلبی طراحی گردید. این سیستم 4 زمینه ورودی و ۲ خروجی داشت. ۶ زمینه ورودی شامل نوع درد قفسه سینه، میزان کلاسترول خون، حداکثر فشار خون، میزان قند خون والد پیک بودند، و دو متغیر خروجی وجود بیماری قلبی در بیمار و اقدامات احتیاطی بر اساس تشخیص بودند. این سیستم قادر بود ۹۲٪ بیماری قلبی را

قسمتی از آن از میان افراد بیمار آلزایمر و یا خویشاوندان این افراد (زیرا ممکن است خود آن‌ها نتوانند بخاطر بیماری خود این فرم‌ها را پر کنند) تهیه می‌شود، و قسمتی دیگر از میان افراد سالم بدست می‌آید. این نسبت بصورت علمی ثابت شده است، که بهتر است ۷۰ درصد جامعه ورودی ما افراد بیمار و ۳۰ درصد افراد سالم باشند و این نسبت‌ها جز بهترین نسبت‌های بررسی می‌باشند.

روش پیشنهادی دارای مراحل مختلفی می‌باشد که به ترتیب عبارتند از:

پیش پردازش داده‌ها

انتقال داده‌ها

آموزش سیستم با استفاده از الگوریتم ژنتیک [16]

ساخت درخت تصمیم‌گیری

تبدیل درخت تصمیم‌گیری

استخراج تصمیمات

در ابتدا مجموعه داده بیماران آلزایمر جمع آوری شده و به آماده‌سازی و پیش‌پردازش داده‌ها پرداخته می‌شود. در آماده‌سازی و پیش‌پردازش داده‌ها، از روش‌های مختلفی استفاده می‌شود. اول این که برخی ویژگی‌ها دارای مقادیر منحصر به فرد هستند. این ویژگی‌ها نمی‌توانند دانش مفیدی را در مجموعه داده ایجاد کنند. لذا این مجموعه ویژگی‌ها باید از داده‌ها حذف شوند. به طور نمونه می‌توان به ویژگی نام و نام خانوادگی اشاره نمود. همچنین ممکن است برخی تراکنش‌ها دارای مقادیر مفقود فراوان باشند. لذا این تراکنش‌ها نیز باید از مجموعه داده‌ها حذف شوند. از طرفی ممکن است، مقادیر برخی ویژگی‌ها دارای مقادیر نویز و مفقود باشند، لذا این مقادیر نیز باید در مجموعه داده اصلاح شوند. مرحله بعدی نوبت به استفاده از ابزار کشف آنومالی است. داده‌هایی که در نقاط خارج از قانون مجموعه داده قرار دارند شناسایی شده و حذف می‌شوند. برای اینکه بتوان روی داده‌ها به عنوان ورودی کار کرد، می‌بایست ویژگی‌هایی را از آن‌ها استخراج نمود. به طور معمول پیش از انتخاب و استخراج ویژگی‌ها، برخی عملیات پیش پردازش بر روی داده‌ها انجام می‌شود.

۳-۱- انتقال داده‌ها

در این قسمت داده‌ها در دامنه‌های درست قرار می‌گیرند. بدین معنا که داده‌ها باید به رنج‌هایی که در سیستم مشخص شده است منتقل شوند و داده‌های خارج از رنج، داده‌های مشکل دار بوده و می‌بایست حذف شوند. داده‌ها می‌بایست در رنج درست قرار بگیرند، که این قسمت بصورت اتوماتیک از روی پرسشنامه تکمیل می‌شود. در نحوه انتقال دیده می‌شود. پارامترهایی که

طراحی شد. علاوه بر آن، تجزیه و تحلیل حساسیت نیز بر روی تست میزان ارتباط ورودی‌های خاصی نیز انجام گردید [9,14]. در پژوهشی که تحت عنوان "تجزیه و تحلیل طبقه بندی شده فشار خون با استفاده از سیستم کنترل منطق فازی در داده کاوی" در سال ۲۰۱۴ توسط میلی ویگانام و کی راجسوری انجام شد، سیستم خبره فازی معرفی گردید که به کاربران اطلاعات دقیقی در مورد فشار خون شان می‌داد. سیستم خبره فازی طراحی شده، بیماری فشار خون را در دو طبقه صفر و یک طبقه بندی می‌کرد [9,15].

