

Texture Analysis of F-18 Fluciclovine PET/CT to Predict Biochemically Recurrent Prostate Cancer: Initial Results

Hakmook Kang^{1,2}, E. Edmund Kim^{3,4}, Sepideh Shokouhi⁵, Kenneth Tokita⁴, and Hye-Won Shin^{4,6}

¹Department of Biostatistics; ²Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, TN; ³Department of Radiological Sciences, University of California, Irvine, CA; ⁴KSK Medical, LLC, KSK Cancer Center of Irvine, and KSK Imaging Center of Irvine, Irvine, CA; ⁵Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN; and ⁶Chiron Total, LLC, Irvine, CA

Correspondence Author:

Hakmook Kang, PhD
2525 West End Ave. Suite #1100,
Nashville, TN 37203;
E-mail: h.kang@vumc.org

Key Words: Positron emission tomography (PET), Axumin, F-18 fluciclovine, prostate cancer, Haralick features

Abbreviations: biochemical recurrence (BCR), least absolute shrinkage and selection operator (LASSO), prostate-specific antigen (PSA), area under the curve (AUC), positron emission tomography (PET), computed tomography (CT), prostate cancer (PCa), gray-level co-occurrence matrix (GLCM), clinical and texture information (CTI), receiver operating characteristic (ROC)

ABSTRACT

Predicting biochemical recurrence of prostate cancer is imperative for initiating early treatment, which can improve the outcome of cancer treatment. However, because of inter- and intrareader variability in interpretation of F-18 fluciclovine positron emission tomography/computed tomography (PET/CT), it is difficult to reliably discern between necrotic tissue owing to radiation therapy and tumor tissue. Our goal is to develop a computational methodology using Haralick texture analysis that can be used as an adjunct tool to improve and standardize the interpretation of F-18 fluciclovine PET/CT to identify biochemical recurrence of prostate cancer. Four main textural features were chosen by variable selection procedure using least absolute shrinkage and selection operator logistic regression and bootstrapping, and then included as predictors in subsequent logistic ridge regression model for prediction ($n = 28$). Age at prostatectomy, prostate-specific antigen (PSA) level before the PET/CT imaging, and number of days between the prostate-specific antigen measurement and PET/CT imaging were also included in the prediction model. The overfitting-corrected area under the curve and Brier score of the proposed model were 0.94 (95% CI: 0.81, 1.00) and 0.12 (95% CI: 0.03, 0.23), respectively. Compared with a model with textural features (TI model) and that with only clinical information (CI model), the proposed model achieved 2% and 32% increase in AUC and 8% and 48% reduction in Brier score, respectively. Combining Haralick textural features based on the PET/CT imaging data with clinical information shows a high potential of enhanced prediction of the biochemical recurrence of prostate cancer.

INTRODUCTION

Prostate cancer (PCa) is the second leading cause of death by cancer among the male population in the United States. Primary PCa is often treated by prostatectomy. However, biochemical recurrence (BCR) of PCa occurs in 27%–53% of patients and is typically detected by the rise in prostate-specific antigen (PSA) serum levels. However, PSA serum levels cannot provide information about the location and extent of the disease (1). One-third of the population experiencing recurrence develops metastasis within 8 years and multiple anatomical scans may be required to evaluate potential metastasis. Anatomical scans may be unable to detect tumors when their size is <1 cm, when PSA serum levels are <10 ng/mL, or when patients are treated with localized therapy (1, 2). Recently, F-18 fluciclovine (anti1-amino-3-F-18

fluorocyclobutane-1-carboxylic acid), also known as Axumin (Blue Earth Diagnostics, Oxford, UK), an FDA-approved positron emission tomography (PET) radiotracer, has been added to the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology for Prostate Cancer. F-18 fluciclovine has received considerable attention because it can be used for early detection and localization of recurrent PCa after prostatectomy (3).

Miller et al. have reported that a specific training program for the interpretation of F-18 fluciclovine PET/computed tomography (CT) images helped readers achieve acceptable diagnostic accuracy and reproducibility for stating recurrent PCa (4). However, the interpretation of F-18 fluciclovine PET/CT images is inherently subjective, leading to inter- and intrareader

variability (4, 5). Furthermore, overall F-18 fluciclovine performance was influenced by PSA levels. For example, true positivity and sensitivity were 19% and 75%, respectively, when PSA levels were ≤ 1.05 ng/mL and 80% and 100%, respectively, when PSA levels were > 3.98 but ≤ 8.90 ng/mL (5). Moreover, it is not trivial for an inexperienced nuclear physician or radiologist to discern between necrotic tissue owing to radiation therapy and tumor tissue, resulting in low predictive power to progression and recurrence of PCa.

