Automatic Lung Cancer Detection and Diagnosis Using Hand Crafted and Deep Learning Features

Winona Richey, Naji Khosravan and Ulas Bagci, *Senior Member, IEEE*

***Abstract*—This paper presents a lung nodule detection and classification system which utilizes a combination of hand crafted and deep learning features. Hand crafted features were obtained from modified methods of bag of frequencies, and taxonomic indices. We included a robust radius estimation algorithm that resulted in an average error of 1.29 pixels. Hand crafted features were obtained from 3D low dose CT scans pertaining to lung cancer patients as well as control subjects, and we used the following specific features: the suspected nodule location, radius of the nodule, spectral signatures, taxonomic diversity and distinctness. Deep learning features were obtained with Inception-v3, a pretrained network trained on *ImageNet*, that is currently the state of the art Convolutional Neural Network (CNN) architecture for the ILSVRC challenge. Our work aims to expand upon existing detection and classification methods to improve lung cancer screening accuracies. Promising results indicated the superiority of combined features for lung tumor detection and diagnosis.**

Keywords*—lung nodules, computed tomography (CT), computer-aided detection, nodule detection, nodule characterization, pulmonary nodules, deep learning

I. INTRODUCTION

For decades, lung cancer has been the leading cause of cancer-related deaths. Diagnosing lung cancer early at a localized stage increases 5-year survival rates dramatically from an average of 17% to an average of 54%. Unfortunately, more than half of lung cancers are diagnosed at distant stages, leading to low 5-year survival rates [1].

Low dose computed tomography (CT) scans are currently the leading imaging technique for lung nodule screening, but missing nodules during screening and over-diagnosis are two common problems in current routines [1], [2]. In a recent study by the Visual Attention Lab at Harvard Medical School, and Brigham and Womens Hospital, researchers found that while performing a familiar lung nodule detection task, 83% of radiologists missed a grey-scale gorilla superimposed on slices of a CT scan [3]. Calling this phenomenon inattentive blindness, the study emphasizes the need for automated detection and clearly displays humans limitations to perception and attention.

II. RELATED WORKS

Herein we summarize the very recent efforts in automated lung cancer detection and diagnosis.
(1) Bag of frequencies is a method of nodule classification that radially samples intensities in three dimensions and then converts the intensity profiles into the frequency domain using a fast Fourier transform. The spectral signatures returned are used as features to create a codebook similar to bag of words [4], from which the authors derived the name. The authors cut two dimensional slices at various angles by taking samples on planes parallel to the faces of an icosahedron circumscribed in the spherical voxel of interest. While their accuracy of classifying nodules is not reported, their algorithm's capability of discriminating blood vessels from spiculated nodules is 87% [5].

(2) Taxonomic indices feature vectors characterize nodules based on texture and patterns, analyzing intensity distribution and their relationships by utilizing phylogenetic trees from ecology [6]. Regions of interest are defined as rings, spheres, and the whole nodule. For each region of interest, the intensities are cast into a tree in ascending order. The resulting tree has leaves that contain intensity values as well as the abundance of that intensity value. Figure 2 shows a visualization of our version of these trees. The taxonomic indices returned as the feature vector are calculated using the leaf properties. The authors reported a mean accuracy for nodule classification of 99.22% using 7 regions of interest and classifying their features with an SVM.

III. METHODS

In this section we discuss our proposed method for classifying lung nodules using variations of [5] and [6]. Each hand crafted feature is three dimensional and customizable through a set of parameters. The functions take as input, the whole CT scan for a patient as an itk image, the nodule radius, and the coordinates of the center of the nodule. Deep learned features were run using GoogLeNet: Inception-v3. Nodules were classified with an SVM using the outputted feature vectors individually, and then concatenated together.

A. Dataset

The LIDC/IDRI, publicly available dataset from the LUng Nodule Analysis (LUNA) 2016 Grand challenge was used for our experiments. The data was annotated for positive findings, regions classified as nodules and negative findings, regions classified as candidates. Areas were considered nodules if accepted by at least 3 out of 4 annotating radiologists and the nodule diameter exceeded 3 mm. Findings were considered candidates, but not nodules, if they were detected by 3 existing lung nodule detection algorithms but were rejected by the annotating radiologists [7].

