Lung 4DCT Image Registration: A Comparison Between ITK Elastix and Monai Deep Learning Approaches

Daniel Tweneboah Anyimadu[123], Taofik Ahmed Suleiman[123]

Email: dtanyimadu@gmail.com, suleimantaofik6@gmail.com
Supervised by: Professor Robert Marti [1] and Josep Quintana [1]

[1] Department of Medical Imaging and Computing, University of Girona, Spain
[2] Department of Electrical and Information Engineering, MAIA, University of Cassino and Southern Lazio, Italy
[3]Graduate School of Science and Technology, Medical Imaging and Applications (MAIA), University of Burgundy, France

ABSTRACT

This project explores five different methods for registering inhale and exhale lung CT scans from the DIR Lab COPDGene dataset. The registration methods implemented include Elastix and Transformix, an intensity-based image registration algorithm, and compared its performance with several pre-trained deep-learning algorithms (SegResNet, Vnet, Autoencoder, and HighResNet). The methodology implemented generally involves raw data handling, data preprocessing and segmentation, then leveraging Elastix/Transformix or the deep-learning registration algorithms and evaluating results using mean Target Registration Error (TRE) and mean standard deviations (std). The comparative analysis of the training and validation datasets (COPD1, COPD2, COPD3, and COPD4) revealed that the Elastix ITK approach utilizing the Par0011, BSP from ModelZoo, and customized BSP (bspline_transfomed) parameters provided a more reliable, and robust TRE compared to the deep learning approaches, and this is a result of a lack of sufficient data for the deep learning approach to learn more representative features. The Elastix ITK method achieved an average TRE (std) of 6.68 (5.98). This project aims to contribute to lung registration methods, offering insights for improved medical image analysis.

Keywords: *Registration, Elastix, Transformix, Deep Learning, SegResNet, Vnet, Autoencoder, HighResNet*

1.0 INTRODUCTION

Image registration is essential for medical image analysis, aligning images for objective feature comparison [1]. The registration of medical images serves various applications, including atlas creation, intra-patient scan alignment, segmentation, and volume quantification [2]. In the medical domain, this technique, whether classical or deep learning-based, revolutionizes diagnostics and treatment planning. Specifically, in the context of chronic obstructive pulmonary disease (COPD), lung CT image registration is vital for detecting lung disorders [3]. Image registration between maximum inhale and exhale phases is relevant for lung diagnosis and characterization and hence this project, utilized the 4DCT DIR-Lab Challenge dataset and focused on leveraging the capabilities of the Insight Segmentation and Registration Toolkit (ITK) and Elastix framework for precise lung CT image registration while comparing the result with several deep learning registration algorithms from Monai. This project aims to enhance diagnostic accuracy, improve treatment planning, and streamline respiratory condition monitoring. The subsequent sections will detail the methodology including the data handling, preprocessing and segmentation, Elastix/Transformix, and deep learning models for registration, as well as evaluation methods, and results for each experiment conducted.

2.0 DATA HANDLING AND PROCESSING

This project has been achieved in three key pipelines. These are data processing and visualization, CT lung preprocessing, and finally, registration using either

Elastix/Transformix or any of the pre-trained deep learning models (SegResNet, Vnet, Autoencoder, and HighResNet). Figure 1 shows a schematic diagram representing the registration of lung CT images (exhale onto inhale).

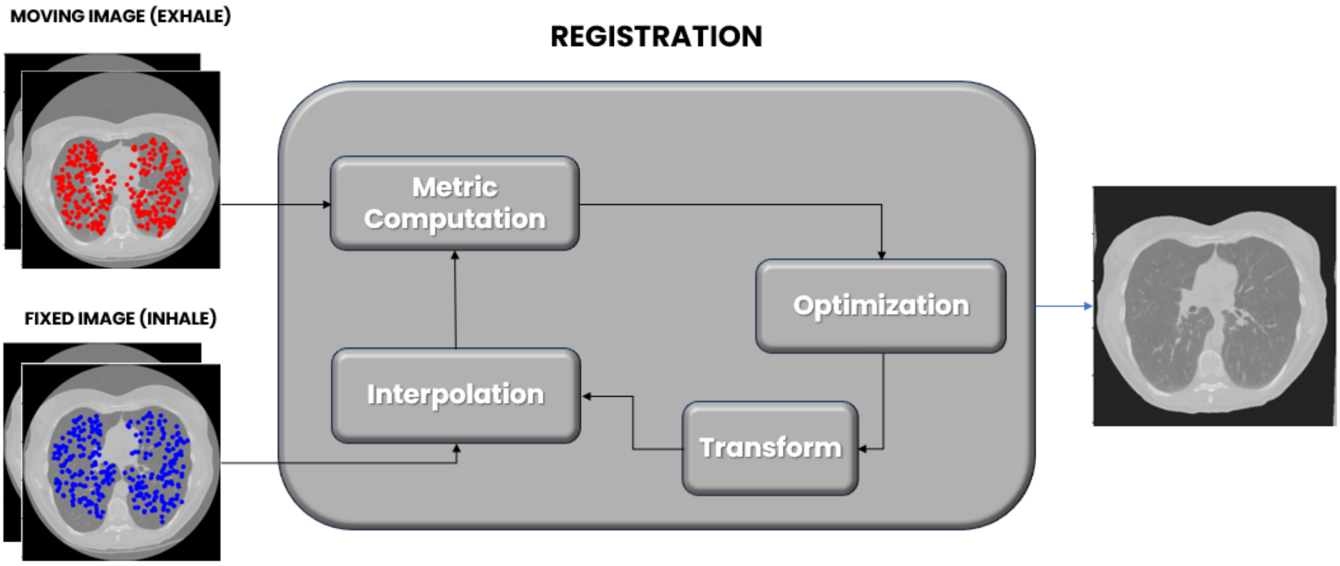


Figure 1: A schematic diagram representing the registration of lung CT images (exhale onto inhale).

