Praca Oryginalna Zdr Publ 2010;120(4):383-387 Original Article

TERESA MAŁECKA-MASSALSKA[1,4], RYSZARD MACIEJEWSKI[2], KRYSTYNA LUPA[1],
PIOTR WĄSIEWICZ[3], WOJCIECH ZAŁUSKA[4], ANDRZEJ KSIĄŻEK[4]

Analiza komputerowa parametrów uzyskanych przy pomocy bioimpedancji elektrycznej u pacjentów hemodializowanych

Streszczenie

Wstęp. Monitorowanie stanu nawodnienia pacjentów dializowanych stanowi istotny aspekt kliniczny jakości ich leczenia. Bioimpedancja elektryczna, jako metoda wykorzystującą właściwości elektryczne tkanki poddanej działaniu prądu zmiennego, jest jednym z narzędzi, które służy określaniu stanu nawodnienia.

Cel. Celem badania było użycie analizy komputerowej parametrów uzyskanych na pacjentach hemodializowanych przy pomocy bioimpedancji elektrycznej do tworzenia algorytmów – drzew decyzyjnych.

Materiał i metodyka. Pomiary zostały przeprowadzone na dwóch grupach pacjentów – 50 przewlekle hemodializowanych (grupa badana) 10 minut przed hemodializą i 46 zdrowych ochotnikach (grupa kontrolna). Badane parametry to TBW (całkowita woda ustroju), ECW (wielkość przestrzeni zewnątrzkomórkowej) i ICW (wielkość przestrzeni wewnątrzkomórkowej). Do pomiarów bioimpedancji użyto analizatora bioimpedancji (model 4000B, Xitron Technologies, San Diego, CA, USA) przy użyciu elektrod (7,7 x 1,9 cm^2). Bioimpedancję mierzono w logarytmicznym spektrum 10 częstotliwości, rozpoczynając od 5 do 500 kHz.

Wyniki. Wyniki obliczeń na normalnych, dyskretyzowanych i normalizowanych danych przetwarzano w środowisku matematyczno-statystycznym R z pomocą wolnodostępnej biblioteki inteligentnych algorytmów Weka, aby wygenerować proste reguły identyfikacji wspomnianych wyżej chorób. Wykonane eksperymenty potwierdziły możliwości uzyskania za pomocą procedur J48 (decision tree) i QDA (quadratic discriminant analysis) dobrych klasyfikatorów automatycznie detekujących właściwego pacjenta po uprzednim wytrenowaniu ich na uzyskanych danych medycznych. Dzięki odpowiedniemu algorytmowi testowania klasyfikatorów uzyskano drzewa decyzyjne z błędem klasyfikacji poniżej 5%. W przyszłości będzie możliwe dalsze zmniejszenie błędu klasyfikacji poprzez zastosowanie na większej ilości danych bardziej złożonych algorytmów tworzenia klasyfikatorów.

Słowa kluczowe: wyszukiwanie danych, drzewko decyzyjne, stan nawodnienia, wielkość przestrzeni zewnątrzkomórkowej, wielkość przestrzeni wewnątrzkomórkowej.

Computer guided analysis of hemodialyzed patients' bioimpedance spectroscopy parameters

Abstract

Introduction. Monitoring the hydration level in dialyzed patients is an important clinical aspect of treatment quality. Bioimpedance is one of the methods using electric properties of body tissues subjected to an alternating multi-frequency amplitude current in order to assess hydration states.

Aim. The aim of the study was to use the computer guided analysis of hemodialyzed patients' bioimpedance spectroscopy parameters for desion trees algorithm.

Material and methods. The measurements were conducted on two groups – 50 patients on chronic hemodialysis (the test group) 10 minutes before hemodialysis and 46 healthy volunteers (the control group). The studied parameters were: TBW (total body water), ECW (extracellular water) and ICW (intracellular water). For bioimpedance measurements a bio-impedance analyzer was used (Xitron Technologies, San Diego, CA, USA, model 4000B Bioimpedance spectroscopy device measuring at 50 frequencies between 5 kHz and 1 MHz) with electrodes (7.7 x 1.9 cm^2). Bioimpedance was measured in a logarithmic spectrum of 10 frequencies starting from 5 to 500 kHz.

