Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال چهارم، شماره اول، بهار و تابستان ۱۴۰۵، صفحات ۱۱۱ – ۹۵

ناحیه‌بندی تومورهای مغزی با استفاده از رمزگذار ترنسفورمر و ماژول‌های توجه انطباقی

نور عصام عبدالنبی[۱]، منصور فاتح[۲*]، سعیده فردوسی[۳]

[۱] دانشجوی دکتری دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران

[۲] دانشیار دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شاهرود، شاهرود، ایران

[۳] استادیار دانشکده ریاضی، آمار و علوم آماری، دانشگاه اسکس، کلچستر، انگلستان

مقاله پژوهشی

تاریخ دریافت:
۱۴۰۴/۰۹/۰۴

تاریخ پذیرش:
۱۴۰۴/۱۲/۰۵

کلیدواژه‌ها:
ناحیه‌بندی، مکانیزم توجه، یادگیری عمیق، تومور مغزی، MRI

نویسنده‌مسئول:
Mansoor_fateh@shahroodut.ac.ir

چکیده

ناحیه‌بندی خودکار و دقیق تومورهای مغزی از تصاویر تشدید مغناطیسی، نقشی حیاتی در تشخیص، برنامه‌ریزی درمان و نظارت بر بیماری دارد. معماری‌های مبتنی بر CNN در استخراج ویژگی‌های محلی توانمند هستند اما در درک زمینه سراسری تصویر با محدودیت مواجه‌اند. مدل‌های ترنسفورمر در مدل‌سازی وابستگی‌های دوربرد و سراسری برتری دارند و مشکلات مدل‌های مبتنی بر CNN را در این زمینه رفع می‌کنند. در این مقاله، ما یک معماری ترکیبی نوین U–شکل پیشنهاد می‌کنیم که از نقاط قوت هر دو رویکرد بهره می‌برد. مدل ما از یک ستون فقرات مبتنی بر Swin–Transformer پیش‌آموخته به عنوان رمزگذار برای استخراج ویژگی‌های سلسله‌مراتبی و غنی از زمینه سراسری استفاده می‌کند. نوآوری اصلی این معماری، معرفی دو ماژول توجه فضایی پیشرفته برای پالایش و تطبیق ویژگی‌های استخراج‌شده از رمزگذار با دامنه پزشکی و ماژول افزایش مقیاس به کمک توجه کانالی در رمزگشا برای باز–وزن‌دهی انطباقی و بهینه اطلاعات دریافتی از اتصالات پرشی است. ارزیابی‌های انجام‌شده بر روی مجموعه داده چالش‌برانگیز BRISC نشان داد که روش پیشنهادی ما با دستیابی به امتیاز ۸۰/۶٪ در IoU وزنی و ۸۸/۶٪ در Dice، از مدل‌های پیشرفته پیشین عملکرد بهتری داشته و کارایی ترکیب ترنسفورمر با مکانیزم‌های توجه دوگانه را اثبات می‌کند.

doi : 10.22034/ABMIR.2026.24013.1189

E-ISSN: 2821-2037

۱- مقدمه

تومور مغزی به رشد غیرطبیعی و تکثیر سلول‌ها در ساختار مغز اطلاق می‌شود. این توده‌ها می‌توانند خوش‌خیم (غیر سرطانی) یا بدخیم (سرطانی) باشند، که در این میان، تومورهای بدخیم مانند گلیوبلاستوما[1] به‌شدت تهاجمی شناخته می‌شوند [۱]. در سطح جهانی، تومورهای مغزی بار بهداشتی قابل‌توجهی را تحمیل می‌کنند. به‌عنوان‌مثال، گزارش‌ها حاکی از آن است که چالش اصلی در مدیریت این بیماری، تشخیص دیرهنگام آن است؛ درصد قابل‌توجهی از سرطان‌های مغز تنها در مراجعات اورژانسی شناسایی می‌شوند و بسیاری از بیماران برای رسیدن به تشخیص قطعی، نیازمند مشاوره‌های متعدد هستند [۲].

تصویربرداری تشدید مغناطیسی (MRI[2]) به دلیل ارائه کنتراست خوب در بافت‌های نرم و نمایش اطلاعات دقیق آسیب‌شناسی و آناتومیک در مُدالیته‌های مختلف (مانند T1-weighted، T2-weighted و FLAIR)، به عنوان تصویربرداری استاندارد شناخته می‌شود [۳, ٤]. رویکرد مرسوم برای برنامه‌ریزی درمان، تعیین دستی مرزهای تومور در این اسکن‌های حجیم است. بااین‌حال، این فرآیند به‌شدت زمان‌بر، وابسته به مهارت اپراتور و مستعد ناهماهنگی‌های قابل‌توجه بین ارزیاب‌های مختلف است [۵]. درنتیجه، نیاز مبرمی به توسعه سیستم‌های ناحیه‌بندی[3] خودکار، دقیق و قابل‌اعتماد احساس می‌شود تا کارایی تشخیص، برنامه‌ریزی جراحی و نظارت بر پاسخ درمانی به‌طور مؤثری بهبود یابد [٦].

ظهور یادگیری عمیق، انقلابی در حوزه تحلیل تصاویر پزشکی ایجاد کرده است و مدل‌های ناحیه‌بندی خودکار، عملکردی در سطح یا حتی فراتر از استانداردهای بالینی از خود نشان داده‌اند [۷]. در ابتدا، شبکه‌های عصبی پیچشی (CNNs[4]) به الگو غالب تبدیل شدند [۸]. معماری‌های مبتنی بر مدل U-Net که با ساختار رمزگذار-رمزگشا و اتصالات پرشی[5] شناخته می‌شوند، به‌طور گسترده‌ای استفاده شده‌اند و کارایی بالایی در وظایف مختلف ناحیه‌بندی تصاویر پزشکی، از جمله تومورهای مغزی، داشتند [۹].

اخیراً، مدل‌های ترنسفورمر[6] که از مکانیزم‌های توجه به خود[7] برای درک وابستگی‌های سراسری[8] بهره می‌برند، به نتایج پیشرفته‌ای در پردازش زبان طبیعی و بینایی کامپیوتر دست‌یافته‌اند [۱۰]. این موفقیت، اقتباس آن‌ها را برای تصویربرداری پزشکی تسریع کرده است و مدل‌هایی مانند Swin-Transformer، ظرفیت قدرتمندی برای مدل‌سازی اطلاعات زمینه‌ای دوربرد در نقشه‌های ویژگی سلسله‌مراتبی از خود نشان داده‌اند [۱۱].

با وجود موفقیت‌های هر یک، هر دو دسته از معماری‌ها دارای محدودیت‌های ذاتی هستند. CNN ها به واسطه کرنل‌های پیچشی خود، در استخراج ویژگی‌های فضایی محلی[9] برتری دارند، اما در مدل‌سازی وابستگی‌های فضایی دوربرد و زمینه سراسری تصویر، با محدودیت ساختاری مواجه هستند [۱۲]. این سوگیری محلی می‌تواند هنگام ناحیه‌بندی تومورهای بزرگ و ناهمگن با مرزهای نامشخص، زیان‌بار باشد [۱۲]. در مقابل، مدل‌های ترنسفورمر اغلب از نظر بازدهی داده ضعیف‌تر عمل می‌کنند و ممکن است برای دستیابی به عملکرد بهینه، نیازمند پیش‌آموزش گسترده بر روی مجموعه داده‌های بسیار بزرگ باشند که این موضوع در حوزه پزشکی (که اغلب با کمبود داده مواجه است) یک چالش محسوب می‌شود [۱۳]. بنابراین، توسعه معماری‌های ترکیبی که به‌طور هم‌افزا، قدرت استخراج ویژگی محلی CNN ها را با توانایی مدل‌سازی زمینه سراسری ترنسفورمرها ترکیب کنند، همچنان یک حوزه تحقیقاتی حیاتی و بسیار فعال باقی‌مانده است. اگرچه معماری‌های پیشگامی نظیر TransUNet و Swin-UNet مسیر استفاده از ترنسفورمرها در ناحیه‌بندی پزشکی را هموار کردند، اما هر یک با چالش‌های ساختاری روبرو هستند. مدل‌های مبتنی بر CNN بهبودیافته مانند Attention U-Net، اگرچه با استفاده از دروازه‌های توجه سعی در تمرکز بر نواحی مهم دارند، اما همچنان به دلیل ماهیت کانولوشنی خود، دارای میدان دید محدود هستند و نمی‌توانند وابستگی‌های سراسری بین پیکسل‌های دور از هم را به

[1] Glioblastoma
[2] Magnetic Resonance Imaging
[3] Segmentation
[4] Convolutional Neural Networks
[5] skip Connection
[6] Transformer
[7] Self-attention
[8] Global Dependencies
[9] Local

خوبی مدل‌سازی کنند .از سوی دیگر، مدل TransUNet از ترنسفورمر تنها در گلوگاه استفاده می‌کند، که باعث می‌شود ویژگی‌های سطح پایین در مراحل ابتدایی رمزگذار از مزایای مدل‌سازی سراسری بهره‌مند نشوند. از سوی دیگر، Swin-UNet که ساختاری تمام–ترنسفورمر دارد، اگرچه در درک زمینه سراسری عالی عمل می‌کند، اما به دلیل حذف کامل کانولوشن‌ها، در استخراج ویژگی‌های فرکانس بالا (مانند لبه‌ها و بافت‌های ظریف تومور) که نقطه قوت ذاتی CNN ها است، دچار ضعف می‌شود [۱۳].

