

دریافت مقاله: ۱۴۰۳/۱۲/۰۹

پذیرش مقاله: ۱۴۰۴/۰۱/۲۳

نوع مقاله: پژوهشی

بهبود تشخیص بیماری آلزایمر با ارائه یک روش ترادیسگر تصویری برای تحلیل تصاویر رزونانس مغناطیسی ساختاری

علیرضا جمیری

دانشکده مهندسی کامپیوتر، واحد قم، دانشگاه آزاد اسلامی، قم، ایران

پست الکترونیکی: alirezajomeiri@iau.ac.ir

احمد حبیبی زاد نوین*

استادیار، دانشکده فنی و مهندسی، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران

پست الکترونیکی: a.habibizad@srbiau.ac.ir

محبوبه شمس

استادیار، دانشکده مهندسی، دانشگاه صنعتی، قم، ایران

پست الکترونیکی: shamsi@qut.ac.ir

چکیده

۹۱/۸ درصد، ویژگی ۹۵/۷ درصد و مساحت زیر منحنی ROC (AUC) برابر با ۰/۹۶ دست یافت. در مقایسه، بهترین روش موجود، یعنی مدل ترکیبی CNN-ViT، به دقت ۹۰/۲ درصد، حساسیت ۸۷/۵ درصد، ویژگی ۹۱/۰ درصد و AUC برابر با ۰/۹۲ رسیده بود، که نشان‌دهنده برتری قابل توجه RAE-ViT است. این مدل همچنین نقشه‌های توجه تفسیرپذیر و هم‌راستا با دانش بالینی تولید می‌کند. برای مقابله با چالش‌های کوچک بودن مجموعه داده و هزینه‌های محاسباتی، RAE-ViT از استخراج ویژگی چندمقیاس و راهبردهای بهینه‌سازی آموزش بهره می‌برد. RAE-ViT رویکردی نویدبخش برای تشخیص زودهنگام آلزایمر ارائه می‌دهد و می‌تواند نتایج بیماران را بهبود بخشد. کلیدواژه‌ها: طبقه‌بندی بیماری آلزایمر، ترادیسگر تصویری، سازوکارهای توجه منطقه‌ای، تصویربرداری رزونانس مغناطیسی، یادگیری عمیق

بیماری آلزایمر نیازمند تشخیص زودهنگام و دقیق برای مداخله مؤثر است. با این حال، روش‌های فعلی مبتنی بر شبکه‌های عصبی هم‌آمیختی و ترادیسگرهای تصویری استاندارد اغلب در تشخیص الگوهای سراسری آتروفی مغزی و تمرکز بر نواحی کلیدی مانند هیپوکامپ با چالش‌هایی مواجهند و دقت و تفسیرپذیری محدودی ارائه می‌دهند. این مقاله مدل ترادیسگر تصویری تقویت شده با توجه منطقه‌ای (RAE-ViT) را معرفی می‌کند که برای طبقه‌بندی آلزایمر با استفاده از داده‌های تصویربرداری رزونانس مغناطیسی ساختاری طراحی شده است. RAE-ViT با بهره‌گیری از سازوکارهای توجه منطقه‌ای، بر نواحی کلیدی مغز (مانند هیپوکامپ و بطن‌ها) تمرکز می‌کند و از توجه سلسله‌مراتبی برای استخراج الگوهای ساختاری محلی و سراسری استفاده می‌نماید. این مدل، ارزیابی شده روی مجموعه داده ادنی، به دقت ۹۴/۲ درصد، حساسیت

مقدمه

بیماری آلزایمر یک اختلال عصبی-تحلیلی پیشرونده است که باعث کاهش توانایی شناختی و از دست دادن حافظه شده و میلیون‌ها نفر را در سراسر جهان تحت تأثیر قرار می‌دهد. تشخیص زودهنگام این بیماری برای بهبود نتایج درمانی بسیار حیاتی است [۱]. تصویربرداری رزونانس مغناطیسی ساختاری ابزاری مؤثر برای شناسایی تغییرات مغزی مرتبط با آلزایمر محسوب می‌شود، به‌ویژه در نواحی مهمی مانند هیپوکامپ و بطن‌های مغزی. با این حال، تفسیر دستی تصاویر رزونانس مغناطیسی فرایندی زمان‌بر است و به دلیل وابستگی به تجربه و قضاوت فردی پزشکان، می‌تواند به نتایج ناسازگار یا متغیر منجر شود، که ضرورت توسعه سامانه‌های نرم‌افزاری تشخیصی خودکار را برجسته می‌سازد. یادگیری عمیق، به‌ویژه شبکه‌های عصبی هم‌آمیختی^۲، در تحلیل تصاویر پزشکی از جمله طبقه‌بندی بیماری آلزایمر موفق عمل کرده است، اما در تشخیص روابط بلند مدت و الگوهای ساختاری سراسری که برای تشخیص بیماری حیاتی هستند، با محدودیت‌هایی روبه‌رو است [۲]. ترادیسگرهای تصویری با استفاده از سازوکارهای خودتوجهی، این محدودیت‌ها را برطرف کرده و پردازش جامع تصاویر با وضوح بالا را امکان‌پذیر می‌سازند، که آن‌ها را به گزینه‌ای ایده‌آل برای تصویربرداری عصبی تبدیل می‌کند [۳]. پیشرفت‌هایی مانند ماژول‌های توجه منطقه‌ای^۳ باعث بهبود تمرکز بر نواحی مهم بیماری (مانند هیپوکامپ و بطن‌ها) شده و دقت و تفسیرپذیری مدل‌ها را افزایش داده‌اند. علاوه بر این، سازوکارهای توجه چندمقیاسی با ترکیب ویژگی‌های جزئی و کلی، دقت طبقه‌بندی آلزایمر را بهبود می‌بخشند [۴].

