VLST: Virtual Lung Screening Trial for Lung Cancer Detection Using Virtual Imaging Trial

Fakrul Islam Tushar[1,2], Liesbeth Vancoillie[1], Cindy McCabe[1,3], Amareswararao Kavuri[1], Lavsen Dahal[1,2], Brian Harrawood[1], Milo Fryling[1], Mojtaba Zarei[1,2], Saman Sotoudeh-Paima[1,2], Fong Chi Ho[1,2], Dhrubajyoti Ghosh[4], Sheng Luo[4], W. Paul Segars[1,2,3], Ehsan Abadi[1,2,3], Kyle J. Lafata[1,2,3,4], Ehsan Samei[1,2,3,*], Joseph Y. Lo[1,2,3,*]

[1] Center for Virtual Imaging Trials, Carl E. Ravin Advanced Imaging Laboratories, Dept. of Radiology, Duke University School of Medicine
[2] Dept. of Electrical & Computer Engineering, Pratt School of Engineering, Duke University
[3] Medical Physics Graduate Program, Duke University
[4] Dept. of Biostatistics and Bioinformatics, Duke University School of Medicine
[5] Dept. of Radiation Oncology, Duke University School of Medicine.
[*] The last two authors contributed equally as co-senior authors.

Abstract

Importance: The efficacy of lung cancer screening can be significantly impacted by the imaging modality used. This Virtual Lung Screening Trial (VLST) addresses the critical need for precision in lung cancer diagnostics and the potential for reducing unnecessary radiation exposure in clinical settings.

Objectives: To establish a virtual imaging trial (VIT) platform that accurately simulates real-world lung screening trials (LSTs) to assess the diagnostic accuracy of CT and CXR modalities.

Design, Setting, and Participants: Utilizing computational models and machine learning algorithms, we created a diverse virtual patient population. The cohort, designed to mirror real-world demographics, was assessed using virtual imaging techniques that reflect historical imaging technologies.

Main Outcomes and Measures: The primary outcome was the difference in the Area Under the Curve (AUC) for CT and CXR modalities across lesion types and sizes.

Results: The study analyzed 298 CT and 313 CXR simulated images from 313 virtual patients, with a lesion-level AUC of 0.81 (95% CI: 0.78-0.84) for CT and 0.55 (95% CI: 0.53-0.56) for CXR. At the patient level, CT demonstrated an AUC of 0.85 (95% CI: 0.80-0.89), compared to 0.53 (95% CI: 0.47-0.60) for CXR. Subgroup analyses indicated CT's superior performance in detecting homogeneous lesions (AUC of 0.97 for lesion-level) and heterogeneous lesions (AUC of 0.71 for lesion-level) as well as in identifying larger nodules (AUC of 0.98 for nodules > 8 mm).

Conclusion and Relevance: The VIT platform validated the superior diagnostic accuracy of CT over CXR, especially for smaller nodules, underscoring its potential to replicate real clinical imaging trials. These findings advocate for the integration of virtual trials in the evaluation and improvement of imaging-based diagnostic tools.

Introduction

Lung cancer ranks as the leading cause of cancer-related deaths, accounting for approximately 1.8 million fatalities in 2020.[1,2] Projections from the American Cancer Society indicate an estimated 238K individuals in the United States are anticipated to be diagnosed with lung cancer during the year 2023.[3] In the realm of timely Lung cancer detection and diagnosis, chest imaging modalities, such as X-rays and Computed Tomography (CT) scans, play a pivotal role. They offer comprehensive insights into abnormalities within the thoracic region.[4]

Associating the early-state cancer detection to the larger likelihood of cure, worldwide multiple lung cancer screening trials such as National Lung Screening Trial (NLST),[4] Nederlands-Leuvens Longkanker Screenings Onderzoek (NELSON) trial,[5] The Multicentric Italian Lung Detection (MILD) trial,[6] The British Thoracic Society Lung Cancer Screening Group (BTSLSG) trial,[7] and The International Early Lung Cancer Action Program (IE-LCAP) trial,[8] contributed valuable insights into the efficacy and implementation of lung cancer screening programs.

