

Robust linear predictor as EEG fluctuation analyzer in diagnosis of Alzheimer's disease

DAGMAR ZACHOVÁ, JAROMIR KUKAL and OLDŘICH VYŠATA

The paper is oriented to EEG signal analysis, which is focused to quasi-stationarity hypothesis that the statistical properties of the channel signal fluctuate in time. Robust linear predictor is used for short segments of EEG as low-pass filter and the difference between the raw EEG and filter output was subject of statistical testing. Novelty is in the fluctuation measurement which enables to classify the Alzheimer's disease patients against controls.

Key words: Alzheimers disease, EEG, quasi-stationarity, linear predictor, robust identification

1. Introduction

Quasi-stationarity of EEG signal can cause difficulties in any signal processing of long sequences. If we divide the original series to short segments of constant length, we can use traditional methods of statistical analysis within any individual segment. Thus, the statistical properties of individual segments can be estimated correctly when the segment length is less then two seconds (in the case of EEG). But the statistical properties of segments vary in time due to the quasi-stationarity of EEG signal. The paper is oriented to statistical analysis of these fluctuations and their robust ranges.

2. EEG in diagnosis of Alzheimer's disease

The EEG is usually used in the diagnosis and evaluation of many cortical and sub-cortical dementias. Often it can help to differentiate between a degenerative disorder such as Alzheimer's disease (AD) and pseudodementia due to psychiatric illness. In AD the temporooccipital alpha rhythm slows down. Power spectra of delta and theta waves increases while beta activity may decreases. Coherence usually decreases mainly in beta and alpha bands.

D. Zachová (e-mail: zachovad@yahoo.com) and J. Kukul are with FNSPE, CTU in Prague, Czech Rep. O. Vyšata is with NEUROPS Ltd., Rychnov nad Kněžnou, Czech Rep.

Received 22.04.2011.

3. Signal description

The multichannel EEG is a traditional tool for the investigation of human brain activity. The electrode signal was recorded with constant frequency $f_s = 200$ Hz and then digitalized to the raw EEG time series X_k for $k = 1, 2, \dots, L$.

The signal was partitioned to nonoverlapping segments of constant length $N \ll L$. Ideal signal should have stationarity property in the meaning that the statistical properties [6] of short segments don't vary in time. From the statistical point of view, the stationarity of EEG is observable only for short sequences up to 2 seconds, thus for $N \ll L < 2f_s$. When the EEG scan is too long then the stationarity hypothesis falls. In this case, the EEG quasi-stationarity was subject of investigation. We used $N \ll 2f_s$ to guarantee interval stationarity of individual segments. Then the robust predictive filter was applied to every segment. The difference between the original data and the prediction was subject of statistical analysis. Various statistics of segment error sample were used and their values changed from segment to segment. Thus, the new time series of length $M = \lfloor L/N \rfloor$ of segment characteristics arisen and its members are R_k for $k = 1, 2, \dots, M$. Statistical analysis of fluctuations is based on various statistical characteristics of R_k series. The process of EEG signal analysis consists of four steps:

- segmentation with X_k as result;
- within segment prediction with e_k as result;
- within segment error analysis with R_k as result;
- fluctuation analysis with Q_k as result.

4. Robust predictive model

We consider a basic linear model [4] in the form

$$Y_{k+S} = \sum_{j=1}^H \beta_j \varphi_j(Y_k, \dots, Y_{k-H+1}) + \varepsilon_{k+S} \quad (1)$$

where

- N is the length of the time series segment (the number of observations);
- H is the history length of time series;
- S is the prediction step length;
- Y_1, Y_2, \dots, Y_N are observations within given segment;
- $\beta_1, \beta_2, \dots, \beta_H$ are unknown coefficients (parameters) of the model;

- $\varphi_1(Y_k, \dots, Y_{k-H+1}), \varphi_2(Y_k, \dots, Y_{k-H+1}), \dots, \varphi_H(Y_k, \dots, Y_{k-H+1})$ are polynomial functions;
- ε_{k+S} is the random noise.