W. Richey is with Tulane University. N. Khosravan and U. Bagci are with CRCV, University of Central Florida.

B. Bag of Frequencies

The concept for our implementation of bag of frequencies was modeled off of the previously discussed work in [5]. The main differences lie in our sampling of two dimensional surfaces and the overall structure of the parameters, both of which are explained in Figure 1.

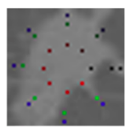
Parameter	Description	Example	
Number of slices	#of 2D surfaces extracted from nodule	10 → [4 slices in XY plane, 3 slices in YZ plane, 3 slices in ZX plane]	`For 10 surfaces, sampling: [45, 90, 135, 180] x angles [60, 120, 180] y angles [60, 120, 180] z angles`
Size of slices	Important if sampling outside of radius	1.2 → slices will be 120% the size of the nodule	Slice below is 40x40 Radius is 17
Sample Radii	Determines #and location of intensity profiles in relation to the radius	[.5, .9, 1.05] → three intensity profiles will be taken at 50%, 90% and 105% of the radius	Each radius is shown in a separate color (RBG) in the image below
Number of intensity Samples	# of points sampled; determines angles b/t points of intensity profile	8 → points will be taken every 45 degrees, counter clockwise (as shown; first sample is black, last is bright)	

Fig. 1. Parameters for our implementation of Bag of Frequencies.

C. Taxonomic Indices

Our version of taxonomic indices was modeled closely off the work in [6]. To the best of our knowledge our algorithm for casting the tree and our indexing system are unique and novel. The authors do not outline their method for inserting intensities into the tree or for indexing the intensity values during calculations, but it is mentioned in their work that all intensities must be converted to positive values. This step is superfluous with our algorithm. A visualization of a tree created by our algorithm is shown in Figure 2.

D. Radius Estimation

We implemented an iterative algorithm to estimate a nodule's radius based on the assumption that the nodule's edge lies between bright intensities inside the nodule, and darker intensities outside the nodule. With the given input of an image patch, the function begins with *start* = image patch center, and *end* = image patch edge. The radius for one circle, halfway between the starting point and the ending point, is sampled using the sampling mechanism from our bag of frequencies methodology. An intensity value threshold is applied to each sampled intensity, then a percentage threshold is applied to the radius to do a binary classification of the sampled radius as either inside or outside the nodule. If the sampled radius is classified as within the nodule, as shown in Figure 3.A, the *start* will update to be the last sampled radius. The next iteration is shown in 3.B. If the sampled radius is classified as outside the nodule radius, the END will update to become the sampled radius. The next sample is taken halfway between the new *start* and *end* points for 5 iterations, after which the maximum precision (within one pixel) has been obtained if the image patch is 40x40 pixels. Figure 4 shows the final estimations for three different nodules.

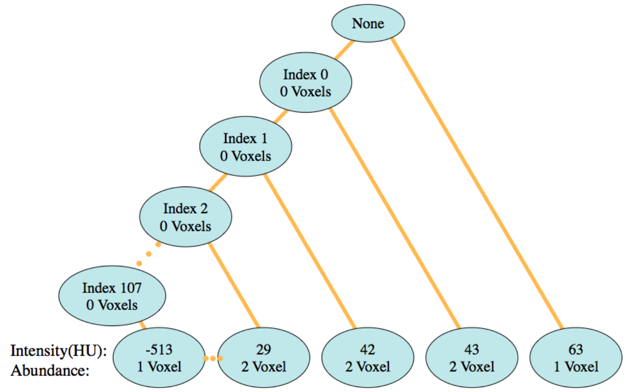


Fig. 2. An example of a tree of intensities for one region of interest; left nodes contain the height for simplistic indexing and easier calculations; right nodes contain the intensity value (in Hounsfield Units). Each node has a property, the abundance, or the number of voxels with that intensity.

