The use of texture-based radiomics CT analysis to predict outcomes in early-stage non-small cell lung cancer treated with stereotactic ablative radiotherapy

Objective: Stereotactic ablative radiotherapy (SABR) is being increasingly used as a non-invasive treatment for early-stage non-small cell lung cancer (NSCLC). A non-invasive method to estimate treatment outcomes in these patients would be valuable, especially since access to tissue specimens is often difficult in these cases.

Methods: We developed a method to predict survival following SABR in NSCLC patients using analysis of quantitative image features on pre-treatment CT images. We developed a Cox Lasso model based on two-dimensional Riesz wavelet quantitative texture features on CT scans with the goal of separating patients based on survival.

Results: The median log-rank p-value for 1000 cross-validations was 0.030. Our model was able to separate patients based upon predicted survival. When we added tumor size into the model, the p-value lost its significance, demonstrating that tumor size is not a key feature in the model but rather decreases significance likely due to the relatively small number of events in the dataset.

Conclusions: The model predicting patient survival following SABR in NSCLC may be useful in future studies by enabling prediction of survival-based outcomes using radiomics features in CT images.

Advances in knowledge: Quantitative image features from NSCLC nodules on CT images have been found to significantly separate patient populations based on overall survival (p = 0.04). In the long term, a non-invasive method to estimate treatment outcomes in patients undergoing SABR would be valuable, especially since access to tissue specimens is often difficult in these cases.

INTRODUCTION
Staging of lung cancer with pathologic tissue has long been the gold-standard to predict recurrence risk in non-small cell lung cancer (NSCLC). When surgery is pursued, patients are often upstaged due to the discovery of occult nodal metastasis.1–3 However, medical comorbidities of patients often limit access to pathologic tissues and surgical management. As an alternative to surgery, non-invasive...
surgical ablative radiation therapy (SABR) is increasingly chosen for both definitive and palliative treatment in early-stage patients and has been linked to improved survival in elderly lung cancer patients.\textsuperscript{4,5} Often SABR and other non-invasive treatments such as conventional radiation, targeted therapy and chemotherapy render surgical staging of the cancer excessive; instead, cancer may be staged by radiological imaging techniques with limited sensitivity and specificity.\textsuperscript{6} For these reasons, image-based biomarkers could play an important role in guiding the treatment of NSCLC patients who are high-risk candidates for invasive procedures and surgery.

Early-stage NSCLC is often curable. For patients who are poor surgical candidates or who refuse surgery, SABR is an alternative treatment with excellent outcomes, highlighted by the Radiation Therapy Oncology Group (RTOG) 0236 clinical trial\textsuperscript{7} and multiple single and multiinstitutional experiences.\textsuperscript{8–10} Although SABR treatment can result in greater than 90% local control, there is a need for methods that can predict which patients are at highest risk of regional or distant metastasis and who may benefit from upfront systemic treatment. Semantic (human-determined) imaging-based biomarkers have recently been described for Stage 1 NSCLC treated with SABR, such as tumor size, contact to mediastinal pleura, and the maximum standard described for Stage 1 NSCLC treated with SABR, such as tumor size, contact to mediastinal pleura, and the maximum standard.

In this study, we hypothesized that quantitative (machine-calculated) image based features from pre-treatment CT scans can predict the risk for tumor recurrence.

Image-based quantitative features from pre-treatment imaging may have the power to predict short-term treatment success and long-term outcomes. Quantitative image features, particularly those based on lesion heterogeneity, have had success in predicting outcomes for a wide range of cancers,\textsuperscript{12–15} including response to chemotherapy,\textsuperscript{16,17} long-term survival in lung cancer,\textsuperscript{18} and long-term survival in other cancers.\textsuperscript{19,20} Our approach uses a quantitative model derived from two-dimensional (2D) Riesz wavelet analysis.\textsuperscript{21} We employed a Cox Lasso model to develop the model, we successfully predicted survival on an expanded cohort of patients, the results were not dependent on tumor size being factored into the model, and we found that both the solid and ground glass opacity (GGO) components provide predictive Riesz data. Our results contribute to the growing literature of radiomics analysis of patient images, which in the context of stereotactic radiation can serve as a methodology to predict outcomes for patients with early-stage NSCLC to guide treatment decisions for patients treated non-invasively.

