Multi-resolution Image Parametrization in Stepwise Diagnostics of Coronary Artery Disease

Matjaž Kukar[1], Luka Šajn[1,*], Ciril Grošelj[2], and Jera Grošelj[2]

[1] University of Ljubljana, Faculty of Computer and Information Science,
Tržaška 25, SI-1001 Ljubljana, Slovenia
`{matjaz.kukar,luka.sajn}@fri.uni-lj.si`
[2] University Medical Centre Ljubljana, Nuclear Medicine Department,
Zaloška 7, SI-1001 Ljubljana, Slovenia
`ciril.groselj@kclj.si`

Abstract. Coronary artery disease is one of the world's most important causes of early mortality, so any improvements of diagnostic procedures are highly appreciated. In the clinical setting, coronary artery disease diagnostics is typically performed in a sequential manner. The four diagnostic levels consist of evaluation of (1) signs and symptoms of the disease and ECG (electrocardiogram) at rest, (2) ECG testing during a controlled exercise, (3) myocardial perfusion scintigraphy, and (4) finally coronary angiography (which is considered as the "gold standard" reference method). In our study we focus on improving diagnostic performance of the third diagnostic level (myocardial perfusion scintigraphy). This diagnostic level consists of series of medical images that are easily obtained and the imaging procedure represents only a minor threat to patients' health. In clinical practice, these images are manually described (parameterized) and subsequently evaluated by expert physicians. In our paper we present an innovative alternative to manual image evaluation – an automatic image parametrization on multiple resolutions, based on texture description with specialized association rules, and image evaluation with machine learning methods. Our results show that multi-resolution image parameterizations equals the physicians in terms of quality of image parameters. However, by using both manual and automatic image description parameters at the same time, diagnostic performance can be significantly improved with respect to the results of clinical practice.

Keywords: machine learning, coronary artery disease, medical diagnosis, image parametrization, association rules, stepwise diagnostic process.

1 Introduction

Coronary artery disease (CAD) is one the world's main cause of early mortality, and there is an ongoing research for improving diagnostic procedures. The usual clinical process of coronary artery disease diagnostics consists of four diagnostic steps (levels): (1) evaluation of signs and symptoms of the disease and ECG (electrocardiogram) at rest; (2) ECG testing during the controlled exercise; (3) stress myocardial scintigraphy; and (4) coronary angiography.

---

* Parts of work presented in this paper are taken from the second author's doctoral dissertation.

In this process, the fourth diagnostic level (coronary angiography) is considered as the "gold standard" reference method. As this diagnostic procedure is invasive, comparatively expensive, and potentially dangerous for the patients, there is a tendency to improve diagnostic performance and reliability of earlier diagnostic levels, especially of myocardial scintigraphy [9, 10]. Approaches used for this purpose include applications of neural networks [1], expert systems [7], subgroup mining [6], statistical techniques and rule-based approaches [11]. In our study we focus on different aspects of improving the diagnostic performance of myocardial scintigraphy.

Results of myocardial scintigraphy consist of series of medical images that are taken both during rest and a controlled exercise. These images are readily available in PC or Mac format by respective SPECT cameras and such and imaging procedure does not threaten patients' mostly frail health.

In clinical practice, expert physicians use their medical knowledge and experience as well as the image processing capabilities provided by various imaging software to manually describe (parameterize) and evaluated the images.

In our paper we present an innovative alternative to manual image evaluation – automatic multi-resolution image parametrization, based on texture description with specialized association rules, and image evaluation with machine learning methods. Our results show that multi-resolution image parametrization equals or even betters the physicians in terms of the quality of image parameters. Additionally, by using both manual and automatic image description parameters at the same time, diagnostic performance can be significantly improved with respect to the results of clinical practice.

2   Methods and Materials

2.1   Stepwise Diagnostic Process

Every medical diagnosis inherently contains some uncertainty and is therefore not completely reliable. Sometimes it is crucial to know the magnitude of diagnosis' reliability in order to minimize risks for patient's health or even life.

