



Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

Darcie A. P. Delzell^{1*}, Sara Magnuson¹, Tabitha Peter¹, Michelle Smith¹ and Brian J. Smith²

¹ Department of Mathematics and Computer Science, Wheaton College, Wheaton, IL, United States, ² Department of Biostatistics, University of Iowa, Iowa City, IA, United States

OPEN ACCESS

Edited by:

Lei Deng,
Jacobi Medical Center, United States

Reviewed by:

Seyedmehdi Payabvash,
Yale School of Medicine, United States

Jongphil Kim,
Moffitt Cancer Center, United States

Jinhua Yu,
Fudan University, China

Hui Li,
University of Chicago, United States

*Correspondence:

Darcie A. P. Delzell
darcie.delzell@wheaton.edu

Specialty section:

This article was submitted to
Cancer Imaging and Image-directed
Interventions,
a section of the journal
Frontiers in Oncology

Received: 30 April 2019

Accepted: 26 November 2019

Published: 11 December 2019

Citation:

Delzell DAP, Magnuson S, Peter T,
Smith M and Smith BJ (2019)
Machine Learning and Feature
Selection Methods for Disease
Classification With Application to Lung
Cancer Screening Image Data.
Front. Oncol. 9:1393.
doi: 10.3389/fonc.2019.01393

As awareness of the habits and risks associated with lung cancer has increased, so has the interest in promoting and improving upon lung cancer screening procedures. Recent research demonstrates the benefits of lung cancer screening; the National Lung Screening Trial (NLST) found as its primary result that preventative screening significantly decreases the death rate for patients battling lung cancer. However, it was also noted that the false positive rate was very high (>94%). In this work, we investigated the ability of various machine learning classifiers to accurately predict lung cancer nodule status while also considering the associated false positive rate. We utilized 416 quantitative imaging biomarkers taken from CT scans of lung nodules from 200 patients, where the nodules had been verified as cancerous or benign. These imaging biomarkers were created from both nodule and parenchymal tissue. A variety of linear, nonlinear, and ensemble predictive classifying models, along with several feature selection methods, were used to classify the binary outcome of malignant or benign status. Elastic net and support vector machine, combined with either a linear combination or correlation feature selection method, were some of the best-performing classifiers (average cross-validation AUC near 0.72 for these models), while random forest and bagged trees were the worst performing classifiers (AUC near 0.60). For the best performing models, the false positive rate was near 30%, notably lower than that reported in the NLST. The use of radiomic biomarkers with machine learning methods are a promising diagnostic tool for tumor classification. They have the potential to provide good classification and simultaneously reduce the false positive rate.

Keywords: radiomics, machine learning, CT image, biomarkers, lung cancer

1. INTRODUCTION

Publication of primary results from the National Lung Screening Trial (NLST) reported that lung cancer screening, especially when performed with low dose computed tomography (CT) scans, can significantly reduce the mortality rate of lung cancer. This result highlights the benefits of lung cancer screening; however, the NLST also found that screening results had a notably high rate of false positive results. Of the total number of low dose CT scans in the NLST, the false

positive rate surpassed 94% (1). The NLST researchers noted that the high false positive rate was a challenge which required further research, and that challenge persists to the present. The negative consequences associated with false positive exam results can include patient anxiety and unnecessary invasive diagnostic procedures such as biopsy (2, 3).

High-throughput extraction of features from imaging data composes the essence of radiomics, an emerging field of research which offers significant improvement to decision-support in oncology (4, 5). Current work examines the predictive power of quantitative imaging biomarkers, which are quantitative features extracted from routine medical images (4, 6, 7), as inputs within predictive classifying models. The information contained in the imaging biomarkers has the potential to improve classification accuracy in a variety of statistical models (2).

Across the literature, quantitative biomarkers taken from imaging data have been used to develop models with the intent to identify and analyze associations between radiomic/nodule features (stages or histological characteristics) and clinical outcomes (survival, recurrence, etc.). Previous work in radiomics aimed at classification of lung nodules has examined a variety of outcomes (5, 8–12). Zhu et al. used outcome categories for lung cancer type with a LASSO classification model (13). Zhang et al. examined outcomes for local/distant failure using several machine learning classifiers (5). Pamar et al. used clusters of biomarkers as predictors in models of overall survival (14). Dilger et al. used an expanded set of radiomic features that included both nodule and parenchymal tissue. They showed an increase in classification performance when the parenchymal tissue was included in feature extraction (3).

In this paper, we investigate the predictive power of biomarkers (computed from both nodule and parenchymal tissue as calculated by Dilger et al. (3)) to classify lung nodule status as malignant/benign while also considering the false positive rate. Our comprehensive approach includes multiple combinations of models and filtering techniques. In particular, combinations of twelve machine learning classifiers along with six feature selection methods were compared, using area under the receiver operating characteristic curve (AUC) as the model performance metric.

