




# Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme

Jan C. Peeken<sup>1,2,3</sup>  | Tatyana Goldberg<sup>4</sup> | Thomas Pyka<sup>5</sup> | Michael Bernhofer<sup>6</sup> |  
 Benedikt Wiestler<sup>7</sup> | Kerstin A. Kessel<sup>1,2,3</sup>  | Pouya D. Tafti<sup>4</sup> | Fridtjof Nüsslin<sup>1</sup> |  
 Andreas E. Braun<sup>4</sup> | Claus Zimmer<sup>7</sup> | Burkhard Rost<sup>5</sup> | Stephanie E. Combs<sup>1,2,3</sup> 

<sup>1</sup>Department of Radiation Oncology, Klinikum rechts der Isar der Technischen Universität München (TUM), München, Germany

<sup>2</sup>Deutsches Konsortium für Translationale Krebsforschung (DKTK), Partner Site Munich, Munich, Germany

<sup>3</sup>Department of Radiation Sciences (DRS), Institute of Innovative Radiotherapy (iRT), Helmholtz Zentrum München, Neuherberg, Germany

<sup>4</sup>Allianz SE, Munich, Germany

<sup>5</sup>Department of Nuclear Medicine, Klinikum rechts der Isar der Technischen Universität München (TUM), Munich, Germany

<sup>6</sup>Department for Bioinformatics and Computational Biology, Technical University of Munich (TUM), Garching, Germany

<sup>7</sup>Department of Neuroradiology, Klinikum rechts der Isar der Technischen Universität, Munich (TUM), München, Germany

## Correspondence

Jan C. Peeken, Klinik für RadioOnkologie und Strahlentherapie, Universitätsklinikum rechts der Isar, Technische Universität München (TUM), München, Germany.  
 Email: jan.peeken@tum.de

## Funding information

DKTK Partner Site Munich German Cancer Consortium

**Background:** For Glioblastoma (GBM), various prognostic nomograms have been proposed. This study aims to evaluate machine learning models to predict patients' overall survival (OS) and progression-free survival (PFS) on the basis of clinical, pathological, semantic MRI-based, and FET-PET/CT-derived information. Finally, the value of adding treatment features was evaluated.

**Methods:** One hundred and eighty-nine patients were retrospectively analyzed. We assessed clinical, pathological, and treatment information. The VASARI set of semantic imaging features was determined on MRIs. Metabolic information was retained from preoperative FET-PET/CT images. We generated multiple random survival forest prediction models on a patient training set and performed internal validation. Single feature class models were created including "clinical," "pathological," "MRI-based," and "FET-PET/CT-based" models, as well as combinations. Treatment features were combined with all other features.

**Results:** Of all single feature class models, the MRI-based model had the highest prediction performance on the validation set for OS (C-index: 0.61 [95% confidence interval: 0.51-0.72]) and PFS (C-index: 0.61 [0.50-0.72]). The combination of all features did increase performance above all single feature class models up to C-indices of 0.70 (0.59-0.84) and 0.68 (0.57-0.78) for OS and PFS, respectively. Adding treatment information further increased prognostic performance up to C-indices of 0.73 (0.62-0.84) and 0.71 (0.60-0.81) on the validation set for OS and PFS, respectively, allowing significant stratification of patient groups for OS.

**Conclusions:** MRI-based features were the most relevant feature class for prognostic assessment. Combining clinical, pathological, and imaging information increased predictive power for OS and PFS. A further increase was achieved by adding treatment features.

## KEYWORDS

biomarker, FET-PET, glioblastoma, machine learning, MRI, prognostic model, VASARI

## 1 | INTRODUCTION

Glioblastoma multiforme (GBM) constitutes the most frequent primary neuronal malignancy. Despite intensive efforts in research, the success of current therapy regimens remains limited with low two-year survival rates of 16.9%.<sup>1</sup> We know that clinical parameters such as younger age, high Karnofsky-performing status (KPS) indices, and female gender correlate with a favorable outcome.<sup>2-8</sup> Several approaches finding molecular determinants of outcome to predict treatment response have been published. However, few have reached clinical relevance.<sup>9-11</sup> MGMT-promoter methylation status and mutational status of IDH appear to carry prognostic value.<sup>11-14</sup>

MRI constitutes the standard imaging modality for pre-therapeutic staging, treatment planning and follow-up diagnostics. Over the years, multiple prognostic relevant semantic properties have been identified. Proposed features quantify the different composites of the tumor (eg, enhancement or edema), or classify characteristics such as multifocality or invasion of brain areas.<sup>3,4,15-17</sup> MRI-based quantification of the extent of resection emerged as prognostic factor.<sup>2,18</sup>

Based on such qualities, the VASARI (Visually Accessible REMBRANDT [Repository for Molecular Brain Neoplasia Data] Images) feature list was defined by the REMBRANDT consortium aiming to standardize the reporting of gliomas. Inter-observer agreement appeared to be high in all but three imaging features.<sup>19</sup> Multiple publications have shown that

VASARI features predict patient outcome, and correlate with mutational status and gene expression patterns.<sup>20-22</sup>