METHODS AND MATERIALS

Patient cohort

A cohort of 116 patients was assembled retrospectively with institutional review board approval from patients treated with thoracic SABR for biopsy-confirmed, primary NSCLC, at Stanford University. We chose the patients based upon a follow-up window described in the methods at our institution. We included all of the patients that met the criteria described where we had the appropriate imaging and contours readily available. A superset of 110 patients of this cohort has been previously described.\textsuperscript{11} This prior article dealt with identifying manually extracted imaging-based predictors of disease progression whereas in this manuscript we used quantitative features to separate patient populations based on overall survival. Only patients with available radiation planning CT images were included. All cancers were Stage I by the seventh edition AJCC staging manual.\textsuperscript{22} Most patients received SABR rather than surgery because of surgical risk due to comorbid conditions, but some refused surgery. Patients were not included if they received chemotherapy prior to a diagnosis of metastatic failure, if they had synchronous tumors, or if they had received a new diagnosis of primary NSCLC following SABR. Patients were also excluded if they had been previously treated for NSCLC, unless it was for prior Stage I disease treated exclusively with lobectomy or pneumonectomy at least 1 year prior to SABR and the second tumor was considered to be a new primary tumor by the treating physicians. SABR treatment dose was determined via a tumor volume-adapted SABR dosing strategy.\textsuperscript{23} All patient images were deidentified prior to analysis, so no informed consent was required from participants.

Patient treatments and outcomes

For each patient, after the decision was made to undergo SABR treatment, patients received a treatment planning CT or PET/CT in the radiation oncology department and typically started treatment within 1 to 2 weeks. Depending upon the clinical and imaging features of the tumor, an appropriate radiation dose was chosen. Patients were treated with a tumor volume-adapted fractionation strategy, with the majority of doses exceeding a BED 100, as described in Shultz et al.\textsuperscript{11} Table 1 describes the patient population, clinical factors, and treatment dose fractionation schemes. Patients were followed with CT and/or PET/CT imaging every 3 months after treatment. Of 116 patients there were a total of 52 events including 12 local failures, 20 regional failures, 18 distant failures and 39 deaths (Table 2).

CT imaging

CT-images, acquired at a single institution using a General Electric CT scanner, were without contrast and were used for treatment planning. Of the 116 patients, 69% ($n = 80$) of CT scans had 1.25 mm slice thickness and 20% ($n = 23$) had 2.5 mm slice thickness. Most patient CT scans (78%, $n = 90$) had in-plane pixel spacing of 0.98 mm; pixel spacing varied in other patients from 0.80 to 1.37 mm.

Outcomes and follow-up

Assessed outcomes included local failure with tumor recurrence in the same lobe of the treated tumor, regional failure with recurrence in regional lymph nodes or the same hemithorax, distant failure consisting of metastatic disease, distant failure without regional failure, progression-free survival and overall survival. Survival outcomes were determined objectively. Local, regional, and distant recurrence was determined using a combination of new findings on imaging, serial progression on imaging, and in many cases confirmed with a biopsy. If progression was convincing based upon the clinical picture treatment was often initiated without a biopsy. Imaging included diagnostic
CT scans, PET scans, and often both were utilized to improve diagnostic assessment. Clinical exam and/or interval imaging was performed at each follow-up evaluation. All follow-up evaluations included imaging that corroborated each clinical outcome, but clinical exam was not performed for some patients who were followed by their local physician. Minimum, median and maximum follow-up time for the entire cohort was 2, 18.5 and 66 months.