However, addressing costs, radiation exposure, and ethical considerations becomes imperative to ensure both financial feasibility and patient safety.[9] Lung cancer screening trials, including the NLST and NELSON, show varying rates of unnecessary procedures (1% to 8%) following positive LDCT scans, with no resulting cancer diagnosis.[4,5,10] Of particular concern are the 4% of patients with 'risk-gap' nodules (3% to 5% cancer probability) who undergo unnecessary procedures, impacting an estimated 7,000 individuals out of 9 million eligible for screening in the U.S. based on NLST criteria.[10]

In the sphere of computational power, virtual simulation, and artificial intelligence, the concept of Virtual Imaging Trials (VITs) emerges as a potent tool to expedite and reshape the landscape of clinical imaging trials. Serving as a swifter, more cost-effective, and safer alternative, it has the potential to revolutionize the traditional approach.[11-13] In their comprehensive review paper, Abadi *et al.* delve extensively into various methods and applications of VITs.[12] Abadi et al. showed coronavirus disease (COVID-19) simulation and Emphysema Quantification utilizing VITs.[14] VITs has also been reported in optimized medical imaging systems, Organ Dosimetry,[15] Furthermore, as the field evolves, acknowledging the potential influence of the evolution of therapies on screening's long-term impact is crucial for informed regulatory decision-making. Badano *et al.* assess the viability of digital breast tomosynthesis (DBT) over digital mammography (DM) through a VIT.[11] Tushar *et al.* showed the effect of data diversity on AI-based diagnosis utilizing VIT.[13]

This research centers around three primary goals: first, the generation of a virtual patient population that reflects the demographics and imaging data characteristics of the clinical Lung Screening Trials (LSTs) for cancer detection cohort; second, the encapsulation of 2000's imaging techniques (CXR and CT) to create virtual images that align with the old imaging technology; and finally, the execution detection task through virtual reader studies using these images. These studies aim to gauge the precision and efficacy of the virtual imaging trial platform. By merging innovative computational approaches with clinical significance, this investigation highlights the transformative capacity of virtual imaging trials in reshaping medical research and practice within a controlled and ethically mindful environment. This, in turn, fosters evidence-based progress.

Methods

Study Design

Fig.1 shows an overview of the different steps performed to accomplish the VLST comprised of simulation virtual population, modeling virtual scanners, development, and evaluation of virtual readers, and finally analysis of the detection and diagnosis performance. In the section below, we will briefly explain each of the steps.

Trial Virtual Population

In any clinical trial, the selection of the targeted population and sample groups holds paramount importance as it directly influences the validation of trial results and outcomes. The composition of the VIT population cohort is established through power calculations, designed to replicate the outcomes of lung cancer detection detailed in eAppendix 1.

Creating a virtual human phantom involves two key steps: constructing the overall body habitus and detailing the specific anatomical abnormalities. In this context, for the VLST, the subjects employed are virtual human phantoms, meticulously generated from authentic clinical CT/PET scans gathered from a singular institutional health system (**Duke University Medical Center**) which encompasses multiple hospitals.

The construction of these simulated human phantoms involves a series of intricate steps: commencing with the segmentation of specific organs, ensuring the quality control of these segmentations, followed by the creation of airways and vessels, texturization of internal organs, and culminating in the voxelization process. A comprehensive outline of this procedure in **Fig. 1**, while further information is available in our previous publication.[16] In this research, we employed both pre-existing and newly developed 4D extended cardiac-torso (XCAT) models representing both sexes at varying age, height, weight, BMI and race combinations.[16-18] Demographically, the mean age stood at approximately 59.5 years, with a near-equal distribution between male (55.68%) and female (44.32%) participants.

Furthermore, the generation of simulated nodules follows a dual-phase approach. Initially, a single-density lesion is formed, adhering to the desired morphology. Subsequently, a convolution process is applied to transform this single density simulated lesion into a multi-density structure.[19,20] while further information is available in our previous publication.[20] Several instances of these simulated nodules, featuring diverse characteristics such as size and morphology, are depicted in **eFigure 2**. These simulated nodules are then randomly inserted into the body habitus of virtual human population, guided by indications derived from prior clinical trial studies.[4] Specifics regarding population demographics, as well as the criteria for inclusion and exclusion, are comprehensively expounded upon in Table 1 and Fig. 2.

