scientific reports

Check for updates

OPEN

MRI radiomics based on machine learning in high-grade gliomas as a promising tool for prediction of CD44 expression and overall survival

Mingjun Yu[1,2], Jinliang Liu[1], Wen Zhou[3], Xiao Gu[4]✉ & Shijia Yu[5]✉

We aimed to predict CD44 expression and assess its prognostic significance in patients with high-grade gliomas (HGG) using non-invasive radiomics models based on machine learning. Enhanced magnetic resonance imaging, along with the corresponding gene expression and clinicopathological data, was downloaded from online database. Kaplan–Meier survival curves, univariate and multivariate COX analyses, and time-dependent receiver operating characteristic were used to assess the prognostic value of CD44. Following the screening of radiomic features using repeat least absolute shrinkage and selection operator, two radiomics models were constructed utilizing logistic regression and support vector machine for validation purposes. The results indicated that CD44 protein levels were higher in HGG compared to normal brain tissues, and CD44 expression emerged as an independent biomarker of diminished overall survival (OS) in patients with HGG. Moreover, two predictive models based on seven radiomic features were built to predict CD44 expression levels in HGG, achieving areas under the curves (AUC) of 0.809 and 0.806, respectively. Calibration and decision curve analysis validated the fitness of the models. Notably, patients with high radiomic scores presented worse OS ($p < 0.001$). In summary, our results indicated that the radiomics models effectively differentiate CD44 expression level and OS in patients with HGG.

Keywords Radiomics, CD44, High-grade glioma, Overall survival, Machine learning

Abbreviations

AUC	The area under the curve
CD44	CD44 antigen
Chr	Chromosome
DCA	Decision curve analysis
DEG	Differentially expressed gene
EMT	Epithelial-to-mesenchymal transition
GO	Gene ontology
HGG	High-grade glioma
ICC	Intraclass correlation coefficient
IDH	Isocitrate dehydrogenase
IHC	Immunohistochemistry
KEGG	The Kyoto Encyclopedia of Genes and Genomes
LASSO	Least absolute shrinkage and selection operator
LGG	Low-grade glioma
LR	Logistic regression

[1]Department of Neurosurgery, Shengjing Hospital of China Medical University, Shenyang 110004, People's Republic of China. [2]Department of Medical Affairs, Shengjing Hospital of China Medical University, Shenyang 110004, People's Republic of China. [3]Department of Pain Management, Dalian Municipal Central Hospital, Dalian 116033, People's Republic of China. [4]Department of Oncology, Shengjing Hospital of China Medical University, Shenyang 110004, People's Republic of China. [5]Department of Neurology, Shengjing Hospital of China Medical University, Shenyang 110004, People's Republic of China. ✉email: tulipbc@163.com; sjyu@cmu.edu.cn

MGMT	O^6-methylguanine DNA methyltransferase promoter
ML	Machine learning
MRI	Magnetic resonance imaging
NBTs	Normal brain samples
NPV	Negative predictive value
OS	Overall survival
PPV	Positive predictive value
PR	Precision recall
Q1	The 25th percentile
Q3	The 75th percentile
ROC	Receiver operating characteristic
RS	Radiomic score
SVM	Support vector machine
T1WI+C	T1WI contrast-enhanced
TCGA	The Cancer Genome Atlas
TCIA	The Cancer Imaging Archive
TME	Tumor microenvironment
VOI	The volume of interest

Glioma is the most common primary malignancy occurring in the central nervous system (CNS)[1]. Despite advancements in surgery, chemotherapy, radiotherapy, and targeted therapy for glioma in recent decades, it remains a highly lethal disease due to limited treatment efficacy. Gliomas are categorized into grade I to IV based on the histologic and molecular features[2]. Generally, grade I and II gliomas are categorized as low-grade glioma (LGG) and have a relatively favorable prognosis. In contrast, grade III and IV gliomas are classified as high-grade gliomas (HGG), which are highly malignant, aggressive and associated with a miserable prognosis. According to the fifth edition of the WHO Classification of Tumors of CNS, the mutational status of IDH is one of the most crucial biomarkers for gliomas, as it helps distinguish GBM from both astrocytoma and oligodendroglioma[3]. Currently, the prognostic indicators of glioma include clinicopathologic features, isocitrate dehydrogenase (IDH) status, O^6-methylguanine DNA methyltransferase promoter (MGMT) status, Chromosome (Chr) 1p19q codeletion and radiomics data of computer tomography (CT) and magnetic resonance imaging (MRI). However, these indicators are insufficient for precision medicine due to the heterogeneity of gliomas. Consequently, there is an urgent need to investigate novel predictors to improve prognostic stratification and personalize medical decision-making.

The protein CD44 antigen (CD44) is a glycoprotein located on the cell surface, which functions as a receptor for extracellular matrix components such as hyaluronan, collagen, and matrix metalloproteinases[4,5]. It participates in various biological functions, including cell adhesion, migration, lymphocyte activation, cell-cell interactions, hematopoiesis, and inflammation[6–9]. A prior study discovered that high CD44 expression is linked to diminished survival rates in patients with grade II and III glioma, indicating its potential as a prognostic marker[10]. Several compounds targeting CD44 yielded promising antitumor effects[11–14].

For instance, a phase II clinical study highlighted the efficacy and safety of Angstrom6 in ovarian cancer by modulating CD44-mediated signaling[11]. Nevertheless, detection of CD44 through blood is inconvenient, expensive, and subjected to spatio-temporal heterogeneity with tumor. Moreover, detection of CD44 by tissue sample is invasive, costly, expensive, and influenced by factors such as antibody specificity and operator variability. Consequently, an efficient, non-invasive, and repeatable method to predict CD44 expression in patients with HGG is urgently needed.

Given its low cost, non-invasiveness, and effectiveness, MRI is routinely implemented in the patients with glioma. Radiomics offers a method for quantitatively, dynamically, and non-invasively reporting tumor characteristics based on advanced imaging features by analyzing a vast array of radiographic characteristics[15]. Machine learning (ML) has been extensively applied in clinical diagnosis and precision medicine, especially in the fields of radiology and oncology. Increasing studies have indicated the potential of non-invasive evaluation via radiomics for early diagnosis and molecular typing of HGG, as well as for predicting the tumor microenvironment (TME) and heterogeneity[16–18]. However, there have been no published studies on predicting CD44 expression in patients with HGG using radiomics.

In the present study, we aimed to build a non-invasive radiomics model by ML to predict CD44 expression in patients with HGG. This was achieved by utilizing data from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA) databases. Additionally, we explored the potential relationship between CD44 and the tumor microenvironment (TME), which could serve as a valuable predictive tool to enhance clinical decision-making and precision medicine.

Materials and methods
Data collection
MRI images, along with corresponding gene expression data and clinicopathological data of patients with HGG, were downloaded from the TCIA (https://www.cancerimagingarchive.net/) and TCGA (https://portal.gdc.cancer.gov/) databases. The inclusion criteria for this study were: (1) newly diagnosed HGG; (2) pathologically confirmed grade III or IV gliomas; (3) availability of axial T1WI contrast-enhanced (T1WI+C) images of diagnostic quality. The exclusion criteria were: (1) absence of gene expression data, clinical information, and survival data; (2) survival time of less than 30 days; and (3) low image quality. The schematic drawing of the criteria was listed in Supplementary Figure S1.

For each patient, we collected the following variables: age (< 60 or ≥60 years), gender (female or male), grade (III or IV), IDH status (wild-type or mutated), Chr 1p19q codeletion status (codeletion or non-codeletion), the methylation status of MGMT promoter (unmethylated/unknown or methylated), chemotherapy (no or yes), radiotherapy (no or yes), expression level of CD44, survival time, and survival outcome (alive or death). The requirement for ethical approval was waived.

Expression analysis of CD44

The RNAseq data (TPM) from TCGA and Genotype-Tissue Expression (GTEx) processed by the TOIL project[19], were accessed from UCSC Xena (https://xenabrowser.net/datapages/). All the patients were divided into CD44high and CD44low groups based on the cutoff expression level of CD44 determined by the survminer R package. Differential CD44 expression analysis was performed between HGG samples and normal brain samples (NBTs), following the log2 transformation of the RNAseq data.

