ORIGINAL ARTICLE

WILEY

# Characterization of immune microenvironment associated with medulloblastoma metastasis based on explainable machine learning

Fengmao Zhao[1]* | Xiangjun Liu[2]* | Jingang Gui[2] | Hailang Sun[1] | Nan Zhang[3] | Yun Peng[2] | Ming Ge[1] | Wei Wang[2]

[1]Department of Neurosurgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China
[2]Laboratory of Tumor Immunology, Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China
[3]Department of Pathology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

**Correspondence**
Wei Wang, Laboratory of Tumor Immunology, Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China.
Email: wei860729@163.com

Ming Ge, Department of Neurosurgery, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China.
Email: geming88@126.com

*These authors have contributed equally to this work.

**Funding source**
National Natural Science Foundation of China, Grant/Award Number: 82101204; Beijing Hospital's Authority Clinical Medicine Development of Special Funding, Grant/Award Number: XMLX202144

Received: 6 June 2024
Accepted: 8 January 2025

## ABSTRACT

**Importance:** Medulloblastoma (MB) is the most common malignant brain tumor in children, with metastasis being the primary cause of recurrence and mortality. The tumor microenvironment (TME) plays a critical role in driving metastasis; however, the mechanisms underlying TME alterations in MB metastasis remain poorly understood.

**Objective:** To develop and validate machine learning (ML) models for predicting patient outcomes in MB and to investigate the role of TME components, particularly immune cells and immunoregulatory molecules, in metastasis.

**Methods:** ML models were constructed and validated to predict prognosis and metastasis in MB patients. Eight algorithms were evaluated, and the optimal model was selected. Lasso regression was employed for feature selection, and SHapley Additive exPlanations values were used to interpret the contribution of individual features to model predictions. Immune cell infiltration in tumor tissues was quantified using the microenvironment cell populations-counter method, and immunohistochemistry was applied to analyze the expression and distribution of specific proteins in tumor tissues.

**Results:** The ML models identified metastasis as the strongest predictor of poor prognosis in MB patients, with significantly worse survival outcomes observed in metastatic cases. High infiltration of CD8+ T cells and cytotoxic T lymphocytes (CTLs), along with elevated expression of the *TGFB1* gene encoding transforming growth factor beta 1 (TGF-$\beta$1), were strongly associated with metastasis. Independent transcriptomic and immunohistochemical analyses confirmed significantly higher CD8+ T cell/CTL infiltration and TGF-$\beta$1 expression in metastatic compared to nonmetastatic MB samples. Patients with both high CD8+ T cell/CTL infiltration and elevated *TGFB1* expression in the context of metastasis exhibited significantly

DOI: 10.1002/ped4.12471

worse survival outcomes compared to patients with low expression and no metastasis.

**Interpretation:** This study identifies metastasis as the key prognostic factor in MB and reveals the pivotal roles of CD8+ T cells, CTLs, and TGF-$\beta$1 within the TME in promoting metastasis and poor outcomes. These findings provide a foundation for developing future therapeutic strategies targeting the TME to improve MB patient outcomes.

**KEYWORDS**

CD8+ T lymphocytes, Machine learning, Medulloblastoma, Metastasis, Transforming growth factor beta 1

## INTRODUCTION

Medulloblastoma (MB) is the most common malignant brain tumor in children and is characterized by significant heterogeneity.[1,2] Integrative genomic analyses have classified MB into four molecular subgroups: Group 3, Group 4, SHH, and WNT, each with distinct metastatic rates and prognostic outcomes.[3,4] The tumor microenvironment (TME), particularly its immune cell components, plays a dual role in tumor progression, either suppressing or promoting tumor growth through direct cell–cell interactions or indirect mechanisms mediated by chemokines and cytokines. Previous studies have revealed that the molecular subgroups of MB exhibit distinct immune cell quantity and proportional profiles. For instance, tumor-associated macrophages and lymphocytes are prominently observed in both murine models and patients with SHH MB.[5–7] Group 3 tumors are characterized by elevated levels of regulatory T cells (Tregs), whereas Group 4 tumors show the highest concentration of natural killer (NK) cells.[8]

Moreover, immune infiltration differences within specific subgroups have been associated with clinical outcomes. For example, high Treg levels in infant SHH MB and elevated monocyte levels in Group 4 MB have been significantly correlated with poorer survival.[8] Despite these insights, most studies have focused primarily on the TME characteristics of specific molecular subgroups, with limited attention given to TME features associated with other critical prognostic factors, such as metastasis. As TME alterations are critical drivers of metastasis, investigating metastasis-associated TME characteristics offers considerable potential for enhancing prognostic accuracy and understanding the mechanisms underlying MB progression.

In recent years, machine learning (ML) models have demonstrated remarkable improvements in predicting various diseases and clinical conditions.[9–11] Compared to traditional statistical methods, ML models excel at capturing complex, nonlinear relationships and identifying previously unrecognized correlations in large datasets.[12] This capability enables ML models to provide deeper insights from clinical data, especially in scenarios with limited prior knowledge about the connections between clinical features and outcomes.[13]

In this study, we implemented an explainable ML model to explore the contributions of diverse clinical and molecular factors to adverse outcomes in MB. Specifically, we sought to identify key factors influencing MB prognosis based on established MB risk stratification schemes.[5–7,14–16] Among these, metastasis emerged as a critical determinant of prognosis. We further utilized the explainable ML model to analyze the contributions of various immune cell populations and effector molecular components within the TME to metastasis. The expression profiles of immune cells and effector molecules driving metastasis progression were systematically identified and characterized.

## METHODS

### Ethical approval

The study was approved by the Human Ethics Committee of Beijing Children's Hospital, affiliated to the Capital Medical University ([2023]-E-113-Y). Research purpose and informed consent had been signed by parents/guardians of all the patients.

### Raw data acquisition and preprocessing

Expression datasets of primary MB tumors, GSE85217 ($n$ = 763) and GSE10327 ($n$ = 62), were obtained from the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/). Raw data were extracted and processed using R software. Probes were mapped to corresponding genes according to the manufacturer's instructions for each microarray platform (e.g., GPL22286, Affymetrix Human Gene 1.1 ST Array). When multiple probes represented the same gene without manufacturer-specific

recommendations, the probe with the highest mean expression level across the dataset was selected for downstream analysis.

