Method of calculation of diagnostic signs based on the results of densitometry of spherical lung formations

I V Molodkin\textsuperscript{1,2}, A G Yakunin\textsuperscript{3} and A M Shayduk\textsuperscript{4}

\textsuperscript{1} Altai State University, Barnaul, Russia
\textsuperscript{2} Altai Regional Oncological Center, Barnaul, Russia
\textsuperscript{3} Polzunov Altai State Technical University, Barnaul, Russia
\textsuperscript{4} Altai State Medical University, Barnaul, Russia

tian.08@mail.ru

Abstract. The article presents a method for obtaining quantitative signs of pulmonary pathology, based on the analysis of medical imaging computed tomography. These signs may be suitable for the differential diagnosis of spherical lung formations. In particular, the possibility of diagnosing cancer and pulmonary tuberculosis was investigated by quantitative statistical characteristics of images of spherical lung formations. Two of these characteristics (the densitometric value and its standard deviation) can be called classical, since modern basic software of computer tomographs can easily calculate them. In addition, characteristics were considered, for which it is necessary to carry out additional calculations using fractal analysis. The article presents the results of a study of the relationship of statistical and fractal characteristics of a tomographic image with the type of pathology in the lungs. The results of a statistical analysis of the fractal dimension of the region of interest and its related parameters are presented, and signs are found suitable for the diagnosis of pulmonary pathologies.

1. Introduction
Doctors often use the descriptive characteristics of the tomographic image as a source of information about a possible diagnosis [1]. At the same time, the doctor mainly relies on their medical experience, besides, characteristics of the tomographic image are usually qualitative and not determined by a strict quantitative procedure. In this paper, we made an attempt to introduce some uniquely determined quantitative characteristics of the tomographic image and reveal which of them do carry statistical information about the pathology type and thus can serve as objective diagnostic characteristics in expert systems of assistance to the doctor.

2. Methods
In tomographic practice, the result of patient's examination is recorded in the DICOM format in files containing information about the density of each pixel in Hounsfield units and is called densitometric value (H):

\[
H = \frac{(\mu_T - \mu_{H,O})}{\mu_{H,O}} \cdot 1000, \quad (1)
\]
where \( \mu_{H_2O} \) – water absorption coefficient, \( \mu_T \) – biological tissue absorption coefficient in the observed pixel. After visualisation, the doctor analyses the collection of these data usually grouped by slices. Currently, modern computer tomographs are able to calculate the densitometric value for any selected area and even for a single pixel. Therefore, the densitometric value and its associated characteristics (mean-square deviation) can be attributed to classic characteristics of the medical image. The doctor can easily get this information, but it will not help in its original form. In order to carry out differential diagnostics, the results of statistical analysis for the confirmed dependence of the densitometric value of the studied diseases on the type of disease (cancer and tuberculosis) are necessary.

After the doctor has chosen an area of interest, one should introduce the quantitative characteristics of this area, which are actually some function of densities of all selected pixels [2].

The average density of tissue in the area of interest

\[
H_{av} = \frac{\sum_{i=1}^{n} H_i}{n},
\]

and standard deviation:

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (H_i - H_{av})^2}{n}},
\]

where \( n \) – number of pixels in this area, these are its obvious characteristics and are always displayed by standard medical image processing programs.

The fractal analysis was applied to solve the problem of searching for additional parameters [3, 4]. If the medical image has fractal properties, the logarithm of the measure linearly depends on the logarithms of the scale \( \varepsilon \) according to the formula:

\[
\ln M = \ln A + (2 - D) \ln \varepsilon,
\]

where \( D \) – fractal dimension. However, the study showed that the behavior of this dependence is not linear, if to take the entire range of scale change.

Therefore, in processing of medical images, we will use a function approximating the dependence of the logarithm of the measure on the logarithm of the scale in the form of \( y = a_1 x + a_2 x^2 \), where \( a_2 \) coefficient will denote the nonlinearity, and \( a_1 \) will denote the fractal behavior. The parameter for the nonlinearity of the dependence will be called "slope".

