Quantitative PET Imaging and Clinical Parameters as Predictive Factors for Patients With Cervical Carcinoma: Implications of a Prediction Model Generated Using Multi-Objective Support Vector Machine Learning

Zhiguo Zhou, PhD[1] 🌐, Genevieve M. Maquilan, MD[2], Kimberly Thomas, MD, MSc[2], Jason Wachsmann, MD[3], Jing Wang, PhD[2], Michael R. Folkert, MD, PhD[2], and Kevin Albuquerque, MD, MS, FRCS[2]

Abstract

Purpose: Quantitative features from pre-treatment positron emission tomography (PET) have been used to predict treatment outcomes for patients with cervical carcinoma. The purpose of this study is to use quantitative PET imaging features and clinical parameters to construct a multi-objective machine learning predictive model. **Materials/Methods:** Seventy-five patients with stage IB2-IVA disease treated at our institution from 2009–2012 were analyzed. Models predicting locoregional and distant failure were generated using clinical parameters (age, race, stage, histology, tumor size, nodal status) and imaging features (12 textural, 9 intensity, 8 geometric features, 2 additional imaging features) from pre-treatment PET. Model features were selected based on a multi-objective evolutionary algorithm to maximize specificity given a fixed moderately high sensitivity using support vector machine learning methods. Model 1 used clinical parameters only (C), Model 2 used imaging features only (I), and Model 3 used clinical and imaging features (C+I). Sensitivity, specificity, area under a receiver-operating characteristic curve (AUC), and p-values were compared to assess ability to predict locoregional and distant failure. **Results:** C+I had the highest performance for both locoregional failure (AUC 0.84, p < 0.01; specificity: 0.86; sensitivity: 0.79) and distant failure (AUC 0.75, p < 0.01; specificity: 0.75; sensitivity: 0.75). **Conclusions:** Based on a moderately high fixed sensitivity and optimized for specificity, the model using both clinical parameters and imaging features (C+I) had the best performance in predicting both locoregional failure and distant failure.

Keywords

cervical carcinoma, radiomics, multi-objective model, PET, clinical parameters

Introduction

[18F] fluoro-2-deoxy-D-glucose (FDG) based positron emission tomography (PET) imaging has become increasingly utilized for radiation therapy treatment planning[1] and to characterize metabolic aspects of the target tumor.[2-4] Allal *et al* demonstrated the prognostic value of standardized uptake value (SUV) from pre-treatment PET in patients with head and neck cancer treated with radiation with or without chemotherapy.[5] Patients with tumors with high tumor SUV max had a significantly lower 3-year local

[1] School of Computer Science and Mathematics, University of Central Missouri, MO, USA
[2] Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, USA
[3] Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX, USA

Corresponding Authors:
Jing Wang and Kevin Albuquerque, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75235, USA.
Emails: Jing.Wang@UTSouthwestern.edu; Kevin.Albuquerque@UTSouthwestern.edu

control (55% vs. 86%, p = 0.01) and disease-free survival (42% vs. 79%, p = 0.005) compared to patients with tumors with low FDG uptake. In the setting of cervical cancer, 287 patients with Stage IA2-IVB disease who underwent pre-treatment PET followed by surgery, chemoradiation, or palliation were assessed for SUVmax, tumor volume, sites of lymph node metastases, and histology, and only SUVmax was identified as an independent predictor of death from cervical cancer.[6]

Development of a range of complex image analytics has led to the expansion of the field of "radiomics," which utilizes data beyond the single value represented by SUVmax, a value limited by patient-specific and image-acquisition-specific factors.[7,8] Radiomics is a method of quantitative data extraction from radiographic imaging to find image correlates to tumor characteristics.[9,10] A potential application is to identify complex image intensity, shape, and textural features that predict a tumor's behavior, response to treatment, and oncologic outcomes, but very few multi-parametric studies have looked at these relationships.[11]

In cervical carcinoma, even with optimal therapy, at least 20% of patients with locally advanced disease confined to the pelvis will fail distantly.[12,13] Identifying these patients early may allow physicians to tailor their treatment to achieve a more durable treatment response and prevent distant failure with additional systemic treatment. The goal of this study is to build a predictive model, using pre-treatment clinical and imaging characteristics, that determines likelihood of locoregional and distant failure for cervical carcinoma patients, enabling the selection of patients who should be considered for further systemic therapy. While radiomics analyses have been explored for treatment outcome prediction for cervical carcinoma after radiation or chemoradiation therapy,[14-19] these analyses often focus on some individual radiomic features or use a single objective during the model training. In this work, we present a multi-objective model to predict distant failure and locoregional failure for cervical carcinoma patients.

Methods and Materials

Patients and Clinical Parameters

Following institutional IRB approval at UT Southwestern Medical Center (approval no. 082013-008), departmental records were reviewed to identify patients treated for cervical carcinoma with definitive intent between 2009 and 2012, allowing time for follow-up. Because this is retrospective review study, informed consent was waived. Patients with stage IB2-IVA disease treated with definitive chemoradiation and high dose rate (HDR) intracavitary brachytherapy (without outback chemotherapy), with complete clinical data and retrievable pre-treatment PET/CT scans were identified (n = 75). A retrospective analysis of clinical parameters, pre-treatment PET/CT imaging characteristics and features, and oncologic outcomes for these patients was performed.

Table 1. Patient Characteristics.

