ARTERYS

Lung AI Nodule Detection

Summary

A major component of Arterys Lung AI is the machine learning-based lung nodule detection algorithm. The algorithm was developed (trained and validated) on chest CT scans from thousands of patients and approximately ten thousand nodules. It is optimized and was clinically evaluated for assisting clinicians in the detection of lung nodules between 5mm and 30mm. The algorithm's clinical evaluation consisted of an 8 reader, 240 case retrospective clinical assessment across primarily low-dose non-contrast exams, representing a lung cancer screening subpopulation and a minority share of standard-dose exams with contrast, representing a subpopulation evaluated for incidental pulmonary nodules.

The assessment primary endpoint measured a performance metric derived from the nodule level sensitivity and false positives (FPs) per scan. The algorithm aided reads' performance metric had a statistically significant increase of 0.042 ($p = 0.0403$) compared to the unaided reads. Timing analysis found the median detection time per reader nodule candidate for algorithm aided reads decreased by 22.8s (95% confidence interval: 19.6s, 25.8s, $p < 0.001$) compared to the unaided reads.

Arterys Lung AI differentiates itself with its viewer-full integration into Arterys Cloud Platform which enables fast and easy, single-click modification, addition, and removal of algorithm and reader proposed nodules. This paradigm also enables clinicians to make analytical decisions and apply more advanced filtering maximizing the study's relevance and impact on the patient's outcome.

Development Materials & Methods

Methodology

Cutting edge deep learning-based methods are utilized in sequence for the nodule detection pipeline:

1. **Proposer pre-processing:** The proposer receives two inputs, data resampled to 0.75mm isotropic resolution, and 5mm maximum intensity projections.
2. **Proposer:** A fully convolutional neural network tuned to maximize nodule sensitivity rather than specificity locates nodule candidates.
3. **Classifier pre-processing:** The classifier receives a 40mm 3D patch centered on a proposed finding which is generated from the original radiological data. The 40mm patch ensures there is sufficient fine-grained details to classify nodules in the intended range for the device.
4. **Classifier:** A neural network designed to increase the specificity of the proposer's results.

Training and Validation Data

The detection algorithm was trained on 1571 series with approximately ten thousand focal abnormalities between approximately 3 and 40mm in size. Validation, hyperparameter tuning, and model selection was performed with 656 series. These data had the following characteristics: in-plane resolution: 0.43 to 0.98mm; slice thickness: 0.3 to 5.0mm; number of images per scan: 66 to 1093 images; effective tube current-time: 20 to 464 mAs.

Training and validation cases were annotated by 2 to 4 radiologists. To optimize the algorithm's high sensitivity

and generalizability, all annotations from at least 1 radiologist were included in training. Nodule morphology was not considered in training, therefore primary lung cancer, metastatic disease, noncancerous processes, and other indeterminate sources are represented in the data.

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 validation. To ensure optimal algorithm performance, follow these guidelines:

- Both lungs are fully visible within the field of view
- Homogeneous slice spacing; there are no gaps in the slice spacing
- Axial direction has the highest in-plane resolution
- Slice thickness less than or equal to 5mm
- No excessive motion artifacts
- No non-physical or scans which may have a processing error

If these guidelines are not met, the detection algorithm is more likely to produce no results or unusual false positives. The clinician should always review the case and nodule detections for veracity.

Clinical Assessment

Assessment Data

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 NLST data represented patients between 55 and 74 years old with a history of smoking and the UCSD data represented patients 18 years of age or older.

The assessment dataset included the OEM of the scanners: GE (n=141), Siemens (n=60), Toshiba (n=27), Phillips (n=12). The 36 standard-dose, contrast scans from UCSD includes a variety of reconstruction kernels and filtered back projection 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

The ground truth (GT) was determined by 4 expert radiologists reviewing the radiological data and identifying nodule candidates, and a separate panel of 2 expert radiologists classifying the detected nodule candidates. All expert radiologists had >10 years accreditation in diagnostic radiology and were thoracic imaging specialists. The assessment's GT contained 580 GT nodules which represented a robust distribution of clinically relevant pulmonary parenchymal abnormalities. The top 8 most common GT nodule classification categories are shown in Table 1.0.

Classification	Count (n, %)
Solid nodule	229 (39.48)
Ground-glass nodule	80 (13.79)
Perifissural nodule	63 (10.86)
Spiculated nodule	57 (9.83)
Pleural nodule	42 (7.24)
Mixed nodule	40 (6.90)
Irregular nodule	10 (1.72)
Thick odular-walled cyst	10 (1.72)

Table 1.0: Summary of the 8 most common GT classification categories.