In a stepwise diagnostic process diagnostic tests are ordered by some pre-determined criteria, such as increasing cost, diagnostic accuracy, and invasiveness (in this order). Key elements of stepwise testing are the estimate of the prior (pre-test) probability of a disease, and the sensitivity and specificity of different diagnostic levels. With this information, test results can be analyzed by sequential use of the Bayes' conditional probability theorem. The obtained post-test probability accounts for the pre-test probability, sensitivity and specificity of the test, and may later be used as a pre-test probability for the next test in sequence (Fig. 1). The process results in a series of tests where
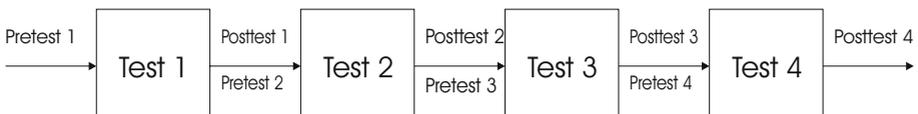
Fig. 1. Increasing diagnostic test levels in stepwise diagnostic process

each test is performed independently. Its results may be interpreted with or without any knowledge of the other test results. In diagnostic problems, the performance of a diagnostic test is described with diagnostic (classification) accuracy (*Acc*), sensitivity (*Se*) and specificity (*Sp*). Test results from earlier levels are used to obtain the final probability of disease. Stepwise diagnostic tests are performed until the post-test probability of disease's presence or absence exceeds some pre-defined threshold value [3].

The Bayes' theorem is applied to calculate the conditional probability of the disease's presence, when the result of a diagnostic test is given. For positive test result the probability $P(d|+) = P(disease|positive\ test\ result)$ is calculated:

$$P(d|+) = \frac{P \cdot Se}{P \cdot Se + (1 - P) \cdot (1 - Sp)} \tag{1}$$

For negative test result the probability $P(d|-) = P(disease|negative\ test\ result)$ is calculated:

$$P(d|-) = \frac{P \cdot (1 - Se)}{P \cdot (1 - Se) + (1 - P_1) \cdot Sp} \tag{2}$$

The post-test probability after a diagnostic test represents the pre-test probability for the subsequent test. This approach may not only incorporate several test results but also the data from the patient's history [3].

2.2   Image Parametrization

Images in digital form are normally described with matrices which are spatially complex and yet do not offer features that could uniformly distinguish between their predefined classes. Determining image features that can satisfactorily discriminate observed image classes is a hard task for which different algorithms exist. They transform the image from the matrix form into a set of numeric or discrete features (parameters) that convey useful information for discrimination between classes.

The ArTeX Algorithm.   The use of association rules used for texture description were first described in [16]. We follow a slightly different approach introduced in [2], where different texture representation and different algorithm for association rules are used.

Fig. 2 illustrates the association rule $(1, 1) \wedge (2, 10) \implies (1, 15) \wedge (3, 5)$, which can be read as follows: if a pixel of intensity 1 is found on distance 1 and a pixel of intensity
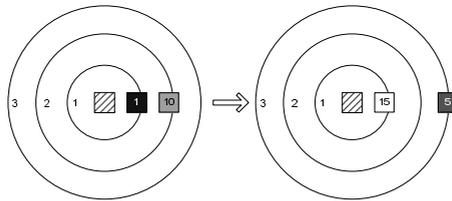
Fig. 2. An illustration of association rule $(1, 1) \wedge (2, 10) \implies (1, 15) \wedge (3, 5)$

10 is found on distance 2, then there is also a pixel of intensity 15 on distance 1 and a pixel of intensity 5 on distance 3.

Using association rules on textures allows to extract a set of features (attributes) for a particular domain of textures. Here is a general description of the ArTeX algorithm:

- *Select a (small) subset of images F for feature extraction.* The subset F can be considerably small. Use at least one example of each typical image in the domain. That is at least one sample per class, or more if the class consists of subclasses.
- *Pre-processing of images in F.* Pre-processing involves the transformation of images to grey scale if necessary, the quantization of grey levels and the selection of proper neighborhood size R. The initial number of grey levels per pixel is usually 256. The quantization process downscales it to say 16 levels per pixel. Typical neighborhood sizes are 3, 4, 5.
- *Generate association rules from images in F.* Because of the representation of texture, it is possible to use any algorithm for association rules extraction. We use the well-known algorithms *Apriori* and *GenRules*.
- *Use generated association rules to extract a set of features.* There are two features associated with each association rule: support and confidence. Use these two attributes of all association rules to construct a feature set. The number of extracted features is twice the number of association rules, which could be quite a lot.

