

1 **Postoperative Karnofsky performance status prediction in** 2 **patients with *IDH* wild-type glioblastoma: a multimodal** 3 **approach integrating clinical and deep imaging features**

4 Tomoki Sasagasako, MD,^{1,2} Akihiko Ueda, MD, PhD,² Yohei Mineharu, MD, PhD,³ Yusuke
5 Mochizuki⁴, Souichiro Doi⁴, Silsu Park, MD,¹ Yukinori Terada, MD, PhD,¹ Noritaka Sano, MD,
6 PhD,¹ Masahiro Tanji, MD, PhD,¹ Yoshiki Arakawa, MD, PhD,¹ Yasushi Okuno, PhD^{2,3}

7 8 **Affiliations:**

9 ¹ Department of Neurosurgery, Kyoto University Graduate School of Medicine, Kyoto, Japan

10 ² Department of Biomedical Data Intelligence, Kyoto University Graduate School of Medicine, Kyoto,
11 Japan

12 ³ Department of Artificial Intelligence in Healthcare and Medicine, Kyoto University Graduate School
13 of Medicine, Kyoto, Japan

14 ⁴ Kyoto University Faculty of Medicine, Kyoto, Japan

15 16 **Correspondence:**

17 Yohei Mineharu, MD, PhD

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

20 +81-75-751-3111

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

22 ABSTRACT

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

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

38 **Results:** Among the 150 patients, 61 had 6-month postoperative KPS scores below 70 and 89
39 scored 70 or higher. We developed three models: a clinical-based model, an MRI-based model, and a
40 multimodal model that incorporated both clinical parameters and MRI features. In the training set, the
41 mean AUC was 0.785 ± 0.051 for the multimodal model, which was significantly higher than the
42 clinical-based model (0.716 ± 0.059 , $P=0.038$) using only clinical parameters and MRI-based model
43 (0.651 ± 0.028 , $P<0.001$) using only MRI features. In the test set, the multimodal model achieved an
44 AUC of 0.810, outperforming the clinical-based (0.670) and MRI-based (0.650) models.

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

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

49 **Abbreviations:**

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

52 **Disclosure of potential conflicts of interest:**

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

54

55 INTRODUCTION

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

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

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

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

76

77 MATERIALS AND METHODS

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

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

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

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

90

91 **Ethics Approval**

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

98

99 **Patients**

100 This single-center retrospective review was conducted between December 2001 and December
101 2022 at our institution. This study included consecutive adult patients aged 18 years and older who
102 were newly histopathologically diagnosed with glioblastoma with isocitrate dehydrogenase (IDH) wild
103 type. The exclusion criteria for the study cohort were as follows: 1) gliomas that were not

104 histopathologically diagnosed as glioblastomas, 2) lack of immunohistochemical testing for the IDH1
105 R132H mutation or absence of IDH2 sequencing, 3) a diagnosis of IDH-mutant gliomas, and 4) a
106 follow-up period of <6 months. A total of 257 patients were pathologically confirmed to have
107 glioblastoma. Among these, 87 patients were excluded based on the following criteria: diagnosis of
108 IDH-mutant gliomas (n=11) or absence of either immunohistochemistry for the IDH1 R132H
109 mutation or IDH2 sequencing (n=76). Additionally, patients lacking sufficient medical records or
110 imaging data (n=18) and those aged <18 years (n=2) were also excluded. Subsequently, patients were
111 divided into training and test sets based on their first operation date. Patients who underwent surgery
112 from December 2001 to October 2018 were assigned to the training set, while those from November
113 2018 to December 2022 were included in the test set. This process resulted in 100 patients in the
114 training set and 50 in the test set (S1 Fig).

115 **Clinical Parameters and Endpoints**

116 The clinical parameters included following 28 variables.

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

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

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

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

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

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

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

133 For the standard concurrent chemoradiotherapy regimen, patients received fractionated focal
134 radiation therapy with a cumulative dose of 60 Gy, accompanied by concomitant TMZ
135 chemotherapy.[6] For elderly patients aged 70 years or older, a hypofractionated radiotherapy schedule
136 of 40 Gy delivered in 15 fractions over 3 weeks was employed.[7] All patients were followed up every
137 1–2 months after surgical treatment.

138

139 **Imaging Acquisition and Preprocessing**

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

146

147 **Tumor Segmentation**

148 To achieve semantic segmentation of the glioblastoma lesion on MRI scans, we utilized a
149 segmentation model based on the U-net architecture, specifically designed for the segmentation of
150 glioma.[9] Using this segmentation model, we highlighted the segmented regions, including enhanced
151 tumors, necrosis, and cystic lesions, which are shown in red. Peritumoral edema and non-enhancing
152 tumor areas are shown in green (Fig 2). Images with a thickness of 3 mm were used for each MRI
153 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.

179

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

185

186 **Development of a KPS Prediction Model Using Neural Networks**

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

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

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

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

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

208

209 **Model Interpretability**

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

219

220 **Statistical Analysis**

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

226

227

228 **RESULTS**

229 Patient Demographics

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

239 **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 (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

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

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

245

246

247 **Model Development**

248 Using the training set, we developed three models to predict 6-month postoperative KPS scores
249 of <70: a clinical-based model, an MRI-based model, and a multimodal model that incorporated both
250 MRI features and clinical data. When both clinical and MRI data were utilized, the area under the
251 curve (AUC) was higher than when using either clinical or MRI data alone (Fig 4A–C, E). The mean
252 AUC was 0.785 (SD 0.051) for the multimodal model using both clinical and MRI features, 0.716 (SD
253 0.059, $P<0.001$) when only clinical parameters were considered, and 0.651 (SD 0.028, $P<0.001$) when
254 only MRI features were used.

