ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ARTIFICIAL INTELLIGENCE

MASTER THESIS

in

Artificial Intelligence for Medicine M

A COMPUTED TOMOGRAPHY ATLAS OF PULMONARY NODULES FOR LUNG CANCER SCREENING

CANDIDATE
Farhad Bayrami

SUPERVISOR
Prof. Stefano Diciotti

CO-SUPERVISOR
Dr. Giulia Raffaella De Luca

Academic Year 2024/2025

3rd Session

Abstract

Lung cancer is the leading cause of cancer-related mortality worldwide, and early detection remains critical for improving survival rates. This thesis proposes an integrated framework for automated construction of a probabilistic nodule atlas using low-dose computed tomography (LDCT) data from the National Lung Screening Trial (NLST). The approach leverages two key methodologies: Lesion Locator, a deep learning model for zero-shot segmentation, and the Advanced Normalization Tools (ANTs) framework for inter-subject registration. Together, these tools address challenges of anatomical variability, segmentation consistency, and inter-patient comparison in LDCT screening data. LesionLocator was applied to the NLST dataset to generate nodule masks. Subsequently, these masks were transformed into a standardized anatomical space defined by a previously obtained lung template via Symmetric diffeomorphic (SyN) non-linear registration. A nodule frequency map in the template space is then constructed by accumulating the warped nodule masks. This map highlights high-density nodule regions in the upper lobes and lung periphery, validating both the registration accuracy and the model's capacity to reproduce clinically known nodule distribution patterns in lung cancer screening. This work establishes a reproducible computational workflow that generates objective spatial priors for population-based analysis. The resulting atlas can serve as a foundational component for future risk stratification tools and detection models, thereby complementing current radiological assessment by automating high-throughput spatial measurements. Overall, this thesis lays the groundwork for next-generation diagnostic systems that integrate machine learning, medical imaging, and population-based analysis to enhance early cancer detection and personalized patient care.

TABLE OF CONTENTS

List of Figures

List of Tables

1. Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, responsible for nearly one in five cancer deaths [1]. Early detection of pulmonary nodules through medical imaging is fundamental for improving prognosis. Computed tomography (CT) offers high-resolution volumetric imaging of lung tissue and has become the gold standard for screening and follow-up [1, 2]. Low-dose CT (LDCT) screening programs enable population-level screening and identification of millimetric nodules that previously remained undetected.

Despite its diagnostic power, LDCT interpretation faces critical challenges in a screening setting: it is time-consuming, prone to inter-reader variability, and the high volume of subtle findings can lead to fatigue. Furthermore, differences in acquisition protocols, such as slice thickness and reconstruction kernels, combined with inherent anatomical variability, hinder consistent quantitative analysis across large patients cohorts [2]. Automated image analysis tools including standardized pre-processing and nodule annotation pipelines are therefore necessary for large-scale lung cancer screening trials. The application of artificial intelligence (AI) in lung cancer screening is illustrated in Figure 1.1.

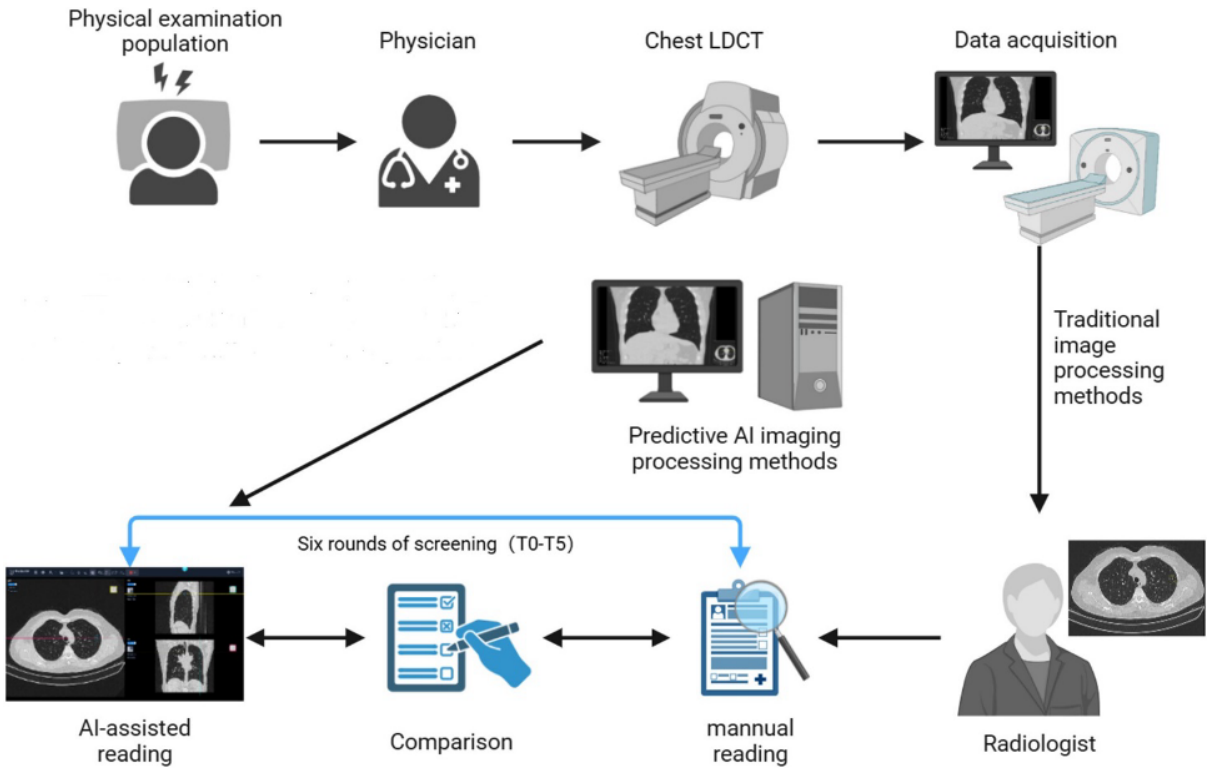


Figure 1-1 Application of Artificial Intelligence in Lung Cancer Screening [1].

The emergence of AI has transformed radiology, with a pronounced impact on thoracic image analysis and lung cancer screening. AI-driven models, functioning as Computer-Aided Detection and Diagnosis (CADe/CADx) systems, directly address the throughput and variability issues of LDCT screening by automating the primary task: pulmonary nodule identification and segmentation [3]. Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in detecting and segmenting pulmonary nodules [4]. Recent transformer-based and hybrid architectures further capture broader contextual information, improving the delineation of challenging nodules adjacent to vessels or pleural surfaces [4]. Comparative evaluations confirm that these AI-based approaches surpass traditional rule-based techniques in both accuracy and efficiency [8]. By automating precise nodule segmentation, AI reduces observer variability and provides reproducible quantitative data for subsequent radiomics-based analyses [1,2].

Despite rapid progress, existing AI systems still face limitations related to dataset bias, lack of generalization, and absence of anatomical standardization [2]. Models trained on single-centre datasets, such as LIDC-IDRI or LUNA16 [4], often fail to generalize to external data, acquired with different scanners or reconstruction settings [5]. Most studies rely on overlap metrics like Dice similarity coefficient (DSC) or intersection over union (IoU), which are insufficient for assessing consistent spatial alignment between subjects. Consequently, robust cross-patient and population-level analyses remain unreliable.

Recent zero-shot and self-supervised frameworks, such as Lesion Locator [6], aim to segment lesions across multiple organs and modalities without retraining. However, these segmentations still depend on consistent spatial references to ensure anatomical validity when aggregating results across cohorts. This critical dependency highlights the need to systematically integrate AI segmentation outputs with robust image registration to achieve reproducible spatial correspondence and unlock reliable population-level statistics.

Image registration is the indispensable tool for overcoming anatomical variability, aligning images from different individuals into a shared coordinate system and enabling voxel-wise correspondence [7]. Affine and diffeomorphic registration algorithms, as implemented in the ANTs ecosystem [5], permit precise and consistent mappings of anatomical structures or lesions, such as lung nodules. This process minimizes positional and morphological variability, which is necessary to support reliable statistical analyses of clinical features.

Atlas construction is the generalization of this registration concept to a cohort level, producing a mean anatomical model (the template) representing the typical lung structure and its spatial variance [7]. Such atlases are critical for identifying deviations related to disease or

demographic factors. Therefore, combining AI segmentation with registration enables mapping of automatically detected lesions into a common lung space, thereby facilitating comparative and longitudinal population studies [7]. The integration of these two components ensures that automatically detected regions are not only segmented consistently but are also spatially normalized and anatomically consistent across the entire population.

The overall objective of this thesis is to develop and evaluate a computational framework that integrates AI-based lung nodule segmentation with image registration to achieve anatomically standardized population-level analyses. The specific aims are as below.

- To implement advanced AI-based lesion segmentation using LesionLocator [8]
- To apply ANTs deformable registration methods for aligning LDCT volumes from NLST to a previously described common template [9].
- To construct and quantify a probabilistic nodule atlas representing the spatial frequency of pulmonary nodules across the cohort.
- To evaluate how registration improves anatomical reproducibility and enables reliable inter-subject comparison of AI segmentation outputs.

This thesis is organized into five chapters. Chapter 1 presents the background, rationale, and specific objectives of the study. Chapter 2 provides an extensive literature review covering AI for lung nodule analysis, image registration techniques, and atlas construction. Chapter 3 describes the datasets, pre-processing, and methodological framework integrating segmentation and registration. Chapter 4 reports the experimental results and discusses their implications. Chapter 5 summarizes conclusions, highlights the novel contributions of this work, and suggests directions for future research.

2. Background

2.1 Pulmonary Nodule Detection and Segmentation

Traditional methods for pulmonary nodule detection and segmentation have been foundational in the evolution of diagnostic imaging, leveraging image processing techniques to exploit contrast and morphological characteristics in computed tomography (CT) scans, including low-dose CT (LDCT) scans used in lung cancer screening trials. Key techniques include thresholding, which isolates nodules based on specific Hounsfield Unit (HU) ranges (e.g., -600 to -400 for dense lesions), and region growing, which iteratively expands from manually selected seed points based on intensity similarity. Subsequent morphological operations, such as erosion and dilation, are then applied to remove small noise artifacts, fill gaps in segmented regions and refine the segmentation boundaries, ensuring smoother contours and reducing irregularities [8].

Despite the inherent simplicity and low computational requirements of these approaches, making them accessible in resource-limited settings where advanced hardware or AI expertise may be unavailable, they are fundamentally semi-automated, requiring radiologist input for seed point selection or parameter tuning. This reliance introduces significant inter-observer variability and limits reproducibility, as evidenced by reported DSC averaging around 0.70 across datasets [10]. These traditional methods also struggle with nodule heterogeneity. For instance, ground-glass opacities (GGOs)—hazy areas of increased attenuation where bronchial and vascular structures remain visible [3]—are difficult to segment due to their low contrast (HU -800 to -500) and scattered borders compared to solid nodules. This results in segmentation errors, with false positives often exceeding 30% due to misclassification of vascular structures, calcifications, or normal lung tissue as nodules [10].

Furthermore, the scalability of these semi-automated tools is severely constrained for massive datasets like the NLST, which includes over 1 million LDCT images from 26,254 participants, where manual adjustments would be impractical and time-consuming, often requiring hours per scan for complex cases [7].

Traditional methods are also highly sensitive to LDCT's lower signal-to-noise ratio, a consequence of reduced radiation doses (1-4 mSv vs. 7-8 mSv for standard CT), which increases error rate in the detection of subtle lesions like GGOs or nodules smaller than 6 mm [3]. These performance deficiencies, consistently highlighted in comparative studies like that of Bianconi et al. [7], necessitate the transition to fully automated, high-throughput AI-based

methods, though traditional techniques retain importance as essential benchmarks and components within sophisticated hybrid segmentation pipelines.

Deep learning has revolutionized pulmonary nodule detection and segmentation by automating feature extraction and significantly improving accuracy on large-scale datasets Gao et al. (2025) [4] conducted a systematic review of 109 studies, identifying convolutional neural network (CNN)-based models as the cornerstone of modern approaches, with leading architectures like Faster R-CNN for detection and U-Net variants for segmentation achieving DSC exceeding 0.90 and sensitivities of 94% for nodules larger than 6 mm.