Earlier experiments [18] show excellent results when using ArTeX-type texture descriptions in conjunction with machine learning algorithms. One of the reasons for the success of ArTeX is that it describes images in a multi-resolution and rotation-invariant manner. This parametrization is also invariant to image brightness which is in our case necessary due to varying radiopharmaceutical agent absorption. These features make ArTeX a promising tool for analyzing myocardial scintigrams.

Multi-resolution Parametrization. Algorithms for image parametrization are suitable either for images (imaging some content of different classes) or textures (representing some repeating patterns). Image illumination, scale and affine transformations often obstruct the parametrization. Algorithms use different pixels' properties and relations between them since images in digital representation are described with pixels. Due to the time and space complexity only a predefined size of pixel neighborhood is observed, which makes detectable relations between pixels quite dependent on image resolution. Not only different image scales require appropriate resolutions, but also when there are more shapes of different size present in a picture more resolutions are desired. By using only a single resolution, we may miss the big picture, and proverbially not see the forest because of the trees.

Another issue is the pattern scale. Not every combination of scale and neighborhood size can guarantee that the pattern would be detected. This yields a solution where more resolutions are simultaneously observed in one image and obtained features for each resolution are combined together in one feature vector.

If we want to use more resolutions it is necessary to determine which ones to use. Many existing applications use fixed resolutions irrespectively of the image content and usually three or four are used [5, 14]. Multi-resolution algorithms usually perform better when using only a few resolutions; more resolutions typically yield worse results.

We have developed an algorithm [17] for determining the resolutions for which more informative features can be obtained. The idea for the algorithm is derived from the well known SIFT algorithm [13]. In this way also resolutions for the hearth scintigraphy are determined.

When detecting the appropriate resolutions the image is consequently resized from 100% down to some predefined lowest threshold at some fixed step. At each resize peaks are counted. Peaks are represented by pixels which differ from their neighborhood either as highest or lowest intensity. This algorithm can be implemented also with DOG (Difference-Of-Gaussian) [13] method which improves the time complexity with lower number of actual resizes required to search the entire resolution space.

Detected peak counts are recorded over all resolutions as a histogram. From the histogram the best resolutions are detected at the highest counts. The number of resolutions we want to use in our parametrization is predefined. When there are more equal counts we chose as diverse resolutions as possible [17]. When optimal resolutions are determined, an image parametrization algorithm (Artex in our case, but could be anything) is used to describe images.

2.3   Medical Data

In our study we used a dataset of 288 patients with suspected or known CAD. All patients had performed proper clinical and laboratory examinations, exercise ECG, stress myocardial perfusion scintigraphy (complete image sets were available for analysis), and coronary angiography. Features for the ECG an scintigraphy data were extracted manually by the clinicians. 10 patients were excluded for data pre-processing and calibration required by multi-resolution ArTeX, so only 278 patients (66 females, 212 males, average age 60 years) were used in actual experiments. In 149 cases the disease was angiographically confirmed and in 129 cases it was excluded. The patients were selected from a population of several thousands patients who were examined at the Nuclear Medicine Department between 2001 and 2004. We selected only the patients with complete diagnostic procedures (all four levels), and for whom the imaging data was readily accessible. Some characteristics of the dataset are shown in Tab. 1.

Table 1. CAD data for different diagnostic levels. Of the attributes belonging to the coronary angiography diagnostic level, only the final diagnosis – the two-valued class – was used in experiments.

Diagnostic level	Number of attributes		
	Nominal	Numeric	Total
1. Signs and symptoms	22	5	27
2. Exercise ECG	11	7	18
3. Myocardial scintigraphy (+9 image series)	8	2	10
4. Coronary angiography	1	6	7
Class distribution	129 (46.40%)		CAD negative
	149 (53.60%)		CAD positive

It must be noted that our patients represent a highly specific population, since many of them had already had performed cardiac surgery or dilatation of coronary vessels. This clearly reflects the situation in Central Europe with its aging population. It is therefore not surprising that both the population and the predictive performance are considerably different than that of our previous study, where data were collected between years 1991 and 1994 [8]. These differences are a consequence of rapidly progressing interventional cardiology. and are therefore not applicable to the general population, but only to comparable population in developed countries. Similarly, general findings about CAD only partially apply to our population.

