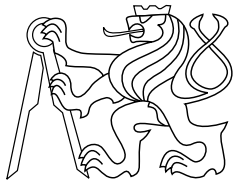




CENTER FOR  
MACHINE PERCEPTION



CZECH TECHNICAL  
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

# Image Texture Analysis of Sonograms in Chronic Inflammations of Thyroid Gland

Daniel Smutek<sup>†</sup>  
Tardi Tjahjadi<sup>‡</sup>  
Radim Šára\*  
Martin Švec\*  
Petr Sucharda<sup>†</sup>  
Štěpán Svačina<sup>†</sup>

smutek@cesnet.cz

<sup>†</sup> 3rd Department of Medicine, 1st Medical Faculty, Charles University  
Prague, Czech Republic

<sup>‡</sup>School of Engineering, University of Warwick, United Kingdom

\*Center for Machine Perception, Czech Technical University  
Prague, Czech Republic

CTU–CMP–2001–15

April 20, 2001

Available at

<ftp://cmp.felk.cvut.cz/pub/cmp/articles/sara/Smutek-TR-2001-15.pdf>

This research has been supported by the Ministry of Health of the Czech Republic under the project NB 5472–3 and from the Ministry of Education of the Czech Republic under projects VS 96049 and MSM 210000012.

**Research Reports of CMP, Czech Technical University in Prague, No. 15, 2001**

Published by

Center for Machine Perception, Department of Cybernetics  
Faculty of Electrical Engineering, Czech Technical University  
Technická 2, 166 27 Prague 6, Czech Republic  
fax +420 2 2435 7385, phone +420 2 2435 7637, www: <http://cmp.felk.cvut.cz>



# Image Texture Analysis of Sonograms in Chronic Inflammations of Thyroid Gland

SMUTEK DANIEL<sup>1)</sup>, TJAHJADI TARDI<sup>2)</sup>, ŠÁRA RADIM<sup>3)</sup>, ŠVEC MARTIN<sup>3)</sup>,  
SUCHARDA PETR<sup>1)</sup>, SVAČINA ŠTĚPÁN<sup>1)</sup>

<sup>1)</sup> 3<sup>rd</sup> Department of Medicine, 1<sup>st</sup> Medical Faculty, Charles University Prague, Czech Republic

<sup>2)</sup> School of Engineering, University of Warwick, United Kingdom

<sup>3)</sup> Centre for Machine Perception, Faculty of Electrical Engineering, Czech Technical University Prague, Czech Republic

## Abstract

*The assessment of sonographic findings of chronic inflamed thyroid tissue in medical praxis is based just on examiner's experience. This paper shows that exact evaluation is possible and that inflamed and healthy tissues can be differentiated by automatic texture analysis of B-mode sonographic images. Classification success rate of 96.6% was achieved with as few as five features automatically selected from a set of 129 spatial and co-occurrence textural characteristics. The rate was evaluated on a validation set of 580 sonograms taken in three types of cross-sections from 39 persons. Two feature selection schemes were proposed. The first was based on compactness and separability criterion computed independently for individual features. The second was based on minimal classification error evaluated on an independent validation set. We observed the first scheme performed significantly worse by achieving only 90.75% recognition success rate and requiring ten features.*

**Keywords:** thyroid gland, Hashimoto's lymphocytic thyroiditis, sonography, texture analysis, feature selection, classification, outlier detection.

## **Introduction**

Hashimoto's lymphocytic thyroiditis, one of the most frequent thyropathies, is a chronic inflammation of the thyroid gland (Wartfsky and Ingbar 1991). It usually leads to hypothyroidism, which causes serious health problems to the patient. The inflammation in the gland changes the structure of the thyroid tissue. These changes are diffuse (i.e., they are not limited to a particular area in the tissue of the gland) affecting the entire gland and can be detected by sonographic imaging (Gooding 1993; Loevner 1996). This disease can also be diagnosed from clinical investigations using immunological, hormonal and metabolic analyses of blood samples, and from cytological investigation using fine needle aspiration biopsy. However, sonographic imaging is the most widely used diagnostic and monitoring tool for this disease.

This work is a first step to quantitative analysis of sonographic images of the thyroid gland. We want to show that it is possible to achieve high classification success rate, when discriminating Hashimoto's lymphocytic thyroiditis and a healthy tissue, with a small number of texture features computed from sonographic images and that the information, which is necessary for classification, is present in sonographic images.

The advantages of using sonographic imaging are its mobility, low cost in comparison with other diagnostic imaging methods (e.g., magnetic resonance imaging) and the direct contact between a patient and a physician. The interaction has been known to have a beneficial effect on the patient and on the diagnostic process. However, the assessment of the diffuse processes is difficult (Simeone et al. 1985; Solbiati et al. 1985) and the diagnosis is made qualitatively from the size of the gland being examined, its perfusion, and the structure and echogenicity of its parenchyma. In making an overall evaluation of a sonogram, the physician uses his/her clinical experience without giving any quantifiable indexes, which are reproducible.

Endeavours for using computers for improving interpretation of medical sonographic images began in the mid-eighties (Kimme-Smith and Jones 1984). These studies also included image texture analysis of thyroid gland (Mailloux et al. 1986), but they were limited only to the comparison of grey-level histograms of different diagnoses. Later works involved mainly localised changes (e.g., nodules, tumours and cysts) in thyroid tissue (Hirning et al. 1989; Morifuji 1989). But there is no published work concerning quantitative analysis of diffuse changes in thyroid gland. However, recent developments in imaging technology considerably improved the quality of sonograms, mainly of subsurface organs such as thyroid gland, and will facilitate a quantitative assessment of the diffuse processes associated with chronic inflammations.

Image texture has two basic properties: the first is concerned with the grey level primitives and the second with the spatial organisation of the grey level primitives (Tuceryan and Jain 1993). The interaction between the two gives the texture characteristics such as fineness, coarseness, smoothness, granulation, randomness, etc. The purpose of texture characterisation is to select a set of measurement or features that will identify the relevant properties of a texture. These features are then used to describe the texture. Numerous features have been proposed and their usefulness evaluated over the years. Tuceryan and Jain (1993) identify five major categories of features for texture identification: statistical, geometrical, structural, model-based and signal processing features. Many of these features represent the texture by second-order correlation structure. Weszka et al. (1976) compared the use of features derived from the Fourier power spectrum, second-order grey-level statistics, co-occurrence statistics

and grey-level run-length statistics. They concluded that the co-occurrence features were the best for classification. The same conclusion was also reported in (Connors and Harlow 1980; Garra et al. 1993; Horng et al. 1996; Mojsilovic et al. 1996; Strand and Taxt 1994; Sujana et al. 1996) for various applications. Based on these conclusions, spatial grey level and co-occurrence features are used in this paper.

The individual features, each of which is a result of the application of an operator to an image, may be grouped as feature vectors. A sufficiently large collection of such vectors including class labels must be available to train a classifier. The trained classifier then compares a vector of unknown class with the training data and assigns the most probable class. Numerous approaches to the classification problem have been reported, see e.g., (Jain et al. 2000) for a comprehensive review. In this paper, the nearest mean classifier combined with majority vote approach is used because of its simplicity.

In this paper the usefulness of a combination of spatial features and second-order statistical texture properties based on co-occurrence matrices for the diagnosis of chronic inflammations in thyroid gland was evaluated. We searched for a combination of smallest possible subset of features that maximises the classification success rate on a validation set.

