The value of nodal information in predicting lung cancer relapse using 4DPET/4DCT

Heyse Li[a)]
Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada

Nathan Becker and Srinivas Raman
Radiation Oncology, UHN Princess Margaret Cancer Centre, 610 University of Avenue, Toronto, Ontario M5T 2M9, Canada

Timothy C. Y. Chan
Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada and Techna Institute for the Advancement of Technology for Health, 124 - 100 College Street, Toronto, Ontario M5G 1P5, Canada

Jean-Pierre Bissonnette
Radiation Oncology, UHN Princess Margaret Cancer Centre, 610 University of Avenue, Toronto, Ontario M5T 2M9, Canada and Techna Institute for the Advancement of Technology for Health, 124 - 100 College Street, Toronto, Ontario M5G 1P5, Canada

(Received 19 January 2015; revised 11 June 2015; accepted for publication 1 July 2015; published 20 July 2015)

Purpose: There is evidence that computed tomography (CT) and positron emission tomography (PET) imaging metrics are prognostic and predictive in nonsmall cell lung cancer (NSCLC) treatment outcomes. However, few studies have explored the use of standardized uptake value (SUV)-based image features of nodal regions as predictive features. The authors investigated and compared the use of tumor and node image features extracted from the radiotherapy target volumes to predict relapse in a cohort of NSCLC patients undergoing chemoradiation treatment.
Methods: A prospective cohort of 25 patients with locally advanced NSCLC underwent 4DPET/4DCT imaging for radiation planning. Thirty-seven image features were derived from the CT-defined volumes and SUVs of the PET image from both the tumor and nodal target regions. The machine learning methods of logistic regression and repeated stratified five-fold cross-validation (CV) were used to predict local and overall relapses in 2 yr. The authors used well-known feature selection methods (Spearman's rank correlation, recursive feature elimination) within each fold of CV. Classifiers were ranked on their Matthew's correlation coefficient (MCC) after CV. Area under the curve, sensitivity, and specificity values are also presented.
Results: For predicting local relapse, the best classifier found had a mean MCC of 0.07 and was composed of eight tumor features. For predicting overall relapse, the best classifier found had a mean MCC of 0.29 and was composed of a single feature: the volume greater than 0.5 times the maximum SUV (N).
Conclusions: The best classifier for predicting local relapse had only tumor features. In contrast, the best classifier for predicting overall relapse included a node feature. Overall, the methods showed that nodes add value in predicting overall relapse but not local relapse. [http://dx.doi.org/10.1118/1.4926755]

Key words: machine learning, node, NSCLC, PET, 4D

1. INTRODUCTION

Positron emission tomography (PET) image features are a promising avenue to inform clinical management of locally advanced nonsmall cell lung cancer (NSCLC).[1–3] Studies have shown that 18F-fluorodeoxyglucose (FDG) PET images contain information that have both prognostic value and predictive capabilities in assessing response to chemoradiation treatment (CRT). However, despite the many studies demonstrating a strong association between PET imaging features and clinical outcomes,[4–15] there has been limited investigation of PET features associated with nodal disease.

In view of the amount of data and derivative features available, researchers are exploring machine learning methods to predict clinical outcomes and response based on image and nonimage features in a range of tumor sites and treatment modalities. Dehing-Oberije *et al.*[16] developed a support vector machine (SVM) classifier for predicting two-year survival of NSCLC patients based on patient characteristics and presenting features of the cancer. Naqa *et al.*[17] used logistic regression to predict failure probabilities of cervical and head-and-neck cancer based on FDG-PET/computed tomography (CT) image features. Their best models for both sites were composed of two features each. Naqa *et al.*[18] found that using SVM along

with a nonlinear kernel resulted in superior performance when the data demonstrated nonlinear behavior for predicting esophagitis, pneumonitis, and xerostomia endpoints. Vaidya *et al.*[19] applied logistic regression to predict the failures of NSCLC patients based on combined FDG-PET/CT features and found that a two-feature classifier performed the best through a bootstrapping feature selection algorithm that analyzed classifiers composed of one to five features. van Stiphout *et al.*[20] developed a SVM classifier for pathological complete response of rectal cancer after chemoradiation therapy based on FDG-PET and clinical features. Zhang *et al.*[21] used FDG-PET information in a SVM classifier with 17 features to evaluate esophageal tumor response to CRT.

In this paper, we build upon these previous studies by developing a logistic regression classifier, which uses 4DPET using FDG and 4DCT image features to predict two-year local and overall relapses in a cohort of LA-NSCLC patients. While nodal PET uptake can be a predictor of distant relapse in head and neck cancers,[22] few studies have investigated its use with machine learning models. We include nodal 4DPET and 4DCT features in our classifiers and explore their predictive value, which has not been considered previously in LA-NSCLC. By varying whether the classifiers can choose nodal features or not, we quantify their predictive value. We also explore how different combinations of features improve the predictive capability of our classifier. Similar to other studies, we use Matthew's correlation coefficient (MCC)[18] to compare the performance of the different classifiers.

