



OPEN Predicting the efficacy of bevacizumab on peritumoral edema based on imaging features and machine learning

Xuexue Bai¹, Ming Feng¹✉, Wenbin Ma¹✉ & Shiyong Wang²✉

This study proposes a novel approach to predict the efficacy of bevacizumab (BEV) in treating peritumoral edema in metastatic brain tumor patients by integrating advanced machine learning (ML) techniques with comprehensive imaging and clinical data. A retrospective analysis was performed on 300 patients who received BEV treatment from September 2013 to January 2024. The dataset incorporated 13 predictive features: 8 clinical variables and 5 radiological variables. The dataset was divided into a training set (70%) and a test set (30%) using stratified sampling. Data preprocessing was carried out through methods such as handling missing values with the MICE method, detecting and adjusting outliers, and feature scaling. Four algorithms, namely Random Forest (RF), Logistic Regression, Gradient Boosting Tree, and Naive Bayes, were selected to construct binary classification models. A tenfold cross-validation strategy was implemented during training, and techniques like regularization, hyperparameter optimization, and oversampling were used to mitigate overfitting. The RF model demonstrated superior performance, achieving an accuracy of 0.89, a precision of 0.94, F1-score of 0.92, with both AUC-ROC and AUC-PR values reaching 0.91. Feature importance analysis consistently identified edema volume as the most significant predictor, followed by edema index, patient age, and tumor volume. Traditional multivariate logistic regression corroborated these findings, confirming that edema volume and edema index were independent predictors ($p < 0.01$). Our results highlight the potential of ML—driven predictive models in optimizing BEV treatment selection, reducing unnecessary treatment risks, and improving clinical decision—making in neuro—oncology.

Keywords Bevacizumab, Peritumoral edema, Machine learning, Predictive model

Peritumoral edema (PTE), characterized by pathological fluid accumulation in the brain parenchyma, is a prevalent and serious complication of metastatic brain tumors that adversely affects both neurological function and survival outcomes¹. As the most frequent malignant intracranial neoplasms in adults, metastatic brain tumors exhibit an incidence rate approximately tenfold higher than primary intracranial tumors². Mechanistically, PTE elevates intracranial pressure, which induces cerebral hypoxia and may progress to life-threatening herniation, ultimately resulting in irreversible neurological deficits or mortality³. The current treatment strategy involves the use of mannitol, diuretics, and steroid medications to alleviate the symptoms of cerebral edema. Previous studies have shown that the control rate of these drugs for refractory cerebral edema ranges from 27 to 39%^{4–6}. It is noteworthy that these medications do not eradicate the underlying causative factors and are often accompanied by various adverse reactions⁷. For instance, long-term use of steroid medications can lead to severe systemic side effects, including but not limited to immunosuppression and avascular necrosis⁸. Similarly, the administration of mannitol may also trigger systemic hypotension, reduced cerebral perfusion, and even acute renal failure⁹.

Vascular endothelial growth factor A (VEGF-A) is a potent inducer of angiogenesis and enhancer of vascular permeability, which is considered to play a pivotal role in the pathogenesis of brain tumor-associated edema¹⁰. As a result, bevacizumab (BEV), a monoclonal antibody targeting VEGF-A, has emerged as an effective therapeutic option for addressing cerebral edema^{11–14}. However, despite the demonstrated efficacy of BEV in treating cerebral edema, its use is accompanied by numerous concerns, such as the potential risks of tumor hemorrhage,

¹Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, No. 1 Shuaifuyuan Hutong, Dongcheng District, Beijing 100730, China. ²Neurosurgery of The First Affiliated Hospital, Jinan University, Guangzhou, China. ✉email: pumchfming@126.com; Wbin718@126.com; Washiyong@126.com

ischemic stroke, and induction of abnormal tumor angiogenesis, as well as potential effects on patients' cognitive function¹⁵. Furthermore, as a high-cost medication, the economic factor must also be considered in the use of BEV. Given the potential risks and its high cost, the ability to accurately predict BEV efficacy in treating cerebral edema holds significant value for clinical decision-making. It can aid in optimizing treatment plans, ensuring therapeutic outcomes while alleviating the financial burden on patients.

Artificial intelligence (AI) includes machine learning (ML) and deep learning (DL), each with its own structure and applications. Through the application of advanced algorithms for deep analysis of diverse data, ML technology is capable of precisely capturing and interpreting the intricate relationships between various clinical features and outcomes. This characteristic endows ML technology with the potential to surpass the limitations of traditional statistical methods in risk prediction, thereby offering clinicians more precise and comprehensive insights to inform decision-making¹⁶. ML models exhibit significant advantages over other prediction methods, primarily in their ability to automatically learn and identify optimal predictive features from training data. This process, devoid of manual intervention, substantially enhances the accuracy and efficiency of predictive models. Additionally, ML models possess robust generalization capabilities, which mean that the knowledge and patterns learned from training data can be effectively applied to new, unseen patient cohorts. This further strengthens their applicability and value in real-world clinical scenarios¹⁷.

Recent advances in neuro-oncology highlight the transformative potential of AI and ML. In this field, ML algorithms extract quantitative imaging features—including shape, grayscale, and texture characteristics—from medical images, which are subsequently processed by classifiers such as support vector machines, random forests (RF), and k-nearest neighbors to significantly improve the efficiency and accuracy of brain tumor detection^{18,19}. Large-scale clinical validations underscore AI's prognostic value in neuro-oncology. Recent studies involving 79,638 participants established robust performance in predicting survival outcomes for high-grade glioma²⁰, while comprehensive meta-analyses confirmed ML's reliability in forecasting tumor recurrence²¹. Beyond diagnostic applications, ML has shown comparable promise in neurovascular disorders, particularly in predicting poor functional outcomes following aneurysmal subarachnoid hemorrhage²². The predictive utility of ML extends to cerebral edema management. Multiple studies have developed MRI radiomics-based models that demonstrate good accuracy in predicting post-stroke cerebral edema^{23–25}, with texture features showing particular value in identifying symptomatic edema after endovascular treatment in high-risk populations²⁶. Notably, ML approaches for post-stroke edema prediction have consistently achieved accuracy rates exceeding 70% in independent validations^{27–29}.

Despite these advances, current research lacks investigations into predicting BEV treatment efficacy for PTE in metastatic brain tumors. This study aims to develop the first ML-based predictive model integrating quantitative imaging biomarkers with clinical parameters. By identifying pretreatment predictors of BEV response, our approach could provide clinicians with two critical decision-support functions: (1) avoiding BEV administration in patients unlikely to benefit, thereby reducing treatment risks and costs, and (2) optimizing therapeutic selection for potential responders. These findings could address an urgent need in neuro-oncology by providing an objective tool for personalizing BEV therapy in metastatic brain tumor patients.