Results. The results of calculations on normal, discredited and normalized data were computed in the R language environment with algorithms from RWeka library to generate simple rules of identification of the aforementioned diseases. The executed experiments affirm a possibility to create good classifiers for detecting a proper patient with the help of J48 (decision tree) i QDA (quadratic discriminant analysis) but only after previous training. Thanks to an appropriate algorithm for testing classifiers, decision trees with the classification error below 5 % were obtained. In the future it will be possible to further diminish the classification error by measuring more variables using more complex algorithms for testing classifiers.

Key words: data retrieval, decision tree, hydrate status, extracellular compartment, intracellular compartment.

[1] Physiology Department, Medical University of Lublin
[2] Human Anatomy Department, Medical University of Lublin
[3] Institute of Electronic Systems, Warsaw University of Technology
[4] Nephrology Department, Medical University of Lublin

INTRODUCTION

Monitoring the hydration level in dialyzed patients is an important clinical aspect of treatment quality. There are many tools to assess the hydrate status. When a current method for monitoring hydration is inadequate, other methods must be considered to provide increased or improved treatment quality for the patient population. The fluid removed from the patient during dialysis is taken away with the use of ultrafiltration mainly from the intravascular space. Knowledge and understanding of other fluid compartments of the body during this dynamic process can be beneficial in reducing complications associated with this therapy. Bioimpedance has been established as a valuable tool in the evaluation of hydration states of various compartment of the body in the dialyzed patient [1-3].

The bioimpedance technique incorporates precise evaluation of hydration levels using physiological data concerning the assessment of water compartment sizes, such as TBW (total body water), ECW (extracellular water), ICW (intracellular water) and interstitial compartments [4]. The specific mechanisms of this technique are based on an elementary principle that electrical resistance of a cylinder is directly proportional to the length and inversely proportional to the cross section area of the cylinder multiplied by the density. This method is based on the evaluation of electrical resistance in body tissues subjected to an alternating multi-frequency amplitude current [5]. Although the principal bioimpedance techniques were first introduced by Thomassett in 1963, an increased interest in this technique appeared in the early seventies of the last century when Nyboer demonstrated a correlation between bioimpedance value assessed with the use of an alternating current and changes in the blood volume [6]. Many articles described the method of the whole body bioimpedance analysis (WBIA). The WBIA method places electrodes on the palm and foot (the wrist and ankle placement of electrodes have also been used). An alternating current, with frequencies from 5 to 500 kHz reaches the electrodes placed at the level of metacarpophalangeal joint in finger III of the upper extremity and at the base of metatarsophalangeal joint in toe II and III of the lower extremity – the voltage is measured between the electrodes placed on the wrist in an imagined line connecting the styloid process of the ulnar bone with the styloid process of the radial bone and the electrode placed in the line connecting medial and lateral condyle. It is possible, utilizing the bioimpedance technique to choose an option of one current frequency usage or a multi-frequency option with an amplitude from a few to a few hundred (500) kHz. It should be noted that the WBIA assessment is dependent on the changes in body position. Therefore, the body position changes must be considered when analyzing results using this method. The segmental bioimpedance technique (SBIA) is an assessment of independent body segments, such as upper extremities, trunk, and lower extremities. The analysis of the results using this technique has been observed to be a more precise evaluation of hydration states and dynamical changes during a dialysis session. As previously stated, the specific mechanisms of the WBIA technique are based on an elementary principle that electrical resistance of the cylinder is directly proportional to the length and inversely proportional to the cross section area of the cylinder multiplied by the density.

$$Z=qL/A \qquad (1)$$

where: Z – impedance (Ohm), g – tissue density (Ohm/cm), L – cylinder's length, A – cross-section area of the cylinder

$$ZxL/L=qxL/A \text{ x } L/L \qquad (2)$$

Since A x L = V (volume, cm^3), we receive a following correlation:

$$V=qxL2/R \qquad (3)$$

where: L (cm) – length of cylinder, R (cm) – electrical resistance and q – cylinder density

The assumption that a human body is a sum of homogenous cylinders and a current can run through extracellular and intracellular space, was an impulse for Hoffer in 1969 to apply this method in measuring the total body water (TBW).

Taking into account the voltage quantity we calculate the electrical resistance (impedance) which is next converted in proportion to the cylinder's volume (upper extremity, lower extremity, trunk) with the use of the formula (3).