مدل‌های ترکیبی که شبکه‌های عصبی هم‌آمیختی و ترانسفورمرها را ادغام می‌کنند، با ترکیب قابلیت استخراج ویژگی‌های فضایی شبکه عصبی هم‌آمیختی و سازوکار

توجه سراسری ترانسفورمرها، عملکرد بهتری نسبت به مدل‌های مستقل ارائه می‌دهند. همچنین، ادغام داده‌های چندوجهی (مانند توموگرافی گسیل پوزیترون^۴، نشانگرهای زیستی مایع مغزی نخاعی^۵ و اطلاعات ژنتیکی) همراه با سازوکارهای توجه میان‌مدالی، دقت تشخیصی را افزایش داده و دید جامع‌تری از آسیب‌شناسی آلزایمر ارائه می‌دهد [۵]. یادگیری خودنظارتی^۶ با بهره‌گیری از داده‌های بدون برچسب، تعمیم‌پذیری مدل را بهبود می‌بخشد، در حالی که یادگیری مشارکتی امکان همکاری و حفظ حریم خصوصی میان مؤسسات مختلف را فراهم کرده و چالش کمبود داده را برطرف می‌کند [۶]. با وجود پیشرفت‌های صورت‌گرفته، بهینه‌سازی ترادیسگرهای تصویری برای کاربردهای بالینی با چالش‌هایی مانند هزینه محاسباتی بالا، نیاز به مجموعه‌های داده‌ای بزرگ و برچسب‌گذاری شده، و مسائل مربوط به تفسیرپذیری مواجه است [۷]. تحقیقات همچنان در حال انجام است تا معماری‌های سبک‌تر توسعه یافته و قابلیت تفسیر مدل‌ها بهبود یابد.

این مقاله، مدل ترادیسگر تصویری تقویت شده با توجه منطقه‌ای (RAE-ViT) را معرفی می‌کند که برای طبقه‌بندی بیماری آلزایمر با استفاده از داده‌های تصویربرداری رزونانس مغناطیسی طراحی شده است. RAE-ViT با ادغام سازوکارهای توجه منطقه‌ای، خودتوجهی سلسله‌مراتبی، و استخراج ویژگی چندمقیاسی، به دنبال دستیابی به عملکردی برتر همراه با حفظ تفسیرپذیری است. این مدل، ارزیابی‌شده روی مجموعه داده ادنی^۷ [۸]، به دقت ۹۴/۲ درصد، حساسیت ۹۱/۸ درصد، و ویژگی ۹۵/۷ درصد و مساحت زیر منحنی ROC (AUC) برابر با ۰/۹۶ دست یافت، در حالی که بهترین روش موجود، یعنی مدل ترکیبی CNN-ViT، به دقت ۹۰/۲ درصد، حساسیت ۸۷/۵ درصد، و ویژگی ۹۱/۰ درصد و AUC برابر با ۰/۹۲ رسیده بود. برخلاف کارهای قبلی که ترادیسگرهای تصویری را برای آلزایمر

4-Positron Emission Tomography (PET)

5-Cerebrospinal Fluid (CSF)

6-Self-supervised learning (SSL)

7-Alzheimer's Disease Neuroimaging Initiative

2-CNN

3-Regional Attention Module (RAM)

تصاویر رزونانس مغناطیسی عملکردی و توموگرافی گسیل پوزیترون به کار رفتند و دقت قابل قبولی (حدود ۸۷-۸۸ درصد) بدون نیاز به مهندسی ویژگی به دست آوردند [۷]. با این حال، شبکه‌های عصبی هم‌آمیختگی استاندارد در درک وابستگی‌های بلندمدت در الگوهای آتروفی مغزی (مانند تغییرات هم‌زمان در هیپوکامپ و بطن‌ها) با محدودیت مواجهند، زیرا میدان پذیرشی محلی آنها نمی‌تواند روابط سراسری را به خوبی مدل‌سازی کند [۱۷]. این کمبود باعث می‌شود که تشخیص زودهنگام، که به شناسایی الگوهای ظریف وابسته است، چالش‌برانگیز شود.

ترادیسگرهای تصویری به‌عنوان جایگزینی برای شبکه‌های عصبی هم‌آمیختگی معرفی شدند و با استفاده از سازوکارهای خودتوجهی^۸، توانایی مدل‌سازی وابستگی‌های سراسری را بهبود بخشیدند [۳]. این مدل‌ها تصاویر را به وصله‌هایی تقسیم می‌کنند و انعطاف‌پذیری بیشتری در نمایش ویژگی‌ها ارائه می‌دهند، که آنها را در وظایف تصویربرداری پزشکی مثل طبقه‌بندی تومور و تقسیم‌بندی ضایعات مؤثر کرده است [۸]. با این حال، ترادیسگرهای تصویری استاندارد در تحلیل تصاویر رزونانس مغناطیسی برای آلزایمر هنوز کاملاً بهینه نشده‌اند. به عنوان مثال، کونانبایف و همکاران [۹] ترادیسگر تصویری را با داده‌های محدود برای طبقه‌بندی آلزایمر آزمایش کردند و به دقت ۸۹/۵ درصد رسیدند، اما فقدان سوگیری‌های القایی دامنه‌محور (مانند تمرکز بر ساختارهای آناتومیکی خاص) دقت را در مقایسه با نیازهای بالینی محدود کرد. علاوه بر این، این مدل‌ها اغلب تفسیرپذیری کافی نداشتند، زیرا نقشه‌های توجه آنها با دانش بالینی هم‌راستا نبودند [۱۸].

مدل‌های ترکیبی که شبکه‌های عصبی هم‌آمیختگی و ترادیسگرهای تصویری را ادغام می‌کنند، تلاش کردند این محدودیت‌ها را جبران کنند. برای مثال، جمیری و همکاران [۷] از ترکیبی از DenseNet و BiLSTM برای

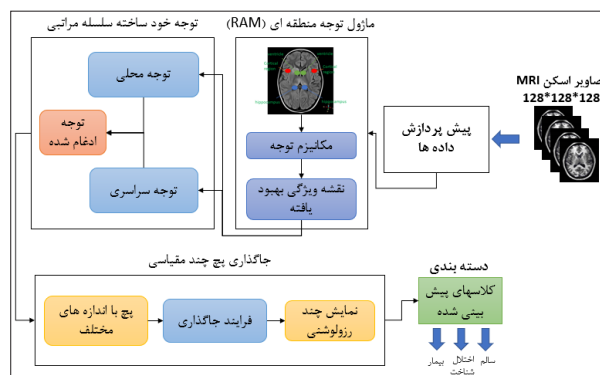
به کار برده‌اند، نوآوری این پژوهش در معرفی RAE-ViT نهفته است که برای اولین بار توجه منطقه‌ای پویا را با تمرکز بر نواحی کلیدی مغز (مانند هیپوکامپ و بطن‌ها) و خودتوجهی سلسله‌مراتبی را در یک چارچوب چندمقیاسی ادغام می‌کند، و نقشه‌های توجه تفسیرپذیری تولید می‌کند که با دانش بالینی هم‌راستاست و عملکرد تشخیصی را به طور قابل‌توجهی بهبود می‌بخشد. این مقاله در ادامه به این ترتیب سازمان‌دهی شده است: بخش دوم کارهای مرتبط را مرور می‌کند، بخش سوم جزئیات روش پیشنهادی RAE-ViT را شرح می‌دهد، بخش چهارم تنظیمات آزمایشی را توضیح می‌دهد، بخش پنجم نتایج و مقایسه‌ها را ارائه می‌دهد، و بخش ششم با نتیجه‌گیری و پیشنهادها برای کارهای آینده پایان می‌یابد.

