گزارش نهایی چالش *MS*

زهرا مهدور*[1]، مرتضی ملک‌نیا[2]

[1]دانشگاه خواجه نصیرالدین طوسی، Zahra.mahdevar@yahoo.com

[2]دانشگاه صنعتی اصفهان، M.maleknia@iut.ac.ir

*سرپرست تیم: دکتر رضا مختاری

چکیده – در این گزارش، به چگونگی تشخیص بیماری *MS* با استفاده از تصاویر *MRI* و بهره گیری از هوش مصنوعی خواهیم پرداخت. به‌طور دقیق‌تر، با توجه به نوع داده‌های آموزش که در دسترس گروه‌های شرکت کننده قرار گرفت و اهداف در نظر گرفته شده، از معماری *U_Net* جهت یادگیری ماشین در تشخیص پلاک‌های متناظر با این بیماری استفاده شده است. در قسمت *Encoder* معماری *U_Net*، شبکه‌های عصبی *MobileNetV2*، *DenseNet121* و *VGG16* به کار گرفته شد. با توجه به نتایج تجربی حاصل از آزمایشات، شامل اعتبار سنجی مدل و نتایج حاصل از مدل بر روی داده‌های آزمون، بهترین عملکرد در میان این سه شبکه عصبی متعلق به شبکه *MobileNetV2* است که جزئیات آن در این گزارش فراهم خواهد شد.

کلید واژه- بیماری MS، تصاویر MRI، شبکه‌های عصبی، هوش مصنوعی

مقدمه

بیماری MS یک بیماری مربوط به سیستم عصبی است که می‌تواند عملکرد مغز و نخاع را تحت تاثیر قرار دهد که باعث ایجاد مشکلات وسیعی از جمله مشکلات بینایی، احساسی و حرکتی می‌شود. علت دقیق این بیماری همچنان مشخص نیست و علائم آن در هر بیمار می‌تواند با شدت‌های مختلف بروز داشته باشد. آسیب‌های متناظر با این بیماری گاهی دائمی و گاهی موقت هستند. تشخیص زودهنگام این بیماری نقش مهمی در روند درمان و جلوگیری از پیشرفت آن دارد.

تصاویر MRI اطلاعات مفیدی در مورد شکل، اندازه و ناحیه رخداد این بیماری که در سیستم عصبی مرکزی انسان شکل می‌گیرد در اختیار پزشک معالج قرار می‌دهد. بررسی دستی (انسانی) تصاویر MRI فرایندی سخت، زمان‌بر و همراه با خطاهای انسانی است. بنابراین، توسعه روش‌های تشخیص و تقسیم بندی خودکار در سال‌های اخیر به یک موضوع مورد توجه در میان محققان در حوزه‌ی پزشکی و هوش مصنوعی بدل شده است.

در چالش پیش‌رو، یافتن ماسک برای پلاک‌های متناظر با بیماری MS و تشخیص نواحی چهار-گانه رخداد آنها از اهمیت ویژه‌ای برخوردار است. به همین منظور، تکنیک Semantic Segmentation را که پیکسل‌های تصاویر MRI را به کلاس‌هایی از پیش تعریف شده (مطابق با داده‌های آموزش) دسته‌بندی می‌کند به‌کار می‌گیریم. این تکنیک، دارای یک پیشینه تحقیقاتی وسیع است که اغلب با استفاده از شبکه‌های عصبی کانولوشن تقسیم‌بندی پیکسل‌ها در تصاویر صورت می‌گرفت. اما، با پیدایش معماری U_Net حوزه Semantic Segmentation دچار تحول و پیشرفت شگرفی شد [1, 2].

شبکه U_Net نوعی معماری برای تقسیم بندی سریع و دقیق تصاویر است که از کاربرد فراوانی در تصاویر مربوط به حوزه پزشکی برخوردار است. نسخه‌های مختلفی از معماری U_Net وجود دارند، از جمله نسخه‌های دو-بعدی و سه-بعدی که به ترتیب مناسب برای نوع داده دو-بعدی و سه-بعدی (حجمی) هستند.

معماری U_Net شامل یک Encoder و یک Decoder می‌باشد که در قسمت Encoder از لایه‌های کانولوشن برای فیلتر کردن و استخراج برجسته‌ترین ویژگی‌های تصویر (عملیات کاهش بعد) که به شبکه داده می‌شود استفاده می‌شود. در قسمت Decoder عملیات بازسازی تصویر با توجه به ویژگی‌های برجسته تصویر صورت می‌پذیرد. در طول این فرایند، ویژگی‌های بین بلوک‌های Encoder و Decoder به یکدیگر متصل می‌شوند.