۳- روش پیشنهادی تشخیص آلزایمر

در اینجا، قصد رسیدن به روشی می‌باشد که به وسیله آن بتوان تا درصد بالایی، بیماری آلزایمر را در بین افراد تشخیص داد؛ زیرا گفته شد که تشخیص بیماری آلزایمر بسیار دشوار می‌باشد. با استفاده از روشی که در اینجا بیان می‌شود قصد داریم تا درختی از تصمیم‌گیری را ارائه کنیم و با استفاده از آن اهمیت هر یک از المان‌ها را نیز در تشخیص بهتر بیماری آلزایمر بیابیم. در اینجا هدف اصلی ارائه درخت تصمیم‌گیری می‌باشد که به وسیله آن بتوان تصمیمات مهمی را در زمینه تشخیص بیماری آلزایمر و همچنین جلوگیری از آن ارائه نمود. ما برای اینکه بتوانیم به دقت بالاتری در روش پیشنهادی دست یابیم، می‌بایست ابتدا با استفاده از داده‌کاوی سیستم را آموزش دهیم. در اینجا ما برای اینکه بتوانیم با داده‌های کمی تصمیم‌گیری‌های موثری را انجام دهیم، می‌بایست سیستم را سریع و با داده‌های کم آموزش دهیم، که از اینرو در این روش پیشنهادی ما از الگوریتم ژنتیک استفاده کردیم [16]. در اینجا الگوریتم پیشنهادی با استفاده از الگوریتم ژنتیک ابتدا مورد آموزش قرار می‌گیرد، و سپس با توجه به قدرت استنتاج معانی از داده‌های پیچیده یا مبهم، سیستم با قدرت بالایی آموزش می‌بیند و قادر است تا با داده‌های پیچیده به راحتی دست و پنجه نرم کند؛ و بعد از آموزش مناسب، سیستم برای تصمیم‌گیری آماده می‌شود. بعد از آموزش ما از درخت تصمیم‌گیری پیشنهادی خود که بسیار بهینه و هوشمند می‌باشد استفاده می‌کنیم، تا بتوانیم با استفاده از روابط منطقی تصمیمات خروجی خود را بیان کنیم. این تصمیمات با توجه به قواعد ریاضی استخراج می‌شود و دارای دقت بالایی می‌باشند. و با استفاده از این سیستم پیشنهادی می‌توان حتی آن‌را هوشمند نیز نمود تا بتواند در شرایط خاص با توجه به تصمیمات استخراج شده خودکار، توصیه‌های خودکاری را نیز بسته به شرایط انجام دهد. ورودی‌های سیستم ما پرسشنامه‌هایی خواهد بود که

سایت‌های مختلفی است که در همین جا گفته شد تا بتوان به مواردی رسید که بتوان با آن بیماری آلزایمر را تشخیص داد. واقع این پارامترها بسیار در بیماری آلزایمر تاثیر گذار می‌باشند و با توجه حساسیت آن‌ها این پارامترها انتخاب شده است.

در این جدول آورده شده است پارامترهای استاندارد می‌باشند که برای رسیدن به این پارامترها از چندین سایت پزشکی، کارشناس‌های مختلف در این زمینه، و همچنین سایت ADNI [۲] که بسیار در زمینه بیماری‌های آلزایمر معروف می‌باشد، کمک گرفته شده است. و جدول زیر ادغامی از مشاوره و مطالعه

