

ROC and reproducibility analysis of designed algorithm for potential diagnosis of Parkinson's disease in ultrasound images

JIRÍ BLAHUTA, TOMÁŠ SOUKUP, PETR ČERMÁK, MICHAL VEČEREK, MILAN JAKEL,
DAVID NOVÁK

Department of Informatics
Silesian University in Opava
Bezručovo náměstí 13, 74601 Opava
CZECH REPUBLIC

jiri.blahuta@seznam.cz tsoukup@centrum.cz petr.cermak@fpf.slu.cz michal.vecerek@vsb.cz
milan@jakel.cz david.novak@fpf.slu.cz

Abstract: - This paper introduces our experimental software which has been developed in MATLAB for potential detection of pathology to detection of Parkinson's disease. The algorithm is based on thresholding of intensity images, using the elliptical ROI followed by computing of area in this ROI for each intensity. The second step is statistical analysis of these data such as ROC curve, variability and Cohen's kappa coefficient. We used DICOM or converted images into JPEG of TCS (transcranial ultrasound images). This statistics is critical for appraisal of repeatability and reproducibility of designed algorithm.

Key-Words: - ROI, ultrasound, reproducibility, MATLAB, algorithm, image processing, ROC

1 Introduction

We work with set of ultrasound images of brain-stem to potential diagnosis of Parkinson's disease (PD). In ultrasound images is characterized by defects in substantia nigra (SN) area in brain-stem. The goal of this paper is to present an application which has been developed for ROI SN for distinction of physiology and pathology based on thresholding and area comparison followed by statistical analysis, especially variability, ROC and level of agreement between 2 examiners. Developed software is helpful for classification of patients who have PD or not. We have tested repeatability and reproducibility of method that is critical factor to using. ROC analysis is very important followed by Cohen's kappa coefficient.

of basal ganglia. Etiology of PD is unknown. More information about PD is available in [2], [3] and [4].

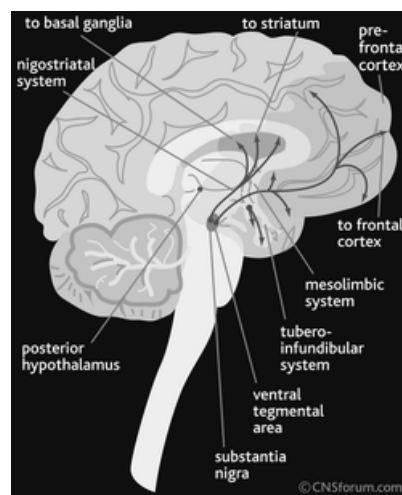


Fig. 1 – Position of SN

2 Substantia nigra and Parkinson's disease

Substantia nigra (SN; in English "black substance") is a brain structure which is located in the mesencephalon (midbrain) and plays an important role for central nervous system. SN produces an important dopamine for correct function of CNS. Parkinson's disease (PD) is caused by the death of dopaminergic neurons. It is a degenerative disease

US is an appropriate method for scanning in neurology, brain is very sensitive and soft tissue. Generally we use diagnostic ultrasound for soft tissues. Diagnostic ultrasound uses frequency up to 12 MHz. General benefits for ultrasound in neurology are:

- no radiation; exposition can be repeated and correlated - repeatability for examiners
- US is well applicable for soft tissues (brain stem density is 34.7 HU)¹
- B-mode scanning is advisable in form brightness differentiation to select ROI SN
- applicable for small areas such as SN

Because ultrasound has no negative influence by ionization, reproducibility is very important factor.

Ultrasound devices for diagnostic ultrasound fully supports and works in international medical standard DICOM (Digital Imaging and Communications in Medicine), which is worldwide standard for imaging, storing and transferring data from various modalities.² MATLAB with Image Processing Toolbox supports DICOM format.³

We work with original ultrasound images in DICOM format and/or with converted images into JPEG format. The following figure shows an example of area of SN on real US image.