Data preprocessing and normalization were performed using the Robust Multiarray Average method implemented in the "affy" package in R. Clinical information, including metastatic status, gender, histology, age, and molecular subgroup, was extracted from associated publications.[17,18] Patients lacking critical clinical data, such as survival or metastasis information, were excluded from the GSE85217 dataset, resulting in a final cohort of 432 MB patients with complete transcriptomic and clinical data for subsequent analysis.

### Risk prediction model development and evaluation

To construct a robust risk prediction model, a voting classifier ensemble was utilized, incorporating multiple ML algorithms, including extreme gradient boosting (XGBoost), bagging decision trees, decision trees, logistic regression, multilayer perceptron, naive Bayes, *k*-nearest neighbors, and support vector machines. Comparative performance evaluation of these models identified XGBoost as the optimal algorithm due to its superior performance on imbalanced datasets and its ability to handle diverse variable types.

The XGBoost algorithm, a nonparametric method based on sequential decision trees, was used to develop predictive models for all-cause mortality and metastasis.[19] XGBoost effectively addresses class imbalance and low event rates without requiring imputation for missing data.[19] Feature selection using Lasso regression identified the top variables contributing to MB survival and metastasis, which were subsequently used as input features for the XGBoost model. Lasso regression improved model interpretability and predictive performance. The dataset was randomly divided into training (70%) and internal validation (30%) subsets using the "caret" package in R. To address class imbalance, an oversampling strategy was applied to enhance the representation of minority classes in the training data.

Hyperparameter tuning was conducted using a grid search with 10-fold cross-validation. Parameters including eta (0.3), max_depth (6), min_child_weight (1), subsample (0.3), and colsample_bytree (1) were optimized over 1000 iterations. Model performance was assessed using the area under the receiver operating characteristic curve (AUC) and accuracy metrics to mitigate biases arising from class imbalance and low event rates.[20]

### Explainable ML model interpretation

To enhance the transparency and interpretability of the XGBoost model, SHapley Additive exPlanations (SHAP) values were employed. SHAP provides a unified framework for quantifying the contribution of each feature to the model's predictions, grounded in cooperative game theory.[21]

For each clinical feature, SHAP values were calculated as the average contribution across all possible feature combinations, enabling detailed insights into the influence of individual variables. Positive SHAP values indicated an increased risk of a specific outcome, whereas negative values signified a decreased risk. SHAP values were visualized at both the global level (across the entire cohort) and the individual level, providing insights into interactions among clinical variables, immune cell markers, and immunological genes. This explainable ML approach facilitated an improved understanding of the mechanisms underlying model predictions.

### Evaluation of immune infiltrating cells

To assess tumor-infiltrating immune and stromal cells in MB, the microenvironment cell populations (MCP)-counter algorithm was utilized. This method estimates the abundance of specific immune and stromal cell populations based on predefined gene expression signatures. MCP-counter generated abundance scores for cell types including CD3+ T cells, CD8+ T cells, cytotoxic T lymphocytes (CTLs), NK cells, B lymphocytes, monocyte lineage cells, myeloid dendritic cells, and neutrophils, as well as stromal components such as endothelial cells and fibroblasts.[22] These abundance scores were calculated independently for each sample, allowing direct comparisons of immune cell enrichment across MB samples.

### Immunohistochemistry

Fourteen MB tumor samples were obtained from Beijing Children's Hospital. Immunohistochemistry was performed to detect the expression of CD8 and transforming growth factor beta 1 (TGF-$\beta$1) proteins in formalin-fixed, paraffin-embedded tissue sections. Briefly, 4 μm-thick sections were deparaffinized with xylene and rehydrated through graded alcohols. Antigen retrieval was performed using a citrate buffer (pH 6.0) in a microwave for 15 min. Endogenous peroxidase activity was blocked with 3% hydrogen peroxide for 10 min. Sections were incubated with primary antibodies at optimized dilutions overnight at 4°C. After washing, slides were treated with biotinylated secondary antibodies and streptavidin-HRP. Immunoreactivity was visualized using 3,3′-diaminobenzidine as the chromogen, followed by counterstaining with hematoxylin. Dehydrated sections were mounted for microscopic evaluation.

Scanned IHC sections were visualized using the Pannoramic 250 Flash III scanner and analyzed with CaseViewer

software (3D HISTECH). Quantitative assessment of staining intensity was performed using Image-Pro Plus version 6.0 software (Media Cybernetics, Inc.), which calculated the area and integrated optical density of the stained sections.

### Statistical analysis

Descriptive statistics were used to summarize patient characteristics in the MB dataset. Group differences were evaluated using the unpaired two-tailed Student's *t*-test or one-way ANOVA for normally distributed variables and the Wilcoxon signed-rank test for non-normally distributed data. A $P$-value $< 0.05$ was considered statistically significant.

Survival analysis was performed using the R software (package "survminer," function "ggsurvplot") to evaluate the overall survival (OS) of MB patients. Kaplan–Meier survival curves were generated to visualize survival probabilities over time. To determine the statistical significance of survival differences between groups, the log-rank test was used as the primary method due to its sensitivity to detecting differences in hazard rates under the assumption of proportional hazards. For survival curves with overlapping segments, which may indicate nonproportional hazards or early differences in survival, the Breslow (Generalized Wilcoxon) test was additionally applied. This test assigns greater weight to early survival events, making it more sensitive in detecting early-phase group differences. By combining these two complementary tests, a comprehensive evaluation of group survival differences was achieved.

## RESULTS

### Clinical factors influencing survival in MB patients

In this study, the publicly available dataset (GSE85217) was analyzed to identify clinical and immune factors influencing prognosis and metastasis in MB patients.[17] The demographic and clinical characteristics, including age, gender, tumor histology, metastatic status, and molecular subgroups, are summarized in Table 1. Kaplan–Meier survival analysis was conducted to evaluate the impact of these characteristics on OS. Prognostic outcomes differed significantly across molecular ($P < 0.01$), histological ($P < 0.01$) subgroups, and metastatic status ($P = 0.02$). Metastatic patients exhibited significantly worse OS compared to nonmetastatic patients. However, no significant differences were observed between age groups ($P = 0.16$) or gender groups ($P = 0.92$) (Figure 1A–E). These findings underscore the critical prognostic influence of molecular subgroups, histological subgroups, and metastatic status, highlighting their potential as prognostic factors in MB.