برای پرداختن به این چالش‌ها، ما یک معماری مبتنی بر UNet را پیشنهاد می‌کنیم. این شبکه ترکیبی رمزگذار–رمزگشا، به‌طور هوشمندانه از نقاط قوت مکمل ترنسفورمرها و مکانیزم‌های توجه در یک چارچوب یکپارچه بهره می‌برد. این معماری از یک Swin-Transformer از پیش‌آموخته به عنوان ستون فقرات رمزگذار[1] استفاده می‌کند که نمایش‌های ویژگی[2] غنی و سلسله‌مراتبی تولید می‌کند. ما یک ماژول نوآورانه به نام توجه فضایی پیشرفته را معرفی می‌کنیم. علاوه بر این، مسیر رمزگشا با بلوک‌های توجه کانالی تقویت شده است. این بلوک‌ها بازتنظیم ویژگی به‌صورت کانال–محور را در طول فرآیند نمونه‌برداری افزایشی[3] انجام می‌دهند تا به‌طور انتخابی، ویژگی‌های آموزنده‌تر دریافت شده از اتصالات پرشی را تقویت کنند. برخلاف TransUNet، ما ستون فقرات ترنسفورمر را در تمام مراحل رمزگذار گسترش داده‌ایم تا سلسله‌مراتب معنایی غنی‌تری ایجاد شود. همچنین برخلاف Swin-UNet، ما به جای حذف کامل کانولوشن، از ماژول‌های توجه فضایی چند–مقیاسه مبتنی بر CNN استفاده کرده‌ایم تا ضعف ترنسفورمر در ناحیه‌بندی مرزهای دقیق تومور را جبران کنیم. این طراحی ترکیبی، عملاً شکاف معنایی[4] میان ویژگی‌های سراسری ترنسفورمر و ویژگی‌های محلی مورد نیاز برای ناحیه‌بندی دقیق را پر می‌کند.

خلاصه نوآوری‌های ما به شرح زیر است:

طراحی یک معماری ناحیه‌بندی ترکیبی نوین که از یک Swin-Transformer به عنوان رمزگذار استفاده می‌کند.

معرفی یک ماژول توجه فضایی پیشرفته که بر روی ویژگی‌های چند مقیاسی رمزگذار اعمال می‌شود تا پالایش ویژگی‌های فضایی را تقویت کند.

استفاده از توجه کانالی در داخل رمزگشا برای انجام بازتنظیم انطباقی و کانال–محور ویژگی‌ها که منجر به بهینه‌سازی ترکیب ویژگی‌های چند مقیاسی می‌شود.

۲- کارهای مرتبط

ناحیه‌بندی خودکار تومورهای مغزی با استفاده از تصاویر MRI، در طول دهه‌های گذشته همواره به عنوان یکی از حوزه‌های تحقیقاتی برجسته مطرح بوده است. رویکردها در این زمینه تکامل چشمگیری را تجربه کرده‌اند و از تکنیک‌های کلاسیک پردازش تصویر به معماری‌های پیشرفته یادگیری عمیق امروزی رسیده‌اند. در این بخش، این سیر تحول را در چهار حوزه کلیدی بررسی می‌کنیم که مشتمل‌بر روش‌های سنتی و یادگیری ماشین، مدل‌های بنیادین یادگیری عمیق، رویکردهای مبتنی بر یادگیری انتقالی و شبکه‌های مبتنی بر ترنسفورمر و توجه هستند.

۲–۱ روش‌های سنتی و یادگیری ماشین

پیش از دوران یادگیری عمیق، ناحیه‌بندی تومور مغزی بر پردازش تصویر سنتی و مدل‌های کلاسیک یادگیری ماشین متکی بود. رویکردهای اولیه بر پایه شدت روشنایی پیکسل‌ها استوار بودند، نظیر روش‌های آستانه‌گذاری[5] و رشد ناحیه‌ای[6] [۱۴]. این روش‌ها با گروه‌بندی پیکسل‌ها بر اساس مقادیر روشنایی، تصویر را ناحیه‌بندی می‌کنند. بااین‌حال، این روش‌ها به نویز تصویر، عدم یکنواختی شدت روشنایی (که در MRI رایج است) و کنتراست ضعیف در مرزهای تومور بسیار حساس هستند که این عوامل آن‌ها را برای وظایف ناحیه‌بندی پیچیده غیرقابل‌اعتماد می‌سازد [۱۵].

مدل‌های بدون نظارت مانند خوشه‌بندی K-Means و Fuzzy C-Means (FCM) برای گروه‌بندی پیکسل‌ها بر اساس ویژگی‌های

[1]Encoder Backbone
[2]Feature Representations
[3]Upsampling
[4]Semantic Gap
[5]Thresholding
[6]Region-growing

شدت روشنایی استفاده شدند [۱۶]. بااین‌حال، این روش‌ها فاقد درک فضایی بودند. برای گنجاندن اطلاعات همسایگی، مدل‌های پیچیده‌تری مانند میدان‌های تصادفی مارکوف (MRFs[1]) توسعه یافتند که روابط فضایی بین پیکسل‌ها را مدل‌سازی می‌کردند [۱۷]. در حوزه یادگیری با نظارت، الگوریتم‌هایی مانند ماشین‌های بردار پشتیبان (SVMs[2]) و به‌ویژه جنگل‌های تصادفی[3] به دلیل توانایی در مدیریت داده‌های با ابعاد بالا، محبوبیت یافتند [۱۸]. نقطه‌ضعف اصلی و مشترک تمام این روش‌ها، اتکای مطلق آن‌ها به فرآیند مهندسی دستی ویژگی[4] بود [۱۹]. در این فرآیند، متخصصان می‌بایست به‌صورت دستی ویژگی‌هایی مانند بافت[5]، شکل[6] و گرادیان‌ها را استخراج کنند. این وابستگی، فرآیند را زمان‌بر، سلیقه‌ای و به‌شدت وابسته به تخصص می‌کرد و تعمیم‌پذیری مدل‌ها را محدود می‌ساخت.

۲-۲ روش‌های یادگیری عمیق

ظهور یادگیری عمیق، مدل و الگو را به‌طور کل تغییر داد. شبکه‌های عصبی پیچشی با قابلیت یادگیری خودکار ویژگی‌های سلسله‌مراتبی مستقیماً از داده‌های خام، نیاز به مهندسی دستی ویژگی را از بین بردند. شبکه عصبی پیچشی به عنوان یکی از اولین مدل‌ها، نشان داد که می‌توان از CNN ها برای تولید نقشه‌های پیش‌بینی پیکسلی[7] برای ناحیه‌بندی معنایی استفاده کرد [۲۰].

بااین‌حال، معماری که به استاندارد طلایی در ناحیه‌بندی تصاویر پزشکی تبدیل شد، U-Net بود [۲۱]. U-Net یک معماری متقارن رمزگذار-رمزگشا را با یک نوآوری کلیدی به نام اتصالات پرشی ارائه داد. این اتصالات، ویژگی‌های با رزولوشن بالا از مسیر رمزگذار را با ویژگی‌های معنایی عمیق از مسیر رمزگشا ترکیب می‌کنند. این امر به U-Net اجازه می‌دهد تا هم اطلاعات مکانی دقیق و هم اطلاعات معنایی را حفظ کند و منجر به مرزهای بسیار دقیق در ناحیه‌بندی شود.

با توجه به ماهیت سه‌بعدی داده‌هایMRI ، معماری U-Net به سرعت به حوزه سه‌بعدی گسترش یافت و مدل‌هایی مانند 3DU-Net و V-Net [۲۲, ۲۳] که از اتصالات باقیمانده[8] نیز بهره می‌برد، توسعه یافتند. این مدل‌ها به جای پردازش برش به برش، کل حجم[9] را تحلیل می‌کنند و زمینه فضایی کامل بین برش‌ها را درک می‌نمایند. در همین راستا، معماری‌های دیگری مانند DeepLab [۲۴] با استفاده از پیچش‌های اتساعی[10] تلاش کردند تا میدان دریافتی[11] را بدون کاهش رزولوشن فضایی افزایش دهند و زمینه چند مقیاسی را ثبت کنند.

۲-۳ یادگیری انتقالی در ناحیه‌بندی

اگرچه مدل‌های بنیادین مانند U-Net قدرتمند هستند، رمزگذارهای آن‌ها معمولاً از ابتدا بر روی مجموعه داده پزشکی هدف آموزش می‌بینند. بااین‌حال، مجموعه داده‌های تصویربرداری پزشکی اغلب به دلیل هزینه و زمان بالای مورد نیاز برای حاشیه‌نویسی تخصصی، کوچک هستند. آموزش شبکه‌های بسیار عمیق بر روی داده‌های محدود می‌تواند منجر به بیش‌برازش[12] و تعمیم‌پذیری ضعیف شود.

یادگیری انتقالی[13] به عنوان یک استراتژی غالب برای غلبه بر این مشکل ظهور کرد [۲۵, ۲۶]. این رویکرد از یک ستون فقرات CNNعمیق مانند VGG19، ResNet، Xception یا MobileNet که قبلاً بر روی یک مجموعه داده تصاویر طبیعی در مقیاس بزرگ، مانند ImageNet، آموزش‌دیده است، استفاده می‌کند [۲۷, ۲۸]. رمزگذار پیش‌آموخته به عنوان یک استخراج‌کننده ویژگی بسیار مؤثر عمل می‌کند، زیرا قبلاً سلسله‌مراتب غنی از ویژگی‌های بصری را آموخته است. این ستون فقرات سپس در یک معماری رمزگذار-رمزگشا به سبک UNet ادغام می‌شود و کل مدل بر روی وظیفه پزشکی خاص، تنظیم دقیق[14] می‌شود [۹, ۲۹]. این رویکرد منجر به همگرایی سریع‌تر می‌شود، به داده‌های آموزشی

[1] Markov Random Fields
[2] Support Vector Machines
[3] Random Forests
[4] Hand-crafted Feature Engineering
[5] Texture
[6] Shape
[7] Pixel-wise

[8] Residual
[9] Volume
[10] Atrous/Dilated Convolutions
[11] Receptive Field
[12] Overfitting
[13] Transfer Learning
[14] Fine-tuned

کمتری نیاز دارد و اغلب با بهره‌گیری از ویژگی‌های تعمیم‌یافته از دامنه منبع، به عملکرد برتری دست می‌یابد.

۲-٤ روش‌های مبتنی بر ترنسفورمر و مکانیزم‌های توجه

محدودیت ذاتی CNN ها در محلی‌گرایی[1] کرنل‌های پیچشی آن‌هاست. آن‌ها برای درک وابستگی‌های دوربرد در یک تصویر (مثلاً ارتباط بین دو بخش دور از هم یک تومور بزرگ) دچار چالش هستند. مکانیزم‌های توجه و معماری‌های ترنسفورمر برای حل این مشکل پدید آمدند.

