

Feature Selection Methods in Image-Based Screening for the Detection of Hashimoto's Thyroiditis in First-Contact Hospitals

Zbigniew Omiotek

Lublin University of Technology, Poland

Andrzej Burda

University of Management and Administration in Zamość, Poland

Abstract

In this paper, results of dimension reduction in feature space for thyroid ultrasound images using the heuristic identification of noisy variables, testing the significance of correlation coefficients and the method of Hellwig index of information capacity, have been compared. The best results were achieved using the Hellwig method. It enabled us to choose only 3 features from a large set of 283 discriminant ones. Classifiers built on the basis of this reduced set of features have the highest classification sensitivity (0,82) and the highest classification specificity (0,83) in comparison to other reduced datasets that we used in our research. Results showed that the Hellwig method can be used as an effective process for dimension reduction in feature space in classification of thyroid ultrasound images.¹

Keywords: feature selection, texture classification, HINoV, Hellwig method, Spearman correlation, Hashimoto's thyroiditis

Introduction

The screening of patients in regional hospitals that are first-contact, is very important in early detection of diseases, including Hashimoto's thyroiditis. A very popular method for diagnosis of the thyroid is ultrasonography because it is relatively cheap in comparison to other procedures of this type and it does not require any specialized devices. Our research aimed at defining the feature vector that would provide high accuracy identification of cases affected by the early stage of Hashimoto's thyroiditis basing on thyroid ultrasound images. It seems that an automatic classification system might be used to support the physician in the diagnosis process (Omiotek, Burda, and Wójcik 2013, 2015). The initial stage of data processing, before clustering or classification, is dimension reduction in feature space. This task can be done using various feature extraction and feature selection methods. This paper presents results obtained by the use of three selected feature selection methods. As a result of the image analysis, thyroid ultrasound images were described with a large set of 283 texture features. Full data set has been reduced to 57 features using the method of heuristic identification of noisy variables. The method of testing the significance of correlation coefficients between texture features and the decision variable made it possible to obtain 48 discriminant features. Finally, full data set was reduced to 3 features using the method of Hellwig index of information capacity. Reduced data sets were used to build and test classifiers. Analysed cases were assigned to one of two classes—sick or healthy. Classification results were used to compare the effectiveness of the feature selection methods.

1. [In the journal European practice of number notation is followed—for example, 36 333,33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style).—Ed.]

Ten different classifier building methods were used. These can be divided into 2 groups in terms of the form of the output model (Liao, Chu, and Hsiao 2012; Ligęza 2006). The first group consisted of the following 6 methods: linear and quadratic discriminant analysis (Fisher 1936), k -nearest neighbours (Enas and Choi 1986), artificial neural networks (Tadeusiewicz 1993), support vector machines (Koronacki and Ćwik 2005) and classification trees (Breiman et al. 1984). These methods made it possible to build classifiers in the form of a single model. Multi-model approach has been used for the second group of 4 methods. Their basis is combining V base models into a single aggregated D^* model. Bootstrap aggregation (Breiman 1996), models strengthening (Freund and Schapire 1997), models binding (Hothorn and Lausen 2005) and random forests (Breiman 2001) belonged to the second group. The multi-model approach is often used to improve the classification accuracy because the aggregated model is generally more accurate when compared to any of the component models.

In Section 1, the procedures for image processing and analysis are described. Feature selection methods and results obtained using these methods are presented in Section 2. Section 3 contains classification results used for the evaluation of feature selection methods. Finally, conclusions are included in Section 4.