2. METHODS AND MATERIALS

2.A. Data

After institutional research ethics board approval, we identified a prospective cohort of LA-NSCLC patients undergoing conventionally fractionated radiation therapy and concurrent chemotherapy with curative intent. These 32 patients consented to 4DPET/4DCT imaging 2 weeks prior to treatment and 3 months after completion of therapy. Of the 32 patients who originally consented to the study, three were rendered ineligible since the first 4DPET/4DCT scan showed evidence of metastatic disease prior to treatment. Two patients withdrew due to toxicity during treatment, and another two died prior to the post-treatment 4DPET/4DCT scan. The remaining 25 patients who contributed to this study are described in Table I. All 25 eligible patients were followed for at least 2 yr (or until death), with a median follow-up of 26 months.

For each patient, we examined two clinical outcomes: local relapse and overall relapse at 2 yr post-therapy. Local relapse was defined by first tumor recurrence being in the planning target volume (PTV). Overall relapse was determined by a recurrence anywhere, including all locoregional and distant failures. Twenty percent of all patients had a local relapse as the first site of disease recurrence and 60% of patients had a relapse by 2 yr. All patients who died during follow-up had a documented recurrence before death. The median progression-free survival was 10.1 months and the overall survival rate at 2 yr was 60%.

TABLE I. Patient population characteristics for our set of 25 patients. "NOS" means the histology could not be definitively classified.

Age	Median	63 (37–79)
Sex	Male	16 (0.64)
	Female	9 (0.36)
Stage	IIA/B	2 (0.08)
	IIIA	12 (0.48)
	IIIB	11 (0.44)
Histology	Adeno	19 (0.76)
	Squamous	4 (0.16)
	NSCLC NOS	2 (0.08)
Primary tumor size	Median	37 mm
	Mean	42.6 mm
Radiotherapy dose	60 Gy	7 (0.38)
	66 Gy	16 (0.64)
	70 Gy	1 (0.04)
	74 Gy	1 (0.04)

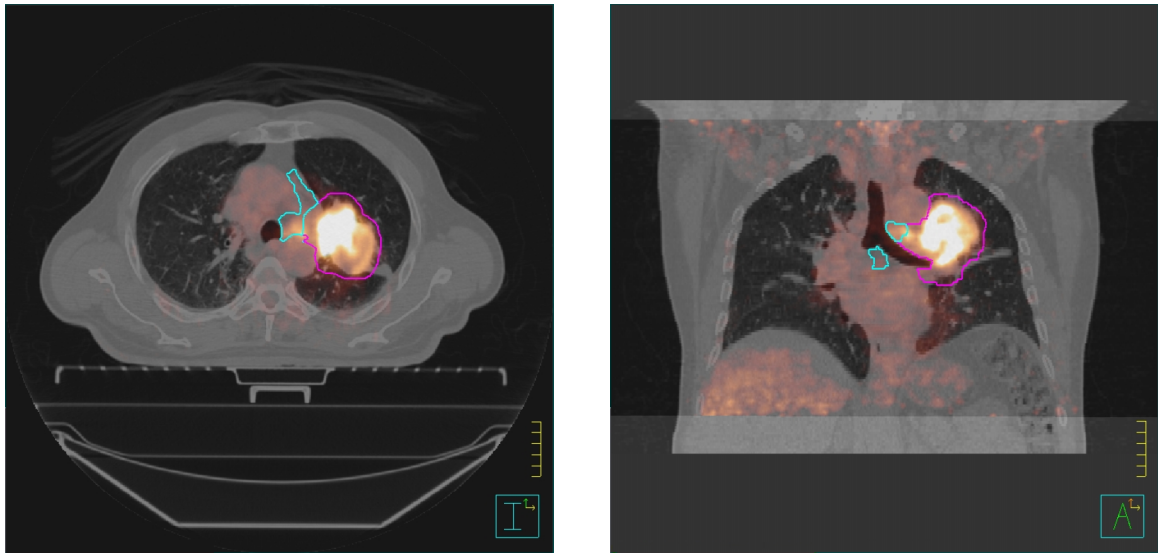
Each patient underwent a planning 4DPET/4DCT scan (Discovery ST, GE Healthcare, Waukesha, WI). The resolution of the 4DCT was $0.78\times0.78\times2$ mm. The resolution of the 4DPET was $3.9\times3.9\times2$ mm and was interpolated to match the resolution of the CT dataset. These images were transferred to our treatment planning system (Pinnacle^3, Philips Radiation Oncology Systems, Madison, WI), where a radiation oncologist contoured the normal tissues and targets on 4DCT. The primary tumor and affected nodes were contoured separately, each consisting of a gross tumor volume (GTV), clinical target volume (CTV), internal target volume (ITV), and PTV. The ITV was derived from the union of the CTV drawn on the inhale and exhale phases of the planning 4DCT dataset. A region of interest for PET subimage analysis was restricted to a mask based on both the tumor and nodal ITVs, as shown in Figure 1.

