

## تحلیل داده‌های مربوط به نشانگرهای زیستی و یافتن نشانگرهای بهینه برای پیش‌بینی بیماری آلزایمر

پدرام علیپور اقدم<sup>۱</sup>، مریم خادمی<sup>۲</sup> و علی برومندنی<sup>۳</sup>

<sup>۱</sup>دانشجوی کارشناسی ارشد، دانشگاه آزاد اسلامی واحد تهران جنوب،

alipourpedram@gmail.com

<sup>۲</sup>استادیار، دانشگاه آزاد اسلامی واحد تهران جنوب،

khademi@azad.ac.ir

<sup>۳</sup>استادیار، دانشگاه آزاد اسلامی واحد تهران جنوب،

broumandnia@azad.ac.ir

### چکیده

با وجود تلاش‌های بسیار، هیچ درمان مؤثری برای بیماری آلزایمر تأیید نشده است و شکست آزمایشات متعدد به‌عنوان‌بخار علمی دنیا تبدیل شده‌است. بیماری آلزایمر از جمله بیماری‌هایی است که تشخیص و درمان آن از نظر پزشکان متخصص حائز اهمیت بوده است. در این مقاله قصد داریم که به بررسی و شناخت و تشخیص بیماری آلزایمر و یافتن کم‌ترین نشانگرهای زیستی که بهترین دقت را در تشخیص بیماری داشته باشند با استفاده از تکنیک‌های داده‌کاوی، بپردازیم. در این پژوهش، به تجزیه و تحلیل نشانگرهای زیستی و تشخیص بیماری آلزایمر به کمک مدل‌های داده‌کاوی پرداخته و نحوه به کارگیری ابزار و الگوریتم‌های متمایز و ویژگی‌های مختلف پایگاه‌های داده گوناگون را مورد تجزیه و تحلیل قرار داده و همچنین نتایج تحلیل داده‌های موجود در پایگاه داده ADNI<sup>۱</sup> نیز مورد بررسی قرار خواهد گرفت.

### کلمات کلیدی

بیماری آلزایمر، آپولیپوپروتئین‌ئی، اختلالات اداری خفیف، گروه کنترل بدون تحلیل سلول‌های مغزی، داده‌کاوی

### مقدمه

پروتئین‌های سازنده آمیلوئید (APP<sup>۲</sup>)، دلیل عمده بیماری آلزایمر در مراحل اولیه آن است. تمایل پژوهش‌ها در زمینه داده‌کاوی پیشگو برای پزشکی بالینی در سال‌های اخیر رشد داشته است. در این حوزه، هدف داده‌کاوی پیشگو ساخت مدل‌هایی است که از اطلاعات پزشکی با ابعاد بالا (تنوع ویژگی‌ها) برای پشتیبانی از تصمیم‌گیری‌های پزشکی و پیشگویی نتایج تشخیص (با داده‌های پزشکی بیمارانی که تا به حال مدل آن‌ها را مشاهده نکرده است)، استفاده می‌کند. مشاهده و کشف زودهنگام بیماری آلزایمر (AD<sup>۳</sup>) مشکلی بزرگی است که نیازمند توجه به مشاهدات بیش‌تری دارد. به‌طور بالقوه ویژگی‌های بسیاری می‌تواند برای تشخیص افراد بدون بیماری آلزایمر و آن‌هایی که در مراحل مختلف بیماری هستند به کار برده شود. این ویژگی‌ها شامل نتایج آزمون‌های شناختی و حافظه‌ای، نتایج تصویربرداری (MRI, PET)،

در دو دهه اخیر، بیماری آلزایمر از جمله بیماری‌هایی است که تشخیص و درمان آن از نظر پزشکان متخصص حائز اهمیت بوده است. مکانیسم‌های دقیق مولکولی که منجر به بروز علائم و تغییرات سیستم عصبی بیماران آلزایمر می‌شود، همچنان ناشناخته است. از دست رفتن برخی نورون‌های مغزی، پلاک‌های آمیلوئید خارج سلولی و درهم پیچیدگی تارهای عصبی درون سلولی (NFT<sup>۴</sup>) و پروتئین tau به صورت مشخص در مغز بیماران آلزایمر مشاهده می‌شود. بنابر "نظریه آمیلوئید" که به صورت گسترده مورد پذیرش قرار گرفته شده است، اجتماع غیرمعمول بتا آمیلوئید پپتیدها (A $\beta$ )، حاصل از شکسته شدن

نهایت قابل فهم از داده‌ها. الگوهای قابل فهم برای تشریح داده‌های موجود، پیش‌بینی و یا طبقه‌بندی داده‌های جدید، جمع‌بندی داده‌ها و در نتیجه کمک به افراد برای کشف الگوهای عمیق‌تر در داده‌ها به کار گرفته می‌شوند [۳]. داده‌کاوی در پزشکی، زمینه‌ای نوظهور و با اهمیت بسیار زیاد برای ارائه تشخیص و همچنین درک بهتر طبقه‌بندی بیماری است.