Virtual Scanners and Imaging Protocols

The developed virtual population was virtually imaged using a validated radiographic simulator (DukeSim)[21] and reconstructed using a vendor-neutral reconstruction tool (MCR toolkit)[22]. The radiographic simulation framework has been detailed in our previous works[14,21,23] and described briefly here. DukeSim generates projection images of voxelized computational models using ray tracing (for primary signal) and monte carlo simulation (for scatter signal and radiation dose). The projections were reconstructed with a weighted filter back projection (WFBP) algorithm using the MCR toolkit. In this study for CT simulation, DukeSim was configured to emulate the physical and geometrical properties of

two generic CT scanners named Duke Legacy W12 and Duke Legacy W20 which are representative of NLST CT techniques.[24] The geometry and acquisition configurations of the scanners are listed in the **eTable 3**.

Due to the lack of direct access to the original NLST CXR datasets, our study did not involve direct referencing images from the NLST. Instead, we examined example images displayed in scientific publications to inform our analysis. Our methodology involved the application of post-processing techniques to CXR images we simulated, to closely mirror the appearance of the example images found in the literature. This post-processing was crucial for ensuring that our images closely matched the visual attributes of older chest radiograph technology as depicted in the scientific papers we consulted.

Virtual Reader

The virtual reader component of our study is designed to emulate the image interpretation process typically performed by radiologists. This study is powered by two RetinaNet models (2D and 3D models), designed for processing CT and CXR images respectively as virtual readers.[25] We have meticulously selected and calibrated readily available machine learning models, optimizing them with publicly accessible clinical datasets (LUNA16,[26] NODE2[27]). This strategic choice ensures widespread accessibility and ease of replication, while also enhancing the models' versatility across a variety of diagnostic scenarios. The core concept was to standardize the model architecture for both CT and CXR to minimize virtual reader variability. The architecture, training methodologies, and patch extraction techniques remained constant across both models, with only the training datasets and data dimensions varying. This approach is akin to having the same radiologist interpret different modalities with consistent training.

The workflow starts with image data augmentation and leverages RetinaNet's feature pyramid network to extract multi-scale features.[25] Anchor boxes generated are matched with ground truth data, refined through regression, and classified to detect objects of interest.[28] The final predicted boxes and classifications are selected and evaluated against the ground truth.[25,28] Distinct from other trials, which commonly employ observer models,[26] this streamlined architecture mirrors the analytical process of a radiologist within a virtual environment. By incorporating such a model, we aim to closely replicate the decision-making process of radiologists, encompassing both the search and the detection phases that are critical in diagnostic imaging. Detailed information regarding the development, validation, and the clinical datasets used for the virtual reader models is comprehensively documented in the **eAppendix3**.

Trial End Point

The evaluation of performance was conducted at two levels: lesion-level and patient-level. At the lesion-level, each lesion was individually assessed to determine the model's accuracy in detecting and characterizing lesions. For patient-level evaluation, the highest predicted probability among all lesions within a single patient was utilized as the representative predictor for that patient, allowing for an assessment of the model's overall diagnostic performance on a per-patient basis. Performance was assessed using the Area Under the Curve (AUC) as a measure of diagnostic accuracy at both the lesion and patient levels, with subgroup analyses for lesion type and size.

The VLST's primary endpoint was the variance in the Area Under the Curve (AUC) between CT and CXR across the entire patient cohort. Additionally, subgroup analyses were conducted to evaluate the AUC variation for two distinct lesion types and for lesions of varying sizes.

Statistical Analysis

Performance was assessed using the AUC. The 95% CIs were calculated using the DeLong method with 2000 bootstrapping samples.

Results

In this VLST, virtual readers assessed 298 simulated CT scans and 313 simulated chest X-rays (CXR) originating from a cohort of 313 virtual human subjects. This cohort included 174 individuals with lung nodules and 139 without. A total of 512 lesions, with the following lung lesion types: 202 (39.4%) homogeneous and 310 (60.5%) heterogenous were evaluated. The mean size (largest axis) of these lesions was found to be 10.09 mm, with a standard deviation of 5.09 mm, indicating variability in lesion sizes across the sample. The smallest lesion measured was 4 mm, while the largest recorded lesion was 34 mm. The distribution of lesion sizes shows that 25% of lesions were 6 mm or smaller (first quartile), the median size was 9 mm, and 75% of the lesions (third quartile) were 12 mm or smaller.

At the lesion level, CT outperformed CXR with an AUC of 0.81 (95% CI: 0.78-0.84) compared to 0.55 (95% CI: 0.53-0.56), reflecting a substantial difference in diagnostic precision between the two modalities. The patient-level evaluation further underscored this disparity, with CT achieving an AUC of 0.85 (95% CI: 0.80-0.89) relative to CXR's 0.53 (95% CI: 0.47-0.60).

