

Zbigniew OMIOTEK
Andrzej BURDA
Waldemar WÓJCIK

METODA KLASYFIKACJI OBRAZÓW USG TARCZYCY Z WYKORZYSTANIEM INDUKCJI DRZEW DECYZYJ

STRESZCZENIE *W artykule przedstawiono metodę klasyfikacji zdjęć USG tarczycy. Metoda ta pozwala zaklasyfikować analizowane przypadki do jednej z dwóch kategorii: chory lub zdrowy. Wyselekcjonowana w trakcie badania grupa przypadków błędnie klasyfikowanych zdaniem autorów może zawierać cechy charakterystyczne dla wczesnego stadium rozwoju choroby Hashimoto. Do budowy modelu klasyfikacji wykorzystano indukcję drzew decyzyjnych. Wyniki testów pokazały, iż zaproponowana metoda może stanowić punkt wyjścia do budowy systemu wspomagającego lekarza w procesie diagnozy.*

Słowa kluczowe: *tarczyca, choroba Hashimoto, przetwarzanie obrazów, klasyfikacja tekstur, drzewa decyzyjne*

mgr inż. Zbigniew OMIOTEK, mgr inż. Andrzej BURDA
e-mail: zomiotek@gmail.com, aburda@wsz.zia.edu.pl

Wyższa Szkoła Zarządzania i Administracji w Zamościu,
Wydział Zarządzania i Administracji

prof. dr hab. inż. Waldemar WÓJCIK
e-mail: waldemar.wojcik@pollub.pl

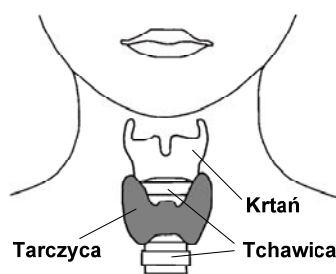
Politechnika Lubelska, Wydział Elektrotechniki i Informatyki

PRACE INSTYTUTU ELEKTROTECHNIKI, zeszyt 260, 2012

1. WSTĘP

Tarczycyca to gruczoł wydzielania wewnętrznego umiejscowiony w przedniej, dolnej części szyi (rys. 1). Zbudowana jest z dwóch płatów bocznych połączonych wąskim przejściem (wąziwą). W sporadycznych przypadkach niektóre osoby posiadają również trzeci płat, tzw. płat piramidowy. Tarczycyca wytwarza hormony zwane trójiodotyroniną (T3), tyroksyną (T4) i kalcytoniną, które wpływają na metabolizm i gospodarkę wapniowo-fosforową organizmu [15].

Jedną z częstszych chorób tego gruczołu jest choroba Hashimoto. Poza jej klasyczną postacią, wyróżnia się kilka odmian tej choroby: poporodowe



Rys. 1. Umiejscowienie gruczołu tarczycy

zapalenie tarczycy, bezobjawowe zapalenie tarczycy, zanikowe autoimmunologiczne zapalenie tarczycy, ogniskowe limfocytarne zapalenie tarczycy. Nazwa choroby pochodzi od nazwiska japońskiego lekarza – Hakuru Hashimoto, który pracując w Berlinie odkrył i zdiagnozował cztery przypadki, a w 1912 r. przedstawił klasyczny obraz tej choroby. Choroba Hashimoto jest chorobą autoimmunologiczną – związaną z zaburzeniami układu odpornościowego. Często jest ona również nazywana przewlekłym limfocyto-

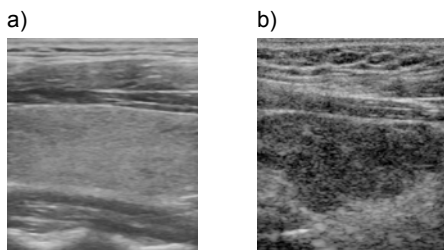
wym zapaleniem gruczołu tarczowego. Choroba ta jest najczęstszą przyczyną niedoczynności tarczycy, a także najczęściej występującym rodzajem zapalenia tego gruczołu. Występuje ona u około 2% populacji, przy czym szacuje się, że kobiety chorują na nią ponad 10 razy częściej niż mężczyźni. Rozpoznaje się ją przeważnie w okresie od 30-go do 50-go roku życia [11].

