ARTERYS

As Accurate as the Radiologist

How Arterys Lung AI provides robust nodule volumes

Introduction

Accurate quantification describing the progression of lung nodules over time is imperative in the diagnosis and treatment of lung cancer. Metrics such as nodule long axis or volume doubling time have been shown to have a clinical prognostic value.[1]

Accurate volumetric masks of nodules can be tedious and time consuming to create. Automatic or semi-automatic methods are therefore desirable. The aim of the clinical assessment was to evaluate the volumetric error of the 3D lung nodule segmentation model available in Arterys Lung AI[2] by comparing it to radiologist-created annotations.

Materials and Methods

69 scans containing 126 unique nodules marked by at least 3 out of 4 radiologists were randomly selected from the Lung Image Database Consortium (LIDC)[3] to evaluate Arterys Lung AI nodule segmentation. The nodule segmentation algorithm requires a single seed point to run; this can be part of a semi-automatic workflow (e.g. a radiologist click) or automatic workflow (e.g. the segmentation is triggered at the point localized by the detection algorithm). The ground truth volume of a nodule was determined by taking the average of the volumes derived from the 3D manual segmentations of each annotator. Using the ground truth volume, both the individual radiologist and the AI model accuracy could be evaluated.

To measure the error between the model and the ground truth a relative absolute volume error (MRAVE) was evaluated (Equation 1).

For radiologist annotations, a comparable error metric called inter-rater RAVE (IRAVE) is determined by dividing the standard deviation of radiologist volumes by the ground truth nodule volumes (Equation 2).

$$\textbf{MRAVE} = \frac{\text{Segmentation volume - Ground truth volume}}{\text{Ground truth volume}} \times 100$$

Equation 1.0
Model relative absolute volume error (MRAVE)

$$\textbf{IRAVE} = \frac{\text{Standard deviation of radiologist volumes}}{\text{Ground truth volume}} \times 100$$

Equation 2.0
Inter-Rater relative absolute volume error (IRAVE)

We evaluate the MRAVE and IRAVE for each nodule using the center of mass of annotator segmentations as the seed point for the algorithm.

Furthermore, because a nodule segmentation can be triggered semi-automatically using a single click inside the nodule or automatically by using a point provided by a detection model, we evaluate the robustness of the segmentation algorithm to the location of the seed point to test robustness a random jitter up to the size of the nodule was applied to the seed point before running inference. This was repeated 12 times for each nodule. The effect of the jitter can then be measured by comparing the variance in the predicted volumes against the ground truth volumes.

Results

For nodules ≥ 6mm, the median IRAVE for radiologists was 21.0% compared to a MRAVE of 11.5%. For nodules ≤ 6mm, the median IRAVE was 24.5% compared to a MRAVE of 16.4%. In all cases the AI produced a smaller median error when compared to the variability between radiologists (Figure 1.0). Example nodule segmentations produced by the AI model and radiologist are shown in Figure 2.0.