These models leverage hierarchical feature learning to identify subtle patterns in CT images—such as texture gradients, edge irregularities, and size variations—that are critical for distinguishing nodules from background structures. For example, Chen et al. [11] introduced the Artificial-Intelligence Lung Imaging Analysis System (ALIAS), which integrates a ReLU cascade Feature Pyramid Network (FPN) for multi-scale nodule detection and a VB-Net architecture for precise segmentation, achieving DSC values above 0.90 on the LIDC-IDRI dataset, a standard benchmark with 1,018 CT scans and 2,669 nodules [11]. ALIAS's strength lies in its end-to-end pipeline, which combines detection, segmentation, and radiomic feature extraction (e.g., HU histograms, texture metrics), supporting clinical decision-making, achieving an accuracy of 87% in the differentiation of benign from malignant nodules [11].

LesionLocator, developed at DKFZ [8], represents a significant advancement in this field byoffering prompt-based zero-shot universal and longitudinal tracking without case-specific retraining. Built on a densely promptable 3D U-Net combined with a UNet-based backbones, it leverages large multi-modal pretraining and fine-tuning on 22,580 annotated lesions across diverse anatomical sites. [6]. Lung-related data are included at multiple stages: thoracic structures appear in the Medical Segmentation Decathlon during pretraining, while DeepLesion contributes numerous pulmonary findings—such as nodules, masses, and metastases—during lesion-specific fine-tuning. For strict zero-shot evaluation of lung nodule segmentation, the model is tested on the Coffee-Break/RAIDER Lung CT dataset, which is held out entirely from training and used to assess generalization to unseen reconstruction protocols. Operating directly on 3D CT volumes and supporting point, box, or prior-mask prompts, LesionLocator achieves state-of-the-art performance across out-of-distribution lesion datasets and provides robust longitudinal propagation from a single initial prompt.

LesionLocator’s prompt-based mechanism allows users to specify nodule centroids (e.g., via coordinates from NLST annotations), enabling precise segmentation even in low-contrast LDCT images, which is critical for early-stage adenocarcinoma detection [8].

Models trained on single-centre datasets, such as LIDC-IDRI or LUNA16, often fail to generalize to external data, acquired with different scanners or reconstruction settings. This is because datasets like LIDC-IDRI (multi-radiologist annotated but primarily from U.S. centers with similar protocols) and LUNA16 (a subset of LIDC-IDRI focused on detection) exhibit biases in ethnicity (predominantly Western populations), scanner types (e.g., Siemens, GE), and reconstruction kernels (e.g., B50f vs. B30f), leading to performance drops from internal DSC >0.90 to ~0.85 externally when applied to diverse cohorts like NLST or NELSON, which have varying slice thicknesses (0.6-2.5 mm) and acquisition parameters (120-140 kVp) [4], [8].

LesionLocator addresses some of these by using a transformer-based model (SAM-Med2D architecture with ViT-b backbone and adapter layers, fine-tuned on 4.6 million medical images) for zero-shot segmentation, reducing dependency on dataset-specific training [12].

Furthermore, computational requirements pose another barrier, as 3D CNNs demand high-end GPUs (e.g., NVIDIA RTX 3090) and significant processing time [12]. Hybrid models, which combine deep learning with traditional processing techniques, offer one partial solution to improve robustness in noisy data.

However, the most critical, yet often overlooked, challenge is the absence of anatomical standardization. While advanced models like LesionLocator provide accurate segmentations, via the combination of deep learning and traditional processing techniques, these masks are still resident in patient-specific coordinate systems. The reliance on simple overlap metrics (DSC/IoU) does not assess consistent spatial alignment between subjects. This means that even perfectly segmented nodules cannot be reliably aggregated for population-level statistical mapping.

In summary, the transition from traditional semi-automated segmentation algorithms to deep learning and zero-shot transformer-based approaches represents a major milestone in pulmonary nodule analysis. While conventional methods remain relevant as benchmarks or pre-processing tools, deep learning models like ALIAS and Lesion Locator achieve superior accuracy, generalizability, and scalability for population-level CT analysis. These advances provide the methodological foundation for subsequent sections of this thesis, which focus on data pre-processing, registration, and atlas-based standardization. To consolidate the discussion above, Table 2.1 provides a comparative summary of the main methodological categories reviewed in this section, ranging from conventional semi-automated algorithms to deep learning and transformer-based frameworks, emphasizing their Dice similarity scores, sensitivity, and current limitations.

Table 2-1 Comparison of Traditional and Deep Learning Methods for Nodule Segmentation

Method Type	Example Techniques	DSC Score	Sensitivity	Limitations	Source
Traditional	Thresholding, Region Growing	~0.70	80-90%	High false positives, manual input	[8]
Deep Learning	CNNs (e.g., U-Net, FPN)	>0.90	94%	Data bias, computational cost	[2], [5]
Zero-Shot DL	Lesion Locator (transformer-based)	0.85-0.92	92%	Prompt dependency	[3]

2.2 Image Registration Techniques

Image registration is the process of aligning different images (in terms of subjects, modality or timepoints) to a common coordinate system. The ANTs, as detailed by Tustison et al. (2021) [5], represent a gold standard open-source ecosystem for this task, offering comprehensive frameworks validated across neuroscience, oncology, and cardiology. ANTs utilizes a multi-stage process to progressively refine alignment, crucial for handling complex anatomical variability in thoracic CT scans. Developed as an open-source toolkit, ANTs is part of the broader ANTsX suite, which includes interfaces like ANTsR (for R integration) and ANTsPy (for Python), allowing seamless incorporation into diverse workflows, such as those involving machine-learning pipelines or statistical-analysis software.

ANTs registration script, included in the ANTsX toolkit, allows the registration of “moving” images to a “fixed” reference image through a multi-stage process that progressively refines the alignment to account for varying levels of deformation. The first stage involves rigid transformation, which corrects for basic misalignments such as rotation and translation using six degrees of freedom (DoF), ensuring that global orientation differences, common in patient-positioning variations, are addressed early. This is followed by the affine stage, which introduces an additional six DoF for scaling and shearing (total 12 DoF), accommodating differences in patient size or scanner calibration, such as variations in voxel spacing across different CT machines (e.g., GE vs Siemens). The final nonlinear stage employs symmetric diffeomorphic mapping (SyN), a hallmark of ANTs, to model local deformations like those caused by respiratory motion or tumour-induced distortions, using gradient steps (e.g., 0.1) and multi resolution shrink factors (e.g., 8×4×2×1) for efficient convergence. The calculated

transforms (e.g., affine matrix and warp field) are then applied to the moving image, and eventually its associated binary masks for given regions or lesions segmentations, to map it into this system.

Tustison et al. (2021) [5] emphasize ANTs' use of symmetric diffeomorphic mapping to ensure invertibility and smoothness, which preserves anatomical topology and reduces landmark errors to sub-millimeter levels in lung applications, making it superior for preserving biological plausibility in deformable tissues like the lungs during inspiration-expiration cycles [13]. ANTs has demonstrated strong performance in benchmarks like the EMPIRE10 challenge for intra-patient chest CT image registration, where it ranked highly in accuracy for aligning scans from the same subject under varying conditions, such as different respiratory phases . The tool's flexibility extends to customizable parameters, such as regularization coefficients (e.g., 3 for smoothing) to prevent folding in warps, and interpolation methods like BSpline (order 3) for continuous-intensity images to avoid aliasing, or Nearest Neighbour for preserving the integrity of binary masks (e.g., nodule segmentations from Lesion Locator [8], ensuring no fractional values are introduced).

Critiques in the literature note ANTs' superiority over rigid registration-only methods (e.g., FSL's FLIRT) in handling respiratory variations, where non-rigid deformations can reduce overlap errors by 20–30% in lung regions [5]. Comparisons to alternatives like Elastix or SimpleITK show ANTs' edge in diffeomorphic mapping for preserving topology, though Elastix may be faster for GPU-accelerated tasks [7]. The ecosystem's open-source nature, hosted on GitHub with over 1,000 citations and regular updates (as of October 08, 2025), underscores its reliability for medical research, supported by an active community that contributes to fixes, updates and extensions. In thoracic applications, ANTs, has been validated in over 500 studies for tasks like emphysema quantification or tumour tracking, with mean Hausdorff distances < 2 mm for lung boundaries [5]. Overall, ANTs, represents a gold standard in deformable registration, bridging traditional imaging challenges with modern AI workflows.

2.3 Normative Lung Atlases and Template Creation

Registration to a common reference space is fundamentally important for comparing images across different time points, modalities, or subjects. Population-based analysis in medical imaging leverages registered data to derive comprehensive statistical insights into diseases, enabling researchers to identify patterns, correlations, and predictive biomarkers across large cohorts. A normative atlas generalizes this registration to a cohort level, aggregating individual patient data into a collective framework to identify the spatial distribution of disease features.

Existing literature supports using this methodology for aggregating lung imaging data. Chen et al. [11] exemplify this methodology through their Artificial-Intelligence Lung Imaging Analysis System (ALIAS), which aggregates nodule segmentations from multiple CT scans into a standardized template space, identified in the CT scan of an arbitrary healthy subject, to build a lung nodule location and volumetric atlas, that could enhance early detection workflows [11].

Their maps in fact enabled voxel-wise comparisons, showing a well-known upper lobe predominance of nodules in a lung cancer screening cohortthereby validating the inter-subject registration capability to preserve the pathological characteristics of the population. Xu et al. [7] applied corrField to low-dose CT (LDCT) data for constructing a chest CT atlas from 853 scans, achieving a 91.7% registration success rate and enabling detailed phenotypic comparisons, such as correlations with body mass index (BMI), chronic obstructive pulmonary disease (COPD) severity, or coronary artery calcium presence .

ANTs multi-step and multi-resolution algorithm is exploited by the *anstMultivariateTemplateConstruction* script to build a population template out of a set of images, and was used to derive a the 3D lung template used in this thesis as reference space [9].

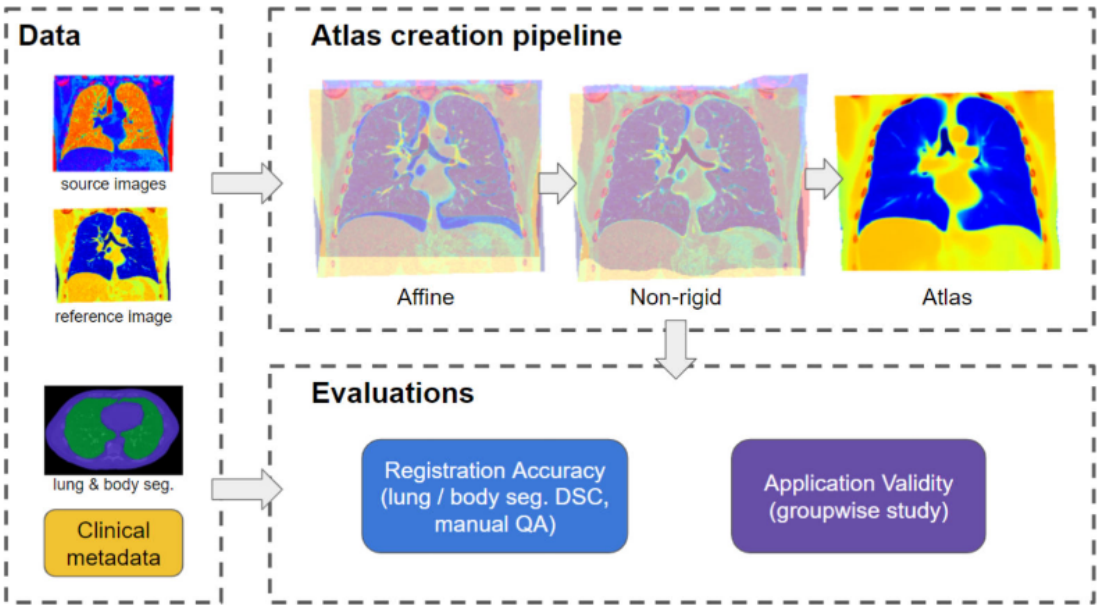


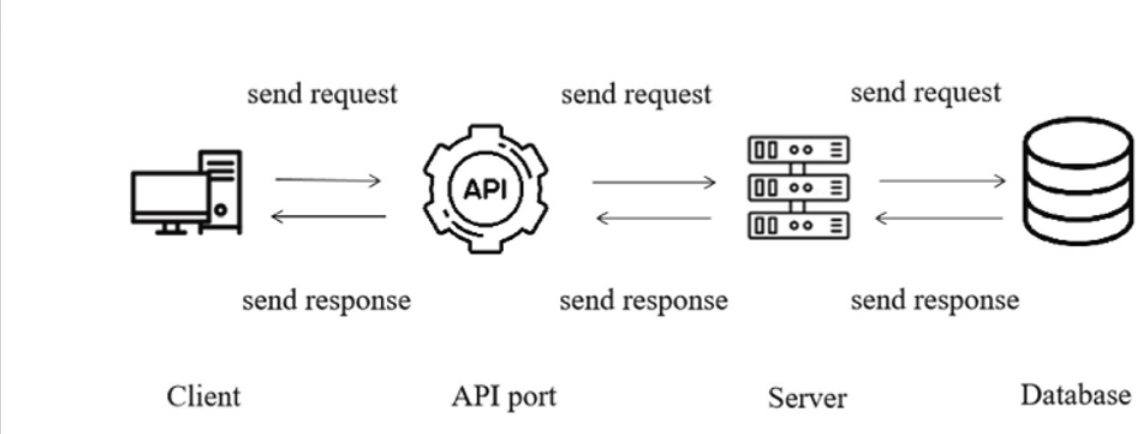
Figure 2-1 Workflow of chest CT atlas creation [7].

