FLT PET Radiomics for Response Prediction to Chemoradiation Therapy in Head and Neck Squamous Cell Cancer

Ethan J. Ulrich[1,2], Yusuf Menda[3], Laura L. Boles Ponto[3], Carryn M. Anderson[4], Brian J. Smith[5], John J. Sunderland[3], Michael M. Graham[3], John M. Buatti[4], and Reinhard R. Beichel[1,6]

Departments of [1]Electrical and Computer Engineering, [2]Biomedical Engineering, [3]Radiology, [4]Radiation Oncology, [5]Biostatistics, and [6]Internal Medicine, University of Iowa, Iowa City, IA

Corresponding Author:
Reinhard R. Beichel, PhD
Department of Electrical and Computer Engineering, University of Iowa,
4016 Seamans Center for the Engineering Arts and Sciences, Iowa
City, IA 52242;
E-mail: reinhard-beichel@uiowa.edu

Radiomics is an image analysis approach for extracting large amounts of quantitative information from medical images using a variety of computational methods. Our goal was to evaluate the utility of radiomic feature analysis from ^{18}F-fluorothymidine positron emission tomography (FLT PET) obtained at baseline in prediction of treatment response in patients with head and neck cancer. Thirty patients with advanced-stage oropharyngeal or laryngeal cancer, treated with definitive chemoradiation therapy, underwent FLT PET imaging before treatment. In total, 377 radiomic features of FLT uptake and feature variants were extracted from volumes of interest; these features variants were defined by either the primary tumor or the total lesion burden, which consisted of the primary tumor and all FLT-avid nodes. Feature variants included normalized measurements of uptake, which were calculated by dividing lesion uptake values by the mean uptake value in the bone marrow. Feature reduction was performed using clustering to remove redundancy, leaving 172 representative features. Effects of these features on progression-free survival were modeled with Cox regression and P-values corrected for multiple comparisons. In total, 9 features were considered significant. Our results suggest that smaller, more homogenous lesions at baseline were associated with better prognosis. In addition, features extracted from total lesion burden had a higher concordance index than primary tumor features for 8 of the 9 significant features. Furthermore, total lesion burden features showed lower interobserver variability.

INTRODUCTION

Concomitant chemoradiation is used as an organ-sparing treatment strategy for advanced oropharyngeal and larynx cancers. Although outcomes vary based on stage, site, and other factors including human papilloma virus status, the 3-year progression-free survival of patients with advanced-stage head and neck cancer after chemoradiation therapy (CRT) is ~60% (1). Patients in whom cancer recurs after initial CRT are considered for salvage surgery; but, patients with presalvage Stage IV disease and those with presalvage Stage III disease at recurrence have poor prognosis with a median survival of <6 month and 14 months, respectively (2). As both new targeted therapies and radiation therapy (RT) delivery methods are developed, there is a need to develop biomarkers that may help stratify patients *a priori* for different treatment modalities or that can predict the likelihood of durable response versus ultimate failure earlier during therapy to allow for adaptive treatment approaches.

Positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose (FDG PET) is widely used in pretreatment staging and post-therapy evaluation of head and neck cancers after RT or CRT. Because of its high negative predictive value in detection of recurrent disease, the National Comprehensive Cancer Network Guidelines now recommend omitting consolidative surgery (neck dissection) if the post-therapy FDG PET obtained at least 12 weeks after initial therapy is negative for residual tumor (3). However, the role of FDG PET in predicting failure of CRT or monitoring treatment response to (chemo)radiation during or early after treatment is not well established (12 weeks after initial therapy is typically required).

Radiation therapy and chemotherapy affect proliferation rates in treated tumors. In addition, pretreatment proliferation

rates may be a determinant of sensitivity to chemotherapy and RT. Assessment of cancer proliferation rates and changes in cell proliferation rate may therefore accurately predict ultimate therapeutic response. 3′-deoxy-3′-^{18}F-fluorothymidine (FLT), a thymidine analogue that is not incorporated into DNA, is the most widely studied PET agent for imaging cell proliferation. The intracellular trapping of FLT is regulated by thymidine kinase 1, a key enzyme in DNA synthesis, with high activity during the proliferative phase of the cell cycle and low activity in the quiescent phase (4). Several studies have shown that untreated head and neck cancers can be imaged with FLT PET with a high tumor-to-background contrast (5–9).

Radiomics is an image analysis approach with the goal of extracting large amounts of quantitative information from medical images using a variety of computational methods. Extracted features include measurements of intensity (uptake), shape, and texture. The objective of this study was to evaluate the utility of FLT PET radiomic features obtained at baseline in the prediction of treatment response in patients with head and neck squamous cell cancer (HNSCC). The present work provides a basis for further optimization of predictive FLT PET features, which can then be further evaluated in future clinical trials.

METHODOLOGY

Patients

A single-center prospective study was performed in patients who had histologically confirmed HNSCC and were scheduled to receive definitive concurrent CRT per standard cancer care. Other eligibility criteria included a Karnofsky score of ≥60, acceptable bone marrow reserve (absolute neutrophil count, ≥1.5 K/mL; platelet count, ≥100 K/mL) and kidney (serum creatinine, ≤2.1 mg/dL), and liver function (bilirubin, ≤1.0 mg/dL; ALT/AST, ≤2.5 times upper limits of normal for the institution). These criteria generally excluded patients who were not robust enough to receive combined modality therapy. Patients were excluded if they had chemotherapy or radiotherapy within 4 weeks before the study (no induction chemotherapy) or were receiving investigational drugs or nucleoside analogues (such as 5-Fluorouracil that could interfere with FLT uptake). All patients were scheduled to undergo a baseline FLT PET scan within 30 days of the initiation of CRT. This was generally done the week before starting treatment. Platinum-based chemotherapy was started the first day of radiotherapy, either with high-dose cisplatinum or a combination of cisplatinum or carboplatinum combined with a taxane. Patients were followed every 3 months with clinical exams for the first year per our clinical routine and 2–4 times per year subsequently. Surveillance FDG PET scans were obtained at 3–4 months after treatment. Subsequent follow-up imaging was individualized on the basis of symptoms and clinical findings. This research was approved by the University of Iowa Institutional Review Board, and all subjects signed an informed consent. The research was conducted according to the principles of the Declaration of Helsinki and Good Clinical Practice.

