*Pavel Troshin*[1]

# MULTIFRACTAL PARAMETRIZATION IN DIAGNOSIS OF LUNGS DISEASES

ABSTRACT. We consider the possibility of spatially asymmetric lungs diseases diagnosis by means of X-ray images processing. The method of multifractal parametrization is implemented. We adjust its parameters and construct a number of classifiers (pathology/normal) using the characteristics of multifractal spectrum.

## 1. INTRODUCTION

One cannot underestimate the problem of diseases detection and computer aided diagnosis. Among all of the lungs diseases (pneumonia, tuberculosis, tumor, emphysema, cancer etc.) and their diagnosis methods (fluorography, radiography, bronchography, lungs volume and contents analysis, pulmonary function tests etc.), we have chosen the most prevalent symptoms: fibrosis and infiltrations viewed on X-ray images.

Such diseases are usually marked by clustering of pocket spots and branching structure of propagation. Therefore they present *fractal* subsets (self-similar, repeating themselves on a number of scales) and in most cases are distributed with varying density, which let us study them as *multifractals*. The features of multifractal are described numerically by its multifractal (MF) spectrum — a set of dimensions of its fractal subsets.

In theory, by means of MF spectrum one should be able to distinguish the presence of pathology with branching structure and estimate its heaviness. In practice however, the second-rate accuracy when calculating MF spectrum (by box-counting method) and difficulties of its analysis impede the task of classification. Certain success was made by different authors using various MF techniques [1]–[5], mostly studying computed tomography (CT) scans.

In this paper we develop simple classifiers for detection of spatially asymmetric lungs diseases, using a suitable set of MF spectrum characteristics for the task of image recognition and

[1]Institute of Mathematics and Mechanics, Kazan Federal University, 420008, Kremlyovskaya str. 35, room 606, Kazan, Russia (*Paul.Troshin@gmail.com*).

--------

classification. We use fluorographic X-ray images which are way simpler to obtain than CT, albeit rougher and contain less information.

In Section 2, basics of MF image parametrization are described, together with the certain selected numerical values of this method.

In Section 3, the development of new classifiers is discussed. We present the results of processing the training set of X-ray images and propose a set of metrics on images, involving their MF spectrum. These metrics serve as descriptors that we use to construct some classifiers for lungs pathologies. For each classifier we find its optimal parameters.

## 2. Multifractal approach to image processing

2.1. **Multifractal formalism.** Here we state briefly the basics of MF formalism (see further [6]–[11]). Let $M \subset L \times L \subset \mathbb{R}^2$, $L > 0$, be a non-empty compact subset of a plane. And let $\mu$ be a finite measure supported on $M$. We divide this set $M$ into a grid $G_\varepsilon$ of square cells $M_i$, $i \in N(G_\varepsilon) \subset \mathbb{N}$, with linear size $\varepsilon \in (0, L)$, and denote $\mu_i = \mu(M_i)$.

We define *general correlation function*

$$\chi(q) = \sum_{i \in N(G_\varepsilon)} \mu_i^q, \qquad q \in \mathbb{R},$$

and its exponent

$$(1) \qquad\qquad \tau(q) = \lim_{\varepsilon \to 0} \frac{\ln \chi(q)}{\ln \varepsilon}, \qquad q \in \mathbb{R}.$$

Indeed $\chi(q) \approx \varepsilon^{\tau(q)}$ whenever $\varepsilon$ is small. We can consider the existence of limit in (1) as the necessary property of multifractal.

Parameter $q$ controls the contribution of areas with small and big measures into the sum $\chi(q)$: when $q \gg 1$, the cells with relatively big values $\mu_i$ play the main role in this sum, and when $q \ll -1$, the cells with smaller $\mu_i$ get important. Thus $q$ is essentially a kind of scaling factor (a "lens"), which reveals the distribution properties of homogenous subsets in $M$.

Consider the following functions:

$$(2) \qquad\qquad \alpha(q) = \frac{d\,\tau(q)}{d\,q}, \qquad q \in \mathbb{R},$$

$$(3) \qquad\qquad f(q) = q\frac{d\,\tau(q)}{d\,q} - \tau(q), \qquad q \in \mathbb{R}.$$

Equations (2)–(3) define parametrically a curve $f = f(\alpha)$ of fractal dimensions ($f(\alpha)$-*spectrum* and $f(q)$-*spectrum*). The value $f(\alpha)$ presents fractal dimension of the homogenous subset $M_\alpha \subset M$ with the same singularity index $\alpha$ of cell measures $\mu_i \approx \varepsilon^\alpha$.