**TABLE 1** Baseline demographics and clinical characteristics of medulloblastoma patients ($n = 432$)

| Characteristics | *n* (%) |
|---|---|
| Age (years) | |
| 0–3 | 62 (14.35) |
| 4–9 | 224 (51.85) |
| 10–17 | 105 (24.31) |
| ≥18 | 41 (9.49) |
| Gender | |
| Female | 154 (35.65) |
| Male | 278 (64.35) |
| Histology | |
| CMB | 292 (67.59) |
| DNMB | 69 (15.97) |
| LCA | 57 (13.19) |
| MBEN | 14 (3.24) |
| Metastasis | |
| No | 303 (70.14) |
| Yes | 129 (29.86) |
| Molecular subgroup | |
| Group 3 | 81 (18.75) |
| Group 4 | 194 (44.91) |
| SHH | 122 (28.24) |
| WNT | 35 (8.10) |

Abbreviations: CMB, medulloblastoma with classic; DNMB, medulloblastoma with desmoplastic/nodular; LCA, large cell/anaplastic; MBEN, medulloblastoma with extensive nodularity.

### ML model analysis identifies key clinical characteristics influencing survival in MB patients

To comprehensively assess the impact of clinical characteristics on MB survival, a machine learning (ML) model was developed. This model incorporated key factors, including age, gender, metastatic status, and molecular and histological characteristics. Eight ML algorithms were evaluated, with the XGBoost algorithm demonstrating the highest accuracy and AUC (Figure 2A). XGBoost was subsequently used to construct a predictive model for MB survival. The model achieved an AUC of 0.68 and an accuracy of 0.71 in the internal validation subsets (Figure 2B).

Global interpretation of the XGBoost model using SHAP values revealed that metastasis was the most critical factor influencing survival outcomes. Among the top 12 features, metastasis, the large cell/anaplastic histological subtype, and the age group 0–3 years emerged as strong predictors of poorer survival outcomes. Conversely, the WNT molecular subtype, MB with desmoplastic/nodular histological

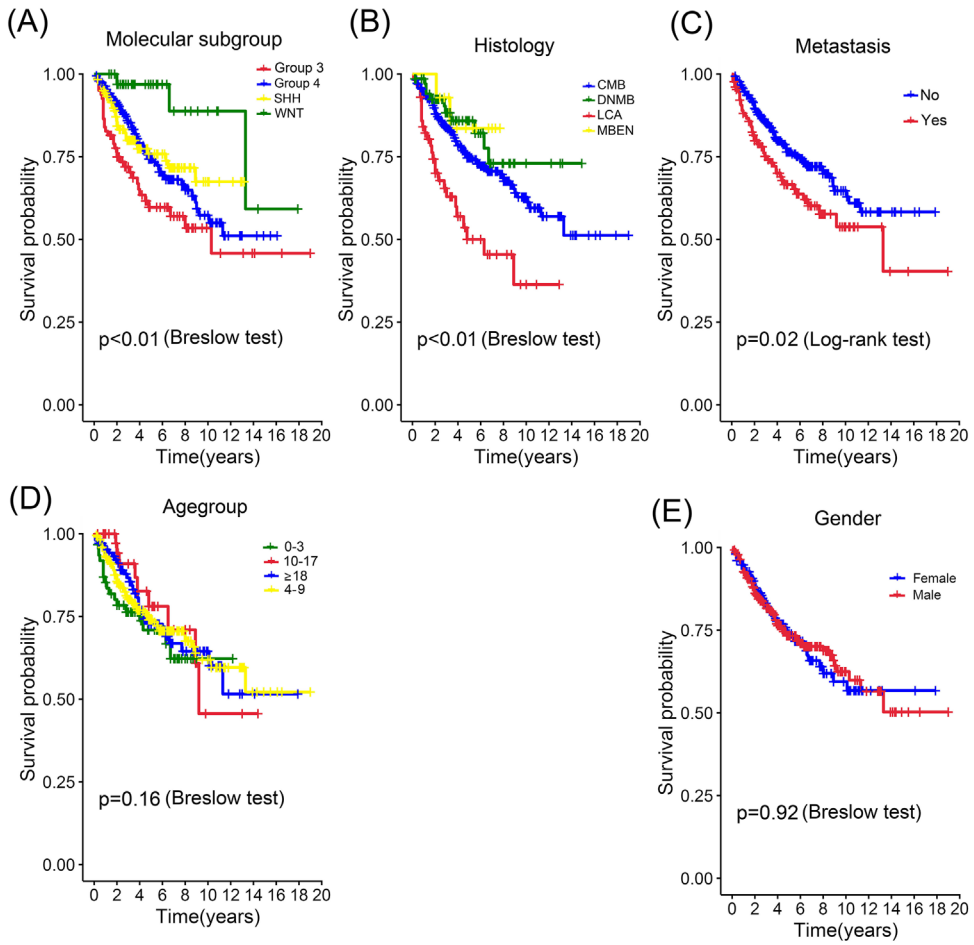


**FIGURE 1** Kaplan–Meier survival curves for overall survival in medulloblastoma patients based on (A) Molecular subgroup, (B) histological subgroups, (C) metastatic status, (D) age, and (E) gender. Statistical significance was tested using the log-rank test, with additional Breslow test applied for curves with overlapping segments.

subtype, and the age group 10–17 years were associated with better survival outcomes (Figure 2C). Other features showed limited predictive value for survival. These results highlight the importance of metastatic status and specific clinical subgroups in shaping survival outcomes and demonstrate the utility of the ML model in identifying high-risk patients to guide personalized treatment strategies.

