LEARNING AND DIAGNOSTICS OF LYMPHOCYTIC THYROIDITIS BASED ON SONOGRAM SCANS

Ludvík Tesař
Department of Adaptive Systems
Inst. of Information Theory and Automation
Czech Academy of Sciences
Prague, Czech Republic
email: tesar@utia.cas.cz

Daniel Smutek, Jan Jiskra and Pavel Maruna
3rd Department of Medicine
1st Medical Faculty, Charles University
Prague, Czech Republic
email: smutek@cesnet.cz

ABSTRACT
The problem of diagnostics of diseases of thyroid gland is addressed. We divide this problem into two stages. In learning stage, the database of sonograms with known diagnosis is used. In diagnostics stage, the sonogram of person with unknown diagnostics is taken and the most probable diagnosis is inferred. The paper uses Bayesian inference with Gaussian clusters of pattern features. The method is tested on Hashimoto's lymphocytic thyroiditis, which is ideally suited for this approach.

KEY WORDS
Pattern recognition, medical image analysis, Bayesian approach, Gaussian mixture modeling, Hashimoto's lymphocytic thyroiditis.

1 Introduction

The problem of diagnosing diseases of thyroid gland is very important, because early diagnosis can help in successful treatment thus improving quality of lives of patients. Following is the situation with Hashimoto's lymphocytic thyroiditis [1]. It is possible to verify it reliably by invasive fine needle espiration biopsy or an expensive magnetic resonance imaging. Cheaper sonogram is harder to evaluate [2, 3]. The inflammation in the gland changes the structure of the thyroid tissue. These changes are affecting the entire gland. Therefore it is likely to be detected by sonographic imaging [4, 5], using pattern recognition. Sonographic imaging is the most widely used diagnostic and monitoring tool for thyroid gland [4, 6, 5]. Developments in automatic texture analysis with computers give the possibility to use image texture analysis methods for medical diagnosis of various diseases. In this particular case, it allows to use texture properties which may not even be visible by the eye.

History of using computers for improving interpretation of medical sonographic images began in the mid-eighties [7, 8]. Studies from eighties also included image texture analysis of thyroid gland, but limited only to the comparison of gray-level histograms of different diagnoses. Another works [9, 10] also localized changes (like tumors and cysts). Except of our publications [11, 12], there is not much work published concerning quantitative analysis of diffuse changes in thyroid gland with a greater number of subjects or involving texture features other than first-order ones derived from gray-level histograms. Recent developments in imaging technology considerably improved the quality of sonograms, mainly of subsurface organs such as the thyroid gland, and will facilitate a quantitative assessment of the diffuse processes associated with chronic inflammations.

2 Description of the diagnostics methods

Diagnostics method described in this paper uses data collected by an experienced physician. Such data consist of the set of sonograph images belonging to persons with known diagnosis. These data are used to construct model in learning stage (Section 2.1). Model constructed from all available expert data is used in diagnostics stage in order to give diagnosis from sonograph image obtained from person with unknown diagnosis.

2.1 Modeling

From each image the region with interior of the thyroid gland is taken, and only this region is considered. As an example of interior of thyroid gland, see Figure 1, where interior is highlighted. The interior of thyroid gland is divided into non-overlapping squares of the size 41×41, as illustrated in the Figure 2 and Figure 3. Feature vectors are calculated from every little square of the size 41x41. Feature vector is the vector of numbers which characterize some aspect of the pattern which is present on given image. In our case the most interesting property, which likely changes in inflamed tissue is the contrast. Therefore Haralick's co-occurance and Muzzolini's spatial features were used. Similar set of features is used as in [11, 12, 13]. The assumption concerning feature vector, is that features contained in this vector carry substantial information about texture in the original image. This assumption is justified due to the fact, that presence of Hashimoto's lymphotitic thyroiditis changes consistence of the thyroid gland and this difference should be detectable by change in pattern in the