Choroba Hashimoto przejawia się powstaniem charakterystycznych zmian zapalnych w obrębie tarczycy i może spowodować zarówno rozrost, jak i zanik tego gruczołu. Może ona przebiegać w postaci ogniskowej lub rozsianej. Ogniskowa postać choroby Hashimoto powoduje umiarkowane zmiany w tarczycy i zwykle nie wywołuje jej zniszczenia. Rozsiana uogólniona postać tej choroby obejmuje całą tarczycę lub znaczną jej część. Zmieniona zapalnie tarczycyca powiększa się stopniowo, tworząc wole, które jest zwykle symetryczne, twarde o mniej lub bardziej gładkiej powierzchni. Leczenie rozsianej postaci choroby jest na ogół nieskuteczne [5].

Rozpoznanie wczesnego stadium rozwoju choroby Hashimoto ma bardzo duże znaczenie dla jej skutecznego leczenia. Pacjenci z takim rozpoznaniem mogą być bowiem znacznie wcześniej kierowani na specjalistyczne badania laboratoryjne i obrazowe. Mogą być także wcześniej wdrożone od-

powiednie procedury leczenia, co jest istotne z punktu widzenia skuteczności terapii. Mimo rozwoju metod diagnozowania choroby Hashimoto, nadal dużym problemem jest trafne rozpoznanie jej wczesnego stadium. Wynika to z faktu, że zaburzenia czynności tarczycy wykazują dużą zmienność objawów i często charakteryzują się powolną dynamiką ich narastania. Choroba rozpoznawana jest zwykle w zaawansowanej postaci, gdy dochodzi już do niedoczynności tarczycy i wdrożenie odpowiedniego leczenia jest bardzo trudne lub wręcz niemożliwe. Dlatego też, niezwykle istotną kwestią jest zwiększenie skuteczności istniejących metod diagnostycznych tak, aby rozpoznanie choroby mogło być dokonane w jak najwcześniejszym stadium [2].

Podstawową metodą diagnostyki tarczycy jest obecnie ciągle badanie ultrasonograficzne. Metoda ta wykorzystuje zjawisko rozchodzenia się, rozpraszania oraz odbicia fali ultradźwiękowej na granicy ośrodków, przy założeniu stałej prędkości fali w różnych tkankach. Różne tkanki w różny sposób odbijają ultradźwięki, dlatego można uzyskać obraz przekroju tarczycy widoczny na monitorze. Dzięki obrazom USG można ocenić: wielkość tarczycy, echogeniczność i niejednorodność mięszu, występowanie guzków, zwapnień i płynów (torbiele). Charakterystyczny obraz ultrasonograficzny dla choroby Hashimoto to obniżona echogeniczność i niejednorodność mięszu (rys. 2).



Rys. 2. Zdjęcia USG tarczycy:

- a) gruczoł zdrowy (mięsz jednorodny, normalna echogeniczność);
- b) gruczoł z rozpoznaniem Hasimoto (mięsz niejednorodny, hypoechogeniczność)

Na podstawie zdjęć USG lekarz specjalista może stosunkowo łatwo rozpoznać przypadki tarczycy zdrowej oraz dotkniętej chorobą w jej zaawansowanym stadium. Jednak rozpoznanie tą metodą początkowego stadium rozwoju choroby nie jest zadaniem łatwym. Dlatego też wydaje się celowe opracowanie komputerowej metody automatycznego rozpoznawania i klasyfikacji zdjęć USG tarczycy. Metoda ta wspomagałaby lekarza w procesie diagnozy przypadków dotkniętych wczesnym stadium rozwoju choroby. W niniejszej pracy zaproponowano koncepcję metody, która w procesie klasyfikacji wykorzystuje indukcję drzew decyzyjnych. Pozwala ona zaklasyfikować analizowane przypadki do jednej z dwóch kategorii: *chory* lub *zdrowy*.

W rozdziale pierwszym scharakteryzowano metodę klasyfikacji za pomocą indukcji drzew decyzyjnych. Rozdział drugi opisuje przetwarzanie i analizę

źródłowych obrazów USG wykorzystanych do badań. Procedurę budowy i testowania modelu klasyfikatora zaprezentowano w rozdziale trzecim. Pracę kończą wnioski zamieszczone w rozdziale czwartym.