Number	75 (100%)
Mean age (range in years)	46.9 (26.2-72.1)
Race	
African American	22 (29%)
Hispanic	27 (36%)
White	23 (31%)
Asian	2 (3%)
Other	1 (1%)
Histology	
Squamous cell carcinoma	63 (84%)
Adenocarcinoma	9 (12%)
Adenosquamous carcinoma	2 (3%)
Other	1 (1%)
FIGO stage	
IB2	21 (28%)
IIA	4 (5%)
IIB	31 (41%)
IIIB	13 (17%)
IVA	6 (8%)
IVB	0 (0%)

These 75 patients (characteristics described in Table 1) were used to build the locoregional and distant failure prediction models. Clinical parameters (age, race, stage, histology, tumor size, and nodal status at diagnosis) were obtained from chart review.

Events were defined as follows: local failure (LF) includes failure in the area receiving high-dose treatment, including cervix, pelvic side wall, parametria, vagina; regional failure (RF) includes failures occurring in areas receiving external beam alone, including pelvic lymph nodes; distant failure (DF) includes distant metastases, including para-aortic lymph nodes (unless included in the treatment field); and locoregional failure (LRF) is any combination of LF and RF events. Only LRF and DF were analyzed as outcomes for the model; the models were designed to predict for these outcomes.

PET/CT images were acquired with a Siemens Biograph 64 (Siemens Medical Solutions USA, Inc. Malvern, PA USA) with 4 detector rings, a spatial resolution of 7-8 mm, and a slice thickness of 5 mm. Segmentation of patients' primary tumor was performed on pre-treatment PET and CT imaging using the imaging informatics system Velocity (Varian, Palo Alto, CA). The clinical target volume (CTV), including the anatomical cervix and PET-positive extension of tumor, was contoured by the clinician investigators using SUV-based thresholding on the primary cervical lesion to include all the PET-avid areas with an SUVmax value of ≥ 4 into the CTV (excluding the bladder) for purposes of subsequent analysis. SUV-based intensity metrics were calculated within the edited ROI, and additional image features (referred to as texture and geometry features) were extracted. For intensity features, the mean, median, standard deviation, maximum and minimum value, skewness, kurtosis, and variance were calculated based on the intensity histogram. Before extracting

Table 2. List of Clinical and Imaging Features.

Clinical features (6)	Texture features (12)	Intensity features (9)	Geometric features (8)	Additional imaging features (2)
Age	Energy	SUV Max	Volume	MTV
Race	Entropy	SUV Min	Major Axis Length	TLG
Stage	Correlation	SUV Mean	Minor Axis Length	
Histology	Contrast	SUV Median	Eccentricity	
Tumor Size	Variance	SUV Standard Deviation	Elongation	
Nodal Status	Sum Mean	SUV Variance	Orientation	
	Inertia	SUV Sum	V Bound	
	Cluster Shade	SUV Skewness	Perimeter	
	Cluster Tendency	SUV Kurtosis		
	Homogeneity			
	Max Probability			
	Inverse Variance			

MTV = metabolic tumor volume. TLG = total lesion glycolysis.

the texture features, a gray level co-occurrence matrix (GLCM) was constructed, using histograms with 64 bins and 3D analysis of the tumor region with 26 neighboring voxels and 13 directions in 3D space. Construction of this GLCM allowed 12 texture features to be extracted. Geometry features (a description of the shape, size, or the relative position of the tumor), metabolic tumor volume (MTV), which is defined as the volume of tumor having at least 40% of max SUV, and total lesion glycolysis (TLG), which is defined as TLG = MTV*mean SUV, were obtained. The complete list of clinical and imaging features used in this study is in Table 2.

Multi-Objective Predictive Model Construction

In most radiomics studies, predictive models are constructed based on a single objective such as overall accuracy or AUC.[11] However, overall accuracy alone may not be a good measure for the predictive models, which can lead to low sensitivity or specificity when positive and negative events are imbalanced in training datasets.[20] Although AUC provides a better measure than overall accuracy by taking both sensitivity and specificity into account, it can be a misleading measure of the predictive model performance.[21-24] Lobo et al summarized 5 drawbacks of AUC measure as follows: "(1) it ignores the predicted probability values and the goodness-of-fit of the model; (2) it summarizes the test performance over regions of the ROC space in which one would rarely operate; (3) it weights omission (falsely predicted positive fraction) and commission errors (falsely predicted negative fraction) equally; (4) it does not give information about the spatial distribution of model errors; and, most importantly, (5) the total extent to which models are carried out highly influences the rate of well-predicted absences and the AUC scores."[21] To overcome the limitations of the conventional single-objective based models, a multi-objective radiomics model was designed to train the predictive model, where both sensitivity and specificity are considered as the objective functions

simultaneously. Assume that sensitivity and specificity are denoted by f_{sen}, f_{spe}, respectively, that is:

$$f_{sen} = \frac{TP}{TP + FN}, \tag{1}$$

$$f_{spe} = \frac{TN}{TN + FP}, \tag{2}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. The goal of the proposed model is to maximize f_{sen} and f_{spe} simultaneously to get the Pareto-optimal solutions:

$$f = \max_{\alpha, \beta} \left(f_{sen}, f_{spe} \right) \tag{3}$$

where $\alpha = \{\alpha_1, \cdots, \alpha_N\}$ denotes the model parameters and $\beta = \{\beta_1, \cdots, \beta_N\}$ denotes the input features. In this work, a support vector machine (SVM) was used to build the predictive model.[25] Using an iterative multi-objective immune algorithm (IMIA),[25] the predictive model is optimized through both feature selection and model parameter optimization. From the Pareto-optimal solutions, the final predictive model was selected to have the highest specificity with a minimum sensitivity 0.75. Five-fold cross validation (80% of patients to train, 20% of patients to test)[20,26] was used to validate the model. Patients were grouped into the training and testing subsets randomly.