2.B. Features

PET features were extracted from the masks defined above on the inhale phase. Standardized uptake value (SUV) data from the masks were exported to MATLAB (MATLAB and Statistics Toolbox Release 2010b, The MathWorks, Inc., Natick, Massachusetts, United States) and were used to construct our initial feature set. These features were preprocessed in order to minimize the effects of large magnitude of any one feature value on the learning algorithm. For each feature, we standardized the values by subtracting the mean and dividing by the variance across all patients.[23] Features with zero variance were removed from the dataset, resulting in the 37 features per patient shown in Table II. We indicate to which site a feature belongs with either a superscript T for tumor (e.g., $\mathrm{SUV}^T_{\mathrm{mean}}$) or N for node (e.g., $\mathrm{SUV}^N_{\mathrm{mean}}$).

2.C. Predictive Model

We developed a logistic regression classifier that used the previously defined 4DPET/4DCT features to predict local and



(a) Axial slice (b) Coronal slice

FIG. 1. Fused 4DPET/4DCT images of a patient from our study. The purple contour is the tumor mask based on the ITV. The teal contour is the node mask based on the nodal ITVs.

overall relapses. The logistic function is defined as

$$f(\mathbf{x}) = \frac{1}{1+e^{-(\mathbf{w}\cdot\mathbf{x}+b)}}, \tag{1}$$

where $\mathbf{w}$ and b are the weights, and $\mathbf{x}$ is a vector of features for a given patient. The function $f(\mathbf{x})$ outputs values in the range [0,1] that represent the probability the patient, as defined through their $\mathbf{x}$ vector, is a member of the positive class, which, in this paper, is relapse. The LIBLINEAR library[24] for MATLAB was used to train and validate the logistic regression classifier. We used a threshold of 0.5 to force a binary output.

2.D. Model comparison and evaluation

We used five-fold stratified cross-validation (CV) to evaluate the results. The data were split into five disjoint sets. One set was held out as the test set and the remaining four were used as the training set. This ensured that our classifier was not evaluated on the data on which it was trained. The test sets were forced to hold at least one patient from each class and each test set had a similar distribution of data points from the positive (relapse) and negative (relapse-free) classes. For every classifier (i.e., different combination of features), we performed CV with 100 repetitions.

We evaluated the performance of each classifier on Matthew's correlation coefficient (MCC)[25] over the 100 repetitions. MCC is defined as

$$\frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{2}$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives. A MCC of +1 indicates a perfect prediction, 0 is equivalent to a fair coin toss, and −1 indicates a prediction that is opposite from the true class. A classifier that naively predicts the majority class for all patients will also have a MCC value of 0. Thus, MCC provides a metric by which to judge the classifiers that is desensitized to class imbalance.

We repeated the 100 repetitions of CV with the number of features held constant to f, for $f = 1,\ldots,8$. We investigated two different methods to select features within each fold. The first method used the Spearman's rank correlation coefficient. For each fold of training data, the univariate Spearman's correlation between each feature and the label was calculated. Then, the top f features with the highest correlation were selected to be included in the classifier. The second method used recursive feature elimination (RFE).[26] Here, a linear SVM was trained on the training fold and we removed the feature with the smallest weight coefficient. We retrained the linear SVM using the trimmed feature set and again removed the feature with the smallest weight. We repeated this until we were left with the top f features with the highest weight.

TABLE II. The set of 37 features for each patient. Each feature listed in the table exists for both the tumor and node regions unless otherwise specified. The intensity volume histogram (IVH) features were defined relative to $\mathrm{SUV}_{\mathrm{peak}}$ because $\mathrm{SUV}_{\mathrm{max}}$ is vulnerable to large outliers.

Feature	Definition
$\mathrm{SUV}_{\mathrm{mean}}$	Mean SUV
$\mathrm{SUV}_{\mathrm{median}}$	Median SUV
$\mathrm{SUV}_{\mathrm{peak}}$	95th percentile of the SUV distribution
$\mathrm{SUV}_{\mathrm{max}}$	Max SUV
IVH_x	Volume with a SUV $\geq x\mathrm{SUV}_{\mathrm{peak}}$, where $x \in [0.25, 0.5, \ldots, 2.25]$ for the tumor and $x \in [0.25, 0.5, \ldots, 3.5]$ for the nodes.
$\mathrm{Vol}_{\mathrm{PET}}$	Volume of the PET mask
$\mathrm{Vol}_{\mathrm{CT}}$	Volume of the CT-based GTV on the exhale phase
$\mathrm{Vol}_{0.5\mathrm{max}}$	Volume with a SUV $\geq 0.5\mathrm{SUV}_{\mathrm{max}}$

TABLE III. Feature composition for best classifiers.

