Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2026.550141.2177
Volume 15, Number 2, Summer 2026, pp. 1-15
Received: 09/29/2025, Revised: 01/06/2026, Accepted: 02/18/2026

Automatic Segmentation of Multiple Sclerosis Lesions on MRI Images Using Adaptive Residual U-Net

Ali Fahmi Jafargholkhanloo[1*], Mousa Shamsi[2] and Mahdi Bashiri Bawil[3]
1*- Department of Engineering Sciences, Faculty of Advanced Technologies, University of Mohaghegh Ardabili, Namin, Iran.
2- Department of Bioelectric, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran.
3- - Department of Bioelectric, Faculty of Biomedical Engineering, Sahand University of Technology, Tabriz, Iran.
[1]a_fahmi@uma.ac.ir, [2]shamsi@sut.ac.ir, and [3] m_bashiri99@sut.ac.ir

Corresponding author's address: Ali Fahmi Jafargholkhanloo, Faculty of Advanced Technologies, University of Mohaghegh Ardabili, Namin, Iran.

Abstract- Multiple Sclerosis (MS) is a chronic and neurodegenerative disorder of the central nervous system, whose diagnosis, classification, and monitoring are primarily performed through Magnetic Resonance Imaging. Accurate segmentation of MS lesion boundaries plays an essential role in early diagnosis, disease severity assessment, and treatment monitoring. However, despite the reliability of manual segmentation, this process is labor-intensive, time-consuming, and requires substantial expertise. Current automated methods show limited sensitivity to small lesions and poor generalization across different acquisition parameters. A residual U-Net architecture incorporating adaptive patch extraction and hybrid loss optimization was developed. The approach combines probabilistic sampling to address class imbalance, residual connections to enhance gradient flow, and a composite Dice-Focal loss function tailored for small object segmentation. The proposed architecture was evaluated on the MSLesSeg dataset comprising 115 FLAIR sequences from 75 patients across multiple imaging centers. Experimental results demonstrate that the proposed architecture has an LTPR of 79.92%, a precision of 72.77%, a recall of 72.22%, and a Dice score of 72.49%, outperforming the classical U-Net and other compared methods in accurate and sensitive detection of MS lesions. Integration of adaptive sampling with residual U-Net architecture effectively addresses fundamental challenges in MS lesion segmentation.

Keywords- Adaptive Patch Extraction, Deep Learning, Multiple Sclerosis Lesions, Medical Image Segmentation, Residual Network, Neuroimaging.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2026.550141.2177
جلد ۱۵، شماره ۲، تابستان ۱۴۰۵، صفحه ۱-۱۵
دریافت: ۱۴۰۴/۰۷/۰۷، بازنگری: ۱۴۰۴/۱۰/۱۶، پذیرش: ۱۴۰۴/۱۱/۲۹

بخش‌بندی خودکار ضایعات مولتیپل اسکلروزیس در تصاویر MRI با استفاده از شبکه U-Net مانده‌ای تطبیقی

علی فهمی جعفرقلخانلو*[۱]، موسی شمسی[۲]، مهدی بشیری باویل[۳]
۱- گروه علوم مهندسی، دانشکده فناوری‌های نوین، دانشگاه محقق اردبیلی، نمین، ایران.
۲- گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی سهند، تبریز، ایران.
۳- گروه بیوالکتریک، دانشکده مهندسی پزشکی، دانشگاه صنعتی سهند، تبریز، ایران.
[1]a_fahmi@uma.ac.ir, [2]shamsi@sut.ac.ir, [3] m_bashiri99@sut.ac.ir

* نشانی نویسنده مسئول: علی فهمی جعفرقلخانلو، نمین، دانشگاه محقق اردبیلی، دانشکده فناوری‌های نوین، گروه علوم مهندسی.

چکیده- بیماری مولتیپل اسکلروزیس (MS) یک اختلال مزمن و زوال عصبی سیستم عصبی مرکزی بوده که بخش‌بندی دقیق مرز ضایعات MS نقش کلیدی در تشخیص زودهنگام، ارزیابی شدت بیماری و پایش روند درمان دارد. بخش‌بندی دقیق مرز ضایعات MS به‌دلیل توزیع نابرابر نمونه‌های ضایعه و پس‌زمینه، اندازه کوچک ضایعات و پروتکل‌های تصویربرداری ناهمگون موضوعی چالش‌برانگیز است. با وجود دقیق بودن بخش‌بندی دستی، این رویکرد فرآیندی زمانبر بوده و نیازمند تخصص بالا است. روش‌های خودکار فعلی در شناسایی ضایعات کوچک MS عملکرد ضعیفی داشته و تعمیم‌پذیری آن‌ها برای داده‌های اخذ شده در شرایط تصویربرداری متفاوت محدود است. برای حل چالش‌های موجود، معماری پیشنهادی ARU-Net، مبتنی بر رویکرد ResNet و U-Net، با ترکیب استخراج پچ تطبیقی و بهینه‌سازی خطای ترکیبی توسعه داده شد، به‌گونه‌ای که رویکرد نمونه‌برداری احتمالی عدم تعادل کلاس را کاهش می‌دهد. اتصالات باقی‌مانده جریان گرادیان را بهبود بخشیده و تابع خطای ترکیبی دایس-فوکال حساسیت مدل به ضایعات کوچک را افزایش می‌دهد. معماری پیشنهادی بر روی پایگاه داده MSLesSeg شامل ۱۱۵ توالی FLAIR از ۷۵ بیمار ارزیابی شد. نتایج آزمایش نشان می‌دهد که معماری پیشنهادی با نرخ تشخیص صحیح به ازای هر ضایعه ۷۹/۹۲٪، دقت ۷۲/۷۷٪، حساسیت ۷۲/۲۲٪ و دایس ۷۲/۴۹٪ نسبت به معماری U-Net سنتی و سایر روش‌های مقایسه شده عملکرد بهتری در شناسایی دقیق و حساس ضایعات MS دارد. ادغام نمونه‌برداری تطبیقی با معماری U-Net مانده‌ای چالش‌های بنیادی بخش‌بندی ضایعات MS را به‌طور مؤثر مورد خطاب قرار می‌دهد.

واژه‌های کلیدی: استخراج پچ تطبیقی، یادگیری عمیق، مولتیپل اسکلروزیس، بخش‌بندی تصاویر پزشکی، شبکه باقی مانده، تصویربرداری عصبی.

۱- مقدمه

بیماری مولتیپل اسکلروزیس (MS) یک اختلال التهابی مزمن، خودایمنی و نورودژنراتیو در سیستم عصبی مرکزی بوده که به‌طور عمده موجب زوال میلین و در موارد پیشرفته، آسیب آکسونی برگشت‌ناپذیر می‌شود. میلین، غلاف چربی محافظ اطراف فیبرهای عصبی، نقش حیاتی در انتقال سریع و مؤثر پیام‌های عصبی ایفا می‌کند. از این‌رو تخریب آن منجر به اختلال در هدایت عصبی و بروز علائم عصبی متنوعی از جمله ضعف عضلانی، اختلال بینایی، مشکلات تعادلی و ناهنجاری‌های شناختی می‌شود. این ضایعه عمدتاً بزرگسالان جوان بین سنین ۲۰ تا ۵۰ سال را درگیر می‌کند. برآوردها نشان می‌دهد که حدود ۲/۵ میلیون نفر در سراسر جهان به این بیماری مبتلا هستند. با این حال، موارد

کودکان نادر بوده و تنها حدود ۵ درصد از کل موارد MS را شامل می‌شوند. بر اساس داده‌های منتشر شده در اطلس MS، حداقل ۳۰۰۰۰ کودک در سراسر جهان از ۵۵ کشور مختلف مبتلا به MS مبتلا تشخیص داده شده‌اند. در طول زمان، سیر بالینی بیماران مبتلا به MS ممکن است به‌صورت پیش‌رونده یا عودکننده-بهبود یابنده[1] (RRMS) باشد. شایع‌ترین نوع این بیماری، یعنی RRMS، با دوره‌های متناوب عود (Relapse) و بهبودی نسبی یا کامل (Remission) مشخص می‌شود؛ به‌طوری‌که در فاز بهبودی، علائم و ضایعات ممکن است به‌طور موقت کاهش یافته یا پنهان شوند [۱، ۲].

تصویربرداری تشدید مغناطیسی[2] (MRI) ابزار استاندارد و قدرتمند برای تشخیص، طبقه‌بندی، و پایش بیماری MS محسوب می‌شود. پایش سیر بیماری از طریق MRI در طول زمان، به‌ویژه برای تشخیص ضایعات جدید، فعال (با جذب ماده حاجب) یا در حال گسترش، ابزار حیاتی در ارزیابی اثربخشی درمان‌های اصلاح کننده انواع بیماری و تصمیم‌گیری‌های بالینی در مورد ادامه یا تغییر درمان محسوب می‌شود. به‌منظور نمایش دقیق و جزئی ضایعات مغزی، انواع مختلفی از توالی‌های تصویربرداری MRI به‌کار گرفته می‌شوند. از جمله توالی‌های پرکاربرد در شناسایی ضایعات MS می‌توان به T_1، T_2 و توالی وارون‌سازی با تضعیف مایع مغزی نخاعی[3] (FLAIR) اشاره کرد. توالی‌های T_1 و T_2 مبتنی بر پاسخ هسته‌های هیدروژن پس از تحریک توسط پالس فرکانس رادیویی هستند. در توالی T_1 زمان بازگشت هسته‌های هیدروژن به حالت تعادل مغناطیسی اندازه‌گیری می‌شود، در حالی که در توالی T_2، زمان لازم برای از بین رفتن هم‌فازی بین اسپین‌های همجوار مورد بررسی قرار می‌گیرد. در تصاویر مبتنی بر T_1 ماده سفید (WM) روشن‌تر از ماده خاکستری (GM) بوده و مایع مغزی نخاعی[4] (CSF) به رنگ تیره ظاهر می‌شود. در مقابل، در تصاویر مبتنی بر T_2، ماده سفید تیره‌تر از ماده خاکستری بوده و CSF روشن دیده می‌شود. توالی FLAIR به‌طور ویژه‌ای برای خاموش‌سازی سیگنال CSF طراحی شده تا مشاهده ضایعات مغزی و نخاعی با وضوح بیشتری انجام گیرد. این فرآیند شامل اعمال یک پالس بازیابی معکوس برای حذف سیگنال CSF و سپس ثبت تصویر مبتنی بر T_2 با تأخیر زمانی مشخص است. تصاویر FLAIR به‌ویژه در برجسته‌سازی ضایعات MS بسیار مؤثر هستند؛ به‌طوری‌که این ضایعات به‌صورت ناحیه‌های پرشدت دیده شده و معمولاً از بافت‌های مجاور به‌خوبی قابل تفکیک هستند. در عمل، توالی FLAIR معمولاً مهم‌ترین ابزار برای تشخیص اولیه ضایعات MS به‌شمار می‌رود، در حالی‌که توالی‌های T_1 و T_2 برای تکمیل تحلیل و تأیید یافته‌های حاصل از FLAIR مورد استفاده قرار می‌گیرند.

همچنین، ضایعاتی که در تصویر مبتنی بر T_1 پس از تزریق گادولینیوم دارای افزایش سیگنال بوده و معمولاً بیانگر پلاک‌های جدید یا فعال شده اخیر هستند، که شاخصی برای فعالیت التهابی در حال جریان محسوب می‌شود [۱].

فرآیند برچسب‌گذاری دستی ضایعات MS در تصاویر MRI، علی‌رغم دقت بالای آن در صورت انجام توسط نورولوژیست‌ها یا رادیولوژیست‌های خبره، فرآیندی زمان‌بر، خسته کننده، و نیازمند مهارت تخصصی بالا است. علاوه بر آن، خطای فردی به‌طور اجتناب‌ناپذیر در نتایج تأثیرگذار بوده و می‌تواند باعث ناهمسانی در برآورد شدت بیماری یا پاسخ به درمان شود. از این‌رو، توسعه و به‌کارگیری روش‌های خودکار مبتنی بر پردازش تصویر و یادگیری ماشین اهمیت ویژه‌ای دارد، زیرا این روش‌ها علاوه بر کاهش زمان و هزینه، دقت و تکرارپذیری بالاتری را فراهم ساخته و می‌توانند به‌عنوان ابزاری کمکی برای متخصصان در تشخیص، پایش پیشرفت بیماری و ارزیابی اثربخشی درمان مورد استفاده قرار گیرند. در سال‌های اخیر، رویکردهای مبتنی بر یادگیری عمیق، نظیر شبکه‌های عصبی پیچشی[5] (CNN) و معماری‌های پیشرفته‌ای مانند U-Net، به‌دلیل توانایی استخراج ویژگی‌های پیچیده و برتری آن‌ها نسبت به رویکردهای سنتی مبتنی بر آستانه‌گذاری، خوشه‌بندی یا فیلترهای دستی مورد توجه بیشتری در زمینه بخش‌بندی تصاویر پزشکی قرار گرفته‌اند. از جمله مزیت این رویکردها می‌توان به بهبود دقت، مقاومت بیشتر در برابر نویز و ناهمگونی تصاویر اشاره کرد. با توجه به این مزایا، رویکردهای یادگیری عمیق، پایه و اساس بسیاری از مطالعات اخیر در حوزه بخش‌بندی خودکار ضایعات MS را تشکیل داده و زمینه‌ای مناسب برای بررسی و تحلیل پیشینه‌ی تحقیقات موجود را فراهم می‌سازند.

