Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 281–286

DOI: 10.15439/2015F361
ACSIS, Vol. 5

Medical diagnosis support and accuracy improvement by application of total scoring from feature selection approach

Wiesław Paja
Faculty of Mathematics and Natural Sciences, University of Rzeszów,
1 Prof. S. Pigonia Street, 35-310 Rzeszów, Poland
Email: wpaja@ur.edu.pl

Abstract— **Melanoma is the most deadly form of skin cancer. Early detection and successful treatment of this disease often is possible. The main goal of this paper is to present results of application of feature selection method to find the most important or all important features that characterize melanocytic spots on the skin and in this way defining of a new Total Dermatoscopy Score formula. Thus, it is possible to decrease dimensionality of that problem. Results gathered during research focus on about six from thirteen descriptive attributes which are the most relevant and are stated as core attributes. Based on these attributes a simple total scoring method could be applied to improve prediction (diagnosis) results, additionally also reducing complexity of problem. Results were acquired by application of six different machine learning algorithms and estimated using several evaluation measures.**

I. 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 well-known ABCD process [2]. ABCD formula devotes to Asymmetry of lesions, their Border, Color and Diversity of structures (or Diameter in other approach), and in some cases the Evolving over time (the ABCDE formula). Based on these features dermatologists could prepare diagnosis by simply observation of investigated lesion.

Meanwhile the incorporation of dermatoscopic techniques, reflectance confocal microscopy and multispectral digital dermatoscopy have greatly enhanced the diagnosis of this cutaneous melanoma. While these devices and techniques could 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, 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 percent and 15 percent, 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. It is also used in development of diagnosis platforms such as DERMA [4] or IMDLS systems [5].

Previous research [5-8] focused on using of data mining and image mining techniques to provide early support to diagnosis of melanocytic lesions. Now, it is proposed to apply feature selection methods to find interesting features inside investigated melanocytic datasets. Thus, we could try to recognize the minimal set of important (relevant) features, but on the other hand we can calculate the importance of each feature used in ABCD formula in the domain of melanoma classification. According to Kohavi and John [9] feature X could be defined to be strongly relevant when removal of X alone from the data always results in deterioration of the prediction accuracy of the ideal Bayes classifier. Feature X is weakly relevant if it is not strongly relevant and there exists a subset of features S, such that the performance of ideal Bayes classifier on S is worse than the performance on $S \cup \{X\}$. A feature is irrelevant if it is neither strongly nor weakly relevant. Improving the performance of machine learning classifiers for diagnosis based on feature selection is often applied [10,11]. In this paper additional application of FS methods is investigated.

II. Dataset used during experiments

The medical dataset which was used in this research concerns melanocytic skin lesions that are a very serious skin and lethal cancer. It is a disease of contemporary time,

This work was partially supported by The Fund of Dean of the Faculty of Mathematics and Natural Sciences, University of Rzeszów

the number of melanoma cases is constantly increasing, due to, among other factors, sun exposure and a thinning layer of ozone over the Earth. Statistical details on this data are given in [12]. Investigated data consist of 326 case of *Benign nevus* and *Blue nevus*, 108 cases of *Suspicious nevus* and 114 cases of *Melanoma malignant*, a total of 548 cases. Descriptive attributes of the data were divided into four categories:

- *Asymmetry*, has three different values: *symmetric spot, one-axial asymmetry* and *two-axial asymmetry,*
- *Border*, is a numerical attribute with values from *0* to *8*,
- *Color* group, has six possible types: *Black, Blue, Dark brown, Light brown, Red* and *White*,
- *Diversity of structures* group, has five possible types: *Pigment dots, Pigment globules, Pigment network, Structureless areas* and *Branched streaks*;

Each of these 11 types of Color and Diversity have values 0 or 1, that is 0 means lack of the corresponding property and 1 means the occurrence of the property. In dermatology this set of features is known as ABCD formula and is also applied to calculated the so-called *Total Dermatoscopy Score (TDS)* [13,14]. The ABCD formula of dermoscopy was the first dermoscopy algorithm created to help differentiate benign from malignant tumors [14]. This algorithm was developed to quantitatively address the crucial question in dermoscopy of whether a melanocytic skin lesion under investigation is benign, suspicious (borderline), or malignant. Based only on four dermatoscopic criteria this method is relatively easy to learn and to apply. The ABCD method has been extensively studied and it has been shown that it improves the diagnostic performance of clinicians evaluating pigmented skin lesions.

