

ارائه یک چارچوب یکپارچه مبتنی بر پیش‌پردازش‌های هدفمند و یادگیری عمیق جهت حذف نویز و

افزایش مقاومت در برابر نمونه‌های متخاصم در تصاویر MR

فرناز حسینی^۱

^۱ استادیار، گروه مهندسی کامپیوتر، دانشگاه ملی مهارت، تهران، ایران

مقاله پژوهشی

چکیده

تاریخ دریافت:

۱۴۰۵/۰۳/۱۴

تاریخ پذیرش:

۱۴۰۵/۰۶/۲۱

کلیدواژه‌ها:

شبکه‌های عصبی عمیق،
نمونه‌های متخاصم، تصاویر
تشدید مغناطیسی (MRI)،
حذف نویز تصویر، بهینه‌سازی
ویژگی مبتنی بر الگوریتم ژنتیک،
طبقه‌بندی تصاویر پزشکی

نویسنده مسئول:

f-hoseini@tvu.ac.ir

شبکه‌های عصبی کانولوشنی مورد استفاده در تحلیل تصاویر پزشکی نسبت به نویز و اغتشاشات متخاصم حساس‌اند. در این پژوهش، چارچوبی چندمرحله‌ای برای طبقه‌بندی دودویی برش‌های تشدید مغناطیسی مغز به دو کلاس دارای تومور قابل مشاهده (Tumor-positive) و فاقد تومور قابل مشاهده (Tumor-negative) ارائه شده است. این چارچوب، پیش‌پردازش غیرخطی، قطعه‌بندی مبتنی بر واترشید، استخراج ویژگی، کاهش بُعد با استفاده از تحلیل مؤلفه‌های اصلی، انتخاب ویژگی مبتنی بر الگوریتم ژنتیک و طبقه‌بندی با شبکه عصبی کانولوشنی را با یکدیگر ترکیب می‌کند. چارچوب پیشنهادی روی ۳۹۲۹ برش متعلق به ۱۱۰ بیمار از مجموعه داده TCGA-LGG/TCIA و با استفاده از اعتبارسنجی متقاطع پنج‌تایی در سطح بیمار ارزیابی شد. مدل روی تصاویر پاک به Accuracy برابر با ۹۷/۲۳٪، F1-score برابر با ۹۶/۲۲٪ و سطح زیر منحنی مشخصه عملکرد گیرنده (AUC) برابر با ۰/۹۸۴ دست یافت. پایداری مدل به صورت جداگانه تحت حمله روش علامت گرادیان سریع (FGSM) و در سه سطح اغتشاش ارزیابی شد. یافته‌ها مؤید کارایی این زنجیره پردازشی برای طبقه‌بندی برش‌های متعلق به بیماران دیده‌نشده هستند؛ با این حال، پیش از تعمیم بالینی نتایج، انجام اعتبارسنجی خارجی ضروری است.

doi : 10.22034/ABMIR.2026.24696.1256

E-ISSN: [2821-2037](#)

/The Author 2026. Published by Yazd University This is an open

access article under the CC BY 4.0 License <https://creativecommons.org/licenses/by/4.0/>.



۱- مقدمه

از نظر میزان دسترسی مهاجم، حملات متخاصم عموماً به دو گروه جعبه‌سفید و جعبه‌سیاه تقسیم می‌شوند. در حمله جعبه‌سفید، مهاجم به معماری، پارامترها یا گرادیان‌های مدل دسترسی دارد؛ درحالی‌که در حمله جعبه‌سیاه، اطلاعات مهاجم به خروجی، برچسب یا امکان پرس‌وجو از مدل محدود می‌شود. مطالعات مربوط به حملات جعبه‌سیاه نشان داده‌اند که حتی بدون دسترسی مستقیم به پارامترهای مدل نیز می‌توان نمونه‌هایی ایجاد کرد که تصمیم طبقه‌بند را تغییر دهند [۸]. بنابراین، ادعای استحکام یک مدل باید در چارچوب یک مدل تهدید مشخص، با تعیین نوع دسترسی مهاجم، دامنه اغتشاش و حملات مورد استفاده ارزیابی شود.

در پاسخ به این آسیب‌پذیری، راهکارهای دفاعی مختلفی پیشنهاد شده‌اند. روش انتشار نویز متخاصم (ANP^v) با اعمال اغتشاش در بازنمایی‌های میانی شبکه و استفاده از آن در فرایند آموزش، تلاش می‌کند پایداری مدل را افزایش دهد [۹]. کوانتیزه‌سازی دفاعی نیز با محدودکردن تقویت خطای کوانتیزه‌سازی در لایه‌های شبکه، به‌طور هم‌زمان کارایی محاسباتی و مقاومت متخاصم را هدف قرار می‌دهد [۱۰]. رویکرد حذف نویز در فضای ویژگی نیز با افزودن بلوک‌های پالایش به feature map های میانی، اثر اغتشاشات متخاصم را در لایه‌های شبکه کاهش می‌دهد [۱۱]. باوجود نتایج امیدوارکننده، اثربخشی این دفاع‌ها به معماری شبکه، نوع حمله، شدت اغتشاش و نحوه تنظیم فرایند آموزش وابسته است.

در کنار روش‌های دفاعی، رویکردهایی برای تشخیص مستقیم نمونه‌های متخاصم ارائه شده‌اند. برای نمونه، Kwon و همکاران از الگوی امتیازهای خروجی طبقه‌بند برای شناسایی ورودی‌های غیرعادی استفاده کردند [۱۲]. بااین‌حال، روش مذکور یک آشکارساز عمومی نمونه متخاصم است و برای تصاویر پزشکی یا تشخیص تومور مغزی طراحی نشده است؛ بنابراین، نمی‌توان آن را رقیب مستقیم روش حاضر در طبقه‌بندی تصاویر MR در نظر

در سال‌های اخیر، شبکه‌های عصبی عمیق ($DNNs^1$) به‌ویژه شبکه‌های عصبی کانولوشنی (CNN^2)، به یکی از ابزارهای مهم تحلیل تصاویر پزشکی تبدیل شده‌اند. قابلیت این شبکه‌ها در یادگیری خودکار بازنمایی‌های سلسله‌مراتبی، زمینه استفاده از آن‌ها را در تشخیص بیماری، طبقه‌بندی تصاویر، قطع‌بندی ضایعات و تحلیل تصاویر سرطان فراهم کرده است [۱]. بااین‌حال، عملکرد مدل‌های یادگیری عمیق به کیفیت، توزیع و نحوه آماده‌سازی داده‌های ورودی وابسته است. در تصویربرداری تشدید مغناطیسی (MRI^3)، عواملی مانند نویز، آرتیفکت‌های تصویربرداری و کم‌نمونه‌برداری می‌توانند کیفیت بازسازی تصویر و در نتیجه قابلیت اعتماد تحلیل‌های بعدی را تحت تأثیر قرار دهند [۲].

یکی از چالش‌های مهم در کاربرد عملی شبکه‌های عصبی عمیق، آسیب‌پذیری آن‌ها در برابر نمونه‌های متخاصم است. این نمونه‌ها با افزودن اغتشاشاتی کوچک و هدفمند به ورودی ایجاد می‌شوند؛ به‌گونه‌ای که ممکن است تغییر ایجادشده برای مشاهده‌گر انسانی محدود یا نامحسوس باشد، اما تصمیم مدل را تغییر دهد [۳]. این آسیب‌پذیری در کاربردهای پزشکی اهمیت بیشتری دارد، زیرا تغییر نادرست تصمیم مدل می‌تواند بر تشخیص، برنامه‌ریزی درمان و تصمیم‌های بالینی اثر بگذارد [۴].

برای تولید و ارزیابی نمونه‌های متخاصم، روش‌های مختلفی ارائه شده‌اند. روش علامت گرادیان سریع ($FGSM^4$) با استفاده از گرادیان تابع زیان، اغتشاشی تک‌مرحله‌ای در جهت افزایش خطای مدل ایجاد می‌کند [۵]. روش تکرارشونده پایه یا I-FGSM/BIM^۵ این فرایند را طی چند گام کوچک اجرا می‌کند و معمولاً حمله قوی‌تری نسبت به FGSM ایجاد می‌کند [۶]. روش گرادیان تصویرشده (PGD^6) نیز با اجرای تکراری به‌روزرسانی گرادینانی و تصویرکردن نمونه در مجموعه مجاز اغتشاش، به یکی از روش‌های مرجع برای ارزیابی استحکام متخاصم تبدیل شده است [۷].

⁵ Basic Iterative Method/Iterative Fast Gradient Sign Method

⁶ Projected Gradient Descent

⁷ Adversarial Noise Propagation

¹ Deep Neural Networks

² Convolutional Neural Network

³ Magnetic Resonance Imaging

⁴ Fast Gradient Sign Method

(negative) ارائه می‌کند. در این تعریف، برش‌های Tumor-negative لزوماً متعلق به افراد سالم نیستند، بلکه برش‌هایی از بیماران مجموعه TCGA-LGG/TCIA هستند که در ماسک متناظر آن‌ها ناحیه تومور قابل مشاهده‌ای وجود ندارد. نوآوری کلیدی این مطالعه در طراحی و پیاده‌سازی یک معماری یکپارچه است که مرحله حذف نویز را نه صرفاً به عنوان یک پیش‌پردازش مجزا، بلکه به عنوان مؤلفه‌ای پیش‌پردازشی با نقش دفاعی بالقوه، در کنار قطعه‌بندی نواحی مغزی و بهینه‌سازی ویژگی‌ها قرار می‌دهد. با این حال، با توجه به ماهیت چندمرحله‌ای چارچوب، عملکرد مدل نهایی به تنهایی برای نسبت‌دادن افزایش پایداری به مرحله حذف نویز کافی نیست و اثر این مؤلفه باید در ارتباط با سایر مراحل پایپ‌لاین تفسیر شود. با این حال، با توجه به ماهیت چندمرحله‌ای چارچوب، عملکرد مدل نهایی به تنهایی برای نسبت‌دادن افزایش پایداری به مرحله حذف نویز کافی نیست و اثر این مؤلفه باید در ارتباط با سایر مراحل پایپ‌لاین تفسیر شود. پس از پیش‌پردازش و قطعه‌بندی، ویژگی‌های استخراج‌شده ابتدا با تحلیل مؤلفه‌های اصلی (PCA^۱) کاهش بُعد می‌یابند و سپس الگوریتم ژنتیک (GA^۲) برای انتخاب زیرمجموعه‌ای از ویژگی‌های مؤثر به کار گرفته می‌شود. در مرحله نهایی، شبکه عصبی کانولوشنی طبقه‌بندی دودویی را انجام می‌دهد. در این چارچوب، نمونه متخاصم کلاس سوم مسئله نیست؛ بلکه حملات FGSM، BIM/I-FGSM و PGD برای ارزیابی ثابتهای پایداری همان طبقه‌بند Tumor-positive/Tumor-negative استفاده می‌شوند. یکی از اهداف اصلی این پژوهش، بررسی نقش یک زنجیره پیش‌پردازشی شامل حذف نویز، قطعه‌بندی و پالایش ویژگی در کاهش حساسیت طبقه‌بند نسبت به نویز و اغتشاشات ورودی است. از آنجاکه این مراحل به صورت یکپارچه اجرا می‌شوند، بهبود مشاهده‌شده به عملکرد کلی چارچوب نسبت داده می‌شود و نه به یک مؤلفه منفرد.

مشارکت‌های اصلی این پژوهش عبارت‌اند از: نخست، طراحی یک زنجیره منسجم شامل حذف نویز، قطعه‌بندی، استخراج ویژگی، PCA، انتخاب ویژگی مبتنی بر الگوریتم ژنتیک و طبقه‌بندی

گرفت. ارتباط این مطالعه با پژوهش حاضر صرفاً به مؤلفه تحلیل خروجی طبقه‌بند محدود می‌شود و نتایج آن در جدول مقایسه کمی روش‌های تشخیص تومور وارد نمی‌شود.

مطالعات انجام‌شده در حوزه تحلیل تصاویر پزشکی نیز نشان داده‌اند که سامانه‌های مبتنی بر یادگیری عمیق در برابر اغتشاشات متخاصم آسیب‌پذیرند و شدت این آسیب‌پذیری می‌تواند به مجموعه داده، وظیفه تشخیصی و معماری مدل وابسته باشد [۱۳]. از این رو، ارزیابی یک روش پزشکی تنها روی تصاویر پاک برای اثبات پایداری آن کافی نیست و لازم است نتایج تصاویر پاک و تصاویر متخاصم به صورت جداگانه گزارش شوند.

در تحلیل تومورهای مغزی، قطعه‌بندی ناحیه تومور می‌تواند به محدود کردن فضای تحلیل و کاهش اثر نواحی نامرتبط تصویر کمک کند. روش‌های مبتنی بر خودرمزگذار نیز برای قطعه‌بندی تومور در تصاویر MR بررسی شده‌اند [۱۴]. با این حال، پژوهش مذکور از مجموعه داده BRATS2020 استفاده کرده است؛ در حالی که مجموعه داده پژوهش حاضر نسخه منتشرشده TCGA-LGG/TCIA است.

مرور پژوهش‌های موجود نشان می‌دهد که مطالعات پیشین عموماً تنها یکی از مؤلفه‌های مسئله را بررسی کرده‌اند؛ به گونه‌ای که برخی بر بازسازی یا حذف نویز تصویر، برخی بر قطعه‌بندی تومور، بعضی بر آموزش متخاصم یا تشخیص نمونه‌های متخاصم و گروهی دیگر بر پالایش بازنمایی‌های میانی شبکه تمرکز داشته‌اند. افزون بر این، بررسی‌های مروری نشان داده‌اند که نبود مدل تهدید یکسان و تفاوت در مجموعه داده، حمله، معماری و معیار ارزیابی، مقایسه مستقیم بسیاری از روش‌های دفاعی را دشوار می‌کند [۱۵]. بنابراین، بررسی یک چارچوب چندمرحله‌ای که پیش‌پردازش، قطعه‌بندی، استخراج و پالایش ویژگی و طبقه‌بندی را در یک زنجیره مشخص ترکیب کند و عملکرد آن را روی تصاویر پاک و متخاصم گزارش دهد، همچنان ضرورت دارد.