#### Data acquisition

Twenty-two patients suffering from Hashimoto's lymphocytic thyroiditis (HLT) in the test group and seventeen test persons with no thyroid disease in the control group were examined. The diagnosis of Hashimoto's lymphocytic thyroiditis was confirmed by fine-needle aspiration biopsy, an increased level of antibodies (anti-TPO, anti-TAG) and by clinical investigations. A total of 1170 sonographic images were acquired as shown in Table 1.

	No. of persons examined	No. of transversal cross-sections	No. of longitudinal cross-sections (both lobes)	No. of images
<b>HLT</b>	22	220	440	660
<b>Healthy</b>	17	170	340	510
<b>Total</b>	39	390	780	1170

**Table 1: The set of test data.**

A sonographic imaging system (Toshiba ECCO-CEE, console model SSA-340A, transducer model PLF-805ST at frequency 8 MHz) was used to generate images with an amplitude resolution of 8 bits (256 grey levels) as shown in Figure 1.

Since the changes in the gland are diffuse it is possible to use global textural characteristics within the image region corresponding to the thyroid gland tissue. The automatic segmentation of such regions is complex and is not the subject of this paper. An interactive tool is used instead by a physician to roughly delineate the boundary of the gland, as illustrated by the outline in Figure 2. The maximum number of  $N \times N$  (21x21, 31x31 and 41x41) non-overlapping rectangular windows within the boundaries are then automatically selected as the *texture samples*. Each sample was assigned a label according to the patient diagnosis (HLT, Healthy) as shown in Table 2.

Size of sample	HLT			Healthy		
	41x41	31x31	21x21	41x41	31x31	21x21
Longitudinal - right lobe	1891	3763	9345	1254	2484	5933
Longitudinal - left lobe	1736	3416	8485	990	1995	4784
Longitudinal - both lobes	3627	7179	17830	2244	4479	10717
Transversal	660	1500	4300	409	957	2590
All types of samples	4287	8679	22130	2653	5436	13307

**Table 2: Details of the texture samples.**

## Methods

### Texture features

Image texture features can be computed by combining pixel grey levels in many different ways (Haralick and Shapiro 1993). By transforming the grey levels, it is possible to enhance some image characteristics that are specific to a particular type of texture. Since the standard practice of diagnosing chronic inflammations in thyroid gland from textures in sonograms is performed subjectively, a computer analysis of texture characteristics which are observable by the human visual system are considered adequate. From the psychophysical evidence it is known that human visual system is capable of pre-attentive texture discrimination from first-order to second-order properties, as defined by the moments of texture primitives (Julesz et al. 1973).

We showed in our previous work (Sara et al. 2000), that local texture properties of sonographic B-mode images measured by the first-order texture statistic are independent of the location in the image and thus are suitable for tissue classification. In this paper, 21 spatial features based on the original grey levels of an image and based on four different grey-level transformations (Muzzolini et al. 1994) were investigated. First-order features are prone to severe distortion as a result of varying image contrast and brightness settings on a sonograph. Their significance is limited to a small neighbourhood of a pixel. Therefore, most of the further 108 features used are second-order statistical texture features based on co-occurrence matrices, which incorporate spatial organisation of texture primitives. This is in agreement with the literature where the second-order statistics are reported as the most successful for texture-based classification of sonographic images.

### Spatial features

Some of the 21 spatial features are based on the original pixel grey levels  $p_{i,j}$ , while others  $p^m_{i,j}$ , are based on transformations of the grey levels, where  $i,j$  denotes the image coordinates of a pixel and  $m$  denotes a transformation. These features were suggested in (Muzzolini et al. 1994) and are summarised in this section. Four grey-level transformations obtained from each of  $S$  samples of  $N \times N$  pixels were used and are defined as follows:

#### 1) gradient magnitude

$$p_{i,j}^{(1)} = |p_{i,j} - p_{i+1,j}| + |p_{i,j} - p_{i,j+1}|$$

2) difference from sample mean

$$p_{i,j}^{(2)} = |p_{i,j} - \bar{p}| \text{ where } \bar{p} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N p_{i,j}$$

3) horizontal curvature

$$p_{i,j}^{(3)} = \left| p_{i,j} - \frac{p_{i-1,j} + p_{i+1,j}}{2} \right|$$

4) vertical curvature

$$p_{i,j}^{(4)} = \left| p_{i,j} - \frac{p_{i,j-1} + p_{i,j+1}}{2} \right|$$

5) original pixel grey levels

$$p_{i,j}^{(5)} = p_{i,j}$$

The Kolmogorov-Smirnov distance (German et al. 1990; Muzzolini et al. 1993) between  $H_i(p^{(m)})$  and  $\bar{H}(p^{(m)})$  is used to derive features  $f_1, \dots, f_5$ , from the transformations  $p^{(m)}$ . The  $H_i(p^{(m)})$  is an estimate of the cumulative distribution function for  $p^{(m)}$  computed from  $N \times N$  sample  $i$  by histogramming and  $\bar{H}(p^{(m)})$  is the robust estimate of the cumulative distribution function mean for  $p^{(m)}$  computed from  $H_i(p^{(m)})$  over all samples  $i$  as follows:

$$\bar{H}(p^{(m)}) = LMS \{H_i(p^{(m)}), i = 1, 2, \dots, S\}$$

The *LMS* is used as a robust statistics instead of a non-robust mean to suppress the influence of outlying values (Rousseeuw and Leroy 1987). *LMS* computes a value  $M$  and a range  $m_T$  for a data set, such that the  $(M - m_T, M + m_T)$  is the shortest interval containing 50% of the original data. It is a common practice to set the estimate of standard deviation  $r$  to the value of  $2.5 \times 1.4826 m_T$  for the case of normal errors (Yang and Levine 1992). Points in the range  $M \pm r$  are called *inliers* and the remaining points are considered as *outliers*. Inliers fall within 98.7% of the samples in a Gaussian distribution.

The Euclidean distance from  $(f_1, \dots, f_5)$  to their mean and median, respectively, are used to compute features  $f_6$  and  $f_7$  as follows:

$$f_6 = \sqrt{(\bar{f}_1 - f_1)^2 + (\bar{f}_2 - f_2)^2 + (\bar{f}_3 - f_3)^2 + (\bar{f}_4 - f_4)^2 + (\bar{f}_5 - f_5)^2},$$

$$f_7 = \sqrt{(\hat{f}_1 - f_1)^2 + (\hat{f}_2 - f_2)^2 + (\hat{f}_3 - f_3)^2 + (\hat{f}_4 - f_4)^2 + (\hat{f}_5 - f_5)^2},$$

where  $\bar{f}_i$  is the mean of feature  $f_i$  for a sample, and  $\hat{f}_i$  is the median of feature  $f_i, i = 1, 2, \dots, 5$ .

Features  $f_8, \dots, f_{12}$  are derived from the transformations  $p^{(m)}$  just like features  $(f_1, \dots, f_5)$  except that the average deviation (*AD*) of the pixel grey level  $p_{i,j}^{(m)}$  is used as the measure, where

$$AD(p^{(m)}) = \frac{1}{N^2} \sum_i \sum_j |p_{i,j}^{(m)} - \bar{p}^{(m)}|, \quad m = 1, 2, \dots, 5.$$

Features  $f_{13}$  and  $f_{14}$  are based on the Euclidean distance from  $(f_8, \dots, f_{12})$  to their mean and median, respectively. They are defined exactly the same way as the features  $f_6$  and  $f_7$  with the exception that the subscripts  $(1, \dots, 5)$  are replaced with subscripts  $(8, \dots, 12)$ .