2.4 Data preparation for AI in lung imaging

Data preparation stands as a foundational and indispensable step in the workflow for AI-driven analysis and image registration, setting the stage for all subsequent processing by ensuring data is in a compatible, standardized format. Li et al. [14] provide a detailed and practical guide to this conversion process, emphasizing the complexities involved in transforming Digital Imaging and Communications in Medicine (DICOM) files, the standard format for medical imaging storage that includes rich metadata like patient information, scanner settings, and acquisition parameters, into the more analysis-friendly Neuroimaging Informatics Technology Initiative (NIfTI) format. This format, optimized for scientific computing, simplified the initial processing stage by bypassing the complex conversion from the proprietary DICOM standard. However, the integrity of the spatial metadata—including voxel dimensions, affine matrices for coordinate transformations, and orientation labels—must still be rigorously maintained to ensure geometric integrity for accurate alignment in downstream tasks [9]. For instance, errors in orientation, such as confusing RAS (Right-Anterior-Superior) with LPS (Left-Posterior-Superior) conventions, can severely skew results by inverting axes and misaligning nodules relative to anatomical landmarks. Therefore, automated parsing and validation scripts were employed to check and enforce consistent metadata, ensuring seamless compatibility with advanced segmentation and registration tools.

Effective dataset handling and pre-processing are foundational to optimizing the performance of AI-driven algorithms in medical imaging, particularly for lung-nodule detection and segmentation tasks, as comprehensively outlined by Wang et al. (2023) [10]. As shown in Figure 2.3, the data preparation pipeline begins with accessing and downloading raw CT scans from the database server and proceeds through a series of pre-processing steps to ensure data quality and consistency [2].

This multifaceted process encompasses a series of critical steps designed to prepare raw imaging data for analysis, ensuring consistency and quality. Normalization of HU values is another vital step, adjusting voxel intensities to a standardized range (e.g., −1000 to +1000 HU for lung CT) to account for scanner manufacturer-specific variations (e.g., GE vs Siemens) and ensure consistent feature extraction across datasets, often involving clipping outliers (e.g., metallic artefacts > 2000 HU) and rescaling using linear or z-score methods to handle batch effects. The characteristics and data availability of four large-scale lung CT datasets are summarized in Table 2.2, highlighting differences in format, slice thickness, demographic coverage, and annotation types [2].



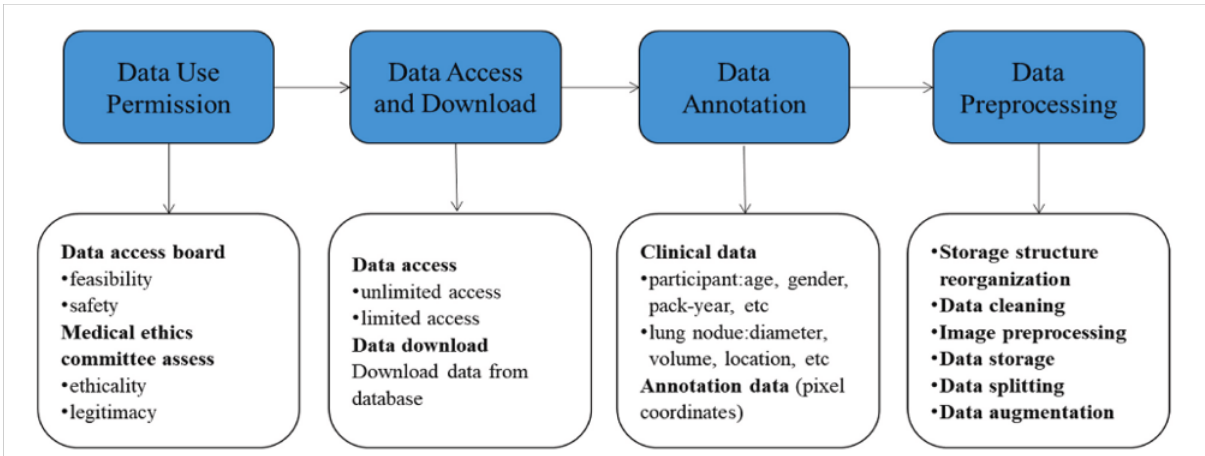


Figure 2-2 Data preparation process illustrating data access, download, and pre-processing workflow for lung CT datasets [2].

Table 2-2 Data availability and characteristics of four large-scale CT imaging datasets commonly used in lung nodule analysis [10]. Legend: "+" Available; "–" Not available.

Dataset	LIDC-IDRI	LUNA16	NLST	NELSON
Total number	1018 scans / 1010 patients	888 scans	54,000 participants	7557 participants
Format of CT images	DICOM	MHD and Raw	DICOM	DICOM
Slice thickness of CT images	0.5–5 mm	≤2.5 mm	1.0–3.2 mm	1.0 mm
Demographic description	+	+	+	+
Follow-up	-	-	+	+
Pixel-based annotation of lung nodule	+	+	–	–

Wang et al. provide detailed insights from four prominent datasets, LIDC-IDRI, LUNA16, NLST, and NELSON used for deep-learning preparation, noting their unique strengths: LIDC-IDRI for annotation variability, LUNA16 for benchmark competitions, NLST for real-world screening, and NELSON for European demographics [2].

Pre-processing for datasets like NLST and LIDC-IDRI requires specific adaptations due to their scale and variability. Resampling to uniform resolutions (e.g., isotropic 1 mm^3 voxels using cubic-spline interpolation via SimpleITK) is critical to align scans from different CT scanners, which may have slice thicknesses ranging from 0.6 to 2.5 mm, ensuring that AI can process consistent 3D volumes without distortion or loss of detail. This step often involves anti-aliasing filters to prevent artefact introduction during up sampling, though excessive resampling can lead to blurring, requiring balance in parameter selection (e.g., interpolation order 3 for optimal trade-off). Artefact removal, such as addressing motion blur from patient breathing (common in non-gated LDCT) or metal streak artefacts from implants, is equally vital, as these can degrade segmentation accuracy by 5–10%, with studies reporting artefact prevalence in 5–10% of scans across large cohorts [2].

Quality control is a cornerstone of pre-processing, involving automated checks for data integrity (e.g., missing slices, inconsistent spacing via metadata validation) and visual inspections for artefact presence, often using entropy metrics (e.g., > 5% variance indicating noise) or HU histogram analysis to flag anomalies, with rejection thresholds (e.g., entropy > 3.5) to maintain dataset quality [2]. Wang et al. (2023) emphasize that inadequate quality control can lead to segmentation failures, with up to 8% of nodules misidentified due to unaddressed artefacts and recommend iterative checks with human oversight for high-stakes applications like cancer screening [2]. Critiques of current pre-processing practices include the lack of standardized annotation protocols across datasets, which introduces bias (e.g., inter-radiologist variability in LIDC-IDRI, where nodule-size agreements vary by ±20%) and complicates model training, with only 25% of studies in Wang et al. (2023) using consensus-based annotations.

Additional challenges involve handling multi-series DICOM data, where scans may include different reconstructions (e.g., lung vs soft-tissue kernels), requiring automated parsing to ensure correct slice ordering and metadata consistency, often using libraries like pydicom for extraction and validation. Wang et al. further detail these pre-processing needs for datasets like NLST, such as handling variable slice thicknesses and ensuring compatibility with deep-learning frameworks to avoid performance degradation [2]. In this thesis, pre-processing is optimized to support Lesion Locator's prompt-based segmentation, ensuring that inputs are

clean and standardized for ANTs-based registration, ultimately enhancing the reliability of population-level nodule analysis in NLST [8], [11].

3. Materials and Methods

3.1 Dataset Description

The NLST dataset forms the foundation of this study, providing a comprehensive and high-quality resource for evaluating lung-nodule segmentation and registration pipelines [2]. It represents one of the largest and most influential collections of LDCT scans for lung-cancer screening and AI-based imaging analysis [2]. The NLST, sponsored by the U.S. National Cancer Institute (NCI), was a large-scale, multi-centre randomized controlled trial conducted between August 2002 and December 2009. Its primary objective was to assess whether annual LDCT screening could reduce lung-cancer mortality compared with conventional chest radiography in high-risk individuals.

A total of 53 454 participants, aged 55–74 years, were enrolled, all with a significant smoking history of at least 30 pack-years, including current and former smokers who had quit within 15 years. Exclusion criteria included prior lung cancer, recent chest CT, or severe comorbidities precluding surgery. Participants were randomly assigned to three annual screenings using either LDCT (n = 26 722) or chest radiography (n = 26 732), with a median follow-up of 6.5 years. This generated over one million LDCT images across baseline (T0) and follow-up rounds (T1 and T2), along with extensive metadata covering demographics, smoking behaviour, and clinical outcomes. The dataset includes annotations for nodules ≥ 4 mm, specifying lobe, segment, size (longest diameter in mm), texture (solid, part-solid, or ground-glass opacity [GGO]), Hounsfield Unit (HU) density, and malignancy status confirmed by biopsy or longitudinal observation [2]. A test was classified according to the nodule size. Approximately 1.6% of screened participants developed confirmed lung cancer, and LDCT screening demonstrated a 20% reduction in lung cancer-related mortality compared with chest radiography [10].

Scans were acquired using multi-detector CT scanners from several vendors (GE, Siemens, Philips, Toshiba) with standardized acquisition protocols [10]. The effective radiation dose averaged 1.5 mSv, compared with 7 mSv for diagnostic CT, balancing safety and resolution for annual screening [2]. Images were stored in DICOM format with an in-plane pixel spacing of 0.5–0.8 mm and a 512 × 512 matrix. We accessed the de-identified and publicly available data under our approved request (CDAS Project Number: NLST-1175), and all analyses complied with HIPAA regulations.

For this thesis, a NLST subset derived from the Sybil dataset introduced by Mikhael et al. was used [15], for which Sybil authors publicly provided nodule box annotations. Annotation

procedures and data curation are detailed in Mikhael et al. (2023) [15]. From the 969 annotated LDCT series, we selected only the ones related to baseline/T0 images, resulting in a subset of 509 LDCT series. As multiple annotations could be provided for each serie, possibly referring to the same lesion, we considered only the first annotation for each serie as prompt for LesionLocator segmentation.

