

# Improved robust handling of electromyograms with mining of new diagnostic signs

Gennady Chuiko<sup>a</sup>, Yevhen. Darnapuk<sup>a</sup>, Olga Dvornik<sup>a</sup>, and Yaroslav Krainyk<sup>a</sup>

<sup>a</sup> Petro Mohyla Black Sea National University, 68 Desantnikov St.,10, Mykolayiv, 54003, Ukraine

## Abstract

The research proposes an improved and more reliable way of computer processing raw electromyograms (EMGs). The main steps of this method are pre-filtering with Haar wavelets, full-wave rectification, followed by post-filtering with a digital filter with a moving median (MM). MM digital filtering is probably the first time applied to EMG.

The proposed method is more reliable than the commonly used Root Mean Square (RMS) method. RMS windowing turns out to be invalid by the effect of massive outliers within the EMG. As shown in the article, this typical deviation from the Gaussian distribution occurs for many untreated EMGs. The paper also suggests a few new diagnostic hallmarks. These are the median and peak values of the smoothed (by MM-filter) EMG and their ratio.

## Keywords 1

electromyograms, massive outliers, robust computer processing, moving median filter, new diagnostic signs

## 1. Introduction

Electromyograms (EMGs) reflect weak electrical signals from muscle and nerve cell arrays [1, 2, 3]. Most medical cues, as well as EMGs, are variable. Such variability is the effect of a human body's dynamical balance with a fickle medium. That is quite natural in the framework of the homeostasis concept of living organisms [3]. So, the study of the variabilities of biomedical signals gives us insight into the adaptive abilities of specific body systems within the healthy state or at pathologies.

Besides, another inborn trait of many biomedical signals [4], EMGs too [3, 5], is the marked non-Gaussianity of their probability distributions. That can manifest itself in various ways: heavy-tailed (high-skewed [4]) distributions, weighty kurtosis, multi-modality, arrays of outliers, etcetera. [6]. Clinicians use the raw EMGs relatively rarely. As a rule, EMGs are processed according to some algorithm, more or less sensitive to non-Gaussianity [7]. Only then are they analyzed, and the clinical decision is made.

The steady request for robust gauges of variability that are low-sensible to the non-Gaussianity exists in clinical practice. An interquartile range and a median are, for instance, the most robust measures for the variability and central tendency among the statistics indexes in this sense [6]. The new and reliable diagnostic signs are welcome, too, allowing the binary sorting of the "norm-pathology" type.

## 2. Related Works and problem statement

Authors [5] have characterized the non-Gaussianity of stimulated surface EMGs by abnormal high kurtosis of the probabilities distributions. The level of deviations from the Gauss distribution depended on the shape of used electrodes and filters. Other visual and convincing ways are offered in [6]. Those are statistical box plots and so-called "probability-probability" (P-P) plots.

---

ITTAP'2021: 1nd International Workshop on Information Technologies: Theoretical and Applied Problems, November 16–18, 2021, Ternopil, Ukraine

EMAIL: genchuiko@gmail.com (A. 1); yevhen.darnapuk@chmnu.edu.ua (A. 2); olga.v.dvornik@gmail.com (A. 3), yaroslav.krainyk@chmnu.edu.ua (A. 4)

ORCID: 0000-0001-5590-9404 (A. 1); 0000-0002-7099-5344 (A. 2); 0000-0002-4545-1599 (A. 3), 0000-0002-7924-3878 (A. 4)



© 2021 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Author [7] has been considered two popular algorithms of EMGs processing. Both include an alternative stage: either full-wave rectification with digital filtration or root mean square (RMS) filter. The author has warned that popular RMS smoothing of EMGs works correctly only for signals with distributions close to normal Gauss one. Under the condition, if outliers occur more often than predicted by the normal distribution, full-wave rectification with the digital filtration by moving average is more robust. The moving median filter is even preferable, but it is practically unused [7].

The handbook [8] consults the range (50-500) ms for the smoothing windows. Authors [9] admit the RMS window lengths from 10 ms up to 1000 ms. Neither statistical significance nor clinical relevance of the smoothed EMG mean was affected by such manipulation of window length (variation did not exceed 3%). The smoothed EMG's peak value turned out a little more sensitive (variation less than 9%). However, the variability indexes were more sensitive to a window length.

It is worth noting that clinicians wide use the mean and peak values of rectified and smoothed EMGs as diagnostic signs [9]. It would be wise to exploit medians of smoothed EMGs instead of habitual mean keeping the robustness in mind.