The goal was to use selected machine learning methods to estimate hierarchy of importance of melanocytic symptoms. These symptoms are part of well-known parameter *TDS* (*Total Dermatoscopy Score*) that is a useful diagnostic tool for melanoma. The *TDS* is computed using the following formula (known as the ABCD formula):

$$TDS = 1.3 * Asymmetry + 0.1 * Border + \\ + 0.5 * \sum Colors + 0.5 * \sum Diversity \quad (1)$$

where A is a description of lesion's asymmetry, B is a description of lesion's border, C is a description of colors appearing in considered lesion, and D is a specification of lesion's diversity.

III. Methods of Experiments

During research a following general procedure was applied:

1. Selection of dataset and features for investigation

(a) Application of set of ranking measures to calculate rank of importance for each feature

(i)With set of contrast features

(ii)Without contrast features

(b) Definition (selection) of the most important feature subset

2. TDS calculation for all original features

3. New TDS calculation based on selected most important features

4. Application of different machine learning algorithms for classification of unseen objects using 10-fold cross validation method

(a) Using all descriptive features

(b) Using only selected, most important features

(c) Using all descriptive features with TDS added

(d) Using only selected, most important features with NewTDS added

5. Comparison of gathered results using different evaluation measures

In the first step, dataset and features for investigation were defined. Then, different ranking measures were applied to estimate importance of each feature. In order to check specificity of the feature selection, the dataset was extended by contrast variables. It means that each original variable was duplicated and it's values were randomly permuted between all objects. Hence a set of non-informative by design shadow variables was added to original variables. The number times when the shadow variables were selected as important gives estimate of the expected level of false discovery. These variables that were selected as important significantly more often than random, were examined further, using different test. To define level of feature importance six well-known ranking measures were applied: *ReliefF*, *Information Gain*, *Gain Ratio*, *Gini Index*, *SVM weight* and *RandomForest*. Additionally, a new parameter, called *RuleQualityFS* (see Table 1), were introduced. It is based on frequency of presence of different feature in rule model generated from dataset and also takes into consideration quality of the rules in which there is. Rank quality of i^{th} attribute could be presented as follow:

$$Q_{A_i} = \sum_{j=1}^{n} Q_{R_j} \{A_i\} \quad (2)$$

where n is a number of rules inside the model, Q_{Rj} defines classification quality of rule R_j and $\{A_i\}$ describe the presence of i^{th} attribute, usually *0* or *1*.

In turn, quality of rule is defined as follow:

$$Q_{R_j} = \frac{E_{corr}}{E_{corr} + E_{incorr}} \quad (3)$$

where E_{corr} depicts number of correctly matched learning examples by j^{th} rule and E_{incorr} depicts number of incorrectly matched learning examples by this rule.

In the second step, the standard *TDS* calculation were performed based on original values of attributes and using formula (1). It is standard procedure utilized by medical specialists.

However, in my research, the third step is crucial. In this point a *NewTDS* value is defined and calculated (see formula

4). According to acquired factors from the first step of experiments (Table 1), a new formula for *TDS* calculation were introduced:

$$\begin{aligned} NewTDS = {} & 38.72 * Border + \\ & + 31.82 * Asymmetry + \\ & + 23.22 * PigmentNetwork + \\ & + 20.00 * BlueColor + \\ & + 16.60 * BranchedStreaks + \\ & + 15.50 * WhiteColor \end{aligned} \tag{4}$$

Six selected attributes were used according to Table 1. For each of them corresponding factor from this table (*RuleQualityFS* column) were inserted. In this way, new attribute which connects others into one value were added to original dataset.

During the fourth step test probing the importance of variables was performed by analyzing the influence of variables used for model building on the prediction quality. Four different combination of attributes were applied.

Six different machine learning model were applied to build different predictors: *Classification Tree (CT), Random Forest (RF), CN2 decision rules algorithm (CN2), Naïve Bayes (NB), k Nearest Neighbors (kNN)* and Support *Vector Machine (SVM)*. During this step a 10-fold cross validation paradigm were used. Ten known evaluation measures were utilized in each predictor: *Classification Accuracy (CA), Sensitivity, Specificity, Area Under ROC curve (AUC), Information Score (IS), F1 score (F1), Precision, Brier measure, Matthew Coefficient Correlation (MCC)* parameter and finally *Informadness ratio* [11].