2. METHODS

2.1. Dataset

This retrospective study analyzed data originally taken from 200 CT scans of the lungs of patients at the University of Iowa Hospital. Pathology and radiology reports were reviewed to identify an analysis set of patients who met eligibility criteria of having (a) a solitary lung nodule (5–30 mm) and (b) a malignant nodule confirmed on histopathology or a benign nodule confirmed on histopathology or by size stability for at least 24 months. Manual segmentations were performed by a graduate student trained in medical image analysis in order to define a region of interest (ROI) around each nodule. The ROIs were defined to include amounts of parenchyma approximately proportional to the nodule sizes. Individual ROI voxels were labeled as belonging to either the nodule or the

TABLE 1 | Demographics of patient cohort.

	Malignant	Benign
Number of patients	110	90
Female	51 (46.4%)	63 (70.0%)
Male	59 (53.6%)	27 (30.0%)
Age, yrs (mean \pm SD)	65.7 \pm 11.2	58.2 \pm 13.2
Pack-years (mean \pm SD)	38.4 \pm 31.2	11.2 \pm 16.9
Nodule size, mm (range, mean \pm SD)	7 – 44, 19.1 \pm 6.3	6 – 30, 15.2 \pm 5.8

parenchyma, with radiomic features calculated separately for each to produce the complete set of 416 (approximately half nodule and half parenchyma) quantitative imaging biomarkers. These biomarkers measured features such as intensity, shape, and texture of the ROI (15). This study is a secondary analysis of de-identified data originally collected with approval from the University of Iowa institutional review board. Demographic information can be found in **Table 1**.

A strength of the dataset is its fairly balanced malignant/benign status breakdown, with 45% of the cases malignant and 55% benign. Many machine learning-based classifying algorithms assume that the outcomes of a data set are balanced, but this assumption is not met when the proportion of outcomes is highly uneven. The data set used in this work has a nearly even ratio of malignant and benign nodules (16).

2.2. Radiomic Features

The 416 radiomic features which were available for this investigation quantified nodule characteristics from CT images acquired from a variety of scanner protocols through the University of Iowa Hospital. The most common CT models used were Siemens SOMATOM Definition, Siemens Sensation 16, Sensation Biograph 40, and Toshiba Aquilion. Using these machines, several protocols were used, including Chest CT scans with and without contrast, CT Angiography scans, Extrenal CT scans, PET/CT scans, and CT: Chest, Abdomen, and Pelvis scans. Slice thickness ranged from 1.0 to 6.0 mm with an average of 3.3 mm (15). From these scans, voxels labeled as parenchyma and nodule were used in the extraction of four classes of features: intensity, shape, border, and texture. The intensity of CT images described the radiodensity of the anatomy [measured using Hounsfield units (HU)] as well as heterogeneity of the nodule. Shape features examined sphericity and the maximum diameter of the nodule. Sphericity was computed by comparing the volume of the nodule to its surface area, and maximum diameter was measured using the Response Evaluation Criteria in Solid Tumors (RECIST). The border features were measured using a rubber band straightening transform (RBST). The texture features were extracted from the nodule and parenchyma regions using Laws' Texture Energy Measures (TEM). From these TEMs, the mean, variance, kurtosis, and skewness of the nodule and parenchyma were extracted. Radiomic features were extracted using a Matlab based CAD tool, and the mathematical definitions for all of the radiomic measurements are described in full in Dilger (17).

TABLE 2 | Summary of feature selection methods.

Feature selection method	Abbreviation
Linear combination	lincom
Pairwise correlation	corr.95
PCA - 0.85 cutoff	pca.85
PCA - 0.90 cutoff	pca.90
PCA - 0.95 cutoff	pca.95
Unfiltered	nofilter

2.3. Feature Selection Methods

As is common in radiomics studies with hundreds of features, many of the biomarkers (features) used as predictors were highly correlated with one another; this challenge necessitated feature selection in order to avoid collinearity, reduce dimensionality, and minimize noise (11, 16, 18, 19). To this end, we considered three feature selection methods: a linear combinations filter, a pairwise correlation filter, and principle component analysis.

