

DOI: 10.5604/01.3001.0013.2541

THE CONSTRUCTION OF THE FEATURE VECTOR IN THE DIAGNOSIS OF SARCOIDOSIS BASED ON THE FRACTAL ANALYSIS OF CT CHEST IMAGES

Zbigniew Omiołek, Paweł Prokop

Lublin University of Technology, Faculty of Electrical Engineering and Computer Science/Institute of Electronics and Information Technology

Abstract. CT images corresponding to the cross-sections of the patients' upper torso were analysed. The data set included the healthy class and 3 classes of cases affected by sarcoidosis. It was a state involving only the trachea – Sick(1), a state including trachea and lung parenchyma – Sick(2) and a state involving only lung parenchyma – Sick(3). Based on a fractal analysis and a feature selection by linear stepwise regression, 4 descriptors were obtained, which were later used in the classification process. These were 2 fractal dimensions calculated by the variation and box counting methods, lacunarity calculated also with the box counting method and the intercept parameter calculated using the power spectral density method. Two descriptors were obtained as a result of a gray image analysis, and 2 more were the effect of a binary image analysis. The effectiveness of the descriptors was verified using 8 popular classification methods. In the process of classifier testing, the overall classification accuracy was 90.97%, and the healthy cases were detected with the accuracy of 100%. In turn, the accuracy of recognition of the sick cases was: Sick(1) – 92.50%, Sick(2) – 87.50% and Sick(3) – 90.00%. In the classification process, the best results were obtained with the support vector machine and the naive Bayes classifier. The results of the research have shown the high efficiency of a fractal analysis as a tool for the feature vector extraction in the computer aided diagnosis of sarcoidosis.

Keywords: fractals, sarcoidosis, computed tomography, image texture analysis

KONSTRUKCJA WEKTORA CECH W DIAGNOSTYCE SARKOIDOZY NA PODSTAWIE ANALIZY FRAKTALNEJ OBRAZÓW CT KLATKI PIERSIOWEJ

Streszczenie. Przeprowadzono analizę obrazów CT górnej części tułowia pacjentów. Zbiór danych zawierał klasę pacjentów zdrowych i 3 klasy przypadków dotkniętych sarkoidozą. Był to stan obejmujący tylko tchawicę – Sick(1), stan obejmujący tchawicę i miąższ płucny – Sick(2) i stan obejmujący tylko miąższ płucny – Sick(3). Na podstawie analizy fraktalnej oraz selekcji cech metodą liniowej regresji krokowej otrzymano 4 deskrytory, które później wykorzystano w procesie klasyfikacji. Były to 2 wymiary fraktalne obliczone za pomocą metod *variation* i *box counting*, *lacunarity* obliczona również za pomocą metody *box counting* oraz parametr *intercept* obliczony za pomocą metody *widmowej gęstości mocy*. W wyniku analizy obrazu szarego otrzymano 2 deskrytory, a 2 kolejne były efektem analizy obrazu binarnego. Skuteczność deskrytorów zweryfikowano za pomocą 8 popularnych metod klasyfikacji. W procesie testowania klasyfikatorów, ogólna dokładność klasyfikacji wyniosła 90,97%, a przypadki zdrowe wykryto z dokładnością 100%. Z kolei, dokładność rozpoznania przypadków chorych była następująca: Sick(1) – 92,50%, Sick(2) – 87,50% i Sick(3) – 90,00%. W procesie klasyfikacji, najlepsze wyniki uzyskano za pomocą maszyny wektorów nośnych i naiwnego klasyfikatora Bayesa. Wyniki badań pokazały wysoką skuteczność analizy fraktalnej jako narzędzia do ekstrakcji wektora cech w komputerowej diagnostyce sarkoidozy.

Słowa kluczowe: fraktale, sarkoidoza, tomografia komputerowa, analiza tekstury obrazu

Introduction

Sarcoidosis is a disease that attacks the body's immune system. A characteristic feature of sarcoidosis is a presence of inflammatory papules, called granulomas, which are on various organs and are not absorbed [1]. Sarcoidosis mainly affects the lungs and lymph nodes of the cavities, which are located in the mediastinum, i.e. in the part of the body which is located within the chest protected by the bony armor created by the ribs and the sternum [2, 9]. As a part of diagnostics of the disease, lung imaging is primarily performed. On this basis, there are 5 stages of sarcoidosis on which the treatment depends. However, to confirm the diagnosis, morphological examinations are performed, allowing the presence of characteristic granulomas to be visible under the microscope [7, 27]. Neglect of treatment of sarcoidosis can lead to respiratory failure and irreversible changes in the lungs, which is why it is very important to detect the disease at an early stage. The use of the computer aided diagnosis system, which based on CT images would provide the physician with a second diagnostic opinion, could be of great benefit here. In the construction of such a system, a fractal analysis could be used. The results we present in the article, confirm the high effectiveness of this method in identifying of healthy cases and different categories of sick cases. It should be noted here, that we are not aware of the works of other authors, where CT images would be used in the computer aided diagnosis system for detecting of sarcoidosis.

1. Material and methods

1.1. Material

CT images corresponding to the cross-sections of the patients' upper torso have been used. The data set included the class of healthy cases and 3 classes of cases diagnosed with sarcoidosis.

These 3 kind of sick cases included only a trachea, a trachea and a pulmonary parenchyma, and only a lung parenchyma. The ground truth of whether a given CT image belongs to the healthy or sick person was determined by a pulmonologist (an expert in the field) based on clinical tests. Each category contained images belonging to 15 patients. In most cases, 4 areas of interest (ROIs) were identified for each patient (2 for each lung). ROIs were selected manually and their size of 80x150 pixels was determined in such a way that they would present the lesion area as accurately as possible (Fig. 1).