1 Image processing and analysis

The research material consisted of sets of ultrasound images belonging to 60 patients (28 healthy and 32 sick). Images of the left and right lobe of the thyroid in a longitudinal section were used. Regions of interest (ROI) of 200×150 pixels during cropping images were isolated (fig. 1). As a result, 126 samples belonging to the cases identified as sick and 108 samples for the healthy ones were obtained. Then, all the ROI were converted from RGB to 8-bit grayscale and normalized within the brightness range with a linear correction. The full data set consisted of 234 cases. The proportion for healthy/sick cases was 108/126 (46% of healthy and 54% of sick patients). This set was divided on training, validation and test part and each of them consisted of 78 cases. The same proportion of healthy and sick cases for each of these 3 sets as for the full data set was maintained. As a result, these sets consisted of 36 healthy and 42 sick cases. Data prepared in such a way were used for the image analysis aimed at defining a set of feature descriptors, with the help of The MaZda computer programme.² It was developed in the Institute of Electronics at the Lodz University of Technology and made available on the Internet, free of charge for research. MaZda allows the analysis of gray texture images and calculates an extensive set of 283 discriminant features. A detailed description of these features can be found in (Haralick 1979; Haralick, Shanmugam, and Dinstein 1973; Hu and Dennis 1994; Lerski et al. 1993), and in the technical documentation of MaZda.

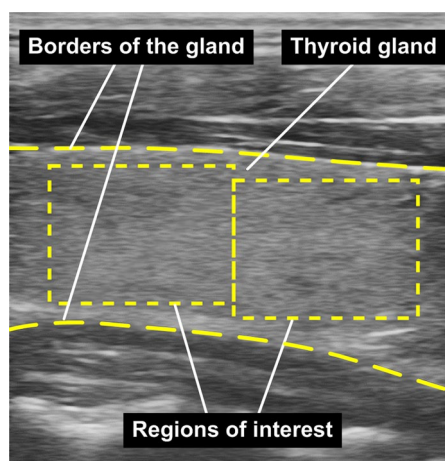


Fig. 1. Location of the ROI on a background of the longitudinal section of the thyroid lobe. The figure reflects real proportions between gland dimensions on the ultrasound image and the size of the ROI

2. See: http://www.elel.p.lodz.pl/programy/cost/progr_mazda.html.

The set of features was obtained basing on:

- Histogram (9 features): histogram’s mean, histogram’s variance, histogram’s skewness, histogram’s kurtosis, percentiles 1%, 10%, 50%, 90% and 99%.
- Gradient (5 features): absolute gradient mean, absolute gradient variance, absolute gradient skewness, absolute gradient kurtosis, percentage of pixels with nonzero gradient.
- Run length matrix (5 features \times 4 various directions): run length nonuniformity, grey level nonuniformity, long run emphasis, short run emphasis, fraction of image in runs.
- Co-occurrence matrix (11 features \times 4 various directions \times 5 between-pixels distances) angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy.
- Autoregressive model (5 features): parameters $\Theta_1, \Theta_2, \Theta_3,$ and $\Theta_4,$ standard deviation.
- Haar wavelet (24 features): wavelet energy (features are computed at 6 scales within 4 frequency bands LL, LH, HL, and HH).

After rejecting features of constant value for all the observations, the final set used for further tests consisted of 281 features. Quotient and interval scales were used during measurement of their numerical values.

2 Feature selection

2.1 Heuristic identification of noisy variables

The method of heuristic identification of noisy variables (HINoV) has been chosen because the results of simulation analysis available in the literature showed that for data generated with a given structure of classes and taking into account noisy variables these results were better in 97% of cases with the use of HINoV method than for all the variables (Walesiak and Gatnar 2009). The algorithm of HINoV method includes 5 following steps (Carmone Jr, Kara, and Maxwell 1999):

1. The input data matrix containing standardized observations for the j -th variable and the i -th object is prepared. It is assumed that the variables describing objects (fig. 2a) are measured in a metric scale. This assumption was fulfilled for variables describing images of the thyroid gland.
2. For each j -th variable a set of objects (A) is clustered using the k -means method on a fixed number of groups (u).
3. The corrected Rand index (R_{jh} for $j, h = 1, 2, \dots, m$) for all the combinations of pairs of divisions ($j \neq h$) is calculated. Half of all the combinations ($m(m - 1)/2$) is calculated because the R_{jh} index is symmetrical.
4. The matrix of $m \times m$ is built on the base of previously calculated corrected Rand indexes (fig. 2b). Then, for each row (or column) the sum of $R_{j\bullet} = \sum_{h=1}^m R_{jh}$ is calculated (fig. 2c).
5. The values of $R_{1\bullet}, R_{2\bullet}, \dots, R_{m\bullet},$ are in a descending order. Then, the scree plot is built and, on its base, a subset of m^* variables is selected. In this way, the number of $m - m^*$ variables disturbing the existing structure of groups in the data set is eliminated.