جدول ۱: متغیرهای ورودی سیستم

پارامترها	علامت اختصاری ما	توضیحات	دامنه مقدر
سن	Age	اینکه سن فرد چقدر است زیرا این یکی از پارامترهای مهم در شناخت بیماری آلزایمر می‌باشد	{بسیار کمسال (۹۱-۱۳)، کمسال (۷۱-۹)، خیلی پیر (۵۵-۷۰)، پیر (۲۴-۵۳)، متوسط (۲۵-۳۵)، جوان (۱۸-۲۴)، نوجوان (۱۱-۱۷)، کودک (۵-۱)، خردسال (۲-۳)، نوزاد (۱-۰)}
داشتن بیماری قبلی	HPI ^۱	فرد دارای چه تعداد بیماری قبلی است. این پارامتر نیز خود می‌تواند روی دچار بیماری آلزایمر بدون تاثیر گذار باشد	{خوبی بیماری (بیش از ۳ بیماری)، کمی بیمارترین ۱ تا ۳ بیماری، سالم (نداشتن هیچ بیماری)}
بهره هوشی	IQ ^۲	بهره هوشی فرد چه میزان است (با استفاده از تست ریون)	{ناپقه (۱۷۰-۱۸۹)، تیز هوش (۱۵۰-۱۶۹)، پر هوش (۱۴۹-۱۳۰)، پاهوش (۱۱۰-۱۲۹)، متوسط (۹۰-۱۰۹)، مرزی (۸۰-۹۰)، مرزی ضعیف (۷۰-۷۹)، کمترین (۵۰-۵۹)، کالیپو (۲۵-۳۵)، کانا (۰-۴۰)}
توانایی صحبت کردن	ATS ^۳	یکی از عواملی که می‌تواند روی تشخیص بیماری آلزایمر تاثیر داشته باشد میزان توانایی در صحبت کردن است.	{خوبی راحت (هیچ مشکلی در صحبت کردن وجود ندارد)، کمی ضعیف (کمی مشکل در صحبت کردن و فراموش کردن بعضی لغات)، خیلی ضعیف (مشکل بسیار در صحبت کردن وجود دارد و فرد نمی‌تواند لغات را بدرستی برای صحبت کردن یاد بیاورد)}
داشتن دستیار مطالعه	HSA ^۴	یک دستیار مطالعه کسی است که به طور متوسط حداقل ۱۰ ساعت در هفته به فرد کمک می‌کند و در دسترس فرد قرار دارد	{خیر، بله}
محیط زندگی	LE ^۵	محیط زندگی در بروز بیماری آلزایمر تاثیر بسیاری دارد.	{بسیار شلوغ، شلوغ، متوسط، کم سر و صدا، بدون هیچ سر و صدایی (تنها زندگی کردن)}
داشتن اختلالات	HD ^۶	آیا فرد در چند وقت گذشته دچار اختلالات افسردگی و دو قطبی شده است؟	{خیر، بله}
استفاده از داروهای تحقیقاتی	URD ^۷	در طی ماه گذشته، فرد نوعی از داروهای تحقیقاتی را به عنوان بخشی از یک آزمایش بالینی به کار برده است	{ز یاد، متوسط، کم، اصلا}
استعمال مواد غیرمجاز	UUM ^۸	در طول دو سال گذشته چه میزان مشروبات الکلی یا مواد مخدر مصرف شده است	{ز یاد، متوسط، کم، اصلا}
داشتن قطعات قلبی در بدن	HMPB ^۹	آیا فرد یک دستگاه ضربان ساز یک دستگاه پزشکی ایمپلنت، قطعات قلبی یا اجسام خارجی در بدن دارد؟ (توجه شود که اگر جسم خارجی زانو یا جایگزینی مفصل ران مصنوعی باشد، جواب خیر است)	{خیر، بله}
تعداد مراجعه به پزشک	NVD ^{۱۰}	در ۵ سال گذشته چندبار بررسی به پزشک مراجعه شده است	{ز یاد، متوسط، کم، اصلا}
تعداد بازی‌های فکری انجام شده	NGP ^{۱۱}	در ۵ سال گذشته چه میزان بازی‌های فکری انجام شده است.	{ز یاد، متوسط، کم، اصلا}
گروه خونی	BT ^{۱۲}		{O-, O+, A-, A+, B-, B+, AB-, AB+}
بررسی اسکن‌های MRI و PET	Scan	آیا تا به حال اسکن‌های MRI و PET از فرد مورد بررسی پزشک قرار گرفته است	{ز یاد، متوسط، کم، اصلا}

یک انتخاب می‌باشد. بدین معنی برای رفتن از گره ریشه به گره پایین‌تر، می‌توان از شاخه‌هایی که به آن گره متصل هستند، یکی انتخاب شود. در انتها هر یک از گره‌های انتهایی یا اصطلاحاً گره برگ تصمیمی را بازگو می‌کند. هر کدام از شاخه‌ها تا رسیدن به برگ دارای سناریویی می‌باشد، که موجب تصمیمی می‌شود. در این روش پیشنهادی، از درخت تصمیم ID3 بهبود یافته استفاده شده است، که بهبود آن موجب سرعت عمل بالای آن شده است. درخت ID3 یک درخت تصمیم‌گیری می‌باشد که دارای یادگیری نیز است و اولین بار توسط راس کوینلن مطرح شد. ایده الگوریتم ID3، ساخت درخت تصمیم‌گیری بالا به پایین می‌باشد که انتخاب گره در آن به وسیله جستجوی حریصانه از میان مجموعه‌ای از صفت‌ها می‌باشد. در اینجا ما برای اینکه قادر باشیم تا مفیدترین صفت را از میان صفات بیابیم، که در کلاسه بندی مفیدتر باشد، از الگویی بخصوص استفاده نمودیم. برای