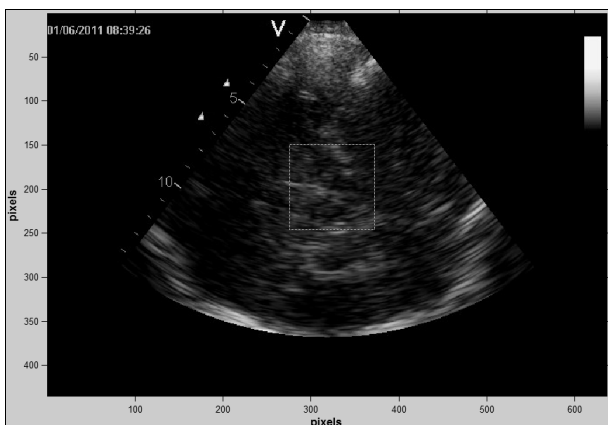


Fig. 2 – Window with ROI SN in US image

3 Implementation and results

In this part we will present how to distinct pathology and physiology practically with using of our application. Furthermore describes this algorithm and mathematical background of used statistical methods. Whole processing we can divide into three phases:

¹ <http://www.fpnotebook.com/Rad/CT/HnsfldUnt.htm>

² <http://www.rsna.org/Technology/DICOM/index.cfm>

³

<http://www.mathworks.com/help/toolbox/images/ref/dicomread.html>

- pre-processing (ROI window, ROI)
- main processing (thresholding with computing of area in ROI)
- statistical analysis of results to repeatability, ROC

3.1 Image pre-processing

Our application can read original DICOM or converted images into common bitmap formats such as BMP, JPEG, PNG, JPEG-2000 and so on. It is possible by two push buttons.

We need to read an image (bitmap or DICOM file) and cut ROI window 50×50 mm. This dimension of window with ROI has been appointed by neurologist. Before cutting an image algorithm checks that input image is in 8-bit or 24-bit depth. For whole processing we need images in gray scale, thus pixel of image has an intensity value $I = \langle 0; 255 \rangle$. If input image is in 24-bit depth (RGB), `rgb2gray` function converts this input by the following formula into grayscale image which is needed for our processing:

$$I = 0,299 \times R + 0,587 \times G + 0,114 \times B. \quad (1)$$

Thus, each pixel p_i in image has an intensity value from range $\langle 0; 255 \rangle$, it is a real matrix with these matrix elements.

3.1.1 Region of interest (ROI)

We require an elliptical ROI with initial area $A = 50 \text{ mm}^2$ with rotating for left/right SN. In application we use 8 presets of ellipses with a different rotation to define ROI. An example of used elliptical ROI shows the following Figure 3. To define ROI we used binary masks which are defined as ROI in image.

```
setVerticesDraggable(maskup,false)
setColor(maskup,'green')
position=wait(maskup);
zoom
binaryImage = maskup.createMask;
```

Method `setVerticesDraggable` sets that we can change the shape of ROI and `maskup.createMask` creates ROI.

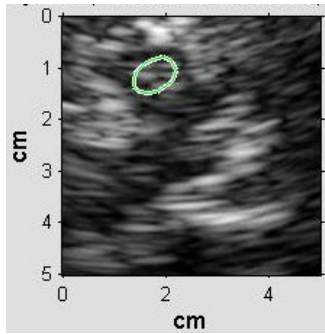


Fig. 3 – An example of elliptical ROI in SN

3.2 Main processing

Phase of main processing is based on binary thresholding in ROI area and computing of area of defects in this ROI for all intensity levels $T \in \langle 0; 255 \rangle$. We will get a graph with computed area which is computed from block processing 1×1 mm. The area has been computed from binary image \Rightarrow number of blocks in this 1×1 mm grid will get the real area in mm^2 . The following figure shows blockprocessed elliptical ROI (Fig. 3). For block processing we use blockproc function:

```
fun = @(block_struct) ...
    mean2(block_struct.data) *
    ones(size(block_struct.data));
reg2 = blockproc(BinImage,[d/10/5 d/10/5],fun);
```

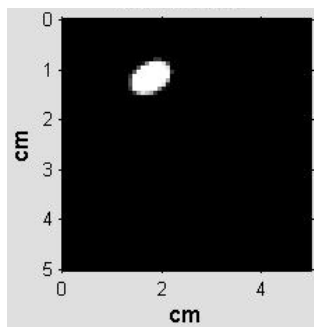


Fig. 4 – Grid 1×1 mm for area

Algorithm of thresholding is based on comparison of each pixel in image with threshold T :

```
for T=0 to 255 do
    compute_area
end
```

For this thresholding we use preallocation to speedup of loop. It is useful if we need to compute with large vector and/or iteration. In this case, preallocation creates a vector with 255 values of 0.