۲- کارهای مرتبط

تلاش‌های اولیه برای طبقه‌بندی بیماری آلزایمر بر مدل‌های آماری و ویژگی‌های مهندسی‌شده از داده‌های تصویربرداری عصبی مانند تصاویر رزونانس مغناطیسی ساختاری، تصاویر رزونانس مغناطیسی عملکردی، و توموگرافی گسیل پوزیترون متکی بود [۱]. مدل‌های سنتی یادگیری ماشین، مانند ماشین بردار پشتیبان و جنگل‌های تصادفی، از ویژگی‌هایی مثل حجم هیپوکامپ و ضخامت قشر مغز برای تمایز بیماران آلزایمری، افراد با اختلال شناختی خفیف، و افراد سالم استفاده می‌کردند [۶]. با این حال، این روش‌ها به دلیل وابستگی به استخراج دستی ویژگی‌ها، تعمیم‌پذیری محدودی به مجموعه داده‌های متنوع داشتند و در برابر تغییرات ظریف ساختاری ناکارآمد بودند.

شبکه‌های عصبی هم‌آمیختگی با استخراج خودکار نمایش‌های سلسله‌مراتبی و ویژگی‌ها از داده‌های خام تصویربرداری، طبقه‌بندی آلزایمر را بهبود بخشیدند [۲]. برای مثال، مدل‌هایی مثل ResNet-50 و DenseNet-121 در پردازش تصاویر رزونانس مغناطیسی ساختاری و

ساختاری طراحی شده است. این مدل با بهره‌گیری از توجه منطقه‌ای^۹، خودتوجهی سلسله‌مراتبی و تعبیه‌های چندمقیاسی^{۱۰}، هم نشانگرهای بیماری در مقیاس محلی و هم الگوهای ساختاری سراسری را استخراج می‌کند. سازوکارهای توجه در تصویربرداری پزشکی به‌طور مؤثر عمل کرده‌اند و منجر به بهبود دقت تشخیصی و افزایش تفسیرپذیری مدل‌ها شده‌اند [۱۴]. شکل ۱ مراحل روش پیشنهادی را نمایش می‌دهد.



شکل ۱: روش پیشنهادی

فرمول‌های زیر ترکیبی از سازوکارهای استاندارد ترادیسگرها و نوآوری‌های خاص این پژوهش هستند که برای بهبود دقت و تفسیرپذیری در تشخیص آلزایمر توسعه یافته‌اند.

۳-۱- ماژول توجه منطقه‌ای

ماژول توجه منطقه‌ای به‌طور پویا بر نواحی حیاتی مغز که با استفاده از اطلاعات آناتومیکی پیشین شناسایی شده‌اند، تمرکز می‌کند. این نواحی شامل هیپوکامپ (Φ_h)، بطن‌ها (Φ_v) و نواحی قشری مغز (Φ_c) هستند. مجموعه نواحی مورد توجه^{۱۱} به صورت زیر تعریف می‌شود:

$$\Phi = \{\Phi_h, \Phi_v, \Phi_c, \dots\} \quad (1)$$

برای هر منطقه $\Phi_i \in \Phi$ ، نقشه توجه Ar با استفاده از سازوکار توجه ضرب نقطه‌ای مقیاس‌یافته محاسبه

تحلیل طولی تصاویر رزونانس مغناطیسی استفاده کردند و به AUC 0.90 رسیدند، اما این روش همچنان در تمرکز بر نواحی کلیدی مغز (مثل هیپوکامپ) ضعف داشت و هزینه محاسباتی بالایی داشت. به‌طور مشابه، مدل Hybrid CNN-ViT [۱۰] با ترکیب ویژگی‌های فضایی شبکه عصبی هم‌آمیختگی و توجه سراسری ترادیسگر تصویری به دقت ۹۰/۲ درصد دست یافت، ولی عدم استفاده از توجه منطقه‌ای خاص، توانایی آن را در تشخیص تغییرات محلی ظریف کاهش داد. این کمبودها نشان می‌دهد که روش‌های موجود در متعادل کردن دقت، تفسیرپذیری، و کارایی محاسباتی ناکاملند.

سازوکارهای توجه در تصویربرداری پزشکی تأثیری زیادی بر بهبود مدل‌ها داشته‌اند [۱۳]. توجه فضایی برای برجسته کردن نواحی مهم (مثل تومورها) و توجه کانالی برای وزن‌دهی نقشه‌های ویژگی به کار رفته و در تشخیص بیماری‌هایی مثل سرطان سینه موفق بوده است [۱۴]. با این حال، در زمینه آلزایمر، کاربرد این سازوکارها کمتر توسعه یافته است. کانگ و لیو [۱۵] مدل 3DNeST را با خودتوجهی محلی برای تشخیص آلزایمر پیشنهاد کردند، اما این روش به الگوهای سراسری کمتر توجه داشت و AUC آن به ۰/۹۱ محدود شد. در مقابل، روش پیشنهادی در این مقاله، برای اولین بار توجه منطقه‌ای پویا (متمرکز بر هیپوکامپ و بطن‌ها) و خودتوجهی سلسله‌مراتبی را با استخراج ویژگی چندمقیاسی ادغام می‌کند تا هم الگوهای محلی و هم سراسری آتروفی مغزی را با دقت ۹۴/۲ درصد و AUC 0.96 مدل‌سازی کند، ضمن این که نقشه‌های توجه تفسیرپذیری ارائه می‌دهد که با دانش بالینی هم‌راستاست و کارایی تشخیصی را بهبود می‌بخشد.