مکانیزم‌های توجه[2] به عنوان افزونه‌هایی برای معماری‌های CNN طراحی شدند. Attention U-Net [۳۰] از دروازه‌های توجه[3] در اتصالات پرشی استفاده می‌کند. این دروازه‌ها یاد می‌گیرند که به‌صورت پویا، نواحی مهم‌تر در ویژگی‌های ورودی را وزن‌دهی کرده و اطلاعات نامربوط را سرکوب کنند، و درنتیجه مدل را بر روی اهداف موردنظر متمرکز سازند.

ترنسفورمرها [۳۱] که در ابتدا برای پردازش زبان طبیعی طراحی شده بودند، به‌طور کامل بر مکانیزم توجه به خود تکیه دارند و می‌توانند ارتباط بین تمام جفت‌های ورودی را به‌صورت سراسری مدل‌سازی کنند. ViT[4] [۳۲] و مشتقات آن این ایده را به بینایی ماشین آوردند.

برای ناحیه‌بندی، مدل‌های ترکیبی مانند TransUNet [۳۳] اولین مدل‌های موفق بودند. TransUNet از یک رمزگذار CNN برای استخراج ویژگی‌های قوی سطح پایین و سپس از یک لایه ترنسفورمر در گلوگاه[5] برای مدل‌سازی زمینه سراسری استفاده می‌کند. پس از آن، مدل‌های مبتنی بر Swin-Transformer مانند Swin-Unet [۳٤] توسعه یافتند که کل معماری U-Net را با بلوک‌های Swin-Transformer (که توجه را به‌صورت سلسله‌مراتبی و در پنجره‌های محلی محاسبه می‌کند) پیاده‌سازی می‌کنند.

۳- روش پیشنهادی

در این بخش، ما معماری نوین خود، را ارائه می‌کنیم که به‌طور دقیق برای ناحیه‌بندی خودکار تومورهای مغزی از MRI طراحی شده است. ابتدا مسئله ناحیه‌بندی را به‌صورت رسمی تعریف می‌کنیم، سپس یک نمای کلی از شبکه را ارائه می‌دهیم و درنهایت هر یک از اجزای اصلی آن را به تفصیل شرح می‌دهیم.

۳-۱ تعریف مسئله

فرض کنید ورودی، یک برش[6] تصویر MRI دوبعدی و تک‌وجهی باشد که آن را با $X \in R^{H\times W\times 1} X \in R^{H\times W\times 1} X \in R^{H\times W\times 1}$ نمایش می‌دهیم. در این تعریف، H (ارتفاع) و W (عرض) ابعاد فضایی تصویر را مشخص می‌کنند و $C = 1$ نشان‌دهنده تک‌کاناله بودن ورودی (مثلاً، فقط مُدالیته T2-weighted) است.

برای هر تصویر ورودی X، یک نقشه ناحیه‌بندی مرجع[7] متناظر به نام $Y \in \{0,1\}^{H\times W}$ وجود دارد. این نقشه، که معمولاً توسط رادیولوژیست‌های متخصص به‌صورت دستی حاشیه‌نویسی شده است، هر پیکسل را به یکی از دو کلاس تخصیص می‌دهد: بافت سالم (کلاس ۰) یا ناحیه تومور (کلاس ۱).

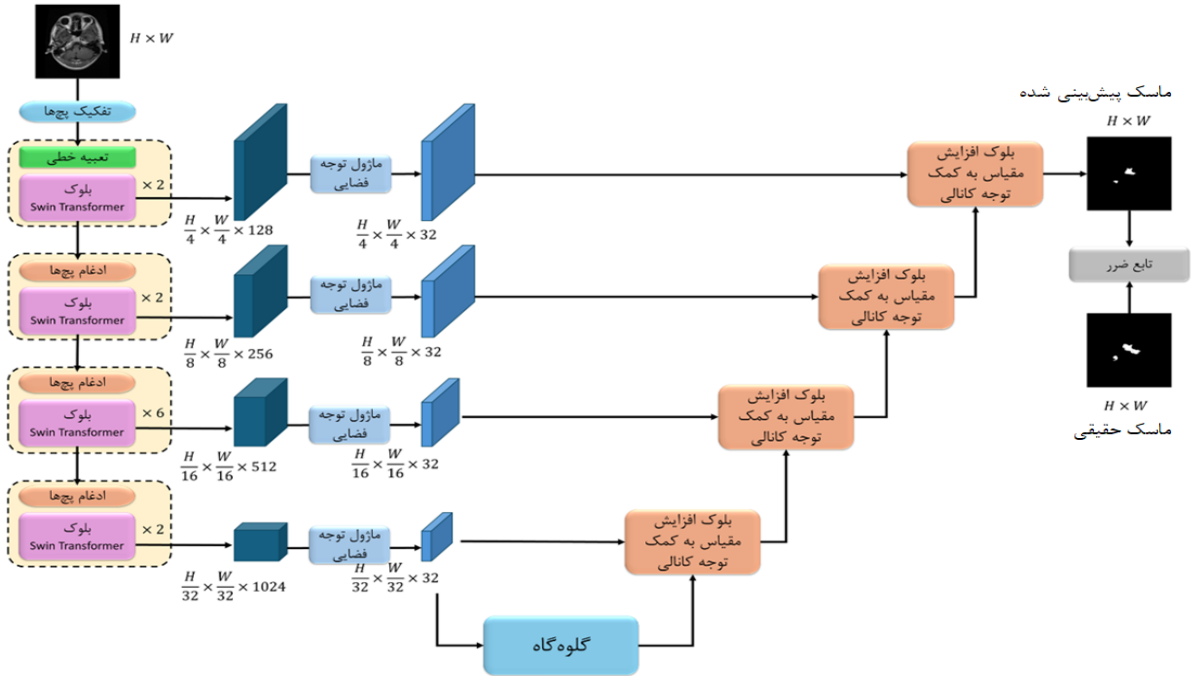
هدف روش پیشنهادی، آموزش یک تابع نگاشت[8] f است که توسط معماری شبکه مدل می‌شود. این تابع، تصویر ورودی Xرا می‌گیرد و یک نقشه ناحیه‌بندی پیش‌بینی‌شده $\hat{Y} = f(X)$تولید می‌کند. نقشه پیش‌بینی‌شده $\hat{Y}$باید دارای ابعاد فضایی یکسانی با ورودی $(H \times W)$ باشد.

درنهایت، هدف نهایی، به حداقل رساندن اختلاف بین نقشه پیش‌بینی‌شده $\hat{Y}$و نقشه مرجع Yاست. به‌عبارت‌دیگر، ما به دنبال توسعه مدلی هستیم که بتواند به‌طور خودکار و با دقت بالا، هر پیکسل در تصویر MRI ورودی را به کلاس صحیح تومور یا بافت سالم نسبت دهد.

[1] Locality
[2] Attention Mechanisms
[3] Attention Gates
[4] Vision Transformer
[5] Bottleneck
[6] Slice
[7] Ground Truth
[8] Mapping Function

۳-۲ کلیات روش پیشنهادی

برای حل مسئله ناحیه‌بندی تومور مغزی، ما یک معماری ترکیبی نوین مبتنی بر چارچوب رمزگذار-رمزگشا U-شکل ارائه می‌دهیم. شمای کلی این معماری در شکل (۱) نشان داده شده است. این شبکه به‌گونه‌ای طراحی شده است که از قدرت استخراج ویژگی سلسله‌مراتبی و چند-مقیاسی ترنسفورمرها در رمزگذار و مکانیزم‌های توجه دقیق فضایی و کانالی در سراسر شبکه بهره ببرد.



شکل (۱): کلیات روش پیشنهادی که یک معماری مبتنی بر UNet است و از Swin-transformer به عنوان ستون فقرات استفاده می‌کند و ماژول‌های توجه فضایی و کانالی، گلوگاه و تابع ضرر هم در آن نشان داده شده است.

مسیر رمزگذار شبکه ما از یک ستون فقرات مبتنی بر Swin-transformer استفاده می‌کند که با وزن‌های از پیش‌آموخته شده بر روی مجموعه داده ImageNet [35] مقداردهی اولیه شده است. این انتخاب به دلیل توانایی اثبات‌شده Swin-transformer در مدل‌سازی وابستگی‌های فضایی دوربرد و درعین‌حال حفظ کارایی محاسباتی از طریق مکانیزم توجه است. رمزگذار به‌صورت سلسله‌مراتبی عمل می‌کند. یک جزء کلیدی در معماری ما، ماژول توجه فضایی است که پس از هر مرحله Swin-transformer و قبل از اتصال پرشی اعمال می‌شود. همان‌طور که در شکل (۱) نشان داده شده است، هر نقشه ویژگی از رمزگذار، وارد ماژول توجه فضایی مربوطه می‌شود. این ماژول وظیفه دارد تا ویژگی‌های فضایی استخراج‌شده توسط ستون فقراتِ از پیش‌آموخته را برای وظیفه خاص ناحیه‌بندی تومور، پالایش و باز-وزن‌دهی کند. این ماژول همچنین ابعاد کانالی را کاهش می‌دهد، که این امر منجر به یک نمایش فشرده و متمرکز بر اطلاعات مکانی کلیدی برای انتقال به رمزگشا می‌شود.

عمیق‌ترین نقشه ویژگی از رمزگذار به گلوگاه که مبتنی برCNN است وارد می‌شود. این گلوگاه وظیفه پردازش بیشتر و استخراج اطلاعات معنایی سطح بالا را قبل از شروع فرآیند بازسازی فضایی در رمزگشا بر عهده دارد. مسیر رمزگشا به‌صورت متقارن با رمزگذار، وظیفه بازسازی نقشه ناحیه‌بندی با رزولوشن کامل را بر عهده دارد. این بخش از چهار بلوک افزایش مقیاس به کمک توجه کانالی تشکیل شده است. هر بلوک، ورودی خود را از لایه قبلی رمزگشا و همچنین نقشه ویژگی پالایش‌شده از اتصال پرشی

دریافت می‌کند. وجود توجه کانالی در این بلوک‌ها به شبکه اجازه می‌دهد تا به‌طور انطباقی، اهمیت ویژگی‌های کانال-محور را تنظیم کند و ترکیب بهینه‌ای از اطلاعات معنایی عمیق و اطلاعات مکانی دقیق را فراهم آورد. در ادامه هر بخش را به تفسیر توضیح می‌دهیم.