1.2. Fractal analysis

A computer representation of a medical image (e.g. radiological, ultrasound or computer tomography) is an image matrix where greyscale (intensity) levels corresponding to elements x, y form a more or less complex surface (Fig. 2). There are many algorithms for estimating the fractal dimension of such surfaces. In this paper, methods based on the power spectral density [10, 28], triangular prism surface area [8, 15] and variation [13, 30] were applied, as well as the box counting method [17, 24], which utilised a binary image in the analysis process. The aforementioned methods gave good results in other studies performed by the authors in which the subject of analysis was a flame area [26] and thyroid ultrasound images [19].

1.3. Image segmentation

The grey images to be analyzed using the box counting method were segmented. As a result of this process, binary images were obtained that were used to calculate the fractal dimension and lacunarity. The results of several segmentation methods were compared, including: local and global Otsu thresholding [23], local Bradley thresholding [3], level set method, adaptive thresholding and active contour method. The basic criterion for the

segmentation quality was a faithful reproduction of disease areas, visible in the CT images in the form of bright spots (Fig. 1). The best results were achieved using the global Otsu thresholding method. Therefore, in further studies, this method was selected for segmentation of grey images. In order to improve the quality of images after segmentation, binary images were subjected to operations of closing (dilation + erosion) and noise removal (removal of objects smaller than 10 pixels). Exemplary segmentation results are shown in Fig. 3.

1.4. Feature selection

Feature selection is a very important process because a properly chosen set of discriminating features allows one to build a high accuracy classifier [20–22]. In this study, a linear stepwise regression method was used for feature selection. This method involves the systematic addition and elimination of features to the set of input attributes given to the input of the linear classifier model depending on their statistical impact on the result of the system operation.

The influence of a feature on the operation of the system is measured by the factor with which it enters the linear model. The method starts from the start-up model, comparing its performance when increasing or decreasing the number of input attributes selected from the full set of potential diagnostic features. At every step (after adding or subtracting a specific attribute), the *F-Snedecor* statistics are determined for the training set. On the basis of a comparison of the *p*-value of this statistic with the assumed *p_enter* tolerance, a decision is made whether a specified feature should be entered into the set of features or not. In turn, as a result of comparing the *p*-value of the *F* statistic with the assumed *p_remove* tolerance, a decision is made to remove (or not) a specified feature from the current set of features. If the specified feature is not in the current set of input attributes, the null hypothesis is tested that its effect on the model's operation is zero. If the hypothesis is not confirmed as a result of the calculation, the attribute is added to the current set of attributes. Conversely, if a particular feature is in the model's input attribute set, the hypothesis is tested that its effect is zero. If this hypothesis is not confirmed, the feature remains in the attribute collection, but if confirmed, it is deleted.

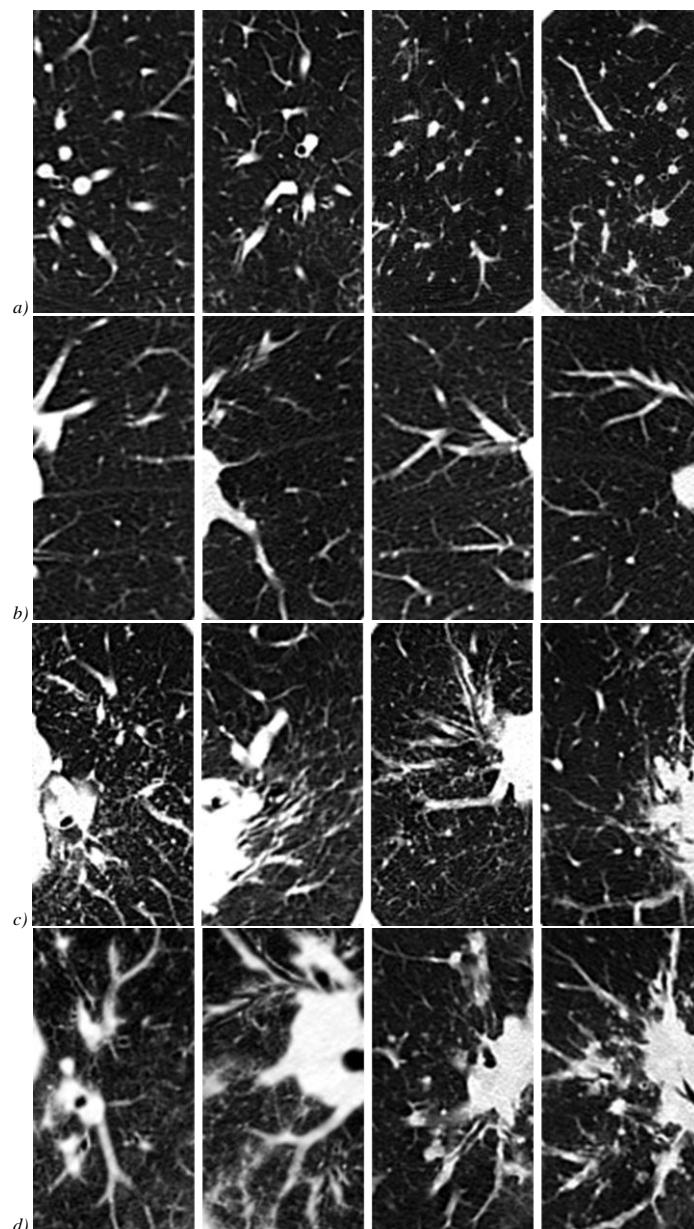


Fig. 1. ROIs characteristic for the 4 analyzed categories: a) healthy cases – Healthy; b) disease includes only a trachea – Sick(1); c) disease includes a trachea and a pulmonary parenchyma – Sick(2); d) disease includes only a pulmonary parenchyma – Sick(3)