Methods

Data sources

This study conducted a retrospective analysis of clinical data from patients with metastatic brain tumors who underwent BEV therapy from September 2013 to January 2024. The dataset consisted of clinical information from a total of 300 patients, including 18 features and 1 label. Among the 18 features, there were 13 clinical characteristics (demographics, medical history), plus 5 radiological features (tumor location, maximum tumor diameter, tumor volume, edema volume, and edema index). Figure 1 illustrates the workflow of this study. This retrospective study was conducted using existing, fully anonymized clinical data without any patient intervention. The Institutional Review Board (IRB) of Peking Union Medical College Hospital, Chinese Academy of Medical Sciences waived the requirement for informed consent in accordance with Article 39 of China's Ethical Review Measures for Biomedical Research Involving Humans (2016), and the study protocol was approved by the hospital's Academic Ethics Committee. The research strictly adhered to the ethical principles of the Declaration

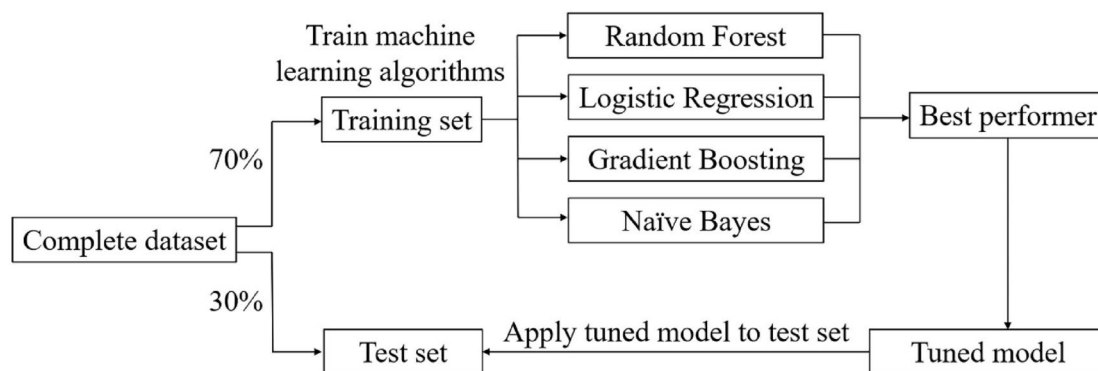


Fig. 1. Workflow diagram of machine learning model training in this study.

of Helsinki and received formal ethical approval exemption from the IRB (Exemption No. ZS-3931). The study design, ethical review process, and informed consent waiver procedures were fully compliant with Article 39 of the aforementioned Ethical Review Measures and Article 13 of China's Personal Information Protection Law (2021).

Efficacy evaluation

The tumor volume and maximum tumor diameter were measured on the T1-weighted contrast-enhanced sequence, while the edema volume was assessed on the T2/FLAIR sequence. Accurate tumor and edema volumes were obtained by summing the segmented volumes from each axial image³⁰. The edema index was calculated based on the edema volume and tumor volume, using the formula: $\text{Edema Index} = (\text{Edema Volume} + \text{Tumor Volume}) / (\text{Tumor Volume})$ ³¹. To evaluate treatment efficacy, we defined a quantitative criterion for edema remission: a reduction in edema volume exceeding 10% post-treatment compared to pre-treatment was considered as edema remission; conversely, a reduction below this threshold was defined as non-remission³². Based on this criterion, patients included in the study were categorized into edema remission and non-remission groups according to the change in edema volume before and after BEV treatment, to further analyze treatment outcomes and potential clinical correlations.

Data analysis

Continuous variables were presented as medians (interquartile ranges), while categorical variables were expressed as numbers/percentages. When analyzing the baseline characteristics of patients, continuous variables were assessed using logistic regression analysis, and categorical variables were analyzed with Chi-square tests. To facilitate the calculation of odds ratios when examining the relationship between independent variables and the dependent variable, continuous variables were first converted into categorical variables (e.g., age > 60 years). Subsequently, univariate logistic regression analysis was performed. Variables with a *P*-value < 0.05 were selected for inclusion in the multivariate logistic regression analysis. All statistical analyses were conducted using SPSS (IBM, Version 26.0).

Model construction and evaluation

The dataset of 300 patients was divided into a training set (70%) and a test set (30%) using stratified sampling. This split ratio was chosen to ensure a robust training set while maintaining a sufficiently large test set for reliable evaluation, particularly given the moderate size of our dataset. Preliminary experiments with alternative split ratios (e.g., 80–20%) showed no significant impact on model performance. Stratified sampling ensured that the distribution of patients in the edema remission group (*n* = 220) and the non-remission group (*n* = 80) was maintained in both subsets. Specifically, the training set consisted of 210 patients (154 from the edema remission group and 56 from the non-remission group), while the test set included 90 patients (66 from the edema remission group and 24 from the non-remission group). (Table 1).

Before training, we conducted extensive data preprocessing to ensure the robustness of our predictive models. Missing values were addressed using the MICE (Multiple Imputation by Chained Equations) method with 10 iterations. Logistic regression (LR) was used as the imputation model for categorical variables, which is particularly effective for handling missing data in both continuous and categorical variables. Outliers were detected using the Interquartile Range (IQR) method and adjusted or removed based on clinical relevance and predefined statistical rules. Feature scaling was performed to normalize the data. Continuous variables were standardized using Z-score normalization, which transforms the data to have a mean of 0 and a standard deviation of 1. This step is particularly important for algorithms like LR and Gradient Boosting Trees (GBT), which are sensitive to the scale of input features. By normalizing the data, we ensured that no single feature dominated the model training process due to its larger scale, while also maintaining robustness to potential outliers.

To further optimize the feature set and improve model performance, we performed feature selection and dimensionality reduction. Feature selection was conducted using Recursive Feature Elimination (RFE) with LR as the underlying model, which iteratively removes the least important features based on model performance evaluated through cross-validation. Dimensionality reduction was performed using Principal Component Analysis (PCA), retaining components that explained 95% of the variance in the data. These steps helped us reduce the feature set from 18 to 13 features, thereby enhancing the predictive performance and generalization ability of the models.

In this study, we selected four distinct ML algorithms to train binary classification models aimed at predicting the efficacy of BEV treatment on PTE in patients with metastatic brain tumors. The choice of these algorithms was based on their unique strengths and potential complementarity, aiming to achieve optimal predictive performance. Specifically, the RF algorithm, with its ensemble learning characteristics, demonstrates robustness in handling large-scale, high-dimensional, and complex medical datasets. The LR model is suitable for dealing

Dataset	Total	Remission	Non-remission	Ratio
Full cohort	300	220	80	2.75:1
Training set	210	154	56	2.75:1
Test set	90	66	24	2.75:1

Table 1. Dataset distribution for model development.

with linearly or approximately linearly separable data patterns, offering clear interpretability. The GBT algorithm is renowned for its powerful ability to mine nonlinear relationships. Lastly, the Naive Bayes (NB) model, grounded in solid probability theory, has advantages in handling small sample sizes and high-dimensional feature spaces. To further ensure model robustness and generalizability, a tenfold cross-validation strategy was implemented during model training, iteratively training on 9 subsets and validating on the remaining subset, with the final performance metrics averaged across all folds.

