Glioma Survival Prediction with Combined Analysis of In Vivo ^{11}C-MET PET Features, Ex Vivo Features, and Patient Features by Supervised Machine Learning

László Papp[1], Nina Pötsch[2], Marko Grahovac[2], Victor Schmidbauer[2], Adelheid Woehrer[3], Matthias Preusser[4,5], Markus Mitterhauser[2,6], Barbara Kiesel[7], Wolfgang Wadsak[2,8], Thomas Beyer[1], Marcus Hacker[2], and Tatjana Traub-Weidinger[2]

[1]QIMP Group, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Vienna, Austria; [2]Division of Nuclear Medicine, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria; [3]Institute of Neurology, Medical University of Vienna, Vienna, Austria; [4]Department of Internal Medicine I, Medical University of Vienna, Vienna, Austria; [5]Comprehensive Cancer Center, Vienna, Austria; [6]Ludwig Boltzmann Institute Applied Diagnostics, Vienna, Austria; [7]Department of Neurosurgery, Medical University of Vienna, Vienna, Austria; and [8]Center for Biomarker Research in Medicine, Graz, Austria

Gliomas are the most common type of tumor in the brain. Although the definite diagnosis is routinely made ex vivo by histopathologic and molecular examination, diagnostic work-up of patients with suspected glioma is mainly done using MRI. Nevertheless, L-S-methyl-^{11}C-methionine (^{11}C-MET) PET holds great potential in the characterization of gliomas. The aim of this study was to establish machine-learning–driven survival models for glioma built on in vivo ^{11}C-MET PET characteristics, ex vivo characteristics, and patient characteristics. **Methods:** The study included 70 patients with a treatment-naïve glioma that was ^{11}C-MET–positive and had histopathology-derived ex vivo feature extraction, such as World Health Organization 2007 tumor grade, histology, and isocitrate dehydrogenase 1 R132H mutational status. The ^{11}C-MET–positive primary tumors were delineated semiautomatically on PET images, followed by the extraction of tumor-to-background–based general and higher-order textural features by applying 5 different binning approaches. In vivo and ex vivo features, as well as patient characteristics (age, weight, height, body mass index, Karnofsky score), were merged to characterize the tumors. Machine-learning approaches were used to identify relevant in vivo, ex vivo, and patient features and their relative weights for predicting 36-mo survival. The resulting feature weights were used to establish 3 predictive models per binning configuration: one model based on a combination of in vivo, ex vivo, and clinical patient information (M36$_{IEP}$); another based on in vivo and patient information only (M36$_{IP}$); and a third based on in vivo information only (M36$_I$). In addition, a binning-independent model based on ex vivo and patient information only (M36$_{EP}$) was created. The established models were validated in a Monte Carlo cross-validation scheme. **Results:** The most prominent machine-learning–selected and –weighted features were patient-based and ex vivo–based, followed by in vivo–based. The highest areas under the curve for our models as revealed by the Monte Carlo cross-validation were 0.9 for M36$_{IEP}$, 0.87 for M36$_{EP}$, 0.77 for M36$_{IP}$, and 0.72 for M36$_I$. **Conclusion:** Prediction of survival in amino acid PET–positive glioma patients was highly accurate using computer-supported predictive models based on in vivo, ex vivo, and patient features.

Key Words: glioma; amino acid PET; survival; radiomics; machine learning

Gliomas are the most common type of tumor in the brain, representing 81% of all cerebral malignancies. The incidence of gliomas as a whole is up to 5.7 per 100,000 people worldwide and increasing (1). Expected patient survival varies with glioma type, with the most frequent and highly malignant type—glioblastoma multiforme—showing the worst 5-y survival rate: about 5%. Clinical evaluation and therapeutic management of glioma patients currently rely on the combined analysis of age, Karnofsky score, and ex vivo tumor grade (1–3). Beyond tumor histology, molecular alterations such as isocitrate dehydrogenase 1 (IDH1) and 2 mutation, as part of the World Health Organization (WHO) 2016 classification system, have additional prognostic value in gliomas (4,5).

Imaging of gliomas is widely performed by MRI (6). Nevertheless, the high sensitivity and specificity of a radiolabeled amino acid PET tracer such as L-S-methyl-^{11}C-methionine (^{11}C-MET) is considered a promising diagnostic approach toward tumor characterization and longitudinal therapeutic monitoring (7–10).

The prognostic value of ^{11}C-MET PET is currently under investigation. Recent reports on the feasibility of dichotomizing tumors by a maximum tumor-to-background ratio (TBR) threshold are contradictory (11–14). In contrast, in vivo features derived from ^{11}C-MET PET SUVs have been reported to hold additive prognostic value (5).

One of the most prominent features of tumors is their heterogeneity across scales (15). It is therefore a logical step to investigate tumor heterogeneity in the context of survival prediction. Nevertheless, heterogeneity cannot be characterized by conventional calculations, such as SUVs, mean or maximum TBR, and metabolic tumor volumes (16). Recent studies have begun to focus on the evaluation of in vivo textural features on PET images, with

promising results for characterizing tumor heterogeneity (*17–19*), therapy response (*20*), and disease-specific survival (*21*). Although a wide range of textural features is available, these calculations are not yet standardized and are subject to variations in acquisition and reconstruction protocols (*15*). Furthermore, textural features derived from textural matrices are affected by the low sample size and the variation in number of bins (NOB) (*22,23*).

Another challenging aspect of textural analysis is feature selection and redundancy (*24*). Here, filtering during preprocessing (*24,25*) or machine learning (ML)–driven approaches to feature selection (*26,27*) can be incorporated to reduce the number of features for predictive analyses. ML approaches have been widely applied in the field of morphologic tissue characterization using large-scale radiomic features (*24,28*). However, these approaches are still underrepresented in the field of molecular imaging.