روش پیشنهادی

در تحقیق حاضر ترادیسگر تصویری تقویت شده با توجه منطقه‌ای (RAE-ViT) برای طبقه‌بندی بیماری آلزایمر با استفاده از داده‌های تصویربرداری رزونانس مغناطیسی

9-Regional Attention
10-Multi- Scale Embeddings
11-Region of Interest (ROI)

$$I'(x, y, z) = T(I(x, y, z)),$$

$$M_r'(x, y, z) = T(M_r(x, y, z)) \quad (۴)$$

با تأکید بر جزئیات خاص هر منطقه، این مراحل تخصیص توجه منطقه‌ای دقیق و مداوم را تضمین کرده و استحکام تشخیصی ماژول‌های توجه منطقه‌ای را افزایش می‌دهند.

۳-۲- توجه خودساخته سلسله‌مراتبی^{۱۲}

RAE-ViT یک سازوکار توجه خودساخته سلسله‌مراتبی را برای شبیه‌سازی الگوهای محلی و توزیعی آتروپی مغزی به کار می‌برد. توجه محلی که برای شناسایی تغییرات ظریف در هر منطقه طراحی شده به صورت زیر تعریف می‌شود:

$$A_{local}^{(r)} = \text{Softmax} \left(\frac{Q_{local}^{(r)} K_{local}^{(r)T}}{\sqrt{d_k}} \right) V_{local}^{(r)} \quad (۵)$$

این رابطه از سازوکار توجه استاندارد [۳] الهام گرفته شده اما به طور خاص برای تحلیل درون منطقه‌ای تطبیق یافته است.

توجه سراسری، که یک نوآوری این پژوهش است اطلاعات را در تمام نواحی مورد توجه جمع می‌کند:

$$A_{global} = \text{Softmax} \left(\frac{Q_{global} K_{global}^T}{\sqrt{d_k}} \right) V_{global} \quad (۶)$$

این رابطه برای مدل‌سازی وابستگی‌های بلند مدت بین نواحی مغزی توسعه یافته است.

سازوکار توجه ترکیبی که ابداع این پژوهش است، به صورت زیر تعریف می‌شود:

$$A = w_{local} \cdot A_{local} + w_{global} \cdot A_{global} \quad (۷)$$

می‌شود که از ترادیسگرهای تصویری اصلی اقتباس شده است:

$$A_r = \text{Softmax} \left(\frac{Q_r K_r^T}{\sqrt{d_k}} \right) V_r \quad (۲)$$

که در آن $Q_r, K_r, V_r \in R^{d_k \times d_v}$ ماتریس‌های پرس و جو، کلید و مقدار برای منطقه Φ_r هستند و d_k ابعاد بردار کلید است. این فرمول استاندارد برای توجه، در اینجا برای تمرکز بر نواحی خاص مغزی تطبیق داده شده است. توجه جمع شده در تمام نواحی مورد توجه، که نوآوری این پژوهش است به صورت زیر محاسبه می‌شود تا اهمیت تشخیصی هر منطقه را وزن‌دهی کند:

$$A_{ROI} = \sum_r w_r \cdot A \quad (۳)$$

که در آن w_r وزن‌های قابل آموزش هستند. این رابطه برای متعادل کردن مشارکت نواحی کلیدی مغز و افزایش تفسیرپذیری مدل طراحی شده است.

تقسیم‌بندی و پیش‌پردازش برای ماژول توجه منطقه‌ای داده‌های تصویربرداری رزونانس مغناطیسی از مراحل پیش‌پردازش زیر عبور می‌کنند تا ورودی با کیفیت بالا برای ماژول توجه منطقه‌ای فراهم شود:

- تقسیم‌بندی: ماسک‌های دوتایی $M_r(x, y, z)$ برای هر منطقه Φ_r با استفاده از FreeSurfer تولید می‌شوند. این مرحله مناطق مرتبط با بیماری را ایزوله کرده و مرزهای واضحی برای استخراج نواحی مورد توجه فراهم می‌آورد.
- نرمال‌سازی: مقادیر شدت $I(x, y, z)$ به $[0, 1]$ نرمال می‌شوند تا یکنواختی در سراسر اسکن‌ها تضمین شده و تغییرپذیری ناشی از پروتکل‌های مختلف تصویربرداری رزونانس مغناطیسی کاهش یابد.

- تقویت: تبدیلات خاص نواحی مورد توجه (مثل چرخش‌ها، تغییرات شدت) برای افزایش تنوع و بهبود تعمیم‌پذیری اعمال می‌شود:

که L_{CE} خطای آنتروپی متقاطع استاندارد و L_{aux} خطای کمکی ابداعی برای هم راستایی نقشه های توجه با دانش بالینی است.

۴-تنظیمات آزمایشی

۴-۱-مجموعه داده

این مطالعه از داده های تصویربرداری رزونانس مغناطیسی ساختاری مجموعه تصاویر ادنی استفاده می کند که شامل ۱/۱۵۲ اسکن است و به نسبت ۷۰ درصد (۸۰۶ اسکن) برای آموزش، ۱۵ درصد (۱۷۳ اسکن) برای اعتبار سنجی، و ۱۵ درصد (۱۷۳ اسکن) برای آزمون تقسیم شده است. همه مدل های پایه (Standard ViT، ResNet-50، DenseNet-121، Hybrid CNN-ViT) تحت RAE-ViT و شرایط یکسان آزمایش شدند تا مقایسه منصفانه ای در جدول ۲ ارائه شود.

جدول ۱: جدول جمعیتی تنوع موجود در مجموعه داده را نشان می دهد، شامل دامنه سنی وسیع، توزیع جنسیتی متوازن و امتیازات MMSE بالینی تایید شده در گروه ها

گروه	تعداد	جنسیت (مرد/زن)	سن (Mean) (±SD)	MMSE (Mean) (±SD)
بیمار آلزایمری	۲۵۵	۱۳۴/۱۲۱	۷۴/۵ ± ۶/۸	۲۲/۳ ± ۳/۱
اختلال شناختی خفیف	۵۲۱	۲۸۴/۲۳۷	۷۱/۸ ± ۷/۲	۲۷/۴ ± ۲/۱
افراد سالم	۳۷۶	۱۹۲/۱۸۴	۶۸/۶ ± ۶/۵	۲۹/۱ ± ۱/۰

اسکن های تصویربرداری رزونانس مغناطیسی با استفاده از اسکنرهای MRI 1.5T و 3T به دست آمده اند که تصویربرداری با وضوح بالا را برای تجزیه و تحلیل ساختاری دقیق فراهم می کنند. تصاویر خام دارای ابعاد $۱۷۶ \times ۲۵۶ \times ۲۵۶$ پیکسل با وضوح پیکسل ۱ میلی متر مربع هستند که اطمینان حاصل می کند که ساختار مغزی به طور دقیق شبیه سازی می شود.