When stratifying the results by lesion type, homogeneous lesions detected via CT had an AUC of 0.97 (95% CI: 0.95-0.98), while CXR reported an AUC of 0.64 (95% CI: 0.61-0.66). For heterogeneous lesions, CT presented an AUC of 0.71 (95% CI: 0.67-0.75) versus CXR's 0.50 (95% CI: 0.48-0.53). Analysis by lesion size (largest axis) also revealed differences in AUC between CT and CXR. For nodules < 8 mm, CT had an AUC of 0.57 (95% CI: 0.52-063), whereas nodules ≥ 8 mm saw a higher AUC of 0.98 (95% CI: 0.96-0.99). In contrast, CXR's AUC for nodules < 8 mm was 0.57 (95% CI: 0.52-0.62), whereas nodules ≥ 8 mm saw a higher AUC of 0.71 (95% CI: 0.67-0.75)

Discussion

VIT methodology replicates three fundamental elements of an imaging trial: patients, scanners, and readers. This study's main purpose was to establish a VIT Platform for lung cancer diagnosis that replicates real-world LSTs' characteristics. We evaluated the diagnostic accuracy of CT and CXR imaging modalities within this virtual trial platform. By simulating a diverse virtual patient population, employing machine learning algorithms as standardized virtual readers, and utilizing validated radiographic simulators for imaging, our approach minimized variability and was closely aligned with real-world radiological assessments. Results demonstrated the CT outperformed the CXR in Nodule detection which is consistent with the performance seen in the real lung screening trials.

The composition of the trial population stands as a pivotal element in the execution and success of any clinical study. The VICTRE trial represents a significant milestone within the realm of virtual imaging trials, exhibiting results that are promising when juxtaposed with those derived from human trials.[11] Our investigation utilized a comparably smaller virtual cohort than that of existing clinical lung screening studies.[4,5,7] Despite this, our study extends beyond the scope of prior VIT research, which typically focuses on simulating specific pathologies or singular organs.[11,12] The construction of our trial population posed unique challenges due to the necessity of replicating the comprehensive thoracic anatomy of a virtual human. This complexity underscores the innovative approach of our study and contributes to the advancement of virtual trials in the simulation of more complex bodily systems.

To mirror the heterogeneity encountered in real-world clinical settings,[4] our study meticulously incorporates diversity within the virtual population's demographics. This inclusion spans across a spectrum of age, gender, and ethnicity, reflecting the multifaceted human population encountered in actual clinical trials.[4] Furthermore, we calibrated our virtual models to account for a range of lesion sizes, drawing upon the vast clinical knowledge and data accumulated from existing clinical trials. This diversity and clinical realism not only enhance the relevance of our findings but also fortify the potential for our virtual trial to serve as a robust model for future research in complex bodily system simulations. For future work, our efforts will pivot towards enriching the virtual population with an even broader spectrum of diverse and underrepresented virtual subjects.

The virtual readers were designed using established deep-learning libraries[29] and were trained on open-access clinical datasets[26,27] to ensure that the development process remained straightforward and broadly applicable. Unlike traditional VITs, which typically employ imaging physics-based observer models that evaluate specific imaging regions for the presence or absence of signal,[11,30] our virtual readers incorporate a search process.[25] This process mirrors the diagnostic approach of a radiologist by assessing the likelihood of a region containing a significant finding. While our deep-learning models are susceptible to the same training and testing biases previously noted in our VIT studies for COVID-19 diagnosis,[13,31,32] we have noted that these biases may be more pronounced in chest X-ray (CXR) analysis.[32] To closely emulate the clinical setting and avoid introducing additional complexities, we have deliberately avoided intricate model architectures and elaborate training methodologies, thereby aligning the virtual readers' functionality as closely as possible with that of human readers across different imaging modalities.

Reported results demonstrated the CT outperformed the CXR in Nodule detection which is consistent with the performance seen in the real lung screening trials. Our simulated homogenous lesions performed higher detection performance in both modalities compared to the heterogeneous. VITs usually focus on the task of lesion detection. However, existing lung cancer clinical imaging trials evaluate the patient-level end-point outcomes as cancer diagnosis.[4] Directly translating VLST's detection patient-level end-point results to NLST's patient-level end-point is not linear as most cases the clinical end-point outcomes are not only driven by findings supported by the imaging but also by additional information not accessible to the VIT's virtual reader. Recognizing this limitation, our future research aims to expand VITs to more closely replicate clinical trial scenarios by considering patient-level outcomes, such as cancer diagnoses, to enhance the clinical relevance of VITs.