In light of the evolving field of texture analysis and radiomic evaluation of PET images, this study was performed with 3 objectives. The first was to propose ML-driven methods of feature selection and weight estimation to identify and compare in vivo ^{11}C-MET PET features, ex vivo features, and patient features of relevance for a 36-mo survival prediction. The second was to establish 4 ML-weighted-feature models to predict 36-mo survival: a model using in vivo, ex vivo, and patient features ($M36_{IEP}$); a model using ex vivo and patient features only ($M36_{EP}$); a model using in vivo and patient features only ($M36_{IP}$); and finally, a model using in vivo features only ($M36_{I}$). The third objective was to validate and compare all 4 models with retrospective survival information.

MATERIALS AND METHODS

Patient Data

Seventy patients with histologically verified treatment-naïve gliomas based on the WHO 2007 classification were collected from a preestablished cohort for this retrospective study (*29*). The patients were examined by ^{11}C-MET PET between 2000 and 2013. All were older than 18 y, and all had accessible medical reports with a follow-up of at least 36 mo. Moreover, the study included only amino acid–positive cases with a known IHD1 R132H mutational status based on immunohistochemical staining (Table 1). Days of survival were dichotomized with a 36-mo threshold and used as a reference label for both the model training and the validation phases of this work (Fig. 1). The study was approved by the local institutional review board. Written informed consent was obtained from all patients before the imaging examinations.

PET Acquisition

PET was performed on an Advance PET system (GE Healthcare) (*30*) 20 min after injection of 770 ± 106 MBq of ^{11}C-MET (mean ± SD; range, 447–972 MBq) produced in-house by a previously described method (*31*). The PET acquisition included a 10-min emission and a 5-min transmission scan. The attenuation- and scatter-corrected emission data were reconstructed by a standardized 3-dimensional filtered backprojection algorithm applying a Hanning filter with a 6.2-mm cutoff. The reconstructed axial matrix size was 128 × 128, with 35 slices per PET acquisition and a 3.125-mm slice thickness. The full width at half maximum of the reconstructed images was 5 mm.

Spatial Normalization and Tumor Delineation

Medical images are commonly resampled to a unified isotropic spatial resolution before radiomic evaluation to standardize voxel size differences in single or multicenter cohorts (*15,32*). To support repeatability, the PET images were resampled to a 1 × 1 × 1 mm voxel size. In choosing the target resolution, we considered the sensitivity of

TABLE 1
Characteristics of Processed Study Cohort

Characteristic	Data
Total patients	70 (100%)
Median age ± SD (y)	48 ± 15
Sex	
Male	42 (60%)
Female	28 (40%)
Histologic type	
Diffuse astrocytoma	17 (24%)
Oligodendroglioma	22 (31%)
Oligoastrocytoma	13 (19%)
Glioblastoma multiforme	17 (24%)
Pilocytic astrocytoma	1 (<2%)
WHO 2007 grade	
I	1 (<2%)
II	21 (30%)
III	31 (44%)
IV	17 (24%)
IDH1 R132H mutation	
Positive	35 (50%)
Negative	35 (50%)

Data are *n* followed by percentage in parentheses, except for age.

various textural features in a small sample (*33,34*). To minimize interpolation artifacts, the resampling was performed with the kriging interpolation (*35*).

The resampled PET images were transferred to commercially available software (Hybrid 3D; Hermes Medical Solutions) for tumor delineation. This process was performed by 2 nuclear medicine physicians in consensus using semiautomated 3-dimensional isocount volume-of-interest (VOI) tools. An optional slice-by-slice manual modifier tool was applied if the tumor boundaries could not be characterized by the isocount VOI tool. An additional reference background cuboid VOI was drawn on the contralateral region for TBR calculations. The coordinates and values of the voxels within the VOIs were exported via the Hybrid 3D software for further processing.

Feature Extraction

The PET voxel values inside each tumor VOI were normalized to the mean of the respective reference background VOI to generate a PET TBR VOI. This step was necessary to correct the individual tracer metabolism of the normal tissue (*36,37*). The PET TBRs were rebinned in 5 different ways considering different NOB and bin size (BS) configurations. Four binning processes were initiated using Equation 1 for a cohort-global bin range with a T_{min} of 1.0 and a T_{max} of 8.5, with NOB equaling 64, 150, 375, and 512. The corresponding BS values were 0.12, 0.05, 0.02, and 0.014, respectively:

$$v_{bin} = NOB \frac{v - T_{min}}{T_{max} - T_{min}}, \qquad \text{Eq. 1}$$

where v is the original TBR voxel value from a given VOI, v_{bin} is the binned TBR value, and NOB is 64, 150, 375, and 512. TBRs with less than a T_{min} of 1.0 were excluded from the VOIs. A T_{max} of 8.5
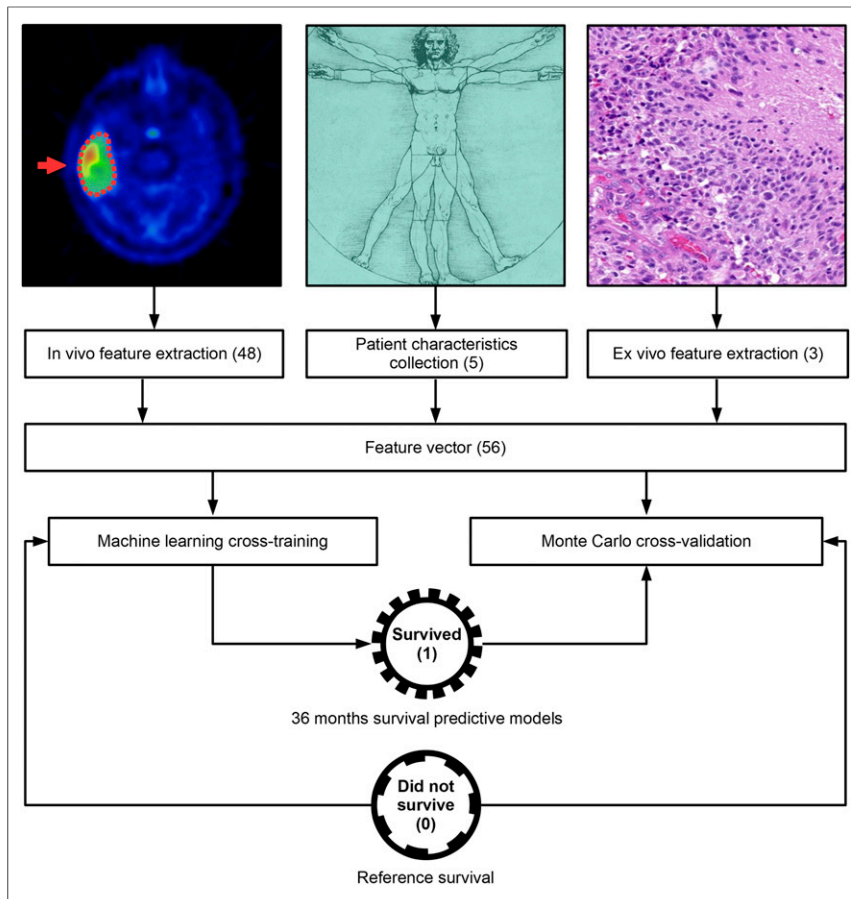
FIGURE 1. Workflow for evaluating glioma cohort. Tumors are manually delineated on ^{11}C-MET PET images, followed by automated extraction of 48 in vivo features. In addition, 5 patient characteristics and 3 ex vivo features are collected to establish 56-feature vector for each case. Extracted features are used for ML cross-training phase, which results in relevant features and their weights for 36-mo survival. Feature weights for 36-mo-survival predictive model are established on basis of ML results. Model is validated with MC cross-validation approach. Reference dichotomized standard survival labels (did not survive [0], survived [1]) are used during both training and validation phases.

arrays to provide a membership probability for each of the $k \in \{\text{survived}, \text{did not survive}\}$ classifier outcomes. In this study, the gaussian distribution was determined by random bootstrapping with replacement (*41,42*). The probability that a feature vector ($\overline{\varphi} \in \mathrm{R}^{56}$) belonged to k classifier outcomes was characterized by MP_k membership probability functions (Eq. 3). This study extended the above approach in 2 ways. First, a binary feature selection array ($\overline{m} \in \{0,1\}^{56}$) was used to describe which features are relevant in the evaluation, and second, a feature weight array ($\overline{w} \in \mathrm{R}^{56}$) was introduced to represent the importance of each selected feature:

$$\mathrm{MP}_k(\overline{\varphi}, \overline{\mu}_k, \overline{\sigma}_k, \overline{m}, \overline{w}) = \sum_{i=1}^{56} m_i\, w_i e^{\frac{\left(\varphi_i - \mu_{k,i}\right)^2}{2\sigma_{k,i}^2}}$$
<div align="right">Eq. 3</div>

The predicted label of a feature vector $\overline{\varphi}$ was provided by the MP_k function with the maximum probability value (Eq. 4):

$$\arg_k \max(\mathrm{MP}_k(\overline{\varphi}, \overline{\mu}_k, \overline{\sigma}_k, \overline{m}, \overline{w})) \qquad \text{Eq. 4}$$

The above predictive model scheme was referred to as M36 in this study. Both the feature selection ($\overline{m}$) and the feature weight ($\overline{w}$) arrays were unknown parameters that were determined by ML approaches.

Model Error Estimation

An estimator was established to characterize the receiver-operator-characteristic distance of the M36 models (Eq. 5) based on number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Supplemental Table 1; supplemental materials are available at http://jnm.snmjournals.org). The measurement compared the model-predicted and reference label values of input feature vectors using the confusion matrix (*43*):

$$\varepsilon = \sqrt{\left(1 - \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\right)^2 + \left(1 - \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}\right)^2}. \qquad \text{Eq. 5}$$

Feature Selection and Weight Estimation

For each of the 5 binning configurations, $\overline{m}$ and $\overline{w}$ were identified in a hierarchical ML-based approach by minimizing the ε model error (Fig. 2). The first ML layer identified relevant features ($\overline{m}$) through an interactive approach using genetic algorithms (*26*), thereby modeling evolutionary processes. The second ML layer then identified the residual weights ($\overline{w}$) based on the content of each input $\overline{m}$ using the Nelder–Mead method (*44,45*). An inherent dependency between the mask and the weight vectors was maintained in such a way that if a feature was not selected ($m_i = 0$), then its weight was zero as well ($w_i = 0$).

To avoid overfitting, the multilayer ML approach was executed in a 14-fold cross-training scheme over the dataset (*46*) with each of the 5 binning configurations. In each fold, 8 different ML algorithms were executed with different genetic algorithm mutation rates managed in

represented the highest TBR in the cohort. The fifth binning technique applied a fixed BS of 0.05 TBR but with a specific NOB per tumor (*38*) as defined by Equation 2.

$$v_{\mathrm{bin}} = \frac{v - \mathrm{VOI}_{\min}}{\mathrm{BS}}, \qquad \text{Eq. 2}$$

where $\mathrm{VOI}_{\min}$ is the minimum voxel value of the given VOI with a BS of 0.05.