Scintigraphic Images. In each patient series of images were taken with the General Electric Millennium SPECT gamma camera, both at rest and after a controlled exercise, thus producing the total of 64 grayscale images in resolution of 64×64 8-bit pixels. Because of patients' movements and partial obscuring the heart by other internal organs, these images are not suitable for further use without heavy pre-processing. For this purpose, a General Electric workstation running eNTEGRA software was used (more specifically we used the Emory Cardiac Toolbox [4]). One of ECToolbox's outputs, a series of 9 polar map (bull's eye) images was used for each patients. Polar maps were chosen because previous work in this field [12] showed that they have useful diagnostic value. The 9 polar map images consist of the following images [4]:

- three raw images (the stress and the rest image, as well as the reversibility image, calculated as a difference between normalized rest and stress images
- three blackout (defect extent) images (which are the stress and the rest image, compared with the respective database of normal images, and suitably processed). Again the reversibility image, calculated as a difference between normalized rest and stress blackout images.
- three standard deviation images that show relative perfusion variance when compared to the respective database of normal images.

An example of polar map images for a typical patient with well-defined CAD is shown in Fig. 3. Unfortunately, in most cases (and especially in our specific population) the differences between images taken during exercise and at rest are not so clear-cut as shown in Fig. 3. Interpretation and evaluation of scintigraphic images therefore requires considerable knowledge and training of expert physicians. Although specialized tools such as the ECToolbox software can aid in this process, they still require a lot of training and medical knowledge for evaluation of results. The aim of our study is to use automatic image parametrization in conjunction with machine learning methods in order to provide additional diagnostic tools.

3   Results

As already mentioned in Sec. 2.3, out of the 288 patients, 10 were excluded for data preprocessing and calibration required by the multi-resolution ArTeX parametrization procedure. These patients were not used in further experiments. The remaining 278 patients with 9 images each were parameterized for three resolutions in advance. The
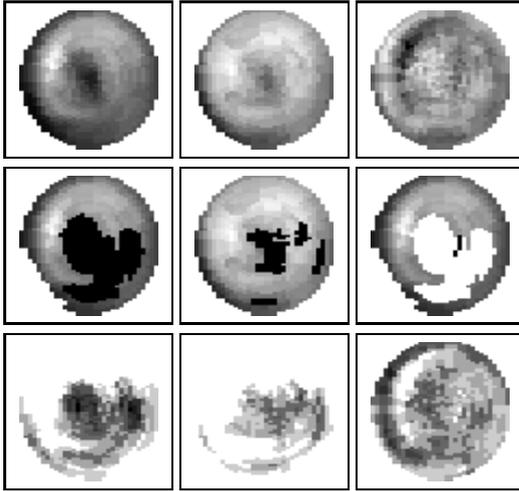
Fig. 3. Typical polar maps taken after exercise (first column), at rest (second column), and their difference (third column). The first row consists of raw images, the second of blackout images, and the third of standard deviations. Black regions indicates less perfused cardiac tissue (a potential defect). Images shown in this figure correspond to the patient with a very clear manifestation of CAD.

proposed three resolutions[1] were $0.95\times$, $0.80\times$, and $0.30\times$ of the original resolution, producing together 2944 additional attributes (features). Since this number is too large for most practical purposes, we filtered[2] it to 200 best features as estimated with the ReliefF algorithm[15]. We also did some experiments with other image parametrization approaches such as wavelet and DFT transform, Gabor filters, and combined them with SIFT-like resolution selection; they, however, mostly performed considerably worse than ArTeX. We omit further analysis of these results due to lack of space.

We applied three popular machine learning algorithms: naive Bayesian classifier, support vector machine, and C4.5 (J4.8) decision tree. We performed experiments with Weka [19] machine learning toolkits.

When necessary, continuous attributes were discretized in advance. Testing was performed in the 10-fold cross-validation setting. Aggregated results of the coronary angiography (CAD negative/CAD positive) were used as the class variable.

Experimental results are compared with diagnostic accuracy, specificity and sensitivity of expert physicians after evaluation of scintigraphic images (Tab. 2). The results of clinical practice were validated by careful blind evaluation of images by an independent expert physician.