### Tumor-infiltrating CD8+ T Cells and CTLs influence MB metastasis and survival outcomes

To investigate the role of TME immune cells in MB metastasis, infiltration analysis of 10 TME cell types was performed using the MCP-counter method based on transcriptomic data from 432 MB primary tumors. Unsupervised hierarchical clustering of infiltration scores failed to distinctly separate metastatic and nonmetastatic groups (Figure 3A). Subsequently, ML models were employed to identify immune cell types with predictive value for metastasis. Among the tested algorithms, XGBoost demonstrated superior performance, achieving the highest accuracy and AUC (Figure 3B). Lasso regression further refined the model by selecting nine critical cell types for analysis, excluding endothelial cells due to their negative impact on model accuracy. The optimized XGBoost model achieved an AUC of 0.83, highlighting its robust predictive capability.

Global SHAP analysis identified CD8+ T cells, CTLs, and dendritic cells as the top three immune cell types influencing MB metastasis (Figure 3C). CD8+ T cells emerged as the most crucial factor in predicting metastatic progression in MB. The SHAP dependence plot revealed a nonlinear relationship between immune cell infiltration scores and SHAP values representing the model's contribution. CD8+ T cells, CTLs, and dendritic cell infiltration scores exhibited a nonlinear positive correlation with SHAP values, as revealed by the SHAP dependence plots. This relationship indicates that as the infiltration scores for these immune cell

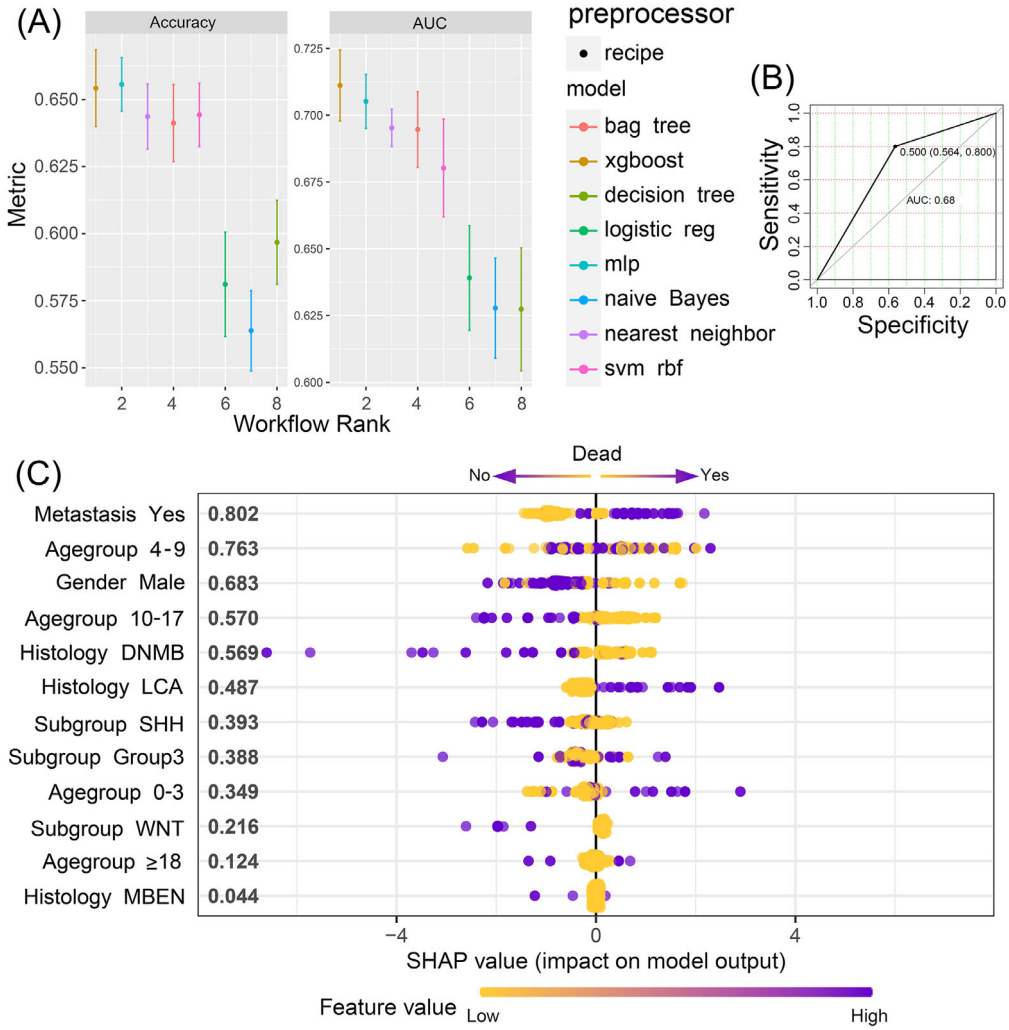


**FIGURE 2** Contribution of clinical features to prognosis prediction evaluated using machine learning (ML) models in medulloblastoma patients. (A) Line plot showing the area under the curve (AUC) and accuracy scores of eight machine learning algorithms based on clinical information. The XGBoost model achieved the highest AUC and accuracy scores among the eight algorithms. (B) Receiver operating characteristic (ROC) curve demonstrating the classification performance of the XGBoost model based on the internal validation subsets for predicting patient mortality (AUC = 0.68). (C) SHAP summary plot highlighting the top 12 clinical features contributing to the XGBoost model for survival prediction. The *y*-axis indicates the rank of feature importance by SHAP values, and the *x*-axis reflects the feature values. DNMB, medulloblastoma with desmoplastic/nodular; LCA, large cell/anaplastic; MBEN, medulloblastoma with extensive nodularity.

types increase, their contributions to the model's metastasis predictions also rise, albeit in a nonlinear manner (Figure 3D).

Further analysis of two MB cohorts (GSE85217 and GSE10327) revealed significantly higher CTL infiltration in metastatic compared to nonmetastatic patients (CTLs $P < 0.01$ and CTLs $P = 0.03$, respectively). However, no significant differences were observed for CD8+ T cells or dendritic cells (Figure 3E,F). We attempted to evaluate the impact of CD8+ T cells, CTLs, and dendritic cells on MB patient survival. There was no significant effect on survival, when stratifying patients based solely on high or low cell infiltration scores (data not shown). The SHAP dependence plot identified thresholds of 3.05 for CD8+ T cells, 4.90 for CTLs, and 3.96 for dendritic cells (SHAP values = 0) as predictive of metastasis (Figure 3D). These thresholds were used to classify high and low infiltration groups, which were combined with metastasis status for survival analysis. The metastatic patients with high infiltration of CD8+ T cells and CTLs had significantly worse OS compared to nonmetastatic patients with low infiltration (CD8+ T cells, $P < 0.01$; CTLs, $P = 0.03$) (Figure 3G). Immunohistochemistry confirmed higher CD8+ T cell levels in metastatic MB samples compared to nonmetastatic samples ($P < 0.01$) (Figure 3H,I). These findings suggest that CD8+ T cells and their activated subset, CTLs, are associated with MB metastasis and poor survival outcomes.

