Postoperative Karnofsky performance status prediction in patients with *IDH* wild-type glioblastoma: a multimodal approach integrating clinical and deep imaging features

Tomoki Sasagasako, MD,[1,2] Akihiko Ueda, MD, PhD,[2] Yohei Mineharu, MD, PhD,[3] Yusuke Mochizuki[4], Souichiro Doi[4], Silsu Park, MD,[1] Yukinori Terada, MD, PhD,[1] Noritaka Sano, MD, PhD,[1] Masahiro Tanji, MD, PhD,[1] Yoshiki Arakawa, MD, PhD,[1] Yasushi Okuno, PhD[2,3]

Affiliations:

[1] Department of Neurosurgery, Kyoto University Graduate School of Medicine, Kyoto, Japan

[2] Department of Biomedical Data Intelligence, Kyoto University Graduate School of Medicine, Kyoto, Japan

[3] Department of Artificial Intelligence in Healthcare and Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan

[4] Kyoto University Faculty of Medicine, Kyoto, Japan

Correspondence:

Yohei Mineharu, MD, PhD

Department of Artificial Intelligence in Healthcare and Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan

+81-75-751-3111

Email: mineharu@kuhp.kyoto-u.ac.jp

ABSTRACT

Background and Purpose: Glioblastoma is a highly aggressive brain tumor with limited survival that poses challenges in predicting patient outcomes. The Karnofsky Performance Status (KPS) score is a valuable tool for assessing patient functionality and contributes to the stratification of patients with poor prognoses. This study aimed to develop a 6-month postoperative KPS prediction model by combining clinical data with deep learning-based image features from pre- and postoperative MRI scans, offering enhanced personalized care for glioblastoma patients.

Materials and Methods: Using 1,476 MRI datasets from the Brain Tumor Segmentation Challenge 2020 public database, we pretrained two variational autoencoders (VAEs). Imaging features from the latent spaces of the VAEs were used for KPS prediction. Neural network-based KPS prediction models were developed to predict scores below 70 at 6 months postoperatively. In this retrospective single-center analysis, we incorporated clinical parameters and pre- and postoperative MRI images from 150 newly diagnosed IDH wild-type glioblastoma, divided into training (100 patients) and test (50 patients) sets. In training set, the performance of these models was evaluated using the area under the curve (AUC), calculated through fivefold cross-validation repeated 10 times. The final evaluation of the developed models assessed in the test set.

Results: Among the 150 patients, 61 had 6-month postoperative KPS scores below 70 and 89 scored 70 or higher. We developed three models: a clinical-based model, an MRI-based model, and a multimodal model that incorporated both clinical parameters and MRI features. In the training set, the mean AUC was 0.785±0.051 for the multimodal model, which was significantly higher than the clinical-based model (0.716±0.059, P=0.038) using only clinical parameters and MRI-based model (0.651±0.028, P<0.001) using only MRI features. In the test set, the multimodal model achieved an AUC of 0.810, outperforming the clinical-based (0.670) and MRI-based (0.650) models.

Conclusion: The integration of MRI features extracted from VAEs with clinical parameters in the

multimodal model substantially enhanced KPS prediction performance. This approach has the potential to improve prognostic prediction, paving the way for more personalized and effective treatments for patients with glioblastoma.

Abbreviations:

KPS, Karnofsky performance status. IDH, isocitrate dehydrogenase. VAE, variational autoencoder. BraTS, Brain Tumor Segmentation challenge

Disclosure of potential conflicts of interest:

The authors declare no conflicts of interest related to the content of this article.

INTRODUCTION

Glioblastoma is a highly malignant brain tumor with a median overall survival of approximately 15–18 months.[1] Despite numerous studies on treatment strategies, it often recurs rapidly, leading to a worsened functional prognosis.

Deep learning has increasingly been applied to detect, diagnose, and predict clinical outcomes in patients with glioblastoma.[2] However, it is still uncertain whether a radiomics approach using deep learning approach can enhance the prediction of clinical outcome or if pre- and postoperative MRI features can reliably predict prognosis.

The Karnofsky Performance Status (KPS) score stands as a robust independent predictor of clinical outcomes within diverse oncology populations affected by malignant tumors.[3] A KPS score of 70 indicates the patient can care for themselves but is unable to carry out daily activities. Among patients with glioblastoma, multiple studies have indicated a correlation between a KPS score of <70 and a poor prognosis.[4] The 6-month postoperative KPS was chosen for its clinical practicality, as it aligns with the standard timeframe for evaluating disease progression after a typical course of adjuvant treatments. Predicting the postoperative KPS, along with the early identification of patients at risk of diminished KPS in the postoperative stage, could lead to improved counseling and more personalized clinical decision-making.[5]

In this study, we developed a multimodal model using clinical parameters and brain MRI to stratify patients into prognostic groups based on their 6-month postoperative KPS. We utilized a deep learning approach to extract imaging features from pre- and postoperative MRI images and investigated their prognostic value.

MATERIALS AND METHODS

The overall study process is illustrated in Fig 1. The proposed algorithm comprises two

primary stages: 1) constructing variational autoencoders (VAEs) to extract the reduced latent features from the pre- and postoperative MRI images, and 2) developing a KPS prediction model by combining the patients' clinical parameters and extracted imaging features. The authors collected patients' data from April 2022 to July 2023.

Fig 1. Outline of the main steps in this study.

(1) Patient surveys and data acquisition from our institute's medical records, (2) brain tumor segmentation using pre- and postoperative MRI, (3) imaging feature extraction using a pretrained variation autoencoder combined with a convolutional neural network, and (4) prediction model development using a neural network and training. The performance metrics in the training set were evaluated using 10 repetition of fivefold cross-validation. GBM=glioblastoma; HGG=high-grade glioma; VAE=variational autoencoder.

Ethics Approval

The Ethics Committee of Kyoto University Hospital approved this study (R2088). Verbal informed consent was obtained from the study participants. Participants were informed before their surgery that their images and clinical information would be used in a retrospective review after being fully anonymized. The ethics committee waived the requirement for written informed consent owing to the retrospective study design. Patients who did not wish to participate were excluded via an opt-out process.

Patients

This single-center retrospective review was conducted between December 2001 and December 2022 at our institution. This study included consecutive adult patients aged 18 years and older who were newly histopathologically diagnosed with glioblastoma with isocitrate dehydrogenase (IDH) wild type. The exclusion criteria for the study cohort were as follows: 1) gliomas that were not

histopathologically diagnosed as glioblastomas, 2) lack of immunohistochemical testing for the IDH1 R132H mutation or absence of IDH2 sequencing, 3) a diagnosis of IDH-mutant gliomas, and 4) a follow-up period of <6 months. A total of 257 patients were pathologically confirmed to have glioblastoma. Among these, 87 patients were excluded based on the following criteria: diagnosis of IDH-mutant gliomas (n=11) or absence of either immunohistochemistry for the IDH1 R132H mutation or IDH2 sequencing (n=76). Additionally, patients lacking sufficient medical records or imaging data (n=18) and those aged <18 years (n=2) were also excluded. Subsequently, patients were divided into training and test sets based on their first operation date. Patients who underwent surgery from December 2001 to October 2018 were assigned to the training set, while those from November 2018 to December 2022 were included in the test set. This process resulted in 100 patients in the training set and 50 in the test set (S1 Fig).

Clinical Parameters and Endpoints

The clinical parameters included following 28 variables.

Preoperative variables: sex, age at diagnosis, dominant hand (right/left), epilepsy (yes/no), aphasia (yes/no), paralysis (yes/no), other neurological findings at onset (yes/no), and preoperative KPS score (%).

Intraoperative variables: surgical strategy (biopsy or tumor removal), awake surgery (yes/no), utilization of 5-aminolevulinic acid (yes/no), photodynamic therapy (yes/no), carmustine wafer placement (yes/no), and motor-evoked or somatosensory-evoked potential monitoring (yes/no).

Immunohistochemical and genetic variables: O-6-methylguanine-DNA methyltransferase (MGMT) methylation (positive/negative), *TERTp* alteration (positive/negative), MIB-1 labeling index (%), and immunohistochemical staining for MGMT (positive/negative).

Postoperative variables: TMZ chemotherapy (yes/no), bevacizumab chemotherapy (yes/no), radiation dose (Gy), number of radiation fractionations (Fr).

Radiological findings: tumor laterality (right/left/bilateral), ependymal involvement (yes/no),

midline shift (yes/no), corpus callosum invasion (yes/no), necrotic or cystic area evident on imaging (yes/no), and extent of resection (1–49%, 50–89%, 90–99%, 100%).

