Melanoma Diagnosis and Classification Web Center System: The Non-invasive Diagnosis Support Subsystem

Wiesław Paja and Mariusz Wrzesień

Institute of Biomedical Informatics,
University of Information Technology and Management in Rzeszów, Poland
{wpaja,mwrzesien}@wsiz.rzeszow.pl

Abstract. In this paper, computer-aided diagnosing and classification of melanoid skin lesions is briefly described. The main goal of our research was to elaborate and to present new version of the developed melanoma diagnosis support system, available on the Internet. It is a subsystem of our complementary melanoma diagnosis and classification web center system. Here, we present functionality, structure and operation of this subsystem. In its current version, five learning models are implemented to provide five independent results of diagnosis. Then, a specific voting algorithm is applied to select the correct class (concept) of the diagnosed skin lesion. Developed tool enables users to make early, non-invasive diagnosing of melanocytic lesions. It is possible using built-in set of instructions that animate diagnosis of four basic lesions types: *benign nevus*, *blue nevus*, *suspicious nevus* and *melanoma malignant*.

Keywords: diagnosis support system, machine learning, learning model, computer aided diagnosis system, teledermatology, Total Dermatoscopy Score, ABCD formula.

1 Introduction

Melanoma is the most deadly form of skin cancer. The World Health Organization estimates that more than 65000 people a year worldwide die from too much sun, mostly from malignant skin cancer [1]. It is an increasingly common tumour, it is the cutaneous tumour with the worst prognosis and its incidence is growing, because most melanomas arise on areas of skin that can be easily examined. Early detection and successful treatment often is possible, most dermatologists can accurately diagnose melanoma in about 80% of cases according to ABCD process [2]. Meanwhile the incorporation of dermatoscopic techniques, reflectance confocal microscopy and multiespectral digital dermatoscopy have greatly enhanced the diagnosis of this cutaneous melanoma. While these devices and techniques cannot diagnose skin cancer, they give dermatologists a closer look at suspicious skin lesions. This, in turn, can help dermatologists find suspicious lesions earlier than before and better determine whether a biopsy is needed. None of these devices can confirm that a suspicious lesion is melanoma. It is,

P. Perner (Ed.): ICDM 2011, LNAI 6870, pp. 96–105, 2011.

however, not yet possible to tell if a patient has melanoma or any type of skin cancer without a biopsy. It is important to combine the classically ABCDs and biopsy to prevention and diagnosis of melanoma.

The five-year survival rate for people whose melanoma is detected and treated before it spreads to the lymph nodes is 99 percent. Five-year survival rates for regional and distant stage melanomas are 65% and 15%, respectively [3]. Thus the curability of this type of skin cancer depends essentially on its early diagnosis and excision. For that reason the ABCD (*asymmetry*, *border*, *color* and *diversity of structure*) clinical rule is commonly used by dermatologists in visual examination and detection of early melanoma [4]. The visual recognition by clinical inspection of the lesions by dermatologists is 75%. Experienced ones with specific training can reach a recognition rate of 80% ([5], [6]).

Recently, some decrease of the illness was observed especially Australia, Scotland and Ireland [7]. Some reasons of this phenomenon can be guessed: (*i*) dissemination of methods for early, non-invasive diagnosing of health risk degree, what creates possibility of self-diagnosing for society of Western Europe and United States; (*ii*) fast access to vast hummer of information sources about symptoms of melanoma malignant, access to the methods of calculating of parameters characterizing health risk degree (based on atypical pigment lesions on the skin, frequency of contacts with solar or ultraviolet radiation, colour of eyes or hair, etc.), or/and (*iii*) access to various methods of calculating chances to survive years by given number of a the patient with diagnosed melanoma [8].

Results of European research in the field discussed have been usually focused on methodology of classification of tumour types, description of selected symptoms and description of pigment lesions, in a phase preceding incurable condition of illness or demanding surgical intervention ([5], [9]).