در راستای این هدف، پژوهش حاضر چارچوبی چندمرحله‌ای برای طبقه‌بندی دودویی برش‌های MR مغزی دارای تومور قابل مشاهده (Tumor-positive) و فاقد تومور قابل مشاهده (Tumor-negative)

¹ Principal Component Analysis

² Genetic Algorithm

به ۱۱۰ بیمار مبتلا به گلیوما درجه پایین را در برمی‌گیرد. این نسخه شامل ۳۹۲۹ جفت تصویر و ماسک است که از میان آن‌ها، ۱۳۷۳ برش دارای ناحیه تومور قابل مشاهده و ۲۵۵۶ برش فاقد ناحیه تومور قابل مشاهده هستند. بنابراین، کلاس منفی نباید «افراد سالم» نامیده شود؛ زیرا برش‌های هر دو کلاس از بیماران مبتلا به LGG استخراج شده‌اند. مسئله طبقه‌بندی پژوهش حاضر نیز

تفکیک دودویی برش‌های Tumor-positive و Tumor-negative است.

برای سازمان‌دهی مقایسه با مطالعات پیشین، ابتدا هم‌خوانی منبع داده، تعریف مسئله و معیار ارزیابی بررسی شد. سپس، برای جدول عددی مقایسات در این مقاله، تنها مطالعاتی انتخاب شدند که از نسخه ۳۹۲۹ تصویری TCGA-LGG/TCIA یا نسخه تقریباً یکسان آن استفاده کرده، مسئله حضور یا فقدان ناحیه تومور در برش را بررسی کرده و Accuracy را برحسب درصد گزارش کرده‌اند.

Gajula و Rajesh چارچوبی شامل ResNet50 برای طبقه‌بندی دودویی و Deep Residual U-Net برای قطعه‌بندی ارائه کردند [۱۷]. این مطالعه از نسخه ۳۹۲۹ تصویری TCGA-LGG/TCIA، شامل ۱۳۷۳ برش Tumor-positive و ۲۵۵۶ برش Tumor-negative، استفاده کرده است. مدل طبقه‌بندی به Accuracy برابر با ۹۴/۲۳٪، دست‌یافت و Precision، Recall و F1-score کلاس تومور به ترتیب برابر با ۹۱، ۹۳ و ۹۲٪ گزارش شدند. این روش از نظر منبع داده و تعریف مسئله به پژوهش حاضر نزدیک است؛ با این حال، استخراج و طبقه‌بندی ویژگی‌ها در آن به صورت انتباه‌انتهای انجام شده و مرحله مستقلی برای کاهش بُد با PCA و انتخاب ویژگی مبتنی بر الگوریتم ژنتیک در نظر گرفته نشده است. افزون بر این، patient-wise بودن تقسیم داده در گزارش مقاله به روشنی اثبات نشده است.

Ramakrishnan و همکاران یک معماری ترکیبی متشکل از InceptionV3، VGG16، ResNet50 و DenseNet را همراه با بهینه‌سازی oneAPI پیشنهاد کردند [۱۸]. این پژوهش نیز از ۳۹۲۹ تصویر مجموعه TCGA-LGG، شامل ۲۵۵۶ برش فاقد تومور قابل مشاهده و ۱۳۷۳ برش دارای تومور قابل مشاهده، برای

CNN؛ دوم، تفکیک روشن مسئله اصلی طبقه‌بندی دودویی از ارزیابی ثانویه استحکام متخاصم؛ سوم، ارزیابی جداگانه مدل روی تصاویر پاک و تصاویر تحت حملات FGSM، BIM/I-FGSM، PGD؛ چهارم، تحلیل سهم مراحل مختلف چارچوب از طریق مطالعه حذف مؤلفه یا ablation؛ و پنجم، مقایسه کمی تنها با مطالعاتی که از همان منبع داده TCGA-LGG/TCIA و مسئله طبقه‌بندی نزدیک استفاده کرده‌اند. با این حال، تفاوت‌های احتمالی در نحوه تقسیم بیماران، پیش‌پردازش، افزایش داده و پروتکل ارزیابی هنگام تفسیر نتایج در نظر گرفته می‌شوند.

ادامه مقاله به این صورت سازمان‌دهی شده است: در بخش دوم، پژوهش‌های مرتبط با تشخیص و قطعه‌بندی تومور، حذف نویز، کاهش بُد، انتخاب ویژگی و ارزیابی متخاصم مرور می‌شوند. بخش سوم چارچوب پیشنهادی و جزئیات مراحل آن را توضیح می‌دهد. در بخش چهارم، مجموعه داده، نحوه تقسیم داده‌ها، تنظیمات الگوریتم ژنتیک، حملات و معیارهای ارزیابی معرفی می‌شوند سپس نتایج، مقایسه با پژوهش‌های پیشین و محدودیت‌ها ارائه می‌گردند و بخش پایانی به جمع‌بندی و پیشنهادهایی برای مطالعات آینده اختصاص دارد.

۲- پیشنهاد تحقیق

در سال‌های اخیر، استفاده از شبکه‌های عصبی عمیق و روش‌های پردازش تصویر برای تحلیل تصاویر تشدید مغناطیسی مغز به طور قابل توجهی گسترش یافته است. تصاویر MR به دلیل کنتراست مناسب بافت نرم و قابلیت نمایش ساختارهای مختلف مغز، یکی از منابع اصلی برای تشخیص و قطعه‌بندی تومورهای مغزی محسوب می‌شوند [۱۶]. با این حال، کیفیت تصاویر، وجود نویز و آرتیفکت، تغییرپذیری بین بیماران، نامتوازن بودن کلاس‌ها و ابعاد بالای فضای ویژگی می‌تواند بر دقت و قابلیت تعمیم مدل‌های یادگیری عمیق اثر بگذارد. از این رو، به کارگیری یک زنجیره پردازشی شامل حذف نویز، قطعه‌بندی، استخراج ویژگی، کاهش بُد و انتخاب ویژگی می‌تواند به ایجاد بازنمایی فشرده‌تر و مؤثرتری از تصاویر MR کمک کند.

مجموعه داده پژوهش حاضر نسخه منتشر شده مجموعه TCGA-LGG/TCIA است که تصاویر و ماسک‌های دوبعدی MR مربوط

فضای با بُعد بالا محدود می‌شود. از این رو، این مرجع پشتیبان روش شناختی مرحله GA-based Feature Selection است.

مطالعات دیگری نیز جنبه‌های مختلف تحلیل مجموعه-TCGA LGG/TCIA را بررسی کرده‌اند. Naser و Deen یک چارچوب مبتنی بر U-Net و VGG16 برای قطعه‌بندی و درجه‌بندی گلیوما درجه پایین پیشنهاد کردند [۲۲]. با توجه به اینکه مسئله اصلی این پژوهش قطعه‌بندی و درجه‌بندی تومور است، نتایج آن را نمی‌توان مستقیماً با Accuracy قطعه‌بندی دودویی-Tumor-positive/Tumor-negative مقایسه کرد.

Asiri و همکاران ترکیبی از CNN، ResNet50 تنظیم‌شده و U-Net را برای تشخیص، طبقه‌بندی و قطعه‌بندی تصاویر TCGA-LGG/TCIA بررسی کردند [۲۳]. این مطالعه اهمیت انتقال یادگیری و ادغام طبقه‌بندی با قطعه‌بندی را نشان می‌دهد؛ با این حال، تفاوت در تعداد بیماران اولیه، نحوه تشکیل زیرمجموعه‌ها و پروتکل ارزیابی مانع از آن است که تمام اعداد گزارش‌شده آن به‌عنوان مقایسه مستقیم با پژوهش حاضر استفاده شوند.

Anaya-Isaza و همکاران عملکرد شبکه‌های کانولوشنی، انتقال یادگیری، افزایش داده و معماری Cross-Transformer را در طبقه‌بندی و تشخیص تومورهای مغزی بر روی دیتاست Figshare و طبقه‌بندی سه نوع تومور بررسی کردند [۲۴]. این پژوهش نشان می‌دهد که معماری شبکه، نوع افزایش داده، روش پیش‌پردازش و ترکیب مجموعه‌های داده می‌تواند بر نتیجه نهایی اثر معناداری داشته باشد.

Sulaiman و همکاران مدل LinkNet-34 با رمزگذار EfficientNetB7 را برای قطعه‌بندی معنایی تومور پیشنهاد کردند [۲۵]. این روش روی مجموعه ۳۹۲۹ تصویری-TCGA LGG/TCIA و مجموعه BRATS2020 ارزیابی شده است؛ با این حال، معیارهای آن عمدتاً در سطح پیکسل محاسبه شده‌اند. بنابراین، Dice و Jaccard گزارش‌شده در این پژوهش نباید با Accuracy طبقه‌بندی برش در روش حاضر مقایسه شوند.

Chatterjee, Rai و Dashkevich نیز معماری ترکیبی UnetResNext-50 را برای تشخیص و قطعه‌بندی تومور روی همان نسخه ۳۹۲۹ تصویری ارائه کردند [۲۶]. این مطالعه از افزایش

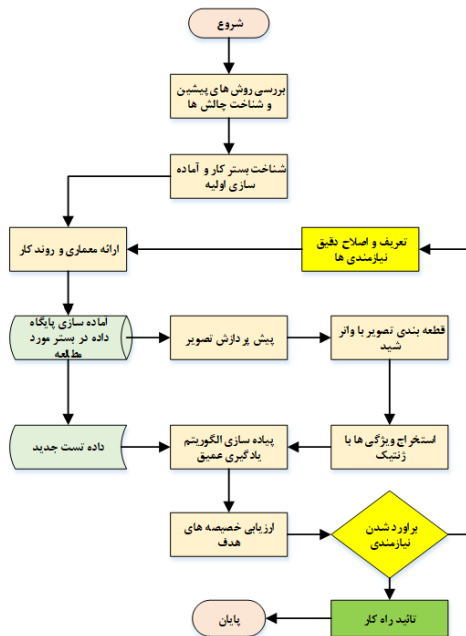
طبقه‌بندی دودویی استفاده کرد. معماری ترکیبی پیشنهادی Accuracy برابر با ۹۶/۲٪ گزارش کرده است. تمرکز اصلی این روش بر ترکیب توان بازنمایی چند معماری عمیق و بهبود کارایی محاسباتی بود. در مقابل، چارچوب پژوهش حاضر پیش از طبقه‌بندی، فضای ویژگی را به‌صورت صریح با PCA و الگوریتم ژنتیک پالایش می‌کند.

Alshomrani یک زنجیره یکپارچه مبتنی بر CNN تنظیم‌شده و Residual U-Net را برای طبقه‌بندی و قطعه‌بندی تومور ارائه کرد [۱۹]. این پژوهش ابتدا روی مجموعه Figshare ارزیابی شد و سپس آزمایش تکمیلی آن روی مجموعه TCGA-LGG/TCIA شامل تقریباً ۳۹۲۹ تصویر و ۱۱۰ بیمار انجام گرفت. در سناریوی طبقه‌بندی دودویی، Accuracy برابر با ۹۶٪ گزارش شد. این مطالعه از نظر منبع داده و ترکیب طبقه‌بندی و قطعه‌بندی به پژوهش حاضر نزدیک است، اما مرحله مستقلی برای کاهش بُعد و انتخاب ویژگی تکاملی ندارد. علاوه بر این، به دلیل استفاده از چند مجموعه داده و گزارش‌نشدن جزئیات کاملاً همسان درباره تقسیم بیماران، مقایسه آن با پژوهش حاضر باید «تقریباً مستقیم» تلقی شود.

پس از استخراج ویژگی، وجود متغیرهای زائد و هم‌بسته می‌تواند پیچیدگی مدل و احتمال بیش‌پرازش را افزایش دهد. تحلیل مؤلفه‌های اصلی با انتقال داده‌ها به مجموعه‌ای از مؤلفه‌های متعامد و مرتب‌کردن آن‌ها بر اساس واریانس، امکان ایجاد یک بازنمایی کم‌بُعدتر را فراهم می‌کند [۲۰]. با این حال، مؤلفه‌های PCA الزاماً بیشترین قدرت تفکیک میان کلاس‌ها را ندارند؛ بنابراین، استفاده از یک مرحله انتخاب ویژگی نظارت‌شده پس از PCA می‌تواند مکمل فرایند کاهش بُعد باشد.

در همین راستا، Roodaki، Sotoodeh و Moosavi روشی مبتنی بر الگوریتم ژنتیک برای انتخاب زیرمجموعه‌ای مؤثر از میان توصیفگرهای تصویری با ابعاد بالا ارائه کردند [۲۱]. تابع برازندگی این روش به‌گونه‌ای طراحی شده است که میان عملکرد و تعداد ویژگی‌های منتخب تعادل برقرار کند. اگرچه کاربرد اصلی این پژوهش بازیابی محتوایی تصویر است و نه تشخیص تومور مغزی، ارتباط آن با چارچوب حاضر دقیقاً به استفاده از الگوریتم ژنتیک برای حذف ویژگی‌های زائد و انتخاب ویژگی‌های مؤثر از یک

ویژگی‌ها انجام می‌شود تا مجموعه‌ای از ویژگی‌های مناسب و کم‌افزونی برای مرحله یادگیری انتخاب گردد.