Each of the binned PET TBR VOIs was subjected to extraction of 48 in vivo features, including general, histogram, and shape features, as well as textural features derived from the gray-level cooccurrence matrix, the gray-level zone-size matrix, and the neighborhood gray-tone difference matrix (*15,23,39*). In addition, 3 ex vivo features and 5 patient characteristics were assigned to the in vivo features to generate a 56-feature vector (48 + 3 + 5) for each tumor (Table 2).

Survival Prediction Model

A model scheme building on all 56 in vivo, ex vivo, and patient features was established on the basis of the principles of geometric probability covering algorithms (*40*). These algorithms model the gaussian distribution of features with $\overline{\mu}_k \in \mathrm{R}^{56}$ means and $\overline{\sigma}_k \in \mathrm{R}^{56}$ deviation

TABLE 2
TABLE 2
Fifty-Six Extracted In Vivo and Ex Vivo Features and Patient Characteristics Assigned to Each Delineated Lesion in Feature Vector

Feature category	Feature name
In vivo general (6)	Minimum, maximum, sum, mean, SD, variance
In vivo histogram (6)	Mean, energy, variance, entropy, skewness, kurtosis
In vivo shape (3)	Compactness, volume, spheric dice coefficient
In vivo GLCM (17)	Inverse difference, inverse difference moment, sum average, sum entropy, difference variance, difference entropy, information correlation, auto correlation, cluster shade, cluster prominence, maximum probability, entropy, contrast, dissimilarity, angular second moment, sum-of-squares variance, correlation
In vivo GLZSM (11)	Emphasis on small zone, large zone, low-gray-level zone, high-gray-level zone, small low-gray-level zone, small high-gray-level zone, large low-gray-level zone, large high-gray-level zone; gray-level nonuniformity; zone-size nonuniformity; zone-size percentage
In vivo NGTDM (5)	Coarseness, contrast, complexity, busyness, texture strength
Ex vivo (3)	Histology, WHO 2007 grade, IDH1 R132H mutational status
Patient (5)	Age, weight, height, body mass index, Karnofsky score

GLCM = gray-level cooccurrence matrix, GLZSM = gray-level zone-size matrix, NGTDM = neighborhood gray-tone difference matrix. Details of listed in vivo features can be found in supplemental materials.

parallel by simulated annealing (*47*). In this way, 112 (14 × 8) feature-mask and feature-weight variants were generated per binning configuration. The final feature mask ($\overline{m}$) was created by a feature-level logical OR operator over the 112 feature-mask variants. The 112 feature-weight variants of the given binning configuration were normalized to the sum of 1.0 and averaged to create the final feature-weight array ($\overline{w}$). The supplemental materials contain a detailed explanation of the ML algorithm and its parameters.

On the basis of the ML-derived features ($\overline{m}$) and their weights ($\overline{w}$), three 36-mo predictive models were established for each of the 5 bin configurations: $M36_{IEP}$, $M36_{IP}$, and $M36_I$. In addition, a binning-independent model, $M36_{EP}$, was created.

Predictive Model Validation

To measure the performance of the established models, Monte Carlo (MC) cross-validation was used (*48*) with 1,000 iterations. In each MC iteration, the dataset was randomly separated into a 60% training dataset (TDS) and a 40% validation dataset (VDS), with stratified selection and no overlapping (TDS ∩ VDS = ∅). The reference gaussian distributions ($\overline{\mu}_k, \overline{\sigma}_k$) of $M36_{IEP}$, $M36_{EP}$, $M36_{IP}$, and $M36_I$ were calculated from the given TDS. The corresponding VDS samples were subsequently evaluated by the configured models in each MC iteration. The predicted and respective reference-value pairs of VDS samples were recorded in a confusion matrix for each model for performance evaluation.

RESULTS

Feature Selection and Weight Estimation

Based on the averaged weights of the 5 binning-specific ML executions, the most prominent features were patient features and ex vivo features, such as age (10.3%), isocitrate dehydrogenase 1 (IDH1) R132H mutational status (8.6%), and WHO 2007 grade (6.8%). TBR sum (5%), spheric dice coefficient (4.7%), volume (4.5%), and coarseness neighborhood gray-tone difference matrix (4.1%) were the most prominent in vivo features (Fig. 3). The same prominent weights were selected by ML regardless of the binning configuration; however, individual weights differed per binning configuration (Supplemental Figs. 1–4).

Predictive Model Validation

The models with highest area under the curve (AUC) in their category based on the MC cross-validation were $M36_{IEP}$ (BS, 0.05), with an AUC of 0.9; $M36_{EP}$, with an AUC of 0.87; $M36_{IP}$ (NOB, 150), with an AUC of 0.77; and $M36_I$ (BS, 0.05), with an AUC of 0.73. The average AUCs of the 4 model types across the different binning configurations were 0.9 for $M36_{IEP}$, 0.87 for $M36_{EP}$, 0.77 for $M36_{IP}$, and 0.7 for $M36_I$. AUC did not significantly differ among different binning configurations; however, the
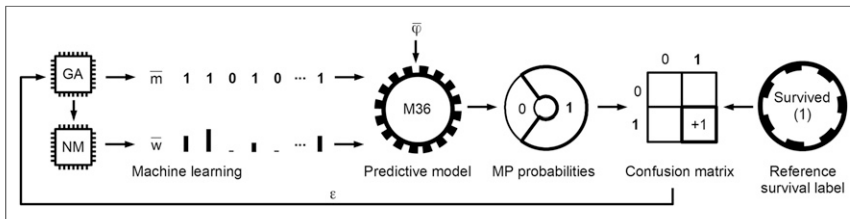