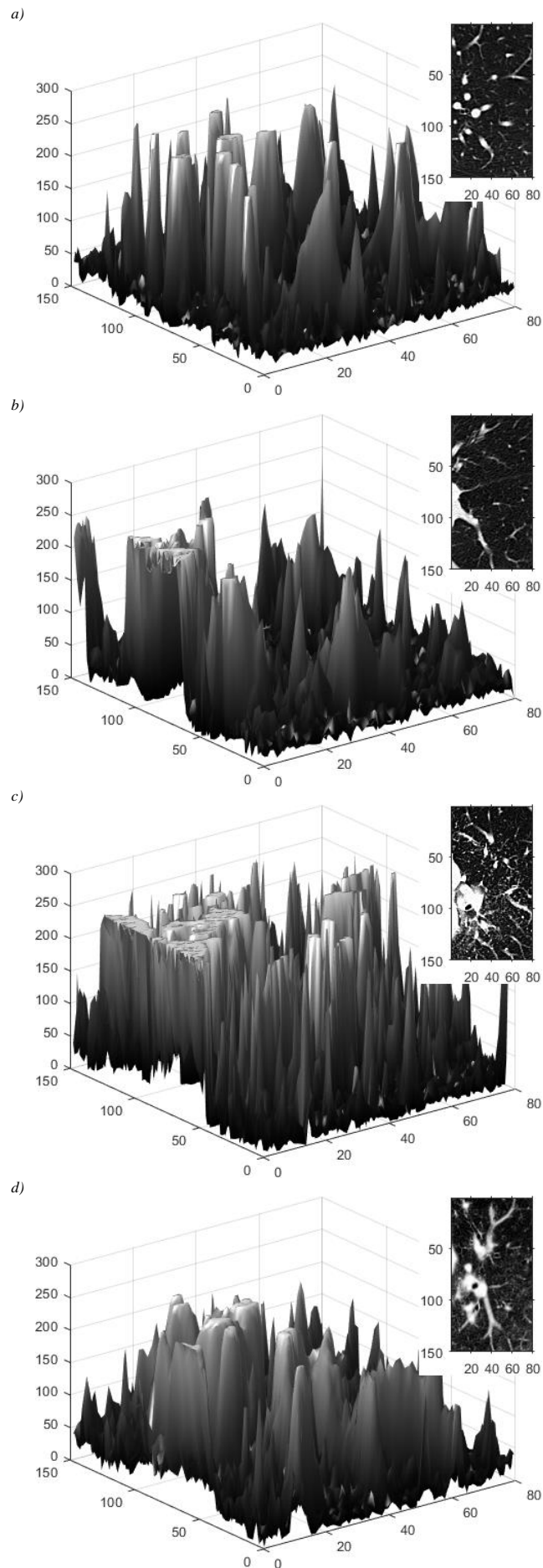


Fig. 2. Sample ROIs and their representation in three-dimensional space: a) Healthy; b) Sick(1); c) Sick(2); d) Sick(3)

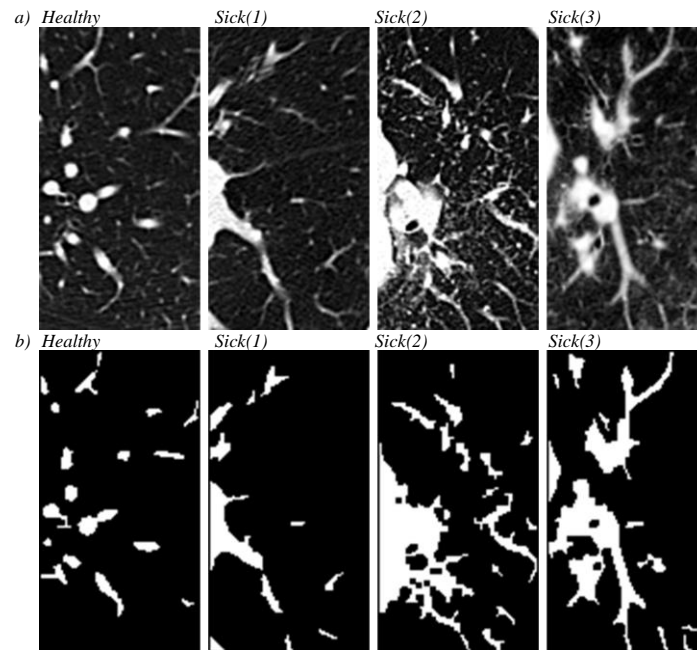


Fig. 3. Sample results of image segmentation for individual categories by the global Otsu thresholding method: a) grey images; b) binary images

2. Results

Images belonging to 4 categories of patients were subjected to fractal analysis to obtain feature descriptors. The following methods were used: power spectral density, triangular prism surface area, variation (filter 1, transect – first differences, transect – second differences) and box counting. On the basis of the power spectral density method, the fractal dimension and intercept of grey images were obtained. The box counting method allowed to estimate the fractal dimension and lacunarity of binary images. The effect of the triangular prism surface area and variation methods was the fractal dimension of grey images.

The results obtained were analyzed for the occurrence of outliers and extreme values. This kind of observations have been removed from the data set. The final number of cases belonging to particular categories was as follows: Healthy – 50, Sick(1) – 50, Sick(2) – 58, Sick(3) – 60. Table 1 presents a summary of average values of individual descriptors and their standard deviation after removal outliers and extreme values. For a binary image, the fractal dimension is in the range of 1-2, and for the grey image in the range of 2-3 (except the power spectral density method for the class Healthy). These values are in line with the theoretical values for this type of images. This situation confirms that the analyzed images have fractal features and justifies the use of fractal analysis as a method for estimating the feature descriptors.

Figure 4 presents box charts prepared on the basis of data from table 1. Observing the dispersion of values of variables defined in Table 1, it can be seen that relatively good separation of observations takes place only between cases belonging to the class Healthy and all the other cases of the sick category (Sick(1), Sick(2), Sick(3)). The best in this respect seem to be the *PSD* and *Int* variables estimated using the power spectral density method. Unfortunately, for observations of the sick category, no variable separates all classes of this category in an unambiguous way. One can only indicate a moderately good separation of one class of cases in relation to other classes of this category. One can give 2 examples in this regard. The first one is the *BC* variable, which can be used to distinguish the class Sick(1) from the classes Sick(2) and Sick(3). The second example are variables *Int* and *Var_FD*, which can be used to distinguish the class Sick(3) from the classes Sick(1) and Sick(2). Box charts in figure 4 don't suggest any variables that could be used to distinguish the class Sick(2) from the classes Sick(1) and Sick(3).