In conclusion, the transformative potential of VITs in advancing evidence-based medicine offers an efficient and ethically conscious approach to medical research and development.
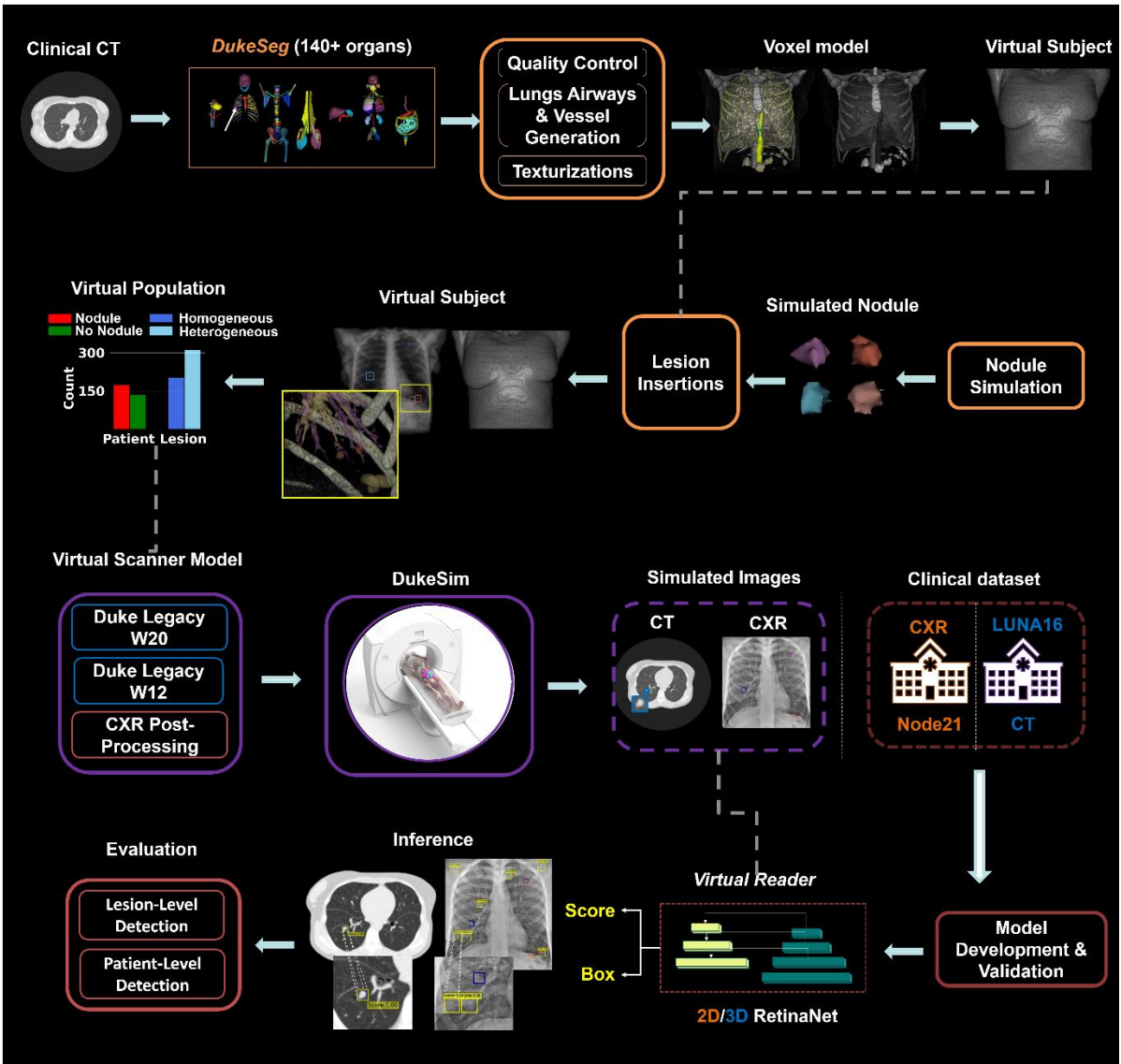
Acknowledgments

Figure 1. Flowchart summarizing the comprehensive workflow of the Virtual Lung Screening Trial (VLST), from the creation of virtual patient models using clinical CT data, through lesion simulation and insertion, to the virtual imaging using DukeSim, and concluding with the evaluation by a virtual reader employing 2D/3D RetinaNet for both lesion-level and patient-level detection evaluation.