Features  $f_{15}, \dots, f_{19}$  are based on the transformations  $p^{(m)}$ , just like features  $(f_1, \dots, f_5)$ , except that the standard deviation ( $SD$ ) of the pixel grey level  $p_{i,j}^{(m)}$  is used as the measure, where

$$SD(p^{(m)}) = \sqrt{Var(p^{(m)})},$$

$$Var(p^{(m)}) = \frac{1}{N^2} \sum_i \sum_j (p_{i,j}^{(m)} - \bar{p}^{(m)})^2, \quad m = 1, 2, \dots, 5$$

Features  $f_{20}$  and  $f_{21}$  are based on the Euclidean distance from  $(f_{15}, \dots, f_{19})$  to the mean and median of  $(f_{15}, \dots, f_{19})$ .

### Co-occurrence matrix features

Co-occurrence matrices can be used to obtain texture features (Haralick and Shapiro 1993; Peckinpugh 1991). For each  $N \times N$  texture sample  $W$  taken from image  $I$ , a set of grey level co-occurrence matrices  $C_d(i, j)$  is calculated for a given separation vector  $\vec{d}$  as follows:

$$C_{\vec{d}}(i, j) = \frac{1}{(N-a)(N-b)} \text{card} \left\{ \{ \vec{r}, \vec{r} + \vec{d} \} : \vec{r}, \vec{r} + \vec{d} \in W \text{ and } I(\vec{r}) = i \text{ and } I(\vec{r} + \vec{d}) = j \right\}$$

where  $\vec{d} = (a, b)$ ,  $I(\vec{r})$  is the grey level of pixel  $\vec{r}$ , from the interval of  $0, 1, \dots, G-1$ . The image resolution of  $G = 256$  was used, the  $\text{card } X$  is the cardinality of set  $X$ . The elements of  $C_d$  represent the frequencies of occurrence of different grey level combinations at a distance  $\vec{d}$ . In this paper, nine Haralick texture features as defined in Table 3 (Haralick and Shapiro 1993) were used.

No (h)	Feature name	Definition
1	Cluster tendency	$\sum_i \sum_j (i - \mu_i + j - \mu_j)^2 C_{\vec{d}}(i, j)$
2	Texture entropy	$-\sum_i \sum_j C_{\vec{d}}(i, j) \log C_{\vec{d}}(i, j)$
3	Texture contrast	$\sum_i \sum_j  i - j  C_{\vec{d}}(i, j)$
4	Texture correlation	$\frac{\sum_i \sum_j (i - \mu_i)(j - \mu_j) C_{\vec{d}}(i, j)}{\sqrt{\text{var}(i)\text{var}(j)}}$
5	Texture homogeneity	$\sum_i \sum_j \frac{C_{\vec{d}}(i, j)}{1 +  i - j }$
6	Inverse difference moment	$\sum_{i,i \neq j} \sum_{j,i \neq j} \frac{C_{\vec{d}}(i, j)}{ i - j }$
7	Maximum probability	$\max_i \max_j C_{\vec{d}}(i, j)$
8	Probability of run length of 2	$\sum_i \frac{(C_i - C_{\vec{d}}(i, i))^2 C_{\vec{d}}(i, i)}{(C_{\vec{d}}(i))^2}$
9	Uniformity of energy	$\sum_i \sum_j (C_{\vec{d}}(i, j))^2$
where $\mu_i = \sum_i \sum_j i C_{\vec{d}}(i, j)$ ; $\mu_j = \sum_i \sum_j j C_{\vec{d}}(i, j)$ ; $C_i = \sum_j C_{\vec{d}}(i, j)$ $\text{var}(i) = \sum_i \sum_j (i - \mu_i)^2 C_{\vec{d}}(i, j)$ ; $\text{var}(j) = \sum_i \sum_j (j - \mu_j)^2 C_{\vec{d}}(i, j)$ $i, j = 0, 1, \dots, G - 1$ .		

**Table 3: Haralick texture features (Haralick and Shapiro 1993).**

Twelve vectors  $\vec{d} = (1,0); (2,0); (3,0); (4,0); (1,1); (2,2); (3,3); (4,4); (0,1); (0,2); (0,3); (0,4)$  as shown in Table 4 were used in the experiments, resulting in twelve grey level co-occurrence matrices for each texture sample of the following sizes: 21x21, 31x31 and 41x41 pixels.

<b>d</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>0</b>		9	10	11	12
<b>1</b>	1	5			
<b>2</b>	2		6		
<b>3</b>	3			7	
<b>4</b>	4				8

**Table 4: The directions of the various vectors  $\vec{d}$ .**

Thus  $D = 108$  co-occurrence matrix features ( $f111-f1129$ ) were generated for each of the sample sizes. These are denoted according to following notation: " $f1dh$ ", where  $d$  and  $h$  are the number of the vector and Haralick feature respectively. For an example,  $f195$  is texture homogeneity for vector  $\vec{d}=(0,1)$ .

## Feature selection

The purpose of feature selection is to reduce the textural description from  $D$  to  $d$  dimensions, where  $d < D$ . Each sample of the classes (HLT, Healthy) can be represented in terms of  $d$  features and be viewed as a vector in  $d$ -dimensional space.

We consider two approaches to feature selection. The goal in the first approach is to select those features that allow feature vectors belonging to any given class to occupy compact and disjoint (separable) regions in  $d$ -dimensional feature space. Compact disjoint clusters can then be well separated by decision boundaries used by a classifier. The goal in the second approach is to select a subset of features that minimise the classification error in a validation set. The validation set must not overlap with the training set used to train the classifier. The first approach is independent of the choice of classifier but is too conservative in what is considered a good feature combination. The second approach is based on a direct classification error minimisation and requires a specific classifier choice. For simple classifiers like Euclidean distance from a class mean, both feature selection schemes should achieve similar results. However, we found that it is not the case as discussed in Discussion Section.

### Compactness and separability criterion

In the first feature selection scheme (Scheme 1), features are ranked based on compactness and separability of the class representations in feature space similar to that used in (Muzzolini et al. 1994). Compactness is measured as the inverse of the number of *LMS* outliers, and separability is measured as the inverse of class overlap between two classes  $C_1$  and  $C_2$ . The rank of feature  $f_i$  is computed as follows:

$$R(f_i) = \sum_{j=1}^2 outliers(f_i, C_j) + w \cdot overlap(C_1, C_2; f_i),$$

where  $outliers(f_i, C_j)$  is the number of outliers (as determined by the *LMS* in Spatial Features Section) for feature  $f_i$  for the texture class  $C_j$ ; and  $overlap(C_1, C_2; f_i)$  is the number of samples in the overlapping region between texture class  $C_1$  and  $C_2$  for feature  $f_i$  (as illustrated in Figure 3), normalised by the sum of the number of inliers for both classes (see Spatial Features Section). This way of computing separability differs from that in (Muzzolini et al. 1994). The  $w$  is a weighting factor and it is chosen so that either the outliers or between-class overlap dominates the rank  $R$ . We chose  $w \in \{0.0001, 0.001, 0.1, 1, 10, 100, 1000, 10000\}$  in our experiment.

