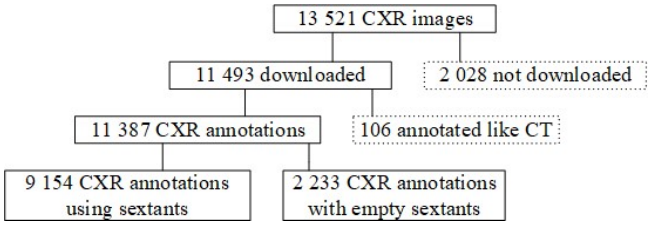

Figure 3: CXR image categories from TB Portals [1] website.

analysis there is a report on the use of the obtained data in solving the prediction of the parameter ”Overall percent of abnormal volume” is given.

II. CXR annotations, as if it were a CT

There are 106 images described like a CT. All from Georgia. These images have 19 parameters with disease descriptions, like a CT scan. Annotation parameters for CXR are absent.

For example, in Fig. 4, patient ID 1534 has one annotated image with a study modality CR (Computed Radiography), but the actual description is as if it were a CT (Computed Tomography).

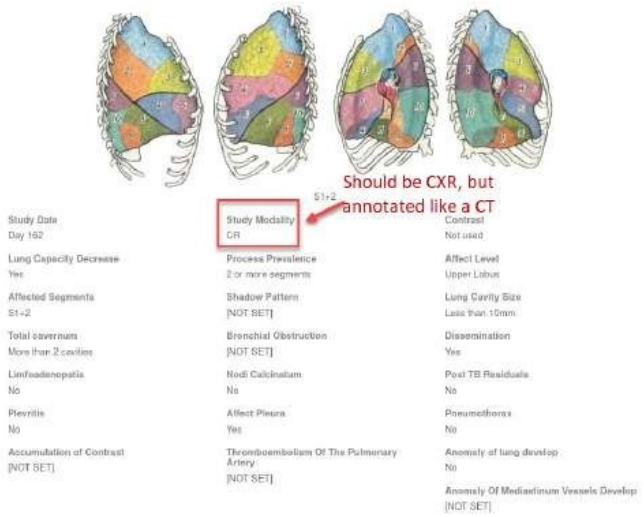

Figure 4: Patient ID 1534 with CXR annotation, as if it were a CT.

It was decided to exclude these images from the further investigations for the following reasons:

- their annotation is different than most other files, meaning that these images cannot participate in neural network training together with everyone else;
- 106 files can participate in separate neural network training, but there is more valuable data to explore;
- it is unclear why these images are described differently, and until this is clarified, it is better to exclude them from further study.

III. CXR annotations of sextants

A. *Lesion names to describe the lungs lobes*

In the CXR image, the lungs are divided into six lobes or sextants:

- Upper Right Sextant;
- Upper Left Sextant;
- Middle Right Sextant;
- Middle Left Sextant;
- Lower Right Sextant;
- Lower Left Sextant.

Each sextant has twenty identical lesion names, which can be found in the corresponding JSON files for each CXR image:

1) Small Cavities
2) Medium Cavities
3) Large Cavities
4) Is any Large cavity belong to a multi-sextant cavity?
5) Can Multiple cavities be seen?
6) Small Nodules
7) Medium Nodules
8) Large Nodules
9) Huge Nodules
10) Is any calcified or partially-calcified Nodule exists?
11) Is any non-calcified Nodule exists?
12) Is any clustered Nodule exists
13) Are multiple Nodules exist?
14) Low/ground glass Density (active fresh nodules)
15) Medium Density (stabilized fibrotic nodules)
16) High Density (calcified nodules, typically sequella)
17) Low/ground glass Density
18) Medium Density
19) High Density
20) Collapse

The data displayed on the CASE BROWSER [2] and TB DEPOT [3] websites and the tag names in JSON files are slightly different, but a mutually unambiguous match can be made even by a non-specialist. The only non- obvious correspondence is that the field ”Infiltrate” on the website should correspond to the tag ”High Density” in JSON files. This match between the ”Infiltrate” field and ”High Density” tag was found by the method of exceptions as the last possible match.

Mutually unambiguous correspondence between the data on the websites and in JSON files is given in Tab. I and Tab. III.

The sequential arrangement of the twenty lesions in Tab. I is the same as on the CASE BROWSER [2] website.

