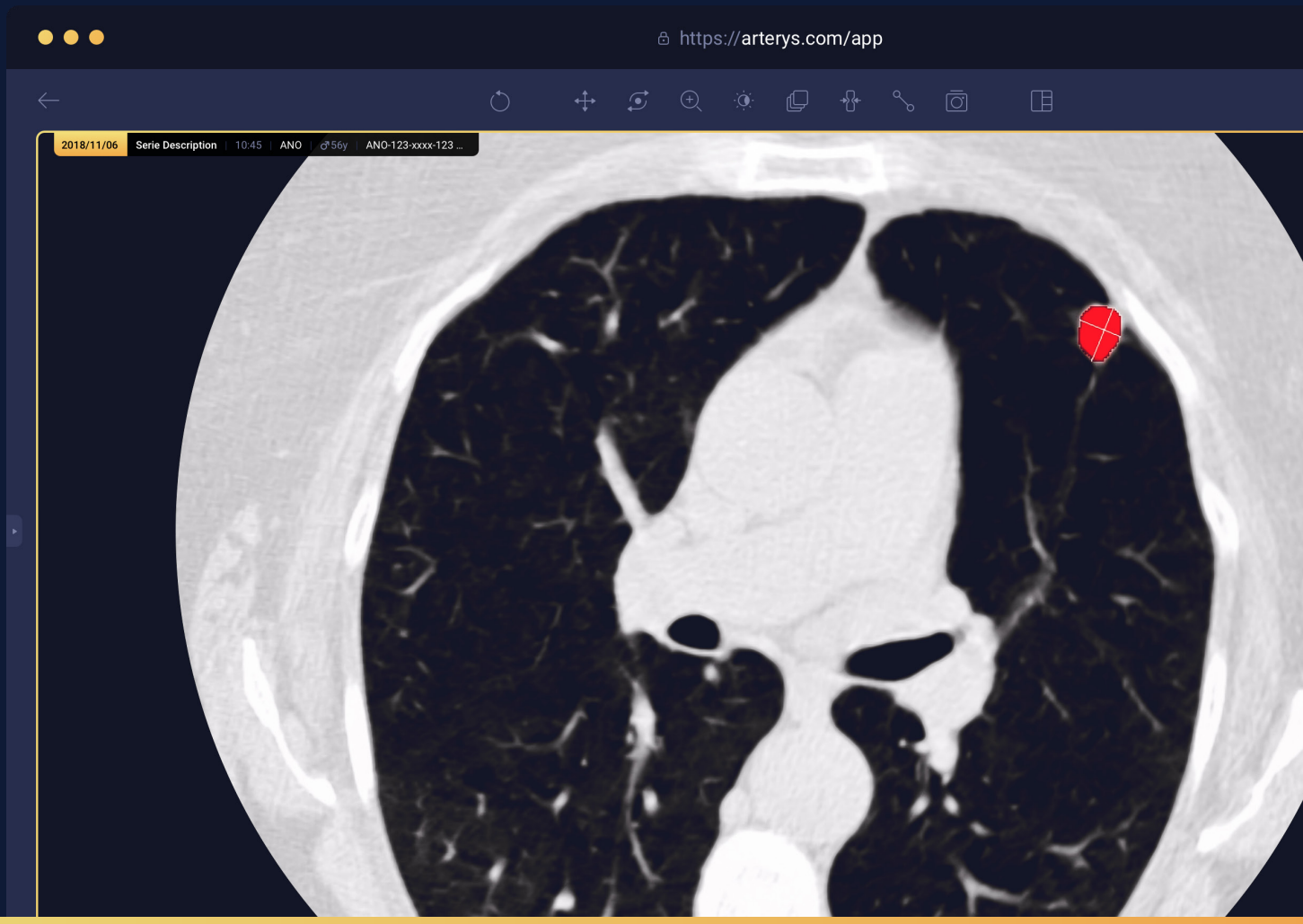

ARTERYS

Lung Nodule Detection Information

Summary

Arterys Lung AI is a deep learning-based CADe algorithm trained and validated on CT scans from thousands of patients and approximately ten thousand nodules. It is designed to assist users in the detection of lung nodules (parenchymal abnormalities irrespective of presumed histology that may represent primary lung cancers, metastatic disease, or non-cancerous processes) between 4mm and 30mm. Arterys Lung AI was evaluated in a retrospective clinical assessment across primarily low-dose non-contrast exams with a minority share of standard-dose exams with contrast from GE, Siemens, Toshiba, or Phillips scanners.

240 Chest CT patients were randomly selected from the National Lung Screening Trial (NLST) and University of California, San Diego (UCSD) for analysis in the retrospective assessment. 204 (NLST / 85%) patients were included to represent lung cancer screening and 36 (UCSD / 15%) patients were included to represent patients that are not specifically being screened for cancer but where the clinical practice is to report any incidental pulmonary nodules.

The study found a clinically meaningful and statistically significant increase in sensitivity of 0.14 (95% confidence interval: 0.11 to 0.17; $p<0.001$) with no statistically significant change to the reading time (+2 seconds with CADe, 95% CI: -4 to 8, $p=0.24$), suggesting that Lung AI increases reader performance without significantly impacting clinical usability.