For each  $w$ , features are ranked from lowest  $R$  (the most significant feature) to highest  $R$  (the least significant feature). From each ranking, the top ten features are then used in turn to compose feature vectors by considering the pair of the best two features, the triplet of the best three features, etc. up to a set of the best 10 features. In this way 80 feature vectors are obtained. Note that the separability in each individual feature is required.

### Minimal classification error criterion

In the second feature selection scheme (Scheme 2), the quality of each selected feature vector  $\vec{v}$  is defined as the classification success rate  $J(\vec{v})$  obtained with  $\vec{v}$  and computed as the percentage of correct classification on a validation set. If the criterion function  $J$  always increases with increasing dimension  $d$  of  $\vec{v}$ , then the optimal feature selection method avoids exhaustive search and may be implemented as a depth-first search algorithm (Jain et al. 2000).

In an independent validation set, however, the criterion function does not show these properties and the search algorithm must therefore be more general (Bishop 1997).

We first perform exhaustive search to find the fifty combinations of two features with the highest criterion function. The following modified depth-first algorithm was then applied to these 50 pairs.

Each of the 50 pairs of features was successively used as the basis ( $\vec{v}$ ) for further feature combinations as follows:

- 1) Add a feature  $f_i$  to the basis ( $\vec{v}$ )  $\Rightarrow$  ( $\vec{v}, f_i$ ).
- 2) Compute and store  $J(\vec{v}, f_i)$ .
- 3) Repeat steps 1 and 2 until all features have been used as  $f_i$ .
- 4) The ( $\vec{v}, f_i$ ) with the highest  $J(\vec{v}, f_i)$  is used as the next  $\vec{v}$ , and repeat steps 1-3 until the new highest  $J(\vec{v}, f_i) \leq J(\vec{v})$

The fifty best combinations of features found in Step 4 are then used in the experiment.

## Classification

All the acquired samples were randomly divided into two non-overlapping sets referred to as the training and the validation sets. The sets are approximately the same size, and all samples from one sonographic image must be in only one of these sets. The training set is used to train a classifier and the validation set is used to evaluate the classification success rate  $J(\vec{v})$ .

The minimum Euclidean distance classifier (Devroye et al. 1996) was used for the classification of the validation set. This classifier was selected as the simplest one to balance the trade-off between classifier complexity and its generalisation ability (Bishop 1997). The classifier is trained by computing the mean value of each selected feature from the training set for both classes (HLT and Healthy) independently. The two vectors of mean values then represent the class prototype. The classification works in two stages. Individual texture samples (e.g., from 21x21 windows) of an image are classified by finding the closest class prototype in the first stage, and the entire sonographic image is classified in the second stage by majority voting. The sonographic image is thus assigned the class  $C$  (HLT, Healthy), which corresponds to the class of most of its samples as illustrated in Figure 4.

In both feature selection schemes, the classification success rate is evaluated by classifying the validation set and computing the percentage of correct classification. This is repeated three times for different divisions of the sample set to the training and validation sets using random indexing. The final reported success rate is the average over the three values.

## Results

A total of 7,287,468 feature values (129 features for each of 56,492 samples) were computed and the different combinations of them were used for classification. Table 5 shows the results of texture classification for various types of cross sections using the features selected by compactness and separability criterion (i.e., Scheme 1). The table includes only three  $n$ -tuples of features (and the value of  $w$  which selected them) with the highest classification success rate for each cross-section and sample size. The best classification rate was 90.75% (shown in bold). This was achieved using a sample size of 41x41 pixels on longitudinal cross-section of

the left lobe, and using the following: Kolmogorov-Smirnov distance based on difference from sample mean; Euclidean distance from means of  $f_1 \dots f_5$ ; maximum probability for co-occurrence matrices computed for separation vectors  $\vec{d} \in \{(1,0), (4,0), (1,1), (4,4), (0,3), (0,4)\}$ ; and inverse difference moment computed for  $\vec{d} \in \{(0,2), (0,3)\}$ . Table 6 shows the properties (number of outliers and overlap) of this best combination of features. The overlap of these features is significant, more than 60%. This means that individual features are not sufficiently separated in the sense of this scheme. Features  $f1106, f1116$  and  $f6$  are very compact (i.e., with the smallest number of outliers), where as features  $f1117, f1127$  are less compact (i.e., with the greatest number of outliers).

Size of sample and type of cross-section	Success rate (%)	Features used for classification	w
41 x 41 right lobe, longitudinal	90.36	f2,f117,f157	100
	89.84	f2,f117	100
	82.01	f10,f124	0.0001
41 x 41 left lobe, longitudinal	<b>90.75</b>	<b>f2,f6,f117,f147,f157,f187,f1106,f1116,f1117,f1127</b>	1000
	85.57	f2,f157	100
	76.28	f124,f193,f195,f1103,f1105,f1113,f1115	1
41 x 41 both lobes, longitudinal	88.12	f2,f5,f6,f137,f157,f1127	100
	87.09	f2,f137	1000
	84.62	f2,f5	100
41 x 41 transversal	85.34	f153,f155,f193,f195,f1105	1
	85.27	f153,f193,f195,f1105	1
	84.56	f153,f195,f1105	1
41 x 41 all cross-sections	85.35	f2,f5,f6,f137,f157	100
	84.72	f2,f5,f137,f157,f1127	1000
	84.65	f2,f5	100
31 x 31 right lobe, longitudinal	81.46	f10,f113,f115,f124,f164	10
	81.12	f10,f124	10
	80.95	f2,f18,f115,f116,f153,f155,f156,f187	100
31 x 31 left lobe, longitudinal	76.58	f164,f195,f1105,f1113,f1115	0.0001
	73.85	f1105,f1125	10
	73.33	f116,f156,f1106,f1107,f1116,f1117	100
31 x 31 both lobes, longitudinal	82.27	f10,f124	10
	76.39	f115,f116,f153,f155,f156,f1106	100
	76.21	f2,f116,f187,f1107,f1117	1000
31 x 31 transversal	85.64	f193,f195	10
	84.79	f154,f163,f164,f196,f1103,f1113,f1115,f1123,f1125	1
	84.10	f114,f154,f164,f174,f196,f1123,f1124,f1125	0.0001
31 x 31 all cross-sections	78.69	f115,f116,f153,f155,f156,f196	100
	78.52	f115,f116,f153	100
	77.09	f10,f124,f164	10
21 x 21 right lobe, longitudinal	85.71	f113,f153,f155,f187	100
	84.35	f113,f155,f187	100
	83.84	f155,f187	100
21 x 21 left lobe, longitudinal	81.37	f124,f134,f164	0.0001
	77.78	f113,f115,f1113	10
	76.92	f124,f164	1
21 x 21 both lobes, longitudinal	82.01	f113,f115,f155	100
	80.48	f124,f164	1
	79.71	f113,f115	100
21 x 21 transversal	88.38	f195,f196,f1105,f1125	10
	87.86	f124,f134,f164,f174,f184,f196,f1124	0.1
	87.52	f124,f164,f174,f184,f196,f1124	1
21 x 21 all cross-sections	83.22	f113,f115	100
	77.99	f10,f124,f134,f164,f184,f196,f1124	10
	77.47	f124,f164	0.0001

**Table 5: The three best classification rates for each type of cross-sections using features selected by compactness and separability criterion (Scheme 1).**

