پیش‌بینی درجات تومور مغزی گلیوما با استفاده از یادگیری ماشین گروهی

حجت امامی*[1]، بابک آذرنوید[2]، محسن عبدالحسین‌زاده[3]
دانشیار دانشکدهٔ فنی و مهندسی، گروه مهندسی کامپیوتر، دانشگاه بناب، بناب، ایران*
استادیار دانشکدهٔ علوم پایه، گروه ریاضی و علوم کامپیوتر، دانشگاه بناب، بناب، ایران
استادیار دانشکدهٔ علوم پایه، گروه ریاضی و علوم کامپیوتر، دانشگاه بناب، بناب، ایران

چکیده

گلیوماها شایع‌ترین تومورهای اولیهٔ دستگاه عصبی‌اند که انواع مختلفی از تومورها با درجات مختلف بدخیمی را شامل می‌شوند. تشخیص و درجه‌بندی دقیق این تومورها به‌دلیل ماهیت تهاجمی و پیش‌روندهٔ برخی از انواع آن، به یک چالش اساسی در حوزهٔ پزشکی تبدیل و این پژوهش بر روی یک مجموعه‌دادهٔ عمومی شامل ۸۳۹ نمونه از پایگاه دادهٔ TCGA (اطلس ژنوم سرطان) انجام شده‌است؛ داده‌ها شامل بیست ویژگی ژنتیکی (وضعیت جهش در ژن‌هایی همانند IDH1، TP53 و ATRX) و سه ویژگی بالینی (نژاد، جنسیت و سن) هستند که هدف، دسته‌بندی تومورها به دو گروه گلیوما درجه پایین و گلیوبلاستوما است؛ تاکنون چندین مدل یادگیری ماشین برای حل مسئلهٔ تشخیص و درجه‌بندی گلیوما ارائه شده‌است که به نتایج قابل قبولی دسته یافته‌اند؛ بااین‌حال، تلاش در این زمینه جهت توسعهٔ مدلی با بیشترین کارایی در دسته‌بندی و درجه‌بندی تومورها نیاز است؛ زیرا کارایی مدل‌های موجود با حالت ایده‌آل فاصلهٔ زیادی دارد. نوآوری این پژوهش، ارائهٔ یک چهارچوب یادگیری ماشین گروهی دولایه (EML) است که برای نخستین‌بار از ترکیب چهار مدل یادگیر پایه شامل یادگیری ماشین بردار پشتیبان (SVM)، تقویت دسته‌بندی (CatBoost)، درخت‌های به‌شدت تصادفی (ERT) و جنگل تصادفی (RF) جهت تنوع‌بخشی و افزایش دقت دسته‌بندی و همچنین یک متامدل شامل رگرسیون لجستیک (LR) استفاده می‌کند؛ همچنین، از یک رویکرد ترکیبی خلاقانه متشکل از الگوریتم‌های Boruta و SHAP برای انتخاب زیرمجموعهٔ بهینهٔ ویژگی‌ها بهره گرفته شده‌است. این رویکرد با هدف کاهش واریانس، جلوگیری از بیش‌برازش و افزایش دقت پیش‌بینی طراحی شده‌است. نتایج آزمایش‌ها روی مجموعه‌دادهٔ آزمون حاکی از آن است که مدل پیشنهادی ELM با کسب مقدار صحت ۸۹.۲۹ درصد روی مجموعه‌دادهٔ آزمون بهترین رتبه را در بین سایر مدل‌های همتا کسب کرده‌است؛ همچنین الگوریتم‌های RF و LR به‌ترتیب با کسب مقدار صحت ۸۷.۲۷ درصد و ۸۶.۸۸ درصد روی داده‌های آزمون در جایگاه دوم و سوم قرار گرفتند.

واژگان کلیدی: تومور مغزی گلیوما، تشخیص گلیوما، داده‌کاوی، یادگیری گروهی، بهینه‌سازی پارامتر.

Predicting glioma brain tumor grades using ensemble machine learning

Hojjat Emami[1]*, Babak Azarnavid[2], Mohsen Abdolhosseinzadeh[3]
Associate Professor, Department of Computer Engineering, Faculty of Engineering, University of Bonab, Bonab, Iran[1]*
Assistant Professor, Department of Mathematics and computer science, Basic Science Faculty, University of Bonab, Bonab, Iran[2]
Assistant Professor, Department of Mathematics and computer science, Basic Science Faculty, University of Bonab, Bonab, Iran[3]

Abstract

Gliomas, or in other words, aggressive and progressive brain tumors, lead to great complexity in the diagnosis and treatment of patients. While recent machine learning models provided encouraging results in glioma diagnosis and grading, the topic is open, and more efforts are needed. Existing models, despite encouraging results, often fall short of the ideal diagnostic state, highlighting the need for further research to develop robust and high-performing predictive models.

This study introduces an optimized ensemble machine learning (EML) model designed to maximize classification (grading) performance and mitigate the pervasive issue of overfitting in glioma grading. Our approach employs a two-layer architecture that synergistically combines diverse weak and base learners. In the first layer, a diverse set of learners, including support vector machine (SVM),

* Corresponding author
*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۴۰۴/۰۵/۲۱ • تاریخ پذیرش: ۱۴۰۴/۰۷/۱۴ • تاریخ انتشار: ۱۴۰۴/۱۲/۲۹ • نوع مطالعه: پژوهشی

categorical boosting (CatBoost), extremely randomized trees (ERT), and random forest (RF), is integrated. This initial ensemble aims to capture a broad spectrum of grading patterns and enhance the overall accuracy by leveraging the complementary strengths of each base model. The outputs from this first layer, representing diversified classification probabilities, are then fed into a second-layer logistic regression (LR) model. This layer refines the predictions, performing the ultimate classification while explicitly addressing and eliminating the overfitting problem, thereby promoting better generalization to unseen data.

To rigorously evaluate the performance of the proposed ELM model, a comprehensive comparison was conducted against its constituent base learners and counterpart machine learning models. All models were assessed using a standard, publicly available glioma dataset. To prevent overfitting, examine the robustness of models, and evaluate models fairly, a 5-fold cross-validation strategy is used in experiments. The effectiveness of models was measured using four performance metrics, including accuracy, recall, precision, and F1-score .

The experimental results demonstrate the superior performance of the proposed EML model. Across all evaluated metrics, our model consistently outperformed the individual base learners and other benchmarked algorithms, securing the top rank in terms of accuracy. Specifically, the LR model operating on the first-layer ensemble predictions proved highly effective in both enhancing accuracy and preventing overfitting. Following our proposed model, the standalone LR and RF models demonstrated commendable performance, ranking second and third, respectively.

The findings of this study underscore the significant potential of an optimized EML model for advancing the field of glioma tumor grading. The proposed model generated promising results and mitigated overfitting through integrating diverse base learners and using an LR model as a meta-model. The results reveal that the proposed model is a reliable and robust tool that can aid Clinical specialists in effectively diagnosing and classifying gliomas, ultimately paving the way for improved patient satisfaction.

Keywords: Brain tumor, tumor detection, glioma, data mining, group learning, parameter optimization.

۱-مقدمه

تومورهای مغزی گلیوما، به‌عنوان شایع‌ترین تومورهای اولیه دستگاه عصبی مرکزی، از سلول‌های گلیال یا پیش‌سازهای آن‌ها سرچشمه می‌گیرند و به‌دلیل ماهیت تهاجمی و روند پیشرفت سریعی که دارند، با نرخ بالای مرگ‌ومیر مواجه‌اند. تومورهای گلیوما به انواع مختلفی تقسیم می‌شوند که شامل آستروسیتوما[1]، اولیگودندروگلیوما[2] و اپندیموما[3] هستند [۱]. دسته‌بندی این تومورها بر اساس درجهٔ خطر و ویژگی‌های بافت‌شناسی برای مدیریت صحیح بیماری، شامل برنامه‌ریزی و نظارت بر درمان و همچنین پیش‌بینی میزان بقای بیمار، از اهمیت بسیار بالایی برخوردار است [۲، ۳].

درمان رایج گلیوما بر اساس درجهٔ تومور تعیین می‌شود و به‌طورکلی شامل بیشینهٔ برداشت جراحی، به‌دنبال آن پرتودرمانی و شیمی‌درمانی است. شیمی‌درمانی بر حسب معمول هم‌زمان یا متوالی با پرتودرمانی تجویز می‌شود [۳، ۴]. روش‌های تشخیص سنتی برای تعیین درجهٔ تومور به‌طور معمول شامل بیوپسی بافتی[4] و تحلیل میکروسکوپی است؛ اما این روش‌ها دارای خطرات و محدودیت‌هایی هستند؛ به همین دلیل، تلاش‌های زیادی برای توسعهٔ روش‌های غیرتهاجمی تصویربرداری مانند تصویربرداری تشدید مغناطیسی[5] (MRI) و استخراج ویژگی‌های رادیومیک صورت گرفته است. این روش‌ها می‌توانند به فراهم‌آوردن داده‌های ارزشمند در خصوص تمایز تومورها و پیش‌بینی احتمال وخامت آن‌ها کمک کنند [۵، ۶]. ارزیابی پیشرفت فعلی تومور به ارزیابی بالینی و تفسیر MRI با استفاده از معیارهای ارزیابی پاسخ در نوروانکولوژی[6] (RANO) بستگی دارد؛ بااین‌حال، این رویکرد نیز دارای موانعی ازجمله مشکل در تفسیر یافته‌های MRI، ناهمگونی و عدم هم‌زمانی رادیوتراپی، ناهنجاری‌های تصویربرداری، پیگیری پراکنده و مدیریت مراقبت‌های بهداشتی است. تصاویر MRI بیماران مبتلا به گلیوما، هنگامی که برای تغییرات در طول زمان (در پایان، پیش و پس از رادیوتراپی) تجزیه‌وتحلیل می‌شوند، می‌توانند ویژگی‌های پیش‌بینی‌کنندهٔ شکست درمان را شناسایی و در هدایت مراقبت‌های بیمارستانی از بیمار، کمک کنند؛ همچنین، به‌منظور بررسی پاسخ به درمان و طبقه‌بندی غیرتهاجمی گلیوما، پژوهش‌گران به بررسی پروتئومیکس[7] روی آورده‌اند تا نشان‌گرهای زیستی پروتئینی مرتبط با مداخلات را شناسایی کنند.

به‌نظر می‌رسد ترکیبی از دو روش تصویربرداری و نشان‌گرهای زیستی مبتنی‌بر آزمایش‌های زیستی ضروری است [۷]. با توجه به حجم وسیع اطلاعات تولیدشده به‌وسیلهٔ هر دو روش، موفقیت در این زمینه به استفاده از روش‌های محاسباتی و الگوریتم‌های هوش مصنوعی برای حل پیچیدگی و ناهمگونی ذاتی در پیشرفت گلیوما بستگی دارد [۳].

[1] Astrocytomas
[2] Oligodendrogliomas
[3] Ependymomas
[4] Tissue Biopsy
[5] Magnetic Resonance Imaging (MRI)
[6] Response Assessment in Neuro-Oncology
[7] Proteomics

براساس طبقه‌بندی سازمان بهداشت جهانی[1] (WHO) در سال ۲۰۱۶، گلیوماها به دو دستهٔ اصلی تقسیم می‌شوند[۸]: گلیوماهای درجهٔ پایین (LGG) و گلیوماهای درجهٔ بالا (HGG). گلیوماهای درجه پایین، ازجمله درجهٔ I (آستروسیتوماهای پیلوسیتیک[2]) و درجهٔ II (آستروسیتوماهای منتشره[3])، به‌طور کلی خوش‌خیم‌اند و رشد کندی دارند؛ درصورت تشخیص زودهنگام قابل درمان‌اند و دارای پیش‌آگهی[4] بهتری برای بیمار هستند؛ درمقابل، گلیوماهای درجهٔ بالا، شامل درجه III(آناپلاستیک آستروسیتوما[5]) و درجهٔ IV (گلیوبلاستوما مالتی‌فرم[6])، تومورهای بدخیم و خطرناک‌تری هستند که نرخ بقای بسیار پایینی دارند و درمان آن‌ها چالش‌های زیادی را به‌همراه دارد. گلیوبلاستوما (GBM)، به‌عنوان برجسته‌ترین و مرگبارترین نوع گلیوما، به‌سرعت رشد می‌کند و احتمال بروز عوارض شدید را افزایش می‌دهد[۹، ۱۰].

گلیوماهای با ویژگی‌های سرطانی متفاوت، واکنش‌های مختلفی به درمان نشان می‌دهند و این امر نیاز به روش‌های تشخیصی دقیق و زودهنگام را برجسته می‌سازد تا بتوان درمان‌های مؤثرتر و هدفمندتری را برنامه‌ریزی کرد؛ برای مثال، برخی از زیرنوع‌های گلیوماهای درجهٔ پایین می‌توانند به‌سرعت به GBM تبدیل شوند؛ لذا تمایز سریع بین گلیوماهای LGG و GBM اهمیت بالایی دارد. موفقیت درمان بستگی به تعریف صحیح این طبقه‌بندی‌ها و شناخت ویژگی‌های بالینی و پاتولوژیک هر نوع دارد.

با توجه به پیچیدگی و تنوع تومورهای مغزی، استفاده از الگوریتم‌های یادگیری ماشین می‌تواند به افزایش دقت در تشخیص و درجه‌بندی آن‌ها کمک کند؛ و درنهایت به طراحی برنامه‌های درمانی بهینه و کاهش عوارض جانبی ناشی از روش‌های تهاجمی منجر شود. این روش‌ها امکان بهبود نتایج بالینی برای بیماران مبتلا به گلیوما را فراهم می‌سازند. در این پژوهش، تمرکز بر طبقه‌بندی تومورهای گلیوما به دو دستهٔ اصلی گلیومای درجهٔ پایین/ خوش‌خیم (LGG) و گلیوبلاستوما/ بدخیم (GBM) است.

به‌تازگی، چندین روش داده‌کاوی و یادگیری ماشین برای درجه‌بندی گلیوما ارائه شده‌است. برخی از برجسته‌ترین روش‌های موجود عبارت‌اند از: SMOTE_SVM [۱۱]، SMOTE_CatBoost [۱۱]، XGBoost [۱۲]، AdaBoost [۷] و RUS_SVM[۱۱].

روش‌های موجود به نتایج امیدبخشی در حل مسئلهٔ درجه‌بندی گلیوما دست یافته‌اند؛ اما کارایی کسب‌شده ایده‌آل نبوده و نیاز است تا روش‌های جدید ارائه شوند و یا روش‌های موجود بهبود یابند. این مقاله توانسته است با ترکیب رویکردهای مختلف یادگیری ماشین، به بهبود قابل‌توجهی در دقت تشخیص و درجه‌بندی گلیوما نسبت به کارهای مشابه پیشین دست یابد. روش ارائه‌شده نه‌تنها دقت تشخیص را افزایش می‌دهد، بلکه با شناسایی مهم‌ترین ویژگی‌های مرتبط با درجه‌بندی گلیوما، امکان تحلیل دقیق‌تر و مؤثرتر داده‌های پزشکی را فراهم می‌کند؛ لذا این رویکرد ترکیبی با بهره‌گیری از نقاط قوت روش‌های مختلف یادگیری، به بهینه‌سازی فرایند تشخیص و ارائهٔ ابزارهای پیشرفته‌تر برای تمایز بین انواع گلیوما، به‌ویژه در راستای بهبود مدیریت بالینی و نتایج درمانی، کمک شایانی می‌کند؛ بنابراین، در پژوهش حاضر، روش یادگیری ماشین گروهی[7] (EML) با هستهٔ پشته‌ای ارائه شده که هدف آن بهبود شاخص‌های کارایی و حل مشکل بیش‌برازش و کم‌برازش از طریق ترکیب الگوریتم‌های یادگیری متنوع است. روش پیشنهادی ساختاری دولایه دارد که در لایهٔ نخست از ترکیبی از مدل‌های تقویت دسته‌بندی (CatBoost)، جنگل تصادفی (RF)، درخت‌های به‌شدت تصادفی (ERT) و ماشین بردار پشتیبان (SVM) و در لایهٔ دوم از مدل رگرسیون لجستیک (LR) جهت ترکیب نتایج مدل‌های لایهٔ نخست و تولید پیش‌بینی نهایی استفاده می‌شود. مدل پیشنهادی تا حد زیادی مشکل بیش‌برازش[8] را حل کرده و موجب بهبود کارایی مدل‌های موجود می‌شود.

به‌اختصار، نوآوری‌های این پژوهش شامل موارد زیر است:

- ارائهٔ یک راه‌کار یادگیری ماشین گروهی برای حل مسئلهٔ درجه‌بندی گلیوما: فرمول‌بندی مسئلهٔ درجه‌بندی تومورهای مغزی گلیوما به‌صورت یک مسئلهٔ یادگیری ماشین نظارتی، موجب می‌شود به‌سهولت بتوان از طیف وسیعی از الگوریتم‌های داده‌کاوی و یادگیری ماشین برای حل مسئله استفاده کرد.
- ارائهٔ مدل یادگیری ماشین گروهی: مدل پیشنهادی EML با رویکردی نوآورانه، کارایی درجه‌بندی گلیوما را به‌شکل قابل‌توجهی بهبود می‌دهد. این مدل از یک ساختار دولایه بهره می‌برد که در لایهٔ نخست، چهار الگوریتم یادگیری پایه شامل SVM، ERT، RF و CatBoost به‌صورت هوشمندانه با هم ترکیب شده‌اند. در لایهٔ دوم، الگوریتم LR برای ترکیب نتایج لایهٔ نخست به‌کار گرفته شده‌است. این رویکرد دولایه باعث

[1] World Health Organization (WHO)
[2] Pilocytic astrocytomas
[3] Diffuse astrocytomas
[4] Prognosis
[5] Anaplastic astrocytomas
[6] Glioblastoma Multiforme
[7] Ensemble machine learnign
[8] Overfitting

افزایش پایداری و بهبود کارایی مدل در مقایسه با مدل‌های پایه می‌شود.