Valverde و همکاران [۳] رویکرد به‌کارگیری یک ساختار آبشاری[6] از دو شبکه عصبی پیچشی سه‌بعدی مبتنی بر پچ را جهت بخش‌بندی خودکار ضایعه WM در بیماران مبتلا به MS ارائه کردند. استفاده از پچ‌های سه‌بعدی امکان درنظرگیری اطلاعات مکانی و ساختاری سه‌بعدی مغز را فراهم می‌آورد. دو مرحله‌ای کردن فرآیند یادگیری باعث می‌شود که در مرحله اول شبکه حساسیت خود را در پیدا کردن پیکسل‌های ضایعه افزایش داده و در مرحله دوم خطای مثبت کاذب را کاهش دهد. عیب اصلی روش ارائه شده در وابستگی آن به تعداد پچ‌ها است. به‌دلیل استخراج یک پچ سه‌بعدی مجزا پیرامون هر وکسل هدف، فرآیند تحلیل نیازمند تکرارهای متعدد محاسباتی بوده که در تصاویر با ابعاد بزرگ منجر به افزایش چشمگیر زمان پردازش و کاهش کارایی محاسباتی می‌شود. Roy و همکاران [۴] به‌منظور

بخش‌بندی ضایعه MS روش مبتنی بر شبکه عصبی تمام پیچشی را ارائه کردند. معماری شبکه شامل چند مسیر پیچشی بوده که پچ‌های چندکاناله دوبعدی را پردازش کرده و خروجی مسیرهای پیچشی ابتدا الحاق شده و سپس یک مسیر پیچشی نهایی به جای لایه تمام متصل[۷] (FC) برای تولید تابع عضویت ضایعه روی کل پچ‌ها اعمال می‌شود. به‌کارگیری فیلترهای کوچک (۳×۳ و ۵×۵) و کاهش تدریجی تعداد فیلترها در هر مسیر پیچشی باعث می‌شود که مرزهای ضایعه بهتر حفظ شوند و تعداد پارامترهای آزاد کاهش یابد. در روش ارائه شده، حذف لایه FC باعث کاهش چشمگیر تعداد پارامترها شده و نیاز به Dropout یا روش‌های مشابه برای کنترل بیش‌برازش کاهش می‌یابد. استفاده از فیلترهای کوچک و مسیر پیچشی نهایی باعث شناسایی دقیق مرزهای ضایعه می‌شود. Aslani و همکاران [۵] یک روش خودکار برای بخش‌بندی ضایعه MS از تصاویر MRI با وزن‌های مختلف ارائه کردند که بر اساس یک شبکه عصبی پیچشی دوبعدی انتها به انتها برای بخش‌بندی مقطع‌های سه‌بعدی طراحی شده است. شبکه ارائه شده شامل مسیرهای موازی متعدد شبکه باقی مانده[۸] (ResNet) بدون اشتراک وزن بوده که هر مسیر به یک مدالیته‌ی مشخص (T_1، T_2 یا FLAIR) اختصاص دارد. برای ادغام اطلاعات حاصل از مدالیته‌های مختلف تصاویر MRI، از بلوک‌های ادغام ویژگی مدالیته‌های مختلف[۹] (MMFF) استفاده شده و برای بازگرداندن ابعاد ویژگی‌ها به رزولوشن اصلی تصویر، بلوک‌های بازنمونه‌برداری چندمقیاسی ویژگی‌ها[۱۰] (MSFU) به‌کار گرفته می‌شوند. خروجی شبکه، نقشه‌های احتمالاتی ضایعه بوده که با اعمال آستانه‌گذاری به یک طبقه‌بندی باینری تبدیل می‌شود. مزیت اصلی روش ارائه شده، به‌کارگیری همزمان از اطلاعات مدالیته‌های مختلف تصاویر MRI و چندرزولوشن است که با استفاده از بلوک‌های MMFF و MSFU امکان ترکیب ویژگی‌های دقیق و فضایی ضایعه را فراهم می‌کند. استفاده از مسیرهای موازی ResNet بدون اشتراک وزن باعث می‌شود که هر مدالیته به‌صورت جداگانه و بهینه ویژگی‌های خود را استخراج کرده و در نتیجه دقت فرآیند بخش‌بندی افزایش یابد. Wang و همکاران [۶]، برای بخش‌بندی سه‌بعدی تومور مغزی رویکرد مبتنی بر ترکیب شبکه عصبی پیچشی سه‌بعدی (3DCNN) و ترنسفورمر را ارائه کردند که با نام TransBTS معرفی شده است. در این روش، ابتدا ورودی تصاویر MRI حجمی با استفاده از 3DCNN به ویژگی‌های فضایی و عمقی کاهش یافته تبدیل می‌شود تا اطلاعات محلی سه‌بعدی استخراج شوند. سپس، این ویژگی‌ها برای ورودی به ترنسفورمر بازسازی می‌شوند. بخش رمزگشا[۱۱] با استفاده از ویژگی‌های تعبیه شده توسط ترنسفورمر و عملیات بازنمونه‌برداری تدریجی، نقشه‌ی بخش‌بندی دقیق سه‌بعدی را تولید می‌کند. برای کاهش پیچیدگی محاسباتی و حفظ اطلاعات محلی، ابتدا رزولوشن ویژگی‌ها با لایه‌های پیچشی کاهش یافته و سپس به شکل توکن[۱۲] به ترنسفورمر داده می‌شوند. مکان‌یابی هر توکن نیز با گنجایش موقعیت[۱۳] یادگیری شده ترکیب می‌شود تا اطلاعات مکانی حفظ شود. در انتها، بخش رمزگشا با بازنمونه‌برداری مرحله‌ای و اتصال‌های پرشی نقشه بخش‌بندی کامل با جزئیات فضایی دقیق تولید می‌کند. ترکیب همزمان اطلاعات محلی و وابستگی‌های ترنسفورمر باعث دقت بالاتر در بخش‌بندی پیچیده سه‌بعدی می‌شود. همچنین، موقعیت‌یابی توکن‌ها با گنجایش موقعیت باعث حفظ اطلاعات مکانی دقیق برای بخش‌بندی می‌شود. Hatamizadeh و همکاران [۷] معماری U-Net مبتنی بر ترنسفورمر (UNETR) را جهت بخش‌بندی سه‌بعدی تصاویر پزشکی معرفی کردند. در معماری‌های کلاسیک شبکه‌های عصبی تمام‌پیچشی، بخش رمزگذار[۱۴] از لایه‌های پیچشی استفاده می‌کند تا ویژگی‌های محلی و بخشی از ویژگی‌های سراسری را بیاموزد، اما به‌دلیل ماهیت محلی فیلترهای پیچشی، این شبکه‌ها در یادگیری وابستگی‌های بلندبرد[۱۵] محدودیت دارند. در UNETR، این مسئله با جایگزینی رمزگذار پیچشی با ترنسفورمر حل شده است. ترنسفورمر رمزگذار، ورودی تصویر سه‌بعدی را به‌صورت توالی پردازش کرده و بازنمایی‌های چندمقیاسی و سراسری از کل تصویر استخراج می‌کند. سپس، این بازنمایی‌ها از طریق اتصال‌های پرشی[۱۶] در سطوح مختلف به رمزگشا متصل می‌شوند. بخش رمزگشا مشابه U-Net طراحی شده و از این اطلاعات برای بازسازی نقشه‌ی نهایی بخش‌بندی استفاده می‌کند. Hatamizadeh و همکاران [۸] با ترکیب رویکرد ترنسفورمر با پنجره‌های جابه‌جا شده با معماری UNETR، نسخه بهبود یافته از این معماری (Swin UNETR) را جهت بخش‌بندی سه‌بعدی تصاویر مربوط به تومور مغزی ارائه کردند. در معماری ارائه شده، تصاویر MRI با مدالیته‌های مختلف ابتدا به توکن‌های سه‌بعدی تقسیم شده و به فضای گنجایش نگاشت شدند. سپس، به جای اعمال مکانیزم خودتوجهی[۱۷] در کل حجم سه‌بعدی (که در UNETR بسیار پرهزینه است)، از مکانیسم پنجره‌های جابه‌جا شونده استفاده شد. این کار باعث می‌شود مکانیزم خودتوجهی فقط در بلوک‌های محلی محاسبه شود، اما به‌دلیل جابجایی پنجره‌ها در لایه‌های متوالی، مدل همچنان قادر به یادگیری وابستگی‌های بلندبرد بین پچ‌های مختلف باقی می‌ماند. در نتیجه، محاسبات بسیار کارا بوده و حافظه کمتری مصرف می‌شود. Lu و همکاران [۹] یک نسخه سبک شده از معماری U-Net را ارائه کردند که با ساده‌سازی هم‌زمان بخش‌های رمزگذار و رمزگشا، پیچیدگی محاسباتی و تعداد پارامترهای شبکه را به‌طور قابل

توجهی کاهش می‌دهد. در این مدل، تعداد کانال‌های ویژگی در تمامی لایه‌ها یکپارچه شده و به‌جای چهار مرحله ادغام ویژگی در مقیاس یکسان در U-Net سنتی، از ادغام ویژگی در مقیاس کامل بهره گرفته شد تا اطلاعات دقیق و سطح‌بندی شده به‌صورت یکپارچه حفظ شود. علاوه بر این، ماژول Ghost برای تولید نقشه‌های ویژگی معادل با هزینه محاسباتی کمتر به‌کار گرفته شده است. در این ماژول، نیمی از نقشه‌های ویژگی از طریق عملیات پیچش استاندارد و نیم دیگر از طریق پیچش تفکیکی[۱۸] تولید شده و سپس با یکدیگر ترکیب شدند. هدف اصلی این معماری، حفظ دقت بخش‌بندی مشابه U-Net بوده به‌طوری که تعداد پارامترها و عملیات محاسباتی کاهش یافته و امکان کاربرد بهینه در وظایف مختلف بخش‌بندی تصاویر پزشکی با مصرف منابع کمتر فراهم شود. Sadeghibakhi و همکاران [۱۰]، یک روش خودکار برای بخش‌بندی ضایعه MS مبتنی بر تصاویر FLAIR ارائه کردند. در روش ارائه شده، از یک شبکه عصبی پیچشی مبتنی بر 3DRes-Net و ماژول توجه فضایی-کانالی[۱۹] استفاده شده است. فرآیند پیشنهادی شامل سه مرحله اصلی است. ابتدا پیش‌پردازش تصاویر با استفاده از رویکرد CLAHE انجام شده و لبه‌های استخراج شده با تصاویر اصلی ترکیب می‌شوند تا تصاویر چهاربعدی ایجاد شود. سپس، پچ‌های تصادفی با اندازه تعریف شده از تصاویر چهاربعدی استخراج شدند. در نهایت، این پچ‌ها وارد شبکه CNN مبتنی بر مکانیزم توجه شده و فرآیند بخش‌بندی ضایعه MS انجام شد. Rondinella و همکاران [۱۱] رویکرد مبتنی بر معماری U-Net تقویت شده از طریق لایه حافظه کوتاه‌مدت بازگشتی با عملیات پیچشی[۲۰] (CLSTML) و ماژول فشرده‌سازی و توجه[۲۱] (SA) را برای بخش‌بندی ضایعه MS از روی تصاویر MRI ارائه کردند. در روش ارائه شده، هر مقطع محوری به همراه دو مقطع مجاور به شبکه داده شده تا وابستگی‌های فضایی بین مقاطع توسط لایه LSTM پردازش شده و اطلاعات ترتیبی برای بهبود بخش‌بندی استخراج شود. مسیر بخش رمزگذار شامل بلوک‌های متراکم، SA و بلوک‌های انتقال است. مسیر بخش رمزگشا به‌صورت متقارن طراحی شده که در آن اطلاعات فضایی جزئی با استفاده از اتصال پرشی حفظ می‌شود. ماژول SA به شبکه اجازه می‌دهد کانال‌های حاوی ویژگی‌های اطلاعاتی تقویت و کانال‌های غیرمفید نادیده گرفته شود و بدین ترتیب توجه دقیق‌تری بر روی مناطق ضایعه اعمال شود. Rondinella و همکاران [۱۲] به بررسی به‌کارگیری مدل‌های انتشار جهت بخش‌بندی ضایعه MS از روی تصاویر MRI پرداختند. این مدل‌ها که نوعی معماری مولد مبتنی بر فرآیند انتشار هستند، با اعمال نویز گام به گام به تصویر اولیه در مرحله پیش‌رونده و بازسازی تصویر اصلی در مرحله معکوس، قادر به یادگیری توزیع داده و تولید تصاویر با جزئیات واقعی و بافت طبیعی هستند. Pasquale و همکاران [۱۳] به ارائه‌ی یک چارچوب مبتنی بر اجماع برای بهبود بخش‌بندی ضایعه MS پرداختند. هدف این چارچوب، ترکیب خروجی‌های چندین مدل یادگیری عمیق پیشرفته موجود است تا دقت و پایداری بخش‌بندی افزایش یابد. در این روش، چندین الگوریتم بخش‌بندی مستقل روی تصاویر T_1 و FLAIR اجرا شده و سپس خروجی‌های هر الگوریتم با استفاده از استراتژی اجماع ادغام شدند. این رویکرد بدون نیاز به آموزش مجدد مدل‌ها، امکان استفاده از مدل‌های آماده و ارزیابی عملکرد آن‌ها روی داده‌های واقعی را فراهم می‌کند. در [۳] باید برای هر وکسل یک پچ سه‌بعدی استخراج شود. این فرآیند باعث افزایش پیچیدگی محاسباتی شبکه می‌شود. همچنین، به‌منظور مقابله با عدم توازن شدید میان کلاس ضایعه و پس‌زمینه، از کاهش نمونه‌برداری کلاس منفی همراه با آموزش آبشاری شبکه استفاده شده است. این رویکرد اگرچه می‌تواند تعادل نسبی در داده‌ها ایجاد کند، اما منجر به از دست رفتن بخشی از اطلاعات مفید و همچنین افزایش پیچیدگی معماری می‌شود. در [۴] جهت جلوگیری از بیش‌برازش، رویکرد Dropout به‌کار گرفته شده است. این کار باعث می‌شود برخی نورون‌ها به صورت تصادفی در هر epoch غیرفعال شوند. پیچیدگی محاسباتی بالای بخش ترنسفورمر در [۶] منجر به مصرف حافظه و زمان پردازش قابل توجهی برای حجم کامل داده MRI می‌شود. از طرف دیگر، نیاز به بازسازی پیچیده ویژگی‌ها برای تبدیل نگاشت ویژگی به توکن و برعکس، مراحل پردازش را طولانی‌تر و پرچالش می‌کند. همچنین، برای حفظ دقت بالا در بخش‌بندی، داده‌ها باید به شکل سه‌بعدی و با ابعاد بزرگ نگه داشته شوند، که می‌تواند سرعت فرآیند ارزیابی را افزایش دهد. به‌دلیل اینکه مکانیزم خودتوجهی در داده‌های سه‌بعدی محاسبات بسیار پرهزینه‌ای دارد، یکی از معایب اصلی UNETR [۷] این است که استفاده از ترنسفورمر به‌عنوان رمزگذار نیازمند حجم بالایی از داده‌های آموزشی و منابع محاسباتی سنگین است. علاوه بر این، ترنسفورمرها به‌صورت ذاتی فاقد بایاس القایی هستند و بدون داده‌های کافی ممکن است در استخراج ویژگی‌های محلی و ساختارهای ریز مانند ضایعات کوچک MS عملکرد ضعیف‌تری نسبت به شبکه‌های پیچشی کلاسیک داشته باشند. همچنین، وابستگی شدید به پچ‌بندی می‌تواند منجر به از دست رفتن جزئیات بافتی در مرزهای نواحی مهم شود. مدل‌های انتشار دقت بالایی در جزئیات و بافت تصاویر دارند، اما پیچیدگی محاسباتی بسیار بالا، نیاز به زمان و حافظه زیاد برای آموزش محدودیت اصلی این مدل‌ها است. علاوه بر این، استفاده مستقیم از مدل‌های انتشار

