Effect of manual correction of an automated tumour segmentation algorithm on the prediction of lung adenocarcinoma prognosis

Khushbu Patel1, Dianna Rowe1, Maria-Fernanda Senosain2, Sarah Atmajanoa1, Sanja Antić1, S.M. Jamshedur Rahman1, Fabien Maldonado1, Heidi Chen3, Pierre P. Massion1,4

1Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN; 2Department of Biostatistics, Vanderbilt University; 3Veterans Affairs Tennessee Valley Healthcare System, Nashville, TN

Background

- Adenocarcinoma is the most common lung cancer among its histological subtypes, and it has wide spectrum of biological, radiological & histological behavior ranging from indolent to aggressive, resulting in overdiagnosis & higher false positive detected cases, which expose patients to unnecessary morbidity and mortality.
- Overdiagnosis is a major challenge in lung cancer screening and can account for up to 18% of all cancer detected in the screening context.
- Thus, distinguishing indolent from malignant nodule is major clinical challenge.
- Computer-Aided Nodule Assessment and Risk Yield (CANARY) offers the computed tomography (CT)-based quantitative imaging tool/scoring system for the non-invasive risk stratification of lung adenocarcinomas with simultaneous prediction of patient survival.
- It provides the Score Indicative of Lesion/Lung Aggression/Abnormality (SILA). Major advantage of SILA is that it provides a continuous variable to predict prognosis.
- Currently, the SILA score is partially automated but the segmentation of the lesions brings a source of variability.

Methods

- Extraction of clinical characteristics (age, gender, smoking history etc.) and sample collection date (bronchial brushing, blood draw, Microbiology – Sputum and Fungal serology, trans-thoracic needle biopsy/Wedge resection or Lobectomy) from the electronic medical record
- Selection of the appropriate CT scan – indeterminate pulmonary nodules 5-30mm, Slice thickness ≤ 3mm, Non-contrast or contrast CT scans, ± 90 days within the blood draw date
- Processing of images (Deidentification, transfer into software)
- Semi-automated segmentation of pulmonary nodule, verified by human

Results

- Nine radiologically distinct clusters (central candidates - "exemplars") were identified with use of an unsupervised clustering algorithm.
- The exemplars were color-coded as violet, indigo, red, orange, blue, cyan, green, yellow, and pink which represent the radiological building blocks of nodules in the AC spectrum.
- Good (G) prognosis: blue-green-cyan, Intermediate (I) prognosis - mixed, Poor (P) prognosis - violet-indigo-red-orange
- SILA score (Score Indicative of Lung Cancer Aggression) is a cumulative aggregate of normalized distributions of ordered CANARY exemplars.
- Major advantage of SILA is that it provides a continuous variable to predict prognosis.

Conclusions

- Manual segmentation based on nodule texture analysis-related prediction reclassifies 14 lung adenocarcinoma and there is a discrepancy between automated & manual segmentation.
- There is a moderate to strong correlation between automated & manual segmentation.
- Whether the manual segmentation improves the prediction accuracy over the automated segmentation has not been tested because of insufficient follow up.

Future Directions

- To validate the effect of manual segmentation on prediction of adenocarcinoma behavior and obtain the imaging biomarkers to improve the machine learning methods
- To develop a better prediction model: Integrate clinical and CT imaging features with cellular and molecular determinants of tumor behavior and determine whether machine learning algorithms can predict behavior of early stage ADCs.

Acknowledgement

Massion Lab
Inspired by Patients, Driven by Science