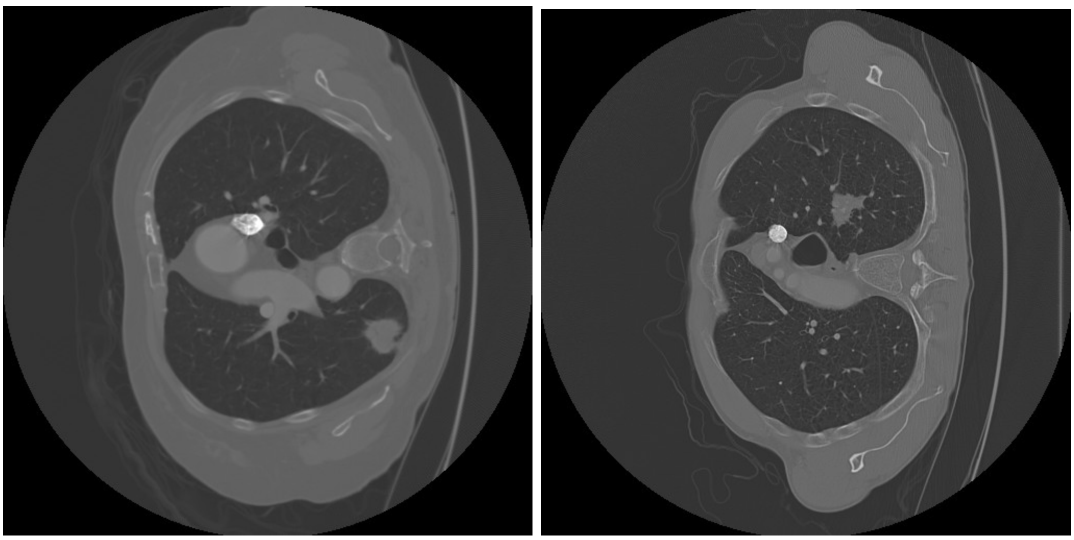
I. Preparation of the database and computing environment

When developing algorithms for automatic lung segmentation, the LIDC-IDRI database was used, described at https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI

The Lung Image Database Consortium (LIDC-IDRI) image collection consists of diagnostic and screening computed tomography (CT) scans of lung cancer with annotated lesions.

Of the 1018 cases collected in the LIDC-IDRI database, 477 imaging studies come from people who have not been diagnosed with lung cancer, and the remaining cases come from people with cancer.

In people who have not been diagnosed with cancer, other types of lesions are present in the lungs, examples of which are shown in Figure 1.



Rys. 1

The database contains a number of non-physiological cases:

- cases with tracheostomy and endotracheal tubes

- cases of fistula between the trachea and the esophagus

- cases of partial or complete lung resection

- artifacts related to the presence of implants or pacemakers

In addition, the database contains images of very different quality, which is due to:

- various slice thicknesses (from 0.45 to 5 mm) - a summary of layer thicknesses is presented in Table 1

- different pixel sizes - a list of pixel sizes is presented in Table 2

- different filters used for reconstruction (both soft tissue and sharp), which is associated with different levels of noise

Table 1

Slice thickness [mm]	Number of cases
do 0.5	11
0.5 – 1.0	274
1.0 – 1.5	225
1.5 – 2.0	105
2.0 – 2.5	285
2.5 – 3.0	115
above 3	3

Table 2

Pixel size [mm]	Number of cases
do 0.5	4
0.5 – 0.6	156
0.6 – 0.7	350
0.7 – 0.8	415
0.8 – 0.9	79
powyżej 0.9	14

Tomographic images were always 512 by 512 pixels in the scan plane. Varying in the range from 0.5 mm to 5 mm, the thickness of the layers affects the number of layers in the CT image. A summary of the number of layers in the images is presented in Table 3.

Table 3

Number of slices of a CT image	Number of cases
do 100	13
100 – 150	397
150 – 200	94
200 – 300	284
300 – 400	64
400 – 500	105
500 – 600	52
600 –700	6
powyżej 700	3

Imaging studies, originally in the DICOM format, were converted to the nifti format, which - unlike the DICOM format - allows the entire CT examination to be saved in one file. Conversion to the nifti format facilitates data management and programmatic access to data, which is implemented using the nibabel library. Listing 1 shows how to read a nifti file into a 3D matrix and write such a matrix to a nifti file.