2.1 Data Source

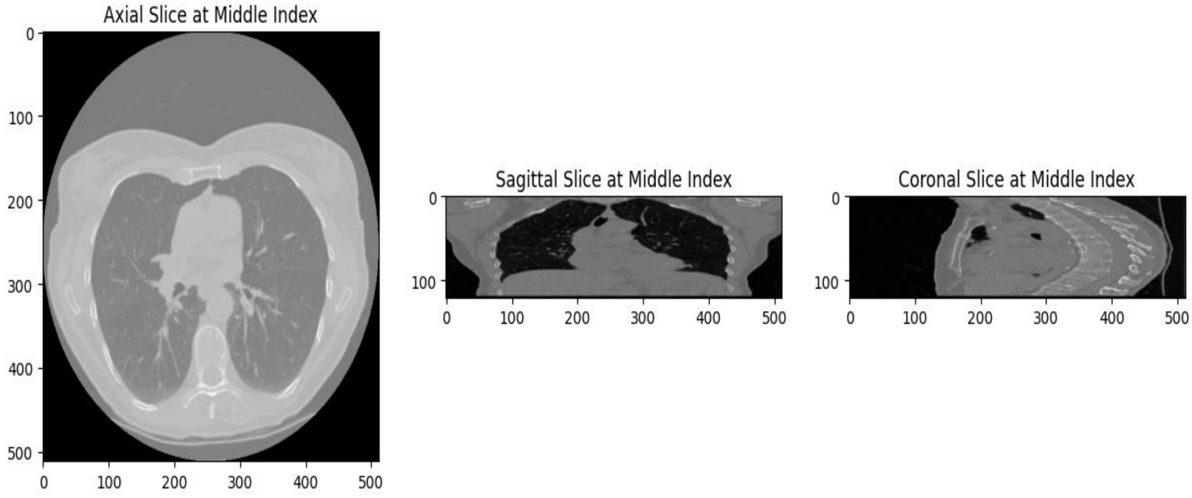
The DIR Lab COPDGene dataset, consisting of 10 pairs of lung CT scans, was used for our project. We focused on cases 1 to 4 for training and validation, each pair comprising of maximum inhalation and exhalation scans. The dataset includes raw images, and landmarks files with 300 coordinates and mean displacement. Since this data was raw, it was important to convert it to ITK readable format such as nifty and hence the need for data handling and conversion.

2.2 Raw Data Handling

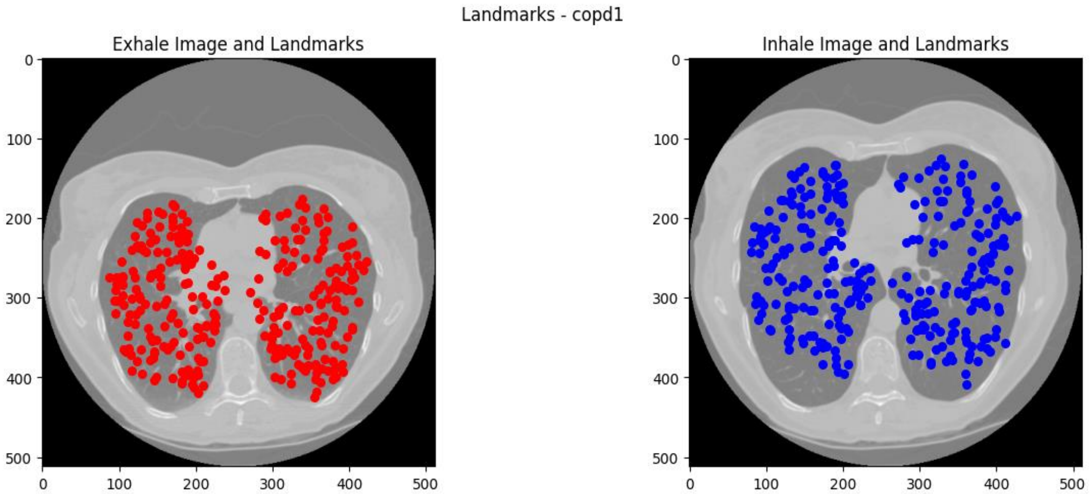
In our lung CT image registration project, a key step involves handling raw binary scalar images. We employed the Python code found in the SimpleITK [4] documentation under "Raw Image Reading" to effectively handle the raw binary scalar images. The "read_raw" function from the documentation serves a dual purpose: it decodes raw binary scalar images and transforms them into the SimpleITK format, to ensure smooth operation with the ITK and Elastix frameworks. We save the resulting images in nifty format to allow further data exploration and analysis.

2.3 Exploratory Data Analysis (EDA)

An exploratory data analysis (EDA) was conducted to understand the characteristics of the 4DCT DIR-Lab Challenge dataset. By visualizing the images alongside their anatomical landmarks, we identified precise landmark locations. This understanding is relevant for aligning inhale and exhale images accurately, enabling detailed feature comparison. One challenge emerged during this exploration — the lungs are embedded in a circular gantry region, making it tricky to distinguish between the intensity distribution and spatial location of the inhale and exhale images. This challenged guided us in making informed decisions on subsequent preprocessing techniques to leverage on the 4DCT DIR-Lab Challenge dataset and remove the circular gantry region. Figure 2 (a and b) shows the CT lung image during exhalation and inhalation for COPD1 at various positions and landmarks.



2a. Axial, sagittal, and coronal views of the CT lung image during exhalation for COPD1.



2b. Original lung CT images with their landmark points, indicated by red and blue dots

Figure 2: Data visualization (2a. Axial, sagittal, and coronal views of the CT lung image, and 2b. Displays the original lung CT images with their landmark points, indicated by red and blue dots.

2.4 CT Lung Preprocessing

To effectively prepare the CT images for subsequent registration with both ITK Elastix and deep-learning frameworks, we have developed the robust “CTPreprocessing” class. This comprehensive class encapsulates key functions: check_fov, segment_kmeans, chest_hole_filling, remove_gantry, and apply_CLAHE. Each function plays a crucial role in enhancing the quality and clarity of CT images, ensuring optimal performance in preprocessing of the lung for the registration algorithm.