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

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

265 KPS=Karnofsky performance status.

266

267 Multimodal models also outperform clinical-based and MRI-based models in terms of
268 accuracy (clinical parameters plus MRI features: 0.728 [SD 0.032]; clinical parameters: 0.674 [SD
269 0.045, $P<0.021$]; MRI features: 0.631 [SD 0.021, $P<0.001$]), sensitivity (clinical parameters plus MRI
270 features: 0.529 [SD 0.091]; clinical parameters: 0.406 [SD 0.117, $P=0.039$]; MRI features: 0.402 [SD
271 0.076, $P=0.01$]), and F1 score (clinical parameters plus MRI features: 0.572 [SD 0.066]; clinical
272 parameters: 0.434 [SD 0.086, $P=0.002$]; MRI features: 0.439 [SD 0.078, $P=0.005$]). The specificity of
273 the multimodal model was 0.847 [SD=0.057], which was not significantly different from the clinical-
274 based model (0.834, [SD=0.115, $P=0.77$]) and the MRI-based model (0.789, [SD=0.042, $P=0.06$]).

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

285 **Fig 5. Prediction performance in the test set**

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

288

289 **Model Interpretability Analysis**

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

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

298

299 **DISCUSSION**

300 **Improved Performance When Incorporating Multimodal Data**

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

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

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

325 In the present study, we utilized a VAE as a feature extractor and demonstrated significant
326 improvements in the performance of the KPS prediction model for patients with glioblastoma. Deep
327 imaging features (i.e., MRI features) were extracted from the latent space and subsequently combined
328 with clinical parameters. When jointly trained on data from MRI and clinical parameters in the
329 training set, the mean AUC for predicting a 6-month postoperative KPS score of <70 was consistently
330 higher (0.785, SD 0.051) than the models trained solely on clinical parameters (0.716, SD 0.059,
331 $P=0.038$) or MRI features (0.651, SD 0.028, $P<0.001$) (Fig 4E). Furthermore, in the test set, the
332 multimodal model's AUC was 0.810, surpassing the clinical-based model's AUC of 0.670 and the
333 MRI-based model's AUC of 0.650 (Fig 5). The important features contributing to the development of
334 the combined model were evaluated using group permutation importance.[13] Among the top five
335 important features, three clinical parameters, namely "age," "radiation dose," and "preoperative KPS,"
336 were included, along with the deep imaging features extracted from MRI, "postoperative mask image,"
337 and "preoperative mask image." Analysis of grouped permutation importance supported the idea that
338 MRI features contributed to model development and improved model performance. We consistently
339 observed an improvement in prediction performance when incorporating clinical parameters and MRI
340 features. This aligns with findings from other studies that have utilized deep learning models to merge
341 diverse data modalities, including scenarios where clinical parameters were integrated with chest X-
342 rays or cancer biomarkers were fused with MRI data.[19,20] The integration of medical imaging data
343 with corresponding medical parameters is proving to be a valuable approach for enhancing model
344 performance.

345

346 **Clinical Implication**

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

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

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

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

377 **Limitations**

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

386

387 **CONCLUSIONS**

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

395

396 **Authors' contributions**

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

404

405 **Code availability**

406 The code used to develop the model described herein is publicly available on GitHub:

407 https://github.com/TomokiSasagasako/GBM_KPS_prediction.git

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

409

410 **Funding**

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

414

415

416 **REFERENCES**

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

- 420 trial. *Lancet Oncol.* 2009;10: 459–466. doi:10.1016/S1470-2045(09)70025-7
- 421 2. Wankhede DS, Selvarani R. Dynamic architecture based deep learning approach for
422 glioblastoma brain tumor survival prediction. *Neurosci Informatics.* 2022;2: 100062.
423 doi:10.1016/j.neuri.2022.100062
- 424 3. West H, Jin JO. Performance status in patients with cancer. *JAMA Oncol.* 2015;1: 998.
425 doi:10.1001/jamaoncol.2015.3113
- 426 4. Barz M, Gerhardt J, Bette S, Aftahy AK, Huber T, Combs SE, et al. Prognostic value of tumour
427 volume in patients with a poor Karnofsky performance status scale – a bicentric retrospective
428 study. *BMC Neurol.* 2021;21: 1–10. doi:10.1186/s12883-021-02424-0
- 429 5. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable
430 prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement.
431 *BMC Med.* 2015;13: 1–10. doi:10.1186/s12916-014-0241-z
- 432 6. Stupp R, Mason WP, van den Bent MJ et al. Radiotherapy plus concomitant and adjuvant
433 temozolomide for glioblastoma in elderly patients. *J Neurooncol.* 2008;88: 97–103.
434 doi:10.1007/s11060-008-9538-0
- 435 7. Roa W, Brasher PMA, Bauman G, Anthes M, Bruera E, Chan A, et al. Abbreviated course of
436 radiation therapy in older patients with glioblastoma multiforme: A prospective randomized
437 clinical trial. *J Clin Oncol.* 2004;22: 1583–1588. doi:10.1200/JCO.2004.06.082
- 438 8. Kleesiek J, Urban G, Hubert A, Schwarz D, Maier-Hein K, Bendszus M, et al. Deep MRI brain
439 extraction: A 3D convolutional neural network for skull stripping. *Neuroimage.* 2016;129: 460–
440 469. doi:10.1016/j.neuroimage.2016.01.024
- 441 9. Kermi A, Mahmoudi I, Khadir MT. Deep convolutional neural networks using U-Net for
442 automatic brain tumor segmentation in multimodal MRI volumes. In: Crimi A, Bakas S, Kujif
443 S, Keyvan F, Reyes M, van Walsum T, eds. *Brain lesion: glioma, multiple sclerosis, stroke and*
444 *traumatic brain inju.* 2019;11384. doi:10.1007/978-3-030-11726-9
- 445 10. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine

- 446 Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall
447 Survival Prediction in the BRATS Challenge. 2018. Available: <http://arxiv.org/abs/1811.02629>
- 448 11. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al.
449 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
450 (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162: W1–W73.
451 doi:10.7326/M14-0698
- 452 12. Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction
453 error estimation. *IEEE Trans Pattern Anal Mach Intell*. 2010;32: 569–575.
454 doi:10.1109/TPAMI.2009.187
- 455 13. Plagwitz L, Brenner A, Fujarski M, Varghese J. Supporting AI-Explainability by Analyzing
456 Feature Subsets in a Machine Learning Model. *Stud Health Technol Inform*. 2022;294: 109–
457 113. doi:10.3233/SHTI220406
- 458 14. Valdebenito J, Medina F. Machine learning approaches to study glioblastoma: a review of the
459 last decade of applications. *Cancer Rep*. 2019;2: 1–15. doi:10.1002/cnr2.1226
- 460 15. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and
461 electronic health records using deep learning: a systematic review and implementation
462 guidelines. *NPJ Digit Med*. 2020;3. doi:10.1038/s41746-020-00341-z
- 463 16. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A Deep learning-based radiomics model for
464 prediction of survival in glioblastoma multiforme. *Sci Rep*. 2017;7: 1–8. doi:10.1038/s41598-
465 017-10649-8
- 466 17. Demircioğlu A. Are deep models in radiomics performing better than generic models? A
467 systematic review. *Eur Radiol Exp*. 2023;7. doi:10.1186/s41747-023-00325-0
- 468 18. Zhang X, Maltbie EA, Keilholz SD. Spatiotemporal trajectories in resting-state fMRI revealed
469 by convolutional variational autoencoder. *Neuroimage*. 2021;244: 105–124.
470 doi:10.1016/j.neuroimage.2021.118588
- 471 19. Reda I, Khalil A, Elmogy M, El-Fetouh AA, Shalaby A, El-Ghar MA, et al. Deep learning role

- 472 in early diagnosis of prostate cancer. *Technol Cancer Res Treat*. 2018;17: 1–11.
- 473 doi:10.1177/1533034618775530
- 474 20. Khader F, Müller-Franzes G, Wang T, Han T, Tayebi Arasteh S, Haarbuerger C, et al.
- 475 Multimodal deep learning for integrating chest radiographs and clinical Parameters: a case for
- 476 transformers. *Radiology*. 2023;309: e230806. doi:10.1148/radiol.230806
- 477 21. McKinnon C, Nandhabalan M, Murray SA, Plaha P. Glioblastoma: Clinical presentation,
- 478 diagnosis, and management. *BMJ*. 2021;374. doi:10.1136/bmj.n1560
- 479 22. Demogeot N, Salleron J, Rech F, Taillandier L, Royer P, Vogin G. Impact of fractionated
- 480 stereotactic radiotherapy on activity of daily living and performance status in
- 481 progressive/recurrent glioblastoma: a retrospective study. *Radiat Oncol*. 2022;17: 1–9.
- 482 doi:10.1186/s13014-022-02169-1
- 483 23. Della Pepa GM, Caccavella VM, Menna G, Ius T, Auricchio AM, Chiesa S, et al. Machine
- 484 learning–based prediction of 6-month postoperative Karnofsky performance status in patients
- 485 with glioblastoma: capturing the real-life interaction of multiple clinical and oncologic factors.
- 486 *World Neurosurg*. 2021;149: e866–e876. doi:10.1016/j.wneu.2021.01.082
- 487 24. Seow H, Tanuseputro P, Barbera L, Earle CC, Guthrie DM, Isenberg SR, et al. Development
- 488 and validation of a prediction model of poor performance status and severe symptoms over time
- 489 in cancer patients (PROVIEW+). *Palliat Med*. 2021;35: 1713–1723.
- 490 doi:10.1177/02692163211019302
- 491 25. Ramakrishna R, Hsu WC, Mao J, Sedrakyan A. Surgeon annual and cumulative volumes
- 492 predict early postoperative outcomes after brain tumor resection. *World Neurosurg*. 2018;114:
- 493 e254–e266. doi:10.1016/j.wneu.2018.02.172
- 494 26. Marcus AP, Marcus HJ, Camp SJ, Nandi D, Kitchen N, Thorne L. Improved prediction of
- 495 surgical resectability in patients with glioblastoma using an artificial neural network. *Sci Rep*.
- 496 2020;10: 1–9. doi:10.1038/s41598-020-62160-2
- 497 27. Mukherjee D, Saha P, Kaplun D, Sinitca A, Sarkar R. Brain tumor image generation using an