۳-۲- آموزش سیستم با استفاده از الگوریتم ژنتیک

الگوریتم ژنتیک، با قابلیت قابل توجه در استنتاج معانی از داده‌های پیچیده یا مبهم می‌تواند برای استخراج الگوها و شناسایی روش‌هایی که آگاهی از آنها برای انسان و دیگر تکنیک‌های کامپیوتری بسیار پیچیده و دشوار است، به کار گرفته شود [16]. الگوریتم ژنتیک به عنوان یکی از ابزارهای داده‌کاوی می‌تواند برای طبقه‌بندی بیماران آلزایمر مورد استفاده قرار گیرد. به طور کلی در این قسمت سیستم مورد آزمایش با استفاده از این الگوریتم ژنتیک مورد آموزش قرار می‌گیرد [16].

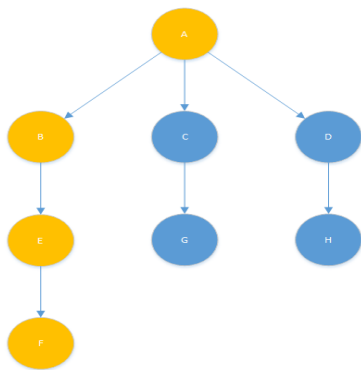
۳-۳- ساخت درخت تصمیم‌گیری

در این قسمت از درخت تصمیم‌گیری استفاده می‌شود. درخت تصمیم‌گیری، درختی است که هر شاخه از آن به عنوان

اینجای مسیر بکاهیم (یعنی باید $G(S,F)=E(S)-$ $(E(A)+E(B)+E(E)+E(F))$ را بدست آورد). البته باید توجه شود که ما می‌بایست در مجموعه A صفاتی که تا به اینجای کار استفاده شده است، به علاوه صفتی که می‌خواهیم قرار دهیم را محاسبه نماییم. بعد از اینکار از بین این مجموعه صفات باقیمانده که برای هر کدام آن‌تروپی را محاسبه نمودیم، صفتی را که دارای G بیشتری است را انتخاب نماییم. در هنگام انتخاب اگر دو صفت دارای G برابر بودند، که احتمال این پیشامد نیز کم نیست، می‌بایست به گره مربوطه، دو یا هر تعداد صفت که دارای بیشترین مقدار G هستند و باهم برابر نیز می‌باشند بیفزاییم. یعنی اگر برای مثال، در گره ای دو صفت دارای G برابر بودند، آنگاه این دو به گره مربوطه دو فرزند می‌افزاییم، و هر کدام از این صفت‌ها به عنوان یک فرزند این گره در نظر گرفته می‌شوند و بعد از این کار روند الگوریتم را برای هر یک از این گره‌ها ادامه می‌دهیم. برای مثال، در 0 (۱) می‌توان مشاهده نمود که مقدار G برای صفت‌های B ، C و D برابر است از این رو همه این صفت‌ها در یک سطح قرار گرفته‌اند.

(۲)

$$G(S,A) = Entropy(S) - \sum_{v \in values(A)} Entropy(v)$$



شکل ۱: مثالی برای انتخاب صفت‌ها با استفاده از درخت تصمیم

این کار باعث می‌شود تا صفت‌هایی که دارای آن‌تروپی بیشتری هستند را بیابیم چرا که این صفت‌ها تأثیر بیشتری را در تصمیم نهایی ما می‌گذارند. این روند جلو رفتن در درخت تصمیم‌گیری تا جایی ادامه می‌یابد که در هر مسیر دیگر صفتی باقی نمانده باشد. در این حالت درخت تصمیم‌گیری کاملاً ساخته شده است و به پایان رسیده است.

برای اینکه بتوان شرط‌هایی را از این درخت استخراج نمود، که بصورت $If...then....$ شود، که در آن بتوان از قوانین

اینکه بتوان کلاسه بندی مفیدی را برای مجموعه یادگیری انجام داد، می‌بایست تعداد سوالات را کاهش داد یا می‌توان گفت می‌بایست عمق درخت تصمیم‌گیری را کاهش داد. از اینرو در این قسمت نیاز به تابعی است که قادر باشد تا متعادل‌ترین تقسیم را انجام دهد، که در این صورت عمق درخت بسیار کاهش می‌یابد و گره‌ها بصورت متعادل، در درخت تقسیم می‌شوند [16].