2.4.1 Field of View Evaluation

The check_fov function serves as the initial step in our CT preprocessing pipeline. It evaluates the field of view (FOV) by examining intensity levels in specific corners of the image. If the intensity in at least three corners falls below the specified threshold, it indicates the presence of FOV. Based on this evaluation, the subsequent segmentation is tailored accordingly, with a K value of 3 if FOV is present and 2 otherwise.

2.4.2 Segmentation with K-Means

Following FOV assessment, the segment_kmeans function employs the K-Means clustering algorithm to segment the CT image. The inverted image is vectorized and subjected to K-Means clustering, with the resulting centroids used to reconstruct the segmented image. This step is crucial in distinguishing anatomical structures and facilitating further processing.

2.4.3 Chest Hole Filling

The chest_hole_filling function addresses the presence of holes or gaps in the segmented image. Utilizing contour detection, the function identifies the contour with the maximum area, creating a mask to fill the corresponding hole. Morphological operations was applied to refine the mask, contributing to a more complete and accurate representation of the lung structure.

2.4.4 Gantry Removal

The remove_gantry function leverages the segmented image and the contours obtained from chest hole filling to selectively remove gantry artifacts. By multiplying the input image with the contours, we effectively eliminate unwanted elements, enhancing the overall quality of the CT image.

2.4.5 Contrast Enhancement with CLAHE

To further improve image contrast and prevent over-saturation in homogeneous areas, the apply_CLAHE function implements the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. This step is crucial for highlighting subtle anatomical features, contributing to more precise image registration and subsequent analysis. Figure 3 illustrates the general workflow of the functions within the class, showcasing the seamless transition of data between each processing step. Figure 4 presents intermediate and final results of the preprocessing, providing a visual representation of the effectiveness of each function in enhancing the CT images for the next registration step.

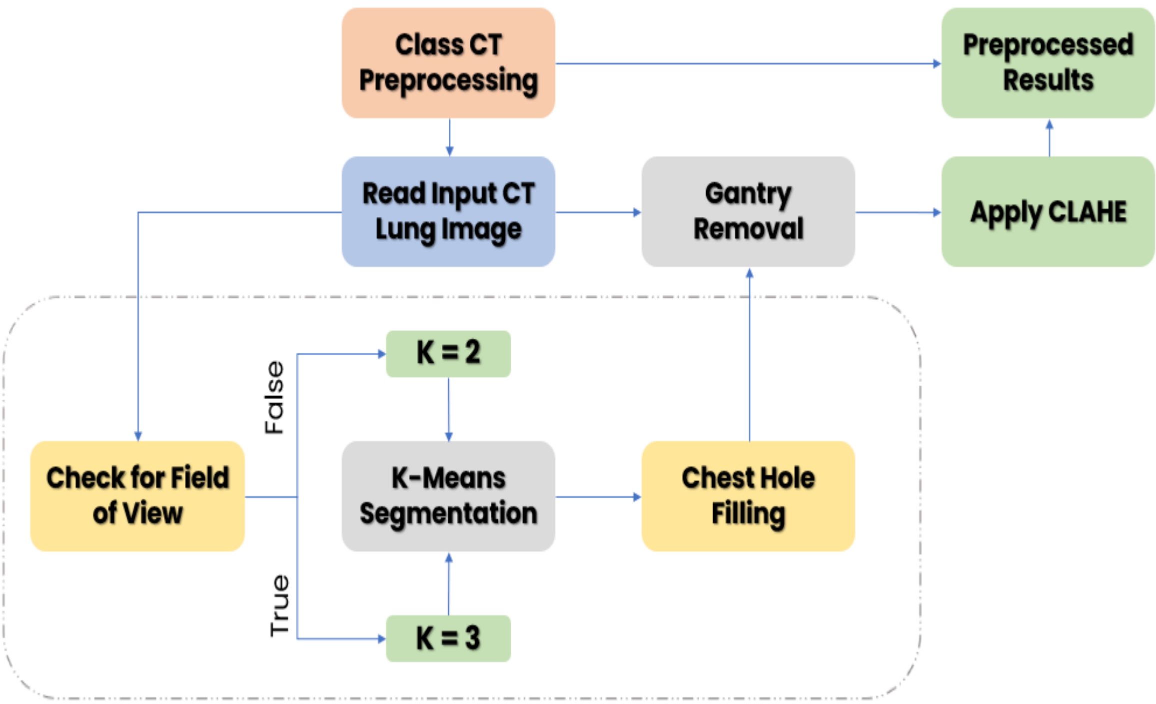


Figure 3 General workflow of the functions within the class, showcasing the seamless transition of data between each processing step.