FIGURE 2. Identification of relevant features and their weights specific for 36-mo survival by ML. Genetic algorithm (GA) and Nelder–Mead (NM) ML methods determine feature-mask ($\overline{m}$) and feature-weight ($\overline{w}$) arrays, respectively. Generic M36 predictive model evaluates input feature vector ($\overline{\varphi}$) and provides its survived (1) and did-not-survive (0) membership probabilities (MP). Highest probability (in current example, "Survived (1)") is chosen as predicted value. Predicted and reference labels of $\overline{\varphi}$ are compared and stored in confusion matrix. Error measurement (ε) from confusion matrix is provided once all feature vectors in training phase are evaluated. ε error value is minimized by ML layers (GA and NM). The above process is performed in iterative cross-training scheme of 14 folds and 8 ML variants in each fold, resulting in 112 feature-mask and feature-weight variants that are used to identify relevant features and their weights for 36-mo glioma survival.
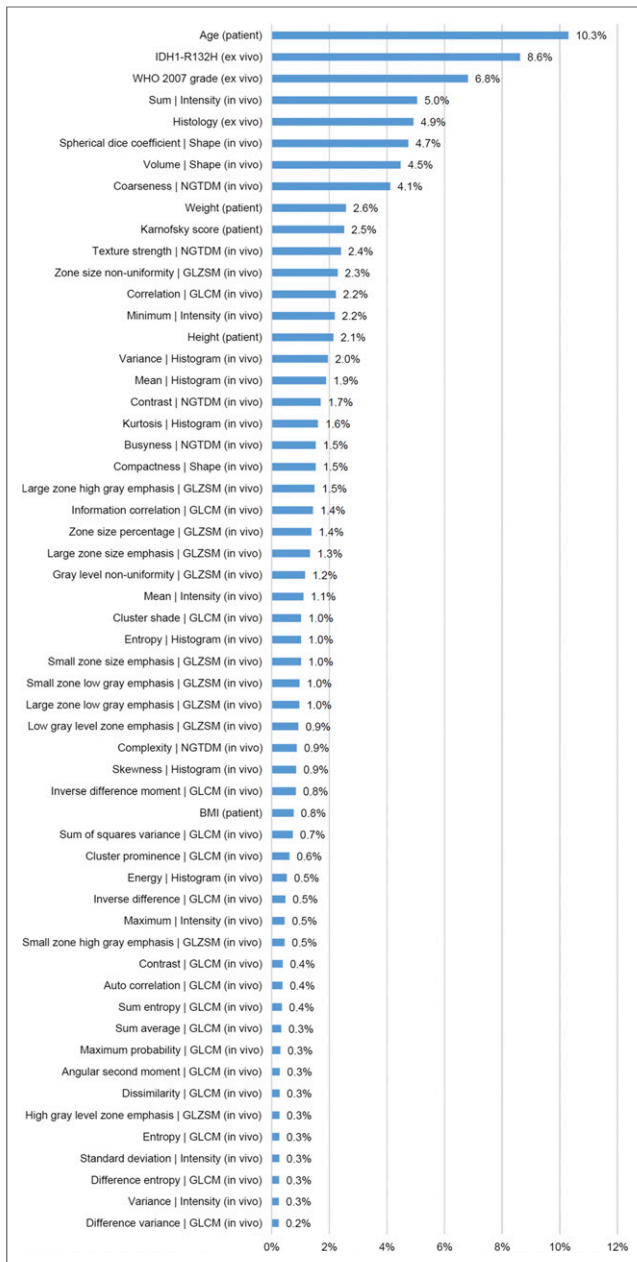
FIGURE 3. ML-derived weights of 56 in vivo, ex vivo, and patient features in descending order. Weights reflect relative importance to one another for predicting 36-mo survival in ^{11}C-MET PET–positive patients. Weights were determined by averaging 560 (5 binning × 112 models) weight variants derived by ML, executed in cross-training scheme. Individual weights of each weight variant were normalized to sum of 1.0 before averaging. BMI = body mass index; GLCM = gray-level cooccurrence matrix; GLZSM = gray-level zone-size matrix; NGTDM = neighborhood gray-tone difference matrix.

largest AUC for models involving in vivo features was achieved with a binning configuration that had a BS of 0.05 (fixed BS, variable NOB per tumor), supporting previous results (38).

A detailed comparison of the models by sensitivity, specificity, accuracy, positive predictive value, negative predictive value, and AUC is shown in Table 3. Figure 4 compares 2 example patients

with prominent features who did and did not survive the 36-mo period.

DISCUSSION

This study investigated the relevance of in vivo, ex vivo, and patient-related features in predicting 36-mo survival for glioma patients and determined the importance of selected features by ML. Although all features were selected at least once during the cross-training phase, the determined feature weights had a nonuniform distribution (Fig. 3 and Supplemental Figs. 1–3).

Across all investigated features, gray-level cooccurrence matrix features such as entropy (0.3%) and angular second moment (0.3), as well as intensity features such as maximum TBR (0.5%), appeared to have low importance in predicting 36-mo survival in patients with amino acid–positive gliomas. In contrast, patient age, IDH1 R132H mutational status, WHO 2007 grade, TBR sum, spheric dice coefficient, and volume and coarseness neighborhood gray-tone difference matrix appeared to be highly important for survival prediction. Our findings regarding TBR sum support previous reports that identified tumor amino acid metabolism as a prominent feature for survival prediction (5,29). In addition, we have shown that tumor-shape features such as volume and spheric dice coefficient further support the accuracy of survival prediction beyond patient and ex vivo values.