برای بخش‌بندی پیکسلی سه‌بعدی به منابع محاسباتی عظیمی نیاز دارد که مقیاس‌پذیری آن در کاربردهای بالینی محدود می‌شود.

به‌منظور حل چالش‌های بخش‌بندی ضایعه MS در مدل‌های سنتی، معماری U-Net مانده‌ای تطبیقی (ARU-Net) پیشنهاد می‌شود که نوآوری‌های اصلی آن به‌صورت زیر خلاصه می‌شود:

- **استخراج پچ تطبیقی:** پراکنده بودن ضایعات MS باعث می‌شود که آموزش مستقیم شبکه‌های عصبی عمیق روی کل تصویر با چالش جدی عدم توازن نمونه‌ها مواجه شده و حساسیت مدل در تشخیص ضایعات کوچک کاهش یابد. برای حل این مشکل، رویکرد نمونه‌برداری پچ‌های هدفمند ارائه می‌شود که در آن تصاویر به پچ‌های کوچک تقسیم شده و فرآیند نمونه‌برداری به‌صورت هوشمند انجام می‌شود. پچ‌هایی که تعداد کافی از پیکسل‌های مربوط به ضایعه دارند همیشه برای آموزش انتخاب می‌شوند، در حالی که پچ‌های پس‌زمینه با احتمال متناسب با نزدیکی به مرز ضایعات انتخاب می‌شوند. همچنین، برای تشخیص بهتر مرزها، لبه‌های ضایعه با آشکارساز کنی شناسایی شده و جهت افزایش احتمال انتخاب پچ‌های مربوط به لبه‌ی ضایعه، عملیات مورفولوژیکی انبساط[۲۲] به‌کار گرفته می‌شود. این رویکرد باعث تمرکز مدل روی ضایعات کوچک و لبه‌های آن شده و همزمان باعث کاهش مصرف حافظه و پیچیدگی محاسباتی می‌شود.
- **معماری ResU-Net:** به‌منظور بهبود جریان گرادیان در شبکه و امکان یادگیری ویژگی‌های سلسله مراتبی و پیچیده ضایعات MS، رویکرد مبتنی بر ترکیب ساختار کلاسیک رمزگذار-رمزگشا U-Net و بلوک‌های باقی‌مانده معرفی می‌شود. در این معماری، هر بلوک باقی‌مانده شامل چند لایه کانولوشن متوالی با نرمال‌سازی دسته‌ای[۲۳] (BN) و تابع فعال‌ساز ReLU بوده که در آن ورودی بلوک مستقیماً به خروجی اضافه می‌شود تا از مشکل ناپدید شدن گرادیان جلوگیری شود. مدل‌های سنتی U-Net اغلب در بازسازی جزئیات بسیار کوچک ضایعات MS عمل‌کرد مناسبی ندارند. با اعمال بلوک‌های باقی‌مانده در هر دو مسیر رمزگذار و رمزگشا، مدل نه‌تنها جریان یادگیری را پایدار نگه می‌دارد، بلکه توانایی بازنمایی سلسله‌مراتبی و پیچیده ویژگی‌های ضایعه را افزایش می‌دهد. به‌کارگیری این رویکرد باعث می‌شود که جزئیات ظریف و مرزهای مبهم ضایعات کوچک بهتر حفظ شده و توانایی مدل در شناسایی تغییرات کوچک تصویر افزایش یابد.
- **تابع خطای ترکیبی:** به‌منظور بهبود دقت مدل در شرایط عدم تعادل کلاس و افزایش حساسیت به پیکسل‌های دشوار مربوط به ضایعه، به‌کارگیری تابع خطای ترکیبی دایس و فوکال[۲۴] است. معیار خطای دایس تضمین می‌کند که همپوشانی بین ماسک پیش‌بینی شده و ماسک واقعی بیشینه شود، در حالی که معیار خطای فوکال بر روی پیکسل‌های چالش‌برانگیز و مناطق کوچک ضایعه متمرکز می‌شود. ترکیب این دو تابع خطا موجب می‌شود مدل هم قادر به حفظ دقت کلی و هم حساس به جزئیات مهم باشد.

۲- معماری پیشنهادی برای بخش‌بندی ضایعه MS

چالش‌های مربوط به بخش‌بندی ضایعات MS شامل اندازه کوچک، شکل نامنظم، پراکندگی در نواحی مختلف مغز و نسبت بسیار پایین در برابر بافت سالم هستند. این مسائل باعث می‌شود که مدل‌های یادگیری عمیق نتوانند به‌خوبی الگوهای مربوط به ضایعات کوچک را تشخیص دهند. همچنین، نسبت پیکسل‌های ضایعه به پس‌زمینه در اکثر تصاویر بسیار ناچیز است. این عدم تعادل سبب می‌شود که مدل‌ها تمایل به یادگیری بافت سالم را داشته و دقت بخش‌بندی ضایعات کاهش یابد. ترکیب رویکردهای مطرح شده باعث می‌شود که معماری پیشنهادی یک راهکار دقیق، مقاوم و کم‌هزینه محاسباتی برای بخش‌بندی ضایعه MS باشد، که در آن محدودیت مدل‌های سنتی معماری U-Net نظیر حساسیت کم به ضایعات کوچک، مصرف بالای حافظه و مشکل عدم تعادل کلاس به‌طور مؤثر اصلاح شود.

معماری U-Net استاندارد، به‌عنوان چارچوب مبتنی بر رمزگذار، رمزگشا با طراحی سلسله‌مراتبی و اتصالات پرشی، زیربنای بسیاری از مدل‌ها در بخش‌بندی تصاویر پزشکی است. با این وجود، ویژگی‌های ساختاری آن شامل کانولوشن‌های متوالی بدون مسیر میان‌بر (Shortcut) و انتقال مستقیم ویژگی‌های سطح پایین از طریق اتصالات میان‌راهی پردازش‌نشده[۲۵]، در مسائل حساس به جزئیات ریز مانند ضایعات کوچک MS محدودیت دارد. نسخه‌های موجود ResUNet با بهره‌گیری از یادگیری باقی‌مانده تلاش کرده‌اند عمق شبکه و پایداری مسیر گرادیان را افزایش دهند، اما عمدتاً بازطراحی آن‌ها به مسیر رمزگذار محدود شده و مسیر رمزگشا بدون بلوک‌های باقی‌مانده و بهینه‌سازی مسئله‌محور برای ویژگی‌های خاص تصاویر MRI (مانند ضایعات MS) حفظ شده است. معماری پیشنهادی (ARU-Net) این محدودیت‌ها را در سه بخش اصلاح می‌کند: (۱) طراحی مجدد بلوک‌های باقی‌مانده در

مسیر رمزگذار و رمزگشا، (۲) سازگارسازی عمق شبکه با اندازه و ویژگی‌های ضایعات کوچک MS، و (۳) استفاده از استخراج پچ تطبیقی برای تمرکز بیشتر روی ساختارهای کم‌کنتراست و با اهمیت.

ساختار کلی مقاله به این صورت سازمان‌دهی می‌شود که در بخش ۲-۱ به ارائه‌ی روش پیشنهادی جهت استخراج پچ خواهیم پرداخت. در بخش ۲-۲ معماری ARU-Net پیشنهادی معرفی می‌شود. در بخش ۲-۳ تابع خطای ترکیبی دایس و فوکال توضیح داده می‌شود. بخش ۳ نتایج معماری پیشنهادی و مقایسه‌ی عمل‌کرد آن با سایر معماری‌ها را ارائه می‌دهد. سرانجام، نتیجه‌گیری و ارائه‌ی پیشنهادها در بخش ۴ ارائه خواهد شد.

۲-۱- روش پیشنهادی برای استخراج پچ

بخش‌بندی ضایعات MS براساس تصاویر MRI یکی از مهمترین کارها در حوزه‌ی پردازش تصویر پزشکی محسوب می‌شود. از دلایل اصلی این موضوع می‌توان به موارد زیر اشاره کرد:

- **ویژگی‌های مورفولوژیکی ضایعات:** ضایعات MS معمولاً دارای اندازه کوچک (معمولاً کمتر از ۱۰ میلی‌متر)، شکل نامنظم و پراکندگی تصادفی در نواحی مختلف مغز هستند. این موضوع باعث می‌شود که مدل‌های مبتنی بر CNN و معماری‌های سنتی U-Net دقت پایینی در شناسایی ضایعات کوچک داشته باشند [۱۰،۱۴-۱۵].
- **عدم تعادل شدید کلاس‌ها:** تعداد پیکسل‌های ضایعه در مقایسه با بافت سالم بسیار کم است. به‌طور معمول نسبت تعداد پیکسل‌های مربوط به ضایعه به تعداد پیکسل‌های پس‌زمینه کمتر از ۱٪ است. این عدم تعادل شدید باعث می‌شود که مدل‌های یادگیری عمیق به سمت پیش‌بینی کلاس اکثریت (بافت سالم) گرایش پیدا کنند و ضایعات کوچک را نادیده بگیرند [۳].
- **محدودیت‌های محاسباتی:** پردازش کل حجم سه‌بعدی تصاویر MRI با رزولوشن بالا نیازمند حافظه و توان محاسباتی زیاد است. این موضوع کاربرد بالینی را محدود می‌کند [۱۰، ۱۵].