Its also useful to consider *generalized Rényi dimensions*:

$$D_q = \frac{\tau(q)}{q-1} = \frac{1}{q-1}(q\alpha(q) - f(q)), \qquad q \in \mathbb{R}.$$

By MF spectrum we denote $f(\alpha)$-spectrum together with Rényi dimensions.

2.2. **Multifractal parametrization.** By means of some standardized procedure of X-ray photography followed by image preprocessing (see [12]), we have got a digital image: a matrix $M$ sized $n \times n$, $n \in \mathbb{N}$ (in our case $n = 512$), with each element $z_{pq} \in M$ being a function of color of the corresponding pixel of the X-ray. In our case $z_{pq} \in [0,1]$ which corresponds to gray scale (black color for dense structures (bones, blood vessels, and pathological formations), white color for thinner matter (lungs and cavities in them)).

In order to find MF characteristics of digital image, we need to examine it on different scales, dividing into the net $G_k$ of cells with linear size $l_k \in scales \subset [1,n] \cap \mathbb{N}$, $k = 1, \ldots, m$, $m \in \mathbb{N}$. The set of scales often defines the look of graphs we get. Besides, each scale is in charge for those image details commensurable with this scale, letting us to pick out some specific scales (in our case $\{4, 8, 16, 32, 64, 128\}$).

Knowing the color value $z_{pq}$ we can define the pixel measure $\mu_{0pq} = z_{pq}/ \sum\limits_{z_{pq}\in M} z_{pq}$ and extend it naturally to finite measure $\mu$ on subsets of $M$.

To each net cell $M_i$ of size $l_k \times l_k$ we assign the measure

$$\mu_{ik} = \frac{1}{\sum\limits_{z_{pq}\in M} z_{pq}} \sum_{z_{pq}\in M_i} z_{pq} = \sum_{z_{pq}\in M_i} \mu_{0pq},$$

the sum of color function over all pixels in this cell, normalized by the measure of the whole set $M$.

For estimation of MF characteristics we need to know the values of $\alpha(q)$ and $f(q)$, yet the implementation of formula (3) for this is inconvenient. We will use the following trick (see [6]). Consider a net $G_\varepsilon$, $\varepsilon > 0$, and the quantities

$$A_\varepsilon(q) = \sum_{i\in N(G_\varepsilon)} \frac{\mu_i^q \ln \mu_i}{\chi(q)} = \left( \sum_{i\in N(G_\varepsilon)} \mu_i^q \ln \mu_i \right) / \sum_{i\in N(G_\varepsilon)} \mu_i^q,$$

$$F_\varepsilon(q) = \sum_{i\in N(G_\varepsilon)} \frac{\mu_i^q \ln \left( \frac{\mu_i^q}{\chi(q)} \right)}{\chi(q)} = \left( \sum_{i\in N(G_\varepsilon)} \mu_i^q \ln \left( \frac{\mu_i^q}{\sum\limits_{i\in N(G_\varepsilon)} \mu_i^q} \right) \right) / \sum_{i\in N(G_\varepsilon)} \mu_i^q.$$

Omitting the justification of passing to the limit, one can derive

$$(4) \quad \alpha(q) = \frac{\mathrm{d}\,\tau(q)}{\mathrm{d}\,q} = \frac{\mathrm{d}\left( \lim\limits_{\varepsilon\to 0} \frac{\ln \chi(q)}{\ln \varepsilon} \right)}{\mathrm{d}\,q} = \lim_{\varepsilon\to 0} \frac{1}{\ln \varepsilon} \left( \frac{\mathrm{d}\ln \chi(q)}{\mathrm{d}\,q} \right) = \lim_{\varepsilon\to 0} \frac{1}{\ln \varepsilon} \sum_{i\in N(G_\varepsilon)} \frac{\mu_i^q \ln \mu_i}{\chi(q)} =$$