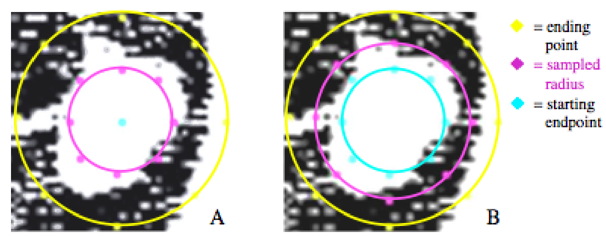


Fig. 3. The first (A) and second (B) iterations of our intensity sampling for nodule radius estimation. The sampled radius is the intermediate radius, shown in color in magenta.

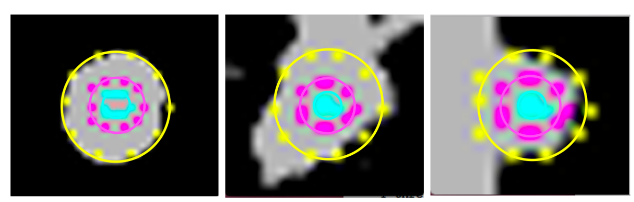


Fig. 4. Output of our radius estimation algorithm, with estimated radius as the outermost circle, shown in yellow.

E. Deep Learning Features

Incorporating hand crafted features reduces the dependency on the deep learned features, and so we opted for a pretrained network: GoogLeNet, Inception-v3. Inception-v3 is trained on the ImageNet dataset and is currently the state-of-the-art CNN architecture for the ILSVRC challenge. Inception-v3 concatenates filters of different sizes and dimensions into a single new filter. The network contains two convolution layers, two pooling layers, and nine Inception layers. Each inception layer of GoogLeNet consists of six convolutional layers with different kernel sizes and one pooling layer.

IV. Results

We ran our features on approximately 950 nodules, and 950 candidates, divided into 80% training data, and 20% testing data. We evaluated the positive accuracy -the number of true positives - and negative accuracy - the number of true negatives - as shown in Figure 5. Each hand crafted feature was tested with the real, annotated radius and the radius returned from the estimation algorithm. This dual testing allows interpretation of features' true descriptiveness, as well as the radius estimation's effect on the accuracy and features' descriptiveness.

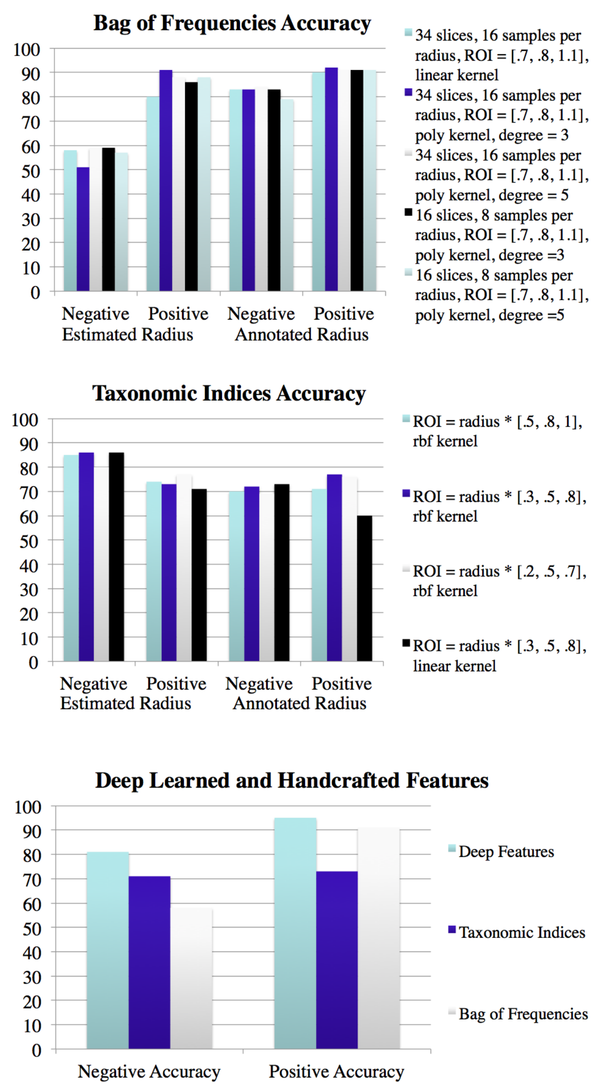


Fig. 5. Accuracy of our features.

