RWTH AACHEN UNIVERSITY

mediT

Proceedings of the 11th German-Russian-Conference on Biomedical Engineering

June 17th - 19th, 2015
Aachen

CLASSIFICATION MODELS APLICATION TO QUANTITATIVE DIAGNOSTICS OF EARLY STAGE PARKINSON'S DISEASE

K.Yu. Obukhov[2], S.A. Nikitov[1], O. S. Sushkova[1],Ivan Kershner[2], I.A. Maliuta[2], Yu.V. Obukhov[2]
[1] Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya 11-7, Moscow, 125009, Russia,
[2] Moscow Institute of Physics and Technology, Institutski per., 9, Dolgoprudny, 141700 Moscow Region,

Abstract **— New approach to quantitative diagnostics of early stage Parkinson's disease was investigated. There are three main differences of EEG wavelet scalograms, which were considered as early stage PD features. The first is a disordering ridge of PD patient scalogram in frequency range more then ~6 Hz in comparison with that of normal volunteer. The second is a more powerful PD patient cortex electrical activity in frequency range ~4-6 Hz. And the third feature is a scalograms asymmetry of the left and right brain semi spheres. The logistic regression model is built with binary target variable – the disease factor. The importance of every feature was analyzed. The learning dataset consists of 5 major features, which were extracted from the EEG data: the maximum of amplitude of theta rhythm divided on the amplitude of alpha rhythm for C3 and C4 EEG electrodes, the amplitude of tremor for left hand divided on the amplitude of tremor for right hand (or otherwise), the correlation values calculated as in equation and the standard deviation of these correlations. There are 53 observations: 33 patients and 20 normal people.**

Keywords— wavelet spectra, electroencephalogram, electromyogram, accelerometer, frequency synchronization, Parkinson's disease, binary classification, logistic regression.

I. INTRODUCTION (*HEADING 1*)

Parkinson's disease (PD) belongs to a wide range class of neurodegenerative diseases caused by the death of dopaminergic neurons of the brain. Particular attention was paid to the mechanisms of the brain plasticity serving to compensate functional insufficiency of the degenerating neurons [1]. From this point of view, the authors consider the dynamic of neurodegenerative diseases, stating the necessity of the development of preclinical diagnostics and the preventive therapy [2]. The main problem of diagnostics PD is to find out markers of disease at pre clinical and early clinical stages [3].

Electroencephalography (EEG) and electromyography (EMG) are the typical investigations of patient brain electrical activity and diseases. Earlier the decreasing of the dominant rhythm frequency and changes of relative Fourier spectral power of different frequency bands were found with the help of EEG and EMG spectral analysis [4-5].

Disorders of different organism systems, such as movement disorders, vegetative, emotional, psychical and so on, are features of PD. It is assumed that such disorders reflect in electrical activity of brain.

Due to such approach the time-frequency features of spontaneous EEG of early stage PD were investigated with the help of wavelet Morlet transform. Particular attention was paid to EEG theta rhythm (~4-6 Hz), and disordering of alpha rhythm (~8-12 Hz) in brain cortical motor zones.

II. METHODS

Continuous wavelet transform (1) with mother function Morlet (2) was used to get EEG signal x(t) time-frequency power density scalogram [1]:

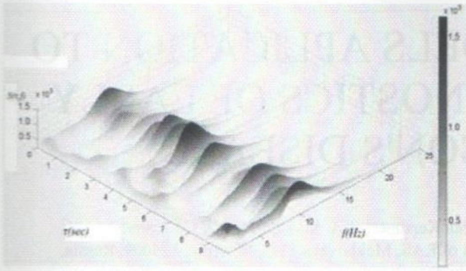
$$S_x(\tau,f) = |W(\tau,f)|^2, \qquad (1)$$

$$W(\tau,T) = \frac{1}{\sqrt{T}}\int x(t)\psi^*\left(\frac{t-\tau}{T}\right)dt \quad , \qquad (2)$$

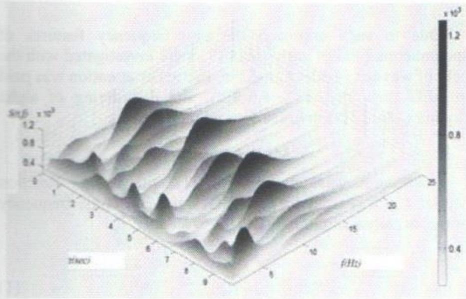
$$\psi(\eta) = \frac{1}{\sqrt{\pi F_b}} e^{2i\pi F_c \eta} e^{-\frac{\eta^2}{F_b}} \quad , \qquad (3)$$

where τ and $f = 1/T$ are time and frequency of scalogram, Fb = Fc = 1.