An average increase of 0.71 false positives per scan with CADe was measured, with the majority (0.54 FP/scan) of this increase concentrated in the 4-6mm range. We identify that inherent variation across readers in determination of what constitutes a nodule, as well the difficulty in exhaustive identification of nodules (particularly those 4-6mm) on part of our ground truth radiologists, inflates the perceived increase in false positive rate. Correct observations not annotated in the ground truth but located by CADe and confirmed by the readers during their aided reads further indicate CADe's upside. Non-exhaustive examination suggests that more than 50% of measured false positives may be attributed to gaps in the ground truth specifically in the 4-6mm range.

Modifying the detection results of Lung AI is fast and easy; nodules can be added or removed with a single click. This paradigm gives the user the power to use what they need for a particular patient and filter out what they deem to be unimportant. As opposed to other CADe software, for which a false positive mark may be directly exported to PACS when detection is performed, Arterys Lung AI has the flexibility to ensure the clinical outcome is firmly in the user's control.

Training Data

The Lung AI algorithm was trained on 1571 series with nearly ten thousand focal abnormalities between 3 and 32mm. Validation, hyperparameter tuning, and model selection was performed with 656 cases. The in-plane resolution of the training and validation data was between 0.43-0.98mm, while the slice thickness varied between 0.3 and 5.0mm. There were between 66 and 1093 images per scan. Effective tube current-time varied between 20 mAs and 464 mAs.

Training and validation cases were annotated by between two and four radiologists. No consensus filtering (e.g. only training on at least ¾ consensus nodules) is performed on the training data. This optimizes the model for high sensitivity and generalizability: given this paradigm, the model is trained on the greatest diversity in focal abnormalities that at least one radiologist would identify.

Nodule origin was not used as grounds for scan inclusion/exclusion and therefore primary lung cancer, metastatic disease, noncancerous processes, and other indeterminate sources are all represented in the dataset; in other words, the model does not distinguish between the possible nodules types.

Algorithmic Backbone

Cutting edge deep learning-based methods are utilized in sequence for the nodule detection pipeline. First, a fully convolutional neural network locates as many nodule candidates as possible. This network is tuned to maximize nodule sensitivity rather than specificity.

The data are resampled to 0.75mm isotropic resolution. The network then takes a thin axial slice and 5mm thick maximum intensity projection simultaneously as inputs and produces a voxel-wise segmentation map of predictions. Due to the fine resolution, nodules as small as 2-2.5mm (just 2-3 voxels across) are within the detection capability of the network. Predictions of sufficient probability are gathered and handed to the second piece of the pipeline: the classification network.

The classification network is designed to increase the specificity of the results. It operates on a three dimensional patch 40mm to a side centered on the image data; this ensures there is sufficient context to classify nodules in the intended range for the device (4-30mm). By not using a larger patch, we ensure that fine-grained image details are “seen” by the network while balancing the limits of memory usage; however, this also means that the classification network has inconsistent results when presented with nodules larger than 40mm, which are clearly visible by the human eye.

Evaluation Data: Patient Population

The patient population designed for the evaluation of Lung AI is representative of a primarily screening-focused institution that also reads some chest CTs from other sources for the presence or absence of lung nodules. The screening population represented in the Lung AI retrospective assessment is composed of patients between 55 and 74 years old with a history of smoking. Other chest CTs utilized in the assessment were acquired for the purpose of detecting pulmonary embolism and come from patients 18 years of age or older.

Detection Limitations

Imaging Conditions and Patients

Low-dose, diagnostic dose, contrast, and non-contrast scans across a range of nodule sizes, resolutions, scanner vendors, and reconstruction methods were used throughout training and hyperparameter tuning. However, it should not be expected that Lung AI will perform equally across the full breadth of CT scans that may be encountered. To succeed, the algorithm requires that both lungs are fully visible within the field of view, there are no gaps in the slice spacing, and there are no excessive motion artifacts. Slice thickness should be less than or equal to 5mm. Non-physical or scans

for which there may have been a processing error (e.g. those wherein slices may not be sorted, or wherein there are duplicate slices) may produce unexpected results. Furthermore, scans wherein the highest in-plane resolution is not in the axial direction may confound the algorithm. In these cases, no results or unusual false positives could be displayed; the user should always review the case and nodule detections for veracity.

Summary of Performance Assessment

Assessment Data

240 chest CTs were selected for evaluation of the effect Lung AI has on reader performance. Manufacturers represented in the dataset include GE (n=141), Siemens (n=60), Toshiba (n=27), and Phillips (n=12). The 36 standard-dose/contrast scans from UCSD includes a variety of reconstruction kernels and filtered back projection (FBP) blended with 40% adaptive statistical iterative reconstruction whereas the 204 low-dose/non-contrast scans from NLST only utilized FBP across a wide range of reconstruction kernels.