The terms: electrical impedance and electrical resistance are often used interchangeably. In fact, impedance comprises the function of inductive resistance (R) and capacitive resistance (Xc).

$$Z^2 = R^2 + Xc^2 \qquad (4)$$

The inductive resistance (R) refers to extracellular fluid impedance, whereas capacitive resistance (Xc) refers to intracellular fluid resistance. Concluding from the formula (1), impedance is a function of the cylinder's length (upper extremity, lower extremity, trunk) and cross-sectional area of the cylinder (A) with a given current frequency. The inductive (R) and capacitive (Xc) resistance values depend on the frequency of alternating current. The correlation between capacitive and inductive resistance is shown in Fig. 1a.

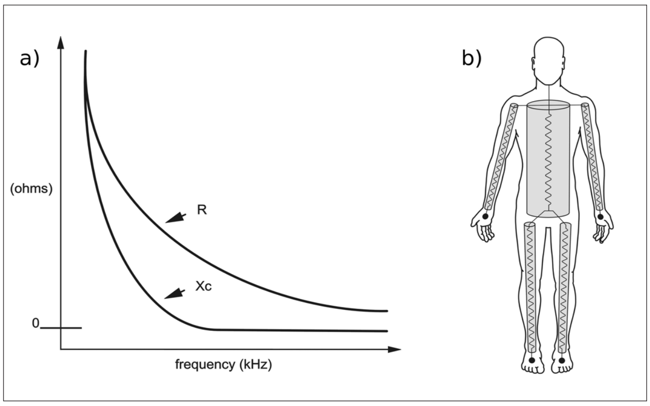

FIGURE 1. a) The relationship between capacitive (R) and inductive (Xc) resistance with a given frequency of an alternating current; R – inductive resistance, Xc – capacitive resistance [7]. b) A human body as a conductor comprising five cylinders [7].

Impedance is a function of conductor geometry which, as a rule, is assumed to be a cylinder. The bioimpedance assumption that a human body is a conductor and consists of five cylinders, is shown in the Fig. 1b.

Based on the analysis of the formula, one may assume that extremities, because of their size, will have a bigger share in the total impedance. Therefore, the total of 80 % of impedance consist of lower and upper extremity resistance, which gives

30 % after conversion into the total body water. With the use of bioimpedance measurement it is difficult to evaluate the total amount of water in the trunk (70 % TBW).

MATERIAL AND METHODS

The study was performed among 50 hemodialyzed patients and 46 healthy volunteers. Inclusion criteria were the following: patients with diagnosed terminal renal insufficiency were included in the study, aged between 18 and 80 years, clinically stable. Exclusion criteria were the following: mental problems that could terminate the study in any way, pregnancy or lactation, amputation of a lower limb, implanted pacemaker, severe hemostatic circulatory insufficiency. The following parameters were measured in each healthy volunteer: body mass (in kg), height (in cm), blood pressure, TBW, ECW, ICW. The following parameters were measured in each patient: body mass before and after hemodialysis (in kg), height of patient (in cm), blood pressure before hemodialysis, TBW, ECW, ICW. The body mass of a patient was measured with the use of a scale with an acceptable deviation of 0.1 kg. The height of a patient (in cm without shoes) was measured with the use of a standard measure. For bioimpedance measurements a bioimpedance analyzer was used (a Xitron Hydra 4200 Bioimpedance spectroscopy device measuring at 50 frequencies between 5 kHz and 1 MHz) with electrodes (7.7 x 1.9 cm^2). All parameters were measured at the beginning of hemodialysis (not during it so that errors in evaluation were avoided) as the greatest fluid distribution occurs within first hour of hemodialysis.

The measurements were performed at the beginning of hemodialysis. The measurements before hemodialysis were performed within 10 minutes after the moment a patient was lain down. Bioimpedance was measured in a logarithmic spectrum of 10 frequencies starting from 5 to 500 kHz (analyzer, Xitron Hydra 4200 Bioimpedance spectroscopy device) with electrodes (7.7 x 1.9 cm^2). Two electrodes inducing an alternating current were placed dorsally on hand (I1) and ankle (I2) of the same body side. The measuring electrodes were placed on the wrist (S1) and ankle (S2). A computer was used to collect and store the data. Its variables were the following: body mass index (BMI), intracellular water (ICW) - water volume inside body cells. (i.e., water in the "living" cells), extracellular water (ECW) - water volume outside the body cell mass (i.e., water in the "inactive" cells), total body water (TBW) - sum of ICW and ECW, ECW_TBW - ECW divided by TBW, ECW_mass - ECW divided by the body mass, height, weight and age.