In recent years, metabolic PET images utilizing amino acid-based tracers gained clinical relevance. For instance, parameters obtained from static [18F]-fluoroethyl-L-tyrosine (FET) PET/CT were shown to inherit prognostic value predicting survival and progression independent of MGMT promoter methylation and clinical factors.<sup>23</sup> Moreover, FET uptake variables significantly correlated with WHO grading.<sup>24,25</sup>

Originally, prognostic models aiming at predicting patients' survival or progression were often based on statistical models. For example, a recently published model by Gittleman et al<sup>26</sup> predicted survival on the basis of age, gender, MGMT-promoter methylation status, and KPS. In recent years, machine learning (ML) approaches have gained significant importance as alternative way for model generation. Due to the improvements in handling large datasets with many input features, ML-based methods may lead the way to clinical decision support systems (CDSS) as the basis for personalized medicine.<sup>27</sup>

In the present study, we sought to determine the prognostic value of ML-based models on the basis of multiple input feature classes including clinical, pathological, semantic MRI features, and FET-PET/CT measurements. Moreover, we analyzed if "multimodal" models combining several feature classes and, specifically, the addition of treatment features further improve the prognostic performance of ML models.

**TABLE 1** Patient characteristics and outcome

	Training set (n = 132)	Validation set (n = 57)
Age (y)	m 58 (min 20, max 85)	m 59 (min 24, max 82)
KPS (%)	m 80 (min 40, max 100)	m 80 (min 40, max 100)
Gender	Male: 82 Female: 50	Male: 39 Female: 18
Positive MGMT-methylation status	33 (25%) (na: 38 [29%])	13 (23%) (na: 16 [32%])
Positive IDH1/2 mutation status	3 (3%) (na: 55 [42%])	1 (5%) (na: 21 [37%])
Ki-67 proliferation index	>20%: 32 (24%) <20%: 31 (24%) (na:69 [52%])	>20%: 16 (28%) <20%: 13 (23%) (na: 28 [49%])
OS (mo)	m 11.9 (min 0, max 75.5)	m 11.7 (min 0.3, max 87.4)
PFS (mo)	m 4.25 (min 0, max 60.8)	m 6.3 (min 0, max 74.0)
Preoperative MRI	124 (94%)	54 (95%)
Postoperative MRI	99 (75%)	46 (81%)

KPS, Karnofsky performance status; m, median; max, maximum; min, minimum; na, not available; OS, overall survival; PFS, progression-free survival.

	Training set (n = 132)	Validation set (n = 57)
Primary Surgery	112 (85%)	50 (88%)
Primary Radiation	20 (15%)	7 (12%)
RTCT	Yes: 103 (78%) No: 8 (6%) (na: 21 [16%])	Yes: 45 (79%) No: 3 (5%) (na: 9 [16%])
Adjuvant CT	Yes: 88 (67%) No: 13 (10%) (na: 31 [23%])	Yes: 41 (72%) No: 4 (7%) (na: 12 [21%])
Therapeutic Interval > 6 weeks <sup>a</sup>	Yes: 32 (24%) No: 79 (60%) (na: 21 [16%])	Yes: 37 (65%) No: 13 (23%) (na: 7 [12%])
PTV (mL)	m: 333.9 (min 29.9, max 701) (na: 4 [3])	m: 340.8 (min 57.9, max 862.4) (na: 1 [2%])
TD (Gy)	m 60 (min 10, max 64)	m 60 (min 18, max 60)
SD (Gy)	m 2.0 (min 1.8, max 5.0)	m 2.0 (min 1.7, max 3.0) (na: 2 [4%])

CT, chemotherapy; m, median; max, maximum; min, minimum; na, not available; PTV, planning target volume; RT, radiotherapy; RTCT, radiochemotherapy; SD, single dose; TD, total dose.

<sup>a</sup>Between surgery and RT.

**TABLE 2** Therapy characteristics

## 2 | MATERIAL AND METHODS

### 2.1 | Patients and study design

In total, 189 patients with GBM treated with radiation therapy (RT) from 2009 to 2016 in our institution were retrospectively analyzed. Patients received established first-line therapy after informed consent following the primary diagnosis. Patient records were assessed for gender, age, and KPS at start of RT (see Table 1). Pathological records were assessed for MGMT-promoter methylation status, IDH1 mutation status, and the Ki67 proliferation index. Immunohistochemistry using an antibody against the IDH1 p.R132H mutation was used to test for IDH1 mutations. The MIB1 antibody was used to test for the Ki67 proliferation index. The Ki67 proliferation index was split at the median of 20% dividing high and low proliferation groups. Promoter methylation was determined using real-time PCR-based methylation-quantification of endonuclease-resistant DNA (MethyQESD) method.<sup>28</sup> MGMT promoter methylation was defined by promoter methylation greater than 8% as described by Reifenberger et al.<sup>13</sup>

Primary tumor resection with sequential radio(chemo)therapy constituted the first line therapy, which was possible in 162 patients. 27 patients were primarily treated with RT after biopsy. 149 patients received concomitant radiochemotherapy following the protocol of Stupp et al<sup>11</sup> with radiotherapy (RT) up to a total dose of 60 Gy (single dose 2 Gy) and temozolomide (75 mg/m<sup>2</sup>) (see Table 2 for therapy characteristics). In addition to temozolomide, one patient received cilengitide and two patients received lomustin. One patient received bevacizumab and three patients were additionally treated with irinotecan.