In total, 30 patients with squamous cell head and neck cancer, including 27 oropharyngeal cancers, 1 unknown primary, and 2 laryngeal cancers, were available for analysis. There were 26 male and 4 female patients with an age range of 36–76

Table 1. Overview of Patients in the FLT PET Study (n = 30)

Patient Characteristics	Categories	Total [%]	Median [Range]
Age at diagnosis (years)			57 [36–76]
Sex	Male	26 [86.7]	
	Female	4 [13.3]	
Site	Oropharynx	27 [90.0]	
	Larynx	2 [6.7]	
	Unknown primary	1 [3.3]	
T-Stage	Tx	1 [3.3]	
	T1	1 [3.3]	
	T2	15 [50.0]	
	T3	7 [23.3]	
	T4	6 [20.0]	
N-Stage	N0	5 [16.7]	
	N1	5 [16.7]	
	N2	16 [53.3]	
	N3	4 [13.3]	
Overall Stage	II	2 [6.7]	
	III	9 [30.0]	
	IVA	13 [43.3]	
	IVB	6 [20.0]	
Follow-Up (Months)			22.0 [4.6–36.0]
Survival Status	Progression-free survival	21 [70]	
	Progression or death	9^a[30]	

^{a} Consists of 4 patients with LR, 4 patients with DM, and 1 patient with LR + DM.

years (median, 57 years). The demographics of the patients including distribution of tumor stages are summarized in Table 1. After a median follow-up of 26 months (range, 7–36 months), 8 patients died of disease, 1 patient was alive with distant metastasis (DM), and 21 patients had no evidence of disease. Among the 8 patients who died from the disease, 4 patients had local recurrence (LR), 1 patient had local recurrence and distant metastasis (LR + DM), and 3 patients had DM alone at the time of initial recurrence or progression. Three patients underwent salvage surgery after completion of radiotherapy because of local recurrence and had no evidence of disease at last follow-up. The median follow-up in patients with no evidence of disease was 25 months.

FLT PET Imaging

For the synthesis of FLT, fluorine-18 fluoride was reacted with 3′-anhydrothymidine-5′-benzoate following the procedure of

Machulla et al. (10). The benzoate protecting group was removed with base hydrolysis and the product purified by semiprep HPLC with 10% ethanol/90% isotonic saline as the mobile phase with typical yields of 5%–8%. FLT was infused via a syringe pump over 2 minutes followed by 10-mL saline flush administered manually. The administered activity of FLT was 2.6 MBq/kg (0.07 mCi/kg) with a maximum dose of 185 MBq (5 mCi). Imaging was performed on a Siemens ECAT EXACT HR + PET scanner (Siemens Medical Solutions USA, Inc., Knoxville, TN) for 40 minutes, starting 60 minutes after injection. Transmission imaging was performed before the injection of FLT. Whole-body scans were obtained for 28 patients, and scans of the head and neck region were obtained for only 2 patients. Images were iteratively reconstructed (2 iterations = 8 subsets, Gaussian 8.0 mm, zoom = 1.2) with a resulting voxel size of 4.29 × 4.29 × 4.29 mm.

Image Analysis

For primary tumors and FLT-avid lymph nodes, volumes of interest (VOIs) defined by high FLT uptake above background were generated by a nuclear medicine physician using a semi-automated segmentation software developed for head and neck tumors in PET (11). Primary tumors were segmented on FLT PET in all patients except for 1 patient who had an unknown primary tumor site. FLT-avid nodal metastases in the neck were identified in 23 patients. In total, 83 lesions/VOIs were identified using the semiautomated PET segmentation tool. Each VOI received an individual label. Subsequently, these labels were used to define 2 different measurement region categories (ie, VOIs) from which radiomic features were extracted. The first measurement region category PT consisted of VOIs representing primary tumor only. The second category LB was the total lesion burden, which corresponds to the primary tumor and all FLT-avid nodes combined. To calculate quantitative features for LB, all lesion previously segmented in a FLT scan were combined into 1 image mask, forming a single VOI. For each measurement region, radiomic features describing intensity, shape, and texture properties were calculated by using the open-source packages PET-IndiC (12) and pyradiomics (13). All features were derived from standardized uptake value (SUV) normalized PET images. A total of 104 quantitative baseline PT features and an additional 99 baseline LB features were extracted from each patient. Note that 5 shape features (ie, slice maximum 2D diameter, column maximum 2D diameter, row maximum 2D diameter, maximum 3D diameter, and sphericity) are meant for single, connected VOIs, so these were excluded from the LB features.

For texture features, the histogram bin size was fixed at 0.25 SUV. The selected bin size follows van Velden et al. (14), where the total number of bins will be ~64 bins, depending on the lesion SUV range. A fixed bin size is used rather than a fixed number of bins because lesion SUV ranges vary among patients and fixing the number of bins is less appropriate for the clinical setting (15).

In addition to SUV-based measurements, normalized measurements of uptake were calculated by dividing lesion SUVs by the mean SUV in the bone marrow. The goal of normalization is to compare the cell proliferation in cancerous tissue to that of a normal structure. Normalization of SUVs was accomplished by

generating a VOI around the largest vertebra completely visible in the field of view using the same segmentation software described above. In total, 30 vertebral VOIs were created using the semiautomated segmentation tool. For most patients, the L5 vertebra was segmented. The L4 vertebra was segmented for 1 patient owing to the L5 vertebra not being completely within the field of view. Because 2 patients had PET scans that did not include lumbar vertebrae, the T4 or T6 vertebra were segmented instead. Patient SUVs were then normalized by dividing by the mean vertebral SUV, and radiomic features were again calculated from the lesion VOIs. In total, 87 vertebra-normalized PT features and 87 vertebra-normalized LB features were generated from each patient. Note that normalization is not applicable for 11 features (ie, shape features) and has no effect on Q1–Q4 distributions, skewness, and kurtosis. For texture features based on normalized uptake, the histogram bin size was fixed at 0.125 (unitless). Note that the bin size is reduced compared with unnormalized texture features (bin size, 0.25), because normalization reduces the lesion intensity ranges compared with unnormalized lesions. In total, 377 baseline radiomic features were extracted from each patient.