Immunohistochemistry (IHC)

We obtained a total of 4 HGG tissue samples and 4 NBTs from patients with epilepsy and brain injury in the Department of Pathology and Neurosurgery, Shengjing Hospital of China Medical University. After fixation in 10% formalin, the samples were paraffin-embedded and sectioned. Subsequently, IHC staining with an anti-CD44 antibody (Proteintech, # 60224-1-Ig, dilution 1:200) was performed to assess CD44 expression. Informed consent was obtained from all patients, and the study was approved by the Ethics Committee of Shengjing Hospital (No. 2019PS105K).

Survival and enrichment analysis of CD44

OS was assessed and plotted using Kaplan–Meier survival curves. Cox univariate and multivariate regression analyses were employed to evaluate the influence of various variables on OS. Moreover, we conducted subgroup analyses and interaction tests to evaluate the effects of CD44. The time-dependent receiver operating characteristic (ROC) curves were performed to estimate the accuracy and discrimination of CD44 in predicting OS at 12, 24, and 36 months of follow-up. To determine the correlations between CD44 and clinical features, we calculated Spearman rank coefficients.

We applied CIBERSORTx (https://cibersortx.stanford.edu/) to deconvolute of the expression matrix of HGG samples, thereby obtaining the distribution of 20 distinct immune cell types in HGG tissues of each sample[20]. We conducted Wilcoxon test to assess the immune cell abundances between the CD44high and CD44low groups by the R package limma[21]. To identify the different functions and relevant pathways between the CD44high and CD44low groups, we conducted gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis on differentially expressed genes (DEGs) by the R package clusterProfiler with q value < 0.05[22].

Development of a radiomics models to predict CD44 expression level in patients with HGG

Totally, we obtained 82 samples by intersecting the TCGA bioinformatics data with TCIA image data. The imaging parameters were as follows: Slice Thickness: 0.9–6.0 mm; Pixel_Spacings: 0.4–1.1 mm; Spacing Between Slices: 1.0–7.5 mm; RepetitionTime: 6.6-3252.8 ms; Echo Time: 2.75-20.0 ms. We applied N4 bias field correction to the MR images to reduce signal inhomogeneity caused by magnetic field variations, which can lead to regions appearing too bright or dark, or inconsistencies in brightness within the same tissue. Images were resampled to achieve isotropy, minimizing variations due to differences in scanning equipment and protocols, as well as lesion size. Image normalization was also implemented to reduce noise and standardize intensity, thereby minimizing differences in image signal intensity across different machines. Patients were divided into CD44high and CD44low groups based on the cutoff expression level of CD44 determined by the survminer R package.

Two experienced physicians manually outlined entire tumor areas in the MRI enhanced T1-weighted images (T1WI)[23] using three-dimensional slicer version 4.10.2. The Volume of Interest (VOI) delineation was performed by two radiologists who were blinded to the clinical data. One radiologist completed the VOI delineation for all cases, while the second radiologist randomly selected 20 samples for independent delineation. A total of 107 radiomic features were extracted by PyRadiomics and subsequently normalized. The intraclass correlation coefficient (ICC) was utilized to evaluate the reliability of the radiomic features between the two physicians, and features with ICC ≥ 0.75 were chosen for subsequent analyses.

Repeat least absolute shrinkage and selection operator (LASSO) was carried out to further screen the radiomic features by R package glmnet[24]. Features that occurred more than 600 times during the 1000 screening were included in the final model. After feature extraction, Z-score normalization was utilized.

Logistic Regression (LR) and Support vector machine (SVM) were used to predict the expression of CD44. The diagnosis discrimination of radiomics models was assessed by plotting the ROC and precision recall (PR) curves. The calibration curve was generated to evaluate the goodness of fit of radiomics models. Brier score was used to determine the overall performance of radiomics models. The clinical benefit was determined by a decision curve analysis (DCA). To identify the predicted CD44 levels by the radiomic models, we calculated radiomic score (RS). The Wilcoxon test was utilized to assess the difference of RS between CD44high and CD44low groups. To determine the correlations among RS, CD44 and genes of EMT, we performed Spearman correlation analysis.

We obtained 170 samples by intersecting TCGA clinical data and TCIA imaging data. The 170 patients were divided into dichotomous variable (i.e., RShigh and RSlow groups) based on the cutoff value of RS derived from SVM model determined by R package survminer. Log-rank tests were used to survival analysis.

Statistics analysis

The quantitative data were described as mean ± standard deviation. The categorical data were described as frequencies and proportions, and then compared using χ^2 test at baseline. Statistical significance was considered as a two-sided p-value < 0.05. The T1WI + C-clinical prediction model was visualized through a nomogram by the rms R package. All statistical analyses were performed by R software version 4.1.1.

Results

Patient characteristics

The flow-chart of this study was presented in Fig. 1. A cohort of 310 HGG patients from the TCGA databases were included in survival analyses. Patients were separated into CD44^{high} (n = 155) and CD44^{low} (n = 155) groups based on the median expression of CD44 (5.043). The baseline characteristics of the patients were detailed in Supplemental Table 1.

Clinical significance of CD44

To elucidate the clinical relevance of CD44, we initially investigated its expression in HGG samples and NBTs. The results showed that CD44 expression was markedly elevated in HGG samples (p < 0.001, Fig. 2A). Furthermore, we conducted immunohistochemistry (IHC) experiments on HGG patient samples and NBTs, confirming that CD44 protein levels were elevated in HGG compared to NBTs (Fig. 2B). The median OS in the CD44^{low} group was 44.6 months, compared to just 16.2 months in the CD44^{high} group. Patients in the CD44^{low} group exhibited better OS than those in the CD44^{high} group (p < 0.001, Fig. 2C). Additionally, we analyzed the impact of various clinical characteristics to the OS of HGG patients. The results indicated that low age (< 60), histologic grade, chemotherapy, IDH mutant, Chr 1p19q codeletion, and methylated MGMT promoter were associated with longer OS (Supplementary Figure S2). The time-dependent ROC curve indicated that the area under the curve (AUC) of the CD44 expression was 0.719 for 1-year OS prediction (Fig. 2D).

Further, the univariate Cox analysis revealed that compared to low CD44 expression, high CD44 expression was significantly correlated with worse OS (HR = 3.025, 95% CI: 2.203–4.154, p < 0.001). Expectedly, low age (< 60), histologic grade, chemotherapy, IDH mutant, Chr 1p19q codeletion, and methylated MGMT promoter presented similar trend. In multivariate Cox analysis, high CD44 expression emerged as an independent factor of shorter OS compared to low CD44 expression (HR = 1.718, 95% CI: 1.189–2.484, p = 0.004, Fig. 2E). Moreover, low histologic grade, chemotherapy, and IDH mutant were identified as independent factors affecting OS in HGG patients. Subgroup analyses revealed that most subgroups with high CD44 expression exhibited unfavorable OS. In exact, significant interaction effects of CD44 expression on the OS were seen in chemotherapy and radiotherapy subgroups (Fig. 2F). The heatmap illustrating the correlation between CD44 expression and clinical characteristics revealed that CD44 expression was negatively and significantly related to IDH status, Chr 1p19q codeletion, and MGMT promoter status (all p < 0.001, Fig. 2G).

Immune infiltration and enrichment analysis

Subsequently, we delved into the distinctions in the tumor immune microenvironment between the CD44^{high} and CD44^{low} groups. As shown in Fig. 3A, the violin plots revealed a higher infiltration of M2 macrophages, monocytes, and CD4 memory resting T cells in the CD44^{high} group compared with the CD44^{low} group (p < 0.001).

To further understand the biological functions and mechanisms of CD44 in HGG, we performed GO and KEGG analyses of DEGs (listed in Supplemental Table 2) between the CD44^{high} and CD44^{low} groups. Synapse organization, focal adhesion, and transcription coregulator activity was significantly enriched in GO analysis (p < 0.05, Fig. 3B). KEGG pathway enrichment analysis unveiled several enriched pathways, including lysosome, Epstein-Barr virus infection, adherens junction, and Human immunodeficiency virus 1 infection (p < 0.05, Fig. 3C).

