Endokrynologia Pediatryczna
Pediatric Endocrinology

Praca oryginalna
Endokrynol. Ped. 2015.14.2.51.9-18.

Original Paper
Pediatr. Endocrinol. 2015.14.2.51.9-18.

Artificial neural networks – a novel tool in modelling the effectiveness of growth hormone (GH) therapy in children with GH deficiency

Sztuczne sieci neuronowe – nowe narzędzie modelowania skuteczności leczenia hormonem wzrostu dzieci z somatotropinową niedoczynnością przysadki

[1]Urszula Smyczyńska, [2]Joanna Smyczyńska, [2]Maciej Hilczer, [2]Renata Stawerska, [2]Andrzej Lewiński, [1]Ryszard Tadeusiewicz

[1]AGH University of Science and Technology; Department of Automatics and Biomedical Engineering, [2]Polish Mother's Memorial Hospital – Research Institute; Department of Endocrinology and Metabolic Diseases

[1]AGH Akademia Górniczo-Hutnicza; Katedra Automatyki i Inżynierii Biomedycznej, [2]Instytut Centrum Zdrowia Matki Polki w Łodzi; Klinika Endokrynologii i Chorób Metabolicznych

Key words
Artificial neural networks, growth hormone, modelling of therapy effectiveness

Słowa kluczowe
Sztuczne sieci neuronowe, hormon wzrostu, modelowanie skuteczności leczenia

Abstract

Introduction. Prediction of recombinant human growth hormone (rhGH) therapy effectiveness seems an important issue, especially in situations when diagnosis is problematic. Until now the problem was solved by multiple regression analysis, while here we propose application of artificial neural networks (ANN), particularly multilayer perceptron (MLP), as a novel tool. **Material and methods.** The analysis was performed on data of 289 patients, treated for at least 2 years due to GH deficiency. First, we automatically classified patients to one of two groups depending on attainment of final height (FH) over or below 10th centile (2-stage classification). Secondly, we introduced third group for FH below 3rd centile (3-stage classification). Finally, we predicted attained FH standard deviation score (SDS) in ANN regression model. **Results.** In 2-stage version MLP networks classified more than 80% of patients correctly. In 3-stage classification, the ratio of right answers was close to 70%. ANN regression model predicted FH SDS with average error of 0.7 SD (4.2 cm), explaining about 45% of its variability. **Conclusion.** The results of the study are promising and the topic definitely worth further consideration. Due to their characteristics, ANN seem to be par-

Streszczenie

Wstęp. Predykcja skuteczności leczenia ludzkim rekombinowanym hormonem wzrostu (rhGH) wydaje się obecnie ważnym zagadnieniem, szczególnie w sytuacjach, gdy postawienie diagnozy jest trudne. Dotychczas problem ten rozwiązywano przy użyciu analizy regresji wielorakiej, natomiast w niniejszej pracy proponujemy zastosowanie nowego narzędzia, jakim są sztuczne sieci neuronowe typu perceptronu wielowarstwowego (MLP). **Materiał i metody.** Analiza została przeprowadzona w oparciu o dane 289 pacjentów, których leczono z powodu niedoboru GH przez co najmniej 2 lata. W pierwszym etapie badania, przeprowadziliśmy automatyczną klasyfikację (2-stopniową) pacjentów do jednej z dwu grup w zależności od tego, czy osiągnęli wzrost końcowy (FH) powyżej 10 centyla. Następnie, wprowadziliśmy dodatkowo trzecią grupę o FH niższym od 3 centyla (klasyfikacja 3-stopniowa). Ostatni eksperyment polegał na predykcji dokładnej wartości wskaźnika odchylenia standardowego (SDS) FH (neuronowy model regresyjny). **Wyniki.** W wersji 2-stopniowej sieci MLP klasyfikowały poprawnie ponad 80% przypadków. W 3-stopniowej odsetek poprawnych odpowiedzi wyniósł niemal 70%. Neuronowy model regresyjny przewidywał FH SDS z przeciętnym błę-

redakcja@pteidd.pl
www.endokrynologiapediatryczna.pl
www.pteidd.pl

Adres do korespondencji / Correspondence address:
Urszula Smyczyńska, al. Mickiewicza 30, 30-059 Kraków, usmycz@agh.edu.pl

ticularly useful in identifying the predictors of growth response to rhGH therapyand in modelling the effectiveness of treatment.