مباحث حول محور فصل مشترک داده‌کاوی و پزشکی، در تشخیص بیماری، به دو بخش اساسی و کلی تقسیم می‌شود: بخش اول و مهم‌ترین بخش، یافتن نشانگرهای زیستی که بیش‌ترین تأثیر را در تشخیص بیماری دارند و به زبان داده‌کاوی یافتن کم‌ترین ویژگی‌ها با بیش‌ترین تأثیر در تشخیص بیماری است. بخش دوم توسعه و یافتن مدلی با دقت بالا با استفاده از داده‌های بیماران و ویژگی‌های یافت شده در بخش نخست است. مقالات موجود در این زمینه عمده تلاش خود را صرف پیدا کردن ویژگی‌های بهینه نموده‌اند چرا که تأثیر و نتیجه این تلاش را در دقت و کارایی مدل توسعه یافته خواهند دید. علت دیگری که اهمیت بخش اول را بیش‌تر کرده ابعاد بالا در داده‌های پزشکی است به عبارت دیگر ویژگی‌های بسیاری را می‌توان برای یک بیمار متصور شد که می‌تواند در آموزش مدل و توسعه آن دخیل باشد. بنابراین کارگیری تمامی این ویژگی‌ها در آموزش مدل قطعاً از کارایی مدل می‌کاهد. علاوه بر کارایی، دقت مدل و قابلیت تعمیم نیز در مدل توسعه یافته مهم است که با کمینه و بهینه کردن ویژگی‌ها هر دو مورد افزایش خواهد یافت. از طرف دیگر، توجه کم‌تر به کاربردی بودن مدل نیز کاهش ویژگی‌ها را توجیه می‌کند. اگر مدل توسعه یافته نیاز به ورودی‌های زیادی برای تشخیص بیماری داشته باشد، ترجیح کلینیک‌های پزشکی استفاده از روش‌های رایج خواهد بود.

تشخیص فاز اولیه‌ی بیماری آلزایمر، در اختلال با مشکلات خفیف شناختی (MCI) است و ترکیب بهینه نشانگرهای زیستی ممکن است بین جمعیت‌ها متفاوت باشد بنابراین نیاز به شخصی‌سازی وجود دارد. همچنین، نشانگرهای زیستی مختلف ممکن است نقش‌های مختلفی در طی مراحل مختلف بیماری و در نتیجه ساخت ترکیبی از مجموعه متغیرهای بهینه با پیشرفت بیماری ایفا کنند [۴].

### انتخاب ویژگی‌ها

در [۵] مارک فَنگیلِز و همکارانش که بخشی از مشارکت مالی اتحادیه اروپا در پروژه پیش‌بینی آلزایمر است، هدف، یافتن بهترین ترکیب از نشانگرهای زیستی برای تشخیص بیماری آلزایمر با استفاده از داده‌های نامگن (تصویربرداری، الکتروفیزیولوژی، سطح مولکولی، آزمون‌های بالینی، جمعیت) و توسعه ابزار بالینی مفید که باعث یکپارچه‌سازی نتایج بهینه نشانگرهای سلولی می‌شوند، است. داده‌های مورد استفاده از پایگاه داده‌ی (ADNI) به‌دست‌آمده‌است. پایگاه داده

اطلاعات مایع نخاعی، نشانگرهای خون و ... می‌باشند. به‌طور کلی مداوی دارو محور، کاهش سرعت پیشرفت علائم بیماری را مدنظر دارد. این روش‌ها در گروه گسترده‌ای از بیماران مؤثر واقع شده است، اما موفقیت این روش‌ها مستقیماً به شناسایی حاملان بیماری در مراحل اولیه آن بستگی دارد. این مسأله نیاز به روش‌های دقیق و به موقع، برای تشخیص این بیماری را توجیه می‌کند.