جدولی را در نظر بگیرید که دارای صفات و کلاسی از صفات نیز می‌باشد. در صورتی به این جدول همگن گفته می‌شود که تنها شامل یک کلاس باشد. توابع زیادی همچون آن‌تروپی، $gini$ $index$ و $classification error$ برای سنجش میزان همگن‌پذیری وجود دارند. در این میان در اینجا از آن‌تروپی استفاده شده است.

(۱)

$$Entropy = \sum_j -p_j \log_2 p_j$$

آن‌تروپی یک جدول صفر است، زیرا احتمال آن مقداری برابر یک است (تنها دارای یک کلاس باشد). آن‌تروپی زمانی به بیشترین مقدار خود می‌رسد، که تمامی کلاس‌های موجود در جدول دارای احتمالی برابر باشند. آن‌تروپی را می‌توان به نوعی معیاری برای سنجش بی‌نظمی در نظر گرفت. هر چه مجموعه منظم‌تر باشد و دارای گوناگونی کمتری باشد، آنگاه آن‌تروپی آن کمتر است و به نوعی بی‌نظمی آن نیز کمتر است و برعکس. البته در اینجا چون ما در مرحله قبل یک کلاسه بندی ابتدایی را انجام دادیم، تقریباً بی‌نظمی نیز پایین می‌باشد، و این خود باعث سرعت عمل بالاتر روش پیشنهادی ما می‌شود. زیرا این قضیه باعث می‌شود که عمق درخت تصمیم‌گیری کم شود و هر چه عمق این درخت کمتر باشد، سرعت تصمیم‌گیری نیز بیشتر می‌شود.

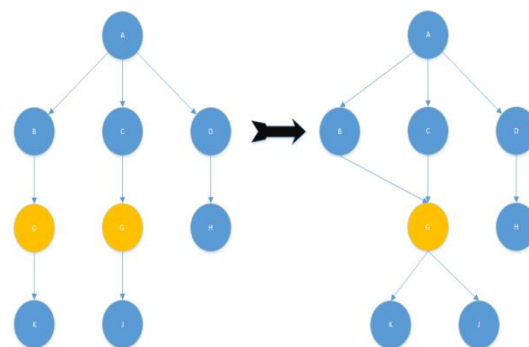
در این قسمت ما از آن‌تروپی استفاده کردیم تا مقدار بی‌نظمی را برای هر صفت از 0 بالا بدست آوریم. برای اینکه بتوانیم صفتی را در درخت تصمیم‌گیری انتخاب کنیم که در رتبه بالاتری از بقیه صفت‌ها باشد، از فرمول ذیل استفاده کردیم. با توجه به این فرمول، ما آن‌تروپی همه صفات را در مجموعه S محاسبه می‌کنیم و مقدار صفت مجموعه A را از آن می‌کاهیم. مجموعه A ، مجموعه همه صفات انتخاب شده از پدر تا به اینجا در یک مسیر خاص می‌باشد.

مثالی برای انتخاب صفت‌ها:

همانطور که در شکل (۱) قابل مشاهده است، ما می‌بایست آن‌تروپی تمامی صفات را از آن‌تروپی صفات انتخاب شده تا به

