توسعه پایدار و ساختمان هوشمند
https://sdsb.buqaen.ac.ir

جلد ۱، شماره ۳، پاییز ۱۴۰۳

بهبود طبقه‌بندی متاستاز مغزی با استفاده از بهینه‌سازی ویژگی‌های رادیومیک

سیما حق‌گزار[۱]، مهدی خزاعی‌پور.[۲*]

۲- گروه مهندسی کامپیوتر- دانشکده فنی و مهندسی- دانشگاه آزاد اسلامی واحد بیرجند- بیرجند- ایران.
۳- گروه مهندسی کامپیوتر- دانشکده فنی و مهندسی- دانشگاه آزاد اسلامی واحد بیرجند- بیرجند- ایران.

چکیده

این مطالعه با هدف بهبود دقت طبقه‌بندی متاستاز مغزی بر روی یک مجموعه دادهٔ جامع از تصاویر MRI مغز همراه با داده‌های بالینی و مورفولوژیکی انجام شده است. افزایش حجم مجموعه داده‌های پزشکی، تعداد خصوصیاتی با ابعاد بالا را ایجاد کرده است که بر طبقه‌بندهای یادگیری ماشین تأثیر منفی می‌گذارد. در یادگیری ماشین، فرایند انتخاب ویژگی برای انتخاب مرتبط‌ترین ویژگی‌ها و کاهش موارد اضافی و نامربوط، اساسی است. الگوریتم‌های بهینه‌سازی، توانایی خود را در حل مسائل انتخاب ویژگی نشان می‌دهند. الگوریتم جستجوی خزندگان، یک الگوریتم جدید بهینه‌سازی الهام‌گرفته از طبیعت است که رفتار احاطه‌کردن و شکار کروکودیل‌ها را شبیه‌سازی می‌کند. جستجوی منحصربه‌فرد الگوریتم RSA نتایج امیدوارکننده‌ای را در مقایسه با سایر الگوریتم‌های بهینه‌سازی به‌دست می‌آورد؛ بنابراین، در مقاله از این الگوریتم برای جستجوی مجموعه ویژگی‌های مهم استفاده شده است. در این مقاله برای طبقه‌بندی داده‌ها از الگوریتم جنگل تصادفی استفاده شده است. یعنی در ابتدا یک مدل جنگل تصادفی بر روی تمام ویژگی‌های رادیومیک موجود در مجموعه داده آموزش داده شده است. سپس با استفاده از الگوریتم بهینه‌سازی فرا اکتشافی جستجوی خزندگان زیرمجموعه‌ای از ویژگی‌های مهم انتخاب شد. سپس نتایج طبقه‌بندی در دو حالت با هم مقایسه شد. حالت اول بدون اعمال الگوریتم RSA، و حالت دوم با اعمال الگوریتم RSA و انتخاب ویژگی‌های مهم. نتایج نشان داد که با اعمال الگوریتم RSA، در طبقه‌بندی داده‌ها بهبود حاصل شده است، و این رویکرد می‌تواند در تشخیص سریعتر و پیش‌بینی متاستاز مغزی کمک کند و همین‌طور می‌توان با استفاده از این بهینه‌سازی قدرتمند، سرعت آموزش مدل و کیفیت آن را نیز به طور قابل توجهی افزایش داد.

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۷/۲۰

تاریخ پذیرش: ۱۴۰۳/۰۹/۱۸

واژه‌های کلیدی:

متاستاز مغزی، ویژگی‌های رادیومیک، انتخاب ویژگی، جنگل تصادفی، الگوریتم جستجوی خزندگان.

* نویسنده مسئول، پست الکترونیکی: Mkhazaiepoor@iau.ac.ir

Journal of Sustainable Development and Smart Building, vol. ۱, no. ۳
Digital Object Identifier https://doi.org/۱۰,۲۲۰۳٤/sdsb.۱,۳,۳٤
THIS WORK IS LICENSED UNDER CC BY-NC ٤,٠ INTERNATIONAL LICENSE.
© ۲۰۲٤ BOZORGMEHR UNIVERSITY OF QAENAT, IRAN

مقدمّه

تشخیص و درمان بیماری به شدت به طبقه‌بندی مجموعه داده‌های زیست‌پزشکی بستگی دارد. طبقه‌بندی چنین مجموعه‌های داده می‌تواند بیماری‌های پیچیده‌ای مانند کووید-۱۹، تومورها و موارد دیگر را شناسایی کند. تشخیص زودهنگام این گونه بیماری‌ها میزان بقا را افزایش می‌دهد [۱]. در علوم زیست‌پزشکی، بیماری‌های دسته‌بندی شده بر اساس ویژگی‌های مختلف طبقه‌بندی می‌شوند [۴,۳,۲]. مجموعه داده‌های زیست‌پزشکی به سرعت در حال رشد هستند، در نتیجه ویژگی‌های با بعد بالا دارند [۵]. در برخی موارد، این ویژگی‌ها اضافی و ناکارآمد هستند، یا همان اثر طبقه‌بندی را در خود جای می‌دهند [۶]. یک طبقه‌بند قوی یادگیری ماشین[۱] برای کاهش پیچیدگی و زمان صرف شده برای طبقه‌بندی این ویژگی‌ها مورد نیاز است [۷]. طبقه‌بندی یادگیری ماشین از ویژگی‌های اضافی، ناکارآمد و مغرضانه دچار مشکل می‌شود [۸]، بنابراین، انتخاب ویژگی، یک جزء مهم از فرایندهای یادگیری ماشین است [۹].

انتخاب ویژگی، نقش مهمی در یادگیری ماشین به عنوان یک مرحله پیش‌پردازش، هرس‌کردن ویژگی‌های اضافی و نامربوط و انتخاب مرتبط‌ترین آن‌ها دارد. این فرایند را می‌توان با حذف ویژگی‌هایی که ممکن است بر عملکرد طبقه‌بندی‌کننده تأثیر منفی بگذارد، مانند ویژگی‌های نامرتبط، زائد و کم اطلاعات انجام داد [۱۰]. انتخاب ویژگی به طور گسترده در بسیاری از برنامه‌های کاربردی استفاده شده است، از جمله تقسیم‌بندی تصویر [۱۱]، پردازش تصویر [۱۲]، تشخیص پزشکی [۱۳]، تشخیص سرطان [۱۴]، تشخیص متن [۱۵] و موارد دیگر. بر اساس مقالات اخیر، مراحل انتخاب ویژگی شامل چهار مرحله اساسی است: (۱) تولید زیرمجموعه‌های ویژگی، (۲) ارزیابی زیرمجموعه‌ها با استفاده از معیارهای مشخص، (۳) تعیین شرط توقف برای فرایند انتخاب، و (۴) اعتبارسنجی نهایی زیرمجموعه انتخاب‌شده. به‌عنوان مثال، در مقاله‌ای توسط Petinrin و همکاران [۱۶]، این مراحل به‌طور جامع مورد بررسی قرار گرفته‌اند. همچنین، در مطالعه‌ای توسط Javed و همکاران [۱۷]، ترکیب روش‌های فیلتر و روکش برای انتخاب ویژگی معرفی شده است که نشان‌دهنده اهمیت این مراحل در بهبود دقت مدل‌های یادگیری ماشین است. علاوه بر این، در مقاله‌ای توسط Suresh و همکاران [۱۸]، تأثیر انتخاب ویژگی بر بهبود دقت طبقه‌بندی‌کننده‌ها مورد بررسی قرار گرفته است. بر اساس معیارهای ارزیابی، روش‌های انتخاب ویژگی به دو رویکرد اصلی تقسیم می‌شوند: روش‌های مبتنی بر فیلتر و روش‌های مبتنی بر روکش،[۲] که هر یک مزایا و محدودیت‌های خاص خود را دارند.

