

# Lung nodule classification: computed tomography (ct) scans

[Technology](#), [Computer](#)



## ABSTRACT

This paper proposes a novel framework for the classification of lung nodules using computed tomography (CT) scans. The proposed framework is based on the integration between (i) the geometric shape features, in terms of construction error of modeled spherical harmonics (SHs); (ii) the appearance feature in terms of Gibbs energy modeled using the 7th order MGRF; and (iii) the size feature using the  $k$ -nearest neighbor ( $k$ -NN) classifier. The final classification is obtained by using the deep autoencoder neural networks. The geometric feature is extracted by calculating the construction error between the original nodule mesh and the SHs-based constructed ones. To calculate this error curve at each point, the surface mesh for each nodule is modeled using different SHs, from 1 to 70, and calculate the difference as the error. Secondly, the appearance feature is modeled using the novel 7th-order Markov-Gibbs random field (MGRF) model in addition to the size feature using  $k$ -NN classifier. Finally, a deep autoencoder

(AE) classifier is applied to distinguish between the malignant and benign nodules. To evaluate the proposed framework, we used the publicly available data from the Lung Image Database Consortium (LIDC). We used a total of 116 nodules that were collected from 60 patients. By achieving a classification accuracy of 96.00%, the proposed system demonstrates promise to be a valuable tool for the detection of lung cancer.

*Index Terms* -CT, 7th-order MGRF, Spherical Harmonics,

Lung nodules

## 1. INTRODUCTION

Lung cancer is considered the leading cause of cancer death among both genders in the United States with about 1 out of 4 cancer deaths resulting from lung cancer [1]. Although there are several imaging modalities used for the diagnosis of lung cancer, e. g., magnetic resonance imaging (MRI), chest radiograph (X-ray), and many other modalities, computed tomography (CT) imaging is the most common and appropriate modality for examining

the lung tissues due to its high resolution and clear contrast compared to other techniques [2]. Recently, the number of lung cancer cases have increased exponentially, and its early detection can increase the chance of survival [3]. Furthermore, an automated assistive tool for the radiologists is of great importance to help in the analysis of the large amount of data available from CT scans. Thus, the computer aided diagnosis systems (CADx) is of great interest and high importance. Recently, a plethora of methods for automated diagnosis of pulmonary nodules in CT scans have been introduced. Various researchers have used image processing and data mining techniques to diagnose the pulmonary nodules. Namely, Macedo et al. [4] have proposed the use of different classifiers, such as the support vector machine (SVM), and rule-based system, to distinguish between malignant and benign lung nodules. They used texture, shape, and appearance features that were extracted from the histogram of oriented gradient (HOG) from the region of interest (ROI). Kumar et al. [5]

used deep features extracted from multi-layer autoencoders for the classification of lung nodules. Although they have proved the effectiveness of extracting high-level features from the input data in their experiments, they disregarded the morphological information, e. g., perimeter, skewness, and circularity of the nodule, which couldn't be extracted by the conventional deep models.

Jia et al. [6] have proposed a rule-based classification system based on growth rate changes and registration technique. Lee et al. [7] have proposed a lung nodule classification system using a random forest classifier aided by clustering. First they merged all the data, then they divided it into two clusters, then divided each cluster into two groups, nodule and non-nodule, based on the training set labels. Finally a random forest classifier was trained for each cluster to distinguish between benign and malignant nodules. Farahani et al. [8] have proposed an ensemble-based system to classify each pulmonary nodule by integrating multiple classifiers like SVM, K-nearest-neighbors, and neural networks.

The classifiers have learned over five morphological features and the output of these classifiers is combined using majority voting.

Huang et al. [9] have proposed a system to differentiate malignant from benign pulmonary nodules based on fractal texture features from Fractional Brownian Motion (FBM) model using (SVM).

Elsayed et al. [10] have proposed a system that uses different classifiers, e. g., Linear, Quadratic, Parzen, Neural Networks, and their different combinations such as mean, median, maximum, minimum, and voting, to enhance the performance of the classification of malignant and benign pulmonary nodules. Kim et

al. [11] have proposed a system which uses a deep neural network to extract abstract information inherent in raw hand-crafted imaging features. Then, the learned representation is used with the raw imaging features to train the classifier. Narayanan et al. [12] also used deep neural network to classify the pulmonary nodules after training on morphological features. Bayanati et al. [13] tried to identify which features on CT images could differentiate between

malignant and benign nodules. They used both texture and shape analysis features and found an enhancement in accuracy but without significant change in the false positive.

The existing methods for the classification of lung nodules have