Pediatr. Endocrinol. 2015.14.2.51.9-18.

dem 0,7 SD (4,2 cm), jednocześnie wyjaśniając około 45% zmienności FH SDS. **Dyskusja.** Wyniki powyższego badania są obiecujące, a temat niewątpliwie wart poszerzonej analizy. Dzięki swoim szczególnym cechom sieci neuronowe wydają się wyjątkowo odpowiednie do identyfikacji optymalnych czynników predykcji i modelowania skuteczności leczenia rhGH.

Endokrynol. Ped. 2015.14.2.51.9-18.

Introduction

Growth hormone (GH) deficiency – diagnosis and goals of therapy

Growth hormone deficiency (GHD) is the main indication for recombinant human GH (rhGH) therapy in children with short stature. The most important goal of treatment is to attain normal final height (FH). For several years, GHD is defined as secondary insulin-like growth factor-I (IGF-I) deficiency [1, 2]. Nevertheless, the main tool in diagnosing GHD are still GH stimulation tests. However, both assessment of GH secretion after pharmacological stimulation and interpretation of the measurement of IGF-I concentration have several limitations. The main problems concerning GH stimulating tests include the arbitrarily established cut-off value for normal and decreased GH peak and the poor reproducibility of GH response to stimulation, leading to the principle of performing at least 2 tests for each patient [3-5]. The interpretation of IGF-I concentration depends on patient's age and gender and should take into account other than GHD possible causes of IGF-I deficiency. Regardless of the importance of documenting decreased GH and/or IGF-I secretion for the diagnosis of GHD, it seems particularly important to identify the patients who may benefit during rhGH therapy. For this purpose, numerous models of growth response to rhGH therapy have been created, either based only on the information available before rhGH therapy onset or including the data on rhGH therapy effectiveness in previous years for predicting growth response in subsequent years [6-8]. Prediction of growth response to rhGH therapy seems to be particularly important in the light of the reports on the effectiveness of rhGH administration in children diagnosed with idiopathic short stature (ISS) and established consensus [9]. The diagnostic criteria of ISS undoubtedly include normal GH secretion in stimulation test, whereas exclusion of IGF-I deficiency is required only by some authors [10], while others diagnose ISS regardless of normal or decreased IGF-I secretion [2].

Up to now, all the created models are based on the multiple linear or non-linear regression that requires a number of presuppositions that must be met. Unfortunately, the relationships between different assessed auxological and hormonal parameters are complex and not fully defined. Furthermore, inclusion of correlated variables in regression models should be avoided, thus the researcher must determine the set of uncorrelated variables used for creating model.

In present study, we propose neural networks modelling as a novel tool for prediction growth response to rhGH therapy in children with GHD. Best of our knowledge, up to now, there are no studies presenting such an approach to this issue.

Neural network modelling

Artificial neural networks (ANN) are complex, biologically inspired computational systems, considered one of the leading tools of machine learning or even artificial intelligence. Their development started when some scientists tried to mathematically model parts of nervous system [11]. Actually, it appeared that it is very complicated to model it with high precision.

However, it has been found that even some simplified models can be quite useful. In fact, there is no need to reproduce whole neuron structure to mimic to some extend its information processing capability. Basically, a neuron is a unit that gathers many signals from outside and processes them to produce only one output signal (Fig. 1). Although those signals are electrical or chemical, they could be simply modelled as numbers. Next property of real neurons that can be relatively easily reproduced is different impact of signals from particular synapses on neurons activation. In nervous system some synaptic signals (potentials)

have stimulatory while other inhibitory impact on postsynaptic neuron; moreover the strength of this influence is not equal for all signals. In artificial neurons this behaviour is simulated by assignment of some numerical weight coefficient to each incoming signal. In such representation, we may consider that inputs with negative weights are inhibitory and with positive – stimulatory; and the greater the absolute value of coefficient the stronger effect the signal has.

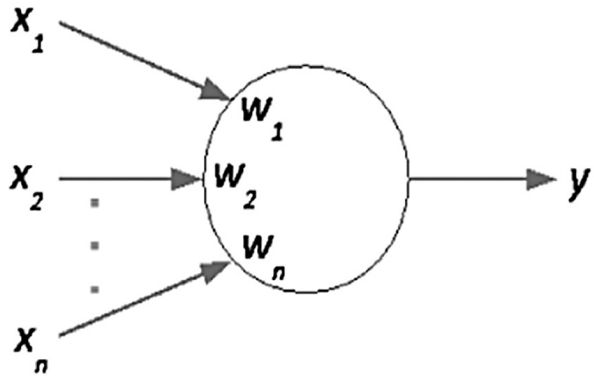

Fig. 1. Scheme of artificial neuron
Ryc. 1. *Schemat sztucznego neuronu*

The problem of building the network, connecting single neurons, may be solved in several ways. The simplest one is to form layers of neurons and make connections between adjacent layers (Fig. 2). In layered network the information flows only forward from input to output. Such a structure is quite similar, for instance, to the one found in retina in eye, although obviously simplified. In more complex networks, the processed information may be sent back in feedback connections.