	Best feature set	Method	Correlated features removed?
Local relapse $\mathcal{T}$	$IVH^{T}_{0.75}$, IVH^{T}_{1}, $IVH^{T}_{1.75}$, $IVH^{T}_{2.25}$, SUV^{T}_{peak}, Vol^{T}_{PET}, Vol^{T}_{CT}, $Vol^{T}_{0.5max}$	RFE	Yes
Local relapse $\mathcal{T} \cup \mathcal{N}$	Vol^{T}_{CT}, $IVH^{N}_{2.5}$, $IVH^{N}_{2.75}$, $IVH^{N}_{3.25}$, $IVH^{N}_{3.5}$	RFE	Yes
Overall relapse $\mathcal{T}$	$IVH^{T}_{0.5}$	RFE	Yes
Overall relapse $\mathcal{T} \cup \mathcal{N}$	$Vol^{N}_{0.5max}$	Spearman	No

We also investigated the removal of correlated features and the use of feature selection within each fold of CV. We calculated the correlation matrix for all 37 features over the 25 examples and removed features that had a correlation greater than 0.75 with another feature.

We then repeated this CV with a reshuffled dataset 100 times and subsequently generated a histogram of the frequency each set of features was chosen. The feature set that had the highest frequency was the final set of features recommended based on this methodology and dataset. The best classifiers were chosen based on their MCC value.

Confidence intervals were constructed using bootstrap resampling methods. Hypothesis testing was conducted using the same bootstrap resampling method.[27] Classifiers were ranked on their MCC value, but we also present their area under the curve (AUC), sensitivity, and specificity values.

3. RESULTS

3.A. Feature composition of best classifiers

Table III shows the feature composition for each of the best models, and Table IV shows the corresponding metrics for the models from Table III. Figure 2 shows the histogram of the top five features for local relapse. For local relapse, the best classifier was composed of eight tumor features and had a MCC of 0.07. For overall relapse, the best classifier was composed of a single nodal feature and had a MCC of 0.29 ($Vol^{N}_{0.5max}$).

3.B. The value of nodal information

Figure 3 shows the histogram of the top five features for overall relapse. For local relapse, restricting the classifier to only tumor features resulted in a better classifier; indeed, the $\mathcal{T}$ classifier had a MCC of 0.07, while the $\mathcal{T} \cup \mathcal{N}$ classifier decreased to a MCC of −0.07 ($p = 0$). The AUC and sensitivity also decreased when going from the best $\mathcal{T}$ classifier to the best $\mathcal{T} \cup \mathcal{N}$ classifier for local relapse. In contrast, for overall relapse, allowing the classifier to consider nodal features increased the MCC by roughly a factor of three, from 0.1 to 0.29 ($p = 0$). The AUC, sensitivity, and specificity also increased.

4. DISCUSSION

For local relapse, the $\mathcal{T}$ classifier demonstrated a higher MCC of 0.07 than the $\mathcal{T} \cup \mathcal{N}$ classifier with a MCC of −0.07. The $\mathcal{T} \cup \mathcal{N}$ classifier had a negative MCC value, meaning that it more often predicted the opposite of what the true class was; in this case, it seems that adding nodal information confounded the classifier. This could have been due to the poor class imbalance for local relapse and the small number of examples.

For overall relapse, we see that the $\mathcal{T} \cup \mathcal{N}$ classifier achieved a higher MCC (0.29) than the $\mathcal{T}$ classifier (0.10). The top single $\mathcal{T} \cup \mathcal{N}$ feature was chosen as $Vol^{N}_{0.5max}$. Using nodal information can significantly improve the classifier when predicting overall relapse, which is consistent with recent correlative studies that also confirm the prognostic value of nodal SUV in NSCLC.[28]

When the classifier's features were chosen using RFE, this resulted in the best model for all classifiers except overall relapse $\mathcal{T} \cup \mathcal{N}$ classifier, which performed the best with the Spearman-based approach. Also, the overall relapse $\mathcal{T} \cup \mathcal{N}$ classifier performed best, even when no correlated features were removed during the preprocessing stage. Despite high intervariable correlation, certain combinations of variables can complement each other when used together.[29] This may

TABLE IV. Mean CV metrics and confidence intervals after 100 repetitions of CV for best classifiers.