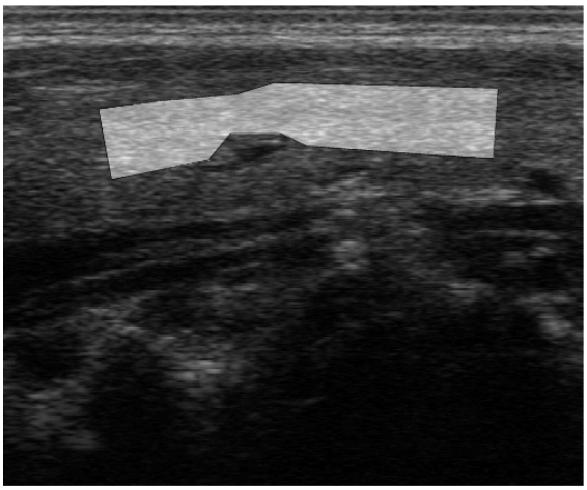

Figure 1. Example of sonograph image with highlighted section of interior of thyroid gland

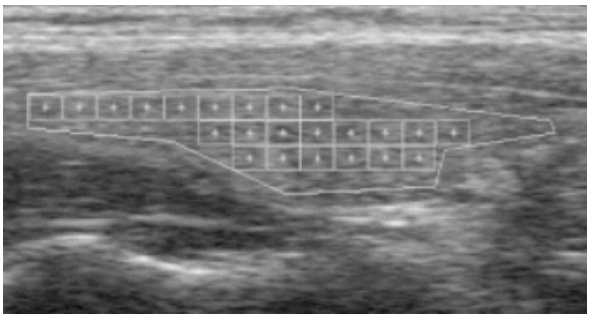

Figure 2. Dividing interesting region of thyroid gland into squares

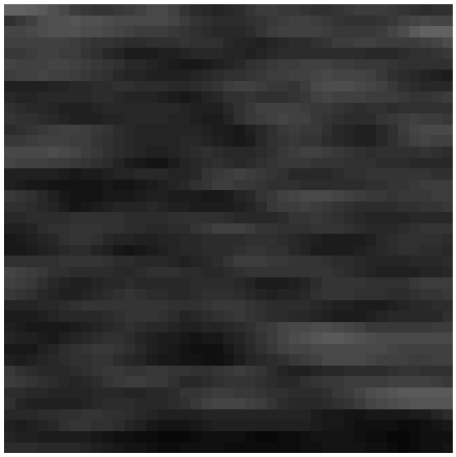

Figure 3. One square from which the feature vector is calculated

sonograph scan.

Feature vectors consist of features, that are calculated from co-occurance matrices. Number of colors in gray scale is reduced from 256 to 64. Haralick features are calculated from 12 different co-occurance matrices (4 horizontally, 4 vertically, 4 diagonally). Nine types of Haralick features for every spatial vector were used. Also, we used 21 Muzzolini spatial features. Whole feature vector consist of 129 features.

The idea behind the method is that feature vectors of inflamed and healthy tissue should form different clusters in the multidimensional feature space (In this case 129-dimensional space). For modeling in the space of feature vectors we use Gaussian Mixture Model. Features calculated for each small image cut are fitted to the Mixture. Two models are to be constructed. Data obtained from the healthy tissue are used to construct the model number 1 and data from the tissue with inflammation are used for constructing the model number 2.

For modeling of feature vector Y, we use Gaussian mixture model:

$$p(Y) = \sum_{i=1}^{n} \frac{\alpha_i \exp\left[-\frac{1}{2}(Y - Y_i)^T C_i^{-1}(Y - Y_i)\right]}{(2\pi)^{\frac{d}{2}} |C_i|^{\frac{1}{2}}} \quad (1)$$

Where n is order of the mixture, d dimension of vector Y, $|\cdot|$ denotes determinant of matrix, symbol $()^T$ denotes transposition. Probability density function $p(Y)$ is thus characterized by scalar parameters α_i (called weights), vectors Y_i (called centers) and positive definite matrices C_i, called covariance matrices, each defined for $i = 1, \ldots, n$. Another condition is, that $\sum_{i=1}^{n} \alpha_i = 1$. Such model is rich enough to represent data even if it is unevenly clustered.