TABLE I.
RANKING OF FEATURES USING SEVEN DIFFERENT MEASURES

Attribute	ReliefF	Inf. gain	Gain Ratio	Gini	SVM weight	RF	RuleQuality FS
Border	0.03	**0.17**	**0.09**	**0.03**	**4.93**	**3.74**	**38.72**
Asymmetry	**0.25**	**0.46**	**0.34**	**0.07**	**7.34**	**10.99**	**31.82**
Pigment network	**0.19**	**0.18**	**0.18**	**0.02**	**1.90**	**3.82**	**23.22**
Blue color	**0.16**	**0.41**	**0.58**	**0.06**	**13.79**	**10.17**	**20.00**
Branched streaks	**0.13**	**0.23**	**0.23**	**0.02**	**2.22**	**3.51**	**16.60**
White color	0.03	**0.06**	**0.07**	**0.01**	**1.64**	**1.12**	**15.50**
Border (contrast)	-0.06	0.01	0.01	0.00	0.08	-0.12	12.50
Black color	-0.05	**0.11**	**0.11**	**0.01**	**2.35**	**2.02**	11.00
Light brown color	-0.02	**0.05**	**0.06**	**0.01**	**1.24**	**1.06**	11.00
Pigment dots	0.08	**0.09**	**0.10**	**0.01**	**1.26**	**1.08**	10.80
Asymmetry (contrast)	0.01	0.01	0.01	0.00	1.16	0.01	10.52
Structureless areas	0.00	**0.04**	**0.07**	**0.01**	**1.24**	**0.48**	9.00
Red color	0.02	**0.08**	**0.08**	**0.01**	1.13	**1.58**	6.50
Black color (contrast)	0.01	0.00	0.00	0.00	0.08	-0.01	5.80
Pigment network (contrast)	-0.08	0.00	0.00	0.00	0.03	-0.05	5.60
Light brown color (contrast)	0.00	0.00	0.00	0.00	0.04	-0.06	5.50
White color (contrast)	0.00	0.00	0.00	0.00	0.01	0.24	5.00
Dark brown color	0.05	**0.06**	**0.07**	**0.01**	0.97	**0.90**	4.80
Pigment dots (contrast)	-0.06	0.00	0.00	0.00	0.04	-0.12	4.80
Branched streaks (contrast)	0.08	0.00	0.00	0.00	0.11	-0.03	4.80
Dark brown color (contrast)	0.03	0.01	0.01	0.00	0.08	0.18	4.00
Blue color (contrast)	0.03	0.00	0.00	0.00	0.01	0.04	3.00
Red color (contrast)	0.05	0.00	0.00	0.00	0.03	-0.08	3.00
Pigment globules	0.02	**0.05**	**0.09**	**0.01**	**1.26**	**0.73**	3.00
Pigment globules (contrast)	-0.02	0.00	0.00	0.00	0.42	0.02	3.00
Structureless areas (contrast)	-0.01	0.00	0.01	0.00	0.14	-0.08	3.00

IV. RESULTS OF EXPERIMENTS

The first experiment revealed six variables, called *core features*, that were indicated as important by all, or nearly all, ranking measures, see Table 1. In this table, we can observe that *Border, Asymmetry, Pigment network*, *Blue color*, *Branched streaks* and *White color* features create stable and core set of features which have the highest values of seven measures of importance, particularly using *RuleQualityFS* measure, introduced in this investigation. In the same table, comparison with importance of contrast values (grey rows colored and *contrast* index) is also presented. The most important contrast feature is *Border (contrast)* for which *RuleQualityFS* measure, defined in earlier section, is equal to *12.50*. In this way, he is also treated as a threshold that separates the *core set* of attributes from all contrast features and other less informative attributes. Most of the measures used in this approach focused that selected core set of features has higher values of these parameters than gathered threshold attribute value. These values are denoted in bold style in Table 1. Hereby,