$$\lim_{\varepsilon\to 0} \frac{A_\varepsilon}{\ln \varepsilon},$$

| | |
|---|---|
| $D_0 = f(q = 0)$ | Fractal dimension of $M = \operatorname{supp} \mu$. |
| $D_1 = \alpha(q = 1)$ | Informational dimension of $M$. |
| $D_2 = 2\alpha(q = 2) - f(q = 2)$ | Correlation dimension of $M$. |
| $D_{-\infty} = D_q(q_{\min})$ $D_{+\infty} = D_q(q_{\max})$ | The bigger are $D_{-\infty}$ and $D_{+\infty}$, the less dense is $M$. |
| $K = D_{-\infty} - D_{+\infty}$ | Describes the amplitude of singularity indices: the bigger is $K$, the less homogeneous is $M$. |
| $f_{-\infty} = f(q_{\min})$ | The bigger is $f_{-\infty}$, the more homogeneous is $M$. Corresponds to fractal dimension of the most thin homogeneous subset in $M$. |
| $f_{+\infty} = f(q_{\max})$ | The bigger is $f_{+\infty}$, the more homogeneous is $M$. Corresponds to fractal dimension of the most dense homogeneous subset in $M$. |
| $\triangle_\infty = D_1 - D_{+\infty}$ | Points to the limit of inner symmetry violation, regarding MF transformation of measure $\mu_i \mapsto \frac{\mu_i^q}{\chi(q)}$. Also points to the measure of order in $M$, and the degree of equilibrium of its structure. |
| $\sigma_{\mathrm{err}\ -\infty} = \sigma_{\mathrm{err}}(q_{\min})$ $\sigma_{\mathrm{err}\ +\infty} = \sigma_{\mathrm{err}}(q_{\max})$ | Mean squared errors when estimating $f(q_{\min})$ and $f(q_{\max})$. |

TABLE 1. Informational content of MF characteristics (see [6]).

$$(5) \quad f(q) = q\frac{\mathrm{d}\,\tau(q)}{\mathrm{d}\,q} - \tau(q) = q\frac{\mathrm{d}\left(\lim\limits_{\varepsilon\to 0}\frac{\ln\chi(q)}{\ln\varepsilon}\right)}{\mathrm{d}\,q} - \lim_{\varepsilon\to 0}\frac{\ln\chi(q)}{\ln\varepsilon} = \lim_{\varepsilon\to 0}\frac{1}{\ln\varepsilon}\left(q\frac{\mathrm{d}\ln\chi(q)}{\mathrm{d}\,q} - \ln\chi(q)\right) =$$

$$\lim_{\varepsilon\to 0}\frac{1}{\ln\varepsilon}\frac{q\sum\limits_{i\in N(G_\varepsilon)}\mu_i^q\ln\mu_i - \sum\limits_{i\in N(G_\varepsilon)}\mu_i^q\ln\chi(q)}{\chi(q)} = \lim_{\varepsilon\to 0}\frac{1}{\ln\varepsilon}\frac{\sum\limits_{i\in N(G_\varepsilon)}\mu_i^q(\ln\mu_i^q - \ln\chi(q))}{\chi(q)} = \lim_{\varepsilon\to 0}\frac{F_\varepsilon}{\ln\varepsilon},$$

and then the values $\alpha(q) = \lim\limits_{\varepsilon\to 0}\frac{\partial A_\varepsilon}{\partial \ln\varepsilon}$ and $f(q) = \lim\limits_{\varepsilon\to 0}\frac{\partial F_\varepsilon}{\partial \ln\varepsilon}$ (L'Hospital rule) are obtained as the slopes of linear regressions of the sets $(\ln\varepsilon, A_\varepsilon)$ and $(\ln\varepsilon, F_\varepsilon)$ correspondingly. Computations (4)–(5) are dependent on the quotient $\varepsilon/L$, so that at another scale $\varepsilon' = K\varepsilon$, $K > 0$, we have $\partial\ln\varepsilon' = \partial(\ln\varepsilon + \ln K) = \partial\ln\varepsilon$, and $\alpha(q)$ and $f(q)$ stay the same. Therefore instead of $\varepsilon$ we will use $scales = \{l_k\}_{k=1}^m$ for linear approximations of points $\{(\ln l_k, A_{l_k})\}_{k=1}^m$ and $\{(\ln l_k, F_{l_k})\}_{k=1}^m$, and take $\mu_i = \mu_{ik}$ for each $l_k \in scales$.