- 498 aggregation of GAN models with style transfer. *Sci Rep.* 2022;12: 1–16. doi:10.1038/s41598-
499 022-12646-y
- 500 28. Pham N, Hill V, Rauschecker A, Lui Y, Niogi S, Fillipi CG, et al. Critical Appraisal of
501 Artificial Intelligence Enabled Imaging Tools Using the Levels of Evidence System. *Am J*
502 *Neuroradiol.* 2023;44: E21–E28. doi:10.3174/ajnr.A7850
- 503

504 **Supporting information**

505 **S1 Fig. Participants flow**

506 **S2 Fig. The architecture of the variational autoencoder**

507 **S3 Fig. The KPS prediction model development using neural network**

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

509

510

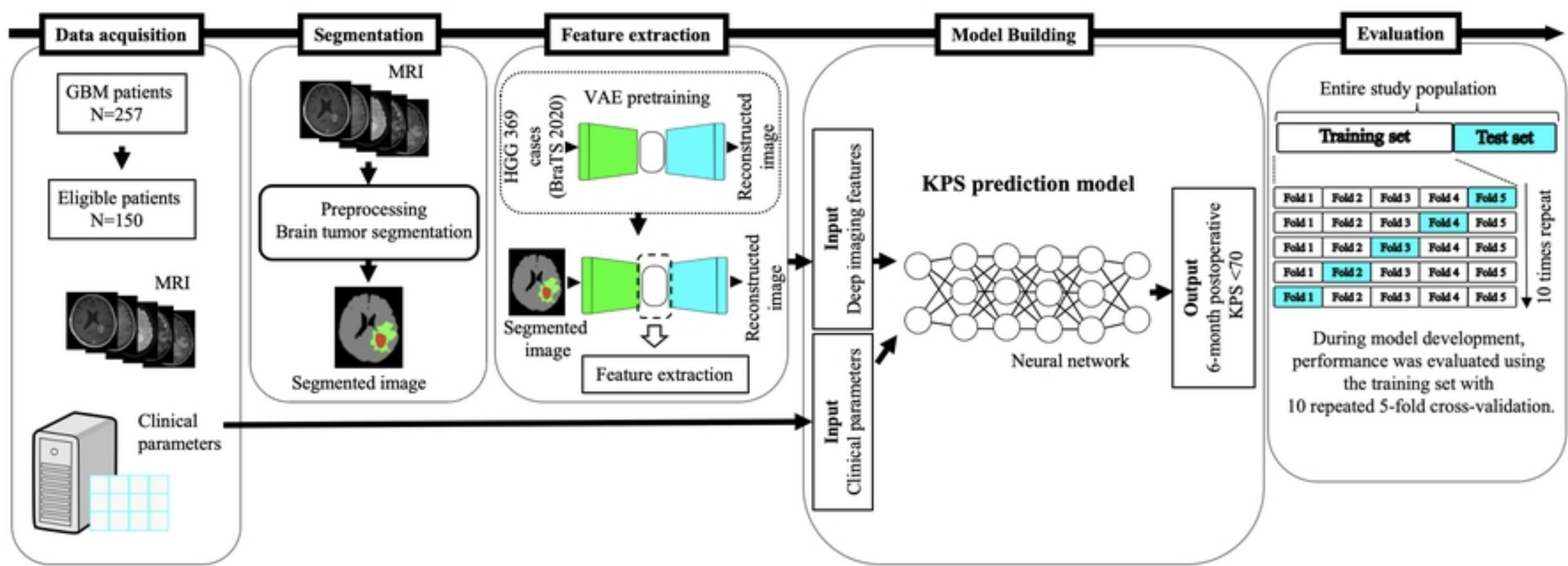


Fig1

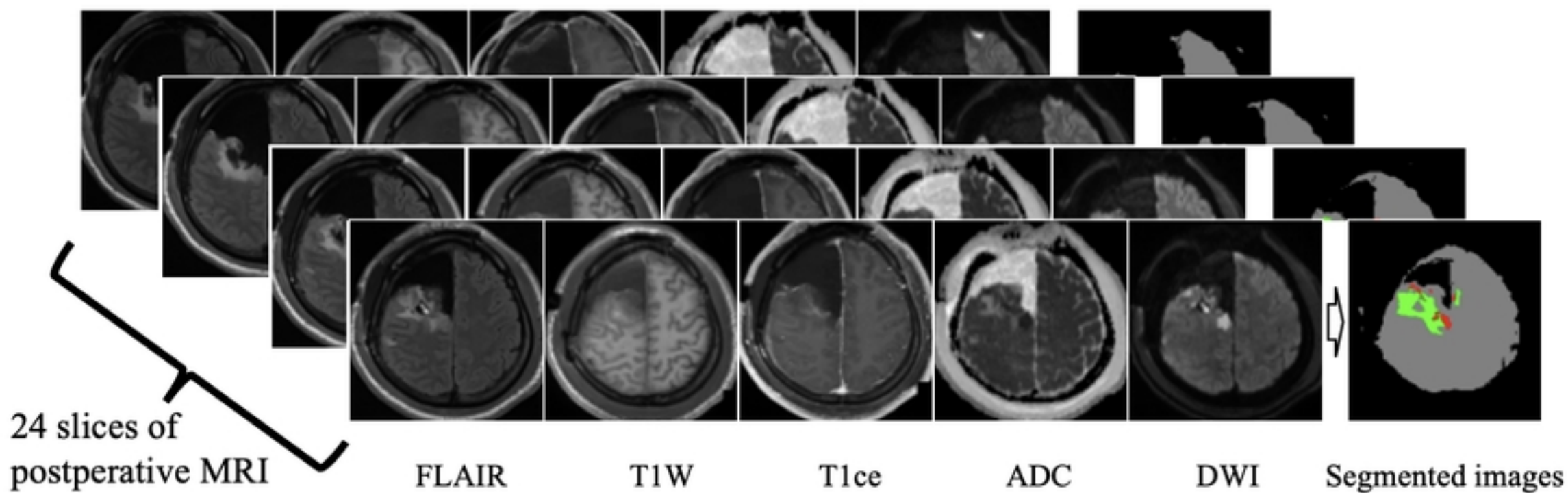
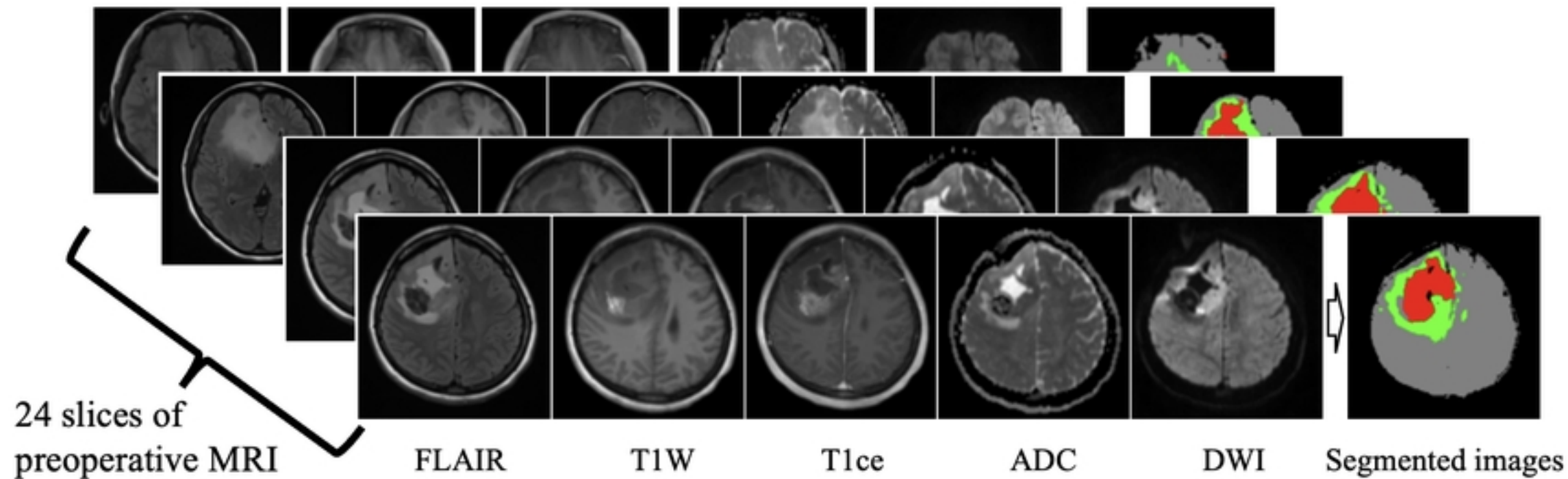