To mitigate the risk of overfitting and ensure the generalizability of our predictive models, we implemented several techniques tailored to the characteristics of each ML algorithm. For LR, we applied L2 regularization (Ridge regularization) and optimized the regularization strength (C) through grid search over the range [0.01, 0.1, 1, 10], while setting the `class_weight` parameter to 'balanced' to address class imbalance. For RF and GBT, we optimized key hyperparameters, including the number of trees and maximum depth of trees, through grid search, and set `class_weight` to 'balanced'. For NB, we adjusted the Laplace smoothing parameter and applied the Synthetic Minority Over-sampling Technique (SMOTE) with `k_neighbors`=5 to the training set, increasing the number of minority class samples to balance class distribution. SMOTE was applied exclusively to the training set, while the test set remained unchanged to ensure unbiased evaluation. Additionally, we employed tenfold cross-validation to further ensure model robustness and generalizability.

In the model performance evaluation phase, various evaluation metrics were employed, including accuracy, precision, recall, and F1-score, to comprehensively and objectively reflect the predictive efficacy of the models³³. To further analyze the models' performance under imbalanced sample scenarios, we calculated the F1-score, precision, and recall for each class ("non-remission" and "remission") separately. Model performance was assessed using a tenfold cross-validation strategy on the training set, followed by evaluation on an independent test set. Furthermore, Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves were plotted, and the corresponding Area Under the Curve (AUC) was calculated to further quantify the models' ability to distinguish between different treatment response groups. Finally, this study leveraged the Python programming language for data preprocessing, visualization analysis, and the training and evaluation of the models.

Results

Retrospective data analysis

This study included a total of 300 patients with metastatic brain tumors who received BEV treatment. Through retrospective analysis of the clinical and radiological characteristics of these two groups, we identified five variables that showed statistically significant differences between the two groups ($P < 0.05$): tumor origin, history of previous surgery, interval between BEV treatments, edema volume, and edema index. This finding suggests that these five variables may be crucial factors influencing the efficacy of BEV treatment on PTE in patients with metastatic brain tumors. The detailed results of the retrospective data analysis are presented in Table 2.

Univariate and multivariate logistic regression analysis

In this study, we performed a univariate logistic regression analysis on 18 variables to initially screen for potential key factors influencing the efficacy of BEV treatment on PTE in patients with metastatic brain tumors. The analysis revealed that five variables—tumor origin, surgical history, maximum tumor diameter, edema volume, and edema index—were statistically significantly associated with the treatment outcome, and thus were considered potential predictors of efficacy. To further validate the independent effects and relative importance of these factors, we subsequently conducted a multivariate logistic regression analysis on the five selected variables. The results indicated that edema volume and edema index emerged as independent risk factors for BEV treatment efficacy. Notably, the multivariate analysis highlighted edema volume as the most critical factor influencing treatment outcomes. Table 3 details the specific results of both the univariate and multivariate logistic regression analyses.

Machine learning prediction models

In this study, we employed four ML algorithms, RF, LR, GBT, and NB, to construct prediction models. Following training and cross-validation, these models were evaluated on an independent test set comprising 90 patients. Table 4 presents the performance metrics of each model in detail. Regarding model accuracy, the RF model achieved the highest accuracy of 0.89, significantly outperforming the LR and NB models, which had accuracies of 0.79 and 0.69, respectively. In terms of precision, the GBT model performed best with a precision of 0.97, while the RF model followed closely with a precision of 0.94. The RF model also attained the highest recall of 0.91, compared with the LR and GBT models, which both had a recall of 0.83. In terms of F1 score, the RF model scored the highest at 0.92, demonstrating its ability to balance precision and recall effectively. Figure 2 showcases the ROC curves and PR curves of the four ML models, providing an in-depth analysis of their classification performance. In Fig. 2A, the ROC curve analysis demonstrates that the RF model performs most effectively, achieving an AUC of 0.91, which is the highest among the four models. Figure 2B presents the PR curves of the models, focusing on evaluating the model's performance on positive class samples. Here, the RF and GBT models demonstrate significant superiority, both achieving an Area Under the PR Curve (AP) of 0.91. Considering all performance metrics comprehensively, the RF model was identified as the best-performing model due to its superior performance in accuracy, recall, F1 score, AUC-ROC, and AUC-PR. Figure 3 illustrates the confusion matrices of the four models, further supporting the aforementioned conclusions.

To further evaluate the models' performance under imbalanced sample scenarios, we analyzed their precision, recall, and F1-scores for the remission and non-remission groups, as shown in Table 5. The RF model demonstrated strong performance across all metrics, achieving high precision (0.95 for remission, 0.92 for non-remission), recall (0.93 for remission, 0.88 for non-remission), and F1-scores (0.93 for remission, 0.89 for non-remission), indicating its robustness in handling class imbalance. The GBT model also performed well,

Parameter	Non-remission	Remission	P-value
Age (Median (IQR))	58.0 years (27.0)	61.5 years (19.0)	0.317
Gender (N)			0.089
Male	33	117	
Female	47	103	
Weight			0.066
BMI < 25.0	28	104	
BMI ≥ 25.0	52	116	
Tumor origin			0.004*
Breast cancer	31	47	
Lung cancer	49	173	
Steroid history			0.068
Yes	71	209	
No	9	11	
Dehydrant history			0.085
Yes	72	9	
No	8	211	
Radiotherapy			0.069
Yes	54	172	
No	26	48	
Operation			0.040*
Yes	56	179	
No	24	41	
Electrolyte disorder			0.102
Yes	27	52	
No	53	168	
New-onset hypertension			0.090
Yes	13	57	
No	67	163	
BEV dose	10.0 mg/kg (5.0)	10.0 mg/kg (0.0)	0.172
BEV interval	2.0 weeks (1.0)	2.0 weeks (1.0)	0.005*
BEV frequency	4.0 times (4.0)	3.0 times (2.0)	0.575
Tumor location			0.215
Frontal lobe	16	62	
Parietal lobe	23	75	
Temporal lobe	19	33	
Occipital lobe	10	18	
Polyencephalic lobe	12	32	
Maximum tumor diameter	3.6 cm (0.7)	3.5 cm (0.7)	0.538
Tumor volume	17.1 ccm (10.0)	15.6 ccm (9.7)	0.383
Edema volume	26.0 ccm (19.8)	51.0 ccm (18.8)	<0.001*
Edema index	2.7 (1.1)	3.9 (2.2)	0.020*

Table 2. Retrospective analysis of clinical and imaging features of all patients. An asterisk (*) indicates a statistically significant *P* value of less than 0.05. *IQR* InterQuartile Range, *BMI* body mass index, *BEV* bevacizumab.

particularly in precision, though its recall for the non-remission group was slightly lower. In contrast, the Logistic Regression (LR) model showed a notable bias towards the remission group, with significantly lower recall and F1-scores for the non-remission group. The NB model performed the weakest across all metrics, indicating its limited effectiveness in imbalanced scenarios. Overall, the RF and GBT models outperformed the other models, demonstrating their superior ability to handle imbalanced datasets. These findings align with the earlier results from the ROC and PR curves, further confirming the RF model's effectiveness in balancing precision and recall.