- انتخاب ویژگی ترکیبی: برای انتخاب ویژگی‌های تأثیرگذار در مدل پیشنهادی از ترکیب هوشمندانهٔ دو الگوریتم SHAP و Boruta استفاده شد. روش پیشنهادی با حذف داده‌های نامربوط و زائد، به ساخت مدل یادگیری ماشین کارآمد، دقیق، سریع و قابل تفسیر کمک می‌کند.
- بهبود کارایی درجه‌بندی گلیوما: آزمایش‌های متعددی جهت ارزیابی روش پیشنهادی و مقایسهٔ آن با سایر مدل‌ها صورت گرفته‌است. تحلیل نتایج آزمایش‌ها تأیید می‌کند که مدل پیشنهادی EML بهبود قابل‌توجهی نسبت به سایر همتایان کسب کرده‌است. مدل EML بر روی مجموعه‌دادهٔ آزمون به‌ترتیب مقادیر ۰.۸۹۲۹، ۰.۸۲۰۵، ۰.۹۴۱۲ و ۰.۸۷۶۷ بر حسب معیارهای صحت، دقت، یادآوری و F1 کسب کرده‌است که از برتری قابل‌توجهی در قیاس با سایر همتایان برخوردار است؛ همچنین تأثیر ویژگی‌های متعدد در درجه‌بندی گلیوما بررسی و مهم‌ترین ویژگی‌ها به‌ترتیب اهمیت مشخص شد تا راهنمای خوبی در تشخیص و درجه‌بندی گلیوما برای پزشکان باشد.

پس از این مقدمه، در بخش دو، کارهای مرتبط در زمینهٔ تشخیص و درجه‌بندی تومورهای مغزی گلیوما مطرح شد. جزئیات روش پیشنهادی به‌همراه مؤلفه‌های آن در بخش سه بررسی شده‌است. در بخش چهارم، به ارزیابی و بحث بر روی نتایج پرداخته شده‌است؛ درنهایت، در بخش پنج به نتیجه‌گیری و ارائهٔ پیشنهادهایی برای کارهای آینده پرداخته شده‌است.

۲-کارهای مرتبط

دسته‌بندی گلیوما برای بهینه‌سازی راه‌بردهای درمانی و بهبود بیماران بسیار مهم است. پیشرفت‌های اخیر در یادگیری ماشینی به‌طور قابل‌توجهی دقت و قابلیت اعتماد این فرایند را فراتر از تحلیل هیستوپاتولوژیک سنتی افزایش داده‌اند. با به‌کارگیری روش‌های یادگیری ماشین تغییر قابل‌توجهی در نحوهٔ ارزیابی گلیوما ارائه شده‌است که می‌تواند به تشخیص‌های دقیق‌تر و تصمیم‌گیری‌های درمانی بهتر منجر شود.

مطالعات مختلفی در زمینهٔ استفاده از الگوریتم‌های داده‌کاوی و یادگیری ماشین برای پیش‌بینی و طبقه‌بندی انواع گلیوما براساس انواع مختلف داده‌ها ازجمله ویژگی‌های رادیولوژیکی، بالینی و مولکولی تمرکز داشته‌اند. پژوهش‌گران از ویژگی‌های رادیومیک استخراج‌شده از پویش‌های MRI برای پیش‌بینی غیرتهاجمی جهش‌های مولکولی و درجات تومور استفاده کرده‌اند؛ برای مثال، *دیپاک و امیر* [۱۳] با استفاده از یادگیری عمیق و انتقال ویژگی‌ها از مدل پیش‌آموزش‌دیدهٔ GoogLeNet برای استخراج ویژگی‌ها از تصاویر MRI برای تمایز بین سه نوع تومور مغزی استفاده کرده‌اند؛ همچنین، *الکساس و همکاران* [۱۴] یک سامانهٔ درجه‌بندی مبتنی‌بر تصویربرداری برای تشخیص دقیق درجات مختلف تومورهای گلیوما با بهره‌گیری از روش‌های تصویربرداری مغناطیسی غیرتهاجمی و استخراج ویژگی‌های متنوع شامل مورفولوژیکی، بافتی و عملکردی توسعه داده‌اند؛ این ویژگی‌ها با کمک الگوریتم‌های پیشرفتهٔ انتخاب ویژگی، بهینه‌سازی شده و درنهایت به یک مدل شبکهٔ عصبی چندلایه جهت طبقه‌بندی و تشخیص نهایی درجهٔ تومورهای گلیوما وارد می‌شوند.

در حوزهٔ کاربرد یادگیری ماشین و یادگیری عمیق برای درجه‌بندی و تشخیص گلیوما پژوهش‌هایی انجام شده‌است، اما تفاوت‌های قابل‌توجهی در روش‌شناسی، داده‌ها و اهداف نشان می‌دهند. *وانگ و همکاران* [۱۵] مدل‌های یادگیری ماشین کلاسیک مانند SVM برای درجه‌بندی گلیوما (درجهٔ II، III و IV) از تصاویر اسلاید کامل (WSI) و نشان‌گر Ki-67 استفاده می‌کند؛ به‌طوری که با تفسیر کمی نتایج و به‌کار بستن ویژگی‌های مورفولوژیکی به دقت نود درصد دست یافته‌اند، اما در هر صورت وابستگی به داده‌های آسیب‌شناسی محدود (۱۴۶ مورد) و ادغام‌نشدن تصاویر چندمدالی مانند MRI است، موجب کاهش تعمیم‌پذیری نتایج می‌شود. ون و همکاران [۱۶] بر چهارچوب یادگیری عمیق گروهی مبتنی‌بر D U-Net 3 برای طبقه‌بندی و درجه‌بندی هم‌زمان گلیوما با MRI چندمدالی تأکید دارد؛ باوجود دقت مناسب، پیچیدگی محاسباتی بالا و نیاز به داده‌های بزرگ برای آموزش در محیط‌های بالینی محدود چالش‌برانگیز است. ژیا و همکاران [۱۷] به تمایز گلیوبلاستوما از متاستاز مغزی منفرد با رادیومیکس و مدل‌های تفسیرپذیر مانند LGBM پرداخته‌اند، با وجود دقت بالا (AUC بیش از ۰.۹)، اما با تمرکز صرف بر تمایز دوگانه (نه درجه‌بندی چندگانه) و نمونه‌های به‌نسبت کوچک (۴۳۴ بیمار) امکان جهت‌گیری نتایج وجود دارد. *جوشی و همکاران* [۱۸] رویکرد گروهی یادگیری ماشین برای پیش‌بینی وجود گلیوما و درجه‌بندی چندگانه با استفاده از نشانگرهای زیستی مانند hTERT و YKL-40 مورد توجه قرار داده‌اند، استفاده از داده‌های غیرتصویری

(۱۳۵ مورد)، موجب سادگی و دقت بالا شده‌است؛ اگرچه نیاز به اعتبارسنجی بیشتر در جمعیت‌های بزرگ‌تر دارد.

گوتا و همکاران [۱۹] نشان دادند که استفاده از ویژگی‌های یادگرفته‌شده به‌وسیلهٔ یک شبکهٔ عصبی کانولوشنی (CNN) برای پیش‌بینی درجهٔ گلیوما از ویژگی‌های رادیومیک رایج بهتر عمل می‌کند. *چنگ و همکاران* [۲۰]، با بهره‌گیری از مجموعه‌ای شامل ۲۱۵۳ ویژگی رادیومیک داخل تومور و اطراف آن استخراج‌شده از تصاویر چندپارامتری MRI پیش‌از عمل ۲۸۵ بیمار و با استفاده از روش‌های یادگیری ماشین مانند LASSO و mRMR برای انتخاب بهترین ویژگی‌ها، توانستند مدلی کارآمد برای پیش‌بینی درجهٔ گلیوما ارائه دهند.

تاسچی و همکاران [۷] روشی سلسله‌مراتبی مبتنی‌بر رأی‌گیری برای انتخاب ویژگی و یادگیری ترکیبی با هدف درجه‌بندی گلیوما ارائه داده‌اند. این پژوهش با بهره‌گیری از داده‌های بالینی و مولکولی از پایگاه‌های دادهٔ عمومی TCGA و CGGA، به بررسی اهمیت فزایندهٔ نشان‌گرهای مولکولی در طبقه‌بندی تومورها پرداخته است. روش پیشنهادی آن‌ها شامل استفاده از چهار روش کاهش ابعاد برای انتخاب ویژگی و پنج مدل یادگیری تحت نظارت بود که در مجموع شانزده ترکیب مختلف را برای یافتن بهترین عملکرد مورد بررسی قرار دادند؛ آن‌ها همچنین عملکرد روش خود را با روش انتخاب ویژگی LASSO به‌صورت جداگانه مقایسه کردند و نشان دادند که رویکرد ترکیبی مبتنی‌بر رأی‌گیری آن‌ها نسبت به روش LASSO برتری دارد. نتایج محاسباتی این مطالعه نشان داد که روش پیشنهادی به‌ترتیب به صحت ۰.۸۷۶۶ و ۰.۷۹۶۶۸ در مجموعه‌داده‌ها دست یافت. داده‌های مورد بررسی *تاسچی و همکاران* شامل ۸۳۹ رکورد داده و ۲۳ ویژگی است، که در پژوهش حاضر نیز مورد توجه قرار گرفته است.

در مطالعه‌ای که توسط *نویاندی و همکاران* [۲۱] انجام شده‌است، یک مدل هوش مصنوعی برای درجه‌بندی گلیوما با استفاده از مدل LightGBM و رویکردهای هوش مصنوعی قابل‌توضیح (XAI) توسعه یافته است. هدف اصلی این پژوهش، افزایش دقت و قابلیت اعتماد در درجه‌بندی گلیوما با استفاده از داده‌های مولکولی و بالینی بیماران از پایگاه دادهٔ اطلس ژنوم سرطان (TCGA) است. مدل LightGBM بهینه‌سازی‌شده در این مطالعه توانست به عملکرد مطلوبی نسبت به سایر مدل‌های یادگیری ماشینی مانند RF، Naive Bayesian و SVM دست یابد. نکتهٔ قابل توجه در این پژوهش، استفاده از روش‌های XAI به‌ویژه مقادیر SHAP است. این روش به پژوهش‌گران امکان داد تا فرایند تصمیم‌گیری مدل را تفسیر کرده و جهش ژن IDH1 را به‌عنوان یک عامل پیش‌بینی‌کنندهٔ مهم در درجه‌بندی گلیوما شناسایی کنند.

سانچز-مارکس و همکاران [۱۱]، در پژوهشی رویکرد داده‌محور در یادگیری ماشینی برای بهبود پیش‌بینی درجه‌بندی گلیوما را مورد بررسی قرار داده‌اند. این مطالعه بر این فرض استوار بود که به‌جای تمرکز صرف بر روی الگوریتم‌ها، بهبود کیفیت داده‌ها می‌تواند منجر به عملکرد بهتر مدل شود. پژوهش‌گران با استفاده از داده‌های بالینی و نشان‌گرهای مولکولی از پایگاه دادهٔ TCGA، به بررسی تأثیر استانداردسازی و روش بیش‌نمونه‌برداری[1] برای طبقهٔ با تعداد کم پرداختند. نتایج تجربی آن‌ها نشان داد که این اقدامات، عملکرد چهار مدل یادگیری ماشینی رایج و دو مدل ترکیبی (Ensemble) را افزایش قابل‌توجهی می‌دهد. مدل‌های ترکیبی RF و CatBoost بهترین امتیازات را در میان تمامی مدل‌ها کسب کرده‌اند؛ یکی از یافته‌های کلیدی این پژوهش، تأیید اهمیت نشان‌گرهای مولکولی در درجه‌بندی گلیوما بود که از طریق تحلیل‌های توصیفی و الگوریتم‌های رتبه‌بندی ویژگی‌ها به‌دست آمد؛ درنهایت، این مطالعه مقایسه‌ای بین رویکرد مدل‌محور و داده‌محور انجام و نشان داد که روش داده‌محور پیشنهادی آن‌ها با صحت ۸۸/۲ درصد (با استفاده از روش بیش‌نمونه‌برداری) نسبت به روش مدل‌محور پیشین با مقدار صحت ۸۷/۶ درصد، عملکرد بهتری دارد.

سامارا و هری [۱۲]، یک رویکرد یک‌پارچه برای درجه‌بندی گلیوما با استفاده از یادگیری ماشینی و روش‌های انتخاب ویژگی Boruta، LASSO و SHAP معرفی کردند. هدف این پژوهش، شناسایی مهم‌ترین نشان‌گرهای ژنتیکی و بالینی و ارائهٔ یک مدل با قابلیت تفسیرپذیری بالینی بود. این پژوهش‌گران با استفاده از یک مجموعه‌داده از TCGA که شامل ۲۳ ویژگی بود، سیزده ویژگی کلیدی ازجمله جهش‌های ژن‌های مهم مانند (IDH1, TP53, ATRX)، سن در زمان تشخیص و نژاد را شناسایی کردند؛ آن‌ها سپس از این ویژگی‌های انتخاب‌شده برای آموزش چهار مدل یادگیری ماشینی XGBoost، SVM، RF و LR استفاده کردند. نتایج نشان داد که مدل LR به بالاترین صحت ۸۸.۰۹ درصد در مجموعه‌داده‌های آزمون دست یافته است.

استفاده از روش‌های هوش محاسباتی، داده کاوی و یادگیری ماشین در زمینهٔ تشخیص و دسته‌بندی سایر انواع بیماری‌ها نیز استفاده شده‌است؛ برای مثال، پژوهش‌گران

[1] Oversampling

برای تشخیص تومورهای سینه، از مدل‌های محاسبات نرم مبتنی بر فازی، تکاملی و هوش جمعی استفاده کرده‌اند [۲۹]. در پژوهشی دیگر، از مدل یادگیری ماشین انباشته برای دسته‌بندی و پیش‌بینی بیماری‌های کبدی استفاده شده‌است [۳۰]. نتایج مدل‌های یادگیری ماشین کارایی مطلوب آن‌ها را در حل مسائل حوزه پزشکی عیان کرده است؛ بنابراین، در ادامه یک روش مبتنی بر یادگیری ماشین برای پیش‌بینی درجات تومور مغزی گلیوما ارائه می‌کنیم.

۳-روش پیشنهادی

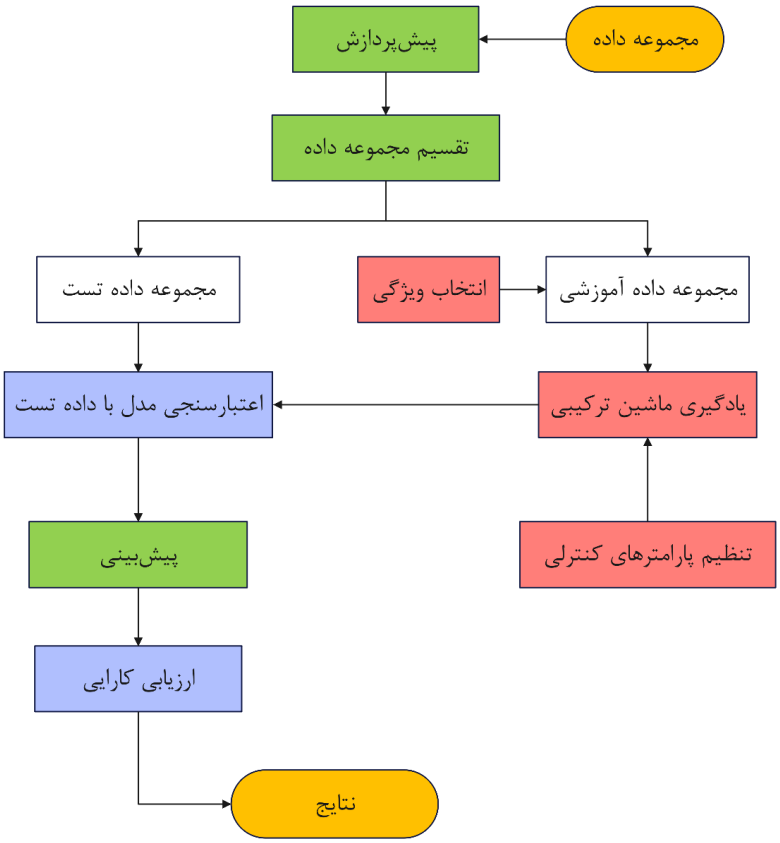
شکل (۱) ساختار کلی رویکرد پیشنهادی را نشان می‌دهد که شامل هشت مؤلفهٔ اصلی است که عبارت‌اند از: پیش‌پردازش، تقسیم داده، انتخاب ویژگی، مدل یادگیرنده، تنظیم پارامترهای کنترلی، اعتبارسنجی مدل، پیش‌بینی و ارزیابی کارایی. در ادامه به‌تفصیل به بررسی این مؤلفه‌ها می‌پردازیم.