Additionally, the 6-month postoperative KPS (%) score, the main endpoint of this study, was collected.

For the standard concurrent chemoradiotherapy regimen, patients received fractionated focal radiation therapy with a cumulative dose of 60 Gy, accompanied by concomitant TMZ chemotherapy.[6] For elderly patients aged 70 years or older, a hypofractionated radiotherapy schedule of 40 Gy delivered in 15 fractions over 3 weeks was employed.[7] All patients were followed up every 1–2 months after surgical treatment.

Imaging Acquisition and Preprocessing

The following 3-mm slices of MRI scans were collected from each patient: T1WI, contrast-enhanced T1W (T1Gd), ADC, DWI, and FLAIR. In the study population, preoperative MRI images were acquired within 2 weeks before surgery, and postoperative MRI images were acquired the day after surgery, whenever the patient's condition allowed. Brain MRI images were processed using a deep brain extractor to remove the skin and cranial bones.[8] The brain parenchyma was extracted based on the T1WI and DWI images (represented in gray in Fig 2).

Tumor Segmentation

To achieve semantic segmentation of the glioblastoma lesion on MRI scans, we utilized a segmentation model based on the U-net architecture, specifically designed for the segmentation of glioma.[9] Using this segmentation model, we highlighted the segmented regions, including enhanced tumors, necrosis, and cystic lesions, which are shown in red. Peritumoral edema and non-enhancing tumor areas are shown in green (Fig 2). Images with a thickness of 3 mm were used for each MRI sequence. After semantic segmentation, the tumor area, represented in red, was automatically

154 measured. We then selected 24 segmented images per patient to ensure that the slice with the largest
155 tumor area was included in the central part of the selected slices.

156 **Fig 2. Tumor segmentation using preoperative and postoperative MRI.**

157 Using a pretrained U-net-based segmentation module, pre- and postoperative segmented images were
158 generated from 24 slices of FLAIR, T1W, T1Gd, ADC, and DWI images. The brain parenchyma is
159 displayed in gray, enhanced tumor lesions and necrosis in red, and peritumoral edema and non-
160 enhancing lesions in green. T1Gd=contrast-enhanced T1W.

161

162 MRI Feature Extraction from the Latent Space of a Variational
163 Autoencoder

164 MRI features were extracted from the segmented brain images using two independently
165 developed VAEs, as shown in Fig 3. VAE 1 was utilized to process segmented tumor lesions from
166 pre- and postoperative MRI images as input data. In contrast, VAE 2 processed data from pre- and
167 postoperative brain parenchyma areas, serving as both input and output. The comprehensive
168 architecture of these VAEs is demonstrated in S2 Fig. From these VAEs, MRI features were extracted
169 from the 48-dimensional latent spaces of both VAE 1 and 2.

170 **Fig 3. Extraction of deep imaging features from the latent space of a variational autoencoder.**

171 Twenty-four slices of the segmented images were separated into tumor lesions (red or green) and brain
172 mask images (gray). The tumor lesion images were processed by VAE 1, where internal 3D
173 convolutional neural networks extracted features and reduced dimensionality using encoder 1. As a
174 result, a 48-dimensional latent space was formed, and the generated deep imaging features were
175 incorporated into the KPS prediction model. Similarly, the brain mask images were processed by
176 being input into VAE 2. Both VAEs 1 and 2 were pretrained using the BraTS 2020 dataset.
177 VAE=variational autoencoder; KPS=Karnofsky performance status; BraTS=brain tumor segmentation
178 challenge.

We pretrained these VAEs using 1,476 MRI datasets of high-grade glioma and glioblastoma MRI images from the BraTS 2020 dataset.[10] BraTS 2020 provided thin-slice MRI images of 369 patients with high-grade gliomas. The slices were extracted at regular intervals to create four different MRI datasets from a single patient. Pretraining of the VAE was conducted using annotated brain mask images and segmented tumor regions as ground truth data (Fig 1).

Development of a KPS Prediction Model Using Neural Networks

To stratify patients with a postoperative KPS score of <70 at 6 months, we developed neural network prediction models using training set (Fig 1).

These prediction models use two distinct inputs: clinical and MRI features. First, we developed a clinical neural network model using clinical parameters. In this clinical-based model, the input consists of clinical features, while the final output layer comprises two neurons, making it suitable for binary classification: predicting whether the 6-month postoperative KPS score <70 or ≥70.

In addition to the clinical-based model, an MRI-based model was constructed using MRI features. Furthermore, clinical and accompanying MRI features were employed as inputs for the multimodal model. The details of the development of the prediction model are shown in S3 Fig.

During the model development with the training set, we assessed predictive performance using the area under the receiver operating characteristic curve (AUC) through 10 repetitions of fivefold cross-validation (Fig 1). This approach served as an appropriate internal validation procedure, especially given the absence of external testing.[11,12] The effectiveness of the model was visualized through the mean ROC curves across all fivefold cross-validation. Furthermore, to evaluate the performance in fivefold cross-validation, the following metrics were computed: accuracy, sensitivity, specificity, and F1 score. The performance of the three developed models—clinical-based, MRI-based, and multimodal—was evaluated using test sets that were held out during the model

development and training process. We also developed the following machine learning models: a Random Forest classifier, XGBoost, and LightGBM. These machine learning models perform hyperparameter tuning using the GridSearch software. We extensively compared and analyzed the predictive capabilities of each classifier by considering the aforementioned metrics (S4 Fig).

Model Interpretability

Grouped permutation feature importance was utilized to determine the feature importance in the neural network-based prediction model.[13] In this analysis, features were optionally grouped into expert-defined subgroups, and a systematic assessment of their importance was conducted. In this study, during the feature extraction process, 196 MRI features were generated from preoperative and postoperative MRI imaging. To enhance interpretability, the variables derived from the preoperative brain mask image were grouped as "pre_mask variables". Similarly, variables from postoperative brain mask, preoperative tumor lesion, and postoperative tumor lesion images were grouped as "post_mask," "pre_lesion," and "post_lesion variables," respectively. Twenty-eight clinical parameters were evaluated separately, without grouping.

Statistical Analysis

All statistical analyses were performed using the SciPy library. Univariate analysis was used to examine the relationship between the 6-month postoperative KPS deterioration and clinical parameters. Fisher's exact test was used to evaluate categorical variables, whereas the Mann–Whitney U test was used for continuous variables. The cross-validated metrics of each prediction model were compared using the paired-samples t test. Statistical significance was set at $P < 0.05$.

RESULTS

Patient Demographics

Of the 150 study population, 61 patients had a 6-month postoperative KPS score of <70 and 89 had ≥70. Baseline clinical parameters are presented in Table 1. Among these 150 patients, 65 (43.3%) were female, with a mean age of 64 years (range, 21–92 years). The median preoperative and 6-month postoperative KPS scores were 80 (range, 20–100) and 70 (range, 0–100), respectively. Out of the 118 patients with a preoperative KPS score of ≥70, 43 had a 6-month postoperative KPS score of <70. Conversely, among the 38 patients with a preoperative KPS score of <70, 14 had a 6-month postoperative KPS score of ≥70. These patients were divided into training and test sets. Patients who underwent surgery from December 2001 to October 2018 were assigned to the training set (n=100), while those from November 2018 to December 2022 were included in the test set (n=50).

Table 1. Baseline clinical characteristics.