Feature Reduction

Redundancy of quantitative features was reduced by using a clustering algorithm. The goal of feature reduction was to replace highly correlated features with a single representative feature. Such a step could be achieved by utilizing a PCA-based feature selection step [eg, FactoMineR (16)]. However, due to the sparseness of our feature space, a more appropriate feature selection method was utilized. First, the similarities of features were calculated by determining the Pearson correlation (r) for all pairs of features. Next, features were clustered according to similarity using an affinity propagation (AP) clustering algorithm (17), an unsupervised dimension reduction technique that others have utilized in the analysis of quantitative imaging features (18–20). An advantage of AP clustering over k-means clustering is that the total number of clusters at the output is automatically determined. Moreover, the algorithm is able to handle infinite dissimilarities, meaning 2 features that are highly dissimilar will not be placed in the same cluster. Therefore, to allow features with only strong correlations defined by $r \geq 0.90$ to be clustered together, all features with pairwise similarity values less than 0.90 were artificially set to have infinite dissimilarity before application of the AP clustering algorithm. As output, the algorithm produces a reduced set of representative exemplar features. An exemplar feature can be either a single feature with no strong correlations with other features or a representative of a cluster containing ≥2 features. The feature reduction step was performed using the *apcluster* package (21) in version 3.2.3 of the R statistical software (22).

Statistical Evaluation

Survival analysis was conducted to estimate and test the effects of quantitative features in the reduced set on progression-free survival (PFS). Time to event for PFS was defined as time from start of treatment to recurrence or death. Effects on survival during the 36-mo, post-treatment period were of primary interest. Hence, subjects who did not experience an event by month
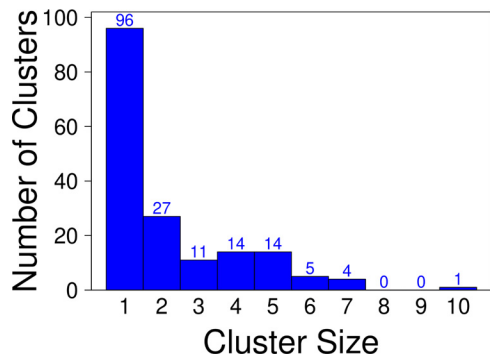
Figure 1. Cluster size distribution for the 172 clusters identified in the feature reduction step.

36 were censored at that point in time for the analysis. Cox regression was used to model the effects of individual quantitative features on survival. Using multiple predictors in a Cox regression model on a small cohort has the potential to overfit the patient data. Therefore, a Cox model with a single predictor was chosen to avoid overfitting. Estimated effects are summarized with hazard ratios (HRs) and the concordance (c)-index. The c-index is an estimate of the probability that, out of 2 randomly selected patients, the model can discriminate which patient will survive longer (23). Values can range from 0.0 to 1.0, with 0.5 indicating absence of discriminant value for the model, 0.7 indicating reasonable discriminant value, and 1.0 indicating perfect discriminant value. Two-sided P-values for tests of significance of features in the models are reported. To account for multiple statistical tests, the false-discovery rate (FDR) was computed using the

Benjamini–Hochberg method (24). Features with a FDR of 10% were identified as significant. All statistical tests were performed using the *survival* package (25) for R.

Interobserver Variability Analysis

To study the variabiltity of feature measurement, a second observer independently generated segmentations (VOIs) for the same 83 lesions and features were calculated as described above. The features extracted from the second observer's VOIs were then compared to the features extracted from the VOIs of the first observer. Agreement in feature measurement was compared using the intraclass correlation coefficient (ICC). To investigate the impact of interobserver segmentations on model performance, a separate model for each predictive feature was generated using segmentations by the second observer. The performance of these models were then compared to the initial models from the first observer. Differences of model performance were reported as changes in c-index values.

RESULTS
Feature Reduction

The feature reduction step took the 377 FLT features as input and clustered similar features together to produce 172 uncorrelated clusters. Figure 1 shows the distribution of cluster sizes. Ninety-six clusters had a size of one, meaning there were 96 features (25.5%) that were not highly correlated with any other feature ($r < 0.9$). The remaining 76 clusters had size of ≥2, with the maximum being a size of 10.

Correlation of Baseline Features With Treatment Outcome

Feature performance was estimated using each feature as a predictor in a univariate Cox regression model. A total of 37

Table 2. Comparison of Predictive FLT Features (Progression-Free Survival) With 3 Commonly Used Features, SUV_{max}, SUV_{peak}, and SUV_{mean}

| Feature (VOI, normalization)	P-Value	HR [95% CI]	FDR	c-Index
Gray-Level Non-Uniformity[a] (LB, N)	0.0002	3.11 [1.70, 5.68]	0.043	0.86
Gray-Level Non-Uniformity[b] (LB, N)	0.0012	3.12 [1.56, 6.24]	0.058	0.72
Spherical Disproportion (LB, U)	0.0012	4.10 [1.56, 10.80]	0.058	0.74
Information Measure of Correlation 2[c] (LB, U)	0.0017	0.32 [0.16, 0.65]	0.058	0.79
Zone Percentage[b] (LB, N)	0.0020	0.18 [0.04, 0.78]	0.058	0.75
Gray-Level Non-Uniformity[a] (LB, U)	0.0020	2.21 [1.40, 3.47]	0.058	0.83
Q1 Distribution (LB, U)	0.0042	0.36 [0.17, 0.75]	0.088	0.78
Volume (LB, U)	0.0043	2.44 [1.38, 4.32]	0.088	0.74
Information Measure of Correlation 1[c] (LB, U)	0.0046	4.07 [1.23, 13.42]	0.088	0.78
SUV_{max} (LB, U)	0.1916	0.60 [0.27, 1.33]	0.395	0.66
SUV_{peak}[d] (LB, U)	0.3341	0.69 [0.32, 1.48]	—	0.63
SUV_{mean}[d] (LB, U)	0.5038	0.76 [0.34, 1.71]	—	0.62