The number of non-empty lesions in the lung sextants is listed in Tab. II. For convenience and to save space, in Tab. II, the word ”Sextant” is omitted in table header. The total number of

Table I:
Twenty lesions to describe the lungs lobes in CXR images

№	Tags in JSON files ”info-all.json”	CASE BROWSER [2] website	TB DEPOT [3] website
		Cavity	
1	Small Cavities	Small Cavities (less than 3 cm)	Small Cavities
2	Medium Cavities	Medium Cavities (3-5 cm)	Medium Cavities
3	Large Cavities	Large Cavities (more than 5 cm)	Large Cavities
4	Is any Large cavity belong to a multi sextant cavity?	Does any Large cavity belong to a multi-sextant cavity?	Large Cavity Multi Sextant
5	Can Multiple cavities be seen?	Can Multiple cavities be seen?	Multiple Cavities Seen
		Nodules	
6	Small Nodules	Small Nodules (less than 3 mm)	Small Nodules
7	Medium Nodules	Medium Nodules (5-15 mm)	Medium Nodules
8	Large Nodules	Large Nodules (15-30 mm)	Large Nodules
9	Huge Nodules	Huge Nodules (more than 30 mm, tuberculoma)	Huge Nodules
10	Is any calcified or partially-calcified Nodule exists?	Any calcified or partially-calcified Nodules?	Partially Calcified Nodule Exists
11	Is any non-calcified Nodule exists?	Any non-calcified Nodules?	Non Calcified Nodule Exists
12	Is any clustered Nodule exists	Any clustered Nodules (nodules 2-5 mm apart)?	Any Clustered Nodule Exists
13	Are multiple Nodules exist?	Can Multiple Nodules be seen?	Multiple Nodule Exists
14	Low/ground glass Density (active fresh nodules)	Low/ground glass Density (active fresh nodules)	Low Ground Glass Density Active Fresh Nodules
15	Medium Density (stabilized fibrotic nodules)	Medium Density (stabilized fibrotic nodules)	Medium Density Stabalized Fibrotic Nodules
16	High Density (calcified nodules, typically sequella)	High Density (calcified nodules, typically sequella)	High Density Calcified Typically Sequella
		Infiltrate	
17	Low/ground glass Density	Low/ground glass Density	Infiltrate Low/Ground Density
18	Medium Density	Medium Density	Infiltrate Medium Density
19	High Density	High Density	Infiltrate
		Collapse	
20	Collapse	Collapse	Collapse

[a]*Lesion names have been left unchanged for accuracy of description.*

Table II:
The number of non-empty lesions in the lung sextants

Lesion name	Upper Right	Upper Left	Middle Right	Middle Left	Lower Right	Lower Left
Small Cavities	6723	5856	4654	4717	2756	2596
Medium Cavities	6720	5856	4653	4714	2755	2596
Large Cavities	6719	5855	4652	4714	2755	2596
Is any Large cavity belong to a multi-sextant cavity?	6732	5868	4663	4727	2768	2608
Can Multiple cavities be seen?	6732	5867	4661	4727	2768	2606
Small Nodules	6724	5856	4656	4719	2757	2599
Medium Nodules	6724	5855	4653	4715	2755	2596
Large Nodules	6720	5853	4652	4714	2755	2596
Huge Nodules	6720	5854	4652	4714	2755	2596
Is any calcified or partially-calcified Nodule exists?	6733	5868	4666	4727	2768	2609
Is any non-calcified Nodule exists?	6731	5868	4665	4728	2768	2609
Is any clustered Nodule exists	6731	5864	4665	4727	2769	2608
Are multiple Nodules exist?	6730	5862	4661	4723	2768	2608
Low/ground glass Density (active fresh nodules)	6726	5856	4656	4718	2757	2599
Medium Density (stabilized fibrotic nodules)	6724	5857	4654	4716	2755	2596
High Density (calcified nodules, typically sequella)	6720	5853	4652	4714	2754	2596
Low/ground glass Density	6724	5856	4657	4715	2760	2598
Medium Density	6724	5858	4655	4716	2757	2597
High Density	6720	5855	4652	4713	2755	2596
Collapse	6720	5855	4652	4714	2754	2598

text annotations is 546,364. For the right lung: 282,817. For the left lung: 263,547.

The division of the lungs into sextants is made by a radiologist according to the following rules.

For CXR:

- Upper sextants are above the lower edge of the aortic arch;
- Middle sextants are between the lower edge of the aortic arch and the right inferior pulmonary vein);
- Lower sextants are below the right inferior pul- monary vein.