۳-۳ ستون فقرات

در معماری پیشنهادی، مسیر رمزگذار به عنوان ستون فقرات شبکه عمل می‌کند و وظیفه استخراج ویژگی‌های سلسله‌مراتبی و چند-مقیاسی از تصویر ورودی را بر عهده دارد. همان‌طور که اشاره شد، معماری‌های مبتنی بر CNN دارای سوگیری ذاتی به سمت ویژگی‌های محلی هستند. برای غلبه بر این محدودیت و بهره‌گیری از توانایی مدل‌سازی وابستگی‌های فضایی دوربرد، ما از معماری پیشرفته Swin-Transformer به عنوان رمزگذار استفاده می‌کنیم. انتخاب Swin-transformer به دو دلیل کلیدی استوار است: اول، توانایی آن در یادگیری نمایش‌های قدرتمند از طریق مکانیزم توجه به خود پنجره‌ای که تعادلی مؤثر بین کارایی محاسباتی و درک زمینه سراسری[1] ایجاد می‌کند. دوم، ما از یک مدل Swin-transformer استفاده می‌کنیم که قبلاً بر روی مجموعه داده عظیم ImageNet پیش‌آموخته شده است. این استراتژی یادگیری انتقالی به شبکه اجازه می‌دهد تا از ویژگی‌های بصری غنی که قبلاً آموخته است، به عنوان نقطه شروع استفاده کند. این امر در حوزه تصاویر پزشکی که اغلب با کمبود داده‌های برچسب‌دار مواجه است، اهمیتی حیاتی دارد و به همگرایی سریع‌تر و تعمیم‌پذیری[2] بهتر مدل کمک شایانی می‌کند.

معماری Swin-transformer ذاتاً سلسله‌مراتبی است و رفتار آن شباهت زیادی به CNN های مدرن دارد. فرآیند استخراج ویژگی در چهار مرحله انجام می‌شود. در ابتدا، تصویر ورودی X با ابعاد $H \times W \times 1$ به یک لایه تعبیه پچ‌ها[3] وارد می‌شود. این لایه، تصویر را به پچ‌های غیر-همپوشان[4] با اندازه 4×4 تقسیم می‌کند و هر پچ را به یک بردار ویژگی با بعد $C = 128$ نگاشت می‌دهد. درنتیجه، ابعاد فضایی به $H/4 \times W/4$ کاهش‌یافته و بعد کانال به C افزایش می‌یابد. این توکن‌ها سپس وارد مرحله اول بلوک‌های Swin-transformer می‌شوند.

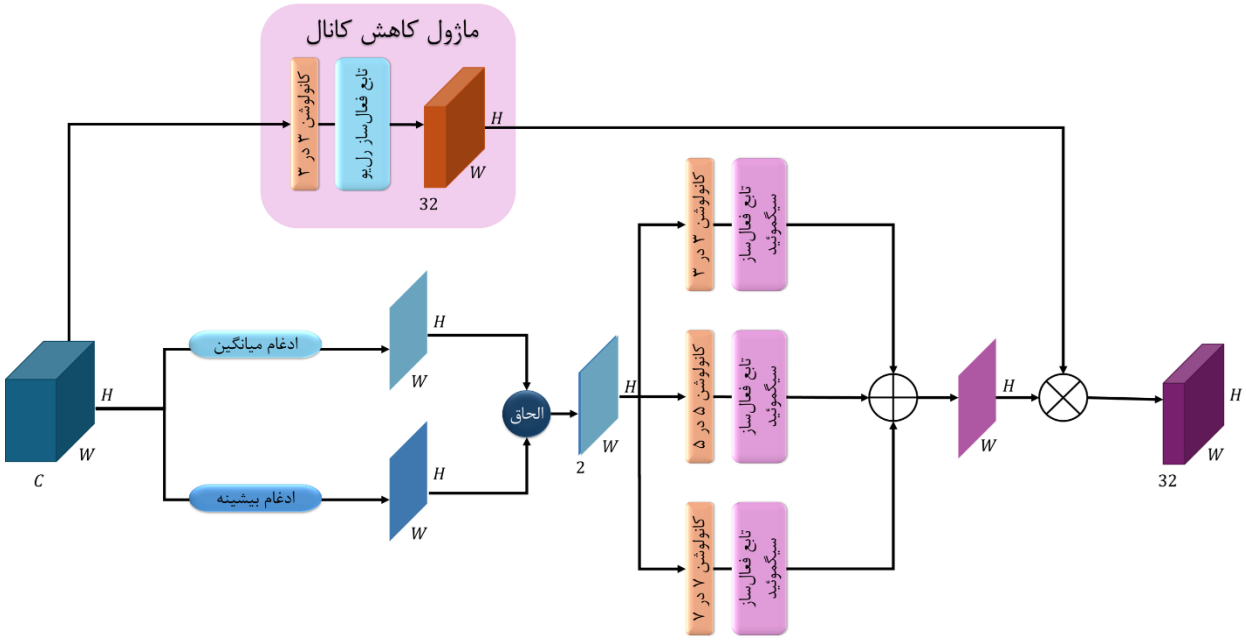
پس از هر مرحله، یک لایه ادغام پچ‌ها[5] قرار دارد. این لایه وظیفه نمونه‌برداری کاهشی[6] را بر عهده دارد. لایه ادغام پچ‌ها، توکن‌های همسایه را با هم ترکیب می‌کند، ابعاد فضایی را به نصف کاهش داده و بعد کانال را دو برابر این فرآیند چهار بار تکرار می‌شود و درمجموع چهار نقشه ویژگی چند-مقیاسی F_1, F_2, F_3, F_4 با رزولوشن‌های فضایی $H/4, H/8, H/16, H/32$ و ابعاد کانالی $C, 2C, 4C, 8C$ (که در مدل پایه $C = 128$ است) تولید می‌کند. همان‌طور که در شکل (۱) نشان داده شده است، خروجی هر یک از این چهار مرحله به عنوان ورودی برای ماژول توجه فضایی و سپس اتصالات پرشی به سمت رمزگشا مورد استفاده قرار می‌گیرد.

۳-۴ ماژول توجه فضایی

همان‌طور که بیان شد، ستون فقرات Swin-transformer که ما از آن استفاده کردیم، بر روی مجموعه داده تصاویر طبیعی ImageNet پیش‌آموخته شده است. اگرچه این ویژگی‌ها بسیار غنی هستند، اما لزوماً برای وظیفه تخصصی ناحیه‌بندی تومور مغزی بهینه نیستند. ویژگی‌های استخراج‌شده ممکن است شامل اطلاعاتی باشند که به تصاویر طبیعی مرتبط هستند اما در MRI نویز یا اطلاعات نامربوط محسوب می‌شوند. بنابراین، نیازمند یک پالایش[7] هست تا این ویژگی‌ها را برای دامنه پزشکی تطبیق دهد و توجه مدل را بر روی نواحی فضایی کلیدی (مانند مرزهای تومور) متمرکز کند. برای این منظور، ما یک ماژول توجه فضایی نوآورانه را طراحی کردیم که معماری دقیق آن در شکل (۲) نشان داده شده است. این ماژول پس از هر مرحله از ستون فقرات رمزگذار و قبل از ارسال ویژگی‌ها به رمزگشا از طریق اتصالات پرشی، اعمال می‌شود. این ماژول دارای دو هدف موازی است: اول، کاهش ابعاد کانالی برای ایجاد یک اتصال پرشی فشرده و کارآمد؛ و دوم، اعمال یک ماسک توجه فضایی برای پالایش ویژگی‌ها.

[1] Global Context
[2] Generalization
[3] Patch Embedding
[4] Non-overlapping
[5] Patch Merging
[6] Downsampling
[7] Refinement

همان‌طور که در شکل (۲) مشاهده می‌شود، ماژول از دو مسیر اصلی تشکیل شده است. مسیر اول، یک ماژول کاهش کانال است که نقشه ویژگی ورودی F_i با ابعاد $H \times W \times C$ را دریافت کرده و با استفاده از یک کانولوشن 3×3 و تابع فعال‌ساز رلیو[1]، ابعاد کانال را به ۳۲ کاهش می‌دهد. این مسیر، تنه اصلی ویژگی‌ها را برای ارسال به رمزگشا آماده می‌کند.



شکل (۲): ماژول توجه فضایی که در آن از ادغام میانگین و ادغام بیشینه استفاده شده است. همچنین از فیلترهای کانولوشنی در ابعاد مختلف بهره گرفته شده و در انتها نیز با کاهش بعد کانال، سعی در کاهش پیچیدگی محاسباتی دارد.

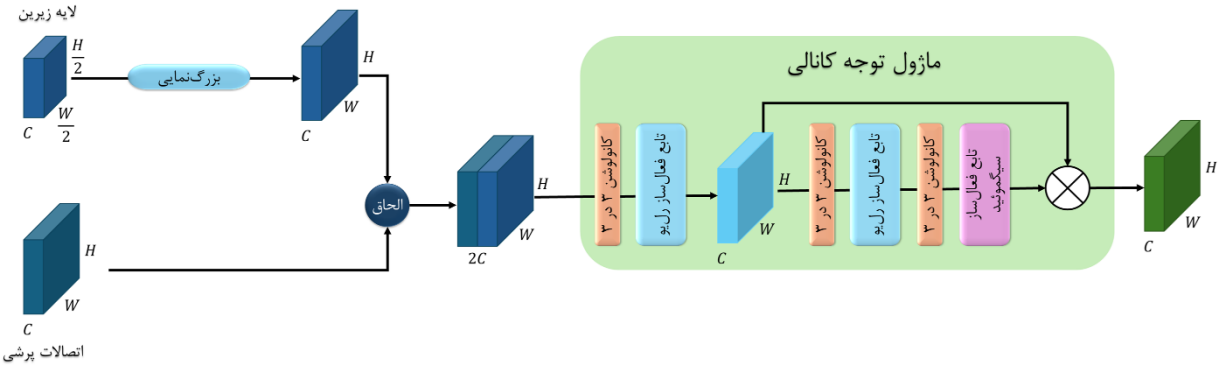
مسیر دوم وظیفه تولید ماسک توجه فضایی را بر عهده دارد. در این مسیر، ابتدا نقشه ویژگی ورودی F_i از طریق دو عملیات ادغام[2] مستقل در امتداد محور کانال‌ها یعنی ادغام میانگین[3] و ادغام بیشینه[4] پردازش می‌شود. این دو عملیات، دو نقشه دوبعدی $H \times W \times 1$ تولید می‌کنند که به ترتیب، خلاصه‌ای از ویژگی‌های میانگین و ویژگی‌های برجسته را در هر موقعیت مکانی ارائه می‌دهند. این دو نقشه سپس به یکدیگر الحاق[5] شده و یک نقشه ویژگی با ابعاد $H \times W \times 2$ را تشکیل می‌دهند.