Table 1. Mean value and standard deviation of fractal descriptors

Image	Method	Descriptor	Variable name	Class	Mean	Std dev
Grey	Power spectral density	Fractal dimension	<i>PSD</i>	Healthy	3.15	0.08
				Sick(1)	2.83	0.18
				Sick(2)	2.91	0.11
				Sick(3)	2.72	0.09
Grey	Power spectral density	Intercept	<i>Int</i>	Healthy	21.76	0.47
				Sick(1)	23.29	0.98
				Sick(2)	23.64	0.40
				Sick(3)	24.43	0.36
Grey	Triangular prism surface area	Fractal dimension	<i>TPSA</i>	Healthy	2.39	0.05
				Sick(1)	2.32	0.06
				Sick(2)	2.37	0.04
				Sick(3)	2.31	0.05
Grey	Variation (Filter 1)	Fractal dimension	<i>Filter</i>	Healthy	2.68	0.04
				Sick(1)	2.57	0.07
				Sick(2)	2.60	0.05
				Sick(3)	2.49	0.06
Grey	Variation (Transect – first differences)	Fractal dimension	<i>Var_FD</i>	Healthy	2.66	0.07
				Sick(1)	2.54	0.06
				Sick(2)	2.56	0.05
				Sick(3)	2.45	0.04
Grey	Variation (Transect – second differences)	Fractal dimension	<i>Var_SD</i>	Healthy	2.51	0.10
				Sick(1)	2.36	0.05
				Sick(2)	2.40	0.09
				Sick(3)	2.28	0.05
Binary	Box counting	Fractal dimension	<i>BC</i>	Healthy	1.01	0.10
				Sick(1)	1.10	0.08
				Sick(2)	1.42	0.15
				Sick(3)	1.46	0.11
Binary	Box counting	Lacunarity	<i>Lac</i>	Healthy	0.68	0.08
				Sick(1)	1.09	0.22
				Sick(2)	0.91	0.27
				Sick(3)	0.83	0.22

The results of the fractal analysis presented in Table 1 and figure 4 do not allow to indicate a set of descriptors that could be used in a system for automatic classification of the 4 analysed categories of observations. Therefore, the data were subjected to statistical analysis aimed at selecting features that would enable to build the classifier. A linear stepwise regression method was applied for this purpose. The *stepwisefit* function of the Matlab program was used, in which the aforementioned method was implemented. The built-in values of the *F* statistics tolerance thresholds were as follows: $p_enter = 0.05$ and $p_remove = 0.1$.

Table 2 presents the results obtained. The meaning of individual columns is as follows:

- B – coefficients with which particular features affect the accuracy of model mapping. The greater the value (as to the module), the greater the influence of a given feature.
- SE – standard deviation of the values of B coefficients for individual features.
- PV – a *p-value* variable containing *p* statistic values for individual features. The smaller the value, the more difficult it is to reject the hypothesis about the low importance of a given feature.
- Status – value 1 means that the given feature is included as the input attribute of the model, 0 – means its elimination.

Table 2. Results of feature selection by linear stepwise regression

Feature	B	SE	PV	Status
PSD	-0.4538	0.7979	0.5702	0
Int	0.2719	0.0568	0.0000	1
TPSA	1.2231	0.7954	0.1256	0
Filter	-2.3626	0.5089	0.0000	1
Var_FD	-0.7321	0.7476	0.3286	0
Var_SD	-0.5577	0.4419	0.2084	0
BC	2.7786	0.2521	0.0000	1
Lac	0.4693	0.1663	0.0052	1

The results obtained show that the features *Int*, *Filter*, *BC* and *Lac* (the smallest *p-value* values) have the greatest influence on the model. They are marked in bold in Table 2. Also, the values of *p* statistics for individual features show a large difference in their impact on the result.

A matrix scatter plot was made for variables selected using the linear stepwise regression method (Fig. 5). On the basis of this plot you can see that observations of the healthy category form a compact group, quite well separated from the observation of the sick category. Apparently this can be seen for the variable pairs *Int-Lac* and *BC-Lac*. Observations of the sick category have a larger dispersion of variable values compared to observations of the healthy category. Therefore, in the scatter plots you can not indicate clear groups of cases that would allow you to unequivocally distinguish between different classes of the sick category. Nevertheless, in the scatter plots one can observe some areas in which observations of individual classes tend to cluster.

It seems that this is best seen for the variables *Filter-BC*. The spatial chart gives better possibilities for observing cases clustering. Figure 6 shows such a graph for variables *Filter-BC-Lac*. Admittedly, there are no separate groups of observations there, but there is a clear tendency to clustering of cases belonging to particular classes. This situation gives a base for the construction of classifiers.

For this purpose, 4 features highlighted in bold in table 2 were used: *Int*, *Filter*, *BC* and *Lac*. Eight popular supervised learning methods were used [4–6, 11, 12, 14, 16, 18, 25, 29]: artificial neural network with multilayer perceptron (MLP), decision tree (DT), *K*-nearest neighbors (*K*-NN), naive Bayes classifier (NBC), quadratic and linear discriminant analysis (QDA, LDA), random forests (RF), support vector machine (SVM). The full data set was randomly divided into a training, validation and test part. The size of each of these subsets was 1/3 of the full set. A combined training and validation sets were used for the final classifiers training. Testing was carried out using an independent test set. Training, validation and classifiers testing have been repeated 10 times. An average value from all tests was taken as the final result of the testing process. The overall classification accuracy (ACC), classification sensitivity and classification specificity was used to assess the accuracy of classifiers. The overall classification accuracy is defined as the ratio of the number of samples correctly classified to the number of all analyzed samples. There were 4 classes of cases in the study, therefore the other 2 measures were generalized to a classifier operating on the principle of "one specific class against all", i.e. assigning results to only one class or beyond it. Therefore, the classifier sensitivity means the probability of correct classification of the sample belonging to the selected class. In turn, the specificity is defined as the probability that samples belonging to other classes will not be assigned to the selected class.

