

Joanna Grabska-Chrzęstowska^{*}, Wiesław Libuszowski^{**}, Waldemar Tomalak^{***}

Weryfikacja użyteczności neuronowych klasyfikatorów wyników badań spirometrycznych****

1. Wprowadzenie

Badanie spirometryczne jest jednym z najpowszechniej wykonywanych badań diagnostycznych w chorobach układu oddechowego. W przypadku astmy i przewlekłej obturacyjnej choroby płuc (POChP) stanowi istotny element w procesie stawiania diagnozy.

Pomiar i ocena tzw. krzywej przepływ – objętość (zależności pomiędzy przepływem wydechowym i objętością wydychanego powietrza podczas manewru forsownego, czyli maksymalnego wydechu) pozwala na ewentualne stwierdzenie obturacji w układzie oddechowym (ograniczenia przepływów wydechowych), bądź podejrzenie zmian o charakterze restrykcyjnym (obniżenia objętości w układzie oddechowym). Więcej informacji na temat badania można znaleźć m.in. w [1] i [6].

W artykule [1] przedstawiono wyniki prac mających na celu ustalenie wpływu liczby wejść do klasyfikatora opartego na sieciach neuronowych typu MLP (*Multi Layered Perception*) na jakość klasyfikacji wyników badania spirometrycznego w odniesieniu do klasyfikacji wykonanej przez eksperta. W kolejnym opracowaniu [3] przedstawiono ideę systemu wnioskującego opartego o zespół trzech sieci neuronowych, który zmniejszał liczbę wyników fałszywie ujemnych niewielkim kosztem wzrostu liczby wyników fałszywie dodatnich, co jest korzystniejsze z medycznego punktu widzenia. Z kolei w artykule [2] porównano wyniki klasyfikacji dotychczas zbadanych sieci i klasycznej metody k -NN. Niniejsza praca ma na celu weryfikację dotychczasowych rezultatów poprzez ocenę działania różnych klasyfikatorów neuronowych na nowej grupie danych i sprawdzenia czułości i swoistości wybranych metod. Dodatkowo – do analiz zastosowano również klasyfikator oparty na metodzie k -NN, czyli metodzie k -najbliższych sąsiadów.

* Katedra Automatyki, Akademia Górniczo-Hutnicza, Kraków; asior@agh.edu.pl

** Doktorant Akademii Górniczo-Hutniczej, Kraków; wlibuszowski@plusnet.pl

*** Zakład Fizjopatologii Układu Oddychania, Rabka-Zdrój; wtomalak@zpigichp.edu.pl

**** Pracę sfinansowano z funduszu badań własnych (umowa AGH nr 10.10.120.39)

Wyniki okazały się zaskakujące. Wszystkie zastosowane metody charakteryzowały się zwiększoną czułością przy dramatycznie zmniejszonej swoistości, a co za tym idzie – również pogorszonej jakości klasyfikacji. Wykazaliśmy, że zarówno metody neuronowe, jak i oparte na klasycznym algorytmie k -NN sprawdzają się w podobnym zakresie. Niektóre sieci, wcześniej nieco gorsze, niespodziewanie znalazły właściwsze rozwiązania. Badania wykazały, że aby uzyskać wiarygodny i uniwersalny klasyfikator, konieczne staje się pozyskanie dodatkowych wyników badań spirometrycznych z różnych okresów czasu i ewentualnie różnych aparatów.

Obliczenia wykonano za pomocą pakietu Statistica Neural Network i programu autorstwa A. Józwika [4].

2. Badanie spirometryczne

Aby wykonać badanie spirometryczne, można zastosować urządzenia mierzące przepływ powietrza w układzie oddechowym i umożliwiające określanie objętości lub jej zmiany na zasadzie integracji przepływu. Jedną z procedur pomiarowych jest rejestracja manewru natężonego wydechu umożliwiającą rejestrację krzywej maksymalny przepływ – objętość oraz wartości z nią związanych.

Zaburzenia wartości charakterystycznych (zwłaszcza FEV1 i FVC) mają wpływ na klasyfikację wyników badania i stanowią podstawę do formułowania wniosków i ustalania diagnozy.