Fig2

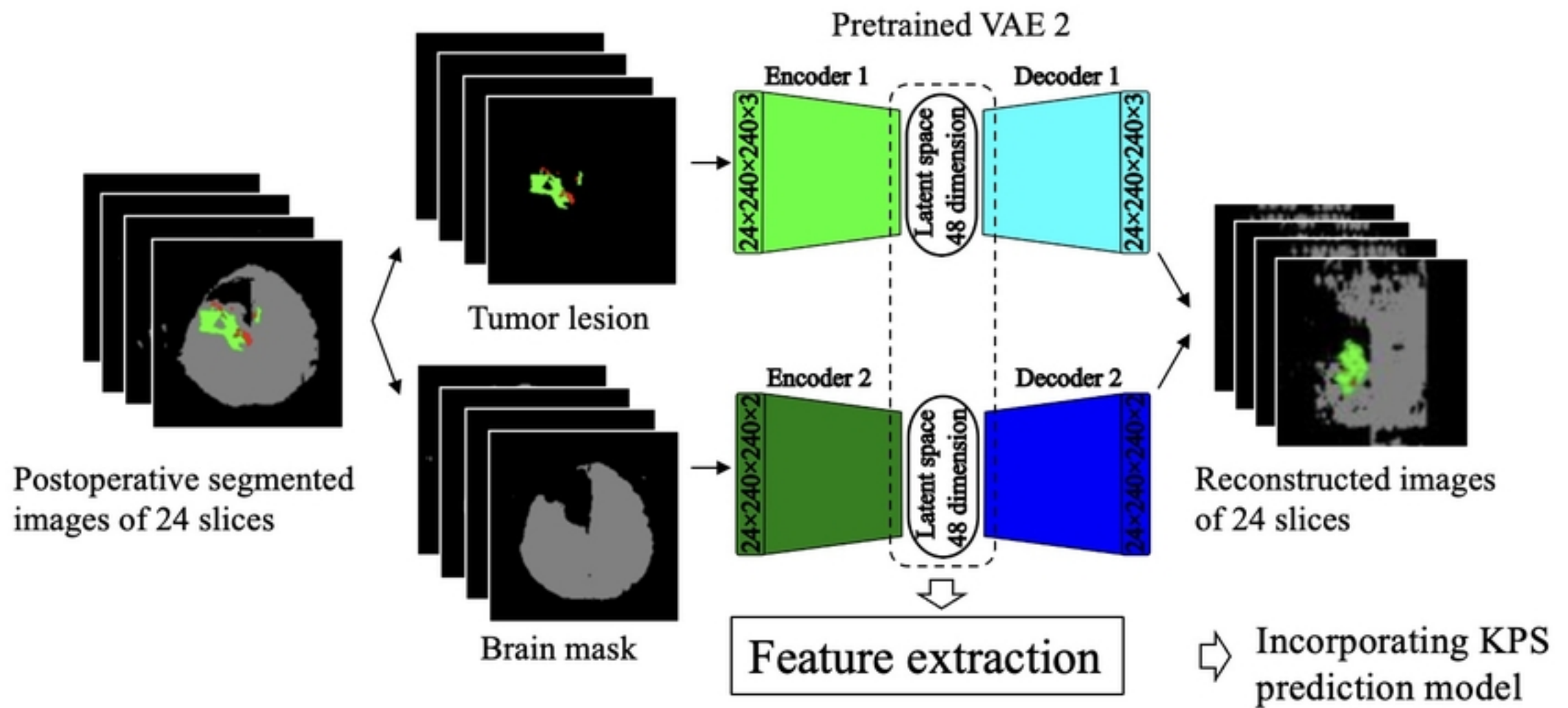
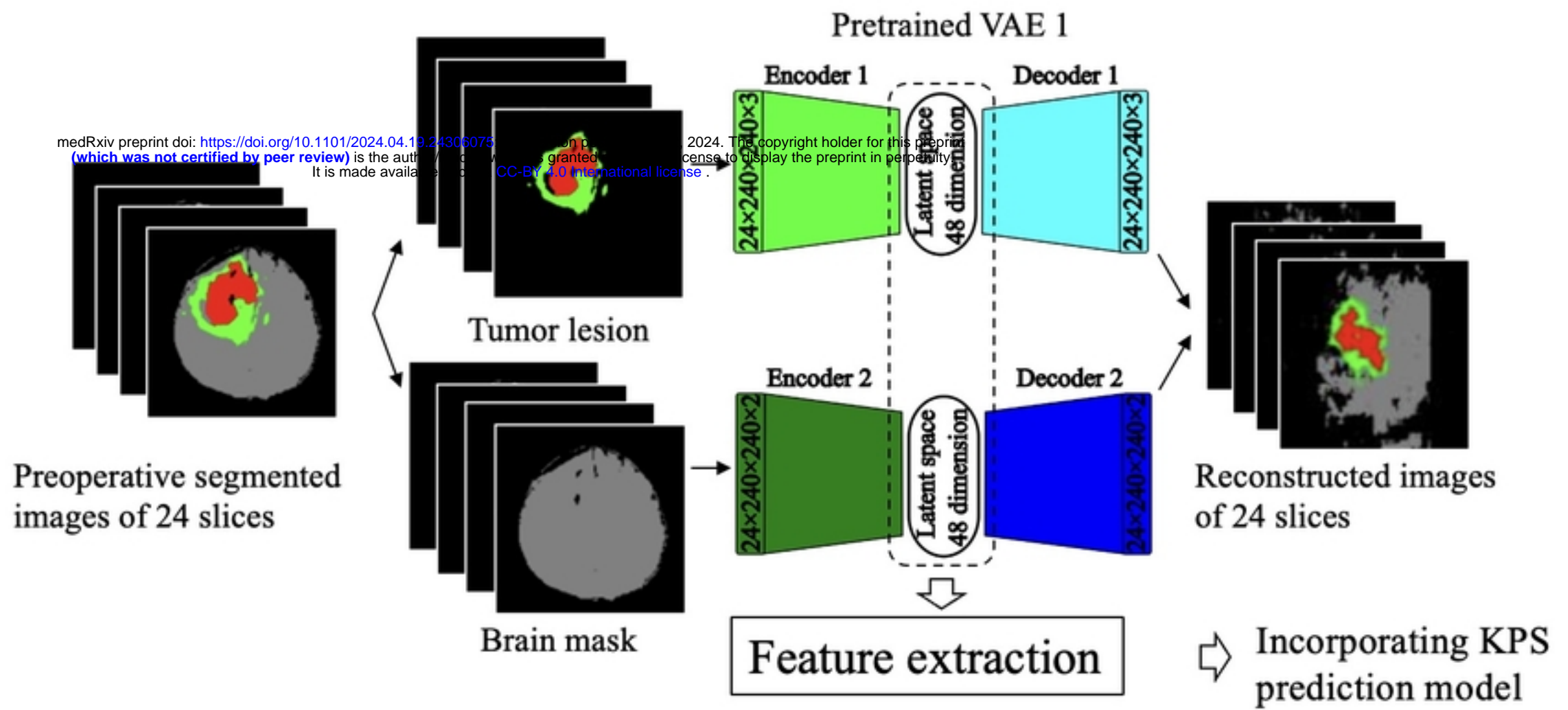