TABLE 3
Performance Values of Predictive Models Evaluated in MC Cross-Validation

Binning configuration	Model	SENS	SPEC	ACC	PPV	NPV	AUC
NOB, 64	M36$_{IEP}$	86%	95%	89%	97%	78%	0.90
	M36$_{IP}$	82%	70%	78%	84%	67%	0.76
	M36$_I$	77%	64%	72%	80%	59%	0.70
NOB, 150	M36$_{IEP}$	88%	95%	90%	97%	81%	0.91
	M36$_{IP}$	81%	74%	79%	85%	67%	0.77
	M36$_I$	77%	65%	73%	81%	60%	0.71
NOB, 375	M36$_{IEP}$	88%	93%	90%	96%	81%	0.91
	M36$_{IP}$	83%	71%	79%	85%	69%	0.77
	M36$_I$	77%	61%	72%	79%	59%	0.69
NOB, 512	M36$_{IEP}$	89%	92%	90%	96%	82%	0.91
	M36$_{IP}$	84%	71%	79%	85%	70%	0.77
	M36$_I$	79%	60%	72%	79%	60%	0.69
BS, 0.05	M36$_{IEP}$	88%	94%	90%	96%	80%	0.91
	M36$_{IP}$	79%	76%	78%	86%	65%	0.77
	M36$_I$	76%	69%	74%	82%	61%	0.73
Nonimaging*	M36$_{EP}$	79%	95%	84%	97%	70%	0.87

*M36$_{EP}$ is presented separately, as it is independent from image binning configurations.

SENS = sensitivity; SPEC = specificity; ACC = accuracy; PPV = positive predictive value; NPV = negative predictive value.
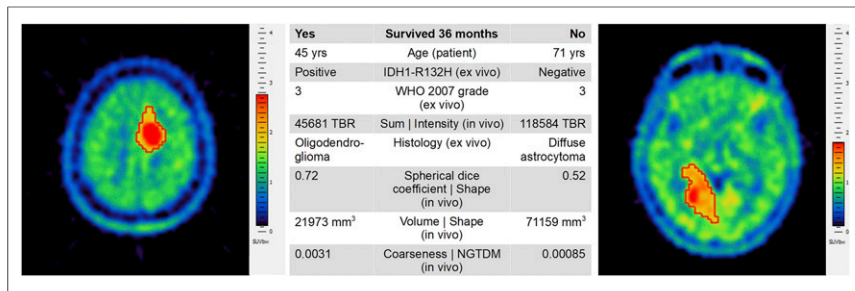
FIGURE 4. Comparison of 2 example patients who did (left) and did not (right) survive 36 mo from time of primary ^{11}C-MET PET scan. Most prominent features as identified by our study are presented in center table. Although both patients had WHO 2007 grade 3 disease, remaining features varied considerably, thus indicating need for combined analysis of multiple features. Axial slices are from Hybrid 3D software visualized by standard spectral palette and with overlaid delineated VOIs (red boundary). NGTDM = neighborhood gray-tone difference matrix.

Overall, 16 predictive models, built on different combinations of binning configurations with ML-determined in vivo, ex vivo, and patient features, were validated in a MC cross-validation scheme. Our results indicate that the highest AUC (0.9) can be achieved with M36$_{IEP}$. The second highest average AUC (0.87) was for M36$_{EP}$. M36$_{IP}$ and M36$_I$ had an average AUC of 0.77 and 0.70, respectively (Table 3). All our combined M36$_{IEP}$ models resulted in a sensitivity and specificity of 86%–98% and 92%–95%, respectively. In contrast, when ex vivo features were excluded, sensitivity was in the range of 79%–84% and specificity decreased to 71%–76% (Table 3), thus indicating that ex vivo features support the identification of patients who are more likely to survive 36 mo. The inclusion of patient features appeared to support a higher sensitivity and specificity across all models.

A literature search demonstrated ML-based glioma predictors; however, these studies relied on MRI-based feature analysis (Table 4). The highest accuracy (96%) for a survival prediction model built on MRI, ex vivo, and patient features was reported by Zhang et al. (49). However, their study was based on a comparatively small patient cohort (n = 28) that included glioblastomas only. Moreover, they used contrast-enhanced T1-weighted MRI–based radiomic features at 3 different time points during the course of the disease. Nie et al.

(50) reported 89% accuracy for their survival model across 69 patients, and the approach of Macyszyn et al. (51) resulted in 76% accuracy (105 patients). Again, both these studies involved only glioblastomas. Emblem et al. (52) established 6-mo, as well as 1-, 2-, and 3-y, survival models based on contrast-enhanced MRI features, tumor volume, and patient features for 235 glioma patients in a multicenter study. They reported 94% sensitivity, 38% specificity, and an AUC of 0.66 for their 36-mo survival model. The highest AUC (0.682) was provided by their 2-y survival model, with 78% sensitivity and 58% specificity.

Our study differed from the above-cited works on multiple accounts: it included the diagnostic, pretherapeutic PET into a combined analysis of in vivo, ex vivo, and patient features. Furthermore, our cohort included not just glioblastoma but various types of glioma (Table 1). We specifically focused on using a statistically accurate resampling (35) to address small-sample–related uncertainties in textural parameters (22,23). Both the feature selection and the feature-weight estimation were ML-driven in a cross-training scheme to minimize the bias of our models. In addition, we compared 5 different binning configuration–based ML executions to investigate their effect on our performance values. Furthermore, the validation was performed by a MC cross-validation scheme with a high iteration to properly estimate the accuracy of our models. Our results outperform previously reported MRI-based results, thus indicating that amino acid PET may hold a prominent role in glioma survival prediction as an alternative or addition to MR-only imaging.