$$\begin{array}{ccc}
 \text{a)} & \text{b)} & \text{c)} \\
 \left[\begin{array}{c} M_1 \\ M_2 \\ \vdots \\ M_j \\ \vdots \\ M_m \end{array} \right] ; & \left[\begin{array}{ccccccc} & R_{12} & R_{13} & \cdots & R_{1h} & \cdots & R_{1m} \\ R_{21} & & R_{23} & \cdots & R_{2h} & \cdots & R_{2m} \\ R_{31} & R_{32} & & \cdots & R_{3h} & \cdots & R_{3m} \\ \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots \\ R_{j1} & R_{j2} & \cdots & \cdots & & \cdots & R_{jm} \\ \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots \\ R_{m1} & R_{m2} & \cdots & \cdots & R_{mh} & \cdots & \end{array} \right] ; & \left[\begin{array}{c} R_{1\bullet} \\ R_{2\bullet} \\ \vdots \\ R_{j\bullet} \\ \vdots \\ R_{m\bullet} \end{array} \right]
 \end{array}$$

Fig. 2. A data structures organization used in the algorithm: a) a vector of variables, b) a matrix of corrected Rand indexes R_{jh} , c) a sum of elements $R_{j\bullet}$ in a row (or column)

To remove variables (features) that disturb the structure of classes of thyroid images the HINoV.Mod function from the clusterSim package of R programme has been used (Koronacki and

Ćwik 2005).³ This function extends capabilities of the original HINoV method because it can also use other than k -means clustering methods. In addition, it can as well be used for non-metric variables (Walesiak 2005). The following parameters for the HINoV.Mod function were used: 2 clusters, squared Euclidean distance and k -medoids clustering method. As a result of calculations, values of the corrected Rand index for each variable in a descending order were obtained. On this base the scree plot was built (fig. 3). This plot made it possible to select a subset of variables for the further analysis. In figure 3 the dashed line is the boundary where the scree begins. Thus, variables to the right of this line can be treated as noisy variables. By the use of the scree criterion 57 significant features were selected.

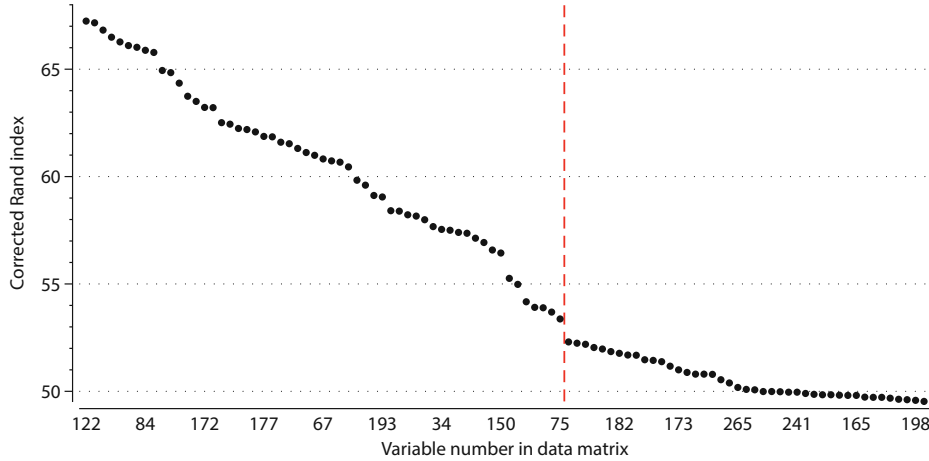


Fig. 3. The initial part of the scree plot showing values of the corrected Rand index for each variable. Red dashed line marks the boundary where the scree begins