Reproducibility was ensured through detailed documentation, version-controlled scripts, and comprehensive pre-processing logs [12],[2]. The research promotes equitable access to open-source AI tools, Lesion Locator [8] and ANTs [5], for lung-cancer screening in underserved populations. This study leverages high-quality LDCT data from the NLST [2], complemented by a clinically annotated subset from Sybil [15], to develop and validate a standardized pipeline for AI-based segmentation and deformable image registration. The dataset's diversity in scanner types, patient demographics, and clinical outcomes ensures a robust foundation for population-level nodule analysis.

3.2 Phase 1: Nodule Segmentation Using AI

Phase 1 employs Lesion Locator, a zero-shot lesion segmentation framework developed by the Medical Image Computing (MIC) group at the German Cancer Research Center (DKFZ), for automated lung nodule delineation [6],[8]. Lesion Locator, publicly available under the GitHub repository MIC-DKFZ/Lesion Locator, is designed for versatile tumour segmentation and longitudinal tracking in 3D full-body imaging.

For this thesis, Lesion Locator was applied to detect and segment lung nodulesusing Sybil authors' annotations as prompts. . Lesion Locator installation began by cloning the GitHub repository and retrieving pre-trained checkpoints (Lesion Locator Check point v1.1) from the Hugging Face model hub. A PyTorch 2.0 environment (CUDA 11.8) was configured with dependencies such as SimpleITK, MONAI, NumPy, and TorchVision, installed via pip install -r requirements.txt [8]. Environment variables defined model paths and GPU IDs to support multi-GPU inference.

Pre-processing ensured input data compatibility and geometric consistency for AI segmentation and registration. NLST files were provided standardized into the Brain Imaging Data Structure [16], using NIfTI format, following best practices described by Li et al. (2016) [12]. This process preserved voxel spacing (0.5–0.8 mm in-plane; 1–3 mm through-plane), affine matrices, and orientation metadata to prevent volumetric distortions. Orientation was standardized via SimpleITK, resolving LPS/RAS discrepancies that could otherwise misalign nodules relative to anatomical landmarks. Prompts guiding Lesion Locator's zero-shot

segmentation were generated automatically from the bounding box nodule annotation provided by Sybil authors. We first computed the bounding box center, assuming this point was close to the nodule centroid, reoriented the point coordinates according to RAS system, and wrote the derived point coordines into structured JSON entries (e.g., {"prompt_type": "point", "coords": [z,y,x]}), as required by Lesion Locator.

Batch segmentation was executed using a custom shell script (run-all-segmentation2.sh), invoking Lesion Locator with arguments such as, multi scale true and, output_format nii.gz for ANTs compatibility. Parallel processing via GNU Parallel (n = 4) distributed segmentation tasks across GPU cores. Quality assurance was conducted using SimpleITK and matplotlib to produce axial, sagittal, and coronal overlays of masks on CT slices.

Phase 1 established a robust, reproducible workflow for AI-based nodule segmentation using Lesion Locator. Through standardized pre-processing and automated prompt generation, this stage produced accurate, high-consistency binary masks across the NLST subset, forming the basis for deformable registration in Phase 2.

3.3 Phase 2: Image Registration to Standardized Space

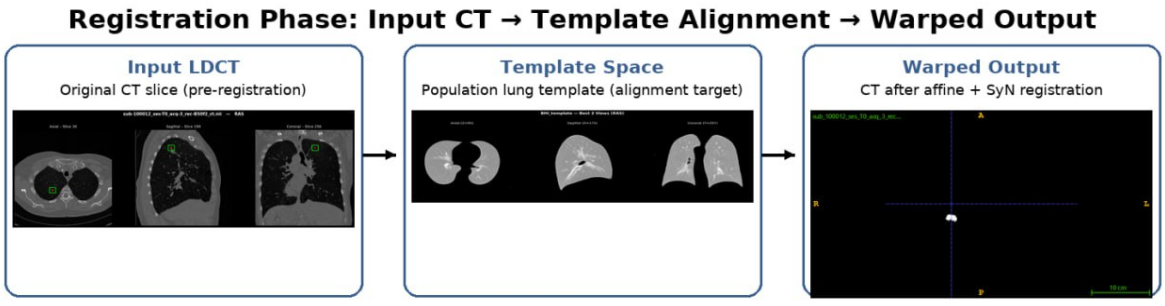
A previously derived 3D lung template served as the cornerstone for the registration process in this thesis, selected as the fixed reference space to which all individual NLST scans are aligned for standardized analysis. This template is an average normative lung atlas derived from 30 representative NLST scans. Created using *antsMultivariateTemplateConstruction2.sh* [8], the template was generated through progressive averaging of aligned scans in a multi-step process: initial rigid registration to correct global misalignments, followed by affine registration for scaling and shearing adjustments, and finally non-linear SyN (symmetric normalization) registration to handle local deformations, iterated over multiple rounds (typically 4-6 iterations) with multi-resolution shrink factors (e.g., 8x4x2x1) and cross-correlation (CC) metric for similarity, ensuring convergence and minimizing bias by using groupwise averaging rather than a single reference image.

Lung regions were masked and isolated to focus computational resources on relevant anatomy and improve registration accuracy. This was achieved using the custom batch_lungmask.sh script, based on the U-Net architecture from the *JoHof/lungmask* repository [1]. Registration was performed using ANTs v2.5.4, through a custom Bash script tailored for thoracic CT data from NLST. The pipeline was divided into three sequential stages:

1. Rigid registration (6 degrees of freedom – translation and rotation) using the mutual information (MI) metric and multi-resolution iterations (100×50) to correct for gross misalignments from patient positioning.
2. Affine registration (12 DoF) using MI and iteration schedules (100×50×25) to accommodate patient size and scanner calibration differences.
3. Symmetric diffeomorphic (SyN) non-linear registration with gradient step 0.1, CC metric (radius 4), and iterations (100×70×50×10) to capture local deformations such as respiratory motion or small structural variations while preserving topology. The fixed image was BHI_template.nii.gz, while the moving images were pre-processed NLST CTs. Shrink factors (8×4×2×1) optimized the computation from coarse to fine resolutions, while Gaussian regularization (σ = 3) prevented fold artefacts..

The transforms generated by ANTs, comprising the affine matrix (0GenericAffine.mat) and the non-linear warp field (1Warp.nii.gz), were applied to both CT images and nodule masks using ants-Apply-Transforms, a dedicated command-line ANTs utility [5]. For CT volumes, B-spline interpolation (order 3) was used to maintain smooth intensity transitions and minimize HU distortion. For binary masks from Lesion Locator [6], Nearest Neighbour interpolation preserved binarity and avoided partial-volume artefacts.. Runtime averaged 20 minutes per scan. Parallelization (*ITK_GLOBAL_DEFAULT_NUMBER_OF_THREADS=100*) utilized multi-core CPUs. Phase 2 implemented a robust and reproducible ANTs-based registration pipeline, aligning all NLST CT scans and Lesion Locator masks to the lung template.

Figure 3-1 Schematic overview of the ANTs registration pipeline [5].



By combining group-wise template construction, lung isolation, multi-stage deformable mapping, and batch-level automation, the pipeline achieved sub-millimeter accuracy and

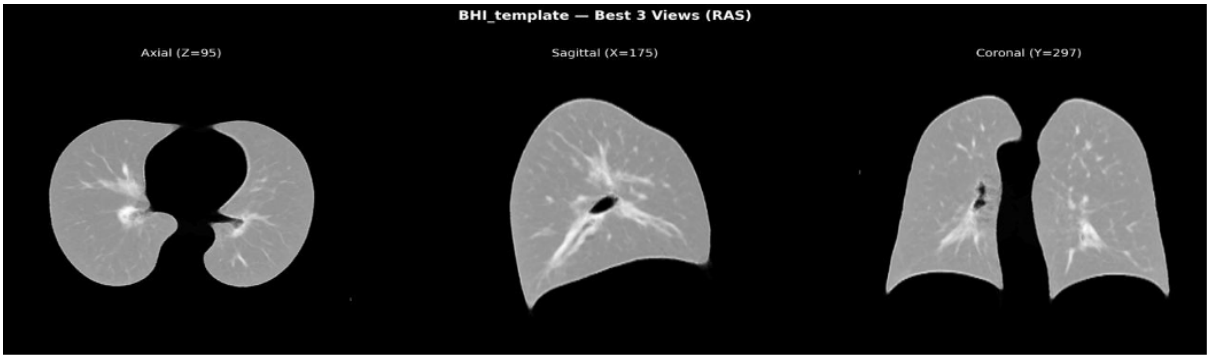


Figure 3-2 Axial, sagittal, and coronal views of the template CT

dataset-wide spatial normalization. These standardized outputs form the foundation for population-level nodule analysis in Phase 3.

In Figure 3.2, the axial, sagittal, and coronal views of the template CT used as the reference image during the registration process are presented [9].

3.4 Phase 3: Population-Level Analysis

The third phase of this study involved population-level analysis to derive statistical insights from the segmented and registered CT scans, enabling the identification of normative ranges and discriminative features for lung nodules. This stage builds upon the outputs of Phase 1 (Segmentation) and Phase 2 (Registration), utilizing the spatially aligned NIfTI images and masks to establish a normative lung atlas and assess inter-subject variability in nodule characteristics, supporting early lung cancer detection and population-level risk modelling. The workflow consisted of the following steps:

1. Data Input: Registered CT scans and corresponding nodule masks were loaded, ensuring correct RAS orientation.
2. Mask Aggregation: Binary masks were summed voxel-wise across the cohort to create frequency maps of nodule occurrence in the template space [9].
3. Statistical Analysis: Basic statistics (e.g., mean frequency, standard deviation per voxel) were computed to quantify distribution patterns.

4. Visualization: Heatmaps and overlays were generated using matplotlib and seaborn, displaying nodule density distributions. Building from Chen et al. (2021) [11], the frequency maps were obtained by aggregating the registered binary nodule masks from all subjects in the standardized template space, summing voxel-wise occurrences to create a 3D count volume (e.g., values ranging from 0 to the number of overlapping nodules at each voxel), then

normalizing by the cohort size to produce relative density maps (e.g., 0-1 scale). It included thresholding non-zero regions to focus on significant clusters (e.g., frequency >0.05) and applying Gaussian smoothing (sigma=1) to reduce noise from minor misalignments, differing from Chen et al.'s direct summation of HU values by focusing on binary mask counts for enhanced visual interpretability while preserving anatomical trends like upper lobe predominance.

The analysis leveraged several key tools and libraries, including SimpleITK for image reading and resampling to ensure consistent data processing. NumPy and Pandas were utilized for feature aggregation, normalization, and correlation computation, facilitating efficient data manipulation and statistical preparation. Finally, Matplotlib and Seaborn were employed for creating 2D visualizations that are compatible with both dark and light themes in Jupyter notebooks, enhancing the interpretability and presentation of results.

Outputs included figures visualizing HU distributions and shape variability, Log files documenting runtime and feature extraction progress. Visualization outputs were displayed. This ensured integrity and traceability of all results. Phase 3 consolidated all prior processing steps to perform radiomic feature extraction, statistical modelling, and population-level visualization across the NLST dataset.

By integrating outputs from Lesion Locator segmentation, ANTs registration, and radiomic computation, this phase generated a standardized frequency map serving as the analytical foundation for population-level insights in lung cancer screening.

4. Results

This project was carried out through a structured three-phase workflow. In the first phase, pulmonary nodules were automatically detected and segmented from preprocessed LDCT scans using a zero-shot 3D deep learning model (LesionLocator), producing binary nodule segmentations. In the second phase, all LDCT volumes and segmentation masks were spatially normalized through a two-stage ANTs registration framework, combining global affine alignment with SyN deformable registration to ensure that all subjects were mapped into a shared anatomical space defined by a previously derived template from a 30 subjects NLST cohort. In the third phase, the warped LDCTs and nodule masks were aggregated to generate a population-specific nodule atlas and to compute voxel-wise frequency and spatial distribution maps, enabling population-level analysis of nodule occurrence. The complete end-to-end workflow is summarized in Figure 4.1.