۴-۲- پروتکل آموزش

پروتکل آموزش برای ترادیسگر تصویری تقویت

که در آن w_{local} و w_{global} وزن های قابل آموزش هستند که سهم توجه محلی و سراسری را تعدیل می بخشند. این رویکرد ترکیبی تضمین می کند که مدل از ویژگی های خاص هر منطقه و توزیعی استفاده می کند و درک جامعی از آسیب شناسی بیماری ارائه می دهد.

۳-۳- ادغام وصله چندمقیاسی

برای پردازش ویژگی ها در سطوح مختلف دقت، RAE-ViT از جاسازی وصله چندمقیاسی استفاده می کند. هر اسکن رزونانس مغناطیسی به وصله های همپوشان با اندازه های مختلف تقسیم می شود:

$$e_{i,j,k} = W_k P_{i,j,k} \quad (8)$$

که در آن $P_{i,j,k}$ پیچ در مقیاس (k) و W_k ماتریس وزن قابل آموزش است. این رابطه، که از ترادیسگرها تصویری اقتباس شده [۳]، برای استخراج ویژگی های چند تفکیکی تطبیق یافته است. نمایش نهایی چند تفکیکی به صورت زیر است که یک نوآوری این پژوهش است:

$$E = \text{contact}(e_{k1}, e_{k2}, \dots, e_{kn}) \quad (9)$$

این رابطه برای شبیه سازی هم زمان تغییرات محلی و الگوهای وسیع تر توسعه یافته است.

۳-۴- سر طبقه بندی^{۱۳}

احتمال طبقه بندی به صورت زیر محاسبه می شود:

$$P(y | x) = \text{Softmax}(W_{clsf}(A_{region}) + b_{cls}) \quad (10)$$

که در آن (f) تابع استخراج ویژگی است و از معیارهای استاندارد اقتباس شده است [۳]. هدف بهینه سازی مدل که ترکیبی از استاندارد و نوآوری است، به صورت زیر است

$$L = L_{CE} + \lambda L_{aux} \quad (11)$$

که در آن A_r نقشه توجه برای منطقه r است، M_r ماسک دوتایی برای منطقه r است که از تقسیم‌بندی FreeSurfer استخراج شده و KL واگرایی کولب-بلر ۱۴ است که شباهت بین M_r و A_r را تحمیل می‌کند.

تقویت داده‌ها

برای بهبود استحکام مدل و جلوگیری از بیش‌برازش، تکنیک‌های تقویت داده‌های گسترده‌ای در طول آموزش اعمال می‌شوند.

۱- تقویت سراسری:

- چرخش‌های تصادفی: زاویه‌های $\pm 10^\circ$ درجه به تصاویر اعمال می‌شود تا تغییرات موقعیتی شبیه‌سازی شود.
- مقیاس‌دهی: تصاویر به‌طور تصادفی $\pm 5\%$ مقیاس‌دهی می‌شوند تا تغییرات بزرگ‌نمایی تصویربرداری شبیه‌سازی شود.

- تغییرات شدت: شدت پیکسل‌ها به‌طور تصادفی $\pm 10\%$ تغییر می‌کند تا تفاوت‌های اسکنر شبیه‌سازی شود.

۲- تقویت خاص مناطق مورد نظر:

ماسک‌های دودویی برای مناطق مورد توجه تغییرات تصادفی هم‌نسبی (مثل چرخش‌ها و جابه‌جایی‌ها) را تجربه می‌کنند تا تضمین شود که مدل سازوکارهای توجه قوی را، صرف‌نظر از تغییرات هم‌راستایی آناتومیکی، بیاموزد. مجموعه داده تقویت‌شده D' به‌صورت زیر تولید می‌شود:

$$D' = \{T(I(x, y, z)), T(M_r(x, y, z)) \mid T \in \mathcal{T}\} \quad (15)$$

که در آن $\mathcal{T}$ تغییرات تقویتی از مجموعه $\mathcal{T}$ است، $I(x, y, z)$ اسکن اصلی تصویربرداری رزونانس مغناطیسی است و $M_r(x, y, z)$ ماسک دوتایی اصلی برای منطقه موردنظر r است.

بهینه‌سازی و منظم‌سازی

برای آموزش مدل از بهینه‌ساز آدام با برنامه زمان‌بندی نرخ یادگیری تطبیقی استفاده می‌شود. نرخ یادگیری اولیه به α تنظیم می‌شود و کاهش کسینوسی در طول اپوک‌های

شده با توجه منطقه‌ای (RAE-ViT) تضمین می‌کند که مدل الگوهای سراسری و منطقه‌ای آتروپی مغزی را که برای طبقه‌بندی بیماری آلزایمر حیاتی هستند، به‌طور مؤثر یاد می‌گیرد. این پروتکل ترکیبی از پیش‌آموزش، تنظیم دقیق و راهبردهای تقویت داده‌ها است که به تقویت استحکام و تعمیم‌پذیری مدل کمک می‌کند.

مرحله پیش‌آموزش

مدل ویژن ترانسفورمر بر روی مجموعه داده CheX-pert پیش‌آموزش داده می‌شود تا ویژگی‌های عمومی تصویربرداری پزشکی مانند لبه‌ها، بافت‌ها و ساختارهای آناتومیکی را بیاموزد. تابع هدف پیش‌آموزش به‌صورت زیر تعریف می‌شود:

$$L_{pretrain} = L_{CE}(P_{pretrain}, Y_{pretrain}) \quad (12)$$

که در آن L_{CE} تابع خطای آنتروپی متقاطع است، $P_{pretrain}$ احتمالات پیش‌بینی شده در طول پیش‌آموزش است و $Y_{pretrain}$ برچسب‌های حقیقت زمینه برای مجموعه داده پیش‌آموزش است.