MobileNetV2 یک شبکه عصبی کانولوشن است که در زیر مجموعه‌ای از پایگاه داده ImageNet آموزش داده شده است. در نتیجه، این شبکه قابلیت ویژه‌ای در تشخیص مهمترین ویژگی‌ها در تصاویر مختلف را دارد. از این‌رو، ما از این شبکه عصبی در قسمت Encoder معماری U_Net استفاده کرده‌ایم.

ایده‌ی انجام این کار از پژوهش تازه شکل گرفته [2] برگرفته شده است.

۲- روش انجام آزمایشات

برای آنکه پیاده‌سازی آزمایشات با هزینه‌های کمتری دنبال شوند و نیاز به پردازنده‌های قدرتمند، حافظه‌های موقتَ با حجم بالا، سوپر کامپیوترها و سرورهای اختصاصی به حداقل برسد، از **برش‌های دو-بعدی** داده‌های حجمی آموزشی که توسط کمیته تخصصی رقابت‌ها در اختیار گروه‌ها قرار گرفت استفاده شد.

پنجاه و پنج داده آموزشی سه-بعدی (حجمی) با ابعاد متفاوت (۱۶۰*۲۵۶*۲۳۶)، (۱۷۶*۲۵۶*۲۵۶)، (۱۶۰*۲۵۶*۲۵۶)، (۱۶۰*۲۵۶*۲۲۴) و ...، با فرمت NIFTI در اختیار قرار گرفت. به همراه هر داده آموزشی، یک تصویر سه-بعدی متناظر با پلاک‌های موجود در تصاویر MRI نیز در دسترس قرار گرفت.

برای هر بیمار حداقل ۶۴۰ برش دو-بعدی و حداکثر ۷۰۰ برش دو- بعدی در دسترس قرار داشت (با توجه به ابعاد متفاوت تصاویر). توجه شود که برای هر بیمار، حداقل ۲۵۰ برش دو-بعدی **فاقد** اطلاعات مربوط به نواحی مغزی می‌شدند، و بنابراین، بک فرایند یادگیری ماشین بهینه، به بخش عمده‌ای از این اطلاعات نیاز ندارد، اگرچه، استفاده از تعداد محدودی از آنها در انجام آزمایشات ضروری به نظر رسید و مورد استفاده نیز قرار گرفت.

با توجه به بحث بالا و محدودیت‌های زمانی و سخت‌افزاری در انجام آزمایشات، از۱۱۶ برش دو-بعدی از هر بیمار در فرایند یادگیری ماشین استفاده شد. این ۱۱۶ برش عبارت‌اند از برش‌های متناظر با بعد دوم (برشی که حاصل از ثابت نگه داشتن بعد دوم از یک داده حجمی است) از تصاویر حجمی که از شماره ۱۰۰ الی ۲۱۶ در نظر گرفته‌شده‌اند. بنابراین، در مجموع، از ۶۳۸۰ برش دو-بعدی جهت فرایند یادگیری ماشین و اعتبار سنجی آن استفاده شد.

۲-۱ پیش پردازش داده‌ها

مهندسی ویژگی به معنای به‌کارگیری دانش و تخصص در جهت بهینه‌سازی ویژگی‌هاست به نحوی که پس از آن داده‌های ورودی به راحتی و به بهترین نحو به مدل در نظر گرفته شده برای یادگیری ماشین ارائه شوند. هر دو مفهوم **نرمال‌سازی** و **استاندارد سازی** بخشی از مهندسی ویژگی هستند که در **مقیاس‌بندی** مورد استفاده قرار می‌گیرند. منظور از مقیاس بندی، قرار دادن ویژگی در دامنه‌ای خاص و محدود به جای قرار داشتن در دامنه‌ای وسیع است. توجه شود که دامنه‌های بزرگ تشخیص ویژگی‌های یک داده را برای یک مدل و معماری مفروض پیچیده می‌کند. از این رو، قبل از شروع فرایند یادگیری، نرمال‌سازی و استاندارد سازی برش‌های دو-بعدی در نظر گرفته شده صورت پذیرفت.

همچنین، همان‌طور که قبلا بیان کردیم، ما از شبکه عصبی MobileNetV2 در قسمت Encoder معماری U_Net استفاده کرده‌ایم. آز آنجا که ورودی‌های این شبکه عصبی با ابعاد از پیش تعریف شده‌ایی کار می‌کنند، قبل از انجام فرایند یادگیری، ابعاد داده‌های ورودی با ابعاد ۱۶۰*۱۶۰ بازسازی شدند.