The method used for estimating parameters was standard Expectation Maximization (EM) algorithm [14]. Learning set of feature vectors $Y^{(h)}(1), Y^{(h)}(2) \ldots, Y^{(h)}(t)$ was taken and estimates $\hat{\alpha}_i^{(h)}$, $\hat{Y}_i^{(h)}$ and $\hat{C}_i^{(h)}$ are calculated by the iterative process (upper index $\cdot^{(h)}$ denotes healthy data or parameters). Strictly speaking, resulting probability density function from (1) is now conditioned by data, therefore the model: $p^{(h)}(Y|Y^{(h)}(1),\ Y^{(h)}(2), \ldots,\ Y^{(h)}(t))$ describes healthy data. Analogically data with Hashimoto's lymphocytic thyroiditis are described by probability density function $p^{(f)}(Y|Y^{(f)}(1),\ Y^{(f)}(2),\ \ldots,\ Y^{(f)}(t))$ with parameter estimates $\hat{\alpha}_i^{(f)}$, $\hat{Y}_i^{(f)}$ and $\hat{C}_i^{(f)}$. Symbolically we can write $p(Y|\mathcal{H})$ as model for healthy and $p(Y|\mathcal{F})$ for tissue sample with inflammation.

During estimation, care must be taken to resolve numeric problems correctly. This is important especially with higher dimension of data and with many samples of data. Our testing application had about 34000 feature vectors from 81 subjects in learning set, with feature vector dimension equal to 129. Numeric issues that must be cared for are:

Positive definiteness of covariance matrices. This matrix must be calculated in decomposition, we used LD

decomposition. It means that $C_i = L^T D L$, where D is positive diagonal, L is lower triangular with ones on diagonal.

Probability density below machine accuracy. All probability densities are to be calculated in logarithmic form. Care must be taken if two quantities in logarithm are to be summed.

Estimation of optimal number of clusters is an interesting question of parameter estimation of Gaussian mixture (1). The more clusters (and bigger n) we have, the more accurately are data modeled. Some kind of Occam's razor should be used if this decision is important. We did not perform any estimation of n, we just manually set this parameter to a few different values and compared performance. As an optimal compromise between accuracy of modeling and speed of estimation we decided to use $n = 8$.

2.2 Inference Step

The task of inference step is to perform classification of any new sonographic image taken from patient, and infer whether the thyroid gland is healthy or it is not, based on model parameters estimated in learning step. For a new sonographic image we calculate the probability that the image belongs to the class of healthy images and probability that it belongs to the class of inflamed images. Using Bayes theorem, we can calculate posterior probabilities of both cases. Information obtained from different images for the same patient can be easily combined giving probability of inflammed thyroid gland for one patient.

During inference step, we are going to tell whether an image represented by its feature vector Y^{new} belongs to the class of healthy or inflamed tissue. We are going to judge this image based on probability densities $p(Y|\mathcal{H})$ and $p(Y|\mathcal{F})$ learned in Section 2.1. Probability density functions $p(Y|\mathcal{H})$ and $p(Y|\mathcal{F})$ can be easily evaluated for Y^{new}. Using Bayes rule we obtain:

$$\begin{aligned} &\Pr(\mathcal{H}|Y^{\mathrm{new}}) = \\ &= \frac{\Pr(\mathcal{H}|Y_0)p(Y^{\mathrm{new}}|\mathcal{H})}{\Pr(\mathcal{H}|Y_0)p(Y^{\mathrm{new}}|\mathcal{H}) + \Pr(\mathcal{F}|Y_0)p(Y^{\mathrm{new}}|\mathcal{F})} \end{aligned} \tag{2}$$

We assume prior information $\Pr(\mathcal{H}|Y_0) = \frac{1}{2}$ and $\Pr(\mathcal{F}|Y_0) = \frac{1}{2}$. Expression $\Pr(\mathcal{H}|Y^{\mathrm{new}})$ is probability that tested sample Y^{new} belongs to healthy class of images.