The computer program calculated all described characteristics of medical images automatically [5]. This program is written in Maple©. The work of the program results in the file containing all the characteristics described above for the corresponding slice. In this case, it was assumed that each slice of the same patient carries information about the disease, so in the work it is supposed that one can obtain information about the disease with each separate slice from the area of interest. If to accept this assumption, one can artificially increase the amount of analyzed data for tuberculosis and lung cancer. During data processing, it was discovered that the fractal dimension of the area of interest is calculated with a very high error if the size of the area of interest is less than \( n \) pixels. For this reason, areas of interest containing less than \( n \) pixels were not considered.

Data of tomographic studies of patients with lung cancer and tuberculosis, 640 and 1851 respectively, were taken as input data [6]. Images were obtained with the help of two computer tomographs without being eliminated by this factor, but even without taking into account the setting of these machines, the differences were recorded. Visualization of SLF was performed by the program MergeeFilmWorkstation [7]. Further, the program ImageJ converted the data to the txt file containing Hounsfield density numbers for a given patient, or more precisely, for each tomographic slice of this patient [8, 9].

At the first stage of the study, it was necessary to find out what the statistical distribution of diagnostic parameters is (2, 3). Standard analysis [10] was applied to solve this issue. The studies were
conducted for patients diagnosed with lung cancer and tuberculosis having spherical lung formations (SLF). The diseases were verified by Vladimir K. Konovalov, Doctor of Medical Sciences, Professor of the Altai State Medical University. These two diseases have visually similar images of the area of interest, but possibly different quantitative characteristics [11].

All quantitative parameters of medical images investigated in this work were checked for normality: densitometric value and its standard deviation, fractal dimension of the area of interest, "slope" function. To carry out statistical analysis, the computer program STATISTICA© was used [12]. This program has a wide range of statistical tools, in particular, those used by us.

As a null hypothesis, we assume that the samples studied belong to the general population. Thus, if the null hypothesis is not rejected, they will not be statistically different and cannot be used in the differential diagnosis of lung cancer and tuberculosis as diagnostic parameters.

3. Results and discussion

When carrying out statistical analysis of the studied distributions for compliance with the normal distribution, ambiguous results were obtained. Visually, the distributions obtained in the study resembled normal, as shown in Figure 1. It depicts one of the typical distributions of the studied characteristics.

![Figure 1. Check of the distribution of the densitometric value for normality for lung cancer (χ² criterion).](image)

Unfortunately, the visual similarity of the studied distributions with normal distributions was not confirmed by statistical analysis. The result of the statistical study is given in Table 1. The probability of an error of the first kind (α errors) in our case is almost always 0, which means a mismatch to the normal distribution of the studied distributions.

In this situation, it is difficult to assert about compliance of the studied characteristics with the normal distribution, so non-parametric analysis was applied to check the belonging of one general population.

Now we check whether these characteristics can be used for diagnostic purposes, i.e. whether they are statistically distinguishable for diseases of lung cancer and tuberculosis. As a null hypothesis, as
mentioned above, we assume that these parameters for these diseases are statistically indistinguishable, i.e. derived from one general population.

Table 1. Results of the study of the distribution of characteristics of medical images of SLF on compliance with normal distribution.

| Type of pathology | Densitometric value | Fractal dimension | MSD | "Slope" |
|-------------------|---------------------|-------------------|-----|---------|
|                   | Mean val. $\chi^2$ | Mean val. $\chi^2$ | Mean val. $\chi^2$ | Mean val. $\chi^2$ |
| Lung cancer       | 33.8 68.3 0         | 2.6 35.9 0.0      | 13.8 165.7 0        | -0.29 73.6 0.0 |
| Tuberculosis      | 30.7 27.8 0         | 2.5 7.8 0.1      | 14.6 134.2 0        | -0.31 20.7 0.01 |

To check the null hypothesis, the Mann–Whitney test was used for distributions that do not comply with a normal distribution with a critical significance level $p<0.05$. The results of the study on verification of the null hypothesis for the studied parameters are given in Table 2.