More about preallocation on the web Matlab Hypertext⁴ with great examples.

```
arth = zeros(0, 255);
```

3.2.1 Pathology and physiology

From these graphs of computed area with decreasing tendency we can classify a pathology and normal physiology.

In practice we verified a basic rule about decrease of area for T for approximate distinction – if decrease is gradual, we can assess that patient is probably diseased (PD features,) and if decrease is very fast, patient is probably healthy (small featureless areas). We have compared it with opinion of expert, doctor. For example, the following figure shows a difference between healthy and diseased patient. Experimentally we set the threshold $T = 10\%$ for analysis of variance.

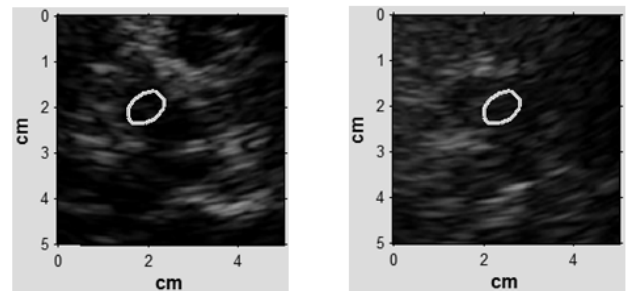


Fig. 5 - Healthy and diseased patient

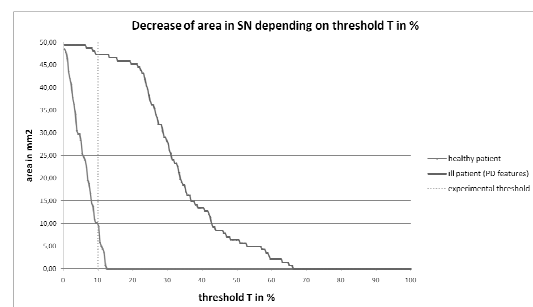


Fig. 6 – Decrease of area for healthy and ill patient

A doctor as expert in neurology appointed who patients are healthy or diseased. It is an initial step for comparison with results of algorithm.

Also we need to ensure a reproducibility of this method. For the same patient we must get the same values of area (or with minimal deviation) to classify pathology/physiology if moderately shift ROI. Reproducibility of this method is the main

4

<http://web.cecs.pdx.edu/~gerry/MATLAB/programming/performance.html>

goal of this work, it is one of the most important notions in medicine.

3.2.2 Repeatability and reproducibility

We must ensure repeatability and reproducibility as we described in previous chapter. We tested 600 samples – each patient was verified by 6 measurements how we described. The following figure demonstrates 6 results of area in ROI of the same patient (diseased), Each of two examiners accomplished 3 measurements with different ROI position to objective results. The deviation is minimal. The following figure shows an example for 2 different measurements. Complete set of 6 measurements is available in [2], [3].

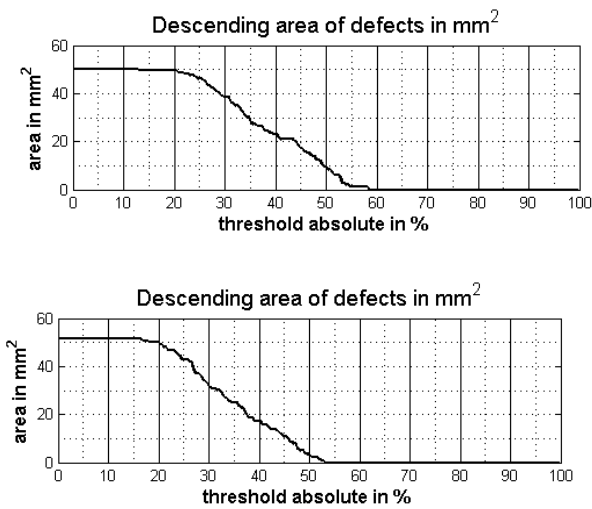


Fig. 7 – An example of deviation of measurement

These minimal changes are criterion for reproducibility of used method. Similar results we got from all patients despite 6 images were not identified well because influence of noise

$$I = u(I) + n(I), \tag{2}$$

where I is an image, $u(I)$ is useful signal and $n(I)$ is noise (for example additive), is too distinct. US images are generally very sensitive to noise, especially to speckle noise.⁵

Graphical outputs such as images or graphs can be saved into bitmap file for next processing. So, all values of area can be saved into Excel XLS file for statistical analysis which is our primary goal.