2.2 Testing the significance of the correlation coefficients

The second method used for dimension reduction in feature space was based on testing the significance of the Spearman correlation coefficients between individual variables and decision variable (healthy or sick). The following relationship for calculating the coefficient of the Spearman rank correlation was used:

$$(1) \quad r_s = \frac{\frac{n^3-n}{6} - T_x - T_y - \sum_{i=1}^n (x_i - y_i)^2}{\sqrt{\left(\frac{n^3-n}{6} - 2T_x\right) \left(\frac{n^3-n}{6} - 2T_y\right)}}$$

where:

n — sample size,

x_i, y_i — rank of the i -th element according to the first and second feature,

T_x, T_y — corrections for combined ranks given by the expression (2), indexes of x and y are used to distinguish between features.

$$(2) \quad T = \frac{1}{12} \sum_{j=1}^k (t_j^3 - t_j),$$

where:

t_j — the number of combined ranks for the j -th combination,

k — the number of combinations.

If $T_x = T_y = 0$ (no combined ranks), then the Spearman rank correlation coefficient can be calculated using the expression

$$(3) \quad r_s = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (x_i - y_i)^2.$$

3. See also: Documentation of the clusterSim package [a] <http://keii.ue.wroc.pl/clusterSim/clusterSim.pdf>.

Combining ranks were in our data, therefore the expression (1) for calculations was used. The following null hypothesis during the test of the Spearman coefficients was accepted:

$$(4) \quad H_0: r_s = 0.$$

Evaluation of this significance depends on the sample size:

- for small samples, special statistical tables are used,
- for samples containing more than 10 cases approximations are used and calculations are independent from statistical tables.

For the analysed dataset, the r_s coefficient based on 234 elements was calculated. In such a situation, the random variable

$$(5) \quad u = \frac{r_s}{\sqrt{1/(n-1)}}$$

has a standardized normal distribution of $N(0, 1)$. Thus, it is sufficient to calculate the u value for the test sample and compare it to the u_α critical value. For appropriate significance levels, the critical value equals to:

- $u_{0,10} = 1,64$ for $\alpha = 0,10$
- $u_{0,05} = 1,96$ for $\alpha = 0,05$
- $u_{0,01} = 2,58$ for $\alpha = 0,01$

If $|u| \geq u_\alpha$ then we reject H_0 hypothesis, while if $|u| < u_\alpha$ then there is no basis for its rejection. As a result of testing the significance of the correlation coefficients, the set of 48 discriminant features for the assumed significance level of $\alpha = 0,05$ was obtained.

2.3 Hellwig index of information capacity

The Hellwig algorithm is one of the most common methods for selecting variables for the econometric model. It was developed by a Polish econometrician Zdzisław Hellwig (1968). The idea of this algorithm is to choose a combination of variables from the set of potential variables. The selected combination of variables should comply with the following conditions. The correlation coefficient for the dependent variable should be as large as possible, while the correlation coefficient for explanatory variables occurring in this combination should be as small as possible.

The Hellwig algorithm uses correlation coefficients for variables. They are: a vector of correlation coefficients (R_0) for the dependent variable (Y) and explanatory variables (X_1, X_2, \dots, X_n), and a matrix of correlation coefficients (R) for explanatory variables (X_1, X_2, \dots, X_n):

$$(6) \quad R_0 = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}; \quad \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{12} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1n} & r_{2n} & \cdots & 1 \end{bmatrix},$$

where n is the number of observations. To determine correlation coefficients the Spearman correlation was used. It was calculated according to the equation (1). The discussed algorithm aims at finding the best combination of explanatory variables (i.e., a combination of the largest integral index of information capacity). Therefore, explanatory variables highly correlated to the dependent variable and poorly correlated to each other are selected. The number of explanatory variables (X_1, X_2, \dots, X_n) equals to $L = 2^n - 1$. Calculations according to the Hellwig algorithm have 3 main steps:

1. Calculation of individual indexes of information capacity for all the combinations:

$$(7) \quad h_{kj} = \frac{r_j^2}{1 + \sum_{\substack{i=1 \\ i \neq j}} |r_{ij}|},$$

where:

k — the number of combination ($k = 1, 2, \dots, l$),

j — the number of variable in a combination ($j = 1, 2, \dots, m$).