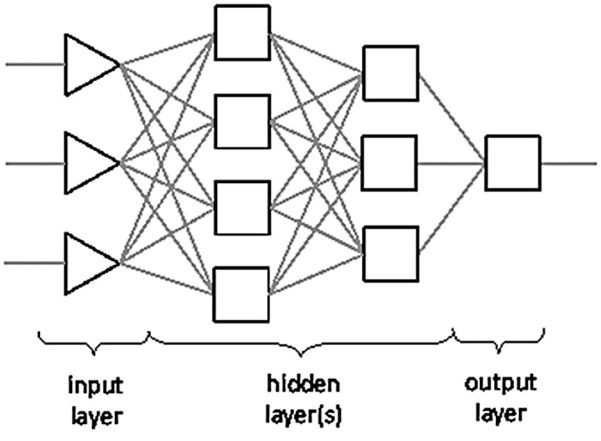

Fig. 2. Schematic representation of layered ANN
Ryc. 2. *Schemat sztucznej sieci neuronowej*

Having described briefly structure of ANN, we should now move to their learning – the process in which ANN model is derived. Generally there are 2 types of learning: supervised (with teacher) or unsupervised (without tutor). In supervised learning, example input data is presented to the network with correct answers that it should produce at the output in response to them. At the beginning the weight coefficients are random and the output far from desired, but then the weights are slowly corrected. The presentation of all examples from learning database is iterated (repeated) many times until the result is satisfactory or stops to improve. In unsupervised learning the knowledge of right answers is not needed and what the network does consists rather in finding similarities in data than reproducing observed dependencies.

The advantages of ANN include possibility of modelling non-linear dependencies without need to fulfil any statistical assumptions concerning data and previous knowledge about character of relation between input and output. For those reasons ANN have beenhitherto applied to solve many problems, among them medical [12]. However, in studies concerning clinical medicine, especially in modelling physical development of children and predicting the response to growth promoting therapies, these tools are relatively not popular.

Material and methods

Patients' cohort, input data and their pre-processing

The models were derived on data collected from 289 patients (218 boys, 71 girls) treated in Department of Endocrinology and Metabolic Diseases in Polish Mother's Memorial Hospital – Research Institute in Lodz with rhGH due to isolated GHD up to fulfilling the withdrawal criteria and no shorter than 2 years. All the information was gathered during routine diagnosis before qualification the patients to rhGH treatment and at the therapy withdrawal. The patients with any concomitant chronic diseases, including multiple pituitary hormone deficiency, genetic syndromes, malnutrition, as well as ones with acquired GHD (due to brain tumours, injuries, cranial irradiation, *etc.*) were excluded from the study.

The retrospective studies on rhGH therapy effectiveness were approved by the Committee of

Ethics of Scientific Research in Polish Mother's Memorial Hospital – Research Institute in Lodz.

In constructing ANN model, only the data available at rhGH therapy onset were used. Among numerous available parameters, the following ones were chosen to create models:

1. Patient's height (H [cm]) at rhGH therapy onset. Before treatment, all the patients were measured by Harpenden stadiometer and their height was converted **to height standard deviation score (H SDS)** for chronological age (CA) and sex, according to the normative data for Polish children[13].

2. Patient's CA at rhGH therapy onset.

3. Patient's bone age (BA) at rhGH therapy onset, assessed on the ground of radiogram of non-writing hand and wrist, according to Greulich-Pyle's standards [14].

4. Patient's sex. This variable differs from other variables, not being quantitative but qualitative one. For the use in model it must be transformed to a numerical value, *e.g.* 0 for males and 1 for females; however the order of assignment has no influence on result and can be arbitrary.

5. Parental heights. Similarly to child's height, the **heights of mother (HM) and father (HF)** may be either given in centimetres or as SDS values and either treated as 2 separate values or combined to **target height (TH)**. Finally, TH may be on its own transformed to TH SDS.