رویکرد مبتنی بر فیلتر[۳] رویکردی برای فیلتر کردن زیرمجموعه‌های ویژگی بر اساس آزمون‌های ارزیابی ایستا است. همچنین، FBA رویکردی برای تغییر زیرمجموعه ویژگی‌ها بر اساس آزمون‌های ارزیابی ایستا است [۲۰,۱۹]. همبستگی پیرسون، آزمون مجذور کای و تجزیه و تحلیل متمایز خطی[۴] نمونه‌هایی از رویکردهای FBA هستند که فیلتر کردن قبل از طبقه‌بند ML برنامه، بدون تماس مستقیم با طبقه‌بند انجام می‌شود [۲۱]. در مقابل، رویکرد مبتنی بر روکش، مستقیماً به طبقه‌بند متصل است و زیرمجموعه‌های ویژگی‌ها را بر اساس عملکرد مدل یادگیری ماشین ارزیابی می‌کند [۲۲]. این رویکرد که به WBA نیز شناخته می‌شود، برای یافتن همبستگی‌های احتمالی میان ویژگی‌ها طراحی شده و با وجود هزینه محاسباتی بالا، معمولاً نسبت به روش مبتنی بر فیلتر نتایج دقیق‌تری ارائه می‌دهد [۲۳, ۲۴].

بر اساس مطالعات، الگوریتم‌های بهینه‌سازی متعددی در قالب WBA برای حل مسئله انتخاب ویژگی[۵] به کار گرفته شده‌اند. برای نمونه، الگوریتم بهینه‌سازی شامپانزه[۶] در حالت پوششی برای انتخاب ویژگی استفاده شده است [۵]، الگوریتم سنجاقک[۷] بهبود یافته با پویایی جمعیت تکاملی و تقاطع تطبیقی به صورت خاص برای بهبود عملکرد در انتخاب ویژگی‌ها طراحی شده است [۲۵]، الگوریتم بهینه‌سازی پروانه[۸] با بهره‌گیری از مفاهیم تئوری اطلاعات، در انتخاب ویژگی نقش مؤثری داشته است [۲۶]، و الگوریتم ازدحام ذرات[۹] با مکانیزم مارپیچی و همچنین الگوریتم بهینه‌سازی نهنگ[۱۰] ترکیب‌شده با بازپخت شبیه‌سازی‌شده[۱۱]، هر دو در حالت پوششی برای بهبود دقت انتخاب ویژگی توسعه یافته‌اند [۲۷, ۲۸]. هدف اصلی این الگوریتم‌ها یافتن مجموعه‌ای بهینه یا نزدیک به بهینه از ویژگی‌ها در زمانی معقول است. این روش‌ها ضمن ارتقاء دقت، به کاهش تعداد ویژگی‌های غیرضروری نیز کمک می‌کنند. در این پژوهش از طبقه‌بند KNN به‌عنوان معیار ارزیابی عملکرد زیرمجموعه‌های انتخاب‌شده استفاده شده است.

این مقاله سعی می‌کند تا از ترکیب رویکرد جنگل تصادفی و الگوریتم جستجوی خزندگان[۱۲] استفاده کند تا کارایی طبقه‌بندی برای مسائل پزشکی در زمینه داده‌های متاستاز مغزی و ویژگی‌های رادیومیک را بهبود ببخشد. از بدن بیماران سرطانی، اطلاعاتی استخراج می‌شوند که نشان‌دهنده تومورهای سرطانی در نقاط مختلف بدن بیماران سرطانی هستند. این خصوصیات با استفاده از الگوریتم‌هایی از تصاویر MRI تومورهای بیماران استخراج شده‌اند. به این خصوصیات، خصوصیات رادیویی[۱۳] گفته می‌شود. برای این منظور، الگوریتم جنگل تصادفی به عنوان الگوریتم پایه برای طبقه‌بندی استفاده شده است. سپس از الگوریتم بهینه‌سازی جستجوی خزندگان برای انتخاب صفات مؤثر در

۱. Machine Learning (ML).
۲. Wrapper Based Approach (WBA).
۳. Filter Based Approach (FBA).
۴. Linear Discriminant Analysis (LDA).
۵. Feature Selection (FS).
۶. Chimp Optimization Algorithm (COA).
۷. Dragonfly Algorithm (DA).
۸. Butterfly Optimization Algorithm (BOA).
۹. Particle Swarm Optimization (PSO).
۱۰. Whale Optimization Algorithm (WOA).
۱۱. Simulated Annealing (SA).
۱۲. Reptile search Algorithm (RSA).
۱۳. Radiomic Features.

طبقه‌بندی و بهبود دقت طبقه‌بندی استفاده شده است تا صفات بهینه در مجموعه صفات رادیومیک در تصاویر MRI متاستاز مغزی را انتخاب کند.

ادامهٔ این مقاله، به صورت زیر سازماندهی شده است: در بخش ۳، روش پیشنهادی بیان می‌شود. مجموعه داده‌ها، نتایج، مقایسه‌ها و ارزیابی در بخش ۴ به تفصیل شرح داده می‌شود. در نهایت در بخش ۵، نتیجه‌گیری کلّی بیان می‌شود.

۳- مبانی و پیش‌زمینه و روش پیشنهادی

الف- متاستاز مغزی

متاستاز مغزی[۱] یا سرطان مغز متاستاتیک[۲] هنگامی اتفاق می‌افتد که سلول‌های سرطانی از محل اصلی خود به مغز پخش شوند [۲۹]. این سلول‌ها ممکن است از طریق جریان خون یا سیستم لنفاوی حرکت کنند و با گسترش آن در مغز، شروع به تکثیر کنند. هر سرطانی می‌تواند پیشرفت کند و در مغز پخش شود اما بعضی از انواع آن‌ها مانند سرطان ریه، سرطان سینه، سرطان کلیه، سرطان پوست و سرطان روده بزرگ، به احتمال زیاد باعث متاستاز مغزی می‌شوند. متاستاز مغزی ممکن است یک یا چند تومور را در مغز ایجاد کند. از طرفی تومورهای متاستاتیک با رشد خود به بافت مغز فشار می‌آورند و عملکرد آن را تغییر می‌دهند. فرد بیمار در نتیجه این اتفاق ممکن است علائمی مانند سردرد، تغییر شخصیت، از دست دادن حافظه و تشنج را تجربه کند.

برای تشخیص و درمان متاستاز مغزی، یکی از کارهایی که انجام می‌گیرد، تهیه تصاویر MRI از تومور موجود در مغز بیمار است. یکی از مواردی که از این تصاویر باید مشخص شود، که به تشخیص و درمان بیماری کمک می‌کند، محل اولیه تومور است. برای این منظور، یک سری خصوصیات از تصاویر MRI استخراج می‌شود. با بررسی این خصوصیات، محل اولیه تومور پیش‌بینی می‌شود.

در این مقاله، روی خصوصیات استخراج شده از تعداد زیادی تصاویر متاستاز مغزی، به‌وسیله الگوریتم جنگل تصادفی، عمل طبقه‌بندی صورت می‌گیرد. در این تصاویر، محل‌های اولیه تومورهای مغزی، چهار ناحیه بوده است که در شبیه‌سازی‌ها با برچسب ۰ تا ۳ شماره‌گذاری شده‌اند. سپس الگوریتم جنگل تصادفی برای طبقه‌بندی تصاویر اجرا شده است تا دقت طبقه‌بندی الگوریتم آزمایش شود. در ادامه برای بهبود طبقه‌بندی الگوریتم جنگل تصادفی، الگوریتم RSA برای انتخاب و کاهش ویژگی‌های تصاویر به کار گرفته شده است تا طبقه‌بندی بهبود یابد.