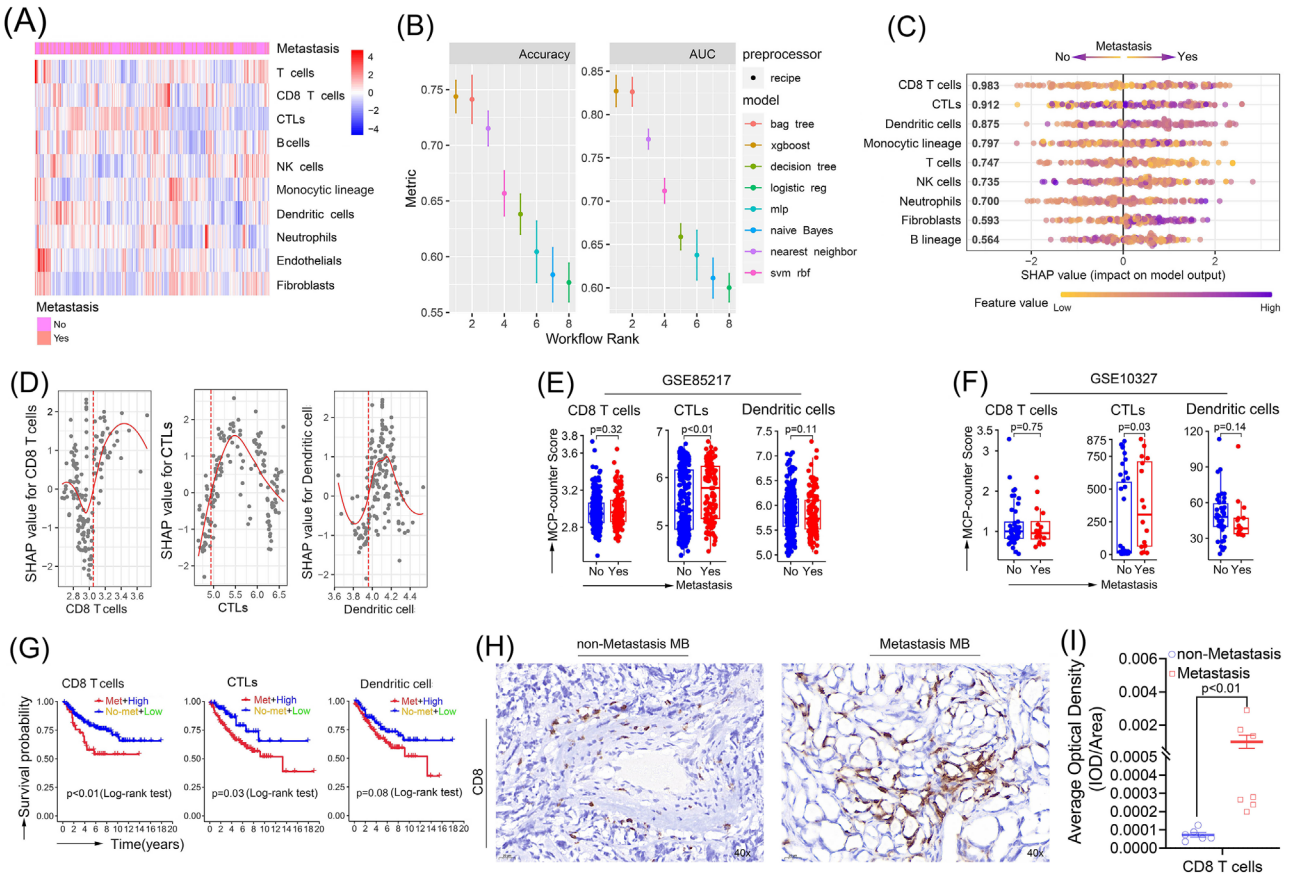


**FIGURE 3** Metastatic-associated immune cell profiles in MB patients. (A) Heatmap showing MCP-counter immune infiltration scores of eight immune cell types and two stromal cell subpopulations, stratified by clinical metastasis status. (B) Line plot illustrating the AUC and accuracy scores of eight machine learning algorithms for predicting metastasis based on MCP-counter immune infiltration scores. (C) SHAP summary plot ranking the importance of immune cell types in the XGBoost model. The plot illustrated the influence of immune cell infiltration scores on SHAP values, with a color gradient (yellow to blue) representing low to high cell infiltration scores. (D) SHAP dependence plot displaying the nonlinear relationship between infiltration scores of the top three immune cell types (CD8+ T cells, CTLs, and dendritic cells) and SHAP values. The red line indicated the cell infiltration score corresponding to a SHAP value of 0. (E and F) Box plots comparing infiltration scores of CD8+ T cells, CTLs, and dendritic cells between metastatic and nonmetastatic MB patients in cohorts (E) GSE85217 and (F) GSE10327. Red points represented metastatic patients, and blue points indicated nonmetastatic patients. (G) Kaplan–Meier survival curves showing overall survival probabilities of MB patients stratified by metastasis status and cell infiltration levels. Statistical significance was tested using the log-rank test. (H) Representative immunohistochemistry staining comparing CD8+ T cell density between metastasis and non-metastasis MB samples. Scale bar, 20 μm. (I) Dot plot comparing CD8+ T cell density scores between MB patients with and without metastasis. Patients with metastasis showed significantly higher CD8+ T cell density scores ($P < 0.01$). AUC, area under the curve; CTLs, cytotoxic T lymphocytes; MB, medulloblastoma; MCP, microenvironment cell populations; NK, natural killer; SHAP, SHapley Additive exPlanations.