برای درک بهتر زمینه فضایی در مقیاس‌های مختلف، این نقشه ۲ کانالی به‌صورت موازی به سه شاخه کانولوشنی با اندازه‌های کرنل متفاوت 3×3، 5×5 و 7×7 وارد می‌شود. خروجی هر یک از این پیچش‌ها از یک تابع فعال‌ساز سیگموید[6] عبور می‌کند تا مقادیر آن به بازه $[0-1]$ نرمال شود و سه نقشه توجه مجزا تولید گردد.

شایان‌ذکر است که این ساختار چند-شاخه، تفاوت بنیادینی با ماژول‌های توجه فضایی استاندارد نظیر آنچه در CBAM یا Attention U-Net به‌کاررفته، دارد. در روش‌های مرسوم، معمولاً از یک کرنل کانولوشن با اندازه ثابت (مثلاً 7×7) برای تولید نقشه توجه استفاده می‌شود. این رویکرد تک-مقیاسی باعث می‌شود مدل تنها بتواند ویژگی‌ها را در یک میدان دید ثابت برجسته کند. اما در ناحیه‌بندی تومورهای مغزی که اندازه تومورها تنوع بسیار زیادی دارد (از ضایعات بسیار کوچک نقطه‌ای تا تومورهای حجیم)، استفاده از یک میدان دید ثابت ناکارآمد است. طراحی پیشنهادی ما با بهره‌گیری از سه شاخه موازی با کرنل‌های 3×3، 5×5 و

[1] ReLU
[2] Pooling
[3] Average Pooling
[4] Max Pooling
[5] Concatenate
[6] Sigmoid

7 × 7، به شبکه اجازه می‌دهد تا به‌صورت هم‌زمان جزئیات ریز (تومورهای کوچک یا لبه‌های پیچیده) و بافت‌های محیطی وسیع‌تر را مشاهده کند. این مکانیزم توجه چند-مقیاسه باعث می‌شود تا ماژول پیشنهادی برخلاف نمونه‌های پیشین، نسبت به تغییرات اندازه تومور مقاوم بوده و پالایش ویژگی‌ها را با دقت بسیار بالاتری انجام دهد.



شکل (۳): بلوک افزایش مقیاس به کمک توجه کانالی

این سه نقشه توجه سپس با یکدیگر جمع عنصری[1] می‌شوند تا یک ماسک توجه فضایی نهایی و چند-مقیاسی ایجاد شود. درنهایت، این ماسک توجه فضایی در نقشه ویژگی کاهش-کانال-یافته (خروجی مسیر اول) به‌صورت عنصری ضرب[2] می‌شود. این عملیات ضرب، به‌طور مؤثری ویژگی‌ها را در نواحی فضایی مهم (که توسط ماسک شناسایی شده‌اند) تقویت کرده و اطلاعات نامربوط در سایر نواحی را تضعیف می‌کند. خروجی نهایی این ماژول، یک نقشه ویژگی پالایش‌شده با ابعاد H × W × 32 است که آماده ارسال به رمزگشا است.

۳-۵ گلوگاه

گلوگاه، عمیق‌ترین نقطه در معماری U-Net ما است و به عنوان پلی میان مسیر رمزگذار و مسیر رمزگشا عمل می‌کند. این بخش، نقشه ویژگی خروجی از آخرین مرحله ستون فقرات (پس از عبور از ماژول توجه فضایی مربوطه) را به عنوان ورودی دریافت می‌کند. این نقشه ویژگی، غنی‌ترین اطلاعات معنایی و سطح بالا را در مورد تصویر دارا می‌باشد، هرچند که کمترین رزولوشن فضایی را دارد. هدف اصلی گلوگاه، پردازش بیشتر و پالایش این اطلاعات معنایی، قبل از شروع فرآیند بازسازی فضایی در رمزگشا است. در معماری ما، این بخش از یک طراحی ساده و درعین‌حال مؤثر مبتنی بر CNN تشکیل شده است. این ماژول شامل دو بلوک کانولوشنی متوالی است که هر بلوک از یک لایه کانولوشن 3 × 3، به دنبال آن یک لایه نرمال‌سازی دسته‌ای[3] برای پایداری آموزش و درنهایت یک تابع فعال‌ساز رلیو برای افزودن غیرخطی بودن، تشکیل شده است. این ساختار به مدل اجازه می‌دهد تا نمایش‌های ویژگی سطح بالا را قبل از ارسال به مسیر رمزگشا، بیشتر ادغام و بهینه کند.

۳-۶ بلوک افزایش مقیاس به کمک توجه کانالی

مسیر رمزگشا وظیفه بازسازی نقشه ناحیه‌بندی با رزولوشن کامل را از طریق فرآیند نمونه‌برداری افزایشی و ادغام ویژگی‌های چند-مقیاسی بر عهده دارد. در هر مرحله از این مسیر، ما از یک بلوک افزایش مقیاس به کمک توجه کانالی استفاده می‌کنیم که ساختار آن در شکل (۳) نشان داده شده است. این بلوک برای ادغام بهینه اطلاعات معنایی سطح بالا با اطلاعات مکانی دقیق (دریافتی از اتصالات پرشی) طراحی شده است.

فرآیند در این بلوک در سه مرحله انجام می‌شود. ابتدا، نقشه ویژگی دریافتی از لایه قبلی رمزگشا (لایه زیرین) که دارای رزولوشن فضایی پایین اما اطلاعات معنایی غنی است، از طریق یک عملیات

[1] Element-wise Addition
[2] Element-wise Multiplication
[3] Batch Normalization

بزرگنمایی عبور می‌کند تا ابعاد فضایی آن دو برابر شود و با ابعاد نقشه ویژگی اتصال پرشی متناظر، یکسان گردد. باید توجه داشت که درصورتی‌که لایه زیرین، گلوگاه باشد، نیازی به بزرگنمایی نیست.

در مرحله دوم، نقشه ویژگی بزرگ‌شده با نقشه ویژگی پالایش‌شده که از اتصال پرشی می‌آید، از طریق عملیات الحاق در امتداد محور کانال‌ها ترکیب می‌شود. این الحاق، اطلاعات معنایی عمیق را با جزئیات فضایی دقیق ادغام می‌کند.

در مرحله سوم و نهایی، نقشه ویژگی الحاق‌شده که اکنون ابعاد کانالی 2C دارد به ماژول توجه کانالی پیشنهادی داده می‌شود. همان‌طور که در شکل (۳) نشان داده شده است، این ماژول ابتدا ابعاد کانال را از طریق یک کانولوشن 3×3 به همراه تابع فعال‌ساز رلیو کاهش داده و نقشه ویژگی میانی F_{mid} را تولید می‌کند. سپس، این نقشه میانی به‌صورت موازی وارد دو شاخه می‌شود: یک شاخه، خود F_{mid} را حفظ می‌کند و شاخه دیگر، یک ماسک توجه از آن استخراج می‌کند. این ماسک از طریق اعمال متوالی یک کانولوشن 3×3، یک تابع فعال‌ساز رلیو، یک کانولوشن 3×3دیگر و نهایتاً یک تابع سیگموید تولید می‌شود. درنهایت، نقشه ویژگی F_{mid} به‌صورت عنصری در این ماسک توجه ضرب می‌شود. این فرآیند به مدل اجازه می‌دهد تا به‌صورت انطباقی، وزن کانال‌های آموزنده‌تر را تقویت و کانال‌های با اطلاعات کمتر را تضعیف کند، که منجر به پالایش نهایی ویژگی‌ها قبل از ارسال به مرحله بعدی رمزگشا می‌شود.

۳-۷ تابع ضرر

برای بهینه‌سازی پارامترهای شبکه، ما از تابع ضرر آنتروپی متقاطع دودویی[1] استفاده می‌کنیم که یک انتخاب استاندارد و مؤثر برای مسائل ناحیه‌بندی دوتایی است. این تابع، اختلاف بین توزیع احتمال پیش‌بینی‌شده توسط مدل در خروجی نهایی و توزیع برچسب واقعی را در سطح هر پیکسل اندازه‌گیری می‌کند.

با توجه به اینکه خروجی نهایی مدل پس از عبور از تابع سیگموید، احتمالی بین ۰ (بافت سالم) و ۱ (تومور) برای هر پیکسل است، تابع ضرر BCE برای یک تصویر به‌صورت نشان داده شده در رابطه (۱) تعریف می‌شود.

$$L_{BCE} = -\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}[y_{ij}\log(\widehat{y_{ij}}) + (1-y_{ij})\log(1-\widehat{y_{ij}})] \quad (۱)$$

در این معادله H و W ابعاد تصویر هستند. y_{ij} برچسب واقعی (۰ یا ۱) برای پیکسل در موقعیت (i, j) و $\widehat{y_{ij}}$ احتمال پیش‌بینی‌شده توسط مدل برای همان پیکسل است. کل فرآیند آموزش بر مبنای کمینه‌سازی این تابع ضرر واحد که بر روی نقشه ناحیه‌بندی نهایی اعمال می‌شود، هدایت می‌گردد.

انتخاب تابع ضرر آنتروپی متقاطع دودویی (BCE) در این پژوهش با در نظر گرفتن ویژگی‌های معماری پیشنهادی انجام شده است. اگرچه توابع ضرری مانند Dice Loss یا Focal Loss برای مدیریت عدم تعادل کلاس‌ها (تفاوت حجم تومور و پس‌زمینه) مرسوم هستند، اما در مدل پیشنهادی، این چالش توسط ماژول‌های توجه فضایی و کانالی مرتفع شده است. این ماژول‌ها با مکانیزم‌های وزن‌دهی درونی، به‌طور خودکار تمرکز شبکه را بر روی نواحی هدف (تومور) معطوف کرده و ویژگی‌های پس‌زمینه را سرکوب می‌کنند؛ درنتیجه نیاز به اعمال وزن‌دهی‌های پیچیده در تابع ضرر کاهش می‌یابد. علاوه بر این، استفاده از BCE به دلیل رفتار هموارتر در محاسبه گرادیان‌ها، پایداری آموزش را برای ستون فقرات مبتنی بر ترنسفورمر تضمین می‌کند.