	Overall (N=150)	KPS < 70 (n=61)	KPS ≥ 70 (n=89)	P value
Sex (female), n (%)	65 (43.3)	24 (39.3)	41 (46.1)	0.50
Median age, year (range)	64 (21 – 92)	72 (32 – 92)	60 (21 – 83)	< 0.001
Dominant hand (right-handed), n (%)	147 (98)	58 (95.1)	89 (100)	0.07
Preoperative epilepsy, n (%)	40 (26.7)	12 (19.7)	28 (31.5)	0.13
Preoperative aphasia, n (%)	42 (28)	24 (39.3)	18 (20.2)	0.02
Preoperative paralysis, n (%)	67 (44.7)	35 (57.4)	32 (36.0)	0.01
Other preoperative neurological findings, n (%)	102 (68)	45 (73.8)	57 (64.0)	0.22
Operation strategy (biopsy), n(%)	27 (18)	20 (32.8)	7(7.9)	< 0.001
Awake surgery, n (%)	54 (36)	19 (31.1)	35 (39.3)	0.39
5-aminelevulinic acid (5-ALA), n (%)	78 (52)	25 (41.0)	53 (59.6)	0.03
Photodynamic therapy, n (%)	7 (4.7)	1 (1.6)	6 (6.7)	0.24
Carmustine wafers placement, n(%)	36 (24)	9 (14.8)	27 (30.3)	0.03

MEP or SEP monitoring, n (%)	**76 (50.1)**	**26 (42.6)**	**50 (56.2)**	**0.14**
Temozolomide chemotherapy, n (%)	**142 (94.7)**	**56 (91.8)**	**86 (89.9)**	**0.27**
Bevacizumab chemotherapy, n (%)	**65 (43.3)**	**32 (52.5)**	**33 (37.1)**	**0.068**
MGMT promoter methylation, n (%)	**51 (34)**	**19 (31.1)**	**32 (36.0)**	**0.60**
TERT promoter mutation, n (%)	**49 (32.7)**	**19 (31.1)**	**30 (33.7)**	**0.86**
Median MIB-1 index, % (range)	**22.5 (0 – 90)**	**21.2 (10 – 80)**	**24.8 (0 – 90)**	**0.45**
IHC staining of MGMT, n (%)	**54 (36)**	**20 (32.8)**	**34 (38.2)**	**0.60**
Radiation dose, Gy (range)	**60 (0 – 63.2)**	**40.05 (0 – 60)**	**60 (0 – 63.2)**	**< 0.001**
Number of radiation fraction (Fr)	**30 (0 – 30)**	**15 (0 – 30)**	**30 (0 – 30)**	**< 0.001**
Tumor laterality, n (%)				
Right	**75 (50)**	**27 (44.3)**	**48 (53.9)**	**0.47**
Left	**63 (42)**	**28 (45.9)**	**35 (39.3)**	**-**
Bilateral	**12 (8.0)**	**6 (9.8)**	**6 (6.7)**	**-**
Ependymal invasion, n (%)	**107 (71.3)**	**51 (83.6)**	**56 (62.9)**	**0.006**
Midline shift, n (%)	**63 (42)**	**29 (47.5)**	**34 (38.2)**	**0.31**
Corpus callosum invasion, n (%)	**41 (27.3)**	**24 (39.3)**	**17 (19.1)**	**0.009**
Necrosis/cysts evident on imaging, n (%)	**138 (0.92)**	**58 (95.1)**	**80 (89.9)**	**0.36**
Extent of resection, n (%)				
1–49%	**30 (20)**	**22 (36.7)**	**8 (9.0)**	**< 0.001**
50–89%	**21 (14)**	**12 (19.7)**	**9 (10.1)**	**-**
90–99%	**43 (28.7)**	**10 (16.4)**	**33 (37.1)**	**-**
100%	**56 (37.3)**	**17 (27.9)**	**39 (43.8)**	**-**
Karnofsky performance status				
Median preoperative KPS, (range)	**80 (20 – 100)**	**70 (20 – 100)**	**80 (40 – 100)**	**< 0.001**
Preoperative KPS ≥ 70, n (%)	**118 (78.7)**	**43 (70.5)**	**75 (84.3)**	**0.002**
Median 6-months postoperative KPS, (range)	**70 (0 – 100)**	**50 (0 – 60)**	**80 (70 – 100)**	**< 0.001**

Categorical variables are presented as the count of patients (percentage), while continuous variables are shown as the median value (range).

MEP, motor evoked potentials. SEP, Somatosensory evoked potentials. MGMT, O-6-Methylguanine-DNA Methyltransferase. TERT, Telomerase Reverse Transcriptase. IHC, Immunohistochemical. KPS, Karnofsky performance status.

Model Development

Using the training set, we developed three models to predict 6-month postoperative KPS scores of <70: a clinical-based model, an MRI-based model, and a multimodal model that incorporated both MRI features and clinical data. When both clinical and MRI data were utilized, the area under the curve (AUC) was higher than when using either clinical or MRI data alone (Fig 4A–C, E). The mean AUC was 0.785 (SD 0.051) for the multimodal model using both clinical and MRI features, 0.716 (SD 0.059, P<0.001) when only clinical parameters were considered, and 0.651 (SD 0.028, P<0.001) when only MRI features were used.

Fig 4. Development of models to predict 6-month postoperative KPS score of <70 using the training set.

During model development, performance was evaluated using the training set with 10 repetitions of fivefold cross-validation. The ROC curves for each repeat of the fivefold cross-validation are represented in gray, whereas the mean ROC curve for the 10 repeats is shown in blue. The area under the curve of the model was 0.715 when using clinical parameters (**A**), 0.651 when using deep imaging features from pre- and postoperative MRI (**B**), and 0.785 when combining clinical parameters with deep imaging features (**C**). (**D**) The top five feature contributions in the multimodal model are evaluated by grouped permutation importance. (**E**) The predicted performance of each model. Data are shown as the mean score ± standard deviation.

KPS=Karnofsky performance status.

Multimodal models also outperform clinical-based and MRI-based models in terms of accuracy (clinical parameters plus MRI features: 0.728 [SD 0.032]; clinical parameters: 0.674 [SD 0.045, P<0.021]; MRI features: 0.631 [SD 0.021, P<0.001]), sensitivity (clinical parameters plus MRI features: 0.529 [SD 0.091]; clinical parameters: 0.406 [SD 0.117, P=0.039]; MRI features: 0.402 [SD 0.076, P=0.01]), and F1 score (clinical parameters plus MRI features: 0.572 [SD 0.066]; clinical parameters: 0.434 [SD 0.086, P=0.002]; MRI features: 0.439 [SD 0.078, P=0.005]). The specificity of the multimodal model was 0.847 [SD=0.057], which was not significantly different from the clinical-based model (0.834, [SD=0.115, P=0.77]) and the MRI-based model (0.789, [SD=0.042, P=0.06]).

The neural network-based prediction model outperformed other machine learning classifiers, including the Random Forest classifier, XGBoost, and LightGBM, when utilizing clinical parameters and MRI features. The mean AUC scores, determined through 10 repeated fivefold cross-validations with optimized hyperparameters, were calculated for each classifier. The mean AUC scores and associated P values compared with the neural network model were as follows: Random Forest classifier, 0.663 (SD 0.020, P<0.001); XGBoost, 0.705 (SD 0.029, P<0.005); and LightGBM, 0.720 (SD 0.23, P=0.01) (S4 Fig). Moreover, the test set evaluation of our neural network models—clinical-based, MRI-based, and multimodal—demonstrated the superiority of the multimodal model with the mean AUC of 0.810. Its performance metrics included accuracy of 0.727, specificity of 0.643, sensitivity of 0.789, and an F1 score of 0.76, mirroring the training set outcomes (Fig 5).

Fig 5. Prediction performance in the test set

The clinical-based, MRI-based, and multimodal models were pretrained using the training set, then their prediction performance was evaluated using the test set.

Model Interpretability Analysis

To improve the understanding, trust, and verification of the model predictions, grouped permutation importance was applied.[13] Grouped permutation importance quantifies the feature contribution, thus providing an interpretable relationship between the incorporated features and the

model prediction. Fig 4D illustrates the importance of the top five relative features for the multimodal model to predict a 6-month KPS score of <70. The most important feature in the model was "age," followed by "radiation dose (Gy)" and "preoperative KPS." Furthermore, "postoperative mask image" ranked fourth, while "preoperative mask image" ranked fifth, and these MRI features also contributed to the model prediction.