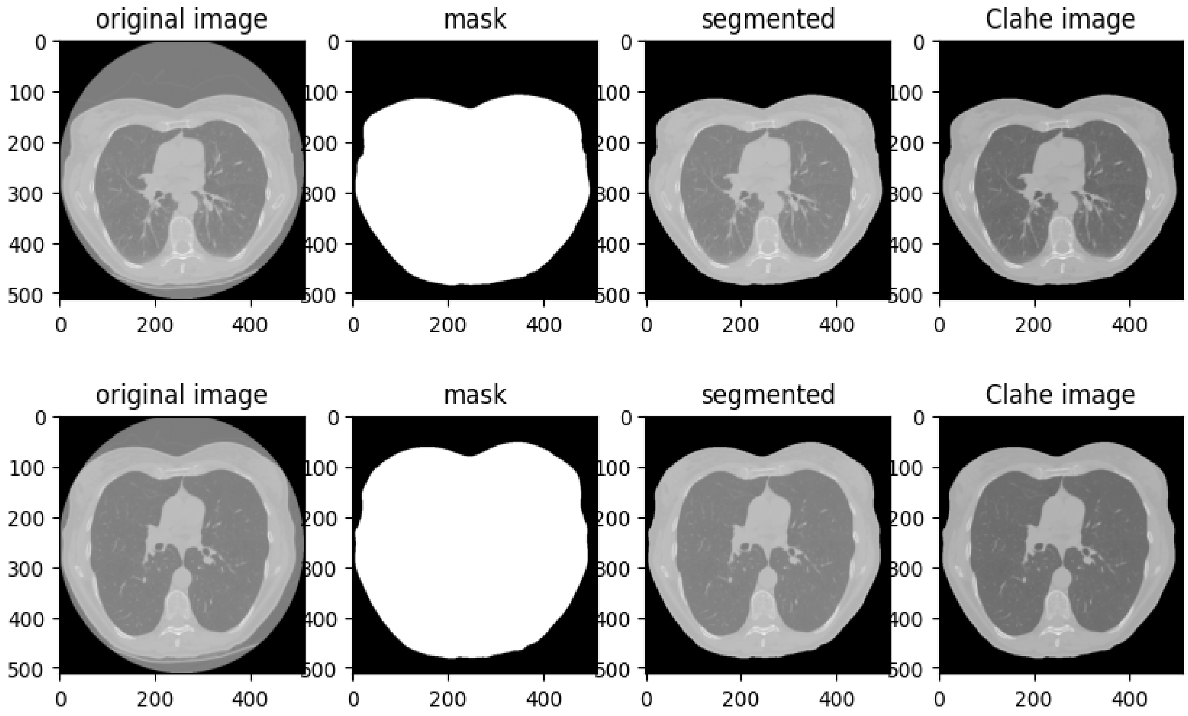


Figure 4: Visualizations of the original image, segmentation mask, segmented image after gantry removal, and the final CLAHE-enhanced image.

2.5 Target Registration Error (TRE)

The primary metric of evaluation for our project is the Target Registration Error (TRE). In the initial phase, we computed the TRE using the provided landmarks without any registration, providing insights into the disparities between exhale and inhale landmarks. Table 1 presents the results of the mean TRE and standard deviation without registration.

Table 1: Mean and standard deviation (std) of target registration error (TRE) before lung CT image registration (exhale to inhale).

Case Folder	**Mean TRE (mm)**	**Std TRE (mm)**
COPD1	26.334214	11.417905
COPD2	21.785988	6.460537
COPD3	12.639169	6.384308
COPD4	29.583560	12.924171

3.0 EXPERIMENTATION: CLASSICAL AND DEEP-LEARNING

The registration process commences by configuring the moving and fixed images, where the exhale image serves as the moving image, and the inhale image as the fixed image. The rationale behind this setup is to execute registration by transforming exhale images to align with inhale images.

3.1 Registration Using Elastix and Transformix

The Elastix transformation is initiated with the input parameters, including the fixed image (inhale), moving image (exhale), transformation parameter, and an output directory. This enables the transformation of exhale images to align with the corresponding inhale images. The resultant output includes the transformed exhale images and their associated transform parameter result. Following the Elastix transformation, the process extends to incorporate the Transformix which uses the moving image (exhale), landmark of the inhale image and the transformed parameter result obtained from the Elastix step. The Transformix transformation produced a landmark which contained several values in different format which are stored in a specified output directory. We developed a function '"clean_output_points" which helped process this result to extract our desired output landmark in the same format as the provided landmarks.

To refine our result and further have a more robust registration, we implemented our Elastix by experimenting it on different parameter files as this allows flexibility with options like multi-resolution registration, image pyramid schedules, and customizable components such as interpolator, optimizer, metric, iterations, spatial samples, etc. Our experiments progressed sequentially, initially utilizing default parameters provided by Elastix and gradually advancing to more intricate, task-specific parameter maps available publicly at the Elastix-ModelZoo documentation.

3.1.1 Registration Parameter Configuration

As stated before, our approach to lung CT image registration using Elastix involved the utilization of baseline parameters, namely the inbuilt Affine and B-spline transformations from ITK. These served as our initial benchmarks in the registration process. Additionally, we explored parameter maps from the Elastix-ModelZoo webpage, specifically curated for Chest/Lung 3D/4D CT image registration. After a detailed review of various parameter maps, we narrowed down our selection to three options: Par0007 which includes intrapatient B-spline transformation cost by mutual information, Par0011 which include intrapatient (sometimes intra-sheep) B-spline transformation with normalized correlation, and Par0049 which is intra-subject B-spline transformation using Mattes mutual information. To streamline our experimentation and focus on promising configurations, we combined multiple parameters, and also customized some of the parameter by modify specific features in the file such as the image pyramid schedules. Our final strategy involved combining the best-performing parameter files, with the goal of optimizing the overall outcomes of the registration process. This systematic exploration aimed to enhance the precision and effectiveness of aligning inhale and exhale lung CT images, and the results of all experiment carried out is detailed in the result and evaluation session.

3.2 Registration Using Deep-Learning (DL)

In this section, we delve into the application of advanced deep learning (DL) models for image registration, specifically focusing on utilizing pre-trained models within the MONAI library to register lung CT images with associated landmarks. Our primary challenge revolved around the preparation of the data in a format suitable for training deep learning registration models.

3.2.1 DL Data Preparation

To address the data preparation challenge, we leveraged our preprocessed images and devised a custom data loading function, "get_files". This function efficiently loads images into training and validation sets, organizing them within a dictionary structure. The dictionary contains keys for fixed and moving images, as well as fixed and moving landmarks.

3.2.2 Custom Transforms for Landmarks

Given that MONAI's transforms are primarily designed for images, with an emphasis on spatial transformations, especially rigid/affine or non-linear transformations of voxel grids, adapting them for point clouds posed a unique challenge. To overcome this limitation, we implemented two custom classes: "LoadKeypoints" and "TransformKeypoints". These classes were instrumental in loading keypoint data from TXT files and ensuring that linear transformations were applied consistently with the underlying image transformations, such as those resulting from affine transformations during augmentation. To provide insight into the effectiveness of the customize classes, Figure 5 showcases an example visualization of images and their corresponding point clouds.