Fig. 1 illustrates the difference in $S(\tau)$ in the brain motor zone C3 (according to the standard 10x20 scheme of electrodes layout) of brain of the normal volunteer (a) and (b) the patient at the first PD according the qualitative stages of PD described by Hoen-Yahr [2]. Below we will take into account two of them. The first takes into account the arising more powerful activity in theta frequency range (4-6 Hz), and the second one deals with the disorder (non stationary) of electrical activity in the frequency range more then 4-6 Hz.



(a)



(b)

Fig. 1 Time-frequency power density scalograms of normal volunteer (a), and of the 1st stage PD patient (b) of the EEG signals in motor cortex zone C3.

To analyze those features we can consider the scalograms extreme time-frequency distribution. The method of scalograms extreme extraction is written in [6].

Fig. 2 illustrates frequency synchronization of C4 EEG, left hand electromyogram (EMG) and measured with the help of accelerometer left wrist tremor of scalograms extreme of the 1st Hoehn-Yahr [7] stage PD patient. It can be considered as an evidence of the role of 4-6 Hz scalograms peaks in movement disorders.

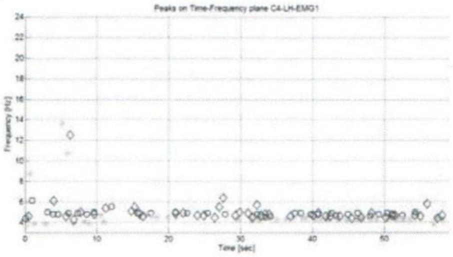
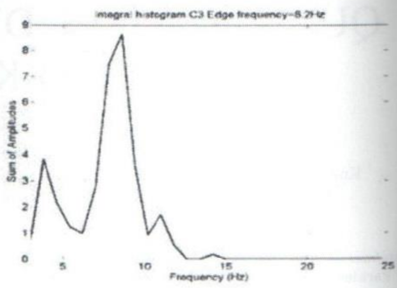
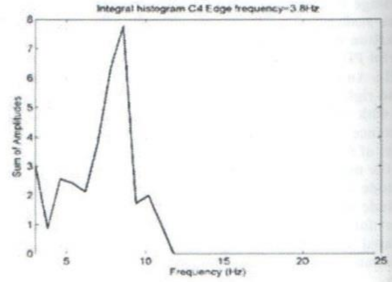


Fig. 2 C4 EEG (circles), left hand electromyogram (rhombs), and left wrist tremor (stars) of scalograms extreme at the 1st stage PD patient

To analyze the scalograms peaks time-frequency distri we consider the histograms of extreme power sums at (Δf, rectangles.



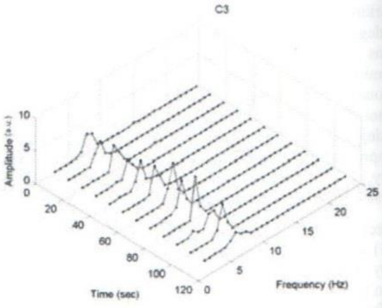
(a)



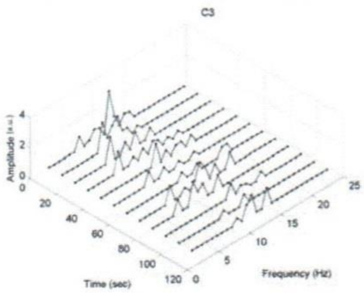
(b)

Fig. 3 Sum of extreme histograms at (0.7 Hz, 180 sec) rectangles for symmetrical C3 (a), and C4(b) EEG electrodes of 1st stage PD patient

Fig. 3 show asymmetry of histograms in 4-5 Hz region of the 1st stage PD patient - the existence of theta rhythm in C3 and the absence of such rhythm in C4 electrodes.