ب- الگوریتم جستجوی خزندگان

RSA یک الگوریتم بهینه‌سازی جدید است که توسط Abualigah و همکاران توسعه یافته است [۳۰]، که شبیه رفتارهای احاطه و شکار کروکودیل است. کروکودیل‌ها، خزندگان نیمه آبزی با ویژگی‌های فیزیکی منحصر به فرد مانند شکل بدن خط دار، توانایی بالا بردن پاهای خود به پهلو هنگام راه رفتن، راه رفتن شکمی و شنا هستند. این ویژگی‌ها به آنها اجازه می‌دهد تا به شکارچیان قدرتمند در طبیعت تبدیل شوند. این بخش قابلیت‌های اکتشاف و بهره‌برداری RSA را شرح می‌دهد که بر اساس محاصره هوشمند و شکار طعمه است. علاوه بر این، توابع ریاضی و شبه کد الگوریتم پوشش داده شده است. RSA یک روش مبتنی بر جمعیت و بدون گرادیان است که می‌تواند مسائل بهینه‌سازی پیچیده و ساده را با توجّه به محدودیت‌های خاص حل کند.

۱) مرحله اولیه

در این مرحله، راه‌حل‌های کاندید اولیه بر اساس نقشه‌های آشفته مانند معادله (۱) تولید می‌شوند. همچنین، فضای جستجو و تابع هدف نیز تعریف شده است. و تمام مقادیر پارامتر قبل از محاسبه تنظیم می‌شوند.

$$X=\begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & x_{1,n-1} & x_{1,n} \\ x_{2,1} & \cdots & x_{2,j} & x_{2,n-1} & x_{2,n} \\ \cdots & \cdots & x_{i,j} & x_{i,n-1} & x_{i,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-1,1} & \cdots & x_{N-1,j} & \cdots & x_{N-1,n} \\ x_{N,1} & \cdots & x_{N,j} & x_{N-1,n} & x_{N,n} \end{bmatrix} \quad (1)$$

که در آن X نشان دهنده راه حل‌های کاندید تولید شده با استفاده از معادله (۲) است، و $x_{i,j}$ موقعیت jامین عامل جستجوی راه حل iام را نشان می‌دهد، و N تعداد راه حل‌های بالقوه است. n نیز نشان‌دهنده اندازه مسئله است.

$$x_{ij}=rand\,(UB-LB)+LB, \quad j=1,2,...,n \quad (2)$$

در معادله (۲) در بالا، rand یک مقدار شروع است. همچنین، UB و LB تعریف شده‌اند که به ترتیب مرزهای بالایی و پایینی مسئله داده شده را مشخص می‌کنند.

۲) فاز اکتشاف (احاطه)

در این قسمت، رفتار اکتشافی (احاطه) RSA مورد بحث قرار می‌گیرد. کروکودیل‌ها دو استراتژی در طول فرایند دور زدن خود انجام می‌دهند: راه رفتن بلند و راه رفتن با شکم. این حرکات به رویکردهای مختلفی اشاره دارد که متعهد به نمایش قابلیت‌های اکتشافی الگوریتم

---

۱. Brain Metastasis .
۲. Metastatic brain cancer.

(جستجوی جهانی) هستند. حرکات تمساح (راه رفتن با ارتفاع بالا و راه رفتن شکم) به دلیل سر و صدای آن‌ها از گرفتن طعمه جلوگیری می‌کند، مگر اینکه مکانیزم جستجوی دیگری را به کار گیرند (مرحله اکتشاف). از این رو، جستجوی اکتشافی فضای جستجوی وسیعی را کشف می‌کند. شاید بعد از چندین جستجو بتواند منطقه امیدوارکننده را پیدا کند.

جستجوی متوازن اکتشاف (محاصره) و بهره‌برداری (شکار) در RSA طبق چهار شرط، تعداد کل تکرارها را به چهار قسمت تقسیم می‌کند. مکانیسم‌های اکتشاف در RSA بر دو استراتژی جستجوی اصلی (پیاده‌روی بالا و راه رفتن شکمی) تمرکز می‌کنند تا فضای جستجو را کشف کنند و راه‌حل بهتری پیدا کنند. استراتژی راه رفتن بالا توسط $t < \frac{T}{۴}$ تعریف می‌شود و استراتژی حرکت راه رفتن شکمی توسط $t \leq \frac{۳T}{۴}$ و $t > \frac{T}{۴}$ تعریف می‌شود. این به این معنی است که شرط ذکر شده، برای تقریباً نیمی از تعداد تکرارها، کاوش (راهپیمایی بالا) و برای نیمی دیگر (راهپیمایی شکمی) را اجرا می‌کند. فرمول به‌روزرسانی موقعیت برای مرحله اکتشاف همان‌طور که در معادله (۳) نشان داده شده، ارائه شده است.

$$x_{(i,j)}(t+1) = \begin{cases} Best_j(t) \times -\eta_{(i,j)}(t) \\ \times \beta - R_{(i,j)}(t) \times rand, & t \leq \frac{T}{4} \\ Best_j(t) \times x_{r1,j} \times ES(t) \times rand, & t \leq 2\frac{T}{4} \; and \; t > \frac{T}{4} \end{cases} \quad (3)$$

که در آن، $Best_j(t)$ jامین موقعیت را در بهترین راه حل به دست آمده تاکنون نشان می‌دهد، rand به یک عدد صحیح بین ۰ و ۱ اشاره دارد، t عدد تکرار فعلی است، و T مخفف حداکثر تعداد تکرار است. $\eta_{(i,j)}$ نشان‌دهنده عملگر اکتشافی است که موقعیت j را در راه حل iام، محاسبه شده توسط معادله (۴)، شناسایی می‌کند. β یک پارامتر حیاتی است که از الگوریتم اصلی RSA گرفته شده و روی مقدار ۰,۱ تنظیم شده است و دقت اکتشاف را برای احاطه کردن (یعنی راه رفتن زیاد) در طول تکرارها، هدایت می‌کند. $R_{(i,j)}$ مقداری است که برای کاهش ناحیه جستجو اعمال می‌شود که با معادله (۵) محاسبه می‌شود. $r_۱$ یک عدد تصادفی بین [۱, N] است و $x_{r۱,j}$ به موقعیت تصادفی راه حل iام اشاره دارد. حس تکاملی ES(t) یک نسبت تصادفی است بین [-۲, ۲] که احتمال کاهش مقادیر در طول تکرارها را توصیف می‌کند، که با معادله (۶) محاسبه می‌شود.

$$\eta_{(i,j)} = Best_j(t) \times P_{(i,j)} \quad (4)$$

$$R_{(i,j)} = \frac{Best_j(t) - x_{(r2,j)}}{Best_j(t) + \varepsilon} \quad (5)$$

$$ES(t) = 2 \times r3 \times \left(1 - \frac{1}{T}\right), \quad (6)$$

که در آن $r_۲$ یک عدد تصادفی بین [۱, N] و ε یک مقدار کوچک است. در معادله (۶)، ۲ مقدار همبستگی است که برای دادن مقادیر بین ۲ و ۰، $r_۳$ استفاده می‌شود که به یک عدد صحیح تصادفی بین [-۱, ۱] دلالت دارد. $P_{(i,j)}$ مربوط به تفاوت بین موقعیت jام بهترین راه حل به دست آمده و موقعیت jام راه حل فعلی است، که با معادله (۷) محاسبه می‌شود.

$$P_{(i,j)} = \alpha + \frac{x_{(i,j)} - M(x_i)}{Best_j(t) \times (UB_{(j)} - LB_{(j)}) + \varepsilon} \quad (7)$$

که در آن $M(x_i)$ برابر موقعیت‌های میانگین راه حل iام است که با معادله (۸) محاسبه می‌شود. $UB_{(j)}$ و $LB_{(j)}$ به ترتیب مرزهای موقعیت jام هستند. α یک پارامتر حیاتی است، که دقت اکتشاف را برای همکاری شکار در طول تکرارها راهنمایی می‌کند، که در این کار مقدار ۰,۱ تنظیم شده است.

$$M(x_i) = \frac{1}{n} \sum_{j=1}^{n} x_{(i,j)} \quad (8)$$

۳) مرحله بهره‌برداری (شکار)