To systematically investigate the influence of the input of different features, 3 versions of the models were built to predict each of the 2 primary outcomes, LRF and DF (total of 6 models). The first model used clinical parameters only (age, race, stage, histology, tumor size, and nodal status at diagnosis); the second model used imaging features only (including intensity, texture, geometric features, and an expanded set of imaging features that included MTV and TLG); the third model used a combination of clinical and all imaging parameters. Not every clinical and imaging parameter was significant for each outcome; the models were built on all available features and the

Table 3. Model Performance.

Locoregional failure

Model	Sensitivity	Specificity	AUC	95%CI
C	0.75 ± 0.03	0.75 ± 0.01	0.80 ± 0.01	[0.55, 0.94]
I	0.79 ± 0.01	0.86 ± 0.03	0.84 ± 0.02	[0.66, 0.95]
C+I	0.80 ± 0.02	0.86 ± 0.02	0.84 ± 0.02	[0.69, 0.96]

Distant Failure

Model	Sensitivity	Specificity	AUC	95%CI
C	0.75 ± 0.02	0.73 ± 0.01	0.75 ± 0.01	[0.64, 0.86]
I	0.75 ± 0.01	0.75 ± 0.02	0.74 ± 0.01	[0.61, 0.88]
C+I	0.75 ± 0.01	0.75 ± 0.02	0.75 ± 0.03	[0.61, 0.87]

C = Model using clinical parameters only. I = Model using imaging features only. C+I = Model using clinical parameters and imaging features.

minimum optimal set was selected during the model optimization by IMIA.

Experimental Setup and Evaluation

This study used IBM SPSS Statistics Software version 24 (IBM, Armonk, NY) to perform correlation and survival analysis and to generate receiver-operating characteristic (ROC) curves. Sensitivity, specificity, AUC, and the p-value of the ROC function were compared for each model. All the experiments were run 10 time, and mean as well as standard deviation values are calculated. ROC curves were compared with the unpaired t-test at a significance level of 0.05. The model was built in MATLAB2019b. All the features are extracted full-automatically. Five-fold cross validation is performed for all the models. In our experiment, each fold in cross validation can be considered as held-off set in each test as this fold has never been seen by the model trained on the other 4 folds.

Results

Median follow-up time for the study population was 27.4 months (range: 3.4-83.5 months, 3 patients with <6 months follow up). Follow-up was short for some patients due to non-compliance. The median number of external beam radiation therapy fractions was 25 at a median dose per fraction of 180 cGy. Patients received a median number of 5 fractions of HDR intracavitary brachytherapy at a median dose per fraction of 600 cGy.

The sensitivity, specificity, area under the ROC curve (AUC), and p value for each predictive model is listed in Table 3. Additionally, we assessed the ability of each clinical parameter and each imaging parameter to predict outcomes. These data are provided in Table 4. When compared to the 3 versions of the models for each outcome, the combined model using both clinical and imaging features as input outperformed the other models that used clinical or imaging features

Table 4. Individual Performance of Each Clinical and Imaging Feature.

Locoregional failure

	Spearman's rho (P value)	AUC (P value)	Hazard ratio (P value)
Age	-0.18 (0.12)	0.39 (0.12)	0.97 (0.16)
Race	0.03 (0.82)	0.52 (0.82)	0.99 (0.96)
Stage	0.18 (0.12)	0.61 (0.14)	1.38 (0.07)
Tumor Size	0.06 (0.62)	0.54 (0.62)	1.05 (0.66)
TLG	0.06 (0.59)	0.54 (0.58)	1.00 (0.79)
SUV_max	-0.07 (0.56)	0.46 (0.07)	0.98 (0.56)
Energy	0.09 (0.46)	0.55 (0.07)	0.00 (0.55)
Entropy	-0.08 (0.52)	0.45 (0.07)	0.87 (0.90)
Contrast	0.04 (0.74)	0.52 (0.07)	1.00 (0.63)
Variance	0.08 (0.48)	0.55 (0.07)	2.52 (0.81)
Max Probability	0.00 (0.99)	0.50 (0.07)	0.00 (0.37)
Inverse Variance	0.09 (0.45)	0.56 (0.07)	3.10 (0.78)

Distant Failure

	Spearman's Rho (P Value)	AUC (P Value)	Hazard Ratio (P Value)
Age	-0.17 (0.15)	0.39 (0.15)	0.98 (0.36)
Race	-0.16 (0.16)	0.40 (0.19)	0.64 (0.10)
Stage	0.09 (0.46)	0.55 (0.47)	1.23 (0.26)
Nodal Status	0.23 (0.05)	0.63 (0.07)	1.86 (0.05)
MTV	-0.03 (0.79)	0.48 (0.78)	1.00 (0.59)
Contrast	-0.07 (0.55)	0.46 (0.54)	1.00 (0.47)
Inertia	-0.06 (0.63)	0.46 (0.63)	0.96 (0.32)
Cluster Shade	-0.15 (0.20)	0.40 (0.19)	0.99 (0.19)
Homogeneity	-0.39 (0.00)	0.25 (0.00)	1.00 (0.00)
Max Probability	0.13 (0.29)	0.58 (0.28)	0.96 (0.84)
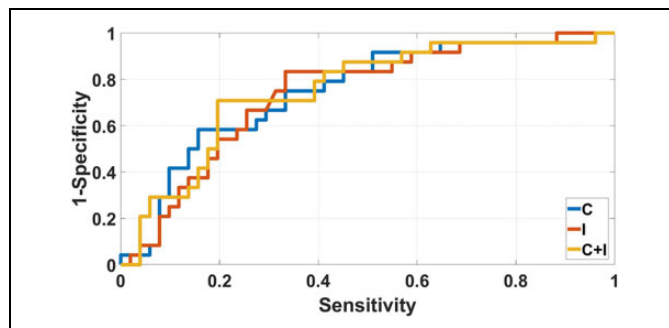
Figure 1. Receiver-operating characteristic (ROC) curves for the 4 models to predict for distant failure. Blue line = C = Model using clinical parameters only. Orange line = I = Model using imaging features only. Yellow line = C+I = Model using clinical parameters and imaging features.