مرحله تنظیم دقیق

پس از پیش‌آموزش، مدل بر روی مجموعه داده ادنی تنظیم دقیق می‌شود تا به‌طور خاص در طبقه‌بندی بیماری آلزایمر تخصص یابد. تابع خطا در طول تنظیم دقیق، ترکیبی وزنی از اهداف سراسری و منطقه‌ای است:

$$L_{fine-tune} = L_{CE} + \lambda L_{ROI} \quad (13)$$

که در آن L_{CE} خطای آنتروپی متقاطع برای دقت کلی طبقه‌بندی است، L_{ROI} خطای خاص منطقه‌ای است که اطمینان می‌دهد نقشه‌های توجه با پیش‌زمینه‌های آناتومیکی هم‌راستا شوند و λ یک هاپرپارامتر است که مشارکت‌های L_{CE} و L_{ROI} را متعادل می‌کند. خطای خاص منطقه‌ای L_{ROI} به‌صورت زیر محاسبه می‌شود:

$$L_{ROI} = \sum_r KL(A_r, M_r) \quad (14)$$

جدول ۲: معیارهای عملکرد RAE-ViT در مقایسه با مدل‌های پایه

مدل	دقت (%)	حساسیت (%)	ویژگی (%)	مساحت زیر منحنی
RAE-ViT	۹۴/۲	۹۱/۸	۹۵/۷	۰/۹۶
Standard ViT	۸۹/۵	۸۶/۳	۹۰/۲	۰/۹۱
ResNet-50	۸۷/۸	۸۴/۷	۸۸/۹	۰/۸۹
DenseNet-121	۸۸/۱	۸۵/۱	۸۹/۳	۰/۹۰
Hybrid CNN-ViT	۹۰/۲	۸۷/۵	۹۱/۰	۰/۹۲

* توضیح: اعداد این جدول از آزمایش مجدد همه مدل‌ها توسط پژوهشگران این مطالعه روی مجموعه داده ادنی (۱۱۵۲ اسکن) با شرایط یکسان (پیش‌پردازش، تقسیم‌بندی داده‌ها به نسبت ۱۵/۱۵/۷۰ برای آموزش/ اعتبارسنجی/ آزمون، و پروتکل آموزش یکسان) به دست آمده‌اند و از مقالات دیگر نقل نشده‌اند. جزئیات بیشتر در بخش ۴ ارائه شده است.*

RAE-ViT دقت ۹۴/۲ درصد را به دست آورد و از ترادیسگر تصویری استاندارد (۸۹/۵ درصد)، ResNet-50 (۸۸/۸ درصد)، DenseNet-121 (۸۸/۱ درصد) و مدل ترکیبی CNN-ViT (۹۰/۲ درصد) پیشی گرفت. ادغام توجه منطقه‌ای و توجه خودساخته سلسله‌مراتبی به‌طور قابل‌توجهی عملکرد طبقه‌بندی را بهبود بخشید. این بهبود حدود ۴ درصد در دقت (از ۹۰/۲ درصد به ۹۴/۲ درصد) و AUC (از ۰/۹۲ به ۰/۹۶) در مقایسه با بهترین روش موجود (Hybrid CNN-ViT)، همراه با افزایش حساسیت از ۸۷/۵ درصد به ۹۱/۸ درصد و ویژگی از ۹۱/۰ درصد به ۹۵/۷ درصد، در تشخیص زودهنگام آلزایمر اهمیت بالینی دارد، زیرا شناسایی دقیق‌تر بیماران و کاهش مثبت‌های کاذب می‌تواند به مداخلات به‌موقع و کاهش هزینه‌های تشخیصی غیرضروری منجر شود. با حساسیت ۹۱/۸ درصد، RAE-ViT در شناسایی موارد آلزایمری برتری دارد که برای تشخیص زودهنگام حیاتی است، و ویژگی ۹۵/۷ درصد آن شناسایی دقیق افراد سالم را تضمین می‌کند و از مثبت‌های کاذب جلوگیری می‌کند. مقدار مساحت زیر منحنی برابر با ۰/۹۶ قدرت تفکیک قوی مدل را نشان می‌دهد و آن را به ابزاری بسیار قابل‌اعتماد برای تشخیص‌های پزشکی تبدیل

آموزشی اعمال می‌شود:

$$\alpha_t = \alpha \cdot \frac{1}{2} \left(1 + \cos \left(\frac{t}{T} \pi \right) \right) \quad (16)$$

که در آن α_t نرخ یادگیری در اپوک t بوده و T تعداد کل اپوک‌های آموزشی است. برای جلوگیری از بیش‌برازش، منظم‌سازی کاهش وزن (γ) به پارامترهای مدل اعمال می‌شود:

$$L_{reg} = \gamma \sum_w w^2 \quad (17)$$

که در آن γ ضریب کاهش وزن است و w نمایانگر پارامترها (وزن‌ها)ی مدل است. هدف نهایی آموزش، ترکیبی از ضررهای طبقه‌بندی، خاص منطقه‌ای و منظم‌سازی است:

$$L_{total} = L_{fine-tune} + L_{reg} \quad (18)$$

۵- نتایج آزمایش‌ها

روش RAE-ViT بر روی مجموعه داده ادنی (۱۱۵۲ اسکن تصویربرداری رزونانس مغناطیسی) در دسته‌های بیمار آلزایمری، اختلال شناختی خفیف و افراد سالم ارزیابی شد. این مدل از شبکه‌های عصبی هم‌آمیختگی و ویژن ترانسفورمرهای استاندارد پیشی گرفت و دقت، حساسیت، ویژگی و مساحت زیر منحنی برتر خود را نشان داد، که قابلیت‌های تشخیصی قوی آن را برای طبقه‌بندی آلزایمر تأیید می‌کند.

۵-۱- معیارهای عملکرد کلی

RAE-ViT عملکرد پیشرفته‌ای را به نمایش گذاشت و دقت، حساسیت و ویژگی بالایی را در طبقه‌بندی افراد مبتلا به آلزایمر، اختلال شناختی خفیف و افراد سالم کسب کرد. همان‌طور که در جدول ۲ خلاصه شده است مقدار مساحت زیر منحنی آن برابر با ۰/۹۶ است که قدرت تفکیک قوی و اطمینان بالای آن در تفکیک دسته‌های تشخیصی را نشان می‌دهد.

می‌کند.