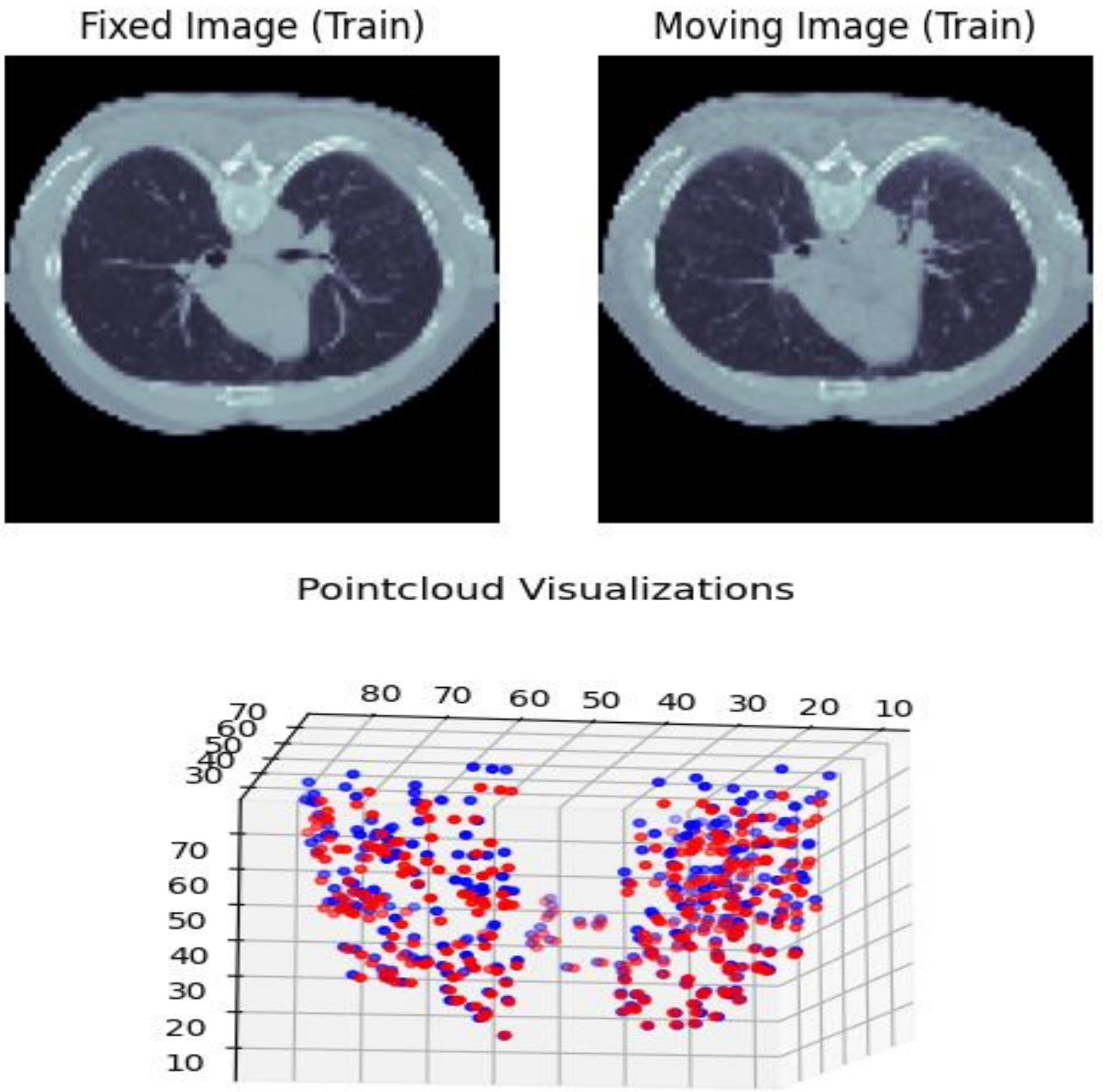


Figure 5: Data visualization from the deep learning data and landmark loader

3.2.3 Hyperparameters and Loss Function

To define the hyperparameters and loss function, we created several functions to achieve this step and the hyperparameters are set in a way to allow the optimization to be driven only by TRE loss since we are interested in points. The "forward" function is responsible for the forward pass of the model. It predicts the Displacement Field (DDF) and warps the moving image based on the predicted DDF while the "collate_fn" function ensures that keypoints are aligned for easy collation during batch processing. The "tre" function computes the TRE loss between fixed and moving landmarks, and finally the "loss_fun" function defines the multi-target loss used for the model optimization. It includes components for TRE, MSE, and Bending Energy and all this are applied to training the model.

3.2.4 Model training and prediction

The models trained are SegResNet, Vnet, Autoencoder, and HighResNet. The general steps involved model initialization, where we specify our pretrained models, and an optimization set up using Adam optimizer with cosine annealing learning rate scheduler. We then created the training and validation loop where the training and prediction is carried out using the TRE as the key loss and best model is saved to enable us to load the pretrained model at anytime to perform prediction. All pretrained models follow this process, the key difference is in their architecture:

- **SegResNet** is a ResNet-based architecture tailored for image segmentation. Here, we employed a warp layer for image warping based on predicted DDF, and incorporates TRE, MSE, and Bending Energy in the loss function.
- **VNet** is a three-dimensional image segmentation network, and we utilized dropout at the various level of the network for regularization.
- **AutoEncoder** consist of an autoencoder architecture with skip connections for feature reuse. We utilized instance normalization and parametric rectified linear unit (PRELU) activation for this model.
- **HighResNet** is a high-resolution network structure for three-dimensional image segmentation. We incorporated padding for channel matching network operation.

4.0 RESULTS, AND EVALUATION METRICS

To assess the performance of the Elastix/Transformix frameworks and pre-trained deep learning models for the lung CT registration task, Target Registration Error (TRE) was computed on both the training and validation sets. TRE quantifies the discrepancy between corresponding points in the fixed (inhale) and transformed moving (exhale) images after the registration process. It serves as a direct measure of the registration accuracy, with smaller TRE values indicating superior alignment precision. The computation of TRE involves identifying anatomical landmarks or fiducial markers in both the fixed and moving images and measuring the Euclidean distance between their corresponding locations post-registration.

4.1 Registration Results for Elastix and Transformix

Figure 6 shows an example of the registered case, while table 2 presents all registration results for different experiments.