Abbreviations: VOI, volume of interest; HR, hazard ratio; CI, confidence interval; FDR, false-discovery rate; PT, primary tumor; LB, lesion burden; U, unnormalized; N, normalized.
[a] Calculated from the gray-level run length matrix (GLRLM).
[b] Calculated from the gray-level size zone matrix (GLSZM).
[c] Calculated from the gray-level co-occurrence matrix (GLCM).
[d] Not selected in feature reduction step, so FDR was not calculated.
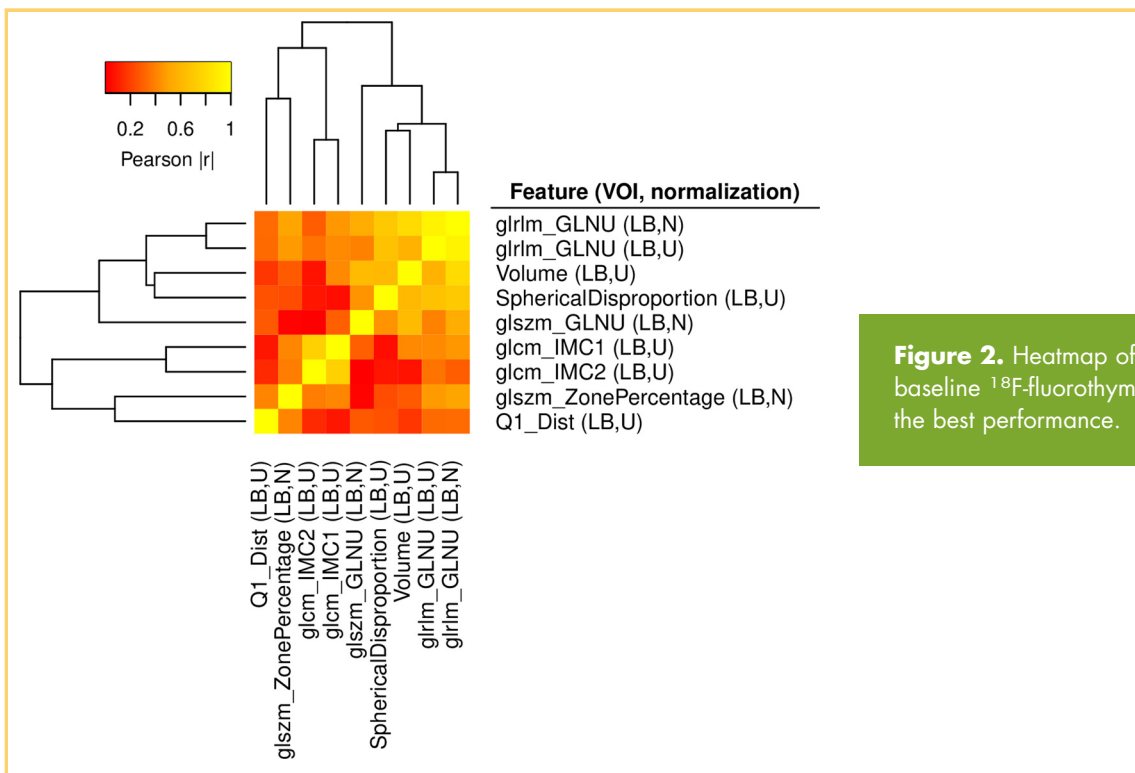
Figure 2. Heatmap of correlations among the 9 baseline ^{18}F-fluorothymidine (FLT) features with the best performance.

exemplar baseline features (21.5%) had *P*-values below the 5% level. After adjusting for multiple testing to control the false-discovery rate, a total of 9 baseline features were identified as significant at the set 10% FDR level. Table 2 summarizes the unadjusted Cox regression *P*-values, estimated hazard ratios with corresponding confidence intervals, FDRs, and *c*-index values for the 9 significant features as well as for 3 commonly used features (ie, SUV_{max}, SUV_{peak}, and SUV_{mean}). SUV_{peak} was defined as the highest average uptake within a 1 cm^3 sphere that is completely contained within the VOI. Note that SUV_{peak} and SUV_{mean} were not selected in the feature reduction step, so univariate analyses were done separately and no FDRs were calculated for SUV_{peak} and SUV_{mean}. The clinical parameter for primary tumor stage (T-stage) was not significantly associated with survival.

Figure 2 shows a heatmap of correlations among the 9 significant features. The feature reduction step used a high correlation threshold ($r \geq 0.90$), so moderate correlations among the best-performing features still exist. By showing the correlations of the features in a heatmap, good-performing lesion characteristics, rather than individual features, may be observed. For example, features that measure lesion size (eg, volume) and shape (eg, spherical disproportion) had good performance. Also, measures of lesion heterogeneity (eg, gray-level nonuniformity and zone percentage) had good performance.

Interobserver Variability Analysis

Table 3 shows the results of the variability analysis for the 9 significant features and the commonly used features SUV_{max}, SUV_{peak}, and SUV_{mean}. Gray-level nonuniformity from the gray-level size zone matrix (GLSZM) had moderate agreement between the 2 observers. The other 8 significant features had strong agreement between the 2 observers. Both SUV_{max} and

SUV_{peak} had perfect agreement between the 2 observers and SUV_{mean} had strong agreement.