۳-۱-مجموعه‌داده‌ها و پیش‌پردازش

مجموعه‌دادهٔ مورداستفاده برای ارزیابی کارایی مدل‌ها از اطلس ژنوم سرطان و مقالهٔ *تاسچی و همکاران* [۷] گرفته شده‌است. این مجموعه‌داده شامل ۸۳۹ رکورد داده و ۲۳ ویژگی است که افراد را در دو دسته گلیوماهای خوش‌خیم (LGG) و گلیوبلاستوما (GBM) طبقه‌بندی کرده‌است. تعداد افراد در دستهٔ LGG برابر ۴۸۷ و تعداد نمونه‌ها در دستهٔ GBM برابر ۳۵۲ است. صفات در دو دستهٔ ژنتیکی (بیست ویژگی) و بالینی (سه ویژگی) قرار دارند. مقادیر متغیرهای ژنتیکی بر حسب معمول به‌صورت جهش‌دار (=۱) و بدون جهش (=۰) نمایش داده می‌شوند. جدول (۱) مشخصات مجموعه‌دادهٔ مورد استفاده را نشان می‌دهد. مقادیر متغیرها که توصیفی هستند، به‌صورت نمایش دودویی یا عددی کدگذاری شده‌اند تا سازگاری مدل‌های یادگیری ماشین را تضمین کنند. توضیحات بیشتر در مورد مجموعه‌دادهٔ گلیوما در مرجع زیر آمده‌است:

https://archive.ics.uci.edu/dataset/759/glioma+grading+clinical+and+mutation+features+dataset

شکل (۲) هم‌بستگی بین متغیرهای وابسته و متغیر هدف Grade را نشان می‌دهد. مقادیر هم‌بستگی پیرسون در شکل بیان‌کنندهٔ ارتباط خطی بین متغیرهای مختلف است. ضریب هم‌بستگی مقداری بین ۱- و ۱ است که در آن ۱- نشان‌دهندهٔ هم‌بستگی خطی به‌طور کامل منفی بین دو متغیر، مقدار ۰ نشان‌دهندهٔ نبود هم‌بستگی خطی بین دو متغیر و مقدار یک نشان‌دهندهٔ هم‌بستگی خطی به‌طور کامل مثبت بین دو متغیر است؛ همان‌گونه که در شکل نشان داده شده‌است، متغیرهای Age و PTEN بیشترین هم‌بستگی را و درمقابل، متغیرهای IDH1 و ATRX کمترین هم‌بستگی را با متغیر هدف دارند.



(شکل-۱): ساختار روش پیشنهادی

(Figure-1): Structure of the proposed approach

(جدول-۱): مشخصات مجموعه‌دادهٔ تومور مغزی گلیوما

(Table-1): Characteristics of the glioma brain tumor dataset

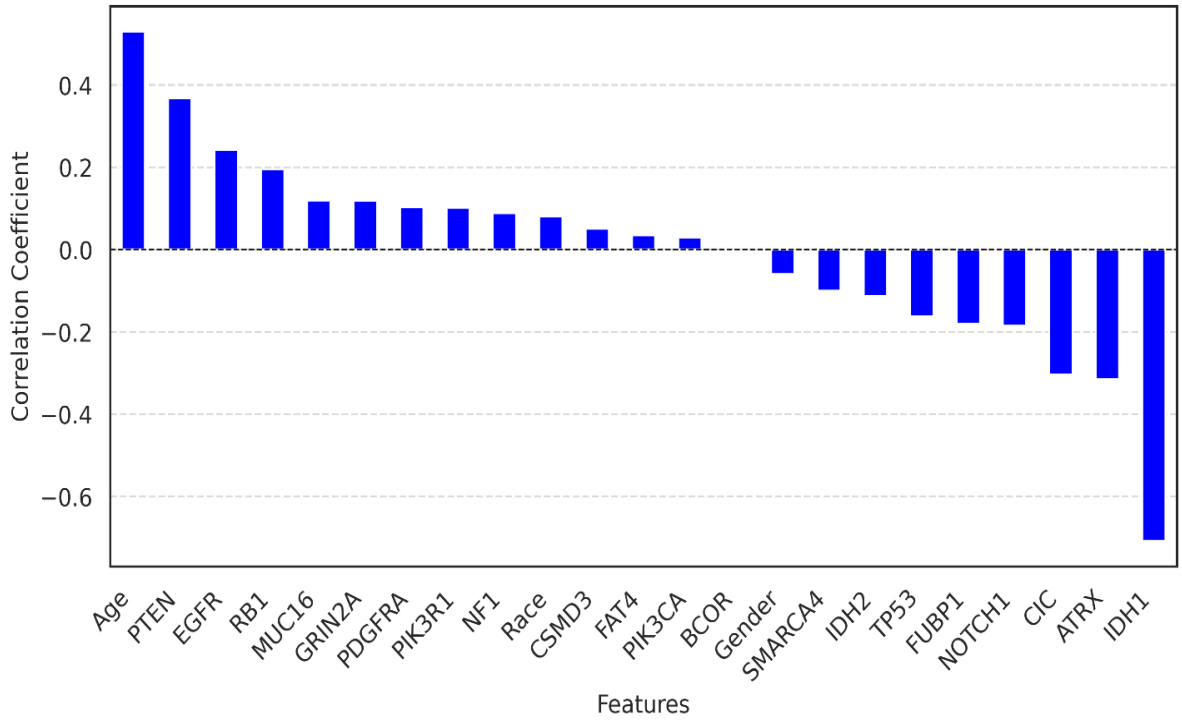
شماره	ویژگی	توصیف	شماره	ویژگی	توصیف
۱	Gender	جنسیت (مرد=۰ و زن=۱)	۱۳	PIK3R1	زیرواحد تنظیمی یک کیناز فسفواینوزیتید-۳
۲	Age at diagnosis	سن (سال)	۱۴	FUBP1	پروتئینی که به عناصر خاص در ناحیهٔ بالادستی ژن‌ها متصل می‌شود و در تنظیم بیان ژن نقش دارد
۳	Race	نژاد (سفید=۰، سیاه یا آفریقایی=۱، آسیایی=۲ و سرخ پوست آمریکایی یا بومی آلاسکا=۳)	۱۵	RB1	پروتئینی در تنظیم رونویسی ژن‌ها و مهار فعالیت‌های رونویسی ۱
۴	IDH1	ژن ایزوسیترات دهیدروژناز ۱	۱۶	NOTCH1	گیرندهٔ کلیدی در انتقال سیگنال‌های سلولی که در تنظیم فرایندهای بیولوژیکی حیاتی و توسعهٔ بافت‌ها نقش دارد
۵	TP53	ژن سرکوب‌گر تومور مرتبط با پروتئین p53	۱۷	BCOR	پروتئینی که به BCL6 متصل می‌شود و فعالیت رونویسی آن را مهار کرده و در تنظیم بیان ژن‌ها، به‌ویژه در دستگاه ایمنی نقش دارد.
۶	ATRX	پروتئین بازسازی‌کنندهٔ کروماتین که نقش مهمی در تنظیم ژن و ثبات ژنوم ایفا می‌کند	۱۸	CSMD3	پروتئینی که در توسعهٔ دستگاه عصبی و تنظیم فعالیت‌های سلولی نقش دارد و شامل چندین دامنهٔ CUBو Sushi است
۷	PTEN	پروتئین سرکوب‌گر تومور و آنزیم فسفاتاز	۱۹	SMARCA4	پروتئینی که در تنظیم ساختار کروماتین و بیان ژن‌ها نقش دارد و به‌عنوان یک بخش کلیدی در کمپلکس‌های تنظیمی SWI/SNF شناخته می‌شود
۸	EGFR	گیرندهٔ فاکتور رشد اپیدرمی	۲۰	GRIN2A	این ژن با تنظیم سیگنال‌دهی عصبی و تأثیر بر رشد و پاسخ به درمان نقش دارد
۹	CIC	پروتئین سرکوب‌گر رونویسی	۲۱	IDH2	ایزوسیترات دهیدروژناز ۲
۱۰	MUC16	ژن مرتبط با سطح سلول	۲۲	FAT4	ژن کدکنندهٔ پروتئینی که به‌عنوان عضوی از خانوادهٔ پروتئین‌های کادهرین شناخته می‌شود و در چسبندگی سلول‌ها به یکدیگر و تعیین قطبیت سلولی نقش دارد
۱۱	PIK3CA	زیرواحد کاتالیتیک آلفا فسفاتیدیل اینوزیتول	۲۳	PDGFRA	گیرندهٔ فاکتور رشد مشتق از پلاکت آلفا
۱۲	NF1	نوروفیبرومین ۱			

مجموعه‌دادهٔ ورودی گلیوما در ابتدا پردازش شده‌است تا برای آموزش و آزمون مدل‌های یادگیری ماشین مناسب باشد. پیش‌پردازش شامل دو مرحله است:

- استانداردسازی مقادیر ویژگی‌های پیوسته برای یک‌دست‌کردن مقیاس آن‌ها: در این مرحله مقادیر ویژگی‌های پیوسته همانند ویژگی Age با استفاده از ماژول مقیاس‌بندی استاندارد در پایتون پردازش شده‌است تا مقادیر ویژگی‌های عددی در مقیاس قابل مقایسه قرار گیرند؛ دلیل این کار برقراری توازن در نقش ویژگی‌ها در فرایند آموزش مدل‌های یادگیری ماشین است.
- مدیریت داده‌های گم‌شده: در صورت وجود داده‌های گم‌شده، برای جایگزینی آن‌ها از روش میانه‌گیری استفاده شده‌است که در آن مقادیر گم‌شده هر ویژگی با میانهٔ سایر مقادیر موجود در همان ویژگی جایگزین می‌شود. گفتنی است، روش‌های متعددی جهت مدیریت داده‌های گم‌شده وجود دارد. دلیل انتخاب روش میانه‌گیری از میان سایر روش‌های متعدد مدیریت داده‌های گم‌شده، مقاومت مطلوب آن درمقابل داده‌های پرت[1] است.

دیگر این‌که است که در برخی مجموعه‌داده‌ها ممکن است نیازی به انجام مراحل پیش‌پردازش نباشد، اما وجود مؤلفهٔ پیش‌پردازش قابلیت استفاده از رویکرد پیشنهادی را تعمیم داده و برای مجموعه داده‌های مختلف آن را قابل‌استفاده می‌کند.

[1] Outlier



(شکل-۲): ضرایب هم‌بستگی بین ویژگی‌ها و متغیر هدف Grade در مجموعه‌دادهٔ گلیوما

(Figure-2): Correlation coefficient between features and target variable Grade in glioma dataset

۳-۲-تقسیم داده‌ها

مجموعه‌دادهٔ ورودی به دو بخش داده‌های آزمایشی (هفتاد درصد) و آزمون (سی درصد) تقسیم شده‌است. مجموعه‌دادهٔ آموزشی برای آموزش مدل‌های یادگیری ماشین استفاده شده و مجموعه‌دادهٔ آزمون برای اعتبارسنجی مدل‌ها و ارزیابی کارایی آن‌ها استفاده می‌شود. در فرایند تقسیم از راهبرد نمونه‌گیری طبقه‌بندی‌شده استفاده شد تا توزیع طبقهٔ اصلی داده‌ها که شامل ۵۸.۰۵ درصد برای طبقهٔ LGG و ۴۱.۹۵ درصد برای طبقهٔ GBM است را حفظ کند. این روش تقسیم داده‌ها که با استفاده از شیوهٔ Stratify کتابخانهٔ train_test_split در پایتون انجام شده‌است، تعداد متوازن داده‌ها در هر دو دسته LGG و GBM را تضمین می‌کند.

۳-۳-انتخاب ویژگی

برای انتخاب ویژگی‌ها، روش ترکیبی که شامل دو الگوریتم Boruta و SHAP است، پیشنهاد شد. در این روش که نسخهٔ تغییریافتهٔ رویکرد مطرح‌شده در مرجع [۱۲]، ابتدا الگوریتم‌های Boruta و SHAP به‌صورت مجزا بر روی مجموعه‌دادهٔ ورودی با تمامی ویژگی‌ها اعمال شده و هرکدام، زیرمجموعه‌ای از ویژگی‌ها را انتخاب می‌کنند؛ سپس از ویژگی‌های انتخاب‌شدهٔ هرکدام از الگوریتم‌ها، اشتراک گرفته می‌شود تا به‌عنوان ویژگی‌های نهایی استفاده شوند.

۳-۳-۱-الگوریتم Boruta

این الگوریتم یک روش انتخاب ویژگی مبتنی‌بر پوشش است که برای تعیین تمام متغیرهای مرتبط با ارزیابی تکراری اهمیت آن‌ها با استفاده از یک دسته‌بند RF طراحی شده‌است؛ برخلاف روش‌های انتخاب ویژگی سنتی که ممکن است متغیرهای با اهمیت، اما با هم‌بستگی ضعیف را کنار بگذارند، Boruta ویژگی‌هایی را که در دسته‌بندی نقش معناداری دارند، حفظ می‌کند [۱۲]. در این روش، میزان اهمیت ویژگی I_i برای هر ویژگی i به‌صورت زیر محاسبه می‌شود:

$$I_i = \frac{1}{T}\sum_{t=1}^{T} G_i^t \qquad (۱)$$

در این رابطه، T تعداد کل درختان و G_i^t اهمیت جینی ویژگی i در درخت t در مدل یادگیری RF است. الگوریتم Boruta امتیاز هر ویژگی را با امتیاز ویژگی‌های سایه[1] مقایسه می‌کند؛ درصورتی که رابطهٔ زیر برای ویژگی I_i برقرار باشد، به‌عنوان ویژگی مرتبط انتخاب می‌شود.

$$I_i \geq \max\left(I_{shadow}\right) \qquad (۲)$$

در این رابطه، $\max\left(I_{shadow}\right)$ بیان‌کنندهٔ بالاترین امتیاز در بین ویژگی‌های سایه است؛ ویژگی‌های سایه، نسخه‌های به‌هم‌ریخته از ویژگی‌های اصلی هستند که برای سنجش اهمیت واقعی آن‌ها در مدل یادگیری استفاده می‌شوند. اگر اهمیت یک ویژگی اصلی از نسخهٔ تصادفی‌شدهٔ

[1] Shadow

آن بیشتر نباشد، ممکن است آن ویژگی بی‌اهمیت باشد. الگوریتم Boruta به‌تعداد پانصد بار اجرا شده و مطابق روش ارائه‌شده در [۱۲] ویژگی‌هایی با امتیاز بالای ۸۵درصد انتخاب شده‌اند. ویژگی‌های انتخاب‌شده به‌وسیلهٔ الگوریتم Boruta عبارت‌اند از:

Age, Race, IDH1, TP53, ATRX, PTEN, EGFR, CIC, MUC16, NF1, PIK3R1, FUBP1, RB1, NOTCH1, SMARCA4, IDH2

۳-۳-۲-الگوریتم SHAP

الگوریتم توضیحات جمعی شیپلی[1] (SHAP) یک رویکرد انتخاب ویژگی مبتنی‌بر قابلیت توضیح است که امتیاز هر متغیر را در پیش‌بینی‌های مدل کمی‌سازی می‌کند[۲۲]؛ برخلاف الگوریتم Boruta که بر اهمیت آماری تمرکز دارند، مقادیر SHAP از نظریهٔ بازی‌ها استخراج می‌شوند و چگونگی تأثیر هر ویژگی بر تصمیمات دسته‌بندی را ارائه می‌دهند[۱۲]. برای محاسبهٔ مقادیر SHAP، از الگوریتم یادگیری XGBoost و میانگین مطلق مقدار SHAP برای هر ویژگی محاسبه شد. در این روش، حد آستانهٔ ۰.۰۲ برای پالایش متغیرهای کم‌اهمیت دنظر گرفته‌شد. مقدار SHAP برای ویژگی i در یک نمونه‌دادهٔ X به‌صورت زیر محاسبه می‌شود:

$$\omega_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S\cup\{i\}) - f(S)] \quad (۳)$$

در این رابطه، F نشان‌گر مجموعهٔ کامل ویژگی‌ها و S زیرمجموعه‌ای از ویژگی‌ها به جز i است. عبارت $f(S\cup\{i\})$ پیش‌بینی مدل پس از افزودن ویژگی i است. عبارت F(S) پیش‌بینی مدل است، درصورتی‌که تنها از زیرمجموعهٔ S بهره ببرد.