یکی از راهکارهای مناسب برای حل مشکلات ذکر شده، به‌کارگیری رویکردهای مبتنی بر استخراج پچ است. در سال‌های اخیر روش‌های مختلفی برای این منظور ارائه شده است. در روش چندمقیاسی [۳]، پچ‌هایی با اندازه‌های مختلف از یک نقطه مرکزی استخراج شده و به‌صورت همزمان به شبکه داده می‌شوند. مزیت این روش ترکیب اطلاعات محلی و کلی است، اما هزینه محاسباتی بالا و پیچیدگی معماری محدودیت‌های اصلی آن محسوب می‌شوند. در [۵] از روش استخراج پچ ثابت استفاده شده است که در آن تصاویر MRI به بلوک‌هایی با اندازه‌ی مشخص تقسیم شده و هر پچ به‌طور مستقل به شبکه داده می‌شود. این رویکرد مزایایی همچون کاهش نیاز به حافظه و ساده‌سازی آموزش شبکه دارد. بااین‌حال، محدودیت‌هایی مانند ناتوانی در شناسایی ضایعات کوچک یا با مرزهای نامنظم نیز گزارش شده است.

روش‌های ارائه شده در سال‌های اخیر عموما از پچ‌های ثابت استفاده کرده‌اند که معمولا در کشف ضایعات کوچک MS عملکرد ضعیفی دارند. برای حل این مسئله، رویکرد استخراج پچ تطبیقی پیشنهاد می‌شود که براساس حضور یا عدم حضور ضایعه و نزدیکی به لبه‌های ضایعه طراحی شده است، به‌گونه‌ای که پچ‌هایی با تعداد کافی پیکسل ضایعه حتماً نگه داشته می‌شوند و پچ‌های بدون ضایعه، در صورت قرارگیری مرکز آن‌ها روی لبه‌ها، با احتمال تقویت شده انتخاب می‌شوند. این رویکرد برخلاف روش‌های ثابت یا چندمقیاسی، حساسیت مدل را نسبت به نواحی کوچک و مرزی ضایعه افزایش داده و به کاهش عدم تعادل کلاس کمک می‌کند. به بیان دیگر، در حالی‌که رویکردهای قبلی با مشکل عدم انعطاف در انتخاب پچ‌ها روبه‌رو بودند، روش پیشنهادی با تعریف یک تابع تطبیقی برای انتخاب اندازه و موقعیت پچ، هم دقت شناسایی ضایعات کوچک را بهبود می‌دهد و هم هزینه‌ی محاسباتی را کاهش می‌دهد. رویکرد پیشنهادی شامل ۶ مرحله اصلی بوده که در ادامه به توضیح جزئیات مراحل موردنظر پرداخته می‌شود:

(۱) اعمال فیلتر گوسین: در ابتدا، ماسک باینری (M) توسط فیلتر گوسین صاف می‌شود. این مرحله موجب کاهش نویزهای با فرکانس بالا و ناهمواری‌های کوچک در ماسک شده که می‌توانند در فرآیند شناسایی لبه باعث ایجاد لبه‌های کاذب شوند. اعمال فیلتر گوسین باعث تشخیص پایدار و دقیق لبه‌ها می‌شود.

$$\widetilde{M} = (255 \times M) * G_\sigma \qquad (۱)$$

که در آن G_σ کرنل گوسین با انحراف معیار σ است.

در این مطالعه، ماسک باینری به‌عنوان یک نگاشت دودویی هم‌اندازه با تصویر ورودی تعریف شده که مستقیماً از برچسب‌های مرجع استخراج می‌شود. در این ماسک، پیکسل‌های متناظر با ناحیه‌ی ضایعه با مقدار یک و سایر پیکسل‌ها با مقدار صفر مشخص شده تا مرز دقیقی میان ناحیه موردنظر و پس‌زمینه ایجاد شود. به‌منظور کاهش اثر نویزهای موضعی و رفع ناپیوستگی‌های مکانی موجود در ماسک اولیه، فرآیند هموارسازی جهت منسجم شدن ساختار فضایی ناحیه مربوط به ضایعه اعمال می‌شود. پس از اعمال فرآیند هموارسازی، ماسک تنها به‌منظور راهنمایی فرآیند

استخراج پچ‌ها به‌کار گرفته شده و نقش ماسک در این مرحله صرفاً تعیین پچ‌هایی است که بر اساس میزان حضور نواحی ضایعه یا تمرکز لبه‌ها باید نگهداری یا حذف شوند. این ماسک مشخص می‌کند کدام پچ‌ها شامل اطلاعات معنادار ضایعه هستند و کدام پچ‌ها تنها نشان دهنده‌ی پس‌زمینه بوده که باید با نرخ نمونه‌برداری کنترل‌شده و هدفمند انتخاب شوند. در نتیجه، ماسک باینری پس از هموارسازی مستقیماً در فرآیند آموزش شبکه یا محاسبه تابع هزینه استفاده نمی‌شود و فقط به‌عنوان یک مکانیزم کنترلی برای هدایت استخراج پچ‌ها و کاهش اثر عدم تعادل کلاس‌ها به‌کار گرفته می‌شود.

(۲) شناسایی مرزهای ضایعه با لبه‌یاب کنی: برای شناسایی مرزهای ضایعه لبه‌یاب Canny اعمال می‌شود. خروجی حاصل از این مرحله نواحی با گرادیان شدت بالا را مشخص می‌کند که معمولاً مرزهای ضایعه هستند. اطلاعات مبتنی بر لبه به الگوریتم امکان می‌دهد تا پچ‌هایی که شامل نواحی مرزی ضایعه هستند را با اولویت بالاتر انتخاب کند، چرا که این نواحی معمولاً بیشترین پیچیدگی را برای مدل‌های بخش‌بندی ایجاد می‌کنند.

$$E_0 = Canny(\widetilde{M}; TH_1, TH_2) \qquad (۲)$$

که در آن TH_1 و TH_1 آستانه‌های پایین و بالا برای لبه‌یاب کنی است.

(۳) تعریف ابعاد پچ و گام‌های لغزش: ابعاد پچ و گام‌های لغزش با مقدار گام متناظر با اندازه‌های پچ (K) به‌صورت مجموعه‌هایی از $\{P_h(k), P_w(k)\}$ و S(k) تعریف می‌شوند.

$$P = \{P_h(k), P_w(k)\}_{k=1}^{K}$$
$$S = \{s(k)\}_{k=1}^{K} \qquad (۳)$$

که در آن مجموعه P شامل ابعاد پچ‌ها و مجموعه S شامل گام‌های متناظر برای حرکت پنجره لغزان است. برای هر پچ، ارتفاع و عرض با (P_h, P_w) مشخص می‌شود. این ابعاد با توجه به اندازه ضایعات و رزولوشن تصاویر انتخاب می‌شوند. برای ضایعات کوچک، پچ‌های کوچکتر انتخاب می‌شوند تا مدل قادر به یادگیری جزئیات ریز و بافت دقیق ضایعه باشد، در حالی که برای ضایعات بزرگ، پچ‌های بزرگ‌تر می‌توانند اطلاعات بیشتری از زمینه را فراهم کنند. مقدار گام لغزشی $s(k)$ مشخص می‌کند که پنجره‌ی پچ هنگام پیمایش تصویر، در هر حرکت چه فاصله‌ای نسبت به موقعیت قبلی جابجا شود، به طوری که پوشش تصویر کامل و همپوشانی بین پچ‌ها حفظ گردد. معمولاً این گام‌ها کمتر از ابعاد پچ انتخاب می‌شوند تا همپوشانی مناسبی بین پچ‌ها ایجاد شود و مرزهای ضایعات از دست نرود. این رویکرد امکان استخراج مجموعه‌ای متنوع و جامع از پچ‌ها را فراهم می‌کند که شامل پچ‌های دارای ضایعه و پچ‌های پس‌زمینه است و در عین حال، حجم داده‌ها و هزینه محاسباتی کاهش می‌یابد.

(۴) پیمایش تصویر و استخراج پچ: با استفاده از اندازه و گام‌های تعریف شده، پنجره لغزان روی تصویر حرکت کرده و در هر موقعیت، پچ تصویر و ماسک متناظر استخراج می‌شود. این مرحله کل تصویر را پوشش داده و باعث می‌شود که یادگیری مدل به‌صورت محلی و جزئی انجام گیرد.

$$\begin{cases} I_{(x,y)}(P) = I[x: x + P_h, y: y + P_w, :] \\ M'_{(x,y)}(P) = \widetilde{M}[x: x + P_h, y: y + P_w, :] \end{cases} \qquad (۴)$$

که در آن $I_{(x,y)}(P)$ و $M'_{(x,y)}(P)$ به‌ترتیب نشان دهنده‌ی پچ تصویر (شامل تمامی کانال‌ها) و پچ ماسک متناظر هموار شده (شامل مقادیر باینری ضایعه و پس‌زمینه) هستند. همچنین، (x, y) مختصات گوشه بالای پچ است. علامت ":" برای انتخاب تمامی کانال‌های تصویر استفاده می‌شود. به این معنی که هر پچ شامل تمام پیکسل‌های محدوده مشخص شده در سطرها و ستون‌ها به همراه تمام کانال‌های موجود است. همین روش برای پچ ماسک نیز به‌کار می‌رود تا مقادیر باینری ضایعه و پس‌زمینه به‌طور کامل و دقیق در محدوده پچ انتخاب شوند.

(۵) محاسبه‌ی تعداد پیکسل‌های ضایعه در پچ: برای هر پچ استخراج شده، تعداد کل پیکسل‌های ضایعه محاسبه می‌شود. این معیار برای تصمیم‌گیری در مورد نگهداری یا حذف پچ به‌کار گرفته می‌شود. پچ‌هایی که تعداد پیکسل‌های ضایعه‌ی بیش از آستانه تعیین شده داشته باشند، به‌طور خودکار حفظ می‌شوند تا مناطقی که ضایعه‌ی بیشتری دارند همیشه در داده‌های آموزشی حضور داشته باشند. برای هر پچ، تعداد پیکسل‌های ضایعه (L) به‌صورت زیر محاسبه می‌شود:

$$L(x, y; P) = \sum_{u=0}^{P_h-1} \sum_{v=0}^{P_w-1} M'\,[x + u, y + h] \qquad (۵)$$

مقدار آستانه برای تعداد پیکسل‌های ضایعه در هر پچ توسط پارامتری تعیین می‌شود که در این مطالعه مقدار پیش‌فرض آن ۱۰ انتخاب شده است. بر اساس این آستانه، پچ‌هایی که شامل حداقل ۱۰ پیکسل ضایعه هستند، بدون اعمال نمونه‌برداری تصادفی مستقیماً در مجموعه آموزشی نگهداری می‌شوند. پچ‌هایی که تعداد پیکسل ضایعه آن‌ها کمتر از آستانه باشد، بسته به موقعیت مرکز پچ نسبت به نواحی لبه ضایعه و احتمال نمونه‌برداری پس‌زمینه، به‌صورت تصادفی انتخاب می‌شوند. این رویکرد باعث می‌شود که مدل در طول آموزش با پچ‌هایی که دارای ضایعات کافی هستند مواجه شود، حساسیت آن نسبت به

نواحی ضایعه افزایش یافته و در عین حال، تعادل بین پچ‌های ضایعه و پس‌زمینه حفظ شود.

(۶) تصمیم‌گیری برای نگه‌داری پچ (نمونه‌برداری تطبیقی): در صورتی که تعداد پیکسل‌های ضایعه در پچ کمتر از آستانه‌ی تعیین‌شده باشد، نگهداری پچ به‌صورت تصادفی با احتمال مشخص‌شده انجام می‌گیرد. این احتمال، بر اساس نقشه‌ی لبه و ضریب تقویت α $(\alpha > 1)$ تعدیل می‌شود، به‌گونه‌ای که پچ‌هایی که مرکز آن‌ها بر روی نواحی مرزی ضایعه قرار دارد، شانس بالاتری برای نگهداری خواهند داشت. تصمیم نهایی برای نگهداری یا حذف پچ با استفاده از نمونه‌برداری برنولی انجام می‌شود. نمونه‌برداری تطبیقی باعث می‌شود که پچ‌های پس‌زمینه به‌صورت انتخابی وارد مجموعه آموزش شوند و همزمان پوشش کافی برای مرزهای ضایعه حفظ شود. این فرآیند باعث بهبود حساسیت مدل نسبت به نواحی مرزی و کاهش اثر عدم تعادل کلاس‌ها می‌شود.