شکل (۱): نمای کلی چارچوب پیشنهادی

در مرحله بعد، باهدف کاهش پیچیدگی داده‌ها و جلوگیری از افزونگی اطلاعات، فرآیند کاهش ابعاد و انتخاب ویژگی‌های مؤثر اعمال می‌شود. این مرحله نقش مهمی در افزایش کارایی مدل و کاهش هزینه محاسباتی دارد. در نهایت، داده‌های پالایش شده به مدل مبتنی بر یادگیری عمیق منتقل می‌شوند تا برش‌های MRI به دو کلاس دارای تومور و فاقد تومور طبقه‌بندی شوند. نمونه‌های متخاصم کلاس مستقلی در خروجی شبکه تشکیل نمی‌دهند، بلکه برای ارزیابی میزان پایداری همین طبقه‌بند در برابر اغتشاشات هدفمند مورد استفاده قرار می‌گیرند.

مزیت اصلی چارچوب پیشنهادی در ادغام منسجم مراحل حذف نویز، پالایش ویژگی‌ها و طبقه‌بندی مبتنی بر یادگیری عمیق است. این ساختار چندمرحله‌ای موجب می‌شود تأثیر نویز و اغتشاشات متخاصم در مراحل اولیه پردازش کاهش یافته و داده‌های ورودی باکیفیت بالاتر و پایداری بیشتر به شبکه عصبی عمیق منتقل شوند؛

داده گسترده و یک معماری عمیق انتهابه‌انتهای استفاده می‌کند، اما فاقد مراحل مستقل PCA و انتخاب ویژگی مبتنی بر الگوریتم ژنتیک است.

درمجموع، مطالعات مبتنی بر TCGA-LGG/TCIA عمدتاً بر طبقه‌بندی مستقیم با CNN، انتقال یادگیری، ترکیب چند شبکه عمیق یا ادغام طبقه‌بندی و قطعه‌بندی متمرکز بوده‌اند. در مقابل، پژوهش حاضر زنجیره‌ای Denoising + Segmentation + Feature Extraction + PCA + GA + CNN را بررسی می‌کند.

۳- روش پیشنهادی

در این پژوهش، یک چارچوب چندمرحله‌ای و مقاوم برای طبقه‌بندی دودویی برش‌های MR مغزی به دو کلاس برش دارای تومور^۱ و برش فاقد تومور^۲ ارائه می‌شود. ارزیابی در برابر نمونه‌های متخاصم، یک تحلیل ثانویه برای سنجش استحکام این طبقه‌بند است و تشخیص نمونه پاک از نمونه متخاصم به‌عنوان یک وظیفه طبقه‌بندی مستقل در نظر گرفته نمی‌شود. ایده اصلی روش پیشنهادی بر این فرض استوار است که افزایش پایداری شبکه‌های عصبی عمیق در برابر اغتشاشات متخاصم صرفاً از طریق بهبود معماری شبکه حاصل نمی‌شود، بلکه نیازمند به‌کارگیری یک فرآیند پردازشی چندمرحله‌ای شامل پیش‌پردازش داده، پالایش ویژگی‌ها و کاهش پیچیدگی داده‌ها پیش از مرحله یادگیری عمیق است. بر این اساس، چارچوب پیشنهادی تلاش می‌کند با ترکیب مراحل مختلف پردازش تصویر و یادگیری ماشینی، داده‌های ورودی را پیش از ورود به مدل یادگیری عمیق به‌صورت ساختاریافته و مقاوم آماده‌سازی کند. نمای کلی این چارچوب در شکل (۱) نشان داده شده است.

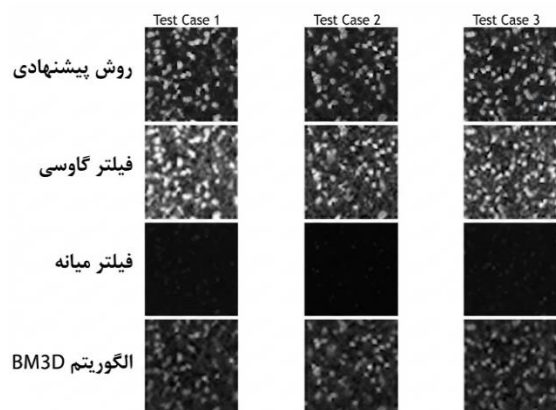
در این چارچوب، ابتدا مرحله پیش‌پردازش باهدف حذف نویز اولیه از تصاویر پزشکی انجام می‌شود تا اثر اغتشاشات موجود در داده‌های ورودی کاهش یابد. در ادامه، فرآیند قطعه‌بندی تصویر به‌منظور جداسازی نواحی مهم و ساختارهای مرتبط در تصاویر پزشکی صورت می‌گیرد. پس‌از آن، ویژگی‌های مؤثر از تصاویر استخراج شده و با استفاده از الگوریتم ژنتیک فرآیند بهینه‌سازی

² Tumor-negative

¹ Tumor-positive

شکل (۳)، سه سناریوی آزمایشی با شدت‌های متفاوت نویز در نظر گرفته شده‌اند تا توانایی روش‌ها در کاهش نویز و حفظ ساختارهای تصویری بررسی شود.

مقایسه نتایج نشان می‌دهد که فیلترهای کلاسیک در برخی سناریوها ممکن است موجب هموارسازی بیش از حد تصویر، حذف ناکافی نویز یا از دست رفتن بخشی از جزئیات بافتی و ساختاری شوند. در مقابل، روش پیشنهادی با هدف ایجاد توازن میان کاهش نویز و حفظ اطلاعات ساختاری تصویر به کار گرفته شده است. باین حال، میزان اثربخشی آن به نوع و شدت نویز و ویژگی‌های تصاویر ورودی وابسته است.



شکل (۳): مقایسه نتایج حاصل از نویزگیری تصاویر با فیلترهای متفاوت

نکته قابل توجه در این مرحله آن است که حذف نویز در این چارچوب با هدف کاهش مؤلفه‌های ناخواسته و محدود کردن انتقال اغتشاشات ورودی به مراحل استخراج ویژگی و طبقه‌بندی به کار گرفته شده است. از این منظر، این مرحله می‌تواند نقش یک مؤلفه دفاعی اولیه را ایفا کند. باین حال، به دلیل تعامل آن با قطعه‌بندی، PCA، انتخاب ویژگی مبتنی بر GA و CNN، عملکرد نهایی را نمی‌توان صرفاً به مرحله حذف نویز نسبت داد. به عبارت دیگر، با پالایش داده‌های ورودی در این مرحله، احتمال انتقال اغتشاشات مخرب به مراحل استخراج ویژگی و یادگیری عمیق کاهش یافته و در نتیجه پایداری و دقت سیستم در مراحل بعدی افزایش می‌یابد.

۲-۳ قطعه‌بندی تصویر

پس از مرحله حذف نویز، تصاویر ورودی به منظور تمرکز بر نواحی معنادار و کاهش پیچیدگی داده‌ها تحت فرآیند قطعه‌بندی قرار

موضوعی که در نهایت منجر به بهبود پایداری و افزایش دقت سیستم در تحلیل تصاویر پزشکی می‌شود.

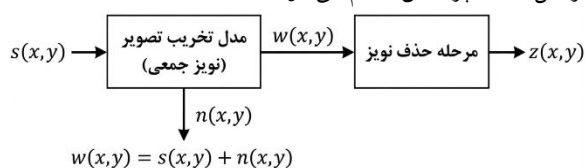
۱-۳ مرحله پیش پردازش و حذف نویز

در مرحله نخست، تصاویر ورودی که شامل تصاویر MR هستند، تحت فرآیند پیش پردازش و حذف نویز قرار می‌گیرند. وجود نویز در تصاویر پزشکی علاوه بر کاهش کیفیت بصری داده‌ها، می‌تواند منجر به استخراج نادرست ویژگی‌ها و افزایش حساسیت مدل‌های یادگیری عمیق نسبت به اغتشاشات ورودی شود. از این رو، انجام یک مرحله حذف نویز مؤثر پیش از فرآیند یادگیری، نقش مهمی در بهبود کیفیت داده‌های ورودی و افزایش پایداری مدل دارد.

در چارچوب پیشنهادی، مدل تخریب تصویر به صورت نویز جمعی در نظر گرفته شده است؛ به طوری که تصویر پاک $s(x, y)$ با مؤلفه نویز $n(x, y)$ ترکیب شده و تصویر تخریب شده $w(x, y)$ را مطابق رابطه زیر تولید می‌کند:

$$w(x, y) = s(x, y) + n(x, y) \quad (1)$$

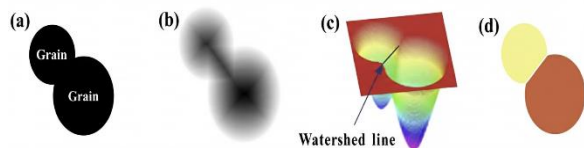
در رابطه (۱)، $n(x, y)$ نمایانگر نویز افزوده شده به تصویر است و هدف از این مدل سازی، شبیه سازی شرایط متداول آلودگی در تصاویر پزشکی و ارزیابی عملکرد روش پیشنهادی در حضور اغتشاشات ورودی است. سپس تصویر $w(x, y)$ وارد مرحله حذف نویز شده و خروجی بازسازی شده $z(x, y)$ تولید می‌شود. روند کلی این فرآیند در شکل (۲) نمایش داده شده است. در چارچوب پیشنهادی، فرآیند حذف نویز به گونه ای طراحی شده است که ضمن کاهش اغتشاشات تصادفی و نویزهای ناخواسته، ساختارهای اصلی و اطلاعات بافتی تصویر حفظ شوند. این مرحله با هدف بهبود کیفیت داده‌های ورودی و جلوگیری از انتقال نویز به مراحل بعدی پردازش انجام می‌شود.



شکل (۲): فرآیند پیشنهادی حذف نویز در تصاویر ورودی

همچنین، به منظور ارزیابی عملکرد مرحله حذف نویز در شرایط مختلف، نتایج روش پیشنهادی با فیلترهای متداول شامل فیلتر گاوسی، فیلتر میانه و الگوریتم BM3D مقایسه شده است. مطابق

اغتشاشات و نویزهای غیرمرتبط را کاهش داده و به افزایش پایداری سیستم در برابر نویز و حملات متخاصم کمک کند.



شکل (۵): قطعه‌بندی تصویر با استفاده از الگوریتم واترشید مبتنی بر آنروپی

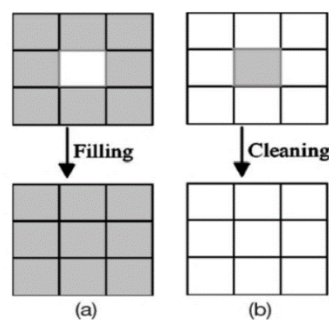
۳-۳ استخراج و انتخاب ویژگی با استفاده از الگوریتم ژنتیک

پس از انجام فرآیند قطعه‌بندی تصویر، مجموعه‌ای از ویژگی‌های اولیه از نواحی استخراج شده محاسبه می‌شود. این ویژگی‌ها می‌توانند اطلاعات ارزشمندی درباره ساختارهای بافتی و الگوهای موجود در تصاویر پزشکی فراهم کنند. باین حال، به دلیل تعداد زیاد ویژگی‌ها و وجود ویژگی‌های کم‌اثر یا دارای افزونگی، استفاده مستقیم از کل مجموعه ویژگی‌ها می‌تواند موجب افزایش پیچیدگی محاسباتی و کاهش کارایی مدل‌های یادگیری عمیق شود. از این رو، انتخاب زیرمجموعه‌ای از ویژگی‌های مؤثر و حذف ویژگی‌های غیرضروری به عنوان یک مرحله مهم در فرآیند پردازش داده‌ها در نظر گرفته می‌شود.

در این پژوهش، به منظور انتخاب و بهینه‌سازی ویژگی‌ها از الگوریتم ژنتیک استفاده شده است. الگوریتم ژنتیک یک روش جستجوی تکاملی است که با الهام از فرآیند انتخاب طبیعی، قادر است در فضای جستجوی بزرگ ویژگی‌ها به دنبال ترکیب‌های بهینه بگردد. در این روش، ابتدا یک جمعیت اولیه از کروموزوم‌ها ایجاد می‌شود که هر کروموزوم نمایانگر یک زیرمجموعه از ویژگی‌های استخراج شده است. سپس کیفیت هر کروموزوم با استفاده از یک تابع برازندگی ارزیابی شده و کروموزوم‌های برتر برای تولید نسل‌های بعدی انتخاب می‌شوند. در پیاده‌سازی حاضر، هر کروموزوم به صورت یک بردار دودویی به طول (d) تعریف شد که (d) تعداد مؤلفه‌های حاصل از PCA است؛ مقدار یک برای هر ژن نشان‌دهنده انتخاب آن ویژگی و مقدار صفر نشان‌دهنده حذف آن بود. اندازه جمعیت برابر با ۵۰ کروموزوم و حداکثر تعداد نسل‌ها

می‌گیرند. هدف از این مرحله جداسازی نواحی مهم تصویر و محدود کردن تحلیل به بخش‌هایی است که اطلاعات تشخیصی مرتبط با ساختارهای بافتی را در خود دارند.

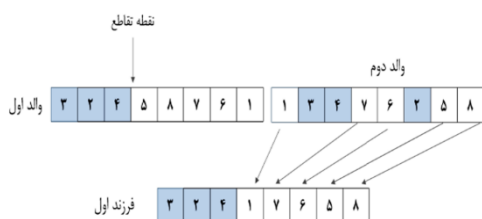
در این پژوهش، برای انجام قطعه‌بندی تصویر از ترکیب عملگرهای مورفولوژیک و فیلتر غیرخطی آب‌پخشان یا واترشید مبتنی بر تابع آنروپی استفاده شده است. در گام نخست، عملگرهای مورفولوژیک بر روی نواحی محلی تصویر اعمال می‌شوند تا مرز ساختارهای برجسته تقویت شده و نواحی غیرضروری یا کم‌اهمیت تضعیف شوند. این مرحله باعث بهبود وضوح مرزهای نواحی موردنظر و آماده‌سازی تصویر برای فرآیند قطعه‌بندی دقیق‌تر می‌شود. نمونه‌ای از اعمال این عملگرها بر روی بخشی از تصویر در شکل (۴) ارائه شده است.