Quantitative features
We assessed the tumors using a three-dimensional region of interest (ROI) that captures the gross tumor volume that was delineated by the treating physician. We chose a single 2D slice on which to extract quantitative imaging features; this slice was the largest 2D ROI that did not contain any pixel with HU > 3000 (to exclude artifacts from fiducial marker placement). Median 2D ROI bounding box size was 20 × 20 pixels.

Multiple techniques, involving Gaussian image filtering and Riesz wavelets (low-pass or high pass),24–26 were used to extract quantitative features to capture lesion texture. Features were extracted from the solid center, the rim or the full tumor (Figure 1). Depending on the parameters used, the number of features extracted from a single 2D ROI bounding box varied between 2 and 30 quantitative values. Features were used with or without nodule size to test it as a confounding factor.

More technical information regarding Riesz wavelet analysis can be found in Supplementary Material 1.

Statistical analysis
For predictive model selection, we used the Cox proportional-hazards Lasso method.23 Based on the model's predictions, patients were partitioned into two exclusive groups, which were

BED, biologically effective dose; NSCLC, non-small cell lung cancer;
then compared with the use of a log-rank test. A low log-rank $p$-value ($p < 0.05$) indicated effectiveness at separating the data into cohorts with significantly different prognoses. For visualizing differences in prognosis between separated cohorts, we used Kaplan-Meier charts.

More technical information regarding Cox Lasso method and statistical analysis can be found in Supplementary Material 1.

Cross-validated predictive model

During cross-validation, we randomly separated 58 patients for model fitting and 58 patients for testing (Figure 2). The goal of the analysis was to train a predictive model on the 58 model fitting patients and to separate the 58 testing patients, using the predictive model, into two groups with significantly different survival curves ($p < 0.05$). The process was performed 1000 times and the median $p$-value was selected as the final result.

More technical information regarding cross-validation method can be found in Supplementary material.

RESULTS

Cross-validated predictive modeling

For this study, we were most interested in the ability of the model to predict overall survival given the number of events. However, predicting many types of treatment failures could be important from a clinical perspective, and each could have a distinct biologic rationale. For example, factors that impact local and distant failure could be very different and could have different texture features. Given our interest in different types of treatment failures, we assessed outcomes that included local failure, regional failure, distant failure, distant failure without regional failure, progression-free survival and overall survival.

Survival results are median 1000 cross-validation $p$-values. Significant results ($p < 0.05$) were found when using overall survival outcomes but no other outcomes (Table 3).

In addition to the above findings, we observed that the use of Gaussian filtering and high-pass Riesz wavelets on full tumor

Table 3. Significant analyses and input parameters for measurement

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Nodule size</th>
<th>Features</th>
<th>Filtering</th>
<th>Riesz parameters</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Full nodule</td>
<td>Gaussian</td>
<td>High-pass, $N = 1$, $J = 1$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center</td>
<td>Gaussian</td>
<td>High-pass, $N = 2$, $J = 3$</td>
<td>0.03</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center</td>
<td>Gaussian</td>
<td>High-pass, $N = 4$, $J = 1–3$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center</td>
<td>Without</td>
<td>High-pass, $N = 2$, $J = 3$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center</td>
<td>Without</td>
<td>High-pass, $N = 2$, $J = 3$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>GGO</td>
<td>Gaussian</td>
<td>High-pass, $N = 2$, $J = 2,4$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center and GGO</td>
<td>Gaussian</td>
<td>High-pass, $N = 2$, $J = 3$</td>
<td>0.04</td>
</tr>
<tr>
<td>Overall survival</td>
<td>Without</td>
<td>Solid center and GGO</td>
<td>Without</td>
<td>High-pass, $N = 1$, $J = 3$</td>
<td>0.04</td>
</tr>
</tbody>
</table>

GGO, ground glass opacity;
nodule without nodule size was significant or nearly significant for outcomes such as overall survival \((p = 0.04–0.07)\), progression-free survival \((p = 0.06–0.09)\) and distant failure without regional failure \((p = 0.09–0.14)\). This confirms that filtering and high-pass Reisz result in separating the data the greatest and that nodule size confounds the model that may be due to the variation in nodule size and the relatively low overall sample size.