شکل (۳) نمودار SHAP را نشان می‌دهد که اهمیت و اثرگذاری هر یک از ویژگی‌ها بر پیش‌بینی مدل را مشخص می‌کند. محور افقی نشان‌دهندهٔ میانگین بزرگی SHAP است که نرخ تأثیر هر ویژگی بر خروجی مدل را نشان می‌دهد. محور عمودی رتبهٔ ویژگی‌ها را بر اساس اهمیت آن‌ها، نشان می‌دهد. مقادیر مثبت و منفی SHAP نشان‌دهندهٔ تأثیر افزایش و کاهش ویژگی‌ها بر پیش‌بینی مدل است؛ همان‌گونه که در شکل (۳) نشان داده شده‌است، ویژگی‌های انتخاب‌شده به‌وسیلهٔ SHAP عبارت‌اند از:

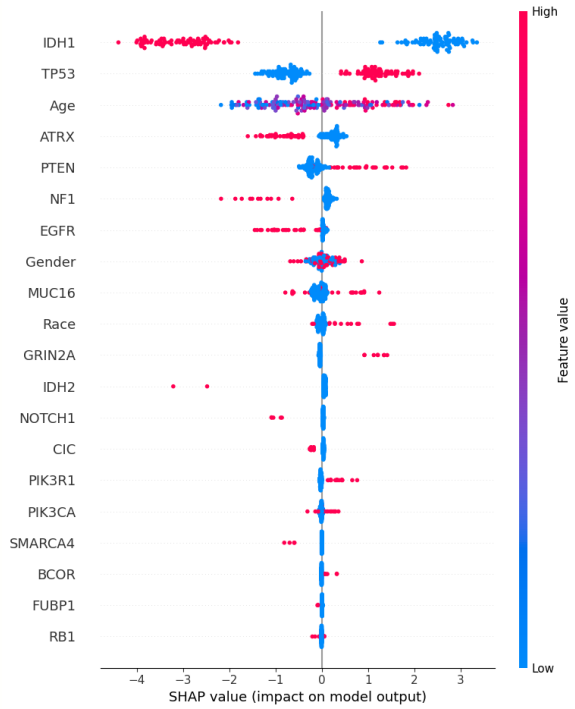
IDH1, TP53, Age, ATRX, PTEN, NF1, EGFR, Gender, MUC16, Race, GRIN2A, IDH2, NOTCH1, CIC, PIK3R1, PIK3CA

[1] SHapley Additive exPlanations

درنهایت، با اشتراک‌گیری از نتایج دو روش SHAP و Boruta، ویژگی‌های منتخب نهایی عبارت‌اند از:

Age, Race, IDH1, TP53, ATRX, PTEN, EGFR, CIC, MUC16, NF1, PIK3R1, NOTCH1, IDH2

پس از اعمال عمل‌گر انتخاب ویژگی، مجموعه‌دادهٔ جدید با ویژگی‌های منتخب ایجاد می‌شود.



(شکل-۳): نمودار SHAP بر روی مجموعه‌دادهٔ آموزشی
(Figure-3): SHAP values on the training dataset

۳-۴-یادگیری ماشین گروهی

مدل یادگیری گروهی (EML) با ساختار پشته‌ای از طریق ترکیب الگوریتم‌های متنوع یادگیری ماشین برای کاهش بایاس و بهبود کارایی پیش‌بینی پیشنهاد شده‌است. شکل (۴) ساختار مدل یادگیری پیشنهادی را نشان می‌دهد. مدل یادگیری شامل دو لایه است که در لایهٔ نخست چهار دسته‌بند قوی شامل SVM، CatBoost، ERT و RF مدل‌های پایه را تشکیل می‌دهند؛ سپس، در لایهٔ دوم، مدل LR به‌عنوان متامدل استفاده شده‌است تا مقادیر پیش‌بینی‌شده از هر مدل پایه را ترکیب کرده و پیش‌بینی نهایی را به‌دست آورد. دلیل استفاده از LR در لایهٔ دوم به‌عنوان یک متامدل ترکیب پیش‌بینی مدل‌های لایهٔ نخست و جلوگیری از مشکل بیش‌برازش است. مجموعه‌دادهٔ آموزشی برای متامدل، مقادیر پیش‌بینی از مدل‌های پایه است که بر روی مجموعه‌دادهٔ آموزشی اصلی تولید شده‌است و مجموعه‌دادهٔ آزمون از میانگین مقادیر پیش‌بینی‌شده برای مجموعهٔ آزمون اولیه تشکیل شده‌است؛ به‌منظور ایجاد مجموعهٔ آزمون که به‌عنوان دادهٔ آموزشی لایهٔ دوم استفاده خواهد شد، مدل‌های پایه بر روی مجموعه‌دادهٔ آموزشی

اولیه آموزش داده می‌شوند؛ به‌منظور جلوگیری از نشت داده‌ها[1]، در فرایند آموزش مدل از استراتژی اعتبارسنجی متقابل با k=5 استفاده شده‌است. در آموزش هرکدام از مدل‌ها، مجموعه‌دادهٔ آموزشی به‌طور تصادفی به پنج بخش تقسیم می‌شود که در هر مرحله، یکی از بخش‌ها به‌عنوان یک مجموعهٔ اعتبارسنجی عمل کرده و هر بخش یک مقدار پیش‌بینی مربوط به داده‌های اعتبارسنجی را تولید می‌کند؛ درنهایت، برای تولید خروجی هرکدام از مدل‌ها، داده‌های اعتبارسنجی استفاده‌شده برای آموزش آن مدل تجمیع و میانگین مقادیر پیش‌بینی‌شده نیز محاسبه می‌شود تا رکورد متناظر مدل در لایهٔ نخست تشکیل شود؛ هنگامی‌که خروجی تمام مدل‌ها تولید شد، رکوردهای حاصل تجمیع و در قالب یک مجموعه‌داده برای آموزش متامدل، به لایهٔ دوم خورانده می‌شوند؛ متامدل در لایهٔ دوم آموزش دیده و پیش‌بینی نهایی را برمی‌گرداند. قابل توجه است که (هفتاد درصد) داده‌ها برای آموزش مدل‌های پایه استفاده می‌شود و نتایج این مدل‌ها به‌عنوان ورودی برای آموزش مدل متا به‌کار می‌رود. در این مرحله، (سی درصد) داده‌های آزمون هیچ‌گونه نقشی ندارند و تنها برای ارزیابی و اعتبارسنجی کارایی مدل استفاده می‌شوند.

در شکل (۴) هر رنگ نمایندهٔ یک بخش کلیدی در این فرایند است. رنگ آبی نشان‌دهندهٔ داده‌های اصلی آموزشی است که برای «آموزش» مدل‌های یادگیری ماشین به‌کار می‌روند. مدل‌های پایه با این داده‌های آبی‌رنگ آموزش می‌بینند. رنگ سبز بخش‌های داده‌ای اعتبارسنجی متقاطع (Cross-Validation) را نشان می‌دهد. در فرایند اعتبارسنجی متقاطع (k-fold Cross-Validation که اینجا k=5 درنظر گرفته شده‌است)، مجموعه‌دادهٔ آموزشی آبی‌رنگ به k بخش (fold) تقسیم شده‌است. در هر مرحله از آموزش، یکی از این قسمت‌ها (که با رنگ سبز مشخص شده‌است) به‌عنوان دادهٔ اعتبارسنجی کنار گذاشته شده و مدل روی k-1 قسمت باقی‌مانده (قطعه‌های آبی‌رنگ) آموزش می‌بینند. رنگ نارنجی نمایان‌کنندهٔ پیش‌بینی‌هایی است که مدل‌های پایه روی داده‌های اعتبارسنجی (سبزرنگ) انجام داده‌اند. قطعه‌های نارنجی رنگ «خروجی‌های میانی» مدل‌های یادگیر پایه هستند. خروجی‌های نارنجی‌رنگ جمع‌آوری شده و درنهایت به‌عنوان مجموعه‌دادهٔ آموزشی برای متامدل در لایهٔ دوم استفاده می‌شوند؛ به‌عبارت دیگر، متامدل روی پیش‌بینی‌های مدل‌های پایه «آموزش» می‌بینند تا بتواند بهترین ترکیب از این پیش‌بینی‌ها رو برای تولید نتیجهٔ نهایی ارائه دهد.

دلایل انتخاب الگوریتم‌های یادگیری به‌عنوان مدل‌های پایه و متامدل به‌صورت زیر است:

- الگوریتم CatBoost [۲۳] مستقیم از ویژگی‌های دسته‌بندی می‌آموزد، دقت و تفسیر نتایج را بدون به‌خطرانداختن عملکرد، بهبود می‌دهد. روش یادشده از یک ساختار درختی متقارن استفاده می‌کند که کارایی فرایند آموزش را بهبود می‌دهد و روند پیش‌بینی را سرعت می‌بخشد؛ مدل CatBoost همچنین مقاومت به‌نسبت بالایی در برابر بیش‌برازش در مجموعه‌های داده با متغیرهای طبقه‌بندی زیاد را دارد. این الگوریتم از روش‌های منظم‌سازی مختلف در طول آموزش استفاده می‌کند که در مقایسه با روش‌های سنتی تقویت گرادیان، حساسیت کمتری نسبت به نوفه و نقاط دورافتاده دارد.
- مدل RF [24] بر اساس اصل ترکیب درختان متنوع و مستقل عمل می‌کند تا طبقه‌بندی‌های قابل‌اعتمادتری به‌دست آورد. این مدل به‌ویژه برای مسائلی مؤثر است که هدف آن‌ها دسته‌بندی نتایج گسسته بر اساس مجموعه‌ای از ویژگی‌های ورودی است؛ همچنین، RF توانایی مطلوبی در شناسایی روابط پیچیدهٔ درون داده‌ها بدون نیاز به تغییرات در ویژگی‌ها را داراست که آن را به‌ویژه برای مجموعه‌های داده با ابعاد بالا و پیچیده مؤثر می‌سازد. رویکرد جمعی کمک می‌کند تا مشکل بیش‌برازش کاهش یابد و پیش‌بینی‌های قوی‌تری نسبت به آن‌هایی که به‌وسیلهٔ درختان تصمیم منفرد تولید می‌شوند، به‌دست آید.
- الگوریتم ERT [۲۵] یک روش یادگیری جمعی است که مجموعه‌ای از درختان تصمیم را ساخته و پیش‌بینی‌های آن‌ها را تجمیع می‌کند تا دقت مدل را بهبود بخشد. اصل اساسی ERT، تصادفی‌سازی شدید در فرایند ساخت درخت است که منجر به ایجاد درختان تصمیم متنوع‌تری می‌شود که به کاهش هم‌بستگی بین درختان کمک کرده و درنهایت، خطر بیش‌برازش را کاهش می‌دهد. مزیت اصلی ERT، کارایی محاسباتی آن‌ها است. این ویژگی آن را به ابزاری قوی و مقاوم برای طیف وسیعی از وظایف یادگیری تحت نظارت تبدیل می‌کند.
- ایدهٔ اصلی پشت SVM [۲۶] این است که یک ابرصفحهٔ بهینه پیدا کند که به‌طور مؤثری طبقهٔ‌های مختلف را در یک مجموعه‌داده جدا کند. ابرصفحهٔ بهینه، نه‌تنها داده‌ها را جدا می‌کند، بلکه فاصله یا حاشیهٔ بین ابرصفحه و نزدیک‌ترین نقاط داده از هر طبقهٔ را نیز حداکثر می‌کند. این نقاط دادهٔ نزدیک، به‌عنوان بردارهای پشتیبان شناخته می‌شوند و نقش حیاتی در تعریف ابرصفحه دارند. با بیشینه‌کردن این حاشیه، مدل می‌تواند تعمیم بهتری ارائه دهد و کمتر در معرض بیش‌برازش به داده‌های جدید و دیده‌نشده باشد. یکی از ویژگی‌های کلیدی SVM، توانایی آن در مدیریت مسائل طبقه‌بندی غیرخطی از طریق یک

[1] Data leakage

روشقدرتمند به نام روش هسته است. توابع هسته به SVM اجازه می‌دهند که به‌طور ضمنی داده‌ها را به یک فضای ویژگی با ابعاد بالاتر نگاشت کنند؛ جایی‌که جداسازی خطی ممکن است، حتی اگر داده‌ها در بعد اصلی خود به‌طور خطی قابل جداسازی نباشند. مدل SVM به‌ویژه در فضاهای با ابعاد بالا مؤثر است و ابزاری مقاوم برای حل مسائل مختلف طبقه‌بندی به‌شمار می‌آیند.

- مدل LR [۲۷] یک الگوریتم پایه‌ای یادگیری ماشین نظارت‌شده است که بیشتر برای وظایف طبقه‌بندی استفاده می‌شود؛ باوجود نامش، هدف آن پیش‌بینی یک نتیجهٔ دسته‌ای یا گسسته به‌جای یک مقدار پیوسته است. این روش به‌ویژه برای طبقه‌بندی دوکلاسه مؤثر است؛ جایی‌که هدف پیش‌بینی یکی از دو نتیجه ممکن است.

از مزایای مدل پیشنهادی ترکیبی می‌توان موارد زیر را برشمرد:

- بهبود دقت پیش‌بینی: از طریق ترکیب هوشمندانهٔ پیش‌بینی‌های مدل‌های پایه که هر یک نقاط قوت متفاوتی دارند؛ برای مثال، الگوریتم CatBoost برای داده‌های طبقه‌بندی‌شده و الگوریتم SVM برای مرزهای غیرخطی مناسب است.
- کاهش بایاس و واریانس در پیش‌بینی‌ها: با استفاده از اعتبارسنجی متقاطع در لایهٔ نخست و میانگین‌گیری از خروجی‌ها، بایاس و واریانس در خروجی مدل کاهش چشم‌گیری دارد.
- حل مشکل بیش‌برازش: با انتخاب یک متامدل خطی ساده همانند LR به‌جای مدل‌های پیچیده‌تر در لایهٔ دوم، مدل پیشنهادی تا حد زیادی مشکل بیش‌برازش را حل می‌کند.
- انعطاف‌پذیری: به‌سادگی می‌توان انواع مختلفی از مدل‌های یادگیری ماشین را به‌عنوان مدل‌های پایه در لایهٔ نخست و نیز به‌عنوان متامدل در لایهٔ دوم به‌کار برد.

از معایب مدل پیشنهادی نیز می‌توان به پیچیدگی محاسباتی بیشتر (نسبت به مدل‌های تشکیل‌دهندهٔ آن) به‌دلیل آموزش چندین الگوریتم پایه و وابستگی به کیفیت الگوریتم‌های تشکیل‌دهندهٔ آن اشاره کرد.

۳-۵-تنظیم پارامتر

ماژول تنظیم پارامتر به‌منظور تنظیم بهینهٔ پارامترهای کنترلی الگوریتم‌های یادگیری درنظر گرفته شده‌است؛ زیرا کارایی مدل‌ها تا حد زیادی به تنظیم بهینهٔ پارامترها وابسته است. بهینه‌سازی پارامترها با استفاده از الگوریتم جست‌وجوی شبکه‌ای تصادفی[1] (RGS) و اعتبارسنجی متقاطع با k=5 در داده‌های آموزشی انجام شده‌است. برای پیاده‌سازی الگوریتم RGS از شیوهٔ RandomizedSearchCV کتابخانهٔ Scikit-Learn پایتون استفاده شد که در هر تکرار، شبکه‌ای از مقادیر پارامترها را تعریف کرده و به‌صورت تصادفی، نمونه‌هایی از مقادیر این شبکه را انتخاب و مورد ارزیابی قرار می‌دهد.

۳-۶-اعتبارسنجی

در این مرحله، مدل آموزش‌دیده بر روی داده‌های آزمون و براساس معیارهای کارایی مختلف مورد ارزیابی قرار می‌گیرد.

۳-۷-پیش‌بینی

پیش‌بینی بر روی مجموعه‌دادهٔ آزمون صورت گرفته و سپس کارایی مدل‌ها بر اساس معیارهای مختلف سنجیده می‌شود. گفتنی است که روش‌های انتخاب ویژگی و تنظیم پارامترهای کنترلی تنها بر روی مجموعه‌دادهٔ آموزشی اعمال شده و مجموعه‌دادهٔ آزمون در کل فرایند توسعهٔ مدل کامل دست‌نخورده باقی می‌ماند. مجموعه‌دادهٔ آزمون تنها برای ارزیابی عملکرد نهایی استفاده می‌شود.

۴-نتایج و بحث

در این بخش، ابتدا مجموعه‌دادهٔ مورد ارزیابی و معیارهای کارایی توضیف می‌شود؛ سپس نتایج آماری کسب‌شده به‌وسیلهٔ الگوریتم‌ها و تحلیل نتایج به‌همراه بحث در مورد آن‌ها پرداخته می‌شود.

۴-۱-معیارهای کارایی

برای ارزیابی مدل‌ها، از پنج معیار کارایی استفاده شده‌است که عبارت‌اند از: صحت (A)، دقت (P)، یادآوری (R)، معیار F1 و ناحیهٔ زیر منحنی ROC (AUC). برای بررسی تعمیم‌پذیری[2] مدل‌ها و جلوگیری از بیش‌برازش، از روش اعتبارسنجی متقابل استفاده شده‌است. توضیحات بیشتر در خصوص معیارهای کارایی در مرجع [۲۸] آمده‌است.