Overall survival (OS) was determined from the end of RT to the time point of death or the time point of censoring (134

reported deaths). Progression-free survival (PFS) was calculated from the end of RT to the first sign of progression, death or time point of censoring (168 reported progressions or deaths), whichever happened first. Progress was defined retrospectively according to MRI- and PET-study reports and/or pathological reports. All clinical and molecular data were collected in the Munich Innovative Radiotherapy (MIRO) database. This study was approved by the ethical committee of the Technical University of Munich (reference number 466/16).

### 2.2 | Magnetic resonance imaging

Pre- and postoperative MRIs were assessed for availability of relevant sequences: T1-weight (T1w)- (or MPRAGE), T1w+Gd-, FLAIR-, T2w- and diffusion imaging with apparent diffusion coefficient (ADC)-maps (see Table S1 for image acquisition parameters). All 27 preoperative features and three postoperative features were determined following the recommendations of the REMBRANDT consortium by a MD with 2 years of experience in radiation oncology (see Table S2 for all VASARI features).<sup>19</sup> The ADC-maps were used to classify "facilitated," "restricted," and "mixed" diffusion. Before analysis, few features were altered to achieve a better patient representation in subgroups (eg, pooling of subgroups) or to retain a binary variable (see Table S3).

### 2.3 | [18F]-fluoroethyl-L-tyrosine (FET) PET studies and analysis

FET-PET/CT scans were performed pre-operatively using a Biograph 16 PET/CT in 68 patients (Siemens Medical Solutions USA, Inc., Malvern, PA, USA). Patients were required to fast for a minimum of 6 hours before undergoing

scanning to achieve standardized metabolic conditions. 190 MBq of FET were intravenously administered and a low-dose CT (24-26 mAs, 120 kV) for attenuation correction was conducted. 30-40 minutes after initial injection, the 10-minute PET acquisitions were performed. Static PET data were reconstructed by two different methods. PET/CTs of 49 patients were reconstructed using filtered back-projection employing a Hann filter with a cut-off frequency of 0.34 Nyquist into  $128 \times 128$  matrices, with a resulting voxel size of 2.1-2.1 mm and slice thickness of 2.4 mm. In 19 patients, an ordered subset expectation maximization algorithm was used ( $200 \times 200$  matrix, 3 iterations, 21 subsets, resulting in a voxel size of 1.6 to 1.6 mm).

Semiautomatic analysis of static FET-PET was performed by an experienced nuclear medicine physician (PT), blinded to histology and clinical outcome, using Matlab (MathWorks, Inc., Natick, MA, USA; Image Processing Toolbox and own code). The procedure was performed as described earlier.<sup>25</sup> Briefly, the images were normalized against background uptake, defined with a region of interest placed into the hemisphere opposite to the tumor as proposed by the German guidelines for brain tumor imaging.<sup>29</sup> Tumor segmentation was conducted by placing seed-points inside the tumor followed by automated region-growing, which was limited, by a margin of 1.3 times the background activity. If necessary, blocking lines were placed manually to prevent the algorithm from growing into surrounding anatomic structures with increased PET signal. Maximum tumor to brain ratio (TBR), mean TBR, metabolic tumor volume (MTV), and the product of mean TBR and MTV, which was defined as total lesion normalized uptake (TLU), were calculated on floating-point data (see data in Table S4).

## 2.4 | Building of ML models

Machine learning modeling and statistical analyses were performed in R (version 3.4.0) (R core team, Vienna, Austria). For all testing purposes, the data set for each of the seven

prediction models was randomly split into one development subset containing 2/3 of all patients ( $n = 132$ ) and one independent test subset containing 1/3 of all patients ( $n = 57$ ). No data from the independent test subset were used for the development of Models 1-7. As ML technique, the random forest algorithm, implemented as an ensemble of decision trees constructed from randomly selected features and training data points, was chosen due to its short training periods, the capability of managing incomplete and noisy data, good interpretability, and high predictive power.<sup>30,31</sup> To predict right censored survival outcomes random survival forest (RSF) models were developed using the *randomForestSRC* package (R core team).<sup>32</sup> The VIMP function was applied to calculate feature permutation importance. Altogether seven prediction models on the basis of different feature classes were trained (see Table 3 for selected features): Firstly, four models were generated on single feature classes including "clinical" (model 1 (M1)), "pathological" (M2), "MRI-based" (M3), and "FET-PET/CT-based" (M4). Secondly, the benefit of combining clinical and pathological features (M5) and all four feature classes combined (M6) was tested. Finally, treatment features were added to all four feature classes (M7).

## 2.5 | Performance evaluation

The performance was assessed on the independent patient test set. The concordance index (C-index) served as performance estimator. Direct comparison of models was performed using the *rcorr.cens* function of the *Hmisc* package. For dichotomization of patient subgroups, the maximally selected rank statistics method was applied to the training set to determine the optimal cut-off value using the *maxstat* package.<sup>33</sup> The same cut-off value was then used to define high-risk and low-risk patients in the validation set. Log-rank tests were conducted to test for statistical significance between patient risk groups. The area under the receiver operator characteristic curve (AUC) was calculated using the *survivalroc* package.