Ground truth was determined by 4 sub-speciality (thoracic imaging) radiologists with >10 years accreditation in diagnostic radiology. Ground truth nodules were determined by markings with >= ¾ consensus for the purposes of our primary endpoint. This ensures that the analysis is focused on the most important nodules. Following the approach of van Ginneken et al.[1] and Jacobs et al.[2], we filtered any predictions on nodules for which the number of ground truth radiologists was fewer than the consensus number of ground truth radiologists.

Clinical reads were conducted by 8 radiologists with >2 years accreditation in diagnostic radiology. A wide variety of reading experiences were represented. Radiologists read all 240 studies twice, once with CADe (aided) and once without CADe (unaided). The reading order was randomized and a minimum of 30 days was set between reads. 4 of the 8 radiologists read without CADe first; the other half read with CADe first. The number of seconds per read was recorded.

Performance

For nodules between 4mm to 30mm, the standalone performance of the model and clinical readers was:

	Sensitivity (95% CI)	Reading Time (95% CI)	False Positives Per Scan (95% CI)
Model - Standalone	0.93 (0.90-0.96)	N/A	1.53 (1.24 - 1.84)
Radiologist - Aided (CADe)	0.85 (0.83-0.87)	149 seconds (145-154)	1.05 (0.98-1.13)
Radiologist - Unaided	0.71 (0.68-0.73)	147 seconds (143-151)	0.34 (0.30-0.38)

The change in individual radiologist performance is displayed in Figure 1. All radiologists aided by CADe showed an increase in their reading sensitivity, with a significant increase noted for five of the eight radiologists. Across the population of readers, a statistically significant increase in sensitivity was noted ($p<0.001$). Secondary analyses performed with respect to other levels of consensus corroborate the same trend identified with ¾ consensus ground truth. Overall, there was no statistically significant change to the reading times with/without CADe. However, individual radiologists varied, saving up to 40 seconds or adding up to 35 seconds per read.

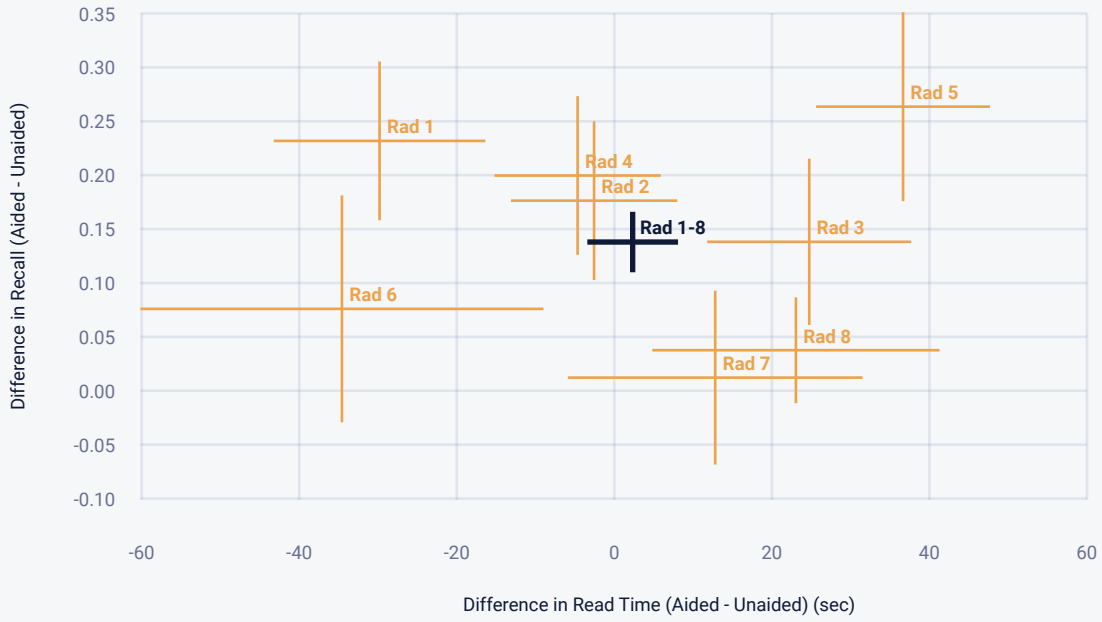

Figure 1. Individual and combined reader performances for nodules between 4-30mm in diameter in the retrospective clinical assessment. Overall, there is a significant improvement in radiologist recall of nodules, with no statistically significant change to the read times. Each element on the plot is the bootstrapped mean and 95% confidence interval for the difference in recall and read time across 100000 iterations for each reader. The "combined" reader is presented as "Rad{1-8}" in black.

1. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study. Bram van Ginneken, Samuel G. Armato, Bartjan de Hoop, Saskia van Amelsvoort-van de Vorst et al. Medical Image Analysis, Volume 14, Issue 6, 2010, Pages 707-722. https://doi.org/10.1016/j.media.2010.05.005

2. Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. Jacobs C, van Rikxoort EM, Murphy K, Prokop M, Schaefer-Prokop CM, van Ginneken B. Eur Radiol. 2016;26(7):2139–2147. https://doi.org/10.1007/s00330-015-4030-7

ARTERYS

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