6. Maximal value of GH peak (max GH) in two stimulation tests: with clonidine 0.15 mg/m^2 orally and with either insulin 0.1 IU/kg *i.v.* or glucagon 30 μg/kg *i.m*. Some patients were qualified to rhGH therapy despite normal GH peak in stimulation tests (due to neurosecretory dysfunction of GH secretion or GH bioinactivity), in some others the previous diagnosis of ISS was verified in repeated GH stimulation tests but still the highest GH peak in first 2 tests was reported The results of the assessment of GH secretion after falling asleep were not taken into account as most of our patients were qualified to treatment before this test become mandatory. Concentrations of GH were measured by the two-site chemiluminescent enzyme immunometric assay (hGH IMMULITE, DPC) for the quantitative measurement of human GH, calibrated to WHO IRP 98/574 standard, with the analytical sensitivity up to 0.01 μg/l, the calibration range up to 40.0 μg/l, the sensitivity of 0.01 μg/l, the intraassay coefficient of variation (CV) of 5.3-6.5% and the interassay CV of 5.5-6.2%.

7. Serum IGF-I concentration before rhGH therapy onset. During period in which the patients were qualified to treatment, the laboratory provided changing normative data, different for boys and girls of the same age, therefore we decided to calculate IGF-I SDS for all measurements, according to the norms given for particular assays. Because it is known that concentrations of IGF-I in human population are log-normally distributed, the transform was performed as proposed by Blum and Schweitzer [15]:

$$\text{IGF–I SDS} = \frac{\log(\text{IGF–I}) - \text{mIGF–I(log–transformed)}}{\text{SD IGF–I(log–transformed)}}$$

where *IGF-I* is serum level in patient, *mIGF-I (log-transformed)* – mean value of IGF-I concentrations in logarithmically transformed population's distribution, *SD IGF-I (log-transformed)* – standard deviation in the same distribution, with all normative values taken for patient's age and gender. Serum IGF-I concentrations were assessed by a solid-phase, enzyme-labelled chemiluminescent immunometric assay, (IMMULITE, DPC), calibrated to WHO NIBSC 1st IRR 87/518, with analytical sensitivity 20.0 μg/l, the calibration range up to 1600.0 μg/l, the intra-assay CV 3.1-4.3% and the inter-assay CV 5.8-8.4%.

The detailed data concerning the whole cohort of patients are presented in Table I.

Table I. Statistical characteristics of patients' cohort
Tabela I. *Charakterystyka statystyczna grupy pacjentów*

	mean	SD	minimum	maximum
H [cm]	134.0	15.2	78.3	165.3
H SDS	-3.07	0.74	-6.88	-1.82
CA [years]	12.0	2.9	3.0	18.1
BA [years]	9.9	2.9	1.3	15.5
HM [cm]	158.7	5.8	141.2	174.6
HM SDS	-1.21	0.97	-4.13	1.43
HF [cm]	171.6	6.8	152.0	192.0
HF SDS	-1.24	1.14	-4.50	2.17
TH [cm]	168.5	7.4	150.5	187.5
TH SDS	-1.23	0.84	-3.63	1.42
max GH [ng/ml]	9.47	5.69	0.00	40.00
IGF-I SDS	-1.12	1.51	-5.42	3.07
FH SDS	-1.54	0.89	-4.80	1.18

Measure of therapy effectiveness (output data)

The most common and appropriate index of rhGH treatment effectiveness is final height (FH) attained by patient in adultness. According to the annual decrees of the President of National Health Fund, the main goal of the therapeutic program of rhGH therapy in children with GHD is to attain FH on the level of at least 10th centile for age and gender. Unfortunately, the therapy withdrawal criteria changed during the period of study. Previously, rhGH administration was ended when: 1) patient's height exceeded the value of 10th centile for adults – 170 cm for boys and 158 cm for girls, according to Polish reference data [13] or 2) patient's BA was over 16 years for boys and 14 years for girls, or 3) patient's height velocity (HV) was below 3.0 cm/year.

Current requirements for therapy ending give much more certainty about attainment of FH and are following: 1) HV below 3.0 cm/year, except for the patients with BA over 16 years for boys and 14 years for girls and 2) BA over 18 years for boys and 16 years for girls.

Since the criteria changed and definitely not all patients attained real FH at rhGH therapy withdrawal, we calculated FH SDS for the patients who really stopped to grow, while for those who still grew, we assumed that they will keep the same height SDS to adultness, adopted as FH SDS, however this assumption might not be fully met by all the analyzed patients.

Finally, we decided that the problem may be treated in different ways – either as classification (qualitative) task in which the network is trained to determine if particular patient can attain satisfactory FH or as regression task in which the network should predict patient's FH SDS. Actually, we performed 3 different experiments:

- 2-stage classification – the networks were trained to predict if patient will attain satisfactory FH (at least10th centile for adults);
- 3-stage classification – 2-stage classification was modified by dividing patients with unsatisfactory FH in two subgroups: between 3rd and 10th centile – acceptable FH and below 3rd centile – remarkably short FH.
- ANN regression model – the networks were trained to predict the exact value of FH SDS.