در این قسمت، رفتار بهره‌برداری (شکار) RSA معرفی می‌شود. کروکودیل‌ها، دو استراتژی در طول فرایند شکار خود انجام می‌دهند: همکاری و هماهنگی. این استراتژی‌ها جستجوی بهره‌برداری (جستجوی محلی) را شبیه‌سازی می‌کنند که در معادله (۹) فرمول‌بندی شده است. استراتژی هماهنگی شکار در این مرحله توسط $t \leq \frac{۳T}{۴}$ و $t > \frac{۲T}{۴}$ مشروط شده است، در غیر این صورت استراتژی همکاری شکار اجرا می‌شود. در RSA اصلی، فرمول به‌روزرسانی موقعیت برای بهره‌برداری در معادله (۹) ارائه شده است.

$$x_{(i,j)}(t+1) = \begin{cases} Best_j(t) \times P_{(i,j)}(t) \times rand, & t \leq 3\frac{T}{4} \; and \; t > 2\frac{T}{4} \\ Best_j(t) - \eta_{(i,j)}(t) \times \varepsilon \\ -R_{(i,j)}(t) \times rand, & t \leq T \; and \; t > 3\frac{T}{4} \end{cases} \quad (9)$$

که در آن $Best_j(t)$ موقعیت jام در بهترین راه حل پیدا شده تاکنون است، $\eta_{(i,j)}$ به پارامتر شکار برای jامین موقعیت در راه حل iام اشاره دارد که با معادله (۳) محاسبه شده است. $P_{(i,j)}$ تفاوت بین موقعیت jام

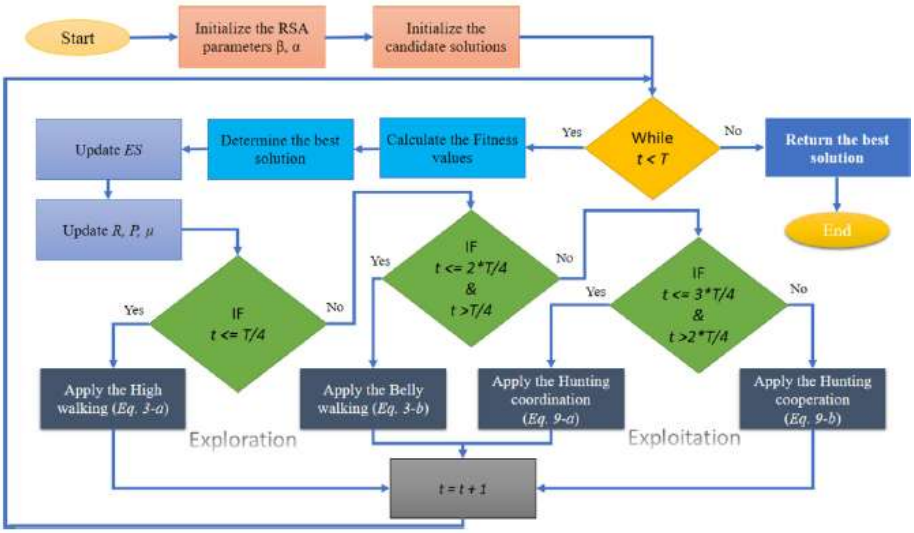
بهترین راه حل پیدا شده و موقعیت jام راه حل فعلی است که با معادله (۶) محاسبه می‌شود. $\eta_{(i,j)}$ دلالت بر پارامتر شکار برای موقعیت jام در راه حل iام دارد که با استفاده از معادله (۳) محاسبه می‌شود. $R_{(i,j)}(t)$ مقداری است که برای کاهش ناحیه جستجو در تکرار جاری اعمال می‌شود که با معادله (۴) محاسبه می‌شود. در شکل ۱، شبه کد الگوریتم RSA ارائه شده است.

Algorithm 1 Pseudo-code of the Reptile Search Algorithm (RSA)

Initialization phase
Initialize RSA parameters α, β, etc.
Initialize the solutions' positions randomly. X: $i = 1, \ldots, N$.
while $(t < T)$ **do**
  Calculate the Fitness Function for the candidate solutions (X).
  Find the Best solution so far.
  Update the ES using Equations (6).
  The beginning of the RSA
  for (i=1 to N) **do**
    for (j=1 to n) **do**
      Update the η, R, P and values using Equations (4), (5) and (7), respectively.
      if $(t \leq \frac{T}{4})$ **then**
        $x_{(i,j)}(t+1) = Best_j(t) \times - \eta_{(i,j)}(t) \times \beta - R_{(i,j)}(t) \times rand$, ▷ {**High walking**}
      else if $(t \leq 2\frac{T}{4}$ and $t > \frac{T}{4})$ **then**
        $x_{(i,j)}(t+1) = Best_j(t) \times x_{(r_1,j)} \times ES(t) \times rand$, ▷ {**Belly walking**}
      else if $(t \leq 3\frac{T}{4}$ and $t > 2\frac{T}{4})$ **then**
        $x_{(i,j)}(t+1) = Best_j(t) \times P_{(i,j)}(t) \times rand$, ▷ {**Hunting coordination**}
      else
        $x_{(i,j)}(t+1) = Best_j(t) - \eta_{(i,j)}(t) \times \epsilon - R_{(i,j)}(t) \times rand$, ▷ {**Hunting cooperation**}
      end if
    end for
  end for
  t=t+1
end while
Return the best solution ($Best(X)$).

شکل ۱: الگوریتم جستجوی خزندگان [۳۰]

همچنین در شکل ۲، الگوریتم RSA به صورت فلوچارت، ارائه شده است.



شکل ۲: فلوچارت الگوریتم جستجوی خزندگان [۳۰]

ج

- جنگل تصادفی

الگوریتم جنگل تصادفی[۱] یک الگوریتم یادگیری ماشین نظارت‌شده است که از الگوریتم‌های درخت تصمیم‌گیری ساخته شده است. ازآنجایی‌که این الگوریتم از تعداد زیادی درخت تصمیم‌گیری ساخته شده است، نام الگوریتم جنگل تصادفی را برای آن در نظر گرفته‌اند. اولین بار، الگوریتم جنگل‌های تصمیم تصادفی با استفاده از روش زیرفضاهای تصادفی به‌وجود آمد [۳۱]. بعدها نسخه‌های دیگری از آن ارتقا پیدا کردند [۳۲].

روش جنگل تصادفی، یک تکنیک یادگیری ماشینی است که برای حل مسائل مربوط به رگرسیون و طبقه‌بندی مورد استفاده قرار می‌گیرد. مبنای روش جنگل تصادفی، یادگیری گروهی است، تکنیکی که بسیاری از طبقه‌بندی‌کننده‌ها را برای ارائه راه‌حل‌هایی برای حل مسائل پیچیده ترکیب می‌کند. یک الگوریتم جنگل تصادفی از درخت تصمیم‌های زیادی تشکیل شده است. "جنگل" تولید شده توسط الگوریتم جنگل تصادفی، از طریق تجمیع bagging یا bootstrap آموزش داده می‌شود. Bagging یک متا الگوریتم مجموعه‌ای است که دقت الگوریتم‌های یادگیری ماشین را بهبود می‌بخشد.

الگوریتم جنگل تصادفی، نتیجه را بر اساس پیش‌بینی درختان تصمیم[۲] با گرفتن میانگین یا میانگین خروجی از درختان مختلف، پیش‌بینی می‌کند. افزایش تعداد درختان، دقت نتیجه را هم افزایش می‌دهد. یک جنگل تصادفی، محدودیت‌های یک الگوریتم درخت تصمیم را از بین برده و برازش بیش از حد مجموعه داده‌ها را کاهش می‌دهد. همچنین، باعث افزایش دقت می‌شود و در نهایت یک پیش‌بینی تولید می‌کند.