For the linear combinations filter (lincom), a QR decomposition along with an iterative procedure is used to determine if some predictors are linear combinations of others. Predictors are sequentially removed until the design matrix is full rank. The pairwise correlation filter removes those predictors whose pairwise correlation is greater than a specified cutoff. The two predictors with the largest absolute correlation are first considered. Of those two, the predictor with the highest average absolute correlation with all other variables is removed. This process continues until all the predictors left have pairwise absolute correlations less than the cutoff. After investigating multiple cutoffs, we chose a cutoff value of 0.95 for the pairwise correlation filter (corr.95) since this cutoff removed highly correlated variables but still retained a large number of features. Principal component analysis reduces dimensionality by creating new, uncorrelated predictors which explain a large proportion of the variance in the predictor space. Principal component analysis was implemented at three different cutoffs (pca.85, pca.90, pca.95), where the number of components accounted for either 85, 90, or 95% of the variance in the predictor space (Table 2).

2.4. Classifiers and Performance Metrics

Combinations of the six feature selection methods and twelve classifiers were investigated by implementing a 10-fold repeated cross-validation framework with five repeats, a standard validation technique (5, 13, 16, 20, 21). The feature selection methods were included in the cross-validation algorithm so that their contribution to the final model fit is reflected in the performance metrics. The classifiers are from three different families: linear, nonlinear, and ensemble (22). Of the linear classifiers, an elastic net (elasticnet), a logistic regression (logistic), a partial least squares model (pls), and a logistic regression with Step AIC were fit. The nonlinear classifiers include a K-nearest neighbors model (knn), a neural network (nnet), and three support vector machines: a linear kernel (svml), a polynomial kernel (svmpoly), and a radial kernel (svmr). The ensemble models used included bagged classification

TABLE 3 | Summary of classifiers.

Model family	Classifier	Abbreviation
Linear	Elastic net	elasticnet
	Logistic regression	logistic
	Partial least squares	pls
	Logistic regression with step AIC	glmStepAIC
	K-nearest neighbors	knn
Nonlinear	Neural network	nnet
	Support vector machine (linear kernel)	svml
	SVM (polynomial kernel)	svmpoly
	SVM (radial kernel)	svmr
	Bagged trees	bag
Ensemble	Random forest	rf
	Stochastic gradient boosting	gbm

trees (bag), random forest (rf), and stochastic gradient boosting (gbm) (Table 3).

The quality of model performance in most machine learning algorithms is dependent upon the choice of various tuning parameters. Some tuning parameters take into account the number of predictors after feature selection. For example, the *mtry* tuning parameter for rf, which determines the number of candidate variables at each branch, is equal to the square root of the number of predictors. Other tuning parameters were chosen based on standard practice (22, 23). For example, the decay tuning parameter for nnet, which helps prevent overfitting, generally takes the values of 0.1, 0.01, and 0.001. All models were fit using the caret R package (24). Our R code implementing the feature selection and classification models is presented as **Supplementary Material**.

3. RESULTS

The linear combinations filter removed 217 biomarkers, leaving a set of 199 predictors. The pairwise correlation filter retained 39 predictors, while principal components analysis retained 12, 14, and 18 components at the 85, 90, and 95% levels, respectively.

Figure 1 gives the predictive performance (AUC) of each feature selection method (in rows) and classifier (in columns), averaged over the 50-folds/repeats in the cross-validation. Logistic regression models cannot be calculated when the number of predictors is larger than the number of observations, so the nofilter row is blank for this classifier. The large number of predictors also caused multiple computing issues with the neural net classifier, so training this classifier without using any feature selection was not considered. Table 4 gives the highest average AUC for each classifier across the various feature selection methods. Principal component analysis yields lower AUC values for all of the classifying models. Using lincom, the top four classification methods perform well, with $AUC \geq 0.728$ (we note that svmr with corr.95 also has an average $AUC = 0.728$). The standard deviation over the folds/repeats is also given, along with sensitivity, specificity, and false positive rate statistics. Specificity and sensitivity were computed using a 0.5 threshold from the