## T cell-related regulator TGF-β1 influences MB metastasis and survival outcomes

Given the worse OS associated with high CD8+ T cell and CTL infiltration in metastatic MB, we screened effector molecules regulating T cell activity. Unsupervised clustering of transcriptional profiles did not distinguish metastatic from nonmetastatic groups (Figure 4A). Using XGBoost, predictive models for metastasis based on gene expression were constructed, with *TGFB1*, *HAVCR2*, and *IFNG* identified as the top three genes influencing predictions (Figure 4B,C). SHAP dependence plots revealed a positive correlation between *TGFB1* expression and metastasis predictions (Figure 4D).

Transcriptomic analysis confirmed significantly higher *TGFB1* expression in metastatic MB samples compared to nonmetastatic samples (GSE85217, $P_{\mathrm{TGFB1}} < 0.01$; GSE10327, $P_{\mathrm{TGFB1}} = 0.01$) (Figure 4E,F). The SHAP dependence plot identified expression thresholds of 6.38 for *TGFB1*, 5.80 for *HAVCR2*, and 3.79 for *IFNG* (SHAP values = 0) as the cutoff points distinguishing metastatic from nonmetastatic events (Figure 4D). Kaplan–Meier survival analysis based on these cutoffs showed that MB

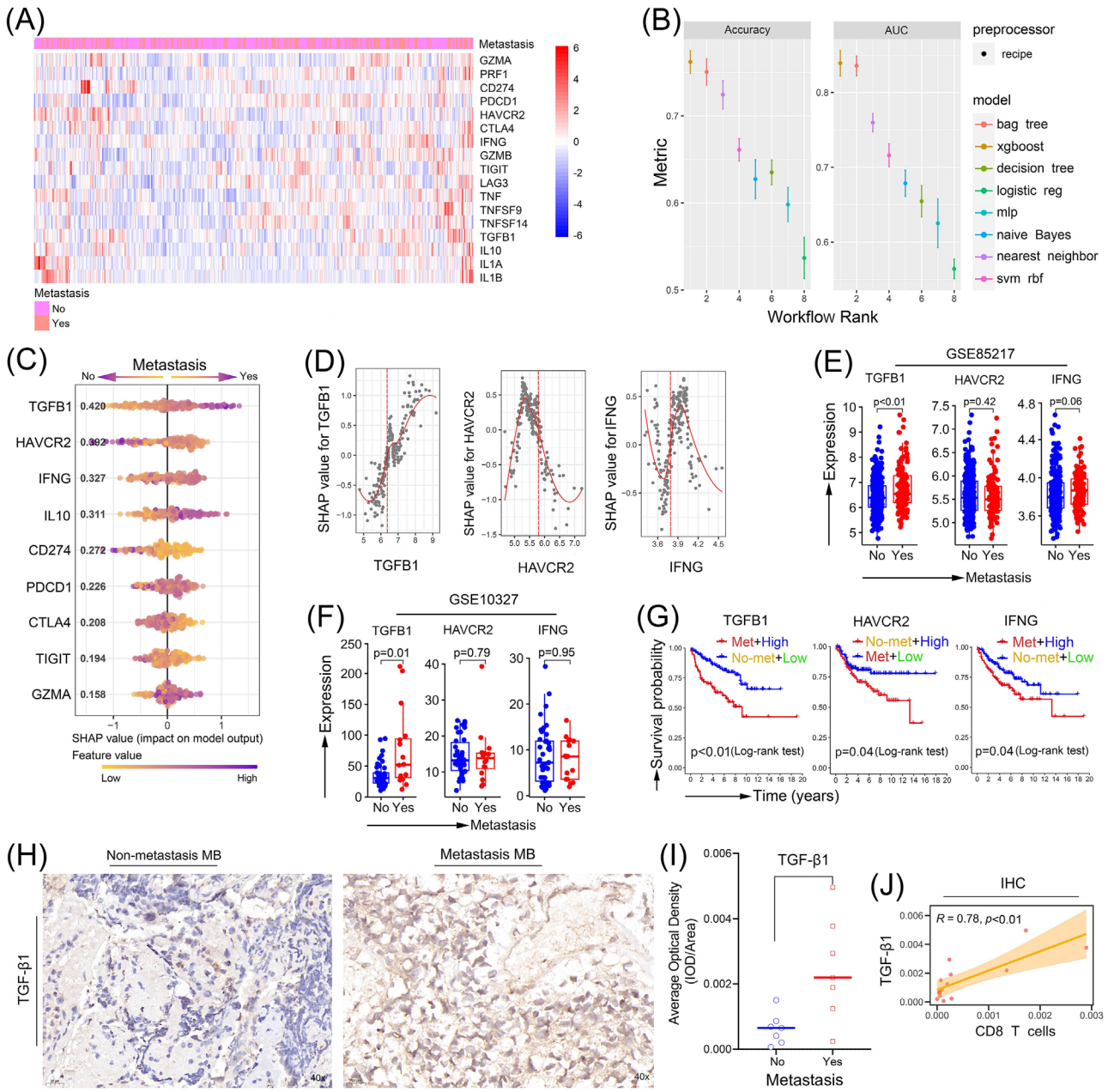


**FIGURE 4** Metastatic-associated T cell-related immune modulators in MB patients. (A) Heatmap displaying the expression levels of 17 immunomodulatory genes, stratified by clinical metastasis status. (B) Line plot illustrating AUC and accuracy scores of eight machine learning algorithms based on the expression of the 17 immunomodulatory genes. (C) SHAP summary plot ranking the importance of the immunomodulatory genes in the XGBoost model. The plot illustrated the influence of immunomodulatory gene expression values on SHAP values, with a color gradient (yellow to blue) representing low to high expression values. (D) SHAP dependence plot displaying the nonlinear relationship between infiltration scores of the top three immunomodulatory genes (*TGFB1*, *HAVCR2*, and *IFNG*) and SHAP values. The red line indicates the gene expression level corresponding to a SHAP value of 0. (E and F) Box plots comparing the expression levels of the top three immunomodulatory genes between metastatic and nonmetastatic MB patients in cohorts (E) GSE85217 and (F) GSE10327. Red points represented metastatic patients, and blue points indicate nonmetastatic patients. (G) Kaplan–Meier survival curves showing overall survival probabilities of MB patients stratified by metastasis status and infiltration levels of *TGFB1*, *HAVCR2*, and *IFNG*. Statistical significance was tested using the log-rank test. (H) Representative immunohistochemistry staining comparing TGF-$\beta$1 expression levels between metastasis and non-metastasis MB samples. Scale bar, 20 μm. (I) Dot plot comparing TGF-$\beta$1 expression density scores between MB patients with and without metastasis. (J) Correlation analysis between CD8+ T cell density scores and TGF-$\beta$1 density scores based on immunohistochemistry results ($r = 0.78$, $P < 0.01$). AUC, area under the curve; MB, medulloblastoma; SHAP, SHapley Additive exPlanations.