alone (Figure 1). The combined model using both clinical and imaging features had excellent prognostic power for locoregional failure, with an AUC of 0.84 (p < 0.01) and specificity of 0.86 at a sensitivity of 0.80, and for distant failure, with an AUC of 0.75 (p < 0.01) and specificity of 0.75 at a sensitivity of 0.75. The combined model also outperformed all individual clinical and imaging parameters. Of note, as shown in Table 4,

Table 5. Selected Features and Importance Analysis of Individual Feature for Locoregional Failure Prediction.

Locoregional failure		Min	Max	AUC-min	AUC-max
C	Age	26	72	0.8103	0.8049
	Race	1	3	0.7832	0.7913
	Stage	0	5	0.7453	0.7832
I	SUV_min	0.0236	1.4979	0.8238	0.8022
	SUV_median	0.3592	20.4168	0.874	0.8293
	SUV_std	0.1312	11.2838	0.8347	0.8238
	SUV_kurtosis	1.5531	29.5791	0.8482	0.8022
	Energy	11.8221	11.9837	0.8564	0.8509
	Entropy	11.8221	11.9837	0.8022	0.8184
	Correlation	24.8516	1331.3048	0.8171	0.7602
	SumMean	0.3884	17.9764	0.8293	0.8401
	Inertia	12.0233	12.5036	0.8672	0.8753
	Cluster Shade	0.0767	1.7816	0.7317	0.7236
	Cluster tendency	2.2313	358.0951	0.8374	0.8753
	Inverse Variance	0.0031	0.0387	0.8469	0.8835
	Orientation	88.8512	87.6894	0.8618	0.8808
C+I	Stage	0	5	0.71	0.7019
	Histology	0	3	0.7832	0.8157
	SUV_median	0.3592	20.4168	0.8076	0.71
	SUV_sum	73.0215	68472.169	0.7154	0.748
	SUV_skewness	0.2611	4.4762	0.7317	0.7818
	SUV_kurtosis	1.5531	29.5791	0.7439	0.7425
	Energy	11.8221	11.9837	0.7371	0.7317
	Variance	11.9434	11.9961	0.7154	0.7751
	Cluster tendency	2.2313	358.0951	0.8103	0.7995
	Homogeneity	30.6399	8124.8148	0.748	0.7398
	V_Bound	34	8670	0.7588	0.7561
	TLG	73.0215	68472.17	0.7182	0.7669
	MTV	188	4759	0.7602	0.7724

AUC-min value and AUC-max value correspond to the results using the minimal or maximal value of the corresponding feature, respectively.

Table 6. Selected Features and Importance Analysis of Individual Feature for Distant Failure Prediction.

| Distant failure		Min	Max	AUC-min	AUC-max
C	Age	26	72	0.6548	0.6674
	Stage	0	5	0.7457	0.7339
	Histology	0	3	0.8096	0.8143
	Tumor size	1	12	0.7191	0.7139
	Nodal status	0	2	0.7343	0.733
I	SUV_var	0.0172	127.3246	0.7809	0.7704
	SUV_kurtosis	1.5531	35.755	0.8061	0.7804
	Energy	11.8221	11.9947	0.8017	0.8174
	Contrast	0.0679	1.7816	0.7948	0.8017
	Variance	11.9435	11.9985	0.7748	0.7757
	Inertia	12.0153	12.5036	0.8122	0.8209
	Orientation	-88.8512	87.6894	0.8122	0.8174
	Perimeter	1.96	61.406	0.8174	0.8104
	MTV	33	4759	0.7922	0.787
C+I	Age	26	72	0.7678	0.7661
	Nodal status	0	2	0.8226	0.8113
	SUV_std	0.1312	11.2838	0.7809	0.7313
	SUV_var	0.0172	127.3246	0.7957	0.7991
	SUV_kurtosis	1.5531	35.755	0.8043	0.7987
	MaxProbability	11.9107	11.9974	0.7939	0.847
	MinorAxisLength	1.1547	14.8646	0.7574	0.7683
	MTV	33	4759	0.7861	0.7704

AUC-min value and AUC-max value correspond to the results using the minimal or maximal value of the corresponding feature, respectively.

using stage alone had very poor predictive value and was not a significant predictor for LRF or DF for the 2 groups of patients. The selected features in 3 models for locoregional failure and distant failure are shown in Tables 5 and 6. When predicting locoregional failure, 3, 13 and 13 features are selected for C, I and C+I models, respectively. In these features, stage is selected in both C and C+I models, while SUV_median, SUV_kurtosis, Energy, Cluster tendency are selected in both I and C+I models. The number of selected features in 3 models for distant failure is 5, 9, 8 for C, I and C+I models, respectively. Age and nodal status are selected in both C and C+I. SUV_var, SUV_kurtosis, and MTV are selected in both I and C+I models. We also evaluated the importance of each individual selected feature, which adopted the same strategy in our previous study[34]. Specifically, for each test sample, we manually changed each selected feature value to their minimal and maximal value, and then the modified test sample is fed into the trained model. The importance of individual feature can be evaluated by the AUC change of prediction model as shown in Tables 5 and 6. A larger difference indicates the greater contribution of this feature on prediction results. The important features for locoregional failure prediction are Stage, Correlation and Variance, while Stage, SUV Kurtosis and MaxProbability are the important features

for distant failure prediction, respectively. To better visualize the change, the magnitude of AUC changes for all the selected features in the 6 models are shown in Figure 2. A larger difference indicates the greater contribution of this feature on prediction results. Since volume is an important feature in many outcome predictions, we also evaluate the performance based on volume alone. We calculated the AUC values for the volume alone, which are 0.41 and 0.39 for locoregional and distant failure prediction, respectively. The corresponding ROC curves are shown in Figure 3.