آنتروپی[۳] و به‌دست‌آوردن اطلاعات، بلوک‌های سازنده درختان تصمیم هستند. مروری بر این مفاهیم اساسی، درک ما را از نحوه ساخت درخت تصمیم بهبود می‌بخشد. بلوک آنتروپی، معیاری برای محاسبه عدم قطعیت و بلوک به‌دست‌آوردن اطلاعات معیاری است که نشان می‌دهد چگونه عدم قطعیت در متغیر هدف با توجه به مجموعه‌ای از متغیرهای مستقل کاهش می‌یابد.

مفهوم به‌دست‌آوردن اطلاعات، شامل استفاده از متغیرهای مستقل (ویژگی‌ها) برای به‌دست‌آوردن اطلاعات درمورد یک متغیر هدف (کلاس) است. از آنتروپی متغیر هدف و آنتروپی شرطی داده شده، برای تخمین سود اطلاعات استفاده می‌شود. در این حالت، آنتروپی شرطی از آنتروپی هدف کم می‌شود. به‌دست‌آوردن اطلاعات در آموزش درختان تصمیم استفاده می‌شود که به کاهش عدم اطمینان در این درختان کمک می‌کند. افزایش اطلاعات به این معنی است که درجه بالایی از عدم قطعیت (آنتروپی اطلاعات) حذف شده است. آنتروپی و کسب اطلاعات در شکافتن شاخه‌ها مهم هستند که فعالیت مهمی در ساخت درخت‌های تصمیم‌گیری است.

در یک جنگل تصادفی، برای یک داده خاص، یک مجموعه درخت تصمیم ایجاد می‌شود که ریشه هر درخت یک خصوصیت را نشان می‌دهد. بر اساس هر خصوصیت، یک تصمیم‌گیری در هر درخت تصمیم ایجاد می‌شود و تصمیم‌گیری نهایی در یک گره برگ صورت می‌گیرد. در نهایت با استفاده از سیستم رأی‌گیری اکثریت، تصمیم نهایی اتخاذ می‌شود.

رگرسیون، وظیفه دیگری است که توسط الگوریتم جنگل تصادفی انجام می‌شود. یک رگرسیون جنگل تصادفی از مفهوم رگرسیون ساده پیروی می‌کند. مقادیر متغیرهای وابسته (ویژگی‌ها) و مستقل در مدل جنگل تصادفی منتقل می‌شوند. میانگین پیش‌بینی هر درخت، خروجی رگرسیون است. این برخلاف طبقه‌بندی تصادفی جنگل است که خروجی آن توسط حالت کلاس درختان تصمیم تعیین می‌شود. اگرچه رگرسیون جنگل تصادفی و رگرسیون خطی از یک مفهوم پیروی می‌کنند، اما از نظر توابع متفاوت هستند. تابع رگرسیون خطی $y=bx+c$ است، که در آن y متغیر وابسته، x متغیر مستقل، b پارامتر تخمین و c یک ثابت است. عملکرد یک رگرسیون تصادفی پیچیده جنگل مانند یک جعبه سیاه است.

د- روش پیشنهادی

در این مطالعه، سعی شده است تا داده‌های تصاویر MRI تومورهای مغزی متاستاز مغزی طبقه‌بندی شوند تا منشأ تومور مشخص شود و بر اساس پیش‌بینی، در یکی از دسته‌ها طبقه‌بندی شوند. در اینجا، ویژگی‌های تصاویر استخراج شده‌اند و به جای کار با تصاویر، با ویژگی‌های تصاویر کار شده است. در اینجا از الگوریتم جنگل تصادفی به عنوان الگوریتم پایه برای طبقه‌بندی استفاده شده است. سپس از الگوریتم RSA برای انتخاب زیرمجموعه‌ای از ویژگی‌های مهم استفاده می‌شود تا دقت طبقه‌بندی افزایش یابد. در شکل ۳، شمای کلّی، به کارگیری الگوریتم RSA برای انتخاب خصوصیات مهم برای طبقه‌بندی نمایش داده شده است.

۴- نتایج شبیه‌سازی

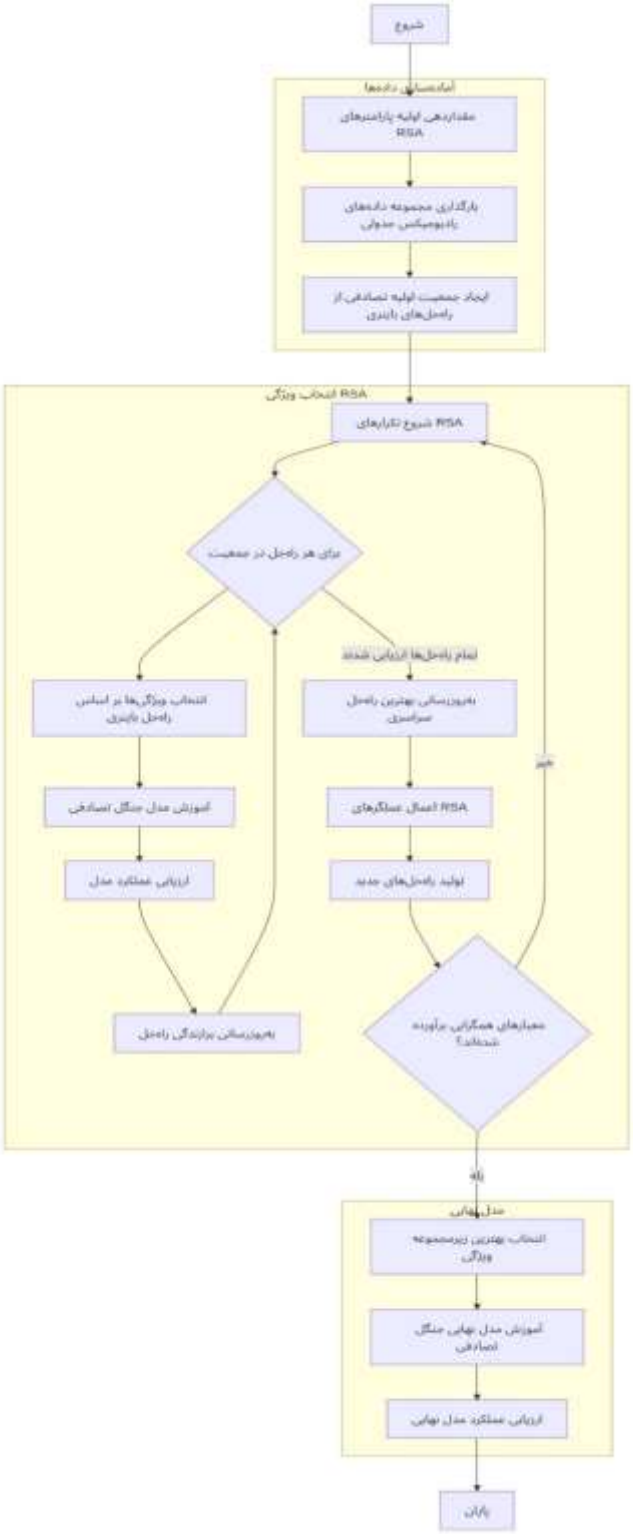
در این بخش، نتایج آزمایش‌ها مورد بحث قرار می‌گیرد. برای ارزیابی و مقایسهٔ اعمال الگوریتم RSA روی انتخاب ویژگی‌های خاص و تأثیر آن در دقت طبقه‌بندی، دو شبیه‌سازی انجام شده است. در شبیه‌سازی اول، الگوریتم جنگل تصادفی، به تنهایی روی داده‌ها اعمال شده است و در شبیه‌سازی دوم، همراه با الگوریتم جنگل تصادفی، الگوریتم RSA نیز به کار گرفته شده است. شبیه‌سازی‌ها در این مقاله با استفاده از زبان برنامه‌نویسی پایتون انجام شده‌اند. در این شبیه‌سازی پس از تعداد

۱. Random Forest.
۲. Decision Tree.
۳. Entropy.

۱۰۰ تکرار، نتایج به دست آمده‌اند. نتایج این شبیه‌سازی در ادامه آمده است.