2. Calculation of integral indexes of information capacity for all the combinations:

$$(8) \quad H_k = \sum_{j=1}^m h_{kj}, \quad k = 1, 2, \dots, l.$$

3. Selection of the optimal combination having the highest integral index of information capacity (H). Variables of this combination should be used to build a future model.

The significant level of the Spearman correlation was set to $\alpha = 0,01$. For this setting 20 significant variables were selected. Then, combinations of these variables were categorized according to the value of the index of information capacity. The set of the first 10 combinations of the highest Hellwig index value is shown in the table 1. The maximum number of variables in analysed combinations was set to 10. As a result of the analysis, it has been found out that the highest Hellwig index value (0,3583) has a combination of 3 variables (tab. 2). These variables were accepted as the result of the Hellwig method and they were used for the further research (Omiotek and Wójcik 2014).

Tab. 1. The set of the first 10 combinations of the highest Hellwig index value

No.	Combination of variables	Hellwig index value
1.	254-275-277	0,3583
2.	254-275-277-281	0,3447
3.	254-273-275-277	0,3320
4.	214-254-275-277	0,3297
5.	225-254-275-277	0,3292
6.	203-254-275-277	0,3281
7.	170-254-275-277	0,3259
8.	254-255-275-277	0,3258
9.	181-254-275-277	0,3256
10.	254-273-275-277-281	0,3240

Tab. 2. Combination of the variables of the highest Hellwig index value

No.	Variable name	Variable meaning
254	Teta2	Parameter (Θ_2) of the autoregression model
275	WavEnLH_s-5	Wavelet energy for the LH subband in the 5th scale
277	WavEnHH_s-5	Wavelet energy for the HH subband in the 5th scale

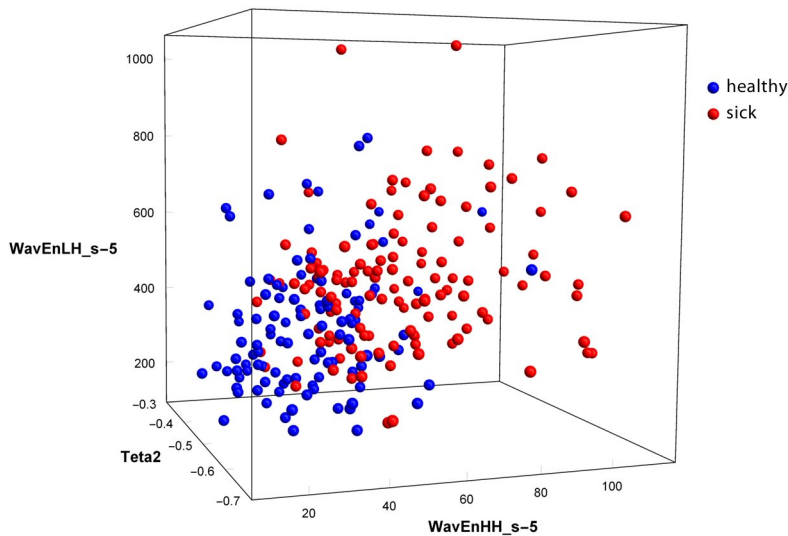


Fig. 4. A three-dimensional map showing objects positioning for variables selected using the Hellwig method

In figure 4, the positioning map of objects for variables selected using the Hellwig method has been shown. This figure shows the tendency to clustering of the cases belonging to the same class. Such a situation gives the potential possibility of using the reduced set of 3 variables to build a classifier.