مدل را اندازه‌گیری می‌کند و مقادیر بالاتر نشان‌دهنده توانایی بهتر در تمایز است. منحنی RAE-ViT (خط آبی) نزدیک‌ترین موقعیت را به گوشه بالا-چپ نمودار دارد و حساسیت بالای ۹۱/۸ درصد و نرخ مثبت کاذب پایین ۴/۳ درصد را در تمام آستانه‌ها نشان می‌دهد. مساحت زیر منحنی برابر با ۰/۹۶ توانایی استثنایی مدل را در تمایز بین بیماری آلزایمر، اختلال شناختی خفیف و افراد سالم نشان می‌دهد و قدرت آن را در شناسایی الگوهای از دست رفتن مغز که برای تشخیص بیماری ضروری هستند، به نمایش می‌گذارد. ویژن ترانسفورمر استاندارد (AUC=0.91): منحنی (سبز) عقب‌تر از RAE-ViT قرار دارد که نشان‌دهنده کاهش حساسیت و ویژگی است. عدم وجود سازوکارهای توجه منطقه‌ای احتمالاً توانایی آن را در تمرکز بر مناطق بحرانی بیماری مانند هیپوکامپ محدود می‌کند: DenseNet-121 (AUC = 0.90) و ResNet-50 (AUC = 0.89) به دلیل توانایی محدودشان در ضبط وابستگی‌های بلندمدت در داده‌های تصویربرداری رزونانس مغناطیسی، عملکرد ضعیف‌تری دارند که منجر به نرخ اشتباهات بالاتر در طبقه‌بندی نشانگرهای زیستی ظریف بیماری آلزایمر می‌شود. منحنی‌های ROC آن‌ها زودتر به یک سطح ثابت می‌رسند، که این محدودیت را منعکس می‌کند. ترکیب شبکه‌های عصبی هم‌آمیختی خالص دارد، اما همچنان از RAE-ViT عقب است و نشان‌دهنده محدودیت‌های ترکیب لایه‌های هم‌آمیختی و ترادیسگر بدون سازوکارهای توجه خاص حوزه است.

نمودارهای بعدی همان توالی را برای مدل‌های استفاده شده نشان می‌دهند.

۶- نتیجه‌گیری و کارهای آینده

این مطالعه مدل ترادیسگر تصویری تقویت شده با توجه منطقه‌ای (RAE-ViT) را برای طبقه‌بندی بیماری آلزایمر با استفاده از داده‌های تصویربرداری رزونانس

۵-۲- معیارهای عملکرد به تفکیک کلاس

برای تحلیل بیشتر عملکرد مدل، معیارها به تفکیک کلاس برای هر دسته تشخیصی بیمار، اختلال شناختی خفیف و سالم محاسبه شدند. این معیارها درک دقیق‌تری از قوت‌ها و ضعف‌های مدل در گروه‌های مختلف تشخیصی فراهم می‌آورند. این معیارها در جدول ۳ نشان داده شده است.

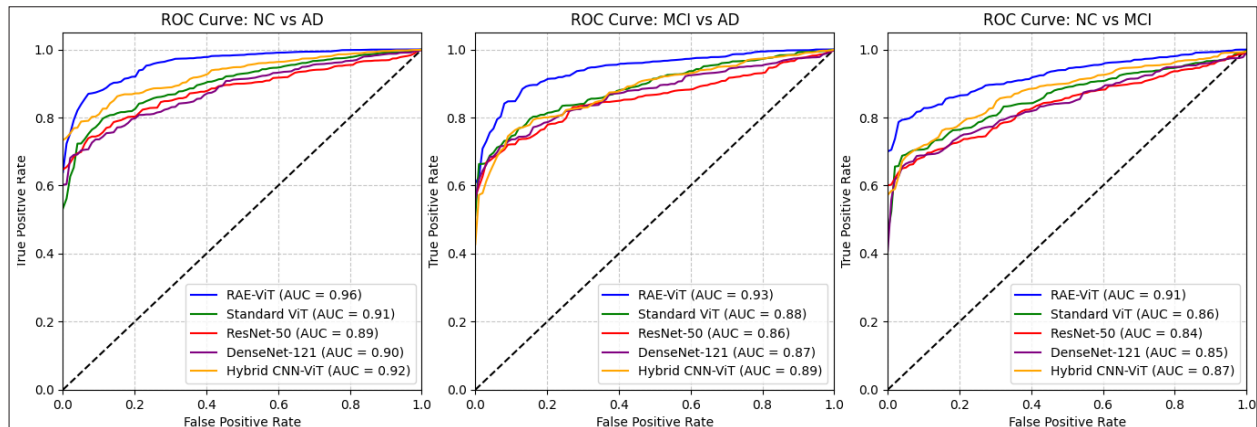
جدول ۳: معیارهای عملکرد به تفکیک کلاس برای RAE-ViT

گروه‌های تشخیصی	حساسیت (%)	ویژگی (%)	دقت (%)	F1-Score (%)
بیمار آلزایمری	۹۱/۸	۹۵/۷	۹۲/۵	۹۲/۱
اختلال شناختی خفیف	۸۹/۵	۹۳/۲	۹۰/۱	۹۸/۸
افراد سالم	۹۴/۳	۹۶/۱	۹۴/۸	۹۴/۵

همان‌طور که در جدول ۲ نشان داده شده است، RAE-ViT عملکرد بالایی در طبقه‌بندی آلزایمر با حساسیت ۹۱/۸ درصد، ویژگی ۹۵/۷ درصد و F1-Score برابر با ۹۲/۱ درصد نشان داد و مثبت‌های کاذب را به حداقل رساند. برای طبقه‌بندی اختلال شناختی خفیف، حساسیت ۸۹/۵ درصد، ویژگی ۹۳/۲ درصد و امتیاز F1 برابر با ۹۸/۸ درصد را به دست آورد و نشان داد که توانایی قوی در شناسایی اختلالات شناختی در مراحل اولیه دارد. این مدل در شناسایی افراد سالم با حساسیت ۹۴/۳ درصد، ویژگی ۹۶/۱ درصد و F1-Score برابر با ۹۴/۵ درصد برتری داشت که قدرت آن را در تمایز بین اختلالات شناختی و افراد سالم تأیید می‌کند.

۵-۳- منحنی ویژگی عملکرد گیرنده (ROC)

شکل ۳ منحنی‌های ROC برای مدل پیشنهادی و مدل‌های پایه (ویژن ترانسفورمر، DenseNet-121، ResNet-50 و Hybrid CNN-ViT) را نشان می‌دهد. منحنی ROC تفاوت بین نرخ مثبت واقعی و نرخ مثبت کاذب را در آستانه‌های مختلف نشان می‌دهد. مساحت زیر منحنی عملکرد کلی



شکل ۳: منحنی‌های ROC برای RAE-ViT و مدل‌های پایه

تصویربرداری رزونانس مغناطیسی عملکردی و ژنتیک) برای بهبود دقت تشخیص و یادگیری خودنظارتی برای تعمیم‌پذیری بهتر با داده‌های برچسب‌خورده محدود تمرکز کنند. یادگیری مشارکتی می‌تواند آموزش با حفظ حریم خصوصی را در میان موسسات تحقیقاتی امکان‌پذیر سازد. همچنین تکنیک‌های پیشرفته توضیح‌پذیری برای ساختن اعتماد در تشخیص‌های مبتنی بر هوش مصنوعی بسیار حیاتی هستند. RAE-ViT گامی بزرگ در طبقه‌بندی خودکار بیماری آلزایمر است و پتانسیل بهبود قابل‌توجه تشخیص زودهنگام و مدیریت این بیماری را دارد، که نتیجه آن بهبود نتایج بیماران و پیشرفت در تشخیص‌های مبتنی بر تصویربرداری عصبی است.