The papers [3,10] considered the denoising or separation of a signal on the low-frequency part (the signal itself) and the high-frequency one (the noise). It needs the additive "signal+noise" model for EMGs, of course. Then such separation is easily accessed with Haar's wavelets [3,10]. The exponential smoothing model of a time series [11], like a raw EMG, for example, permits to get the "Errors (noise)-Trend-Seasonality" (ETS) pattern for this series. Starting from the Maple 18 version, this model is well-programmed within the program package "Time Series Analysis." [12].

Good is namely that way of medical signal processing, accompanied by the recommendation about a single one or even a short vector of diagnostics features [13]. If one shares a practiced clinician point of view, of course. Thus, our approach intends to ensure reliable methods for processing medical signals with direct indications of diagnostic signs.

This research aims to work out improved robust processing of EMGs with the account of their non-Gaussianity. The following stages have to be included in that work:

1. Testing and estimation of non-Gaussianity of probability distributions of raw EMGs with estimations of outliers fractions as a possible diagnostic sign.
2. Definition of the ETS model for raw EMFs.
3. Separation of signal and noise and estimation of Signal-to-Noise ratio (SNR) as a possible diagnostic sign.
4. Full-wave rectification and first-time robust moving median smoothing of EMGs in comparison with the RMS method.
5. Estimating and comparing medians and peak-values, peak-to-medians ratios, and variabilities as clinical diagnostic signs for RMS and moving medians (MM) windowing methods.

### **3. Methodology side**

#### **3.1 Initial data provenance and description**

The database [14] was the source of three samples of raw EMGs. Data were collected with a Medelec Synergy N2 EMG Monitoring System (Oxford Instruments Medical, Old Woking, United Kingdom).

A healthy patient (age 44-), a patient with verified myopathy (age 57), and a patient with verified neuropathy (age 62), all-male, were investigated. The EMG records have different lengths: 50860, 110336, and 147858 samples, respectively [3,14]. The pointers "Healthy," "Myopathy," and "Neuropathy" will be in use further in the text to denote these patients and their indicators

A 25 mm concentric needle electrode was placed into the anterior tibialis muscle of each subject. The patient was then asked to dorsiflex the foot gently against resistance. The needle electrode was repositioned until motor unit potentials with a rapid rise time were identified by the device. Data were then gathered for some short time, at which point the patient was asked to relax, and the needle removed.

The data were recorded at 50 kHz and then downsampled to 4 kHz as the final sample rate of raw EMGs. Two analog filters were used during the recording process: a 20 Hz high-pass filter and a 5KHz

low-pass filter. ADC (analog to digital converter) resolution: was 16 bits. The gain: 10000 ADC units/mV [14].

### 3.2 Computer processing toolbox

Authors have performed this research in the general framework of Maple 2020, one of the last versions of a well-known computer Math system [12]. According to the list of project tasks from section 2, the following program packages were in use: "Statistics," "TimeSeriesAnalysis," and "SignalProcessing."

Sometimes, authors needed to write short programs (procedures) using a high-level handy Maple programming language. For instance, it was so at estimations of outliers weights or the separation of raw EMGs on inner data and outliers. That was necessary for the framework of the accepted approach (see section 2): robust estimations and search for diagnostic hallmarks.

## 4. Results and discussions

### 4.1 Normality testing of raw EMGs and outliers weight's estimation

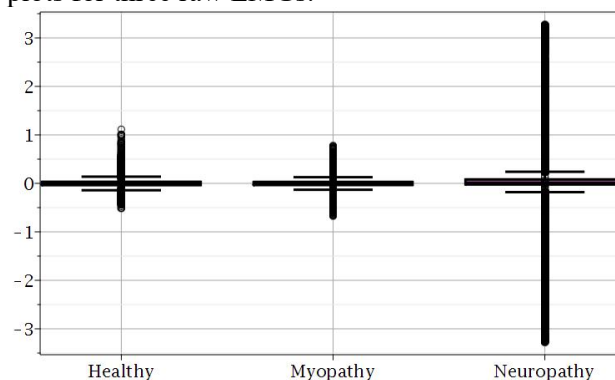
Let start the normality testing with kurtosis of probabilities distributions following [ 5]. There are different definitions of kurtosis. Let use here the quantity excess kurtosis, which is zero equal for the normal Gauss distribution. The excess kurtosis can be easily obtained by subtracting three from ordinary kurtosis. The ratio of the fourth cumulant of distribution to its variance defines this last.