We found that the combined model still had the best performance regardless of whether stage was used or not (Table 8). Again, the highest performance (after excluding stage) was seen in the combined model. For LRF, the AUC was 0.7 (p < 0.01) and specificity was 0.67 at a fixed sensitivity of 0.75, and for DF, the AUC was 0.78 (p < 0.01) and specificity was 0.73 at a fixed sensitivity of 0.75. Bivariate analysis showed that a high probability of distant failure as determined by the combined model (probability > 0.7) correlated significantly with death (p < 0.01).

A log-rank test was performed to compare survival of patients who were predicted to have a low probability of distant metastases (probability < 0.5) compared to the survival of patients predicted to have a high probability of distant metastases (probability ≥ 0.5) by the combined model. Patients with a low probability of distant metastases had a mean survival time (median not reached) of 57.8 months (95% CI: 50.5-65.1) while patients with a high probability of distant metastases had a median survival time of 19.0 months (95% CI: 12.6-25.4, p < 0.01) (Figure 4). Figure 5 shows the
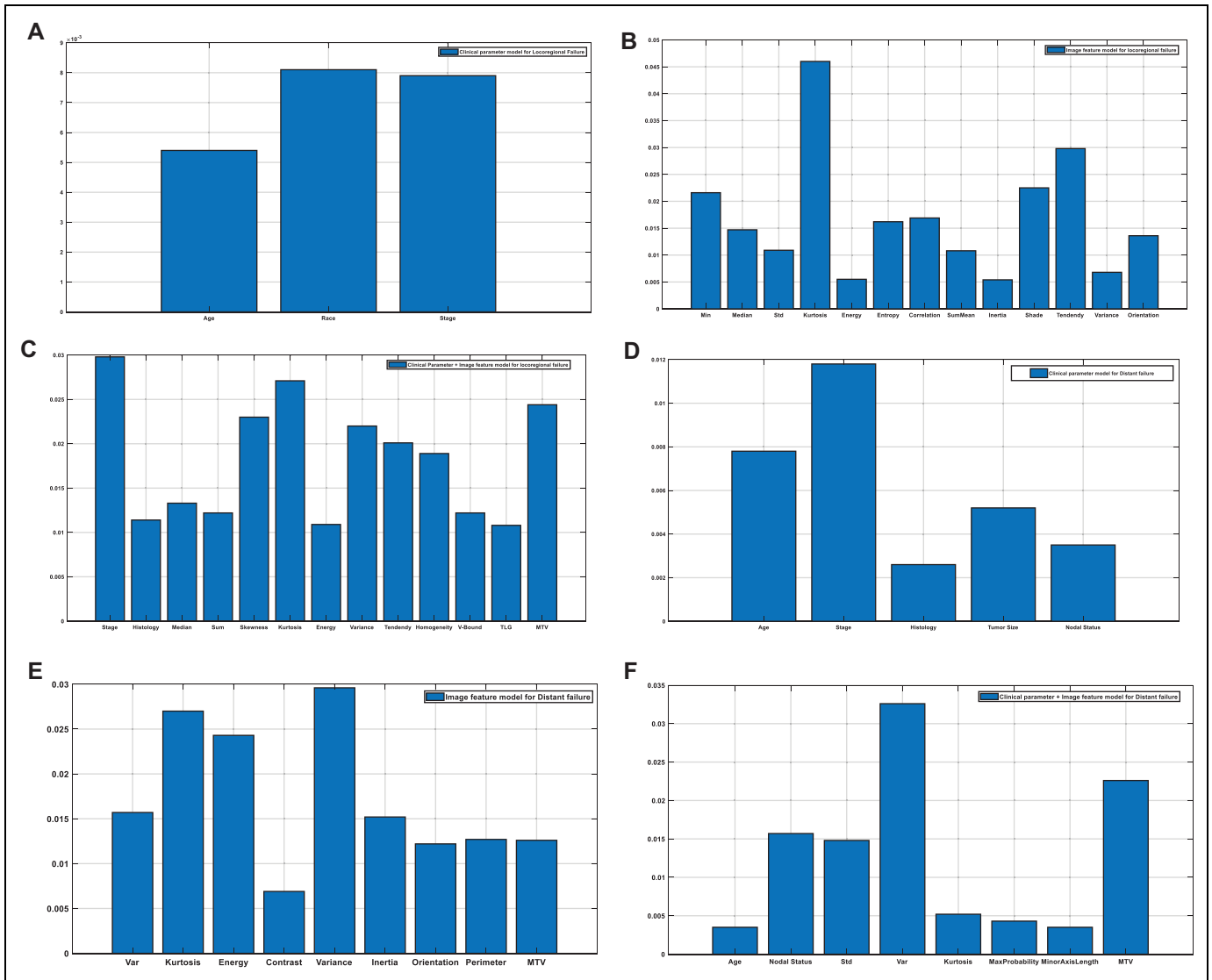
Figure 2. The magnitude of AUC changes for selected features in 6 models.
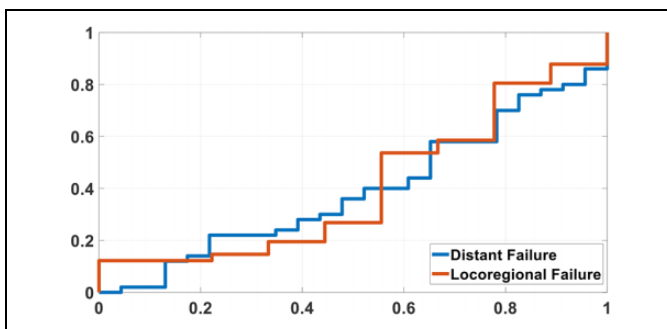
Figure 3. Receiver-operating characteristic (ROC) curves for volume with distant failure and locoregional failure prediction.