Nevertheless, our study was limited in several regards. Although IDH1 R132H mutational status—as one of the essential biomarkers in the 2016 update of the WHO glioma classification—was present, tumor typing was based on the older, 2007, WHO standard. Furthermore, because the in vivo feature extraction relied on PET-identifiable VOI analysis, only amino acid–positive gliomas could be investigated. Last, the current work was built on single-center studies; thus, the effect of acquisition and reconstruction variations on our in vivo features could not be

TABLE 4
Studies Correlating Glioma Survival with In Vivo, Ex Vivo, or Patient Features by ML Approaches

Study	Cohort	Feature source	Analysis method	Validation method	Result
Zhang et al. (49)	GBM (n = 28)	MRI, ex vivo, patient	Logistic regression, SVM, decision tree, neural network	Leave-one-out cross-validation	96% ACC
Nie et al. (50)	GBM (n = 69)	MRI	Deep learning	Cross-validation	89% ACC
Macyszyn et al. (51)	GBM (n = 105)	MRI	Feature preselection, SVM	Cross-validation	76% ACC
Emblem et al. (52)	Glioma (n = 235)	MRI, tumor volume, patient	SVM	Multicenter cross-validation	94% SENS, 38% SPEC

GBM = glioblastoma; SVM = support vector machine; ACC = accuracy; SENS = sensitivity; SPEC = specificity.

evaluated (*15*). Looking ahead, a logical next step in identifying key features that correlate with glioma survival would be to extract features from multicenter PET/MR images. The proposed ML methods together with the predictive model are highly generic, as they do not consider any prior knowledge about the modalities or extracted features. Thus, evaluation of these ML methods for other cancers can be envisaged.

CONCLUSION

The results of the current study support the application of ML using in vivo, ex vivo, and patient features to predict survival in amino acid PET–positive glioma patients.

DISCLOSURE

Thomas Beyer is founder of cmi-experts GmbH. No other potential conflict of interest relevant to this article was reported.

REFERENCES

1. Ostrom QT, Bauchet L, Davis FG, et al. The epidemiology of glioma in adults: a state of the science review. *Neuro-Oncol.* 2014;16:896–913.
2. Daniels TB, Brown PD, Felten SJ, et al. Validation of EORTC prognostic factors for adults with low-grade glioma: a report using intergroup 86-72-51. *Int J Radiat Oncol Biol Phys.* 2011;81:218–224.
3. Gorlia T, Stupp R, Brandes AA, et al. New prognostic factors and calculators for outcome prediction in patients with recurrent glioblastoma: a pooled analysis of EORTC Brain Tumour Group phase I and II clinical trials. *Eur J Cancer.* 2012;48:1176–1184.
4. Eckel-Passow JE, Lachance DH, Molinaro AM, et al. Glioma groups based on 1p/19q, *IDH*, and *TERT* promoter mutations in tumors. *N Engl J Med.* 2015;372:2499–2508.
5. Lopci E, Riva M, Olivari L, et al. Prognostic value of molecular and imaging biomarkers in patients with supratentorial glioma. *Eur J Nucl Med Mol Imaging.* 2017;44:1155–1164.
6. Fouke SJ, Benzinger T, Gibson D, Ryken TC, Kalkanis SN, Olson JJ. The role of imaging in the management of adults with diffuse low grade glioma: a systematic review and evidence-based clinical practice guideline. *J Neurooncol.* 2015;125:457–479.
7. Ueda S, Tsuda H, Asakawa H, et al. Clinicopathological and prognostic relevance of uptake level using ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography fusion imaging (^{18}F-FDG PET/CT) in primary breast cancer. *Jpn J Clin Oncol.* 2008;38:250–258.
8. Yoo MY, Paeng JC, Cheon GJ, et al. Prognostic value of metabolic tumor volume on ^{11}C-methionine PET in predicting progression-free survival in high-grade glioma. *Nucl Med Mol Imaging.* 2015;49:291–297.
9. Galldiks N, Law I, Pope WB, Arbizu J, Langen KJ. The use of amino acid PET and conventional MRI for monitoring of brain tumor therapy. *Neuroimage Clin.* 2016;13:386–394.
10. Yamane T, Sakamoto S, Senda M. Clinical impact of ^{11}C-methionine PET on expected management of patients with brain neoplasm. *Eur J Nucl Med Mol Imaging.* 2010;37:685–690.
11. Smits A, Westerberg E, Ribom D. Adding ^{11}C-methionine PET to the EORTC prognostic factors in grade 2 gliomas. *Eur J Nucl Med Mol Imaging.* 2008;35:65–71.
12. Takano K, Kinoshita M, Arita H, et al. Diagnostic and prognostic value of ^{11}C-methionine PET for nonenhancing gliomas. *AJNR Am J Neuroradiol.* 2016;37:44–50.
13. Nariai T, Tanaka Y, Wakimoto H, et al. Usefulness of L-[methyl-^{11}C] methionine-positron emission tomography as a biological monitoring tool in the treatment of glioma. *J Neurosurg.* 2005;103:498–507.
14. Kobayashi K, Hirata K, Yamaguchi S, et al. Prognostic value of volume-based measurements on ^{11}C-methionine PET in glioma patients. *Eur J Nucl Med Mol Imaging.* 2015;42:1071–1080.
15. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging.* 2017;44:151–165.
16. Visvikis D, Hatt M, Tixier F, Le Rest CC. The age of reason for FDG PET image-derived indices. *Eur J Nucl Med Mol Imaging.* 2012;39:1670–1672.
17. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline ^{18}F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011;52:369–378.
18. Tixier F, Hatt M, Cheze-Le Rest C, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in ^{18}F-FDG PET. *J Nucl Med.* 2012;53:693–700.
19. Werner RA, Kroiss M, Nakajo M, et al. Assessment of tumor heterogeneity in treatment-naïve adrenocortical cancer patients using ^{18}F-FDG positron emission tomography. *Endocrine.* 2016;53:791–800.
20. Bundschuh RA, Dinges J, Neumann L, et al. Textural parameters of tumor heterogeneity in ^{18}F-FDG PET/CT for therapy response assessment and prognosis in patients with locally advanced rectal cancer. *J Nucl Med.* 2014;55:891–897.
21. Pyka T, Bundschuh RA, Andratschke N, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol.* 2015;10:100.
22. Desseroit M-C, Tixier F, Weber WA, et al. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non–small cell lung cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J Nucl Med.* 2017;58:406–411.
23. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern Syst.* 1973;3:610–621.
24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* 2016;278:563–577.
25. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
26. Jung M, Zscheischler J. A guided hybrid genetic algorithm for feature selection with expensive cost functions. *Procedia Comput Sci.* 2013;18:2337–2346.
27. Mi H, Petitjean C, Dubray B, Vera P, Ruan S. Robust feature selection to predict tumor treatment outcome. *Artif Intell Med.* 2015;64:195–204.
28. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441–446.
29. Poetsch N, Woehrer A, Gesperger J, et al. Visual and semiquantitative ^{11}C-methionine PET: an independent prognostic factor for survival of newly diagnosed and treatment-naïve gliomas. *Neuro-Oncol.* 2018;20:411–419.
30. DeGrado TR, Turkington TG, Williams JJ, Stearns CW, Hoffman JM, Coleman RE. Performance characteristics of a whole-body PET scanner. *J Nucl Med.* 1994;35:1398–1406.
31. Mitterhauser M, Wadsak W, Krcal A, et al. New aspects on the preparation of [^{11}C]methionine: a simple and fast online approach without preparative HPLC. *Appl Radiat Isot.* 2005;62:441–445.
32. Yip SSF, Parmar C, Kim J, Huynh E, Mak RH, Aerts HJWL. Impact of experimental design on PET radiomics in predicting somatic mutation status. *Eur J Radiol.* 2017;97:8–15.
33. Brooks FJ, Grigsby PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med.* 2014;55:37–42.
34. Hatt M, Majdoub M, Vallieres M, et al. ^{18}F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med.* 2015;56:38–44.
35. Stytz MR, Parrott RW. Using kriging for 3D medical imaging. *Comput Med Imaging Graph.* 1993;17:421–442.
36. Vander Borght T, Asenbaum S, Bartenstein P, et al. EANM procedure guidelines for brain tumour imaging using labelled amino acid analogues. *Eur J Nucl Med Mol Imaging.* 2006;33:1374–1380.
37. 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;50:167–173.