Continuous spectrum of $q \in \mathbb{R}$ should also be sampled. We chose to take $q \in Q = \{0, \pm 1, \ldots, \pm 70\}$, and then in each particular classifier we specialize it to a lesser subset $\{q\}_{q=q_{\min}}^{q_{\max}} \subset Q$.

It is customary to pick out several characteristics of $f(q)$-spectrum, those having understandable informational sense, see Table 1.
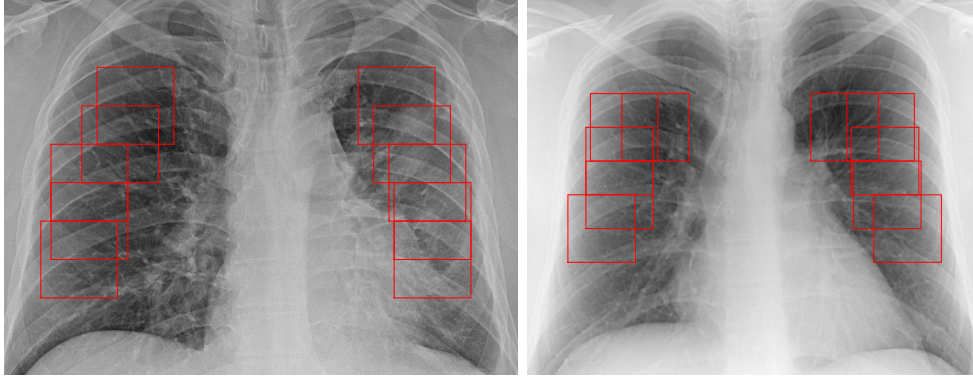
FIGURE 1. Cutting out symmetrical square ROIs (lungs with pathology (left) and normal (right)).
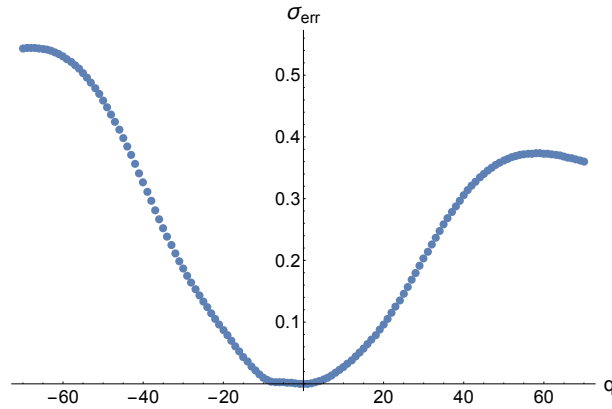


FIGURE 2. Typical graph of mean squared error of regression when estimating $f(q)$.

## 3. DATA ANALYSIS AND AN APPROACH TO CLASSIFICATION

In this paper, we restrain ourselves to diagnosis of horizontally asymmetrical pathologies of lungs (with reference to spine on x-ray images), in which case one of the lungs is damaged more than the other.

Our goal is to find suitable characteristics of MF spectrum and adjust the method of MF parametrization, in order to diagnose asymmetrical pathology by a given pair of symmetrically cut regions of interest (ROIs).

For this purpose, we use the training set of 55 pairs of ROIs with asymmetric pathologies (the set $P$) and 51 pairs of ROIs from healthy lungs (the set $H$) (see Fig. 1), and denote corresponding statistical populations by $P' \supset P$ and $H' \supset H$ (the X-ray database of republic clinical antitubercular health center, Tatarstan, Russia). We assume the presence of scanning procedure generating admissible ROIs, which lay inside pulmonary fields (see [13] for spotting the features on X-rays of lungs).

Let $P_{\text{ill}}$ be a set of all ROIs with pathologies taken from the pairs in $P$, and $(P \cup H)_{\text{norm}}$ be a set of ROIs without pathologies taken from the pairs in $P$ and $H$.

We have chosen empirically the following values for MF parametrization:

- $n \times n = 512 \times 512$ — pixel size of each ROI;
- $scales = \{4, 8, 16, 32, 64, 128\}$ — pixel sizes of scales;
- $Q = \{-70, -69, \ldots, 69, 70\}$ — sample of the set $\mathbb{R} \ni q$.