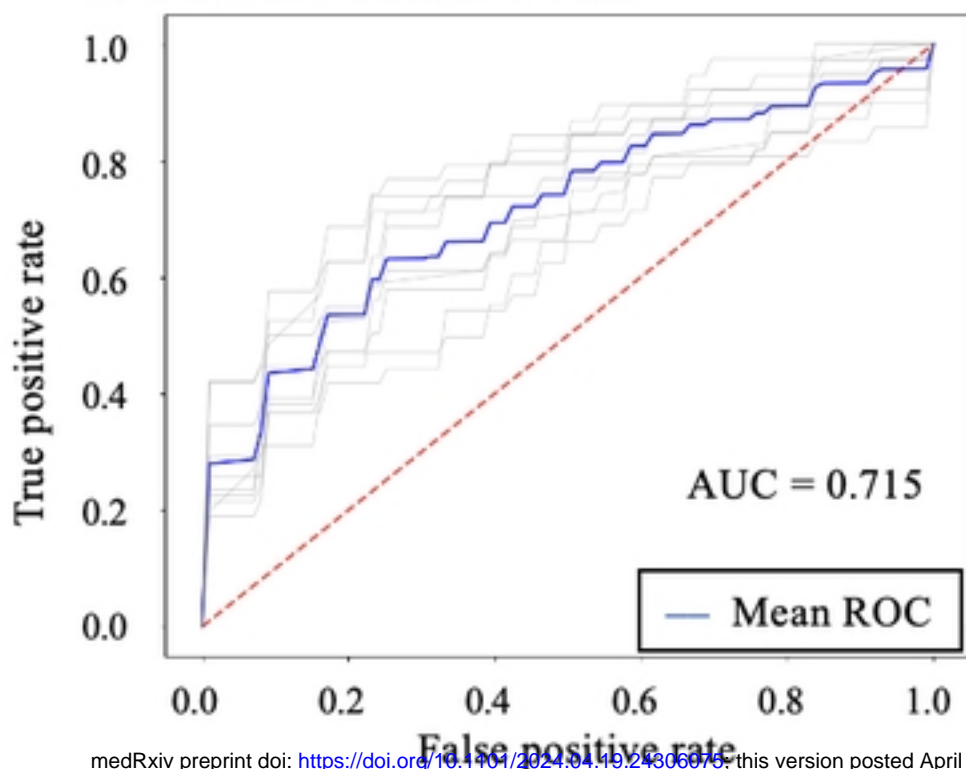
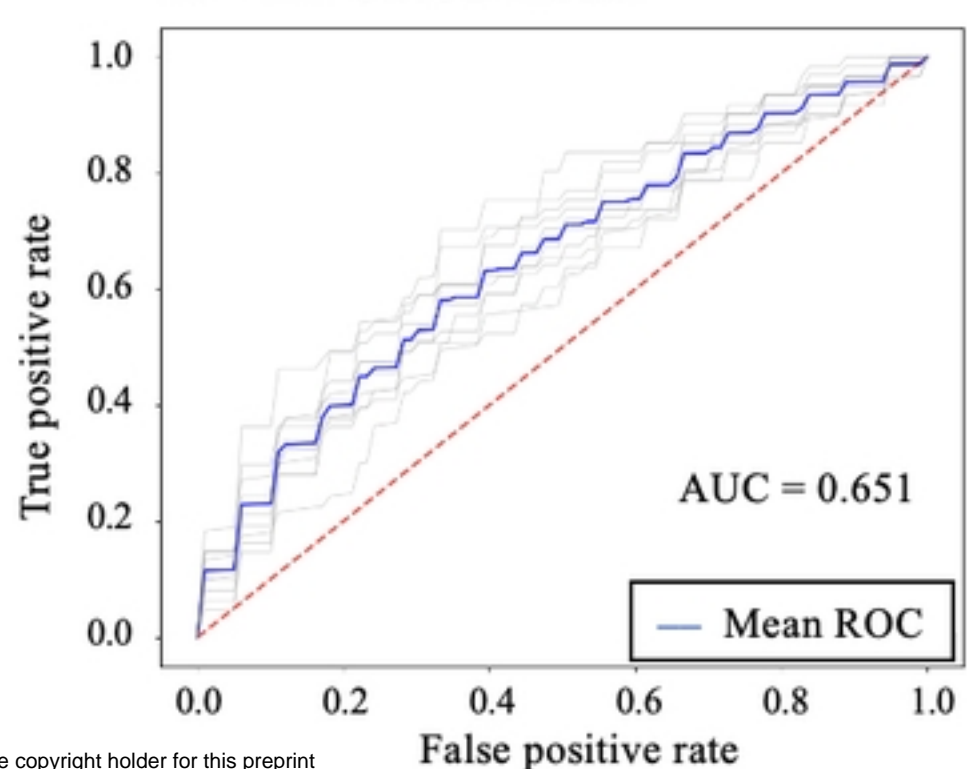