TABLE II.
AVERAGE CLASSIFICATION RESULTS GATHERED USING DIFFERENT CLASSIFICATION QUALITY MEASURES APPLIED TO SIX MACHINE LEARNING MODELS FOR FOUR INVESTIGATED SETS OF FEATURES COMBINATION

Model	CA	Sens	Spec	AUC	IS	F1	Prec	Brier	MCC	Informadness
All original feature set										
CT	0.79	0.78	0.92	0.92	1.30	0.78	0.78	0.34	0.70	0.70
RF	0.83	0.79	0.93	0.97	1.11	0.80	0.85	0.27	0.75	0.72
CN2	0.82	0.79	0.93	0.94	1.32	0.81	0.84	0.27	0.75	0.72
NB	0.78	0.77	0.92	0.96	1.24	0.78	0.80	0.27	0.71	0.69
kNN	0.81	0.82	0.93	0.94	1.40	0.82	0.81	0.29	0.75	0.76
SVM	0.84	0.83	0.94	0.97	1.37	0.84	0.85	0.21	0.78	0.78
AVG	**0.81**	**0.80**	**0.93**	**0.95**	**1.29**	**0.80**	**0.82**	**0.28**	**0.74**	**0.73**
Selected core feature set										
CT	0.77	0.73	0.91	0.91	1.19	0.73	0.75	0.34	0.65	0.64
RF	0.75	0.69	0.90	0.94	0.99	0.68	0.72	0.33	0.61	0.58
CN2	0.76	0.70	0.90	0.92	1.09	0.72	0.78	0.34	0.64	0.60
NB	0.73	0.70	0.90	0.94	1.12	0.71	0.73	0.33	0.61	0.59
kNN	0.77	0.75	0.91	0.92	1.26	0.75	0.76	0.34	0.67	0.66
SVM	0.75	0.71	0.90	0.93	1.09	0.72	0.74	0.33	0.63	0.61
AVG	**0.75**	**0.71**	**0.90**	**0.93**	**1.12**	**0.72**	**0.75**	**0.33**	**0.64**	**0.61**
All original feature set with TDS parameter										
CT	1.00	1.00	1.00	1.00	1.85	1.00	1.00	0.00	1.00	1.00
RF	0.99	0.98	0.99	1.00	1.59	0.98	0.99	0.06	0.98	0.98
CN2	1.00	1.00	1.00	1.00	1.62	1.00	1.00	0.03	1.00	1.00
NB	0.92	0.90	0.97	0.99	1.53	0.90	0.91	0.13	0.88	0.87
kNN	0.86	0.87	0.95	0.96	1.51	0.86	0.86	0.21	0.81	0.82
SVM	0.94	0.92	0.98	1.00	1.64	0.93	0.93	0.08	0.91	0.90
AVG	**0.95**	**0.94**	**0.98**	**0.99**	**1.62**	**0.95**	**0.95**	**0.09**	**0.93**	**0.93**
Selected core feature set with NewTDS parameter										
CT	1.00	1.00	1.00	1.00	1.85	1.00	1.00	0.00	1.00	1.00
RF	1.00	1.00	1.00	1.00	1.61	1.00	1.00	0.04	1.00	1.00
CN2	1.00	1.00	1.00	1.00	1.59	1.00	1.00	0.04	1.00	1.00
NB	0.99	0.98	1.00	1.00	1.80	0.98	0.98	0.02	0.98	0.98
kNN	0.94	0.92	0.98	0.99	1.68	0.93	0.93	0.10	0.91	0.90
SVM	0.98	0.98	0.99	1.00	1.74	0.98	0.98	0.04	0.97	0.97
AVG	**0.98**	**0.98**	**1.00**	**1.00**	**1.71**	**0.98**	**0.98**	**0.04**	**0.98**	**0.98**

we can observe that different measures give different threshold, and it also shows that some other measures than *RuleQualityFS* include all original variables in core sets, e.g. *Information Gain, Gain Ratio, Gini index* and *Random Forest* application. Thus, we cannot extract smaller set of relevant attributes than the original one.

The second part of experiments focused on calculation of standard *TDS* and *NewTDS* defined earlier. Based on formula 1 and formula 4, these two values were obtained. Then, two datasets that include *TDS* and *NewTDS* respectively were investigated in next part of experiment.

method. Average results are collected in Table 2. Procedure were utilized to four specified sets:

(i) original set containing all descriptive features,

(ii) only selected core feature set based on its importance calculated in the first step,

(iii) original set containing all descriptive features with added standard TDS parameter,

(iv) core feature set with added NewTDS parameter.