### عوامل موثر در بیماری آلزایمر

بیماری آلزایمریک بیماری پیش‌رونده تحلیل سلول‌های عصبی است که امروزه ۴/۵ میلیون نفر در آمریکا به آن مبتلا هستند. پیش‌بینی‌می‌شود این تعداد به ۱۱ تا ۱۶ میلیون نفر در سال ۲۰۵۰ برسد که در این دوره، نسل اوج جمعیتی به سالمندی می‌رسند [۱]. بیماری آلزایمر در مرحله مقدماتی آن، با اختلالات ادراکی خفیف (MCI) و تغییرات گسترده بیمارگونه در مغز شناسایی شده و مسأله‌ای بزرگ در بهداشت و درمان محسوب می‌شود [۲]. به نظر می‌رسد بیماری آلزایمر به میزان زیادی ارثی بوده و از طریق عوامل ژنتیکی پیچیده‌ای که زمان‌بندی بیماری را تحت تأثیر قرار می‌دهند منتقل می‌شود. متأسفانه قابلیت‌های پیش‌بینی گرانه‌ی عوامل خطر ساز ژنتیکی بسیار اندک است [۲]. بهترین عامل ژنتیک شناسایی شده،  $\epsilon 4$  که ژن ناهمگن *Apolipoprotein E (APOE)* است، در ۲۵٪ تا ۵۰٪ از بیماران آلزایمر وجود نداشته است [۲]. قابلیت‌های ادراکی بیمار آلزایمری در طول زمان به‌صورت قابل برگشت تحلیل می‌رود و در مراحل پایانی بیماری، مراقبت کامل در فعالیت‌های عادی روزانه نیز لازم خواهد بود. در حال حاضر درمانی برای بیماری آلزایمر وجود ندارد، اما روش‌های دارویی و غیردارویی برای کنترل آن ایجاد شده است.

بیماری آلزایمر بین ۶۰٪ تا ۷۰٪ از موارد زوال عقل را پوشش می‌دهد و تا زمانیکه داروهای آینده و استراتژی‌های پیشگیرنده دسترس باشند، تشخیص اولیه و نظارت بر پیشرفت آلزایمر بخش اصلی درمان است. علاوه بر طبقه‌بندی داده‌ها، انتخاب ویژگی و تکنیک‌های کاهش ویژگی در بیوانفورماتیک، اخیراً یکی از پیش‌نیازها برای ساخت مدل‌های پیشگو شده است، تا ابعاد داده‌های پزشکی را به طوری مؤثر کاهش دهد. طبیعت ابعاد بالا در بیوانفورماتیک، به سرمایه‌ای برای تکنیک‌های انتخاب ویژگی تبدیل شده است. فرصت‌های فراژنتیکی بسیاری برای درک جزئیات مکانیسم مولکولی بیماری آلزایمر وجود دارد که هنوز از آن‌ها استفاده نشده است.

### داده‌کاوی کاربرد آن در پزشکی

داده‌کاوی و یا کشف دانش از داده عبارت است از فرآیند غیرتکراری شناسایی الگوهای صحیح، جدید، ناشناخته، با ظرفیت کاربردی و در

میانگین خطا	۳۰/۵۹٪	۱۵/۱۴٪
کم‌ترین خطا	۱۷٪	۶/۲٪
بیش‌ترین خطا	۴۰/۵٪	۲۵/۷٪
انحراف استاندارد	۶/۲۱	۵/۲۳۶
میانگین دقت	۶۶/۷٪	۸۳/۵٪

جالب توجه است که همه مدل‌های ریاضیاتی و الگوریتم‌هایی که مورد استفاده قرار گرفته است، به پروتئین-6 Interleukin اشاره داشته و آن را در امضای ۱۰ پروتئینی قرار داده‌اند. راوتی و مُسکاتو با استفاده از یک روش بیوانفورماتیکیکپارچه، به یک امضای کوچکتر دست یافتند. امضای ۶ پروتئینی از طریق تحلیل گراف روابط پروتئین‌ها بدست آمده و بسیار جالب است که IL-6 نیز در این امضای کوچک‌تر در نظر گرفته شده‌است. در نهایت در امضای ۵ پروتئینی، IL-6 کنار گذاشته می‌شود و این ۵ پروتئین اکنون به یک زیرمجموعه مناسب از ۱۸ پروتئین اولیه شناسایی شده توسط ری و همکاران تبدیل شده است. زمانی که در مجموعه آزمون نمونه‌های بیماران آلزایمر مدنظر قرار داده شده‌است، نتایج میانگین (بر روی ۲۴ طبقه‌بندی‌کننده) برای امضای ۵ پروتئین به‌دست می‌آید؛ ۹۶٪ در پیش‌بینی آلزایمر و ۹۰٪ در پیش‌بینی گروه تحت کنترل، بدون تحلیل سلول‌های مغزی. این مورد قابل توجه‌ترین نقطه ضعف، کیفیت‌پیش‌بینی بیماری اختلال خفیف شناختی MCI است و از آن‌جا که مدل‌ها برای تشخیص بیماران MCI تمرین ندیده‌اند، استفاده از پروتئین‌های بیش‌تر باعث می‌شود عملکرد اندکی بهبود بیابد. راوتی و مُسکاتو از الگوریتم گسسته‌سازی فیاض و ایرانی برای انتخاب و کمی‌سازی استفاده کرده‌اند.