	Local relapse		Overall relapse	
	$\mathcal{T}$	$\mathcal{T} \cup \mathcal{N}$	$\mathcal{T}$	$\mathcal{T} \cup \mathcal{N}$
Num features	8	5	1	1
Mean MCC [CI]	0.07 [0.04, 0.11]	−0.07 [−0.11, −0.04]	0.10 [0.06, 0.13]	0.29 [0.26, 0.32]
Mean AUC [CI]	0.55 [0.53, 0.58]	0.44 [0.41, 0.47]	0.52 [0.50, 0.53]	0.69 [0.68, 0.71]
Mean sensitivity [CI]	0.57 [0.53, 0.61]	0.34 [0.31, 0.38]	0.56 [0.53, 0.59]	0.57 [0.55, 0.59]
Mean specificity [CI]	0.52 [0.50, 0.53]	0.58 [0.56, 0.60]	0.48 [0.45, 0.52]	0.63 [0.60, 0.65]

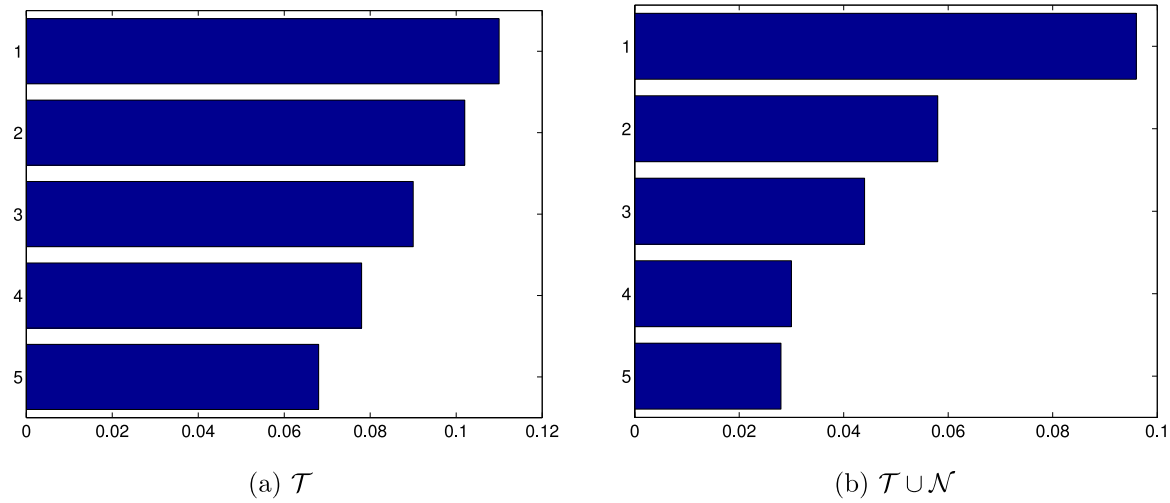


FIG. 2. Histogram of top five feature sets across 100 repetitions of CV for local relapse. Full label names are located in the Appendix (Tables V and VI).

explain why the overall relapse classifier performed the best by including all the features.

The poor performance in predicting local relapse could also be explained by the variability in defining of local relapse or in-field failure.[30] Also, all the patients who had a local recurrence in our study were diagnosed with a recurrence only based on imaging; no pathologic sampling was performed. Using CT alone to diagnose recurrence can be inaccurate due to the post-treatment effects and this phenomenon is well established in the literature.[31]

4.A. Dataset

Our feature dataset was extracted from 4DPET/4DCT images. Respiratory motion causes the max and peak SUV to decrease on 3DPET images, and the use of 4DPET alleviates this motion blur to provide sharper images.[32] Although 4DPET is not standard-of-care, nor is it available at most institutions, we suspect that the use of 4DPET features improved the quality of our prediction over the use of 3DPET features alone. 4DCT also helps with the accurate contouring of the mask. The inhale and exhale phases of the PTV were contoured and the ITV was generated from that. Most of our features were derived from the ITV, which is the GTV plus some margin, so the margins are large enough that we can be sure the features are robust to the autocontouring process. The contours were also reviewed afterward by a single expert to ensure consistency and that no other organs were included in the ITV contours.

The patient population used in this study was composed of 25 patients, so we were limited to using cross-validation methods to measure the effectiveness of our classifiers. The results of this study should be validated on a larger independent dataset to evaluate the effectiveness of the classifiers and the relevance of the features selected.[33]

Our ongoing work is focused on quantifying the additional benefit that 4DPET scans provide over 3DPET scans with these machine learning methods. We are also studying the change of 4DPET/4DCT features during treatment through the use of serial imaging, and any improvements this could bring to our models. Second-order image features such as contrast, correlation, homogeneity, and entropy are also being investigated.