Figure 7 shows the average value of the overall classification accuracy obtained for the test set. The largest value, equal to 90.97%, was achieved for the SVM classifier. Comparable accuracy was also obtained by the MLP classifier, for which ACC was equal to 90.83%. The difference from the SVM classifier was only 0.14%. It should be noted that ACC exceeded 82% for all classification methods.

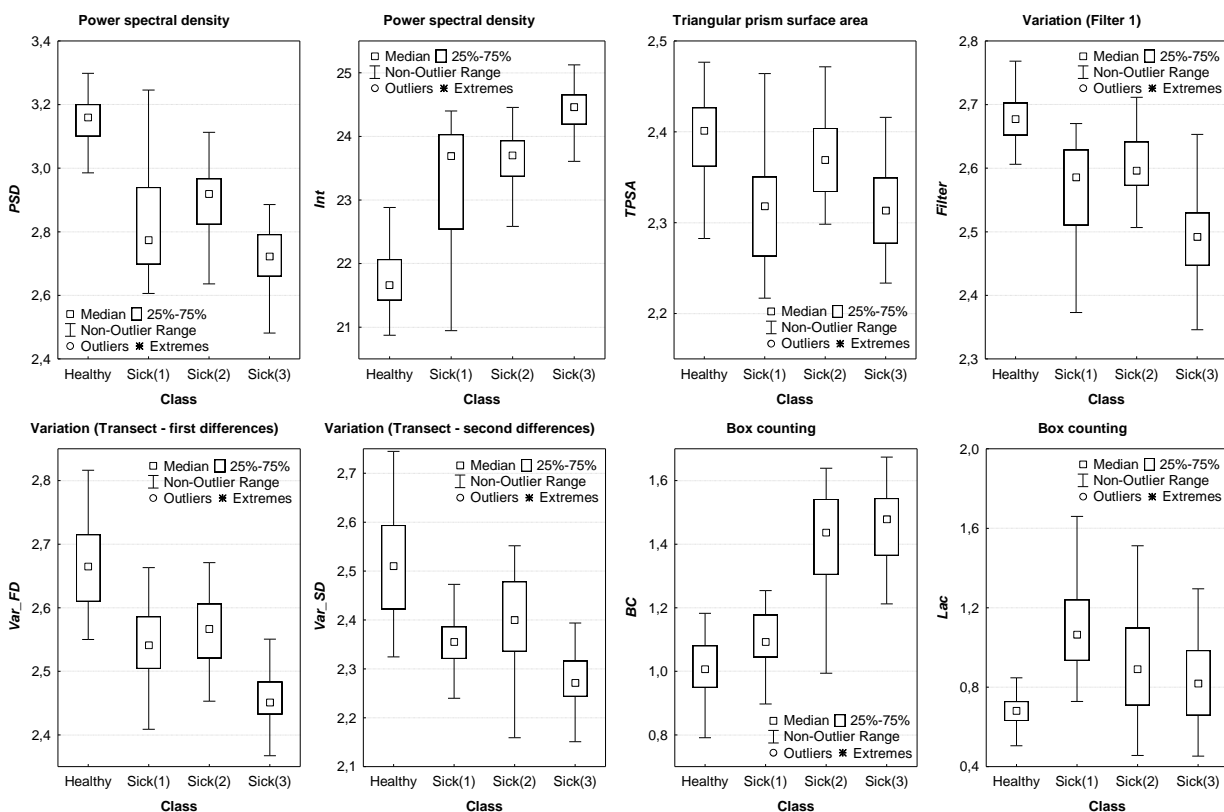


Fig. 4. Dispersion of a mean value of individual variables. A title of each chart shows a name of a method used. A meaning of variables is explained in Table 1

The results regarding the estimation of the classification sensitivity (separately for each class) are shown in Table 3. For the class Healthy, the highest sensitivity of 100% was achieved for 4 classifiers, i.e. SVM, NBC, LDA and 7-NN. For classes belonging to the sick category, the results obtained were as follows (the best classifier was given in brackets): Sick(1) – 92.50% (SVM); Sick(2) – 87.50% (MLP); Sick(3) – 90.00% (NBC).

Table 3. Classification sensitivity

Classifier	Healthy	Sick(1)	Sick(2)	Sick(3)
SVM	100	92.50	86.50	87.00
MLP	99.38	91.88	87.50	86.50
QDA	98.75	82.50	85.00	85.00
NBC	100	80.63	80.00	90.00
RF	95.63	90.63	80.00	84.50
LDA	100	75.00	84.00	83.50
7-NN	100	81.25	79.00	80.50
DT	93.13	80.63	74.00	83.00

Table 4 presents the classification specificity for each of the 4 observation classes. For the class Healthy, the best result (100%) was obtained for SVM and NBC classifiers. A similar specificity value (99.82%) was also provided by the MLP classifier. In turn, for the other classes the results were as follows: Sick(1) – 97.86% (SVM); Sick(2) – 95.19% (MLP and LDA); Sick(3) – 96.15% (NBC).