For machine learning experiments we considered several different settings: evaluation of attributes extracted by physicians; evaluation of all attributes extracted by multi-resolution ArTeX; evaluation of all attributes extracted by multi-resolution ArTeX in

---

[1] A resolution of $0.30\times$ means $0.30 \cdot 64 \times 0.30 \cdot 64$ pixels instead of 64×64 pixels.

[2] Even better results could be expected if the wrapper approach were used instead of filtering. We chose not to follow this lead for now because of its time consumption.

Table 2. Diagnostic results of the physicians compared with results of machine learning classifiers obtained from the original attributes, as extracted by physicians. Results that are significantly ($p < 0.05$) different from clinical results are emphasized.

	All basic attributes		
	Accurracy	Specificity	Sensitivity
Physicians	64.00	71.10	55.80
Naive Bayes	**68.34**	69.80	**67.10**
SVM	65.10	62.80	**67.10**
J4.8	**57.19**	53.50	**60.40**

conjunction with attributes, extracted by physicians (Tab. 3); as well as the above variants reduced to 200 best attributes with ReliefF (Tab. 4). Significance of differences to clinical results was evaluated by using the McNemar's test.

From Tab. 2 we can see that machine learning algorithms are approximately on level with expert physicians when evaluating the original data, as collected by physicians. Naive Bayesian classifier even achieves significantly higher diagnostic accuracy and slightly lower sensitivity than physicians, while the J4.8 decision tree achieves significantly lower diagnostic accuracy. However, for physicians, improvements of specificity are more important than improvements of sensitivity or overall accuracy, since increased specificity decreases the number of unnecessarily performed higher-level diagnostic tests, and consequently shorter waiting times for truly ill patients.

Table 3. Experimental results of machine learning classifiers on parameterized images obtained by using all available attributes. Results that are significantly ($p < 0.05$) better than clinical results are emphasized.

	All image and basic attributes			All image attributes		
	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
Physicians	64.00	71.10	55.80	64.00	71.10	55.80
Naive Bayes	**70.50**	**72.10**	**69.10**	**70.14**	**72.10**	**68.50**
SVM	**69.40**	69.80	**69.10**	61.15	58.10	**63.80**
J4.8	65.10	60.50	**69.10**	59.71	63.80	55.00

In Tab. 3 we can see that some machine learning algorithms have difficulties when handling a huge number (2944) of attributes, with only 278 learning examples. This can lead to overfitting the learning data and thus lower their diagnostic performance. Only naive Bayesian classifier is significantly better than physicians when using all 2944 attributes. However, using these 2944 attributes together with the original attributes invariably improves the physicians' results, in two of three cases even significantly.

Tab. 4 depicts the situation where machine learning algorithms considerably benefit from attribute filtering. In all cases the results are significantly better than the results of physicians. Especially nice results are that of naive Bayesian classifier, which improves diagnostic accuracy, sensitivity and specificity.

Table 4. Experimental results of machine learning classifiers on parameterized images obtained by selecting only the best 200 attributes. Results that are significantly better ($p < 0.05$) than clinical results are emphasized.

	200 best image and basic attributes			200 best image attributes		
	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
Physicians	64.00	71.10	55.80	64.00	71.10	55.80
Naive Bayes	**74.10**	**79.80**	**69.10**	**72.30**	**79.80**	**65.80**
SVM	**69.42**	65.90	**72.50**	**70.14**	**72.90**	**67.80**
J4.8	**67.62**	63.60	**71.10**	**68.34**	63.60	**72.50**

We also experimented with machine learning classifiers in stepwise process, as shown in Fig. 1 and described in Sec. 2.1. By the stepwise diagnostic process, after the third diagnostic level we get the following percentages reliable diagnoses (post-test probability ≥ 0.90), which are practically the same as the results of expert physicians:

- 30.94% reliable true positive diagnoses and 10.50% erroneously reliable false positive diagnoses
- 19.88% reliable true negative diagnoses and 7.83% erroneously reliable false negative diagnoses

Our preliminary experiments show, that by using additional attributes from parameterized images, we can increase the number of reliable positive and negative diagnoses by almost 10% while keeping the number of incorrect diagnoses lower than the physicians in clinical practice.