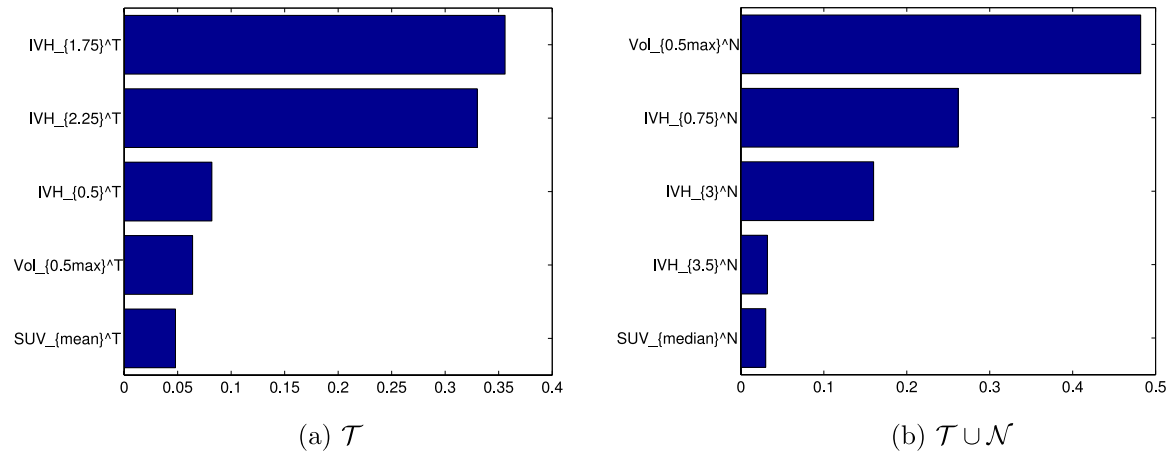


FIG. 3. Histogram of top five features across 100 repetitions of CV for overall relapse.

5. CONCLUSION

Our results showed that it is possible to predict local relapse and overall relapse at the two-year time point. For local relapse, the best classifier found through feature selection methods was composed of eight tumor features. For overall relapse, the best classifier found through feature selection methods was composed of a single nodal feature: the volume greater than 0.5 times the maximum SUV (N).

The results demonstrated that adding nodal features to a classifier improved MCC for predicting overall relapse. In addition, adding nodal features to a classifier did not improve MCC for predicting local relapse. Future work will explore higher order statistics and time trends in the hope of identifying features to facilitate individualized care and help improve outcomes for those patients.

ACKNOWLEDGMENTS

We would like to thank Dr. Katy Clarke, Dr. Victoria Ford, Dr. Paula McCloskey, and Dr. Mei Ling Yap for the CT contours and Dr. Vladimir Pekar for the automated PET thresholding algorithms. We would also like to thank Dr. Issam El Naqa for his advice. We also thank the High Performance Computing Virtual Laboratory (HPCVL) for providing the computational infrastructure used in this paper. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR) through the Collaborative Health Research Projects (CHRP) Grant No. 398106-2011.

APPENDIX: FULL FEATURE SETS FOR LOCAL RELAPSE

TABLE V. Full list of features for the top five $\mathcal{T}$ feature sets for local relapse.

Rank	Feature set
1	$\mathrm{IVH}^T_{0.75}$, IVH^T_{1}, $\mathrm{IVH}^T_{1.25}$, $\mathrm{IVH}^T_{1.75}$, SUV_T, $\mathrm{Vol}^T_{\mathrm{PET}}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$
2	$\mathrm{IVH}^T_{0.5}$, $\mathrm{IVH}^T_{0.75}$, IVH^T_{1}, $\mathrm{IVH}^T_{1.25}$, $\mathrm{SUV}^T_{\mathrm{peak}}$, $\mathrm{Vol}^T_{\mathrm{PET}}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$
3	$\mathrm{IVH}^T_{0.75}$, IVH^T_{1}, $\mathrm{IVH}^T_{1.25}$, $\mathrm{IVH}^T_{2.25}$, $\mathrm{SUV}^T_{\mathrm{peak}}$, $\mathrm{Vol}^T_{\mathrm{PET}}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$
4	$\mathrm{IVH}^T_{0.5}$, $\mathrm{IVH}^T_{0.75}$, $\mathrm{IVH}^T_{1.75}$, $\mathrm{IVH}^T_{2.25}$, SUV_T, $\mathrm{SUV}^T_{\mathrm{peak}}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$
5	$\mathrm{IVH}^T_{0.75}$, IVH^T_{1}, $\mathrm{IVH}^T_{1.75}$, $\mathrm{IVH}^T_{2.25}$, SUV_T, $\mathrm{SUV}^T_{\mathrm{peak}}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$

TABLE VI. Full list of features for the top five $\mathcal{T} \cup \mathcal{N}$ feature sets for local relapse.