Table 4. Classification specificity

Classifier	Healthy	Sick(1)	Sick(2)	Sick(3)
SVM	100	97.86	94.81	95.00
MLP	99.82	97.68	95.19	94.81
QDA	98.39	97.32	92.88	94.42
NBC	100	94.46	92.31	96.15
RF	98.57	95.54	94.42	94.04
LDA	94.82	95.89	95.19	94.62
7-NN	97.14	95.54	92.31	94.23
DT	98.93	94.82	89.62	92.50

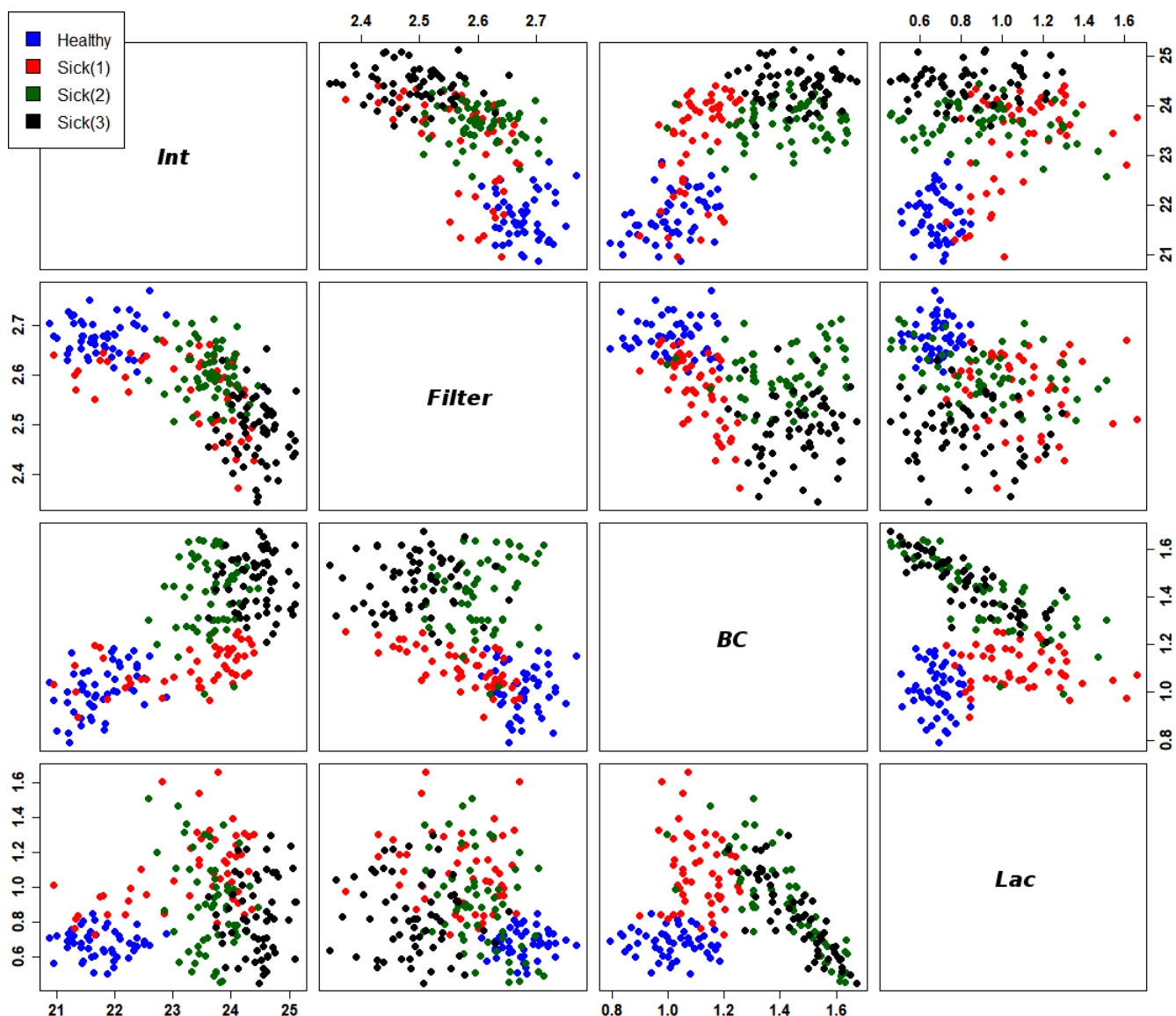


Fig. 5. Scatter plots for the variables selected by the linear stepwise regression method. Variables of the specified symbols are described in Table 1

3. Discussion

The study used 8 fractal descriptors estimated by various methods. Their values confirm the fractal nature of the studied CT images (Table 1) and justify the use of fractal analysis as a research tool. Using the linear stepwise regression method, 4 descriptors were selected from this set for constructing classifiers. These were: intercept – calculated using the power spectral density method; 2 fractal dimensions – calculated by

the variation and box counting methods; lacunarity – also calculated using the box counting method. The first 2 descriptors (*Int*, *Filter*) were calculated based on a grey image analysis, while the next 2 (*BC*, *Lac*) based on a binary image.

Table 5 gives a summary of the best results obtained while classifiers testing. The highest overall classification accuracy was obtained for the SVM classifier (ACC=90.97%). In the case of classification sensitivity and specificity, the test results depend on the class of observation. The best results have been

achieved here for the class Healthy. Both sensitivity and specificity are at the level of 100%. Healthy cases are the easiest to distinguish from the other classes, because they not contain large, bright areas corresponding to the disease state (Fig. 1a). Taking into account the Sick cases, the best accuracy was achieved for the class Sick(1), where the disease state includes the trachea (Fig. 1b). The sensitivity was 92.50% here and the specificity was 97.86%.