2. KLASYFIKACJA OBIEKTÓW ZA POMOCĄ INDUKCJI DRZEW DECYZJI

Klasyfikacja jest jedną z najstarszych i zarazem najważniejszych metod eksploracji danych. Polega ona na znajdowaniu odwzorowania danych w zbiór predefiniowanych klas. Na podstawie zawartości bazy danych budowany jest model, który służy do klasyfikowania nowych obiektów w bazie danych lub lepszemu zrozumieniu istniejącego podziału obiektów na predefiniowane klasy. Klasyfikacja znalazła szereg zastosowań, jest to np. automatyczne rozpoznawanie obiektów w dużych bazach danych obrazów, wspomaganie decyzji przyznawania kredytów bankowych, rozpoznawanie trendów na rynkach finansowych. Ogromne zastosowanie znalazła także w systemach medycznych, gdzie w bazie danych medycznych mogą być znalezione reguły klasyfikujące poszczególne schorzenia, a następnie przy pomocy znalezionych reguł automatycznie może być przeprowadzone diagnozowanie nowych pacjentów. Klasyfikacja jest procesem dwuetapowym. W pierwszym etapie konstruowany jest model (klasyfikator), opisujący predefiniowany zbiór klas danych lub zbiór pojęć. Dane wejściowe stanowi tutaj treningowy zbiór próbek (przypadków, obserwacji), będących listą wartości atrybutów opisowych (deskryptorów) i wybranego atrybutu decyzyjnego. Klasyfikator *uczy się* zadania polegającego na tym, aby w oparciu o wartości deskryptorów, każdemu przypadkowi ze zbioru uczącego przydzielić taką wartość atrybutu decyzyjnego, aby zminimalizować błąd klasyfikacji na tym zbiorze. W drugim etapie klasyfikacji otrzymany model stosowany jest do klasyfikacji/predykcji nowych obiektów.

W literaturze zaproponowano wiele metod klasyfikacji [13]. Można wśród nich wyróżnić: klasyfikację poprzez indukcję drzew decyzyji, sztuczne sieci neuronowe, zbiory przybliżone, klasyfikatory Bayes'owskie, metaheurystyki (np. algorytmy genetyczne), metodę *k*-najbliższych sąsiadów, analizę statystyczną oraz wiele innych metod, które są ciągle rozwijane. Przykłady zastosowań wybranych metod uczenia maszynowego w medycynie opisano w [3] i [6]. Wśród wymienionych wcześniej metod klasyfikacji najbardziej popularna jest metoda wykorzystująca indukcję drzew decyzyji. Jest ona szczególnie atrakcyjna ze względu na intuicyjny, zrozumiały dla człowieka, a zatem przyjazny sposób reprezentacji wiedzy [14].

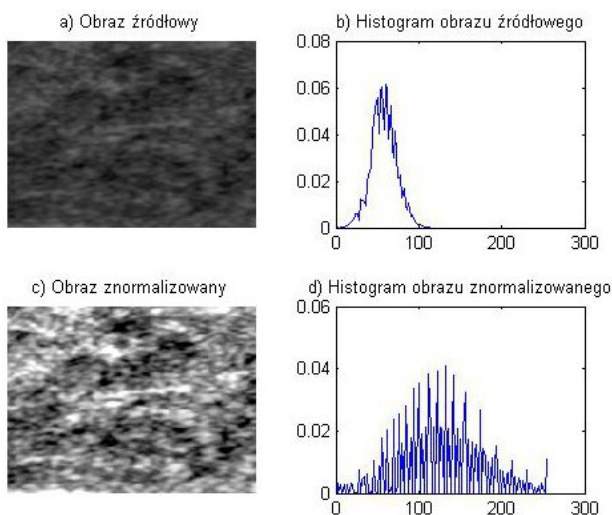
Drzewa decyzji mogą być konstruowane stosunkowo szybko w porównaniu z innymi metodami klasyfikacji. Główną ich zaletą jest jawna reprezentacja wiedzy, skalowalność przy wykorzystaniu dużych zbiorów danych oraz możliwość użycia danych wielowymiarowych. Ponadto dokładność tej metody jest porównywalna z dokładnością innych metod klasyfikacji. Większość dostępnych na rynku narzędzi komercyjnych do eksploracji danych opiera się właśnie na modelu drzew decyzji. Natomiast zasadniczą wadą omawianej metody, w jej podstawowej postaci, jest duża wrażliwość na brakujące wartości atrybutów, gdyż u ich podstaw istnieje niewyrażone jawnie założenie o pełnej dostępności informacji zgromadzonych w bazie przypadkach. Do wad należy zaliczyć również brak możliwości wychwycenia korelacji pomiędzy atrybutami.