When we transcribe (1), we obtain an equation system that could be described in matrix form as

$$\begin{bmatrix} Y_{H+S} \\ Y_{H+1+S} \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} \varphi_{1,2,\dots,H}(Y_H, \dots, Y_1) \\ \varphi_{1,2,\dots,H}(Y_{H+1}, \dots, Y_2) \\ \vdots \\ \varphi_{1,2,\dots,H}(Y_{N-S}, \dots, Y_{N-S-H+1}) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_H \end{bmatrix} + \begin{bmatrix} \varepsilon_{H+S} \\ \varepsilon_{H+1+S} \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad (2)$$

in other words

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (3)$$

It is significant that the number of equations (degrees of freedom) must be higher than the number of estimated coefficients, i.e. $N - H - S + 1 > H$. Further, supposed that $E(\boldsymbol{\varepsilon}) = \bar{\mathbf{0}}$, where symbol E indicates the expected value. Providing this we can express estimated values Y_{k+S} (for $k = H, H + 1, \dots, N - S$) through the following formula

$$E(Y_{k+S}) = \sum_{j=1}^H \beta_j \varphi_j(Y_k, \dots, Y_{k-H+1}). \quad (4)$$

These estimated values are equal to functional values of selective regression function

$$\hat{Y}_{k+S} = \sum_{j=1}^H b_j \varphi_j(Y_k, \dots, Y_{k-H+1}) \quad (5)$$

where

- b_j is the scatter estimate of unknown parameter β_j (for $j = 1, 2, \dots, H$);
- \hat{Y}_{k+S} is the predicted value Y_{k+S} (for $k = H, H + 1, \dots, N - S$).

Equation system (5) can be described in matrix form as

$$\begin{bmatrix} \hat{Y}_{H+S} \\ \hat{Y}_{H+1+S} \\ \vdots \\ \hat{Y}_N \end{bmatrix} = \begin{bmatrix} \varphi_{1,2,\dots,H}(Y_H, \dots, Y_1) \\ \varphi_{1,2,\dots,H}(Y_{H+1}, \dots, Y_2) \\ \vdots \\ \varphi_{1,2,\dots,H}(Y_{N-S}, \dots, Y_{N-S-H+1}) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_H \end{bmatrix}, \quad (6)$$

in other words

$$\hat{\mathbf{y}} = \mathbf{\Phi}\mathbf{b}. \quad (7)$$

We can use robust methods for the coefficient estimating of model (3), i.e. for vector \mathbf{b} specification (see 4.1).

The difference between observed and predicted value is called residue and denoted as vector \mathbf{e} . The residue in given point is equal to $e_i = \hat{Y}_i - Y_i$, therefore for the model (1) the residual vector has the form of

$$\mathbf{e} = \begin{bmatrix} e_{H+S} \\ e_{H+1+S} \\ \vdots \\ e_N \end{bmatrix}. \quad (8)$$

4.1. Robust identification techniques

Robust techniques of parameter estimating represent the alternative to classical statistic methods that are very sensitive to outliers in input data. We know several types of robust estimates, namely: L-estimates, R-estimates and M-estimates. It is most suitable to apply M-estimates, the pioneer of which was Huber [2]. M-estimate of model coefficients $\mathbf{\beta}$ is defined via function minimization (with respect to \mathbf{b})

$$\psi(\mathbf{b}) = \sum_{i=H}^{N-S} \rho\left(\frac{e_{i+S}}{\sigma}\right) = \sum_{i=H}^{N-S} \rho\left(\frac{Y_{i+S} - \Phi_{i-H+1}^T \mathbf{b}}{\sigma}\right) \quad (9)$$

where

- T is transposition symbol;
- ρ is a penalty function (see Tab. 1);
- σ is standard deviation;
- Φ_i is i -th row of the matrix Φ .