$$L_{New}(x,y;P) = \begin{cases} 1, L(x,y;P) \geq \tau \\ \rho.\left(1+(\alpha-1).E\left(c_{x,y}(P)\right)\right), L(x,y;P) < \tau \end{cases} \quad (۶)$$

$$c_{x,y}(P) = \left(x + \left\lfloor \frac{P_h}{2} \right\rfloor, y + \left\lfloor \frac{P_w}{2} \right\rfloor\right) \quad (۷)$$

$$E\left(c_{x,y}(P)\right) = \begin{cases} 1, if\ the\ patch\ center\ c_{x,y}(P) \in edge\ region\ E \\ 0, \quad otherwise \end{cases} \quad (۸)$$

$$Final\ Decision: keep \sim Bernoulli\left(L_{New}(x,y;P)\right) \quad (۹)$$

که در آن $E(.)$ ، τ و ρ به‌ترتیب نشان دهنده‌ی عضویت در ناحیه‌ی لبه، آستانه حداقل تعداد پیکسل ضایعه در پچ و احتمال نگهداری پچ فاقد ضایعه است. مقدار ρ نشان‌دهنده‌ی احتمال پایه نگه‌داری پچ‌هایی است که شامل ضایعه کافی نیستند. در این مطالعه، مقدار پیش‌فرض آن ۵٪ در نظر گرفته شده است. بدین ترتیب، پچ‌هایی که تعداد پیکسل ضایعه آن‌ها کمتر از آستانه تعیین شده باشد، با احتمال ρ به‌صورت تصادفی وارد مجموعه آموزش می‌شوند. این تنظیم به گونه‌ای انجام شده است که مدل در طول آموزش با تعداد مناسبی از نمونه‌های پس‌زمینه مواجه شود، در حالی که پوشش پچ‌های دارای ضایعه حفظ شده و حساسیت مدل نسبت به نواحی ضایعه کاهش نمی‌یابد. به علاوه، این رویکرد به تعادل بین پچ‌های ضایعه و پس‌زمینه کمک کرده و اثر نامتعادل بودن کلاس‌ها را کاهش می‌دهد. مقادیر مربوط به پارامترهای تابع خطای فوکال به‌صورت تجربی انتخاب شده‌اند. احتمال پیش‌بینی هر پیکسل توسط مدل، از خروجی تابع فعال‌سازی سیگموید شبکه حاصل شده و به‌طور طبیعی بین صفر و یک قرار دارد، که نشان دهنده‌ی میزان تعلق پیکسل به کلاس ضایعه است. ضریب تعدیل α برای متعادل‌سازی سهم پیکسل‌های ضایعه و پس‌زمینه برابر ۰/۲۵ و ضریب تمرکز γ برای افزایش حساسیت به پیکسل‌های دشوار و نواحی کوچک ضایعه برابر ۲ انتخاب شده است. این مقادیر ثابت در تمامی مراحل آموزش اعمال می‌شوند تا مدل ضمن کاهش اثر عدم تعادل کلاس‌ها، توجه بیشتری به نمونه‌های چالش‌برانگیز داشته باشد.

۲-۲- معماری ResU-Net پیشنهادی

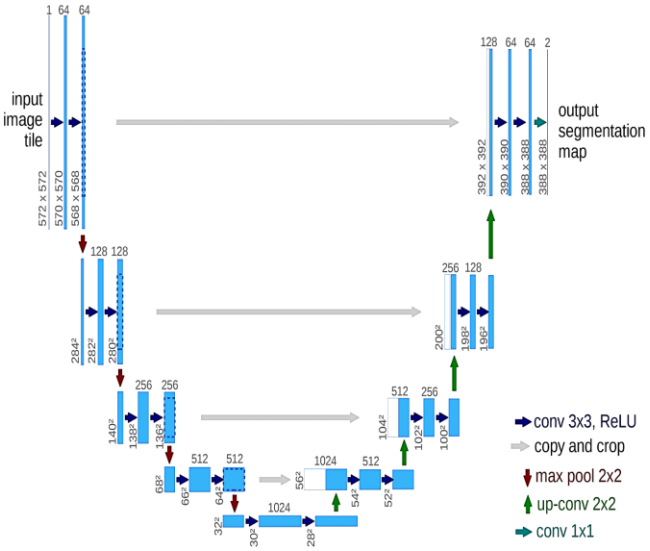
U-Net شبکه‌ی عصبی پیچشی جهت بخش‌بندی تصویر بوده که اولین بار توسط Ronneberger و همکاران [۱۶] در سال ۲۰۱۵ معرفی شد. همان‌طوری که در شکل ۱ نیز نشان داده شده است، این شبکه به شکل یک ساختار متقارن از رمزگذار-رمزگشا با اتصال پرشی طراحی شده و قادر است با داده‌های محدود نیز فرآیند بخش‌بندی تصویر را با دقت بالا انجام دهد. این معماری شامل ۴ بخش اصلی است [۱۶]:

(۱) بخش رمزگذار: فرآیند استخراج ویژگی‌های چندسطحی و فشرده‌سازی تصویر را انجام می‌دهد. هر بلوک رمزگذار شامل چند لایه‌ی پیچشی ۳×۳ با تابع فعال‌ساز ReLU است. پس از انجام فرآیند پیچش، جهت افزایش عمق ویژگی‌ها و کاهش پیچیدگی محاسباتی، یک لایه‌ی ادغام (Max Pooling) ۲×۲ به تصویر اعمال می‌شود. این بخش معمولاً شامل ۴ سطح بوده که هر سطح به استخراج ویژگی‌های سطح پایین‌تر تا سطح بالاتر کمک می‌کند.

(۲) بخش گره میانی (Bottle Neck): وظیفه اتصال بخش رمزگذار به بخش رمزگشا و استخراج ویژگی‌های سطح بالاتر را دارد. این بخش معمولاً شامل چند لایه پیچش با تابع فعال‌ساز ReLU و نرمال‌سازی دسته‌ای بوده که نمایانگر عمیق‌ترین و فشرده‌ترین سطح ویژگی‌ها است و اطلاعات استخراج شده از بخش رمزگذار را برای بخش رمزگشا آماده می‌کند.

(۳) بخش رمزگشا: وظیفه این بخش بازگرداندن نقشه ویژگی‌های فشرده شده در بخش گره میانی به ابعاد اصلی تصویر است. به‌عبارت دیگر، این بخش فرآیند افزایش ابعاد و بازیابی جزئیات مکانی را انجام می‌دهد.

(۴) بخش اتصال پرشی: برای جلوگیری از حذف جزئیات مکانی، ویژگی‌های بزرگنمایی شده با ویژگی‌های سطح متناظر در مسیر رمزگذار الحاق می‌شوند. این مرحله باعث می‌شود که رمزگشا علاوه بر اطلاعات فشرده‌ی سطح بالا، به جزئیات مرزی، لبه‌ها و ساختارهای ریز تصویر نیز دسترسی داشته باشد.



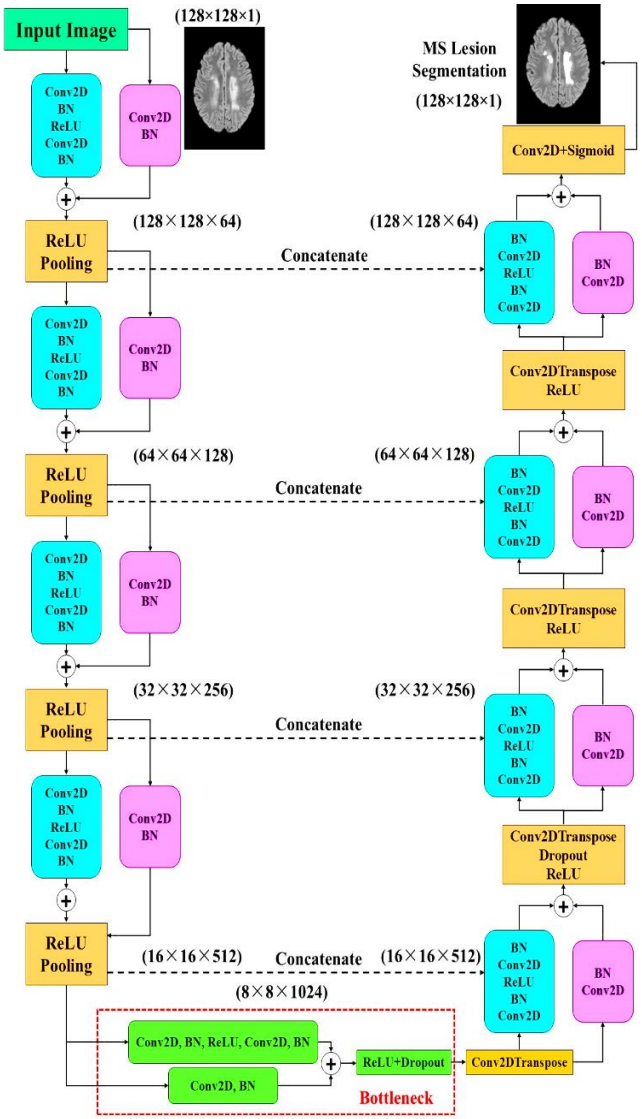
شکل ۱: نمایی از معماری شبکه U-Net [۱۶].

شبکه U-Net سنتی محدودیت در عمق شبکه، مشکل یادگیری ویژگی‌های پیچیده و حساسیت به نویز و عدم تعادل کلاس‌ها را دارد. شبکه‌های عمیق‌تر بدون اتصالات مستقیم قادر به یادگیری ویژگی‌های پیچیده و جزئیات ریز مانند ضایعه MS نیستند. از طرف دیگر، افزایش عمق شبکه بدون به‌کارگیری رویکرد Residual باعث ناپدید شدن گرادیان و کاهش دقت می‌شود. برای حل مشکلات مطرح شده، معماری ARU-Net پیشنهاد می‌شود که ساختار آن در شکل ۲ نشان داده شده است. در این ساختار، رنگ‌های مختلف به‌منظور تمایز اجزای کلیدی شبکه استفاده شده‌اند. بلوک‌های باقی‌مانده با رنگ آبی فیروزه‌ای نمایش داده شده‌اند و شامل دو کانولوشن متوالی به همراه BN و تابع فعال‌ساز ReLU هستند که جریان گرادیان را بهبود بخشیده و یادگیری ویژگی‌های سلسله‌مراتبی را تسهیل می‌کنند. کانولوشن‌های استاندارد همراه با BN در مسیر رمزگذار و رمزگشا با رنگ صورتی مشخص شده‌اند تا تقویت ویژگی‌ها و هم‌ترازی کانال‌ها را امکان‌پذیر سازند. لایه‌های ادغام، Conv2DTranspose و توابع فعال‌ساز مرتبط، که وظیفه کاهش و افزایش ابعاد را بر عهده دارند، با رنگ زرد نشان داده شده‌اند. بخش گره میانی شبکه، شامل دو کانولوشن متوالی و نرمال‌سازی با اتصال میانبر و Dropout برای افزایش پایداری یادگیری و جلوگیری از بیش‌برازش، با رنگ سبز مشخص شده است. همچنین، اتصالات پرشی با خطوط نقطه‌چین نشان داده شده‌اند تا اطلاعات مکانی و ساختاری سطح پایین را به رمزگشا منتقل کنند.

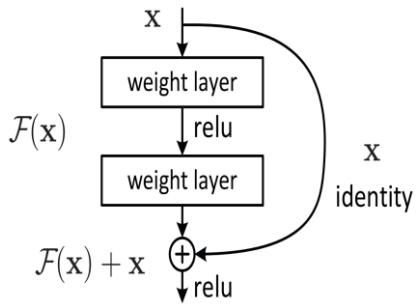
بلوک Residual یکی از نوآوری‌های معماری پیشنهادی بوده که در سال ۲۰۱۶ توسط He و همکاران [۱۷] معرفی شد. ساختار این بلوک در شکل ۳ نشان داده شده است. در معماری پیشنهادی، هر بلوک Residual شامل دو لایه‌ی پیچش ۳×۳ متوالی بوده که پس از هر پیچش، عملیات نرمال‌سازی دسته‌ای (BN) و در اولین لایه، تابع فعال‌ساز ReLU اعمال می‌شود. به‌منظور جلوگیری از افت اطلاعات، خروجی این مسیر اصلی با ورودی اولیه بلوک از طریق یک اتصال میان‌بر جمع شده که به‌صورت زیر تعریف می‌شود:

$$y = x + F(x) \qquad (۱۰)$$

که در آن $F(x)$ نشان دهنده‌ی توالی عملیات پیچش، نرمال‌سازی دسته‌ای و فعال‌سازی است. در مواردی که تعداد کانال‌های ورودی و خروجی یکسان نباشد، برای همسان‌سازی ابعاد، یک لایه پیچشی ۱×۱ همراه با نرمال‌سازی دسته‌ای بر روی مسیر میان‌بر اعمال می‌شود. در نهایت، یک تابع فعال‌ساز ReLU بر خروجی جمع اعمال می‌شود تا فرآیند غیرخطی‌سازی و یادگیری عمیق امکان‌پذیر شود.



شکل ۲: معماری Net-ARU پیشنهادی جهت بخش‌بندی ضایعه MS از روی تصاویر MRI.



شکل ۳: نمایی از ساختار بلوک Residual [۱۷].