To assess model performance stability, the segmentations of the second observer were used to produce a second model for

Table 3. Interobserver Agreement for Predictive FLT Features and 3 Commonly Used Features, SUV_{max}, SUV_{peak}, and SUV_{mean}

Feature (VOI, normalization)	Measurement Agreement
Gray-level Non-Uniformity[a] (LB, N)	0.99
Gray-level Non-Uniformity[b] (LB, N)	0.75
Spherical Disproportion (LB, U)	0.96
Information Measure of Correlation 2[c] (LB, U)	0.98
Zone Percentage[b] (LB, N)	0.91
Gray-level Non-Uniformity[a] (LB, U)	0.99
Q1 Distribution (LB, U)	0.90
Volume (LB, U)	0.99
Information Measure of Correlation 1[c] (LB, U)	0.95
SUV_{max} (LB, U)	1.00
SUV_{peak} (LB, U)	1.00
SUV_{mean} (LB, U)	0.94

Measurement agreement was calculated as the Intraclass Correlation Coefficient (ICC) between the feature values of the first and second observer.

Abbreviations: VOI, volume of interest; LB, lesion burden; U, unnormalized; N, normalized.
[a] Calculated from the gray-level run length matrix (GLRLM).
[b] Calculated from the gray-level size zone matrix (GLSZM).
[c] Calculated from the gray-level co-occurrence matrix (GLCM).

Table 4. Differences of Model Performance Due to Interobserver Segmentation Variability

Feature (VOI, normalization)	Δc-index
Gray-Level Non-Uniformity[a] (LB, N)	0.00
Gray-Level Non-Uniformity[b] (LB, N)	−0.01
Spherical Disproportion (LB, U)	−0.03
Information Measure of Correlation 2[c] (LB, U)	0.03
Zone Percentage[b] (LB, N)	0.01
Gray-Level Non-Uniformity[a] (LB, U)	0.01
Q1 Distribution (LB, U)	−0.07
Volume (LB, U)	−0.01
Information Measure of Correlation 1[c] (LB, U)	0.03

Change Calculations are the Difference (Δ) of the c-Indices Between the Model of the First Observer and the Model of the Second Observer. Abbreviations: VOI, volume of interest; LB, lesion burden; U, unnormalized; N, normalized.
[a] Calculated from the gray-level run length matrix (GLRLM).
[b] Calculated from the gray-level size zone matrix (GLSZM).
[c] Calculated from the gray-level co-occurrence matrix (GLCM).

each of the predictive features shown in Table 2. Table 4 shows the performance differences of the second model in reference to the first model for the univariate predictors with the best performance. For most features, only small changes in performance (c-index) were observed, indicating that model performance was stable. Only 1 feature (Q1 distribution) had a change in c-index >5 percentage points.
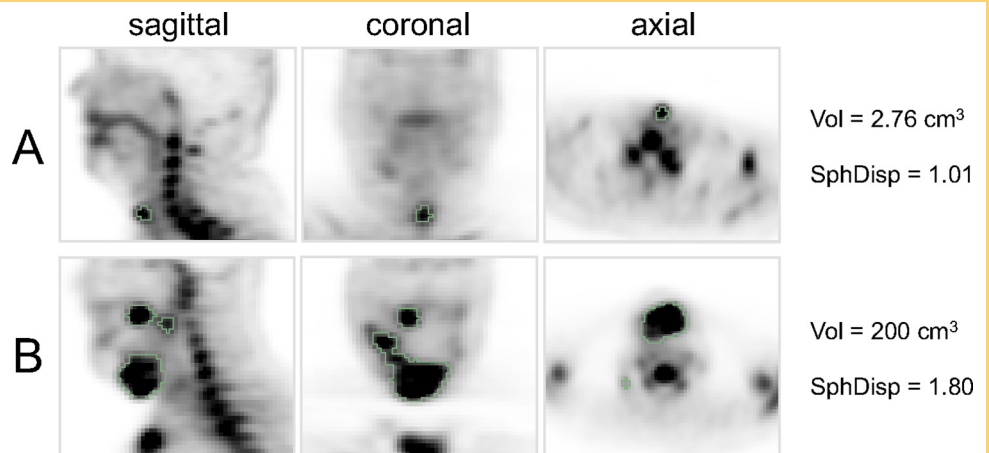
DISCUSSION
Performance
In this work, we investigated associations of patient outcomes with radiomic features derived from FLT PET lesion segmentations. Radiomics generates many features that can be highly correlated from each subject, so a feature reduction step was included to remove redundancies from the feature space. Despite this reduction, a large number of features were not highly correlated and tested in the performance analysis, so controlling the false-discovery rate was used to reduce false positives. A total of 9 FLT features were considered significant.
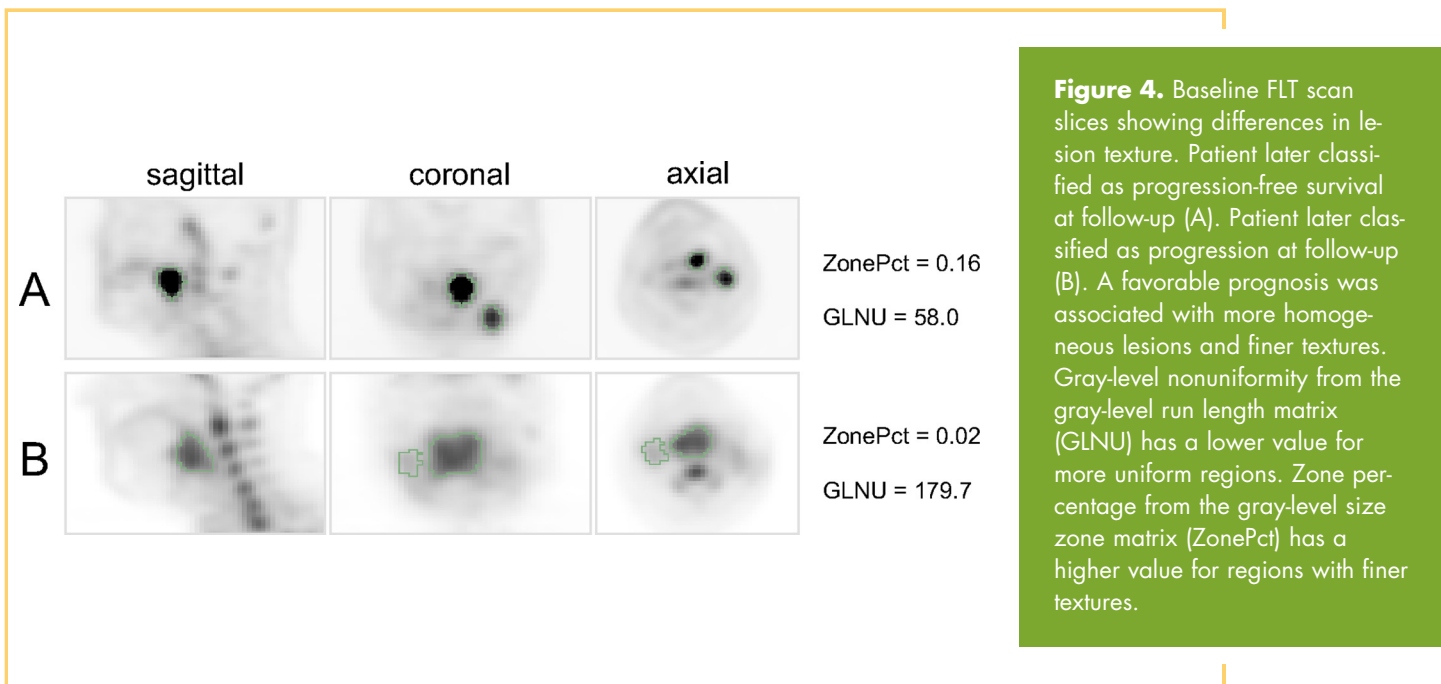
Our results suggest that a favorable prognosis is associated with a small lesion size, a more sphere-like lesion shape, and homogeneous intensity. Figure 3 shows the baseline scans of 2 patients with different outcomes and different FLT-avid lesion shapes. The surviving patient (Figure 3A) has a small, sphere-like lesion. The patient later classified with progressive disease (Figure 3B) has large lesions with a large, irregular surface area. Our results also suggest that lesion texture/homogeneity of intensity may be an indicator of outcome. Figure 4 shows the baseline scans of 2 patients with different outcomes and different lesion textures. The surviving patient (Figure 4A) has lesions with smaller regions of more uniform texture. The patient later classified with progressive disease (Figure 4B) has large regions and an overall nonuniform texture.