DISCUSSION

Improved Performance When Incorporating Multimodal Data

The current study demonstrated that combining imaging features with clinical parameters is effective in improving the performance of clinical-based models, leading to the construction of clinically implementable models. There has been a notable increase in machine learning-based models to solve medical challenges in glioblastoma[14]; however, these models typically use data from only one modality (e.g., clinical parameters). Recently, several researchers have successfully improved the performance of models designed for clinical implementation by combining multiple modalities rather than relying on a single modality.[15] To construct a multimodal model incorporating medical imaging data into clinical-based models, it is necessary to extract imaging features from the radiographic images. The imaging features used in this process can be broadly categorized into two types: handcrafted and deep imaging features.[16] In general, handcrafted features are defined by the use of explicit formulas and are often derived from morphological, statistical, and textural properties. On the other hand, deep imaging features are generated through a deep learning using transfer learning.[17]

Lao et al. examined the importance of deep imaging and handcrafted features in the development of an overall survival prediction model for 112 patients with glioblastoma.[16] They compared 1,403 handcrafted features with 98,304 deep imaging features, which were extracted using a convolutional neural network from the preoperative MRIs. They concluded that deep imaging features

contributed more to the model's performance. Recently, reviewing 69 studies of radiomic models, Demircioglu reported that clinical-based models constructed based on deep imaging features often outperform those relying on handcrafted features. Additionally, the author suggested that combining the two into a fused model could potentially enhance model performance.[17] When processing three-dimensional MRI data using a pretrained convolutional neural network, a very large number of features were generated compared to the number of patients. Therefore, strong feature selection and shrinkage are required to develop reliable clinical-based models and increase interpretation.[18]

In the present study, we utilized a VAE as a feature extractor and demonstrated significant improvements in the performance of the KPS prediction model for patients with glioblastoma. Deep imaging features (i.e., MRI features) were extracted from the latent space and subsequently combined with clinical parameters. When jointly trained on data from MRI and clinical parameters in the training set, the mean AUC for predicting a 6-month postoperative KPS score of <70 was consistently higher (0.785, SD 0.051) than the models trained solely on clinical parameters (0.716, SD 0.059, P=0.038) or MRI features (0.651, SD 0.028, P<0.001) (Fig 4E). Furthermore, in the test set, the multimodal model's AUC was 0.810, surpassing the clinical-based model's AUC of 0.670 and the MRI-based model's AUC of 0.650 (Fig 5). The important features contributing to the development of the combined model were evaluated using group permutation importance.[13] Among the top five important features, three clinical parameters, namely "age," "radiation dose," and "preoperative KPS," were included, along with the deep imaging features extracted from MRI, "postoperative mask image," and "preoperative mask image." Analysis of grouped permutation importance supported the idea that MRI features contributed to model development and improved model performance. We consistently observed an improvement in prediction performance when incorporating clinical parameters and MRI features. This aligns with findings from other studies that have utilized deep learning models to merge diverse data modalities, including scenarios where clinical parameters were integrated with chest X-rays or cancer biomarkers were fused with MRI data.[19,20] The integration of medical imaging data with corresponding medical parameters is proving to be a valuable approach for enhancing model performance.

Clinical Implication

Prediction of health status and functional impairment is critical for clinical and personal decision-making in patients with glioblastoma. A KPS score of ≥70 indicated that the patients were capable of independent self-care. Identifying patients who will require nursing or caregiving 6 months postoperatively or patients who are currently in need of care but are expected to recover independent living within 6 months is crucial for providing personalized medical management.

A low KPS was significantly correlated with a poor prognosis. For patients presenting with a KPS score of <70, less invasive treatments may be considered as an alternative to the standard protocol, which generally includes tumor removal followed by chemoradiotherapy.[21] A recent retrospective analysis revealed that the mean overall survival for patients with a postoperative KPS score of <70 was 8 months.[4] Moreover, in clinical practice, when managing recurrent glioblastoma, their performance status can significantly influence therapeutic decision-making, which may involve options like surgical re-intervention, re-irradiation, or best supportive care.[22] Due to the limited availability of publicly accessible, precise clinical databases, research predicting the course of performance status in patients with malignant tumors is much scarcer compared to the development of prediction models for overall survival or progression-free survival. However, in recent years, several studies have reported that machine learning approaches have successfully predicted changes in KPS 6 months postoperatively in patients with glioblastoma, as well as poor performance status in patients with cancer 6 months after diagnosis.[23,24] The development of a KPS prediction model could help stratify patients based on their anticipated clinical course, resulting in significant implications for optimizing the balance between preserving quality of life and pursuing a more aggressive treatment approach.

The prediction of prognosis using pre- and postoperative MRI may also contribute to surgical planning. The planning and outcome of brain tumors are influenced by the surgeon's experience and involve weighing the benefits of resection against the risk of neurological impairment.[25] It has

recently been reported that preoperative T1Gd MRI images can accurately predict surgical resectability using a neural network.[26] Furthermore, a recent study has demonstrated that generative artificial intelligence models are capable of producing fine-quality images of brain tumors and normal parenchyma.[27] We assumed that the present KPS prediction model could enhance surgical simulation when integrated with a reliable surgical resectability prediction model and by using generative artificial intelligence to produce postoperative MRI images from preoperative scans.

Limitations

This study was limited by its relatively small patient cohort. Therefore, well-powered studies are required. As a retrospective study conducted at a single center, its external validity may have been limited by patient selection bias in our department. In accordance with the previously reported critical appraisal guidelines for AI research, our study aligns with Level 5B: one retrospective study with only internal data used for final performance reporting.[28] Our understanding of surgical techniques and adjuvant therapy has gradually evolved. Patients treated in the latter years of this study likely benefited from our greater knowledge and improved treatments that are not included as clinical parameters in this study.

CONCLUSIONS

Imaging features extracted from MRI scans using VAEs may provide valuable representations reflecting the prognosis of patients with wild-type IDH glioblastoma. The integration of these imaging features, achieved through the development of a multimodal model, significantly enhanced the performance of the neural network-based prediction model. Predicting the 6-month postoperative KPS score has the potential to impact personalized treatment decisions, including the selection of treatment intensity and consideration of early palliative care. The future clinical implementation of the KPS prediction model offers the possibility of tailored medical interventions.

Authors' contributions

Conception and design: Tomoki Sasagasako, Yohei Mineharu, Akihiko Ueda; Acquisition of data: Tomoki Sasagasako, Yohei Mineharu; Analysis and interpretation of data: Tomoki Sasagasako, Yohei Mineharu, Akihiko Ueda, Yasushi Okuno; Drafting the article: Tomoki Sasagasako; Critically revising the article: Akihiko Ueda, Yohei Mineharu, Yoshiki Arakawa, Yasushi Okuno; Administrative/technical/material support: Yusuke Mochizuki, Souichiro Doi, Silsu Park, Yukinori Terada, Noritaka Sano, Masahiro Tanji; Study supervision: Yasushi Okuno and Yoshiki Arakawa; Approval of the final version of the manuscript: all authors.

Code availability

The code used to develop the model described herein is publicly available on GitHub: https://github.com/TomokiSasagasako/GBM_KPS_prediction.git

The clinical data in this study are available, on reasonable request, from the corresponding author.

Funding

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) under the RIKEN joint research and collaboration fund for "Translational Research in Basic and Clinical Sciences for the Construction of an AI Pharmaceutical Platform".

REFERENCES

1. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC

trial. Lancet Oncol. 2009;10: 459–466. doi:10.1016/S1470-2045(09)70025-7

2. Wankhede DS, Selvarani R. Dynamic architecture based deep learning approach for glioblastoma brain tumor survival prediction. Neurosci Informatics. 2022;2: 100062. doi:10.1016/j.neuri.2022.100062

3. West H, Jin JO. Performance status in patients with cancer. JAMA Oncol. 2015;1: 998. doi:10.1001/jamaoncol.2015.3113

4. Barz M, Gerhardt J, Bette S, Aftahy AK, Huber T, Combs SE, et al. Prognostic value of tumour volume in patients with a poor Karnofsky performance status scale – a bicentric retrospective study. BMC Neurol. 2021;21: 1–10. doi:10.1186/s12883-021-02424-0

5. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. BMC Med. 2015;13: 1–10. doi:10.1186/s12916-014-0241-z

6. Stupp R, Mason WP, van den Bent MJ et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma in elderly patients. J Neurooncol. 2008;88: 97–103. doi:10.1007/s11060-008-9538-0

7. Roa W, Brasher PMA, Bauman G, Anthes M, Bruera E, Chan A, et al. Abbreviated course of radiation therapy in older patients with glioblastoma multiforme: A prospective randomized clinical trial. J Clin Oncol. 2004;22: 1583–1588. doi:10.1200/JCO.2004.06.082

8. Kleesiek J, Urban G, Hubert A, Schwarz D, Maier-Hein K, Bendszus M, et al. Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. Neuroimage. 2016;129: 460–469. doi:10.1016/j.neuroimage.2016.01.024

9. Kermi A, Mahmoudi I, Khadir MT. Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. In: Crimi A, Bakas S, Kujif S, Keyvan F, Reyes M, van Walsum T, eds. Brain lesion: glioma, multiple sclerosis, stroke and traumatic brain inju. 2019;11384. doi:10.1007/978-3-030-11726-9

10. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine

Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. 2018. Available: http://arxiv.org/abs/1811.02629

11. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162: W1–W73. doi:10.7326/M14-0698

12. Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell. 2010;32: 569–575. doi:10.1109/TPAMI.2009.187

13. Plagwitz L, Brenner A, Fujarski M, Varghese J. Supporting AI-Explainability by Analyzing Feature Subsets in a Machine Learning Model. Stud Health Technol Inform. 2022;294: 109–113. doi:10.3233/SHTI220406

14. Valdebenito J, Medina F. Machine learning approaches to study glioblastoma: a review of the last decade of applications. Cancer Rep. 2019;2: 1–15. doi:10.1002/cnr2.1226

15. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3. doi:10.1038/s41746-020-00341-z

16. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A Deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Sci Rep. 2017;7: 1–8. doi:10.1038/s41598-017-10649-8

17. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. Eur Radiol Exp. 2023;7. doi:10.1186/s41747-023-00325-0

18. Zhang X, Maltbie EA, Keilholz SD. Spatiotemporal trajectories in resting-state FMRI revealed by convolutional variational autoencoder. Neuroimage. 2021;244: 105–124. doi:10.1016/j.neuroimage.2021.118588

19. Reda I, Khalil A, Elmogy M, El-Fetouh AA, Shalaby A, El-Ghar MA, et al. Deep learning role

in early diagnosis of prostate cancer. Technol Cancer Res Treat. 2018;17: 1–11. doi:10.1177/1533034618775530

20. Khader F, Müller-Franzes G, Wang T, Han T, Tayebi Arasteh S, Haarburger C, et al. Multimodal deep learning for integrating chest radiographs and clinical Parameters: a case for transformers. Radiology. 2023;309: e230806. doi:10.1148/radiol.230806

21. McKinnon C, Nandhabalan M, Murray SA, Plaha P. Glioblastoma: Clinical presentation, diagnosis, and management. BMJ. 2021;374. doi:10.1136/bmj.n1560

22. Demogeot N, Salleron J, Rech F, Taillandier L, Royer P, Vogin G. Impact of fractionated stereotactic radiotherapy on activity of daily living and performance status in progressive/recurrent glioblastoma: a retrospective study. Radiat Oncol. 2022;17: 1–9. doi:10.1186/s13014-022-02169-1

23. Della Pepa GM, Caccavella VM, Menna G, Ius T, Auricchio AM, Chiesa S, et al. Machine learning–based prediction of 6-month postoperative Karnofsky performance status in patients with glioblastoma: capturing the real-life interaction of multiple clinical and oncologic factors. World Neurosurg. 2021;149: e866–e876. doi:10.1016/j.wneu.2021.01.082

24. Seow H, Tanuseputro P, Barbera L, Earle CC, Guthrie DM, Isenberg SR, et al. Development and validation of a prediction model of poor performance status and severe symptoms over time in cancer patients (PROVIEW+). Palliat Med. 2021;35: 1713–1723. doi:10.1177/02692163211019302

25. Ramakrishna R, Hsu WC, Mao J, Sedrakyan A. Surgeon annual and cumulative volumes predict early postoperative outcomes after brain tumor resection. World Neurosurg. 2018;114: e254–e266. doi:10.1016/j.wneu.2018.02.172

26. Marcus AP, Marcus HJ, Camp SJ, Nandi D, Kitchen N, Thorne L. Improved prediction of surgical resectability in patients with glioblastoma using an artificial neural network. Sci Rep. 2020;10: 1–9. doi:10.1038/s41598-020-62160-2

27. Mukherkjee D, Saha P, Kaplun D, Sinitca A, Sarkar R. Brain tumor image generation using an

aggregation of GAN models with style transfer. Sci Rep. 2022;12: 1–16. doi:10.1038/s41598-022-12646-y

28. Pham N, Hill V, Rauschecker A, Lui Y, Niogi S, Fillipi CG, et al. Critical Appraisal of Artificial Intelligence Enabled Imaging Tools Using the Levels of Evidence System. Am J Neuroradiol. 2023;44: E21–E28. doi:10.3174/ajnr.A7850

Supporting information

S1 Fig. Participants flow

S2 Fig. The architecture of the variational autoencoder

S3 Fig. The KPS prediction model development using neural network

S4 Fig. Comparison between the neural network and other machine learning algorithm

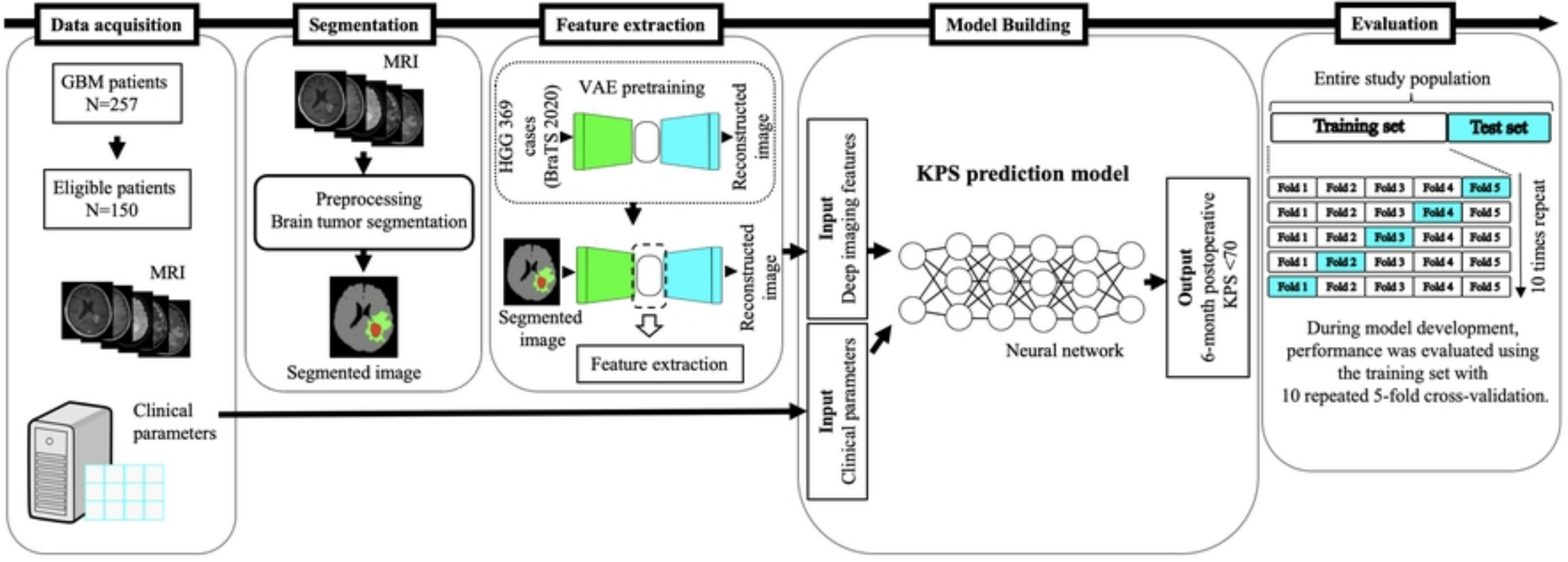
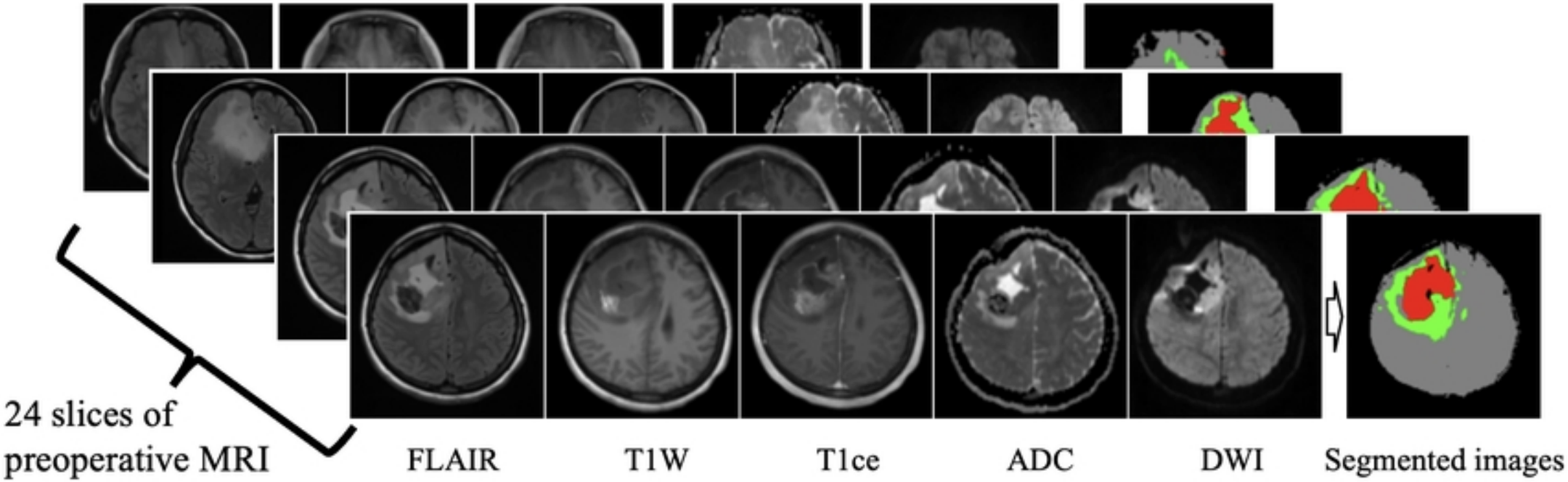