Construction and evaluation of radiomics models

Totally, 82 patients from the intersection of TCGA and TCIA datasets were enrolled for radiomics analysis. Patients were separated into CD44^{high} and CD44^{low} groups based on the median expression of CD44 (5.043). The ICCs of the features extracted by PyRadiomics were listed in Supplemental Table 3. The median of ICC was 0.938, and 93 features (86.9%) had an ICC of ≥ 0.75, which were enrolled for further feature selection. After repeat LASSO, seven features with frequencies > 600 remained in the model, including shape_sphericity, shape_sphericity, gray level size zone matrix (glszm)_gray level non uniformity normalized, and gray level run length matrix (glrlm)_long run high gray level emphasis, among others (Fig. 4).

Subsequently, we constructed a radiomics model based on LR. The importance of seven features were showed in Fig. 5A. The radiomics score (RS) was calculated using the formula: RS = -0.01933 + 0.880206×shape_sphericity + 0.072922×glszm_gray level non uniformity normalized + 0.165961 × glrlm_long run high gray level emphasis-1.05154 × neighborhood gray tone difference matrix (ngtdm)_coarseness-0.5184 × glszm_large area low gray level emphasis-0.44085 × shape_surface area + 0.483907×first order mean. The model demonstrated an accuracy of 0.768, with sensitivity of 0.829, specificity of 0.707, positive predictive value (PPV) of 0.739, and negative predictive value (NPV) of 0.806. The predictive performance of the model was satisfactory, achieving an AUC of 0.806 (Fig. 5B). Moreover, the AUC in the 10-fold cross-validation and PR curve reached 0.751 and 0.772 (Fig. 5C,D).

The calibration curve and Hosmer-Lemeshow test indicated a favorable consistency between the model-predicted probabilities of CD44 high expression and the actual values (p = 0.161, Fig. 5E). The DCA curves suggested our radiomics model provided higher net benefit than all treatments and none treatment within a threshold probability was 0.1–0.7 (Fig. 5F). Notably, the RS was significantly higher in the CD44^{high} group than in the CD44^{low} group (p < 0.001, Fig. 5G).
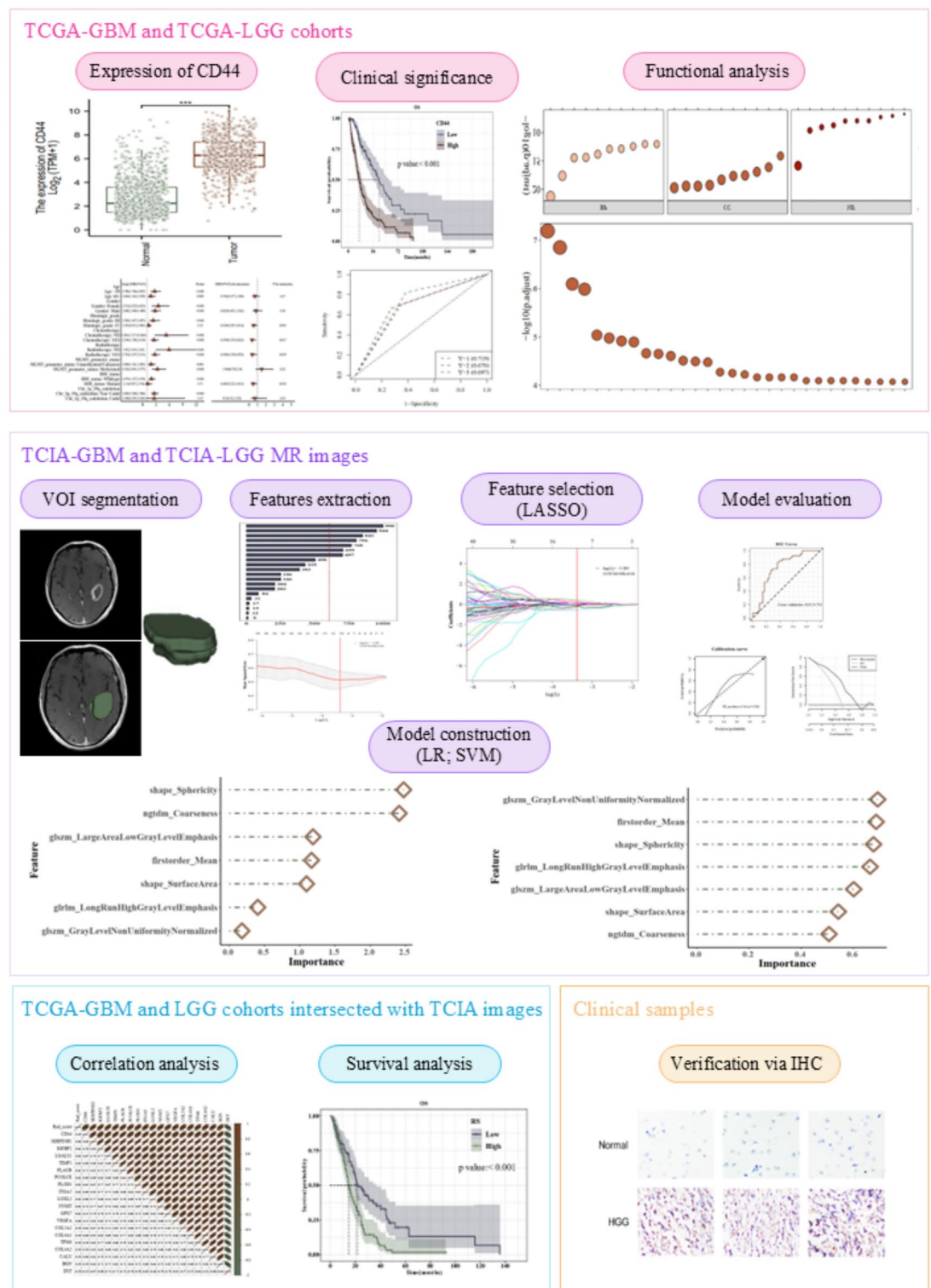
Fig. 1. The brief flowchart of the study process.