**TABLE 3** Machine learning models with multimodal feature classes

Model and feature class	Features
M1: "Clinical"	Age, KPS, Gender
M2: "Pathological"	MGMT-promoter-methylation, IDH-mutational status, Ki 67%-PI
M3: "MRI-based"	VASARI features
M4: "FET PET/CT-based"	TBRmax, TBRmean, MTV, TLU
M5: "Clinical/Pathological"	M1 + M2
M6: "Clinical/Pathological/Imaging"	M1 + M2 + M3 + M4
M7: "Clinical/Pathological/Imaging" + treatment features	M1 + M2 + M3 + M4 + RTCT, surgery, PTV, TD, SD, adjuvant CT, therapeutic Interval <sup>a</sup>

CT: chemotherapy; KPS: Karnofsky performance status; MTV: mean tumor volume; PI: proliferation index; PTV: planning target volume; RT: radiotherapy; RTCT: radiochemotherapy; SD: single dose; TBR: tumor to brain ratio; TD: total dose; TLU: total lesion uptake.

<sup>a</sup>Between surgery and RT.

### 3 | RESULTS

#### 3.1 | Prediction of patients' OS

First, the predictive value of ML models based on single variable classes was tested for OS (see Table 4 for C-index values and confidence intervals of all models). The clinical (M1) and pathological (M2) models showed low predictive capability in both development (C-index: M1 0.74 [95% confidence interval: 0.67-0.81], M2: 0.64 [0.58-0.71]) and independent test set with no significant difference from random in the latter (C-index: M1 0.59 [0.48-0.70], M2: 0.49 [0.37-0.60]). With a C-index of 0.93 (0.87-1.00), the MRI-based model (M3) had the highest prediction performance on the development set, which could be validated, with a reduced C-index of 0.61 (0.51-0.72) significantly different from random. The FET-PET/CT-based model (M4) showed high prediction capacity on the development set (C-index: 0.93 [0.86-1.00]), which was, however, not reproducible on the validation set (C-index: 0.54 (0.44-0.65)).

Combining clinical and pathological variables (M5) did confer a non-significant improvement (test-set, p-value: 0.18) above the performance of the clinical model (C-index: development: 0.75 [0.68-0.82], testing: 0.64 [0.53-0.75]). The combination of all feature classes (M6) did trigger a further performance increase (C-index: development: 0.94 [0.88-1.00], testing: 0.70 [0.59-0.81]) significantly better than all single feature class models alone (test-set: M1  $P = 0.018$ , M2  $P = 0.003$ , M3  $P = 0.044$ , M4  $P = 0.003$ ).

In the final model, treatment features describing the delivered therapeutic regimens were added to the model M7. This led to a further rise above the best performing combined model (M6) with a C-index of 0.96 (0.89-1.00) in the development set and 0.73 (0.62-0.84) in the independent test set without reaching significance in direct comparison ( $P = 0.34$ ).

The recently proposed nomogram from the work of Gittleman et al<sup>26</sup> was tested on the independent test set using the 12-month survival probabilities. It achieved a predictive performance with an AUC of 0.64 in comparison with AUCs of 0.75 and 0.80 for models M6 and M7.

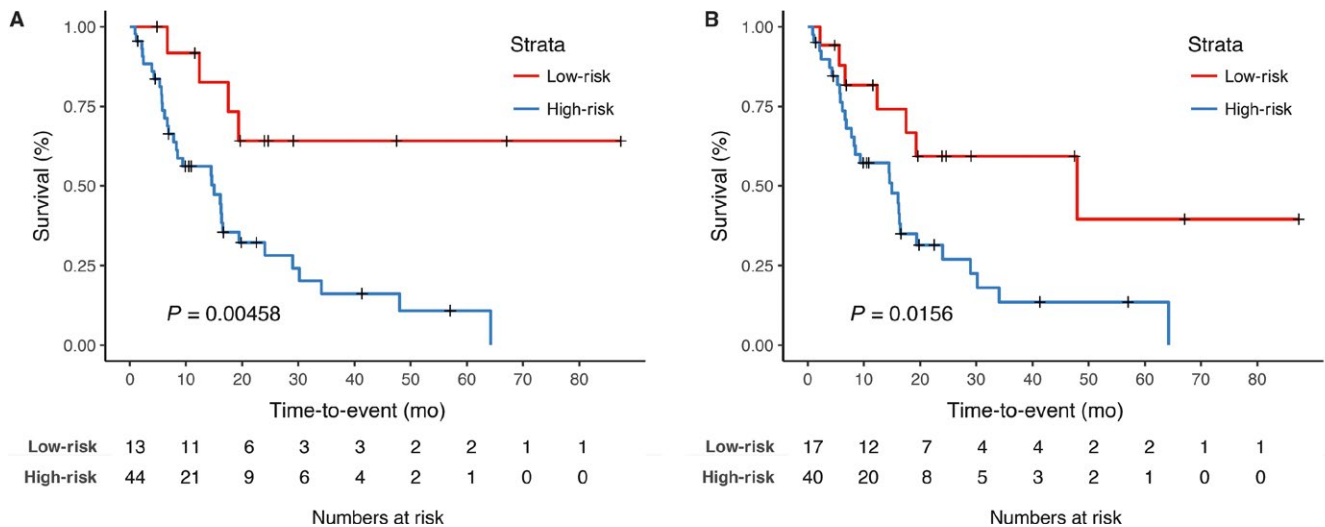
Next, two risk groups were defined by the predictors of the models M6 and M7 using cut-off points optimized on the training set. Kaplan-Meier survival curves of the validation cohort are shown in Figure 1. Both models significantly discerned high-risk from low-risk patients (M6  $P = 0.0048$ , M7  $P = 0.0156$ ).