Rank	Feature set
1	IVH^T_{1}, $\mathrm{IVH}^T_{1.75}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{IVH}^N_{2.5}$, $\mathrm{IVH}^N_{3.5}$
2	$\mathrm{IVH}^T_{0.5}$, $\mathrm{IVH}^N_{2.5}$, $\mathrm{IVH}^N_{2.75}$, $\mathrm{IVH}^N_{3.25}$, $\mathrm{IVH}^N_{3.5}$
3	$\mathrm{IVH}^T_{0.75}$, $\mathrm{Vol}^T_{\mathrm{CT}}$, $\mathrm{Vol}^T_{0.5\mathrm{max}}$, $\mathrm{IVH}^N_{2.5}$, $\mathrm{IVH}^N_{3.5}$
4	$\mathrm{Vol}^T_{\mathrm{CT}}$, IVH^N_{2}, $\mathrm{IVH}^N_{2.5}$, $\mathrm{IVH}^N_{3.25}$, $\mathrm{IVH}^N_{3.5}$
5	$\mathrm{Vol}^T_{0.5\mathrm{max}}$, $\mathrm{IVH}^N_{2.75}$, IVH^N_{3}, $\mathrm{IVH}^N_{3.25}$, $\mathrm{IVH}^N_{3.5}$

[a]Author to whom correspondence should be addressed. Electronic mail: heyse.li@mail.utoronto.ca

[1]D. De Ruysscher *et al.*, "Individualised isotoxic accelerated radiotherapy and chemotherapy are associated with improved long-term survival of patients with stage III NSCLC: A prospective population-based study," Radiother. Oncol. **102**, 228–233 (2012).
[2]RTOG 1106 Protocol Information, http://www.rtog.org/ClinicalTrials/ProtocolTable/StudyDetails.aspx?study=1106, accessed November 27, 2014.
[3]Dose Escalation by Boosting Radiation Dose Within the Primary Tumor Using FDG-PET-CT Scan in Stage IB, II and III NSCLC (PET Boost), http://clinicaltrials.gov/show/NCT01024829, accessed November 27, 2014.
[4]V. Ahuja, R. E. Coleman, J. Herndon, and E. F. Patz, "The prognostic significance of fluorodeoxyglucose positron emission tomography imaging for patients with nonsmall cell lung carcinoma," Cancer **83**, 918–924 (1998).
[5]R. J. Downey, T. Akhurst, M. Gonen, A. Vincent, M. S. Bains, S. Larson, and V. Rusch, "Preoperative F-18 fluorodeoxyglucose-positron emission tomography maximal standardized uptake value predicts survival after lung cancer resection," J. Clin. Oncol. **22**, 3255–3260 (2004).
[6]R. Sasaki *et al.*, "[18F] fluorodeoxyglucose uptake by positron emission tomography predicts outcome of non–small-cell lung cancer," J. Clin. Oncol. **23**, 1136–1143 (2005).
[7]J. F. Vansteenkiste *et al.*, "Prognostic importance of the standardized uptake value on 18F-fluoro-2-deoxy-glucose–positron emission tomography scan in non–small-cell lung cancer: An analysis of 125 cases," J. Clin. Oncol. **17**, 3201–3206 (1999).
[8]F. M. S. Kong, K. A. Frey, L. E. Quint, R. K. T. Haken, J. A. Hayman, M. Kessler, I. J. Chetty, D. Normolle, A. Eisbruch, and T. S. Lawrence, "A pilot study of 1 fluorodeoxyglucose positron emission tomography scans during and after radiation-based therapy in patients with non–small-cell lung cancer," J. Clin. Oncol. **25**, 3116–3123 (2007).
[9]C. J. Hoekstra *et al.*, "Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non–small-cell lung cancer," J. Clin. Oncol. **23**, 8362–8370 (2005).
[10]C. Pöttgen *et al.*, "Value of 18F-fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in non–small-cell lung cancer for prediction of pathologic response and times to relapse after neoadjuvant chemoradiotherapy," Clin. Cancer Res. **12**, 97–106 (2006).
[11]W. van Elmpt, M. Öllers, A. M. C. Dingemans, P. Lambin, and D. De Ruysscher, "Response assessment using 18F-FDG PET early in the course of radiotherapy correlates with survival in advanced-stage non–small cell lung cancer," J. Nucl. Med. **53**, 1514–1520 (2012).
[12]W. A. Weber, V. Petersen, B. Schmidt, L. Tyndale-Hines, T. Link, C. Peschel, and M. Schwaiger, "Positron emission tomography in non–small-cell lung cancer: Prediction of response to chemotherapy by quantitative assessment of glucose use," J. Clin. Oncol. **21**, 2651–2657 (2003).
[13]J. L. L. Guerra, G. Gladish, R. Komaki, D. Gomez, Y. Zhuang, and Z. Liao, "Large decreases in standardized uptake values after definitive radiation are associated with better survival of patients with locally advanced non–small cell lung cancer," J. Nucl. Med. **53**, 225–233 (2012).
[14]R. J. Hicks, M. P. Mac Manus, J. P. Matthews, A. Hogg, D. Binns, D. Rischin, D. L. Ball, and L. J. Peters, "Early FDG-PET imaging after radical radiotherapy for non–small-cell lung cancer: Inflammatory changes in normal tissues correlate with tumor response and do not confound therapeutic response evaluation," Int. J. Radiat. Oncol., Biol., Phys. **60**, 412–418 (2004).
[15]M. P. Mac Manus, R. J. Hicks, J. P. Matthews, A. Wirth, D. Rischin, and D. L. Ball, "Metabolic (FDG–PET) response after radical radiotherapy/chemoradiotherapy for non-small cell lung cancer correlates with patterns of failure," Lung Cancer **49**, 95–108 (2005).
[16]C. Dehing-Oberije *et al.*, "Development and external validation of prognostic model for 2-year survival of non–small-cell lung cancer patients treated with chemoradiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **74**, 355–362 (2009).
[17]I. E. Naqa *et al.*, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," Pattern Recognit. **42**, 1162–1171 (2009).
[18]I. E. Naqa, J. D. Bradley, P. E. Lindsay, A. J. Hope, and J. O. Deasy, "Predicting radiotherapy outcomes using statistical learning techniques," Phys. Med. Biol. **54**(18), S9–S30 (2009).
[19]M. Vaidya, K. M. Creach, J. Frye, F. Dehdashti, J. D. Bradley, and I. E. Naqa, "Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer," Radiother. Oncol. **102**, 239–245 (2012).