(a)



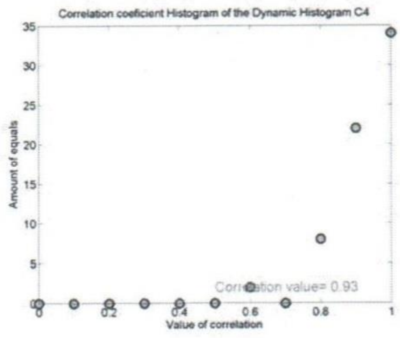
(b)

Fig. 4 Histogram dynamics for the normal volunteer (a), and for the 1st stage PD patient. Histograms was calculated for (0.7 Hz, 10 sec) rectangles

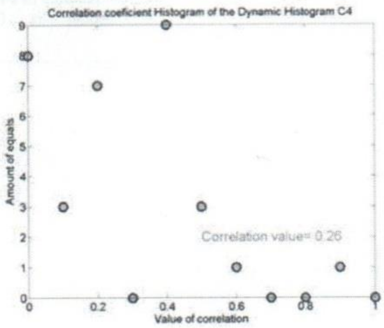
The dynamical histograms calculated at (0.7 Hz, 10 sec) rectangles show the disordering of electrical activity in more then 6 Hz for the 1st stage PD (see fig. 4). Such disordering can be evaluated with the help of dynamical histograms correlation matrixes. This evaluation can be done by histograms of correlation values. Fig. 5 shows the difference of such histograms for the normal volunteer and the 2nd PD patient.

The quantitative feature C can be considered as a weighted sum:

$$C = \frac{\sum_{R_i} R_i \cdot N(R_i)}{\sum_{R_i} N(R_i)}, \qquad (4)$$



(a)



(b)

Fig. 5 Histograms of correlation values for the normal volunteer (a), and the 2nd stage PD patient (b).

where $N(Ri)$ is a quantity of correlation values Ri in correlation matrix. The average correlation value C indicated in fig. 5.

To aggregate the features into one value, which could indicate the probability of PD, the logistic regression model for binary prediction was built. The probability of a particular outcome is linked to the linear prediction function [8]

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_m x_{m,i} \qquad (5),$$

where p_i is the probability of positive outcome for observation i, given $x_{m,i}$ – the feature m in a dataset. Because logistic regression predicts probabilities, rather than just classes, we can fit it using likelihood [9]. For each training data-point, we have a vector of features x_m, and an observed class, y_i. The probability of that class was either p, if $y_i = 1$, or $1 - p$, if $y_i = 0$. The likelihood is then:

$$L = \prod_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{1-y_i} \qquad (6),$$

Maximizing the likelihood function, the weight of each feature can be computed.

To analyze the model efficiency, ROC analysis was made. The Receiver Operating Characteristics (ROC curve) indicated the dependence between Sensitivity – True Positive Rate, and Specificity – True Negatives Rate [10]. The Area Under Curve (AUC) can show the quality of the classification model.

III. CLINICAL RESULTS

The learning dataset consists of 5 major features, which were extracted from the EEG data: the maximum of amplitude of theta rhythm divided on the amplitude of alpha rhythm for C3 and C4 EEG electrodes named AC, the amplitude of tremor for left hand divided on the amplitude of tremor for right hand (or otherwise) named Trem, the correlation values calculated as in equation (4) named C and the standard deviation of these correlations named S. There are 53 observations: 33 patients and 20 normal people.

The ROC curve for model based on all features in shown in Table 1. The AUC of the model is 0.9924.