V. Discussion

Given the few samples taken and the brevity of our refinement process, our results show promise. The accuracy reported from the original publications of these works shows that there is a room for improvement on both of our handcrafted features. Combined results were obtained by concatenating the feature vectors, but the extreme disparity in size led to lower accuracy than anticipated. It is also important to notice that each feature preformed best under different SVM conditions. The functions for bag of frequencies and taxonomic indices are easily customizable with parameters and can easily be adjusted to include more samples for more descriptive feature vectors.

VI. Future Work

Further tuning of our hand crafted features' parameters to maximize individual results will increase overall accuracy. Next steps for this research include pairing finely tuned parameters with an improved combination method.

a) Voting: Combining these features using a majority voting system, giving equal votes to each classifier should improve accuracy. An SVM would be run individually on each set of features and patches would be considered positive findings if classified as nodules by two out of the three features. Because each feature would be classified with a separate SVM, the SVMs could be finely tuned to that specific set of features, making it easier for the SVM to discriminate between classes in the feature space.

b) Pipeline: Implementing the combination of these features in the form of a pipeline would likely increase accuracy and improve time efficiency. If the SVMs were weighted to favor nodule candidates, one could see better results with applying a mask, classifying with deep learned features, and then with features from bag of frequencies, followed finally by a classification using taxonomic indices.

c) Radius Estimation: Improving the accuracy of the radius estimation will improve the hand crafted features' accuracy and the process of improving the estimation could help tune the feature parameters.Accounting for a few special cases that are commonly miscalculated in our algorithm could reduce our radius estimation error, and our overall error accordingly.

While the taxonomic indices proved more valuable with radii estimated with our algorithm, bag of frequencies was more accurate with the true, annotated radius. Because our features focus on different characteristics of the nodules, it would follow that different features may find different radii most descriptive. However, this difference should be accounted for by adjusting parameters. The information from these discrepancies may help tune parameters according to feature vectors' classification results achieved with various estimations. For example, our generally overestimated radii achieved better results for taxonomic inidices. This would indicate that we should increase the sampled radii when setting the function's parameters.

d) Custom Neural Network: A customized neural network could most accurately complement the hand crafted features

References

[1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1):5–29, 2015.

[2] Feng Li, Shusuke Sone, Hiroyuki Abe, Heber MacMahon, Samuel G Armato, and Kunio Doi. Lung cancers missed at low-dose helical ct screening in a general population: Comparison of clinical, histopathologic, and imaging findings 1. *Radiology*, 225(3):673–683, 2002.

[3] Trafton Drew, Melissa L-H Võ, and Jeremy M Wolfe. The invisible gorilla strikes again sustained inattentional blindness in expert observers. *Psychological science*, 24(9):1848–1853, 2013.

[4] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 524–531. IEEE, 2005.

[5] Francesco Ciompi, Colin Jacobs, Ernst Th Scholten, Mathilde MW Wille, Pim A de Jong, Mathias Prokop, and Bram van Ginneken. Bag-of-frequencies: A descriptor of pulmonary nodules in computed tomography images. *IEEE transactions on medical imaging*, 34(4):962–973, 2015.

[6] Antonio Oseas de Carvalho Filho, Aristófanes Corrêa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassú Nunes, and Marcelo Gattass. Lung-nodule classification based on computed tomography using taxonomic diversity indexes and an svm. *Journal of Signal Processing Systems*, pages 1–18, 2016.

[7] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.