[20]R. G. P. M. van Stiphout *et al.*, "Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging," Radiother. Oncol. **98**, 126–133 (2011).
[21]H. Zhang, S. Tan, W. Chen, S. Kligerman, G. Kim, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal ^{18}F-FDG PET features, clinical parameters, and demographics," Int. J. Radiat. Oncol., Biol., Phys. **88**, 195–203 (2014).
[22]G. J. Kubicek, C. Champ, S. Fogh, F. Wang, E. Reddy, C. Intenzo, R. W. Dusing, and M. Machtay, "FDG-PET staging and importance of lymph node SUV in head and neck cancer," Head Neck Oncol. **2**, 19 (2010).
[23]T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagation for classification," Int. J. Comput. Theory Eng. **3.1**, 1793–8201 (2011).
[24]R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: A library for large linear classification," J. Mach. Learn. Res. **9**, 1871–1874 (2008).
[25]B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," Biochim. Biophys. Acta, Protein Struct. **405**, 442–451 (1975).
[26]I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Mach. Learn. **46**, 389–422 (2002).
[27]J. Fox, *Applied Regression Analysis and Generalized Linear Models* (Sage Publications, 2008).
[28]V. H. F. Lee, W. W. Chan, E. Y. P. Lee, T. S. Choy, P. P. Ho, D. K. Leung, K. O. Lam, D. L. W. Kwong, T. W. Leung, and P. L. Khong, "Prognostic significance of standardized uptake value of lymph nodes on survival for stage iii non-small cell lung cancer treated with definitive concurrent chemoradiotherapy," Am. J. Clin. Oncol. (2014) [Epub ahead of print].
[29]I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res. **3**, 1157–1182 (2003).
[30]M. Machtay, R. Paulus, J. Moughan, R. Komaki, J. Bradley, H. Choy, K. Albain, B. Movsas, W. T. Sause, and W. J. Curran, "Defining local-regional control and its importance in locally advanced non-small cell lung carcinoma: An rtog analysis," J. Thorac. Oncol. **7**, 716–722 (2012).
[31]H. G. Colt, S. D. Murgu, R. J. Korst, C. G. Slatore, M. Unger, and S. Quadrelli, "Follow-up and surveillance of the patient with lung cancer after curative-intent therapy: Diagnosis and management of lung cancer: American college of chest physicians evidence-based clinical practice guidelines," Chest J. **143**, e437S–e454S (2013).
[32]M. Aristophanous, R. I. Berbeco, J. H. Killoran, J. T. Yap, D. J. Sher, A. M. Allen, E. Larson, and A. B. Chen, "Clinical utility of 4d fdg-pet/ct scans in radiation treatment planning," Int. J. Radiat. Oncol., Biol., Phys. **82**, e99–e105 (2012).
[33]P. Lambin *et al.*, "Predicting outcomes in radiation oncology multifactorial decision support systems," Nat. Rev. Clin. Oncol. **10**, 27–40 (2012).