در معماری U-Net استاندارد، هر بلوک تنها از دو کانولوشن ۳×۳ بدون مسیر میان‌بر تشکیل شده است. این ساختار با افزایش عمق شبکه باعث تشدید محوشدگی گرادیان و از دست رفتن اطلاعات سطح پایین می‌شود. در اغلب نسخه‌های گزارش‌شده از ResUNet برای بخش‌بندی ضایعات MS، بلوک‌های یادگیری باقی‌مانده عمدتاً محدود به مسیر رمزگذار بوده و مسیر رمزگشا ساختار کلاسیک U-Net را بدون تقویت نگه می‌دارد. در نتیجه، مرحله‌ی بازسازی در فرآیند Up-Sampling فاقد مکانیسم‌های تقویت ویژگی بوده و توان کافی برای بازنمایی دقیق ساختارهای با کنتراست پایین و اندازه کوچک ضایعات MS را ندارد. در معماری پیشنهادی، بلوک‌های باقی‌مانده به‌طور هماهنگ در هر دو مسیر رمزگذار و رمزگشا بازطراحی شده‌اند. این بلوک‌ها شامل توالی Conv , Normalization , ReLU , Conv , Normalization به همراه میانبر فرافکنی[۲۶] مبتنی بر کانولوشن ۱×۱ برای تطابق کانال‌ها هستند، که امکان عبور مستقیم اطلاعات و گرادیان‌ها از لایه‌های عمیق به سطح بالاتر را فراهم می‌کند. به‌کارگیری میانبر فرافکنی در مسیر رمزگشا به ویژه برای بازسازی دقیق لبه‌ها و ویژگی‌های با فرکانس بالا نقش کلیدی در آشکارسازی ضایعات کوچک MS دارد. علاوه بر این، بلوک‌های رمزگشا با تعداد فیلترهای متناسب و استفاده از Normalization طوری طراحی شده‌اند تا فرآیند بازسازی پایدارتر شده و احتمال نوسان گرادیان کاهش یابد.

۲-۳- تابع خطای ترکیبی دایس و فوکال

به‌منظور بهبود عملکرد شبکه در شناسایی ضایعات کوچک و دشوار MS و مقابله با عدم تعادل شدید کلاس‌ها، تابع خطای ترکیبی (Focal-Dice) به‌کار گرفته می‌شود که شامل هزینه دایس و هزینه فوکال است. این ترکیب باعث پوشش دقیق شبکه با تصاویر مرجع شده و حساسیت آن نسبت به پیکسل‌های دشوار می‌شود. معیار هزینه دایس به‌صورت زیر تعریف می‌شود [۱۸]:

$$\mathcal{L}_{Dice} = 1 - \frac{\sum_i y_i . \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i} \qquad (۱۱)$$

که در آن y_i و $\hat{y}_i$ به‌ترتیب نشان دهنده‌ی مقادیر واقعی و پیشبینی شده‌ی پیکسل i است.

برای مقابله با عدم تعادل شدید بین پیکسل‌های ضایعه و پس‌زمینه، معیار خطای فوکال به‌کار گرفته می‌شود [۱۹]:

$$\mathcal{L}_{Focal} = -\frac{1}{N}\sum_{i=1}^{N} \alpha(1-p_i)^{\gamma}.\log(p_i) \qquad (۱۲)$$

که در آن p_i احتمال پیشبینی پیکسل i، α ضریب تعدیل بین نمونه‌های مثبت (پیکسل‌های متعلق به ضایعه) و منفی (پیکسل‌های متعلق به پس‌زمینه) و γ ضریب تمرکز بر پیکسل‌های دشوار (لبه‌ها یا نواحی کوچک ضایعه) است.

براساس روابط (۱۱) و (۱۲) تابع خطای ترکیبی دایس-فوکال به‌صورت زیر تعریف می‌شود:

$$\mathcal{L}_{Hybrid} = \lambda_1 . \mathcal{L}_{Dice} + \lambda_2 . \mathcal{L}_{Focal} \qquad (۱۳)$$

که در آن λ_1 و λ_2 به‌ترتیب نشان دهنده‌ی ضریب وزن‌دهی برای بخش خطای دایس و خطای فوکال است. این ترکیب باعث می‌شود شبکه هم دقت کلی بالایی داشته باشد و هم نسبت به ضایعات کوچک و پیکسل‌های دشوار حساس باقی بماند. مقدار ضریب $\lambda_1 = 2$ برای مقیاس‌بندی مقدار دایس به منظور هم‌تراز کردن دامنه آن با وزن‌دهی بخش فوکال انجام می‌شود، تا هنگام ترکیب دو تابع خطا، سهم هر کدام متناسب و پایدار باقی بماند. بخش فوکال برای مدیریت عدم تعادل کلاس و تمرکز روی پیکسل‌های دشوار به کار می‌رود. انتخاب $\lambda_2 = 5$ باعث می‌شود که پیکسل‌های کوچک، لبه‌ها و نواحی چالش‌برانگیز ضایعه وزن بیشتری دریافت کنند و مدل حساسیت بیشتری نسبت به آن‌ها پیدا کند. این مقادیر براساس آزمون و خطا و ارزیابی عملکرد روی داده‌های آموزشی و اعتبارسنجی انتخاب شده است.

۳- نتایج پیاده‌سازی

در این مطالعه پایگاه داده‌ی MSLesSeg [۱] جهت بخش‌بندی ضایعات MS به‌کار گرفته شده است. این پایگاه شامل داده‌های MRIمربوط به ۷۵ بیمار (۴۸ زن و ۲۷ مرد) در بازه‌ی سنی ۱۸ تا ۵۹ سال است. هر بیمار بین یک تا چهار نوبت تصویربرداری (Timepoint)داشته که در مجموع ۱۱۵ سری تصویربرداری را تشکیل می‌دهد. برای هر نوبت تصویربرداری، سه مدالیته MRI شامل T_1، T_2 و FLAIR ثبت گردیده است. فرآیند تصویربرداری در بیمارستان گاریبالدی کاتانیا (ایتالیا) انجام شده و بخشی نیز به صورت پس‌نگرانه[۲۷] تهیه شده است. به‌عبارت دیگر، بخشی از تصاویر MRI به‌طور مستقل توسط بیماران از مراکز خصوصی تهیه و به پایگاه داده اضافه شده است. داده‌ها با شدت میدان ۱/۵ و ۳

تسلا به‌دست آمده و از آنجا که پروتکل یکسانی در تصویربرداری‌ها وجود نداشته است، ناهمگونی قابل توجهی در داده‌های خام قابل مشاهده است. در این مطالعه از تصاویر مبتنی بر وزن FLAIR استفاده شده است، چراکه این توالی با سرکوب سیگنال CSF، کنتراست ضایعات مولتیپل اسکلروزیس را افزایش داده و دقت بخش‌بندی را بهبود می‌دهد. تصاویر FLAIR دارای وضوح میانگین ۴۵۵/۸×۴۵۵/۸ پیکسل، فاصله پیکسلی ۰/۷۵×۰/۷۵ میلی‌متر، ضخامت اسلایس ۲/۶۱ میلی‌متر و میانگین مقادیر زمان وارونگی، زمان اکو و زمان تکرار به‌ترتیب برابر ۹۰/۵۵، ۲۰۷۲/۶۲ و ۵/۹۶ میلی‌ثانیه هستند. این پایگاه داده به دو بخش مجموعه آموزشی (۵۳ بیمار) و مجموعه آزمون(۲۲ بیمار) تقسیم شده است. مجموعه آموزشی شامل ۵۰ بیمار مبتلا به MS از نوع RRMS و ۳ بیمار مبتلا به MS پیشرونده ثانویه (SPMS) بوده، در حالی که مجموعه آزمون شامل ۲۱ بیمار RRMS و یک بیمار مبتلا به MS پیشرونده اولیه (PPMS) است.

جهت ارزیابی عمل‌کرد الگوریتم‌ها در فرآیند بخش‌بندی، چهار معیار نرخ تشخیص صحیح به ازای هر ضایعه[۲۸] (*LTPR*)، دقت[۲۹] (*PRE*)، حساسیت[۳۰] (*SEN*) و تشابه دایس (*DS*) به‌کار گرفته می‌شوند.

LTPR معیاری برای ارزیابی توانایی مدل‌های مختلف در شناسایی ضایعات جداگانه در تصاویر است. این معیار نشان می‌دهد که چه نسبتی از ضایعات واقعی موجود در تصویر، توسط مدل به‌درستی شناسایی شده‌اند.

$$PRE = \frac{TP}{TP + FP} \quad (۱۴)$$

$$SEN = \frac{TP}{TP + FN} \quad (۱۵)$$

$$DS(I_1, I_2) = \frac{2|I_1 \cap I_2|}{|I_1| + |I_2|} \quad (۱۶)$$

که در آن *TP*، *TN*، *FP*، *FN*، I_1 و I_2 به‌ترتیب نشان دهنده‌ی موارد مثبت درست شناسایی شده، موارد منفی درست شناسایی شده، موارد منفی به اشتباه مثبت شناسایی شده، موارد مثبت به اشتباه منفی شناسایی شده، تصویر مرجع و تصویر بخش‌بندی شده است.

۳-۱- تنظیم پارامترها

مقادیر تمامی پارامترها براساس آزمون و خطا و ارزیابی عملکرد روی داده‌های آموزشی و اعتبارسنجی انتخاب شده‌اند. جهت فراهم کردن شرایط مناسب برای پیاده‌سازی، اندازه واقعی تصاویر به ۱۲۸×۱۲۸ تغییر اندازه داده شد. در قسمت اعمال فیلتر گوسین، از کرنل ۳×۳ با مقدار $\sigma = 0.1$ استفاده شد. در بخش استخراج پچ، مقدار آستانه‌های پایین و بالای لبه‌یاب کنی به‌ترتیب برابر ۳۰ و ۱۰۰ درنظر گرفته شد. گام جابه‌جایی برابر با ۶۴ پیکسل در نظر گرفته شد تا همپوشانی کافی میان نواحی مجاور حفظ شود. برای اطمینان از حضور کافی اطلاعات مربوط به ضایعه در هر پچ، حداقل تعداد پیکسل‌های ضایعه در یک پچ برابر با ۱۰ درنظر گرفته شد. جهت مقابله با عدم تعادل داده‌ها میان پس‌زمینه و ضایعه، نرخ نمونه‌گیری از نواحی پس‌زمینه برابر ۰/۰۵ انتخاب شد، تا در نواحی بدون ضایعه، احتمال انتخاب پچ محدود شود. همچنین، جهت ایجاد تعادل بهتر میان پچ‌های حاوی ضایعه، حاشیه‌ی ضایعه و پس‌زمینه، نرخ نمونه‌برداری در نواحی مرزی با ضایعه پنج برابر بیشتر از نرخ نمونه‌برداری معمول پس‌زمینه انجام گرفت.

معماری شبکه‌ی پیشنهادی شامل چهار سطح در بخش رمزگذار بوده که تعداد فیلترها به‌ترتیب ۶۴، ۱۲۸، ۲۵۶ و ۵۱۲ افزایش یافته و پس از هر سطح، یک لایه MaxPooling2D اعمال می‌شود تا ابعاد فضایی کاهش یابد. در قسمت گره میانی نیز یک بلوک باقی‌مانده با ۱۰۲۴ فیلتر قرار دارد که جهت کاهش بیش‌برازش، از یک لایه Dropout با نرخ ۰/۳ استفاده شده است. در بخش رمزگشا، از لایه‌های Conv2DTranspose با اندازه هسته ۲×۲ و گام جابه‌جایی برابر با ۲ برای افزایش ابعاد فضایی استفاده شده و پس از هر لایه Transpose، نرمال‌سازی دسته‌ای جهت افزایش پایداری فرآیند یادگیری، کاهش نوسانات گرادیان و افزایش سرعت همگرایی شبکه اعمال شد. هر مرحله رمزگشا شامل الحاق با خروجی متناظر رمزگذار و سپس یک بلوک باقی‌مانده است. تعداد فیلترها در بخش رمزگشا به‌ترتیب ۵۱۲، ۲۵۶، ۱۲۸ و ۶۴ کاهش می‌یابد و لایه‌های Dropout با نرخ ۰/۲ برای سطوح بالاتر اعمال می‌شوند. خروجی شبکه از یک کانولوشن ۱×۱ با تابع فعال‌سازی سیگموید تشکیل شده که برای مسئله‌ی بخش‌بندی باینری مناسب است.

بهینه‌ساز Adam با نرخ یادگیری اولیه 1e-04 جهت آموزش شبکه به‌کار گرفته شد. به‌منظور تسریع همگرایی و جلوگیری از افت یادگیری، از کاهش خودکار نرخ یادگیری استفاده شد؛ به گونه‌ای که در صورت عدم بهبود در ۵ دوره متوالی اعتبارسنجی، نرخ یادگیری به نصف کاهش یافته و حداقل نرخ یادگیری برابر 1e-07 درنظر گرفته شد. همچنین، برای پیشگیری از بیش‌برازش و صرفه‌جویی در زمان آموزش، رویکرد Early Stopping فعال شد تا در صورت عدم بهبود در ۱۰ دوره متوالی، آموزش متوقف شده و وزن‌های بهترین دوره بازیابی شوند.