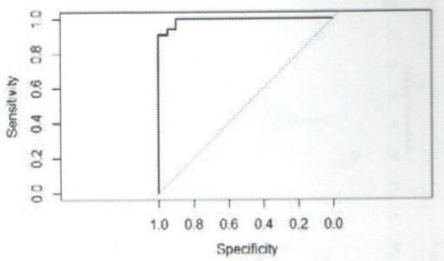


Fig. 6 Receiver Operating Characteristic for logistic regression model

The resulted coefficients in a regression models with Akaike Information Criterion is provided below.

```
Call:  glm(formula = RES ~ AC + Trem + C + S, family = binomial(link
= logit),
    data = Data)

Coefficients:
(Intercept)           AC         Trem            C            S
     -9.224       54.687        3.483       -2.402        2.315

Degrees of Freedom: 52 Total (i.e. Null);  48 Residual
Null Deviance:          70.25
Residual Deviance: 12.05 AIC: 22.05
```

Table 1. Logistic Regression model results: the coefficient weights and AIC

To calculate the importance of every feature in the model, the ROC curve was built for every feature with PD dependent variable. The table of AUC for every feature is shown in Table 2..

Feature	AUC
Amplitude theta/alpha rhythm	0.79
Amplitude of Tremor	0.94
Correlation Value	0.59
Correlations st.dev.	0.48

Table 2. AUC for every feature based models

In addition, various models were built with different features. The results are indicated in Table 3.

Feature	Set 1	Set 2	Set 3	Set 4
Amplitude theta/alpha rhythm	+	-	+	+
Amplitude theta/alpha rhythm	+	+	-	+
Correlation Value	-	+	+	+
Correlations st.dev.	-	-	+	-
AUC Result	0.99	0.95	0.85	0.99

Table 3. Model results on different sets of features

IV. CONCLUSION

Three main features of early stage Parkinson's disease were analyzed: a disordering ridge of PD patient scalogram, powerful PD patient cortex electrical activity in frequency range ~4-6 Hz and a scalograms asymmetry of the left and right brain semi spheres for PD patients. To quantify the discrimination power of these features, classification model was build (logistic regression) with binary dependent variable. The model was built on the dataset of 53 observations: 33 patients and 20 control cases. The highest importance shows amplitude of tremor (AUC 0.94). The model shows AUC 0.99 on training dataset. The results show very high accuracy and low Type I and Type II errors. The probability interpretation of the model can be used as a unique variable that considers all features of PD extracted from EEG data.

This research was supported by Russian Foundation for Basic Research, the project #15-07-07846.

REFERENCES

[1] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Hundington. Clinical, neurological and neurochemical correlations". J. Neurol. Sci. 1973. v. 20. № 4. p. 415-455.

[2] "Neurodegenerative Diseases: Fundamental and Applied Issues" / ed. by M.V. Ugrumov. – Moscow: Nauka, 2010. – ISBN 978-5-02036710-4 (in Russian).

[3] E. Bezard, C.E. Gross, "Compensatory mechanisms in experimental and human parkinsonism: towards a dynamic approach". Prog. Neurobiol. 1998. v. 55. № 2. p. 93-116.

[4] A.C. England, R.S. Schwab, E. Peterson,"The electroencephalogram in Parkinson's syndromee". EEG Clin. Neurophysiol. 1959. v. 11. № 4. p. 723–731.

[5] H.W. Berendse, C.J. Stam, "Stage-dependent patterns of disturbed neural synchrony in Parkinson's disease". Parkinsonism and Related Disorders. 2007. v. 13, Suppl. 3. p. 440–445.

[6] Yu. V. Obukhov, M.S. Korolev, A.V. Gabova, G.D. Kuznetsova, M.V. Ugrumov, "Method of early stage Parkinson's disease electroencephalography diagnostics" // RF patent. - 2484766, 20.06.2013. (in Russian).

[7] M.M. Hoehn, M.D. Yahr, "Parkinsonism: onset, progression and mortality"., // Neurology. - 1967, V. 17, pp. 427-442, PMID 6067254.

[8] Hosmer, David W.; Lemeshow, Stanley (2000). Applied Logistic Regression (2nd ed.). Wiley. ISBN 0-471-35632-8.

[9] Menard, Scott W. (2002). Applied Logistic Regression (2nd ed.). SAGE. ISBN 978-0-7619-2208-7.

[10] Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". Remote Sensing of Environment 62 (1): 77–89. doi:10.1016/S0034-4257(97)00083-7