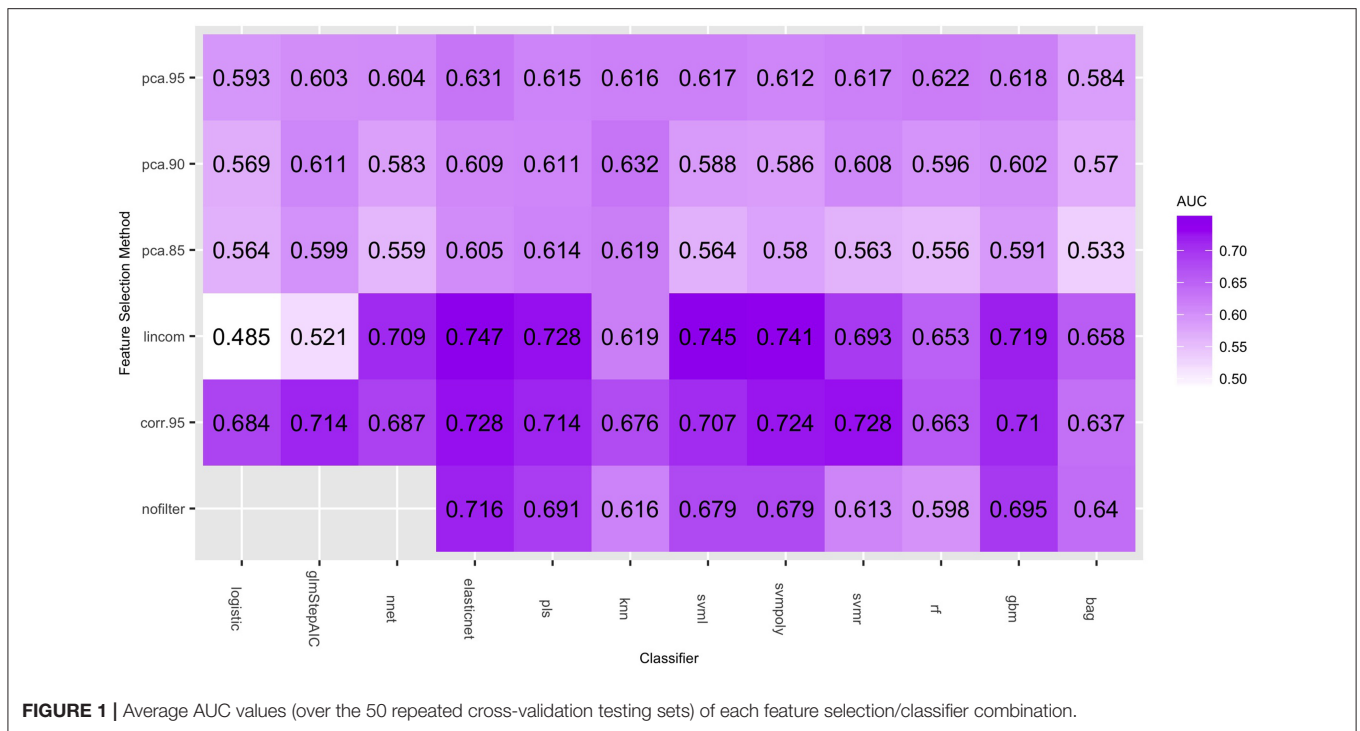


FIGURE 1 | Average AUC values (over the 50 repeated cross-validation testing sets) of each feature selection/classifier combination.

TABLE 4 | AUC values for classifiers with highest predictive performance (SD taken over the 50 cross-validation testing sets).

Classifier	Feature method	AUC	SD	Sensitivity	Specificity	False positive rate	SD
elasticnet	lincom	0.747	0.111	0.616	0.729	0.271	0.136
svml	lincom	0.745	0.112	0.549	0.765	0.235	0.126
svmpoly	lincom	0.741	0.113	0.569	0.781	0.219	0.132
pls	lincom	0.728	0.111	0.627	0.707	0.293	0.126
svmr	corr.95	0.728	0.106	0.542	0.780	0.220	0.148
gbm	lincom	0.714	0.106	0.596	0.733	0.267	0.140
glmStepAIC	corr.95	0.714	0.110	0.636	0.684	0.316	0.130
nnet	lincom	0.709	0.113	0.620	0.707	0.293	0.143
logistic	corr.95	0.684	0.108	0.600	0.689	0.311	0.116
knn	corr.95	0.676	0.109	0.482	0.738	0.262	0.117
rf	corr.95	0.663	0.124	0.473	0.730	0.270	0.127
bag	lincom	0.658	0.106	0.529	0.702	0.298	0.146

model predicted class probabilities. The AUC standard deviations are fairly similar, while sensitivity and specificity have larger variation. The false positive rates are more variable than the AUC values, and the mean false positive rates are all notably lower (all less than 32%) than the 94% found in the results of the NLST.

Figure 2 shows the distribution of the AUC scores for the four best performing classifiers: elasticnet, svml, svmpoly, and pls. Among all feature selection methods, corr.95 and lincom yielded the highest AUC values on average across these four classifiers.

The lincom feature selection with the elasticnet classifier has the best overall predictive performance (AUC = 0.747), followed by the svml classifier with the lincom feature selection (AUC = 0.745). As has been observed in other radiomic studies, support vector machines perform well with respect to predictive performance (21).

The boxplots in Figure 3 show the distribution of the false positive rates for the four best performing classifiers. These distributions show that the lowest false positive rates were achieved in combination with either the lincom or corr.95 feature selection methods for all four of these classifiers. These two feature selection methods result in both the highest average AUC values and the lowest false positive rates.