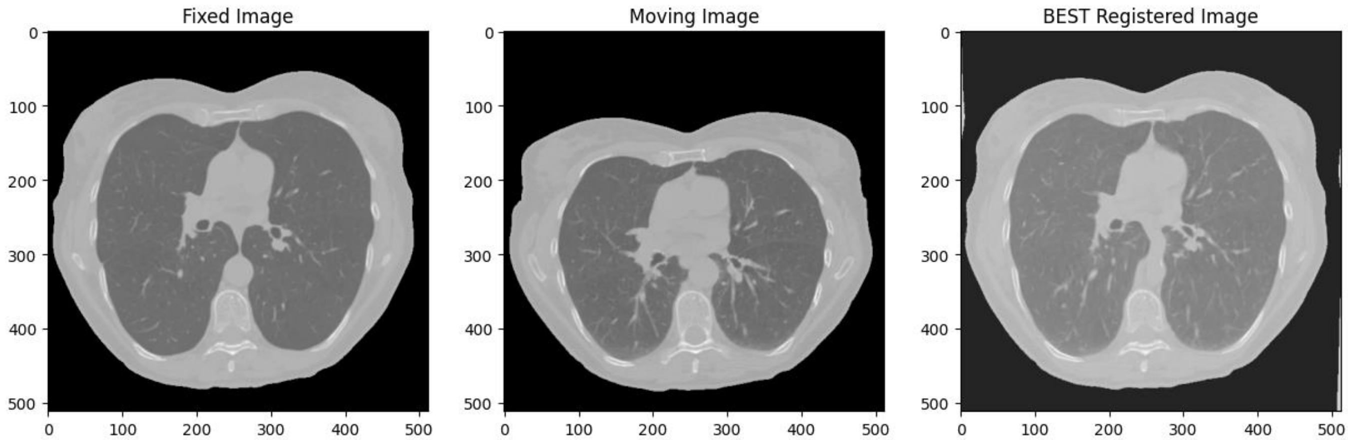


Figure 6: Visualization of the Elastix Registration from COPD1

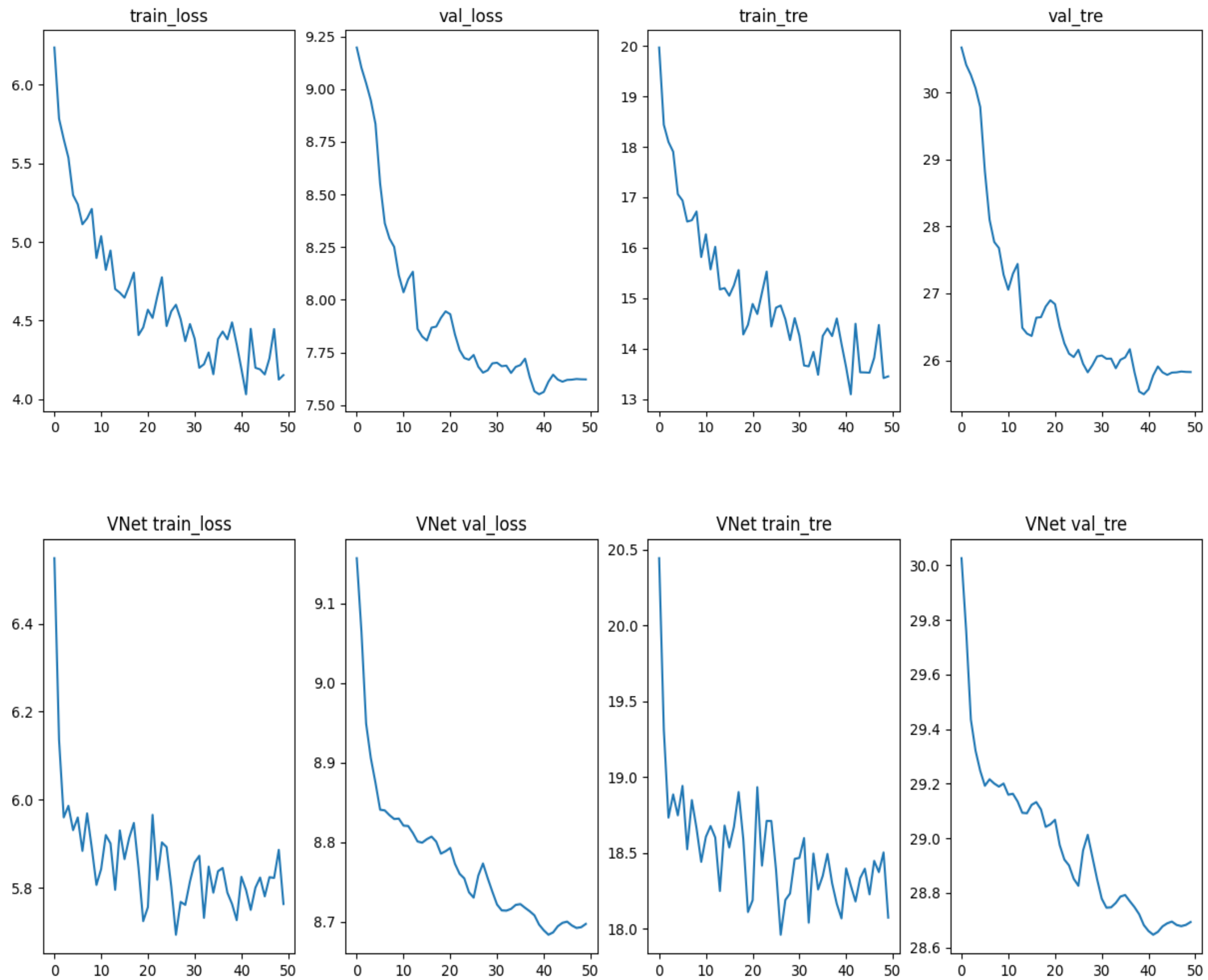
Table 2: Registration results for various parameter configurations using Elastix and Transformix. The values represent the mean TRE and standard deviation in millimetres (mm) of the registration performance for different COPD datasets.