Feature	% of total outliers	% of overlap
<b>f2</b>	15.04	60.97
<b>f6</b>	7.67	77.75
<b>f117</b>	20.76	76.30
<b>f147</b>	23.99	64.82
<b>f157</b>	23.18	63.56
<b>f187</b>	19.15	71.87
<b>f1106</b>	4.22	76.64
<b>f1116</b>	4.84	77.49
<b>f1117</b>	26.60	71.56
<b>f1127</b>	26.23	69.67

**Table 6: The percentage of outliers and overlapping regions of the set of features (chosen using Scheme 1) which gives the best classification rate (samples size 41x41 pixels, longitudinal cross-section of left lobe, classification success rate 90.75%).**

Table 7 shows the classification results for various types of cross sections using the features selected by minimal classification error criterion (i.e., Scheme 2). The best classification rate was 96.6% (shown in bold), which is significantly better than that achieved with Scheme 1. Furthermore, the number of features used for the best success rate is half that used in Scheme 1. The classification was achieved using a sample size of 41x41 pixels on longitudinal cross-section of the right lobe, and using the following: uniformity of energy for the co-occurrence matrices with vector  $\vec{d}=(2,0)$ ; Kolmogorov-Smirnov distance based on difference from sample mean; texture contrast for co-occurrence matrices with vector  $\vec{d}=(4,4)$ ; Euclidean distance from means of  $f_1...f_5$ ; and probability of run length of 2 for co-occurrence matrices with vector  $\vec{d}=(4,4)$ . The order of features (in each row) in Table 7 is the same as the order in which they were selected by the scheme (i.e., it corresponds with the importance of the features for classification).

Size of sample and type of cross-section	Success rate (%)	Features used for classification
41 x 41 right lobe, longitudinal	<b>96.60</b>	<b>f129,f2,f183,f6,f188</b>
	96.26	f15,f2,f6,f17
	96.09	f123,f2,f6,f11,f1124,f1128
41 x 41 left lobe, longitudinal	96.39	f183,f2,f6,f141,f156,f1122
	96.39	f184,f2,f6,f191,f1102
	96.39	f182,f2,f178,f6,f1127,f169
41 x 41 both lobes, longitudinal	95.56	f154,f2,f6,f15
	95.30	f18,f2,f6,f1111,f174,f137
	95.30	f1104,f2,f6,f1106,f178,f132
41 x 41 transversal	92.19	f113,f125,f194,f161
	<i>91.84</i>	<i>f125,f194</i>
	91.67	f113,f194,f123,f1105,f8
41 x 41 all cross-sections	92.48	f195,f2,f6,f191,f4,f132
	92.20	f1125,f2,f6,f9,f1127,f4,f194,f167
	92.14	f2,f4,f6,f183
31 x 31 right lobe, longitudinal	87.59	f148,f17,f194,f135,f171
	87.41	f113,f115,f194,f179
	87.24	f125,f194,f113,f161
31 x 31 left lobe, longitudinal	83.76	f114,f124,f139,f134,f115
	83.42	f1104,f124,f19,f1129
	83.25	f121,f124,f1119,f154,f19,f11,f1129,f1127
31 x 31 both lobes, longitudinal	84.31	f113,f193,f124,f17,f19,f21,f157
	84.14	f124,f127,f19,f121
	84.06	f124,f131,f19,f145,f111,f141
31 x 31 transversal	92.14	f153,f155,f1105,f194,f128,f159,f1129,f169,f161,f127
	91.79	f113,f8,f155,f194,f9,f178,f119
	91.28	f1103,f113,f157,f194,f178,f169
31 x 31 all cross-sections	87.18	f2,f4,f184,f139,f172
	85.30	f113,f189,f13,f194,f115
	84.96	f113,f9,f194,f1111,f179
21 x 21 right lobe, longitudinal	89.63	f15,f185,f124,f19,f119,f151
	88.78	f113,f173,f194,f115
	88.27	f123,f194,f13,f135
21 x 21 left lobe, longitudinal	84.44	f1106,f6,f124,f8,f1122,f162
	84.44	f124,f143,f16,f1128,f15,f1123
	83.25	f124,f19,f145,f183
21 x 21 both lobes, longitudinal	87.64	f116,f6,f196,f171
	85.00	f124,f16,f197,f15,f113,f199,f115,f1121
	84.83	f1125,f124,f19,f187
21 x 21 transversal	95.04	f138,f196,f6,f116,f156,f1124
	94.87	f156,f6,f136,f196,f162,f135,f148
	93.16	f10,f153,f147,f127,f193
21 x 21 all cross-sections	86.01	f124,f19,f169
	85.55	f123,f194,f119,f177
	85.44	f1105,f115,f153,f194,f119

**Table 7: The three best classification rates for each type of cross-sections based on features selected by minimal classification error criterion (Scheme 2).**

In both feature selection schemes, a better classification success rate is obtained for larger size of texture samples. This could be attributed to the higher stability of features in larger samples in describing the spatial organisation of texture primitives.

The differences in classification success rates between longitudinal right and longitudinal left cross-sections are not significant, and the results of longitudinal cross-sections of both lobes are similar to the mean values of both individual longitudinal cross-sections. So it is concluded that it is not important whether the sonogram of longitudinal cross-section is acquired from the right or left lobe. In fact, in practice the findings in the sonographic examination in both lobes are identical. This also confirms that we are dealing with a diffuse process, which affects the whole tissue and that the method used for classification is stable without any artefacts.

The image area of displayed thyroid tissue is much smaller in transversal cross-sections than in longitudinal ones, cf. Figure 2. This means that a smaller number of texture samples fit within the boundary of thyroid gland in transversal cross-section as compared to the other cross-section. This could be the reason for the differences in the results obtained using both cross-sections.

We observed better success of small samples in transversal cross-sections as compared to larger samples in the same cross-section type. We attributed this to the fact that larger texture samples do not cover the small areas of the transversal cross-sections well, and thus there is less information to be used for recognition.

The mean values and standard deviations of features in the best combination are shown in Figure 5. In Figure 6 we show the feature space for the best pair of features selected using Scheme 2. It was f125, f194 (i.e., texture homogeneity for co-occurrence matrices with vector  $\vec{d}=(2,0)$  and texture correlation for co-occurrence matrices with vector  $\vec{d}=(0,1)$ ) with classification success rate 91.84%, achieved using sample size of 41x41 pixels and transversal cross-section (indicated in italic in Table 7). Note the class clusters are quite well separated even though individual features are not. The significance of this will be discussed in the Discussion Section.

## **Discussion**

The assessment of sonographic findings of chronic inflamed thyroid tissue in present medical praxis is based just on physician's clinical experience - the diagnosis is made only qualitatively. We showed that it is possible to gain exact, quantitative characteristics of the thyroid gland and to use it for tissue classification. The most difficult part of automatic classifier design is the selection of a small set of measurements (features) that are the most suitable for the classification task. We focused in this paper on two approaches to this selection. The compactness and separability criterion was originally used in (Muzzolini et al. 1994). But even with the improved Scheme 1 proposed in this paper, the number of selected features are more than that selected by the minimal classification error criterion (Scheme 2), which also achieved higher success rate. The best combination selected by Scheme 1 contains all 10 features, 5 features more than those selected by Scheme 2. The criterion in Scheme 1 requires separability in each individual feature. As demonstrated in Figure 6, this is a misleading criterion since the separability in higher-dimensioned space can be quite good even if individual features are not separated at all. Our experimental results show Scheme 2

achieved significantly better results with fewer features than Scheme 1. Therefore, this method should be used for feature selection in sonographic texture classification.