شکل (۴): اعمال عملگرهای مورفولوژیک بر روی ناحیه‌ای از تصویر در ادامه، الگوریتم واترشید برای تفکیک دقیق نواحی مختلف تصویر اجرا می‌شود. در این روش، تصویر به صورت یک سطح توپوگرافی در نظر گرفته می‌شود که در آن شدت پیکسل‌ها به عنوان ارتفاع سطح تعریف می‌شود. با استفاده از تابع آنروپی، مرزهای نواحی با تغییرات اطلاعاتی بیشتر شناسایی شده و فرآیند قطعه‌بندی با دقت بالاتری انجام می‌شود. نتیجه این فرآیند در شکل (۵) نشان داده شده است.

اجرای مرحله قطعه‌بندی موجب تمرکز مراحل بعدی پردازش بر نواحی دارای اطلاعات تشخیصی می‌شود و در نتیجه فضای جستجوی ویژگی‌ها کاهش می‌یابد. این موضوع علاوه بر کاهش پیچیدگی محاسباتی، موجب افزایش دقت استخراج ویژگی‌ها و بهبود عملکرد مدل یادگیری عمیق در مراحل بعدی می‌شود. در نهایت، محدود شدن تحلیل به نواحی مهم تصویر می‌تواند اثر

جمعیت، محاسبه تابع برازندگی، انتخاب ویژگی یا تعیین شرط توقف استفاده نشدند. پس از تعیین زیرمجموعه بهینه ویژگی‌ها روی داده‌های آموزشی، همان زیرمجموعه بدون اجرای مجدد الگوریتم ژنتیک روی مجموعه آزمون اعمال شد.



شکل (۷): عملگر تقاطع یکنواخت مورد استفاده

جواب فعلی	۴	۲	۹	۵	۱۱	۳	۱	۷	۱۲	۸	۶	۱۰
تعویض جفتی	۴	۲	۷	۵	۱۱	۳	۱	۹	۱۲	۸	۶	۱۰
معکوس سازی	۴	۲	۷	۱	۳	۱۱	۵	۹	۱۲	۸	۶	۱۰

شکل (۸): عملگرهای جهش مورد استفاده در فرآیند تکامل

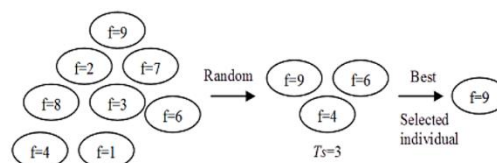
برای جلوگیری از نشت اطلاعات، برازش PCA و اجرای الگوریتم ژنتیک در هر fold به صورت مستقل و صرفاً با استفاده از تصاویر و برجسب‌های ۸۸ بیمار مجموعه آموزش انجام شد. ارزیابی تابع برازندگی نیز از طریق اعتبارسنجی درونی سه‌قسمتی و در سطح بیمار، درون همین مجموعه آموزش صورت گرفت. ۲۲ بیمار مجموعه آزمون در تشکیل جمعیت اولیه، برازش PCA، محاسبه تابع برازندگی، انتخاب ویژگی، تنظیم پارامترها یا تعیین شرط توقف دخالت نداشتند. پس از انتخاب کروموزوم بهینه، پارامترهای PCA و ماسک ویژگی‌های منتخب ثابت نگه‌داشته شدند و بدون برازش یا انتخاب مجدد بر مجموعه آزمون اعمال گردیدند. همچنین، تمام برش‌های متعلق به هر بیمار صرفاً در یکی از مجموعه‌های آموزش یا آزمون قرار گرفتند.

۳-۴ ایجاد نمونه‌های متخاصم و سناریوی حمله

به منظور ارزیابی میزان پایداری چارچوب پیشنهادی در برابر اغتشاشات هدفمند، در این پژوهش از نمونه‌های متخاصم تولیدشده به صورت الگوریتمی استفاده شد. در این چارچوب، نمونه متخاصم به تصویری اطلاق می‌شود که با افزودن یک اغتشاش کوچک اما جهت‌دار به تصویر اولیه ایجاد می‌شود؛

برابر با ۱۰۰ در نظر گرفته شد. جمعیت اولیه به صورت تصادفی تولید شد و هر کروموزوم حداقل شامل یک ویژگی فعال بود. انتخاب والدین با روش مسابقه‌ای و اندازه مسابقه برابر با ۳ انجام شد. برای تولید فرزندان از تقاطع یکنواخت با احتمال ۰/۸ و از جهش معکوس‌سازی بیت با احتمال ۰/۱ استفاده شد. همچنین، دو کروموزوم برتر هر نسل بدون تغییر به نسل بعد منتقل شدند. الگوریتم در صورت رسیدن به ۱۰۰ نسل یا عدم بهبود بهترین مقدار برازندگی طی ۲۰ نسل متوالی متوقف شد. برای افزایش بازتولیدپذیری، بذر مولد اعداد تصادفی برابر با ۴۲ تنظیم گردید. کیفیت هر کروموزوم با استفاده از تابع برازندگی ترکیبی ارزیابی شد که به طور هم‌زمان قدرت تفکیک ویژگی‌های منتخب و اندازه زیرمجموعه ویژگی‌ها را در نظر می‌گیرد.

در مرحله انتخاب، از استراتژی انتخاب به روش مسابقه استفاده شده است. در این روش، چند کروموزوم به صورت تصادفی از جمعیت انتخاب شده و بهترین آن‌ها بر اساس مقدار تابع برازندگی به عنوان والد انتخاب می‌شود. نمونه‌ای از این فرآیند در شکل (۶) نمایش داده شده است. برای تولید فرزندان از تقاطع یکنواخت با احتمال ۰/۸ استفاده شد. در این عملگر، مقدار هر ژن فرزند با احتمال برابر از یکی از دو والد انتخاب می‌شود. این روش با نمایش دودویی کروموزوم در مسئله انتخاب ویژگی سازگار است. برای ترکیب اطلاعات ژنتیکی کروموزوم‌ها به کار گرفته می‌شود. در این پژوهش، متناسب با کدگذاری دودویی کروموزوم‌ها، از عملگر تقاطع یکنواخت استفاده شد که نمونه‌ای از آن در شکل (۷) نشان داده شده است. علاوه بر این، برای حفظ تنوع جمعیت و جلوگیری از همگرایی زود هنگام الگوریتم، عملگر جهش نیز بر روی برخی ژن‌ها اعمال می‌شود که نمونه‌هایی از آن در شکل (۸) ارائه شده است.



شکل (۶): استراتژی انتخاب به روش مسابقه در الگوریتم ژنتیک
الگوریتم ژنتیک در هر fold به صورت مستقل و صرفاً بر ویژگی‌ها و برجسب‌های مجموعه آموزش اجرا شد. داده‌های آزمون در تشکیل

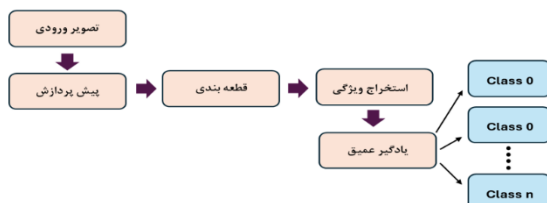
مجاز اغتشاش، جست‌وجوی قوی‌تری برای یافتن نمونه متخاصم انجام می‌دهد. از این رو، این حملات ممکن است افت عملکرد بیشتری نسبت به FGSM ایجاد کنند. ارزیابی تجربی حاضر به FGSM محدود است و نتایج آن نباید به BIM، PGD یا سایر مدل‌های تهدید تعمیم داده شود.

به منظور تحلیل رفتار مدل در شدت‌های مختلف حمله FGSM، بودجه اغتشاش در سه سطح آزمایشی تنظیم شد و تأثیر افزایش آن بر خروجی مدل بررسی گردید. تمام نتایج این بخش صرفاً به حمله FGSM و تنظیمات به کاررفته در همین مطالعه مربوط‌اند. تصاویر متخاصم تولیدشده همانند تصاویر پاک وارد زنجیره کامل پردازش شامل حذف نویز، قطعه‌بندی، استخراج ویژگی، بهینه‌سازی ویژگی‌ها با الگوریتم ژنتیک و طبقه‌بندی نهایی شدند.

۳-۵ طبقه‌بندی برش‌های دارای تومور و فاقد تومور

با یادگیری عمیق

در مرحله نهایی، بازنمایی‌های حاصل از زنجیره پیش‌پردازش، قطعه‌بندی و بهینه‌سازی ویژگی‌ها به شبکه عصبی کانولوشنی سفارشی ارائه می‌شوند تا هر برش MRI به یکی از دو کلاس Tumor-positive یا Tumor-negative طبقه‌بندی شود. برچسب هر برش از ماسک مرجع متناظر آن استخراج می‌شود؛ به طوری که وجود حداقل یک پیکسل توموری در ماسک، نشان‌دهنده کلاس Tumor-positive و خالی بودن کامل ماسک، نشان‌دهنده کلاس Tumor-negative است. نمونه‌های متخاصم تنها در مرحله آزمون استحکام و با حفظ برچسب اصلی برش به کار گرفته می‌شوند. معماری کلی روش پیشنهادی و ارتباط مراحل آن در شکل (۹) ارائه شده است.



شکل (۹): بلوک‌دیگرام روش پیشنهادی برای طبقه‌بندی برش‌های دارای تومور و فاقد تومور

به گونه‌ای که تغییر ایجادشده از نظر بصری محدود بوده، اما بتواند تصمیم مدل را دچار اختلال کرده یا دقت طبقه‌بندی را کاهش دهد. هدف از به کارگیری این نوع نمونه‌ها، شبیه‌سازی شرایطی است که در آن مهاجم تلاش می‌کند با حداقل تغییر در داده ورودی، سامانه تشخیص را گمراه سازد. در این مطالعه، تولید نمونه‌های متخاصم با استفاده از روش «علامت گرادیان سریع» (FGSM)^۱ انجام شده است. این حمله یکی از روش‌های استاندارد و پرکاربرد در ارزیابی آسیب‌پذیری مدل‌های یادگیری عمیق محسوب می‌شود و بر مبنای گرادیان تابع هزینه نسبت به ورودی عمل می‌کند. در این روش، با محاسبه جهت افزایش خطای مدل، اغتشاشی کنترل شده به تصویر افزوده می‌شود تا احتمال برچسب‌گذاری نادرست افزایش یابد. اگر x تصویر ورودی، y برچسب واقعی، θ پارامترهای مدل و $J(\theta, x, y)$ تابع هزینه باشند، نمونه متخاصم x_{adv} مطابق رابطه زیر تولید می‌شود:

$$x_{adv} = x + \epsilon \text{sign}(\nabla_x J(\theta, x, y)) \quad (2)$$

که در آن ϵ شدت اغتشاش اعمال شده را کنترل می‌کند و $\text{sign}(\cdot)$ علامت مؤلفه‌ای گرادیان را نشان می‌دهد. به منظور حفظ مقادیر شدت روشنایی در بازه مجاز، پس از اعمال اغتشاش، مقادیر پیکسل‌ها در دامنه معتبر تصویر محدود شدند. این سازوکار سبب می‌شود که نمونه‌های متخاصم تولیدشده ضمن حفظ شباهت ظاهری با تصاویر اصلی، از دید مدل یادگیری عمیق دارای اثر مخرب باشند. در سناریوی ارزیابی این پژوهش، حمله از نوع جعبه‌سفید در نظر گرفته شد؛ بدین معنا که فرآیند تولید نمونه متخاصم با استفاده از اطلاعات داخلی مدل، شامل پارامترها و گرادیان تابع هزینه، انجام می‌شود. انتخاب سناریوی سفیدجعبه باهدف ارزیابی مستقیم حساسیت طبقه‌بند نسبت به گرادیان تابع هزینه انجام شد. با این حال، FGSM یک حمله تک‌مرحله‌ای است و سفیدجعبه بودن آن به معنای قوی‌ترین یا سخت‌گیرانه‌ترین حمله ممکن نیست. در حملات تکرارشونده‌ای مانند BIM/i-FGSM، اغتشاش طی چند گام متوالی اعمال می‌شود؛ درحالی که PGD با استفاده از به‌روزرسانی تکرارشونده و تصویرکردن نمونه در ناحیه

¹ Fast Gradient Sign Method

نویز و نمونه‌های متخصص شود. بر همین اساس، در چارچوب پیشنهادی، PCA و الگوریتم ژنتیک به‌عنوان دو مؤلفه مکمل پس از استخراج ویژگی عمیق به کار گرفته شده‌اند تا فرآیند تصمیم‌گیری صرفاً مبتنی بر انبوهی از ویژگی‌های خام نباشد، بلکه بر پایه یک بازنمایی فشرده، پالایش شده و هدفمند انجام پذیرد.