The computing gave such a series: 11.1, 6.4, and 22.0 in the order "Healthy," "Myopathy," "Neuropathy." One can see the positive excess kurtosis that is inherent in leptokurtic distribution. Leptokurtic distributions predict heavy tails on either side. That indicates outliers' great arrays.

The statistical box plot can visualize arrays of outliers. In the box plot method, a data set is split into quartiles, and the difference between the first and the third quartiles defines the height of the box [6,15]. Box plots (also called box-and-whisker plots or box-whisker plots) give an excellent graphical image of the concentration of the data. They also show how far the extreme values (outliers) are from most of the data.

A box plot is constructed from five values: the minimum value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum value. We use these values to compare how close other data values are to them. The two vertical lines (called whiskers) outside the box extend to the smallest and largest observations within the range  $Q1(Q3) \pm 1.5 \times IQR$  (interquartile range).

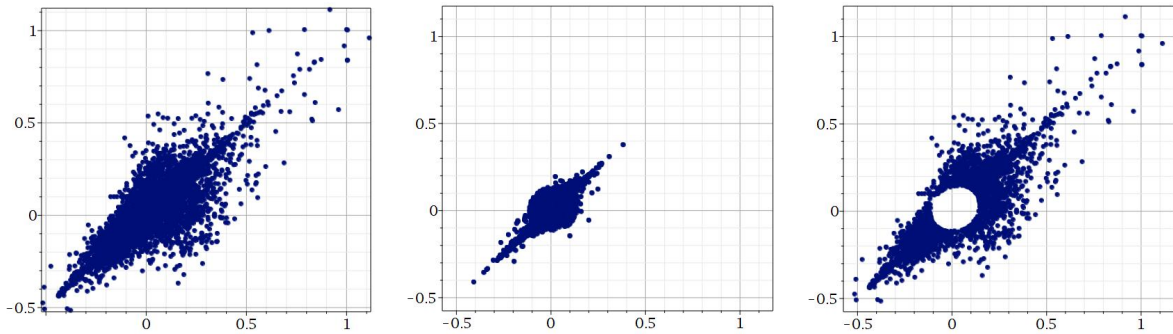
If there are no outliers, then the whiskers extend to the min and max values. Observations  $1.5 \times IQR$  greater than Q3 or less than Q1 should be considered outliers (these are so-called inner fences [6]). Outliers can point us to exciting distribution patterns and should not be removed as "bad" data from the start. Fig 1 shows the box plots for three raw EMGs.



**Figure 1:** Box-and-whisker plots for three raw EMGs showing outliers as bold vertical lines from separate points. The medians are practically unseen as horizontal lines near zero, but one can insight narrow interquartile ranges (variabilities) and relatively tight ranges between maximal and minimal values {whiskers}.

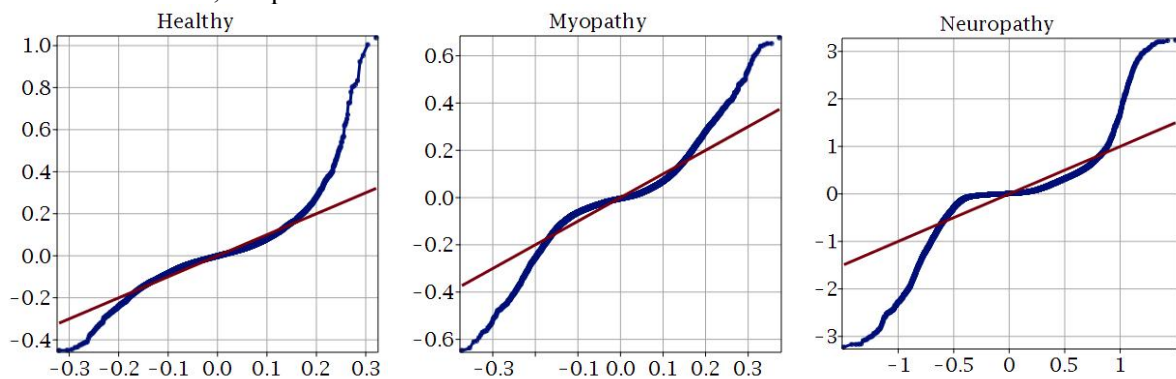
The interquartile ranges (IQRs), as the robust measure of variability, forms such a series: 70, 65, and 105 (in  $\mu\text{V}$ ) in the order "Healthy," "Myopathy," "Neuropathy." Fig. 1 shows many outliers; they are the values out of setting ranges. The lot of outliers is the first hallmark of a non-Gauss distribution of probability.