Clinical Reads

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 Arterys Lung AI's nodule detection algorithm (CADe aided reads) and once without the algorithm (unaided reads). To avoid biases, the read type and reading order were randomized and a minimum of 30 days was set between read types.

Results

The assessment's primary endpoint measured the area under the curve (AUC) of the alternative free-response operating characteristic (AFROC), a performance metric derived from the nodule level sensitivity, and false positives (FPs) per scan. The CADe aided reads' AFROC AUC had an increase that was statistically significant of 0.042 (95% confidence interval: 0.0002, 0.0822, p = 0.0403) compared to the unaided reads. Table 2.0 summaries the performance for the mean CADe aided reads, mean unaided reads, and standalone CADe algorithm. Figure 1.0 shows the AFROC across the same groups.

Performance Group	AFROC AUC	Sensitivity	FPs/scan
Mean CADe Aided Reads	0.657	0.605	1.226
Mean Unaided Reads	0.615	0.452	0.618
Standalone CADe Algorithm	0.682	0.667	1.492

Table 2.0: Performance summary for the mean CADe aided reads, mean unaided reads and standalone CADe algorithm

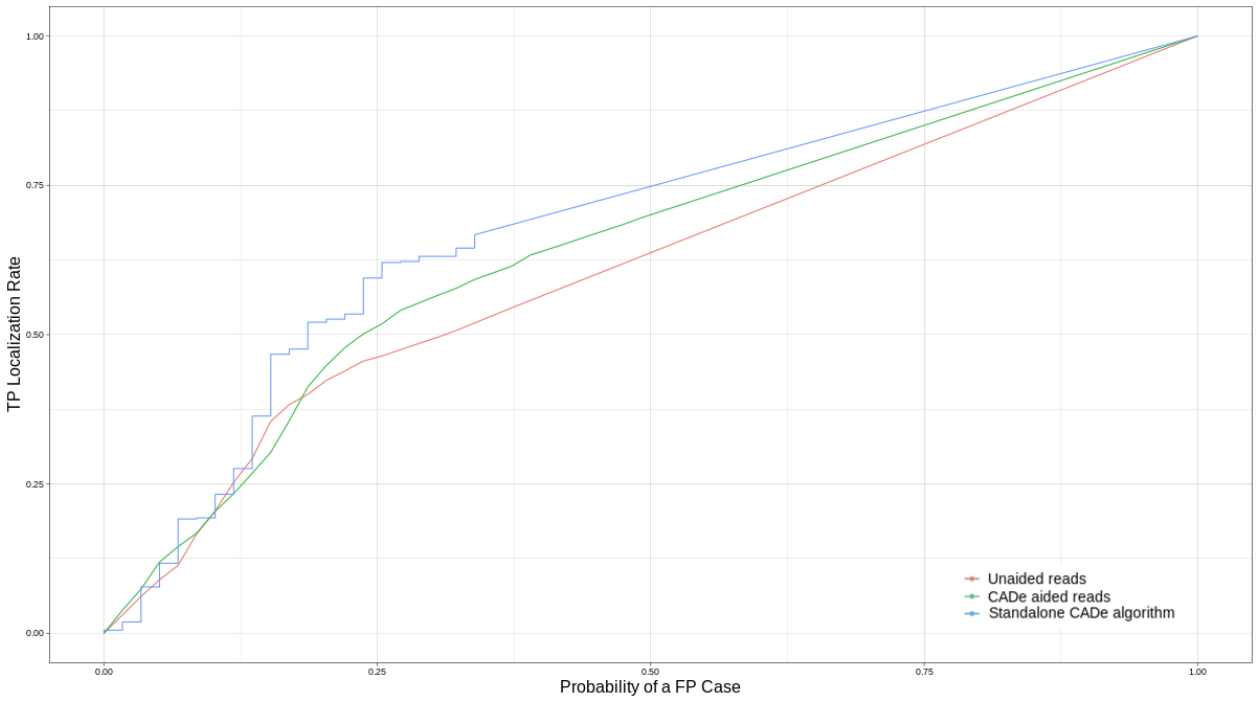

Figure 1.0: The AFROC curves for the mean CADe aided reads, mean unaided reads and standalone CADe algorithm

6 of the 8 clinical assessment readers CADe aided reads' AFROC AUCs were greater than their unaided reads. The largest increase among these 6 clinical assessment readers was 17.75%.

Timing analysis found the median detection time per reader nodule candidate for CADe aided reads decreased by 22.8s (95% confidence interval: 19.6s, 25.8s, $p < 0.001$) compared to the unaided reads.

The clinical performance assessment found consistent improvements for the AFROC AUC across the patient and OEM of the scanner subpopulations. For the nodule size subpopulation, sensitivity and FP per scan were consistent for the standalone CADe algorithm.