جدول (۱): مشخصات معماری شبکه CNN سفارشی

نوع لایه	جزئیات / پارامترها	تابع فعال‌سازی
ورودی	تصویر نرمال شده با ابعاد $(224 \times 224 \times 1)$	-
لایه کانولوشن ۱	۳۲ فیلتر (3×3) ، Stride ۱	ReLU
لایه Pooling	Max Pooling (2×2)	-
لایه کانولوشن ۲	۶۴ فیلتر (3×3) ، Stride ۱	ReLU
Batch Normalization	اعمال روی خروجی لایه کانولوشن ۲	-
لایه کانولوشن ۳	۱۲۸ فیلتر (3×3) ، Stride ۱	ReLU
لایه Fully Connected	۵۱۲ نورون	ReLU
لایه Dropout	نرخ ۰/۵	-
لایه خروجی	۱ نورون (Sigmoid)	Sigmoid

جدول (۲): تنظیمات هایپرپارامترهای آموزش مدل

پارامتر آموزشی	مقدار تنظیم شده
تابع بهینه‌ساز (Optimizer)	Adam
نرخ یادگیری اولیه (Initial Learning Rate)	۰/۰۰۱
اندازه دسته (Batch Size)	۳۲
تعداد دوره‌های آموزشی (Epochs)	۵۰
تابع هزینه (Loss Function)	Binary Cross-Entropy
نرخ Dropout	۰/۵
روش ارزیابی	5-Fold Cross-Validation

در این چارچوب، PCA برای نگاشت ویژگی‌های استخراج شده به یک زیرفضای کم‌بعدتر و معنادارتر مورد استفاده قرار می‌گیرد تا مؤلفه‌های دارای واریانس مؤثر حفظ و وابستگی‌ها و افزونگی‌های

مشخصات معماری شبکه عصبی مورد استفاده در این پژوهش به‌گونه‌ای طراحی شده است که تعادلی میان پیچیدگی مدل و قابلیت تعمیم‌پذیری ایجاد نماید. این شبکه شامل سه لایه کانولوشن با تابع فعال‌سازی ReLU است که برای استخراج ویژگی‌های سلسله‌مراتبی، با یک لایه Batch Normalization و یک لایه Max Pooling ترکیب شده‌اند. در بخش انتهایی شبکه، یک لایه تمام متصل^۱ با ۵۱۲ نورون و یک لایه خروجی تک‌نورونی با تابع فعال‌سازی Sigmoid قرار گرفته است. خروجی این لایه، احتمال تعلق برش ورودی به کلاس Tumor-positive را برآورد می‌کند و طبقه‌بندی نهایی با مقایسه این احتمال با آستانه تصمیم انجام می‌شود. بنابراین، دو برچسب خروجی مدل Tumor-positive و Tumor-negative هستند و نمونه متخصص یک کلاس خروجی مستقل محسوب نمی‌شود.

آموزش شبکه با استفاده از تابع هزینه آنتروپی متقاطع دودویی^۲ و الگوریتم بهینه‌سازی Adam انجام می‌شود. نرخ یادگیری اولیه برابر با ۰/۰۰۱ تنظیم شده است که در صورت عدم کاهش تابع هزینه پس از ۱۰ دوره آموزشی^۳، این نرخ با ضریب ۰/۱ کاهش می‌یابد. همچنین، برای ارزیابی تعمیم‌پذیری مدل، از اعتبارسنجی متقاطع پنج‌تایی^۴ گروه‌بندی شده در سطح بیمار استفاده شد. شناسه بیمار به‌عنوان متغیر گروه‌بندی در نظر گرفته شد؛ بنابراین، تمام برش‌های متعلق به هر بیمار، شامل برش‌های Tumor-positive و Tumor-negative، در یک fold واحد قرار گرفتند و هیچ بیماری به‌طور هم‌زمان در مجموعه‌های آموزش و آزمون حضور نداشت. جزئیات دقیق پیکربندی لایه‌ها و تنظیمات آموزش مدل در جداول (۱) و (۲) آورده شده است.

لازم به ذکر است که هرچند شبکه‌های عصبی کانولوشنی به‌طور ذاتی قابلیت استخراج ویژگی‌های سطح پایین و سطح بالا را از داده‌های تصویری دارا هستند، اما خروجی این لایه‌ها لزوماً فاقد افزونگی، همبستگی یا حساسیت به اغتشاشات ورودی نیست. در کاربردهای پزشکی، که تغییرات ظریف بافتی اهمیت بالایی دارند، وجود ویژگی‌های همپوشان یا ناپایدار می‌تواند منجر به افزایش پیچیدگی مدل، کاهش قابلیت تعمیم و افت عملکرد در مواجهه با

³ Epoch

⁴ 5-Fold Cross-Validation

¹ Fully Connected

² Binary Cross-Entropy

از منظر روش شناختی، قابلیت تعمیم چارچوب پیشنهادی با تکیه بر طراحی ماژولار آن تقویت شده است؛ به گونه‌ای که ترکیب مراحل پیش‌پردازش و حذف نویز، قطعه‌بندی مبتنی بر بازسازی مورفولوژیک، استخراج ویژگی چندمقیاسی با استفاده از تبدیل موجک گسسته (DWT)، و کاهش بُعد از طریق تحلیل مؤلفه‌های اصلی (PCA)، موجب کاهش حساسیت مدل نسبت به تغییرات شدت، کنتراست و نویز موجود در تصاویر می‌شود. افزون بر این، به‌منظور کاهش خطر بیش‌برازش ناشی از محدودیت تعداد نمونه‌ها و افزایش پایداری مدل در مواجهه با داده‌های جدید، از تکنیک افزایش داده^۱ استفاده شد. این فرآیند شامل چرخش‌های تصادفی در بازه $30 \pm$ درجه، بازتاب افقی و عمودی، و تغییرات جزئی در شدت روشنایی بود. ابتدا بیماران هر fold از یکدیگر تفکیک شدند و سپس عملیات افزایش داده فقط روی برش‌های متعلق به بیماران مجموعه آموزش انجام شد. هیچ تصویر اصلی یا افزایش‌یافته‌ای از بیماران آزمون در فرایند آموزش حضور نداشت. همچنین، نسخه‌های افزایش‌یافته هر برش در همان partition مربوط به بیمار مبدأ باقی ماندند.

با توجه به وجود چندین برش هم‌بسته برای هر بیمار، اعتبارسنجی متقاطع پنج‌تایی در سطح بیمار انجام شد. ابتدا ۱۱۰ بیمار، بر اساس شناسه بیمار، به پنج fold بدون هم‌پوشانی تقسیم شدند؛ به‌طوری‌که هر fold شامل ۲۲ بیمار بود. در هر تکرار، تمام برش‌های متعلق به ۲۲ بیمار یک fold صرفاً به‌عنوان مجموعه آزمون و تمام برش‌های متعلق به ۸۸ بیمار باقی‌مانده به‌عنوان مجموعه آموزش استفاده شدند. تمام برش‌های یک بیمار، اعم از Tumor-positive و Tumor-negative، همواره در یک partition واحد باقی ماندند. در نتیجه، هیچ شناسه بیماری میان مجموعه‌های آموزش و آزمون مشترک نبود و مدل در هر تکرار روی بیمارانی ارزیابی شد که هیچ برشی از آن‌ها را در مرحله آموزش مشاهده نکرده بود. در این مطالعه مجموعه اعتبارسنجی مستقلی استفاده نشد و معماری شبکه و هایپارامترهای آموزشی پیش از شروع ارزیابی متقاطع ثابت شدند. پس از اجرای پنج تکرار، معیارهای عملکرد برای هر fold

غیرضروری کاهش یابد. این امر نه‌تنها موجب کاهش بار محاسباتی می‌شود، بلکه از منظر یادگیری نیز مرز طبقه‌بندی را ساده‌تر و پایدارتر می‌سازد. سپس الگوریتم ژنتیک باهدف جستجوی تطبیقی در فضای زیرمجموعه‌های ویژگی، مجموعه‌ای از ویژگی‌های بهینه را انتخاب می‌کند که بیشترین توان تفکیک‌پذیری را در کنار حداقل همپوشانی اطلاعاتی داشته باشند. بدین ترتیب، ترکیب CNN، PCA و الگوریتم ژنتیک در این پژوهش یک زنجیره مکمل برای «استخراج، فشرده‌سازی و پالایش مقاوم ویژگی‌ها» را شکل می‌دهد، نه مجموعه‌ای از مراحل تکراری یا زائد.

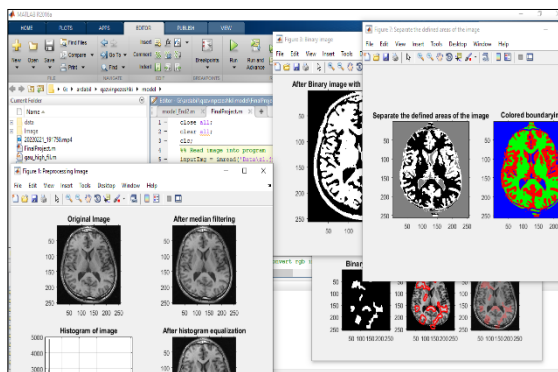
۴- نتایج تجربی

۴-۱ معرفی مجموعه داده

برای ارزیابی عملکرد و میزان پایداری چارچوب پیشنهادی، از مجموعه‌داده عمومی تصاویر مغزی بیماران مبتلابه گلیومای درجه پایین (LGG) استفاده شد. این مجموعه شامل تصاویر ماسک‌های مرجع مربوط به ۱۱۰ بیمار است و هر بیمار دارای چندین برش دوبعدی MRI است. در این مطالعه، واحد پیش‌بینی یک برش MRI بود؛ باین‌حال، واحد تقسیم داده بیمار در نظر گرفته شد. برشی که ماسک مرجع آن حداقل شامل یک پیکسل توموری بود، به‌عنوان Tumor-positive و برشی که ماسک آن کاملاً خالی بود، به‌عنوان Tumor-negative برچسب‌گذاری شد. برش‌های Tumor-negative نیز متعلق به بیماران LGG هستند و نشان‌دهنده افراد سالم یا یک گروه کنترل مستقل نیستند. انتخاب این مجموعه‌داده بر پایه دو ملاحظه اساسی صورت پذیرفت: نخست، دسترسی عمومی آن که بازتولیدپذیری نتایج و امکان مقایسه مستقیم با مطالعات پیشین را فراهم می‌سازد؛ دوم، وجود برچسب‌ها و ماسک‌های مرجع دقیق برای نواحی توموری که ارزیابی کمی و معتبر مراحل قطعه‌بندی و استخراج ویژگی را امکان‌پذیر می‌نماید. در این پژوهش، تعداد ۱۴ ویژگی متمایز از تصاویر MRI استخراج شد که قادرند جنبه‌های بافتی و ساختاری نواحی توموری را به‌طور مؤثر بازنمایی نمایند.

¹ Data Augmentation

کانولوشنی عمیق است. نمای کلی مدل اجرایی و نحوه تعامل میان اجزای مختلف چارچوب در شکل (۱۰) ارائه شده است.



شکل (۱۰): مدل قابل اجرا از روش پیشنهادی با ابزار متلب

آزمایش‌ها بر روی سیستمی با پردازنده Intel Core i7، حافظه اصلی ۱۶GB RAM و کارت گرافیک NVIDIA GTX ۱۰۶۰ با ۶GB حافظه اختصاصی انجام شده‌اند. آموزش شبکه عصبی با استفاده از بهینه‌ساز Adam، نرخ یادگیری اولیه ۰/۰۱، اندازه دسته (Batch Size) برابر با ۳۲ و به مدت ۵۰ دوره آموزشی (Epoch) انجام شده است. به منظور جلوگیری از بیش‌برازش، از لایه Dropout با نرخ ۰/۵ و همچنین اعتبارسنجی متقاطع برای ارزیابی تعمیم‌پذیری مدل استفاده شده است.

به‌کارگیری مدل اجرایی در این محیط آزمایشی، علاوه بر ارزیابی شاخص‌های عملکردی نظیر دقت، حساسیت و ویژگی، امکان تحلیل معیارهای غیرعملکردی ازجمله زمان پردازش، پایداری محاسباتی و سربار سیستم را نیز فراهم کرده است. بررسی هم‌زمان این شاخص‌ها تصویری جامع از کارایی عملی چارچوب پیشنهادی ارائه می‌دهد و قابلیت کاربرد آن را در سناریوهای واقعی پردازش تصاویر پزشکی مورد ارزیابی قرار می‌دهد.

۴-۳ تحلیل عملکرد قطعه‌بندی تصاویر MR

قطعه‌بندی دقیق تصاویر MR یکی از مراحل کلیدی در افزایش پایداری و کارایی مدل‌های یادگیری عمیق در تشخیص تومورهای مغزی است. هدف از این مرحله، استخراج نواحی آناتومیکی معنادار، حذف ساختارهای غیرمرتبط و کاهش پیچیدگی داده ورودی پیش از آموزش شبکه CNN سفارشی است.

به صورت جداگانه محاسبه و نتایج نهایی به شکل میانگین $\pm$ انحراف معیار پنج fold گزارش شدند.

تمام مراحل وابسته به داده در هر تکرار اعتبارسنجی متقاطع به‌طور مستقل و صرفاً بر اساس مجموعه آموزش همان fold انجام شدند. عملیات پیش‌پردازش و افزایش داده پس از تفکیک بیماران اجرا شد و افزایش داده فقط به برش‌های آموزشی محدود بود. در مواردی که یک پارامتر پیش‌پردازشی نیازمند برآورد از توزیع داده بود، پارامتر مربوط تنها از مجموعه آموزش تخمین زده شد و بدون تخمین مجدد بر مجموعه آزمون اعمال گردید.

پس از استخراج ویژگی‌ها، PCA صرفاً بر ماتریس ویژگی‌های مجموعه آموزش همان fold برازش شد. برادر میانگین، جهت مؤلفه‌های اصلی و تعداد مؤلفه‌های حفظ‌شده بدون استفاده از تصاویر یا برچسب‌های مجموعه آزمون تعیین شدند. سپس همان تبدیل PCA آموزش‌دیده، بدون برازش مجدد، بر ویژگی‌های مجموعه آزمون اعمال شد.

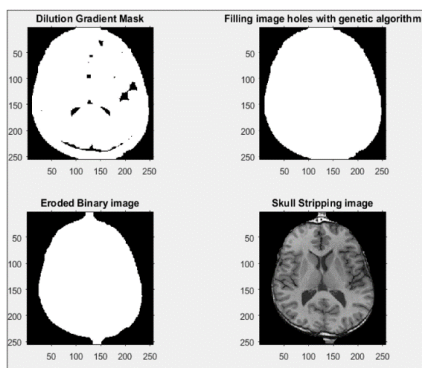
انتخاب ویژگی مبتنی بر الگوریتم ژنتیک نیز در هر fold فقط با استفاده از ویژگی‌ها و برچسب‌های مجموعه آموزش انجام شد. محاسبه تابع برازندگی، انتخاب والدین، تقاطع، جهش و تعیین زیرمجموعه نهایی ویژگی‌ها هیچ دسترسی به داده‌های آزمون نداشتند. پس از پایان فرایند انتخاب ویژگی، زیرمجموعه منتخب ثابت نگه‌داشته شد و برای تبدیل ویژگی‌های مجموعه آزمون همان fold به کار رفت. بدین ترتیب، مجموعه آزمون تا پایان آموزش، کاهش ابعاد و انتخاب ویژگی کاملاً کنار گذاشته شد و صرفاً برای ارزیابی نهایی استفاده گردید.