W metodzie indukcji drzew decyzji wynikiem klasyfikacji jest skierowany acykliczny graf o strukturze drzewiastej. Każdy wierzchołek (węzeł) grafu reprezentuje test na atrybutach, każdy łuk reprezentuje wynik testu, a każdy liść reprezentuje pojedynczą klasę lub rozkład wartości klas. Najwyższy węzeł nazywany jest korzeniem drzewa. Drzewo decyzji dzieli rekurencyjnie zbiór treningowy na partycje do momentu, w którym każda partycja zawiera dane należące do jednej klasy lub do momentu, gdy w ramach partycji dominują dane należące do jednej klasy. Należy tutaj zaznaczyć, iż rozmiar partycji jest ograniczony. Każdy wierzchołek wewnętrzny drzewa stanowi tzw. punkt podziału, realizujący test na atrybutach, który dzieli zbiór danych na partycje.

Podstawowym algorytmem konstrukcji drzew decyzji jest algorytm zachłanny. Tworzy on drzewo rekurencyjnie techniką top-down w sposób „dziel i rządź”. Istnieje wiele wariantów tego algorytmu, wśród których najczęściej stosowanymi są algorytmy ID3 oraz C4.5. Inne metody, to metoda C&RT, wywodząca się ze statystyki, czy metoda CHAID związana z rozpoznawaniem obrazów. Powyższe algorytmy różnią się głównie przyjętym kryterium podziału, czyli sposobem, w jaki tworzone są nowe węzły wewnętrzne w drzewie. Kryterium podziału powinno maksymalizować dokładność konstruowanego drzewa lub innymi słowy, minimalizować błędną klasyfikację danych. Drzewo decyzji jest zwykle konstruowane w dwóch fazach. W fazie pierwszej, zwanej fazą budowania, wzrostu lub indukcji, drzewo jest tworzone na podstawie treningowej bazy danych. W fazie drugiej, zwanej fazą przycinania lub redukcji, następuje przycinanie drzewa w celu poprawy dokładności i uniezależnienia się od efektu przetrenowania. W fazie przycinania następuje identyfikacja i usunięcie gałęzi reprezentujących punkty osobliwe i szum występujący w danych. Z przycinaniem drzewa wiążą się dwie główne strategie: *postpruning*, w ramach której konstruujemy pełne drzewo decyzji i usuwamy z niego zawodne części oraz *prepruning*, gdzie gałąź przestaje być rozwijana, gdy informacje zaczynają być zawodne. W praktyce preferowany jest *postpruning*, ponieważ *prepruning* często powoduje efekt tzw. „wczesnego stopu” [4].

3. PRZETWARZANIE I ANALIZA OBRAZÓW ŹRÓDŁOWYCH

Materiał do badań zawierał serie zdjęć USG tarczycy należące do 66-ciu pacjentów. Wśród nich było 31 osób zdrowych oraz 35 chorych, u których rozpoznano chorobę Hashimoto. Dla każdej serii zdjęć wyodrębniono od jednego do trzech regionów zainteresowania o rozmiarach 200x150 pikseli, które obejmowały centralne obszary lewego i prawego gruczołu tarczycy. W efekcie otrzymano 54 próbki dla przypadków rozpoznanych jako zdrowe oraz 85 dla chorych (w sumie 139 przypadków). Następnie wyodrębnione fragmenty zdjęć poddano normalizacji obrazu z korekcją liniową w programie Matlab [16]. Na rysunku 3 przedstawiono przykładową próbkę obrazu przed i po normalizacji.