incidence of distant metastases for patients predicted to have low probability of distant failure (probability < 0.5) to that of patients predicted to have high probability of distant failure (probability ≥ 0.5) by the combined model.
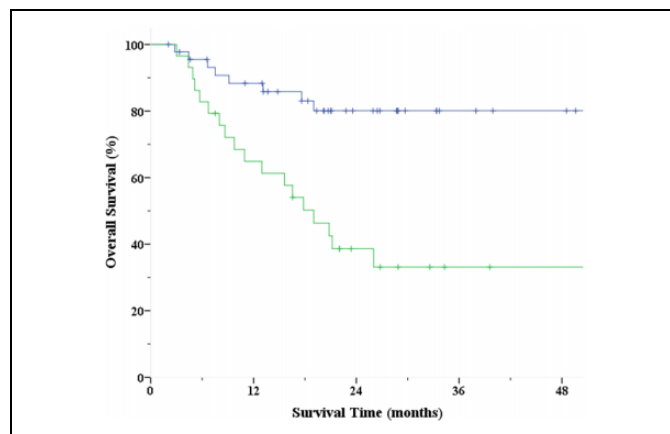
Discussion and Conclusion

Radiotherapy with concurrent chemotherapy is a standard of care for patients with stage IB2-IVA cervical cancer.[27] A meta-analysis showed chemoradiation was associated with a 5-year DFS improvement of 8% over patients receiving radiation alone.[28] However, relapses are common in the setting of standard therapy (with studies showing 5-year distant failure rates of 23%[29] and 27%[30]), indicating the need for additional intensified therapy to achieve optimal outcome. Adjuvant chemotherapy has additional survival benefits, with a meta-analysis showing its association with a 54% reduction in the risk of death and an absolute benefit of 19% at 5 years (60%-79%) when used after a course of definitive chemoradiation[28] This treatment is associated with increased toxicity,[31] and randomized studies have been initiated to address its benefit. The OUTBACK trial is a phase III protocol including unselected patients with stage IB2- IVA cervical cancer who will receive definitive concurrent

Table 7. The Parameters for SVM in all the Models.

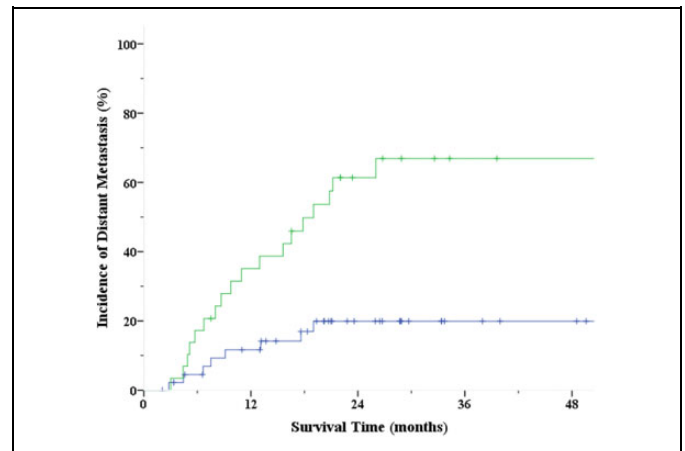|		c	g
Locoregional failure	C	-7 ± 2	8 ± 0
	I	13 ± 1	8 ± 2
	C+I	12 ± 2	6 ± 1
Distant failure	C	-7 ± 2	7 ± 0
	I	-8 ± 2	6 ± 1
	C+I	-8 ± 2	5 ± 1

Table 8. Model Performance Without Stage as a Clinical Parameter.

Locoregional failure

Model	Sensitivity	Specificity	AUC
C	0.75	0.63	0.7
C+I	0.75	0.67	0.7

Distant Failure

Model	Sensitivity	Specificity	AUC
C	0.75	0.61	0.65
C+I	0.75	0.73	0.78

Figure 5. Incidence of distant metastases for patients predicted to have low probability of distant failure (probability < 0.5, blue) compared to patients predicted to have high probability of distant failure (probability ≥ 0.5, green) by C+I (model using clinical parameters and imaging features).

Figure 4. Survival of patients predicted to have low probability of distant failure (probability < 0.5, blue) compared to survival of patients predicted to have high probability of distant failure (probability ≥ 0.5, green) by C+I (model using clinical parameters and imaging features).

chemoradiation and then be randomized to receive additional chemotherapy with 4 cycles of adjuvant carboplatin and pacli-taxel versus no further therapy.[32] Unfortunately, due to diverse eligibility criteria, inclusion of patients at low risk for DF may result in an inability to show a benefit for OUTBACK chemotherapy, and unselected administration of chemotherapy may result in excess toxicity for minimal benefit.