الف- جزئیات مجموعه داده‌ها

مجموعه دادهٔ استفاده شده در مقاله از در تاریخ ۲۰۲۳/۰۴/۱۱ در پلت‌فرم figshare پست شده است. این مجموعه داده با عنوان A Comprehensive dataset of annotated brain metastasis images with clinical and radiomic data در این پلت‌فرم ارسال شده است [۳۳]. دو فایل اصلی در این مجموعه قرار دارد که خصوصیات رادیومیک بیماران در آن‌ها قرار دارد. تعداد این خصوصیات برای هر تصویر ۱۱۳۰ عدد است. در این مجموعه داده، تعداد ۱۲۷۶ رکورد وجود دارد که هر رکورد آن اطلاعات یک بیمار است. در این مجموعه داده برای برخی بیماران چند رکورد وجود دارد، بنابراین تعداد بیماران از تعداد رکوردها کمتر است.

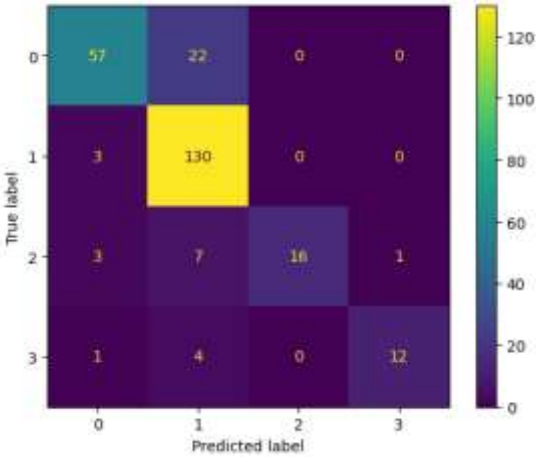


شکل ۳: شمای کلّی به کارگیری الگوریتم RSA در مقاله.

ب- نتایج و بحث

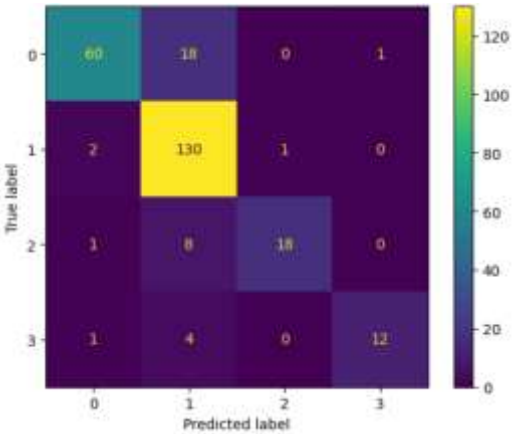
در ابتدا لازم به ذکر است که پس از شبیه‌سازی با ۱۰۰ تکرار در الگوریتم RSA، تعداد خصوصیات مؤثر در طبقه‌بندی، از ۱۱۳۰ عدد در ابتدا، به ۳۵۱ عدد کاهش یافته است. یعنی تعداد ۷۷۹ خصوصیت حذف شده‌اند که کاهش قابل توجهی را نشان می‌دهد.

در مدل‌سازی اول، یعنی بدون استفاده از RSA، ماتریس سردرگمی[۱] شکل ۴ را داریم. در این ماتریس، قطری که از بالا سمت چپ به سمت پایین سمت راست کشیده شده، تعداد پیش‌بینی‌های درست را نشان می‌دهد و سایر خانه‌ها، تعداد پیش‌بینی‌های اشتباه را نشان می‌دهند.



شکل ۴: ماتریس سردرگمی در مدلسازی بدون RSA

در شکل ۵ در ماتریس سردرگمی مربوط به شبیه‌سازی با RSA، می‌بینیم که با افزودن RSA، بهبودهایی در پیش‌بینی حاصل شده است. مثلاً در قطر اصلی ماتریس، تعداد پیش‌بینی درست برای برچسب ۱، از ۵۷ در مدل‌سازی بدون RSA به ۶۰ در مدل‌سازی با RSA رسیده است. همچنین تعداد پیش‌بینی درست برای برچسب ۳، از ۱۶ در مدل‌سازی بدون RSA به ۱۸ در مدل‌سازی با RSA افزایش یافته است.



شکل۵: ماتریس سردرگمی در مدل‌سازی با RSA

در جدول ۱، مقادیر precision، recall، و f۱-score برای هر چهار کلاس طبقه‌بندی، و نیز برای پارامترهای accuracy، macro avg، و weighted avg برای شبیه‌سازی بدون RSA ذکر شده‌اند. در جدول در جدول ۲ نیز، مقادیر این پارامترها، برای شبیه‌سازی همراه با RSA نشان داده شده‌اند.

جدول ۱: مقادیر precision، recall، و f۱-score برای هر چهار کلاس طبقه‌بندی، و نیز برای پارامترهای accuracy، macro avg، و weighted avg برای شبیه‌سازی بدون RSA.

	precision	recall	F۱-score	support
۰	۰/۸۹	۰/۷۲	۰/۸۰	۷۹
۱	۰/۸۰	۰/۹۸	۰/۸۸	۱۳۳
۲	۰/۹۵	۰/۵۹	۰/۷٤	۲۷
۳	۰/۹۲	۰/۷۱	۰/۸۰	۱۷
accuracy	-	-	۰/۸٤	۲۵٦
macro avg	۰/۹۰	۰/۷۵	۰/۸۰	۲۵٦
weighted avg	۰/۸٦	۰/۸٤	۰/۸۳	۲۵٦

جدول ۲: مقادیر precision، recall، و f۱-score برای هر چهار کلاس طبقه‌بندی، و نیز برای پارامترهای accuracy، macro avg، و weighted avg برای شبیه‌سازی همراه با RSA.

	precision	recall	F۱-score	support
۰	۰/۹٤	۰/۷٦	۰/۸٤	۷۹
۱	۰/۸۱	۰/۹۸	۰/۸۹	۱۳۳
۲	۰/۹۵	۰/٦۷	۰/۷۸	۲۷
۳	۰/۹۲	۰/۷۱	۰/۸۰	۱۷
accuracy	-	-	۰/۸٦	۲۵٦
macro avg	۰/۹۱	۰/۷۸	۰/۸۳	۲۵٦
weighted avg	۰/۸۷	۰/۸٦	۰/۸٦	۲۵٦

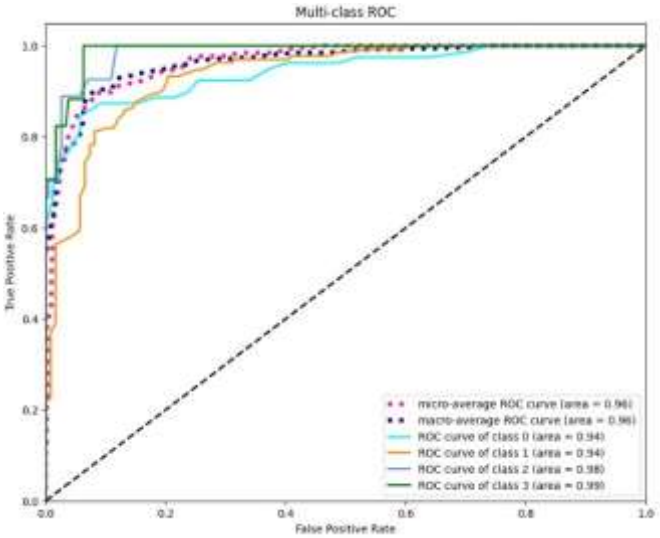
[۱]. Confusion Matrix.