The Haralick texture analysis (6) was developed in 1970s for extracting spatial information from images. Over time, texture analysis was adopted for diagnostic and prognostic applications in medical imaging to characterize tissue properties (eg, heterogeneity) (5–10), which is believed to influence the outcome of cancer treatment. This study is particularly focused on evaluating the utility of several Haralick texture features including contrast, variance, and correlation for predicting PCa recurrence. Haralick texture features are calculated from a gray-level co-occurrence matrix (GLCM) of an image. GLCM is defined as the distribution of co-occurring voxel values at different offsets.

We hypothesize that the quantitative extraction of high-dimensional mineable data (or image biomarkers) of F-18 fluciclovine images may be sensitive to changes in pattern or spatial distribution of F-18 fluciclovine and thus may enhance the prediction of recurrence, thereby making for better clinical decision. The goal of this study is to develop a computational methodology using Haralick texture analysis that can be used as an adjunct tool to improve and standardize the interpretation of F-18 fluciclovine PET/CT to identify BCR of PCa, particularly for inexperienced nuclear physicians and radiologists.

MATERIALS AND METHODS

Study Population and Data Set

The data for the present study were collected in a retrospective way. F-18 fluciclovine PET/CT images were obtained from 28 prostatectomy patients seen at the KSK Cancer Center of Irvine from February 2017 to October 2017. These 28 patients received an F-18 fluciclovine injection for detection of suspected BCR. To protect patients' confidentiality and privacy, only minimally required clinical information was collected for the study. These include age at prostatectomy, initial Gleason score, PSA levels prior to radiation (if occurred), prior cancer treatments, PSA levels prior to F-18 fluciclovine scan, and number of days between PSA measurement and F-18 fluciclovine scan. The *P*-values for the difference in those demographic variables between F-18 fluciclovine-positive and -negative groups were computed by Wilcoxon rank sum tests for continuous variables and chi-square tests for categorical variables. The Institutional Review Board at Vanderbilt University approved the retrospective study (IRB #171811), and the requirement to obtain informed consent was exempted.

F-18 Fluciclovine PET/CT

Imaging Protocol. Each patient received 370 MBq (10 mCi) of F-18 fluciclovine as a bolus intravenous injection according to the manufacturer's instruction. The PET/CT images were acquired using a PET/CT scanner (GE Discovery, GE Healthcare, Chicago,

IL). The patients were requested to have nothing by mouth for at least 4 hours prior to F-18 fluciclovine administration; 10 mCi of F-18 fluciclovine in < 5 mL saline was administered to patients who were in supine position. The administered dose was calculated using a dose calibrator. Once F-18 fluciclovine was administered, the patients were instructed to raise their arms above their head, and CT scans were started for attenuation correction and anatomical correlation purposes. PET images were acquired within 3–5 minutes post administration of F-18 fluciclovine from the mid-thigh area to the base of the skull. PET/CT images were reconstructed following the standard methods set forth by the imaging facility.

Image Interpretation. F-18 fluciclovine PET/CT images were evaluated by a highly experienced (> 40 years) nuclear medicine physician certified by the American Board of Nuclear Medicine and an interpretation program by Blue Earth (Oxford, UK). Specific anatomical locations (lesions) of focal increased activity with greater intensity than adjacent background, bone marrow, or cardiac blood pool activity were classified as positive (mild, moderate, and marked) or negative for BCR based on visual assessment. Anatomical regions of interest included both prostate bed (residual prostate, prostate bed, and seminal vesicles) and extraprostatic sites such as pelvic lymph node and extrapelvic metastasis.

Preprocessing of Imaging Data. We used the DCM2NII function implemented in MRICron software (7) to convert patients' DICOM-format PET images into NIFTI-format images. Without further preprocessing, those NIFTI-format images were imported into the R Statistical Software (version 3.4.3, R Foundation for Statistical Computing, Vienna, Austria) for subsequent statistical analyses.