Listing 1

```
##############################################
# import library for nifti file operations
import nibabel as nib

# example image in nifti format
scan_path = ‘niftiImage.nii.gz’
# read nifti file to an array
img = nib.load(scan_path).get_fdata()

# create nifti image to be saved
niftiImage = nib.Nifti1Image(img, affine=np.eye(4))
# save image to nifti file
nib.save(niftiImage,’image.nii.gz’)
##############################################
```

II. Body mask segmentation

The body mask segmentation method has been described, for example, in the work of John and Mini [1, 2, 3]. In this work, the body mask is found based on iterative global thresholding, using the concept of density differences and the morphological closure operation.

[1] John J, Mini MG. Multilevel thresholding based segmentation and feature extraction for pulmonary nodule detection. Procedia Technol 2016; 24: 957-63.

[2] Shabana Rasheed Ziyad, Venkatachalam Radha and Thavavel Vayyapuri, Overview of Computer Aided Detection and Computer Aided Diagnosis Systems for Lung Nodule Detection in Computed Tomography, Current Medical Imaging, 2020, 16, 16-26.

[3] Tom Doel, David J. Gavaghan, Vicente Grau, Review of automatic pulmonary lobe segmentation methods from CT, Computerized Medical Imaging and Graphics 40 (2015) 13–29.

The implemented airway segmentation method is based on thresholding. When imaging a typical patient's chest, the CT image usually covers the area from the top of the shoulders to the upper abdomen. In physiological cases, in this area of the body, there is no direct communication between the area occupied by the airway and the area outside the patient's body volume.

Since the database includes non-physiological cases, e.g. cases with tracheostomy, the above condition is not met and in order to correctly segment the airways, a body mask must first be extracted, which will then allow to limit the search area for the airways.

The main part of the body mask segmentation program is presented in Listing 1, which also includes comments on the performed operations. Tissues within the body have a Hounsfield number above the threshold of TH = -191. When implementing body mask segmentation, you can/should use morphological operations and single-connected component labeling operations.

OpenCV: https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html

scikit-image: https://scikit-image.org/docs/stable/api/skimage.morphology.html

scipy.ndimage: https://docs.scipy.org/doc/scipy/reference/ndimage.html

Listing 1

```
###############################################


# read CT image from nifti file to a 3D array
```

```
orgIm = ……

# create a placeholder for body mask
thresholded = ……

# the body mask is determined in a loop for subsequent 2D slices of the 3D study
for sl in range(0,orgIm.shape[2]):
        dum = segmentThorax(orgIm[:,:,sl],………..)
        np.copyto(thresholded[:,:,sl],dum)

# saving segmentation results to a file
…………….
###################################################
```

Fig. 2 shows an example CT image and the corresponding body mask segmentation.

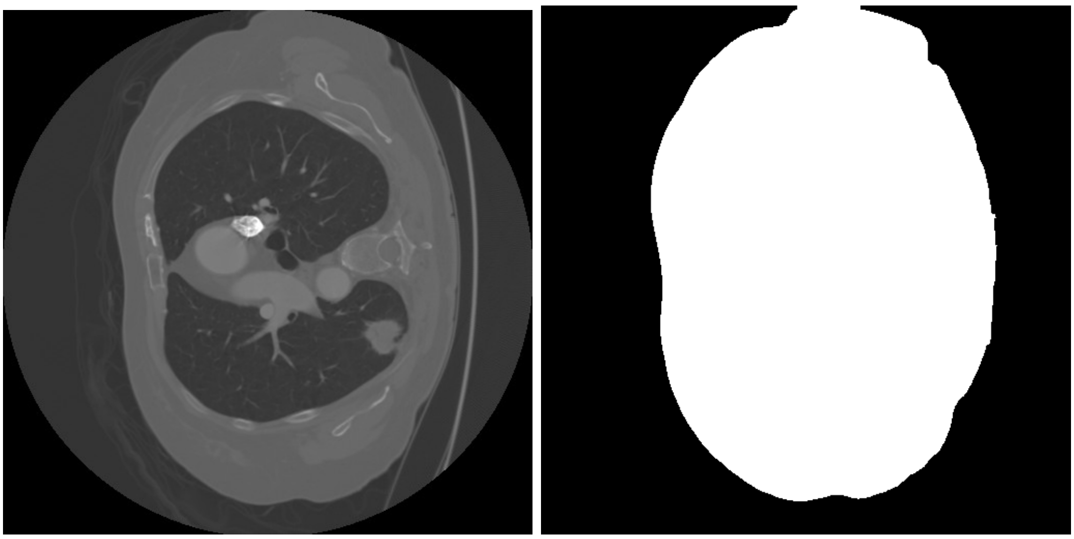


Fig. 2 CT image (left) and corresponding body mask (right)