In Scheme 1, the combination of features selected with a higher weighting factor  $w$  were generally more successful than a combination acquired with a lower  $w$ . This means the selected feature set was based more on separability (between-class overlap) than on compactness (within-class outliers). This finding is also in conformity with the medical sonographic examination where a physician's decision depends mostly on differences between individual diagnoses than on tiny inaccuracies in the sonographic image.

The properties of a sonographic image are dependent on local changes of density of the tissue. Since the shape of thyroid gland follicles is spherical, their properties should be isotropic. Therefore the differences between longitudinal and transversal cross-sections observed in our experiments are not caused by different properties of the texture in individual cutting planes. However it is also expected that there are differences among individual sonograms, but these typically occur in all cross-sections independently of the cutting plane. These differences are caused by camber of patient's neck, angle and press of transducer, amount of gel used, varying distances of the tissue from transducer, and noise and artefacts in sonographic images. These differences cannot be avoided because they are caused by the physical nature of sonographic examining methods. In order to reduce the errors in texture classification we used the robust statistics, by assuming that these errors will correspond to outliers and are thus removed.

## **Conclusions**

This paper shows that chronic inflamed thyroid tissue can be differentiated from tissue of healthy thyroid gland by means of image texture analysis of sonograms with a classification success rate of 96.6%. Since this can be obtained with a small number of features (five), the result confirms that the information related to diagnosis can be extracted well from sonographic images of subsurface organs. This is the reason we propose that features with the best classification results are also suitable as quantitative indicators of sonographic examination of thyroid gland. These enable reproducibility of the sonographic diagnosis, facilitate assesment of changes of the diasease and make possible the comparison of different physicians' findings of a sonographic examination. To developpe such a diagnostic system is the subject of an ongoing research.

Our results show that feature selection for sonographic texture classification should be based on minimal classification error criterion computed on a validation set, since it achieves significantly better classification success rate than selection method based on class separability.

Our experiment further confirmed that differences in textural properties of sonograms between longitudinal and transversal cross-sections are negligible. However, longitudinal cross-sections provide greater amount of image data from a larger contiguous area of the gland tissue, therefore they are more useful for automatic texture analysis. This is particularly important for large sample sizes (that achieve better classification success rate).

It is known from literature that among chronic inflammations of thyroid gland several sub-groups can be distinguished (Wartfsky and Ingbar 1991). This suggests that texture features in the HLT class should exhibit larger within-class variations than in the Healthy class. Our

experiment shows that three of the best five features have this property. The experiment however cannot prove the existence of internal structure within the HLT class. To design a targeted experiment to prove or disprove this hypothesis is a topic for our future research.

### ***Acknowledgement***

This research has been supported by grants from the Ministry of Health of the Czech Republic under the project MZ ČR NB 5472-3 and from the Ministry of Education of the Czech Republic under projects MŠ ČR VS 96049 and J04/98:210000012. The authors would like to thank the University of Warwick (UK) for providing a MiMI research fellowship which supported the international collaboration. The authors are also indebted to Dr. Zdeňka Límanová for her careful performing cytological investigation of the thyroid tissue of patients used in our research.

## References

- Bishop CM.: Neural Networks for Pattern Recognition. Oxford: University Press, 1997.
- Connors RW, Harlow CA. A theoretical Comparative Study of Texture Algorithms. *IEEE T Pattern Anal* 1980;2: 204-221.
- Devroye L, Györfi L, Lugosi G. A Probabilistic Theory of Pattern Recognition. New York: Springer Verlag, 1996.
- Garra B, Krasner B, Horii S et al. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrason Imaging* 1993;15(4):267-285.
- German D, German S, Graffigne C, Dong P. Boundary detection by constrained optimization, *IEEE T Pattern Anal* 1990;12:609-638.
- Gooding GA. Sonography of the thyroid and parathyroid. *Radiol Clin North Am* 1993;31(5):967-989.
- Haralick RM, Shapiro LG. Computer and Robot Vision. Addison-Wesley Publishing Company, 1993: volume 1,453-507.
- Hirning T, Zuna I, Schlaps D et al. Quantification and classification of echographic findings in the thyroid gland by computerized B-mode texture analysis. *Eur J Radiol* 1989;9(4):244-247.
- Hornig MH, Sun YN, Lin XZ: Texture feature coding method for classification of liver sonography. In: Buxton B, Cipolla R, ed. Proc. Fourth European Conf. on Computer Vision. ECCV '96, Berlin, 1996;1:209-218.
- Jain RK, Duin RPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE T Pattern Anal* 2000;22(1):4-37.
- Julesz B, Gilbert EN, Shepp LA, Frish HL. Inability of humans to discriminate between visual textures that agree in second-order statistics - revisited. *Perception* 1973;2:391-405.
- Kimme-Smith C, Jones JP. The relative effects of system parameters on texture in gray-scale ultrasonograms. *Ultrasound Med Biol* 1984;10(3):299-307.
- Loevner LA. Imaging of the thyroid gland, *Semin Ultrasound CT* 1996;17(6):539-562.
- Mailloux GE, Bertrand M, Stampfler R, Ethier S. Computer analysis of echographic textures in Hashimoto disease of the thyroid, *J Clin Ultrasound* 1986;14(7):521-527.
- Mojsilovic A, Popovic M, Sevic D. Classification of the ultrasound liver images with the 2N multiplied by 1-D wavelet transform. In: Proc. 1996 IEEE Int. Conf. on Image Processing, ICIP'96, Los Alamitos, CA, USA, 1996;1: 367-370.
- Morifuji H. Analysis of ultrasound B-mode histogram in thyroid tumors, *Nippon Geka Gakkai Zasshi* 1989;90(2):210-221.
- Muzzolini R, Yang YH, Pierson R. Multiresolution texture segmentation with application to diagnostic ultrasound images, *IEEE T Med Imaging* 1993;12(1):108-123.
- Muzzolini R, Yang YH, Pierson R. Texture characterization using robust statistics. *Pattern Recogn* 1994;27(1):119-134.
- Peckinpugh SH: An improved method for computing gray-level co-occurrence matrix based texture measures. *Graph Model Im Proc* 1991;53:574-580.
- Rousseeuw PJ, Leroy AM. Robust Regression and Outlier Detection. New York: John Wiley and Sons, Inc., 1987.
- Sara R, Svec M, Smutek D, Sucharda P, Svacina S. Texture Analysis of Sonographic Images for Diffusion Processes Classification in Thyroid Gland Parenchyma. In: Proc. Conf. Analysis of Biomedical Signals and Images. Brno University of Technology VUTUM Press, 2000:210-212.
- Simeone FJ, Daniel GH, Müller PR et al. High-resolution real-time sonography. *Radiology* 1985;155:431-439.
- Solbiati L, Volterrani L, Rizzato G et al. The Thyroid Gland with Low-Uptake Lesions: Evaluation by Ultrasound. *Radiology* 1985;155:187-196.
- Strand J, Taxt T. Local Frequency Features for Texture Classification, *Pattern Recogn* 1994;27(10):1397-1406.
- Sujana H, Swarnamani S, Suresh S. Application of artificial neural networks for the classification of liver lesions by image texture parameters. *Ultrasound Med Biol* 1996;22(9):1177-1181.
- Tuceryan M, Jain AK. Texture Analysis. In Chen CH, Pau LF, Wang PSP, ed. Handbook Pattern Recognition and Computer Vision, Singapore, World Scientific, 1993:235-276.
- Wartfsky L, Ingbar SH. Disease of the Thyroid. In Harrison's Principles of Internal Medicine, 12th Ed. New York, McGraw-Hill Inc., 1991:1712.
- Weszka JS, Dyer CR, Rosenfeld A. A comparative study of texture measures for terrain classification, *IEEE T Syst Man Cyb* 1976;6:269-285.
- Yang YH, Levine MD. The background primal sketch: An approach for tracking moving objects, *Mach Vision Appl* 1992;5:17-34.