۵-نتایج عددی

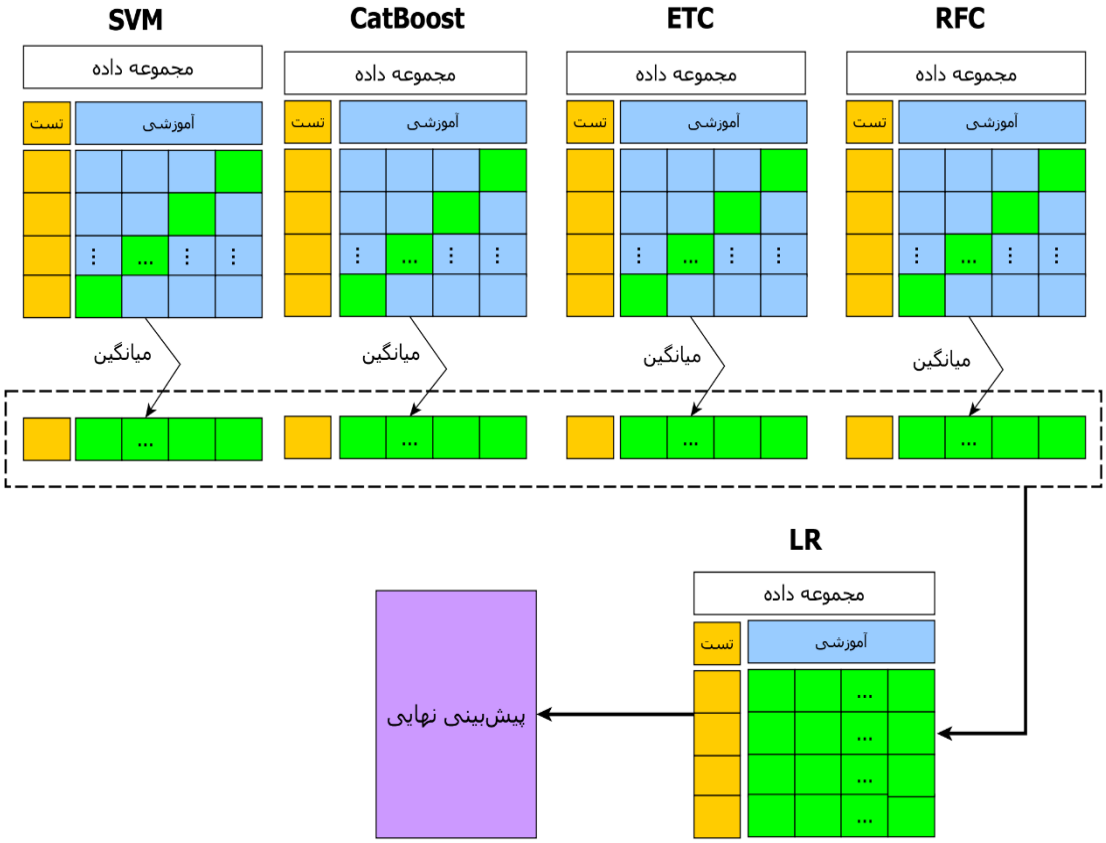
جدول (۲) نتایج به‌دست‌آمده به‌وسیلهٔ مدل‌های پیش‌بینی‌کننده بر روی مجموعه‌دادهٔ آموزشی را نشان می‌دهد. برحسب معیار صحت، مدل‌های CatBoost و EML به‌صورت مشترک بهترین نتایج را در قیاس با سایر مدل‌ها بر روی مجموعه‌دادهٔ آموزشی کسب کرده‌اند. رتبهٔ نخست برحسب معیارهای دقت و F1 به مدل EML و برحسب معیار یادآوری به مدل CatBoost تعلق دارد. رتبهٔ دوم و سوم برحسب معیار دقت به الگوریتم‌های CatBoost و RF تعلق دارد. برحسب معیار یادآوری نیز رتبهٔ دوم به مدل EML و رتبهٔ سوم به مدل ERT اختصاص دارد.

[1] Randomized grid search

[2] Generalization

الگوریتم‌های R و SVM به‌عنوان مدل‌های دسته‌بند شناخته شده و مقدار دقت قابل‌قبولی کسب کرده‌اند، اما بر حسب معیار یادآوری، نتوانسته‌اند چندان موفق عمل کنند. این موضوع موجب شده‌است تا برحسب معیار F1، کارایی متوسطی کسب کنند که در قیاس با الگوریتم‌های برتر، نتیجهٔ مطلوبی محسوب نمی‌شود. اختلاف بین مدل پیشنهادی EML با مدل Catboost برحسب معیارهای یادآوری و F1 بر روی مجموعه‌دادهٔ آموزشی ناچیز است، اما نتایج جدول (۳) نشان می‌دهد که مدل EML بر روی مجموعه‌دادهٔ آزمون، فاصلهٔ قابل‌توجهی با مدل CastBoost دارد و معیارهای کارایی را به‌نحو چشم‌گیری بهبود داده‌است. مدل RF عملکرد مطلوبی را از نظر معیارهای کارایی بر روی مجموعه‌دادهٔ آموزشی و آزمون کسب کرده‌است. بر روی مجموعه‌دادهٔ آزمون، رتبهٔ دوم از نظر عملکرد کلی به مدل RF تعلق دارد؛ این روش، تعادل مناسبی بین معیارهای دقت و یادآوری از خود نشان داده‌است. مدل SVM در آموزش به مقدار صحت ۰.۸۸۲۲ رسیده‌است، اما در مجموعه‌دادهٔ آزمون این مقدار به ۰.۸۵۷۱ کاهش یافته است که قابلیت اعتماد مدل را کاهش می‌دهد. مدل ETR در کنار مدل‌های LR و CatBoost بر روی مجموعه‌داده‌های آموزشی و آزمون عملکرد قابل‌قبولی کسب کرده‌است و تعادل مناسبی بین دقت و یادآوری فراهم کرده که برای شناسایی گلیوما مؤثر است. مدل Catboost و LR عملکرد پایداری، هم در فاز آموزش و هم در مرحلهٔ آزمون نشان دادند؛ این مدل‌ها اگرچه بهترین کارایی را کسب نکرده‌اند، اما به‌طور مؤثر مقادیر دقت و یادآوری را متعادل کرده و برای درجه‌بندی گلیوما مناسب‌اند.

.



(شکل-۴): ساختار مدل پیشنهادی EML

(Figure-4): The structure of the proposed EML model

با تحلیل نتایج مدل‌ها درمی‌یابیم که مدل پیشنهادی EML بر روی مجموعه‌دادهٔ آزمون، بهترین کارایی را برحسب تمامی معیارهای کارایی کسب کرده‌است که منجر به توانایی بهتر در شناسایی صحیح تومورها (کاهش منفی‌های کاذب) و حفظ تعادل مناسب بین دقت و یادآوری می‌شود. در فاز آزمون، برحسب معیار دقت مدل LR رتبهٔ دوم را کسب کرده‌است. مدل پیشنهادی EML بهترین انتخاب از منظر تشخیص پزشکی است؛ زیرا نرخ دقت، یادآوری و همچنین صحت بالایی دارد. در بررسی وضعیت تومور، اگر تومور تشخیص داده نشود خطر جانی و تبعات جبران‌ناپذیری دارد. معیار دقت نیز بسیار مهم است، اما اشتباه در تشخیص تومور در بیمار سالم (مثبت کاذب) در قیاس با تشخیص‌ندادن تومور، قابل‌تحمل است.

(جدول-۲): نتایج آماری به‌دست‌آمده به‌وسیلهٔ مدل‌های پیش‌بینی‌کننده بر روی مجموعه‌دادهٔ آموزشی

(Table-2): Results obtained by predictive models on training dataset

مدل	صحت	دقت	یادآوری	F1
CatBoost	۹۵۲۳.۰	۹۲۹۱.۰	۹۵۹۳.۰	۹۴۴۰.۰
RF	۹۰۱۶.۰	۹۰۴۴.۰	۹۰۱۶.۰	۹۰۲۰.۰
ERT	۸۸۴۹.۰	۸۱۲۰.۰	۹۳۱۴.۰	۸۶۷۶.۰
SVM	۸۸۲۲.۰	۸۹۶۲.۰	۸۸۲۲.۰	۸۸۹۰.۰
LR	۸۷۹۲.۰	۸۸۶۳.۰	۸۸۰۹.۰	۸۸۱۶.۰
EML	۹۵۲۳.۰	۹۴۰۵.۰	۹۴۸۰.۰	۹۴۴۲.۰

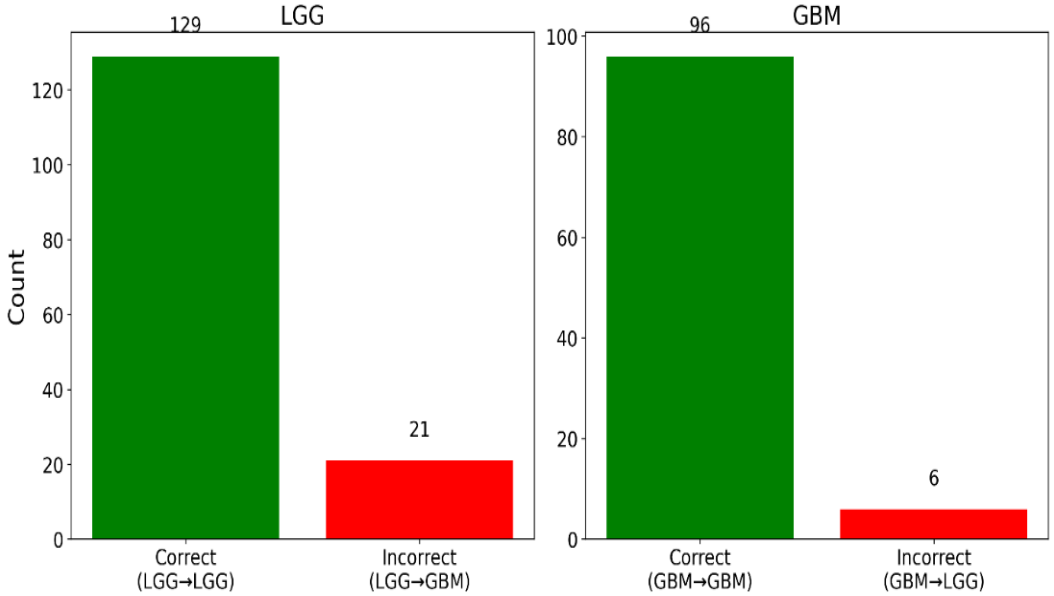
از طرفی مدل LR نیز عملکرد مطلوبی دارد؛ یادآوری (۰.۹۴۱۲) و نرخ دقت مطلوبی (۰.۸۰۶۷) دارد که منجر به کاهش مثبت کاذب می‌شود و آن را به جایگزینی قوی برای محیط‌های بالینی که دقت مدل ارزشمند است، تبدیل می‌کند. مدل SVM باوجود یادآوری بالا (۰.۹۰۲۰)، به‌دلیل دقت پایین (۰.۷۷۹۷) نامناسب است؛ زیرا در تشخیص تورمورها ضعیف عمل می‌کند. مدل‌های درختی شامل CatBoost، RF و ERT عملکرد قابل‌قبولی دارند، درحالی که RF کمی بهتر عمل کرده‌است. به‌اختصار، اگر شناسایی بیشینهٔ تومورها (یادآوری بالا)، دقت و تفسیرپذیری اولویت باشد می‌توان با قاطعیت مدل EML را به‌عنوان بهترین مدل مطرح کرد؛ همچنین در جایگاه بعدی، مدل‌های درختی CatBoost، RF، ERT و همچنین مدل LR گزینهٔ جایگزین خواهند بود. به‌نظر می‌رسد استفادهٔ هم‌زمان از مدل‌ها و ترکیب هوشمندانهٔ آن‌ها همانند آنچه در مدل EML رخ داده‌است، در تشخیص و درجه‌بندی تومورها بسیار مفید است.

(جدول-۳): نتایج آماری به‌دست‌آمده به‌وسیلهٔ مدل‌های پیش‌بینی‌کننده بر روی مجموعه‌دادهٔ آزمون

(Table-3): Results obtained by predictive models on testing dataset

مدل	صحت	دقت	یادآوری	F1
CatBoost	۸۸۱۰.۰	۸۰۵۱.۰	۹۳۱۴.۰	۸۶۳۶.۰
RF	۸۸۸۹.۰	۸۱۳۶.۰	۹۴۱۲.۰	۸۷۲۷.۰
ERT	۸۸۱۰.۰	۸۰۰۰.۰	۹۴۱۲.۰	۸۶۴۹.۰
SVM	۸۵۷۱.۰	۷۷۹۷.۰	۹۰۲۰.۰	۸۳۶۴.۰
LR	۸۸۴۹.۰	۸۰۶۷.۰	۹۴۱۲.۰	۸۶۸۸.۰
EML	۸۹۲۹.۰	۸۲۰۵.۰	۹۴۱۲.۰	۸۷۶۷.۰

شکل (۵) نمودار خطاهای پیش‌بینی طبقهٔ‌ها برای مدل پیشنهادی را بر روی مجموعه‌دادهٔ آزمون نشان می‌دهد. محور افقی بیان‌کنندهٔ طبقهٔ‌های واقعی است که در اینجا شامل طبقهٔ‌های LGG و GBM می‌شود. محور عمودی تعداد پیش‌بینی‌های اشتباه (یا صحیح) برای هر طبقه را نشان می‌دهد. مدل در پیش‌بینی طبقهٔ GBM به‌طور واضح درصد بالاتری از پیش‌بینی‌های درست را به‌خود اختصاص داده و این نشان‌دهندهٔ عملکرد بهتر مدل در این طبقه است. در مورد طبقهٔ LGG به‌نظر می‌رسد که مدل دقت کمتری داشته باشد و نیاز به بهبود عملکرد در شناسایی این طبقه وجود دارد.

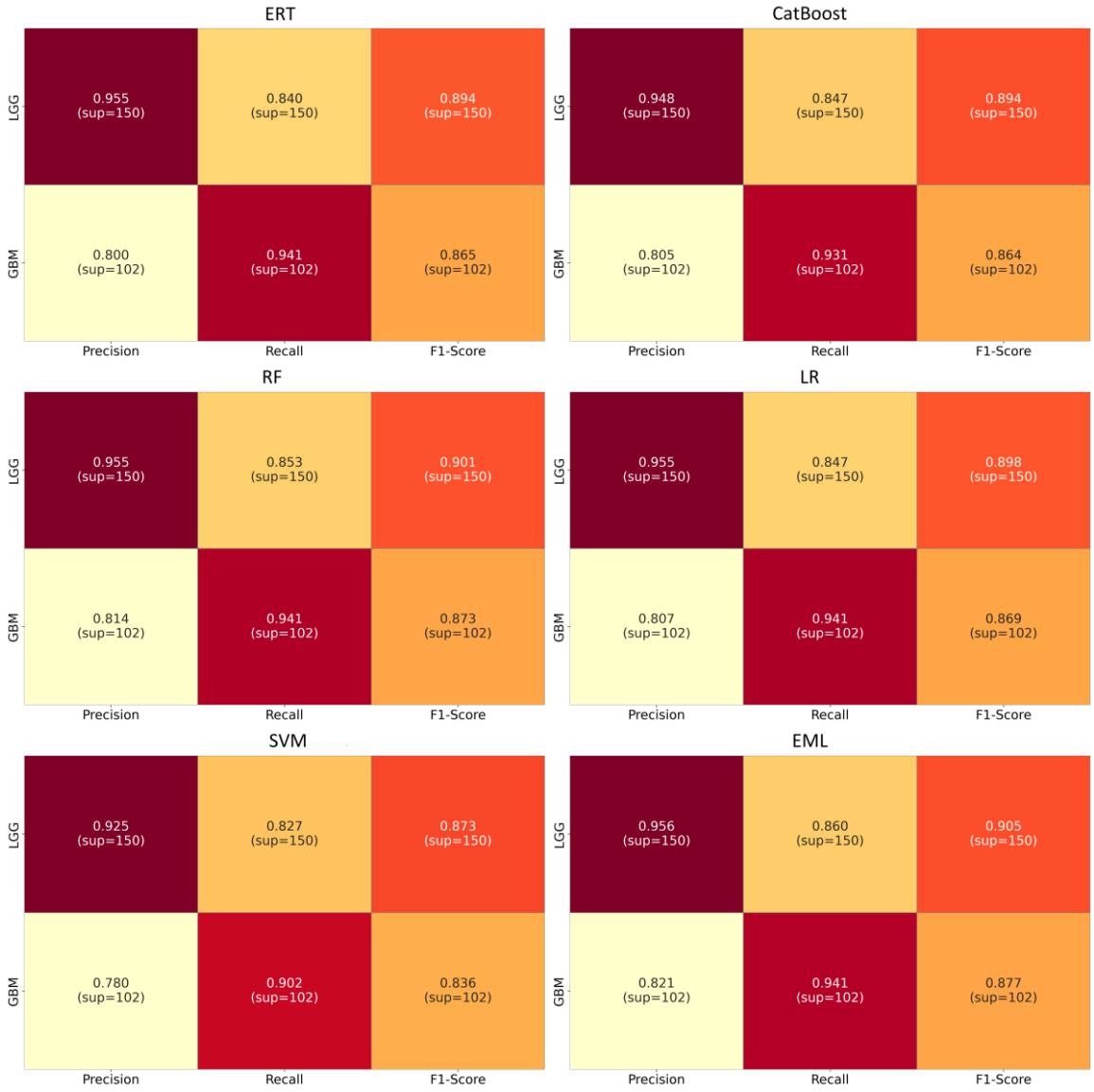


(شکل-۵): خطای پیش‌بینی به‌وسیلهٔ مدل پیشنهادی برحسب درجهٔ تومور بر روی مجموعه‌دادهٔ آزمون

(Figure-5): Model prediction errors (by tumor grade) obtained by the proposed model on the testing dataset

شکل (۶) عملکرد مدل‌ها را در تفکیک طبقه‌ها برحسب معیارهای دقت، یادآوری، F1 و معیار پشتیبان (sup) نشان می‌دهد. نتایج عملکرد مطلوب مدل را در تمایز بین دو طبقهٔ LGG و GBM نشان می‌دهد. از نظر معیار دقت، مدل

EML در طبقه‌بندی طبقهٔ LGG در قیاس با طبقهٔ GBM دقت بالاتری دارد؛ بنابراین امکان دارد در طبقهٔ GBM موارد مثبت کاذب ایجاد شوند. مدل در تشخیص طبقهٔ GBM یادآوری بالا، اما دقت پایینی کسب کرده‌است. این موضوع تأیید می‌کند که مدل در تشخیص موارد GBM بهتر عمل کرده، اما با دقت موارد تشخیص‌داده‌شده پایین‌تر که ناشی از نامتوازنی داده‌ها است. از منظر معیار F1، مدل برای هر دو طبقه امتیازات قابل‌قبولی کسب کرده‌است (نتایج بهتر برای طبقهٔ LGG) که نشان‌دهندهٔ تعادل خوب بین دقت و یادآوری برای هر طبقه است. مقادیر پشتیبان نشان می‌دهد که یک نامتوازنی در تعداد داده‌های هر طبقه وجود دارد (۱۰۲ نمونه از طبقهٔ GBM در مقابل ۱۵۰ مورد از طبقهٔ LGG). این موضوع می‌تواند بر عملکرد مدل‌ها تأثیر بگذارد، به‌ویژه بر دقت طبقهٔ اکثریت (طبقهٔ LGG). به‌منظور ارزیابی بهتر مدل پیشنهادی و مقابله با نامتوازن‌بودن طبقه‌ها، در کار آینده در نظر است از روش‌های بازنمونه‌گیری استفاده شود.