Bayes rule (2) allows to classify one image based on learned models. We have group of feature vectors based on several different images from the same patient (denoted as $Y^{\mathrm{new}}(1), Y^{\mathrm{new}}(2), \ldots, Y^{\mathrm{new}}(n)$). We will put them together by calculating joint probability density function for one patient:

$$\begin{aligned} &p(\mathbf{Y}|\mathcal{H}) = \\ &= p(Y^{\mathrm{new}}(1), Y^{\mathrm{new}}(2), \ldots, Y^{\mathrm{new}}(n)|\mathcal{H}) = \\ &= \prod_{k=1}^{n} p(Y^{\mathrm{new}}(k)|\mathcal{H}) \end{aligned} \tag{3}$$

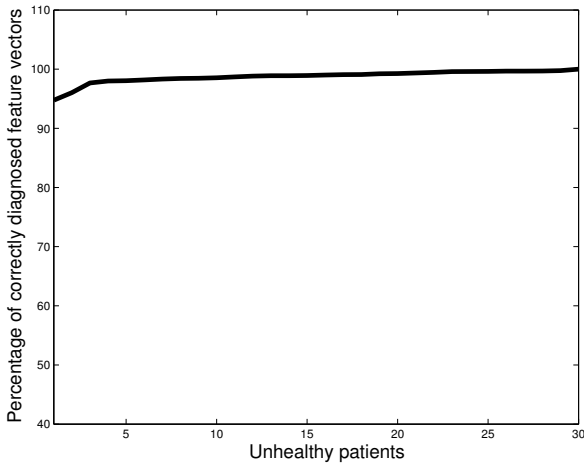

Figure 4. Sorted non-healthy patients (all 30) with percentage of correctly diagnosed feature vectors

Analogically we can obtain $p(\mathbf{Y}|\mathcal{H})$.

We can make similar formula as (2) for the series of images:

$$\begin{aligned} &\Pr(\mathcal{H}|\mathbf{Y}) = \\ &= \frac{\Pr(\mathcal{H}|Y_0)p(\mathbf{Y}|\mathcal{H})}{\Pr(\mathcal{H}|Y_0)p(\mathbf{Y}|\mathcal{H}) + \Pr(\mathcal{F}|Y_0)p(\mathbf{Y}|\mathcal{F})} \end{aligned} \tag{4}$$

Formula (4) gives the probability that given patient has healthy or inflamed gland.

3 Example

To test the method, we used a set of sonograms from 81 patients, from which 30 were healthy and 51 had Hashimoto's lymphocytic thyroiditis. From each patient we had between 100 to 1000 images. Total number of feature vectors was approximately 34000. We used leave-one-out method to ensure that method was trained with different data than tested. Leave-one-out method also avoids the possible bias introduced by relying on any one particular division into test and train data sets. We performed 81 experiments, in each of them we removed one patient, tried to learn from the rest, and tried to diagnose removed patient. Results are in table below. Column "total" denote total number of patients, "diag=" denotes diagnosis from the algorithm.

patients	total	diag=non-healthy	diag=healthy
non-healthy	30	30	0
healthy	51	1	50

One patient from healthy-group was incorrectly diagnosed as non-healthy, which might be border-case.

Figures 4 and 5 show non-healthy and healthy patients sorted by percentage of correctly diagnosed feature vectors. In the start of the graph in Figure 5 we can see that the incorrectly diagnosed patient has 40% of feature vectors diagnosed correctly. There is also one patient diagnosed correctly with only 60% of feature vectors diagnosed

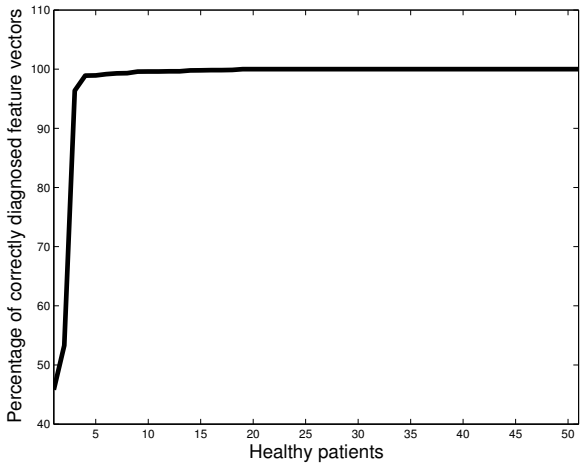

Figure 5. Sorted healthy patients (all 51) with percentage of correctly diagnosed feature vectors

correctly. All other patients have more than 95% of feature vectors diagnosed correctly.