In addition, we also conducted feature importance analysis using the RF model and the GBT model. For the RF model, the analysis revealed that edema volume, edema index, age, tumor volume, and maximum tumor diameter were the top five most important features. Among these, edema volume stood out as the most critical factor, with an importance score of 0.28. Furthermore, in the feature importance assessment of the GBT model, edema volume maintained its dominant position, achieving an impressive importance score of 0.61. This finding

Parameter	UV OR (95% CI)	UV p	MV OR (95% CI)	MV p*
Age	1.51 (0.90–2.53)	0.118		
Gender	0.62 (0.37–1.04)	0.069		
Weight	1.67 (0.98–2.83)	0.059		
Tumor origin	0.43 (0.25–0.75)	0.003	0.56 (0.27–1.13)	0.104
Steroid	0.42 (0.17–1.04)	0.061		
Dehydrant	0.38 (0.14–1.03)	0.058		
Radiotherapy	0.58 (0.33–1.02)	0.059		
Operation	0.53 (0.30–0.96)	0.036	0.62 (0.28–1.35)	0.225
Electrolyte disorder	1.65 (0.94–2.88)	0.080		
New-onset hypertension	0.56 (0.29–1.08)	0.083		
BEV dose	1.07 (0.99–1.17)	0.103		
BEV interval	0.58 (0.34–1.00)	0.052		
BEV frequency	0.85 (0.72–1.00)	0.058		
Tumor location	0.68 (0.24–1.87)	0.450		
Maximum tumor diameter	0.57 (0.32–1.00)	0.048	0.97 (0.36–2.58)	0.945
Tumor volume	0.96 (0.92–1.00)	0.053		
Edema volume	1.13 (1.10–1.16)	<0.001	1.09 (1.04–1.13)	<0.001*
Edema index	3.63 (2.48–5.33)	<0.001	1.87 (1.12–3.12)	0.017*

Table 3. Univariate and multivariate logistic regression analysis. An asterisk (*) indicates a statistically significant P value of less than 0.05. UV, Univariable; MV, Multivariable; CI, Confidence Interval; OR, Odds Ratio; BEV, Bevacizumab.

Factor	Random forest	Logistic regression	Gradient boosting	Naive Bayes
True positives (TP)	60	55	55	44
True Negatives (TN)	20	16	22	18
False Positives (FP)	4	8	2	6
False Negatives (FN)	6	11	11	22
Accuracy	0.89	0.79	0.86	0.69
Precision	0.94	0.87	0.96	0.88
Sensitivity (Recall)	0.91	0.83	0.83	0.67
Specificity	0.83	0.67	0.92	0.75
F1 Score	0.92	0.85	0.89	0.76
AUC-ROC	0.91	0.82	0.89	0.74
AUC-PR	0.91	0.88	0.91	0.84

Table 4. Machine learning model performance. AUC, Area Under the Curve; ROC, Receiver Operating Characteristic; PR, Precision-Recall.

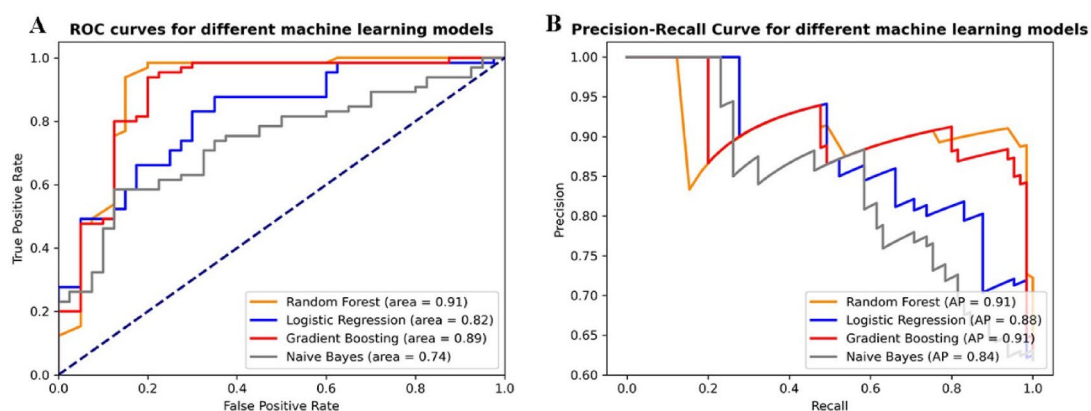


Fig. 2. ROC curves and PR curves of all machine learning models in this study. Subfigure (A) ROC curves of four machine learning models. Subfigure (B) PR curves of four machine learning models.