## Figure Legends

Figure 1: Sonographic images of HLT and healthy thyroid gland. The gland in the longitudinal cross-sections is the area of uniform texture in the upper half of the image. In the transversal cross-section, it is a two-lobe formation similar to spectacles, also in the upper half of the image.

Figure 2: Sonographic images (longitudinal cross-section on left, transversal on right) with drawn boundary of thyroid gland and texture samples (rectangular windows of 21x21 pixels).

Figure 3: The overlapping region between texture class  $C_1$  and  $C_2$  (Muzzolini et al. 1994). The overlap in the shaded area results when inliers from texture class  $C_1$  (HLT) intersects inliers from texture class  $C_2$  (Healthy).

Figure 4: The classification of an image according to the majority of samples classified by a minimum distance classifier (MDC).

Figure 5: Vertical bar chart of the features, which give the best classification rate, showing mean value, standard deviations, maximum and minimum (samples of size 41x41 pixels, longitudinal cross-section of right lobe, classification success rate 96.6%).

Figure 6: Distribution of all measurements in feature space for the best pair (f125 – 2.62% outliers, 71.8% overlap; f194 – 2.9% outliers, 91.23% overlap) of features - classification success rate 91.84%.

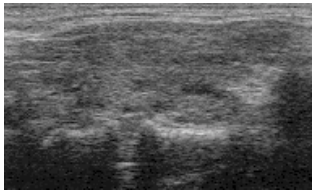
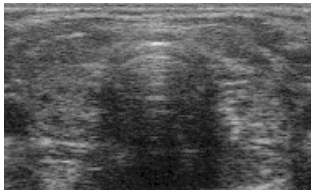
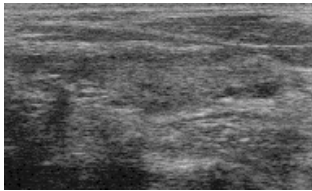
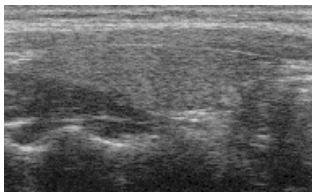
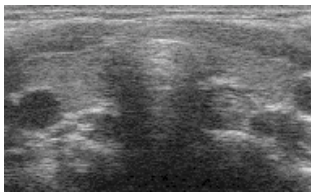
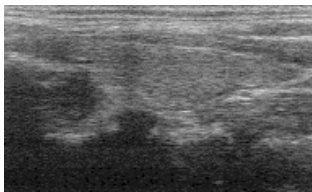
	Longitudinal cross-section of the right lobe	Transversal cross-section	Longitudinal cross-section of the left lobe
HLT			
Healthy			