Figure 4 gives the ROC curve for the best performing classifier/feature selection combination (elasticnet/lincom). Although the NLST did not report false negative rates, the ROC curve displays the tradeoff between specificity and sensitivity. While the classifiers have reduced the false positive rate, the tradeoff is an increase in the false negative rate, which would be estimated to be near 0.38 for this particular classifier. This natural tradeoff between specificity and sensitivity for classifiers would suggest that radiomic methods should not be the sole diagnostic tool in lung cancer diagnosis. However, the reduction of the false positive rate for a non-invasive procedure is a substantial improvement and supports the inclusion of these methods in clinical practice.

4. DISCUSSION

While awareness of the benefits of preventative screening for lung cancer has increased in recent years, there is still a need

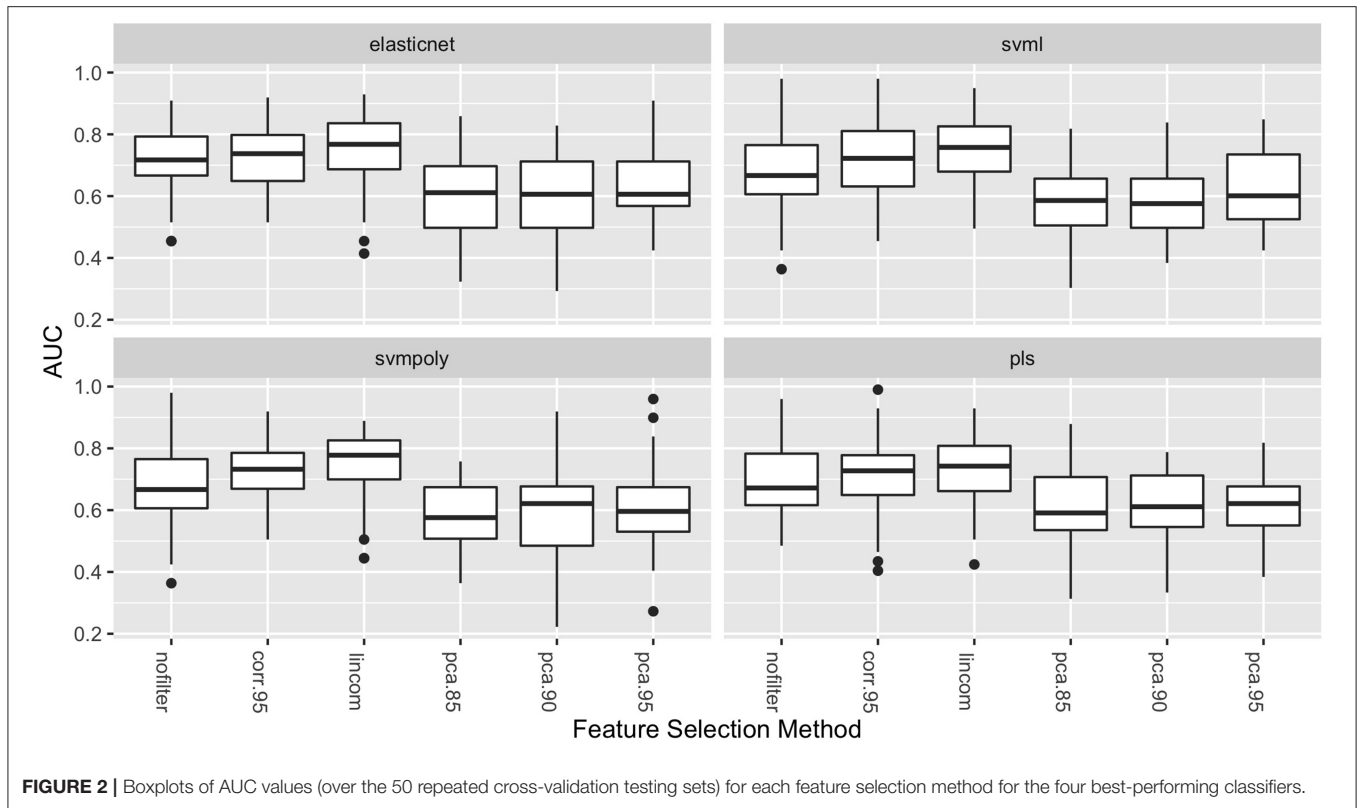


FIGURE 2 | Boxplots of AUC values (over the 50 repeated cross-validation testing sets) for each feature selection method for the four best-performing classifiers.