علاوه‌براین، با توجه به اینکه ورودی شبکه عصبی مذکور تصاویری سه-کاناله هستند (قرمز، سبز و آبی)، تصاویر را از طیف خاکستری به طیف سه-کاناله مذکور تبدیل کردیم.

در نهایت، ۹۰ درصد برش‌های دو-بعدی در نظر گرفته شده برای فرایند یادگیری (Learning) مورد استفاده قرار گرفت و ۱۰ درصد مابقی جهت فرایند اعتبار سنجی (Validation) منظور گردید.

۲-۲- کمینه‌کردن تابع زیان

مجموعه‌ی داده‌های آموزش برای بهینه‌سازی متغیرهای شبکه در نظر گرفته شده، که ترکیبی از معماری U_Net و MobileNetV2 است، به واسطه کمینه کردن یک تابع زیان (Loss function) مورد استفاده قرار می‌گیرند.

تابع زیان مورد استفاده در این چالش، عبارت است از:
Sparse categorical cross entropy

یکی از چالش‌های اساسی در کمینه‌سازی تابع زیان، توزیع ناهمگن پیکسل‌های تصویر در یک کلاس خاص است. به طور دقیق‌تر، درصد بسیار زیادی از پیکسل‌ها می‌بایست به کلاس << تصویر پیش‌زمینه>> تعلق بگیرند. پیکسل‌هایی که در این کلاس قرار می‌گیرند بدان معنا هستند که در این قسمت از تصویر پلاکی از بیماری مشاهده **نشده** است. بنابراین، بیش از ۹۰ درصد از پیکسل‌های یک تصویر در این کلاس قرار میگیرند، که سبب کوچک شدن بیش از اندازه تابع زیان در حین فرایند یادگیری می‌گردد. در حقیقت، تابع زیان بیش از هر کلاس دیگری متاثر از این کلاس است. برای ایجاد تعادل میان کلاس‌ها، به کمک پیکسل‌های هر کلاس، **یک بردار وزن اصلاحی** به منظور متعادل سازی اثر کلاس‌ها در مقدار تابع زیان در نظر گرفته شد.

بهینه‌یاب‌های متعددی برای کمینه‌سازی یک تابع زیان موجود هستند. از جمله این بهینه‌یاب‌ها می‌توان به ADAM ، SGD ، RMSprop و ... اشاره کرد. با توجه به نتایج تجربی به دست آمده با استفاده از بهینه‌یاب‌های مختلف، بهترین عملکرد، هنگامی که از بهینه‌یاب ADAM استفاده کردیم مشاهده شد.

۲-۳- خروجی آزمایشات

آزمایشات در بستر Google Colab و با همان سرویس پایه که مبتنی بر یک حافظه موقت ۱۲,۷ گیگابایتی و پردازنده پیش فرض آن در حالت رایگان انجام شدند. کد نویسی‌ها نیز همگی به زبان Python صورت گرفتند و از توابع و ابزارهای موجود در کتابخانه Tensorflow استفاده شد.

خروجی آزمایشات به این شکل است که مدل پس از یادگیری باید هر پیکسل از یک برش دو -بعدی را به یکی از کلاس‌های پنج‌گانه نسبت دهد، به این

صورت که کلاس شماره 0 نماینده آن قسمت‌هایی است که پلاک بیماری در آنها تشخیص داده نمی‌شود (کلاس تصویر پیش‌زمینه) و کلاس‌های 1 الی 4 نماینده پلاک‌های بیماری در نواحی چهار-گانه هستند که هر کدام با رنگ مخصوص به خود مشخص می‌گردند.

بخش اول: نتایج حاصل بر داده‌های اعتبار سنج (Validation)

همان‌طور که در بخش‌های قبل اشاره شد، فرایند یادگیری و اعتبار سنجی به وسیله ۶۳۸۰ برش دو-بعدی متناظر با بعد دوم از داده‌های حجمی و برگرفته از همه بیماران انجام شد. ۹۰ درصد از این داده‌ها برای فرایند یادگیری و ۱۰ درصد دیگر جهت اعتبار سنجی مدل استفاده شدند. نتایج حاصل از مدل آموزش دیده شده بر روی داده‌های اعتبار سنجی توسط تابع Dice score برای پنج کلاس در نظر گرفته شده از قرار جدول ۱ هستند:

جدول ۱: دقت کلاس بندی پنج‌گانه بر اساس تابع Dice score بر روی داده‌های اعتبار سنج.