Statistical Analysis

Haralick Features. Haralick textural features associated with patterns or spatial distribution of pixel intensities in each 2-dimensional image were calculated using GLCM (6). For each patient, GLCMs of 21 selected axial slices of PET images covering the prostate bed and extraprostatic sites were computed using the function GLCM in the *glm* package in R Statistical Software using shifts of 0° , 45° , 90° , and 135° with a 3×3 window size defining the neighbors of a reference pixel for texture calculation. The function calculated each texture statistic using the specified 4 directions (ie, shift angles) and then returned the mean value of the neighbors' texture values for each pixel. Then each Haralick feature was calculated for each slice by taking the average of all pixel-level feature values. We calculated 8 Haralick features for each slice of PET image: mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation. Details are described in Haralick et al.'s study (6).

In total, 168 features (8 features/slice \times 21 slices) were entered into the variable selection procedure using least absolute shrinkage and selection operator (LASSO) logistic regression and bootstrapping, where the recurrence status (binary variable) was the dependent variable. The *glmnet* package (8) in R was used to fit the LASSO logistic regression model. In this procedure, patients were resampled 500 times with replacement, and a LASSO logistic regression model was fit for each resampled data set. A LASSO logistic regression performs variable selection to

determine which texture variables are important in explaining the dependent variable, shrinking less-important variables to zero (9). Then the texture features were ranked by how often they were included in the chosen LASSO logistic regression models. The top 4 texture features were then included as predictors in the final logistic ridge regression models. Because of the number of subjects being fewer ($n = 28$) than the number of explanatory variables, we used logistic ridge regression models to avoid overfitting (10, 11).

Define Models

Patients were identified as having “recurrence” (probability of recurrence of prostate cancer = 1) if there was presence of viable tumor cells in the primary tumor bed or the lymph nodes identified in the F-18 fluciclovine PET image, and as nonrecurrence (probability of recurrence of prostate cancer = 0) if otherwise. This binary response was modeled with chosen texture variables and 3 clinical information. The “Clinical and Texture Information” (CTI) model is expressed as follows:

$$\text{logit}[\text{Pr}(pRP = 1)] = \beta_0 + \beta_1 h_{\text{texture}_1} + \beta_2 h_{\text{texture}_2} + \dots + \beta_m h_{\text{texture}_m} + \beta_{m+1} \text{AgeP} + \beta_{m+2} \text{PSA} + \beta_{m+3} \text{DaysPA}, \quad (1)$$

where h_{texture_i} is the i th texture variable; *AgeP* is the age at prostatectomy; *PSA* denotes PSA level before fluciclovine PET imaging; *DaysPA* denotes the number of days between the PSA measurement and fluciclovine PET imaging; $m = 4$ is the number of texture features.

For comparisons, 2 other models were analyzed: a model with only clinical information (CI model) including *AgeP*, *PSA*, and *DaysPA*, and a model with only texture information (TI model) as predictors. In each logistic ridge regression, the tuning parameter was estimated as described in by Cule et al. (12). The performance of the 3 models was compared in terms of overfitting-corrected AUC (area under the receiver operating characteristic [ROC] curve) and Brier scores (11, 13).

The statistical significance of the difference in overfitting-corrected AUCs and Brier scores was investigated by generating a bootstrap distribution of the difference with 500 replicates. The difference for overfitting-corrected AUC is defined by subtracting AUC of the CI (or TI) model from that of the CTI model. Therefore, a positive value of the difference in AUC implies that the CTI outperforms the other models in terms of AUC. The difference for overfitting-corrected Brier score is also defined in the same fashion—by subtracting the Brier score of the CI (or TI) model from that of the CTI model, meaning that a negative value of the difference is indicative of the CTI model outperforming the other models in terms of Brier score.

Overfitting-Corrected Statistics

With each logistic ridge regression model, the probability of recurrence of prostate cancer was computed for all patients and was used to construct an ROC curve for the model. The AUC was then estimated using the trapezoidal rule. Also, the sum of the squared difference between the predicted probability of recurrence and the observed recurrence was computed, and the average of this difference was defined as the Brier score for the