We can notice that the mean squared error $\sigma_{\text{err}}(q)$ of linear regression when estimating $f(q)$ at each $q \in Q$ for every sample in $P_{\text{ill}} \cup (P \cup H)_{\text{norm}}$ growths when $|q|$ is large (see Fig. 2). Thus the result of calculating $f(q)$ (and $\alpha(q)$ as well) becomes less reliable. Therefore we will adjust $Q \subset \{-70, \ldots, 70\}$ for each classifier in order to maximize the *specificity* (see further) of the method.

To each image $I \in P_{\text{ill}} \cup (P \cup H)_{\text{norm}}$ we can assign a set of its MF characteristics (see Table 1):

$$I \mapsto (D_0, D_1, D_2, D_{-\infty}, D_{+\infty}, K, f_{-\infty}, f_{+\infty}, \triangle_\infty, \sigma_{\text{err}\ -\infty}, \sigma_{\text{err}\ +\infty}).$$

Statistical properties of MF spectrum of ROIs in $P_{\text{ill}} \cup (P \cup H)_{\text{norm}}$ are given in the Table 2 (left). We can derive, that $D_0$ is nearly constant and can be excluded from our consideration. Besides, the differences in values in $P_{\text{ill}}$ and in $(P \cup H)_{\text{norm}}$ are not big, which obstructs us to use them for classification by one image (not by a pair).

Let $I = (I', I'') \in P \cup H$ be a pair of ROIs (the order is not important since we do not know which side of lungs contains pathology). Consider the following set of metrics $d_s$, $s \in S = \{1, \ldots, 9\}$, (chosen from the bigger amount by maximal values of *sensitivity* and *specificity* when changing $Q \subset \{-70, \ldots, 70\}$):

$$d_1 = \left((D_1' - D_1'')^2 + (D_2' - D_2'')^2 + (D_{-\infty}' - D_{-\infty}'')^2 + (D_{+\infty}' - D_{+\infty}'')^2 + (K' - K'')^2 + \right.$$
$$\left. (f_{-\infty}' - f_{-\infty}'')^2 + (f_{+\infty}' - f_{+\infty}'')^2 + (\triangle_\infty' - \triangle_\infty'')^2\right)^{1/2},$$

$$d_2 = |D_1' - D_1''| + |D_2' - D_2''| + |D_{-\infty}' - D_{-\infty}''| + |D_{+\infty}' - D_{+\infty}''| + |K' - K''| +$$
$$|f_{-\infty}' - f_{-\infty}''| + |f_{+\infty}' - f_{+\infty}''| + |\triangle_\infty' - \triangle_\infty''|,$$

$$d_3 = \sqrt{\sum_{q=q_{\min s}}^{q_{\max s}} (D_q' - D_q'')^2 / (q_{\max s} - q_{\min s} + 1)},$$

$$d_4 = \sqrt{\sum_{q=q_{\min s}}^{q_{\max s}} (f'(q) - f''(q))^2 / (q_{\max s} - q_{\min s} + 1)},$$

$$d_5 = |\sigma_{\text{err}\ -\infty}' - \sigma_{\text{err}\ -\infty}''| + |\sigma_{\text{err}\ +\infty}' - \sigma_{\text{err}\ +\infty}''|,$$

$$d_6 = |\triangle_\infty' - \triangle_\infty''|, \qquad d_7 = |f_{+\infty}' - f_{+\infty}''|, \qquad d_8 = |D_{+\infty}' - D_{+\infty}''|, \qquad d_9 = |K' - K''|.$$

These metrics are new descriptors of the pairs $I$:

$$I = (I', I'') \mapsto (d_1 = d_1(I', I''), \ldots, d_9 = d_9(I', I'')).$$

| | | E | $\sigma$ |
|---|---|---|---|
| $D_0$ | $P_{\text{ill}}$ | 2.147 | 0.000 |
| | $(P \cup H)_{\text{norm}}$ | 2.147 | 0.000 |
| $D_1$ | $P_{\text{ill}}$ | 2.146 | 0.001 |
| | $(P \cup H)_{\text{norm}}$ | 2.146 | 0.001 |
| $D_2$ | $P_{\text{ill}}$ | 2.144 | 0.002 |
| | $(P \cup H)_{\text{norm}}$ | 2.144 | 0.002 |
| $D_{-\infty}$ | $P_{\text{ill}}$ | 2.245 | 0.044 |
| | $(P \cup H)_{\text{norm}}$ | 2.239 | 0.046 |
| $D_{+\infty}$ | $P_{\text{ill}}$ | 2.119 | 0.013 |
| | $(P \cup H)_{\text{norm}}$ | 2.126 | 0.007 |
| $K$ | $P_{\text{ill}}$ | 0.126 | 0.050 |
| | $(P \cup H)_{\text{norm}}$ | 0.113 | 0.050 |
| $f_{-\infty}$ | $P_{\text{ill}}$ | 0.381 | 0.333 |
| | $(P \cup H)_{\text{norm}}$ | 0.471 | 0.338 |
| $f_{+\infty}$ | $P_{\text{ill}}$ | 1.343 | 0.314 |
| | $(P \cup H)_{\text{norm}}$ | 1.570 | 0.227 |
| $\triangle_\infty$ | $P_{\text{ill}}$ | 0.028 | 0.013 |
| | $(P \cup H)_{\text{norm}}$ | 0.021 | 0.007 |
| $\sigma_{\text{err } -\infty}$ | $P_{\text{ill}}$ | 0.309 | 0.132 |
| | $(P \cup H)_{\text{norm}}$ | 0.283 | 0.105 |
| $\sigma_{\text{err } +\infty}$ | $P_{\text{ill}}$ | 0.205 | 0.113 |
| | $(P \cup H)_{\text{norm}}$ | 0.175 | 0.116 |

| | | E | $\sigma$ |
|---|---|---|---|
| $d_1$ | P | 0.555 | 0.291 |
| | H | 0.439 | 0.269 |
| $d_2$ | P | 0.783 | 0.378 |
| | H | 0.597 | 0.327 |
| $d_3$ | P | 0.021 | 0.014 |
| | H | 0.015 | 0.012 |
| $d_4$ | P | 0.327 | 0.164 |
| | H | 0.270 | 0.158 |
| $d_5$ | P | 0.244 | 0.119 |
| | H | 0.172 | 0.123 |
| $d_6$ | P | 0.015 | 0.012 |
| | H | 0.004 | 0.003 |
| $d_7$ | P | 0.400 | 0.285 |
| | H | 0.193 | 0.151 |
| $d_8$ | P | 0.015 | 0.012 |
| | H | 0.004 | 0.003 |
| $d_9$ | P | 0.039 | 0.029 |
| | H | 0.026 | 0.021 |

TABLE 2. Statistics for MF characteristics (left) and for metrics involving them (right). Here E is a mean value over the corresponding sample, and $\sigma$ is a standard deviation.

Let $C_s = [\min_{I \in P \cup H} d_s(I), \max_{I \in P \cup H} d_s(I)] \subset \mathbb{R}$ be a range of values of descriptor $d_s$, $s \in S$.

We can notice from the statistical properties of given metrics (see Table 2, right) that the descriptors of healthy pairs are slightly smaller and stay closer to 0 than those of pairs with pathologies. This justifies using a simple classifier by two areas.

We consider the simplest 2-dimensional classifier with parameters $s = (s_1, s_2) \in S = \{(s_1, s_2) \mid s_1, s_2 \in \{1, \ldots, 9\}, s_1 < s_2\}$ and $c = (c_1, c_2) \in C_s = C_{s_1} \times C_{s_2}$:

$$(6) \qquad I \in P' \cup H' \mapsto \begin{cases} P', & \text{if } d_{s_1} \geq c_1 \ \text{ or } \ d_{s_2} \geq c_2 \quad \text{(classified as pathology)}, \\ H', & \text{if } d_{s_1} < c_1 \ \text{ and } \ d_{s_2} < c_2 \quad \text{(classified as normal)}. \end{cases}$$

As an additional parameter we will use the set $Q_s = \{q\}_{q=q_{\min s}}^{q_{\max s}} \subset Q$, used for estimation of MF spectrum, with possibility to vary $q_{\min s}$ and $q_{\max s}$ to obtain the best results.