مراجع

1. Jack, Clifford R. 2022. "Advances in Alzheimer's disease research over the past two decades." *The Lancet Neurology* 21.10: 866-869.
2. Malik, I., Iqbal, A., Gu YH, Al-Antari, MA. 2024. Deep Learning for Alzheimer's Disease Prediction: A Comprehensive Review. *Diagnostics* (Basel). 2024 Jun 17; 14(12): 1281. doi: 10.3390/diagnostics14121281. PMID: 38928696; PMCID: PMC11202897.
3. Dosovitskiy, A. 2020. "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv: 2010.11929*.
4. Poudel, Sahadev, and Sang-Woong Lee. 2021. "Deep multi-scale attentional features for medical image segmentation." *Applied Soft Computing* 109 (2021): 107445.
5. Moamen, Z. et al. 2025. "Fusing CNNs and attention-mechanisms to improve real-time indoor Human Activity Recognition for classifying home-based physical reha-

مغناطیسی معرفی می‌کند. با ترکیب توجه منطقه‌ای، توجه خود-مقیاس‌پذیر و استخراج ویژگی‌های چندمقیاسی، RAE-ViT عملکردی پیشرفته را در مجموعه داده ادنی به دست می‌آورد و در مقایسه با مدل‌های مبتنی بر شبکه‌های عصبی هم‌آمیزی و ویژن ترانسفورمر استاندارد در شبیه‌سازی الگوهای آتروفی مغز محلی و سراسری بهتر عمل می‌کند. نقشه‌های توجه آن با دانش بالینی هم‌راستا هستند و از اعتبار آن در کمک به تشخیص این بیماری پشتیبانی می‌کنند. با این حال، چالش‌هایی باقی‌مانده است. نیازهای محاسباتی بالای ترادیسگرهای تصویری، از جمله RAE-ViT، که به دلیل استفاده از توجه منطقه‌ای و چندمقیاسی پیچیدگی بیشتری نسبت به Hybrid CNN-ViT (با حدود ۲۰ درصد پارامتر کمتر) دارد، مانع از استفاده آن در زمان واقعی در محیط‌های بالینی می‌شود. با این حال، بهبود ۴ درصد در دقت و AUC، به‌ویژه در تشخیص زودهنگام، ارزش این پیچیدگی را در کاربردهای بالینی توجیه می‌کند، زیرا تشخیص دقیق‌تر می‌تواند هزینه‌های بلندمدت درمان را کاهش دهد. برای رفع این چالش، پیشنهاد می‌کنیم در کارهای آینده از تکنیک‌هایی مانند کاهش ابعاد وصله‌ها (از 16×16 به 8×8)، هرس مدل، یا استفاده از معماری‌های سبک‌تر مثل MobileViT استفاده شود تا هزینه محاسباتی کاهش یابد، در حالی که عملکرد بالا حفظ شود. تحقیقات آینده باید بر روی یکپارچه‌سازی داده‌های چندمدالی (مانند توموگرافی گسیل پوزیترون،

- bilitation exercises.” *Computers in Biology and Medicine* 184: 109399.
6. Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Trojanowski, J. Q. 2017. The Alzheimer’s Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer’s & Dementia*, 13(6), e1-e85. DOI: 10.1016/j.jalz.2016.11.007
 7. Jomeiri, A., Habibizad Navin, A., & Shamsi, M. 2024. “Longitudinal MRI analysis using a hybrid DenseNet-BiLSTM method for Alzheimer’s disease prediction.” *Behavioural Brain Research* 463: 114900.
 8. Emerald, H., Onyeka Emebob, U., & Asotie Omonhinmin. C. 2022. “Vision transformers in medical imaging: A review.” *arXiv preprint arXiv:2211.10043*.
 9. Kunanbayev, Kassymzhomart, Vyacheslav Shen, and Dae-Shik Kim. 2024. “Training ViT with Limited Data for Alzheimer’s Disease Classification: An Empirical Study.” *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland.
 10. Muthalakshmi, M., et al. 2024. “Federated Learning for Secure and Privacy-Preserving Medical Image Analysis in Decentralized Healthcare Systems.” *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE.
 11. Shamshad, F. et al. 2023. “Transformers in medical imaging: A survey.” *Medical Image Analysis* 88: 102802.
 12. Gonçalves, T. et al. 2022. “A survey on attention mechanisms for medical applications: are we moving toward better Algorithms?.” *IEEE Access* 10: 98909-98935.
 13. Zhang, Y., Yu, X., Hu, Q., Zhang, X., Yang, Y., & Xiao, H. 2024. An attention mechanism-based lightweight UNet for musculoskeletal ultrasound image segmentation. *Med Phys*. 2025 Jan; 52(1):400-413. doi: 10.1002/mp.17503. Epub 2024 Dec 2. PMID: 39620487.
 14. Shaffi, N., Viswan, V. & Mahmud, M. 2024. Ensemble of vision transformer architectures for efficient Alzheimer’s Disease classification. *Brain Inf.* 11, 25. <https://doi.org/10.1186/s40708-024-00238-7>.
 15. Xiaopeng, K., & Liu, Y. 2023. “3DNesT: A Hierarchical Local Self-Attention Model for Alzheimer’s Disease Diagnosis.” *2023 International Conference on Neuromorphic Computing (ICNC)*. IEEE.
 16. Schwenke, Leonid, & Atzmueller, M. 2023. “Extracting Interpretable Local and Global Representations from Attention on Time Series.” *arXiv preprint arXiv:2312.11466*.
 17. Qiu, Shangran, et al. “Development and validation of an interpretable deep learning framework for Alzheimer’s disease classification.” *Brain* 143.6 (2020): 1920-1933.
 18. Lozupone, G. et al. 2024. “AXIAL: Attention-based eXplainability for Interpretable Alzheimer’s Localized Diagnosis using 2D CNNs on 3D MRI brain scans.” *arXiv preprint arXiv:2407.02418*.