Dice loss score “0”	99%
Dice loss score “1”	84%
Dice loss score “2”	86%
Dice loss score “3”	79%
Dice loss score “4”	94%

علاوه‌بر این، نتایج حاصل از مدل بر روی یک داده اعتبار سنج دو-بعدی در شکل ۱ نمایش داده شده است (تصاویر در صفحه آخر نمایش داده شده‌اند) . در این شکل، شکل سمت چپ یک برش دو-بعدی از تصویر MRI را نشان می‌دهد، شکل میانی نواحی ماسک گذاری شده صحیح هستند و شکل سمت راست نواحی ماسک گذاری شده توسط مدل را نشان می‌دهند. همانگونه که از شکل دیده میشود، مدل توسعه داده شده با دقت بالایی ماسک گذاری را انجام داده است.

بخش دوم: نتایج حاصل بر زیرمجموعه‌ای از داده‌های آموزش استفاده نشده:

این بخش از آن لحاظ حائز اهمیت است که می‌توان عملکرد مدل را بر روی داده‌های آموزش دو-بعدی که لزوما از برش‌های متناظر با بعد دوم داده‌های حجمی **نیستند** مشاهده کرد. دقت شود که این داده‌ها متمایز با داده‌های اعتبار سنج هستند، در حالی که همچنان مکان صحیح ماسک‌ها را در هر یک از این داده‌های دو-بعدی میشناسیم. در حقیقت، با این آزمایش در می‌یابیم که انتخاب برش‌های دو-بعدی از بعد دوم تصاویر حجمی می‌تواند نتایج مطلوبی هنگام تشخیص پلاک‌ها از روی یک برش که لزوما مستخرج از بعد دوم نیستند بدست آورد یا خیر. نتایج تابع Dice score بر روی ۳۵۸۵ داده از این نوع از قرار جدول ۲ هستند:

جدول۲: دقت کلاس بندی پنج‌گانه بر اساس تابع Dice score بر روی داده‌های استفاده نشده دو-بعدی متناظر با بعدهای اول و سوم

Dice loss score “0”	99%
Dice loss score “1”	61%
Dice loss score “2”	81%
Dice loss score “3”	64%
Dice loss score “4”	95%

دقت شود که این نتایج تنها بر روی برش‌های دو بعدی متناظر با بعدهای اول و سوم بدست آمده‌اند در حالی که، در عمل، هر داده‌ی حجمی شامل برش‌هایی متناظر با هر سه بعد است. بنابراین، در عمل شاهد درصدهای بهتری نیز باید باشیم. شکل ۲ نتایج حاصل از مدل را بر روی چنین داده‌ای نشان میدهد.

بخش سوم: نتایج حاصل بر داده‌های آزمون

سرانجام مدل توسعه یافته را بر روی تمام برش‌های دو-بعدی متناظر با بعد سوم داده‌های آزمون به‌کارگرفتیم و ماسک‌های تولید شده برای هر برش استخراج گردید. سپس، ماسک‌های تولید شده در تصاویر دو-بعدی را با اتصال به یکدیگر به یک تصویر سه-بعدی با فرمت NII تبدیل کرده‌ایم.

شکل ۳ نتایج حاصل از مدل را بر روی یک داده آزمون نشان میدهد.