Table 2. Results of the study of the characteristics of medical images of SLF on the belonging to one general population.

| Parameters under study | $\alpha$ errors |
|------------------------|-----------------|
| Densitometric value    | 0.0             |
| MSD of densitometric value | 0.02        |
| Fractal dimension      | 0.0             |
| "Slope" function       | 0.56            |

4. Conclusion

The new method of calculation of quantitative parameters of SLF allows to distinguish two classical (densitometric value and its standard deviation) and two additional parameters (fractal dimension of the area of interest, "slope" function). The statistical analysis of the above quantitative characteristics of medical images allows to assert:

- The mean densitometric value of the area of interest is statistically different for lung cancer and pulmonary tuberculosis, which suggests that this quantitative parameter can be used in differential diagnosis. On average, the densitometric value for patients with lung cancer is higher than for patients with tuberculosis, and this difference is statistically significant.

- The standard deviation of the densitometric value of the area of interest has a probability of error of the first kind close to the level of critical significance. The obtained result does not give a sure answer to the question about the possibility of differential diagnosis of lung cancer and pulmonary tuberculosis on the basis of this parameter. Perhaps, the increase in the sample size will solve this issue.

- The mean value of the fractal dimension of the area of interest is different for patients with lung cancer and pulmonary tuberculosis. On average, the fractal dimension of the area of interest for patients with lung cancer is higher than for patients with tuberculosis, and this is a statistically significant difference.

- The mean value of the "slope" function in the area of interest is the same for patients with lung cancer and pulmonary tuberculosis. This parameter cannot be a significant diagnostic parameter.
It should be noted that the differences in these quantitative characteristics are statistical. For each individual patient, these characteristics may differ significantly from the average and the ratio between them may not be maintained.

References
[1] Ulin S E, Mikhajlov V N, Nikitaev V G et al. 2009 Physical methods of medical introscopy: study guide (M: MIFI) (In Russian)
[2] Belikova T P, Yashunskaya N I and Kogan E A 1992 Digital processing of tomograms and measuring of statistical features in the issue of early differential diagnostics of spherical lung formations (M: Inst-t of problems of inf. Transmission) (In Russian)
[3] Oczeretko E, Borowska M, Kitlas A et al. 2008 Fractal analysis of medical images in the irregular region of interest BioInformatics and BioEngineering, BIBE 2008. 8th IEEE International Conference
[4] Frederick E D 2001 Computer aided diagnosis of acute pulmonary embolism: Ph.D. – Durham.
[5] Vorona O I, Leonov S L, Konovalov V K and Shaiduk A M 17.04.2013 Certificate of state registration of computer program No. 2013613849
[6] Vorona O I 2013 Obtaining of parameters of computer images (In Russian)
[7] System eFilm Workstation 3.0, Merge Technologies Inc., USA, www.merge.com
[8] Aivazyan S A, Enukov I S and Meshalkin L D 1983 Applied statistics: bases of modeling and primary data processing (M.: Finances and statistics) (In Russian)
[9] Shilov B V and Engilevsky N A 2013 ImageJ Program for the study of medical and biological images LAP LAMBERT Academic Publishing (In Russian)
[10] Orlov A I 2004 Mathematics of occurrence: Probability and statistics: Study guide (M.: MH-Press) (In Russian)
[11] Konovalov V K, Lobanov M N, Leonov S L, Shaiduk A M, Kolmogorov V G and Dombrovsky A A 2013 Densitometry of spherical lung formations with the use of artificial neuron Bulletin of Altai Science 2-1 (In Russian)
[12] Borovikov V P 2013 Popular introduction to the modern data analysis in the STATISTICA system: Study guide for universities (M.: Hot line – Telecom) (In Russian)