Data acquisition
GBM patients
N=257
Eligible patients
N=150
MRI
Clinical parameters
Segmentation
MRI
Preprocessing
Brain tumor segmentation
Segmented image
Feature extraction
VAE pretraining
HGG 369 cases (BraTS 2020)
Reconstructed image
Segmented image
Reconstructed image
Feature extraction
Model Building
Input
Deep imaging features
Input
Clinical parameters
KPS prediction model
Neural network
Output
6-month postoperative KPS <70
Evaluation
Entire study population
Training set
Test set
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
10 times repeat
During model development, performance was evaluated using the training set with 10 repeated 5-fold cross-validation.

Fig1


24 slices of
preoperative MRI
FLAIR
T1W
T1ce
ADC
DWI
Segmented images

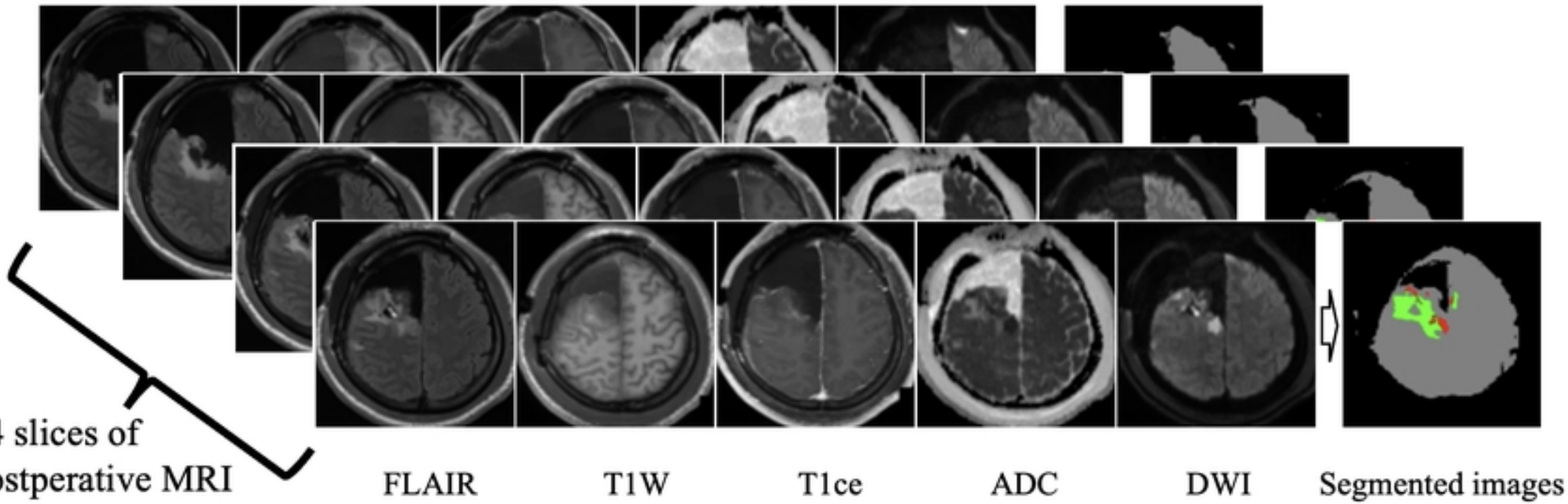
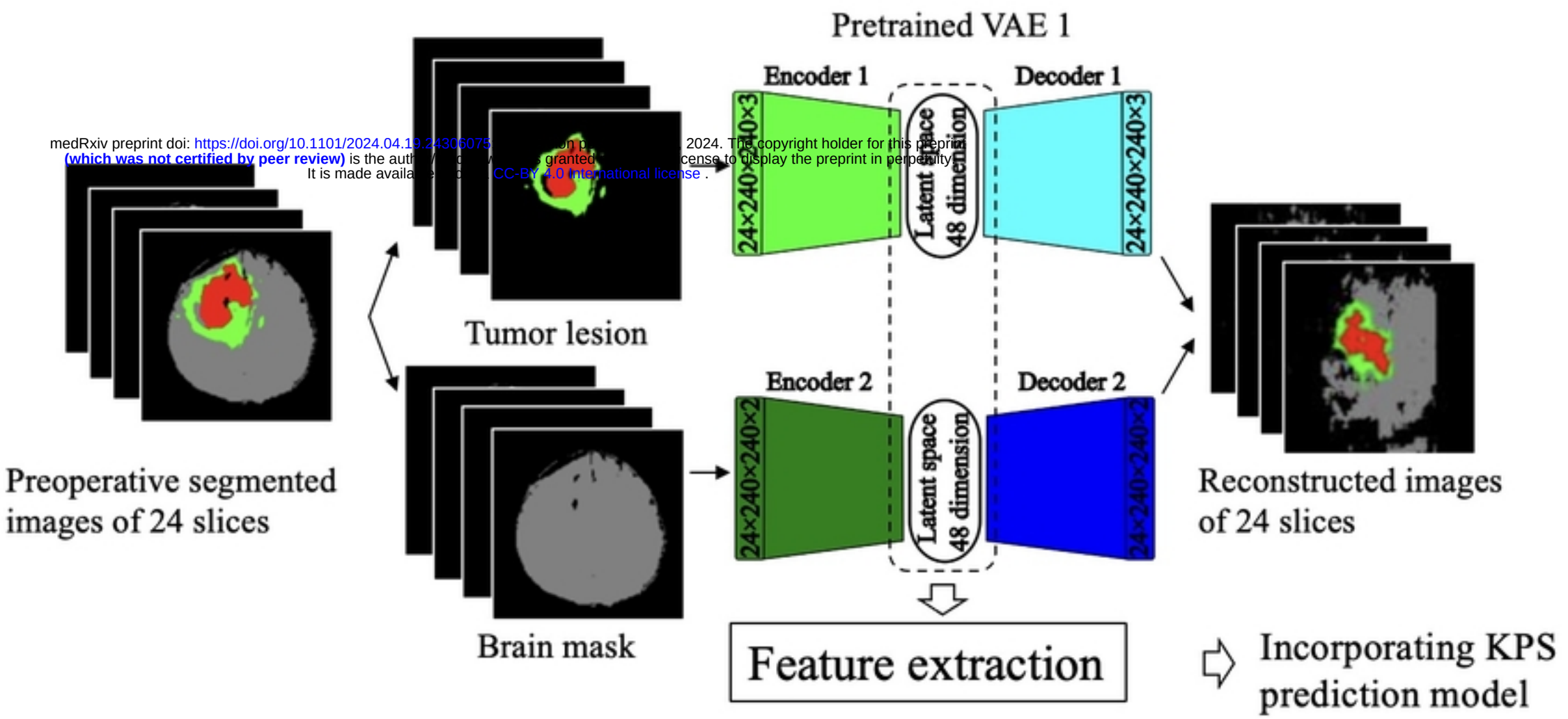