When implementing the weight function defined as $w(\xi) = \frac{d\rho(\xi)}{d\xi} \frac{1}{\xi}$ (see Tab. 1), satisfying $w(0) = 1$ and substituting to the Taylor series of (9) we obtain a method of weighted least squares (WLS) [4]

$$\sum_{i=H}^{N-S} w\left(\frac{e_{i+S}}{S_N}\right) Y_{i+S} \Phi_{(i-H+1)j} = \sum_{i=H}^{N-S} \sum_{k=1}^H w\left(\frac{e_{i+S}}{S_N}\right) \Phi_{(i-H+1)j} \Phi_{(i-H+1)k} b_k$$

where $j = 1, \dots, H$. The method of WLS consists in implementation of the following operations:

1. initial estimate of \mathbf{b} by means of method of least squares, iteration counter set to $l = 1$.

2. residue specification \mathbf{e} in l th iteration;
3. calculation of weights and then $l = l + 1$;
4. specification of parameters $\mathbf{b}^{(l)}$ (estimate of vector \mathbf{b} in l th iteration) and residue specification.

If the estimates $\mathbf{b}^{(l)}$ a $\mathbf{b}^{(l-1)}$ are not close enough, we repeat the steps 3 and 4. It is important when calculating the balance in step 3 that the robust estimate of standard deviation σ is not recalculated, i.e. it's specified on the basis of error residue \mathbf{e} after the least squares method application. Such a \mathbf{b} , by which the penalty function reached the lowest value, is considered as the best estimate of parameter β .

The question is how to get the robust estimate of standard deviation σ . There is statistics $\sigma^* = MAD_E/0.6745$ most frequently used in practice, where MAD_E stands for median of E_1, E_2, \dots, E_N and $E_i = |e_i - \tilde{E}|$, \tilde{E} is median of e_1, e_2, \dots, e_N .

Table 1. Robust approaches

method	$\rho(\xi)$	$w(\xi)$	range	constant
Tukey	$B^2 \left(1 - \left(1 - (\xi/B)^2\right)^3\right) / 6$	$\left(1 - (\xi/B)^2\right)^2$	$ \xi \leq B$	$B=4.865$
	$B^2/6$	0	$ \xi > B$	
Huber	$\xi^2/2$	1	$ \xi \leq k$	$k=1.345$
	$k \xi - k^2/2$	$k/ \xi $	$ \xi > k$	
Andrews	$A^2 (1 - \cos(\xi/A))$	$(A/\xi)\sin(\xi/A)$	$ \xi \leq A\pi$	$A=1.339$
	$2A^2$	0	$ \xi > A\pi$	
Welsch	$W^2 \left(1 - \exp\left(-(\xi/W)^2\right)\right) / 2$	$\exp\left(-(\xi/W)^2\right)$	—	$W = 2.985$
Talwar	$\xi^2/2$	1	$ \xi \leq k$	$k = 2.795$
	$k^2/2$	0	$ \xi > k$	

4.2. Statistical analysis of prediction error and time fluctuation

Let us have signal of length L , divided into segments of fixed length N and values H, S being set. Afterwards, we effect suitable robust identification of model (1), coefficient and indicate residue vector $\mathbf{e} = (e_1, e_2, \dots, e_p)^T$. Now, it's time to think of how to characterize error prediction in one segment and how best to characterize variability of error prediction of the whole signal in time. In kind of criterion featuring as total error prediction in one segment the following two characteristics can be used. The first one can be described through the relation

$$R = (E|\mathbf{e}|^q)^{1/q} = \left(\frac{1}{p} \sum_{k=1}^p |e_k|^q\right)^{1/q} \tag{10}$$

where $q \in \langle 0, \infty \rangle$ a $p = N - H - S + 1$. Let's identify this method as a method of root of expected value of residues (MREVR(q)).

In order to make description of the second characteristic easier let us set $a_k = |e_k|$ and let us arrange a_k in such a way that $a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(p)}$. Afterwards, the total signal error prediction in one segment will be calculated as

$$R = a_{(\lfloor qp \rfloor)} \quad (11)$$

where parameter $q \in (0, 1)$ and $p = N - H - S + 1$. This approach we call the method of quantiles of the residues (MQR(q)).