Wartości uzyskane na podstawie badania i użyte jako wejścia do klasyfikatora to:

- FEV₁: objętość powietrza wydmuchnięta w czasie pierwszej sekundy natężonego wydechu;
- FVC: największa objętość powietrza wydmuchnięta przy maksymalnym wysiłku wydechowym po uprzednim największym możliwym wydechu;
- PEF: szczytowy przepływ wydechowy zarejestrowany w trakcie badania maksymalnie natężonego wdechu;
- MEF_{x%}FVC: maksymalny przepływ powietrza, gdy $x\%$ FVC pozostało do wydmuchnięcia; przepływ zmierzony podczas maksymalnie natężonego wydechu, w punkcie odpowiadającym $x\%$ pozostałej do wydmuchnięcia natężonej pojemności życiowej płuc.

Dotychczasowe eksperymenty przeprowadzono w oparciu o 1803 wyniki badań spirometrycznych wykonanych w Zakładzie Fizjopatologii Układu Oddychania Instytutu Gruźlicy i Chorób Płuc w Rabce. Jako dane weryfikujące zastosowano 172 wektory odnoszące się do wyników samych chłopców, pozyskanych w powyższym ośrodku w innym czasie. Wybór tylko jednej płci został spowodowany chęcią uniknięcia dodatkowych nieznanymi elementami wnoszonych przez różnice w budowie płuc u dziewczynek i chłopców. Dodatkowo pokazano rezultat testów dla wszystkich 337 nowych danych obu płci.

3. Zastosowanie sieci neuronowych do klasyfikacji badań spirometrycznych

3.1. Sposób przedstawienia wyników

W zastosowaniach medycznych zazwyczaj wyniki przedstawia się w formie tabeli 2×2 (tab. 1), w której wpisuje się liczbę pacjentów odpowiadającej jednemu z czterech wymienionych w opisie tabeli przypadków.

Tabela 1
Sposób przedstawienia wyników rozpoznania normy od zaburzenia dla wybranej metody

	NORMA sieć	ZABURZENIE sieć
NORMA ekspert	<i>a</i>	<i>c</i>
ZABURZENIE ekspert	<i>b</i>	<i>d</i>

OBJAŚNIENIA:

- a* – liczba przypadków klasyfikowanych w normie przez eksperta i sieć – przypadki prawdziwie ujemne
- b* – liczba przypadków uznanych przez eksperta za normę, a przez sieć za zaburzenie – fałszywie ujemne
- c* – liczba przypadków klasyfikowanych przez eksperta jako zaburzenie a przez sieć jako norma – fałszywie dodatnie
- d* – liczba przypadków kwalifikowanych przez eksperta i sieć jako zaburzenie – prawdziwie dodatnie

Jakość rozpoznania, czyli liczbę poprawnych rozpoznań, oblicza się za pomocą następującego wzoru

$$\text{liczba poprawnych rozpoznaw} = (a + d) * 100\% \quad (1)$$

Oprócz jakości rozpoznania wprowadza się także dwa parametry statystyczne charakteryzujące wartość diagnostyczną zaproponowanego testu. Zdolność testu do wykrywania choroby u rzeczywiście chorych pacjentów nazywamy czułością metody, a umiejętność testu wykluczenia choroby u osób rzeczywiście zdrowych jej swoistością. Wyniki podane przez eksperta przyjmuje się za całkowicie pewne i wiarygodne. Wartości parametrów czułości i swoistości oblicza się według następujących wzorów:

$$\text{czulosc} = \frac{d}{d + b} * 100\% \quad (2)$$

$$\text{swoistosc} = \frac{a}{a + c} * 100\% \quad (3)$$

3.2. Dotychczasowe wyniki zastosowania sieci neuronowych

Sieci neuronowe typu feedforward, uczone za pomocą algorytmu ze wsteczną propagacją błędów, są chętnie stosowane do zadań klasyfikacji [5]. Aby znaleźć właściwą strukturę sieci, trzeba określić liczbę jej wejść i wyjść. W efekcie badania spirometrycznego uzyskujemy 6 wartości charakteryzujących przepływ w drogach oddechowych: FVC, FEV1, PEV, MEF75%FVC, MEF50%FVC oraz MEF25%FVC. Dodatkowo, pacjenta mogą charakteryzować dane antropometryczne: wzrost, waga i wiek.