The minor program (procedure) scripted on Maple language has allowed us to count the percentages of outliers for each of the EMGs. At a glance, they are not extremely high: 7.45, 12.05, and 21.45 in the previous order. Another procedure has allowed us to separate EMGs on the inner data and outliers. Unexpectedly, it turned out that outliers define the form and area of Poincare plots. Figure 2 illustrates this statement.



**Figure 2:** Poincare plots for the healthy patient with a minimal fraction of outliers. All axis are calibrated in mV. The left-hand graph shows the plot with outliers; the middle graph is without outliers, and the right-hand graph only shows outliers without the central part. The variability descriptors linked with Poincare plots have to be quite sensitive to the presence of the outliers.

Let estimate the above-said deviations of normality with the normal plots. Such a graph is a probability-probability plot for two distributions: the normal (the horizontal axis) against the factual one (the vertical axis) [6]. Both samples of random values, the normal and the factual, have the same length, and their cumulative probabilities functions are compared. The thin straight line matches the normal distribution, the points – factual one.



**Figure 3:** Probability-probability (P-P, or normal) plots for raw EMGs. The horizontal axes show the expected from the normality hypothesis values, the vertical axes – the factual ones. The thin straight line meets the Gauss distribution, the bold line from points meets the factual one.

Figure 3 shows that the factual distribution is far from the Gauss pattern, especially for larger absolute values, for outliers. Thus, the results of this section are well agreed among themselves and prove significant non-Gaussinity of probability distributions of raw EMGs.

## 4.2. Exponential smoothing model of raw EMGs and their denoising

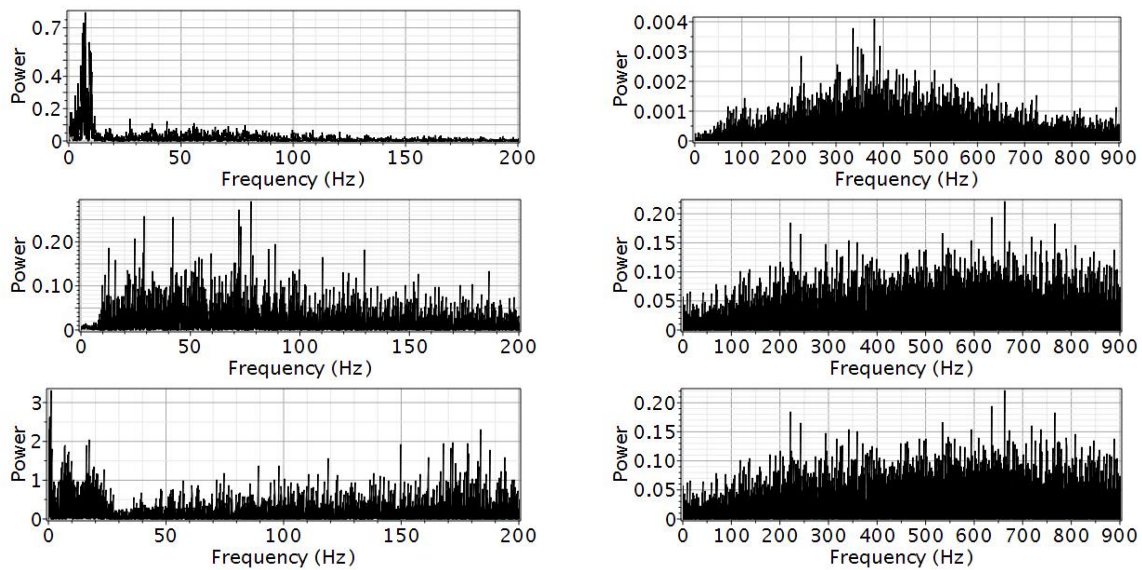
The authors got the same ETS models for all three raw EMGs. Those were the simplest possible models with additive errors (noises) without any trends or seasonalities. We had used the computer

math system Maple 2020. Precisely, it was the algorithms of its "Time Series Analysis" program package.