Table 1. Cohort Characteristics of the VLST Virtual Population of 264 out of 313* patients With Cases Corresponding to with and without simulated nodule.

Characteristics	All	With Lung Nodule (n=174)	Without Lung Nodule (n=90)
Age (year)			
Mean ± Std	59.52 ± 14.44	59 ± 15	60 ± 13
Sex			
Male	147 (55.68%)	95 (54.60%)	52 (57.78%)
Female	117 (44.32)	79 (45.40%)	79 (42.22%)
Weight (kg)			
Mean ± Std (min-max)	78.33 ± 20.31 (17. 00- 148.00)	77.14 ± 20.52 (17. 00- 148.00)	80.62 ± 19.82 (30. 00- 133.90)
BMI			
Mean ± Std (min-max)	27.17 ± 5.99 (13-50.49)	26.69 ± 6.08 (13-50.49)	27.97 ± 5.76 (18.37-41.65)
Race			
White	198 (75 %)	128 (73.56%)	70 (70.78%)
Black or African American	56 (21.21%)	36 (20.69%)	20 (22.22%)
Other/Unknown	(10 3.79%)	10 (5.75 %)	
Ethnic			
Not Hispanic or Latino	260 (98.48)	171 (98.28%)	89 (98.89 %)
Hispanic/ Unknown	4 (1.52 %)	3 (1.73 %)	1 (1.11%)

*Note: Characteristics for the existing 49 phantoms are currently unavailable.
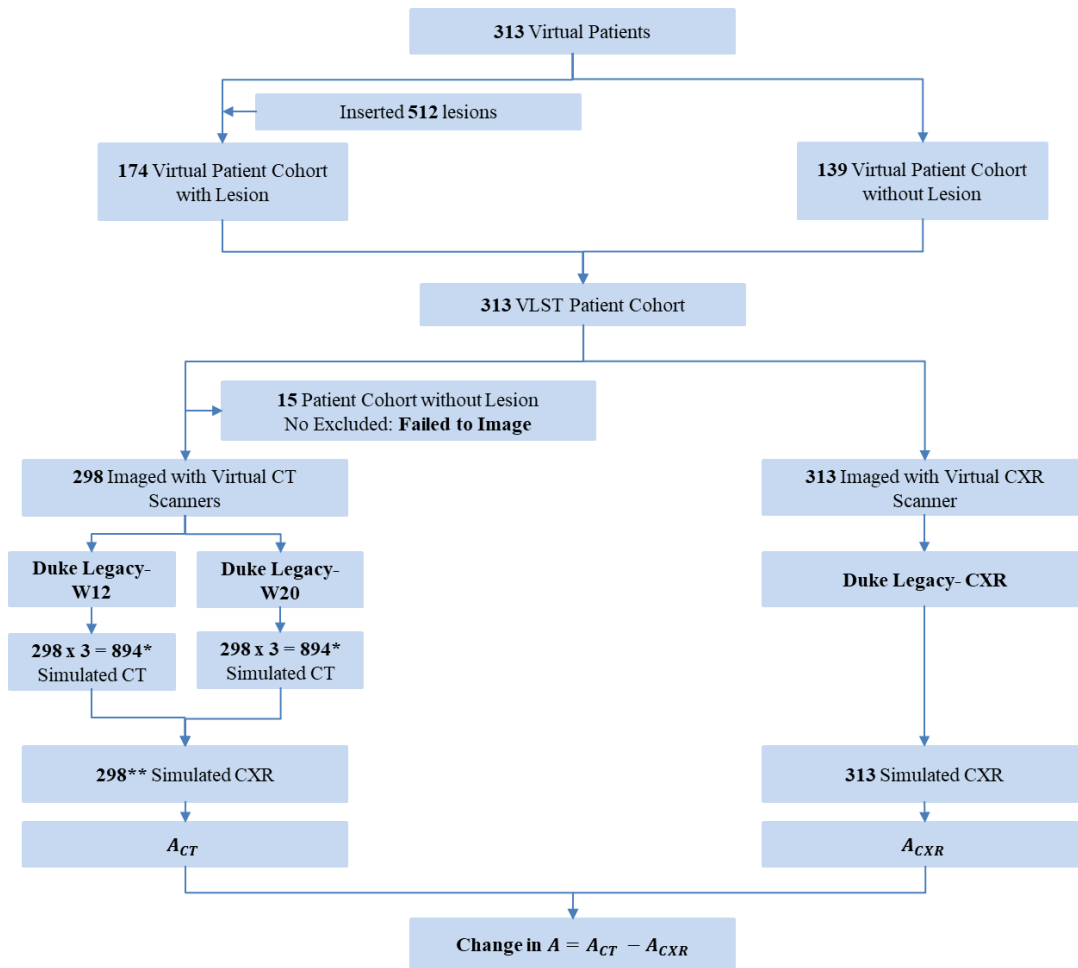
Figure 2. Virtual Patient's inclusion and exclusion criteria and progress through the study. Within the study, 313 virtual patients were assessed, with 174 of them having a total of 512 lesions, varying from homogeneous to heterogeneous in nature. All 313 virtual patients underwent both virtual CT and CXR scans. For the CT cohort, 15 virtual patients without nodule were excluded due to imaging failures. *The remaining 298 virtual images were processed through two distinct scanners, the Duke Legacy-12 and the Duke Legacy-20, with each scanner producing three unique imaging configurations per patient (298 x 3=894). **From these six configurations, one CT image per patient, total 298 was randomly selected for evaluation. As for the CXR cohort, all 313 virtual patients were successfully imaged using the Duke Legacy-CXR scanner. A indicate the area under the receiver operating characteristic curve.