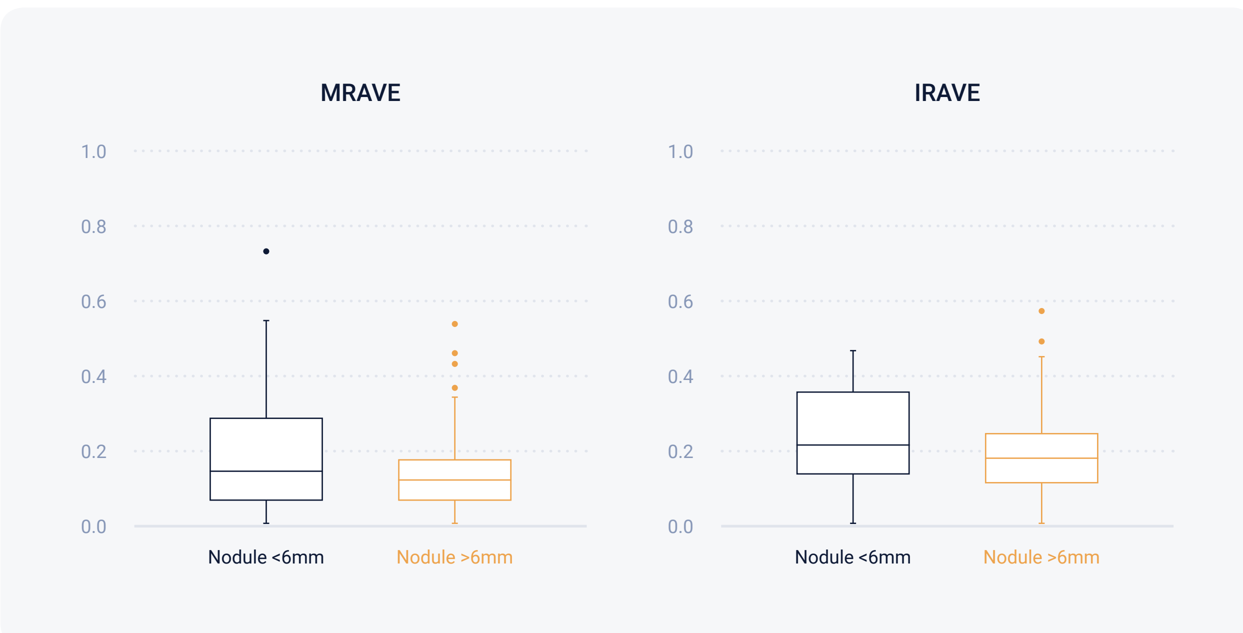

Figure 1.0
Boxplots of the relative volume error for model segmentations, MRAVE, (left) and inter-rater variability, IRAVE (right) for nodules greater or smaller than 6mm in diameter. Note note that the distribution of the error is tighter and smaller for the MRAVE relative to the IRAVE

For seed point jitter the standard deviation of the predicted volumes were calculated and compared relative to the ground truth volume. For a click within ½ the size of the nodule, the median difference was 3.6%; for a click anywhere within the nodule the median value was 5.4%. This implies that an offset seed point introduced negligible error.

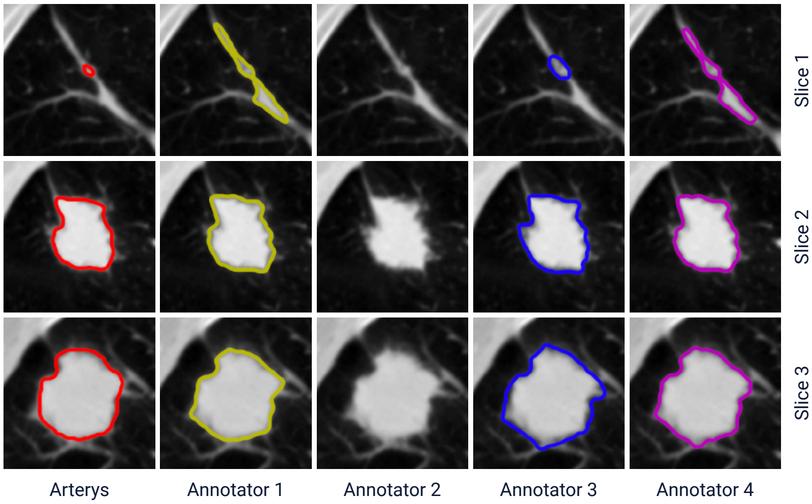

Figure 2.0
A comparison of the segmentation provided by Arterys LungAI (red contours, left) and the ground truth provided by four radiologists for a large nodule abutting a vessel. Three axial slices of 40mm x 40mm are shown. The segmentation provided by Arterys LungAI does not “spill over” into the vessel. The segmentation truth provided by annotators 1 and 4 includes the vessel, perhaps as the result of an overzealous or inattentive use of a 3D flood fill tool. No annotation was provided for this nodule by annotator 2.

Conclusion

Arterys LungAI provides reproducible segmentations and volumes. The volumes are as accurate as radiologists and are robust to the segmentation seed point. Using Arterys Lung AI can provide radiologists with robust and reproducible nodule metrics for the diagnosis and tracking of lung cancer.

1. R. Larici, A. Farchione, P. Franchi, M. Ciliberto, G. Cicchetti, L. Calandriello, A. del Ciello, L. Bonomo (2017) “Lung nodules: size still matters“ European Respiratory Review, 26 (146) 170025.
2. Arterys Lung AI v18.09.
3. Armato et al (2011) “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans.“ Med Phys. Feb;38(2):915-31.

ARTERYS

info@arterys.com
51 Federal St. Suite 305, San Francisco, CA 94107
www.arterys.com