Our current research in the classification of medical images of skin lesions presents developed internet-based system for diagnosing of four categories of melanoma: *benign nevus*, *blue nevus*, *suspicious nevus*, and *melanoma malignant* [10]. Our system supports five different methods (learning models) of diagnosing: ***(i) classic ABCD rule*** (based on TDS parameter) ([11], [12]), ***(ii) optimized ABCD rule***, (based on our own New TDS parameter [13]), ***(iii) decision tree*** (based on ID3 algorithm) [14], ***(iv) genetic dichotomization***, based on a linear learning machine with genetic searching for the most important attributes [15], and ***(v)*** application of a new classifier from the family of ***belief networks***. Based on these five partial results, system suggests the final result, using the specific evaluation and voting algorithm.

2 Structure and Operation of the System

Our diagnosing support system provides user interface in the form of a website to get the access to its three main working modules (Fig. 1). The first module is dedicated to persons without medical background, and serves to self-diagnosing. This module allows to determine - in a very simple and clear way - all symptoms required for correct classification of a given skin lesion (Fig. 2). Thus, using

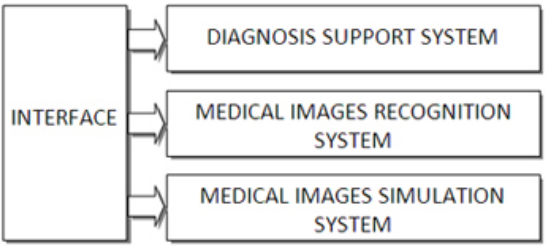

Fig. 1. The main structure of the system

this module, user can easily acquainted with the knowledge, required for correct recognition of symptoms, related to a given lesion. Next, this module can be treated as an advanced calculator for non-invasive diagnosing of melanocytic lesions. Input values for this module create a vector containing values of 13 descriptive attributes: *asymmetry*, *border*, six *colors* (*white*, *blue*, *black*, *red*, *dark brown*, *light brown*) and five different *structures* (*pigment globules*, *pigment dots*, *structureless areas*, *branched streaks*, *pigment network*). These values, provided by the user, are used to calculate the 14-th attribute known in medicine as *TDS* parameter [7] (*Total Dermatoscopy Score*) and additionally also a *NewTDS* [8]