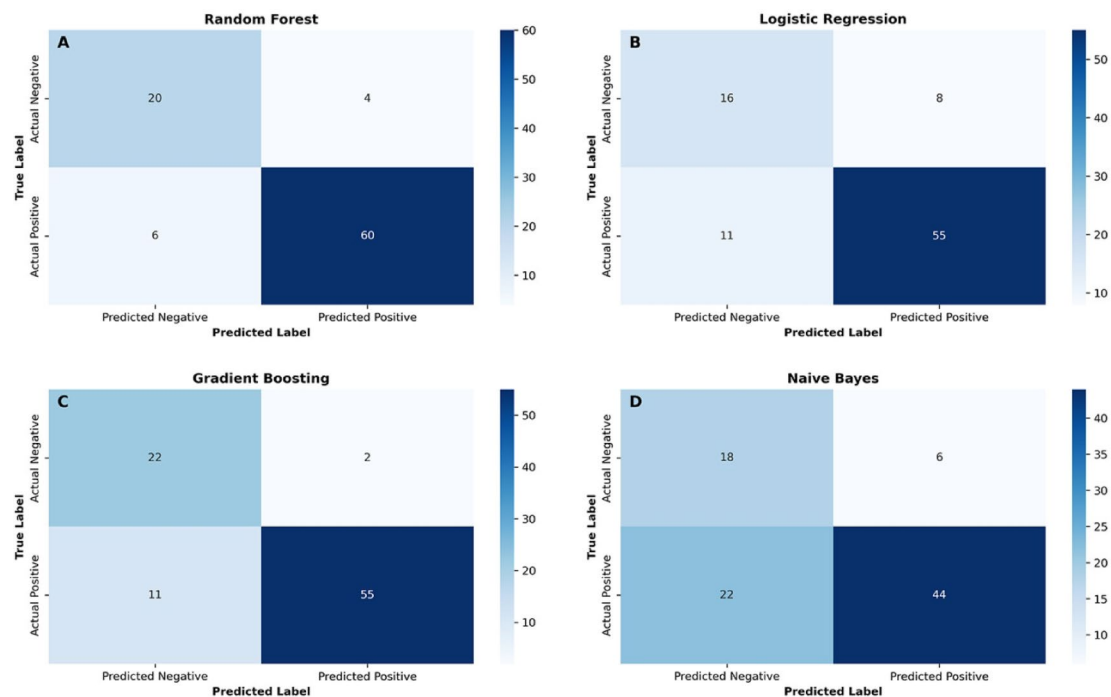


Fig. 3. Confusion Matrices of All Machine Learning Models Utilized in This Study. Subfigure (A) Confusion Matrix of Random Forest Model. Subfigure (B) Confusion Matrix of logistic regression Model. Subfigure (C) Confusion Matrix of Gradient Boosting Tree Model. Subfigure (D) Confusion Matrix of Naive Bayes Model.

Factor	Overall	Remission	Non-remission
Random forest			
Precision	0.94	0.95	0.92
Sensitivity (Recall)	0.91	0.93	0.88
F1 Score	0.92	0.93	0.89
Logistic regression			
Precision	0.87	0.89	0.82
Sensitivity (recall)	0.83	0.88	0.72
F1 score	0.85	0.90	0.78
Gradient boosting			
Precision	0.97	0.98	0.96
Sensitivity (Recall)	0.83	0.88	0.76
F1 score	0.89	0.92	0.85
Naive Bayes			
Precision	0.88	0.93	0.84
Sensitivity (recall)	0.67	0.69	0.63
F1 Score	0.76	0.81	0.69

Table 5. Model performance on overall, remission, and non-remission groups.

underscores the pivotal role of edema volume in the predictive models. Figure 4 visually presents the ranking of feature importance for both the RF and GBT models.

To delve deeper into the potential associations among various features, we employed ML techniques to conduct a feature correlation analysis and plotted a feature correlation heatmap. The heatmap in Fig. 5 shows the correlation between various features. The analysis revealed a strong positive correlation between tumor volume and maximum tumor diameter, with a correlation coefficient of 0.81. Additionally, we observed a moderate positive correlation ($r=0.62$) between tumor origin and gender. The edema index and edema volume exhibited a positive correlation of 0.57. Furthermore, a significant negative correlation was observed between the edema index and both tumor volume and maximum tumor diameter, with correlation coefficients of -0.65 and -0.59 , respectively.

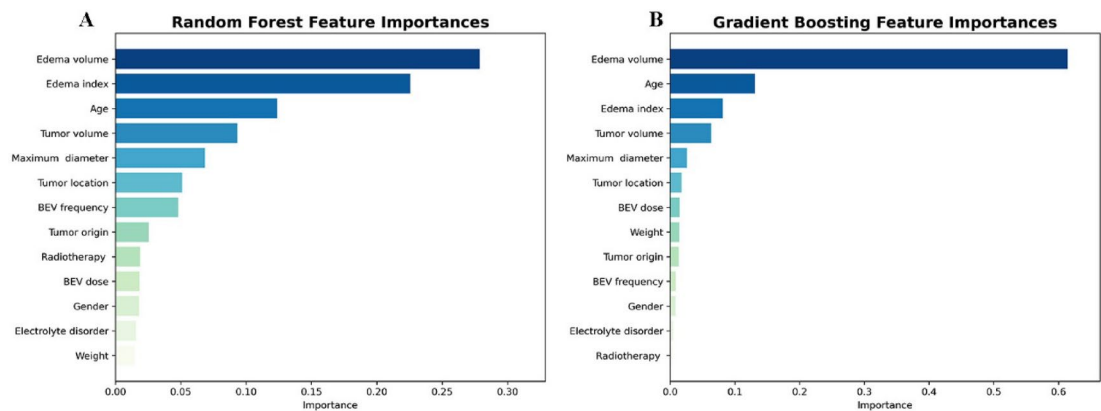


Fig. 4. Feature Importance Analysis Based on Machine Learning Models. Subfigure (A) Feature Importance Plot of Random Forest Model. Subfigure (B) Feature Importance Plot of Gradient Boosting Tree Model.