Table. 2: VLST End Points (AUC and Change in AUC) for the CT and CXR in lesion and patient-level, sub-group analysis with lesion type and lesion size (largest axis).

VLST	AUC (95% CI)		p-value
	CT	CXR	
Lesion-level	0.81 (0.78-0.84)	0.55 (0.53-0.56)	$< 2.2e^{-16}$
Patient-level	0.85 (0.80-0.89)	0.53 (0.47- 0.60)	$< 6.554e^{-14}$
By lesion-type			
Homogeneous lesion (n=202)	0.97 (0.95-0.98)	0.64 (0.61-0.66)	$< 2.2e^{-16}$
Heterogeneous lesion (n=310)	0.71 (0.67-0.75)	0.50 (0.48-0.53)	$< 2.2e^{-16}$
By lesion-size			
Nodule < 8 mm (n=79)	0.57 (0.52-0.63)	0.57 (0.52-0.62)	*0.961*
Nodule > = 8 mm (n=433)	0.98 (0.96-0.99)	0.71 (0.67-0.75)	$< 2.2e^{-16}$

References

1.   Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. *International Journal of Cancer*. 2021;149(4):778-789. doi:10.1002/ijc.33588

2.   Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209-249. doi:10.3322/caac.21660

3.   Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *Ca Cancer J Clin*. 2023;73(1):17-48.

4.   Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine*. 2011;365(5):395-409. doi:10.1056/nejmoa1102873

5.   De Koning HJ, Van Der Aalst CM, De Jong PA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *New England Journal of Medicine*. 2020;382(6):503-513. doi:10.1056/nejmoa1911793

6.   Pastorino U, Sverzellati N, Sestini S, et al. Ten-year results of the Multicentric Italian Lung Detection trial demonstrate the safety and efficacy of biennial lung cancer screening. *European Journal of Cancer*. 2019;118:142-148. doi:10.1016/j.ejca.2019.06.009

7.   Field JK, Duffy SW, Baldwin DR, et al. The UK Lung Cancer Screening Trial: a pilot randomised controlled trial of low-dose computed tomography screening for the early detection of lung cancer. *Health Technology Assessment*. 2016;20(40):1-146. doi:10.3310/hta20400

8.   Welch HG, Woloshin S, Schwartz LM, et al. Overstating the Evidence for Lung Cancer Screening. *Archives of Internal Medicine*. 2007;167(21):2289. doi:10.1001/archinte.167.21.2289

9.   Heleno B, Thomsen MF, Rodrigues DS, Jørgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. *BMJ : British Medical Journal*. 2013;347:f5334. doi:10.1136/bmj.f5334

10.   Sundaram V, Gould MK, Nair VS. A Comparison of the PanCan Model and Lung-RADS to Assess Cancer Probability Among People With Screening-Detected, Solid Lung Nodules. *Chest*. 2021/03/01/ 2021;159(3):1273-1282. doi:https://doi.org/10.1016/j.chest.2020.10.040

11.   Badano A, Graff CG, Badal A, et al. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Network Open*. 2018;1(7):e185474. doi:10.1001/jamanetworkopen.2018.5474

12.   Abadi E, Segars W, Tsui BM, et al. Virtual clinical trials in medical imaging: a review. *Journal of Medical Imaging*. 2020;7(4):042805.