Dotychczasowe eksperymenty przedstawione szczegółowo w pracy [1] wskazały na parametr wzrostu jako najważniejszy z wszystkich cech antropometrycznych. Sieć neuronowa z sześcioma wejściami będącymi rezultatem badania spirometrycznego i dodanym parametrem wzrostu okazała się wstępnie najlepszym klasyfikatorem dla tego typu badań. Przy próbach redukcji liczby wejść dopuszczających także zmianę liczby parametrów oddechowych otrzymano podobne rezultaty. Należy dodać, że minimalna struktura wejściowa, którą otrzymano, zawierała tylko cztery elementy, oprócz wzrostu również wagę pacjenta przy dwóch wartościach oddechowych: MEF50%FVC i MEF25%FVC.

Następne próby – uzyskania klasyfikatora o lepszej czułości – zaowocowały opracowaniem systemu trzech sieci, o nieco lepszych wynikach rozpoznawania i możliwością poprawy tych rezultatów. Dobór optymalnych progów klasyfikacji każdej z sieci składowych systemu pozostał do dalszych badań.

Wyniki jakości klasyfikacji, czułości i swoistości przedstawiono w tabeli 2.

Tabela 2
Wyniki jakości poszczególnych metod dla 511 wyników (chłopców)

	Liczba prawidłowych rozpoznań [%]	Czułość metody [%]	Swoistość metody [%]
Sieć nr 1: 6 wejść* (próg = 0,5)**	85,7	72,9	92,5
1-NN***: 6 wejść	79,5	70,6	85
Sieć nr 2: 6 wejść (próg = 0,9)	83,1	59,3	98,9
23-NN: 6 wejść	82,8	66,5	93
Sieć nr 3: 6 wejść + wzrost	88,6	82,5	92,3
1-NN: 6 wejść + wzrost	82,6	74,1	87,9
23-NN: 6 wejść + wzrost	84,7	65,5	96,8
System sieci	90,2	92	87

OBJAŚNIENIA:

- * – 6 parametrów oddechowych: FVC, FEV1, PEV, MEF75%FVC, MEF50%FVC i MEF25%FVC
- ** – próg = n – w przypadku jednego wyjścia binarnego wynik uznaje się za prawidłowy, gdy dla 0 odpowiedź sieci $< n$, dla 1 odpowiedź sieci $0 \geq n$
- *** – 1-NN – metoda k -NN opierająca się na jednym sąsiedzie

Dla porównania pokazano rezultaty otrzymane za pomocą metody k -najbliższych sąsiadów dla tych samych elementów na wejściu. Wyraźnie uwypukla się tendencja sieci neuronowych, bez parametru wzrostu na wejściu (sieć nr 1 i 2), do zwiększonej możliwości wykrywania zdrowych pacjentów (pojęcie swoistości testu) przy bardzo obniżonej zdolności wykrywania chorych osób (czułość testu). Dla metody k -NN (gdzie k oznacza liczbę sąsiadów) przy zwiększonej liczbie sąsiadów (do 23) też można zaobserwować podobną tendencję, czyli zwiększanie swoistości kosztem wyraźnego spadku czułości. Cały nacisk przy powstawaniu koncepcji i stworzeniu systemu trzech sieci był nastawiony na zmianę tej tendencji na przeciwną. Wysiłki opłaciły się, system sieci charakteryzuje się najlepszą czułością przy dopuszczalnej swoistości.

3.3. Wyniki testowania sieci neuronowych nowymi danymi

Aby przedstawić szczegółowe odpowiedzi poszczególnych klasyfikatorów na nowy zestaw danych, konieczne stało się pokazanie tabeli z liczbami określającymi klasyfikację eksperta i rezultaty przyjętej metody badawczej oraz sprawdzenie, czy czułość i swoistość testu jest wiarygodna.