Figure 1

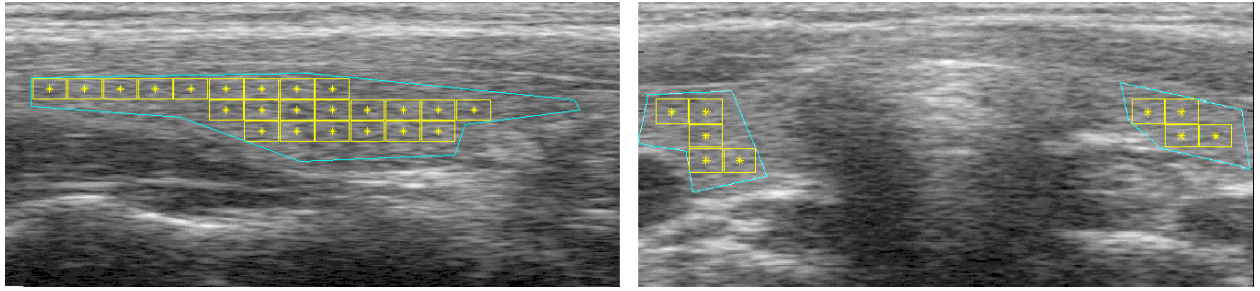


Figure 2

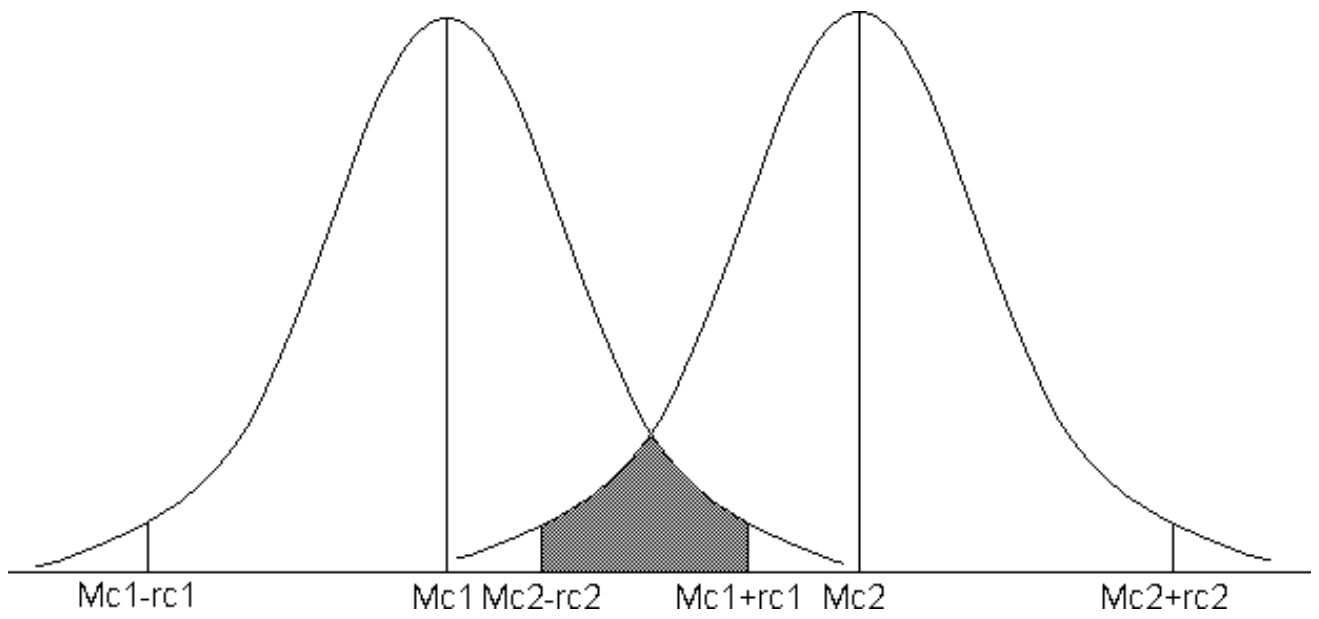


Figure 3

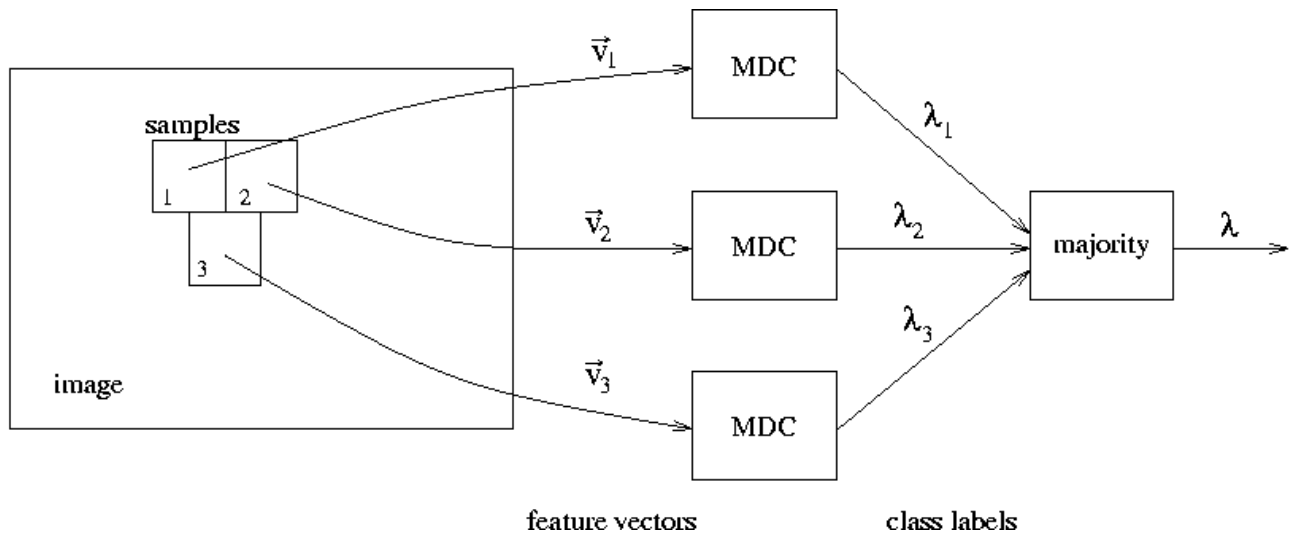


Figure 4

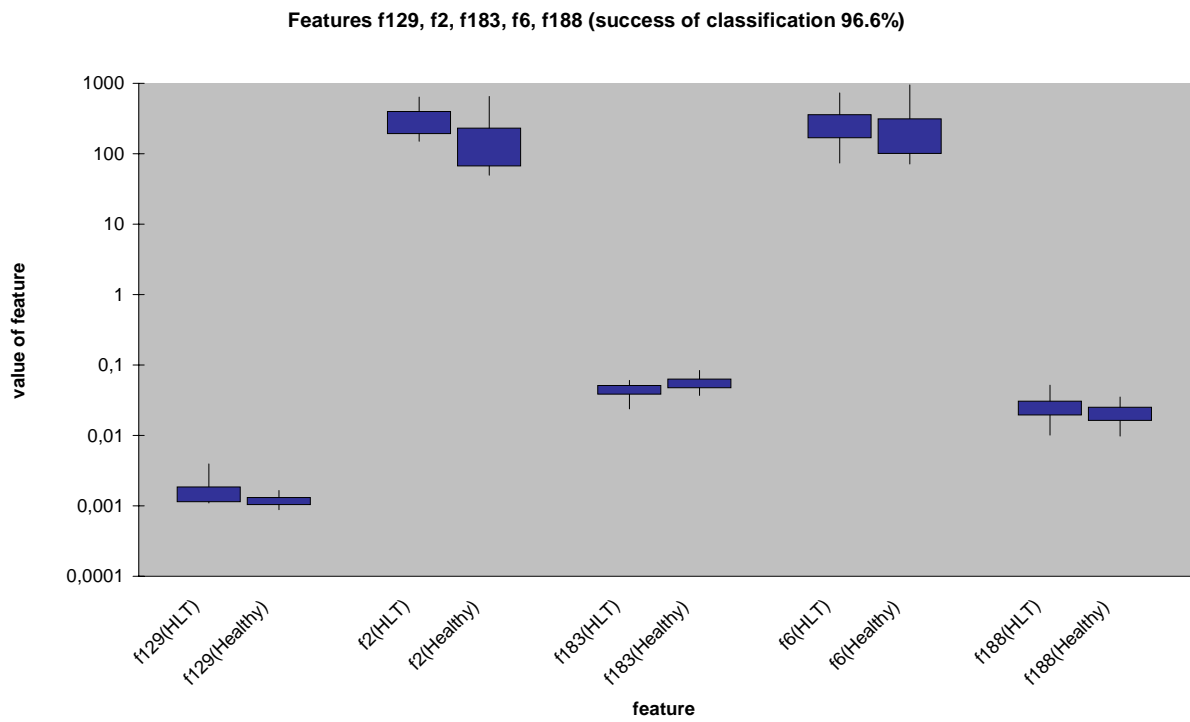


Figure 5

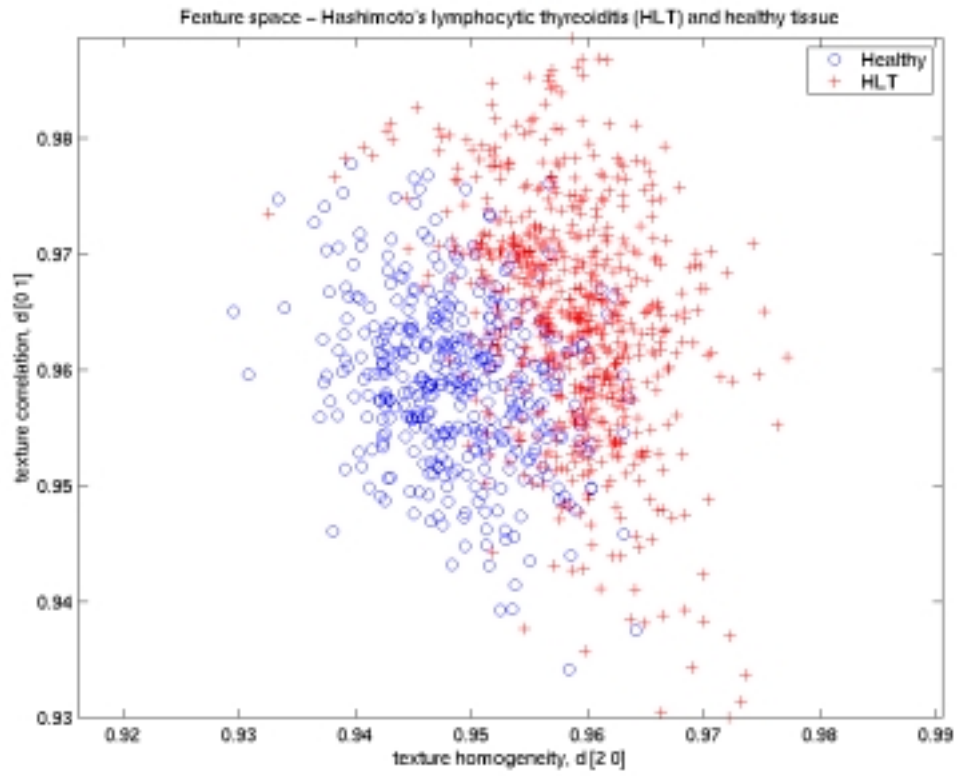


Figure 6