To further investigate the optimal radiomics model for predicting CD44 expression in HGG, we developed a radiomics model utilizing SVM. The importance of seven features were showed in Fig. 6A. The SVM-based radiomics model achieved an accuracy of 0.780, with sensitivity at 0.854, specificity at 0.707, PPV at 0.745, and NPV at 0.829. Similarly, the predictive performance of the model was good with an AUC of 0.809 (Fig. 6B). In addition, the AUC of 10-fold cross-validation and the PR curve reached 0.778 and 0.771. (Fig. 6C,D). The calibration curve and Hosmer-Lemeshow test indicated a favorable consistency between the predictive CD44 expression and the true value ($p = 0.194$, Fig. 6E). The DCA curves suggested that our SVM-based radiomics
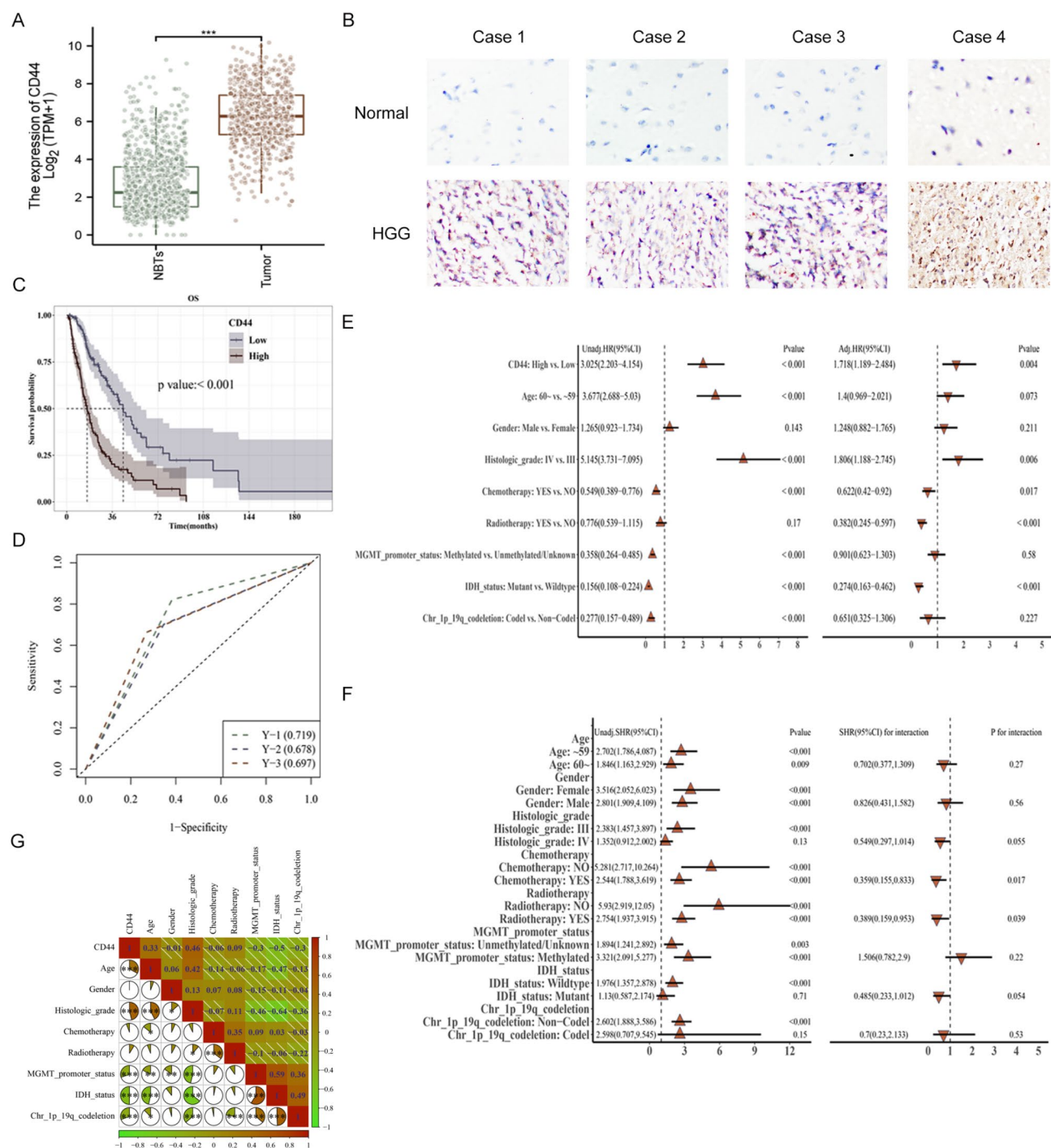
Fig. 2. Expression of CD44 elevated and predicted poor OS in HGG. (**A**) CD44 was high expression in HGG tissues. (**B**) IHC analysis revealed that CD44 was upregulated in HGG compared to that in normal brain tissues. Images were captured at × 100. (**C**) Patients in the CD44^{low} group had better OS. (**D**) The ROC curves of CD44 in predicting OS at 12, 24, and 36 months of follow-up. (**E**) Cox univariate and multivariate regression analyses of the effects of clinicopathological variables on OS. (**F**) Subgroup analyses and interaction tests of the effects of CD44 expression level on OS. (**G**) The correlations between CD44 expression and clinical characteristics of HGG. CI, confidence interval; HGG, high-grade glioma; HR, hazard ratio; IDH, isocitrate dehydrogenase; IHC, immunohistochemistry; MGMT, O^6-methylguanine-DNA-methyltransferase; OS, overall survival; ROC, time-dependent receiver operating characteristic. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.
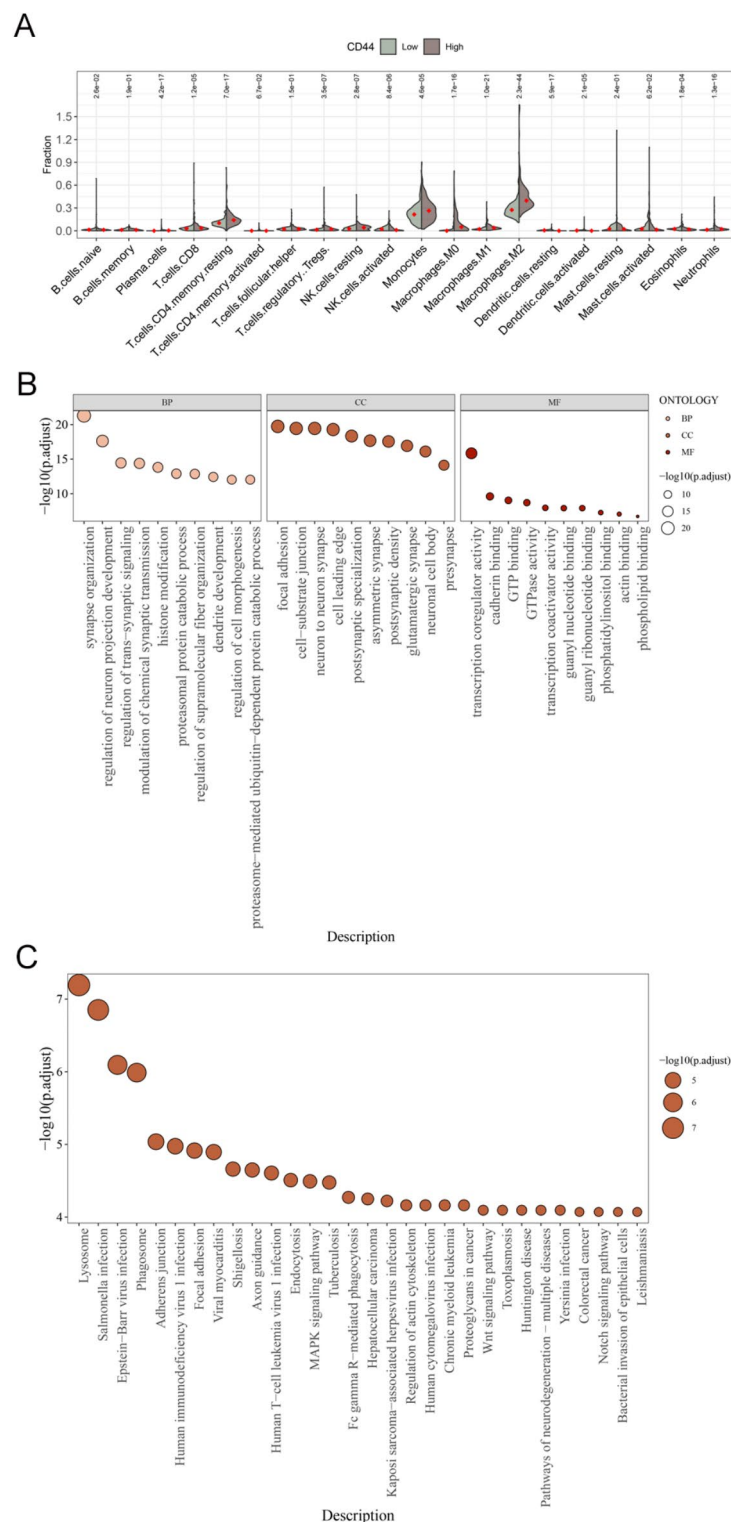
Fig. 3. The immune and functional analyses of DEGs between the CD44^{high} and CD44^{low} groups. (**A**) The correlations of CD44 expression and immune cell infiltration in HGG. (**B**) The GO analyses of DEGs between the CD44^{high} and CD44^{low} groups. (**C**) The KEGG analyses of DEGs between the CD44^{high} and CD44^{low} groups. BP, biological process; CC, cell component; DEGs, differentially expressed genes; GO, gene ontology; HGG, high-grade glioma; KEGG, the Kyoto Encyclopedia of Genes and Genomes; MF, molecular function.
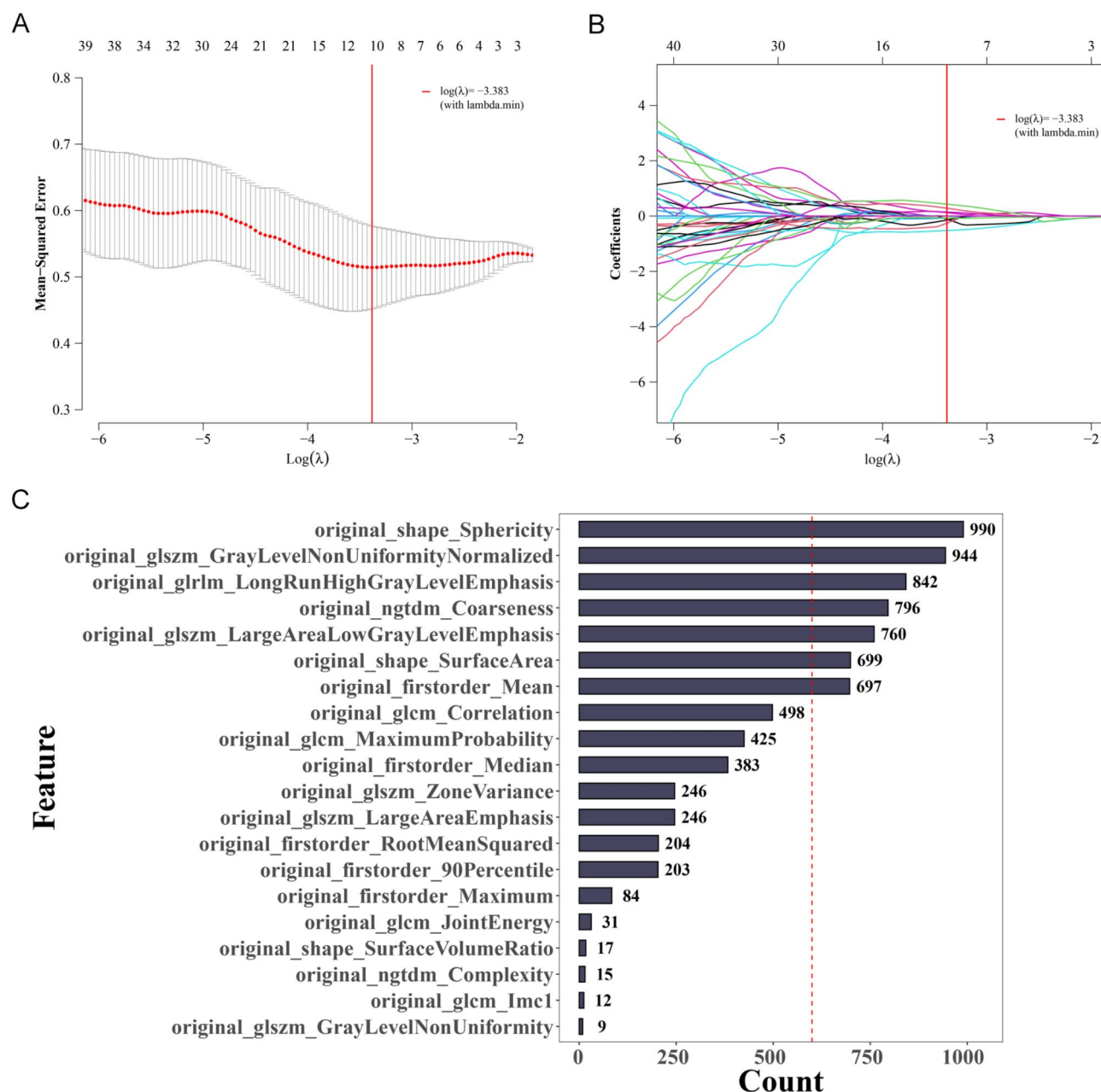
Fig. 4. Selection of radiomic features by LASSO. (**A**) Plot of the LASSO coefficient profiles of the selected features. The vertical line presented the optimal λ value. (**B**) Selection of tuning parameter (λ) in the LASSO model. The vertical line presented the optimal λ value. (**C**) Visualization the top 20 radiomic features with the most counts. The dashed line presented 600 counts. LASSO, least absolute shrinkage and selection operator.