با مقایسه مقادیر دو جدول فوق مشاهده می‌شود که در چهار مورد، مقادیر دو جدول یکسان بوده است و در تمامی موارد دیگر، در مدل‌سازی با کاهش خصوصیات به‌وسیله RSA، مقادیر رشد داشته‌اند. این موضوع نشان می‌دهد که ما هم خصوصیات مورد استفاده در طبقه‌بندی را کاهش چشمگیری داده‌ایم و هم به دقت‌های بالاتر در طبقه‌بندی دست یافته‌ایم. در جدول ۳ نیز مقادیر اعتبارسنجی متقاطع K-Fold را برای دو شبیه‌سازی می‌بینیم. که در شبیه‌سازی با RSA مقداری بهبود مشاهده می‌شود.

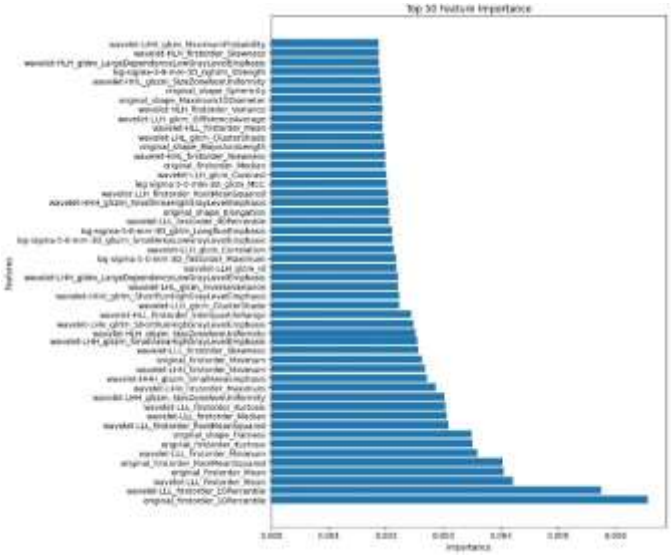
جدول ۳: مقادیر اعتبارسنجی متقاطع K-Fold

شبیه‌سازی بدون RSA	شبیه‌سازی با RSA
۰/۸۴۷۲	۰/۸۴۹۵

در شکل ۶ که منحنی ROC را نشان می‌دهد، مشاهده می‌شود که پس از اعمال RSA برای طبقه‌بندی، برای دسته‌های طبقه‌بندی، اعداد از ۰/۹۴ به بالا هستند که مقادیر کاملا قابل قبولی هستند. در شکل ۷، میزان اهمیت ۵۰ خصوصیت از ۳۵۱ خصوصیت انتخاب شده، نشان داده شده است.



شکل ۶: منحنی ROC



شکل ۷: میزان اهمیت ۵۰ خصوصیت از ۳۵۱ خصوصیت انتخاب شده.

۵- بحث و نتیجه‌گیری

در این مطالعه، با هدف بهبود دقت تشخیص متاستاز مغزی با بهره‌گیری از الگوریتم‌های یادگیری ماشین، داده‌های بیماران تحلیل و

طبقه‌بندی چندکلاسه انجام شد. نتایج به‌دست‌آمده از مدل‌های پیاده‌سازی‌شده نشان داد که عملکرد مدل‌ها در شناسایی کلاس‌های مختلف متاستاز مغزی قابل توجه بوده است. نمودار ROC چندکلاسه، که یکی از مهم‌ترین ابزارهای ارزیابی عملکرد مدل‌های طبقه‌بندی محسوب می‌شود، در این پژوهش مورد استفاده قرار گرفت تا توان تفکیک مدل بین کلاس‌ها را نشان دهد. همان‌طور که در نمودار مشاهده می‌شود، منحنی‌های ROC برای هر کلاس جداگانه رسم شده‌اند و مساحت زیر منحنی (AUC) برای کلاس‌های ۰، ۱، ۲، و ۳ به ترتیب ۰/۹۴، ۰/۹۴، ۰/۹۸ و ۰/۹۹ است که بیانگر دقت بسیار بالای مدل در تمایز بین این کلاس‌ها است. علاوه بر آن، میانگین‌های کلان (macro)و خرد (micro) هر دو AUC برابر با ۰/۹۶ را نشان می‌دهند که نشان از تعادل مناسب بین دقت در کل کلاس‌ها و تعداد نمونه‌ها دارد. این نتایج نشان می‌دهند که مدل پیشنهادی، نه تنها در تشخیص کلاس‌هایی با فراوانی بالا عملکرد خوبی دارد، بلکه در تشخیص کلاس‌های کم‌فراوانی نیز دقت مناسبی را ارائه می‌دهد. چنین سطحی از عملکرد، حاکی از آن است که ویژگی‌های استخراج‌شده از تصاویر یا داده‌های بالینی به خوبی توانسته‌اند نماینده‌ای مناسب از الگوهای پنهان در داده‌ها باشند. با توجه به اهمیت تشخیص زودهنگام و دقیق متاستاز مغزی در تصمیم‌گیری‌های درمانی، استفاده از این روش می‌تواند در بهبود کیفیت مراقبت و افزایش شانس بقاء بیماران مؤثر واقع شود. در نهایت، پیشنهاد می‌شود در مطالعات آینده از داده‌های بیشتر و متنوع‌تر و همچنین تکنیک‌های پیشرفته‌تر یادگیری عمیق استفاده شود تا مدل‌های دقیق‌تر و تعمیم‌پذیرتری توسعه یابند.

منابع

[۱] Siegel, L. Rebecca, D. Kimberly, Miller, Hannah, E. Fuchs and J. Ahmedin, "Cancer statistics," CA: a cancer journal for clinicians ۷۲, vol. ۱, ۲۰۲۲.

[۲] B. Remeseiro and V. B. Canedo, "A review of feature selection methods in medical applications," Computers in biology and medicine, vol. ۱۱۲, pp. ۱۰۳-۳۷۵, ۲۰۱۹.

[۳] P. Shakeel, A. T. Mohamed, Z. Al-Makhadmeh, and M. Musa Jaber, "RETRACTED ARTICLE: Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks," Neural Computing and Applications, vol. ۳۲, no. ۳, pp. ۷۷۷-۷۹۰, ۲۰۲۰.

[٤] T. A. Khan, S. Ho Ling, N. Tram and A. Mohan Sanagavarapu, "A modified particle swarm optimization algorithm for solving DNA problem," In ۲۰۱۹ ٦۰th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), pp. ۱-۵. IEEE, ۲۰۱۹.

[۵] E. Pashaei and E. Pashaei. "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification," Neural Computing and Applications, vol. ۳٤, no. ۸, pp. ٦٤۲۷-٦٤۵۱, ۲۰۲۲.

[٦] P. Jaegyun, P. Min-Woo, D. Kim and J. Lee, "Multi-population genetic algorithm for multilabel feature selection based on label complementary communication," Entropy ۲۲, no. ۸, ۲۰۲۰.

[۷] M. Alweshah, S. Al Khalaileh, B. B. Gupta, A. Almomani, A. I. Hammouri, and M. Azmi Al-Betar. "The monarch butterfly optimization algorithm for solving feature selection problems," Neural Computing and Applications, pp. ۱-۱۵, ۲۰۲۲.

[۸] W. Muhammad Hammad, M. Sajjad, A. Nadeem, A. Abbas, A. Shaheen, W. Aziz, A.Anjum, U. Manzoor, M. A. Balubaid and Seong-O. Shim. "On the feature selection methods and reject option classifiers for robust cancer prediction," IEEE Access ۷, PP. ۱٤۱۰۷۲-۱٤۱۰۸۲, ۲۰۱۹.

[۹] Sh. S. Salameh, H. Turabieh, S. Al Azwari and F. Baothman. "Enhanced binary genetic algorithm as a feature selection to predict student performance," Soft Computing ۲٦, no. ٤, pp. ۱۸۱۱-۱۸۲۳, ۲۰۲۲.

[۱۰] H. Gang, B. Du, X. Wang and G. Wei, "An enhanced black widow optimization algorithm for feature selection," Knowledge-Based Systems ۲۳۵, pp. ۱۰۷-٦۳۸, ۲۰۲۲.