A pattern that was statistically significant \((p = 0.05)\) or nearly significant \((p < 0.10)\) was observed with the use of Gaussian filtering, high-pass Riesz wavelets, without nodule size feature included in the model for outcomes such as overall survival, progression-free survival and distant failure without regional failure.

Kaplan–Meyer charts show examples of statistically significant (Figure 3A,B) and non-significant (Figure 3C,D) results generated with the use of our cross-validated predictive model.

**DISCUSSION**
Quantitative image analysis provides useful information to evaluate disease and guide treatment decisions based upon assessed risks. In this study, we investigated the association of treatment failure with quantitative features based on Riesz wavelets that represent lesion texture. Features of this form have been shown in prior work to correlate with tumor type, progression and response to therapy. Our quantitative features show statistically significant results in separating patients with early-stage NSCLC into groups of different survival. Other interesting results that need further investigations have been found during cross-validation in outcomes such as overall survival, progression-free survival and distant failure without regional failure.

Figure 3. (A) Kaplan–Meyer chart \((p = 0.021)\) evaluating progression-free survival, full nodule, with Gaussian filtering, high-pass Riesz wavelets \((N = 1, J = 1)\) (B) Kaplan–Meyer chart \((p = 0.028)\) evaluating distant failure, full nodule, with Gaussian filtering, high-pass Riesz wavelets \((N = 1, J = 1)\) (C) Kaplan–Meyer chart \((p = 0.96)\) showing lack of significance for progression-free survival, full nodule, with Gaussian filtering, high-pass Riesz wavelets \((N = 1, J = 1)\) (D) Kaplan–Meyer chart \((p = 0.92)\) showing lack of significance for distant failure, full nodule, with Gaussian filtering, high-pass Riesz wavelets \((N = 1, J = 1)\).
We believe that our approach has great potential to predict the biologic nature of cancers thereby distinguishing risk for local, regional, and distant failure. Our method is also thought to be generalizable, and will be applying it to additional malignancies. A limitation of our study is the limited number of events in our data set, which could hinder our ability to develop a robust model. In addition, we recognize that the size of our patient population is limited, though we included all patient data available at our institution that met inclusion criteria. The goal of our analysis was to develop the model and explore its potential clinical value. Exploring larger data sets will be needed to validate the predictive capability of our models, and we are actively pursuing larger validation sets. Such analysis may also be able to understand the relative importance of the solid and GGO components of a tumor for their clinical and biological significance. Nevertheless, our current results appear promising.

In addition to evaluating larger data sets, our future work will be to assess the potential clinical benefit of implementing this methodology in prospective clinical settings and to incorporate more sophisticated analyses. Additional modeling could include three-dimensional image features, which have been shown to be more effective at predicting malignancy in breast lesions than 2D features.24

All analyses were performed on treatment planning CT scans with the aim of decreasing variability. However, slice thickness, the timing of contrast, and level of breath-hold was not consistent throughout the years of treatment, so this variation could limit our results. Standardization using pretreatment diagnostic thoracic CT scans may have a positive impact on future results.

The quantitative techniques employed here are novel and worth considering in future analyses that involve quantitative features as predictors of longitudinal censored outcomes. Particularly, our use of Cox Proportional Hazards Lasso regression is a powerful method of determining which quantitative features would be viable predictors of survival in future cohorts.

**FUNDING**

This work was supported by the Swiss National Science Foundation (under grant PZ00P2 154891) and grants from the National Cancer Institute, National Institutes of Health, U01CA142555, 1U01CA190214, and 1U01CA187947.

**REFERENCES**