model provided a higher net benefit than all treatments and none treatment within a threshold probability was 0.1–0.8 (Fig. 6F). The RS was significantly higher in the CD44[high] group than in the CD44[low] group ($p < 0.001$, Fig. 6G). Although the AUC for the SVM model was higher than that for LR model, no significant difference was found between the two models by Delong test. The RS of SVM model was chosen for further analyses due to its relatively higher accuracy.

In addition to CD44, Spearman correlation analysis illustrated that RS was also significantly correlated to some molecules involved in EMT ($p < 0.05$), including serpin H1 (SERPINH1), insulin like growth factor binding protein 2 (IGFBP2), galectin-1 (LGALS1). (Fig. 6H). To investigate the prognostic implications of the radiomics model, we obtained 170 patients' data with HGG from TCGA database for subsequent survival analysis. All the patients were divided into RS[high] ($n = 85$) and RS[low] ($n = 85$) groups based on the cutoff value of 0.532. The baseline clinicopathological characteristics of the patients were presented in Supplemental Table 4. As shown in Fig. 6I, the median OS was significantly elevated in RS[low] group (21.6 months) compared to RS[high] group (14.9 months), indicating that patients in RS[high] group had poorer OS than those in RS[low] group ($p < 0.001$).
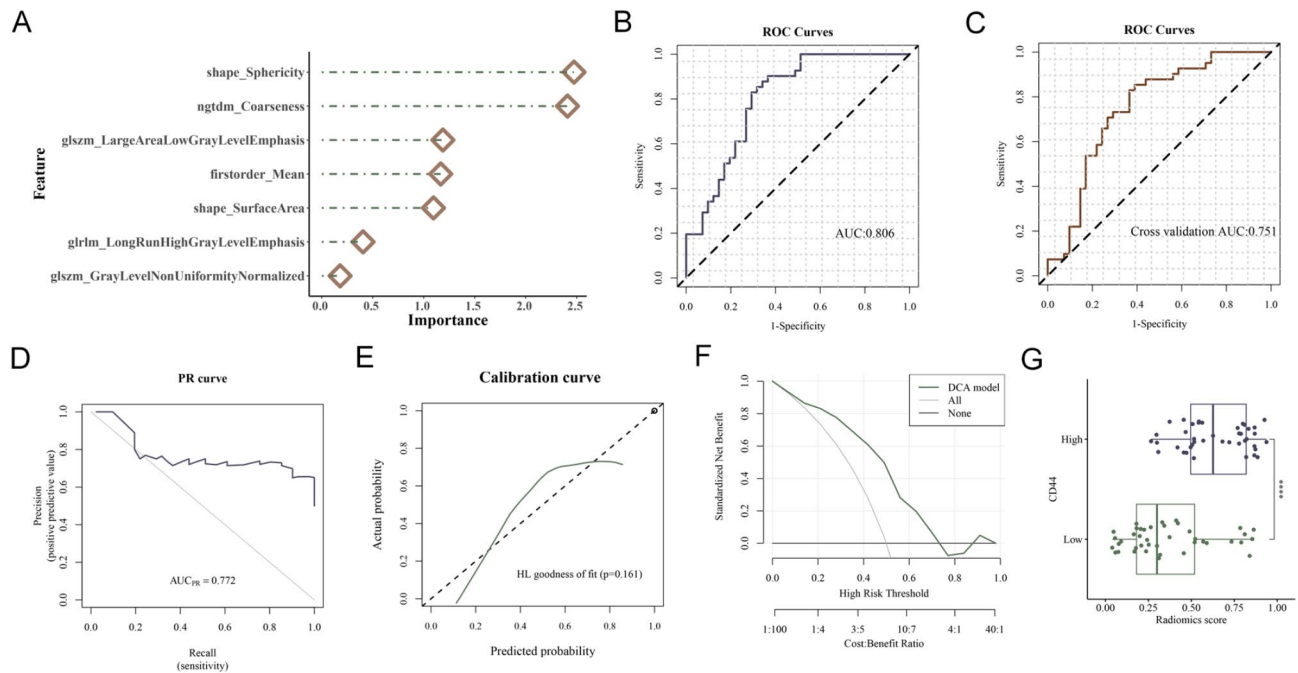
Fig. 5. Construction and assessment of the radiomics model for predicting CD44 expression based on LR. (**A**) Seven radiomic features and the corresponding importance in the radiomics model based on LR. (**B**) ROC curve. (**C**) ROC curve of 10-fold cross-validation. (**D**) PR curve. (**E**) Calibration curve. (**F**) Plot of DCA. (**G**) Predicted probabilities of RS in CD44high and CD44low groups. AUC, the area under the curve; DCA, decision curve analysis; Glrlm, gray level run length matrix; Glszm, gray level size zone matrix; HL, Hosmer-Lemeshow test; LR, logistic regression; Ngtdm, neighborhood gray tone difference matrix; PR, precision recall; ROC, receiver operating characteristic. ****, $p < 0.0001$.