Rys. 3. Przykładowa próbka obrazu przed i po normalizacji

Do ekstrakcji cech dyskryminacyjnych analizowanych obrazów wykorzystano program MaZda opracowany w Instytucie Elektroniki Politechniki Łódzkiej i udostępniony bezpłatnie do celów naukowych w Internecie [17]. Program MaZda (nazwa od „**m**acierz **z**darzeń”) pozwala analizować szare obrazy teksturowe i wyznaczać obszerny zbiór 283 deskryptorów cech. Ich dokładny opis można znaleźć w pozycjach [7, 8, 10, 12], a także w dokumentacji programu MaZda.

Zbiór cech otrzymano na podstawie:

1. Wektora histogramu obrazu (9 cech): średnia, wariancja, skośność, kurtoza, percentyle 1%, 10%, 50%, 90% i 99%.
2. Macierzy gradientu obrazu (5 cech): średnia, wariancja, skośność, kurtoza, odsetek pikseli z niezerowym gradientem.

3. Macierzy rozkładu długości pasm (5 cech x 4 kierunki): nierównomierność pasm, nierównomierność szarości, uwydatnienie krótkich pasm, uwydatnienie długich pasm, część obrazu w pasmach.
4. Macierzy przejść tonalnych (11 cech x 4 kierunki x 5 odległości): drugi moment kątowy, kontrast, korelacja, suma kwadratów, odwrotny moment różnicowy, sumaryczna średnia, sumaryczna wariancja, sumaryczna entropia, entropia, wariancja różnicowa, entropia różnicowa.
5. Modelu autoregresji (5 cech): parametry θ_1 , θ_2 , θ_3 , θ_4 , odchylenie standardowe.
6. Dyskretnej transformaty falkowej (24 cechy): energia współczynników falkowych (falki Haara) dla czterech podpasm (LL, LH, HL, HH) i sześciu skal.

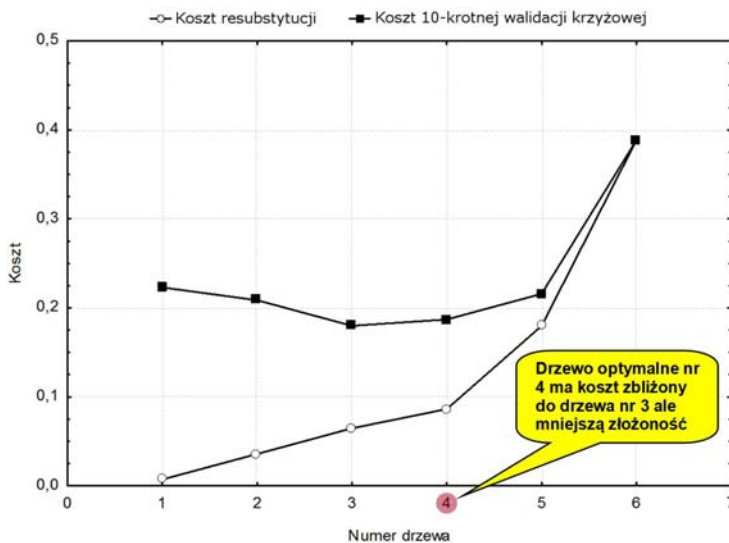
4. BUDOWA I TESTOWANIE KLASYFIKATORA

Opis eksperymentu

Eksperyment przeprowadzono z wykorzystaniem algorytmu C&RT [1]. Do oceny jakości modelu, z uwagi na ograniczoną liczbę przypadków w analizowanym zbiorze, zastosowano metodę dziesięciokrotnej walidacji krzyżowej. Ze względu na wspomniany wstępny etap badań, z zamysłem pominięto procedury eliminowania nieistotnych, ze względu na cel klasyfikacji, atrybutów opisowych, a uznano za zasadne ocenić ich ważność w procesie parametryzacji drzewa decyzji. Przyjęto też równe koszty błędnej klasyfikacji, miarę Giniego jako kryterium dobroci dopasowania i szacowane na podstawie liczebności obu kategorii, prawdopodobieństwo ich występowania *a priori*. Minimalną liczbę przypadków w węźle określono na 13, a maksymalną liczbę węzłów na 1000. Jako regułę stopu wybrano przycięcie przy błędzie złej klasyfikacji [9].