The authors are not aware of any publications that normalize lesion uptakes with the mean vertebral uptake before analysis of response prediction for HNSCC with FLT PET. Three out of the 5 best-performing intensity-based features were normalized with the mean vertebral uptake. Table 5 compares the c-indices of intensity-based FLT features with and without normalization. Texture features from the gray-level co-occurrence matrix have poorer performance after normalization. Texture features from the gray-level run length matrix and the GLSZM have a small increase in performance after normalization. Due to our small cohort of patients, more analysis is needed on a larger patient population to determine if these differences are significant.

All features identified as having an association with patient outcome were calculated from the total lesion burden (Table 2). This suggests that important information about the disease is found not only in the primary tumor, but also in the FLT-avid lymph nodes. Table 6 compares the c-indices of the 9 best-performing FLT features calculated from the primary tumor and the total lesion burden. All but 1 feature (ie, information measure of correlation 1) had higher performance when calculated from the total lesion burden. Furthermore, the interoperator agreement (ICC) average and standard deviation of the 9 best-performing FLT features for primary tumor and the total lesion burden was 0.88 ± 0.13 and 0.94 ± 0.08, respectively. Thus, FLT PET features derived from total lesion burden show higher agreement, and 8 out of the 9 best-performing features had strong agreement between different observers (Table 3). As

Figure 3. Baseline FLT scan slices showing differences in lesion size and shape. Patient later classified as progression-free survival at follow-up (A). Patient later classified as progression at follow-up (B). A favorable prognosis was associated with small tumor volume (Vol) and a lower spherical disproportion (SphDisp).

sagittal   coronal   axial

A   Vol = 2.76 cm³   SphDisp = 1.01

B   Vol = 200 cm³   SphDisp = 1.80

Figure 4. Baseline FLT scan slices showing differences in lesion texture. Patient later classified as progression-free survival at follow-up (A). Patient later classified as progression at follow-up (B). A favorable prognosis was associated with more homogeneous lesions and finer textures. Gray-level nonuniformity from the gray-level run length matrix (GLNU) has a lower value for more uniform regions. Zone percentage from the gray-level size zone matrix (ZonePct) has a higher value for regions with finer textures.

stated before, more analysis is needed on a larger patient population to determine if these differences are significant.

Related Work

The association of standard FLT features and outcome has been previously studied. For example, Hoshikawa et al. reported that baseline FLT tumor volume and total lesion proliferation (TLP) were predictive of locoregional tumor control in 32 patients with HNSCC treated with CRT and surgery (26). We found similar results to their findings for the total lesion burden volume ($P = .004$) and total lesion burden TLP ($P = .012$) for predicting 3-y progression–free survival. Note that total lesion burden TLP in our analysis was not selected during the feature reduction step. Hoshikawa et al. later reported that baseline FLT tumor volume, TLP, and SUV$_{max}$ were predictive of locoregional tumor control in 53 patients with HNSCC treated with RT or CRT (27). Our results are not similar to their findings for unnormalized SUV$_{max}$ ($P = .192$). This may be due to our smaller patient cohort (30 vs. 53). However, Linecker et al. reported earlier that

high FLT uptake is associated with poor outcome in 20 patients treated with RT and CRT (8).