With the goal of identifying a subset of patients for which benefits of intensified therapy might outweigh its additional risks, prior studies have attempted to evaluate the use of various clinical and imaging parameters individually in prognosis and in predictive models. Studies examining the prognostic ability of individual parameters include one which found tumor spatial heterogeneity could predict patient outcomes for sarcoma (p < 0.001).[33] Among various FDG PET-CT features, intensity-volume-histogram variables had the highest association for locoregional recurrence after radiotherapy in non-small cell lung carcinoma.[34] A study using predictive models created support vector machine and logistic regression models that used a combination of clinical and FDG PET-CT parameters to predict pathologic tumor response to chemoradiation in esophageal carcinoma.[35] The support vector machine model achieved very high accuracy (AUC 1.00) when spatial-temporal PET features were combined with conventional PET-CT measures and clinical parameters. Most of the PET prognostic studies for cervix cancer are qualitative correlative studies[36-38] or have focused on single quantitative measures like textural analysis.[39] Using a cohort of 14 cervical cancer patients treated at a single institution, El Naqa et al found that a combination of intensity-volume histogram (IVH) metrics and texture features extracted from PET images had high predictive power for response to treatment.[11] The models in our study were constructed using a larger dataset consisting of 75 patients for model training and testing. Additionally, we used both clinical and a more expansive set of imaging parameters to build our model with the highest predictive power, which had an AUC of 0.84 for locoregional failure prediction. Nevertheless, for distant failure prediction, the model achieved lower AUC of 0.75. The performance difference between distant and locoregional failure prediction could be caused by the regions used to extract features. In this work, all radiomic features were extracted from segmented primary lesion. For distant metastasis, regions surrounding the tumor or "Shell"[40] may provide more prognostic information on tumor metastasis potential, which was not considered in this work. Incooperating features extracted from the shell may improve the model performance for distant failure prediction, which is worthy of investigation in a future study.

The strengths of this study include a relatively large sample of patients treated uniformly over a short period with sufficient risk factors to have local and systemic failures. In addition, the combined clinical and imaging predictive model approach helps maximize the prognostic capability. Limitations to our study are the absence of an external validation set and verification at another institution. The utilization of magnetic resonance imaging-based features, which we did not address in this study, may also add to the predictive capability. Our analysis was performed using some patients with short follow-up, which could be seen as another limitation; however, only 3 patients had <6 months of follow up and would not be expected to have a significant impact on the results. Additionally, we are focused on early failures, since these patients would theoretically benefit most from an early intervention that could be predicted by pre-treatment imaging. Therefore, we feel a longer follow-up may not be as critical to address this clinical question.

When provided a high sensitivity and then optimized to maximize specificity, we found that the most complex model using clinical parameters and imaging parameters had the best results, compared to the other 2 models and to all the individual clinical and imaging parameters, for locoregional failure and distant failure. The combined model could be used to select patients at high risk for distant or locoregional failure who would potentially benefit from additional adjuvant therapy as administered in the OUTBACK protocol.

Declaration of Conflicting Interests

Funding

ORCID iD

Zhiguo Zhou https://orcid.org/0000-0002-8867-6907

References

1. MacManus M, Nestle U, Rosenzweig KE, et al. Use of PET and PET/CT for radiation therapy planning: IAEA Expert Report 2006-2007. *Radiother Oncol*. 2009;91:85-94. doi:10.1016/j.radonc.2008.11.008

2. Heron DE, Andrade RS, Beriwal S, Smith RP. PET-CT in radiation oncology: the impact on diagnosis, treatment planning, and assessment of treatment response. *Am J Clin Oncol*. 2008;31:352-362. doi:10.1097/COC.0b013e318162f150

3. Nestle U, Weber W, Hentschel M, Grosu AL. Biological imaging in radiation therapy: role of positron emission tomography. *Phys Med Biol*. 2009;54:R1-25. doi:10.1088/0031-9155/54/1/r01

4. Kwon SH, Yoon JK, An YS, et al. Prognostic significance of the intratumoral heterogeneity of (18) F-FDG uptake in oral cavity cancer. *J Surgical Oncol*. 2014;110:702-706. doi:10.1002/jso.23703

5. Allal AS, Dulguerov P, Allaoua MG, et al. Standardized uptake value of 2-[(18)F] fluoro-2-deoxy-D-glucose in predicting outcome in head and neck carcinomas treated by radiotherapy with or without chemotherapy. *J Clin Oncol*. 2002;20(5):1398-1404.

6. Kidd EA, Siegel BA, Dehdashti F, Grigsby PW. The standardized uptake value for F-18 fluorodeoxyglucose is a sensitive predictive biomarker for cervical cancer treatment response and survival. *Cancer*. 2007;110;1738-1744. doi:10.1002/cncr.22974

7. Beaulieu S, Kinahan P, Tseng J, et al. SUV varies with time after injection in (18)F-FDG PET of breast cancer: characterization and method to adjust for time differences. *J Nucl Med*. 2003;44(7):1044-1050.

8. Keyes JW Jr. SUV: standard uptake or silly useless value? *J Nucl Med*. 1995;36(10):1836-1839.

9. Bourgier C, Colinge J, Aillères N, et al. Radiomics: definition and clinical development [in French]. *Cancer Radiother*. 2015;19:532-537. doi:10.1016/j.canrad.2015.06.008

10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278:563-577. doi:10.1148/radiol.2015151169

11. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit*. 2009;42:1162-1171. doi:10.1016/j.patcog.2008.08.011

12. Rose PG, Ali S, Watkins E, et al. Long-term follow-up of a randomized trial comparing concurrent single agent cisplatin, cisplatin-based combination chemotherapy, or hydroxyurea during pelvic irradiation for locally advanced cervical cancer: a gynecologic oncology group study. *J Clin Oncol*. 2007;25:2804-2810. doi:10.1200/jco.2006.09.4532

13. Whitney CW, Sause W, Bundy BN, et al. Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: a gynecologic oncology group and southwest oncology group study. *J Clin Oncol*. 1999;17(5):1339-1348.