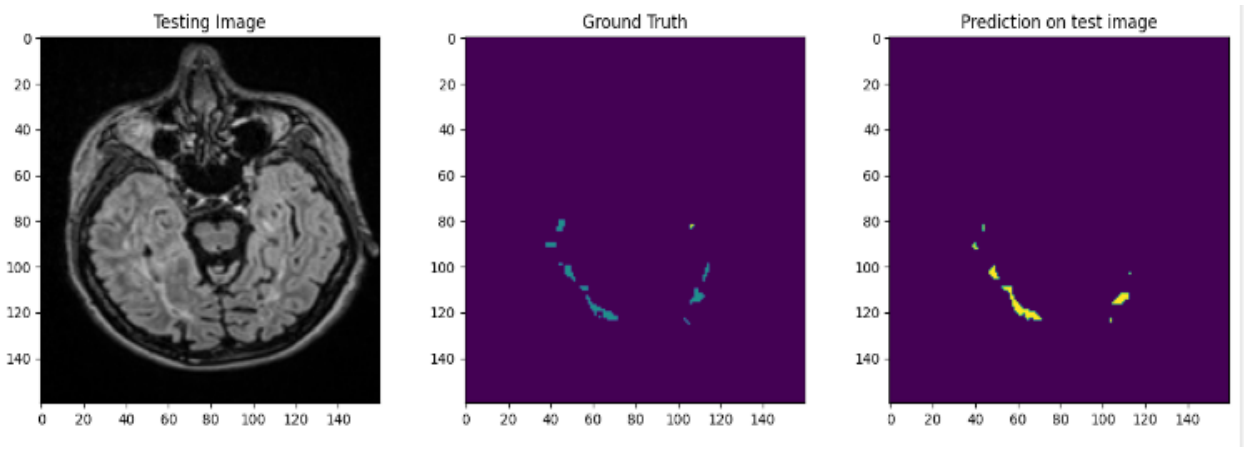
نتیجه‌گیری

نتایج بدست آمده از مدل توسعه یافته نشان می‌دهند که با دقت نسبتا خوبی ماسک‌ها در محل خود و با رنگ متناسب با نواحی چهارگانه قرار گرفته‌اند. تیم توسعه دهنده بر این باور هستند که با استفاده از امکانات سخت افزاری بهتر و با تکیه بر همین مدل و این‌بار با استفاده از برش‌هایی متناظر با هر سه بعد در فرایند یادگیری، نتایج حاصل را می‌توان بهبود داد. در مجموع، بر این باور هستیم، که با توجه به حجم پایین داده‌های استفاده شده در فرایند یادگیری، نتایج خوبی با حداقل هزینه استخراج شده است.

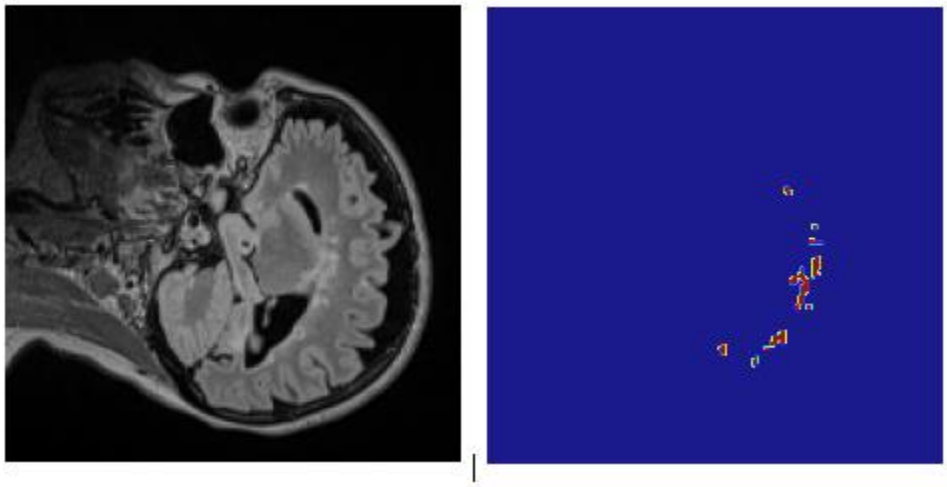
منابع

[1] BC. Bag and HK. Maity, C. Koley. UNET MOBILENETV2: MEDICAL IMAGE SEGMENTATION USING DEEP NEURAL NETWORK (DNN).

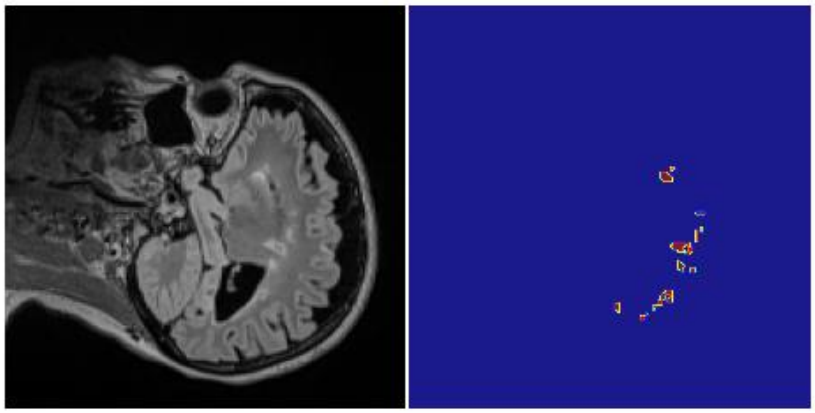
[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer, 234–241



شکل ۱: چپ: یک برش دو بعدی از یک داده اعتبار سنج. میانی: محل صحیح ماسک‌ها. راست: ماسک‌های تولید شده توسط مدل توسعه یافته.



شکل ۲: چپ: یک برش دو- بعدی متناظر با بعد سوم از یک داده استفاده نشده. راست: ماسک‌های تولید شده توسط مدل توسعه یافته متناظر با این برش.



شکل ۳: چپ: یک برش دو- بعدی از یک داده آزمون. راست: : ماسک‌های تولید شده توسط مدل توسعه یافته متناظر با این برش.