Additionally, to compare results, average values (*AVG*) of all evaluation measures were calculated.

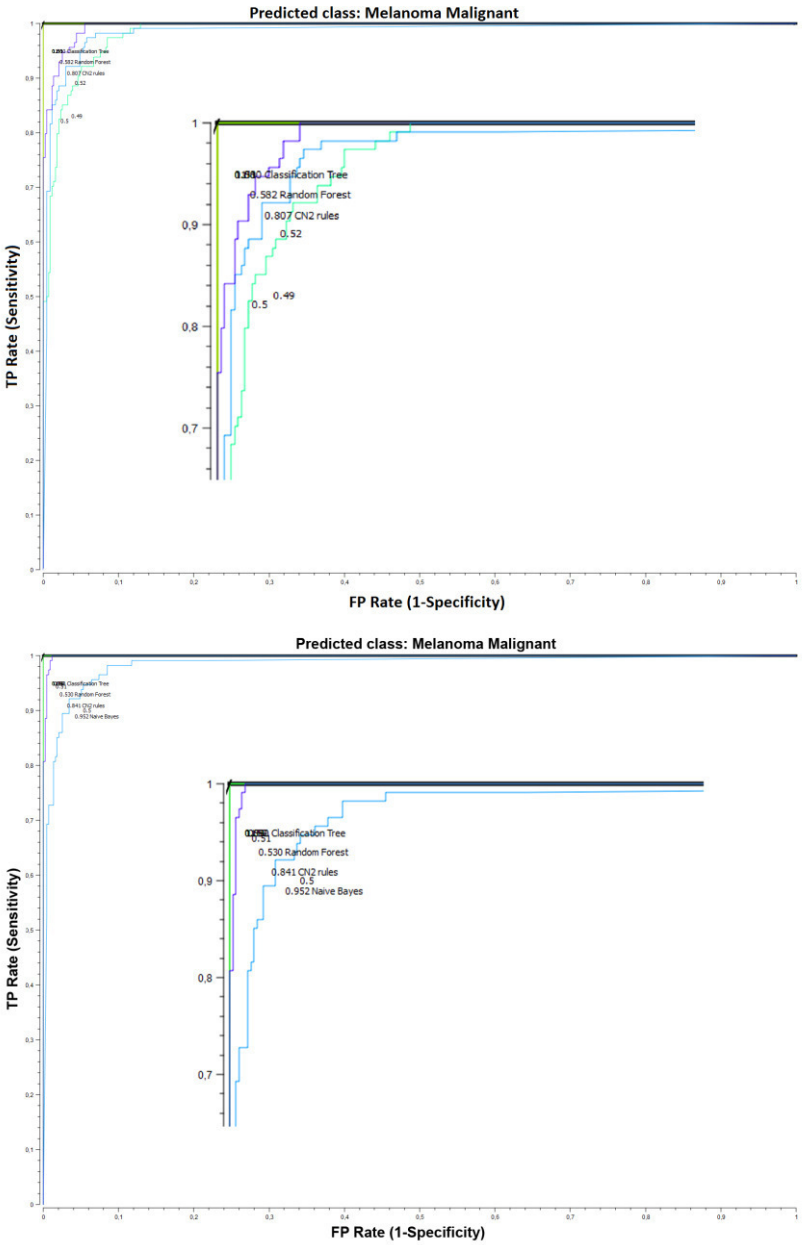

Fig. 1 Comparison of ROC curves gathered for Melanoma malignant class using six learning algorithms by investigation of original dataset (top chart) and selected core features with added *NewTDS* attribute (bottom chart)

The third part of experiment devoted to estimation of prediction quality of utilized machine learning algorithms described in section III. During this step six different algorithms were applied using 10-fold cross validation

Based on acquired results (see Table 2), it could be stressed that core set of features which contains only 6 from 13 attributes has very similar prediction quality as it was observed with all original 13 attributes. For instance, popular measure in data analysis AUC decreased on average only from 0.95 to 0.93. However, average Classification Accuracy decreased rather significantly from 0.81 to 0.75, and also Informadness, which in itself connects Sensitivity and Specificity, decrease on average from 0.73 to 0.61. Next, if we tried to add calculated standard TDS values it is observed that AUC reached better average value, 0.99. This outcome could be also observed in form of Receiver Operating Characteristic curve. Comparison of ROC curves for original and with NewTDS feature set generated only for Melanoma malignant class is presented on figure 1. In turn, all other measures also increased significantly. By adding TDS values to dataset Informadness measure increased on average from 0.73 to 0.93. Thus, it could be said that this approach could be positively applied in other, different medical issues.