Table 1. Clinicopathological Characteristics of Study Subjects

Characteristic	Subjects (n = 28)		P-value
	F-18 Fluciclovine-Positive (n = 17)	F-18 Fluciclovine-Negative (n = 11)	
Age at prostatectomy, years (Mean ± SD)	66 ± 6	61 ± 8	.11
Prior cancer therapies (N)			
Prostatectomy only	17	11	1.00
Prostatectomy + Radiation therapy	4/17 (24%)	2/11 (18%)	
Initial Gleason score (N)			
No records or missing	2 (12%)	2 (18%)	.22 ^b
6	1 (6%)	0 (0%)	
7	9 (53%)	7 (64%)	
8	4 (24%)	2 (18%)	
9	1 (6%)	0 (0%)	
PSA ^a , ng/mL			
PSA < 1 ng/mL	14 (82%)	9 (82%)	1.00 ^b
PSA 1 ≤ 2 ng/mL	0 (0%)	1 (9%)	
PSA ≥ 2 ng/mL	3 (18%)	1 (9%)	
Interval between PSA prior to F-18 fluciclovine and F-18 fluciclovine scan, days, median (range)	29 (6, 309)	13 (0, 68)	.07 ^c

^a PSA before F-18 fluciclovine scan.

^b Cells with zero observation were combined with other categories to compute the P-value.

^c P-value was computed after removing 2 outliers, that is, 149 and 309 days.

Table 2. F-18 Fluciclovine–Positive Region Identified by an Experienced Nuclear Medicine Physician

Region (n = 17)	n (%)
Prostate bed only	8 (47%)
Prostate bed + pelvic lymph nodes	4 (24%)
Extrapelvic sites ^a	5 (29%)

^aThe extrapelvic sites include retroperitoneal lymph node or bone lesion.

model. We then used the bootstrap to correct for overfitting in each case with 100 repetitions (13, 14); in particular, for each resampled data set, each model was validated against the original data to compute overfitting-corrected statistics (see online supplemental Appendix A for more details). The overfitting-corrected AUC (or Brier score) was used because this approach will not yield inflated standard errors, which would have been the case if we had instead used leave-one-out cross-validation (15). We repeated this procedure 500 times to generate a bootstrap distribution of each overfitting-corrected statistic of interest and 95% confidence intervals.

RESULTS

Demographic Characteristics

Clinicopathological characteristics of the study subjects are presented in Table 1. The mean age was 66 years (SD: 6) and 61 years (SD: 8) for F-18 fluciclovine–positive and F-18 fluciclovine–negative subjects, respectively. F-18 fluciclovine PET/CT was positive in 17 subjects (61%) and negative in 11 (39%). All 28 study subjects had prior undergone prostatectomy; 4 out of 17 (24%) F-18 fluciclovine–positive subjects received radiation therapy for the salvage of recurrent tumor or iliac lymph node metastasis compared with 2 out of 11 (18%) F-18 fluciclovine–negative subjects. Initial Gleason score of ≥ 7 was noted in 14

(82%) and 9 (82%) F-18 fluciclovine–positive and F-18 fluciclovine–negative subjects, respectively. In total, 82% of study participants had PSA serum level <1 ng/mL than that obtained before F-18 fluciclovine PET scans. Note that F-18 fluciclovine–positive patients tend to have a slightly longer interval between PSA measurement before the PET/CT imaging and F-18 fluciclovine scan owing to two outliers, that is, 149 and 309 days.

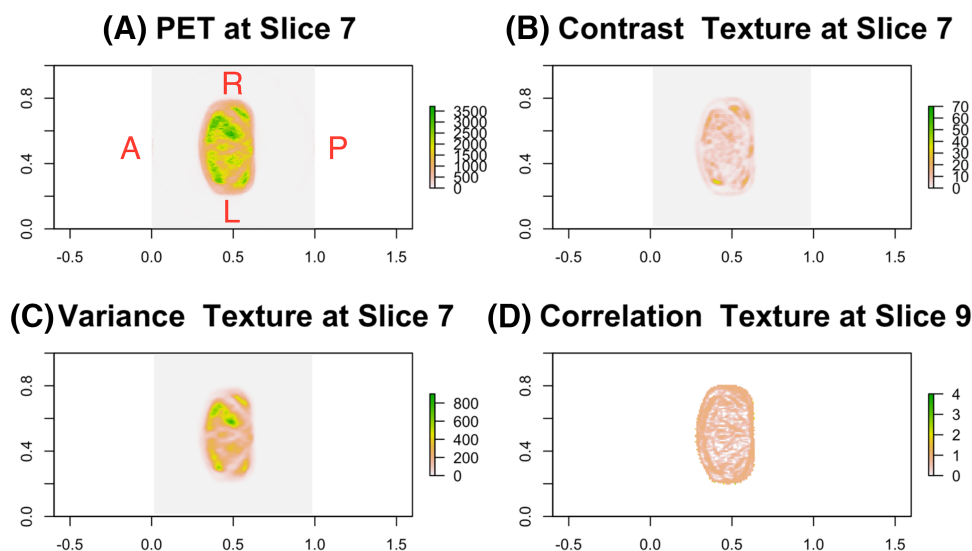
F-18 Fluciclovine–Positive Regions. F-18 fluciclovine–positive regions had been identified by the experienced nuclear medicine radiologist and are presented in Table 2. The main recurrence sites with F-18 fluciclovine–positive scans are prostate bed (47%) and prostate bed plus pelvic lymph nodes (24%); in 5 patients with F-18 fluciclovine–positive scans (29%) the recurrent and metastatic sites were extrapelvic.