To enhance the clinical applicability, we developed a predictive nomogram incorporating radiomic score, age, gender, chemotherapy, radiotherapy, the methylation status of MGMT promoter, and Chr 1p19q codeletion status. The clinical application nomogram score sheet displayed the clinical predictive scheme at 12, 36, and 60 months (Fig. 6J).

Discussion

The present study aimed to elucidate the prognostic value of CD44 in HGG, and to develop radiomics models by ML to predict CD44 expression non-invasively. Our study corroborated that CD44 expression level was an independent predictor of OS, and high-expressed CD44 was associated with poor OS among HGG patients. In addition, we constructed two favorable radiomics models to predict CD44 expression level in HGG patients, achieving AUC values of 0.809 and 0.806, respectively. Furthermore, HGG patients with higher RS predicted relatively shorter OS.

CD44, a cell-surface receptor with multifaceted roles in cellular functions, has been implicated as a risk factor in glioma, as demonstrated in previous studies[25,26]. For instance, a meta-analysis revealed that elevated CD44 expression adversely impacts OS in patients with glioma[25], which aligns with our findings. Accumulating evidence suggests that CD44 is responsible for growth and metastasis in HGG[27,28]. Interestingly, our analysis found that the most significant GO enrichment terms associated with CD44 include synapse organization, focal adhesion, cell-substrate junction, and neuron to neuron synapse. In addition, CD44 was recently reported to participate in TME[29], immunotherapy[30], and stemness[31] in HGG. Notably, our research also discovered a propensity for M2 macrophages infiltration in the CD44 high-expressed HGG, indicating a potential role of CD44 in driving M2 macrophage polarization and remodeling TME[29].

Numerous molecules, including CD44, contribute to HGG progression in distinct ways and at different stages, contributing to intratumoral and intertumoral heterogeneity[32]. However, assessing this heterogeneity pathologically is invasive, subject to sampling variability, and impractical in clinical practice for each sample. Thus, it is essential to construct an efficient approach to quantify the heterogeneity. MRI is routinely utilized for HGG diagnosis and follow-up in clinical practice, offering the capability to non-invasively assess the entire lesion and dynamic monitoring it. Several evidences have inspired us to explore the potential of MRI-based radiomics for diagnosis and classification of HGG[33,34].

In the current research, we identified CD44 as an independent risk factor for OS by multivariate COX regression analysis. To further predict CD44 expression in HGG, we applied repeat LASSO to screen the radiomic features of MRI images, and constructed two radiomics models based on LR and SVM methods, respectively. Interestingly, both models exhibited satisfactory predictive power of CD44 expression. As precision medicine continues to advance, an increasing number of molecular targets have been identified. Notably, accurate
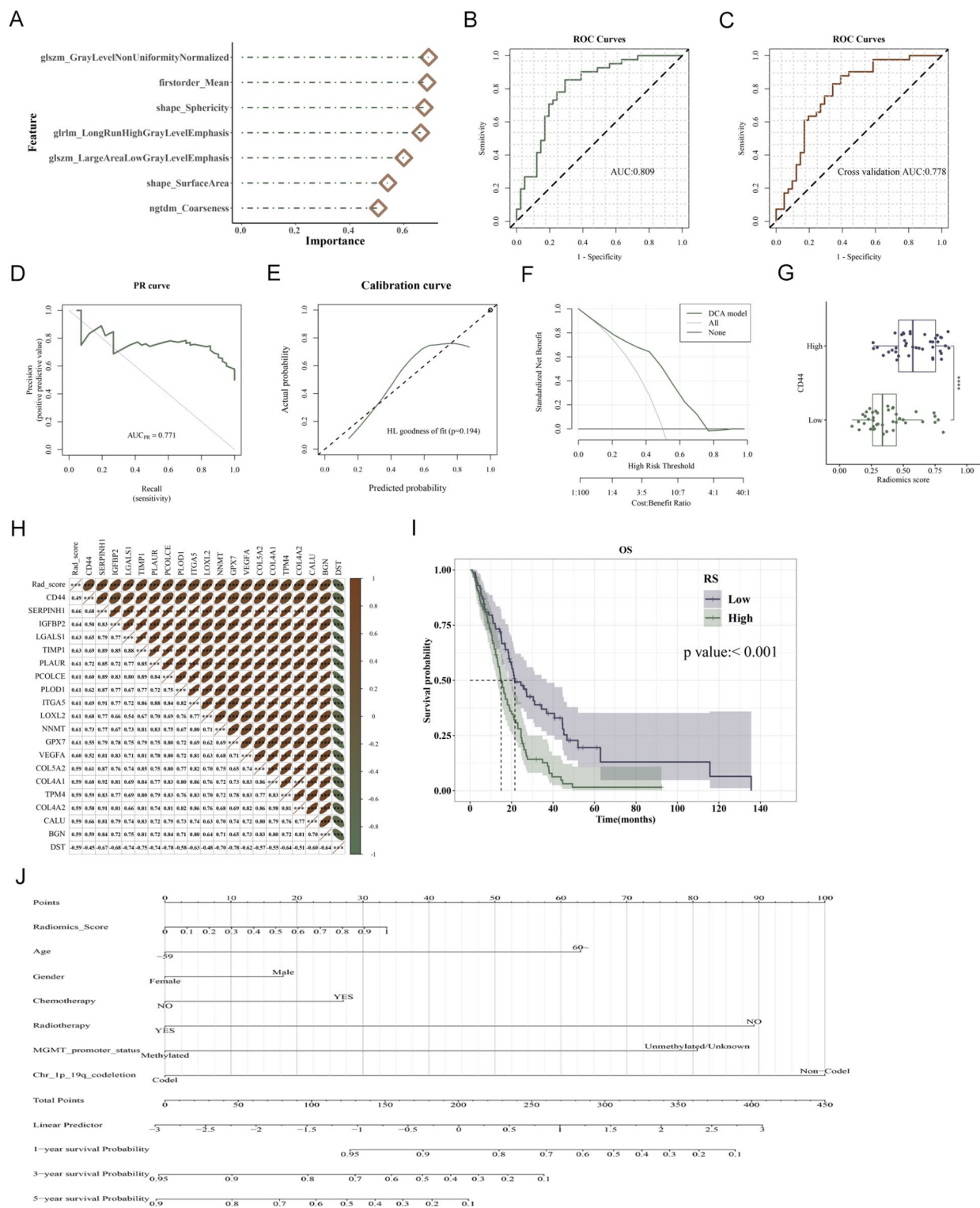
Fig. 6. Construction and assessment of the radiomics model for predicting CD44 expression based on SVM. (**A**) Seven radiomic features and the corresponding importance in the radiomics model based on SVM. (**B**) ROC curve. (**C**) ROC curve of 10-fold cross-validation. (**D**) PR curve. (**E**) Calibration curve. (**F**) Plot of DCA. (**G**) Predicted probabilities of RS in CD44high and CD44low groups. (**H**) Spearman correlation analysis of RS and EMT molecules. (**I**) Survival analysis of RS for OS. (**J**) Clinical application of radiomics model in HGG. AUC, the area under the curve; DCA, decision curve analysis; Glrlm, gray level run length matrix; Glszm, gray level size zone matrix; HL, Hosmer-Lemeshow test; Ngtdm, neighborhood gray tone difference matrix; OS, overall survival; PR, precision recall; ROC, receiver operating characteristic; RS, Radiomic score; SVM, support vector machine; ***, $p < 0.001$; ****, $p < 0.0001$.

prediction of these targets has become critical for diagnosis and therapy, forming the foundation of precision oncology applications. Consequently, many investigators are committed to build radiomics models to predict molecule status intratumor precisely and generalizability[35,36]. Thus, radiomics models are emerging as a popular research direction for heterogeneity measure and biomarker prediction.