⁵

http://www.gehealthcare.com/usen/ultrasound/education/docs/whitepaper_SRI.pdf

4 Statistical analysis

Statistical analysis for our case is based on variance, standard deviation and coefficient of variability. For experimental results we appointed $T = 10\%$, this threshold can be appropriate to testing. With reproducibility is statistical analysis needed to testing of deviations between cases. It is a stability of algorithm and we require minimal errors.

4.1 Variance and standard deviation

Variance and standard deviation are basic statistical characteristics for testing. We appoint X as statistical measurement (area for $T = 10\%$), arithmetic mean

$$mean(X) = \frac{1}{n} \sum_{i=1}^n x_i \tag{3}$$

for all values of area for $T = 10\%$ and 600 samples. Furthermore, we computed variance

$$var(X) = \frac{\sum(x_i - mean(X))^2}{n} \tag{4}$$

and consequently standard deviation σ which is defined as square root of variance.

$$\sigma = \sqrt{var(X)}. \tag{5}$$

If we get variance and standard deviation we can compute coefficient of variability, which is significant characteristic defined in percent by the following formula:

$$v(X) = \frac{\sigma}{mean(X)} \times 100. \tag{6}$$

From these formulas we can express that variance is a square of standard deviation, thus variance is σ^2

$$\sigma^2 = var(X). \tag{7}$$

4.2 ROC curve and Cohen's kappa

(10)

For ROC analysis and Cohen's kappa coefficient we must consider the 95th percentile of measured results as a standard for medical applications. Percentile has been computed from mean area of physiological patients. Mathematically, n-th percentile P_n , we suppose a series of measurements x_1, x_2, \dots, x_N and then:

$$P_n = (100/N)(n-1/2), \tag{8}$$

where n is n-th percentile and N is number of measurements. For our 95th percentile we obtain

$$P_{95} = (100/N)(95-1/2). \tag{9}$$

ROC curve shows a specificity and sensitivity of measurements. Mathematically is based on categorization of possible results and number of cases. We computed a differences between this percentile and average from 6 measurements. Sum of these differences is nonnegative or negative. For each patient we computed it and compare with visual assessment by neurological specialist. We suppose 4 different cases:

- if sum of differences > 220 ⇒ **patient is securely physiological**
- if sum of differences ∈ (220; 0) ⇒ **patient is probably physiological**
- if sum of differences ∈ (0; -220) ⇒ **patient is probably pathological**
- if sum of differences < -220 ⇒ **patient is securely pathological**

These results are followed by computing of count of patient in each category to ROC curve.

Very important is Cohen's kappa coefficient. In medicine, Cohen's kappa coefficient is very important result, which says level of agreement between two examiners. It is critical to appraisal of reproducibility. Mathematically we express by the following formula:

$$K = \frac{\sum_{i=1}^r n_{ii} - \sum_{i=1}^r m_{ii}}{n - \sum_{i=1}^r m_{ii}},$$

where $m_{ii} = \frac{n_{i+}n_{+i}}{n}$ for matrix $r \times r$. We use the Cohen's version of kappa, because we are two raters only, no more. Formally, Cohen's kappa is defined by the following formula:

$$\frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \tag{11}$$

where $\text{Pr}(a)$ is the relative observed agreement among raters and $\text{Pr}(e)$ is the hypothetical probability of chance agreement. Kappa coefficient is between $\kappa \in (0; 1)$.

4.3 Practical results of analysis

This part discusses practical results and conclusions from experimental application.

For each patient we computed values of variance followed by standard deviation (STD) that is described in previous section, from values of area with $T = 10\%$. This threshold can be determining for comparison. Also we computed values for another threshold, variance and deviation have a minimal error. For example, the following chart shows changes in variance and standard deviation analysis for all 100 patients.

Deviation for experimental measure is relatively low – mean of STD for 100 patients is $\sigma = 3,87$, it is applicable as deviation of $3,87 \text{ mm}^2$ in ROI. This analysis is needed for verifying of repeatability, thus for the same patient we will get a similar values.