٤- ارزیابی روش پیشنهادی

در فصل پیشین، مدل پیشنهادی خود را که برای ناحیه‌بندی تومور مغزی طراحی شده است، به تفصیل شرح دادیم. در این فصل، به ارزیابی جامع عملکرد مدل پیشنهادی می‌پردازیم. برای این منظور، ابتدا مجموعه داده مورد استفاده در آزمایش‌ها را معرفی می‌کنیم، سپس معیارهای ارزیابی و جزئیات پیاده‌سازی مدل را تشریح کرده و درنهایت، نتایج به‌دست‌آمده را با سایر روش‌های پیشرفته مقایسه و تحلیل می‌نماییم.

[1] Binary Cross-Entropy

٤-١ مجموعه دادگان

ارزیابی و آموزش مدل‌های یادگیری عمیق به یک مجموعه داده استاندارد، چالش‌برانگیز و باکیفیت نیاز دارد. در این پژوهش، ما از مجموعه داده BRISC[1] استفاده کردیم [٣٦]. این مجموعه داده، یک مجموعه جدید از اسکن‌های MRI است که با هدف رفع محدودیت‌های مجموعه داده‌های موجود، مانند عدم تعادل کلاس‌ها و ناهماهنگی در حاشیه‌نویسی‌ها، توسعه یافته است.

مجموعه داده BRISC شامل ٦٠٠٠ تصویر MRI از نوع T1-weighted است که به‌طور رسمی به دو بخش آموزشی (٥٠٠٠ تصویر) و آزمایشی (١٠٠٠ تصویر) تقسیم شده است. اگرچه این مجموعه داده در نسخه اصلی خود شامل برچسب‌هایی برای سه نوع تومور شایع مغزی (گلیوما، منتژیوما و پیتویتاری) و همچنین تصاویر سالم (بدون تومور) است که در هر تصویر، یک نوع از این تومورها وجود دارد.

یکی از ویژگی‌های برجسته این مجموعه داده، فرآیند دقیق حاشیه‌نویسی مجدد آن است. تمام ماسک‌های تومور توسط رادیولوژیست‌ها و پزشکان متخصص تأیید شده‌اند تا از بالاترین سطح دقت اطمینان حاصل شود. علاوه بر این، این مجموعه داده از نظر صفحات تصویربرداری متعادل است. از آنجایی که معماری پیشنهادی ما مبتنی بر ورودی‌های دوبعدی است، ما از تمام برش‌ها در هر سه نما به عنوان تصاویر آموزشی مستقل استفاده کردیم.

٤-٢ جزئیات پیاده‌سازی

تمام آزمایش‌ها و شبیه‌سازی‌ها با استفاده از چارچوب یادگیری عمیق PyTorch پیاده‌سازی شدند. فرآیند آموزش مدل‌ها بر روی پلتفرم Kaggle و با بهره‌گیری از قدرت پردازشی یک پردازنده گرافیکی (GPU) مدل Tesla T4 انجام پذیرفت. جزئیات کامل هایپرپارامترها[2]، شامل نرخ یادگیری، بهینه‌ساز، اندازه دسته و تعداد تکرار آموزش، در جدول (١) خلاصه شده است.

تصاویر موجود در مجموعه داده با فرمت JPEG و ماسک‌های متناظر با فرمت PNG ذخیره شده‌اند. این کار در این دیتاست به این دلیل انجام شده بود که به خاطر فشرده‌سازی، ماسک ممکن بود خراب شود فلذا فرمت آن را با PNG ذخیره کرده‌اند. در مرحله آموزش و آزمون، این تصاویر پس از فراخوانی به تنسورهای نرمال‌سازی شده تبدیل می‌شوند تا دقت محاسبات اعشاری در طول فرآیند یادگیری حفظ شود. درنهایت، ماسک‌های خروجی تولید شده توسط مدل نیز جهت حفظ کیفیت و جلوگیری از فشرده‌سازی مخرب، با فرمت PNG ذخیره می‌گردند.

جدول (١): هایپرپارامترهای روش پیشنهادی

هایپرپارامترها	مقدار
چارچوب یادگیری عمیق	PyTorch
محیط آموزش	Kaggle
واحد پردازش گرافیکی[3]	Tesla T4
ابعاد تصویر ورودی	512×512
اندازه دسته	٨
تعداد تکرار	٦٠
نرخ آموزش	1×10^{-4}

پیش از ورود داده‌ها به شبکه، عملیات پیش‌پردازش شامل نرمال‌سازی شدت روشنایی پیکسل‌ها با استفاده از روش Min-Max به بازه $[0, 1]$ انجام شد تا پایداری عددی و سرعت همگرایی افزایش یابد. همچنین به‌منظور افزایش تنوع داده‌های آموزشی و جلوگیری از بیش‌برازش، تکنیک‌های افزونگی داده شامل چرخش تصادفی، قرینه‌سازی افقی و عمودی و تغییرات جزئی در کنتراست بر روی داده‌ها اعمال گردید.

٤-٣ معیارهای ارزیابی

به‌منظور ارزیابی کمی و مقایسه عملکرد ناحیه‌بندی مدل پیشنهادی، ما از دو معیار استاندارد و پرکاربرد در حوزه ناحیه‌بندی تصاویر پزشکی، یعنی ضریب شباهت دایس[4] (Dice) و اشتراک بر روی اجتماع[5] (IoU)، استفاده می‌کنیم. هر دوی این معیارها، میزان همپوشانی فضایی بین ماسک پیش‌بینی‌شده توسط مدل ($\widehat{M}$) و ماسک برچسب واقعی (M_{gt}) را اندازه‌گیری می‌کنند.

Dice: این معیار، میزان همپوشانی بین دو مجموعه را اندازه‌گیری می‌کند و به‌ویژه برای ارزیابی عملکرد ناحیه‌بندی در مواردی که

[1] BRain Tumor Image Segmentation and Classification
[2] Hyperparameters
[3] GPU
[4] Dice Similarity Coefficient
[5] Intersection Over Union

عدم تعادل کلاس (مانند کوچک بودن ناحیه تومور نسبت به کل تصویر) وجود دارد، بسیار مؤثر است. همان‌طور که در رابطه (۲) نشان داده شده است، Dice به‌صورت دو برابر مساحت اشتراک، تقسیم بر مجموع مساحت دو ماسک تعریف می‌شود. مقدار این ضریب بین ۰ (عدم همپوشانی) و ۱ (همپوشانی کامل) قرار دارد.

$$Dice(\widehat{M}, M_{gt}) = \frac{2 \times |\widehat{M} \cap M_{gt}|}{|\widehat{M}| + |M_{gt}|} \quad (۲)$$

اشتراک بر روی اجتماع: این معیار که با نام ضریب ژاکارد[1] نیز شناخته می‌شود، همپوشانی بین ماسک پیش‌بینی‌شده و ماسک واقعی را محاسبه می‌کند. مطابق رابطه (۳) IoU به‌صورت مساحت اشتراک تقسیم بر مساحت اجتماع دو ماسک تعریف می‌گردد. مشابه Dice، مقدار این ضریب نیز در بازه ۰ تا ۱ متغیر است.

$$IoU(\widehat{M}, M_{gt}) = \frac{|\widehat{M} \cap M_{gt}|}{|\widehat{M} \cup M_{gt}|} \quad (۳)$$

جدول (۲): مقایسه با روش‌های پیشین. بیشترین مقدار در هر ستون برجسته و دومین مقدار زیرخط دارد

مدل	گلیوما IoU	منتژیوم IoU	هیپوفیز IoU	وزنی IoU	Dice
Unet	٦٩/٧	٧٧/١	٧٩/٣	٧٥/٧	٨٦/٢
++Unet	٧١/٧	٧٤/٢	٧٩/٧	٧٥/٣	٨٥/٩
MANet	٧٢/٤	٧٧/٥	٧٨/٠	٧٦/٣	٨٦/٥
EINet	٧٣/٦	٧٨/٤	٨٠/٣	٧٧/٧	٨٧/٤
TransUNet	٧٣/٨	٧٩/١	٨١/٥	٧٨/٩	٨٧/٢
Swin-UNet	٧٤/٥	٧٩/٨	٨٠/٠	٧٩/٣	٨٧/٥
DAD	٧٥/٢	٨٠/٤	٨٢/٣	٧٩/٥	٨٨/١
ABANet	٧٢/٤	٨٠/٤	٨٤/٧	٧٩/٥	٨٨/١
روش پیشنهادی	٧٤/٢	٨٢/٣	٨٤/٤	٨٠/٦	٨٨/٦