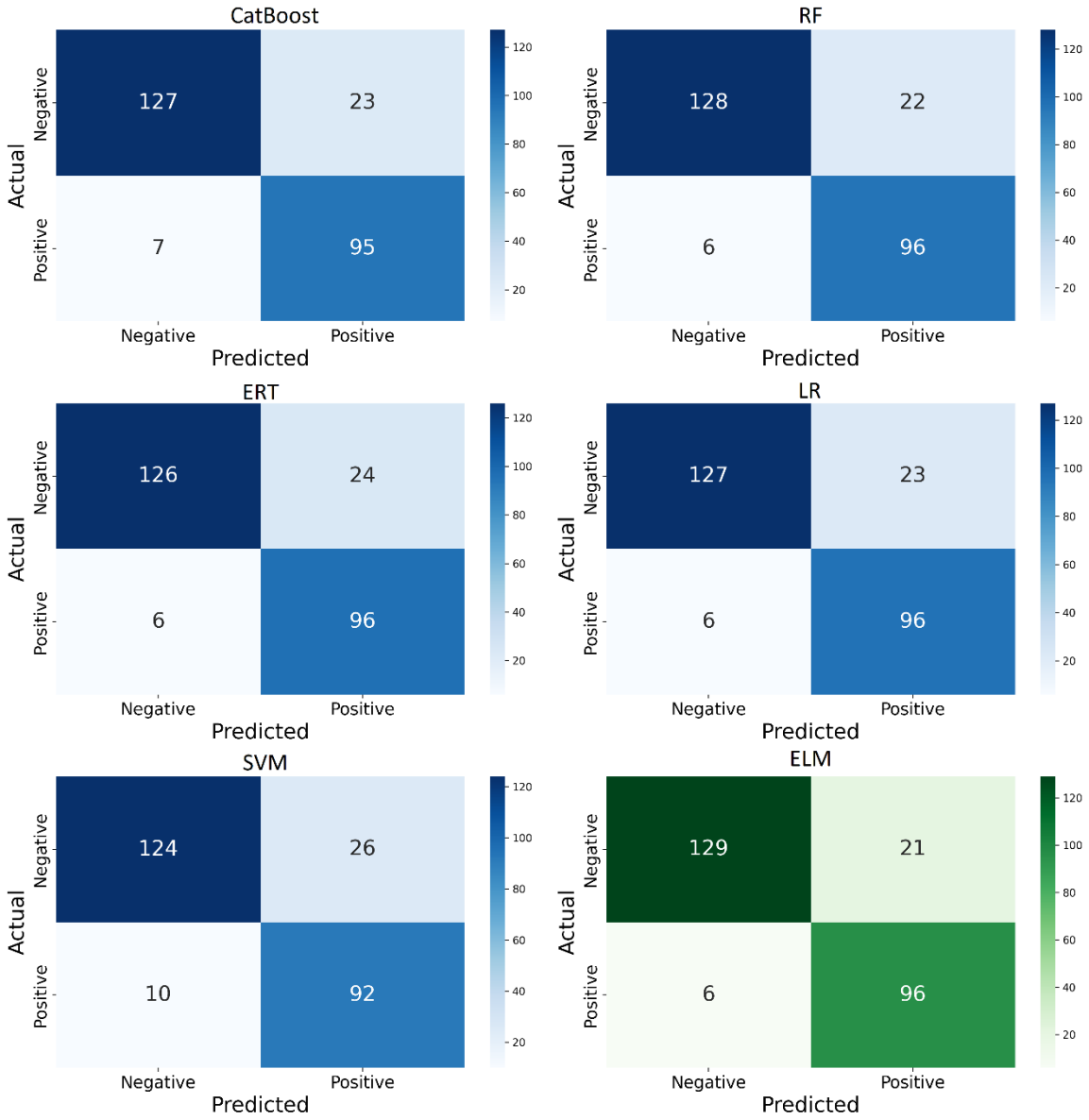


(شکل-۶): کارایی درجه‌بندی مدل‌ها بر روی طبقه‌های LGG و GBM بر روی مجموعه‌دادهٔ آزمون

(Figure-6): Classification performance of models on LGG and GBM classes on the test dataset

شکل (۷) ماتریس آشفتگی[1] برچسب‌های پیش‌بینی‌شده را در برابر برچسب‌های واقعی نشان می‌دهد. ماتریس نشان می‌دهد که مدل‌های یادگیری تمایز بین دو طبقه را به‌خوبی انجام داده‌اند و توازن مطلوبی بین حساسیت[2] (۰.۹۴۱۲) و صراحت[3] (۰.۸۶) نشان می‌دهد، به‌ویژه در شناسایی صحیح نمونه‌های طبقهٔ GBM به‌خوبی عمل کرده‌است. گفتنی است که در تشخیص گلیوما و سایر بیماری‌های حاد، بیشتر به‌دنبال مدلی هستیم که دارای حساسیت بسیار بالایی باشد؛ به‌عبارت دیگر، ترجیح داده می‌شود کمی تعداد مثبت کاذب (FP) پذیرفته تا اطمینان حاصل شود هیچ منفی کاذب (FN) و درنتیجه هیچ بیمار واقعی از دست نرود.

[1] Confusion matrix
[2] Sensitivity
[3] Specificity



(شکل-۷): ماتریس آشفتگی تولیدشده به‌وسیلهٔ مدل‌ها بر روی مجموعه‌دادهٔ آزمون

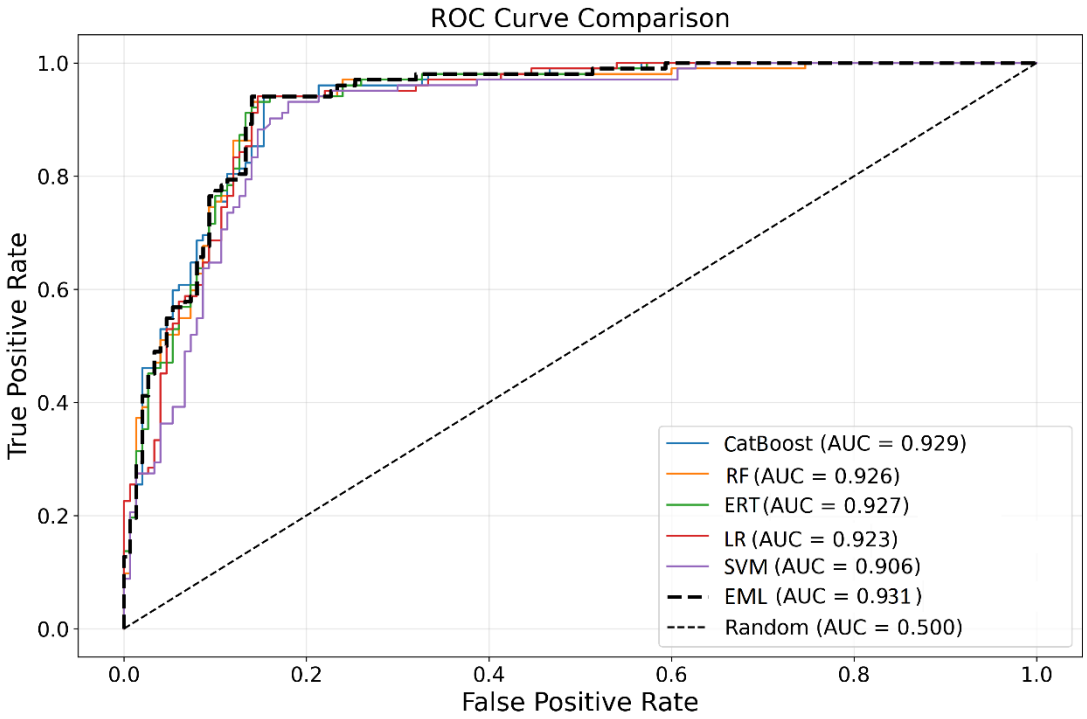
(Figure-7): The confusion matrix generated by the models on the testing dataset

مقدار حساسیت کسب‌شده به‌وسیلهٔ مدل پیشنهادی مبین کاربردپذیری آن در محیط واقعی است؛ همچنین، مدل پیشنهادی در کاهش مثبت کاذب و منفی کاذب بسیار موفق بوده که موجب بهبود کارایی آن در مقایسه با سایر مدل‌ها شده و مدل SVM ضعیف‌ترین نتایج را تولید کرده‌است. سایر مدل‌ها نیز نتایج قابل‌قبولی کسب کرده‌اند.

شکل (۸) منحنی مشخصهٔ عملکرد (ROC) را برای مدل‌ها نشان می‌دهد. منحنی نشان می‌دهد که مدل پیشنهادی با کسب AUC=0.931 توانسته است بین دو طبقهٔ تومور LGG و GBM، با نرخ مطلوب مثبت واقعی (TP) و نرخ پایین مثبت کاذب (FP)، تمایز مؤثر قائل شود. نتایج نشان می‌دهد برای کارهایی که نیاز به طبقه‌بندی دو طبقه دارند، مدل پیشنهادی انتخاب قابل اعتمادی است؛ همچنین مدل‌های CatBoost (۰.۹۲۹)، RF (۰.۹۲۶) و ERT (۰.۹۲۷) نیز نتایج بسیار نزدیکی به مدل پیشنهادی ارائه کرده‌اند. مدل‌های LR (۰.۹۲۳) و SVM (۰.۹۰۶) عملکرد خوبی دارند، اما نسبت به مدل‌های برتر ضعیف‌ترند. تمام مدل‌ها (به‌جز مدل تصادفی با AUC=0.5) با مقدار AUC بالای ۰.۹، قدرت تشخیص عالی را نشان می‌دهند که بیان‌کنندهٔ توانایی بالای همهٔ مدل‌ها در تمایز بین طبقه‌های مثبت و منفی است. تفاوت ناچیز بین مدل‌های برتر پیشنهاد می‌کند که عوامل دیگری مانند سرعت اجرا یا سادگی مدل نیز در انتخاب نهایی باید درنظر گرفته شوند.

شکل (۹) منحنی دقت-یادآوری را برای مدل‌های یادگیری نشان می‌دهد. بیشتر مدل‌ها به‌ویژه مدل پیشنهادی عملکرد مطلوبی را نشان داده‌اند و توازن مناسبی بین دقت و یادآوری برقرار کردند. مدل پیشنهادی با کسب دقت متوسط (AP) ۰.۸۸۷ عملکرد خوبی را در پیش‌بینی

طبقه‌های مثبت و حفظ نرخ یادآوری فراهم کرده‌است. منحنی نوسانات کمی را نشان می‌دهد، که نشان‌دهندهٔ توانایی طبقه‌بند برای حفظ دقت است؛ زیرا مثبت واقعی (TP) بیشتری را ثبت می‌کند. مدل‌های ET و RF با کسب دقت متوسط ۰.۸۷۸ به‌صورت مشترک در جایگاه سوم از نظر کارایی قرار می‌گیرند. مدل LR با دقت متوسط ۰.۸۷۳ عملکرد به‌نسبت خوبی دارد. مدل SVM با دقت ۰.۸۴۲ ضعیف‌ترین مدل است، به‌ویژه که در یادآوری بالاتر اُفت قابل‌توجهی در میزان دقت نشان می‌دهد.



(شکل-۸): منحنی ROC مدل‌ها بر روی مجموعه‌دادهٔ آزمون
(Figure-8): ROC curve of models on the testing dataset

شکل (۹) منحنی دقت-یادآوری را برای مدل‌های یادگیری نشان می‌دهد. بیشتر مدل‌ها به‌ویژه مدل پیشنهادی عملکرد مطلوبی را نشان داده‌اند و توازن مناسبی بین دقت و یادآوری برقرار کردند. مدل پیشنهادی با کسب دقت متوسط (AP) ۰.۸۸۷ عملکرد خوبی را در پیش‌بینی طبقه‌های مثبت و حفظ نرخ یادآوری فراهم کرده‌است. منحنی نوسانات کمی را نشان می‌دهد، که نشان‌دهندهٔ توانایی طبقه‌بند برای حفظ دقت است؛ زیرا مثبت واقعی (TP) بیشتری را ثبت می‌کند. مدل‌های ET و RF با کسب دقت متوسط ۰.۸۷۸ به‌صورت مشترک در جایگاه سوم از نظر کارایی قرار می‌گیرند. مدل LR با دقت متوسط ۰.۸۷۳ عملکرد به‌نسبت خوبی دارد. مدل SVM با دقت ۰.۸۴۲ ضعیف‌ترین مدل است، به‌ویژه که در یادآوری بالاتر اُفت قابل‌توجهی در میزان دقت نشان می‌دهد.

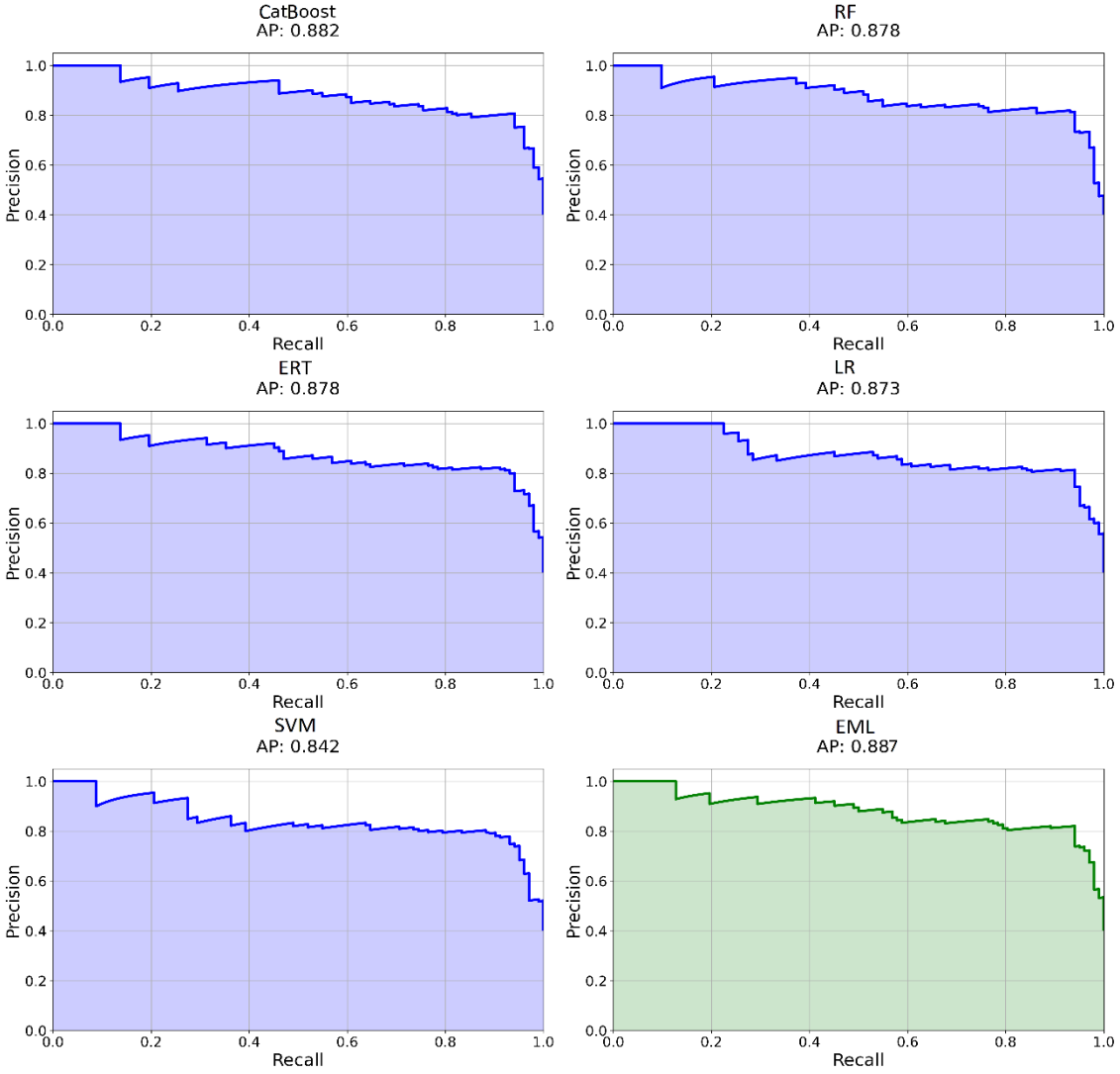
شکل (۱۰) منحنی‌های کالیبراسیون (Calibration Curves) برای شش مدل مختلف را نشان می‌دهد. این منحنی‌ها ابزاری مفید برای ارزیابی کالیبره‌بودن پیش‌بینی‌های احتمالاتی یک مدل هستند که میزان تطابق احتمالات پیش‌بینی‌شده با مشاهدات را ارزیابی می‌کند که برای ارزیابی ریسک بالینی قابل اعتماد اهمیت زیادی دارد.