4   Discussion

Although our study is still in early stages, the results are promising. We have shown that multi-resolution ArTeX parametrization in conjunction with machine learning techniques can be successfully used as an intelligent tool in image evaluation, as well as as a part of the stepwise diagnostic process. Automatic image parametrization and machine learning methods can help less experienced physicians evaluate medical images and thus improve their combined performance (in terms of accuracy, sensitivity and specificity).

From the practical use of described approaches two-fold improvements of the diagnostic procedure can be expected. Due to higher specificity of tests (by almost 9%), fewer patients without the disease would have to be examined with coronary angiography which is invasive and therefore dangerous method. Together with higher sensitivity this would also save money and shorten the waiting times of the truly ill patients.

The most significant result of our study may well be the improvement in the predictive power of the stepwise diagnostic process. The almost 10% improvement of positive and negative patients who would not need to be examined with costly further tests, represents a significant improvement in the diagnostic power as well as in the rationalization of the existing CAD diagnostic procedure without danger of incorrectly diagnosing more patients than in current practice.

However, it should be emphasized that the results of sour study are obtained on a significantly restricted population and therefore may not be generally applicable to the normal population, i.e. to all the patients coming to the Nuclear Medicine Department of the University Clinical Centre in Ljubljana.

Acknowledgements

References

[1] Allison, J.S., Heo, J., Iskandrian, A.E.: Artificial neural network modeling of stress single-photon emission computed tomographic imaging for detecting extensive coronary artery disease. J Nucl Cardiol 95(2), 178–181 (2005)

[2] Bevk, M., Kononenko, I.: Towards symbolic mining of images with association rules: Preliminary results on textures. Intelligent Data Analysis (2006)

[3] Diamond, G.A., Forester, J.S.: Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. New England Journal of Medicine 300, 1350 (1979)

[4] General Electric. ECToolbox Protocol Operator's Guide (2001)

[5] Ferreira, C.B.R., Borges, D.L.: Automated mammogram classification using a multi-resolution pattern recognition approach. In: SIBGRAPI01, vol. 00, pp. 76 (2001)

[6] Gamberger, D., Lavrac, N., Krstacic, G.: Active subgroup mining: a case study in coronary heart disease risk group detection. Artif Intell Med 28(1), 27–57 (2003)

[7] Garcia, E.V., Cooke, C.D., Folks, R.D., Santana, C.A., Krawczynska, E.G., De Braal, L., Ezquerra, N.F.: Diagnostic performance of an expert system for the interpretation of myocardial perfusion spect studies. J Nucl Med 42(8), 1185–1191 (2001)

[8] Grošelj, C., Kukar, M., Fettich, J., Kononenko, I.: Impact of machine learning to the diagnostic certainty of the patient's group with low coronary artery disease probability. In: Proc. Computer-Aided Data Analysis in Medicine, Bled, Slovenia, pp. 68–74 (1997)

[9] Kukar, M.: Transductive reliability estimation for medical diagnosis. Artif. intell. med. 81–106 (2003)

[10] Kukar, M., Kononenko, I., Grošelj, C., Kralj, K., Fettich, J.: Analysing and improving the diagnosis of ischaemic heart disease with machine learning. Artificial Intelligence in Medicine 16(1), 25–50 (1999)

[11] Kurgan, L.A., Cios, K.J., Tadeusiewicz, R.: Knowledge discovery approach to automated cardiac spect diagnosis. Artif Intell Med 23(2), 149–169 (2001)

[12] Lindahl, D., Palmer, J., Pettersson, J., White, T., Lundin, A., Edenbrandt, L.: Scintigraphic diagnosis of coronary artery disease: myocardial bull's-eye images contain the important information. Clinical Physiology 6(18) (1998)

[13] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)

[14] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)

[15] Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learning 53, 23–69 (2003)

[16] Rushing, J.A., Ranganath, H.S., Hinke, T.H., Graves, S.J.: Using association rules as texture features. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(8), 845–858 (2001)

[17] Šajn, L.: Multiresolution parameterization for texture classification and its application in analysis of scintigrafic images. PhD thesis, Faculty of Computer and Information Science, University of Ljubljana, in Slovene (2007)

[18] Šajn, L., Kukar, M., Kononenko, I., Milčinski, M.: Computerized segmentation of whole-body bone scintigrams and its use in automated diagnostics. Computer Methods and Programs in Biomedicine 80(1), 47–55, 10 (2005)

[19] Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)