Wyniki badań

W efekcie przycinania powstało 6 drzew decyzji, dla których koszty błędnej klasyfikacji: *Koszt resubstytucji* oraz *Koszt 10-krotnej walidacji krzyżowej* przedstawiono na rysunku 4. Graf drzewa optymalnego (nr 4), ze względu na wielkość błędu klasyfikacji na zbiorze testowym, pokazano na rysunku 5, a macierz klasyfikacji przypadków analizowanego zbioru – w tabeli 1. Z kolei tabela 2 przedstawia ważność atrybutów opisowych ze względu na przyjęty cel klasyfikacji, przy zastosowaniu drzewa optymalnego.



Rys. 4. Sekwencja kosztów błędnej klasyfikacji

TABELA 1
Macierz klasyfikacji

Wielkość	Obserwowane	Przewidywane		Łącznie w wierszu
		Chory	Zdrowy	
Częstość	Chory	76	9	85
Udział [%]		89,41	10,59	100
Częstość	Zdrowy	3	51	54
Udział [%]		5,56	94,44	100

TABELA 2
Ważność atrybutów opisowych

Lp.	Nazwa atrybutu	Znaczenie	Ważność
1.	Teta2	Par. Θ_2 mod. autoregr.	100
2.	WavEnLH_s-6	Energia współczynników falkowych	96
3.	WavEnHL_s-6		87
4.	WavEnHH_s-5		87
5.	WavEnLL_s-4		84
6.	S(5,5)DifVarnc		Wariancja różnicowa
7.	S(5,-5)DifVarnc	74	
8.	WavEnLH_s-3	Energia współczynników falkowych	73
9.	WavEnLL_s-1		73
10.	WavEnLL_s-2		71

Dyskusja wyników

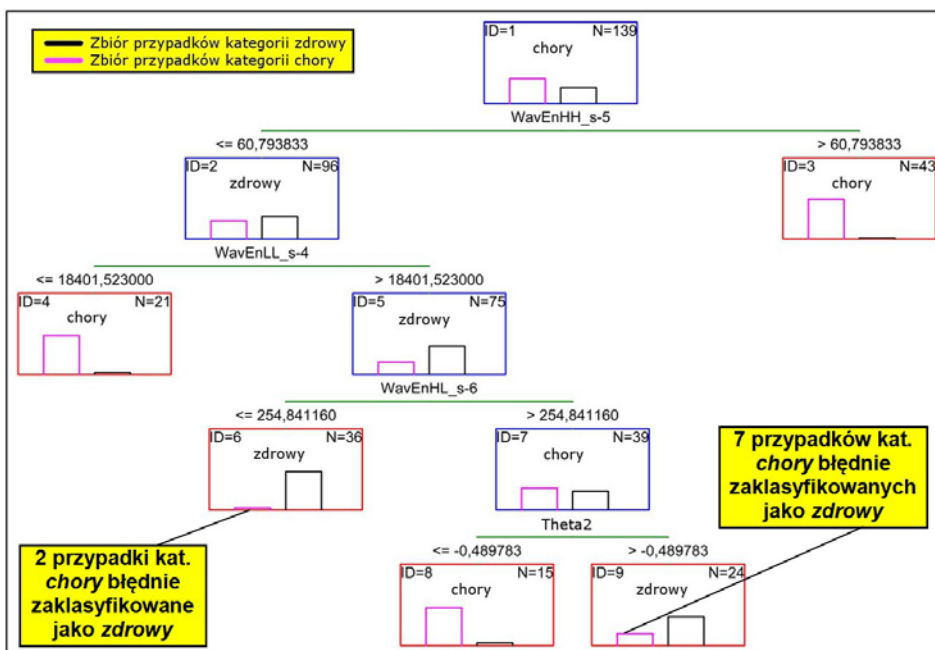
Wybrany model optymalny – drzewo decyzji C&RT nr 4, klasyfikuje przypadki pacjentów, u których rozpoznano chorobę Hashimoto, z dokładnością 89,41%, a błędnie przyporządkowuje do tej kategorii 5,56% przypadków zdrowych. Dwukrotnie większy błąd klasyfikacji fałszywie negatywnej od fałszywie pozytywnej nie jest korzystny z medycznego punktu widzenia. Należy jednak zaznaczyć, że niedogodność tę w dalszych badaniach będzie można zniwelować poprzez zróżnicowanie kosztów błędnej klasyfikacji na etapie budowy modelu. Jeżeli jednak uwzględnimy fakt, że błąd klasyfikacji poniżej 20% został osiągnięty na niezależnym zbiorze testowym (wyznaczonym metodą dziesięciokrotnej walidacji krzyżowej), to w odniesieniu do typowych standardów uczenia maszynowego, wynik ten można uznać za stwarzający możliwości wiarygodnego zastosowania użytej metodyki w planowanym systemie doradczym wspomagającym lekarza w diagnostyce tej choroby.