Models' derivation

Among many available types on ANN we have chosen to use multilayer perceptron (MLP), that can solve both classification and regression problem. Each neuron in MLP processes data in two stages. Firstly, each input x_i is multiplied by its weight w_i and then they are summed, what can be expressed as following formula:

$$s = \sum_{i=0}^{n} w_i \cdot x_i$$

Secondly, the above sum is passed to the activation function, which in MLP is logistic function:

$$y = \frac{1}{1 + e^{-\beta \cdot s}}$$

This function produces outputs close to 0 for low values of s and close to 1 for high$s$.

All our models were derived in *STATISTICA Neural Networks PL* with the use of automatic creator that allows analysis of numerous networks with the same input and output, but different number of hidden layers (1 or 2) and neurons in those layers. We also allowed elimination of insignificant inputs, to which negligible weight coefficients were assigned.

Before the models were built, we had to divide our database into three sets. First – learning set – is used directly to train the network, second – validation set – to control learning by checking that quality of results for validation and training set is comparable (if it were not it would indicate that the ANN "memorized" trainings set, rather than learned a dependence), third – testing set – to control the networks performance on independent data. In our case, the three sets contained following number of cases:

- learning set: 150 (111 boys, 39 girls),
- validation set: 70 (53 boys, 17 girls),
- testing set: 69 (54 boys, 15 girls).

The detailed data concerning the subgroups (sets) of patients are presented in Table II.

Before being introduced to model all the parameters were automatically normalized by minimax transform, what is a common practice in neural network modelling.

Models' quality measures

We characterized all our models with measures of predictive quality. Quality of classification was expressed as percentage of correct answers. This parameter was calculated for whole dataset and for each of three subsets separately. Among obtained values the most important one is percentage of right classifications in testing set, since it represents how the models function on new data, not being

Table II. Statistical characteristics of patients' subsets
Tabela II. *Charakterystyka statystyczna pacjentów w poszczególnych podzbiorach*

	mean±SD (minimum, maximum)		
	Learning set	Validation set	Testing set
H [cm]	**133.4**±15,1 (89.1, 162.0)	**135.1**±14.6 (78.3, 157.4)	**134.1**±16.2 (83.9, 165.3)
H SDS	**-3.02**±0,72 (-6.88, -1.81)	**-3.16**±0.85 (-5.92, -1.89)	**-3.06**±0.69 (-4.62, -1.82)
CA [years]	**11.9**±2,8 (3.6, 17.1)	**12.3**±2.6 (3.5, 16.2)	**12.0**±3.1 (3.0, 18.1)
BA [years]	**9.8**±2.9 (2.0, 15.5)	**10.2**±2.8 (1.5, 14.0)	**9.8**±3.0 (1.3, 14.5)
HM [cm]	**159.6**±5.4 (146.0, 174.6)	**157.6**±5.9 (142.9, 171.0)	**158.1**±6.2 (141.2, 171.2)
HM SDS	**-1.07**±0,91 (-3.33, 1.43)	**-1.40**±0.98 (-3.85, 0.83)	**-1.32**±1.04 (-4.13, 0.87)
HF [cm]	**171.7**±6,6 (152.0, 192.0)	**172.2**±6.6 (154.0, 190.0)	**170.7**±7.6 (152.0, 188.0)
HF SDS	**-1.22**±1,1 (-4.50, 2.17)	**-1.13**±1.11 (-4.17, 1.83)	**-1.39**±1.26 (-4.5, 1.5)
TH [cm]	**168.7**±7,3 (151.5, 187.5)	**168.3**±7.2 (150.5, 184.7)	**168.0**±7.9 (151.4, 185.5)
TH SDS	**-1.15**±0,8 (-3.63, 1.42)	**-1.26**±0.82 (-2.90, 0.95)	**-1.36**±0.93 (-3.63, 1.08)
max GH [ng/ml]	**9.2**±4.9 (0.6, 31)	**10.5**±7.7 (0.0, 40.0)	**9.00**±4.78 (2.19, 27.7)
IGF-I SDS	**-1.10**±1.48 (-4.54, 3.07)	**-1.31**±1.47 (-5.42, 1.78)	**-0.96**±1.62 (-4.78, 2.53)
FH SDS	**-1.55**±0,86 (-4.80, 1.18)	**-1.53**±0.96 (-3.67, 0.25)	**-1.55**±0.91 (-4.33, -0.12)

involved in learning process. In the case of 2-stage classifier one may expect also calculation of specificity and sensitivity; however here they are not given, because the models were designed in a way that ensures that both those parameters were practically equal to classification accuracy.