#### 3.2 | Prediction of patients' progression-free survival

The same model categories were developed to predict PFS (see Table 4 for C-index values and confidence intervals).

**TABLE 4** Performance estimates as concordance index (C-index) and 95% confidence intervals of all seven models predicting OS and PFS

	M1	M2	M3	M4	M5	M6	M7	
OS	Training set	0.74 (0.67-0.81)	0.64 (0.58-0.71)	0.93 (0.87-1.00)	0.93 (0.86-1.00)	0.75 (0.68-0.82)	0.94 (0.88-1.00)	0.96 (0.89-1.00)
	Test set	0.59 (0.48-0.70)	0.49 (0.37-0.60)	0.61 (0.51-0.72)	0.54 (0.44-0.65)	0.64 (0.53-0.75)	0.70 (0.59-0.81)	0.73 (0.62-0.84)
PFS	Training set	0.68 (0.62-0.75)	0.63 (0.56-0.69)	0.73 (0.66-0.79)	0.80 (0.73-0.87)	0.70 (0.63-0.77)	0.81 (0.74-0.88)	0.79 (0.72-0.85)
	Test set	0.56 (0.45-0.66)	0.50 (0.40-0.61)	0.61 (0.50-0.71)	0.45 (0.35-0.56)	0.61 (0.51-0.71)	0.68 (0.57-0.78)	0.71 (0.60-0.81)



**FIGURE 1** Kaplan-Meier curves for overall survival showing the performance model 6 (M6) and model 7 (M7) in the internal validation cohort. The developed classifiers for overall survival M6 and M7 were used to assign patients to a “high-risk” and “low-risk” group in the validation patient cohort. The log-rank test was applied to test for significant separation of survival curves and calculation of  $P$ -values. Model 6 did divide significantly patient subgroups on the validation test set ( $P = 0.00458$ ). Model 7 significantly divided high-risk from low-risk patients ( $P = 0.0156$ )

Compared to the survival models, predictive performance was lower for the single feature class models clinical M1, pathological M2, MRI-feature-based M3, and FET-PET/CT-based M4 in the development set (C-indices of 0.68, 0.63, 0.73, and 0.80, respectively). On the independent test set, prediction performance achieved even further reduced results (C-indices 0.56, 0.50, 0.60, and 0.45 for M1, M2, M3, and M4, respectively).

The multimodal models predicted comparably to the best single feature class model M3. The combined clinical and pathological model 5 achieved a C-index of 0.70 (0.63-0.77) in the development and 0.61 (0.51-0.71) in the independent test sets, which was non-significantly compared to the clinical model M1 (test-set,  $P = 0.096$ ). The combined model M6 achieved C-indices of 0.81 (0.74-0.88) on the development set and 0.68 (0.57-0.78) on the independent test set significantly outperforming the single feature class models M1 and M3 (M1  $P = 0.016$ , M2  $P = 0.196$ , M3  $P = 0.014$ , M4  $P = 0.233$ ).

Finally, adding therapeutic information to M7 increased prognostic performance further up to a C-index of 0.79 (0.72-0.84) on the development set and 0.71 (0.60-0.81) on the independent test set with a significantly better prediction than all other models (M1  $P = 0.0013$ , M2  $P = 0.023$ , M3  $P = 0.00057$ , M4  $P = 0.0001$ , M5  $P = 0.002$ , M6  $P < 0.0001$ ). The Gittleman nomogram, which was generated to predict survival at 12 months, showed worse prognostic capacity with an AUC of 0.67 in comparison with AUCs of 0.82 and 0.83 for models M6 and M7. Kaplan-Meier curves plotting PFS for patient subgroups separated by the prediction models did not show significant separations for M6 and