Table 5. A summary of the best classification results

Index	Class	Value (%)	Classifier
ACC	All classes	90.97	SVM
Classification sensitivity	Healthy	100	SVM, NBC, LDA, 7-NN
	Sick(1)	92.50	SVM
	Sick(2)	87.50	MLP
	Sick(3)	90.00	NBC
Classification specificity	Healthy	100	SVM, NBC
	Sick(1)	97.86	SVM
	Sick(2)	95.19	MLP, LDA
	Sick(3)	96.15	NBC

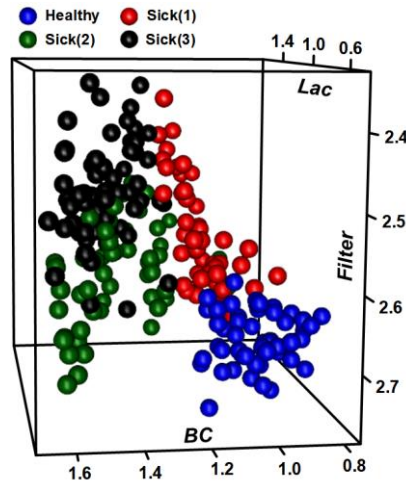


Fig. 6. Spatial scatter plot for the variables Filter-BC-Lac

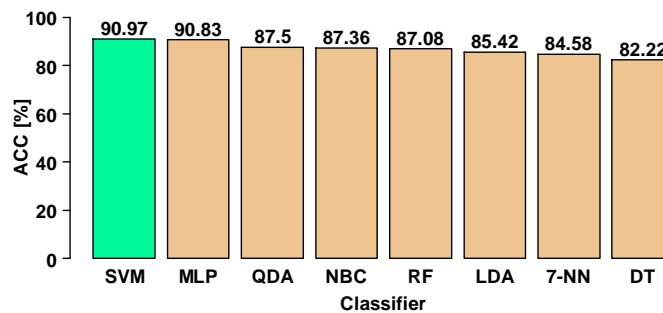


Fig. 7. Overall classification accuracy. The meaning of the classifier symbols is as follows: SVM – support vector machine; MLP – artificial neural network with multilayer perceptron; QDA – quadratic discriminant analysis; NBC – naive Bayes classifier; RF – random forests; LDA – linear discriminant analysis; 7-NN – K-nearest neighbours for K=7; DT – decision tree

The second place in terms of classification accuracy is occupied by the class Sick(3), for which the disease state includes pulmonary parenchyma (fig. 1d). For this class of observations, the sensitivity was at the level of 90.00%, and the specificity was 96.15%. The worst results were obtained for the class Sick(2), where the disease state includes both trachea and pulmonary parenchyma (fig. 1c). In this case, the sensitivity was 87.50%, and the specificity was 95.19%. The class Sick(2) has proved to be the most difficult to classify because it includes symptoms occurring both in Sick(1) and Sick(3) class. The effectiveness of the SVM and NBC classifiers is also worth noting. Out of 9 classification accuracy indicators, the SVM classifier turned out to be the best in 5 cases, and the NBC classifier in 4 ones (table 5). Such results provide the basis for the use of these classifiers in a computer system for the automatic diagnosing of sarcoidosis.

The results obtained also show that in the case of observations of the sick category, for each of the classes, the classification sensitivity is a few percent worse than the specificity. These differences for the classes Sick(1), Sick 2) and Sick(3) are respectively 5.36%, 7.69% and 6.16%. This means that the probability of a correct classification of observations belonging to a given class is lower than the probability that observations belonging to other classes will not be assigned to a given class. Unfortunately, this relation between the values of both classification quality parameters is not beneficial for the patient

from a medical point of view. Therefore, further research is required to increase the classification sensitivity at least to the level of specificity. It seems that good results would be obtained by combine the effects of fractal analysis with the results of other methods for calculating the feature descriptors of image textures, e.g. autoregression model, wavelet transform and statistical methods (greyscale histogram, co-occurrence matrix, higher-order statistics, run-length matrix, matrix gradient) [14].

The results obtained are difficult to compare with the results of other authors, because (as noted in the introduction) there are no known publications regarding the automatic detection of sarcoidosis based on CT images. The advantages of the presented method include the fact that it is non-invasive and offers relatively high accuracy in recognizing cases belonging to particular classes. Let us remind that the worst of the results (87.50%), concerning the classification sensitivity of the class Sick(2), was at a relatively high level. Another advantage is the fact that the method is based on only a few features selected solely on the basis of the fractal analysis. This situation should simplify the construction of the future computer system. On the other hand, a certain drawback of the applied method of analysis is the fact that there is a lack of standardization of methods for estimating the fractal dimension. Consequently, the values of this parameter for the same image, estimated by different methods, are similar but not identical. In addition, the estimation range of the regression line slope can be

determined in various ways, which influences the value of the estimated parameter. Finally, the presented method must be tested on data sets of a larger size so as to obtain statistical significance of the results. Let us recall that the number of observations for particular classes was: Healthy – 50, Sick(1) – 50, Sick(2) – 58, Sick(3) – 60. The tests carried out showed that in order to obtain statistical significance, the number of individual subsets should exceed 100 observations. Future studies should also consider the impact of changes in volume, turnover and ROI position on classification results. Despite the problems indicated, the fractal analysis seems to be an interesting tool in the study of CT image textures. The relatively high classification accuracy achieved thanks to fractal descriptors gives the basis for their use in a computer system for the automatic diagnosis of sarcoidosis.

4. Conclusions

The research has shown the high effectiveness of the fractal analysis as a tool for the feature vector extraction in diagnosis of sarcoidosis. The following results were obtained: overall classification accuracy – 90.97%; classification sensitivity – 100% (Healthy), 92.50% (Sick(1)), 87.50% (Sick(2)), 90.00% (Sick(3)); classification specificity – 100% (Healthy), 97.86% (Sick(1)), 95.19% (Sick(2)), 96.15% (Sick(3)). Achieving high classification accuracy was possible due to the simultaneous use of fractal descriptors of grey and binary images. These were: intercept – calculated using the power spectral density method; 2 fractal dimensions – calculated by the variation and box counting methods; lacunarity – also calculated using the box counting method. Fractal analysis can be an alternative approach in studies of CT image textures aimed at automatic diagnosis of sarcoidosis. Its results, combined with such classification methods as support vector machine or naive Bayes classifier, give interesting possibilities in the construction of an automated computer system that could help the physician in the preliminary diagnosis of difficult cases based on CT images. Thanks to this, the doctor would have a quick and objective additional opinion that would be very valuable at the initial diagnosis stage. In the case of sarcoidosis, this is particularly important because the detection of the disease at an early stage increases the patient's chances for effective therapy.