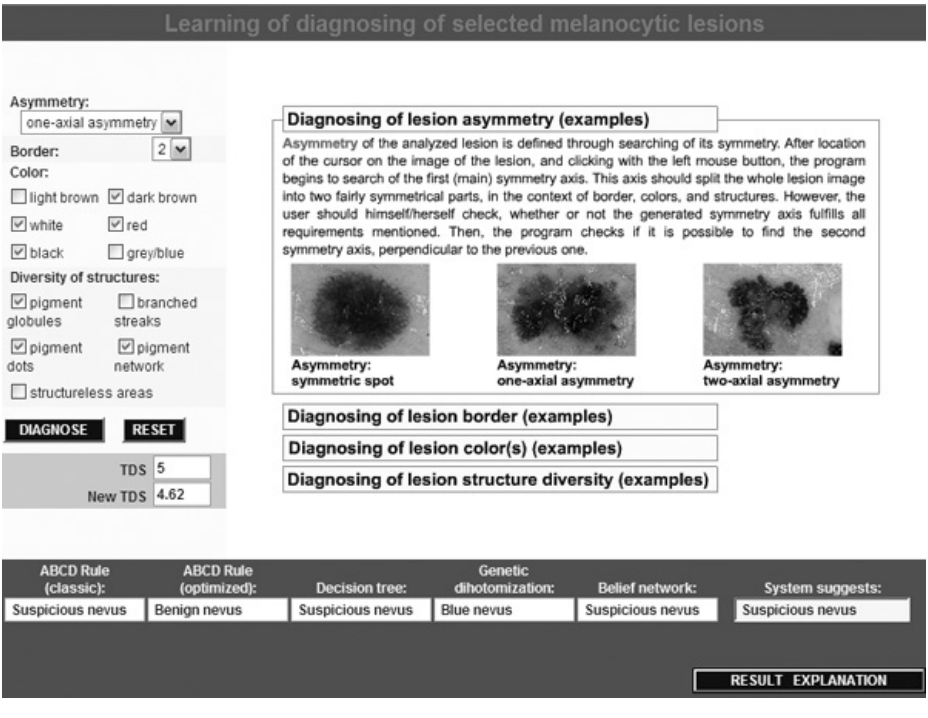

Fig. 2. Graphical interface of the diagnosis support subsystem

(Fig. 3). Then, five different algorithms previously mentioned, are responsible for development of five learning models (five partial classifiers). The classification process based on these models is described in Section 3.

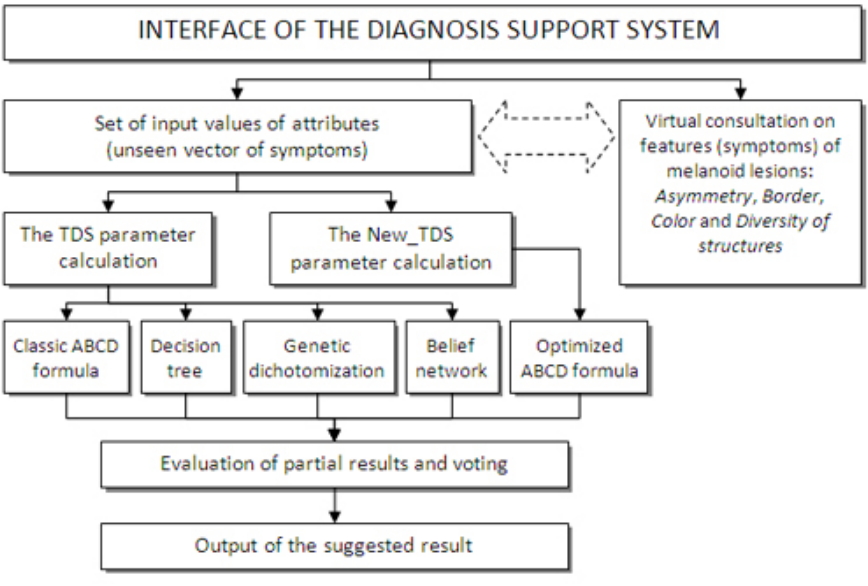

Fig. 3. Structure of the diagnosis support subsystem

The second module (Fig. 1), is based on automatic analysis and recognition of medical images. This approach consists of a system solution designed to analyze photographs of the patient's injury by means of image processing techniques where the dermatologists will capture the image of a melanoma using a digital dermatoscope, and a set of algorithms will process the image and provide an output diagnosis in an automated manner. The first results gathered along this line were presented in other articles [16].

In turn, the third module enables to generate the exhaustive number of simulated images, which considerably broaden the informational source database, and can be successfully used in the process of training less experienced medical doctors. It contains algorithms of semantic conversion of textual description of melanocytic lesion into respective image of the lesion. Detailed description of this approach could be found in earlier publications [17]. This module is currently developed as a component inside research project of polish Ministry of Science and Higher Education.

3 Recognition Algorithms and Classification Process

During our previous studies, a wide range of different learning algorithms were evaluated and tested. Thus, five different learning models were developed. Next,

these models were implemented in form of a web services and embedded in our system, inside first module. The details of each models are presented bellow.