۴-۲. پیاده‌سازی مدل پیشنهادی و محیط آزمایش

چارچوب پیشنهادی به‌طور کامل با استفاده از نرم‌افزار MATLAB (نسخه ۲۰۲۳a) و جعبه‌ابزارهای Image Processing Toolbox و Deep Learning Toolbox پیاده‌سازی شده است. این پیاده‌سازی شامل تمامی مراحل اصلی روش، از جمله پیش‌پردازش و حذف نویز تصاویر MR، قطعه‌بندی نواحی معنادار، استخراج ویژگی، بهینه‌سازی ویژگی‌ها با استفاده از الگوریتم ژنتیک، و در نهایت طبقه‌بندی مبتنی بر شبکه عصبی

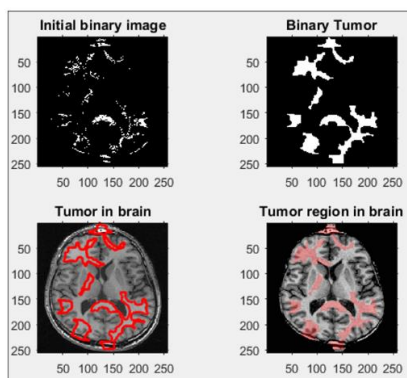
ارائه یک چارچوب یکپارچه مبتنی بر پیش‌پردازش‌های هدفمند و یادگیری عمیق جهت حذف نویز و

افزایش مقاومت در برابر نمونه‌های متخاصم در تصاویر MR

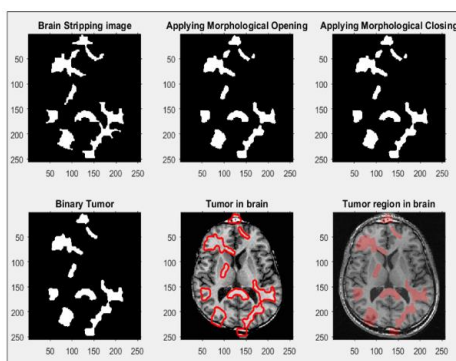


شکل (۱۳): جداسازی جمجمه از مغز

نتیجه نهایی قطعه‌بندی که شامل حذف نویز باقی‌مانده و برجسته‌سازی ناحیه تومور است، در شکل (۱۴) قابل مشاهده است. همچنین تحلیل مورفولوژیک ناحیه توموری شامل بررسی ویژگی‌هایی نظیر مساحت، پیرامون، میزان فشردگی (Compactness) و ناهمواری مرزی در شکل (۱۵) ارائه شده است.

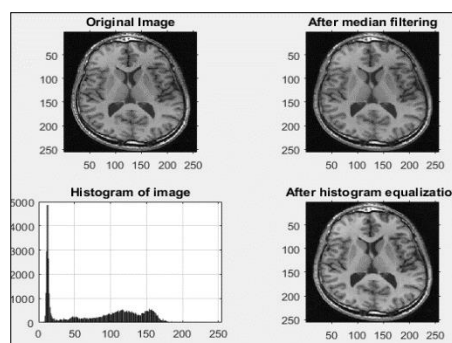


شکل (۱۴): قطعه‌بندی تصویر MR برای حذف نویز و تشخیص تومور

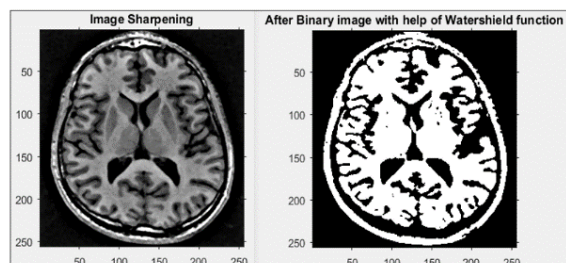


شکل (۱۵): بررسی مورفولوژیک تومور مغزی

در چارچوب پیشنهادی، ابتدا مرحله پیش‌پردازش شامل نرمال‌سازی شدت روشنایی، حذف نویز و بهبود کنتراست تصویر اعمال شده است. نتایج این مرحله در شکل (۱۱) ارائه شده است. این مرحله موجب افزایش نسبت سیگنال به نویز (SNR) و کاهش ناهمگنی شدت روشنایی می‌شود.



شکل (۱۱): پیش‌پردازش اولیه تصویر MR توسط مدل قابل اجرا در گام بعد، قطعه‌بندی تصویر با استفاده از الگوریتم واترشید مبتنی بر تابع آنتروپی انجام شده است. استفاده از معیار آنتروپی باعث افزایش حساسیت روش نسبت به تغییرات بافتی در ناحیه تومور شده و از بیش‌قطعه‌بندی جلوگیری می‌کند. خروجی این مرحله در شکل (۱۲) نمایش داده شده است.

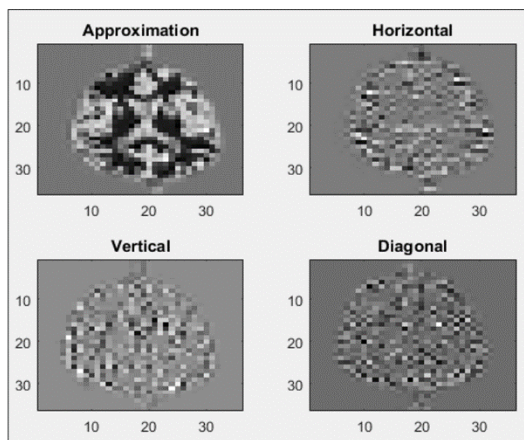


شکل (۱۲): قطعه‌بندی تصویر بر اساس واترشید مبتنی بر تابع آنتروپی

در ادامه، فرآیند جداسازی جمجمه از مغز^۱ با استفاده از استخراج شاخص‌های بافتی و بهینه‌سازی مرزی انجام شده است (شکل ۱۳). این مرحله نقش مهمی در جلوگیری از ورود ویژگی‌های غیرمرتبط به مرحله یادگیری عمیق دارد. در مرحله حذف جمجمه، جداسازی بافت مغز با استفاده از آستانه‌گذاری شدت و عملیات مورفولوژیک شامل بازشدن، بسته‌شدن و پرکردن حفره‌ها انجام می‌شود.

¹ Skull Stripping

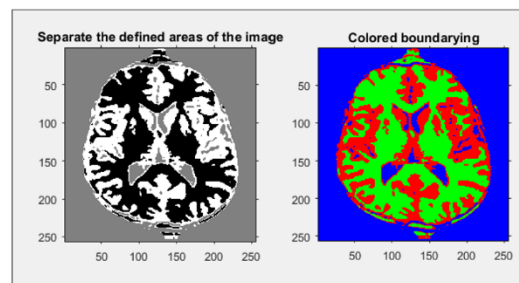
تولیدشده با FGSM نیز بررسی شدند. هدف این تحلیل، تعیین رفتار زنجیره پیش‌پردازش و طبقه‌بندی در حضور اغتشاشات هدفمند مبتنی بر گرادین بود. نتایج حاصل، شواهدی درباره عملکرد مدل تحت سناریوی مشخص FGSM فراهم می‌کنند؛ باین‌حال، به دلیل نبود ارزیابی تجربی با حملات تکرارشونده، از تعمیم این یافته‌ها به BIM، PGD یا سایر حملات خودداری می‌شود.



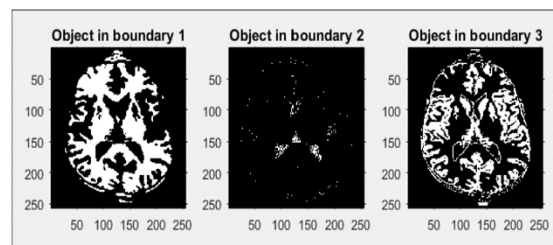
شکل (۱۸): نتایج مرحله حذف نویز برای تصاویر پاک و نمونه‌های متخاصم

شکل (۱۸) مؤلفه‌های عمودی، افقی، مورب و تقریبی تصاویر را پس از اعمال مرحله کاهش نویز نشان می‌دهد. مقایسه کیفی مؤلفه‌ها بیانگر کاهش بخشی از اغتشاشات و حفظ ساختارهای اصلی تصویر است. نتایج کیفی نشان می‌دهد که مرحله حذف نویز، تحت نمونه‌های FGSM بررسی شده، بخشی از اغتشاشات ورودی را کاهش داده و هم‌زمان ساختارهای اصلی تصویر را حفظ کرده است. این مشاهده به سناریوی FGSM مطالعه حاضر محدود است و برای نتیجه‌گیری درباره مقاومت مدل در برابر حملات تکرارشونده کافی نیست و نباید به‌عنوان اثبات مستقل کارایی denoising در دفاع متخاصم تفسیر شود.

پس از استخراج نواحی معنادار، این نواحی جهت آموزش شبکه عصبی کانولوشنی آماده‌سازی شده‌اند. فرآیند تفکیک نواحی و لایه‌بندی تصویر برای یادگیری عمیق در شکل‌های (۱۶) و (۱۷) نمایش داده شده است. این مرحله موجب تمرکز شبکه CNN سفارشی بر نواحی تشخیصی و کاهش نویز یادگیری می‌شود.



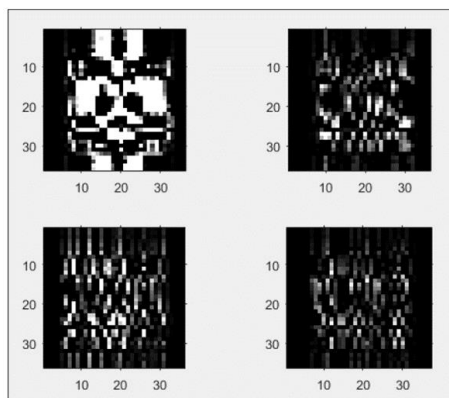
شکل (۱۶): جداسازی نواحی مختلف در تصویر MR برای آموزش توسط یادگیری عمیق



شکل (۱۷): بخش‌بندی لایه‌های تصویر برای آموزش و یادگیری عمیق

۴-۴ ارزیابی استحکام طبقه‌بند در برابر نمونه‌های متخاصم

به‌منظور ارزیابی استحکام چارچوب پیشنهادی، علاوه بر تصاویر پاک مجموعه آزمون، نسخه‌های متخاصم همان تصاویر با استفاده از حمله FGSM و در سه سطح شدت اغتشاش تولید شدند. نمونه‌های متخاصم کلاس جدیدی ایجاد نمی‌کنند و برچسب اصلی Tumor-positive یا Tumor-negative هر برش پس از اعمال حمله بدون تغییر باقی می‌ماند. هدف این ارزیابی، بررسی میزان حفظ عملکرد طبقه‌بند تومور در حضور اغتشاشات هدفمند است. تصاویر متخاصم پس از تولید، از همان زنجیره پردازشی مورد استفاده برای تصاویر پاک عبور داده شدند و عملکرد مدل روی نمونه‌های پاک و متخاصم ارزیابی شد. برای ارزیابی مقاومت چارچوب پیشنهادی، علاوه بر تصاویر پاک، نسخه‌های متخاصم



شکل (۲۰): نتایج نهایی پس از اعمال PCA

۴-۶ معیارهای ارزیابی و تحلیل زمان پردازش

در این پژوهش، به منظور ارزیابی دقیق و چندبعدی عملکرد چارچوب پیشنهادی، از معیارهای متداول طبقه‌بندی شامل دقت (Accuracy)، صحت (Precision)، امتیاز F1-score، حساسیت (Sensitivity)، ویژگی (Specificity) و شاخص AUC استفاده گردید. این شاخص‌ها بر پایه مقادیر استخراج‌شده از ماتریس درهم‌ریختگی محاسبه می‌شوند و هر یک از جنبه‌ای متفاوت، توانایی مدل را در تفکیک صحیح نمونه‌های مثبت و منفی نشان می‌دهند. روابط ریاضی این معیارها در ادامه و در قالب روابط (۳) تا (۷) ارائه شده است.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

در این روابط:

TP (True Positive): برش Tumor-positive که مدل آن را به درستی Tumor-positive پیش‌بینی کرده است.

۴-۵ کاهش ابعاد داده‌ها با تحلیل مؤلفه‌های اصلی

به منظور کاهش ابعاد داده‌ها و حذف همبستگی میان ویژگی‌های استخراج‌شده، از روش تحلیل مؤلفه‌های اصلی (PCA)^۱ استفاده شده است. این روش با تبدیل داده‌ها به مجموعه‌ای از مؤلفه‌های متعامد، امکان نمایش اطلاعات اصلی در فضایی با ابعاد کمتر را فراهم می‌کند. برای جلوگیری از نشت اطلاعات، PCA در هر تکرار اعتبارسنجی متقاطع به صورت مستقل و تنها بر ویژگی‌های مجموعه آموزش برآزش شد. پارامترهای حاصل از این برآزش بدون محاسبه مجدد برای تبدیل ویژگی‌های مجموعه آزمون همان fold به کار رفتند. بنابراین، داده‌های آزمون در تعیین مؤلفه‌های اصلی یا تعداد مؤلفه‌های منتخب دخالت نداشتند. نمونه‌ای از مؤلفه‌های استخراج‌شده در شکل (۱۹) و نتایج نهایی پس از اعمال PCA در شکل (۲۰) نمایش داده شده است.