PARAMETER(S)	COPD1	COPD2	COPD3	COPD4	Average
Inbuilt Affine + Bspline	12.69 ± 5.94	18.50 ± 7.48	5.14 ± 4.11	15.31 ± 7.24	12.91 ± 6.19
Par0049	18.72 ± 9.68	18.10 ± 8.13	4.52 ± 4.34	17.45 ± 6.98	14.70 ± 7.28
Par0011	9.43 ± **5.43**	15.86 ± 8.06	4.29 ± 3.48	10.59 ± 5.00	10.04 ± 5.50
Inbuilt Affine + customized Bspline	7.51 ± 6.43	8.19 ± 6.45	3.15 ± 3.62	15.29 ± 11.13	8.54 ± 6.91

Par0011 Affine + customized Bspline	7.31 ± 6.39	8.43 ± 6.93	3.08 ± 3.50	**8.58** ± 6.85	6.85 ± 5.92
Customized Bspline	8.03 ± 6.60	8.61 ± **6.40**	3.14 ± 3.57	18.00 ± 12.23	9.45 ± 7.2
Par0007	10.97 ± 5.85	14.57 ± 6.64	5.13 ± **3.38**	12.02 ± **4.30**	10.67 ± **5.04**
Customized Bspline, and Bspline	7.63 ± 6.66	8.51 ± 6.63	2.99 ± 3.45	17.86 ± 12.11	9.25 ± 7.21
Combining all best parameters	**6.87 ± 6.39**	**8.33** ± 7.29	**2.94** ± 3.41	8.59 ± 6.82	**6.68** ± 5.98

4.2 Registration Results for Deep Learning

These results are gotten from using the model to predict the training and validation set. This is not the best approach but since we do not have enough data, we did this step and rely on the loss during training and validation to identify potential over fitting. Figure 7 shows the visualization of the training and validation losses for all the models while Figure 8 shows an example of one of the registered cases using HighResNet for both images and landmarks using overlay for before and after registration. Table 3 presents all registration results for the different models.



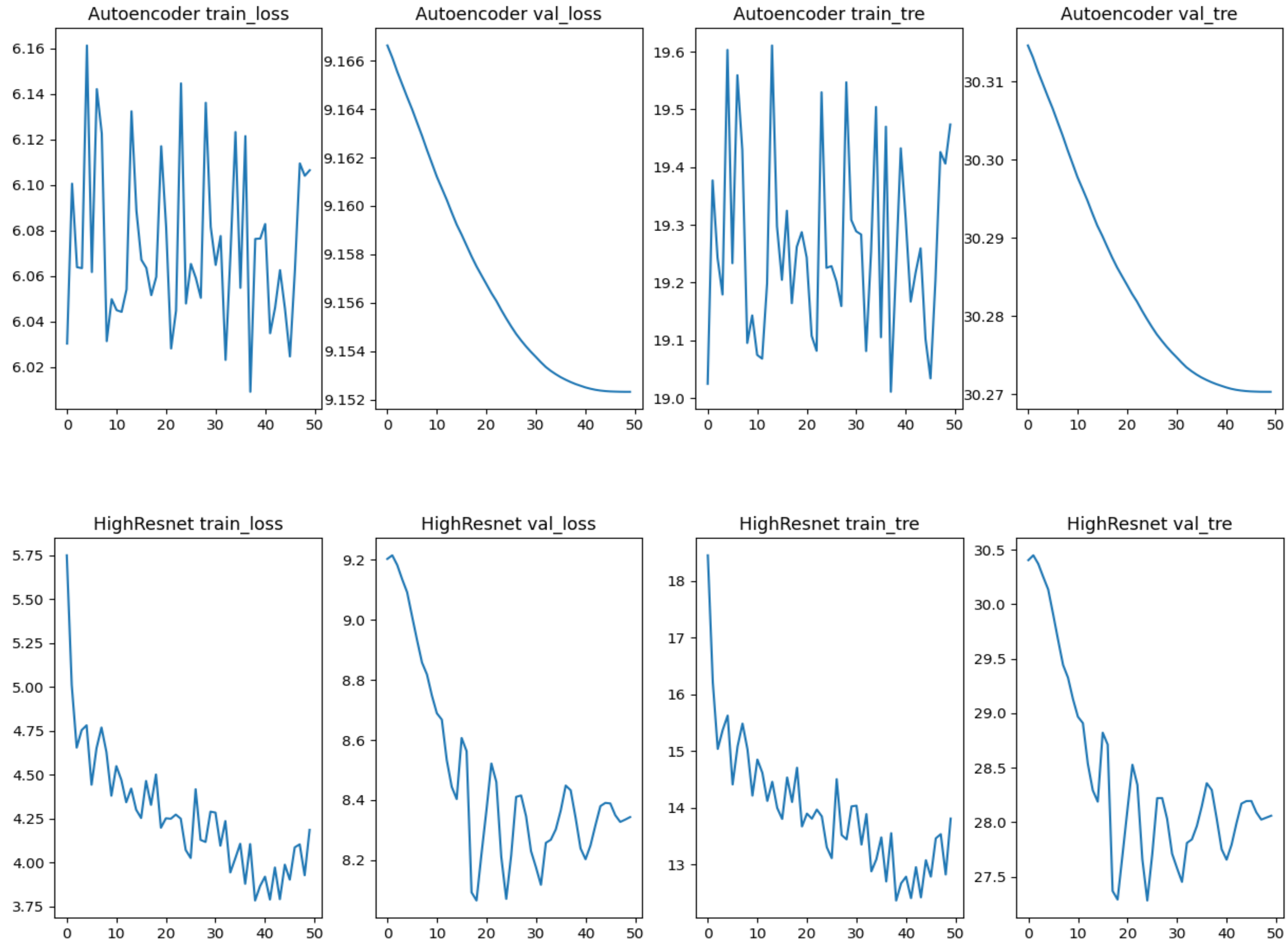


Figure 7: Training and Validation Losses for each model (first plot is the SegResNet model)