اما چانگ سولی و همکارانش [۸] روشی ابتکاری ارائه دادند که نتوانست به دقت کار راوتی و مُسکاتو برسد با این حال آن‌ها از روشی ترکیبی استفاده کردند که برخلاف بیش‌تر روش‌ها، قواعد شرطی خواناتری (که رابطه بین نشانگرهای زیستی و نتایج تشخیصی را برقرار می‌کند) برای انسان تولید می‌کند، ضمن این‌که آن‌ها تنها از چهار نشانگر زیستی (چهار پروتئین) برای آموزش مدل خود استفاده کردند. آن‌ها روش مبتکرانه‌ای برای تعیین نشانگرهای زیستی و طبقه‌بندی با استفاده از روش ترکیبیراف و منطق فازیبر روی داده‌های بیماری آلزایمر ارائه کردند و هم‌چنین تئوری مجموعه راف (Rough) برای کاهش تعداد ویژگی‌ها و تولید قواعد به‌طور مؤثر به‌کار گرفته شده است. در واقع روش ارائه شده توسط چانگ سولی و همکارانش در مقایسه با دیگر روش‌های موجود می‌تواند روشی مناسب به عنوان ابزار روش‌شناسی جدید برای تعیین و طبقه‌بندی داده‌های بیماری آلزایمر و به‌طور کل، داده‌های بیولوژی باشد. در واقع با این‌که آن‌ها نتوانستند به دقتی بالاتر دست پیدا کنند اما روشی جدید برای طبقه‌بندی

ADNI شامل داده‌های بیماران تشخیص آلزایمر داده شده و نیز مواردی که تحت کنترل سلامتی می‌باشند، است. داده‌ها از ۸ مطالعه مختلف وجود دارند (غربال‌گری، پایه، و در ماه‌های ۶، ۱۲، ۱۸، ۲۴، ۳۰ و ۳۶ دنبال می‌شود) در مجموع ۵۰۵ ویژگی در بیش از ۱۷ دسته تقسیم و استفاده شده‌اند. مشاهده می‌شود که ابعاد داده‌ها بسیار بالا است.

برای مقایسه ویژگی‌های اسمی از آزمون شاخص نیکویی برازش مجذور کای (Chi-square goodness-of-fit) استفاده شده و سطح اهمیت ۰/۰۱ از نظر آماری با اهمیت محسوب شده است.

### تلاش‌ها برای تشخیص آلزایمر

ری و همکاران [۶] خدمت چشم‌گیری به تلاش‌های جاری برای یافتن آزمایشی در سطح مولکولی به منظور تشخیص سریع‌تر بیماری آلزایمر نمودند. به نظر می‌رسد این روش در سطح جدیدترین فن‌آوری‌ها، بهبود چشم‌گیریافته است و در نتیجه، نتایج آن‌ها توجه بسیاری را در سطح دنیا به خود جلب نموده است. آن‌ها با استفاده از فراوانی ۱۲۰ پروتئین در مجموعه آموزش از ۸۳ نمونه پلاسما ذخیره شده، یک امضای ۱۸ پروتئینی ایجاد کردند. بر روی دو مجموعه مجزای MCI و Alzheimer's Disease (AD) (نمونه‌های افراد با بیماری آلزایمر در برابر افراد تحت کنترل) و ۴۷ نفره (نمونه‌های با اختلالات ادراکی خفیف) این امضا توانست به ترتیب، کارایی کلی ۸۱٪ و ۹۱٪ در پیش‌بینی بیماری آلزایمر نشان دهد.

اما در [۷] راوتی و مُسکاتو با استفاده از داده‌هایی که ری و همکارانش جمع‌آوری کرده بودند به یک امضا ۶ پروتئینی دست‌یافتند که دقت حدود ۹۶٪ در تشخیص آلزایمر دارد. بدترین نتایج عملکرد ممکن که یک مجموعه k پروتئینی می‌تواند داشته باشد، در حالتی است که آن‌ها به صورت تصادفی (از بین ۱۲۰ پروتئین تحت بررسی) انتخاب شده باشند. اما راوتی و مُسکاتو برای درک بهتر این قضیه دو آزمایش جداگانه انجام دادند. به این صورت که به صورت تصادفی ۲۰ بار، ۱۸ پروتئین به صورت تصادفی انتخاب کردند و خطای حاصل از آن‌ها را محاسبه کردند. برای این آزمایش از یک الگوریتم جنگل تصادفی (RF, Random Forests) به عنوان طبقه‌بندی‌کننده پایه استفاده کردند. از آن‌جایی که طبقه‌بندی‌کننده انتخاب شده نیز طبیعتی تصادفی دارد، برای هر امضا، ۱۰ آزمایش با سیدهای مختلف اجرا شده است. آزمایش دوم را مانند آزمایش اول ولی با انتخاب ۶ پروتئین به صورت تصادفی تکرار کردند و نتایج کلی به صورت جدول (۱) به‌دست آمد.