The authors are aware of 2 other publications that report correlations of FLT based radiomic features and patient outcomes. Willaime et al. reported that radiomic features were predictive of treatment response in 11 breast cancer patients treated with chemotherapy (28). However, the different cancer site and treatment type does not allow for a meaningful comparison with our results. Majdoub et al. (29) reported that tumor proliferative volume and textural features are predictive of disease-free survival in 45 patients with HNSCC treated with RT and CRT. They found that large, more heterogeneous lesions

Table 5. Comparison of c-Index Values for Unnormalized and Normalized Features

Feature	Unnormalized	Normalized
Gray-Level Non-Uniformity[a]	0.83	**0.86**
Gray-Level Non-Uniformity[b]	0.66	**0.72**
Information Measure of Correlation 2[c]	**0.79**	0.63
Zone Percentage[b]	0.73	**0.75**
Information Measure of Correlation 1[c]	**0.78**	0.56

Higher c-Index Values for Each Feature are Indicated in Bold.
[a] Calculated from the gray-level run length matrix (GLRLM).
[b] Calculated from the gray-level size zone matrix (GLSZM).
[c] Calculated from the gray-level co-occurrence matrix (GLCM).

Table 6. Comparison of c-Index Values for Features Calculated from the Primary Tumor and the Total Lesion Burden

Feature (Normalization)	Primary Tumor	Lesion Burden
Gray-Level Non-Uniformity[a] (N)	0.71	**0.86**
Gray-Level Non-Uniformity[b] (N)	0.50	**0.72**
Spherical Disproportion (U)	0.49	**0.74**
Information Measure of Correlation 2[c] (U)	0.75	**0.79**
Zone Percentage[b] (N)	0.68	**0.75**
Gray-Level Non-Uniformity[a] (U)	0.71	**0.83**
Q1 Distribution (U)	0.64	**0.78**
Volume (U)	0.59	**0.74**
Information Measure of Correlation 1[c] (U)	**0.79**	0.78

Higher c-Index Values for Each Feature are Indicated in Bold.
Abbreviations: U, unnormalized; N, normalized.
[a] Calculated from the gray-level run length matrix (GLRLM).
[b] Calculated from the gray-level size zone matrix (GLSZM).
[c] Calculated from the gray-level co-occurrence matrix (GLCM).

were associated with a less favorable prognosis, which is consistent with our findings.

In current clinical practice, FDG PET imaging is commonly utilized for assessment of response to treatment. For this purpose, simple quantitative image features like SUVmax, SUVpeak (30), metabolic tumor volume (MTV), or total lesion glycolysis (TLG) have been proposed, out of which SUVmax is most widely adopted. In a recent study, Castelli et al. (31) summarized the results of 45 studies regarding the predictive value of such FDG PET features with respect to clinical outcome in HNC treatment with chemoradiotherapy (CRT). The study concluded that MTV and TLG in pretreatment PET scans showed good correlation with disease free survival (DFS) or overall survival (OS). In this work, we have investigated FLT PET derived image features. At this stage, it is unclear which imaging approach (ie, tracer) results in better predictive performance. For example, the volume defined by above normal tracer uptake showed good performance on FLT data (Table 2) as well as in FDG PET studies (31). However, to decide which approach is preferable, a dedicated study is needed.

Limitations

This study has several limitations. The HPV (human papilloma virus) status, which is now a well-known prognostic factor in oropharyngeal cancers, was not available for this cohort as it was not routinely obtained when subjects were enrolled in this study. Furthermore, the effects of repeated scans and image reconstruction parameters on FLT-based radiomic features was not determined. Willaime et al. did investigate test–retest variability of texture features in breast cancer using FLT PET (28). They report similar results to a study by Tixier et al., which investigated the test-retest variability of FDG PET texture features using 16 patients with esophageal cancer (32). Both studies

found that measures of tumor homogeneity and entropy had good repeatability. Leijenaar et al. investigated the repeatability of FDG PET texture features in non–small cell lung cancer (33). A majority of features (71%) were stable during test-retest analysis.

Yan et al. reported that zone percentage of the GLSZM was sensitive to image reconstruction parameters and should be used with caution (34). Their work used 20 patients with lung lesions imaged with FDG PET. Zone percentage was associated with patient outcome in our results, and it is a measure of fine textures. It is reasonable to expect that high variability of zone percentage calculations by different image reconstruction parameters would also occur in FLT PET. Reconstruction parameters were held constant for the images in our study.

CONCLUSION

In conclusion, radiomics is a useful approach for extracting large amounts of information from tumor images. We investigated the association of patient outcomes with radiomic features extracted from tumors imaged with FLT PET. Radiomics features performed favorably compared to standard clinical stage. We found that smaller, more homogenous lesions at baseline were associated with a better prognosis in 30 patients with head and neck cancer. Therefore, for future studies of FLT-based prediction of outcome, we recommend including radiomic features of lesion size, shape, and texture features that measure lesion homogeneity. We also recommend that radiomic features be calculated from the total lesion burden, rather than the primary tumor only, so that the largest amount of disease information is used for analysis. Our findings enable future optimization of FLT-based features which can then be assessed in validation studies.

REFERENCES

1. Ang KK, Zhang Q, Rosenthal DI, Nguyen-Tan PF, Sherman EJ, Weber RS, Galvin JM, Bonner JA, Harris J, El-Naggar AK, Gillison ML, Jordan RC, Konski AA, Thorstad WL, Trotti A, Beitler JJ, Garden AS, Spanos WJ, Yom SS, Axelrod RS. Randomized phase III trial of concurrent accelerated radiation plus cisplatin with or without cetuximab for stage III to IV head and neck carcinoma: RTOG 0522. J Clin Oncol. 2014;32:2940–2950.

2. Goodwin JWJ. Salvage surgery for patients with recurrent squamous cell carcinoma of the upper aerodigestive tract: when do the ends justify the means? Laryngoscope. 2000;110(3 Pt 2 Suppl 93):1–18.

3. Sadowski SM, Neychev V, Millo C, Shih J, Nilubol N, Herscovitch P, Pacak K, Marx SJ, Kebebew E. Prospective study of 68Ga-DOTATATE positron emission tomography/computed tomography for detecting gastro-entero-pancreatic neuroendocrine tumors and unknown primary sites. J Clin Oncol. 2016;34:588–596.

4. Rasey JS, Grierson JR, Wiens LW, Kolb PD, Schwartz JL. Validation of FLT uptake as a measure of thymidine kinase-1 activity in A549 carcinoma cells. J Nucl Med. 2002;43:1210–1217.