3.1 Learning Model Based on a Classic and Optimized ABCD Rule

Logical values of symptoms, provided by user in the first module, are processed using two different algorithms: *(i)* Calculation of TDS, and *(ii)* Calculation of NewTDS. It is worth to say, that both algorithms are based on a constructive induction mechanism [18], a very important methodology in machine learning. Then, the enlarged solution space (13+1 dimensions) is defined using the classic ABCD formula for calculation TDS parameter (see Equation 1),

$$\begin{aligned}\mathbf{TDS} = (1.3 * \mathbf{A}symmetry) + (0.1 * \mathbf{B}order) + (0.5 * \Sigma\mathbf{C}olors) + \\ +(0.5 * \Sigma\mathbf{D}iversity)\end{aligned} \tag{1}$$

where **A** is a description of lesion's *asymmetry*, **B** is a description of lesion's *border*, **C** is a description of *colors* that occur in investigated lesion, and **D** is a specification of lesion's *diversity of structure*. The variable *Asymmetry* has three different values: *symmetric spot* (counted as *0*), *one-axial asymmetry* (counted as *1*), *two-axial asymmetry* (counted as *2*). *Border* is a numerical attribute, with values from *0* to *8*. A lesion is partitioned into eight segments. The border of each segment is evaluated: the sharp border contributes *1* to *Border*, the gradual border contributes *0*. The total amount of border values should be between *0* and *8*. *Color* has six possible values: *black*, *blue*, *dark brown*, *light brown*, *red* and *white*. Similarly, *Diversity* has five values: *pigment dots*, *pigment globules*, *pigment network*, *structure-less areas* and *branched streaks*. In our data set *Color* and *Diversity* were replaced by binary single-valued variables: *present* (value is equal to *1*) or *absent* (value is equal to *0*), for example, the pigment dots structure is absent, the black color is present, etc. In this way, our dataset contains objects described by 13 descriptive attributes. Simultaneously optimized formula was used to calculate the **NewTDS** (see Equation 2)

$$\begin{aligned}\mathbf{NewTDS} = (0.8 * Asymmetry) + (0.11 * Border) + \\ +(0.5 * ColorWhite) + (0.8 * ColorBlue) + (0.5 * ColorDarkBrown) + \\ +(0.6 * ColorLightBrown) + (0.5 * ColorBlack) + (0.5 * ColorRed) + \\ +(0.5 * PigmentNetwork) + (0.5 * PigmentDots) + \\ +(0.6 * PigmentGlobules) + (0.6 * BranchedStreaks) + \\ +(0.6 * StructurelessAreas)\end{aligned} \tag{2}$$

Learning model, developed using standard **TDS** parameter, classified unseen objects with average error rate equal 11%, however learning model, developed using optimized **NewTDS** parameter, and classified the same set of unseen objects with average error rate about 5%.

3.2 Learning Model in Form of Decision Tree

The third way to diagnose lesions is by using a decision tree (Fig. 4). This model was developed using the source data set presented earlier. In the process

of developing the decision tree the **ID3/C4.5** algorithm was used. It was stated that developed decision tree classified new, unseen melanoma cases with error rate equal exactly 1.4%. The developed tree is shown below:

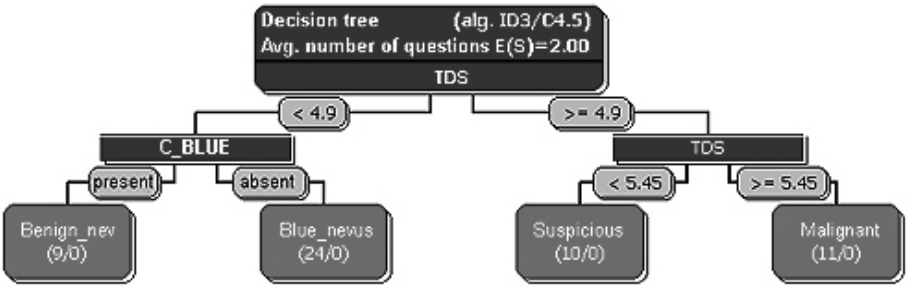

Fig. 4. Learning model in form of decision tree

3.3 Learning Model Based on the Genetic Dichotomization

This learning model contains **n(n-1)/2** number of vectors of diagnosed melanocytic lesions (where generally **n** is the number of identified concepts, in our case **n=4**), capable to classify correctly four classes of melanoid lesions. These vectors were developed in learning process, outside of described system. These vector are able to correct classifications of lesions that always belongs to two classes. Next, vectors are crossed to increase their classification quality. In our research for four classes learning model contains six described dichotomous vectors. Recognition process of unseen cases is executed automatically (see Table 1): system assigns to unseen case a category, pointed out by the maximal number of vectors. Classification process of unseen cases, is related to assigning to category which was indicated by the biggest number of vectors. Implemented genetic dichotomization learning model has optimal control parameters [19], which make possible to obtain average error rate equal to 6%.

Table 1. Illustration of an example recognition process, realized by the genetic dichotomization model

Vector no.	Capable to recognize unseen case: (Melanoma malignant)	Class assigned to example	Final decision
1	Benign nevus or Blue nevus	Benign nevus or Blue nevus	
2	Benign nevus or Malignant	Malignant	
3	Benign nevus or Suspicious	Benign nevus or Suspicious	Melanoma
4	Blue nevus or Malignant	Malignant	Malignant
5	Blue nevus or Suspicious	Blue nevus or Suspicious	
6	Malignant or Suspicious	Malignant	

3.4 Learning Model in Form of Belief Network

Bayesian classification machine describe interaction between nodes, that allow to develop learning model in form of belief network presented on 5. This network has average error rate equal to 4%. The most important attributes that directly impact on decision were: *pigment network* (D_PIGM_NETW), classic *TDS* (TDS), *asymmetry* (ASYMMETRY) and *color blue* (C_BLUE). Classification process is based on determining of all attributes, network nodes and achieving of probability of decision categories. Unseen case is assigned to a category, which displays the highest value of marginal likelihood.

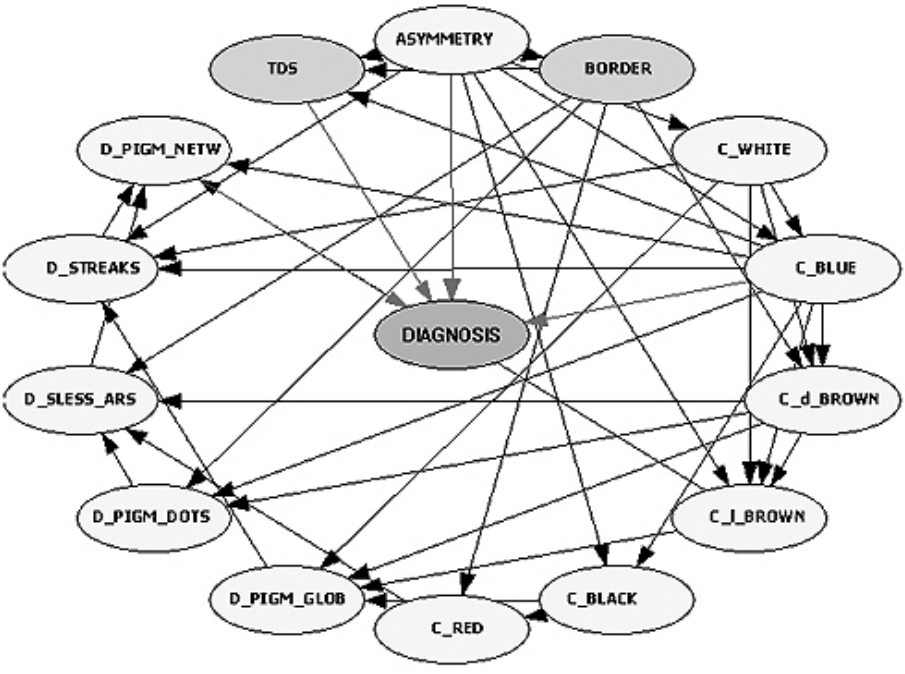

Fig. 5. Learning model in form of belief network

3.5 Algorithm for Optimal Diagnosis Selection

Presented system suggests five independent diagnosis gathered from five learning models: classic ABCD rule, optimized ABCD rule, decision tree, genetic dichotomization and belief network. Achieved results are input data into optimized diagnosis selection block (see Figure 3). Each result has own weight parameter dependent on error rate assigned to given learning model (see Table 2). These weight parameters are defined in Equation 3