For CT:

- Upper sextants are above keel level;
- Middle sextants are between keel and right inferior pulmonary vein;
- Lower sextants are below right inferior pulmonary vein.

Approximately, *with some reservations*, the lung may be considered to be divided into three equal lobes vertically and into two equal lobes horizontally. It was decided to determine the boundaries of the sextants by calculating the third part of the total area of the lungs. With this choice of boundaries, the sizes of the sextants will be approximately the same in area. In this case the boundaries of the sextant are approximate and not related to lung anatomy. The left lung is where the heart and left arm are located. The right lung is where the right arm is located.

Tab. II shows that the highest number of lung diseases was registered in the upper lung lobes and the lowest number of lung diseases was registered in the lower lung lobes. At the same time, on average, more diseases were registered in the right lung than in the left lung.

According to Radiopaedia ”post-primary infections have a strong predilection for the upper zones” [4]. Right versus left lesions asymmetry is not so obvious as top versus bottom and depends on the type (class) of the lesion. Statistical Atlas of Lung Lesions [5], [6], which was made by our team, is still in research. The difficulty in studying the statistical distribution of lesions is that they are of different nature and different biological substrates.

B. Analysis of sextant lesion names

According to the CASE BROWSER [2] website 20 lesion names are divided on the four groups (classes): ”Cavity”, ”Nodules”, ”Infiltrate” and ”Collapse” (see Tab. I).

Additional analysis of the sextant lesions list is shown in Fig. 5.

In Fig. 5, equally significant sections of text are highlighted with the same color. The 20 parameters are specially grouped in such a way that the same significant values are next to each other. Here are the definitions for these diseases.

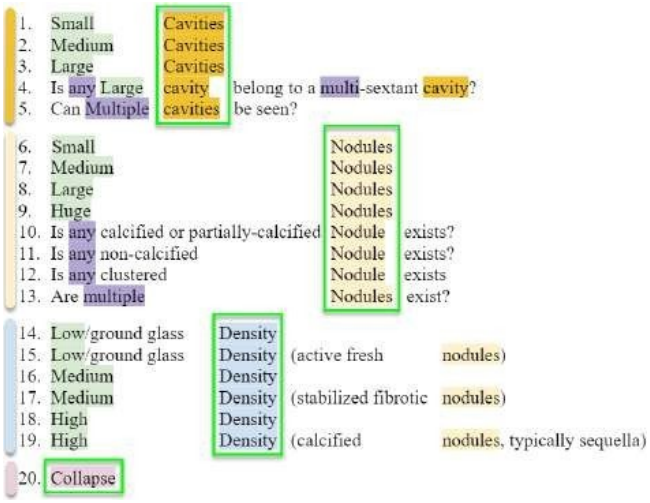

Figure 5: Graphical result of lesion names analysis for lung lobes.

A lung cavity or pulmonary *cavity* is an abnormal, thick-walled, air-filled space within the lung. Cavities in the lung can be caused by infections, cancer, autoimmune conditions, trauma, congenital defects, or pulmonary embolism. The most common cause of a single lung cavity is lung cancer. Bacterial, mycobacterial, and fungal infections are common causes of lung cavities [7].

According to the glossary of terms for chest imaging proposed by the Fleischner Society, a lung nodule is defined as an approximately rounded opacity more or less well-defined measuring up to 3 cm in diameter [8]. Rounded lesions measuring more than 3 cm in diameter are termed lung masses and should be considered indica- tive of lung cancer until histologically proven otherwise. Lung mass approach differs from that of nodules [9].

The *density* of the lung reflects the total mass of fluid, air, and dry lung tissue per unit volume of the lung. Lung density can be measured by evaluation of attenuation of an electron beam with CT. This technique has been shown to be sufficiently reliable and sensitive to distinguish normal from abnormal lung water [10].

A pulmonary *infiltrate* is a substance denser than air, such as pus, blood, or protein, which lingers within the parenchyma of the lungs. Pulmonary infiltrates are associated with pneumonia, tuberculosis, and sarcoido- sis [11].

A *collapsed* lung (pneumothorax) occurs when air escapes from the lung. The air then fills the space outside of the lung between the lung and chest wall. This buildup of air puts pressure on the lung, so it cannot expand as much as it normally does when you take a breath. The medical name of this condition is pneumothorax [12].

Four general categories (classes) of disease can be distinguished from the definitions above:

1) class ”Cavity” in the presence of one or more abnormal thick-walled spaces in the lung filled with air (gas);