جدول (۱): گزارش آزمایش‌های تصادفی

امضا تصادفی با ۶ ژن      امضا تصادفی با ۱۸ ژن

سفید (WM) و مایع مغزی نخاعی (CSF<sup>V</sup>) دسته‌بندی شد. سپس هر تصویر دسته‌بندی شده براساس بافت از مغز، با استفاده از یک روش تابدادن تصویر با ابعاد بالا، به صورت فضایی در یک فضای قالب، نرمال‌سازی شدند. در زمان تغییر تصویر، کل حجم بافت در هر ناحیه حفظ شده است، این کار با افزایش چگالی در حالت فشرده‌سازی یک ناحیه و بالعکس، انجام شده است. نقشه‌های چگالی بافت در فضای قالب ایجاد شد که نشان‌دهنده اندازه‌های حجمی متناظر با ماده خاکستری، ماده سفید و مایع مغزی نخاعی بود.

روش [۲] شامل مؤلفه‌های زیر است: (۱) تخمین تفکیک‌کننده چند متغیره مبتنی بر تصویر بین گروه تحت کنترل و بیماران آلزایمر؛ (۲) تخمین تفکیک‌کننده چند متغیره چندزنی بین گروه تحت کنترل و بیماران آلزایمر؛ (۳) یکپارچه‌سازی و تجمیع تفکیک‌کننده‌های تصویری و چندزنی برای به دست آوردن یک شناسه تصویری-ژنتیکی برای پیش‌بینی تبدیل به بیماری آلزایمر در بیماران اختلالات خفیف ادراکی.

### تحلیل داده‌ها

موارد آزمون در بخش شناختی ADAS<sup>A</sup> به ترتیب زیر انجام می‌شود: ابتدا آزمون به خاطر آوردن کلمات انجام می‌شود، بخش شناخت کلمات در انجام می‌شود و دیگر آزمون‌های شناختی در بین آن‌ها انجام می‌شود. با جدا سازی این دو قسمت حافظه کلمه‌ای شانس اینکه فرد کلمات را در دوبرخ با هم اشتباه بگیرد به حداقل می‌رسد. در ابتدای جلسه آزمون، قبل از آزمون به خاطر آوردن کلمات، آزمونگر باید یک گفت‌وگوی کوتاه با فرد درباره موضوعات طبیعی مثل آب‌وهوا، آمدن فرد به کلینیک، یا فرد صبحانه چه خورده است، داشته باشد. این گفت‌وگو کمک می‌کند تا فرد در یک راحتی پیش از شروع آزمون قرار بگیرد و به آزمونگر فرصت می‌دهد تا مشاهده کند که فرد چگونه می‌تواند از زبان استفاده کند و زبان را درک کند. آزمون ADAS یک آزمون زمان‌دار نیست و نمره فرد به این بستگی ندارد که تا چه حد سریع آزمون انجام شده است. موارد شناختی باید طوری انجام شود که جلسه به طور یکنواخت و سریع پیش برود، اما نه طوری که فرد احساس فشار کند که باید سریع پاسخ دهد.

### نتایج تحلیل داده‌های آزمون ADAS

در جدول ADASSCORE تعدادی مقدار از دست رفته در حوزه‌های مختلف وجود داشت که راه حل حذف رکوردها (الگوها) مدنظر قرار گرفت چرا که تعداد این الگوها ۹۲ مورد بود که در مقایسه با تعداد کل رکوردها که برابر ۳۸۵۶ است تعداد کمی است بنابراین الگوها با مقادیر از دست‌رفته حذف شدند. ویژگی هدف در تمامی گزارش‌های این پژوهش ویژگی DXCURREN است که دارای سه مقدار ۱، ۲ و ۳

داده‌های بیماری آلزایمر ارائه دادند. ابزاری برای مصور ساختن گسستگی مقادیر مربوط به یک نشانگر زیستی در گروه بیمار در مقابل گروه غیربیمار توسعه یافته شده است.