$$W = (100\% - ErrorRateOfTheModel)/100\% \tag{3}$$

Table 2. Weight parameters for each learning model

Learning model	Weight parameter
Classic ABCD formula	W1=(100%-11%)/100%=0.89
Optimized ABCD formula	W2=(100%-5%)/100%=0.95
Decision tree	W3=(100%-1.4%)/100%=0.986
Genetic dichotomy process	W4=(100%-6%)/100%=0.94
Belief network	W5=(100%-4%)/100%=0.96

The final result is prepared depending on total amount of weight parameters for suggested diagnosis. It should be stressed that learning models with lower error rate have greater influence on final result.

On the following case (see Table 3), two from five learning models generate *Benign nevus* result, two others generate *Blue nevus* result, and the last learning model generates *Suspicious nevus*. Added weight parameters show that the most credible result is *Blue nevus* which has the greatest total weight parameter equal to **1.946**. Thus, system suggests the *Blue nevus* as final diagnosis.

Table 3. Calculation of the weight parameters

Diagnosis	Benign nevus	Blue nevus	Suspicious nevus	Melanoma malignant
Classic ABCD formula	0.89	0	0	0
Optimized ABCD formula	0.95	0	0	0
Decision tree	0	0.986	0	0
Genetic dichotomy process	0	0	0.94	0
Belief network	0	0.96	0	0
Total weight parameters	1.84	1.946	0.94	0

4 Conclusions and Future Remarks

Correct classification of pigment skin lesions is possible using histopatological research of lesion. The newest trend of diagnosing devoted to using non-invasive methods, has become cause of disseminating of information technology tools supporting this process.

In this paper, practical development of a new internet information system for classification of melanocytic lesions, are briefly described. This system has also some teaching functions, improves analyzing of datasets based on calculating of values of *Total Dermatoscopy Score* parameter. Inside this system, a five different methods were applied to determine correct diagnosis of skin lesions. As it was stated, each method is characterized by different error rate. It was indispensable to take its influence on final diagnose into consideration. Developed internet-based tool enables users to make early, non-invasive diagnosis of melanocytic lesions. The latest version of our system is available in the Internet: www.melanoma.pl.

Future research focuses on overall implementation of all three functional subsystems described in Section 2. It should be also stressed that in the future development of mobile version of our system is planned.

Acknowledgments. This research has been supported by the grant No. N N516 482640 from the National Research Center in Cracow, Poland.