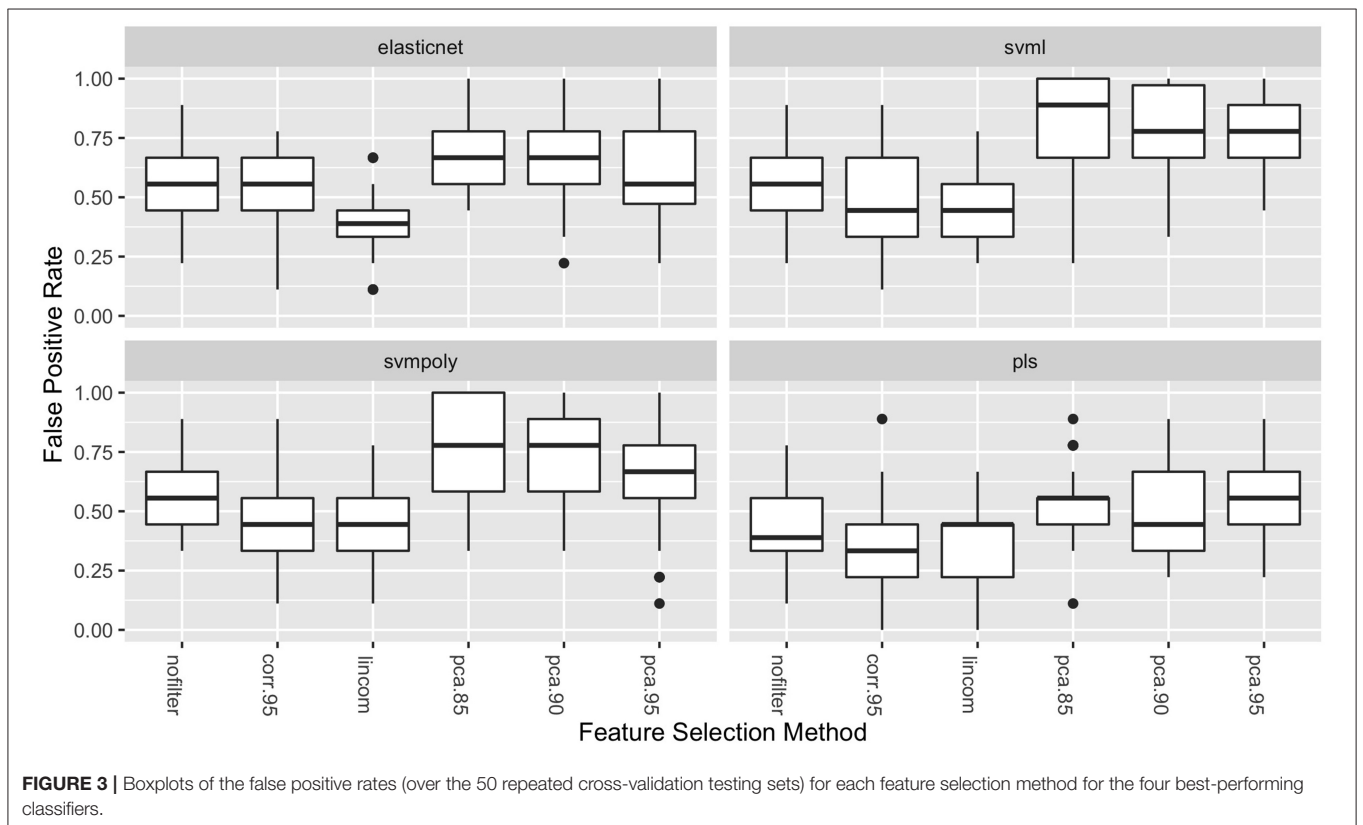


FIGURE 3 | Boxplots of the false positive rates (over the 50 repeated cross-validation testing sets) for each feature selection method for the four best-performing classifiers.

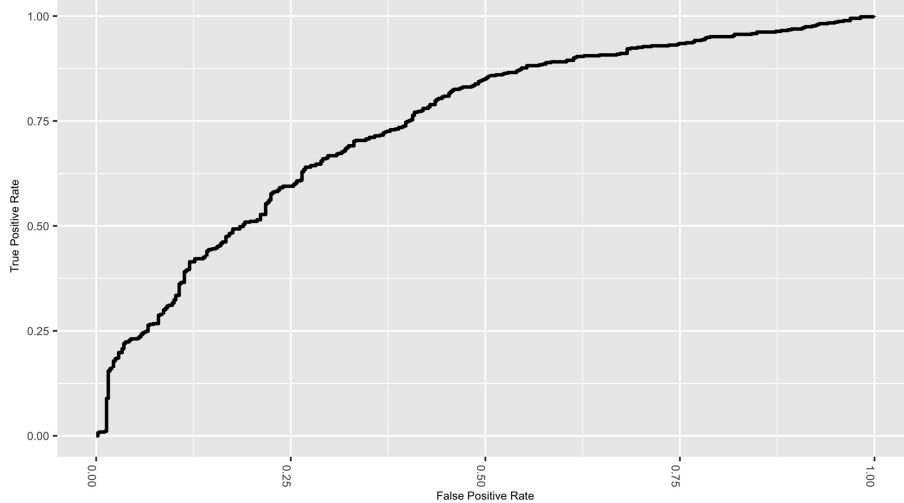


FIGURE 4 | ROC curve for the elastic net classifier with the linear combinations filter.