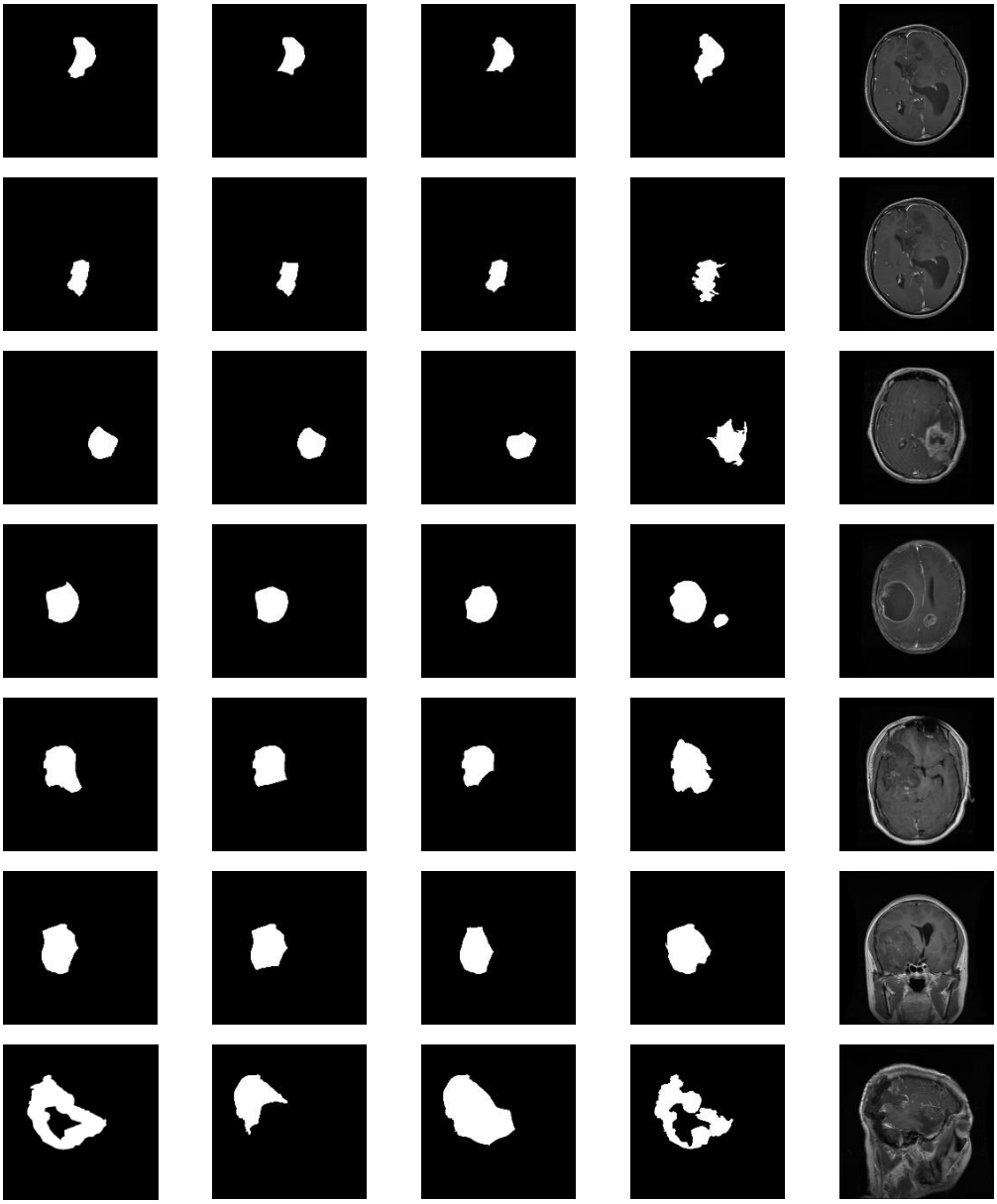
٤-٤ مقایسه با روش‌های پیشین

به‌منظور سنجش دقیق‌تر کارایی معماری پیشنهادی، ما نتایج آن را با شش مدل پیشرفته و شناخته‌شده در حوزه ناحیه‌بندی تومور مغزی، از جمله U-Net، U-Net++، MANet، EINet، TransUnet، Swin-UNet، DAD و ABANet مقایسه کردیم. تمام مدل‌های رقیب با استفاده از همان مجموعه داده و تحت شرایط آزمایشی یکسان آموزش و ارزیابی شدند. جدول (۲) نتایج کمی این مقایسه را بر روی مجموعه داده BRISC نشان می‌دهد. تحلیل نتایج ارائه شده در جدول (۲) به‌وضوح برتری مدل پیشنهادی ما را در معیارهای کلی نشان می‌دهد. روش ما موفق به کسب بالاترین امتیاز در هر دو معیار IoU وزنی و Dice شده است. این نتایج، نشان‌دهنده توانایی برتر مدل در دستیابی به یک تعادل کارآمد در ناحیه‌بندی هر سه نوع تومور است.

هنگام بررسی نتایج به تفکیک هر کلاس، مدل پیشنهادی ما بهترین عملکرد را در ناحیه‌بندی منتژیوم با IoU% ۸۲/۳ کسب کرده است که بهبود قابل‌توجهی نسبت به سایر روش‌ها نشان می‌دهد. در مورد ناحیه‌بندی هیپوفیز، روش ما با امتیاز ٪۸٤/٤ در جایگاه دوم و با اختلاف بسیار ناچیزی نسبت به ABANet قرار دارد که نشان‌دهنده عملکرد بسیار رقابتی آن است. در ناحیه‌بندی چالش‌برانگیز گلیوما، مدل ما IoU٪۷٤/۲ را کسب کرده و پس از مدل DAD در رتبه

[1] Jaccard Index

دوم قرار می‌گیرد. بعلاوه، مدل پیشنهادی با اختلاف ٪۱/۷ و ٪۱/۳ در معیار IoU وزنی، عملکرد بهتری نسبت به TransUNet و Swin-UNet داشته است. مدل Swin-UNet اگرچه نتایج خوبی دارد، اما به دلیل حذف کانولوشن‌ها و تکیه صرف بر ترنسفورمر، در بازیابی دقیق لبه‌های تومور با چالش مواجه شده است. در مقابل، مدل پیشنهادی ما با حفظ ساختار U-Net و تقویت آن توسط ماژول‌های توجه کانولوشنی، توانسته است تعادل بهینه‌ای میان زمینه سراسری و دقت محلی برقرار کند که نتیجه آن برتری در تمام شاخص‌های کلی است.



شکل (٤): مقایسه کیفی نتایج ناحیه‌بندی. از راست به چپ: الف) تصویر ورودی، ب) ناحیه‌بندی مرجع، ج) مدل ABANet د) مدل DAD و ه) روش پیشنهادی

درمجموع، دستیابی به رتبه اول در معیارهای جامع IoU وزنی و Dice و همچنین کسب رتبه اول یا دوم در هر سه کلاس تومور به‌صورت مجزا، مؤید آن است که ترکیب ستون فقرات Swin-Transformer با ماژول‌های توجه فضایی و کانالی، یک استراتژی مؤثر برای افزایش دقت ناحیه‌بندی بوده است.

علاوه بر مقایسه کمی، ارزیابی کیفی عملکرد مدل نیز برای درک بهتر رفتار آن در مواجهه با چالش‌های ساختاری تومورها ضروری است. شکل (٤) نمونه‌هایی از نتایج ناحیه‌بندی را بر روی تصاویر آزمایشی مجموعه داده BRISCنشان می‌دهد. در این شکل، ستون اول تصویر ورودی، ستون دوم ناحیه‌بندی مرجع و ستون پنجم خروجی مدل پیشنهادی را نمایش می‌دهد. همچنین برای مقایسه دقیق‌تر، خروجی دو مدل DAD و ABANet که طبق جدول (۲) نزدیک‌ترین عملکرد کمی را به روش ما داشته‌اند، در ستون‌های سوم و چهارم آورده شده است. همان‌طور که مشاهده می‌شود، اگرچه مدل‌های DAD و ABANet در تشخیص کلی ناحیه تومور موفق عمل کرده‌اند، اما در نواحی با کنتراست پایین و مرزهای نامنظم دچار خطا شده‌اند. در مقابل، روش پیشنهادی به لطف بهره‌گیری از ویژگی‌های سراسری استخراج‌شده توسط ترنسفورمر و پالایش دقیق توسط ماژول‌های توجه فضایی و کانالی، همپوشانی بیشتری با برچسب حقیقی داشته و مرزهای تومور را با دقت و پیوستگی بیشتری بازیابی کرده است.

جدول (۳): نتایج مطالعه فرسایشی برای بررسی تأثیر ماژول‌های توجه فضایی و کانالی بر روی مجموعه داده BRISC

	ماژول توجه فضایی	توجه کانالی	IoU	Dice
مدل پایه			۷٦/۱	۸٥/٥
	✓		۷۷/٥	۸٦/٥
		✓	۷۸/٤	۸۷/۱
	✓	✓	۸۰/٤	۸۸/٦

٤-٥ مطالعه فرسایشی

همان‌طور که در جدول (۳) مشاهده می‌شود، مدل پایه که تنها از ساختار رمزگذار Swin-Transformer و یک رمزگشای استاندارد (بدون مکانیزم‌های توجه پیشنهادی) بهره می‌برد، به IoU٪۷٦/۱ دست یافته است. اگرچه این عملکرد به دلیل قدرت ذاتی ستون فقرات ترنسفورمر قابل‌قبول است، اما همچنان پتانسیل بهبود را نشان می‌دهد.

با افزودن ماژول توجه فضایی به مسیرهای پرشی، شاخص IoU با رشدی ۱/٤ درصدی به ٪۷۷/٥ افزایش یافت. این بهبود تأیید می‌کند که پالایش ویژگی‌های مکانی قبل از انتقال به رمزگشا، به مدل کمک می‌کند تا تمرکز دقیق‌تری بر روی مرزهای تومور و نواحی موردنظر داشته باشد. از سوی دیگر، ادغام ماژول توجه کانالی در بلوک‌های افزایش مقیاس، تأثیر بیشتری داشته و IoU را به ٪۷۸/٤ ارتقا داده است؛ که نشان‌دهنده اهمیت باز وزن‌دهی ویژگی‌های معنایی و سرکوب کانال‌های حاوی اطلاعات نامربوط در فرآیند بازسازی است.

نکته حائز اهمیت، عملکرد مدل پیشنهادی است که با بهره‌گیری هم‌زمان از هر دو ماژول، به IoU وزنی ٪۸۰/٦ و ضریب دایس ٪۸۸/٦ دست یافته است. این جهش عملکرد (که بیش از مجموع بهبودهای تکی هر ماژول است)، بیانگر وجود یک اثر هم‌افزایی قوی میان این دو مکانیزم است. درواقع، توجه فضایی با مشخص کردن کجا (مکان‌های مهم) و توجه کانالی با مشخص کردن چه چیزی (ویژگی‌های مهم)، یکدیگر را تکمیل کرده و منجر به دقیق‌ترین ناحیه‌بندی شده‌اند.