References

1. Lucas, R., McMichael, T., Smith, W., Armstrong, B.: Solar ultraviolet radiation. In: Global Burden of Disease from Solar Ultraviolet Radiation. Environmental Burden of Disease Series, vol. 13. World Health Organization, Genewa (2006)
2. Rigel, D.S., Russak, J., Friedman, R.: The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs. A Cancer Journal for Clinicians 60, 301–316 (2010)
3. American Cancer Society. Cancer Facts & Figures 2009. American Cancer Society, Atlanta (2009)
4. Friedman, R.J.: Early detection of malignant melanoma: the role of the physician examination and self examination of the skin. Cancer Journal for Clinicians 35, 130–151 (1985)
5. Schmid-Saugeon, P., Guillod, J., Thiran, J.P.: Towards a computer-aided diagnosis system for pigmented skin lesions. Computerized Medical Imaging and Graphics 27, 65–78 (2003)
6. Blanzieri, E.: Exploiting classifier combination for early melanoma diagnosis support. In: Proc. of the 11th European Conference on Machine Learning, Barcelona, May 31-June 2, pp. 55–62 (2000)
7. Kirn, T.F.: Reasons Unclear for Worlwide Decline in Melanoma. Skin & Allergy News 31(5), 41–42 (2000)
8. `http://www.dermoncology.med.nyu.edu`
9. Taouil, K., Chtourou, Z., Ben Romdhane, N.: A robust system for melanoma diagnosis using heterogeneous image databases. Journal of Biomedical Science and Engineering 3, 576–583 (2010)
10. Grzymała-Busse, J.W., Hippe, Z.S., Knap, M., Paja, W.: Infoscience Technology: The Impact of Internet Accessible Melanoid Data on Health Issues. In: Smith, F.J. (ed.) Data Science Journal, vol. 4, pp. 77–81 (2005)
11. Braun-Falco, O., Stolz, W., Bilek, P., Merkle, T., Landthaler, M.: Das Dermatoskop. Eine Vereinfachung der Auflichtmikroskopie von pigmentierten Hautveranderungen. Hautartzt 40, 131–136 (1990)
12. Stolz, W., Harms, H., Aus, H.M., Abmayr, W., Braun-Falco, O.: Macroscopic diagnosis of melanocytic lesions using color and texture image analysis. J. Invest. Dermatol. 95, 491–497 (1990)
13. Alvarez, A., Bajcar, S., Brown, F.M., Grzymała-Busse, J.W., Hippe, Z.S.: Optimization of the ABCD Formula Used for Melanoma Diagnosis. In: Kłopotek, M.A., Wierzchoń, S.T. (eds.) Advances In Soft Computing (Intelligent Information Systems and Web Mining), pp. 233–240. Physica-Verlag, Heidelberg (2003)
14. Hippe, Z.S., Grzymała-Busse, J.W., Błajdo, P., Knap, M., Mroczek, T., Paja, W., Wrzesień, M.: Classification of Medical Images In the Domain of Melanoid Skin Lesions. In: Kurzyński, M., Puchała, E., Woźniak, M., Żołnierek, A. (eds.) Computer Recognition Systems. Advances in Soft Computing, pp. 519–526. Springer, Heidelberg (2005)

15. Hippe, Z.S., Wrzesień, M.: Some Problems of Uncertainty of Data after the Transfer from Multi-category to Dichotomous Problem Space. In: Burczyski, T., Cholewa, W., Moczulski, W. (eds.) Methods of Artificial Intelligence, pp. 185–189. Silesian University of Technology Edit. Office, Gliwice (2002)
16. Cudek, P., Grzymała-Busse, J.W., Hippe, Z.S.: Multistrategic Classification System of Melanocytic Skin Lesions: Architecture and First Results. In: Kurzyński, M., Woźniak, M. (eds.) Advances in Intelligent and Soft Computing. Computer Recognition Systems, vol. 3, pp. 381–387. Springer, Heidelberg (2009)
17. Hippe, Z.S., Grzymała-Busse, J.W., Piatek, Ł.: Synthesis of Medical Images in the Domain of Melanocytic Skin Lesions. In: Kłopotek, M.A., Wierzchoń, S.T., Trojanowski, K. (eds.) Advances in Soft Computing (Information Technologies in Biomedicine), pp. 225–231. Springer, Heidelberg (2008)
18. Paja, W., Wrzesień, M.: Medical datasets analysis: A constructive induction approach. In: Perner, P. (ed.) ICDM 2010. LNCS (LNAI), vol. 6171, pp. 442–449. Springer, Heidelberg (2010)
19. Hippe, Z.S., Wrzesień, M.: Some problems of uncertainty of data after the tran-sfer from multi-category to dichotomaous problem space. In: Burczyński, T., Cholewa, W., Moczulski, W. (eds.) Methods of Artificial Intelligence, pp. 185–189. Silesian University of Technology Edit. Office, Gliwice (2002)