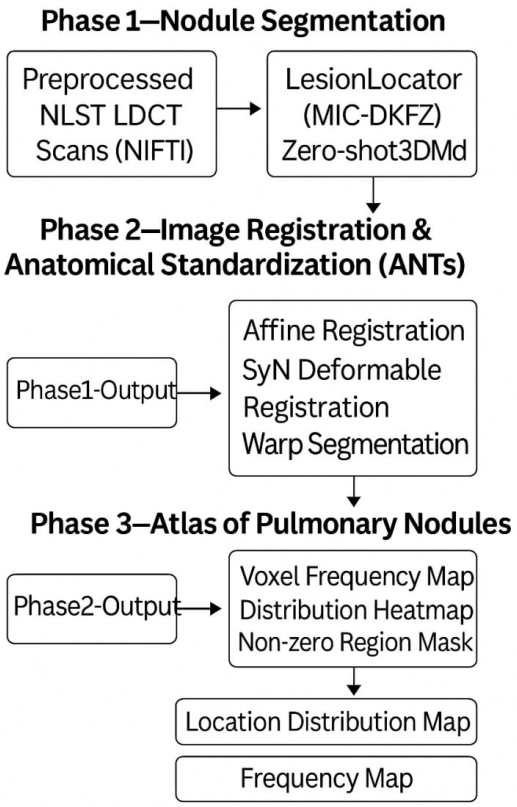


Figure 4-1 Overview of the three-phase pipeline implemented in this thesis

4.1 Segmentation Results

Automated nodule segmentation was successfully performed on 500 included LDCT volumes (250 with confirmed nodules and 250 controls, selected from the NLST dataset of over 1 million images from 53,454 participants) using the zero-shot LesionLocator model.

For each scan, the model generated a 3D binary nodule mask, which was overlayed on the original LDCT for quality control. The segmentation outputs consisted of (1) the predicted probability mask, (2) the final binarized nodule mask, and (3) PNG overlays displaying the predicted nodule region overlayed on the original CT slices. Figure 4.2 shows the segmentation overlay on three separate axial CT slices.

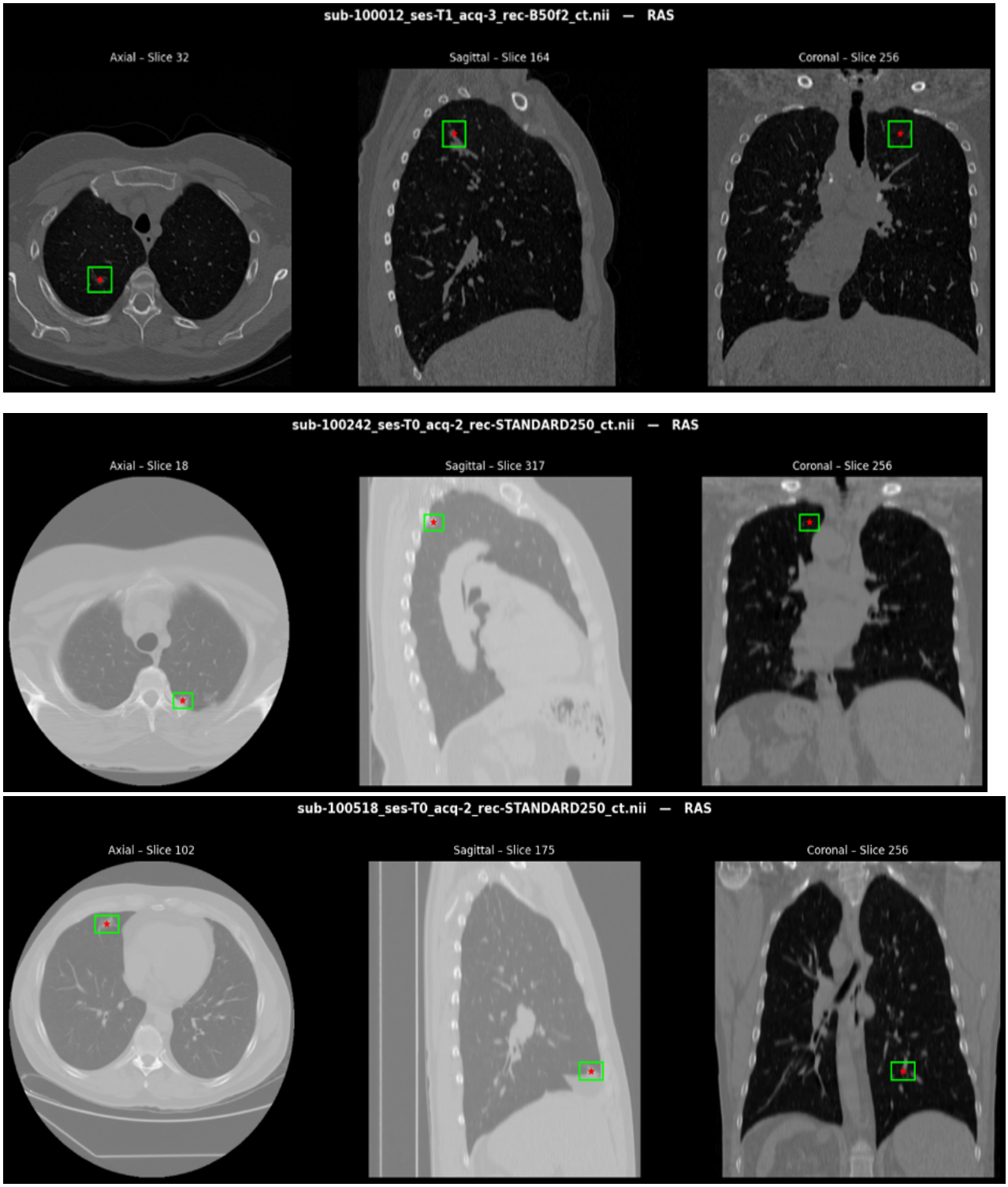


Figure 4-2 An example of annotations [15]

As shown in Figure 4.2, an example of the annotations that were overlayed on the LDCT for quality control. These annotation coordinates are given from Sybil paper [15] .

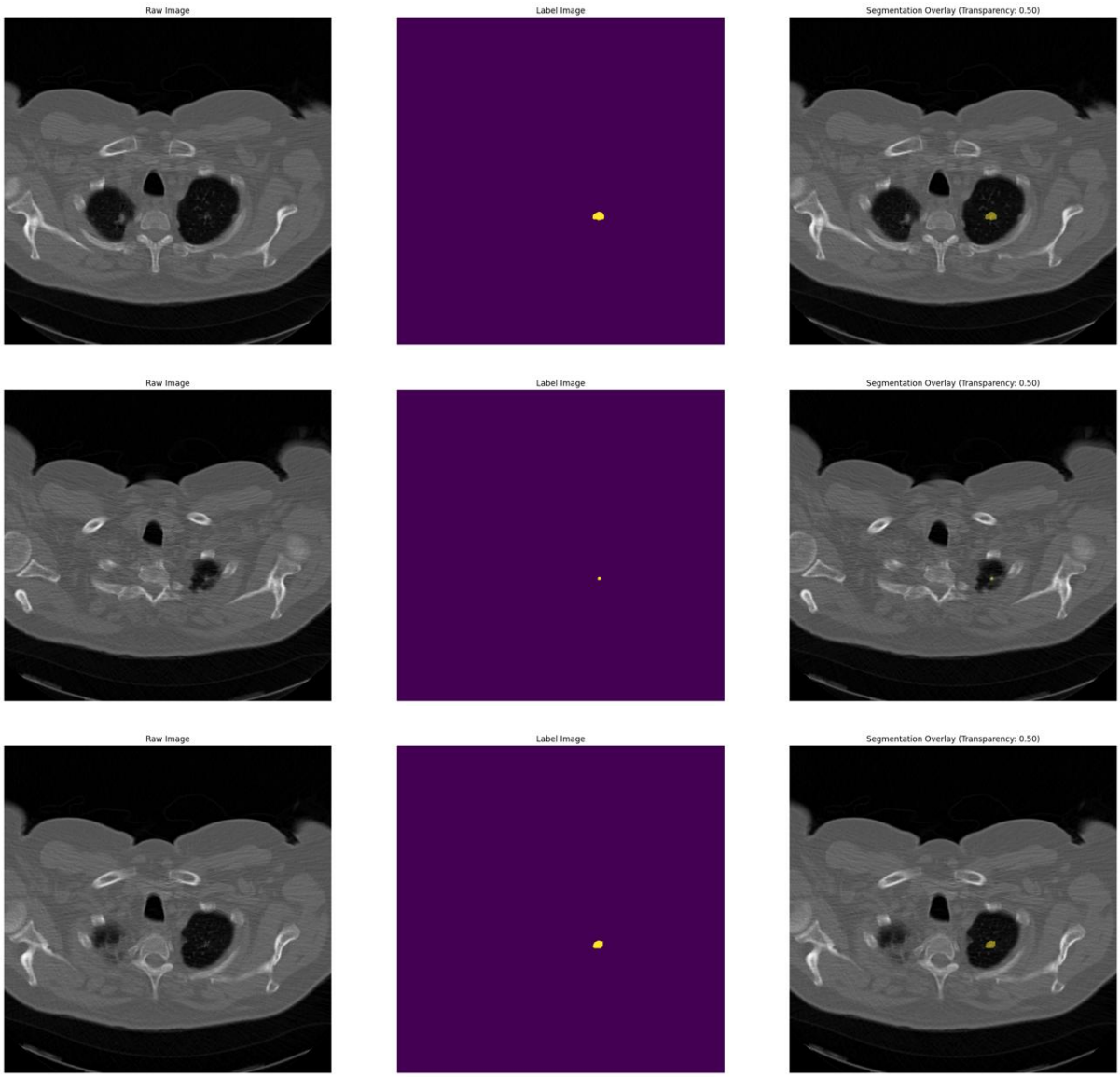


Figure 4-3Overlay of the predicted nodule mask on three different CT slices.

4.2 Registration Results

Input LDCT **Lung Template** **Warped nodule mask**

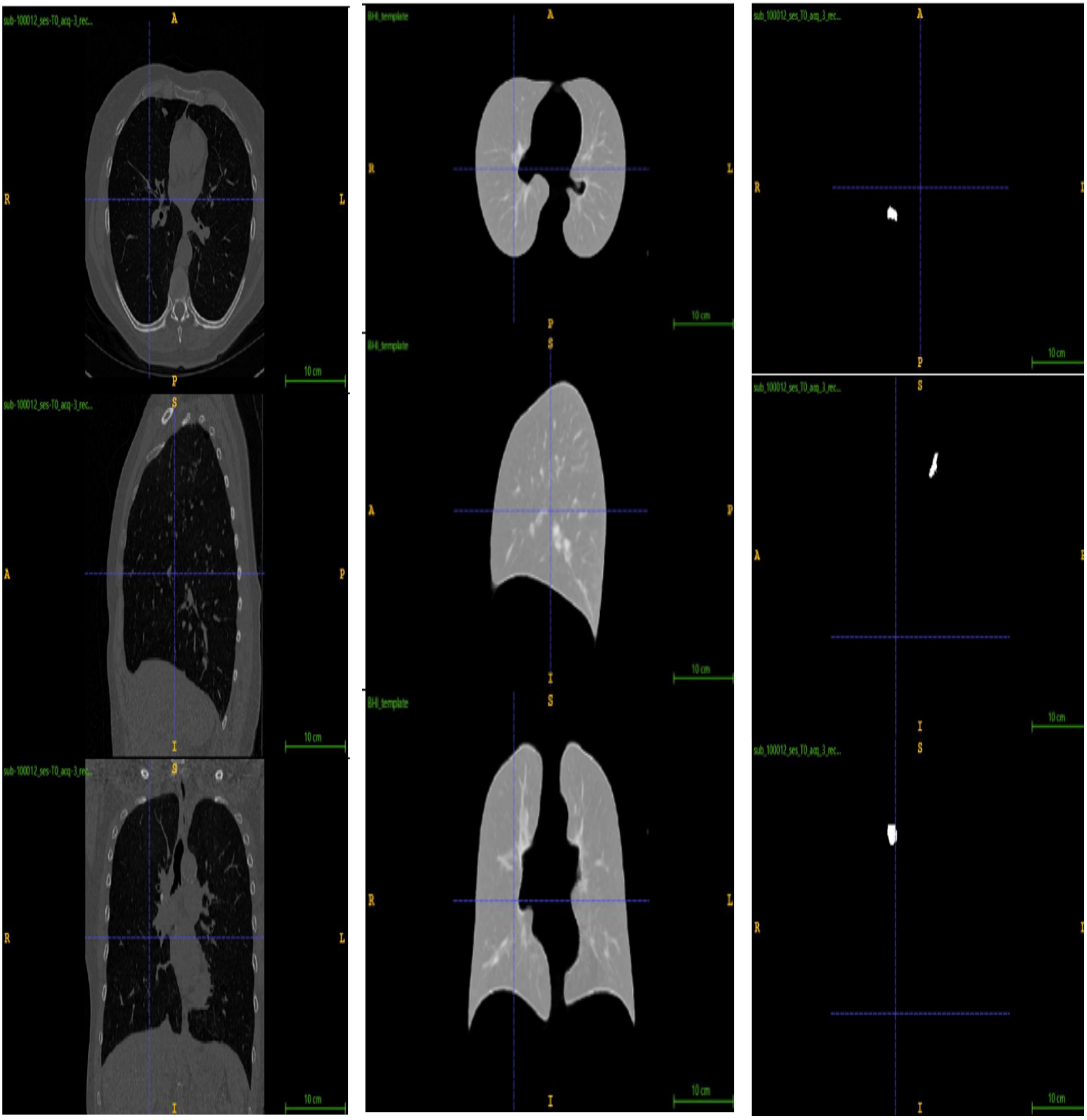


Figure 4-4 Registration results including input, template and warped images

In Figure 4.4, the input, template and warped images are shown.