Regression models require completely different assessment of quality. We decided to use 2 common measures. First one was root mean square error (RMSE), calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y_{di})^2}{n}}$$

where y_i is predicted value and y_{di} is known correct value of patient FH SDS and n is number of patients taken into account in calculation. Next parameter was coefficient of determination (R^2) that is often interpreted as the amount of variability in data that is explained by model. For ANN model (and any non-linear) it is actually calculated by subtracting amount of unexplained variance from 1:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_{di})^2}{\sum_{i=1}^{n} (y_i - y_m)^2}$$

where y_m is average of patients' FH SDS. We decided to calculate R^2 for whole data and for testing set separately.

Results

Collective models' characteristics

We performed classification experiments with 4 sets of input data, differing with form of patients' and parents' heights: 1) all heights in centimetres and parental heights as 2 separate variables (HM and HF) – marked later as CM.HM.HF, 2) all heights in centimetres and parental heights as TH – CM.TH, 3) all heights as SDS and parental heights as 2 variables (HM SDS and HF SDS) – SDS.HM.HF, 4) all heights as SDS and parental heights as TH SDS – SDS.TH.

In regression models we decided to consequently express all heights as SDS (3rd and 4thposition inlist of inputs, above), since as explained above it seemed necessary for FH.

For all versions we designed 100 models and then chose the best and most interesting examples that are presented in following sections. The architecture of our MLP networks is described in

following way: *MLP: number (N°) of inputs: N° of input neurons – N° of neurons in 1st hidden layer – N° of neurons in 2nd hidden layer (if applicable) – N° of output neurons: N° of outputs*.

2-stage classification

Our best 2-stage classifiers gave about 80% of correct answers. We observed that contrary to more traditional statistical methods, MLP models can process raw heights in centimetres as well as (or even a bit betterthan) corresponding SDS values. Usually introduction of raw heights resulted in less complex models (with less hidden neurons). What is important, all models shown in Table III presented similar accuracy in learning and testing set. The only variable eliminated in some of most accurate models was maximal value of GH peak instimulation tests.

3-stage classification

In case of 3-stage classification the best obtained result was almost 70% of correct answers (Table IV). Similarly to previous experiment, transforming heights to SDS did not improve models accuracy. In this case, we may also be interested in more detailed information about network's answers, which is given for one of our models in Table V (for other networks, results were similar). We marked FH above 10th centile as *>10c*,between 3rd and 10th centile as *<10c* and below 3rd centile as *<3c*. We observe that 16 patients (5.5%) were qualified incorrectly to the group with exceptionally short FH, what seems to be the most serious kind of error. The results indicate also that models eliminating max GH as redundant variable had similar accuracy to those using max GH.

ANN regression

In MLP regression models we were able to predict patient's FH SDS with average error (RMSE) lower than 0.7 SD in testing set, what is equivalent to 4.2 cm (SD of height in population of Polish adults is 6 cm). In the best model we obtained $R^2=0.49$ for whole dataset and $R^2=0.45$ for testing

Table III. Characteristics of best 2-stage classification models
Tabela III. *Charakterystyka najlepszych klasyfikatorów 2-stanowych*

Network architecture	Correct classifications in particular sets [%]				Type of inputs	Eliminated inputs
	L	V	T	All		
MLP 8:8-5-4-1:1	82.7	82.9	79.7	81.0	CM.HM.HF	-
MLP 7:7-4-1:1	83.3	82.9	78.3	81.0	CM.HM.HF	max GH
MLP 7:7-13-11-1:1	81.3	80.0	82.6	81.3	CM.TH	-
MLP 8:8-20-10-1:1	80.7	80.0	78.3	79.9	SDS.HM.HF	-
MLP 7:7-3-6-1:1	78.0	78.6	75.4	77.5	SDS.HM.HF	max GH
MLP 7:7-8-6-1:1	78.0	80.0	76.8	77.5	SDS.TH	-

Table IV. Characteristics of best 3-stage classification models
Tabela IV. *Charakterystyka najlepszych klasyfikatorów 3-stanowych*

Network architecture	Correct classifications in particular sets [%]				Type of inputs	Eliminated inputs
	L	V	T	All		
MLP 7:7-10-3:1	69.3	70.0	69.6	69.6	CM.HM.HF	max GH
MLP 7:7-8-3:1	69.3	67.1	69.6	68.9	CM.TH	-
MLP 8:8-5-3:1	64.7	65.7	63.8	64.7	SDS.HM.HF	
MLP 6:6-4-3:1	65.3	67.1	68.1	66.4	SDS.TH	max GH

Table V. Classification statistics for MLP 7:7-10-3:1
Tabela V. *Statystyki klasyfikacyjne dla sieci MLP 7:7-10-3:1*

		Correct answer		
		>10c	<10c	<3c
Classifier answer	>10c	**128**	27	13
	<10c	14	**31**	18
	<3c	4	12	**42**

set only. Again, max GH was the variable most frequently eliminated from analysis. For detailed data see Table VI.