Decision tree, a typical data mining (retrieval) method is described in an internet medical dictionary as "a graphic construct showing available choices at each decision node of managing a clinical problem along with probabilities (if known) of possible outcomes for patient's freedom from disability, life expectancy, and mortality." In computer science its properties are the following: each internal node tests an attribute (a column in the data collection), each branch corresponds to an attribute value, each leaf node assigns a classification.

In a more general sense, a classifier is mapping from a (discrete or continuous) variable space *X* to a discrete set of classes denoted by labels *Y*. Learning classifiers are divided into unsupervised learning and supervised classifiers. The former need training sets labeled by experts in order to obtain knowledge about classes e.g. about a patient class. The latter may be able to make the proper classification only with the help of raw data and special distance measurements between examined patients – points in the multivariable space [8]. This space can be visualized in two dimensions after the scaling based e.g., on principle component analysis (PCA), which transforms real dimensions to artificial ones, where the first ones have the most correlations within and the rest of dimensions may be omitted with small resulting errors.

A Bayesian network, belief network is a directed acyclic probabilistic graphical model that represents a set of random variables and their conditional dependencies. For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. For given symptoms such a network can be used to compute the probabilities of the presence of various diseases.

RESULTS

During the measurements there were sets obtained from two groups (healthy volunteers – the control group and hemodialyzed patients – the test group). The test group consisted of 50 patients and the test group of healthy young medical staff (46 people). Raw data were collected by medical equipment. Its variables were the following: body mass index (BMI) – a ratio of weight to height used as a quick measure of health status, BMI values from 18-24.9 are desirable [kg/m^2], intracellular water (ICW) – water volume inside the body cells. (i.e., water in the "living" cells), extracellular water (ECW) - water volume outside the body cell mass (i.e., water in the "inactive" cells), total body water (TBW) – sum of ICW and ECW, ECW_TBW – ECW divided by TBW, ECW_mass – ECW divided by the body mass, height, weight and age.

Standardized (z-score by age and sex), discretized data (three levels) and data retrieval results were computed in the R language environment [9] in order to find a simple rule for recognizing health problems. Z-score, so called standard score, is equal to raw score minus mean of this raw score population, the subtraction result divided by its standard deviation, used in statistics. All the data, except age, were standardized separately for women and men and for age – five intervals from infinity to mean plus standard deviation, from mean plus standard deviation to mean and so on up to minus infinity.

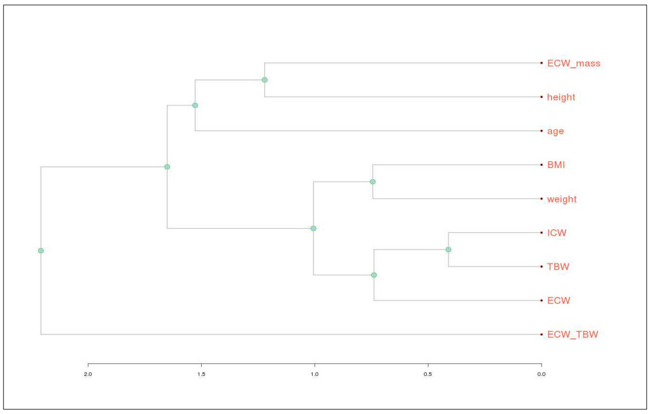

FIGURE 2. Correlation clusters dendrogram of standardized data.

In Fig. 2 all the standardized variable correlation cluster dendrogram is depicted. The shorter arms, the stronger correlation between joint variables. The separate analyses for control and test group show the similar correlations, so in the dendrogram the data obtained from both groups were used. The strongest correlation is between ICW and TBW. ECW_TBW has the smallest correlation, the next one is ECW_mass together with height. The 'age' variable, after being standardized, is not strongly correlated with other variables.

In next experiments the group marker variable was also added: P – for patients, and K – for control group. In the search for effective classifiers, several were tested on the training set sampled randomly from the original data without replacements, usually 50%. The rest of our data was used for testing purposes.