Thus, we get for each channel time series prediction errors $\{R_1, R_2, \dots, R_M\}$, respectively structured selection $\{R_{(1)}, R_{(2)}, \dots, R_{(M)}\}$ where $M = \lfloor L/N \rfloor$. For assess the variability of the prediction errors EEG signal in time can be used such as one of the following sample (segment) characteristics :

- maximum $R_{\max} = \max \{R_1, R_2, \dots, R_M\}$;
- minimum $R_{\min} = \min \{R_1, R_2, \dots, R_M\}$;
- range $R_R = R_{\max} - R_{\min}$;
- mean $\bar{R} = \frac{1}{M} \sum_{k=1}^M R_k$;
- standard deviation $\sigma = \sqrt{\frac{1}{M-1} \sum_{k=1}^M (R_k - \bar{R})^2}$;
- median $\tilde{R} = \frac{1}{2}(R_{(M/2)} + R_{(M/2+1)})$ for the even M , respectively $\tilde{R} = R_{((M+1)/2)}$ for the odd M ;
- median absolute deviation $MAD_Z = \tilde{Z}$ where \tilde{Z} stands for median of Z_1, Z_2, \dots, Z_M and $Z_i = |R_i - \tilde{R}|$ for $i = 1, 2, \dots, M$;
- 1st quartile (lower quartile) $R_{0.25} = R_{(\lfloor 0.25M \rfloor)}$;
- 3rd quartile (upper quartile) $R_{0.75} = R_{(\lfloor 0.75M \rfloor)}$;
- interquartile range $IQR = R_{0.75} - R_{0.25}$.

These aggregating characteristics will be denoted as Q in the next text.

4.3. Quality of classification

There is a direct relationship between the quality of parameter setting and the quality of classification. In our case, the optimal parameter setting has greatest differences in Q between groups AD and CN. The quality of parameter setting was driven by the apparatus of statistical hypothesis testing. Variability of the prediction error we calculated for each channel and each person. Thus, there are two samples

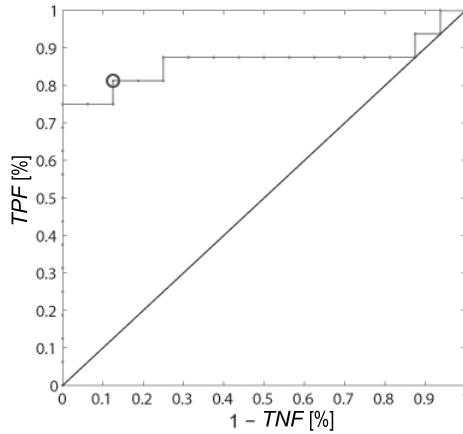


Figure 1. ROC curve of 2nd channel for interquartile range (Welsch's method, MREVR(q), N=150, H=10, S=1)

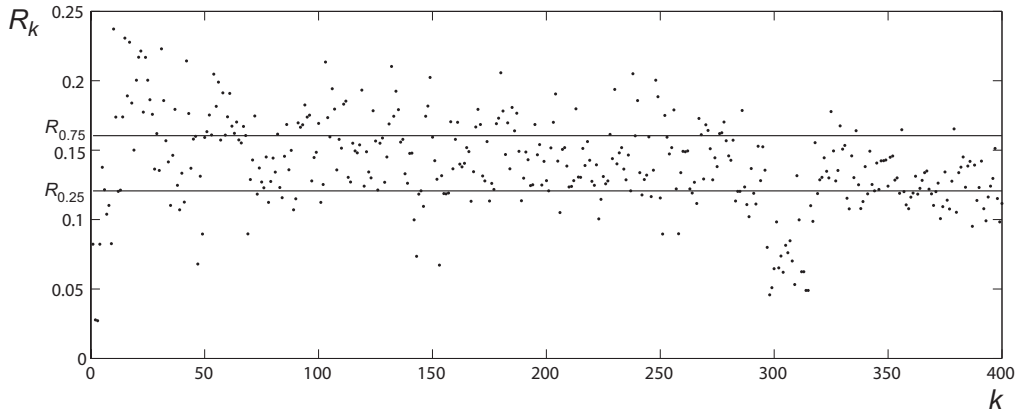


Figure 2. Fluctuation of MREVR in time for a healthy person ($IQR = R_{0.75} - R_{0.25}$)