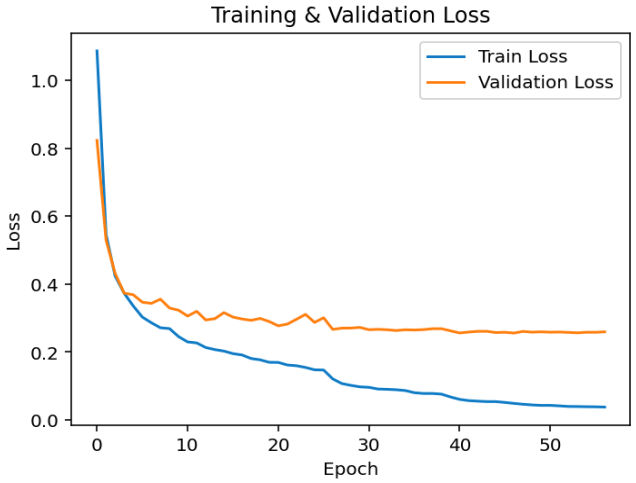
معماری پیشنهادی بر روی سیستمی با سخت‌افزار شامل پردازنده Intel Core i7–9750H با سرعت ۲/۶۰ گیگاهرتز، حافظه‌ی رم ۱۶ گیگابایت و یک پردازنده‌ی گرافیکی NVIDIA GeForce GTX 1650 پیاده‌سازی و آموزش داده شد.

۳-۲- نتایج پیاده‌سازی بر روی پایگاه داده‌ی MSLesSeg

روند تغییرات خطای مربوط به داده‌های آموزش و اعتبارسنجی مدل ReU-Net پیشنهادی در طول فرآیند یادگیری در شکل ۴ نشان داده شده است. در شروع فرآیند آموزش، مقدار تابع هزینه به‌دلیل عدم بهینه بودن وزن‌های اولیه زیاد بوده و با افزایش تعداد تکرار به سرعت کاهش می‌یابد. این روند نزولی نشان دهنده فرآیند همگرایی شبکه و بهبود تدریجی در پارامترهای بهینه‌سازی بوده که باعث همگرایی تدریجی شبکه و توانایی آن در استخراج الگوهای معنادار از داده است. در مقابل، خطای اعتبارسنجی نیز در مراحل اولیه کاهش چشمگیری داشته و سپس در حوالی مقدار epoch برابر ۵۰، به یک سطح پایدار رسیده که بیانگر تعادل مناسب بین یادگیری و تعمیم‌دهی مدل است. به‌کارگیری رویکرد خطای ترکیبی باعث می‌شود که مدل بتواند علاوه بر به حداقل رساندن عدم توازن بین پیکسل‌های ضایعه و پس‌زمینه، روی نواحی سخت‌تر تمرکز بیشتری داشته باشد. این موضوع در رفتار منحنی اعتبارسنجی نیز قابل مشاهده است، چرا که کاهش سریع در مراحل اولیه نشان دهنده حساسیت مدل نسبت به پیکسل‌های ضایعه کوچک و پراکنده است. به‌کارگیری رویکرد استخراج پچ تطبیقی باعث می‌شود که مجموعه داده آموزشی شامل توزیع متعادلی از نواحی ضایعه و پس‌زمینه باشد. این توازن در نمونه‌برداری منجر به پایداری نسبی منحنی خطای اعتبارسنجی می‌شود. به‌طور کلی، روند نزولی خطای آموزش همراه با پایداری خطای اعتبارسنجی نشان می‌دهد که معماری پیشنهادی به سطح مطلوبی از همگرایی و پایداری رسیده است. این رفتار نشان می‌دهد که شبکه پیشنهادی قادر به استخراج بازنمایی‌های تفکیک‌پذیر از نواحی ضایعه و پس‌زمینه بوده و در عین حال، با حفظ شکاف کنترل شده میان خطاهای آموزش و اعتبارسنجی، از بروز بیش‌برازش احتمالی در فرآیند آموزش جلوگیری نموده است. نتایج کیفی معماری پیشنهادی بر روی نمونه‌ای از تصاویر پایگاه داده‌ی MSLesSeg در شکل ۵ نشان داده شده است. ردیف اول شامل برش‌های انتخاب شده از تصاویر مبتنی بر وزن FLAIR مربوط به ضایعه MS است. ردیف دوم مربوط به ماسک‌های مرجع بوده که توسط متخصصان به‌صورت دستی برچسب‌گذاری شده‌اند. ردیف سوم نشان دهنده‌ی خروجی شبکه پیشنهادی است.

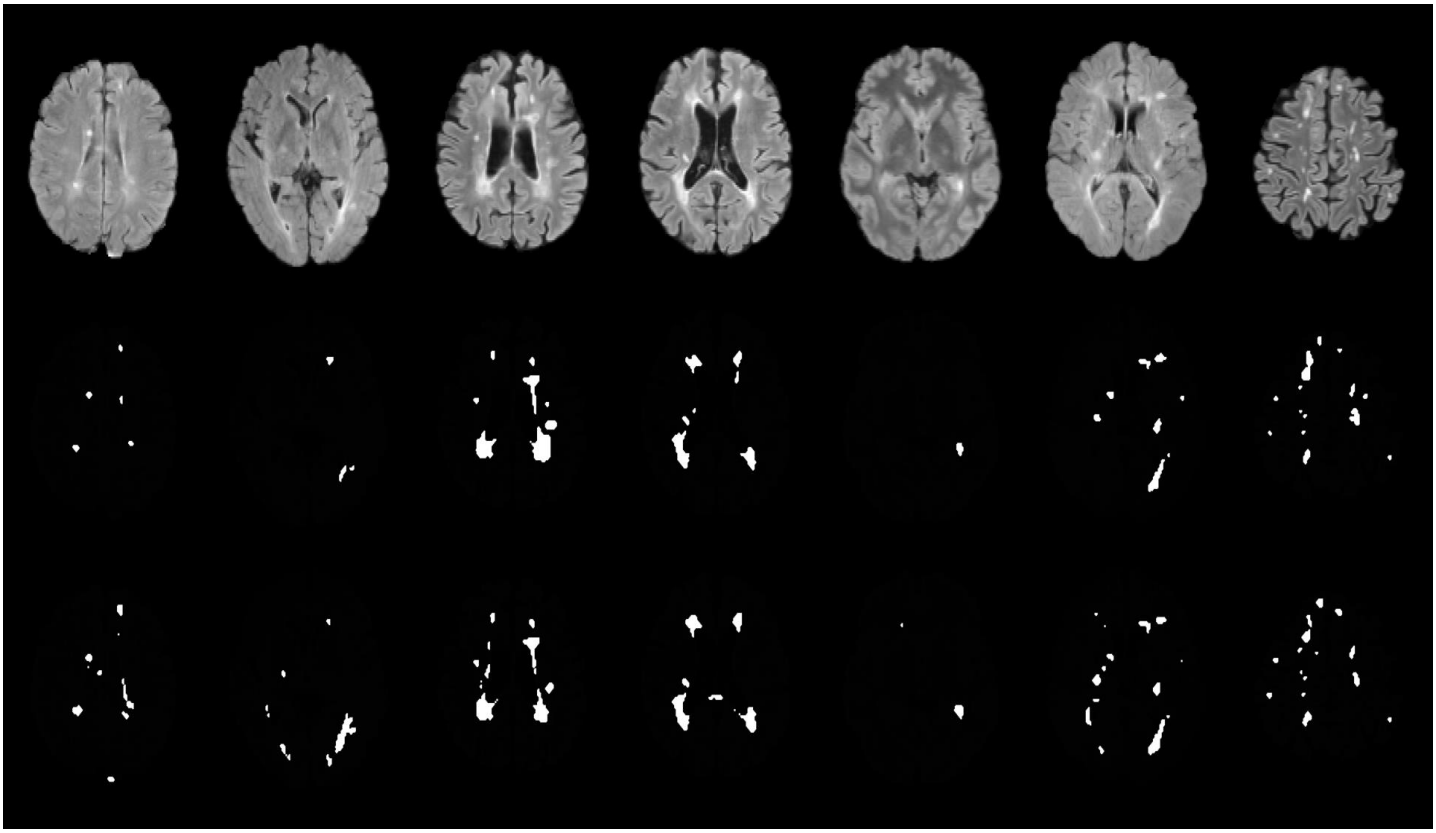
همچنین، نتایج کمی معماری‌های مختلف در بخش‌بندی ضایعه MS در جدول ۱ آورده شده است.

در اکثر برش‌ها، شبکه پیشنهادی توانسته است ضایعه MS را با دقت قابل‌توجهی شناسایی و محدوده‌ی آن‌ها را تقریباً منطبق با برچسب‌گذاری‌های دستی مشخص کند. همان‌گونه که در برخی برش‌ها نیز قابل مشاهده است، مدل قادر به شناسایی ضایعات کوچک و پراکنده نیز بوده است. این موضوع از اهمیت ویژه‌ای برخوردار است، چرا که ضایعات با اندازه محدود در اغلب رویکردهای خودکار با مشکل مواجه می‌شوند. عملکرد مناسب مدل در این زمینه نشان دهنده‌ی کارایی به‌کارگیری نوآوری‌های مطرح شده در بهبود حساسیت مدل نسبت به ضایعات کم حجم است. شکل ۵ نشان می‌دهد که معماری پیشنهادی در برش‌های میانی توانسته است مرزبندی ضایعات وسیع در کنار بطن‌های جانبی را با دقت مناسبی بخش‌بندی کند. اگرچه مدل پیشنهادی در حالت کلی عملکرد مناسبی در بخش‌بندی ضایعه MS دارد، اما در برخی موارد عملکرد نامناسبی نیز مشاهده می‌شود. به‌طور مثال، در بعضی برش‌ها بخش‌هایی از ضایعات کوچک به‌طور کامل شناسایی نشده‌اند (FN) و در مقابل، در برخی نواحی خروجی مدل پیشنهادی پیکسل‌هایی را بخش‌بندی کرده است که در برچسب‌های مرجع وجود ندارند (FP). این تفاوت‌ها می‌تواند ناشی از محدودیت کیفیت داده‌های ورودی، تنوع شدت سیگنال در تصاویر FLAIR، یا خطای برچسب‌گذاری انسانی باشند.



شکل ۴: نمایی از روند مقدار تابع هزینه بر روی داده‌های آموزش و اعتبارسنجی براساس معماری ARU-Net پیشنهادی.

مقایسه به‌دست آمده در جدول ۱ نشان می‌دهد که مسئله‌ی عدم توازن داده به‌طور مستقیم و در دو سطح داده و تابع هزینه در معماری ARU-Net پیشنهادی مدیریت شده است. در سطح داده، با به‌کارگیری نمونه‌برداری وفقی، احتمال انتخاب پچ‌های حاوی



شکل ۵: نمایی از خروجی معماری ARU-Net پیشنهادی بر روی پایگاه داده‌ی MSLesSeg جهت بخش‌بندی ضایعه MS.
(ردیف اول): تصاویر اصلی. (ردیف دوم): تصویر مرجع نشان دهنده‌ی نواحی مربوط به ضایعه MS. (ردیف سوم): خروجی معماری پیشنهادی.

ضایعه یا نواحی مجاور مرز آن افزایش یافته و در مقابل نمونه‌های پس‌زمینه با نرخ کنترل شده وارد فرآیند آموزش می‌شوند. در سطح مدل نیز، استفاده از تابع خطای ترکیبی دایس و فوکال باعث می‌شود که شبکه توجه بیشتری به نمونه‌های دشوار و نواحی با حجم کوچک ضایعه داشته باشد و در نتیجه، حساسیت مدل نسبت به ضایعات پراکنده و کوچک به‌طور مؤثر افزایش یابد. در تمامی معیارها، معماری شبکه پیشنهادی توانسته نسبت به سایر روش‌ها عملکرد بهتری در بخش‌بندی ضایعه MS داشته باشد. روش پیشنهادی نسبت به معماری معرفی شده در [۸] سبک‌تر و سریع‌تر است، چراکه نیازی به توکن‌سازی سه‌بعدی و مکانیزم توجه پیچیده‌ای ندارد. معماری پیشنهادی، با به‌کارگیری نمونه‌برداری وفقی و خطای ترکیبی معرفی‌شده، به‌طور خاص روی نواحی کوچک و دشوار تمرکز می‌کند، در حالی‌که معماری Swin UNETR برای رسیدن به چنین عملکردی نیازمند داده‌های بیشتر و تنظیمات دقیق است. همچنین، مدل‌های انتشار [۱۲] دقت بالایی در جزئیات و بافت تصاویر دارند، اما پیچیدگی محاسباتی بالا، نیاز به زمان و حافظه زیاد برای آموزش محدودیت اصلی آن‌هاست. علاوه بر این، استفاده مستقیم از این مدل‌ها برای

جدول ۱: مقایسه عملکرد معماری ARU-Net پیشنهادی با سایر معماری‌ها در بخش‌بندی ضایعه MS (میانگین ۵ فولد).