چیک‌یوچن و همکارانش [۹] روشی محاسباتی برای رتبه‌بندی پروتئین‌های مرتبط با آلزایمر، براساس لیست اولیه از ژن‌های مرتبط با آلزایمر و داده‌های عمومی مربوط به تعاملات پروتئین انسانی ایجاد نمودند. در این روش، لیست اولیه‌ای شامل ۶۵ ژن مرتبط با آلزایمر از پایگاه داده OMIM گردآوری شده است و به ۷۰ پروتئین مربوط به آلزایمر نگاشت شده‌اند. در مجموعه ابتدایی بیماری آلزایمر، لیست اولیه‌ای از ۶۵ رکورد ژن‌های مرتبط با آلزایمر از پایگاه داده به دست آوردند. این رکوردها در ادامه با استفاده از جدول‌های نگاشت پروتئین از پایگاه داده HGNC، به ۷۰ پروتئین نگاشت شدند. این تفاوت تعداد به دلیل نگاشت یک به چند ژن‌ها و بسط پروتئین آن‌ها به وجود آمده است. با استفاده از پایگاه داده OPHID و روش گسترش نزدیک‌ترین همسایه، به ۷۷۵ تعامل پروتئین انسانی مرتبط با آلزایمر دست یافتند. این مجموعه تعاملات گسترش یافته بیماری آلزایمر شامل ۶۵۷ پروتئین انسانی است. یکی از پروتئین‌های مهم بیماری آلزایمر، پروتئین tau عاملی شناخته شده در تحلیل سلول‌های عصبی مرتبط با بیماری آلزایمر است. پروتئین اتصال‌دهنده آمیلوئید بتا (APPB1 (رتبه ۳۳) پروتئینی است که زوج تعاملی شناخته شده برای پروتئین APP است اما در پایگاه داده OMIM رابطه ژنتیک به بیماری آلزایمر تنها برای دیگر عضو این خانواده، یعنی APPB2 (رتبه ۳۲) گزارش شده بود. با این وجود روش چیک‌یوچن و همکارانش پیش‌بینی کرد که APPB1 نیز نقشی در بیماری آلزایمر دارد که در تحقیقات بعدی این امر اثبات شده است

رومن فیلیپویچ و همکارانش [۲] روشی چندگانه تصویری و ژنتیکی برای پیش‌بینی تبدیل بیماری اختلالات خفیف ادراکی (MCI) به بیماری آلزایمر ارائه دادند. آن‌ها نشان دادند که تجمیع اطلاعات چندزنی و تصویربرداری عصبی می‌تواند پیش‌بینی تبدیل به بیماری آلزایمر را بهبود بخشد. آن‌ها یک روش تشخیص الگوی چندمتغیره برابیه‌دست‌آوردنیک تفکیک‌کننده بین افراد سالم و بیماران آلزایمر به کار گرفته‌اند. هم‌چنینیک تفکیک‌کننده چندمتغیره ژنتیک بین جمعیت‌های آلزایمر و تحت کنترل ایجاد کرده‌اند سپس این تفکیک‌کننده‌های تصویربرداری عصبی و چندزنی را در یک مدل خطی مجتمع کرده و یک پیش‌بینی‌کننده ترکیبی تصویری-ژنتیکی برای تبدیلی بیماری اختلالات خفیف ادراکی به بیماری آلزایمر ساخته‌اند. آن‌ها نیز از اطلاعات ژنتیک ADNI استفاده نمودند. تصاویر MR، براساس چارچوب تغییر شکل بدون تغییر حجم، پردازش شدند. هر تصویر MR بدون جمجه از مغز در ابتدا از طریق روش دسته‌بندی بافت مغز که دیژانگ و همکارانش پیشنهاد کرده‌اند، به ماده خاکستری (GM)، ماده

داده‌ها با استفاده از فرمول (۱) نرمال شدند و کمینه و بیشینه به ترتیب ۱- و ۱ انتخاب شد. این کار با استفاده از SQLServer 2012 انجام شد.

$$x'_{ij} = \left( \frac{x_{ij} - x_{min,j}}{x_{max,j} - x_{min,j}} \right) (x'_{max,j} - x'_{min,j}) + x'_{min,j} \quad (1)$$

که در آن  $x'_{min,j}$  مقدار کمینه (-۱)،  $x'_{max,j}$  مقدار بیشینه (+۱)،  $x_{min,j}$  مقدار کمینه،  $x_{max,j}$  مقدار بیشینه حوزه مورد نظر،  $x_{ij}$  مقدار حوزه در هر رکورد و  $x'_{ij}$  مقدار حاصل از نرمال‌سازی است.

#### جدول (۴): نتایج پیش‌بینی ابتلا به بیماری آلزایمر

با استفاده از داده‌های آزمون ADAS بعد از نرمال‌سازی داده‌های

مجموعه	مجموعه تست	مجموعه آموزش	مجموعه آموزش	مجموعه آموزش
71.24%	275	73.25%	2,464	صحیح
28.76%	111	26.75%	900	غلط
	386		3,364	کل

توجه به نمودارهای توزیع مشخص است که توزیع بیش‌تر ویژگی‌ها توزیع نرمالی نیست بنابراین می‌توان برخی از آن‌ها را حذف کرد. باید توجه داشت که هدف ساخت مدل با کم‌ترین ورودی و بیشترین دقت است. بنابراین با استفاده از نمودارهای توزیع ویژگی‌ها، ویژگی‌های حذف شده عبارتند از: Q2، Q3، Q5، Q6، Q9، Q10، Q11، Q12. در این حذف‌ها نیم‌نگاهی به اهمیت ویژگی‌ها که با استفاده از SVM به دست آمده است، داشتیم.