٥- نتیجه‌گیری

در این پژوهش، ما به چالش حیاتی ناحیه‌بندی خودکار تومورهای مغزی از تصاویر MRI پرداختیم. این فرآیند، به دلیل تنوع ظاهری تومورها و مرزهای اغلب نامشخص با بافت‌های سالم، یک وظیفه پیچیده در تحلیل تصاویر پزشکی است. با اذعان به محدودیت‌های ذاتی معماری‌های CNN (که در درک زمینه سراسری ضعیف هستند) و مدل‌های ترنسفورمر (که ممکن است جزئیات محلی را نادیده بگیرند)، هدف اصلی این مقاله، طراحی و ارزیابی یک

معماری ترکیبی نوین بود که به‌طور هم‌افزا از نقاط قوت هر دو رویکرد بهره ببرد. ما یک شبکه رمزگذار–رمزگشا U–شکل پیشنهاد دادیم. ستون فقرات رمزگذار این مدل از یک Swin-Transformer پیش‌آموخته تشکیل شده است تا نمایش‌های ویژگی قدرتمند و چند–مقیاسی استخراج کند. نوآوری اصلی این معماری در دو بخش کلیدی نهفته است: اول، معرفی ماژول توجه فضایی پیشرفته که بر روی خروجی‌های رمزگذار اعمال می‌شود تا ویژگی‌ها را به‌صورت فضایی پالایش کرده و آن‌ها را برای وظیفه تخصصی پزشکی تطبیق دهد. دوم، استفاده از بلوک‌های افزایش مقیاس به کمک توجه کانالی در مسیر رمزگشا که به‌طور هوشمندانه، اطلاعات دریافتی از اتصالات پرشی و لایه‌های عمیق‌تر را برای بازسازی دقیق مرزها باز–وزن‌دهی می‌کند. ارزیابی‌های انجام‌شده بر روی مجموعه داده چالش‌برانگیز BRISC، کارایی بالای مدل پیشنهادی را به اثبات رساند. نتایج تجربی نشان داد که معماری ما در مقایسه با چندین روش پیشرفته به عملکرد برتری دست یافته است.

با وجود نتایج امیدوارکننده، این پژوهش مسیر را برای تحقیقات آتی هموار می‌کند. معماری فعلی بر روی تصاویر دوبعدی عمل می‌کند؛ گسترش آن به یک ساختار کاملاً سه‌بعدی می‌تواند با درک زمینه بین–برشی، دقت را به‌ویژه در مرزهای پیچیده تومور افزایش دهد. همچنین، این مدل بر روی تصاویر تک‌وجهی (T1-weighted) ارزیابی شد؛ در آینده، ادغام مُدالیته‌های مکمل مانند T2-weighted و FLAIR می‌تواند اطلاعات غنی‌تری را برای ناحیه‌بندی فراهم آورد. درنهایت، می‌توان ماژول‌های توجه پیشنهادی را در سایر وظایف تصویربرداری پزشکی یا بر روی مجموعه داده‌های دیگر ارزیابی کرد تا از تعمیم‌پذیری آن‌ها اطمینان حاصل شود.

References

[1] E. Goceri, "An efficient network with CNN and transformer blocks for glioma grading and brain tumor classification from MRIs," Expert Systems with Applications, vol. 268, p. 126290, 2025.

[2] T. AL-SHEHARI, M. KADRIE, M. AL-RAZGAN, and T. ALFAKIH, "TumorGANet: A Transfer Learning and Generative Adversarial Network-Based Data Augmentation Model for Brain Tumor Classification," 2024.

[3] N. Altini et al., "A Comparison Between Unimodal and Multimodal Segmentation Models for Deep Brain Structures from T1-and T2-Weighted MRI," Machine Learning and Knowledge Extraction, vol. 7, no. 3, p. 84, 2025.

[4] M. A. Ilani, D. Shi, and Y. M. Banad, "T1-weighted MRI-based brain tumor classification using hybrid deep learning models," Scientific Reports, vol. 15, no. 1, p. 7010, 2025.

[5] R. Li et al., "DeepGlioSeg: advanced glioma MRI data segmentation with integrated local-global representation architecture," Frontiers in Oncology, vol. 15, p. 1449911, 2025.

[6] M. Rajabghane, A. Bahrololoum, and M. Eftekhari, "Improving Unet Networks for Medical Image Segmentation by adding Attention Mechanism Layers," Journal of Machine Vision and Image Processing, vol. 10, no. 4, pp. 49-59, 2024.

[7] K. C. Pasunoori, C. R. Prasad, and K. R. Kumar, "A systematic review on deep learning based brain tumor segmentation and detection using MRI: Past insights, present techniques and future trends," Computational Biology and Chemistry, p. 108696, 2025.

[8] N. Huda and K. R. Ku-Mahamud, "CNN-Based Image Segmentation Approach in Brain Tumor Classification: A Review," Engineering Proceedings, vol. 84, no. 1, p. 66, 2025.

[9] P. K. Tiwary, P. Johri, A. Katiyar, and M. K. Chhipa, "Deep Learning-Based MRI Brain Tumor Segmentation with EfficientNet-Enhanced UNet," IEEE Access, 2025.

[10] Y. Lyu and X. Tian, "MWG-UNet++: Hybrid transformer U-Net model for brain tumor segmentation in MRI scans," Bioengineering, vol. 12, no. 2, p. 140, 2025.

[11] T. M. Angona and M. R. H. Mondal, "An attention based residual U-Net with swin transformer for brain MRI segmentation," Array, vol. 25, p. 100376, 2025.

[12] D. J. Ghadimi et al., "Deep Learning-Based Techniques in Glioma Brain Tumor Segmentation Using Multi-Parametric MRI: A Review on Clinical Applications and Future Outlooks," Journal of Magnetic Resonance Imaging, vol. 61, no. 3, pp. 1094-1109, 2025.

[13] N. Rasool and J. I. Bhat, "A critical review on segmentation of glioma brain tumor and prediction of overall survival," Archives of Computational Methods in Engineering, vol. 32, no. 3, pp. 1525-1569, 2025.

[14] R. C. Gonzalez, Digital image processing. Pearson education india, 2009.
[15] M. M. Saleh, M. E. Salih, M. A. Ahmed, and A. M. Hussein, "From traditional methods to 3d u-net: A comprehensive review of brain tumour segmentation techniques," Journal of Biomedical Science and Engineering, vol. 18, no. 1, pp. 1-32, 2025.
[16] J. C. Bezdek, L. Hall, and L. P. Clarke, "Review of MR image segmentation techniques using pattern recognition," Medical physics, vol. 20, no. 4, pp. 1033-1048, 1993.
[17] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Muller-Gartner, "Markov random field segmentation of brain MR images," IEEE transactions on medical imaging, vol. 16, no. 6, pp. 878-886, 1997.
[18] D. Zikic et al., "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in International conference on medical image computing and computer-assisted intervention, 2012: Springer, pp. 369-376.
[19] A. Criminisi and J. Shotton, Decision forests for computer vision and medical image analysis. Springer Science & Business Media, 2013.
[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431-3440.
[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, 2015: Springer, pp. 234-241.
[22] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in International conference on medical image computing and computer-assisted intervention, 2016: Springer, pp. 424-432.
[23] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 fourth international conference on 3D vision (3DV), 2016: Ieee, pp. 565-571.
[24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2017.S. M. Rasa et al., "Brain tumor classification using fine-tuned transfer learning models on magnetic resonance imaging (MRI) images," Digital health, vol. 10, p. 20552076241286140, 2024.
[25] D. Rastogi et al., "Brain tumor detection and prediction in MRI images utilizing a Fine-Tuned transfer learning model integrated within deep learning frameworks," Life, vol. 15, no. 3, p. 327, 2025.
[26] S. Anari, G. G. De Oliveira, R. Ranjbarzadeh, A. M. Alves, G. C. Vaz, and M. Bendechache, "EfficientUNetViT: efficient breast tumor segmentation utilizing UNet architecture and pretrained vision transformer," Bioengineering, vol. 11, no. 9, p. 945, 2024.
[27] A. Mukasheva, D. Koishiyeva, G. Sergazin, M. Sydybayeva, D. Mukhammejanova, and S. Seidazimov, "Modification of U-net with pre-trained ResNet-50 and atrous block for polyp segmentation: Model TASPP-UNet," Engineering Proceedings, vol. 70, no. 1, p. 16, 2024.
[28] A. Sharma and P. K. Mishra, "Inception UNet architecture for breast tumor segmentation and detection using hybrid deep learning approach," Multimedia Tools and Applications, vol. 84, no. 24, pp. 28225-28263, 2025.
[29] O. Oktay et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.
[30] A. Vaswani et al., "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.
[31] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
[32] J. Chen et al., "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.
[33] H. Cao et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," in European conference on computer vision, 2022: Springer, pp. 205-218.
[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009: Ieee, pp. 248-255.
[35] A. Fateh, Y. Rezvani, S. Moayedi, S. Rezvani, F. Fateh, and M. Fateh, "BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification with Swin-HAFNet," arXiv preprint arXiv:2506.14318, 2025.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 4, No. 1, Spring & Summer 2026, pp. 95 - 111

Tumor Segmentation in MRI using a Transformer Encoder and Adaptive Attention Modules

Noor Isam Abdulnabi[1], Mansoor Fateh[2*], Saideh Ferdowsi[3]

[1] PHD Candidate, Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

[2] Associate Professor, Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

[3] Assistant Professor, School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, United Kingdom

Article Information

Original Research Paper

Received:
2025 November 25

Accepted:
2026 February 24

Keywords:
Segmentation, Attention Mechanism, Deep Learning, Brain Tumor, MRI

Corresponding Author*:
Mansoor_fateh@shahroodut.ac.ir

Abstract

Accurate and automatic segmentation of brain tumors from Magnetic Resonance Imaging (MRI) plays a vital role in diagnosis, treatment planning, and disease monitoring. While Convolutional Neural Network (CNN)-based architectures excel at extracting local features, they are limited in comprehending global image context; conversely, Transformer models are superior in modeling long-range and global dependencies, thereby addressing this CNN limitation. In this paper, we propose a novel hybrid U-shaped architecture that effectively combines the strengths of both approaches by utilizing a pre-trained Swin-Transformer backbone as the encoder to extract hierarchical and context-rich global features. The key innovation is the introduction of two sophisticated spatial attention modules to refine and adapt the encoder features specifically for the medical domain, along with a channel attention-aided upsampling module in the decoder to adaptively and optimally re-weight the information received from the skip connections. Evaluations conducted on the challenging BRISC dataset show that our proposed method outperforms previous state-of-the-art models, achieving an 80.6% score in Weighted IoU and 88.6% in the Dice Coefficient, thereby proving the efficiency of combining the Transformer with dual attention mechanisms.

: 10.22034/ABMIR.2026.24013.1189

E-ISSN: 2821-2037