13.   Tushar FI, Dahal L, Sotoudeh-Paima S, et al. Data diversity and virtual imaging in AI-based diagnosis: A case study based on COVID-19. *arXiv preprint arXiv:230809730*. 2023;

14.   Abadi E, Paul Segars W, Chalian H, Samei E. Virtual Imaging Trials for Coronavirus Disease (COVID-19). *AJR Am J Roentgenol*. Feb 2021;216(2):362-368. doi:/10.2214/AJR.20.23429

15.   Fu W, Sharma S, Abadi E, et al. iPhantom: A Framework for Automated Creation of Individualized Computational Phantoms and Its Application to CT Organ Dosimetry. *IEEE J Biomed Health Inform*. Aug 2021;25(8):3061-3072. doi:10.1109/JBHI.2021.3063080

16.   Lavsen D, Yuqi W, Fakrul Islam T, et al. Automatic quality control in computed tomography volumes segmentation using a small set of XCAT as reference images. 2023:1246342.

17.   Segars WP, Bond J, Frush J, et al. Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Medical Physics*. 2013;40(4):043701. doi:https://doi.org/10.1118/1.4794178

18.   Data from: American Association of Physicists in Medicine, Truth-Based CT (TrueCT) Reconstruction Challenge

19.   Sauer TJ, Bejan A, Segars P, Samei E. Development and CT image-domain validation of a computational lung lesion model for use in virtual imaging trials. *Medical Physics*. 2023;50(7):4366-4378. doi:https://doi.org/10.1002/mp.16222

20.   McCabe C, Solomon J, Segars WP, Abadi E, Samei E. *Synthesizing heterogeneous lung lesions for virtual imaging trials*. vol 12925. SPIE Medical Imaging. SPIE; 2024.

21.   Abadi E, Harrawood B, Sharma S, Kapadia A, Segars WP, Samei E. DukeSim: A Realistic, Rapid, and Scanner-Specific Simulation Framework in Computed Tomography. *IEEE Transactions on Medical Imaging*. 2019;38(6):1457-1465. doi:/10.1109/tmi.2018.2886530

22.   Clark DP, Badea CT. MCR toolkit: A GPU-based toolkit for multi-channel reconstruction of preclinical and clinical x-ray CT data. *Medical Physics*. 2023;

23.   Jadick G, Abadi E, Harrawood B, Sharma S, Segars WP, Samei E. A scanner-specific framework for simulating CT images with tube current modulation. *Physics in Medicine & Biology*. 2021;66(18):185010.

24.   Cagnon CH, Cody DD, McNitt-Gray MF, Seibert JA, Judy PF, Aberle DR. Description and implementation of a quality control program in an imaging-based clinical trial. Elsevier; 2006.

25.   Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. 2017:2980-2988.

26.   Setio AAA, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images:

The LUNA16 challenge. *Medical Image Analysis*. 2017;42:1-13. doi:10.1016/j.media.2017.06.015

27.	Sogancioglu E, Murphy K, Ginneken Bv, Scholten E, Schalekamp S, Hendrix W. Data from: Node21. Grand Challenge Competition. 2021.

28.	Zhang S, Chi C, Yao Y, Lei Z, Li SZ. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. 2020:9759-9768.

29.	Cardoso MJ, Li W, Brown R, et al. Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:221102701*. 2022;

30.	Samei E, Krupinski EA. *The handbook of medical image perception and techniques*. Cambridge University Press; 2018.

31.	Tushar FI, Abadi E, Mazurowski MA, Segars WP, Samei E, Lo JY. Virtual vs. Reality: External Validation of COVID-19 Classifiers using XCAT Phantoms for Chest Computed Tomography. presented at: Proc Spie; 2022;

32.	Dahal L, Tushar FI, Abadi E, et al. Virtual vs. Reality: External Validation of COVID-19 Classifiers using XCAT Phantoms for Chest Radiography. presented at: Medical Imaging 2022: Physics of Medical Imaging; 2022;