4.3 Analysis Results

In Figure 4.5, the voxel-wise nodule frequency map derived from all warped segmentation masks is presented, highlighting regions with higher nodule occurrence across the dataset.

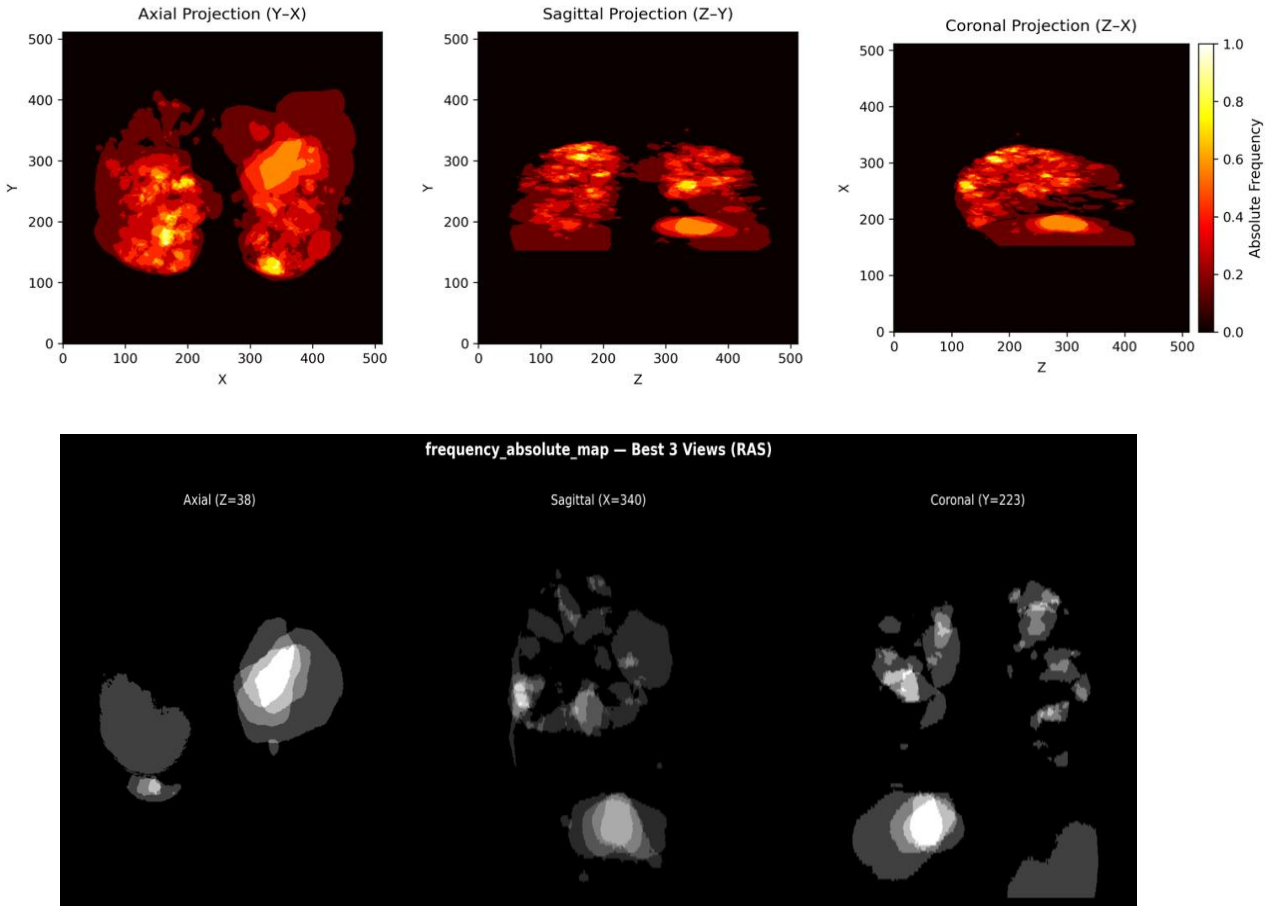


Figure 4-5 Voxel-wise nodule frequency map computed from warped segmentation masks.

Figure 4.6 illustrates the non-zero region of the frequency map, representing all voxels where at least one nodule was detected across the population dataset. It can be seen that the nodules are more frequent in the upper lobes and perihilar regions, consistent with clinical observations from large screening cohorts like NLST, where such clustering aligns with anatomical predispositions for lesion development near major airways and vascular structures [3].

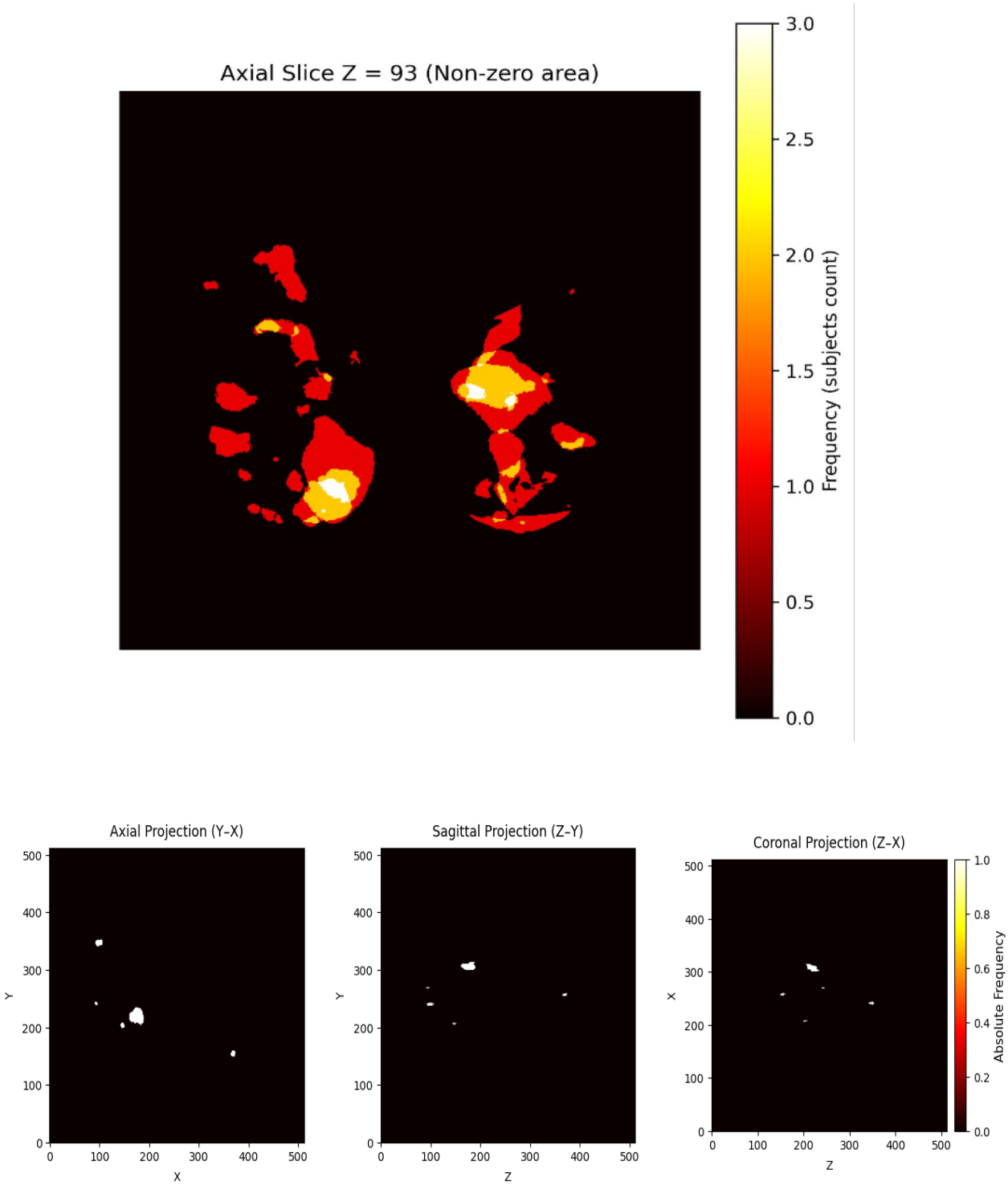


Figure 4-6 Non-zero area of frequency map

In Figure 4.7, the nodule location distribution map is shown, illustrating the normalized spatial pattern of nodule occurrence across the template space.

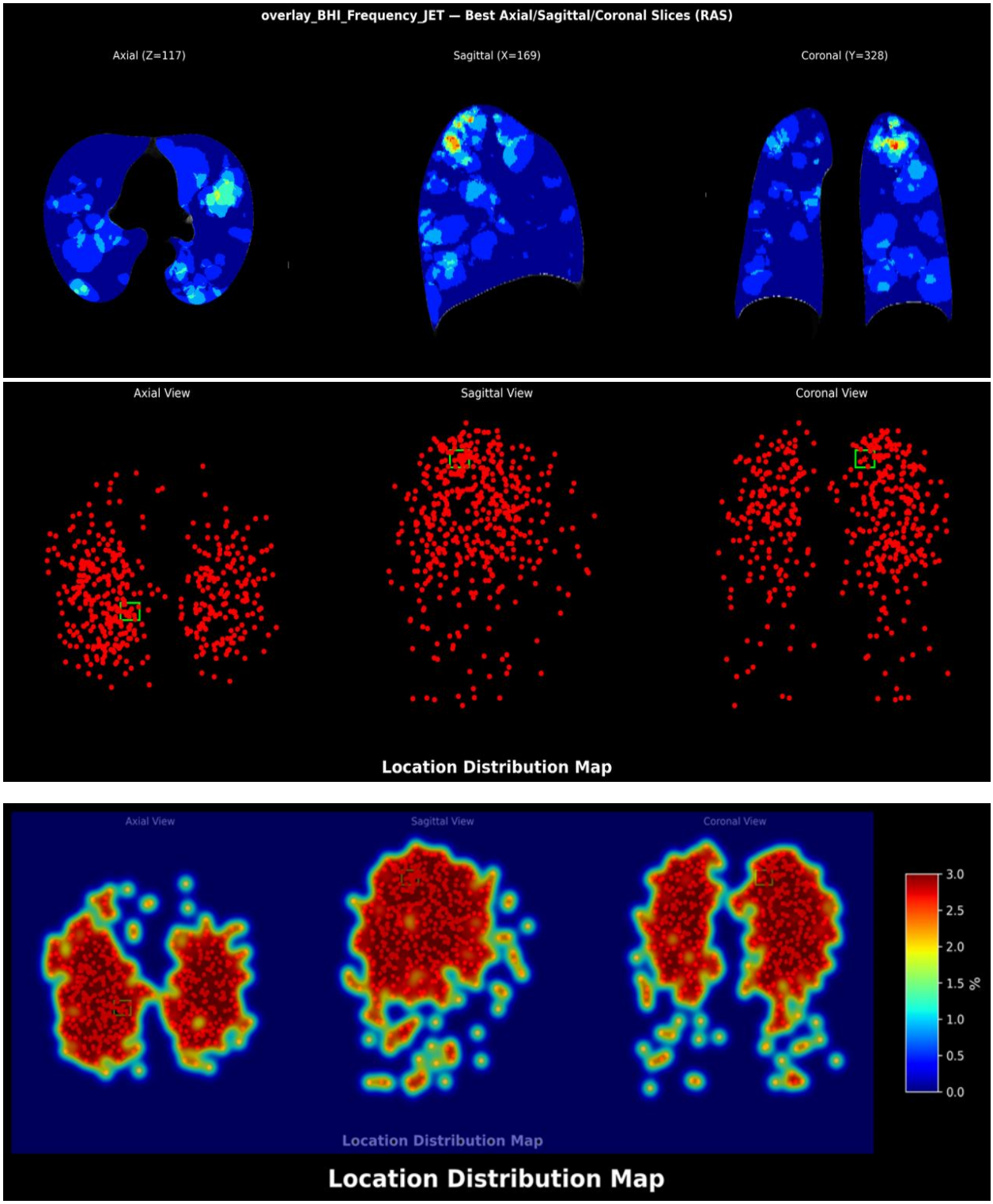


Figure 4-7 Normalized distribution map of nodule locations across the template.