Name	Value
se3	1x1 strl
segImage	256x256 logical
segmented_image	1x3 cell
sha_image	256x256 uint8
sizeTumor	[256,256]
Skewness	0.5363
Smoothness	0.9418
Standard_Deviation	0.0833
stats	1x1 struct
stripBrain	256x256 logical
stripBrain1	256x256 logical
striplImage	256x256 uint8
striplImage1	256x256 uint8
sum_b1b2	72
sum_b1b2b3	108
temp	-2.3226e-56
thresh	10
time	1x15 double
Variance	0.0069

شکل (۱۹): نمونه‌ای از مؤلفه‌های استخراج‌شده

نتایج نشان می‌دهد که استفاده از PCA ضمن حفظ بخش عمده‌ای از اطلاعات مهم داده‌ها، موجب کاهش پیچیدگی محاسباتی و افزایش کارایی فرآیند آموزش شبکه عصبی عمیق شده است. این کاهش ابعاد علاوه بر تسریع فرآیند یادگیری، احتمال بیش‌برآزش مدل را نیز کاهش داده و به بهبود قابلیت تعمیم آن کمک می‌کند.

¹ Principal Component Analysis

بیانگر عملکرد طبقه‌بندی در سطح برش برای بیماران دیده‌نشده هستند.

جدول (۳): معیارهای ارزیابی عملکرد طبقه‌بندی چارچوب پیشنهادی

مقدار (درصد)	معیار ارزیابی
۹۷/۲۳٪	دقت کل (Accuracy)
۹۶/۶۷٪	صحت (Precision)
۹۵/۷۸٪	بازیابی (Recall)
۹۸/۱۵٪	ویژگی (Specificity)
۹۶/۲۲٪	F1-score
۰/۹۸۴	سطح زیر منحنی (AUC)

عملکرد طبقه‌بندی چارچوب پیشنهادی در جدول (۳) ارائه شده است. مقدار Recall برابر با ۹۵/۷۸٪ نشان می‌دهد که مدل بخش عمده برش‌های Tumor-positive را به‌درستی شناسایی کرده است و در نتیجه نرخ خطای منفی کاذب آن نسبتاً پایین است. همچنین، Specificity برابر با ۹۸/۱۵٪ بیانگر توانایی مدل در طبقه‌بندی صحیح برش‌های Tumor-negative و کاهش پیش‌بینی‌های مثبت کاذب است. این نتایج به عملکرد مدل در سطح برش مربوطاند و نباید به تشخیص بیمار مبتلا از فرد سالم تعبیر شوند. مقدار AUC توانایی طبقه‌بندی را در تفکیک برش‌های Tumor-positive و Tumor-negative در دامنه‌ای از آستانه‌های تصمیم نشان می‌دهد. AUC برابر با ۰/۹۸۴، بیانگر قدرت تفکیک مناسب مدل روی داده‌های ارزیابی شده است؛ باین‌حال، این معیار به‌تنهایی اطلاعاتی درباره پایداری مدل در برابر نویز یا حملات متخاصم فراهم نمی‌کند. ارزیابی چنین پایداری‌ای مستلزم گزارش جداگانه Clean Accuracy و Robust Accuracy در بودجه‌های اغتشاش و حملات مشخص است.

تحلیل زمان اجرای مراحل مختلف چارچوب پیشنهادی در شکل (۲۱) ارائه شده است. به‌منظور ارزیابی هزینه محاسباتی روش، زمان اجرای مراحل اصلی در محیط MATLAB R2023a اندازه‌گیری شده است. طبق شکل (۲۱)، زمان صرف‌شده برای مراحل چهارگانه پیش‌پردازش به‌ترتیب برابر با ۰/۸۸۱۷، ۱/۱۶۸۰، ۱/۶۷۳۲ و ۰/۶۹۶۷ ثانیه گزارش شده است. در ادامه، زمان اجرای قطعه‌بندی

TN (True Negative): برش Tumor-negative که مدل آن را به‌درستی Tumor-negative پیش‌بینی کرده است.

FP (False Positive): برش Tumor-negative که به‌اشتباه Tumor-positive پیش‌بینی شده است.

FN (False Negative): برش Tumor-positive که به‌اشتباه Tumor-negative پیش‌بینی شده است.

معیار دقت (Accuracy) نسبت کل پیش‌بینی‌های صحیح به کل نمونه‌ها را نشان می‌دهد.

معیار صحت (Precision) بیانگر میزان صحت پیش‌بینی‌های مثبت است و مشخص می‌کند چه سهمی از نمونه‌های مثبت اعلام‌شده واقعاً مثبت هستند.

معیار بازیابی (Recall) که به‌عنوان نرخ مثبت واقعی نیز شناخته می‌شود، توانایی مدل در شناسایی صحیح نمونه‌های مثبت را اندازه‌گیری می‌کند.

علاوه بر معیارهای فوق، شاخص F1-score، به‌عنوان میانگین هارمونیک صحت و بازیابی، معیار متعادل‌تری را برای ارزیابی عملکرد مدل فراهم می‌کند؛ به‌ویژه در شرایطی که توزیع کلاس‌ها ممکن است کاملاً متقارن نباشد، این شاخص نشان‌دهنده توانمندی هم‌زمان مدل در حفظ دقت و قدرت فراخوانی است.

معیار ویژگی (Specificity) یا نرخ منفی واقعی، توانمندی مدل را در تشخیص صحیح نواحی پاک و جلوگیری از اعلام مثبت کاذب اندازه‌گیری می‌کند.

درنهایت، شاخص AUC که بیانگر سطح زیر منحنی مشخصه عملکرد سیستم (ROC) است، به‌عنوان یک معیار جامع مستقل از آستانه، قدرت تفکیک کلی مدل را در جداسازی کلاس‌های مثبت و منفی در تمامی سطوح تصمیم‌گیری نشان می‌دهد. مقدار نزدیک‌تر این شاخص به عدد یک، گویای پایداری بالاتر و خطای دسته‌بندی کمتر در شرایط عملیاتی و نویزی است.

از آنجاکه چندین برش متعلق به یک بیمار هستند، تفکیک مجموعه‌های آموزش و آزمون پیش از محاسبه معیارها در سطح بیمار انجام شد. باین‌حال، معیارهای Accuracy، Precision، Recall، Specificity، F1-score و AUC بر اساس پیش‌بینی برش‌های خارج از fold محاسبه شدند؛ بنابراین نتایج گزارش‌شده

جدول (۴): مقایسه کمی عملکرد روش‌های قطعه‌بندی

Jaccard	Dice	روش قطعه‌بندی
۰/۷۸	۰/۸۶	واترشیپ کلاسیک
۰/۸۲	۰/۸۹	واترشیپ + پیش‌پردازش
۰/۸۸	۰/۹۳	روش پیشنهادی

به‌طور مشخص، مقدار Dice از ۰/۸۶ در روش پایه به ۰/۹۳ در روش پیشنهادی افزایش یافته است که بیانگر بهبود حدود ۷ درصدی در همپوشانی ناحیه تومور است

به‌منظور اطمینان از تکرارپذیری و اعتبار سنجی نتایج، تمامی مراحل آموزش و آزمون مدل ۵ مرتبه با استفاده از زیرمجموعه‌های متفاوت داده (K=5) تکرار شد. همان‌طور که در جدول (۵) مشاهده می‌شود، انحراف معیار پایین در روش پیشنهادی (SD=0.41) نشان‌دهنده پایداری بالای چارچوب در مواجهه با داده‌های مختلف است.

جدول (۵): تأثیر مراحل پیش‌پردازش و قطعه‌بندی بر عملکرد

طبقه‌بندی CNN سفارشی (میانگین $\pm$ انحراف معیار)

Accuracy (%)	نوع ورودی به شبکه CNN سفارشی
۹۴/۱ $\pm$ ۰/۸۵	تصویر خام
۹۵/۳ $\pm$ ۰/۶۲	تصویر پس از پیش‌پردازش
۹۷/۲۳ $\pm$ ۰/۴۱	خروجی قطعه‌بندی پیشنهادی

هم‌چنین، برای ارزیابی معناداری آماری بهبود حاصل‌شده، آزمون t-test تک‌نمونه‌ای بین دقت روش پیشنهادی و بهترین روش رقیب انجام گرفت. مقدار $p\text{-value}=0.027$ حاصل شد که با توجه به سطح معناداری ($\alpha=0.05$)، فرضیه صفر رد شده و برتری عملکرد روش پیشنهادی از نظر آماری کاملاً معنادار تلقی می‌شود. این تحلیل تأیید می‌کند که بهبود دقت حاصله، ناشی از طراحی منسجم چارچوب چندمرحله‌ای است و صرفاً حاصل تنظیمات تصادفی وزن‌های شبکه نیست. برای بررسی معناداری آماری بهبود در

با بازسازی مورفولوژیک ۰/۷۸۶۱ ثانیه و تعیین مرزها ۱/۰۲۴۵ ثانیه ثبت گردید. هم‌چنین، زمان استخراج ویژگی با تبدیل موجک (DWT^۱) برابر با ۰/۴۰۰۳ ثانیه و کاهش ابعاد با PCA برابر با ۰/۴۸۹۹ ثانیه است. نهایتاً، زمان طبقه‌بندی عمیق ۴۹۹۳/۱ ثانیه اندازه‌گیری شد. در کنار این کارایی زمانی، شاخص‌های ارزیابی عملکرد شامل دقت ۹۷/۲۳٪، precision برابر با ۹۶/۶۷٪ و recall برابر با ۹۵/۷۸٪ حاصل شده است که نشان‌دهنده توازن مطلوب میان پیچیدگی محاسباتی و دقت مدل است.

The processing time of each stage is:
Pre-processing time: 0.8817 1.1680 1.6732 0.6967
Processing time of segmetation by morphological reconstruction: 0.7861
Processing time of segmetation by boundary: 1.0245
Processing time of DWT: 0.4003
Processing time of PCA: 0.4899
Classification processing time: 1.4993
Accuracy: 97.2300
precision: 96.6700
Recall: 95.7800

شکل (۲۱): زمان پردازش مراحل مختلف روش پیشنهادی

۷-۴ ارزیابی کمی قطعه‌بندی و تحلیل تأثیر آن بر

عملکرد طبقه‌بندی

پس از آماده‌سازی داده‌ها برای آموزش شبکه عصبی، عملکرد مرحله قطعه‌بندی به‌صورت کمی ارزیابی شد تا دقت استخراج ناحیه تومور بررسی گردد. برای این منظور از معیارهای استاندارد ارزیابی قطعه‌بندی شامل (DSC^۲) و Jaccard Index استفاده شد. معیار Dice که یکی از مهم‌ترین شاخص‌های همپوشانی بین ناحیه قطعه‌بندی‌شده و ناحیه مرجع است، مطابق رابطه (۸) محاسبه می‌شود.

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (۸)$$

در این رابطه TP، FP و FN به‌ترتیب نشان‌دهنده مثبت صحیح، مثبت کاذب و منفی کاذب هستند. مقدار بالاتر Dice بیانگر همپوشانی بیشتر با ناحیه مرجع و عملکرد دقیق‌تر الگوریتم قطعه‌بندی است. نتایج مقایسه روش پیشنهادی با دو روش پایه شامل واترشیپ کلاسیک و واترشیپ همراه با پیش‌پردازش در جدول (۴) ارائه شده است. نتایج نشان می‌دهد که چارچوب پیشنهادی در تمامی معیارهای ارزیابی عملکرد بهتری داشته است.

^۲ Dice Similarity Coefficient

^۱ Discrete Wavelet Transform

الگوریتم ژنتیک به‌طور مستقل حذف و ارزیابی نشده است. افزون بر این، معیار جدول دقت طبقه‌بندی روی داده‌های ارزیابی شده است و به‌تنهایی نقش اختصاصی denoising در افزایش adversarial robustness را اثبات نمی‌کند. بنابراین، بهبود مشاهده‌شده باید به اثر تجمعی مراحل پیش‌پردازش و قطع‌بندی نسبت داده شود و نه منحصرأ به مرحله حذف نویز.

۴-۸ مقایسه با روش‌های مشابه

برای جلوگیری از مقایسه نتایج حاصل از مجموعه داده‌ها یا مسائل متفاوت، جدول زیر فقط شامل مطالعاتی است که از نسخه ۳۹۲۹ تصویری مجموعه TCGA-LGG/TCIA یا نسخه تقریباً یکسان آن استفاده کرده‌اند، مسئله طبقه‌بندی دودویی حضور یا فقدان ناحیه تومور را بررسی کرده‌اند و Accuracy گزارش شده آن‌ها برحسب درصد بوده است.

مرحله قطع‌بندی، آزمون t زوجی بین مقادیر Dice روش واترشید کلاسیک و روش پیشنهادی انجام شد. مقدار $p = 0.004$ به‌دست آمده، که بسیار کوچک‌تر از سطح معناداری ۰/۰۱ است، نشان می‌دهد که هم‌افزایی تابع آنتروپی و پیش‌پردازش هدفمند، منجر به بهبودی سیستماتیک و معتبر در دقت جداسازی نواحی توموری شده است.