$Q_{AD} := \{Q_1^{AD}, Q_2^{AD}, \dots, Q_n^{AD}\}$ and $Q_{CN} := \{Q_1^{CN}, Q_2^{CN}, \dots, Q_m^{CN}\}$ where n and m indicate the number of individuals in AD and CN groups. The null and the alternative hypothesis were constructed as follows:

- H_0 : expected value of random variables Q_{AD} , Q_{CN} are not different, i.e. $\mu_{AD} = \mu_{CN}$;
- H_1 : expected value of random variables are different, i.e. $\mu_{AD} \neq \mu_{CN}$.

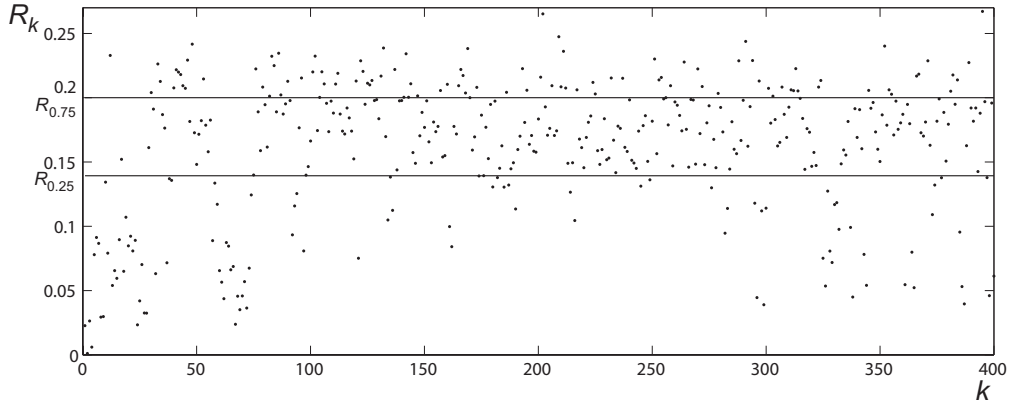


Figure 3. Fluctuation of MREVR in time for a patient with Alzheimer's disease ($IQR = R_{0.75} - R_{0.25}$)

Assuming equal variances in both groups, we can use the two-sample two-sided t-test [4], where the test criterion is calculated as

$$T = \frac{\overline{Q}_{AD} - \overline{Q}_{CN}}{\sqrt{(n-1)\sigma_{AD}^2 + (m-1)\sigma_{CN}^2}} \sqrt{\frac{mn(n+m-2)}{n+m}}. \quad (12)$$

\overline{Q}_{AD} and \overline{Q}_{CN} denote the sample means, σ_{AD}^2 a σ_{CN}^2 are sample variances.

The criterion (12) has Student's t-distribution with $df = n + m - 2$ degrees of freedom. We calculated adequate p -value for given T and df .

Another possible tool for assessing the quality of classifiers is the sensitivity and specificity. Sensitivity reflects the probability of correct classification of positive sample (AD) and specificity reflects the probability of correct classification of negative sample (CN).

Let TP (true positive) be number of samples that the classifier correctly classified into AD, let FP (false positive) be number of samples that the classifier incorrectly classified into AD, let TN (true negative) be number of samples that the classifier correctly classified into CN and let FN (false negative) be number of samples that the classifier incorrectly classified into CN. The sensitivity and the specificity can be estimated as follows:

- sensitivity (*true positive fraction*) $TPF = \frac{TP}{TP+FN}$;
- specificity (*true negative fraction*) $TNF = \frac{TN}{FP+TN}$.