4 Conclusion

We have shown that we are able to automatically diagnose Hashimoto's lymphocytic thyroiditis using feature vectors and Gaussian Mixture modeling with Bayesian decision-making. We showed, that we are able to tell, whether set of images from given patient belong to given class and we are able to calculate the probability. The method was tested on sonograms of 81 patients with total number of about 34000 image cuts. All tested patients except for one healthy patient were diagnosed correctly, using leave-one-out method to ensure that training data set differs from testing set.

5 Acknowledgments

The paper was supported by Grant agency of Academy of Sciences of the Czech Republic project 1ET101050403 and Czech Ministry of Health NO/7742-3 and NR/8352-3 projects.

References

[1] L. Wartfsky and S. H. Ingbar, "Disease of the thyroid", in *Principles of Internal Medicine*, Harrison, Ed., number 1712. McGraw-Hill Inc., New York, 1991, 12th Ed.

[2] F. J. Simeone, G. H. Daniel, and P. R. Mueller, "High-resolution real-time sonography", *Radiology*, vol. 162, no. 155, pp. 431–439, 1985.

[3] L. Solbiati, L. Volterrani, and G. Rizzatto et al., "The thyroid gland with low-uptake lesions: Evalution by ultrasound", *Radiology*, vol. 162, no. 155, pp. 187–196, 1985.

[4] G. A. Gooding, "Sonography of the thyroid and parathyroid", *Radiol Clin North Am*, vol. 5, no. 31, pp. 967–989, 1993.

[5] L. A. Loevner, "Imaging of the thyroid gland", *Semin Ultrasound CT MRI*, vol. 17, no. 6, pp. 539–562, 1996.

[6] C. R. Hopkins and C. C. Reading, "Thyroid and parathyroid imaging", *Semin Ultrasound CT MRI*, vol. 16, no. 4, pp. 279–295, 1995.

[7] C. Kimme-Smith and J. P. Jones, "The relative effects of system parameters on texture in gray-scale ultrasonograms", *Ultrasound Med Biol*, vol. 10, no. 3, pp. 299–307, 1984.

[8] G. E. Mailloux, M. Bertrand, R. Stampfler, and S. Ethier, "Computer analysis of echographic textures in Hashimoto disease of the thyroid gland", *J Clin Ultrasound*, vol. 14, no. 7, pp. 521–527, 1986.

[9] T. Hirning, I. Zuna, and D. Schlaps, "Quantification and classification of echographic findings in the thyroid gland by computerized b-mode texture analysis", *Eur J Radiol*, vol. 9, no. 4, pp. 244–247, 1989.

[10] H. Morifuji, "Analysis of ultrasound B-mode histogram in thyroid tumors", *Nippon Geka Gakkai Zasshi*, vol. 90, no. 2, pp. 210–221, 1989.

[11] D. Smutek, R. Šára, P. Sucharda, and L. Tesař, "Different types of image texture features in ultrasound of patients with lymphocytic thyroiditis", in *Proceedings of ISICT 2003*, Aleksy M. et al., Ed., Trinity College Dublin, Ireland, September 2003, pp. 100–102, Computer Science Press, Trinity College Dublin.

[12] D. Smutek, R. Šára, P. Sucharda, T. Tjahjadi, and M. Švec, "Image texture analysis of sonograms in chronic inflammations of thyroid gland", *Ultrasound in Medicine and Biology*, vol. 29, no. 11, pp. 1531–1543, 2003.

[13] R. Šára, D. Smutek, P. Sucharda, and Š. Svačina, "Systematic construction of texture features for hashimoto's lymphocytic thyroiditis recognition from sonographic images", in *Proceedings of Artificial Intelligence in Medicine (AIME'01)*, Cascais, Portugal, July 2001, pp. 339–348, Springer Verlag LNAI 2101.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.