3 Evaluation of the feature selection methods

If objects are classified into 2 classes, then the classifier accuracy is often described by means of 2 parameters. This is especially significant for medical applications, where we have 2 classes—healthy and sick patients. The first parameter, sensitivity (SE), is defined by the formula

$$(9) \quad SE = \frac{TP}{TP + FN},$$

where TP is the number of intrinsically sick patients classified to the class of sick patients (true-positive classification) and FN is the number of intrinsically sick patients classified to the class of healthy patients (false-negative classification). It gives the probability estimation of the correct case classification as sick, provided that it belongs to the class of sick patients. The second parameter, specificity (SP), is defined by the formula

$$(10) \quad SP = \frac{TN}{FP + TN},$$

where TN is the number of intrinsically healthy patients classified to the class of healthy patients (true-negative classification) and FP is the number of intrinsically healthy patients classified to the class of sick patients (false-positive classification). It gives the probability estimation of the correct case classification as healthy, provided that it belongs to the class of healthy patients.

The results of the classification are presented in the next part of this section. Ten different classification methods and 3 data sets reduced using methods described in Section 2 have been used. All the reduced data sets were divided into training, validation and test part according to rules described in Section 1. Training and validation sets were employed to build classifiers. Following classification methods were used: linear (LDA) and quadratic (QDA) discriminant analysis, classification trees (CT), k -nearest neighbours (K-NN), artificial neural networks (ANN), support vector machines (SVM), bootstrap aggregation (BAGG), models strengthening (STREN), models binding (BIND) and random forests (RF). Next, the best models for each feature selection method were chosen. These models were evaluated using the sensitivity (10) and specificity (11) on the basis of the independent test set. These parameters give the valuable information about such classifier properties that are very important, especially in medical applications. Analysed cases were assigned to one of two classes—sick or healthy. Classification results were used for the comparison of

Tab. 3. Classification sensitivity for the test data set (the best results for each classification method marked in bold)

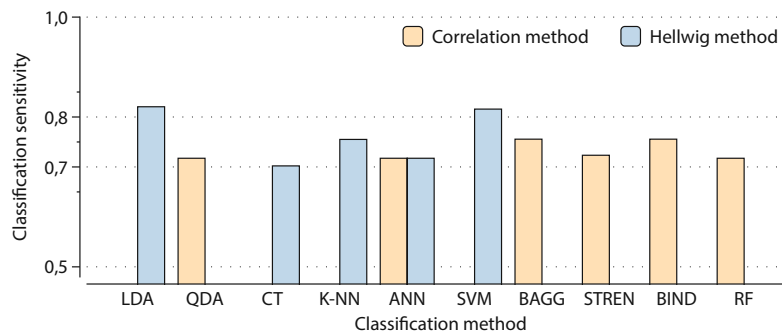
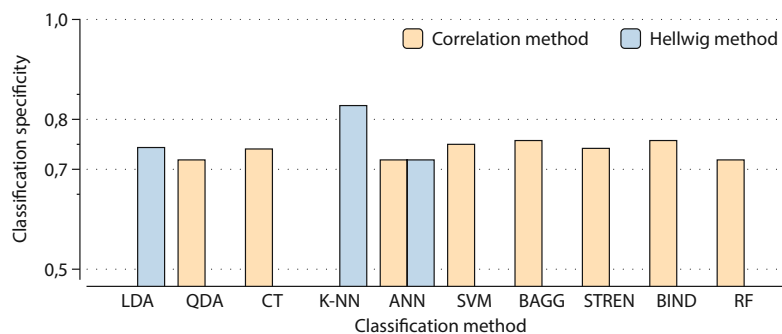
Classification method	Feature selection method		
	HINoV	Correlation	Hellwig
LDA	0,6591	0,6389	0,8205
QDA	0,6102	0,7174	0,7143
CT	0,5758	0,6863	0,7021
K-NN	0,5897	0,6939	0,7551
ANN	0,6667	0,7174	0,7174
SVM	0,6889	0,7391	0,8158
BAGG	0,6744	0,7556	0,7174
STREN	0,6596	0,7234	0,6667
BIND	0,6744	0,7556	0,7174
RF	0,6809	0,7174	0,6744

Tab. 4. Classification specificity for the test data set (the best results for each classification method marked in bold)

Classification method	Feature selection method		
	HINoV	Correlation	Hellwig
LDA	0,6176	0,5476	0,7436
QDA	0,6842	0,7188	0,6667
CT	0,6667	0,7407	0,7097
K-NN	0,5128	0,7241	0,8276
ANN	0,6667	0,7188	0,7188
SVM	0,6667	0,7500	0,7250
BAGG	0,6286	0,7576	0,7188
STREN	0,6452	0,7419	0,6111
BIND	0,6286	0,7576	0,7188
RF	0,6774	0,7188	0,6286

effectiveness of the feature selection methods (tab. 3 and 4). The best results for each classification method are presented in figure 5 (classification sensitivity) and in figure 6 (classification specificity).