The last step of experiment shows that the feature space could be probably significantly reduced. It means, that we can use only six from thirteen descriptive attributes in connection with new total score parameter could be successfully applied. In Table 2, the average value of all evaluation measure increased significantly reaching almost limits. For example AUC and Specificity reached 1.0, in turn Informadness achieve 0.98, what is very good result. According to this results it could be stressed that this methodology improves prediction of learning models and additionally simplifies space of problems by reducing its dimensionality.

References

[1] R. Lucas, A. McMichael, B. Armstrong, and W. Smith, "Estimating the global disease burden due to ultraviolet radiation exposure.," *Int. J. Epidemiol.*, vol. 37, pp. 6546–67, 2008.

[2] F. R. Rigel D.S., Russak J., "The evolution of melanoma diagnosis: 25 years beyond the ABCDs," *CA. Cancer J. Clin.*, vol. 60, no. 5, pp. 301–316, 2010.

[3] American Cancer Society, "Cancer Facts & Figures," *Cancer Facts Fig.*, 2014.

[4] R. Nicolas, A. Fornells, E. Golobardes, G. Corral, S. Puig, and J. Malvehy, "DERMA: A melanoma diagnosis platform based on collaborative multilabel analog reasoning," *Sci. World J.*, vol. 2014, 2014.

[5] J. W. Grzymala-Busse, Z. S. Hippe, M. Knap, and W. Paja, "Infoscience technology: the impact of internet accessible melanoid data on health issues," *Data Sci. J.*, vol. 4, pp. 77–81, 2005.

[6] P. Cudek, W. Paja, and M. Wrzesien, "Automatic System for Classification of Melanocytic Skin Lesions Based on Images Recognition," in *Man-Machine Interactions 2, Proceedings of the 2nd International Conference on Man-Machine Interactions, ICMMI 2011, The Beskids, Poland, October 6-9, 2011*, vol. 103, pp. 189–196.

[7] P. Cudek, W. Paja, and M. Wrzesien, "Image Recognition System for Diagnosis Support of Melanoma Skin Lesion," in *Security and Intelligent Information Systems - International Joint Conferences, SIIS 2011, Warsaw, Poland, June 13-14, 2011, Revised Selected Papers*, 2011, vol. 7053, pp. 217–225.

[8] W. Paja and M. Wrzesien, "Medical Datasets Analysis: A Constructive Induction Approach," in *Advances in Data Mining. Applications and Theoretical Aspects, 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14, 2010. Proceedings*, 2010, vol. 6171, pp. 442–449.

[9] R. Kohavi and R. Kohavi, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, 1997.

[10] A. Wosiak and D. Zakrzewska, "Feature Selection for Classification Incorporating Less Meaningful Attributes in Medical Diagnostics," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, 2014, vol. 2, pp. 235–240.

[11] N. Pérez, M. A. Guevara, A. Silva, I. Ramos, and J. Loureiro, "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, 2014, vol. 2, pp. 209–217.

[12] Z. S. Hippe, S. Bajcar, P. Blajdo, J. P. Grzymala-Busse, J. W. Grzymala-Busse, M. Knap, W. Paja, and M. Wrzesien, "Diagnosing Skin Melanoma: Current versus Future Directions," *TASK Q.*, vol. 7, no. 2, pp. 289–293, 2003.

[13] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions.," *J. Am. Acad. Dermatol.*, vol. 30, no. 4, pp. 551–559, 1994.

[14] U. Weigert, W. H. C. Burgdorf, and W. Stolz, "ABCD rule," in *An Atlas of Dermoscopy, Second Edition*, A. A. Marghoob, J. Malvehy, and R. P. Braun, Eds. CRC Press, 2012, pp. 113–117.