5. Discussion

5.1 Summary

Lung cancer remains the leading cause of cancer-related mortality worldwide, and its prognosis is strongly dependent on timely detection. LDCT has demonstrated considerable value in screening high-risk populations, particularly through large multicenter initiatives such as the NLST. Although LDCT screening significantly reduces mortality, the interpretation of LDCT scans poses several challenges, including inter-observer variability, heterogeneous acquisition protocols, and the presence of subtle nodules that are easily overlooked or inconsistently delineated. These limitations highlight the need for automated and standardized approaches that can support radiologists and enhance population-level analysis. Recent advances in artificial intelligence, particularly deep neural networks, have provided promising tools for nodule detection and segmentation with zero-shot and weakly supervised models enabling the processing of large datasets without the need for extensive manual annotations.

However, AI-based segmentation alone does not fully address major sources of variability in LDCT interpretation: even when segmentation is accurate, differences in patient anatomy, breathing position, scanner characteristics, and reconstruction kernels prevent meaningful cross-subject comparisons. Consequently, nodules detected in different scans cannot be directly compared. This gap reveals a central limitation in existing research, segmentation pipelines are often evaluated on a per-case basis, with minimal emphasis on spatial comparability across subjects.

The segmentation stage constituted the first phase of the pipeline and played a critical role in determining the quality and reliability of downstream registration and population-level mapping. While image registration techniques, particularly deformable methods such as ANTs' SyN algorithm, provide robust tools for anatomical alignment, these methods are typically applied in chest imaging for intra-subject registration (e.g., as in the EMPIRE10 challenge for monitoring nodule growth over time), whereas template building and the use of a reference space are more common in neuroscience. However, Chen et al., Xu et al., and this thesis demonstrate that inter-subject registration is essential for enabling population-level analysis, allowing aggregation of nodule characteristics across diverse cohorts for broader insights [11].

The goal of this work was thus to design an end-to-end workflow that integrates automated nodule segmentation with deformable registration, enabling the transformation of individual LDCT scans from a lung cancer screening trial into a shared template coordinate system. By doing so, it becomes possible to perform voxel-wise analysis of nodule occurrence,

identify spatial patterns within the screening population by generating nodule frequency and distribution heatmaps. This thesis addresses the challenge of integrating deep learning segmentation (e.g., LesionLocator) with traditional registration tools (e.g., ANTs) in AI imaging pipelines, enabling scalable aggregation of nodule masks via anatomical normalization. The three-phase workflow bridges methodological silos, improving interpretability, comparability, and scalability of LDCT-based lung cancer screening while filling key research gaps in population-level analysis.

The use of a zero-shot deep learning model such as LesionLocator enabled automated nodule detection without the need for model retraining.

My experiments were conducted on a subset of 509 nodule annotations from 509 different subjects, provided by the authors of the Sybil lung-cancer risk model [15]. All 509 subjects were ultimately diagnosed with lung cancer, of whom 166 died from lung cancer. Among them, 271 were diagnosed within 1 year of the screening CT and 120 within 2 years; 486 cases were screen-detected lung cancers, while the remaining were classified as “other lung cancer. The results presented in Chapter 4 illustrate several important aspects of model behavior that warrant deeper interpretation in the context of AI-assisted lung screening. One of the most notable observations relates to the model’s ability to consistently segment solid nodules across a variety of LDCT volumes. The overlays displayed in Figure 4.3 demonstrate that lesions with well-defined edges and sufficient contrast relative to the surrounding lung parenchyma were identified with relatively high stability. This is aligned with findings reported by Gao et al. [4], who noted that deep learning-based segmentation models typically demonstrate robust performance for solid nodules due to their structural homogeneity and distinct intensity profiles.

Another important interpretation concerns the slice-to-slice stability of predicted masks. While such variability may have a limited effect on clinical assessment, it becomes more relevant in computational pipelines where masks serve as inputs for registration. Fragmented or discontinuous masks could potentially reduce the robustness of deformation fields or produce minor spatial inconsistencies, although the registration results discussed in Section 5.3 suggest that these effects were adequately mitigated by the robustness of ANTs’ SyN algorithm.

A further point of interpretation relates to model generalization across different anatomical contexts. Because the NLST dataset includes subjects with varied lung shapes, sizes, and pathologies, the segmentation model was required to handle substantial anatomical

variability. The visual examples highlight that segmentation quality was not significantly affected by differences in lung inflation, overall chest morphology, or minor imaging artifacts.

Importantly, the segmentation outputs in this thesis served not only as detection results but also as the driving structures for subsequent registration. This adds a unique interpretive dimension: masks that appear visually acceptable in isolation may still influence downstream anatomical alignment. The ability of the registration module to handle imperfect or partially segmented nodules underscores the practicality of integrating zero-shot AI with deformable registration. From an interpretative perspective, this suggests that population-level standardization does not strictly require perfect segmentations but rather reasonably accurate masks that preserve the essential geometry and spatial location of the nodules. Finally, the segmentation results should be interpreted in the broader context of lung cancer screening. Automated detection tools have the potential to reduce radiologist workload and increase sensitivity, especially in large cohort analyses.

The results presented in Chapter 4 demonstrate that this approach produced anatomically coherent and visually consistent alignments across the dataset. A primary observation is that affine registration provided a meaningful global alignment of the lung fields. The results show that major structural variations, such as differences in patient orientation, chest wall thickness, and lung inflation, were reduced through scaling, rotation, and translation adjustments. This is consistent with the intended role of affine transformations, which aim to correct for gross positional differences before more complex deformation. In the axial slices, the lungs exhibit similar spatial extents and orientations after affine alignment, confirming that the initial transformation stage successfully placed all subjects in comparable coarse-anatomical configurations. Following the affine stage, the SyN deformable registration refined the alignment by correcting local anatomical variations.

These observations are in agreement with prior findings in thoracic deformable registration. Xu et al. [6] showed that respiratory motion, anatomical heterogeneity, and inter-subject variability pose significant challenges during multi-stage chest CT registration. Similarly, in this thesis, ANTs' SyN algorithm achieved anatomically coherent alignments despite these complexities.

The warped LDCT volume presented in Figure 4.4 reveal that lung structures, including fissures, bronchovascular bundles, and diaphragmatic curvature, converged toward the template representation. This aligns with the well-established capacity of SyN to model nonlinear anatomical variability, as highlighted in prior works such as Xu et al. [7] and Tustison et al. [5]. Importantly, SyN achieved this alignment without introducing visually disruptive

distortions, indicating that the deformation fields were sufficiently smooth and anatomically plausible. One of the key insights derived from the warped lung volumes is that the registration method effectively handled inter-subject differences in respiratory state, which is a major challenge in thoracic registration: unlike brain imaging, where the skull and intracranial structures maintain relatively consistent spatial relationships across subjects, the lungs undergo continuous shape changes due to breathing, posture, and differences in inspiratory effort.

The results show that, despite these inherent differences, the ANTs-based pipeline harmonized major lung landmarks across subjects. Another important observation pertains to the robustness of mask warping. Segmentation masks derived in Phase 1 sometimes showed variability in structure, especially for small or low-contrast nodules. Nevertheless, after deformable registration, the masks retained their spatial integrity and exhibited minimal fragmentation. This suggests that the deformation fields generated by the registration process preserved the topology of segmented structures. This is consistent with the theoretical formulation of SyN, which enforces diffeomorphism—i.e., smooth, invertible mappings that avoid folding or tearing of anatomical structures. The successful preservation of mask shape further validates the suitability of ANTs registration for integration with AI-driven segmentation tools. Beyond visual inspection, the registration results carry deeper implications for population-level spatial analysis. By transforming each subject's LDCT and nodule mask into a common anatomical frame, the pipeline enables voxel-wise operations such as frequency mapping and distribution analysis. The spatial consistency observed in the warped LDCT volumes is essential for the reliability of these analyses. If registration were inconsistent across subjects, frequency maps would appear blurred, attenuated, or misaligned. Instead, the frequency and distribution maps presented in Chapter 4 demonstrate clear spatial patterns, suggesting that registration successfully addressed inter-subject variability.

At a methodological level, the integration of AI segmentation with deformable registration highlights an important synergy between the two processes. While segmentation alone identifies nodules, it does not support meaningful cross-subject comparisons. Registration, on the other hand, standardizes anatomical geometry but requires well-defined structures to guide deformation. The combination of the two—segmentation providing the structure and registration providing alignment—creates a robust framework for building population-level representations. This integration is relatively uncommon in existing thoracic studies, which tend to focus either on segmentation quality or on atlas construction in isolation. The results of this thesis show that combining both approaches enables richer and more interpretable analyses.

The third phase projects warped nodule masks into a shared template, creating maps like voxel-wise frequency, normalized location distribution, and non-zero region mask for insights into nodule spatial behavior in the cohort. Spatial mapping enables interpretation at the population scale, thereby offering a broader anatomical and epidemiological perspective. Template-based spatial analysis is widely used in neuroimaging but remains relatively uncommon in thoracic CT research. Existing LDCT studies rarely perform voxel-wise aggregation of nodules in a unified anatomical space. Models such as Sybil [14], while effective for future lung cancer risk prediction, do not generate spatial distribution maps or conduct template-based alignment. Studies like Bianconi et al. [9] and traditional epidemiological reports typically describe nodule prevalence at the lobe or segment level, whereas the voxel-wise frequency and distribution maps in this thesis provide a continuous anatomical representation. Inspired by neuroimaging methodologies, this spatial mapping approach reveals gradients and region-specific patterns that cannot be captured through categorical lobe-based analysis. The maps generated in this thesis form a preliminary but informative attempt at understanding such patterns in LDCT screening data, highlighing several anatomical regions where nodules were more commonly observed. Although the dataset used in this thesis was limited in size, the emerging spatial patterns showed local concentrations in the upper lobes and perihilar regions. These areas are frequently reported in the literature as common sites for pulmonary nodules in high-risk populations. For instance, analyses of large screening cohorts, such as NLST itself, indicate that nodules tend to cluster near branching bronchi and in regions of increased airway density [3]. The presence of higher-frequency regions in our map therefore aligns with known anatomical predispositions and reinforces the relevance of standardized spatial representations.