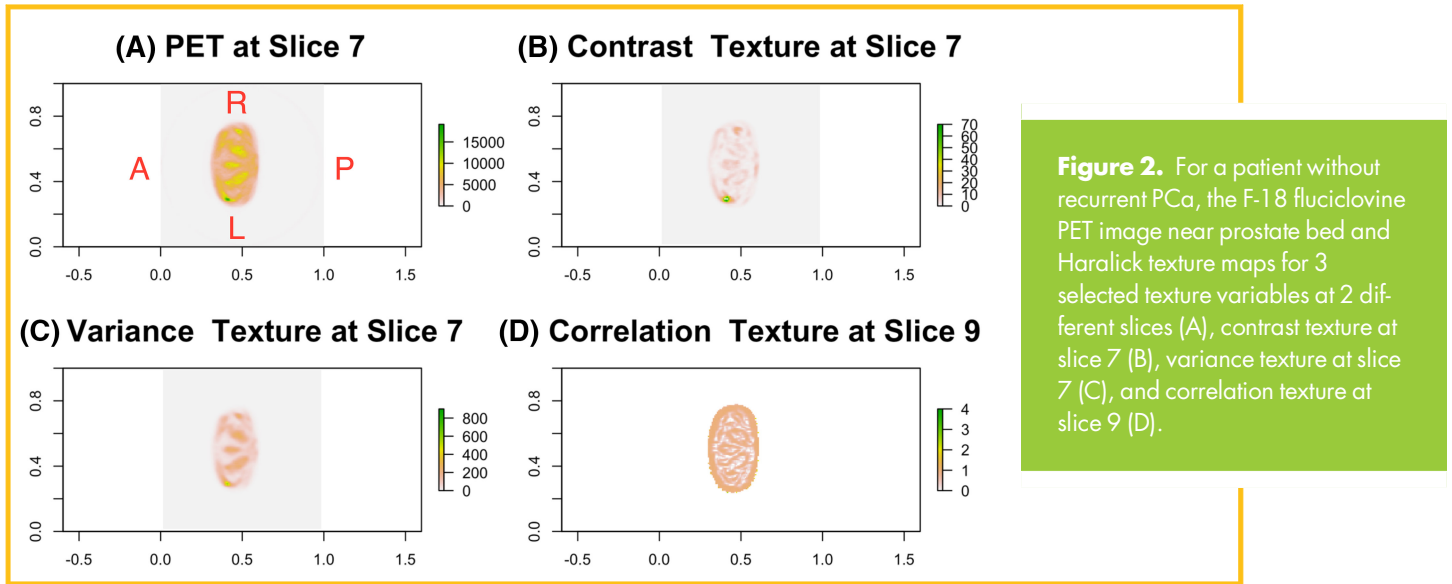
Selected Haralick Text Features. The top 4 texture features selected via 500 bootstrapped data included variance and contrast textures at the 7th slice and correlation textures at the 9th and 14th slices. For illustration purpose only, the representative axial image of PET/CT scans with 3 texture features at the 7th and 9th slices for a patient with recurrent PCa is presented in Figure 1. A similar image for a patient without recurrent PCa is summarized in Figure 2. Although each texture feature has its own meaning and interpretation, our goal is to not interpret them in our model but to incorporate the feature information into the model to enhance the ability to predict BCR. The estimated logistic ridge regression coefficients of the 3 models, that is, CTI, TI, and CI models, are given in the online supplemental Appendix B.