The simplicity and uniformity of ETS models for raw EMGs allow the pre-filtration onto low-frequency part (or a signal by itself) and high-frequency one (a noise) with Haar wavelets, as it has been proposed in [3, 10]. Fig.4 shows the periodograms of pre-filtered signals and noises for three EMGs.

Periodograms of Fig.4 can be a diagnostic tool by themselves, but one should admit them as excessive complex for clinic practice. Let introduce more digital and straightforward indexes, such as the known signal-to-noise ratio (SNR). One can define it by the logarithm of the RMS amplitudes ratio (in decibels, dB). These values of SNR are the following: 12.75 dB, 6.96 dB, and 6.25 dB in the same order as in Fig.1. Note that all raw EMGs are pointedly noisy. Still, EMGs of pathological states are about twice noisier than the healthy ones. Thus, the pre-filtering sounds reasonable.

It should be marked that pre-filtered EMGs comprise twice fewer counts (samples), shorter than raw EMGs. It is equal to a twofold decrease of sample rate to 2 kHz.



**Figure 4:** Periodograms of pre-filtered signals (graphs on the left) and noises (on the right). The order of the periodograms - "Healthy," "Myopathy," "Neuropathy" from the top pair to the bottom. Note that the EMG of a healthy patient is the least noisy. The power scale is absolute (in arbitrary units).

Curiously, the pre-filtering changes the percentage of outliers barely (shifts are within a few tenths of a single percent). Hence, this parameter, least for the healthy patient, refers to the signal itself, no way to the additive noise. Then this index most likely is diagnosis-caused. If so, then the Poincare plot shape also is diagnosis-dependent.

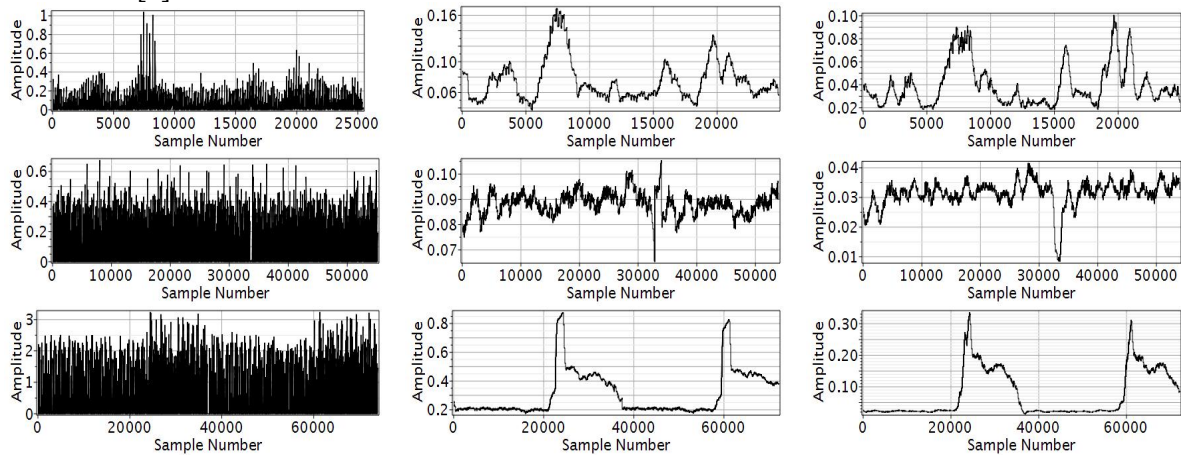
### 4.3 Full-wave rectification and smoothing of pre-filtered EMGs

An all-over accepted means to capture the envelope of raw EMGs is the root mean square (RMS) values within a window that "slides across" the signal. The RMS value is plotted at the center of the window to avoid time shifts in the envelope relative to the signal.

However, this way well works only if the number of outliers is minor [7]. That assumption looks quite doubted for raw and pre-filtered EMGs (see sections 4.1, 4.2). The more robust tool can be the moving median (MM) digital filter applied to the rectified signal. Although this filter (MM) never was used in EMG smoothing afore [7].

This section would like to show the full-wave rectified pre-filtered FMGs and the results of two alternative smoothing methods: RMS and MM. The smoothing windows lengths were roughly proportional to the lengths of the signals: 300, 600, and 900 ms, respectively. Figure 5 reflects this computing. Let recall the fact that RMS-windowing need not preliminary rectification of a signal, opposite to MM one.