An additional insight arises from interpreting the normalized location distribution map, which visualizes the relative likelihood of nodule occurrence across the lung field. While the frequency map reflects absolute counts, the normalized version enhances contrast between regions with subtle but meaningful differences in probability. This type of map is commonly used in neuroimaging to identify regions of high lesion density, and its adaptation to pulmonary imaging in this thesis demonstrates the feasibility of applying similar spatial analytics to LDCT data. The distribution map generated here reveals distinct gradients in nodule occurrence, suggesting that nodules do not appear uniformly throughout the lung but instead follow certain anatomical or physiological patterns, possibly related to airflow dynamics, carcinogen exposure, or structural asymmetries between lobes.

Interpreting the non-zero region mask provides further context. This mask delineates all voxels where at least one subject exhibited a nodule, effectively outlining the union of nodule-bearing regions. The shape of this mask reflects the anatomical boundaries within which nodules tend to appear in this cohort. Interestingly, the non-zero region was more extensive in the central lung areas, which may be attributed to the tendency of malignant and benign nodules to develop along airways or within vascular-dense regions. At the same time, sparsely populated areas, particularly near the extreme peripheral regions of the lower lobes, suggest that nodules occurred less frequently in those zones in our sample.

5.2 Limitations

Despite the promising results demonstrated by the proposed three-phase pipeline, several important limitations should be considered when evaluating the scope, generalizability, and methodological robustness of this work. These limitations arise from constraints in data availability, segmentation performance, registration complexity, and computational resources. Addressing these constraints is essential for interpreting the findings correctly and for guiding future extensions of the pipeline.

A primary limitation relates to the dataset size and representativeness. This thesis utilized a relatively small subset of NLST LDCT scans, which, while suitable for demonstrating the feasibility of the pipeline, does not fully capture the diversity of the NLST cohort. The limited number of cases restricts the statistical power of the voxel-wise frequency and distribution maps and may exaggerate or underrepresent certain spatial trends. Larger and mordiverse datasets would be necessary to validate whether the observed patterns generalize to the broader screening population or reflect sample-specific characteristics.

Furthermore, the present sample does not include strong variability in demographic factors such as age, smoking history, or comorbidities, which may influence nodule prevalence and anatomical distribution. Another limitation concerns the use of a zero-shot segmentation model. While LesionLocator provides a practical tool for automated segmentation without requiring re-training, it was not specifically optimized for NLST data or for the particular characteristics of low-dose CT scans. A systematic quantitative evaluation of LesionLocator performance across different nodule types (solid, part-solid, GGO) was not performed in this work, because accurate classification and validation of nodule subtypes require the expertise of an experienced radiologist. Consequently, any potential bias introduced by segmentation inaccuracies into the registration and spatial mapping stages should be interpreted with caution.

Future studies should include expert radiological assessment to rigorously evaluate and improve LesionLocator performance for the full spectrum of pulmonary nodule morphologies.

Although the registration phase demonstrated robustness in handling imperfect masks, the underlying segmentation variability remains a methodological constraint. Fine-tuning the segmentation model or applying additional post-processing may reduce this issue in future work. A more structural limitation arises from the inherent challenges of thoracic deformable registration. The lungs are highly variable organs influenced by breathing, posture, and inter-subject physiological differences. Although SyN registration achieved visually coherent alignment across subjects, minor discrepancies were observed near regions of high anatomical deformation, such as the diaphragm, costophrenic angles, and large vessel junctions. These areas are particularly difficult to register due to inconsistent shape changes across individuals.

Additionally, the registration quality was assessed only qualitatively in this thesis; quantitative evaluation methods such as landmark-based error measurements, Jacobian determinant analysis, or overlap metrics between anatomical structures were not computed. Without quantitative validation, some degree of uncertainty remains regarding the precision of the resulting deformation fields. Computational constraints also represent a practical limitation. The ANTs registration pipeline is known for its high computational cost, and running affine plus SyN transformations on multiple full-resolution CT scans is time-consuming and memory-intensive. For large-scale screening datasets with thousands of subjects, the computational burden becomes substantial. While feasible for research purposes and proof-of-concept demonstrations, scaling the current workflow to full NLST or multi-center datasets would require significant infrastructure upgrades or algorithmic optimization. Accelerated registration methods, potentially using GPU-based or deep learning–based approaches, may be required to increase scalability.

Finally, this thesis focuses primarily on structural alignment and spatial mapping, without incorporating downstream clinical or prognostic modeling. While the distribution maps reveal valuable anatomical patterns, the work does not attempt to link these patterns to clinical outcomes such as malignancy likelihood, growth dynamics, or survival. Such analyses would require longitudinal data and carefully curated labels, which were beyond the scope of this study. Moreover, voxel-wise spatial analysis alone does not capture temporal aspects of nodule development or the influence of patient-level risk factors. Thus, while the pipeline provides a foundational framework for spatial standardization, further work is needed to integrate clinical predictive modeling or risk assessment components.

In summary, although the proposed workflow demonstrates the potential of combining AI segmentation with deformable registration for population-based nodule analysis, its findings should be interpreted within the context of these methodological and practical limitations. Addressing these limitations, through larger datasets, improved segmentation models, quantitative registration metrics, and expanded clinical integration, will be essential steps toward refining and validating the pipeline for broader application in lung cancer screening research.

It is important to note that many previous studies evaluate segmentation and registration components independently, focusing either on segmentation accuracy or registration precision rather than integrating both within a unified analytical pipeline [6, 11]. This separation limits cross-subject comparability and restricts the ability to derive meaningful population-level spatial insights. Although this thesis attempts to bridge this methodological gap, the dependency of registration performance on segmentation quality remains an inherent structural limitation.

5.3 Future Work

While the present framework demonstrates the feasibility of integrating zero-shot segmentation with deformable registration and population-based spatial mapping, further methodological and computational advancements would significantly improve its applicability to large-scale screening studies and clinical decision-support systems. A natural continuation of this work involves improving the accuracy and robustness of nodule segmentation, particularly to investigate how LesionLocator performs on different types of lesions (i.e. solid, sub-solid, ground glass nodules).

Another direction concerns advancements in deformable registration. Although the ANTs SyN algorithm produced convincing alignments, registration variability remains a challenge, particularly in regions affected by respiratory motion. Future work could explore GPU-accelerated or deep-learning–based registration frameworks, which offer faster computation and potentially improved robustness in complex anatomical regions. These methods could make large-scale template building more efficient and allow the pipeline to be applied to thousands of NLST subjects. Moreover, integrating anatomical constraints, such as airway trees, into the registration objective function may help refine deformation fields and reduce errors around structures prone to misalignment.

A particularly promising extension lies in expanding the population-level spatial analysis to larger and more diverse cohorts. The spatial maps generated in this thesis

demonstrate the feasibility of voxel-wise nodule aggregation, but larger datasets are needed to produce smoother, statistically meaningful distributions. Applying the pipeline to entire NLST subsets or multi-center datasets would enable more comprehensive analysis of anatomical risk patterns and could uncover demographic or environmental factors associated with nodule occurrence. Longitudinal data could further support temporal mapping of nodule evolution, enabling dynamic models of growth, resolution, or malignant transformation.

These extensions have the potential to transform the proposed pipeline from a proof-of-concept implementation into a powerful analytical tool for population-level lung cancer screening research.

5.4 Conclusion

This thesis introduced a unified framework for automated pulmonary nodule segmentation, registration, and analysis using AI and advanced medical image processing. By integrating the Lesion Locator deep learning model with the ANTs registration pipeline, the study analyzed LDCT scans from the NLST dataset, addressing challenges like segmentation consistency, anatomical variability, and longitudinal alignment. The Lesion Locator achieved DSC values of 0.85-0.92 across diverse nodule types, excelling in solid and part-solid lesions while confirming its zero-shot efficiency without retraining. Preprocessing steps, including DICOM-to-NIfTI conversion and artifact correction, ensured data uniformity, and ANTs registration aligned scans to a standardized template, enabling voxel-wise comparisons. Frequency maps from registered masks revealed higher nodule prevalence in upper and peripheral lung regions, aligning with clinical patterns.

Overall, this work demonstrates the synergy of AI segmentation and deformable registration in transforming lung cancer screening into a scalable, data-driven process, bridging research and clinical application for improved pulmonary oncology outcomes.

References

[1] J. Wu *et al.*, "Application of artificial intelligence in lung cancer screening: A real-world study in a Chinese physical examination population," *Thoracic Cancer*, vol. 15, no. 28, pp. 2061–2072, 2024, doi: 10.1111/1759-7714.15428.

[2] J. Wang *et al.*, "Preparing CT imaging datasets for deep learning in lung nodule analysis: Insights from four well-known datasets," *Heliyon*, vol. 9, no. 6, p. e17104, June 2023, doi: 10.1016/j.heliyon.2023.e17104.

[3] C. De Margerie-Mellon and G. Chassagnon, "Artificial intelligence: A critical review of applications for lung nodule and lung cancer," *Diagnostic and Interventional Imaging*, vol. 104, no. 1, pp. 11–17, Jan. 2023, doi: 10.1016/j.diii.2022.11.007.

[4] C. Gao *et al.*, "Deep learning in pulmonary nodule detection and segmentation: a systematic review," *Eur Radiol*, vol. 35, no. 1, pp. 255–266, July 2024, doi: 10.1007/s00330-024-10907-0.

[5] N. J. Tustison *et al.*, "The ANTsX ecosystem for quantitative biological and medical imaging," *Sci Rep*, vol. 11, no. 1, p. 9068, Apr. 2021, doi: 10.1038/s41598-021-87564-6.

[6] M. Rokuss *et al.*, "LesionLocator: Zero-Shot Universal Tumor Segmentation and Tracking in 3D Whole-Body Imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025, pp. 30872–30885.

[7] K. Xu *et al.*, "Development and characterization of a chest CT atlas," in *Medical Imaging 2021: Image Processing*, B. A. Landman and I. Išgum, Eds., Online Only, United States: SPIE, Feb. 2021, p. 48. doi: 10.1117/12.2580800.

[8] *MIC-DKFZ/LesionLocator*. (Oct. 01, 2025). Python. MIC-DKFZ. Accessed: Oct. 16, 2025. [Online]. Available: https://github.com/MIC-DKFZ/LesionLocator

[9] G. R. D. Luca, M. Mascalchi, and S. Diciotti, "Development of a 3-D standardized lung template from low-dose CT scans," presented at the IEEE-EMBS International Conference on Biomedical and Health Informatics 2025, Oct. 2025. Accessed: Nov. 21, 2025. [Online]. Available: https://openreview.net/forum?id=O6vCD7wMrT#discussion

[10] F. Bianconi *et al.*, "Comparative evaluation of conventional and deep learning methods for semi-automated segmentation of pulmonary nodules on CT," *Quant Imaging Med Surg*, vol. 11, no. 7, pp. 3286–3305, July 2021, doi: 10.21037/qims-20-1356.

[11] L. Chen *et al.*, "An artificial-intelligence lung imaging analysis system (ALIAS) for population-based nodule computing in CT scans," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101899, Apr. 2021, doi: 10.1016/j.compmedimag.2021.101899.

[12] B. Li, G. E. Christensen, E. A. Hoffman, G. McLennan, and J. M. Reinhardt, "Establishing a Normative Atlas of the Human Lung," *Academic Radiology*, vol. 19, no. 11, pp. 1368–1381, Nov. 2012, doi: 10.1016/j.acra.2012.04.025.

[13] K. Murphy *et al.*, "Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge," *IEEE Transactions on Medical Imaging*, vol. 30, no. 11, pp. 1901–1920, Nov. 2011, doi: 10.1109/TMI.2011.2158349.

[14] B. Li, G. Christensen, E. A. Hoffman, and J. M. Reinhardt, "Establishing a Normative Atlas of the Human Lung:," *Academic Radiology*, vol. 10, no. 3, 2003.

[15] P. G. Mikhael *et al.*, "Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk from a Single Low-Dose Chest Computed Tomography," *Journal of Clinical Oncology*, vol. 41, no. 12, pp. 2191–2200, Apr. 2023, doi: 10.1200/JCO.22.01345.

[16] X. Li, P. S. Morgan, J. Ashburner, J. Smith, and C. Rorden, "The first step for neuroimaging data analysis: DICOM to NIfTI conversion," *Journal of Neuroscience Methods*, vol. 264, pp. 47–56, May 2016, doi: 10.1016/j.jneumeth.2016.03.001.