Analiza parametrów drzewa nasuwa także przypuszczenie, które podane zostanie weryfikacji w kolejnym etapie badań, że węzeł ID = 9 (liść drzewa), który zawiera stosunkowo dużą liczbę (7) błędnie zaklasyfikowanych przypadków kategorii *chory*, może zawierać bardzo interesujące, z medycznego punktu widzenia, przypadki występowania symptomów choroby we wczesnym jej stadium. Na uwagę zasługuje również fakt, że spośród 283 deskryptorów cech, szczególnie istotne, ze względu na cel klasyfikacji, okazały się te uzyskane przy pomocy dyskretnej transformaty falkowej (wśród 10-ciu najlepszych deskryptorów znalazło się aż 7 z nich).

5. WNIOSKI / PODSUMOWANIE

Przedmiotem badań były wyodrębnione fragmenty zdjęć USG tarczycy. Były to 54 próbki tarczycy zdrowej oraz 85 próbek z rozpoznąną chorobą Hashimoto. Zdjęcia poddano normalizacji obrazu oraz ekstrakcji cech dyskryminacyjnych, w wyniku czego wyznaczono zbiór 283 deskryptorów. Na tej podstawie, wykorzystując metodę indukcji drzew decyzji, przy pomocy algorytmu C&RT, zbudowano model klasyfikatora. Testy modelu wykonane na dostępnych próbkach zdjęć wykazały, że może on być z powodzeniem stosowany jako element systemu doradczego, wspomagającego lekarza w procesie diagnozy. Istotna dla prowadzenia dalszych badań, wydaje się grupa przypadków błędnie zaklasyfikowanych w węzle nr 9 drzewa decyzji. Może ona, zdaniem autorów, zawierać informacje interesujące z punktu widzenia roz-

poznawania wczesnego stadium rozwoju choroby Hashimoto. Celem dalszych badań będzie potwierdzenie tej tezy oraz porównanie wyników klasyfikacji uzyskanych przy pomocy algorytmu C&RT z modelami uzyskanymi innymi metodami uczenia maszynowego, a w tym z zastosowaniem sieci przekonań Bayes'a, sztucznych sieci neuronowych oraz innych algorytmów generowania drzew decyzyjnych.



Rys. 5. Optymalne drzewo decyzyjne nr 4 (liczba podziałów równa 4, liczba węzłów końcowych równa 5)

LITERATURA

1. Breiman L., Friedman J.H., Olshen R.A., Stone C.J.: Classification and regression trees. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, 1984.
2. Butlewski T., Franek E.: Diagnostyka obrazowa chorób tarczycy. Choroby Serca i Naczyń, Nr 6, 37-41, 2009.
3. Castanho M.J.P., Hernandez F., De Ré A.M., Rautenberg S., Billis A.: Fuzzy expert system for predicting pathological stage of prostate cancer. Expert Systems with Applications, Vol. 40, 466-470 (w druku).
4. Cichosz P.: Systemy uczące się. WNT, Warszawa 2000.
5. Fauci A.S., Braunwald E., Isselbacher K.J., Wilson J.D., Martin J.B., Kasper D.L., Hauser S.L., Longo D.L.: Interna Harrisona. Tom III. Wydawnictwo Czelej, Lublin 2009.

