Lung Cancer 115 (2018) 34–41

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Review

Radiomics and radiogenomics in lung cancer: A review for the clinician

Rajat Thawani[a,b,*,1], Michael McLane[b,1], Niha Beig[b], Soumya Ghose[b], Prateek Prasanna[b], Vamsidhar Velcheti[c], Anant Madabhushi[b]

[a] Maimonides Medical Center, 4802 Tenth Avenue, Brooklyn, NY 11219, United States
[b] Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, United States
[c] Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, United States

ARTICLE INFO

Keywords:
Radiomics
Radiogenomics
Image analysis
Lung cancer

ABSTRACT

Lung cancer is responsible for a large proportion of cancer-related deaths across the globe, with delayed detection being perhaps the most significant factor for its high mortality rate. Though the National Lung Screening Trial argues for screening of certain at-risk populations, the practical implementation of these screening efforts has not yet been successful and remains in high demand. Radiomics refers to the computerized extraction of data from radiologic images, and provides unique potential for making lung cancer screening more rapid and accurate using machine learning algorithms. The quantitative features analyzed express subvisual characteristics of images which correlate with pathogenesis of diseases. These features are broadly classified into four categories: intensity, structure, texture/gradient, and wavelet, based on the types of image attributes they capture. Many studies have been done to show correlation between these features and the malignant potential of a nodule on a chest CT. In cancer patients, these nodules also have features that can be correlated with prognosis and mutation status. The major limitations of radiomics are the lack of standardization of acquisition parameters, inconsistent radiomic methods, and lack of reproducibility. Researchers are working on overcoming these limitations, which would make radiomics more acceptable in the medical community.

1. Introduction

Lung cancer is a leading cause of cancer-related deaths worldwide. In 2015, more than 3 million cases of lung cancer and 1.7 million lung cancer-related deaths were documented across the globe [1]. With 70% of lung cancer diagnoses being made after the onset of symptoms from advanced local or metastatic disease, the five-year survival rate following diagnosis is roughly 16% [2,3]. The survival rate is above 50%, but only when the disease is diagnosed while it is still localized [4]. Unfortunately, only 15% of lung cancers are diagnosed at early stages and an accurate and affordable screening method remains in significant demand.

The National Lung Screening Trial (NLST) concluded with a recommendation of using low-dose computed tomography (LDCT) for screening at-risk populations for lung cancer, and demonstrated a 20% reduction in lung cancer mortality in these patients [5]. Despite the survival benefits offered by lung cancer screening, this protocol creates an economic burden due to poor specificity. According to data from the NLST, 18% of those with identified lung nodules on LDCT were overdiagnosed as having lung cancer when their lesions were in fact benign [6]. This leads to a significant number of subsequent imaging studies and surgical excisions that are costly and likely unnecessary.

Radiomics refers to the extraction of sub-visual, yet quantitative, image features with the intent of creating mineable databases from radiological images [7]. In oncology, features of tumors identified from radiological data (e.g. CT and MRI scans) can be used to reveal diagnostic, predictive, and prognostic associations in cancer patients via correlations to objective response criteria like survival, or response to treatment [8,9]. Image features that are often targeted for extraction and analysis in this setting include nodule volume, nodule shape, intensity patterns, and a range of “texture” features [9,10]. The image analysis is completely computerized and these features are extracted automatically, and with high throughput [10]. There has been substantial interest in the use of radiomics in the context of lung cancer screenings with the goal of maximizing sensitivity and specificity, while minimizing time burden imposed upon radiologists. Apart from diagnosis, radiomics is also being used to predict prognosis and response to certain therapies in the field of precision medicine. Some features have even been shown to identify genomic alterations within tumor DNA, a field that is now called ‘radiogenomics’. These features can identify the

* Corresponding author at: Maimonides Medical Center, 4802 Tenth Avenue, Brooklyn, NY 11219, United States.
E-mail address: rajat.thawani@case.edu (R. Thawani).
[1] Equal contribution.

https://doi.org/10.1016/j.lungcan.2017.10.015
Received 1 June 2017; Received in revised form 14 October 2017; Accepted 29 October 2017

presence of specific mutations and alterations in biological pathways, which in turn affects the management and outcome of patients.

In this review, we describe radiomics from a clinical perspective. Specifically, we review work that speaks to the utility of radiomics and radiogenomics for diagnosing and predicting prognosis of individuals with non-small cell lung cancer (NSCLC), and we discuss the potential and the importance of integrating these methods into everyday clinical management of these patients. Although some groups have been applying radiomics to PET/CT and MRI data in lung cancer, we limit this review to radiomics in CT scans, since CT is currently the primary means for screening and monitoring lung cancer in clinic. Lung cancer screening scans typically utilize low dose CT images, while diagnostic scans are more often high quality and with contrast enhancement.

Past reviews on the topic of radiomics in lung cancer have focused on comparing the accuracy of different methodologies and specific algorithms, but have not adequately discussed the potential utility and practical limitations of this work with regard to clinical practice. We aim to present the current applications of radiomics to lung cancer research, and the potential implications for its integration into clinical oncology, in a way that medical professionals with little familiarity with computer science can understand.

2. Overview of radiomic and radiogenomic feature analysis

The process of radiomics or radiogenomics first involves conversion of radiographic images into mineable data and occurs through four steps: (a) image acquisition and reconstruction, (b) region of interest segmentation, (c) feature extraction and quantification, (d) building predictive and prognostic models (Fig. 1).