Dla najlepszej pojedynczej sieci o 7 wejściach, 6 neuronach w warstwie ukrytej i jednym wyjściu otrzymane rezultaty przedstawiono w tabeli 3. Wyniki w zakresie diagnozowania normy nie są dobre. Sieć podała tylko 2/3 prawidłowych odpowiedzi. Z tego wynika, że 1/3 zdrowych pacjentów trafiałaby bez żadnego klinicznego uzasadnienia do dalszych badań. Natomiast nadspodziewanie dobrze klasyfikator radzi sobie z pacjentami o udokumentowanych nieprawidłowościach w układzie oddechowym. Aż 94% chłopców z zaburzeniami oddechowymi zostałoby zdiagnozowanych poprawnie. Z medycznego punktu widzenia taki wynik można by uznać za zadowalający, gdyż najważniejszym celem badań jest wychwycenie przypadków prawdziwie dodatnich.

Interesującym przypadkiem było zastosowanie sieci bez parametru wzrostu (tab. 4), która w poprzednich badaniach ustępowała innym klasyfikatorom zwłaszcza w diagnozowaniu chorych pacjentów.

Niespodziewanie okazało się, że zastosowanie tej samej sieci z innym progiem klasyfikacji – 0,9 – daje najlepsze wyniki. Przyjęcie progu 0,9 oznacza, że odpowiedzi z przedziału (0, 0,9) uznano za będące w normie, a z przedziału <0,9, 1) sklasyfikowano jako zaburzenie (tab. 5). Klasyfikacja zdrowych pacjentów bardzo się poprawiła bez znaczącego pogorszenia liczby źle zdiagnozowanych chorych chłopców. Niestety, nie można z góry przewidzieć, że gorsza sieć będzie w przypadku innej grupy pacjentów lepszym klasyfikatorem.

Tabela 3
Wyniki testowania sieci o 6 parametrach oddechowych i parametrze wzrostu (172 chłopców)

	NORMA sieć	ZABURZENIE sieć
NORMA ekspert	81	42
ZABURZENIE ekspert	3	46

Tabela 4

Wyniki testowania sieci o 6 parametrach oddechowych na wejściu – próg 0,5 (172 chłopców)

	NORMA sieć	ZABURZENIE sieć
NORMA ekspert	95	28
ZABURZENIE ekspert	7	42

Tabela 5

Wyniki testowania sieci o 6 parametrach oddechowych na wejściu – próg 0,9 (172 chłopców)

	NORMA sieć	ZABURZENIE sieć
NORMA ekspert	110	13
ZABURZENIE ekspert	8	41

Tabela 6

Wyniki testowania systemu sieci (172 chłopców)

	NORMA sieć	ZABURZENIE sieć
NORMA ekspert	79	44
ZABURZENIE ekspert	0	49

System sieci w przypadku nowych danych wychwycił wszystkich chorych pacjentów, ale zupełnie nie sprawdził w diagnozowaniu normy (tab. 6).

4. Porównanie wyników sieci z metodą k -NN

Aby stwierdzić, czy dotychczasowe modele zawiodły, czy to zbiór nowych danych stanowi problem, porównano wyniki sieci neuronowych z metodą k -NN.

Zestawienie wszystkich rezultatów testowania kolejnych metod nowym zbiorem pacjentów przedstawiono w tabeli 7.

Metoda k -najbliższych sąsiadów w dotychczasowych badaniach ustępowała nieco wszystkim neuronowym klasyfikatorom [2].

Jednakże z trzech metod uznanych poprzednio za najlepsze klasyfikatory (druga część tabeli), najbardziej stabilna wydaje się metoda 1-NN. Idea systemu sieci powstała z myślą

o polepszeniu czułości metody kosztem spadku swoistości. Dla nowych danych zaowocowało to 100% czułością, ale niestety z nie do przyjęcia swoistością – poniżej 70%.

Można zauważyć globalną tendencję we wszystkich klasyfikatorach: wzrostu czułości i spadku swoistości. Najlepiej na tym wyszły metody z dotychczas niedopuszczalnie niską wartością czułości i wysoką swoistością. Nie można w żaden sposób stwierdzić, co w nowym zestawie wektorów spowodowało taką tendencję.