#### جدول (۵): نتایج پیش‌بینی ابتلا به بیماری آلزایمر پس از حذف ویژگی‌ها

مجموعه	مجموعه تست	مجموعه آموزش	مجموعه آموزش	مجموعه آموزش
70.47%	272	71.91%	2,419	صحیح
29.53%	114	28.09%	945	غلط
	386		3,364	کل

مشاهده می‌شود که ۸ ویژگی حذف شد و دقت مدل کم‌تر از ۱ درصد کاهش یافته است.

#### جدول (۶): نتایج پیش‌بینی ابتلا به بیماری آلزایمر با استفاده از

داده‌های آزمون ADAS پس از حذف رکوردهای دورافتاده

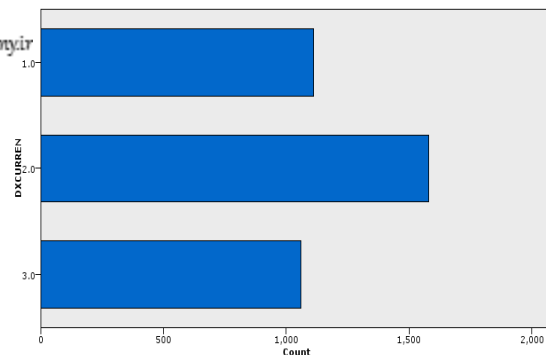
مجموعه	مجموعه تست	مجموعه آموزش	مجموعه آموزش	مجموعه آموزش
69.81%	259	96.93%	3,378	صحیح
30.19%	112	3.07%	107	غلط
	371		3,485	کل

است. مقدار ۱ به معنی سالم، مقدار ۲ به معنی اختلالات خفیف ادراکی و مقدار ۳ به معنی ابتلا به بیماری آلزایمر است. نمودارهای با استفاده از نرم‌افزار Clementine ترسیم‌شده و نتایج حاصل از پیش‌بینی بیماری آلزایمر با استفاده از مدل SVM<sup>۹</sup> به دست آمده است.

#### جدول (۲): فیلدهای موجود در جدول ADASSCORE

نام فیلد	شرح فیلد
Q1	نمره بخش به خاطر آوردن کلمات
Q2	نمره بخش انجام دستورات
Q3	نمره بخش طراحی
Q5	نمره بخش به یاد آوردن اسامی
Q6	نمره بخش انگاره‌سازی
Q7	نمره بخش جهت‌یابی
Q8	نمره بخش تشخیص کلمات
Q9	نمره بخش به خاطر سپردن دستورات آزمون
Q10	نمره بخش توانایی زبانی که صحبت می‌شود
Q12	نمره بخش قوه‌ادراک

در شکل (۱) توزیع ویژگی هدف را مشاهده می‌کنید که مشخص است که افراد مبتلا به اختلالات خفیف ادراکی بیش‌تر هستند.



شکل (۱): توزیع متغیر هدف DXCURREN در جدول ADASSCORE

جدول (۳): نتایج پیش‌بینی ابتلا به بیماری آلزایمر با استفاده از داده‌های آزمون ADAS. لازم به ذکر است که داده‌ها نرمال نشده‌اند.

مجموعه	مجموعه تست	مجموعه آموزش	مجموعه آموزش	مجموعه آموزش
69.81%	259	96.93%	3,378	صحیح
30.19%	112	3.07%	107	غلط
	371		3,485	کل

(ICMLC), 2010 Second International Conference on, pp. 218-222.

- [4] Mark van Gils, Juha Koikkalainen, Jussi Mattila, Sanna-Kaisa Herukka, Jyrki Lotjonen, Hilka Soininen and the Alzheimer's Disease Neuroimaging Initiative, "Discovery and use of efficient biomarkers for objective disease state assessment in Alzheimer's disease". 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31- September 4, 2010.
- [5] Predict Alzheimer's disease, 2013, www.predictad.eu .
- [6] Sandip Ray, Markus Britschgi, Charles Herbert, Yoshiko Takeda-Uchimura, Adam Boxer, Kaj Blennow, Leah F Friedman, Douglas R Galasko, Marek Jutel, Anna Karydas, Jeffrey A Kaye, Jerzy Leszek, Bruce L Miller, Lennart Minthon, Joseph F Quinn, Gil D Rabinovici, William H Robinson, Marwan N Sabbagh, Yuen T So, D Larry Sparks, Massimo Tabaton, Jared Tinklenberg, Jerome A Yesavage, Robert Tibshirani, and Tony Wyss-Coray, "Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins." Second International Conference on Machine Learning and Computing, Oct.2007.
- [7] Martin Gomez Ravetti, Pablo Moscato, "Identification of a 5-Protein biomarker molecular signature for predicting Alzheimer's disease." Centre for Bioinformatics, Biomarker Discovery & Information-Based Medicine, The University of Newcastle. 2008.
- [8] L. ChangSu, L. Chiou-Peng, and M. Masek, "Rough-Fuzzy Hybrid Approach for Identification of Biomarkers and Classification on Alzheimer's Disease Data," in Bioinformatics and Bioengineering (BIBE), 2011 IEEE 11th International Conference on, pp. 84-91.
- [9] Jake Yue Chen, Changyu Shen, Andrey Y. Sivachenko, "Mining Alzheimer disease relevant proteins from integrated protein interactome data." Pacific Symposium on Biocomputing, pp. 367-378. World Scientific, (2006).