Overfitting-Corrected AUC and Brier Score

In Table 3, the AUCs and Brier scores (corrected for overfitting) from the 3 logistic ridge regression models are reported along with the corresponding 95% confidence intervals. The overfitting-corrected AUC of 0.94 from the CTI model outperformed the AUC of 0.71 from the CI model, which is about a 32% ($= 0.23/0.71$) increase in the AUC of the CTI model compared to that of

Figure 1. For a patient with recurrent prostate cancer (PCa), the F-18 fluciclovine positron emission tomography (PET) image near prostate bed and Haralick texture maps for 3 selected texture variables at 2 different slices (A), contrast texture at slice 7 (B), variance texture at slice 7 (C), and correlation texture at slice 9 (D).





the CI model. It is noteworthy that both CTI and TI models achieved very similar AUC and 95% confidence intervals, although the CTI model slightly outperformed the TI model. This indicates that adding the Haralick texture information into the model significantly improved the performance of the model. Moreover, the CTI model achieved the smallest overfitting-corrected Brier score of 0.12, indicating that the CTI model predicted the recurrence of PCa more accurately than the TI (Brier score = 0.13) and CI (Brier score = 0.23) models. The additional contribution of the texture information was assessed by comparing the Brier score of the CTI model with that of the CI model, achieving reduction of 48% (= 0.11/0.23) in the Brier score. Note that the Brier score of the CTI model also slightly outperformed that of the TI model (ie, 8% [= 0.01/0.13] improvement) owing to its use of clinical information in addition to the texture information. Therefore, it is evident that incorporating texture information has a much stronger effect on both the AUC and Brier score with respect to the predictive performance of the model.

In Figure 3, the bootstrap distribution of the differences in overfitting-corrected AUCs between CTI and CI, and between CTI and TI, is graphically summarized for illustration purposes. The area of the shaded region in Figure 3 (ie, ~0 and 0.40) indicates how the overfitting-corrected AUCs resulting from the CTI model are likely smaller than those resulting from the CI and TI models. That is, the CTI model would be at least 99% more likely to outperform the CI model and ~60% more likely to outperform the TI model in terms of the overfitting-corrected AUC value. Table 4 summarizes how the first model (CTI) likely outperformed the second model (CI) in terms of AUCs and Brier scores (corrected for overfitting). Note that the model with additional texture information is about 98% more likely to outperform the CI model in terms of the overfitting-corrected AUC and Brier score. However, the CTI model slightly outperformed the TI model in terms of overfitting-corrected AUC and Brier score, indicating that texture information, compared with clinical information, significantly contributes toward improving the performance of the model.

The bootstrap distribution (not reported here) of the differences in overfitting-corrected Brier scores between CTI and CI and between CTI and TI is very similar to the graphs shown in Figure 3. The comparisons between the models in terms of overfitting-corrected Brier scores are summarized in the last column of Table 4. Note that the CTI model is about 67% (98%) more likely to have smaller Brier scores than the TI model (CI model).

DISCUSSION

Haralick texture features extracted from longitudinal 18F-florbetapir PET scans have been used to successfully differentiate between subject groups (eg, normal vs patients with mild cognitive impairment) without normalizing PET intensity using a reference region (16); it has been shown that first- and higher-order textual features with low-level variations are identified for reproducible solid tumor segmentation using FDG-PET/CT scans (17). In this study, we have, to the best of our knowledge, shown for the first time that a statistical model (CTI model) combining Haralick texture features computed from F-18 fluciclovine PET/CT images with patients' clinical information may improve the chances of accurately detecting BCR. The overfitting-corrected

Table 3. The AUC and Brier Scores (with Bootstrap 95% Confidence Intervals) Based on Overfitting-Correction

	Overfitting-Corrected	
	AUC	Brier Score
CTI model	0.94 (0.81, 1.00)	0.12 (0.03, 0.23)
TI model	0.92 (0.79, 1.00)	0.13 (0.04, 0.24)
CI model	0.71 (0.44, 0.89)	0.23 (0.15, 0.31)

The scores are for the model with texture information and clinical data (CTI model) (A), the same model without clinical data (TI model) (B), and the model with only clinical data (CI model) (C).