patients with high *TGFB1* expression and metastasis had significantly worse survival outcomes compared to non-metastatic patients with low *TGFB1* expression ($P < 0.01$) (Figure 4G). Immunohistochemistry further validated elevated TGF-$\beta$1 levels in metastatic samples ($P = 0.02$) (Figure 4H,I). Correlation analysis showed a positive relationship between TGF-$\beta$1 and CD8+ T cell levels ($r = 0.78$, $P < 0.01$) (Figure 4J), suggesting that elevated TGF-$\beta$1 may contribute to increased CD8+ T cell infiltration in metastatic MB.

## DISCUSSION

Considering the ambiguous contribution and nonlinear interactions of multiple features, this study presented an explainable ML-based risk prediction model for identifying patients with MB at risk of death and metastasis.[21] The use of explainable ML is particularly important in clinical studies due to the need for transparency and interpretability. Although non-explainable models, such as deep neural networks, often deliver higher predictive accuracy by capturing complex data relationships, their "black-box" nature presents challenges in clinical settings, where understanding the decision-making process is crucial for ethical and safety reasons. In contrast, explainable models, like those used in this study, offer greater interpretability, enabling clinicians to understand and trust predictions. By using explainable ML to predict outcomes based on clinical features, immune cell profiles, and cytokine expression, this study provides a clear rationale for model predictions, enhancing confidence in its application for clinical diagnosis and prognostic evaluation.

Among the various clinical features analyzed, metastasis emerged as the most significant factor in predicting mortality in MB patients. Survival analysis consistently demonstrated a poor prognosis for patients with metastatic MB. Although previous studies have largely emphasized the prognostic influence of MB molecular subgroups (Group 3, Group 4, WNT, and SHH) due to their distinct molecular characteristics,[23] our ML-based decision model identified metastasis as the strongest predictor of MB prognosis. This finding highlights the critical role of metastasis in shaping prognosis and offers a valuable tool for enhancing clinical diagnosis and prognostic evaluation in MB patients.

TME played a crucial role in the metastatic progression of tumors.[24] Although increasing evidence suggests that immune cell subsets vary significantly based on the clinical characteristics of MB, limited progress has been made in elucidating the TME factors associated with MB metastasis.[6] The ML model in this study identified CD8+ T cells and CTL subpopulations as the primary immune cell types influencing MB metastasis. This observation is consistent with prior findings showing that Group 3 tumors, which have a higher incidence of metastasis, contain elevated proportions of CD8+ T cells.[5,25] Additionally, earlier studies have explored immune cell infiltration scores and their association with MB outcomes using conventional cutoff thresholds, such as median or average values, to predict adverse events.[6] However, these studies rarely identified that specific immune cell infiltration levels have independent prognostic value. The SHAP dependence plots generated by the ML model provided an objective method for defining threshold values that distinguish predictive events, allowing for precise feature grouping. In our study, the grouping of immune cell characteristics combined with metastasis status revealed significant differences in MB prognosis.

In the metastatic TME, the expression of immunoregulatory molecules often shifts toward an imbalance dominated by inhibitory factors.[26] TGF-$\beta$1, in particular, frequently acts as an immunosuppressive mediator, promoting tumor metastasis and progression by inhibiting the antitumor activity of immune cells.[27,28] Consistent with this, high *TGFB1* expression was identified as a positive contributor to metastasis in the XGBoost prediction model, suggesting its critical role in MB metastasis. Furthermore, the inability of highly infiltrated CD8+ T cells and CTLs in metastatic patients to prevent tumor progression may be partially attributed to the suppressive effects of TGF-$\beta$1. This observation provides a plausible explanation for why these immune cells fail to counteract tumor advancement in MB despite their high infiltration levels.

In conclusion, this study utilized interpretable ML models to evaluate the effects of clinical and TME characteristics on the risks of mortality and metastasis in MB patients. The findings identified metastasis as a key clinical factor associated with poor prognosis, whereas CD8+ T cells, CTLs, and the immunosuppressive factor TGF-$\beta$1 within the TME were revealed to significantly influence MB progression and prognosis. These results not only deepen our understanding of the mechanisms underlying MB progression but also provide potential avenues for improving prognostic evaluation and guiding therapeutic strategies in MB.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

1. Packer RJ, Gajjar A, Vezina G, Rorke-Adams L, Burger PC, Robertson PL, et al. Phase III study of craniospinal radiation therapy followed by adjuvant chemotherapy for newly diagnosed average-risk medulloblastoma. *J Clin Oncol.* 2006;24:4202-4208. DOI: 10.1200/JCO.2006.06.4980