Discussion

In Poland, the main goal of the rhGH therapy in children with GHD is the attainment of normal FH (*i.e.* exceeding the value of 10th centile for adults). Accurate estimation before treatment if the patient should meet this criterion at completion of growth seems particularly important as it may ensuremore effective implementation ofthe objectives of the therapy. Prediction of growth response allows to be realistic with regard to the expected effects of treatment. Of course, from the point of view of the physicians, it may unethical not qualify for treatment the patients with the most severe growth retardation and more advanced age, who may significantly improve their growth rate and centile position during rhGH therapy but fail to attain normal FH exceeding the cut-off value of 10th centile.

The best 2-stage models presented in current study ensured proper classification of over 80% of patients with respect to the attained FH over or below 10th centile. In 3-stage classification the obtained accuracy equalled to almost 70%, being lower than in previous case due to increased complexity of problem. Still, it seems that introducing the differentiation between short but acceptable (between 3rd and 10th centile) and exceptionally short (below 3rd centile) FH can be beneficial. The authors of present paper did not find previous studies focused on such an approach to the problem, nonetheless introducing such a diagnostic tool seems to be quite useful in practice. In future, it may be possible to modify models characteristics in order to increase their sensitivity, although it usually results in decreased specificity.

The best ANN regression models explained about 45% of variability of the attained FH in testing data, with the mean error of FH SDS below 0.7 SD. In comparison with classic regression models [6-8] these results seem to be quite promising, taking into account the relatively low number of input variables (including only the data from the diagnostics performed before rhGH therapy onset) as well as changing criteria of the therapy withdrawal. It seems that inclusion of the data form patients' history (birth weight), pubertal stage at therapy onset as well as the information on rhGH dose may further improve the predictions of models.

The accuracy of models based on the raw data, expressing patients' and parents' height in cm and those with heights expressed as SDS was similar. Thus the transformation of data, required for typical comparisons of patients of both genders and different ages, seems not absolutely necessary for ANN models, because of their exceptional ability to aggregate input data, among which the age and gender are also included. The possibility of entering the direct data from measurements substantially facilitates the use of models in practice. Also, development of other ANN

Table VI. Characteristics of best MLP regression models
Tabela VI. *Charakterystyka najlepszych regresyjnych modeli MLP*

Network architecture	RMSE [SD]				R2		Type of inputs	Eliminated inputs
	L	V	T	All	T	All		
MLP 8:8-6-1:1	0.61	0.69	0.68	0.65	0.44	0.48	SDS.HM.HF	-
MLP 7:7-4-1:1	0.61	0.65	0.67	0.63	0.45	0.49	SDS.HM.HF	max GH
MLP 5:5-20-12-1:1	0.65	0.69	0.67	0.67	0.44	0.44	SDS.TH	max GH, sex
MLP 6:6-7-1:1	0.64	0.70	0.68	0.67	0.43	0.44	SDS.TH	max GH

models, including the data concerning the initial response to rhGH therapy as predictors of its total effectiveness, seems worth of further studies.

In all experiments (2-stage and 3-stage classification and ANN regression) some of the models automatically removed max GH from the set of input data as a negligible variable, using all the auxological parameters and IGF-I level. It is to some extend surprising, as decreased GH peak in stimulation tests is still the basis of GHD diagnosis. On the other hand, in the regression model of Carel et al. [7] the regression coefficient for GH peak was very low (-0.08). Similarly, in the model of prediction of FH at start of treatment, derived by Ridder et al. [8], GH peak was a predictor of FH only for children who were prepubertal at therapy onset but not for pubertal ones. It is well known that the reproducibility of GH stimulation tests is unsatisfactory, with the relatively low sensitivity and specificity [16]. It has been documented previously by endocrinologists from our study group that measurements of IGF-I concentrations are much more reproducible than GH peaks in stimulation tests [17]. Moreover, rhGH therapy effectiveness in children with normal GH peak in stimulation tests has also been documented [3, 9, 10]. In present study, some patients were treated despite normal results of GH stimulation tests during 1st assessment, as in the repeated stimulation tests – performed due to further decrease of centile position during follow-up – GH secretion turned out to be decreased. All these children had IGF-I deficiency, responding to rhGH administration in generation test, performed soon after 1st assessment of GH secretion. We are convinced that they could not be GH-sufficient during 1st assessment while GH-deficient during 2nd one.