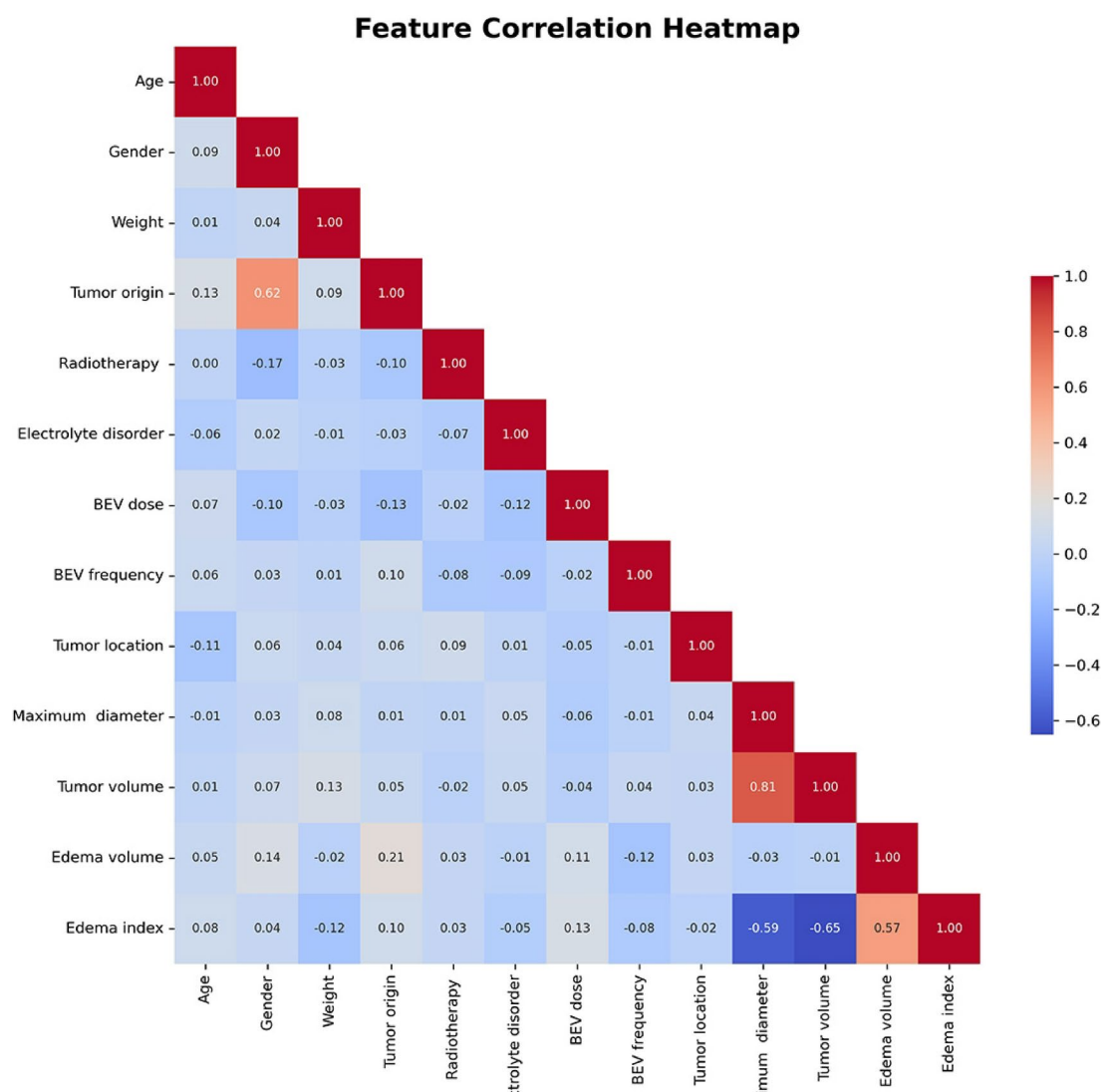


Fig. 5. Heatmap of the correlation between features based on the machine learning model.

Discussion

The present study aimed to predict the efficacy of BEV in treating PTE in patients with metastatic brain tumors by integrating clinical characteristics, imaging features, and ML models. Univariate and multivariate logistic regression analyses identified edema volume and edema index as the main factors affecting BEV's efficacy, a finding further supported by feature importance analyses in both RF and GBT models, where edema volume emerged as the most significant predictor. Edema volume serves not only as an imaging biomarker of mechanical compression but also reflects tumor biological behavior, particularly VEGF-A-driven angiogenic activity and vascular permeability—a dual role supported by prior studies linking VEGF-A to edema severity^{34,35}. The strong positive correlation between tumor volume and maximum diameter ($r=0.82$) reflects their shared role in quantifying tumor burden, suggesting that either parameter could suffice for clinical assessment. Additionally, the correlation between tumor origin and gender ($r=0.62$) likely stems from the cohort composition: all breast cancer patients were female, while lung cancer patients were predominantly male. This observation warrants further research into gender-specific differences in VEGF signaling or angiogenic activity, which may guide personalized BEV therapy.

Furthermore, by applying various ML algorithms, including RF, LR, GBT, and NB, we successfully developed a model capable of accurately predicting the efficacy of BEV in treating PTE in patients with metastatic brain tumors. Among these models, the RF model excelled in multiple performance metrics, such as the AUC value, accuracy, precision, and F1 score, indicating its robust predictive power and stability. Notably, the RF model demonstrated exceptional performance in handling imbalanced data, which can be attributed to its ensemble learning mechanism, inherent randomness³⁶, and adaptive weighting that balances precision and recall, making it robust for clinical applications³⁷. This characteristic highlights RF's suitability for medical studies, where class imbalance is common and accurate identification of minority classes (e.g., the non-remission group) is clinically critical. Our results confirm that imbalance significantly impacts model performance, with RF and GBT outperforming LR and NB in such scenarios. Future studies could further improve performance via advanced sampling (e.g., ADASYN), cost-sensitive learning, or ensemble techniques like stacking^{38,39}. In conclusion, the results of this study not only validate the superiority of RF model in imbalanced scenarios but also provide important references for model selection and optimization under similar data distributions.

The predictive model developed in this study offers significant clinical utility by stratifying patients with PTE into probable responders and non-responders to BEV treatment based on integrated clinical and imaging characteristics. This stratification enables clinicians to optimize therapeutic strategies: prioritizing BEV for likely responders to accelerate edema remission, while considering alternative approaches for predicted non-responders to avoid unnecessary exposure to BEV's risks (e.g., hemorrhage, hypertension) and reduce financial burdens—particularly valuable in resource-limited settings⁴⁰. Unlike traditional clinical judgment alone, the AI model leverages advanced ML to analyze complex, multi-dimensional data, providing more objective and reproducible predictions. The model's ability to guide targeted use of this high-cost biologic addresses a critical need for cost-effective precision medicine in neuro-oncology. The AI model's implementation holds transformative potential across multidisciplinary clinical practice. For neurosurgeons involved in the long-term management of PTE, the model offers data-driven guidance to optimize medical therapy (e.g., BEV initiation or alternative strategies), supporting clinical decisions during patient follow-up. Neuro-oncologists can leverage its predictions to personalize treatment regimens, minimizing BEV-related risks for non-responders while optimizing outcomes for responders. Radiologists benefit from the model's standardized quantification of imaging biomarkers (e.g., edema volume), reducing inter-observer variability and improving consistency in tumor board discussions. Through translating complex ML outputs into actionable clinical insights and integrating predictive analytics into routine practice, this approach not only bridges the gap between computational analytics and clinical decision-making in neuro-oncology but also fosters a collaborative framework for PTE management across multidisciplinary teams (neurologists, oncologists, and radiologists).

It is important to acknowledge the limitations of this study. Firstly, as a retrospective study, there is a potential risk of data bias. Secondly, despite selecting four representative algorithms, the range of ML algorithms employed in this study was limited; there may be other, more suitable algorithms for this particular research that were not included. Lastly, some subjectivity in the feature selection process cannot be overlooked. These limitations underscore the need for future studies to address these issues by incorporating larger, prospective datasets, exploring a wider range of algorithms, and adopting more rigorous and objective feature selection methods in order to further validate and refine the predictive models.

In future research, we plan to expand the scope of our predictive model by incorporating additional imaging features, such as perfusion and diffusion-weighted MRI parameters, which could provide deeper insights into the pathophysiology of PTE and further enhance predictive performance. For instance, perfusion MRI could offer valuable information on tumor vascularity and blood–brain barrier integrity, while diffusion-weighted imaging (DWI) might reveal microstructural changes in the brain parenchyma associated with edema formation⁴¹. Additionally, integrating molecular biomarkers and gene expression data, such as VEGF-A and other angiogenesis-related genes, could reveal underlying biological mechanisms driving edema formation and response to BEV⁴². By combining imaging and molecular data, we hope to develop a more comprehensive and multi-modal predictive framework. The integration of this model into clinical practice has the potential to significantly enhance patient management. To ensure successful implementation, we would collaborate with multidisciplinary teams, including neuro-oncologists, radiologists, and clinical researchers, to develop standardized protocols for data collection, model application, and outcome evaluation. Regular feedback from clinicians would be incorporated to refine the model and ensure its alignment with real-world clinical needs. By bridging the gap between predictive modeling and clinical decision-making, this approach has the potential to not only improve treatment outcomes but also optimize resource utilization and enhance the overall quality of care for patients with metastatic brain tumors.