A major advantage of radiomics is that it, ideally, does not significantly disrupt clinical workflow. The images acquired for diagnostic purposes by a physician undergo reconstruction into 2D or 3D images by radiology technicians. After this, the region of interest, i.e. the nodule, is segmented, and it is from this region that quantitative features are extracted and analyzed. Segmentation can be done manually by a radiologist, but although this is a good way to ensure accuracy, the process can be quite cumbersome. Alternatively, segmentation can be automated using abrupt changes in gray level (corresponding to edges) and similarity in gray level to measure homogeneity, though this may not reach the accuracy of manual segmentation. Segmentation is more critical in measuring the shape of the nodule, in contrast to something like 2D texture. It is also critical for being able to identify textural patterns within the nodule, as opposed to the peri-nodular space and the lung parenchyma. From the region of interest, quantitative features can then be extracted automatically and with high throughput in order to build machine learning-based models.

The quantitative features utilized in radiomic analysis refer to algorithms that can be used to describe local regions within a radiologic image. Several algorithms exist for this purpose, though common radiomic features are currently divided into the following classes: intensity-based, structural, texture/gradient-based, and wavelet. Semantic features are in contrast to radiomic features in that they are not subvisual, but rather are observed and described directly by radiologists. These features are included in this review for completeness, as they may also be used to build classifiers. The application of these features depend on preferences of the researchers and area of application [11]. For example, structural features characterize the shape and size of lung nodules, with more spiculated and irregular tumors correlating with malignant nature. Intensity, texture, and gradient-based features, on the other hand, characterize tumor heterogeneity, which is known to correlate with poor prognosis. The specifics of these feature classes are discussed in greater detail in a subsequent section.

Once all of the features have been extracted, a wide range of statistical models are often used to choose a subset of top features that correlate with the outcome specified by the hypothesis during feature selection [12]. Selection of top ranking features reduces dimensionality of the problem and improves prediction accuracies. The process of feature selection can occur either via univariate or multivariate statistical models. Univariate methods only depend on feature association, ignoring redundancy, whereas multivariate methods investigate interactions within different features and select them after weighing both association and redundancy. Often, Fisher score, Chi-squared test, and Wilcoxon, amongst other statistical models are used for feature selection. For example, in a chi-squared test based feature selection, chi-squared statistics between every feature variable and the target variable (the label/outcome) is computed. If the target variable is independent of the feature variable, the feature variable is discarded. Fischer score gives a higher rank to features with higher variance (implying more

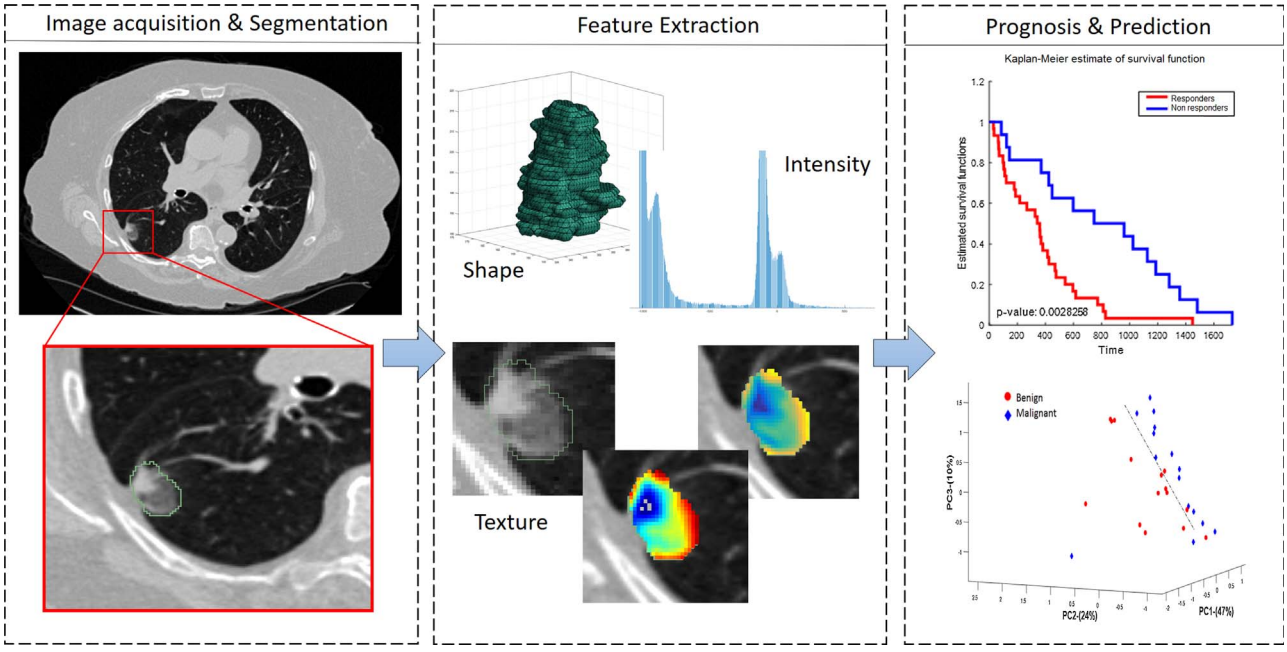

Fig. 1. Overview of a typical Radiomic workflow. CT image acquisition is followed by automatic or manual expert segmentation. Radiomic features (such as texture, shape) are extracted in a quantitative manner to build machine learning based models. The output of the machine learning models can be predictive or prognostic in nature, depending on the clinical question.