Pay attention that RMS and MM smoothing results are similar but not identical. In particular, the amplitudes of MM-envelopes are less than for RMS ones and materially shifted down. That is logical because the median, as an indicator of the central tendency, is least sensitive to the effect of the outliers. Opposite, massive outliers rise the envelopes' amplitudes markedly within RMS-windowing as predicted in [7].



**Figure 5:** Full-wave rectified pre-filtered EMGs (the left column), their RMS-windowing (the middle column), and smoothing by MM (the right column). All amplitudes are given in mV. The order of patients (from upper to down) is the same as in Fig.4.

Mean, and peak values of smoothed EMGs are now in use among clinicians [5]. However, median and peak values should be a more robust set considering all the above argumentation. That is why Table 1 collects these and a few extra values, suggested as robust diagnostics signs.

**Table 1**

The comparison of results between two alternative smoothing methods: RMS and MM.

Methods ↓	Patients →	Healthy	Myopathy	Neuropathy
RMS	Medians	69	89	213
	Peak values	169	106	879
	Peak/median	2.4	1.2	4.1
	Variabilities	29	5	220
MM	Medians	33	32	26
	Peak values	101	42	337
	Peak/median	3.0	1.3	13.0
	Variabilities	21	4	125

Table 1 presents the values usable in diagnostics for two different smoothing methods. Almost all values are given in  $\mu\text{V}$  (microvolts), except for peak-to-median ratios. Interquartile ranges defined variabilities.

Consider the peak value to median ratio within the MM method. This ratio strongly correlates with the kurtosis mentioned afore. The correlation coefficient equals 0.9869 and is statistically significant on the confidence level equal to 0.99 despite small numbers (three) of patients. The correlation between variabilities and the peak-median ratio is even better (0.9997).

Thus, the variability of the correctly smoothed EMG might be a new diagnosis sign. It is a pair of dozen microvolts for healthy patients, while only a few microvolts for myopathy and about a hundred microvolts for neuropathy. The variability correlates strongly with the kurtosis of probabilities distribution and with peak-to-median ratios. Both correlations are significant statistically on the confidence level of 0.99.

The handling way with full-wave rectification and following filtering (smoothing) by moving median (MM) is more robust than RMS-windowing under the condition of massive outliers. Such

conditions seem to be the case in many laboratory measurements of EMGs [7]. Based on this conclusion, the authors recommend only the bottom half of Table 1 as the diagnostic information.

## 5. Conclusions

Let make conclusions pushing off from the list of tasks from section 2.

1. The non-Gaussianity of raw EMGs was proved. Positive kurtosis and massive outliers (with the relative weight from a few up to a few dozen percentages) cause pointed deviations from the Gaussian distribution.

2. All raw EMGs have a simple ETS model: additive noises, no trends, and no seasonalities.

3. Thus, the noises can be easily removed by pre-filtering with Haar wavelets. Pathological EMGs are twice as noisy as for healthy patients. The pre-filtering (denoising) has to be a mandatory procedure at raw EMGs handling.

4. Full-wave rectification, and first-time robust moving median smoothing of EMGs were performed compared to commonly used RMS-windowing.

5. We have suggested a few new robust diagnostic signs based on this research's results (see Table 1).

So, the list of tasks was carried out; hence the aim of the study was achieved. We have offered the new robust processing method for raw EMGs with two attainable and straightforward diagnostic hallmarks.

## 6. Acknowledgments

The authors are grateful to Maplesoft (a division of Waterloo Maple Inc.) for granting access to the licensed copy of Maple 2020.

The results of this paper are a part of studies on the scientific and research project "Development of a hardware and software complex for non-invasive monitoring of blood pressure and heart rate for dual purpose" (State registration number No. 0120U101266, 2020–2021), financed from the state budget.