References

- [1] Arkema E. V., Cozier Y. C.: Epidemiology of sarcoidosis: current findings and future directions. *Ther Adv Chronic Dis* 9(11)/2018, 227–240.
- [2] Baughman R. P., Culver D. A., Judson, M. A.: A Concise Review of Pulmonary Sarcoidosis. *Am J Respir Crit Care Med* 183/2011, 573–581.
- [3] Bradley D., Roth G.: Adaptive Thresholding Using the Integral Image. <http://www.scs.carleton.ca/~roth/iit-publications-iti/docs/gerh-50002.pdf> (available: 19.05.2019).
- [4] Breiman L.: Bagging Predictors. *Mach Learn* 24/1996, 123–140.
- [5] Breiman L.: Random Forests. *Mach Learn* 45/2001, 5–32.
- [6] Breiman L., Friedman J., Olshen R., et al.: *Classification and Regression Trees*. CRC Press, London 1984.
- [7] Bonifazi M., Gasparini S., Alfieri V., Renzoni E. A.: Pulmonary Sarcoidosis. *Semin Respir Crit Care Med* 38/2017, 437–449.
- [8] Clarke K. C.: Computation of the fractal dimension of topographic surfaces using the triangular prism surface area method. *Comp Geosci* 12/1986, 713–722.
- [9] Criado E., Sánchez M., Ramírez J., Arguis P., de Caralt T. M., Perea R. J., et al.: Pulmonary Sarcoidosis: Typical and Atypical Manifestations at High-Resolution CT with Pathologic Correlation. *RadioGraphics* 30/2010, 1567–1586 [DOI: 10.1148/rg.306105512].
- [10] Dennis T. J., Dessipris N. G.: Fractal modelling in image texture analysis. *IEEE Proc.-F* 136/1989, 227–235.
- [11] Enas G. G., Chai S. C.: Choice of the smoothing parameter and efficiency of the k-nearest neighbor classification. *Comput Math Appl* 12/1986, 235–244.
- [12] Freund Y., Schapire R. E.: A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci Int* 55/1996, 119–139.
- [13] Gneiting T., Sevcikova H., Percival D.: Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data. *Statistical Science* 27(2)/2012, 247–277.
- [14] Hothorn T., Lausen B.: Bundling classifiers by bagging trees. *Comput Stat Data An* 49(4)/2005, 1068–1078.
- [15] Iftekharuddin K. M., Jia W., Marsh R.: Fractal analysis of tumor in brain MR images. *Mach Vision Appl* 13(5-6)/2003, 352–362.
- [16] Liao S. H., Chu P. H., Hsiao P. Y.: Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Syst Appl* 39/2012, 11303–11311.
- [17] Mandelbrot, B., [The fractal geometry of nature], W. H. Freeman and Company, New York, (1983).
- [18] Omiotek Z.: Improvement of the classification quality in detection of Hashimoto's disease with a combined classifier approach. *P I Mech Eng H* 231(8)/2017, 774–782.
- [19] Omiotek Z.: Fractal analysis of the grey and binary images in diagnosis of Hashimoto's thyroiditis. *Biocybern Biomed Eng* 37(4)/2017, 655–665.
- [20] Omiotek Z., Wójcik W.: Zastosowanie metody Hellwiga do redukcji wymiaru przestrzeni cech obrazów USG tarczycy. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS* 3/2014, 14–17 [DOI: 10.5604/20830157.1121333].
- [21] Omiotek Z., Wójcik W.: An efficient method for analyzing measurement results on the example of thyroid ultrasound images. *Przegląd Elektrotechniczny* 11/2016, 15–18.
- [22] Omiotek Z., Stepanchenko O., Wójcik W., Legieć W., Szatkowska M.: The use of the Hellwig's method for feature selection in the detection of myeloma bone destruction based on radiographic images. *Biocybern Biomed Eng* 39(2)/2019, 328–338.
- [23] Otsu N.: A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9(1)/1979, 62–66.
- [24] Plotnick R. E., Gardner R. H., O'Neil R. V.: Lacunarity indices as measures of landscape texture. *Landscape Ecol* 8(3)/1993, 201–211.
- [25] Quinlan J. R.: Induction of decision trees. *Mach Learn* 1/1986, 81–106.
- [26] Sawicki D., Omiotek Z.: Evaluation of the possibility of using fractal analysis to study the flame in the co-firing process. *Proc. SPIE* 10808/2018.
- [27] Soto-Gomez N., Peters J. I., Nambiar A. M.: Diagnosis and Management of Sarcoidosis. *American Family Physician* 93(10)/2016, 840–848.
- [28] Super B. J., Bovik A. C.: Localized measurement of image fractal dimension using Gabor filters. *J Visual Commun Image Represent* 2(2)/1991, 114–128.
- [29] Venables W. N., Ripley B. D.: *Modern Applied Statistics with S-PLUS*. Springer, Berlin 1998.
- [30] Zhu Z., Stein M.: Parameter estimation for fractional Brownian surfaces. *Statistica Sinica* 12/2002, 863–883.

Ph.D. Eng. Zbigniew Omiotek

e-mail: z.omiotek@pollub.pl

A graduate of the Faculty of Electronics at the Military University of Technology in Warsaw. In the years 1995-2013 – research and teaching assistant. Since 2015 – assistant professor at the Lublin University of Technology. His scientific interests are focused around methods for digital image processing, machine learning and computer aided diagnosis.

ORCID ID: 0000-0002-6614-7799

M.Sc. Eng. Paweł Prokop

e-mail: pawel.prokop@pollub.edu.pl

In 2012 he graduated with a degree in Computer Science of the Faculty of Electrical Engineering and Computer Science at the Lublin University of Technology. Currently a Ph.D. student at the Faculty of Electrical Engineering and Computer Science. His scientific interests are focused around computer science and biomedical engineering.

ORCID ID: 0000-0002-3078-8287

otrzymano/received: 20.05.2019

przyjęto do druku/accepted: 15.06.2019