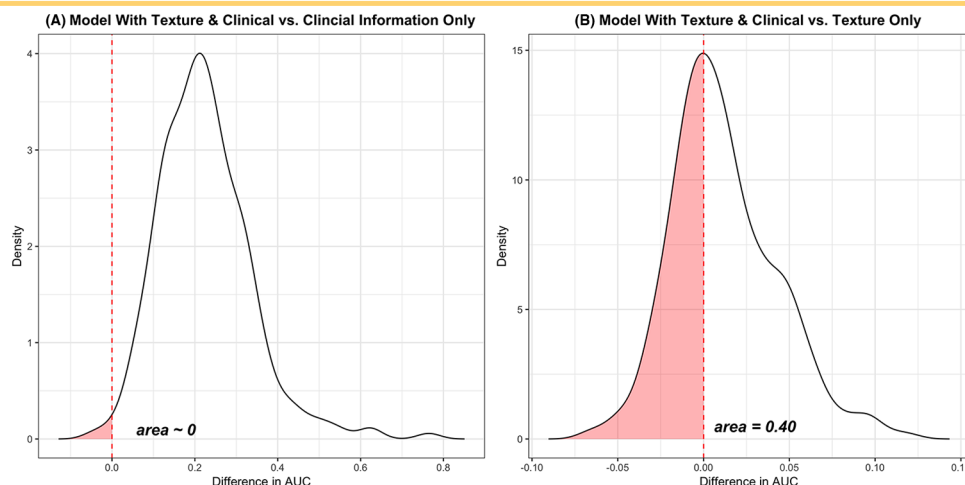


Figure 3. Differences in overfitting-corrected area under the curve (AUCs) between CTI model and CI model (A) and between CTI model and TI model (B). The area, which is ~ 0 and 0.4, in each plot indicates how likely the overfitting-corrected AUCs based on the CTI model are smaller than those based on the CI and TI models. That is, the CTI model would at least be 99% (or 60%) more likely to outperform the CI (or TI) model in terms of overfitting-corrected AUC. CTI, TI, and CI models denote a model with texture information and clinical information, a model with only texture information, and a model only with clinical information, respectively.

AUC of 0.94 and Brier score of 0.12 of the CTI model outperformed the AUC of 0.71 and Brier score of 0.23 of the CI model (the model with only clinical information) and the AUC of 0.92 and Brier score of 0.13 of the TI model (the model with only texture information). Our proposed CTI model can subdue reader variability and PSA-dependent F-18 fluciclovine performance in detecting BCR of PCa.

Although our primary goal is to incorporate the textural feature information into the model to enhance the ability to predict BCR and to not infer on the regression coefficients of the models, we found that the top 4 texture features are negatively associated with the probability of BCR. That is, a patient with higher variance, contrast, and correlation at the chosen axial slices tends to be free from BCR compared with a patient with lower values of these values. Because age at prostatectomy, number of days

between the PSA measurement and fluciclovine-PET imaging, and PSA serum level before fluciclovine PET/CT imaging were included in the model owing to their clinical significance instead of statistical significance, interpretation of the corresponding regression coefficients is not of interest and thus not included.

Although our primary finding is encouraging, the major limitation of this study includes the moderate sample size ($n = 28$). To validate our models in an independent cohort of patients with PCa with and without recurrence, we aim to collaborate with multiple F-18 fluciclovine research and/or clinical groups to have access to a similar data set with a large sample size. Because of the inability to access such external data sets, we internally validated our model using the bootstrap. The second limitation is that the true recurrence status of each patient was assessed by only 1 highly experienced (of >40 years) nuclear medicine physician. Although we strongly believe that the assessed recurrence status is very close to the underlying truth, it would be still better to have multiple experienced physicians assess the recurrence status. The third limitation is that we did not fully account for patient-level variability in pelvic anatomy, indicating that each axial slice out of the chosen 21 slices per patient may not be perfectly aligned with the corresponding slice of another patient. Also, we did not mask the background region in each slice, which may reduce the signal-to-noise ratio in each feature of interest. However, we believe that our results are still promising as “initial results.” In the future, with a larger-scale data set after masking the background noise, we are planning to extend our approach to accommodate voxel-level texture information extracted from 3D volume covering the pelvic region of each patient, which can result in more reliable conclusion, free from patient-level variability in pelvic anatomy. The fourth limitation is that we assumed the linear relationship between the logit of the probability of recurrence and the explanatory variables as shown in

Table 4. Summary of CTI outperforming CI or TI Model in terms of AUCs and Brier Scores

	Overfitting-Corrected	
	Prob (AUC)	Prob (Brier Score)
CTI vs CI	0.99	0.98
CTI vs TI	0.60	0.67