Finally, it should be stressed here that ANN models – in contrast to the majority of standard statistic tools – may provide information on hitherto unclear relationships whose mathematical form is not fully known (or even completely unknown). Actually, neural networks can find complex, non-linear dependencies, selecting at the same time only those input variables that are relevant to analysed problem. Those abilities are considered the most important advantages of ANN when compared with multiple regression models. Moreover, ANN modelling is less restricting with respect to analysed data, not requiring fulfilment of any statistical conditions. It seems that in future AAN models could be useful during qualifying individual short children to treatment with rhGH, monitoring and optimising its effectiveness and identifying non-responders in whom the therapy outcome is substantially worse than expected. Such an approach could contribute to applying the rules of personalized medicine to rhGH administration.

References / Piśmiennictwo

1. Wit J.M., Ranke M.B., Kelnar C.J.H.: The ESPE classification of paediatric endocrine diagnoses – Short stature. Horm. Res., 2007:68, 1-9.
2. Savage M.O., Burren C.P., Rosenfeld R.G.: The continuum of growth hormone-IGF-I axis defects causing short stature: diagnostic and therapeutic challenges. Clin. Endocrinol. (Oxf), 2010:72, 721-728.
3. Richmond E., Rogol A.D.: Current indications for growth hormone therapy for children and adolescents. Endocr. Dev., 2010:18, 92-108.
4. GH Research Society:Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. J. Clin. Endocrinol. Metab., 2000:85, 3990-3993.
5. Rosenfeld R.G., Albertsson-Wikland K., Cassorla F. et al.: Diagnostic Controversy – the Diagnosis of Childhood Growth-Hormone Deficiency Revisited. J. Clin. Endocrinol. Metab., 1995:80, 1532-1540.
6. Ranke M.B., Lindberg A., Chatelain P. et al.: Derivation and validation of a mathematical model for predicting the response to exogenous recombinant human growth hormone (GH) in prepubertal children with idiopathic GH deficiency. J. Clin. Endocrinol. Metab., 1999:84, 1174-1183.
7. Carel J.C., Ecosse E., Nicolino M. et al.: Adult height after long term treatment with recombinant growth hormone for idiopathic isolated growth hormone deficiency: observational follow up study of the French population based registry. BMJ, 2002:13, 325-370.
8. de Ridder M.A.J., Stijnen T., Hokken-Koelega A.C.S.: Prediction of adult height in growth-hormone-treated children with growth hormone deficiency. J. Clin Endocrinol. Metab., 2007:92, 925-931.
9. Cohen P., Rogol A.D., Deal C.L. et al.: Consensus Statement on the Diagnosis and Treatment of Children with Idiopathic Short Stature: A Summary of the Growth Hormone Research Society, the Lawson Wilkins Pediatric Endocrine Society, and the European Society for Paediatric Endocrinology Workshop. J. Clin. Endocrinol. Metab., 2008:93, 4210-4217.
10. Navarro R., Dunn J.D., Lee P.A. et al.: Translating clinical guidelines into practice: the effective and appropriate use of human growth hormone. Am. J. Manag. Care., 2013:19,281-289.
11. Tadeusiewicz R., Chaki R., Chaki N.: Exploring Neural Networks with C#. CRC Press, Taylor & Francis Group, Boca Raton 2014.
12. Yardimci A.: Soft computing in medicine. Appl. Soft. Comput., 2009:9, 1029-1043.
13. Palczewska I., Niedźwiecka Z.: Somatic development indices in children and youth of Warsaw. Med. Wieku Rozwoj., 2001:5, 18-118.
14. Greulich W.W., Pyle S.I.: Radiographic Atlas of Skeletal Development of the Hand and Wrist. Stanford University Press,Stanford 1993.
15. Blum W.R., Schweitzer R.: Insulin-like growth factors and their binding globulins. [In:] Diagnostic of endocrine function in children and adolescents 3rd revised and extended edition. Ranke M.B.(ed),Karger, Basel 2003, 166-199.
16. Webb E.A., Dattani M.T.: Diagnosis of growth hormone deficiency. Endocr. Dev., 2010:18, 55-66.
17. Hilczer M., Smyczynska J., Stawerska R. et al.:Stability of IGF-I concentration despite divergent results of repeated GH stimulating tests indicates poor reproducibility of test results. Endocr. Regul., 2006:40, 37-45.