38. Leijenaar RTH, Nalbantov G, Carvalho S, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep.* 2015;5:11075.

39. Miyamoto E, Merryman T Jr. *Fast Calculation of Haralick Texture Features.* Pittsburgh, PA: Human Computer Interaction Institute; 2005.

40. Zhang J, Li SZ, Wang J. Geometrical probability covering algorithm. In: Wang L, Jin Y, eds. *Fuzzy Systems and Knowledge Discovery.* New York, NY: Springer; 2005:223–231.

41. Beleites C, Baumgartner R, Bowman C, et al. Variance reduction in estimating classification error using sparse datasets. *Chemometr Intell Lab Syst.* 2005;79: 91–100.

42. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2014;13:8–17.

43. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett.* 2006;27: 861–874.

44. Nelder JA, Mead R. A simplex method for function minimization. *Comput J.* 1965;7:308–313.

45. Singer S, Singer S. Efficient implementation of the Nelder-Mead search algorithm. *Appl Numer Anal Comput Math.* 2004;1:524–534.

46. Ypsilantis P-P, Siddique M, Sohn H-M, et al. Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. Anto RJ, ed. *PLoS One.* 2015;10:e0137036.

47. Moseley HN, Lane AN, Belshoff AC, Higashi RM, Fan TW. A novel deconvolution method for modeling UDP-N-acetyl-D-glucosamine biosynthetic pathways based on ^{13}C mass isotopologue profiles under non-steady-state conditions. *BMC Biol.* 2011;9:37.

48. Li T, Tang W, Zhang L. Monte Carlo cross-validation analysis screens pathway cross-talk associated with Parkinson's disease. *Neurol Sci.* 2016;37:1327–1333.

49. Zhang H, Molitoris J, Tan S, et al. SU-F-R-04: radiomics for survival prediction in glioblastoma (GBM) [abstract]. *Med Phys.* 2016;43:3373.

50. Nie D, Zhang H, Adeli E, et al. 3D deep learning for multi-model imaging-guided survival time prediction of brain tumor patients. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016.* Vol. 9901. Cham, Switzerland: Springer International Publishing; 2016:212–220.

51. Macyszyn L, Akbari H, Pisapia JM, et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-Oncol.* 2016;18:417–425.

52. Emblem KE, Pinho MC, Zöllner FG, et al. A generic support vector machine model for preoperative glioma survival associations. *Radiology.* 2015;275:228–234.