The table depicts the probability of how likely the first model would outperform the second model for each comparison in terms of AUC and Brier scores based on overfitting-correction. Prob (AUC) and Prob (Brier score) denotes the probability that the first model (ie, CTI) outperforms the second model (ie, CI or TI) in terms of overfitting-corrected AUC and Brier score, respectively. CTI, TI, and CI models denote a model with texture information and clinical information, a model only with texture information, and a model only with clinical information, respectively.

model 1. There may exist nonlinear relationships between the logit of the probability of recurrence and the explanatory variables, which may not be fully captured by our proposed model with moderate sample size. Our current study uses traditional statistical linear model although machine learning (ML) techniques, including random forest and deep learning, may outperform our model because ML techniques are designed to capture hidden and nonlinear association between outcome and explanatory variables. However, given the modest sample size, ML techniques are not preferred over traditional statistical linear model because of the potential for overfitting. Moreover, a recent systematic review reveals that the benefit of ML over traditional logistic regression for clinical prediction would be minimal even with a

large-scale data set (18), although that review did not focus on clinical prediction using biomedical imaging data. We are planning to compare our statistical model with ML techniques including the deep learning approach in terms of AUC and Brier score when we have access to much larger data sets in the future.

Supplemental Materials

Supplemental Appendix A: <https://doi.org/10.18383/j.tom.2020.00029.sup.01>

Supplemental Appendix B: <https://doi.org/10.18383/j.tom.2020.00029.sup.02>

ACKNOWLEDGMENTS

We thank Mark Fiecas at University of Minnesota for helpful discussion.

Disclosures: No disclosures to report.

Conflict of Interest: None reported.

REFERENCES

1. Zanoni L, Bossert I, Matti A, Schiavina R, Pultrone C, Fanti S, Nanni C. A review discussing fluciclovine ([18F] PET/CT imaging in the detection of recurrent prostate cancer. *Future Oncol*. 2018;11:1101–1115.
2. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, Fossati N, Gross T, Henry AM, Joniau S, Lam TB, Mason MD, Matveev VB, Moldovan PC, van den Bergh RCN, Van den Broeck T, van der Poel HG, van der Kwast TH, Rouvière O, Schoots IG, Wiegel T, Cornford P. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol*. 2017;71:618–629.
3. Savir-Baruch B, Zanoni L, Schuster DM. Imaging of prostate cancer using fluciclovine. *PET Clin*. 2017;12:145–157.
4. Miller MP, Kostakoglu L, Pryma D, Yu JQ, Chau A, Perlman E, Clarke B, Rosen D, Ward P. Reader training for the restaging of biochemically recurrent prostate cancer using [18F]fluciclovine PET/CT. *J Nucl Med*. 2017;58:1596–1602.
5. European Medicines Agency. Assessment report, 2017. 2017 [Available from: https://www.ema.europa.eu/en/documents/assessment-report/axumin-epar-public-assessment-report_en.pdf.
6. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern*. 1973;SMC-3:610–621.
7. Rorden C, Brett M. Stereotaxic display of brain lesions. *Behav Neurol*. 2000;12:191–200.
8. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1–22.
9. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58:267–288.
10. Le Cessie S, van Houwelingen JC. Ridge estimators in logistic regression. *J R Stat Soc Ser C Appl Stat*. 1992;41:191–201.
11. Kang H, Hainline A, Arlinghaus LR, Elderidge S, Li X, Abramson VG, Chakravarthy AB, Abramson RG, Bingham B, Fakhoury K, Yankeelov TE. Combining multiparametric MRI with receptor information to optimize prediction of pathologic response to neoadjuvant therapy in breast cancer: preliminary results. *J Med Imaging*. 2017;5:011015.
12. Cule E, De Iorio M. Ridge regression in prediction problems: automatic choice of the ridge parameter. *Genet Epidemiol*. 2013;37:704–714.
13. Frank E HJ. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. 2nd ed. NY: Springer; 2015.
14. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc*. 1983;78:316–331.
15. Breiman L, Spector P. Submodel selection and evaluation in regression: the x-random case. *International Statistical Review*. 1992;60:291–319.
16. Campbell DL, Kang H, Shokouhi S. Application of Haralick texture features in brain [18F]florbetapir positron emission tomography without reference region normalization. *Clin Interv Aging*. 2017;12:2077–2086.
17. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol*. 2010;49:1012–1016.
18. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.