We need two indices of effectiveness for our classification method, *sensitivity* (true positive rate) and *specificity* (true negative rate):

| $(d_{s_1}, d_{s_2})$ | $(d_6, d_1)$ | $(d_2, d_3)$ | $(d_7, d_6)$ | $(d_1, d_3)$ | $(d_8, d_7)$ |
|---|---|---|---|---|---|
| $Q_s$ | {-6,…,70} | {0,…,70} | {-42,…,46} | {0,…,70} | {-50,…,70} |
| $C_s$ | (0.005,0.378) | (0.312,0.005) | (0.144,0.006) | (0.293,0.005) | (0.005,0.381) |
| $\mathrm{SNS}_s$ | 0.909 | 0.909 | 0.909 | 0.909 | 0.909 |
| $\mathrm{SPC}_s$ | 0.804 | 0.784 | 0.765 | 0.765 | 0.765 |
| $(d_{s_1}, d_{s_2})$ | $(d_8, d_5)$ | $(d_8, d_4)$ | $(d_9, d_4)$ | $(d_7, d_3)$ | $(d_6, d_5)$ |
| $Q_s$ | {-50,…,42} | {-6,…,70} | {-2,…,46} | {-10,…,34} | {-42,…,34} |
| $C_s$ | (0.005,0.248) | (0.005,0.171) | (0.005,0.167) | (0.105,0.005) | (0.005,0.223) |
| $\mathrm{SNS}_s$ | 0.909 | 0.909 | 0.909 | 0.909 | 0.909 |
| $\mathrm{SPC}_s$ | 0.745 | 0.745 | 0.725 | 0.706 | 0.706 |

TABLE 3. Best 2-dimensional linear classifiers.

$$\mathrm{SNS} = \frac{\#\{I \in P \mid I \text{ classified as belonging to P}\}}{\#(P)},$$

$$\mathrm{SPC} = \frac{\#\{I \in H \mid I \text{ classified as belonging to H}\}}{\#(H)},$$

where $\#(\cdot)$ is a cardinality of a set. In case of classifier (6), for $s \in S, c \in C_s$ it turns to be

$$\mathrm{SNS}_s(c) = \frac{\#\{I \in P \mid d_{s_1} \geq c_1 \text{ or } d_{s_2} \geq c_2\}}{\#(P)},$$

$$\mathrm{SPC}_s(c) = \frac{\#\{I \in H \mid d_{s_1} < c_1 \text{ and } d_{s_2} < c_2\}}{\#(H)}.$$

Since in mass fluorographic (screening) examination, pulmonary diseases are relatively rarely detected (4 patients in 1000), we set high value of $\gamma = 0.9$ as the lower limit of sensitivity and search for classifiers with maximal specificity.

In order to find the most effective pairs of metrics $(d_{s_1}, d_{s_2})$ we solved numerically the following task:

(7) $$\hat{s} = \arg\max_{s \in S} \max_{Q_s \subset Q} \max_{c \in C_s} \{\mathrm{SPC}_s(c) \mid \mathrm{SNS}_s(c) \geq \gamma\}.$$

**Result.** *In the Table 3 we present* 10 *best solutions (classifiers* (6)*) to the problem* (7)*. And on the Fig. 3 we depict the best classification rule* $(d_6, d_1)$ *with parameters* $Q_{(6,1)} = \{-6, \ldots, 70\}$, $C_{(6,1)} = (0.005, 0.378)$ *and efficiency indices* $\mathrm{SNS}_{(6,1)} = 0.909$, $\mathrm{SPC}_{(6,1)} = 0.804$ *(red points for samples from* $I \in P$, *and blue ones for* $I \in H$*).*

Existence of outliers in training set is mostly due to irregularity of X-ray image exposure, and image asymmetry caused by the body structure and ineffectual angle of projection.

As it turned up, solutions to the problem (7) for 1- and 3-dimensional classifiers give worse results ($\mathrm{SPC}_s < 0.72$).
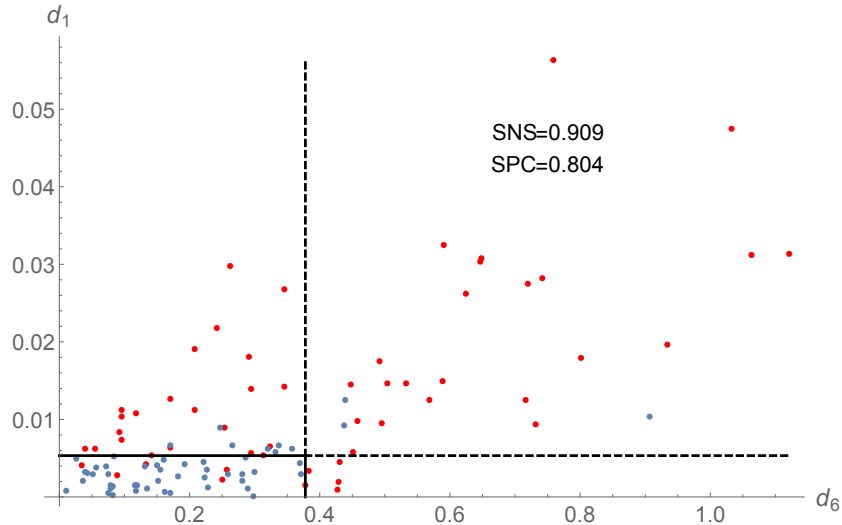
FIGURE 3. Best classifier with descriptors $(d_6, d_1)$.