Conclusion

This study pioneers a predictive model integrating imaging features and ML to forecast BEV efficacy in treating PTE in metastatic brain tumor patients. Beyond demonstrating high predictive accuracy, our model identifies edema volume as the most critical predictor, offering neurosurgeons and radiologists a data-driven tool to optimize therapeutic strategies. The clinical relevance lies in stratifying patients into likely responders and non-responders before BEV initiation, addressing two major neuro-oncology challenges: avoiding unnecessary treatment risks and reducing the economic burden of ineffective therapy. For clinicians, this model provides actionable insights—enabling personalized treatment plans that prioritize BEV for high-probability responders, while exploring alternatives for predicted non-responders. Integration of such AI-driven decision support systems could significantly enhance precision medicine, particularly in resource-constrained settings where cost-effectiveness is paramount. Future efforts should focus on prospective validation and real-world implementation to refine the model's utility in diverse clinical scenarios.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 9 August 2024; Accepted: 30 April 2025

Published online: 08 May 2025

References

- Nag, S., Manias, J. L. & Stewart, D. J. Pathology and new players in the pathogenesis of brain edema. *Acta Neuropathol* **118**, 197–217. <https://doi.org/10.1007/s00401-009-0541-0> (2009).
- Karimi, E. et al. Single-cell spatial immune landscapes of primary and metastatic brain tumours. *Nature* **614**, 555–563. <https://doi.org/10.1038/s41586-022-05680-3> (2023).
- Bosoi, C. R. & Rose, C. F. Brain edema in acute liver failure and chronic liver disease: Similarities and differences. *Neurochem Int* **62**, 446–457. <https://doi.org/10.1016/j.neuint.2013.01.015> (2013).
- Tripathi, M. et al. The safety and efficacy of bevacizumab for radiosurgery-induced steroid-resistant brain edema; not the last part in the ship of theseus. *Neurol India* **67**, 1292–1302. <https://doi.org/10.4103/0028-3886.271242> (2019).
- Roth, P., Wick, W. & Weller, M. Steroids in neurooncology: Actions, indications, side-effects. *Curr Opin Neurol* **23**, 597–602. <https://doi.org/10.1097/WCO.0b013e32833e5a5d> (2010).
- Taranova, II & Kokhno, V. N. Brain edema treatment procedure using continuous controlled infusion of mannitol in neurosurgical patients. *Anesteziol Reanimatol*, 29–34 (2010).
- Walcott, B. P., Kahle, K. T. & Simard, J. M. Novel treatment targets for cerebral edema. *Neurotherapeutics* **9**, 65–72. <https://doi.org/10.1007/s13311-011-0087-4> (2012).
- Dietrich, J., Rao, K., Pastorino, S. & Kesari, S. Corticosteroids in brain cancer patients: Benefits and pitfalls. *Expert Rev Clin Pharmacol* **4**, 233–242. <https://doi.org/10.1586/ecp.11.1> (2011).
- Velho, V. L., Bhide, A. A. & Ansari, S. A. H. Effect of mannitol irrigation on brain edema in a live rat model. *Asian J Neurosurg* **13**, 766–768. https://doi.org/10.4103/ajns.AJNS_11_18 (2018).
- Tamura, R., Tanaka, T., Miyake, K., Yoshida, K. & Sasaki, H. Bevacizumab for malignant gliomas: Current indications, mechanisms of action and resistance, and markers of response. *Brain Tumor Pathol* **34**, 62–77. <https://doi.org/10.1007/s10014-017-0284-x> (2017).
- Kurkjian, C. & Kim, E. S. Risks and benefits with bevacizumab: Evidence and clinical implications. *Ther Adv Drug Saf* **3**, 59–69. <https://doi.org/10.1177/2042098611430109> (2012).
- Lu, V. M. et al. Efficacy and safety of bevacizumab for vestibular schwannoma in neurofibromatosis type 2: A systematic review and meta-analysis of treatment outcomes. *J Neurooncol* **144**, 239–248. <https://doi.org/10.1007/s11060-019-03234-8> (2019).
- Sadraei, N. H. et al. Treatment of cerebral radiation necrosis with bevacizumab: The Cleveland clinic experience. *Am J Clin Oncol* **38**, 304–310. <https://doi.org/10.1097/COC.0b013e32829c3139> (2015).
- Pillay Smiley, N., Alden, T., Hartsell, W. & Fangusaro, J. Severe radiation necrosis successfully treated with bevacizumab in an infant with low-grade glioma and tumor-associated intractable trigeminal neuralgia. *Pediatr. Blood Cancer*. **63**, 1671–1673. <https://doi.org/10.1002/pbc.26055> (2016).
- Lee, I. et al. Bevacizumab use and the risk of arterial and venous thromboembolism in patients with high-grade gliomas: A nested case-control study. *Pharmacotherapy* **39**, 921–928. <https://doi.org/10.1002/phar.2310> (2019).
- Bertsimas, D. & Wiberg, H. Machine learning in oncology: Methods, applications, and challenges. *JCO Clin Cancer Inform* **4**, 885–894. <https://doi.org/10.1200/cci.20.00072> (2020).
- Hollon, T. C. et al. A machine learning approach to predict early outcomes after pituitary adenoma surgery. *Neurosurg Focus* **45**, E8. <https://doi.org/10.3171/2018.8.Focus18268> (2018).
- Zhan, Y., Hao, Y., Wang, X. & Guo, D. Advances of artificial intelligence in clinical application and scientific research of neuro-oncology: Current knowledge and future perspectives. *Crit Rev Oncol Hematol* **209**, 104682. <https://doi.org/10.1016/j.critrevonc.2025.104682> (2025).
- Bhinder, B., Gilvary, C., Madhukar, N. S. & Elemento, O. Artificial intelligence in cancer research and precision medicine. *Cancer Discov* **11**, 900–915. <https://doi.org/10.1158/2159-8290.Cd-21-0090> (2021).
- Mohammadzadeh, I. et al. Application of artificial intelligence in forecasting survival in high-grade glioma: Systematic review and meta-analysis involving 79,638 participants. *Neurosurg Rev* **48**, 240. <https://doi.org/10.1007/s10143-025-03419-y> (2025).
- Mohammadzadeh, I. et al. Can we rely on machine learning algorithms as a trustworthy predictor for recurrence in high-grade glioma? A systematic review and meta-analysis. *Clin Neurol Neurosurg* **249**, 108762. <https://doi.org/10.1016/j.clineuro.2025.108762> (2025).
- Mohammadzadeh, I. et al. Machine learning for predicting poor outcomes in aneurysmal subarachnoid hemorrhage: A systematic review and meta-analysis involving 8445 participants. *Clin Neurol Neurosurg* **249**, 108668. <https://doi.org/10.1016/j.clineuro.2024.108668> (2025).
- Deng, Q. et al. Predictive value of machine learning models for cerebral edema risk in stroke patients: A meta-analysis. *Brain Behav* **15**, e70198. <https://doi.org/10.1002/brb3.70198> (2025).
- Wu, S. et al. Predicting the emergence of malignant brain oedema in acute ischaemic stroke: A prospective multicentre study with development and validation of predictive modelling. *EclinicalMedicine* **59**, 101977. <https://doi.org/10.1016/j.eclinm.2023.101977> (2023).
- Pu, M. et al. Predictors and outcome of malignant cerebral edema after successful reperfusion in anterior circulation stroke. *J Stroke Cerebrovasc Dis* **32**, 107139. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2023.107139> (2023).