6. Grzymała-Busse J.W., Hippe Z.S.: Data Mining Methods Supporting Diagnosis of Melanoma. W: Proc. 18th IEEE Symposium on Computer-Based Medical Systems, IEEE Comp. Soc., Los Alamitos (California, USA), 371-373, 2005.
7. Haralick R.: Statistical and Structural Approaches to Texture. Proceedings of the IEEE, Vol. 67, No 5, 786-804, 1979.
8. Haralick R., Shanmugam K., Dinstein I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics, Vol. 3, No 6, 610-621, 1973.
9. Hill T., Lewicki P.: STATISTICS: Methods and Applications. StatSoft, Tulsa 2007.
10. Hu Y., Dennis T.: Textured Image Segmentation by Context Enhanced Clustering. IEE Proc.-Visual Image and Signal Processing, Vol. 141, No 6, 413-421, 1994.
11. Kokot F. (red.): Choroby wewnętrzne. Podręcznik akademicki. Tom 2. Wydawnictwo Lekarskie PZWL, Warszawa 2006.
12. Lerski R., Straughan K., Shad L., Boyce D., Bluml S., Zuna I.: MR Image Texture Analysis – An Approach to Tissue Characterization. Magnetic Resonance Imaging, Vol. 11, 873-887, 1993.
13. Liao S.-H., Chu P.-H., Hsiao P.-Y.: Data mining techniques and applications – A decade review from 2000 to 2011. Expert Systems with Applications, Vol. 39, 11303-11311, 2012.
14. Ligęza A.: Logical Foundations for Rule-Based Systems. Uczelniane Wydawnictwo Naukowo-Dydaktyczne Akademii Górniczo-Hutniczej w Krakowie, Kraków 2005.
15. Sokołowska-Pituchowa J. (red.): Anatomia człowieka. Podręcznik dla studentów medycyny. Wyd. VII. Wydawnictwo Lekarskie PZWL, Warszawa 2006.
16. Wróbel Z., Koprowski R.: Praktyka przetwarzania obrazów z zadaniami w programie Matlab. Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008.
17. www.eletel.p.lodz.pl/programy/cost/progr_mazda.html.

Rękopis dostarczono dnia 19.07.2012 r.

METHOD FOR CLASSIFICATION
OF ULTRASOUND THYROID IMAGES
BY DECISION TREE INDUCTION

Zbigniew OMIOTEK, Andrzej BURDA
Waldemar WÓJCIK

ABSTRACT *The article presents a method for classification of ultrasound thyroid images. This method allows to classify the analyzed cases as sick and healthy. It also allows to separate fairly large group of incorrectly classified cases. According to the authors, this*

group may include characteristics of the early stage of Hashimoto's disease. Decision tree induction has been used to build a classification model. Test results showed that the proposed method can provide a starting point to build a support system in the process of medical diagnosis.

Keywords: *thyroid, Hashimoto's disease, image processing, texture classification, decision trees*

Mgr inż. Zbigniew OMIOTEK – absolwent Wydziału Elektroniki WAT w Warszawie. W latach 1995-2000 asystent w Instytucie Podstaw Elektroniki WAT. Od 2000 r. asystent w Katedrze Informatyki i Inżynierii Wiedzy WSZiA w Zamościu, a od 2011 r. doktorant na Wydziale Elektrotechniki i Informatyki Politechniki Lubelskiej. Zainteresowania naukowe skupiają się wokół metod analizy i rozpoznawania obrazów medycznych.



Mgr inż. Andrzej BURDA – absolwent Wydziału Maszyn Górniczych i Hutniczych AGH w Krakowie. Od 1999 r. asystent w Katedrze Informatyki i Inżynierii Wiedzy WSZiA w Zamościu, a od 2009 r. doktorant na Wydziale Elektrotechniki, Automatyki, Informatyki i Elektroniki AGH. Zainteresowania naukowe skupiają się wokół metod klasyfikacji obiektów.

Prof. dr hab. inż. Waldemar WÓJCIK – ukończył studia na Politechnice Wrocławskiej – Wydział Elektroniki (1975), doktorat obronił na Politechnice Lubelskiej (1985), a habilitację na Politechnice Lwowskiej (2001). W 2009 otrzymał tytuł Profesora nauk technicznych. W swoim dorobku ma 14 monografii, ponad 100 prac naukowych oraz 3 patenty. Jest współautorem około 20 opracowań dla przemysłu dotyczących światłowodowych sieci kontrolno-pomiarowych i teleinformatycznych.