5. Cobben DC, van der Laan BF, Maas B, Vaalburg W, Suurmeijer AJ, Hoekstra HJ, Jager PL, Elsinga PH. 18F-FLT PET for visualization of laryngeal cancer: comparison with 18F-FDG PET. J Nucl Med. 2004;45:226–231.

6. Hoshikawa H, Kishino T, Mori T, Nishiyama Y, Yamamoto Y, Inamoto R, Akiyama K, Mori N. Comparison of 18F-FLT PET and 18F-FDG PET for detection of cervical lymph node metastases in head and neck cancers. Acta Otolaryngol. 2012;132:1347–1354.

7. Hoshikawa H, Nishiyama Y, Kishino T, Yamamoto Y, Haba R, Mori N. Comparison of FLT-PET and FDG-PET for visualization of head and neck squamous cell cancers. Mol Imaging Biol. 2011;13:172–177.

8. Linecker A, Kermer C, Sulzbacher I, Angelberger P, Kletter K, Dudczak R, Ewers R, Becherer A. Uptake of 18F-FLT and 18F-FDG in primary head and neck cancer correlates with survival. Nuklearmedizin. 2008;47:80–85; quiz N12.

9. Menda Y, Boles Ponto LL, Dornfeld KJ, Tewson TJ, Watkins GL, Schultz MK, Sunderland JJ, Graham MM, Buatti JM. Kinetic analysis of 3′-deoxy-3′-18F-fluorothymidine (18F-FLT) in head and neck cancer patients before and early after initiation of chemoradiation therapy. J Nucl Med. 2009;50:1028–1035.

10. Machulla HJ, Blocher A, Kuntzsch M, Piert M, Wei R, Grierson JR. Simplified labeling approach for synthesizing 3′-deoxy-3′-[18F]Fluorothymidine ([18F]FLT. J Radioanal and Nucl Chem. 2000;243:843–846.

11. Beichel RR, van Tol M, Ulrich EJ, Bauer C, Chang T, Plichta KA, Smith BJ, Sunderland JJ, Graham MM, Sonka M, Buatti JM. Semiautomated segmentation of head

and neck cancers in 18F-FDG PET scans: a just-enough-interaction approach. Med Phys. 2016;43:2948–2964.

12. Ulrich EJ, van Tol M, Bauer C, Fedorov A, Beichel RR, Buatti JM. PET-IndiC extension documentation. In: 3D Slicer Wiki Internet. Available from: https://www.slicer.org/wiki/Documentation/Nightly/Extensions/PET-IndiC.

13. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77:e104–e107.

14. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, Hoekstra OS, Smit EF, Boellaard R. Repeatability of radiomic features in non-small-cell lung cancer 18F]FDG-PET/CT studies: impact of reconstruction and delineation. Mol Imaging Biol. 2016;18:788–795.

15. Leijenaar RT, Nalbantov G, Carvalho S, Van Elmpt WJ, Troost EG, Boellaard R, Aerts HJ, Gillies RJ, Lambin P. The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep. 2015;5:11075.

16. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw. 2008;25:1–18.

17. Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007;315:972–976.

18. Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, Decker PA, Robb RA. Quantitative CT imaging of interstitial lung diseases. J Thoracic Imaging. 2013;28.

19. Maldonado F, Boland JM, Raghunath S, Aubry MC, Bartholmai BJ, Hartman TE, Karwoski RA, Rajagopalan S, Sykes AM, Yang P. Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)—a pilot study. J Thoracic Oncol. 2013;8:452–460.

20. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, Ji Y. Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. Sci Rep. 2015;5:17787.

21. Bodenhofer U, Kothmeier A, Hochreiter S. APCluster: an R package for affinity propagation clustering. Bioinformatics. 2011;27:2463–2464.

22. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: https://www.R-project.org/.

23. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247:2543–2546.

24. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300.

25. Therneau TM, Lumley T. Package 'survival' version 2.42-6 [Internet]; 2018 July 13. Available from: https://cran.r-project.org/package=survival.

26. Hoshikawa H, Yamamoto Y, Mori T, Kishino T, Fukumura T, Samukawa Y, Mori N, Nishiyama Y. Predictive value of SUV-based parameters derived from pretreatment 18F-FLT PET/CT for short-term outcome with head and neck cancers. Ann Nucl Med. 2014;28:1020–1026.

27. Hoshikawa H, Mori T, Yamamoto Y, Kishino T, Fukumura T, Samukawa Y, Mori N, Nishiyama Y. Prognostic value comparison between (18)F-FLT PET/CT and (18)F-FDG PET/CT volume-based metabolic parameters in patients with head and neck cancer. Clin Nucl Med. 2015;40:464–468.

28. Willaime JMY, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. Phys Med Biol. 2012;58:187.

29. Majdoub M, Visvikis D, Texier F, Hoeben B, Visser E, Cheze-Le Rest C, Hatt M. Proliferative 18F-FLT PET tumor volumes characterization for prediction of locoregional recurrence and disease-free survival in head and neck cancer. In: SNMMI 2013: Society of Nuclear Medicine and Molecular Imaging Annual Meeting. Vancouver, Canada; 2013. Available from: https://hal.archives-ouvertes.fr/hal-00936229.

30. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1:122S–50S.

31. Castelli J, De Bari B, Depeursinge A, Simon A, Devillers A, Roman Jimenez G, Prior J, Ozsahin M, de Crevoisier R, Bourhis J. Overview of the predictive value of quantitative 18 FDG PET in head and neck cancer treated with chemoradiotherapy. Crit Rev Oncol Hematol. 2016;108:40–51.

32. Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med. 2012;53:693.

33. Leijenaar RT, Carvalho S, Velazquez ER, Van Elmpt WJ, Parmar C, Hoekstra OS, Hoekstra CJ, Boellaard R, Dekker ALAJ, Gillies RJ, Aerts HJWL, Lambin P. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol. 2013;52:1391–1397.

34. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, Tham IW, Townsend D. Impact of image reconstruction settings on texture features in 18F-FDG PET. J Nucl Med. 2015;56:1667–1673.