Fig3

A. Clinical-based model

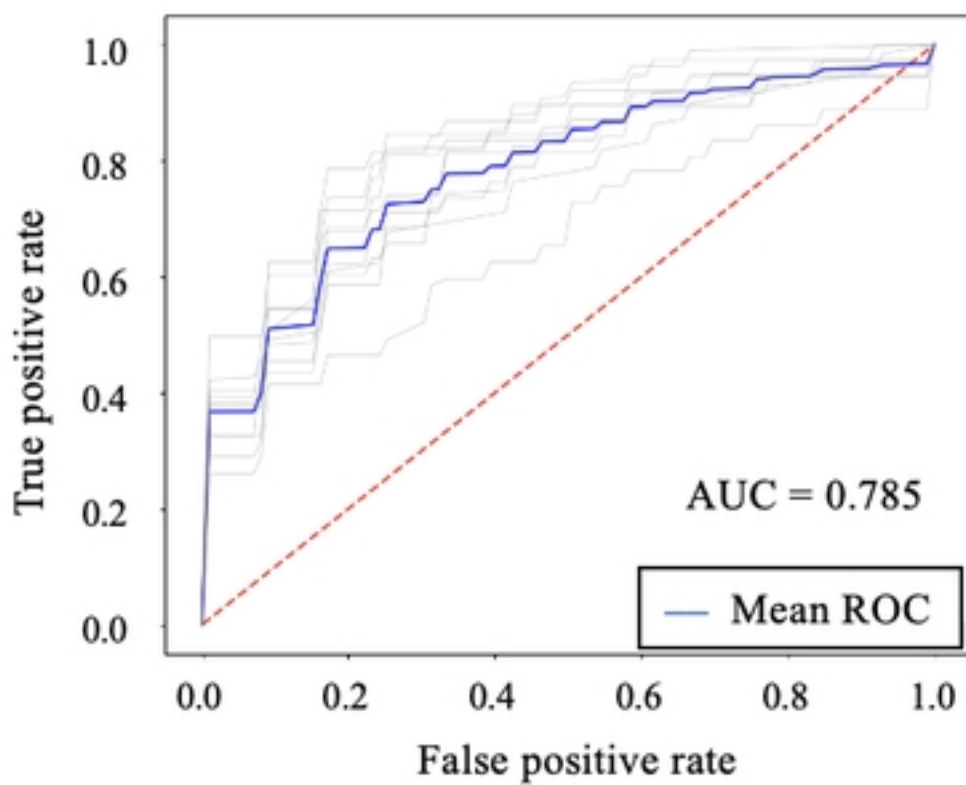


B. MRI-based model

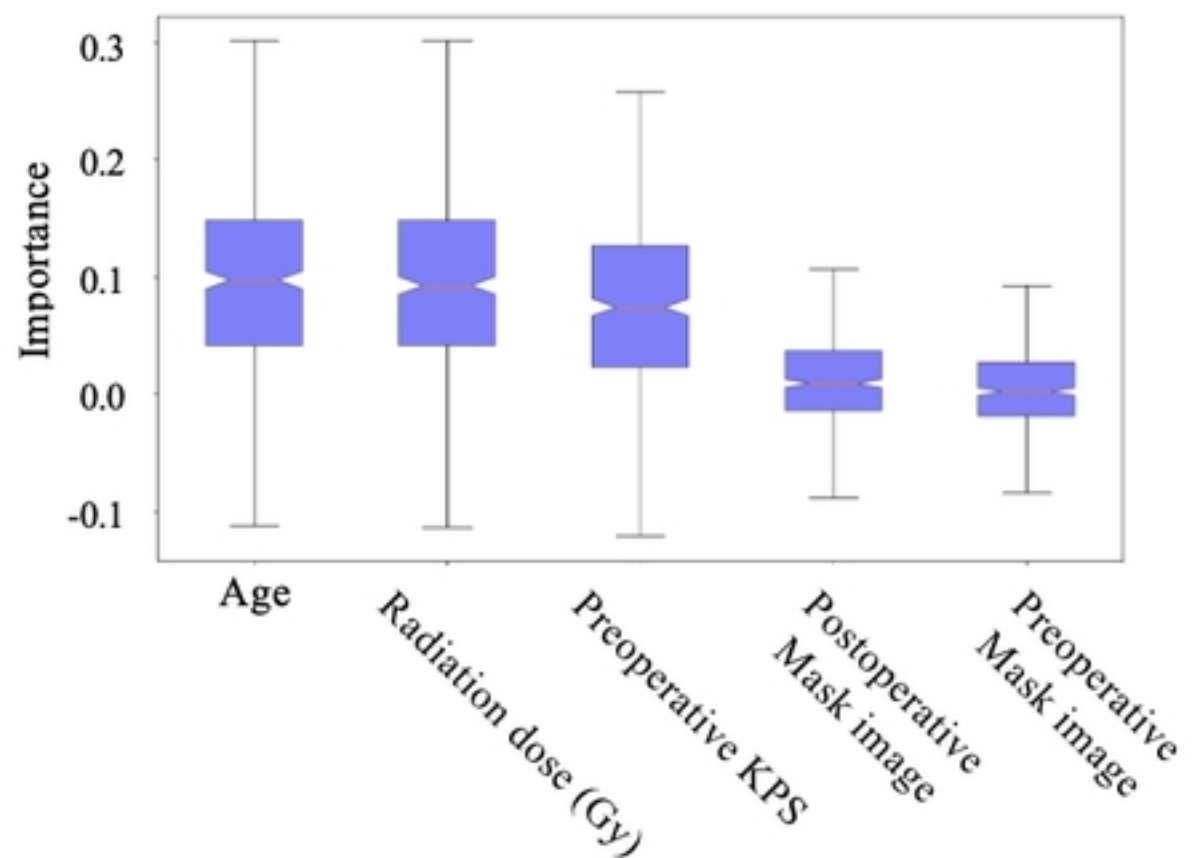


medRxiv preprint doi: <https://doi.org/10.1101/2024.04.19.24306676>; this version posted April 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

C. Multimodal model



D. Top 5 feature contributions in the multimodal model



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

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