	آموزش	تست	تست
صحیح	2,415	71.94%	276
غلط	942	28.06%	110
کل	3,364		386

افزایش دقت حدود ۱ درصدی در این نتایج نسبت به حالت قبل مشخص است. مشاهده می‌شود که با استفاده از داده‌های تست ADAS که شامل پرسش‌هایی از فرد مورد مطالعه است، به دقتی حدود ۷۲٪ رسیدیم. که البته این دقت با حذف موارد اختلالات خفیف ادراکی بیش‌تر خواهد شد.

### نتیجه‌گیری

طبق مطالب بیان شده ملاحظه کردیم که همه محققین جهت تشخیص بیماری آلزایمر به انتخاب ویژگی‌های مناسب اهمیت زیادی داده‌اند. برخی تحقیقات تنها از یک نوع داده مثلاً داده‌های مربوط به پروتئین‌های موجود در پلاسما خون استفاده کرده‌اند و برخی از انواع مختلف داده مثل اطلاعات ژنتیکی، تصویربرداری‌ها از مغز، اطلاعات مایع نخاعی و دیگر از انواع اطلاعات استفاده کرده‌اند. آن چه که مواضعست این است که هر چه ورودی‌های مدل برای انجام پیش‌بینی کم‌تر باشد مدل ساده‌تر و کاربردی‌تر خواهد بود با این حال دقت مدل را نباید از دست داد. اما به نظر می‌رسد ترکیب این اطلاعات دقت بهتری به مدل خواهد بخشید. ولی آن چه که مورد توجه قرار نگرفته است استفاده از اطلاعاتی است که مربوط به نحوه زندگی افراد و نوع تغذیه آن‌ها است. اطلاعاتی که می‌تواند تأثیر به‌سزایی در ابتلاء افراد به این بیماری داشته باشد. با این که ابتلاء افراد به این بیماری با ژن افراد بسیار در رابطه است اما به نظر می‌رسد نحوه زندگی افراد در دهه‌های اخیر و ماشینی شدن زندگی افراد و افزایش این بیماری نیز تأثیر کمیندارد و باید مورد توجه قرار بگیرد. استفاده از دخیانیات و مصرف مشروبات الکلی نیز می‌تواند مورد تحقیق قرار گیرد.

### مراجع

- [1] Alzheimer's association, 2013, www.alz.org .
- [2] R. Filipovych, B. Gaonkar, and C. Davatzikos, "A Composite Multivariate Polygenic and Neuroimaging Score for Prediction of Conversion to Alzheimer's Disease," in Pattern Recognition in NeuroImaging (PRNI), 2012 International Workshop on, pp. 105-108.
- [3] S. Joshi, D. Shenoy, G. G. Vibhudendra Simha, P. L. Rrashmi, K. R. Venugopal, and L. M. Patnaik, "Classification of Alzheimer's Disease and Parkinson's Disease by Using Machine Learning and Neural Network Methods," in Machine Learning and Computing



## زیر نویس ها

Alzheimer's Disease Neuroimaging Initiative<sup>۱</sup>  
Neurofibrillary Tangles<sup>۲</sup>  
Amyloid Precursor Protein<sup>۳</sup>  
Alzheimer's Disease<sup>۴</sup>  
Mild Cognitive Impairment<sup>۵</sup>، اختلالات ادراکی خفیف  
Apolipoprotein E<sup>۶</sup>، نوعی ژن که بر روی کروموزوم ۱۹ واقع است.  
Cerebrospinal Fluid<sup>۷</sup>، مایع مغزی نخاعی  
Alzheimer's Disease Assessment Scale<sup>۸</sup>، مقیاس ارزیابی بیماری  
آلزایمر  
Support Vector Machine<sup>۹</sup>، ماشین بردار پشتیبانی

dataacademy.ir