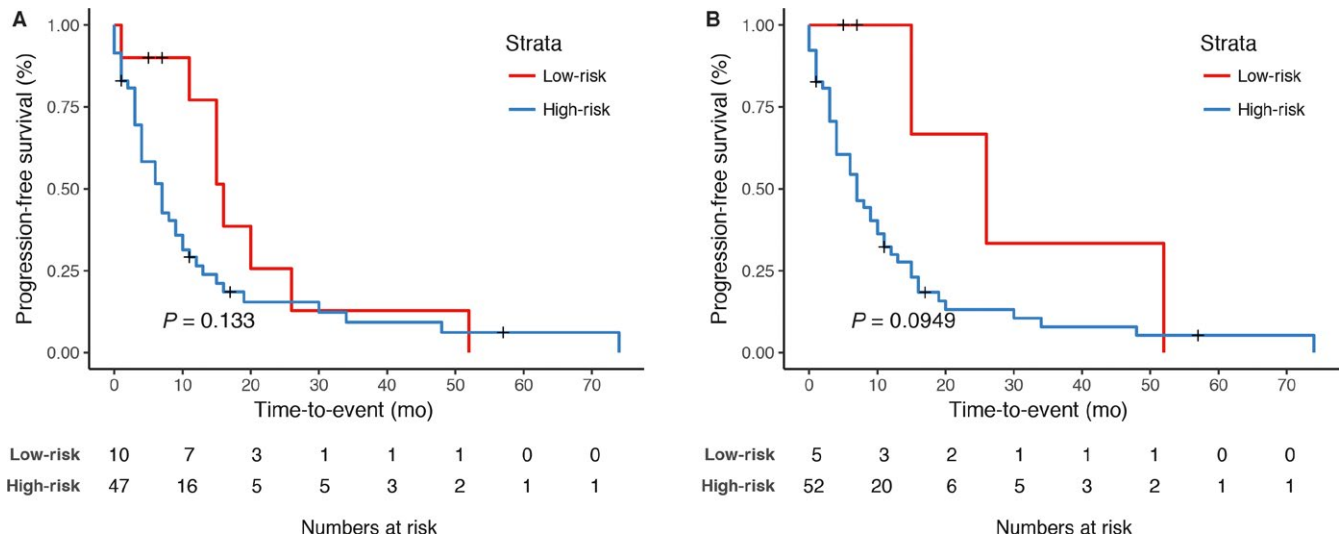
M7 (see Figure 2). For M7, however, there was a trend toward significance ( $P = 0.095$ ).

### 3.3 | MRI-based and therapy-related feature dominate OS prediction performance

In order to evaluate the importance of single features for the performance of the combined models M6 and M7, the permutation feature importance was assessed (see Table S5). Consistent with the observation of model M3 as best single feature class model, MRI-based features were the most important features besides patients' age. In contrast, model M7 was dominated by treatment features, such as single and total radiation dose, PTV volume and surgery. Besides, the feature classes MRI-based, clinical, and pathological were among the 10 best performing features. The most important MRI-based features in both models included "Satellites" and "Thickness of CE margin."

In PFS prediction models, M6 and M7 MRI-based features appeared to be the most important feature class beside the known prognostic factors "age," "KPS," "gender," and "MGMT" status (see Table S6). Model M7 showed a similar feature importance distribution as for OS with treatment features providing the most important features. The MRI-based features "deep white matter invasion," "ependymal invasion," "proportion of resection of enhancing tumor," and of "edema" were among the 10 most important features in both models.

Consistent with the low performance of model M4, FET-PET/CT-based features did not show a high permutation importance for prediction of OS and PFS.



**FIGURE 2** Kaplan-Meier curves for progression-free survival showing the performance of model 6 (M6) and model 7 (M7) in the internal validation cohort. The developed classifiers M6 and M7 for progression-free survival were used to assign patients to a “high-risk” and “low-risk” group in the validation patient cohort. The log-rank test was applied to test for significant separation of survival curves and calculation of  $P$ -values. No significant separation of PFS curves could be observed for model 6 ( $P = 0.133$ ). For model 7, there was a separation of survival curve without reaching statistical significance ( $P = 0.0949$ )

## 4 | DISCUSSION

In this work, we have demonstrated the potential of ML-based predictive models for the prognostic classification of GBM patients. Exploiting the capabilities of RF models to deal with large feature numbers and missing data, we assessed the value of certain feature classes and their combinatory effect. Moreover, we evaluated the benefit of integrating treatment features.

OS prediction was dominated by MRI-based features having the highest single model performance (M3) and showing high feature representation in the combined models M6 and M7. The feature classes in clinical and pathological models showed lower predictive performance. However, clinical and pathological were consistently selected for the combined models M6 and M7. For both prediction tasks, combining all pre-therapeutic feature classes inside model M6 did increase prognostic performance above the best single feature class model.

Finally, adding therapeutic information led to a further increase in prognostic performance for OS and PFS with a higher predictive performance compared to the formerly proposed nomogram by Gittleman et al.<sup>26</sup> Kaplan-Meier survival analysis showed a significant separation of high-risk from low-risk patients for the OS model.

Multiple previous studies have analyzed the prognostic potential of semantic MRI-based features alone or in conjunction with clinical or pathological features. A simple model based only on three semantic imaging features

"volume," "T1/FLAIR-ratio," and "hemorrhage" achieved a 12-month AUC of 0.67 for survival inferior to the combined models M6 and M7 (AUCs of 0.75 and 0.80, respectively) and similar to the clinical nomogram by Gittleman et al.<sup>22</sup> Two further studies could demonstrate an incremental benefit by combining clinical features with VASARI features yielding a C-index of 0.69, respectively.<sup>20,34</sup> In our study, model M6 combining MRI-based features with clinical, pathological and FET PET/CT-based features showed a similar performance of 0.69. This may indicate that clinical and semantic imaging may be sufficient for pre-therapeutic prognostic assessment.