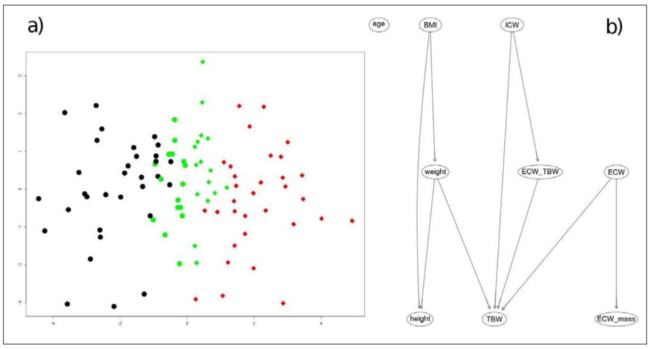

FIGURE 3 a) Multidimensional scaling of a given population generated two k-means clusters (rhombs and circles denotes separated clusters, and black, green, red colors – TBW levels: higher, mean, lower); b) An exemplary bayesian network calculated for standardized data (TBW is the main attribute connected with other attributes).

Based on the PCA k-mean clustering without previous training performed on standardized data except sex, group and age, the parameters revealed that artificial clusters are connected with TBW discretized levels: lower, mean, higher (in Fig. 3a), but not with the ECW_TBW distribution. Therefore, further experiments with different distance measures and cluster methods (e.g., fuzzy, hierarchical, divisive, agglomerative, with dissimilarities or raw data) prove that the supervised (without training) classification method of calculating patient and control group split does not exist and such clusters are associated with TBW levels, even after TBW variable's removal from the given data set. The calculated (with the help of bnlearn R package) on standardized data Bayesian network (in Fig. 3b) confirms this idea: almost all connections go to the TBW node. The 'age' attribute after score standardizing is not connected with others.

The decision tree classifier from Fig. 4a was calculated from 6 raw data variables TBW, ECW, ICW, ECW_TBW, ECW_mass, BMI with the mentioned grouping marker values in the leaves. Attributes were chosen for nodes in this tree by J48 algorithm from RWeka library after creating trees from all combinations of two, three, four, five, and six input parameters and after five construction trials for each combination choosing one with the least error mean (constructions on 70% of all data sets and testing on 30% of the data – measuring the classification error mean). Thus, the tree from Fig. 4a is one of the most efficient tree classifiers tested also on 100% population of two sets with only 1 incorrectly classified patient and 3 incorrectly classified control group individuals – this means that classification error is below 5 %. The node ECW_ TBW is the „must have" node in the majority of generated decision trees (especially these generated for the standardized data, but unfortunately, with a greater mean classification error) despite being not correlated with other variables in the control group and the examined group.

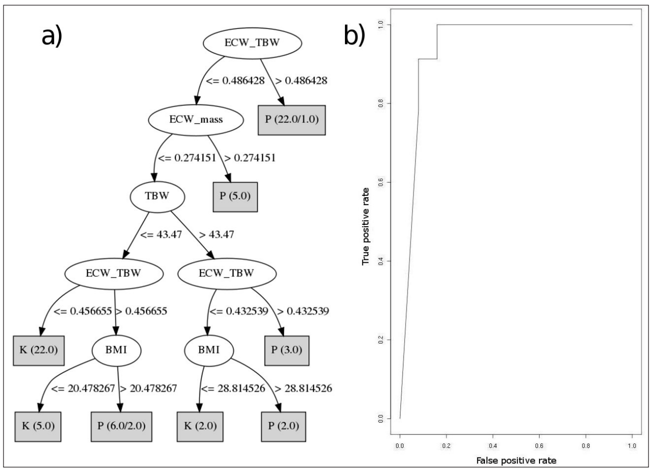

FIGURE 4 a) The best decision tree generated by J48 Rweka procedure on raw data. In leaves (6.0/2.0) means 6 individuals and 2 of them incorrectly classified. b) ROC curve from the quadratic discriminant analysis classifier for standardized data.

So after obtaining a very good decision tree, the next learning-based classifiers were tested. The k-nearest neighbor and naive Bayes classifiers were not appropriate for the given data set, though the naive Bayes for raw data were sometimes better than the previous ones.

Finally, linear discriminant analysis classifiers and quadratic discriminant analysis classifiers were tested. The former has better results for raw data, and the more precise latter – for both raw and standardized data. The former was not better than the J48 tree as is proved in the depicted in Fig. 4b ROC curve for the quadratic discriminant analysis (QDA) classifier. The area under the ROC curve called AUC was equal to 0.944. The larger AUC, the better the classifier is. After many trials it was verified that single tree classifiers can be as good as the QDA classifiers.