for improved accuracy in nodule classification. Moreover, a high false positive rate for the diagnostic outcome of lung cancer screening remains a major challenge. Nodule characteristics (biomarkers) calculated from CT scans offer the possibility of improved nodule classification through various modeling techniques. Machine learning algorithms have the potential to harness the predictive power in nodule characteristics. However, little work has been done to compare the performance of various machine learning methods used in conjunction with different feature selection methods, especially as they relate to lung cancer tumor diagnosis.

However, models to predict pulmonary nodule status have been developed and evaluated in other studies. Chen et al. extracted 750 imaging features and compared the performance of a support vector machine (SVM) trained with all to an SVM trained with a sequential forward selection of 4 features (2). Leave-one-out cross-validation demonstrated superior accuracy of 84% for the 4-feature model vs. 56% for all features. Alahmari et al. studied the prognostic performance of radiomics features and found the addition of feature changes over time (delta radiomics) to improve AUC performance from 0.773 to 0.822 (25). SVM and random forest models as well as different feature selection algorithms were considered in their analysis. Final results are presented for random forest models and ReliefF feature selection, suggesting that these were the optimal choices, although comparisons to the others were not presented. A computer-aided lung nodule detection system was proposed by Ma et al. (26). In their approach, multiscale nodule and vessel enhancement filters were applied to patient images prior to extracting 979 radiomics features for training of a random forest classifier. Comparisons to other modeling approaches were not made. Uthoff et al. used a set of 922 radiomics features that is an extension of ours with both nodule features and parenchyma features calculated in 25, 50, 75, and 100% bands around the maximal in-plane diameter of the nodule (27). They used k-medoids clustering to select features for

training of an artificial neural network. K-medoids feature selection is similar in spirit to the high correlation selection approach we used in that both reduce the number of features by selecting representative ones from those that are similar. Comparisons to other modeling approaches are not presented in their publication.

In this study, we considered the ability of nodule biomarkers to accurately predict malignant/benign status. The elastic net, support vector machines with polynomial and linear kernels, and partial least squares were the most predictive classifiers. When combined with the linear combination and correlation feature selection methods, these four classifiers had AUC values comparable in accuracy to the most predictive models studied in previous radiomic analyses (14, 16, 21). Furthermore, we observed that these classifiers greatly reduced the false positive rate from that given in the NLST results.

The observations from this investigation suggest that classifiers such as support vector machines and elastic net perform well with quantitative imaging biomarkers as their predictors. We also show that the chosen feature selection method will impact model performance, and we recommend using linear combination or a correlation-based reduction method over principal components. Different CT modalities and/or different patient population characteristics may yield different results. In order to recommend a particular model for application in a clinical setting, these results would need to be externally validated.

As a comparison, the two best classifier/feature selection combinations were fit with both the 416 biomarkers, as well as the demographic variables of sex, age, and pack-years (the number of packs smoked per day multiplied by the number of years smoked). Elastic Net with the Linear Combination filter had an average AUC of 0.747 (see Table 4) without the demographic variables included. This number was increased to 0.854 when these variables were added. The Linear Support Vector Machine with the Linear Combination filter had an average AUC of

0.745 without the demographic variables included. This number was increased to 0.820 when these variables were added. This suggests that radiomic features, while having good predictive performance, can be enhanced when other patient characteristics are included in the model.

Taken together, a number of common themes emerge from our present work and the past work of others. First, methods that reduce the number of features prior to model training appear to improve predictive performance. We believe this is especially true in the field of radiomics where large numbers of features tend to be highly correlated. Oftentimes, there are many features that do not provide additional information because they are linear combinations of others and may be removed with a linear combination filter. In addition, radiomics features tend to exhibit strong clustering for which high correlation or k-medoid selection seems to improve prediction even when in the cases of models, like random forests and gradient boosting, that perform automatic feature selection. Second, our work suggests that SVM performs well in the radiomics setting and supports its use by others. Furthermore, we found the commonly used random forest model to have poor performance; whereas, the less commonly used in radiomics—but commonly used in genomics—elastic net model was our top performer. Thus, we encourage consideration and reporting of more than one modeling approach in radiomics research. Finally, there is strong evidence that pulmonary features derived from the parenchyma and that reflect changes over time help with prediction. Likewise,

as is the case in many fields, improvements in prediction are often achieved when utilizing subject matter expertise in the development of features and modeling approaches.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