14. Altazi BA, Fernandez DC, Zhang GG, et al. Investigating multi-radiomic models for enhancing prediction power of cervical cancer treatment outcomes. *Physica Medica*. 2018;2(46):180-188.

15. Fang M, Kan Y, Dong D, et al. Multi-habitat based radiomics for the prediction of treatment response to concurrent chemotherapy and radiation therapy in locally advanced cervical cancer. *Front Oncol*. 2020;10:563.

16. Gao S, Du S, Lu Z, et al. Multiparametric PET/MR (PET and MR-IVIM) for the evaluation of early treatment response and prediction of tumor recurrence in patients with locally advanced cervical cancer. *Eur Radiol*. 2020;30(2):1191-1201.

17. Lucia F, Visvikis D, Marie-Charlotte D, et al. Prediction of outcome using pretreatment 18 F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging*. 2018;45(5):768-786.

18. Meng J, Liu S, Zhu L, et al. Texture analysis as imaging biomarker for recurrence in advanced cervical cancer treated with CCRT. *Sci Rep*. 2018;8(1):1-9.

19. Takada A, Yokota H, Nemoto MW, et al. A multi-scanner study of MRI radiomics in uterine cervical cancer: prediction of in-field tumor control after definitive radiotherapy based on a machine learning method including peritumoral regions. *Jpn J Radiol*. 2020;38(3):265-273.

20. Zhou Z, Folkert M, Cannon N, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. *Radiother Oncol*. 2016;119:501-504. doi:10.1016/j.radonc.2016.04.029.

21. Lobo JM, Jimenez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol Biogeography*. 2008;17(2):145-151.

22. Wald NJ, Bestwick JP. Is the area under an ROC curve a valid measure of the performance of a screening or diagnostic test? *Journal of Medical Screening*. 2014; 21(1): 51-56.

23. Jiménez-Valverde A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecol Biogeography*. 2012;21(4):498-507.

24. Adams NM, Hand DJ. Comparing classifiers when the misallocation costs are uncertain. *Pattern Recognit*. 1999;32:1139-1147.

25. Zhou Z, Folkert M, Iyengar P, et al. Multi-objective radiomics model for predicting distant failure in lung SBRT. *Phys Med Biol*. 2017;62:4460-4478. doi:10.1088/1361-6560/aa6ae5

26. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20:273-297. doi:10.1007/bf00994018

27. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Cervical Cancer (Version 1.2017). https://www.nccn.org/professionals/physician_gls/pdf/cervical.pdf

28. Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration. Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: a systematic review and meta-analysis of individual patient data from 18 randomized trials. *J Clin Oncol*. 2008;26:5802-5812. doi:10.1200/jco.2008.16.4368

29. Fortin I, Jürgenliemk-Schulz I, Mahantshetty UM, Lindegaard J, Kirchheiner K, Pötter R. Distant metastases in locally advanced cervical cancer pattern of relapse and prognostic factors: early results from the EMBRACE study. *Int J Radiat Oncol Biol Phys*. 2015;93:S8-S9. doi:10.1016/j.ijrobp.2015.07.026

30. Schmid MP, Franckena M, Kirchheiner K, et al. Distant metastasis in patients with cervical cancer after primary radiotherapy with or without chemotherapy and image guided adaptive brachytherapy. *Gynecol Oncol*. 2014;133:256-262. doi:10.1016/j.ygyno.2014.02.004

31. Peters WA III, Liu PY, Barrett RJ II, et al. Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *J Clin Oncol*. 2000; 18:1606-1613. doi:10.1200/jco.2000.18.8.1606

32. Sudeep G. Adjuvant chemotherapy in locally advanced cervical cencer: the ceiling remains unbroken. *Journal of Gynecologic Oncology*. 2019;30(4):1-3. doi:10.3802/jgo.2019.30.e97

33. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med*. 2008;49:1973-1979. doi:10.2967/jnumed.108.053397

34. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, Naqa IE. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother Oncol*. 2012;102:239-245. doi:10.1016/j.radonc.2011.10.014

35. Zhang H, Tan S, Chen W, et al. Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal 18F-FDG PET features, clinical parameters, and demographics. *Int J Radiat Oncol Biol Phys*. 2014;88:195-203. doi:10.1016/j.ijrobp.2013.09.037

36. Yilmaz M, Adli M, Celen Z, Zincirkeser S, Dirier A. FDG PET-CT in cervical cancer: relationship between primary tumor FDG uptake and metastatic potential. *Nucl Med Commun*. 2010;31:526-531. doi:10.1097/MNM.0b013e32833800e7

37. Yoo J, Choi JY, Moon SH, et al. Prognostic significance of volume-based metabolic parameters in uterine cervical cancer determined using 18F-fluorodeoxyglucose positron emission tomography. *Int J Gynecol Cancer*. 2012;22:1226-1233. doi:10.1097/IGC.0b013e318260a905

38. Sharma DN, Rath GK, Kumar R, et al. Positron emission tomography scan for predicting clinical outcome of patients with recurrent cervical carcinoma following radiation therapy. *J Cancer Res Ther*. 2012;8:23-27. doi:10.4103/0973-1482.95169

39. Shang-Wen C, Wei-Chih S, Te-Chun H, et al. Textural features of cervical cancers on FDG-PET/CT associate with survival and local relapse in patients treated with definitive chemoradiotherapy. *Sci Rep*. 2018;8:11859. doi:10.1038/s41598-018-30336-6

40. Hao H, Zhou Z, Li S, et al. Shell feature: a new radiomics descriptor for predicting distant failure after radiotherapy in non-small cell lung cancer and cervix cancer. *Phys Med Biol*. 2018;63(9):095007.