CONCLUSION

The executed experiments affirm possibilities of creating good classifiers for detecting a proper patient with the help of medical data sets, but only after previous training. In the improved test environment (evaluating decision trees generated for all attribute combinations) we obtained excellent tree classifiers, even better than QDA classifiers. Supervised methods, especially clustering, are not even adequate in the case of body water parameters. Maybe for larger sets, with a greater number of variables and rows, it would be possible. In the near future we are going to measure more variables with the help of more sophisticated medical equipment and evaluate all possible classifiers once again.

In this paper we proved that computer science techniques can be a great help for medical physicians in: improving predictive abilities of different tests and making a better differential diagnosis.

Praca prezentowana na Międzynarodowej Konferencji pt. "Zdrowie Publiczne wyzwaniem XXI wieku", Lublin, 20-22 października 2010 r.

REFERENCES

1. Perez-Garcia R, Lopez-Gomez JM, Jofre R. Haemodialysis dose, extracellular volume control and arterial hypertension. Nephrol Dial Transplant 2001;16:98-101.
2. Tattersall J. Bioimpedance analysis in dialysis: state of the art and what we can expect. Blood Purif. 2009;27(1):70-4.
3. Wizemann V, Wabel P, Chamney P, Załuska W, Moissl U, Rode C, Małecka-Massalska T, Marcelli D. The mortality risk of overhydration in haemodialysis patients. Nephrol Dial Transplant 2009;24(5):1574-9.
4. Yi-Chou Ch, Han-Hsiang Ch, Jui-Chi Y. Postdialysis extracellular volume is rational for evaluating dry weight in hemodialysis patients. Nephron 2002;90:109-110.
5. Schwenk A, Beisenherz A, Romer K. Phase angle from bioelectrical impedance analysis remains an independent predictive marker in HIV-infected patients in the era or highly active antiretroviral treatment. American Journal of Clinical Nutrition 2000;72:496-501.
6. Spinale FG, PhD, Reines HD, Cook MC. Noninvasive estimation of extravascular lung water using bioimpedance. Journal of Surgical Research 1989;47:535-540.
7. Kushner R. Bioelectrical Impedance Analysis: A Review of Principles and Applications. J Ani Coll of Nutr 1992;11(2):199-206.
8. Ripley B. D. Pattern Recognition and Neural Networks, Cambridge Univ. Press, (1996).
9. VenablesWN, RipleyBD, "Modern Applied Statistics with S", Fourth edition, Springer, (2002).

Informacje o Autorach
Dr n. med. Teresa Małecka-Massalska – adiunkt, Katedra i Zakład Fizjologii Człowieka, Katedra i Klinika Nefrologii, Uniwersytet Medyczny w Lublinie; prof. dr hab.n.med. Ryszard Maciejewski – Katedra i Zakład Anatomii Człowieka, Uniwersytet Medyczny w Lublinie; prof. dr hab. n. med. Krystyna Lupa – kierownik, Katedra i Zakładu Fizjologii Człowieka, Uniwersytet Medyczny w Lublinie; dr inż. Piotr Wąsiewicz – adiunkt, Wydział Elektroniki i Technik Informacyjnych, Instytut Systemów Elektronicznych, Politechnika Warszawaska, Warszawa; prof. dr hab. n. med. Wojciech Załuska – adiunkt, Katedra i Klinika Nefrologii, Uniwersytet Medyczny w Lublinie; prof. dr hab. n. med. Andrzej Książek – kierownik, Katedra i Klinika Nefrologii, Uniwersytet Medyczny w Lublinie.

Adres do korespondencji
Dr n. med. Teresa Małecka-Massalska
Katedra i Zakład Fizjologii Człowieka, Uniwersytet Medyczny
ul. Radziwiłłowska 11, 20-080 Lublin
Tel.: +48 (81) 528 84 05, 609304934, Fax: + 48 (81) 442 90 30
e-mail: tmalecka@gmail.com

Dr inż. Piotr Wąsiewicz
Wydział Elektroniki i Technik Informacyjnych, Instytut Systemów Elektronicznych, Politechnika Warszawaska
ul. Nowowiejska 15/19, 00-665 Warszawa,
pwasiewi@elka.pw.edu.pl