Routine MRI diagnosis relies on morphological features of the lesions, which is lack of objective, quantitative information. Radiomics provides high-dimensional quantitative information that might have been overlooked or unnoticed by the naked eye, thereby aiding diagnosis based on imaging features. It is often referred to as "digital biopsy". Recently, extensive studies have confirmed the importance of radiomics in predicting biomarkers in glioma. For instance, Sha et al.[37] constructed radiomics models to predict the IDH mutation combined with MGMT methylated molecular subtype in glioma, achieving an impressive AUC of 0.9. Another study built an MRI-based radiomic signature that had promising ability to assess the promoter mutation status of TERT in glioblastoma patients[38]. Similarly, a recent study built two radiomics models based on MRI to predict CD44 expression in LGG[39]. In this research, we applied MRI radiomic features to predict CD44 expression in HGG, and our prediction models achieved AUCs of 0.806 and 0.809 in LR and SVM models, respectively. These results provide an efficient, cost-effective, repeatable, non-invasive method to predict CD44 expression in HGG.

Following the screening of 107 radiomic features, seven radiomic features were finally retained to build our predictive model, including shape_sphericity, glszm_gray level non uniformity normalized, glrlm_long run high gray level emphasis, ngtdm_coarseness, glszm_large area low gray level emphasis, shape_surface area, and first order mean. Glszm, a texture feature, represents the number of connected voxels with the same gray level intensity and arrangement pattern in an image[40]. A higher value indicates greater heterogeneity within regions of interest. Several studies confirmed that glszm-derived features are common and important radiomic factors for the research in glioma[39,41]. Wang et al. built a predictive model for CD44 expression in LGG, in which a feature from glszm category showed an odds ratio of 2.634 ($p = 0.008$)[40]. Consistent with their findings, our study also highlighted the crucial role of the glszm category in both models. In the SVM model, glszm_gray level non uniformity normalized emerged as the most important feature, while in the LR model, glszm_large area low gray level emphasis ranked as the third most important feature. The above results underscore that the glszm radiomic features played an important role in predicting the CD44 expression level in HGG.

A multitude of studies have highlighted the potential of radiomics models as a promising tool for the diagnosis, management, and prognostication across various cancer types, including ovarian cancer, hepatocellular carcinoma, and meningioma[42–44]. Several studies have specifically focused on constructing radiomics models to predict the survival of glioma patients[45,46]. Li et al.[47] established a prognostic radiomics model for patients with glioma, which demonstrated that patients with high radiomic feature scores had shorter OS. Similarly, our radiomics model, incorporating texture features, first order features, and shape features from radiomic images, consistently exhibited discriminatory performance to predict OS in HGG patients. According to Kaplan–Meier curves, the median OS of RS[low] group was 21.6 months, which was 6.7 months longer than that of RS[high] group. These findings collectively demonstrate the fantastic prognostic power of our radiomics models.

While our radiomics models demonstrated strong performance in predicting CD44 expression and OS in patients with HGG, several limitations of the present research provide orientations for future study. Firstly, the research was an exploratory study, further validation study should be reported. In addition, the present study was based on MRI enhanced T1WI images, and future studies should consider incorporating additional functional MRI sequences. Moreover, the dataset used in this study was obtained from public database and the sample capacity was relatively small. Multicenter prospective studies are still needed to combat bias and enhance the generalizability of our findings.

Conclusion

Elevated CD44 expression was found to be strongly associated with shorter OS in patients with HGG. Our radiomics model, referred to as RS, demonstrated the ability to accurately discern the level of CD44 expression and predict OS in patients with HGG. Moreover, RS showed a significant correlation with some EMT related genes. This non-invasive radiomics model exhibited reliable and feasible performance, offering a potential powerful tool for predicting outcomes and assisting clinical decision-making in the era of individualized precision medicine and target therapy.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

References

1. Lapointe, S., Perry, A. & Butowski, N. A. Primary brain tumours in adults. *Lancet* **392**, 432–446. https://doi.org/10.1016/S0140-6736(18)30990-5 (2018).
2. Yang, K. et al. Glioma targeted therapy: Insight into future of molecular approaches. *Mol. Cancer* **21**, 39. https://doi.org/10.1186/s12943-022-01513-z (2022).
3. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro. Oncol.* **23**, 1231–1251. https://doi.org/10.1093/neuonc/noab106 (2021).
4. Casalino-Matsuda, S. M., Monzon, M. E., Day, A. J. & Forteza, R. M. Hyaluronan fragments/CD44 mediate oxidative stress-induced MUC5B up-regulation in airway epithelium. *Am. J. Respir. Cell Mol. Biol.* **40**, 277–285. https://doi.org/10.1165/rcmb.2008-0073OC (2009).

5. Midgley, A. C. et al. Transforming growth factor-beta1 (TGF-beta1)-stimulated fibroblast to myofibroblast differentiation is mediated by hyaluronan (HA)-facilitated epidermal growth factor receptor (EGFR) and CD44 co-localization in lipid rafts. *J. Biol. Chem.* **288**, 14824–14838. https://doi.org/10.1074/jbc.M113.451336 (2013).

6. Vikesaa, J. et al. RNA-binding IMPs promote cell adhesion and invadopodia formation. *EMBO J.* **25**, 1456–1468. https://doi.org/10.1038/sj.emboj.7601039 (2006).

7. Crosby, H. A., Lalor, P. F., Ross, E., Newsome, P. N. & Adams, D. H. Adhesion of human haematopoietic (CD34+) stem cells to human liver compartments is integrin and CD44 dependent and modulated by CXCR3 and CXCR4. *J. Hepatol.* **51**, 734–749. https://doi.org/10.1016/j.jhep.2009.06.021 (2009).

8. Yoshida, T., Matsuda, Y., Naito, Z. & Ishiwata, T. CD44 in human glioma correlates with histopathological grade and cell migration. *Pathol. Int.* **62**, 463–470. https://doi.org/10.1111/j.1440-1827.2012.02823.x (2012).

9. Funaro, A., Spagnoli, G. C., Momo, M., Knapp, W. & Malavasi, F. Stimulation of T cells via CD44 requires leukocyte-function-associated antigen interactions and interleukin-2 production. *Hum. Immunol.* **40**, 267–278. https://doi.org/10.1016/0198-8859(94)90026-4 (1994).

10. Hou, C. et al. Overexpression of CD44 is associated with a poor prognosis in grade II/III gliomas. *J. Neurooncol.* **145**, 201–210. https://doi.org/10.1007/s11060-019-03288-8 (2019).

11. Ghamande, S. A. et al. A phase 2, randomized, double-blind, placebo-controlled trial of clinical activity and safety of subcutaneous A6 in women with asymptomatic CA125 progression after first-line chemotherapy of epithelial ovarian cancer. *Gynecol. Oncol.* **111**, 89–94. https://doi.org/10.1016/j.ygyno.2008.06.028 (2008).

12. Finlayson, M. Modulation of CD44 activity by A6-peptide. *Front. Immunol.* **6**, 135. https://doi.org/10.3389/fimmu.2015.00135 (2015).

13. Valentine, A. et al. FKBPL and peptide derivatives: novel biological agents that inhibit angiogenesis by a CD44-dependent mechanism. *Clin. Cancer Res.* **17**, 1044–1056. https://doi.org/10.1158/1078-0432.CCR-10-2241 (2011).

14. Yakkundi, A. et al. The anti-migratory effects of FKBPL and its peptide derivative, AD-01: Regulation of CD44 and the cytoskeletal pathway. *PLoS One* **8**, e55075. https://doi.org/10.1371/journal.pone.0055075 (2013).

15. Fan, B., Fan, B., Sun, N., Zou, H. & Gu, X. A radiomics model to predict gammadelta T-cell abundance and overall survival in head and neck squamous cell carcinoma. *FASEB J.* **38**, e23529. https://doi.org/10.1096/fj.202301353RR (2024).