Good results have been achieved by using 2 presented methods for dimension reduction in feature space. The first of them was the correlation method and the second one—the Hellwig method. On the basis of a set of features selected by the correlation method classifiers of a highest classification sensitivity for 6 classification methods were built (tab. 3). For the correlation method, the best classifiers of sensitivity equal to 0,7556 were built, using bootstrap aggregation and models binding methods. In terms of the specificity, the best results have been achieved for 8 classification methods using the correlation method for the feature selection (tab. 4). The best classifiers of sensitivity equal to 0,7576 were built using bootstrap aggregation and models binding methods as well. Good results were also achieved for a set of features selected by the Hellwig method. For this feature selection method, 5 classifiers had the highest classification sensitivity (tab. 3) and 3 of them had the best classification specificity (tab. 4). For the Hellwig method, the model of

**Fig. 5.** The best models in terms of the classification sensitivity**Fig. 6.** The best models in terms of the classification specificity

the highest classification sensitivity (equal to 0,8205) using the linear discriminant analysis was built and the model of the best classification specificity (equal to 0,8276) using the k -nearest neighbours method was built. It should be emphasized that the set of features reduced using the Hellwig method made it possible to build 2 models of classification sensitivity equal to approx. 0,82, which proved to be the best for all the built models (fig. 5). These classifiers have been built using linear discriminant analysis and support vector machines methods. By means of the features selected by the Hellwig method the classifier of the highest classification specificity equal to approx. 0,83 (fig. 6) has been built as well. The k -nearest neighbours method for this case was used. The worst results were achieved for the HINoV method. For this case, both the sensitivity and the specificity of the built classifiers did not exceed the level of 0,7.

Conclusions

Results of dimension reduction in feature space using the method of heuristic identification of noisy variables, testing the significance of the correlation coefficients and the method of Hellwig index of information capacity have been compared in this work. An input set consisted of 283 discriminant features. It was reduced to 57 features using the first method, to 48 features using the second one and to 3 features using third method for feature selection. Next, reduced data sets were used to build and test popular classifiers. Results showed the usefulness of the correlation and Hellwig index of information capacity methods, which have been mainly used so far in the study of economic phenomena. We believe that these methods can be successfully used not only to study of the textures to detect the Hashimoto's disease, but also to other classification issues based on the texture image analysis. Results obtained using the Hellwig method should be pointed. Classifiers built using a set of data reduced by this method had the highest classification sensitivity, equal to approx. 0,82 (LDA and SVM methods), and the highest classification specificity, equal to approx. 0,83 (K-NN method). These results suggest that the Hellwig method can be used as an effective method for dimension reduction in feature space in classification of thyroid ultrasound images.