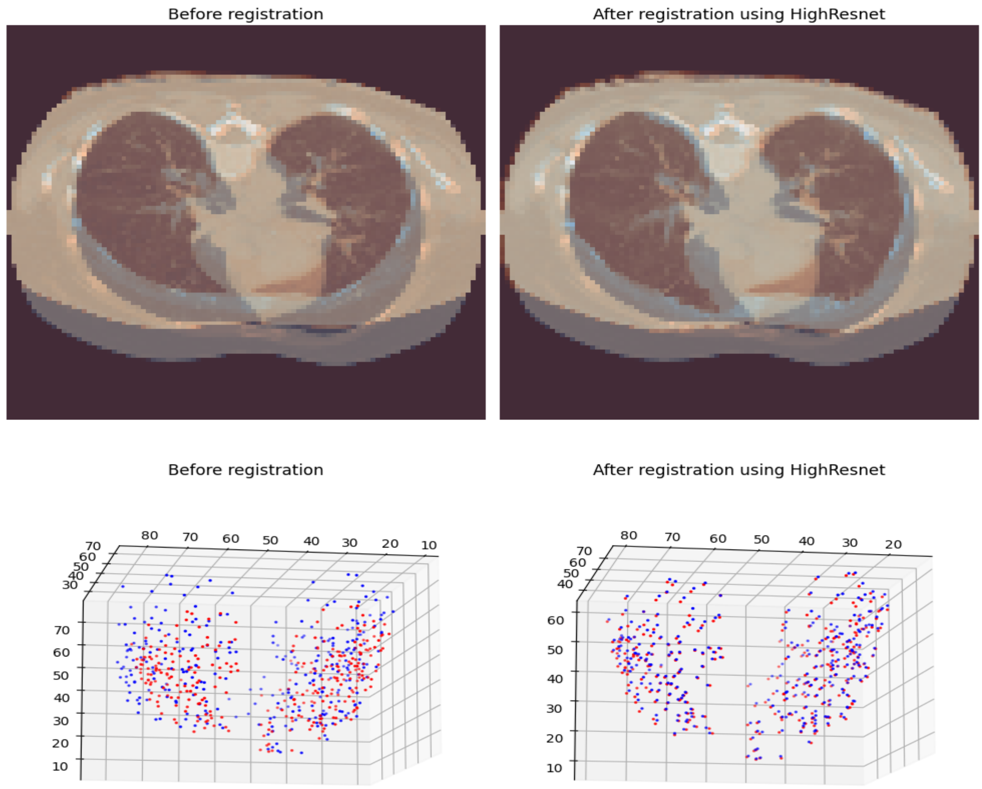


Figure 8: Visualization of the result of the HighResNet model for both images and landmarks using overlay for before and after registration

Table 3: This table presents the mean Target Registration Error (TRE) in mm for different deep learning models on COPD datasets.

MODEL	COPD1	COPD2	COPD3	COPD4	Average
SegResNet	7.2533	10.2973	6.0914	8.4222	8.0161
VNet	3.773	3.2669	4.0019	4.1112	3.7883
Autoencoder	0.5788	0.5662	0.5765	0.5530	0.5686
HighResNet	8.2021	11.3182	3.9731	6.1212	7.4037

5.0 RESULTS DISCUSSION

The results of our comprehensive study comparing classical Elastix-based image registration with advanced deep learning (DL) models present valuable insights into the effectiveness of these methodologies for lung CT image registration. Here, we delve into the key findings, emphasizing the strengths and limitations of each approach.

5.1 Elastix and Transformix

The Elastix-based image registration, employing various parameter configurations, demonstrated commendable performance across multiple COPD datasets. Noteworthy observations include:

- Inbuilt Affine + Bspline: This baseline configuration yielded competitive results, establishing it as a robust starting point for lung CT image registration.
- Optimized Parameter Configurations: Parameter maps such as Par0049, Par0011, and customized Bspline showcased promising results. The systematic exploration and combination of these parameters resulted in improved precision and effectiveness in aligning inhale and exhale lung CT images.
- Customized Bspline and Combining Best Parameters: Notably, configurations involving customized Bspline transformations exhibited enhanced performance. Furthermore, combining the best parameters led to superior outcomes, emphasizing the significance of parameter tuning in refining the registration process.

5.2 Deep Learning Registrations

The exploration of pre-trained DL models, including SegResNet, VNet, Autoencoder, and HighResNet, introduced a dynamic dimension to our project. Key insights from the DL-based image registration are as follows:

- SegResNet and HighResNet Performance: Both SegResNet and HighResNet, designed for distinct purposes, showcased competitive results. SegResNet's balance between precision and computational efficiency, and HighResNet's handling of high-resolution images, contributed to the diversity of DL strategies.
- Autoencoder: The results from the autoencoder however is a huge overfitting as the model could not learn intricate features due to the complexity of autoencoder and less available data.
- VNet displayed promising results, and its performance indicates the adaptability of segmentation-focused models for broader medical image processing and registration tasks.

5.3 Comparative Analysis

For the challenge, we decided to use our elastix result as they are more reliable and have been tested using several parameter files and experimentation. The deep learning on the other hands can only perform better than the elastix with sufficient amount of data. A promising approach would be to combine these results, as monai support the results of a deep learning registration model to be converted into an affine transformation parameter which can further be passed to the elastix as a parameter file for a well refined registration and this hybrid method is a very promising research area to look into.

6.0 CONCLUSION

In this project, we evaluated and compared key medical image registration methods, focusing on the specific case of lung computed tomography (CT) registration. Our comparison included the classical Elastix/Transformix approach and pre-trained deep learning methods, SegResNet, VNet, Autoencoder, HighResNet. Through thorough exploration of different configurations and preprocessing techniques for each method and given the limited case to study, Elastix/Transformix registration algorithm demonstrated more robust and reliable performance compared to other methods.

7.0 REFERENCES

[1] M. Chen, N. J. Tustison, R. Jena, and J. C. Gee, “Image Registration: Fundamentals and Recent Advances Based on Deep Learning,” *Neuromethods*, vol. 197, pp. 435–458, 2023, doi: 10.1007/978-1-0716-3195-9_14/FIGURES/4.

[2] J. Antoine Maintz, “A survey of medical image registration,” *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, 1998, doi: 10.1016/s1361-8415(98)80001-7.

[3] L. Zhang, B. Jiang, H. J. Wisselink, R. Vliegenthart, and X. Xie, “COPD identification and grading based on deep learning of lung parenchyma and bronchial wall in chest CT images,” *Br. J. Radiol.*, vol. 95, no. 1133, May 2022, doi: 10.1259/BJR.20210637.