## 7. References

- [1] K. R. Mills, "The basics of electromyography." *Neurol. Pract.*, 76, 2, (2005): 32–35, doi: 10.1136/jnnp.2005.069211
- [2] I. Saad, N. H. Bais, C. Bun Seng, H. Mohd Zuhir, and N. Bolong, Electromyogram (EMG) signal processing analysis for clinical rehabilitation application, *Proc. - AIMS 2015, 3rd Int. Conf. Artif. Intell. Model. Simul.*, pp. 105–110, 2016. doi: 10.1109/AIMS.2015.76
- [3] G. Chuiko, O. Dvornik, Y. Darnapuk, and Y. Baganov, "Devising a new filtration method and proof of self-similarity of electromyograms," *Eastern-European J. Enterp. Technol.*, 4, 9(112), (2021):. 15–22, doi:10.15587/1729-4061.239165
- [4] D. Bratashov, O. Grishin, A. Kozlova, A. Abdurashitov, R. Verkhovskii, E. V. Shashkov, O. Inozemtseva, V.P. Zharov, Statistics peculiarities of photoacoustic signals in flow cytometry of B16F10 cells, 20th Int. Conf. Photoacoustic. Photothermal Phenom., no. July, p. 45, 2019. Moscow, URL: [https://www.researchgate.net/publication/334721619\\_Statistics\\_peculiarities\\_of\\_photoacoustic\\_signals\\_in\\_flow\\_cytometry\\_of\\_B16F10\\_cells](https://www.researchgate.net/publication/334721619_Statistics_peculiarities_of_photoacoustic_signals_in_flow_cytometry_of_B16F10_cells)
- [5] Noureddine Messaoudi, Raïs El'hadi Bekka, Philippe Ravier, Rachid Harba, "Assessment of the non-Gaussianity and non-linearity levels of simulated sEMG signals on stationary segments." *Journal of Electromyography and Kinesiology*, 32, (2017):70-82, doi:10.1016/j.jelekin.2016.12.006.
- [6] A Field, *Discovering Statistics Using SPSS (3rd Ed.)*. Los Angeles-London-New Delhi-Singapore-Washington DC: SAGE; 2009, URL: [https://www.academia.edu/24632540/Discovering\\_Statistics\\_Using\\_SPSS\\_Introducing\\_Statistical\\_Method\\_3rd\\_edition](https://www.academia.edu/24632540/Discovering_Statistics_Using_SPSS_Introducing_Statistical_Method_3rd_edition)
- [7] W. Rose, KAAP686 Mathematics and Signal Processing for Biomechanics. Electromyogram analysis, 2019, URL: <http://www1.udel.edu/biology/rosewc/kaap686/notes/EMG%20analysis.pdf>

- [8] P. Konrad. ABC of EMG – A Practical Introduction to Kinesiological Electromyography. © 2005 by Noraxon U.S.A., Inc. URL: <https://www.noraxon.com/wp-content/uploads/2014/12/ABC-EMG-ISBN.pdf>
- [9] A. M. Burden, S. E. Lewis, E. Willcox, "The effect of manipulating root mean square window length and overlap on reliability, inter-individual variability, statistical significance and clinical relevance of electromyogram." *Manual Therapy*, 19, 6,(2014): 595-601, doi: 10.1016/j.math.2014.06.003
- [10] G. Chuiko, O. Dvornik, Y. Darnapuk, Y. Krainyk , Principal Component Analysis, Quantifying, and Filtering of Poincaré Plots for time series typal for E-health, in R.Patgiri, A. Biswas, P, Roy. (eds) *Health Informatics: A Computational Perspective in Healthcare. Studies in Computational Intelligence*, Springer, Singapore. 2021, v.932, pp.61-76, doi: 10.1007/978-981-15-9735-0\_4
- [11] J. Brownlee, *A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python*, 2020. URL: <https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/>
- [12] *Time Series Analysis*, © Maplesoft, a division of Waterloo Maple Inc. 2019, URL: [https://www.maplesoft.com/products/maple/new\\_features/maple18/TimeSeriesAnalysis.aspx](https://www.maplesoft.com/products/maple/new_features/maple18/TimeSeriesAnalysis.aspx)
- [13] V. Martsenyuk, A. Sverstiuk, A. Kłos-Witkowska, A. Horkunenکو and S. Rajba, "Vector of Diagnostic Features in the Form of Decomposition Coefficients of Statistical Estimates Using a Cyclic Random Process Model of Cardiosignal," 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2019, pp. 298-303, doi: 10.1109/IDAACS.2019.8924398.
- [14] A. L. Goldberger, , L. A. N. Amaral, L. Glass, J. M. Hausdorff,. P. C. Ivanov, R. G. Mark, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101 (23), 2000: e215–e220. doi: 10.1161/01.cir.101.23.e215
- [15] C. Thirumalai, M. Vignesh, and R. Balaji, Data analysis using box and whisker plot for lung cancer, 2017 *Innovations in Power and Advanced Computing Technologies (i-PACT)*, 2017, pp. 1-6, doi: 10.1109/IPACT.2017.8245071.