Tabela 7
Wyniki jakości poszczególnych metod (172 chłopców)

	Liczba prawidłowych rozpoznań [%]	Czułość metody [%]	Swoistość metody [%]
Sieć nr 1: 6 wejść (próg 0,5)	79,7	85,7	77,2
1-NN: 6 wejść	74,4	85,7	69,9
Sieć nr 2: 6 wejść (próg 0,9)	87,8	83,7	89,4
23-NN: 6 wejść	74,4	85,7	69,9
Sieć nr 3: 6 wejść + wzrost	73,8	93,9	65,9
1-NN: 6 wejść + wzrost	79,1	89,8	74,8
System sieci	74,4	100	64,2

Dla ostatecznej weryfikacji wyników dla wszystkich nowych danych (337 dziewcząt i chłopców) pokazano w tabeli 8 wartości jakości, czułości i swoistości sieci z dodanym parametrem wzrostu i wskaźnikiem płci i metody k -NN też o 8 elementach wejściowych.

Tabela 8
Wyniki jakości dwóch metod (172 chłopców + 165 dziewcząt)

	Liczba prawidłowych rozpoznań [%]	Czułość metody [%]	Swoistość metody [%]
Sieć: 6 wejść + wzrost + płeć	72,1	97,3	60
1-NN: 6 wejść + wzrost + płeć	75,1	91,1	67,1

Podobnie jak w przypadku samych chłopców, sieci neuronowe prawidłowo sklasyfikowały mniej przypadków niż metoda 1-NN, ale wykazały się bardzo dobrą czułością. Dla obu metod wartość parametru swoistości testu jest zbyt niska.

5. Wnioski

W pracy skrótowo przypomniano dotychczasowe wyniki badań nad znalezieniem uniwersalnego klasyfikatora do automatycznego oddzielania normy od zaburzenia w przypadku badań spirometrycznych. Następnie przedstawiono wyniki testowania wszystkich otrzymanych wcześniej modeli nowym niezależnym zbiorem danych.

Wydaje się, że metody k -NN dające dotychczas nieco gorsze wyniki klasyfikacji okazały się bardziej stabilne od wcześniej preferowanych wybranych sieci neuronowych. Mimo że grupa nowych danych wydaje się w jakiś sposób niepodobna do poprzedniej grupy wektorów, to jednak wyniki metody znajdowania najbliższego sąsiada charakteryzują się dobrą czułością, ale przy stosunkowo niskiej swoistości.

Z kolei zbudowany system sieci neuronowych na razie nie sprawdził się przy klasyfikacji zdrowych pacjentów, ale zgodnie z przeznaczeniem, w 100% rozpoznał wszystkich chorych z testowej grupy.

Jednak najważniejsza konkluzja jest następująca: aby osiągnąć dalekosiężny cel zbudowania automatycznego i uniwersalnego klasyfikatora badań spirometrycznych, konieczne staje się pozyskanie jeszcze większej liczby danych najlepiej z różnych okresów i być może różnych aparatów.

Literatura

- [1] Grabska-Chrzęstowska J., Libuszowski W., Tomalak W.: *Analiza wpływu parametrów wejściowych na klasyfikację wyników badania spirometrycznego przy pomocy sieci neuronowych*. Półrocznik AGH Automatyka, t. 8, z. 3, 2004, 469-475, ISSN 1429-3447
- [2] Grabska-Chrzęstowska J., Libuszowski W., Tomalak W.: *Porównanie metody k NN i techniki sieci neuronowych w klasyfikacji badań spirometrycznych*. Materiały konferencyjne „Modelowanie Cybernetyczne Systemów Biologicznych”, Kraków, UJ i AGH 2005
- [3] Grabska-Chrzęstowska J., Tomalak W.: *Zastosowanie neuronowego systemu wnioskującego do klasyfikacji wyników badania spirometrycznego*. Materiały konferencyjne „Sztuczna Inteligencja w Inżynierii Biomedycznej”, Kraków, AGH 2004
- [4] Jóźwik A., Serpico S., Roli F.: *A parallel network of modified 1-NN and k -NN classifiers - application to remote-sensing image classification*. Pattern Recognition Letters, 19, 1998, 57-62,
- [5] Tadeusiewicz R.: *Sieci neuronowe*. Warszawa, PWN 1993
- [6] Tomalak W.: *Badania czynnościowe układu oddechowego*. Materiały III Sympozjum „Modelowanie i Pomiary w Medycynie”, Krynica 2001, 37-49