References

- BREIMAN, L. 1996. "Bagging Predictors." *Machine Learning* no. 24 (2):123–140. doi: 10.1023/A:1018054314350.
- . 2001. "Random Forests." *Machine Learning* no. 45 (1):5–32. doi: 10.1023/A:1010933404324.
- BREIMAN, L., J.H. FRIEDMAN, R.A. OLSHEN, and C.J. STONE. 1984. *Classification and Regression Trees, The Wadsworth Statistics/Probability Series*. Belmont, Calif.: Wadsworth International Group.
- CARMONE JR, F.J., A. KARA, and S. MAXWELL. 1999. "HINoV: A New Model to Improve Market Segment Definition by Identifying Noisy Variables." *Journal of Marketing Research* no. 36 (4):501–509. doi: 10.2307/3152003.
- ENAS, G.G., and S.C. CHOI. 1986. "Choice of the Smoothing Parameter and Efficiency of K-Nearest Neighbor Classification." *Computers & Mathematics with Applications-Part A* no. 12 (2):235–244. doi: 10.1016/0898-1221(86)90076-3.
- FISHER, R.A. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* no. 7 (2):179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x.
- FREUND, Y., and R.E. SCHAPIRE. 1997. "A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* no. 55 (1):119–139. doi: 10.1006/jcss.1997.1504.
- HARALICK, R.M. 1979. "Statistical and Structural Approaches to Texture." *Proceedings of the IEEE* no. 67 (5):786–804. doi: 10.1109/Proc.1979.11328.
- HARALICK, R.M., SHANMUGA.K, and I. DINSTEN. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems Man and Cybernetics* no. Smc3 (6):610–621. doi: 10.1109/Tsmc.1973.4309314.
- HELLWIG, Z. 1968. "On the Optimal Choice of Predictors." In *Toward a System of Quantitative Indicators of Components of Human Resources Development*, edited by Z. Gostkowski. Paryż: UNESCO.

- HOTHORN, T., and B. LAUSEN. 2005. "Bundling Classifiers by Bagging Trees." *Computational Statistics & Data Analysis* no. 49 (4):1068–1078. doi: 10.1016/j.csda.2004.06.019.
- HU, Y., and T.J. DENNIS. 1994. "Textured Image Segmentation by Context Enhanced Clustering." *IEEE Proceedings-Vision Image and Signal Processing* no. 141 (6):413–421. doi: 10.1049/ip-vis:19941548.
- KORONACKI, J., and J. ĆWIK. 2005. *Statystyczne systemy uczące się*. Warszawa: Wydawnictwa Naukowo-Techniczne.
- LERSKI, R.A., K. STRAUGHAN, L.R. SCHAD, D. BOYCE, S. BLUML, and I. ZUNA. 1993. "MR Image Texture Analysis—an Approach to Tissue Characterization." *Magnetic Resonance Imaging* no. 11 (6):873–887. doi: 10.1016/0730-725x(93)90205-R.
- LIAO, S.H., P.H. CHU, and P.Y. HSIAO. 2012. "Data Mining Techniques and Applications—A Decade Review from 2000 to 2011." *Expert Systems with Applications* no. 39 (12):11303–11311. doi: 10.1016/j.eswa.2012.02.063.
- LIGEZA, A. 2006. *Logical Foundations for Rule-Based Systems, Studies in Computational Intelligence*. Berlin – New York: Springer.
- OMIOTEK, Z., A. BURDA, and W. WÓJCIK. 2013. "The Use of Decision Tree Induction and Artificial Neural Networks for Automatic Diagnosis of Hashimoto's Disease." *Expert Systems with Applications* no. 40 (16):6684–6689. doi: 10.1016/j.eswa.2013.03.022.
- . 2015. "Application of Selected Classification Methods for Detection of Hashimoto's Thyroiditis on the Basis of Ultrasound Images." In *Computational Intelligence, Medicine and Biology: Selected Links*, edited by K. Pancierz and E. Zaitseva, 23–37. Cham u.a.: Springer.
- OMIOTEK, Z., and W. WÓJCIK. 2014. "Zastosowanie metody Hellwiga do redukcji wymiaru przestrzeni cech obrazów USG tarczycy." *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Środowiska* (3):14–17.
- TADEUSIEWICZ, R. 1993. *Sieci neuronowe*. 2nd ed., *Problemy Współczesnej Nauki i Techniki Informatyka*. Warszawa: Akademicka Oficyna Wydawnicza RM.
- WALESIAK, M. 2005. "Problemy selekcji i ważenia zmiennych w zagadnieniu klasyfikacji." *Prace Naukowe AE we Wrocławiu. Taksonomia* 12 (1076):106–118.
- WALESIAK, M., and E. GATNAR. EDS. 2009. *Statystyczna analiza danych z wykorzystaniem programu R*. Warszawa: Wydawnictwo Naukowe PWN.