در تحلیل کالیبراسیون به این سؤال پاسخ می‌دهیم که وقتی یک مدل می‌گوید احتمال رخداد تومور GBM برابر۸۰درصد است، آیا درواقع ۸۰درصد از دفعاتی که این پیش‌بینی را می‌کند، تومور از نوع GBM است یا خیر. نمودارها نشان می‌دهند که همهٔ مدل‌ها، کالیبراسیون مطلوبی با امتیاز Brier از بازه ۰.۸۳۴۲ تا ۰.۹۶۳۸ کسب کرده‌اند، که در آن امتیازهای پایین‌تر نشان‌دهندهٔ کالیبراسیون دقیق‌تر است.

با بررسی نمودارها می‌توان دریافت که مدل پیشنهادی EML با امتیاز ۰.۸۳۴۲، بهترین کالیبراسیون را به‌دست آورده‌است و پس از آن با اختلاف کمی مدل‌های CatBoost (۰.۸۴۹۷) و RF (۰.۸۹۲۵) قرار گرفتند. ضعیف‌ترین نتایج نیز متعلق به مدل SVM با امتیاز ۰.۹۶۳۸ است.

همان‌گونه که پیش‌تر در شکل (۳) نشان داده شد، نمودار SHAP نشان می‌دهد که ویژگی‌های IDH1، TP53،

Age و ATRX تأثیرگذارترین پارامترها در پیش‌بینی گلیوما در مجموعه‌دادهٔ مورد بررسی هستند. شکل (۱۱) (الف) نمودار SHAP برای بررسی تأثیر ویژگی‌ها در درجه‌بندی بیماران در دستهٔ LGG و شکل (۱۱) (ب) تأثیر ویژگی‌ها در درجه‌بندی به‌عنوان GBM نشان می‌دهد. این شکل‌ها نشان می‌دهد که چگونه ترکیب ویژگی‌ها بر تصمیمات دسته‌بندی برای هر بیمار تأثیر می‌گذارند و تأثیر شفافی از هر ویژگی را برای پیش‌بینی در اختیار پزشکان قرار می‌دهد. مقادیر مثبت و منفی SHAP برای هر ویژگی نشان‌دهندهٔ افزایش یا کاهش احتمال تشخیص هستند. برای مورد LGG، پیش‌بینی مدل به میزان چشم‌گیری به ویژگی‌های IDH1 و سن کوچک‌تر وابسته است. سن بالا احتمال تشخیص LGG را کاهش می‌دهد. در مورد GBM ویژگی‌های تأثیرگذار، مقادیر بزرگتر IDH1، PTEN، TP53 و سن بالاتر هستند.



(شکل-۹): منحنی دقت-یادآوری تولیدشده به‌وسیلهٔ مدل‌ها بر روی مجموعه‌دادهٔ آزمون

(Figure-9): Precision-recall curve generated by predictive models on the test dataset

پیچیدگی محاسباتی مدل EML به‌طور تقریبی برابر مجموع پیچیدگی سایر مدل‌هاست؛ یعنی:

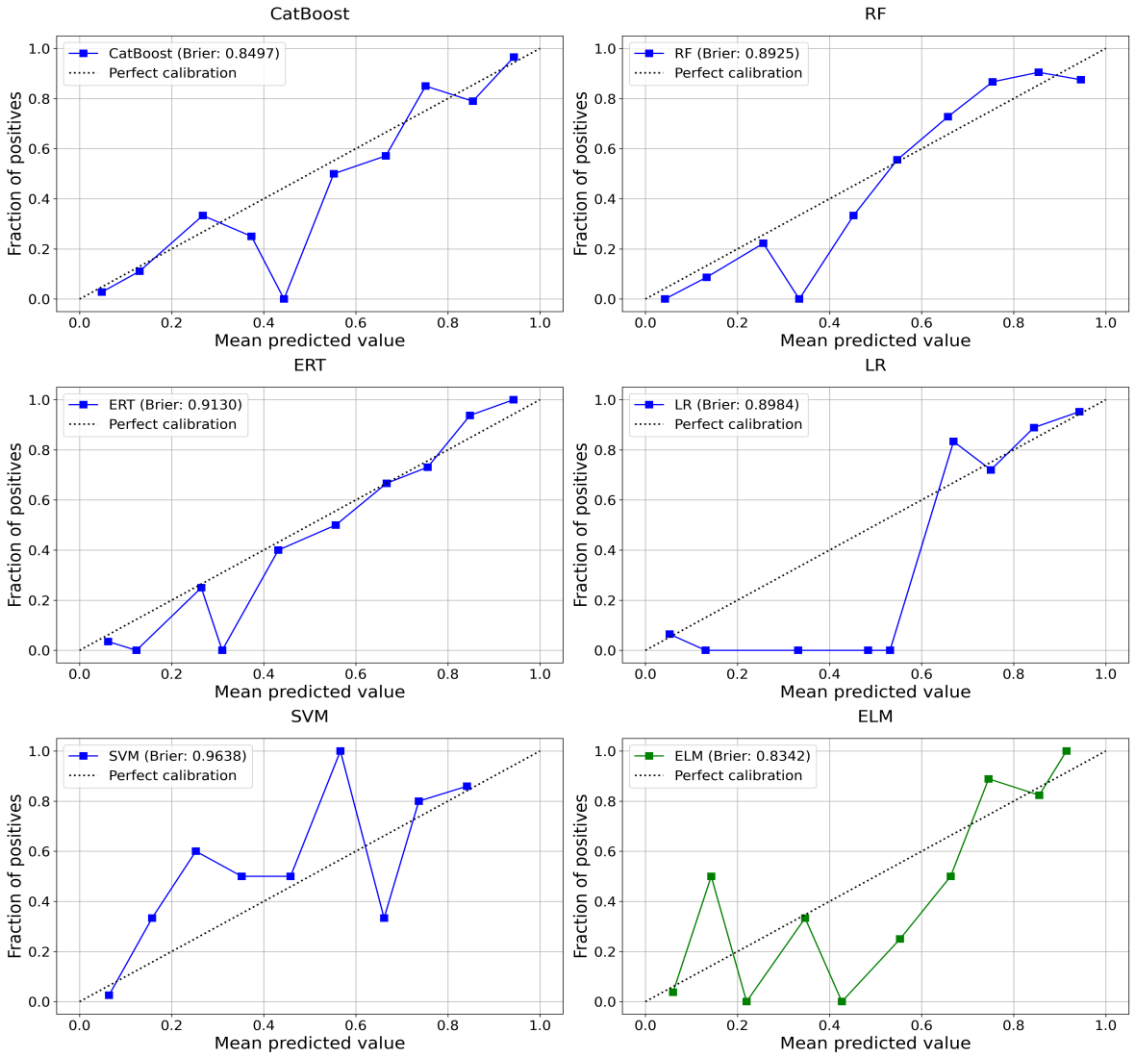
$O(EML) \approx O(SVM + CatBoost + RF + ERT + LR)$

جدول (۴) پیچیدگی محاسباتی مدل پیشنهادی و اجزای تشکیل‌دهندهٔ آن را نشان می‌دهد.

در جدول (۴) مفهوم متغیرها به‌شرح زیر است:

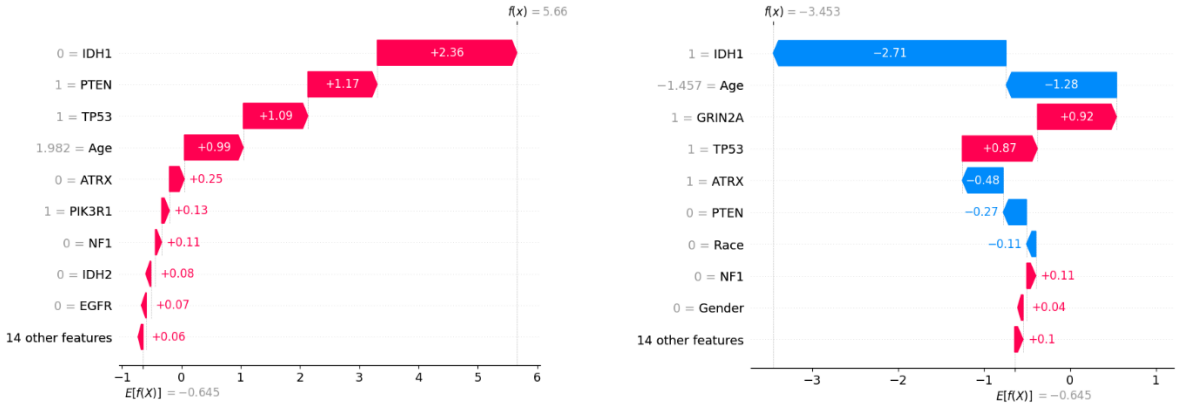
- n: تعداد نمونه‌ها
- d: تعداد ویژگی‌ها
- t: تعداد درختان
- i: تعداد تکرارها
- nodes: تعداد کل گره‌های درختان
- n_sv: تعداد بردارهای پشتیبان
- log(n): عمق درخت

مدل EML پیچیدگی مکانی و زمانی بالاتری نسبت به سایر مدل‌های پایه دارد؛ اما در مقابل، دقت و پایداری بسیار بهتری کسب کرده‌است. برای یک مسئلهٔ حیاتی همانند درجه‌بندی تومور مغزی که دقت و قابلیت اطمینان از اولویت برخوردار است، استفاده از مدل EML به‌طور کامل توجیه‌پذیر است.



(شکل-۱۰): منحنی‌های کالیبراسیون تولیدشده به‌وسیلهٔ مدل‌ها بر روی مجموعه‌دادهٔ آزمون

(Figure-10): Calibration curves generated by the predictive models on the test dataset



(ب) تأثیر ویژگی‌ها در درجه‌بندی بیماران در دستهٔ GBM

(الف) تأثیر ویژگی‌ها در درجه‌بندی بیماران در دستهٔ LGG

(شکل-۱۱): تأثیر ویژگی‌ها در درجه‌بندی گلیوما در تحلیل SHAP

(Figure-11): The feature importance score on glioma grading in the SHAP

به‌منظور بررسی معناداری آماری برتری مدل پیشنهادی، آزمون فریدمن به‌همراه آزمون پس‌هولم (با سطح معناداری ۰/۰۵) بر روی نتایج حاصل از ۵- fold cross-validation برای مدل پیشنهادی EML و مدل‌های پایه (SVM, CatBoost, RF, ERT, LR) که در این پژوهش پیاده‌سازی و ارزیابی شدند، انجام گرفت. نتایج این آزمون که در جدول (۵) ارائه شده‌است، نشان می‌دهد که اختلاف عملکرد مدل EML با تمامی مدل‌های پایه از نظر آماری معنادار است (p-value < 0.05)

(جدول-۴): مقایسهٔ پیچیدگی محاسباتی مدل پیشنهادی EML و سایر مدل‌ها

(Table-4): Comparison of computational complexity of the proposed EML model and its constituent models

مدل	پیچیدگی زمانی (آموزش)	پیچیدگی زمانی (استنتاج)	پیچیدگی مکانی
SVM	$O(n^3 \times d)$	$O(s \times d)$	$O(s \times d)$
CatBoost	$O(t \times n \times d \times \log(n))$	$O(t \times \log(n))$	$O(t \times nodes)$
RFR	$O(t \times n \times d \times \log(n))$	$O(t \times \log(n))$	$O(t \times nodes)$
ERT	$O(t \times n \times d \times \log(n))$	$O(t \times \log(n))$	$O(t \times nodes)$
LR	$O(i \times n \times d)$	$O(d)$	$O(d)$
EML	$\approx O(t \times n \times d \times \log(n) + n^3 \times d)$	$\approx O(4 \times t \times \log(n) + d)$	$\approx O(3 \times t \times nodes + s \times d)$

(جدول-۵): نتیجهٔ آزمون پس‌هولم بر اساس معیار Accuracy برای مدل پیشنهادی و مدل‌های پایه بر روی مجموعه‌دادهٔ (Table-5):

Results of Holm's Contr

جفت مدل‌های مقایسه‌شده	p-value	معناداری(α=0.05)
EML vs. SVM	۰.۰۰۳	دارد
EML vs. LR	۰.۰۱۸	دارد
EML vs. RF	۰.۰۲۶	دارد
EML vs. CatBoost	۰.۰۱۱	دارد

نتایج آزمون Friedman که در جدول (۶) آمده‌است، نشان می‌دهد که مدل پیشنهادی EML تفاوت آماری بسیار معناداری با سایر مدل‌های تشکیل‌دهندهٔ آن دارد.

(جدول-۶): نتایج آزمون Friedman برای مدل پیشنهادی و مدل‌های پایه بر روی مجموعه‌دادهٔ (Table-6): Friedman test results for the proposed model and the baseline models on the dataset

مدل	رتبهٔ میانگین
EML	۱.۲
CatBoost	۲.۶
RF	۳.۰
LR	۳.۲
SVM	۵.۰

۵-۱-مقایسه با روش‌های موجود

مدل پیشنهادی EML با چندین مدل مطرح و جدید بر ساس معیارهای کارایی مقایسه شده‌است. نتایج مقایسه برحسب معیارهای کارایی در جدول (۷) آمده‌است. تحلیل نتایج برتری مدل EML را در قیاس با سایر روش‌های همتا نشان می‌دهد. باتوجه به نتایج، مدل پیشنهادی در معیار صحت نسبت به دیگر مدل‌ها برتری دارد که بیان‌کنندهٔ این است که درصد پیش‌بینی‌های صحیح این مدل به‌طور قابل‌توجهی بیشتر از سایر همتایان است. این مدل در قیاس با سایر مدل‌ها، بالاترین کارایی را از نظر یادآوری کسب کرده‌است. این موضوع نشان می‌دهد که مدل EML در شناسایی در بالاترین میزان تومورهای گلیوما به‌صورت موفق عمل کرده‌است که منجر به کاهش تبعات تشخیص‌ندادن تومور می‌شود. مدل EML توانسته است نمونه‌های مثبت بیشتری را شناسایی کند و تعداد بسیار کمی از منفی‌های کاذب را گزارش دهد. مدل SMOTE_SVM در مرجع [۱۱] از نظر معیارهای دقت و F1 بهتر از سایرین عمل کرده‌است. گفتنی است که در مرجع [۱۱] از روش SMOTE برای ایجاد توازن بین تعداد رکوردها در دو طبقهٔ LGG و GBM استفاده شده‌است. برای انجام این کار داده‌های مصنوعی تولید می‌شود که ممکن است ساختار طبیعی داده‌ها را از بین می‌برد. در روش SMOTE نمونه‌داده‌ها به‌صورت خطی بین دو نقطه تولید می‌شوند. این کار ساختار طبیعی و پیچیدهٔ داده‌ها را در طبقهٔ اقلیت نادیده می‌گیرد؛ به‌خصوص اگر مرز تصمیم‌گیری غیرخطی باشد. در برخی موارد، این موضوع می‌تواند منجر به تولید نمونه‌هایی شود که در واقعیت منطقی نیستند و منجر به دقت کاذب شود. بدیهی است که معیار دقت بسیار مهم است، اما دربارهٔ مدل پیشنهادی EML می‌توان این‌گونه بیان کرد شناسایی بیشترین میزان بیماران (یادآوری بالا) مفید است؛ زیرا مدل حتی اگر به اشتباه در بیمار سالم تومور تشخیص دهد (مثبت کاذب) در قیاس با تشخیص‌ندادن تومور، قابل تحمل‌تر است.