In recent years, quantitative computational imaging features ("radiomics") have been shown to add prognostic value above clinical and molecular factors in GBM patients.<sup>35,36</sup> Interestingly, a combined clinical and radiomic model achieved similar prognostic performances with C-indices of 0.70 and 0.65 for OS and PFS, respectively. A radiomic model would have the great advantage of being less operator-dependent. However, multiple technical hurdles including the dependency of image acquisition parameters, equipment, preprocessing, and feature extraction need to be solved before safe clinical applications. Until then, semantic imaging feature may constitute a valuable alternative that is less dependent on technical variances.

It should be noted that this study was performed on the basis of a retrospective patient cohort. However, treatment regimens were overall rather homogenous with primary therapy following the recommendation of Stupp et al.<sup>11</sup> in 74% of patients. Prognostic performance was tested on an internal

validation cohort lowering available patients for model generation. In contrast to the above-mentioned benefits of ML-based models, model generation requires relatively large training sets. Limited patient numbers thus foster instability of model performances, which may further be increased by missing data.

In this study, we analyzed the predictive value of FET-PET/CT-based features in model M4. M4 showed high prognostic performances for both prediction tasks in the training set, which could not be reproduced in the independent test set. In contrast, previous studies have shown prognostic potential for OS and PFS.<sup>23-25</sup> There are two reasons that might explain an underestimation of the FET-PET/CT prognostic effect. First, the patient number with available PET data was relatively low compared to the total patient number. Second, two distinct reconstruction methods were used that may have led to inconsistencies in PET measures. A prospective study should be performed to evaluate the effect of FET-PET/CT features. In the future, prognostic performance might be enhanced by including texture features or dynamic FET-PET/CT measures.<sup>25,37</sup>

Current prognostic models are often based on clinical information. In recent years, a large number of novel prognostic imaging and molecular-based biomarkers have been identified.<sup>38-41</sup> Incorporating treatment features into a CDSS may increase the prognostic efficacy by quantifying the effect of partially given or omitted therapies.<sup>27</sup>

In summary, we demonstrated the applicability of ML models for the prediction of patients' OS and PFS. Semantic MRI-based features for OS and PFS showed relevant prognostic value. The inclusion of treatment data further increased predictive performance and may help to optimize follow-up procedures or 2nd line therapy regimens as CDSS.

## ACKNOWLEDGMENTS

The work was funded in part by the German Cancer Consortium (DKTK).

## CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## ORCID

Jan C. Peeken  <https://orcid.org/0000-0003-2679-9853>

Kerstin A. Kessel  <https://orcid.org/0000-0002-6544-8151>

Stephanie E. Combs  <https://orcid.org/0000-0002-5233-1536>

## REFERENCES

- Ostrom QT, Gittleman H, Xu J, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2009–2013. *Neuro Oncol.* 2016;18(Suppl. 5):v1-v75.
- Lacroix M, Abi-Said D, Fourney DR, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *J Neurosurg.* 2001;95(2):190-198.
- Zinn PO, Sathyan P, Mahajan B, et al. A novel volume-age-KPS (VAK) glioblastoma classification identifies a prognostic cognate microRNA-gene signature. *PLoS ONE.* 2012;7(8):1-9.
- Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, Cloughesy TF. MR imaging correlates of survival in patients with high-grade gliomas. *AJNR Am J Neuroradiol.* 2005;26(10):2466-2474.
- Curran WJ, Scott CB, Horton J, et al. Recursive partitioning analysis of prognostic factors in three Radiation Therapy Oncology Group malignant glioma trials. *J Natl Cancer Inst.* 1993;85(9):704-710.
- Filippini G, Falcone C, Boiardi A, et al. Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma. *Neuro Oncol.* 2008;10(1):79-87.
- Burth S, Kickingereder P, Eidel O, et al. Clinical parameters outweigh diffusion- and perfusion-derived MRI parameters in predicting survival in newly diagnosed glioblastoma. *Neuro Oncol.* 2016;18(12):1673-1679.
- Combs SE, Niyazi M, Adeberg S, et al. Re-irradiation of recurrent gliomas: pooled analysis and validation of an established prognostic score-report of the Radiation Oncology Group (ROG) of the German Cancer Consortium (DKTK). *Cancer Med.* 2018;7(5):1742-1749.
- Combs SE, Edler L, Rausch R, Welzel T, Wick W, Debus J. Generation and validation of a prognostic score to predict outcome after re-irradiation of recurrent glioma. *Acta Oncol (Madr).* 2013;52(1):147-152.
- Kessel KA, Hesse J, Straube C, et al. Validation of an established prognostic score after re-irradiation of recurrent glioma. *Acta Oncol (Madr).* 2017;56(3):1-5.
- Stupp R, Hegi ME, Mason WP, 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(5):459-466.
- Combs SE, Rieken S, Wick W, et al. Prognostic significance of IDH-1 and MGMT in patients with glioblastoma: one step forward, and one step back? *Radiat Oncol.* 2011;6(1):115.
- Reifenberger G, Hentschel B, Felsberg J, et al. Predictive impact of MGMT promoter methylation in glioblastoma of the elderly. *Int J Cancer.* 2012;131(6):1342-1350.
- Weller M, Felsberg J, Hartmann C, et al. Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network. *J Clin Oncol.* 2009;27(34):5743-5750.
- Park JK, Hodges T, Arko L, et al. Scale to predict survival after surgery for recurrent glioblastoma multiforme. *J Clin Oncol.* 2010;28(24):3838-3843.
- Hammoud MA, Sawaya R, Shi W, Thall PF, Leeds NE. Prognostic significance of preoperative MRI scans in glioblastoma multiforme. *J Neurooncol.* 1996;27(1):65-73.