2. Tarbell NJ, Friedman H, Polkinghorn WR, Yock T, Zhou T, Chen Z, et al. High-risk medulloblastoma: a pediatric oncology group randomized trial of chemotherapy before or after radiation therapy (POG 9031). *J Clin Oncol.* 2013;31:2936-2941. DOI: 10.1200/JCO.2012.43.9984
3. Northcott PA, Korshunov A, Witt H, Hielscher T, Eberhart CG, Mack S, et al. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol.* 2011;29:1408-1414. DOI: 10.1200/JCO.2009.27.4324
4. Northcott PA, Shih DJ, Peacock J, Garzia L, Morrissy AS, Zichner T, et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature.* 2012;488:49-56. DOI: 10.1038/nature11327
5. Pham CD, Mitchell DA. Know your neighbors: different tumor microenvironments have implications in immunotherapeutic targeting strategies across MB subgroups. *Oncoimmunology.* 2016;5:e1144002. DOI: 10.1080/2162402X.2016.1144002
6. Bockmayr M, Mohme M, Klauschen F, Winkler B, Budczies J, Rutkowski S, et al. Subgroup-specific immune and stromal microenvironment in medulloblastoma. *Oncoimmunology.* 2018;7:e1462430. DOI: 10.1080/2162402X.2018.1462430
7. Margol AS, Robison NJ, Gnanachandran J, Hung LT, Kennedy RJ, Vali M, et al. Tumor-associated macrophages in SHH subgroup of medulloblastomas. *Clin Cancer Res.* 2015;21:1457-1465. DOI: 10.1158/1078-0432.CCR-14-1144
8. Grabovska Y, Mackay A, O'Hare P, Crosier S, Finetti M, Schwalbe EC, et al. Pediatric pan-central nervous system tumor analysis of immune-cell infiltration identifies correlates of antitumor immunity. *Nat Commun.* 2020;11:4324. DOI: 10.1038/s41467-020-18070-y
9. Macesic N, Bear Don't Walk OJ IV, Pe'er I, Tatonetti NP, Peleg AY, Uhlemann AC. Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data. *mSystems.* 2020;5:e00656-19. DOI: 10.1128/mSystems.00656-19
10. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, et al. Predicting clinical outcomes of large vessel occlusion before mechanical thrombectomy using machine learning. *Stroke.* 2019;50:2379-2388. DOI: 10.1161/STROKEAHA.119.025411
11. Quesada JA, Lopez-Pineda A, Gil-Guillén VF, Durazo-Arvizu R, Orozco-Beltrán D, López-Domenech A, et al. Machine learning to predict cardiovascular risk. *Int J Clin Pract.* 2019;73:e13389. DOI: 10.1111/ijcp.13389
12. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. *Nat Med.* 2020;26:71-76. DOI: 10.1038/s41591-019-0724-8
13. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018;2:749-760. DOI: 10.1038/s41551-018-0304-0
14. Zannoni GF, Ciucci A, Marucci G, Travaglia D, Stigliano E, Foschini MP, et al. Sexual dimorphism in medulloblastoma features. *Histopathology.* 2016;68:541-548. DOI: 10.1111/his.12770
15. Zeltzer PM, Boyett JM, Finlay JL, Albright AL, Rorke LB, Milstein JM, et al. Metastasis stage, adjuvant treatment, and residual tumor are prognostic factors for medulloblastoma in children: conclusions from the Children's Cancer Group 921 randomized phase III study. *J Clin Oncol.* 1999;17:832-845. DOI: 10.1200/JCO.1999.17.3.832
16. Morfouace M, Cheepala S, Jackson S, Fukuda Y, Patel YT, Fatima S, et al. ABCG2 transporter expression impacts group 3 medulloblastoma response to chemotherapy. *Cancer Res.* 2015;75:3879-3889. DOI: 10.1158/0008-5472.CAN-15-0030
17. Cavalli F, Remke M, Rampasek L, Peacock J, Shih D, Luu B, et al. Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell.* 2017;31:737-754. DOI: 10.1016/j.ccell.2017.05.005
18. Ramaswamy V, Taylor MD. Bioinformatic strategies for the genomic and epigenomic characterization of brain tumors. *Methods Mol Biol.* 2019;1869:37-56. DOI: 10.1007/978-1-4939-8805-1_4
19. Ghahremanloo M, Choi Y, Lops Y. Deep learning mapping of surface MDA8 ozone: the impact of predictor variables on ozone levels over the contiguous United States. *Environ Pollut.* 2023;326:121508. DOI: 10.1016/j.envpol.2023.121508
20. Freeman EA, Moisen GG, Coulston JW, Wilson BT. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance. *Can J Forest Res.* 2016;46:323-339. DOI: 10.1139/cjfr-2014-0562
21. Fahmy AS, Csecs I, Arafati A, Assana S, Yankama TT, Al-Otaibi T, et al. An Explainable machine learning approach reveals prognostic significance of right ventricular dysfunction in nonischemic cardiomyopathy. *JACC Cardiovasc Imaging.* 2022;15:766-779. DOI: 10.1016/j.jcmg.2021.11.029
22. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17:218. DOI: 10.1186/s13059-016-1070-5
23. Goddard J, Castle J, Southworth E, Fletcher A, Crosier S, Martin-Guerrero I, et al. Molecular characterisation defines clinically-actionable heterogeneity within Group 4 medulloblastoma and improves disease risk-stratification. *Acta Neuropathol.* 2023;145:651-666. DOI: 10.1007/s00401-023-02566-0
24. Quail DF, Joyce JA. The microenvironmental landscape of brain tumors. *Cancer Cell.* 2017;31:326-341. DOI: 10.1016/j.ccell.2017.02.009
25. Pham CD, Flores C, Yang C, Pinheiro EM, Yearley JH, Sayour EJ, et al. Differential immune microenvironments and response to immune checkpoint blockade among molecular subtypes of murine medulloblastoma. *Clin Cancer Res.* 2016;22:582-595. DOI: 10.1158/1078-0432.CCR-15-0713
26. Liu W, Almo SC, Zang X. Co-stimulate or co-inhibit regulatory T cells, which side to go. *Immunol Invest.* 2016;45:813-831. DOI: 10.1080/08820139.2016.1186690

27. Li MO, Wan YY, Sanjabi S, Robertson AK, Flavell RA. Transforming growth factor-beta regulation of immune responses. *Annu Rev Immunol.* 2006;24:99-146. DOI: 10.1146/annurev.immunol.24.021605.090737
28. Ouyang W, Rutz S, Crellin NK, Valdez PA, Hymowitz SG. Regulation and functions of the IL-10 family of cytokines in inflammation and disease. *Annu Rev Immunol.* 2011;29:71-109. DOI: 10.1146/annurev-immunol-031210-101312

**How to cite this article:** Zhao F, Liu X, Gui J, Sun H, Zhang N, Peng Y, et al. Characterization of immune microenvironment associated with medulloblastoma metastasis based on explainable machine learning. *Pediatr Investig*. 2025;9:59–69. https://doi.org/10.1002/ped4.12471