به‌عنوان یکی از کارهای آینده درنظر است تا از نسخه‌های پیشرفته SMOTE و روش‌های بازنمونه‌گیری در رویکرد پیشنهادی استفاده شود تا کارایی مدل با حفظ ساختار داده‌های اصلی هرچه بیشتر بهبود یابد. گفتنی است که به‌دلیل در دسترس‌نبودن داده‌های خام و توزیع نتایج مدل‌های ارائه‌شده در مطالعات دیگر [۷, ۱۱, ۱۲]، امکان انجام آزمون آماری مستقیم با آن مدل‌ها میسر نیست و مقایسهٔ با آن‌ها براساس مقادیر گزارش‌شده در مقالات یادشده صورت گرفته است.

(جدول-۵): مقایسهٔ مدل پیشنهادی EML با سایر همتایان مطرح بر روی مجموعه‌دادهٔ آزمون

(Table-5): Comparison of the proposed EML model with state-of-the-art models on the testing dataset

مدل	مرجع	صحت	دقت	یادآوری	F1
XGBoost	[12]	۰.۸۶۳	۰.۸۷۱	۰.۸۶۳۰	۰.۸۶۷
RF	[12]	۰.۸۴۵	۰.۸۵۲	۰.۸۴۵	۰.۸۴۸
LR+SVM+RF	[7]	۰.۸۷۴	۰.۸۰۸	۰.۹۱۴	۰.۸۵۵
LR+SVM+AdaBoost	[7]	۰.۸۷	۰.۸۰۲	۰.۹۱۵	۰.۸۵۲
SMOTE_SVM	[11]	۰.۸۸۲	۰.۸۸۶	۰.۸۸۲	۰.۸۸۲
SMOTE_CatBoost	[11]	۰.۸۸۱	۰.۸۸۵	۰.۸۸۱	۰.۸۸۱
RUS_CatBoost	[11]	۰.۸۸۱	۰.۸۴۷	۰.۹۲۹	۰.۸۸۶
RUS_SVM	[11]	۰.۸۷۹	۰.۸۴۵	۰.۹۲۹	۰.۸۸۵
EML	--	۰.۸۹۳	۰.۸۲۵	۰.۹۴۱	۰.۸۷۷

۶-نتیجه‌گیری

در این مقاله، روش یادگیری ماشین ترکیبی برای حل مسئلهٔ درجه‌بندی تومورهای مغزی گلیوما ارائه شد. مدل پیشنهادی برای بهبود معیارهای کارایی، از ترکیب هوشمندانهٔ مدل‌های یادگیری ماشین در دو سطح بهره برده‌است. در سطح نخست، چهار مدل قوی یادگیری شامل جنگل تصادفی، ماشین بردار پشتیبان، درخت‌های به‌شدت تصادفی و دسته‌بندی تقویتی به‌صورت موازی بر روی مجموعه‌دادهٔ آموزشی اجرا شد و خروجی آن‌ها به مدل رگرسیون ترابری در دستهٔ دوم خورانده می‌شود تا خروجی نهایی تولید شود. این ترکیب، تا حد زیادی مشکل بیش‌برازش را حل کرده و کارایی درجه‌بندی گلیوما را بهبود داده‌است. نتایج ارزیابی بر روی مجموعه‌دادهٔ آزمون نشان داد که مدل پیشنهادی ترکیبی EML با کسب صحت ۰.۸۹۳ کارایی روش‌های لبهٔ دانشی در ادبیات موضوع را بهبود داده‌است؛ همچنین این مدل، در فاز آموزش و آزمون، پایداری مطلوبی در عملکرد دارد که می‌تواند برای کاربردهای واقعی، درک زیست‌شناسی تومور و شناسایی اهداف دارویی مفید باشد. در آزمایش‌ها مشخص شد که توازن مناسبی بین تعداد اعضای تومورهای LGG و GBM وجود ندارد. ویژگی‌های IDH1، TP53، Age و ATRX تأثیرگذارترین پارامترها در پیش‌بینی گلیوما در مجموعه‌دادهٔ مورد بررسی تشخیص داده شدند. برای مورد LGG، پیش‌بینی مدل به میزان چشم‌گیری به ویژگی‌های IDH1 و سن کوچک‌تر وابسته است. سن بالا احتمال تشخیص LGG را کاهش می‌دهد. در مورد GBM ویژگی‌های تأثیرگذار، مقادیر بزرگ‌تر IDH1، PTEN، TP53 و سن بالاترند؛ به‌هرحال، باتوجه به نتایج حاصل، از جهت شناسایی بیشترسن میزان تومورها (یادآوری بالا)، دقت و تفسیرپذیری، می‌توان با قاطعیت مدل EML را به‌عنوان بهترین مدل مطرح کرد. از طرفی در میان مدل‌های موجود SVM عملکرد مطلوبی نشان نمی‌دهد.

مدل پیشنهادی، با وجود عملکرد بسیار مطلوب، دارای چندین محدودیت است؛ یکی از مهم‌ترین آن‌ها، بررسی تأثیر نامتوازن‌بودن طبقه‌ها بر عملکرد مدل‌ها است؛ پس از تحلیل مجموعه‌داده، مشخص شد که تعداد نمونه‌های طبقهٔ GBM کمتر از LGG است. این موضوع باعث می‌شود مدل پیشنهادی باوجود کسب نرخ یادآوری بالا در تشخیص GBM دقت پایین‌تری کسب کند که منجر به افزایش احتمال خطای مثبت کاذب می‌شود که در مسائل پزشکی حائز اهمیت است؛ همچنین، انتخاب حد آستانهٔ خاص برای الگوریتم‌های Boruta و SHAP در مرحلهٔ انتخاب ویژگی، باید به‌صورت بهینه تنظیم شود تا ریسک حذف ویژگی‌های مفید را کاهش دهد؛ درنهایت، پیچیدگی مدل پیشنهادی به‌دلیل ترکیب چندین مدل پایه و یک متامدل بیشتر از مدل‌های پایه است که چالش‌هایی را در پیاده‌سازی مدل ایجاد می‌کند.

به‌منظور انجام پژوهش بیشتر و بهبود مدل پیشنهادی موارد زیر به‌عنوان کارهای آینده درنظر گرفته شده‌است:

- استفاده از روش‌های بازنمونه‌گیری برای مقابله با نامتوازن‌بودن طبقه‌ها و بهبود عملکرد مدل در تشخیص هر دو نوع تومور
- بهینه‌سازی پارامترهای کنترلی الگوریتم‌های انتخاب ویژگی مانند Boruta و SHAP، می‌تواند منجر به انتخاب بهینهٔ ویژگی‌ها و ارتقای هرچه بیشتر عملکرد مدل شود.
- ارزیابی مدل روی مجموعه‌داده‌های متنوع و بزرگ‌مقیاس جهت بررسی قابلیت تعمیم‌پذیری، پایداری کاربردپذیری و همچنین نقاط قوت و ضعف مدل پیشنهادی در محیط‌های بالینی ضروری

۷-مراجع 7-References

[1] D. N. Louis *et al.*, "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathol*, vol. 114, no. 2, pp. 97–109, 2007, doi: 10.1007/s00401-007-0243-4.

[2] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images," *Comput Biol Med*, vol. 121, p. 103758, 2020, doi: 10.1016/j.compbiomed.2020.103758.

[3] A. Krauze, "Using Artificial Intelligence and Magnetic Resonance Imaging to Address Limitations in Response Assessment in Glioma," *Oncology Insights*, vol. 1, pp. 1–14, 2022, doi: 10.55085/oi.2022.616.

[4] S. Pereira, R. Meier, V. Alves, M. Reyes, and C. A. Silva, "Automatic brain tumor grading

from MRI data using convolutional neural networks and quality assessment," in *International Workshop on Machine Learning in Clinical Neuroimaging*, Cham: Springer International Publishing, 2018, pp. 106–114.

[5] K. O. Almansory and F. Fraioli, "Combined PET/MRI in brain glioma imaging," *Br J Hosp Med*, vol. 80, no. 7, pp. 380–386, 2019, doi: 10.12968/hmed.2019.80.7.380.

[6] J. Tiefenbach *et al.*, "The use of advanced neuroimaging modalities in the evaluation of low-grade glioma in adults: a literature review," *Neurosurg Focus*, vol. 56, no. 2, 2024, doi: 10.3171/2023.11.FOCUS23649.

[7] E. Tasci, Y. Zhuge, H. Kaur, K. Camphausen, and A. V. Krauze, "Hierarchical Voting-Based Feature Selection and Ensemble Learning Model Scheme for Glioma Grading with Clinical and Molecular Characteristics," *Int J Mol Sci*, vol. 23, no. 22, 2022, doi: 10.3390/ijms232214155.

[8] P. Wesseling and D. Capper, "WHO 2016 Classification of gliomas," *Neuropathol Appl Neurobiol*, vol. 44, no. 2, pp. 139–150, 2018, doi: 10.1111/nan.12432.

[9] M. Platten and D. A. Reardon, "Concepts for Immunotherapies in Gliomas," *Semin Neurol*, vol. 38, no. 1, pp. 62–72, 2018, doi: 10.1055/s-0037-1620274.

[10] R. L. Soiza, A. I. C. Donaldson, and P. K. Myint, "Treatment of glioblastoma in adults," *Ther Adv Vaccines*, vol. 9, no. 6, pp. 259–261, 2018.

[11] R. Sánchez-Marqués, V. García, and J. S. Sánchez, "A data-centric machine learning approach to improve prediction of glioma grades using low-imbalance TCGA data," *Sci Rep*, vol. 14, no. 1, pp. 1–16, 2024, doi: 10.1038/s41598-024-68291-0.

[12] M. N. Samara and K. D. Harry, "Integrating Boruta , LASSO , and SHAP for Clinically Interpretable Glioma Classification Using Machine Learning," *BioMedInformatics*, vol. 5, no. 34, pp. 1–26, 2025.

[13] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Comput Biol Med*, vol. 111, p. 103345, 2019, doi: 10.1016/j.compbiomed.2019.103345.

[14] A. Alksas *et al.*, "A Novel System for Precise Grading of Glioma," *Bioengineering*, vol. 9, no. 10, pp. 1–21, 2022, doi: 10.3390/bioengineering9100532.

[15] X. Wang *et al.*, "Machine learning models for multiparametric glioma grading with quantitative result interpretations," *Front Neurosci*, vol. 13, no. JAN, p. 432416, Jan. 2019, doi: 10.3389/FNINS.2018.01046/BIBTEX.

[16] L. Wen, H. Sun, G. Liang, and Y. Yu, "A deep ensemble learning framework for glioma segmentation and grading prediction," *Sci Rep*, vol. 15, no. 1, pp. 1–15, Dec. 2025, doi: 10.1038/S41598-025-87127-Z;SUBJMETA.

[17] X. Xia, W. Wu, Q. Tan, and Q. Gou, "Interpretable Machine Learning Models for Differentiating Glioblastoma From Solitary Brain Metastasis Using Radiomics," *Acad Radiol*, vol. 32, no. 9, pp. 5388–5400, Sep. 2025, doi: 10.1016/J.ACRA.2025.05.016.

[18] R. C. Joshi, R. Mishra, P. Gandhi, V. K. Pathak, R. Burget, and M. K. Dutta, "Ensemble based machine learning approach for prediction of glioma and multi-grade classification," *Comput Biol Med*, vol. 137, p. 104829, Oct. 2021, doi: 10.1016/J.COMPBIOMED.2021.104829.

[19] S. Gutta, J. Acharya, M. S. Shiroishi, D. Hwang, and K. S. Nayak, "Improved Glioma Grading Using Deep Convolutional Neural Networks," *American Journal of Neuroradiology*, vol. 42, no. 2, pp. 233–239, 2021.

[20] J. Cheng, J. Liu, H. Yue, H. Bai, Y. Pan, and J. Wang, "Prediction of Glioma Grade Using Intratumoral and Peritumoral Radiomic Features From Multiparametric MRI Images," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 19, no. 2, pp. 1084–1095, 2022, doi: 10.1109/TCBB.2020.3033538.

[21] T. R. Noviandy, G. M. Idroes, and I. Hardi, "Integrating explainable artificial intelligence and light gradient boosting machine for glioma grading," *Informatics and Health*, vol. 2, no. 1, pp. 1–8, 2025, doi: 10.1016/j.infoh.2024.12.001.

[22] S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, 2017, pp. 1–10.

[23] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J Big Data*, vol. 7, no. 94, 2020, doi: 10.1186/s40537-020-00369-8.

[24] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests : A survey and results of new tests," *Pattern Recognit*, vol. 44, no. 2, pp. 330–349, 2011, doi: 10.1016/j.patcog.2010.08.011.

[25] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach Learn*, vol. 63, no. 1, pp. 3–42, 2006, doi: 10.1007/s10994-006-6226-1.

[26] S. Lavanya, S. V. Annlin Jeba, P. Bhuvaneswari, and F. H. Shajin, "Support vector machine classifier optimized with seagull optimization algorithm for brain tumor classification," *Concurr Comput*, vol. 35, no. 1, pp. 1–13, 2023, doi: 10.1002/cpe.7396.

[27] M. Maalouf, "Logistic regression in data analysis: An overview," *International Journal of Data Analysis Techniques and Strategies*, vol. 3, no. 3, pp. 281–299, 2011, doi: 10.1504/IJDATS.2011.041335.

[28] I. M. De Diego, A. R. Redondo, R. R. Fernández, J. Navarro, and J. M. Moguerza, "General Performance Score for classification problems," *Applied Intelligence*, vol. 52, no. 10, pp. 12049–12063, 2022, doi: 10.1007/s10489-021-03041-7.

[۲۹] خدادای، الناز، حسینی، راحیل، مزینانی مهدی، «ارائهٔ مدل‌های محاسبات نرم مبتنی بر فازی، تکاملی و هوش جمعی در تحلیل تصاویر ماموگرافی جهت تشخیص تومورهای سینه»، فصلنامهٔ پردازش علائم و داده ها، دورهٔ۱۶، شمارهٔ ۲، صفحات ۱۶۵-۱۴۷، ۱۳۹۸.

[29] E. Khodadadi, R. Hosseini, M. Mazinani, "Soft Computing Methods based on Fuzzy, Evolutionary and Swarm Intelligence for Analysis of Digital Mammography Images for Diagnosis of Breast Tumors," Journal of Signal and Data Processing, vol. 16, no. 2, pp. 147-165, 2019, doi: 10.29252/jsdp.16.2.147.

[۳۰] آذرنوید، بابک، عبدالحسین زاده، محسن، امامی، حجت، «مدل یادگیری ماشین انباشته برای دسته‌بندی و پیش‌بینی بیماری‌های کبدی»، فصلنامهٔ پردازش علائم و داده ها، دورهٔ۲۲، شمارهٔ ۲، صفحات ۷۹-۹۶، ۱۴۰۴.

[30] B. Azarnavid, M. Abdolhosseinzadeh, H. Emami, "Stacking machine learning model for classification and prediction of liver diseases," Journal of Signal and Data Processing, vol. 22, no. 2, pp. 79-96, 2025, doi: 10.61882/jsdp.22.2.79.

حجت امامی دانش‌آموختهٔ رشتهٔ مهندسی کامپیوتر گرایش هوش مصنوعی و دانشیار گروه مهندسی کامپیوتر دانشگاه بناب است. زمینه‌های پژوهشی وی شامل داده‌کاوی، یادگیری ماشین، سامانه‌هاای چندعاملی، الگوریتم‌های اکتشافی و فرااکتشافی، هوش جمعی و کاربردهای آن است. هم‌اکنون او در زمینهٔ استفاده از یادگیری ماشین در حوزهٔ مدیریت ترافیک هوایی، تشخیص بیماری در حوزهٔ پزشکی و جایابی گره‌ها در شبکه‌های نوری کار پژوهشی خود را ادامه می‌دهد.
نشانی رایانامهٔ ایشان عبارت است از:
emami@ubonab.ac.ir

بابک آذرنوید دانش‌آموختهٔ رشتهٔ ریاضی کاربردی و هم‌اکنون استادیار گروه ریاضی و علوم کامپیوتر دانشگاه بناب است. زمینه‌های پژوهشی او شامل روش‌های عددی حل معادلات دیفرانسیل، آنالیز عددی، یادگیری ماشین و کاربردهای آن‌ها است. در حال حاضر، وی در زمینهٔ استفاده از یادگیری ماشین برای تشخیص بیماری در حوزهٔ پزشکی و همچنین روش‌های عددی و فراابتکاری در حل معادلات دیفرانسیل به فعالیت‌های پژوهشی خود ادامه می‌دهد.
نشانی رایانامهٔ ایشان عبارت است از:
babakazarnavid@ubonab.ac.ir

محسن عبدالحسین‌زاده دانش‌آموختهٔ رشتهٔ ریاضی کاربردی گرایش تحقیق در عملیات است. او استادیار گروه ریاضی و علوم کامپیوتر دانشگاه بناب است. زمینه‌های پژوهشی وی شامل بهینه‌سازی شبکه، بهینه‌سازی ترکیبیاتی، الگوریتم‌های ابتکاری و فراابتکاری، تصمیم‌گیری غیرقطعی، هوش مصنوعی و یادگیری ماشین است. هم‌اکنون او در زمینهٔ استفاده از یادگیری ماشین در حوزهٔ مسیریابی، تشخیص بیماری در حوزهٔ پزشکی و الگوریتم‌های ترکیبی کار پژوهشی خود را ادامه می‌دهد.
نشانی رایانامهٔ ایشان عبارت است از:
mohsen.ab@ubonab.ac.ir