The optimum, threshold for AD / CN classification is obtainable from ROC curve [1] as compromise between maximum values of TPF and TNF. We prefer to maximize $\min(TPF, TNF)$ according to minimax decision principle.

5. Results

There were 32 EEG records included in our study. The groups of AD and CD consist of 16 and 16 patients. We used international 10-20 electrode system [5]. During the measurement of electrical activity, our subject were examined in the bed having with closed eyes and without any external stimulus. Manually selected low-artifact segments of EEG data were approximately 300 seconds long with sampling frequency of 200 Hz. Only anti-aliasing analogue filter was used during data preprocessing. Electric potential was measured in millivolts.

Table 2. Minimum p -values for the default setting

channel	p -value	characteristic	method (q)
1	0.003301	IQR	MQR(1/2)
2	0.000201	IQR	MREVR(2)
3	0.070778	R_{\min}	MQR(1/2)
4	0.077238	IQR	MREVR(2)
5	0.013298	IQR	MREVR(2)
6	0.001757	IQR	MREVR(2)
7	0.002733	MAD	MQR(1/2)
8	0.182820	R_{\min}	MQR(1/2)
9	0.081693	R_{\min}	MREVR(2)
10	0.118012	IQR	MREVR(2)
11	0.113751	IQR	MQR(1/2)
12	0.045587	IQR	MKR(1/2)
13	0.012805	R_R	MREVR(2)
14	0.047399	R_{\min}	MREVR(2)
15	0.052538	R_{\min}	MKR(1/2)
16	0.378503	R_{\min}	MREVR(2)
17	0.231664	R_{\min}	MREVR(2)
18	0.101802	σ	MREVR(2)
19	0.137812	R_{\min}	MREVR(2)

During computer experiments, which we aimed to optimum parameter setting, we used model (1) with fixed functional base $\varphi_j(Y_k, \dots, Y_{k-H+1}) = Y_{k-j+1}$ for $j \in \{1, 2, \dots, H\}$. The following procedure was used:

- signal was divided into segments ($N = 100$);

- standardization of each segment was performed $Y_i^* = \frac{Y_i - \bar{Y}}{\sigma}$;
- default values of model parameter were used ($H = 10, S = 1$);
- Tukey's method was used as default robust method;
- default value for MREVR was $q = 2$;
- default value for MQR was $q = 1/2$;
- two most suitable channels were chosen on the basis of two-sample two-sided t-test at significance level of 0.05;
- with the help of p -value, optimal values for parameters N, H, S and q were found, and most suitable robust method was chosen.

Results of numerical calculations are included in the Tab. 2 using default parameters. Bold font was used for p -value below critical probability (0.05). The best in AD / CN resolution are channel 2 and 6, which were subject consequential analysis. The second aim was to study the influence of processing parameters (N, H, S) to p -value. Following parameter values were involved in the combination with Tukey's method:

- length of the segment $N=100, 125, 150, 200$;
- history length of time series $H=6, 8, 10$;
- length of the prediction step $S=1, 2, 3$.

The results of testing are summarized in the Tabs. 4, 5. The best results were obtained for $N = 150, H = 8$ or $10, S = 1$ or 2 in the case of Turkey's method and channels 2 and 6. The parameter setting was then used for the other methods and channels. As seen in the Tab. 3, the p -values of robust methods are lower than in the squares approach (LSQ) in the case of channel 6. Similar result (except Andrew's and Huber's method) is valid in the case of channel 2 (see Tab. 3). The method MREVR is recommended for the segment error evaluation. The methods MAD and IQR are the best for the fluctuation analysis.