17. Diehn M, Nardini C, Wang DS, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A*. 2008;105(13):5213-5218.
18. Stummer W, Reulen HJ, Meinel T, et al. Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias. *Neurosurgery*. 2008;62(3):564-576.
19. The Cancer imaging Archive. VASARI Research Project. 2014. Available from: <https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project>.
20. Mazurowski MA, Desjardins A, Malof JM. Imaging descriptors improve the predictive power of survival models for glioblastoma patients. *Neuro Oncol*. 2013;15(10):1389-1394.
21. Gutman DA, Cooper L, Hwang SN, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology*. 2013;267(2):560-569.
22. Rao A, Rao G, Gutman DA, et al. A combinatorial radiographic phenotype may stratify patient survival and be associated with invasion and proliferation characteristics in glioblastoma. *J Neurosurg*. 2016;124(4):1008-1017.
23. Suchorska B, Jansen NL, Linn J, et al. Biological tumor volume in 18F-FET-PET before radiochemotherapy correlates with survival in GBM. *Neurology*. 2015;84(7):710-719.
24. Gempt J, Bette S, Ryang Y-M, et al. 18F-fluoro-ethyl-tyrosine positron emission tomography for grading and estimation of prognosis in patients with intracranial gliomas. *Eur J Radiol*. 2015;84(5):955-962.
25. Pyka T, Gempt J, Hiob D, et al. Textural analysis of pre-therapeutic [18F]-FET-PET and its correlation with tumor grade and patient survival in high-grade gliomas. *Eur J Nucl Med Mol Imaging*. 2016;43(1):133-141.
26. Gittleman H, Lim D, Kattan MW, et al. An independently validated nomogram for individualized estimation of survival among patients with newly diagnosed glioblastoma: NRG Oncology RTOG 0525 and 0825. *Neuro Oncol*. 2017;19(5):669-677.
27. Lambin P, van Stiphout R, Starmans M, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol*. 2013;10(1):27-40.
28. Reifenberger G, Malzkorn B, Acker T, et al. Results of the international interlaboratory comparison of mgmt promoter methylation analysis involving twenty-three academic centers in germany, austria and the netherlands. *Neuro Oncol*. 2014;16(Suppl. 3):iii49-iii50.
29. Langen K-J, Bartenstein P, Boecker H, et al. German guidelines for brain tumour imaging by PET and SPECT using labelled amino acids. *Nuklearmedizin*. 2011;50(4):167-173.
30. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
31. Liu M, Xu X, Tao Y, An Wang X. Improved random forest method based on RELIEFF for Medical Diagnosis. BT - 2017 IEEE International Conference on Computational Science and Engineering, CSE 2017, and IEEE International Conference on Embedded and Ubiquitous Computing, EUC 2017, Guangzh. 2017. p. 44–9. Available from: <http://doi.ieeecomputersociety.org/10.1109/CSE-EUC.2017.19>.
32. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. *Biostatistics*. 2014;15(4):757-773.
33. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Comput Stat Data Anal*. 2003;43(2):121-137.
34. Peeken JC, Kessel KA, Nüsslin F, Braun AE, Combs SE. Semantic imaging features predict disease progression and survival in glioblastoma multiforme patients. *Strahlenther Onkol*. 2018;194(9):824-834.
35. Bae S, Kang S. Radiomic MRI phenotyping of glioblastoma: improving survival prediction. *Radiology*. 2018;289(3):797-806.
36. Kickingereder P, Burth S, Wick A, et al. Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology*. 2016;280(3):880-889.
37. Jansen NL, Suchorska B, Wenter V, et al. Prognostic significance of dynamic 18F-FET PET in newly diagnosed astrocytic high-grade glioma. *J Nucl Med*. 2015;56(1):9-15.
38. Szopa W, Burley TA, Kramer-Marek G, Kaspara W. Diagnostic and therapeutic biomarkers in glioblastoma: current status and future perspectives. *Biomed Res Int*. 2017;2017:1–13.
39. Narang S, Lehrer M, Yang D, Lee J, Rao A. Radiomics in glioblastoma : current status, challenges and potential opportunities. *Transl Cancer Res*. 2016;5(2):383-397.
40. Kickingereder P, Radbruch A, Burth S, et al. MR perfusion-derived hemodynamic parametric response mapping of bevacizumab efficacy in recurrent glioblastoma. *Radiology*. 2016;279(2):542-552.
41. Piroth MD, Holy R, Pinkawa M, et al. Prognostic impact of post-operative, pre-irradiation 18F-fluoroethyl-l-tyrosine uptake in glioblastoma patients treated with radiochemotherapy. *Radiother Oncol*. 2011;99(2):218-224.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Peeken JC, Goldberg T, Pyka T, et al. Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med*. 2019;8:128–136. <https://doi.org/10.1002/cam4.1908>