26. Wen, X., Hu, X., Xiao, Y. & Chen, J. Radiomics analysis for predicting malignant cerebral edema in patients undergoing endovascular treatment for acute ischemic stroke. *Diagn Interv Radiol* **29**, 402–409. <https://doi.org/10.4274/dir.2023.221764> (2023).
27. Miao, J. et al. Predictors of malignant cerebral edema in cerebral artery infarction: A meta-analysis. *J Neurol Sci* **409**, 116607. <https://doi.org/10.1016/j.jns.2019.116607> (2020).
28. Meinel, T. R. et al. Multivariable prediction model for futile recanalization therapies in patients with acute ischemic stroke. *Neurology* **99**, e1009–e1018. <https://doi.org/10.1212/wnl.000000000000200815> (2022).
29. Zeng, W. et al. Predicting futile recanalization, malignant cerebral edema, and cerebral herniation using intelligible ensemble machine learning following mechanical thrombectomy for acute ischemic stroke. *Front Neurol* **13**, 982783. <https://doi.org/10.3389/fneur.2022.982783> (2022).
30. Baris, M. M., Celik, A. O., Gezer, N. S. & Ada, E. Role of mass effect, tumor volume and peritumoral edema volume in the differential diagnosis of primary brain tumor and metastasis. *Clin Neurol Neurosurg* **148**, 67–71. <https://doi.org/10.1016/j.clineuro.2016.07.008> (2016).
31. Bitzer, M. et al. Angiogenesis and brain oedema in intracranial meningiomas: Influence of vascular endothelial growth factor. *Acta Neurochir (Wien)* **140**, 333–340. <https://doi.org/10.1007/s007010050106> (1998).
32. Kreisl, T. N. et al. A phase II trial of single-agent bevacizumab in patients with recurrent anaplastic glioma. *Neuro Oncol* **13**, 1143–1150. <https://doi.org/10.1093/neuonc/nor091> (2011).
33. Yuan, W., Sui, L., Xin, H., Liu, M. & Shi, H. Discussion on machine learning technology to predict tacrolimus blood concentration in patients with nephrotic syndrome and membranous nephropathy in real-world settings. *BMC Med Inform Decis Mak* **22**, 336. <https://doi.org/10.1186/s12911-022-02089-w> (2022).
34. Liu, X. et al. Intraoperative goal-directed fluid management and postoperative brain edema in patients having high-grade gliomas resections: A randomized trial. *Int J Surg* <https://doi.org/10.1097/jvs9.0000000000001969> (2024).
35. Minichmayr, I. K. et al. Distribution of bevacizumab into the cerebrospinal fluid of children and adolescents with recurrent brain tumors. *Paediatr Drugs* **26**, 429–440. <https://doi.org/10.1007/s40272-024-00624-y> (2024).
36. Wu, Z., Yang, Y., Chen, M. & Zha, Y. Matrix metalloproteinase 9 expression and glioblastoma survival prediction using machine learning on digital pathological images. *Sci Rep* **14**, 15065. <https://doi.org/10.1038/s41598-024-66105-x> (2024).
37. Zhao, Y. et al. SPCNet: Deep self-paced curriculum network incorporated with inductive bias. *IEEE Trans. Neural Netw. Learn. Syst.* (2025). <https://doi.org/10.1109/tnnls.2025.3544724>
38. Nafi, M. M. I. Predicting C- and S-linked Glycosylation sites from protein sequences using protein language models. *Comput Biol Med* **189**, 109956. <https://doi.org/10.1016/j.combiomed.2025.109956> (2025).
39. Liu, X. et al. Segment any tissue: One-shot reference guided training-free automatic point prompting for medical image segmentation. *Med Image Anal* **102**, 103550. <https://doi.org/10.1016/j.media.2025.103550> (2025).
40. Sharma, A., Low, J. & Mrugala, M. M. Neuro-oncologists have spoken - the role of bevacizumab in the inpatient setting: A clinical and economic conundrum. *Neurooncol Pract* **6**, 30–36. <https://doi.org/10.1093/nop/npy011> (2019).
41. Lin, R. et al. Depleting parenchymal border macrophages alleviates cerebral edema and neuroinflammation following status epilepticus. *J Transl Med* **22**, 1094. <https://doi.org/10.1186/s12967-024-05912-2> (2024).
42. Zhuang, H., Shi, S., Yuan, Z. & Chang, J. Y. Bevacizumab treatment for radiation brain necrosis: mechanism, efficacy and issues. *Mol Cancer* **18**, 21. <https://doi.org/10.1186/s12943-019-0950-1> (2019).

Acknowledgements

The authors acknowledge the following financial support: the National High Level Hospital Clinical Research Funding (2022-PUMCH-C-012); CAMS Innovation Fund for Medical Sciences (2023-I2M-C&T-B-008), and the Teaching Reform Project of Peking Union Medical College (2022ZLGC0120).

Author contributions

All authors made a contribution to the work reported and agreed to be accountable for all aspects of the work. M.F. and W.B.M. designed this study and were responsible for subsequent revisions of the paper. X.X.B. completed related literature retrieval, data acquisition and data analysis. X.X.B. drafted the manuscript. X.X.B., W.B.M., S.Y.W. and M.F. gave critical feedback during the study or during the manuscript. All authors provided final approval of the version and agreed on the journal for publication.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and informed consent

This retrospective study was conducted using existing, fully anonymized clinical data without any patient intervention. The Institutional Review Board (IRB) of Peking Union Medical College Hospital, Chinese Academy of Medical Sciences waived the requirement for informed consent in accordance with Article 39 of China's Ethical Review Measures for Biomedical Research Involving Humans (2016), and the study protocol was approved by the hospital's Academic Ethics Committee. The research strictly adhered to the ethical principles of the Declaration of Helsinki and received formal ethical approval exemption from the IRB (Exemption No. ZS-3931). The study design, ethical review process, and informed consent waiver procedures were fully compliant with Article 39 of the aforementioned Ethical Review Measures and Article 13 of China's Personal Information Protection Law (2021).

Consent for publication

All participating authors give their consent for this work to be published.

Additional information

Correspondence and requests for materials should be addressed to M.F., W.M. or S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025