24 slices of
postperative MRI
FLAIR
T1W
T1ce
ADC
DWI
Segmented images

Fig2



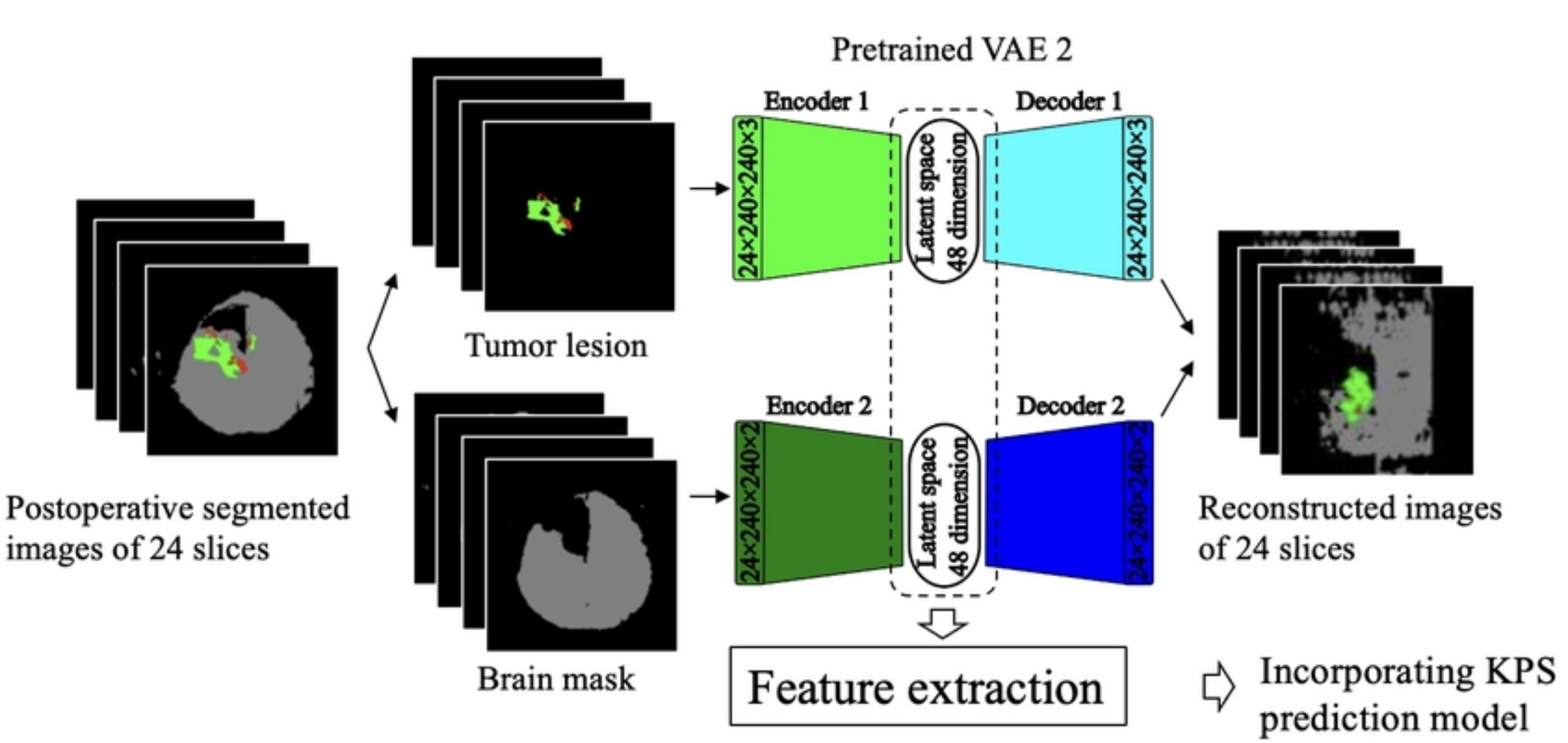
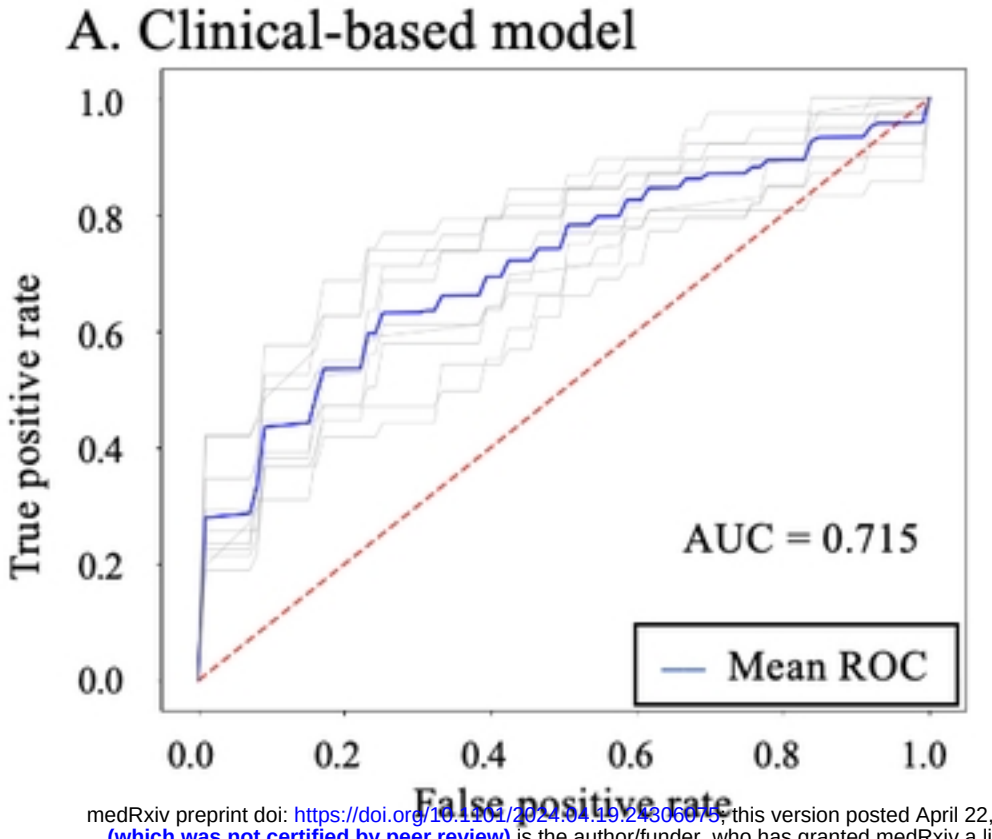
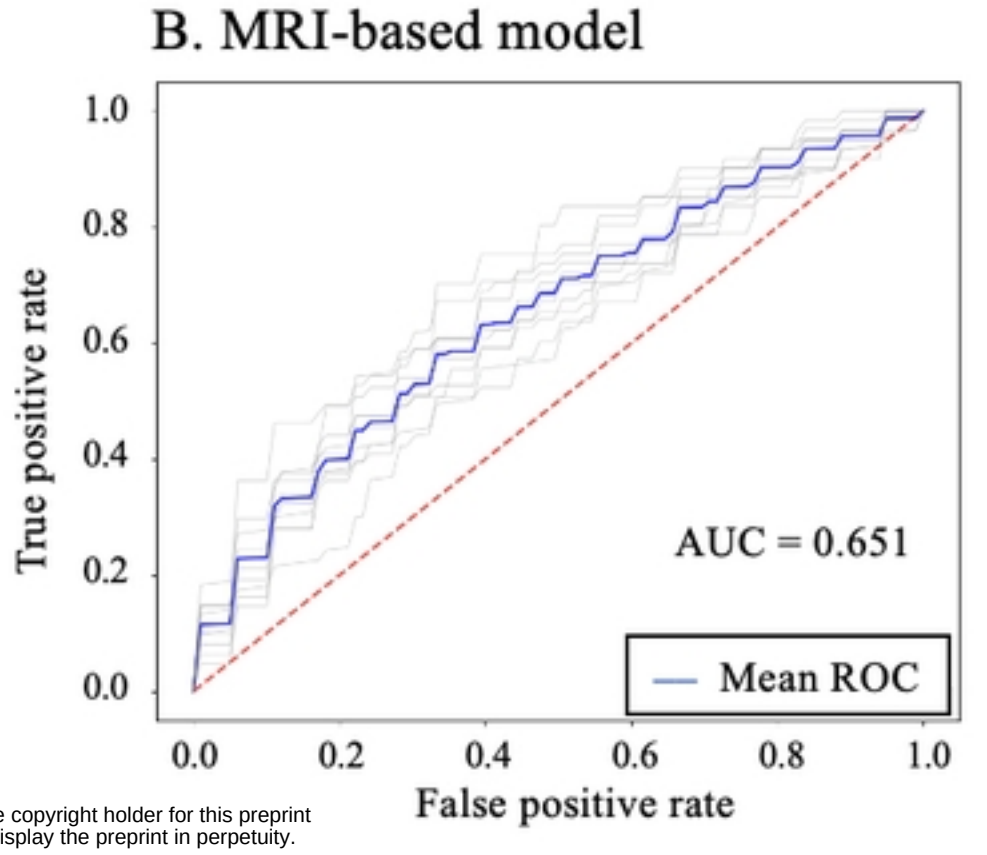
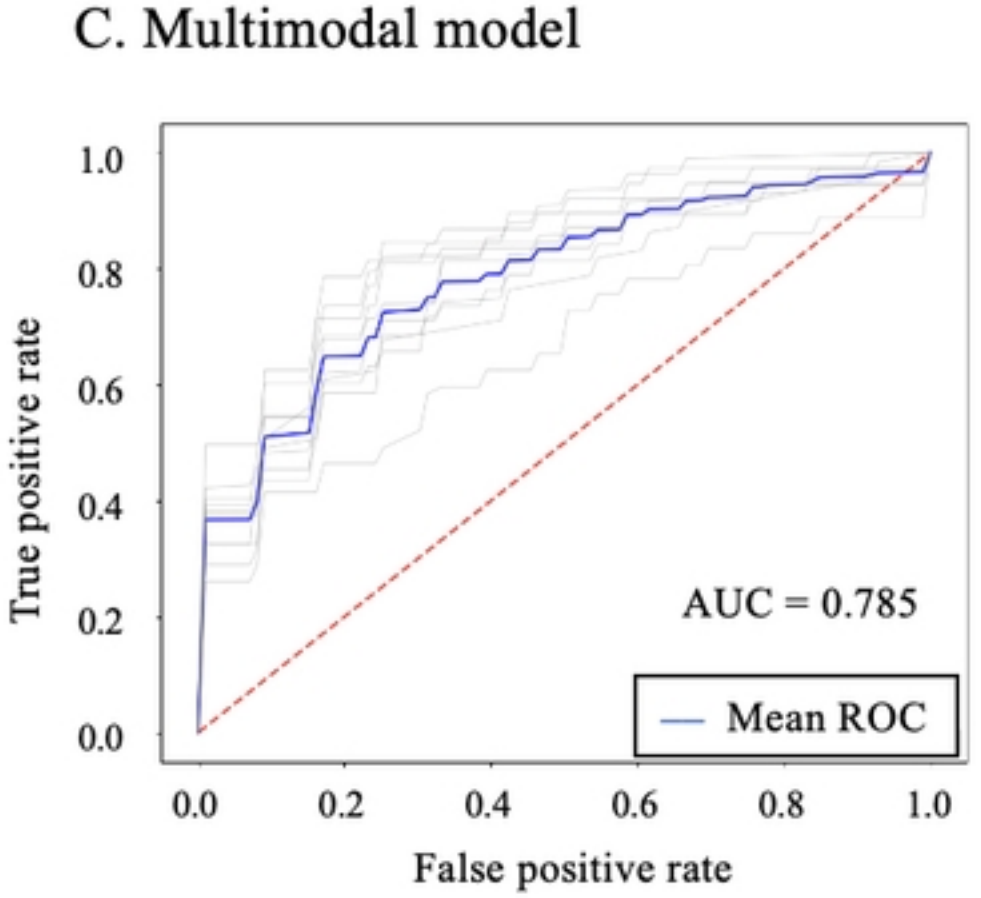
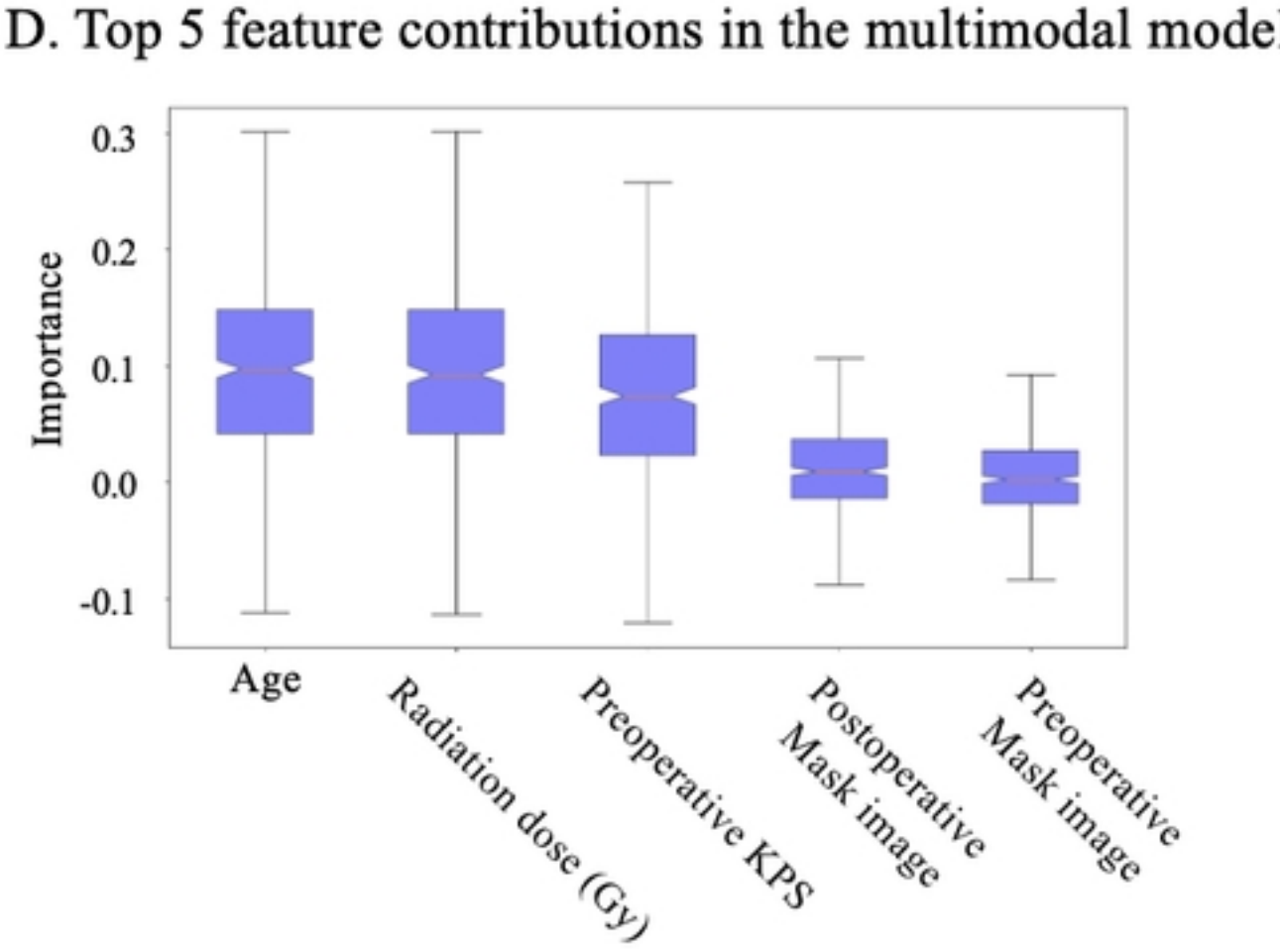


Fig3









E. Predictive performance of each model in the training set

Metrics	Clinical-based model	MRI-based model	Multimodal model	P value†	P value*
AUC	0.716 ± 0.059	0.651 ± 0.028	0.785 ± 0.051	**0.038**	**<0.001**
Accuracy	0.674 ± 0.045	0.631 ± 0.021	0.728 ± 0.032	**0.021**	**<0.001**
Specificity	0.834 ± 0.115	0.789 ± 0.042	0.847 ± 0.057	0.77	0.06
Sensitivity	0.406 ± 0.117	0.402 ± 0.076	0.529 ± 0.091	**0.039**	**0.01**
F1 score	0.434 ± 0.086	0.439 ± 0.078	0.572 ± 0.066	**0.002**	**0.005**

Mean score ± Standard deviation
† Multimodal model versus clinical-based model
* Multimodal model versus MRI-based model

Fig4

Metrics	Clinical-based model	MRI-based model	Multimodal model
AUC	0.670	0.650	0.810
Accuracy	0.660	0.540	0.727
Specificity	0.740	0.815	0.643
Sensitivity	0.565	0.217	0.789
F1 score	0.604	0.303	0.769

Fig5