- Advanced Research in ISSN, Vol. 3 , No. 12, PP.2278-6252, 2014.
- [3] S.Kumar and G. Kaur , "Detection of Heart Diseases using Fuzzy Logic", International Journal of Engineering Trends and Technology , Vol.4 , pp.2694-2699, 2013.
- [4] O.DeZhi and P.Butt, "C Software & Data Engineering", Global Journal of Computer Science and Technology, Vol. 14 , 2014.
- [5] J. Irani, N. Pise and M. Phatak, "Clustering Techniques and the Similarity Measures used in Clustering: A Survey", pp. 10-11, 2016.
- [6] S. Krishna Anand, R. Kalpana and S. Vijayalakshmi, "Design and Implementation of a Fuzzy Expert System for Detecting and Estimating the Level of Asthma and Chronic Obstructive Pulmonary Disease", Middle-East Journal of Scientific Research Vol. 14, pp. 1435-1444, 2013.
- [7] Mayilvaganan M and K.Rajeswari "Human Blood Pressure Classification Analysis using Fuzzy Logic Control System in Datamining", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, 2014.
- [8] M. Darbandi, "Kalman Filtering for Estimation and Prediction Servers with Lower Traffic Loads for Transferring High-Level Processes in Cloud Computing", HCTL International Journal of Technology Innovations and Research, Vol. 23, pp. 10-20, 2017.
- [9] V.Chandra, P.Singh, "Fuzzy Based High Blood Pressure Diagnosis", International Journal of Advanced Research in Computer Science & Technology (IJARCST), Vol. 2, 2014.
- [10] O. Onuwa, "Fuzzy Expert System for Malaria Diagnosis", Oriental Journal of Computer Science & Technology , Vol.7, No. 2, pp. 273-284, 2014.
- [11] R.Kaur, A.Kaur, "Hypertension Diagnosis Using Fuzzy Expert System", International Journal of Engineering Research and Application (IJERA), 2014.
- [12] h. Kim, B. Kim, "Computerized recognition of Alzheimer disease-EEG using genetic algorithms and neural network", Future Generation Computer Systems, Elsevier B.V. vol.21, pp.1124-1130, 2005.
- [13] Z. Yin, Y. Zhao, X. Lu, and H.Duan, 2014, "A Hybrid Intelligent Diagnosis Approach for Quick Screening of Alzheimer's Disease Based on Multiple Neuropsychological Rating Scales", Hindawi Publishing Computational and Mathematical Methods in Medicine, 2015.
- [14] S. Cheong, S. H. oh, S. Y. Lee, "Support Vector Machines with Binary Tree Architecture for Multiclass Classification", Korea Advanced Institute of Science and Technology, Daejeon, 2004.
- [15] S. Kuma, "Diagnosis of heart disease using Advanced Fuzzy resolution Mechanism" International Journal of Science and Applied Information Technology (IJSAIT), Vol.2 , No.2, PP. 22-30 , 2013.
- [16] Latha k.c., Madhu b, Ayesha s., Ramya r., balasubramanan "Visualization of Risk in Breast Cancer using Fuzzy Logic in Matlab Environment", International Journal of Computational Intelligence Techniques, Vol.4, pp.114-117, 2013.

انجمنی نیز کمک گرفت، ما درخت را تبدیل نمودیم، که در موارد بسیاری این درخت تبدیل به گراف می شود، و از حالت درخت خارج می شود. البته اگر این درخت از نظر ظاهری تبدیل به گراف شود، برای ما همچنان بصورت درخت در نظر گرفته می شود. بنابراین می توان گفت چیزی مابین گراف و درخت بدست می آید. در این قسمت، ما در هر سطح گره‌های همانم را با هم ادغام می کنیم و فرزندان آن‌ها به این گره ادغام شده افزوده می شود. می توان در شکل (۲) این کار را مشاهده نمود تا بهتر بتوان آنرا درک نمود.



شکل ۲: مثالی از تبدیل درخت تصمیم گیری

همانطور که در (۲) می توان مشاهده نمود گره‌های B و C دارای فرزند مشترک G بودند که این دو گره G در یک سطح نیز قرار داشتند. در این حالت این دو گره تبدیل به یک گره شده و فرزندان آن‌ها نیز به گره جدید افزوده می شود.

۴- نتیجه گیری

در حال حاضر بیماری آلزایمر یکی از شایع‌ترین دلایل زوال عقل در افراد سالخورده است. تحقیقات به عمل آمده نشان می دهد که تعداد افراد مبتلا به این بیماری تا ۲۸ سال آینده به دو برابر تعداد فعلی خواهد رسید، و همچنین در سال ۲۰۵۰ از هر ۸۵ نفر یک نفر به این بیماری مبتلا خواهد بود. بنابراین تشخیص این بیماری در اوایل شکل گیری آن بسیار با اهمیت تلقی می گردد. در این مقاله در ابتدا به بررسی مقدمه ای از این بیماری و علم پیش بینی و تشخیص بیماری پرداختیم، و سپس پس از بررسی تاریخچه این علم، به پیشنهاد روش جدیدی در جهت تشخیص بیماری آلزایمر بر مبنای بکارگیری درخت تصمیم گیری که دارای دقت بالایی است پرداختیم.

مراجع

- [1] A.Adeli and M.Neshat, "A Fuzzy Expert System for Heart Disease Diagnosis", Proc. Int. Conf. on Engineers and Computer Scientists , pp. 17-19, 2010.
- [2] N. Mishra and Dr. P. Jha, "A Review on the Applications of Fuzzy Expert System for Disease Diagnosis", International Journal of