مدل	*LTPR*	*PRE*	*SEN*	*DS*
MSSegDiff [12]	۶۲/۵۰	۷۲/۴۱	۶۷/۱۹	۶۸/۵۱
UNETR [7]	۷۳/۳۲	۶۴/۹۶	۶۷/۵۱	۶۴/۲۱
Swin UNETR [8]	۷۰/۱۷	۷۱/۵۷	۶۶/۷۶	۶۷/۸۶
TransBTS [6]	۵۱/۱۰	۵۰/۲۷	۵۴/۴۷	۴۹/۱۷
Proposed ARU-Net	**۷۹/۹۲**	**۷۲/۷۷**	**۷۲/۲۲**	**۷۲/۴۹**

۴- نتیجه‌گیری و ارائه‌ی پیشنهادها

مقاله ارائه شده معماری U-Net مبتنی بر شبکه باقی مانده (ResU-Net) را جهت بخش‌بندی ضایعه MS از روی تصاویر MRI مبتنی بر وزن FLAIR معرفی می‌کند. معماری پیشنهادی با

ترکیب بلوک‌های باقی‌مانده در ساختار U-Net و استفاده از رویکرد استخراج پچ تطبیقی، توانست چالش‌های اساسی این حوزه را به‌طور مؤثر مدیریت کند. عدم تعادل بین کلاس‌ها، پراکندگی فضایی ضایعات و کوچک بودن نواحی آسیب دیده از جمله چالش‌های اصلی بود که در این مطالعه حل شدند. علاوه بر این، به‌کارگیری تابع خطای ترکیبی دایس-فوکال نیز به بهبود دقت در شناسایی پیکسل‌های دشوار و افزایش حساسیت مدل نسبت به ضایعات کوچک کمک کرد. اغلب رویکردهای موجود در بخش‌بندی تصاویر پزشکی از پچ‌های ثابت یا برش‌های منظم بهره می‌برند و فرآیند انتخاب پچ معمولاً بدون درنظرگرفتن تراکم ضایعه، فاصله از مرز ضایعه یا میزان اطلاعات مرتبط با کلاس مثبت انجام می‌شود. در مقابل، روش پیشنهادی یک رویکرد انتخاب پچ مبتنی بر محتوا را ارائه می‌دهد که تفاوت اساسی با کارهای پیشین دارد. در این رویکرد، پچ‌هایی که شامل تعداد قابل قبولی از پیکسل‌های ضایعه هستند به‌صورت قطعی در مجموعه‌ی آموزش قرار می‌گیرند. پچ‌های پس‌زمینه با احتمالی متناسب با فاصله آن‌ها از مرز ضایعه انتخاب شده و با استخراج مرز ضایعه و به‌کارگیری عملیات مورفولوژیک، احتمال قرارگیری پچ‌های حاوی مرز در فرآیند آموزش افزایش می‌یابد. نتایج تجربی بر روی پایگاه داده MSLesSeg نشان داد که مدل توسعه یافته عملکرد قابل توجهی در معیارهای ارزیابی مختلف داشته و بخش‌بندی مناسبی میان ضایعات و بافت‌های سالم ایجاد کرده است. ارزیابی مقایسه‌ای نشان داد که معماری پیشنهادی در تمامی معیارهای سنجش نسبت به روش‌های موجود برتری قابل توجهی دارد. معماری ARU-Net پیشنهادی از لحاظ محاسباتی سبک‌تر و سریع‌تر از معماری SwinUNETR است، زیرا نیازی به توکن‌سازی سه‌بعدی و مکانیزم‌های توجه پیچیده‌ای ندارد. معماری ResU-Net پیشنهادی با به‌کارگیری نمونه‌برداری تطبیقی و تابع خطای ترکیبی، تمرکز ویژه‌ای بر نواحی کوچک و دشوار دارد، در حالی که معماری SwinUNETR برای دستیابی به عملکرد مشابه نیازمند داده‌های بیشتر و تنظیمات دقیق‌تر است. در مقایسه با مدل‌های انتشار، اگرچه این مدل‌ها دقت بالایی در جزئیات و بافت تصاویر دارند، اما دارای محدودیت‌های اساسی هستند. پیچیدگی محاسباتی بالا، نیاز به زمان و حافظه زیاد برای آموزش، و نیاز به منابع محاسباتی عظیم برای بخش‌بندی پیکسلی سه‌بعدی از جمله این محدودیت‌ها هستند که مقیاس‌پذیری آن‌ها در کاربردهای بالینی را محدود می‌سازد. این مقایسه‌ها نشان دهنده‌ی تعادل مناسب معماری پیشنهادی بین دقت، کارایی محاسباتی و قابلیت پیاده‌سازی عملی است. با وجود عملکرد مطلوب معماری پیشنهادی در بخش‌بندی ضایعه MS، چندین محدودیت نیز وجود دارد که باید مورد توجه قرار گیرد. به‌کارگیری استخراج تطبیقی پچ‌ها باعث بهبود تعادل بین کلاس‌ها می‌شود، اما ممکن است باعث از دست رفتن برخی ارتباطات ساختاری در سطح کل تصویر نیز شود. این مسئله می‌تواند بر شناسایی ضایعات گسترده و پیوسته اثر منفی بگذارد. عدم استفاده از اطلاعات تکمیلی سایر مدالیته‌ها مانند T_1 و T_2 می‌تواند دقت نهایی مدل را محدود کند. معماری پیشنهادی صرفاً بر روی یک پایگاه داده‌ی مشخص آموزش و ارزیابی شده است. بنابراین، کارایی آن در داده‌های واقعی کلینیکی هنوز به‌طور کامل بررسی نشده است. برای بهبود عملکرد معماری پیشنهادی چندین کار برای تحقیقات آتی پیشنهادی می‌شود. (۱) ادغام مدالیته‌های مختلف تصاویر MRI می‌تواند حساسیت و دقت مدل را در تشخیص ضایعات MS افزایش دهد. (۲) گسترش معماری شبکه از طریق اضافه کردن ماژول‌های پیشرفته‌ای همچون مکانیزم توجه یا بلوک‌های متراکم به‌منظور مدل‌سازی بهتر وابستگی‌های مکانی و زمانی، می‌تواند موجب ارتقای کیفیت بخش‌بندی شود. (۳) به‌کارگیری رویکردهای نیمه‌نظارتی و خودنظارتی، از طریق استفاده از داده‌های فاقد برچسب یا دارای برچسب‌های ناقص، می‌تواند به‌عنوان راهکاری کارآمد در جهت حل مشکلات مربوط به کمبود برچسب‌های دقیق مورد استفاده قرار گیرد. (۴) مدیریت بهبودیافته‌ی عدم تعادل داده از طریق معماری‌های پیشرفته مانند شبکه‌های مولد متخاصم می‌تواند تعادل میان نمونه‌ها را حفظ کرده و قابلیت مدل در تشخیص ضایعات دشوار را افزایش داده تا امکان تولید داده‌های مصنوعی متعادل فراهم شود. انجام مطالعات چندمرکزی و ارزیابی عملکرد بر روی داده‌های متنوع کلینیکی برای تأیید تعمیم‌پذیری و کاربردپذیری عملی روش پیشنهادی ضروری است.

معماری ARU-Net پیشنهادی گامی مؤثر در جهت بهبود بخش‌بندی خودکار ضایعات مولتیپل اسکلروزیس محسوب شده و نتایج حاصل از این تحقیق نشان دهنده‌ی عملکرد مطلوب این روش برای کمک به تشخیص بالینی است. با به‌کارگیری پیشنهادهای ارائه شده، می‌توان انتظار داشت که نسل‌های بعدی این معماری عملکرد بهتری در محیط‌های بالینی واقعی داشته و به ابزاری کارآمد برای پزشکان در تشخیص و پیگیری بیماری مولتیپل اسکلروزیس تبدیل شوند.

مراجع

[1] F. Guarnera, A. Rondinella, E. Crispino, G. Russo, C. Di Lorenzo, D. Maimone, ... and S. Battiato, "MSLesSeg: baseline and benchmarking of a new Multiple Sclerosis Lesion Segmentation dataset," *Scientific Data*, vol. 12, no. 920, 2025. doi: https://doi.org/10.1038/s41597-025-05250-y.

[2] M. Popa, G. A. Vișa, and C. R. Șofariu, "PediMS: A Pediatric Multiple Sclerosis Lesion Segmentation Dataset," *Scientific*

Data, vol. 12, no. 1184, 2025. doi: https://doi.org/10.1038/s41597-025-05346-5.

[3] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, ... and X. Lladó, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach," *NeuroImage*, vol. 155, pp. 159-168, 2017. doi: https://doi.org/10.1016/j.neuroimage.2017.04.034.

[4] S. Roy, J. A. Butman, D. S. Reich, P. A. Calabresi and D. L. Pham, "Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks," *arXiv preprint arXiv:1803.09172*, 2018. doi: https://doi.org/10.48550/arXiv.1803.09172.

[5] S. Aslani, M. Dayan, L. Storelli, M. Filippi, V. Murino, M. A. Rocca and D. Sona, "Multi-branch convolutional neural network for multiple sclerosis lesion segmentation," *NeuroImage*, *196*, 1-15, 2019. doi: https://doi.org/10.1016/j.neuroimage.2019.03.068.

[6] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha and J. Li, "Transbts: Multimodal brain tumor segmentation using transformer," In *International conference on medical image computing and computer-assisted intervention*, pp. 109-119, 2021. doi: https://doi.org/10.1007/978-3-030-87193-2_11.

[7] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, ... and D. Xu, "Unetr: Transformers for 3d medical image segmentation," In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 574-584, 2022. doi: https://doi.org/10.1109/WACV51458.2022.00181.

[8] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," In *International MICCAI brainlesion workshop*, pp. 272-284, 2021. doi: https://doi.org/10.1007/978-3-031-08999-2_22.

[9] H. Lu, Y. She, J. Tie and S. Xu, "Half-UNet: A simplified U-Net architecture for medical image segmentation," *Frontiers in neuroinformatics*, vol. 16, 2022. doi: https://doi.org/10.3389/fninf.2022.911679.

[10] M. Sadeghibakhi, H. Pourreza, and H. Mahyar, "Multiple sclerosis lesions segmentation using attention-based CNNs in FLAIR images," *IEEE journal of translational engineering in health and medicine*, vol. 10, pp. 1-11, 2022. doi: https://doi.org/10.1109/JTEHM.2022.3172025.

[11] A. Rondinella, E. Crispino, F. Guarnera, O. Giudice, A. Ortis, G. Russo, ... and S. Battiato, "Boosting multiple sclerosis lesion segmentation through attention mechanism. *Computers in Biology and Medicine*, vol. 161, pp. 107021, 2023. doi: https://doi.org/10.1016/j.compbiomed.2023.107021.

[12] A. Rondinella, F. Guarnera, O. Giudice, A. Ortis, G. Russo, E. Crispino, ... and S. Battiato, "Enhancing multiple sclerosis lesion segmentation in multimodal mri scans with diffusion models," In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 3733-3740, 2023. doi: https://doi.org/10.1109/BIBM58861.2023.10385334.

[13] A. P. De Rosa, M. Benedetto, S. Tagliaferri, F. Bardozzo, A. D'Ambrosio, A. Bisecco, ... and F. Esposito, "Consensus of algorithms for lesion segmentation in brain MRI studies of multiple sclerosis," *Scientific Reports*, vol. 14, no. 1, 2024. doi: https://doi.org/10.1038/s41598-024-72649-9.

[14] S. R. Hashemi, S. S. M. Salehi, D. Erdogmus, S. P. Prabhu, S. K. Warfield, and A. Gholipour, "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection," *IEEE Access*, vol. 7, pp. 1721-1735, 2018. doi: https://doi.org/10.1109/access.2018.2886371.

[15] E. Essa, D. Aldesouky, S. E. Hussein and M. Z. Rashad, "Neuro-fuzzy patch-wise R-CNN for multiple sclerosis segmentation," *Medical & Biological Engineering & Computing*, vol. 58, no. 9, pp. 2161-2175, 2020. doi: https://doi.org/10.1007/s11517-020-02225-6.

[16] 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*, pp. 234-241, 2015. doi: https://doi.org/10.48550/arXiv.1505.04597.

[17] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016. doi: https://doi.org/10.1109/CVPR.2016.90.

[18] N. Abraham and N.M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," *In 2019 IEEE 16th international symposium on biomedical imaging*, pp. 683-687, 2019. doi: https://doi.org/10.1109/ISBI.2019.8759329.

[19] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980-2988, 2017. doi: https://doi.org/10.1109/ICCV.2017.324.

پاورقی‌ها:

[1] Relapsing-Remitting
[2] Magnetic Resonance Imaging
[3] Fluid-Attenuated Inversion Recovery
[4] Cerebrospinal Fluid
[5] Convolutional Neural Network
[6] Cascade Structure
[7] Fully-Connected Layer
[8] Residual Network
[9] Multi-Modal Feature Fusion
[10] Multi-Scale Feature Upsampling
[11] Decoder Section
[12] Token
[13] Position Embedding
[14] Encoder Section
[15] Long-Range Dependencies
[16] Skip Connections
[17] Self-Attention
[18] Depthwise Separable Convolution
[19] Spatial-Channel Attention
[20] Convolutional Long Short-Term Memory Layer
[21] Squeeze-and-Attention
[22] Dilation
[23] Batch Normalization
[24] Focal-Dice Loss
[25] Raw Skip-Connections
[26] Projection Shortcut
[27] Retrospective
[28] Lesion-wise True Positive Rate
[29] Precision
[30] Sensitivity