Authors acknowledge financial support from the National Institute of Health (NIH R25HL131467) and the National Cancer Institute (NCI P30CA086862). This research was also supported by the G. W. Aldeen Fund at Wheaton College.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2019.01393/full#supplementary-material>

REFERENCES

1. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality rate with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
2. Chen CH, Chang CK, Tu CY, Liao WC, Wu BR, Chou KT, et al. Radiomic features analysis in computed tomography images of lung nodule classification. *PLoS ONE*. (2018) 13:e0192002. doi: 10.1371/journal.pone.0192002
3. Dilger SK, Uthoff J, Judisch Aea. Improved pulmonary nodule classification utilizing quantitative lung parenchyma features. *J Med Imaging*. (2015) 2:041004. doi: 10.1117/1.JMI.2.4.041004
4. Gillies R, Kinahan P, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169
5. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. (2017) 403:21–7. doi: 10.1016/j.canlet.2017.06.004
6. Kumar V, Gu Y, Basu S. Radiomics: the process and the challenges. *Magn Reson Imaging*. (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
7. Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol*. (2018) 28:4514–23. doi: 10.1007/s00330-018-5463-6
8. Kuruvilla J, Gunavathi K. Lung cancer classification using neural networks for CT images. *Comput Methods Prog Biomed*. (2014) 113:202–9. doi: 10.1016/j.cmpb.2013.10.011
9. Lambin P, Rios-Velazquez E, Leijenaar Rea. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
10. Lin Y, Leng Q, Jiang Z, Guarnera MA, Zhou Y, Chen X, et al. A classifier integrating plasma biomarkers and radiological characteristics for distinguishing malignant from benign pulmonary nodules. *Int J Cancer*. (2017) 141:1240–8. doi: 10.1002/ijc.30822
11. Parmar C, Leijenaar RTH, Grossmann P, Velazquez ER, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci Rep*. (2015) 5:1–10. doi: 10.1038/srep11044
12. Parmar C, Grossmann P, Rietveld Dea. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. *Front Oncol*. (2015) 5:272. doi: 10.3389/fonc.2015.00272
13. Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol*. (2018) 28:2772–8. doi: 10.1007/s00330-017-5221-1
14. Parmar C, Grossmann P, Bussink Jea. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. (2015) 5:13087. doi: 10.1038/srep13087
15. Dilger SKN. *Pushing the Boundaries: Feature Extraction From the Lung Improves Pulmonary Nodule Classification*. Iowa City, IA: University of Iowa (2016).
16. Zhang Y, Oikonomou A, Wong Aea. Radiomics-based prognosis analysis for non small cell lung cancer. *Sci Rep*. (2017) 7:46349. doi: 10.1038/srep46349
17. Dilger SKN. *The Use of Surrounding Lung Parenchyma for the Automated Classification of Pulmonary Nodules*. Iowa City, IA: University of Iowa (2013).
18. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234
19. Lambin P, Leijenaar R, Deist Tea. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. (2017) 14:749. doi: 10.1038/nrclinonc.2017.141
20. Krafft SP, Briere TM, Court LE, Martel MK. The utility of quantitative ct radiomics features for improved prediction of radiation pneumonitis. *Med Phys*. (2018) 45:5317–24. doi: 10.1002/mp.13150
21. Sun T, Wang J, Li Xea. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Comput Methods Prog Biomed*. (2013) 111:519–24. doi: 10.1016/j.cmpb.2013.04.016

22. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York, NY: Springer (2013).
23. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw Articles*. (2008) 28:1–26. doi: 10.18637/jss.v028.i05
24. Kuhn M, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, et al. *caret: Classification and Regression Training*. R package version 6.0-80 (2018). Available online at: <https://CRAN.R-project.org/package=caret>
25. Alahmari SS, Cherezov D, Goldgof DB, Hall LO, Gillies RJ, Schabath MB. Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening. *IEEE Access*. (2018) 6:77796–806. doi: 10.1109/ACCESS.2018.2884126
26. Ma J, Zhou Z, Ren Y, Xiong J, Fu L, Wang Q, et al. Computerized detection of lung nodules through radiomics. *Med Phys*. (2017) 44:4148–58. doi: 10.1002/mp.12331
27. Uthoff J, Stephens MJ, Newell Jr JD, Hoffman EA, Larson J, Koehn N, et al. Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *Med Phys*. (2019) 46:3207–16. doi: 10.1002/mp.13592

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Delzell, Magnuson, Peter, Smith and Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.