## 4. CONCLUSION AND ACKNOWLEDGEMENTS

Binary classification method (pathology/normal) proposed here on practice can be combined with another image recognition methods. We have confirmed the hypothesis of possibility to use MF characteristics for the given problem of classification.

Effectively analogous classifiers one can obtain from metrics $d_s$, using the relevant vector machine method to separate areas on the graphs like Fig. 3. It is also interesting to implement Mahalanobis distance for this purpose.

Often, pathological structure in lungs has multifractal properties. The existence of pathology should come out in multifractal spectrum of the X-ray image. Here we have studied diagnosis by a pair of symmetrically cut ROIs, and we leave the question of analogous classification by a single ROI for further research.

## REFERENCES

[1] *Rocchisani J.M., Lévy-Vehel J., Paycha F.*, Multifractal Measurements of Lung Diseases. First International Conference on Fractals in Biology and Medicine, Ascona, Switzerland, 1993.

[2] *Hemsley A., Mukundan R.*, Multifractal measures for tissue image classification and retrieval. In Proceedings of the IEEE International Symposium on Multimedia, 2009, pp. 618–623.

[3] *Musibau I., Mukundan R.*, Multifractal techniques for emphysema classification in lung tissue images. 3rd International conference on environment, chemistry and biology IPCBEE, 2014, vol. 78.

[4] *Tay C., Mukundan R., Racoceanu D.*, Multifractal Analysis of Histopathological Tissue Images. Auckland, New Zealand: 26th International conference Image and Vision Computing New Zealand (IVCNZ 2011), 29 Nov – 1 Dec 2011, pp. 80–85.

[5] *Sørensen L., Shaker S.B., de Bruijne M.*, Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Transactions on Medical Imaging, 2010, vol. 29(2), pp. 559–569.

[6] *Vstovsky, G.V., Kolmakov, A.G., Bunin, I.J.*, Introduction to multifractal parametrization of materials structures. Moscow-Izhevsk: Regular and Chaotic Dynamics, 2001.

[7] *Vstovsky, G.V.*, Elements of information Physics. Moscow, Edition-issuing center of Moscow state industrial university, 2002.

[8] *de Souza J., Rostirolla S.P.*, A fast MATLAB program to estimate the multifractal spectrum of multidimensional data: Application to fractures. Computers & Geosciences, vol. 37, 2011, pp. 241—249.

[9] *Reljin I., Reljin B., Pavlovic I.*, Multifractal analysis of grayscale images. In Proceedings of the Mediterranean Electro-technical Conference, 2000, vol. 2, pp. 490—493.

[10] *Reljin I.S., Reljin B.D.*. Fractal geometry and multifractals in analyzing and processing medical data and images. Archive of Oncology, 2002, vol. 10(4), pp. 283–293.

[11] *Véhel J.L., Mignot P.*, Multifractal Segmentation of Images. Fractals, 1994, vol. 2(3), pp. 371–378.

[12] *Zakirov A.N., Kuleev R.F., Timoshenko A.S., Vladimirov A.V.* Advanced approaches to computer-aided detection of thoracic diseases on chest X-rays. Applied Mathematical Sciences, vol. 9, 2015, no. 88, pp. 4361–4369.

[13] *Belhaouair S.B., Kuleev R.F.* On a new approach to the automated detection of thoracic organs diseases using the spot feature in the analysis of digital X-ray images. Applied Mathematical Sciences, vol. 8, 2014, no. 164, pp. 8171–8177.