6. Conclusion

Robust linear predictive filter was used for the characterization of signal variability within individual segments. The quasi-stationarity analysis is recommended as a tool for the classification of Alzheimer's disease against controls. The best results were obtained on EEG channel 2 with sampling period 200 Hz, segment length $N = 150$, history depth $H = 10$, step of prediction $S = 1$, Welsch's method, MREVR (method of root of expected value of residues) characteristics of EEG fluctuations. Then, the adequate optimum values are:

Table 3. Minimum p -values for the 6th channel and different robust methods

robust method	p -value	q	method	$N - H - S$	characteristic
LSQ	0.000339	3/2	MREVR	150 - 8 - 2	IQR
Tukey	0.000326	2	MREVR	150 - 8 - 2	IQR
Andrews	0.000315	2	MREVR	150 - 8 - 2	IQR
Huber	0.000212	9/4	MREVR	150 - 8 - 2	IQR
Welsch	0.000299	2	MREVR	150 - 8 - 2	IQR
Talwar	0.000337	5/4	MREVR	150 - 8 - 2	IQR

Table 4. Minimum p -values for the 2nd channel and Tukey's method

N	p -value	H	S	characteristic	method(q)
100	0.000236	10	2	R_{\min}	MREVR(2)
	0.000320	8	1	IQR	MQR(1/2)
125	0.000261	10	1	IQR	MREVR(2)
	0.000283	10	2	σ	MQR(1/2)
150	0.000071	10	1	IQR	MREVR(2)
	0.000225	8	1	IQR	MQR(1/2)
175	0.000127	10	1	IQR	MREVR(2)
	0.000427	10	1	IQR	MQR(1/2)
200	0.000171	10	1	MAD	MREVR(2)
	0.000495	8	1	IQR	MQR(1/2)

- p -value = 6.11×10^{-5} ;
- sensitivity $TPF = 81.3\%$;
- specificity $TNF = 87.5\%$.

From the biomedical point of view the novel method gives better sensitivity and specificity comparing conventionally used quantitative methods. It will be verified on bigger sample of AD patients. Sensitivity to the degree of mental deficit will be established. Comparison of the groups of cortical and subcortical dementia and pseudodementia in depression will be made. This method may improve early diagnosis of Alzheimer's disease.

Table 5. Minimum p -values for the 6th channel and Tukey's method

N	p -value	H	S	characteristic	method (q)
100	0.000683	8	1	IQR	MREVR(2)
	0.001112	8	1	IQR	MQR(1/2)
125	0.001047	8	1	IQR	MREVR(2)
	0.002533	10	1	σ	MQR(1/2)
150	0.000326	8	2	IQR	MREVR(2)
	0.000764	8	1	IQR	MQR(1/2)
175	0.000591	8	2	MAD	MREVR(2)
	0.002549	8	1	IQR	MQR(1/2)
200	0.001146	6	3	$R_{0.25}$	MREVR(2)
	0.004374	8	1	IQR	MQR(1/2)

Table 6. Minimum p -values for the 2nd channel and different robust methods

robust method	p -value	q	method	$N - H - S$	characteristic
LSQ	6.59×10^{-5}	7/8	MQR	150 - 8 - 1	IQR
Tukey	6.32×10^{-5}	9/4	MREVR	150 - 10 - 1	IQR
Andrews	6.74×10^{-5}	9/4	MREVR	150 - 10 - 1	IQR
Huber	6.66×10^{-5}	9/4	MREVR	150 - 10 - 1	IQR
Welsch	6.11×10^{-5}	9/4	MREVR	150 - 10 - 1	IQR
Talwar	6.47×10^{-5}	7/8	MQR	150 - 8 - 1	IQR

References

- [1] T. FAWCETT: An Introduction to ROC analysis. In '*Pattern Recognition Letters (Amsterdam, 2006)*', 27 Elsevier Science, (2006), 861-876.
- [2] P.J. HUBER and E.M. RONCHETTI: Robust Statistic. John Wiley & Sons, Hoboken, 2009.
- [3] R.G. LEHR and A. PONG: ROC Curve. In '*Encyclopedia of Biopharmaceutical Statistics (New York, 2003)*', Marcel Dekker, (2003), 884-891.
- [4] M. MELOUN and J. MILITKÝ: The statistical analysis of experimental data . Academia, Praha, 1998.

- [5] E. NIEDERMEYER and F. LOPES DA SILVA: *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, Philadelphia, 2004.
- [6] M.B. PRIESTLEY: *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London, 1988.