16. Gihr, G. et al. Diffusion weighted imaging in gliomas: A histogram-based approach for tumor characterization. *Cancers (Basel)* **14**, 3393. https://doi.org/10.3390/cancers14143393 (2022).

17. Park, Y. W. et al. A fully automatic multiparametric radiomics model for differentiation of adult pilocytic astrocytomas from high-grade gliomas. *Eur. Radiol.* **32**, 4500–4509. https://doi.org/10.1007/s00330-022-08575-z (2022).

18. Kim, A. R. et al. Absolute quantification of tumor-infiltrating immune cells in high-grade glioma identifies prognostic and radiomics values. *Cancer Immunol. Immunother.* **70**, 1995–2008. https://doi.org/10.1007/s00262-020-02836-w (2021).

19. Vivian, J. et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **35**, 314–316. https://doi.org/10.1038/nbt.3772 (2017).

20. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782. https://doi.org/10.1038/s41587-019-0114-2 (2019).

21. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. https://doi.org/10.1093/nar/gkv007 (2015).

22. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).

23. Huang, W. Y. et al. Radiological model based on the standard magnetic resonance sequences for detecting methylguanine methyltransferase methylation in glioma using texture analysis. *Cancer Sci.* **112**, 2835–2844. https://doi.org/10.1111/cas.14918 (2021).

24. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).

25. Wu, G. et al. Expression of CD44 and the survival in glioma: A meta-analysis. *Biosci. Rep.*, **40**, doi: 10.1042/BSR20200520 (2020).

26. Anido, J. et al. TGF-beta receptor inhibitors target the CD44(high)/Id1(high) glioma-initiating cell population in human glioblastoma. *Cancer Cell* **18**, 655–668. https://doi.org/10.1016/j.ccr.2010.10.023 (2010).

27. Zhang, H. et al. RUNX1/CD44 axis regulates the proliferation, migration, and immunotherapy of gliomas: A single-cell sequencing analysis. *Front. Immunol.* **14**, 1086280. https://doi.org/10.3389/fimmu.2023.1086280 (2023).

28. Feng, S. et al. MiR-373/miR-520s-CD44 axis significantly inhibits the growth and invasion of human glioblastoma cells. *Arch. Med. Res.* **53**, 550–561. https://doi.org/10.1016/j.arcmed.2022.08.003 (2022).

29. Zhao, T. et al. Chitinase-3 like-protein-1 promotes glioma progression via the NF-kappaB signaling pathway and tumor microenvironment reprogramming. *Theranostics* **12**, 6989–7008. https://doi.org/10.7150/thno.75069 (2022).

30. Xiao, Y. et al. CD44-mediated poor prognosis in glioma is associated with M2-polarization of tumor-associated macrophages and immunosuppression. *Front. Surg.* **8**, 775194. https://doi.org/10.3389/fsurg.2021.775194 (2021).

31. Wang, R. et al. Bio-fabricated nanodrugs with chemo-immunotherapy to inhibit glioma proliferation and recurrence. *J. Control Release* **354**, 572–587. https://doi.org/10.1016/j.jconrel.2023.01.023 (2023).

32. Orozco-Mera, J., Peralta-Pizza, F., Escobar-Vidarte, O., Coll-Tello, B. & Ariza, A. Signaling pathways in the relapse of glioblastoma. *Oncology (Williston Park)* **37**, 107–117. https://doi.org/10.46883/2023.25920986 (2023).

33. Chilaca-Rosas, M. F., Garcia-Lezama, M., Moreno-Jimenez, S. & Roldan-Valadez, E. Diagnostic performance of selected MRI-derived radiomics able to discriminate progression-free and overall survival in patients with midline glioma and the H3F3AK27M mutation. *Diagnostics (Basel)* **13**, 849. https://doi.org/10.3390/diagnostics13050849 (2023).

34. Khalid, F. et al. Multimodal MRI radiomic models to predict genomic mutations in diffuse intrinsic pontine glioma with missing imaging modalities. *Front. Med. (Lausanne)* **10**, 1071447. https://doi.org/10.3389/fmed.2023.1071447 (2023).

35. Monaco, L. et al. The "digital biopsy" in non-small cell lung cancer (NSCLC): A pilot study to predict the PD-L1 status from radiomics features of [18F]FDG PET/CT. *Eur. J. Nucl. Med. Mol. Imaging* **49**, 3401–3411. https://doi.org/10.1007/s00259-022-05783-z (2022).

36. Ghaffari Laleh, N., Ligero, M., Perez-Lopez, R. & Kather, J. N. Facts and hopes on the use of artificial intelligence for predictive immunotherapy biomarkers in cancer. *Clin. Cancer Res.* **29**, 316–323. https://doi.org/10.1158/1078-0432.CCR-22-0390 (2023).

37. Sha, Y. et al. Prediction of the molecular subtype of IDH mutation combined with MGMT promoter methylation in gliomas via radiomics based on preoperative MRI. *Cancers (Basel)* **15**, 1440. https://doi.org/10.3390/cancers15051440 (2023).

38. Zhang, H. et al. Deep learning radiomics for the assessment of telomerase reverse transcriptase promoter mutation status in patients with glioblastoma using multiparametric MRI. *J. Magn. Reson. Imaging* https://doi.org/10.1002/jmri.28671 (2023).

39. Wang, Z. et al. Radiomics features based on T2-weighted fluid-attenuated inversion recovery MRI predict the expression levels of CD44 and CD133 in lower grade gliomas. *Future Oncol.* **18**, 807–819. https://doi.org/10.2217/fon-2021-1173 (2022).

40. Xu, F. et al. Radiomic-based quantitative CT analysis of pure ground-glass nodules to predict the invasiveness of lung adenocarcinoma. *Front. Oncol.* **10**, 872. https://doi.org/10.3389/fonc.2020.00872 (2020).

41. Suarez-Garcia, J. G., Hernandez-Lopez, J. M., Moreno-Barbosa, E. & de Celis-Alonso, B. A simple model for glioma grading based on texture analysis applied to conventional brain MRI. *PLoS One* **15**, e0228972. https://doi.org/10.1371/journal.pone.0228972 (2020).

42. Gao, L. et al. Radiomic model to predict the expression of PD-1 and overall survival of patients with ovarian cancer. *Int. Immunopharmacol.* **113**, 109335. https://doi.org/10.1016/j.intimp.2022.109335 (2022).
43. Bo, Z. et al. Prediction of response to lenvatinib monotherapy for unresectable hepatocellular carcinoma by machine learning radiomics: A multicenter cohort study. *Clin. Cancer Res.* https://doi.org/10.1158/1078-0432.CCR-22-2784 (2023).
44. Patel, R. V., Yao, S., Huang, R. Y. & Bi, W. L. Application of radiomics to meningiomas: A systematic review. *Neuro. Oncol.* https://doi.org/10.1093/neuonc/noad028 (2023).
45. Sun, Q. et al. Biologic pathways underlying prognostic radiomics phenotypes from paired MRI and RNA sequencing in glioblastoma. *Radiology* **301**, 654–663. https://doi.org/10.1148/radiol.2021203281 (2021).
46. Yan, J. et al. Deep learning features from diffusion tensor imaging improve glioma stratification and identify risk groups with distinct molecular pathway activities. *EBioMedicine* **72**, 103583. https://doi.org/10.1016/j.ebiom.2021.103583 (2021).
47. Li, G. et al. An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas. *Brain* **145**, 1151–1161. https://doi.org/10.1093/brain/awab340 (2022).

Acknowledgements

Author contributions

Mingjun Yu and Shijia Yu conceived and designed the research; Mingjun Yu, Xiao Gu, Wen Zhou and Jinliang Liu performed the research and analyzed the data; Xiao Gu and Shijia Yu wrote the paper. All authors read and approved the final manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

The study was performed in accordance with the relevant guidelines and regulations of the Declaration of Helsinki. The studies involving human participants were reviewed and approved by the Ethics Committee of the Shengjing Hospital (No. 2019PS105K).

Consent to participate

Informed consent was obtained from all subjects and/or their legal guardian(s).

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-90128-7.

Correspondence and requests for materials should be addressed to X.G. or S.Y.

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.