Coefficient of variability $v(X)$ is a characteristic which is described by formula (6). For example we will show value $v(X)$ for three patients. We suppose that X_1, X_2 and X_3 are three different samples (patients).

$$\begin{aligned} v(X_1) &= 10,269 \\ v(X_2) &= 28,349 \\ v(X_3) &= 1,755 \end{aligned}$$

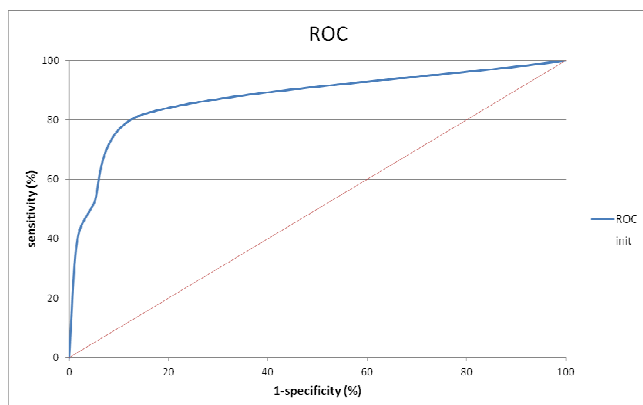
Detailed information about used methods of statistics is available on the web of MIT.⁶

6

http://web.mit.edu/10.001/Web/Course_Notes/Statistics_Notes/Visualization/node4.html

4.3.1 ROC curve

ROC curve for described case shows the following graph.



4.3.2 Cohen's kappa coefficient

In the previous chapter we introduced how to use the Cohen's kappa coefficient for 2 raters only. We only have two raters; each rater accomplished 3 measurements in ROI to distinction pathology or physiology in SN.

intra-raters kappa coefficient 1 = 0.947

intra-raters kappa coefficient 2 = 0.943

inter-raters kappa coefficient = 0.880

The level of agreement is "almost perfect", interpretation of kappa value is described in the publication "*Understanding Interobserver Agreement: The Kappa Statistic*"⁷. With these results we can adjudge that reproducibility of designed method is very good.

5 Conclusion

The main goal of this work was to show an image processing of medical ultrasound images with ROC analysis to reproducibility of method.

In the first part we outlined our algorithm based on thresholding in elliptical ROI to computing area for ROI SN. These outputs are primary results for statistical analysis of measure. Our application can read DICOM or converted DICOM images into common bitmap format.

In the second part we demonstrated how to show the repeatability and reproducibility of algorithm with statistical analysis include ROC and Cohen's

kappa coefficient. We evaluated the results from 2 examiners.

The next processing will be based on automatic searching of defects by using of artificial neural networks.

References:

[1] Becker, G., "Degeneration of substantia nigra in chronic Parkinson's disease visualized by transcranial color-coded real-time sonography", 1995, *Journal of Neuroimaging* 45.

[2] Blahuta J., Soukup T., Čermák P., "The image recognition of brain-stem ultrasound images with neural network based on PCA", EMEH'11, 2011, *Recent Researches in Communications, Electrical and Computer Engineering*, ISBN: 978-960-474-286-8, pp. 134-142.

[3] Blahuta J., Soukup T., Čermák P., "Image processing of medical diagnostic neurosonographical images in MATLAB", EMEH'11 2011, *Recent Researches in Communications, Electrical and Computer Engineering*, ISBN: 978-1-61804-019-0, pp. 85-90.

[4] Gelb, D., O. E. G.-S.: 1999, "Diagnostic criteria for Parkinson's disease, *Archives of Neurology*" 56(1), 33–39.

[5] Montgomery, C. D., Runger, C. G., "Applied Statistics and Probability for Engineers", 2006, Wiley; 4th Edition, ISBN-13: 978-0471745891.

[6] Schreiber, J., Sojka, E., Ličev, L., Škňouřilová, P., Gaura, J., Školoudík, D.: "A new method for the detection of brain stem in transcranial ultrasound images", *Proceedings of Biosignals 2008*, 2008.

[7] Školoudík, D.: "Reproducibility of sonographic measurement of the substantia nigra", *Ultrasound in Medicine & Biology* (9), 2007, pp. 1347–1352.

[8] Školoudík, D., Bártová, P., Rössner, P., Herzig, R., Kaňovský, P., Langová, K.: "Correlation between substantia nigra features detected by sonography and Parkinson's disease symptoms.", *Journal of ultrasound in medicine*, 2010, pp. 37–42.

[9] Tabachnick, G. B., Fidell, S. L., "Using Multivariate Statistics, 5th Edition", 2006, Pearson Education; 5th Edition, ISBN-13: 978-0205459384.

7

<http://www.stfm.org/fmhub/fm2005/May/Anthony360.pdf>