نتایج جدول (۵) یک تحلیل مرحله‌ای محدود از اثر مراحل ابتدایی چارچوب ارائه می‌کنند. استفاده از تصویر پس از پیش‌پردازش، دقت را از ۹۴/۱۰ به ۹۵/۳۰ درصد افزایش داده است که معادل ۱/۲۰ واحد درصد بهبود نسبت به ورودی خام است. پس از افزودن قطع‌بندی پیشنهادی، دقت به ۹۷/۲۳٪ رسیده است که ۱/۹۳ واحد درصد بیشتر از حالت پیش‌پردازش و ۳/۱۳ واحد درصد بیشتر از ورودی خام است. کاهش انحراف معیار از ۰/۸۵ به ۰/۴۱ نیز بیانگر نوسان کمتر نتایج در تنظیمات گزارش شده است. با این حال، این مقایسه یک ablation جامع محسوب نمی‌شود؛ زیرا اثر PCA و

جدول (۶): مقایسه عملکرد روش پیشنهادی با روش‌های پژوهش‌های پیشین

مرجع	سال	تکنیک	مجموعه داده	تعداد تصاویر	Accuracy (%)
[۱۷]	۲۰۲۳	ResNet50 + Deep Residual U-Net	TCGA-LGG/TCIA-Kaggle	۳۹۲۹	۹۴/۲۳
[۱۸]	۲۰۲۴	Hybrid CNN: InceptionV3 + ResNet50 + VGG16 + DenseNet همراه با oneAPI	TCGA-LGG-Kaggle	۳۹۲۹	۹۶/۲۰
[۱۹]	۲۰۲۴	Fine-tuned CNN + Residual U-Net	TCGA-LGG/TCIA آزمایش دوم	تقریباً ۳۹۲۹	۹۶/۰۰
روش پیشنهادی	—	Denoising + Segmentation + Feature Extraction + PCA + GA + CNN	TCGA-LGG/TCIA-Kaggle	۳۹۲۹	۹۷/۲۳

پیشنهادی پیش از طبقه‌بندی، حذف نویز، قطع‌بندی، کاهش بُعد با PCA و انتخاب ویژگی مبتنی بر الگوریتم ژنتیک را اجرا می‌کند. در مقابل، روش‌های مقایسه‌شده عمدتاً به یادگیری انتهابه‌انتها یا ترکیب چند معماری عمیق متکی هستند.

در این زیرمجموعه غربال‌شده، روش پیشنهادی به‌ترتیب ۳/۰۰، ۱/۰۳ و ۱/۲۳ واحد درصد Accuracy بیشتری از روش‌های [۱۷]، [۱۸] و [۱۹] نشان می‌دهد. این تفاوت می‌تواند با پالایش چندمرحله‌ای ورودی و فضای ویژگی مرتبط باشد؛ زیرا چارچوب

۵- نتیجه‌گیری و پیشنهادها آتی

در این پژوهش، یک چارچوب چندمرحله‌ای و مقاوم مبتنی بر یادگیری عمیق برای تحلیل تصاویر MR مغزی ارائه شد که با تلفیق مراحل پیش‌پردازش، حذف نویز، قطعه‌بندی، استخراج و بهینه‌سازی ویژگی، و طبقه‌بندی نهایی، توانست عملکرد مناسبی در طبقه‌بندی برش‌های MR به دو کلاس Tumor-positive و Tumor-negative ارائه دهد و در عین حال، امکان ارزیابی پایداری این طبقه‌بندی را در برابر اغتشاشات متخاصم فراهم سازد. نتایج نشان دادند که استفاده تجمعی از پیش‌پردازش و قطعه‌بندی با بهبود دقت طبقه‌بندی نسبت به ورودی خام همراه بوده است. با این حال، به دلیل یکپارچه بودن مراحل مختلف چارچوب، سهم مستقل حذف نویز، PCA و الگوریتم ژنتیک از عملکرد نهایی در مطالعه حاضر تعیین نشده است. نتایج به‌دست‌آمده نشان دادند که چارچوب پیشنهادی، ضمن حفظ دقت بالا، از نظر ساختاری نیز برای تحلیل تصاویر پزشکی پیچیده و ناهمگن مناسب است. همچنین، ارزیابی انجام‌شده با FGSM نشان داد که ترکیب پیش‌پردازش و حذف نویز می‌تواند تحت تنظیمات بررسی‌شده، حساسیت مدل را نسبت به اغتشاشات مبتنی بر گرادین کاهش دهد. این نتیجه صرفاً به حمله FGSM و تنظیمات آزمایشی مطالعه حاضر مربوط است.

بهره‌گیری از ترکیب روش‌های پردازش تصویر و یادگیری عمیق در این پژوهش، زمینه‌ای مناسب برای توسعه سامانه‌های دقیق‌تر و مقاوم‌تر در کاربردهای پزشکی فراهم می‌سازد.

با وجود نتایج مطلوب، روش پیشنهادی با محدودیت‌هایی نیز همراه است. ارزیابی این چارچوب بر روی یک مجموعه داده مشخص و یک معماری معین انجام‌شده است؛ از این رو، بررسی آن بر روی داده‌ها و معماری‌های متنوع‌تر می‌تواند به ارزیابی دقیق‌تر تعمیم‌پذیری روش کمک کند. همچنین عملکرد روش به کیفیت تصاویر ورودی، شدت نویز و دقت مرحله قطعه‌بندی وابسته است و در تصاویر با کنتراست پایین، مرزهای مبهم یا نویز شدید ممکن است کاهش یابد. افزون بر این، استفاده از الگوریتم واترشیف، با وجود بهبود کیفیت پردازش، موجب افزایش زمان پردازش و هزینه محاسباتی می‌شود. همچنین ارزیابی پایداری متخاصم عمده‌تاً بر پایه حمله FGSM انجام‌شده است و بررسی روش در برابر حملات متنوع‌تر می‌تواند در مطالعات آینده دنبال شود.

در ادامه این مسیر، پیشنهاد می‌شود ارزیابی روش بر روی مجموعه داده‌های مستقل‌تر و متنوع‌تر، معماری‌های عمیق مختلف، و سناریوهای بالینی واقعی‌تر انجام شود تا دامنه تعمیم‌پذیری آن با دقت بیشتری مشخص گردد. همچنین، استفاده از خروجی‌های دوبعدی برای بازسازی سه‌بعدی تومور، افزودن قابلیت تمایز میان تومورهای خوش‌خیم و بدخیم، بررسی الگوریتم‌های فراابتکاری دیگر برای انتخاب و بهینه‌سازی ویژگی‌ها، و بهره‌گیری از روش‌های پیشرفته‌تر استخراج ویژگی و معماری‌هایی نظیر FCN می‌تواند به ارتقای بیشتر عملکرد روش منجر شود. پیگیری این مسیرها می‌تواند زمینه‌ساز توسعه چارچوب‌هایی دقیق‌تر، سریع‌تر، هوشمندتر و مقاوم‌تر برای تحلیل تصاویر پزشکی در محیط‌های واقعی بالینی باشد.

References

- [1] Joel, M. Z., Avesta, A., Yang, D. X., Zhou, J.-G., Omuro, A., Herbst, R. S., Krumholz, H. M., & Aneja, S. (2023). Comparing detection schemes for adversarial images against deep learning models for cancer imaging. *Cancers*, 15(5), 1548. <https://doi.org/10.3390/cancers15051548>
- [2] Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R., & Rosen, M. S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697), 487–492. <https://doi.org/10.1038/nature25988>
- [3] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. J., & Fergus, R. (2014). Intriguing properties of neural networks. *International Conference on Learning Representations*. <https://arxiv.org/abs/1312.6199>
- [4] Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, 363(6433), 1287–1289. <https://doi.org/10.1126/science.aaw4399>
- [5] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *International Conference on Learning Representations*. <https://arxiv.org/abs/1412.6572>

- [6] Kurakin, A., Goodfellow, I. J., & Bengio, S. (2017). Adversarial examples in the physical world. International Conference on Learning Representations Workshop. <https://arxiv.org/abs/1607.02533>
- [7] Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. International Conference on Learning Representations. <https://arxiv.org/abs/1706.06083>
- [8] Papernot, N., McDaniel, P., Goodfellow, I. J., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security (pp. 506–519). Association for Computing Machinery. <https://doi.org/10.1145/3052973.3053009>
- [9] Liu, A., Liu, X., Yu, H., Zhang, C., Liu, Q., & Tao, D. (2021). Training robust deep neural networks via adversarial noise propagation. IEEE Transactions on Image Processing, 30, 5769–5781. <https://doi.org/10.1109/TIP.2021.3082317>
- [10] Lin, J., Gan, C., & Han, S. (2019). Defensive quantization: When efficiency meets robustness. International Conference on Learning Representations. <https://arxiv.org/abs/1904.08444>
- [11] Xie, C., Wu, Y., van der Maaten, L., Yuille, A. L., & He, K. (2019). Feature denoising for improving adversarial robustness. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 501–509). IEEE. <https://doi.org/10.1109/CVPR.2019.00059>
- [12] Kwon, H., Kim, Y., Yoon, H., & Choi, D. (2021). Classification score approach for detecting adversarial example in deep neural network. Multimedia Tools and Applications, 80, 10339–10360. <https://doi.org/10.1007/s11042-020-09167-z>
- [13] Bortsova, G., González-Gonzalo, C., Wetstein, S. C., Dubost, F., Katramados, I., Hogeweg, L., Liefers, B., van Ginneken, B., Pluim, J. P. W., Veta, M., & van Tulder, G. (2021). Adversarial attack vulnerability of medical image analysis systems. Medical Image Analysis, 73, 102141. <https://doi.org/10.1016/j.media.2021.102141>
- [14] Hoseini, F., Shamlou, S., & Ahmadi-Gharehtoragh, M. (2024). Segmentation of MR images for brain tumor detection using autoencoder neural network. Discover Artificial Intelligence, 4(1), 71. <https://doi.org/10.1007/s44163-024-00180-x>
- [15] Chakraborty, A., Alam, M., Dey, V., Chattopadhyay, A., & Mukhopadhyay, D. (2021). A survey on adversarial attacks and defences. CAAI Transactions on Intelligence Technology, 6(1), 25–45. <https://doi.org/10.1049/cit.2.12028>
- [16] Buda, M., Saha, A., & Mazurowski, M. A. (2019). Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. Computers in Biology and Medicine, 109, 218–225. <https://doi.org/10.1016/j.compbiomed.2019.05.002>
- [17] Gajula, S., & Rajesh, V. (2023). Performance analysis of classification and detection of brain tumor MRI images using ResNet50 deep residual U-Net. Journal of Iranian Medical Council, 6(1), 101–110. <https://doi.org/10.18502/jimc.v6i1.11848>
- [18] Ramakrishnan, A. B., Sridevi, M., Vasudevan, S. K., Manikandan, R., & Gandomi, A. H. (2024). Optimizing brain tumor classification with hybrid CNN architecture: Balancing accuracy and efficiency through oneAPI optimization. Informatics in Medicine Unlocked, 44, 101436. <https://doi.org/10.1016/j.imu.2023.101436>
- [19] Alshomrani, F. (2024). A unified pipeline for simultaneous brain tumor classification and segmentation using fine-tuned CNN and residual UNet architecture. Life, 14(9), 1143. <https://doi.org/10.3390/life14091143>
- [20] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- [21] Roodaki, A., Sotoodeh, M., & Moosavi, M. R. (2025). Genetic algorithm-based feature selection from high-dimensional descriptors for improved content-based image retrieval. IEEE Access, 13, 198034–198053. <https://doi.org/10.1109/ACCESS.2025.3633906>
- [22] Naser, M. A., & Deen, M. J. (2020). Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. Computers in Biology and Medicine, 121, 103758.

- <https://doi.org/10.1016/j.compbimed.2020.103758>
- [23] Asiri, A. A., Shaf, A., Ali, T., Aamir, M., Irfan, M., Alqahtani, S., Mehdar, K. M., Halawani, H. T., Alghamdi, A. H., Alshamrani, A. F. A., & Alqhtani, S. M. (2023). Brain tumor detection and classification using fine-tuned CNN with ResNet50 and U-Net model: A study on TCGA-LGG and TCIA dataset for MRI applications. *Life*, 13(7), 1449. <https://doi.org/10.3390/life13071449>
- [24] Anaya-Isaza, A., Mera-Jiménez, L., Verdugo-Alejo, L., & Sarasti, L. (2023). Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network. *European Journal of Radiology Open*, 10, 100484. <https://doi.org/10.1016/j.ejro.2023.100484>
- [25] Sulaiman, A., Anand, V., Gupta, S., Al Reshan, M. S., Alshahrani, H., Shaikh, A., & Elmagzoub, M. A. (2024). An intelligent LinkNet-34 model with EfficientNetB7 encoder for semantic segmentation of brain tumor. *Scientific Reports*, 14, 1345. <https://doi.org/10.1038/s41598-024-51472-2>
- [26] Rai, H. M., Chatterjee, K., & Dashkevich, S. (2021). Automatic and accurate abnormality detection from brain MR images using a novel hybrid UnetResNext-50 deep CNN model. *Biomedical Signal Processing and Control*, 66, 102477. <https://doi.org/10.1016/j.bspc.2021.102477>

An Integrated Framework Based on Targeted Pre-processing and Deep Learning for Denoising and Enhancing Robustness Against Adversarial Attacks in MR Images

Farnaz Hoseini^{1*}

¹Assistant Professor, Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran

Article Information

Original Research Paper

Received:

2026 June 04

Accepted:

2026 September 12

Keywords:

Deep Neural Networks,
Adversarial Examples,
Magnetic Resonance Imaging
(MRI), Image Denoising,
Genetic Algorithm-based
Feature Optimization, Medical
Image Classification

Corresponding Author*:

f-hoseini@tvu.ac.ir

Abstract

Convolutional neural networks used in medical image analysis are sensitive to noise and adversarial perturbations. This study presents a multistage framework for binary classification of brain magnetic resonance slices as tumor-positive, indicating a visible tumor, or tumor-negative, indicating no visible tumor. The framework integrates nonlinear preprocessing, watershed segmentation, feature extraction, dimensionality reduction using principal component analysis, genetic-algorithm-based feature selection, and convolutional neural network classification. It was evaluated on 3,929 slices from 110 patients in the TCGA-LGG/TCIA dataset using patient-wise five-fold cross-validation. On clean images, the model achieved 97.23% accuracy, a 96.22% F1-score, and an area under the receiver operating characteristic curve (AUC) of 0.984. Robustness was assessed separately under the fast gradient sign method (FGSM) attack at three perturbation levels. The findings support the utility of this processing pipeline for classifying slices from unseen patients; however, independent external validation remains necessary before clinical generalization to other populations or imaging environments.

 : 10.22034/ABMIR.2026.24696.1256

E-ISSN: [2821-2037](#)

/The Author 2026. Published by Yazd University This is an open access article under the CC BY 4.0 License <https://creativecommons.org/licenses/by/4.0/>.