[۱۱] S. Mahalakshmi, and T. Velmurugan, "Detection of brain tumor by particle swarm optimization using image segmentation," Indian Journal of Science and Technology ۸, pp. ۱-۲۲, ۲۰۱۵.

[۱۲] E. Ahmed A, M. Al-Qaness, L. Abualigah, D. Oliva, Z. Yahya Algamal, A. M. Anter, R. Ali Ibrahim, R. M. Ghoniem and M. Abd Elaziz, "Boosting arithmetic optimization algorithm with genetic algorithm operators for feature selection: Case study on cox proportional hazards model," Mathematics ۹, no, pp. ۲۱-۲۳, ۲۰۲۱.

[۱۳] A. Ahmed M and M. Ali, "Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems," Soft Computing ۲٤, no. ۳, pp. ۱۵٦۵-۱۵۸۴, ۲۰۲۰.

[۱٤] P.M., L. Agilandeeswari and K. Ganesan, "An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier," Journal of Ambient Intelligence and Humanized Computing ۱۰, pp. ۲٦۷-۲۹۳, ۲۰۱۹.

[۱۵] M. Tubishat, A. Mohammad AM, I. Norisma and A. Ibrahim, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," Applied Intelligence ٤۹, pp. ۱٦۸۸-۱۷۰۷, ۲۰۱۹.

[۱٦] H. Shilan S, O. O. Petinrina, A. O. Hashi and F. Saeed, "Filter-Wrapper Combination and Embedded Feature Selection for Gene Expression Data," International Journal of Advances in Soft Computing & Its Applications ۱۰, no. ۱, ۲۰۱۸.

[۱۷] J. Farrukh, I. Thomas and M. Memedi, "A comparison of feature selection methods when using motion sensors data: a case study in Parkinson's disease," In ۲۰۱۸ ٤۰th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. ۵٤۲٦-۵٤۲۹, ۲۰۱۸.

[۱۸] S. S, D. T. Newton, T. H. Everett, G. Lin and B. S. Duerstock, "Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia. Front Neuroinform," vol. ۱٦, ۲۰۲۲.

[۱۹] J. Alan, K. Brkić and N. Bogunović. “A review of feature selection methods with applications,” In ۲۰۱۵ ۳۸th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. ۱۲۰۰-۱۲۰۵, ۲۰۱۵.

[۲۰] U. Ali Muhammad, U. K. Yusof and M. Sabudin, “Filter-based feature selection using information theory and binary cuckoo optimisation algorithm,” Journal of Information Technology Management ۱٤, no. Special Issue: ٥th International Conference of Reliable Information and Communication Technology (IRICT ۲۰۲۰), pp. ۲۰۳-۲۲۲, ۲۰۲۲.

[۲۱] W. Jie, J. Xu, Ch. Zhao, Y. Peng and H. Wang, “An ensemble feature selection method for high-dimensional data based on sort aggregation,” Systems Science & Control Engineering ۷, no. ۲, pp. ۳۲-۳۹, ۲۰۱۹.

[۲۲] E. Zenab Mohamed, N. Mohd Yasin, A. Qalid Md Sabri, R. Sihwail, M. Tubishat and H. Jarrah, “Improved equilibrium optimization algorithm using elite opposition-based learning and new local search strategy for feature selection in medical datasets,” Computation ۹, no. ٦, vol. ٦۸, ۲۰۲۱.

[۲۳] M. Tubishat, I. Norisma, L. Shuib, M. AM Abushariah, and S. Mirjalili, “Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection,” Expert Systems with Applications ۱٤٥, pp. ۱۱۳-۱۲۲, ۲۰۲۰.

[۲٤] T. Jingwei and S. Mirjalili, “General learning equilibrium optimizer: a new feature selection method for biological data classification,” Applied Artificial Intelligence ۳٥, no. ۳, pp. ۲۴۷-۲۶۳, ۲۰۲۱.

[۲٥] L. Jiahui, H. Kang, G. Sun, T. Feng, W. Li, W. Zhang and B. Ji, “IBDA: improved binary dragonfly algorithm with evolutionary population dynamics and adaptive crossover for feature selection,” IEEE Access ۸, pp. ۱۰۸۰۳۲-۱۰۸۰۵۱, ۲۰۲۰.

[۲٦] T. Anurag and A. Chaturvedi, “A hybrid feature selection approach based on information theory and dynamic butterfly optimization algorithm for data classification,” Expert Systems with Applications ۱۹٦, pp. ۱۱٦-٦۲۱, ۲۰۲۲.

[۲۷] Ch. Ke, F. Y. Zhou and X. Feng Yuan, “Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection,” Expert Systems with Applications ۱۲۸, pp. ۱٤۰-۱٥۶, ۲۰۱۹.

[۲۸] S. Rami, Kh. Omar, Kh. A. Zainol Ariffin and M. Tubishat, “Improved harris hawks optimization using elite opposition-based learning and novel search mechanism for feature selection,” Ieee Access ۸, pp. ۱۲۱۱۲۷-۱۲۱۱٤٥, ۲۰۲۰.

[۲۹] B. Adrienne, P. K. Brastianos, L. Garzia and M. Valiente, “Brain metastasis,” Nature Reviews Cancer ۲۰, no. ۱, pp. ٤-۱۱, ۲۰۲۰.

[۳۰] A. Laith, M. Abd Elaziz, P. Sumari, Z. Woo Geem and A. H. Gandomi. “Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer,” Expert Systems with Applications ۱۹۱, pp. ۱۱٦-۱٥۸, ۲۰۲۲.

[۳۱] H. Tin Kam, “Random decision forests,” In Proceedings of ۳rd international conference on document analysis and recognition, vol. ۱, pp. ۲۷۸-۲۸۲, ۱۹۹٥.

[۳۲] B. Leo, “Random forests,” Machine learning ٤٥, pp. ٥-۳۲, ۲۰۰۱.

[۳۳]A. comprehensive dataset_of_annotated brain metastasis_images_with_clinical_and_radiomic_data/٦۱۹٤۱۰ ۴/۱. (□□□□□://□□□□□□□□□□□□□□□□.□□□□□□□□□.□□□/□□□□□□□□□□□□)

Improving brain metastasis classification using radiomic feature optimization

Sima Haghgozaar,[۱] Mehdi khazaiepoor[*۲]

۱-Department of Computer Engineering- Faculty of Technology and Engineering - Birjand Islamic Azad University - Birjand – Iran.

۲-Department of Computer Engineering- Faculty of Technology and Engineering - Birjand Islamic Azad University - Birjand – Iran.

Abstract

This study aimed to improve the accuracy of brain metastasis classification on a comprehensive dataset of brain MRI images along with clinical and morphological data. The increasing size of medical datasets has created a number of high-dimensional features that negatively impact machine learning classifiers. In machine learning, the feature selection process is fundamental to selecting the most relevant features and reducing redundant and irrelevant ones. Optimization algorithms demonstrate their ability to solve feature selection problems. The Reptile Search Algorithm is a new nature-inspired optimization algorithm that simulates the surrounding and hunting behavior of crocodiles. The unique search of the RSA algorithm yields promising results compared to other optimization algorithms; therefore, this algorithm is used in this paper to search for the set of important features. In this paper, the random forest algorithm is used to classify the data. That is, initially a random forest model was trained on all radiomic features in the dataset. Then, a subset of important features was selected using the metaheuristic optimization algorithm for reptile search. Then, the classification results were compared in two cases. The first case without applying the RSA algorithm, and the second case with applying the RSA algorithm and selecting important features. The results showed that by applying the RSA algorithm, data classification was improved, and this approach can help in faster diagnosis and prediction of brain metastasis. Also, by using this powerful optimization, the speed of model training and its quality can be significantly increased.

Keywords: Brain metastasis, radiomic features, feature selection, random forest, reptile search algorithm.