Image Analysis and CADx System for Mucosal Lesions

Artur Chodorowski, Chitta R Choudhury and Tomas Gustavsson

***Abstract*— A computer aided diagnosis (CADx) system for oral mucosal lesions has been developed using clinical cases from India as training examples. The investigated classifiers were Support Vector Machine (SVM) and Bayes Point Machine (BPM), and the task was to discriminate potentially precancerous lesions from non-precancerous lesions. The discriminating features consisted of color differences and lesions' shape properties. The overall classification accuracy was 85% (29 out of 34) for both SVM and BPM classifiers.**

I. INTRODUCTION

Oral cancer is a serious condition which requires an early detection for a successful treatment. An important part of the clinical routine examination at the dentist is visual and palpatory examination of the oral tissues. The goal is early detection and discrimination of potentially precancerous lesions, in this way decreasing the oral cancer mortality rates.

Today the common practice at odontological clinics and mobile registration units is the registration of human oral cavity using still picture or video CCD color cameras. The recorded images are stored to archive patient data, which permits comparative follow-up of the clinical appearance of a lesion from a longitudinal follow-up aspect.

The digital color images are not diagnostic, a biopsy is frequently taken of lesional tissue and the specimen is examined by an experienced histopathologist. The further process may include the molecular DNA investigations [9][14]. Nevertheless, the recording and computerized analysis of true color images is still important, as a cost-effective method for early detection of suspicious lesions to be monitored.

One of the areas most suffering from oral cancer is South Asia and in particular India, with around 100 000 oral cancer cases each year [3]. In this work we have evaluated digital true color images of human oral cavity, recorded on subjects in the Indian population.

The images were evaluated with respect to color and shape properties and we investigated supervised learning algorithms for clinical decision making. The task was to discriminate between potentially precancerous lesions and the usually harmless lesions. The lesions and the results were compared with the results of the previous study of oral lesions among Nordic population [6]. Fig. 1 shows examples of the recorded lesions: an Oral Leukoplakia and Oral Submucous Fibrosis (OSMF), which are potentially precancerous lesions, and an Oral Lichenoid Reaction (OLR), which is usually a harmless lesion. The pre-cancerous lesions may develop into Oral Squamous Cell Carcinoma (OSCC).

Fig. 1. (left-upper) An oral squamous cell carcinoma (right-upper) oral leukoplakia, a potentially precancerous lesion (left-lower) an oral submucous fibrosis, potentially precancerous lesion (right-lower) an oral lichenoid reaction, usually harmless lesion.

II. MATERIAL AND METHODS

A. Material

The material analyzed in this study was recorded at the Institute of Dental Sciences, Deralakatte, India, and other linked centers of the Institute. The digital true color images were recorded using Kodak DX6490 digital still picture camera, with typical resolution of 2304x1728 pixels. In total, 34 images have been analyzed, 23 images of oral leukoplakia, which is regarded as potentially precancerous lesion, and 11 images of lichenoid reactions/lichen planus, usually harmless lesions. The clinical diagnoses have been histopathologically verified.

B. Image analysis

The regions of interest consisted of lesion area and the area regarded as normal (healthy) tissue. The lesion area has been marked semi-automatically using the live-wire technique [1]. The reason for a semi-automatic segmentation was to reduce the delineation time and avoid inter- and intra-operator variations. However, the fully automatic segmentation of the lesions is still an unsolved problem due to the high variability

Manuscript received July 15, 2008. This project was supported by SIDA (the Swedish International Development Cooperation Agency), grant no. SWE-2005-430.

A. Chodorowski is with the Department of Signals and Systems, Chalmers University of Technology, Sweden, (e-mail: artur@s2.chalmers.se)

C. R. Choudhury is with Oral Biology and Centre for Oral Disease Prevention and Control, AB Shetty Memorial Institute of Dental Sciences, Deralakatte, Mangalore, India and International Centre For Tropical Oral Health, Bournemouth University and Poole NHS, England, (e-mail: cr_choudhury@yahoo.co.uk)

T. Gustavsson was with the Department of Signals and Systems, Chalmers University of Technology, Sweden

of the lesions' appearance and lack of well-defined lesion prototypes.

The local live-wire cost function $C(p,q)$ from pixel p to the neighboring pixel q was defined as:

$$C(p,q) = w_Z f_Z(q) + w_C f_C(q) + w_G f_G(q) + w_D f_D(p,q) \quad (1)$$

where $f_Z(q), f_C(q), f_G(q)$ and $f_D(p,q)$ represent the Laplacian zero crossing edge detection, Canny edge detection, gradient magnitude and gradient direction cost terms, respectively, weighted by a corresponding weight constants. The cost term for the gradient magnitude at pixel q was defined as:

$$f_G(q) = 1 - G(q)/max(G) \quad (2)$$

where $G(q)$ is the magnitude of the color gradient and $max(G)$ is the highest gradient magnitude from the entire image. The cost term for gradient direction from pixel p to pixel q was defined as:

$$f_D(p,q) = acos\left\{ \frac{D_x(p)D_x(q) + D_y(p)D_y(q)}{G(p)G(q)} \right\} / \pi \quad (3)$$

where $D_x(p)$ and $D_y(p)$ are the eigenvectors corresponding to the largest eigenvalue for the x and y gradient directions of pixel p, respectively.

Currently, the weights w_Z, w_C, w_G, and w_D are set by trial and error to the constant values 5, 3, 1, and 1, respectively. A more advanced weight learning algorithm is required in the future. The extracted by the live-wire boundary corresponded to the minimum of the cost function defined in (1).

The normal tissue area was selected manually, as it is more difficult to decide what tissue serves as a normal (reference) area. Fig. (2) below shows an example of the delineated regions of interest.

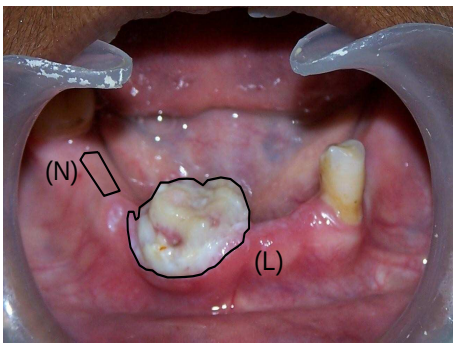

Fig. 2. Regions of interest: (L) lesion area (N) normal tissue area.

The image RGB values were transformed to Hue-Saturation-Intensity color system using the following equations: $Hue = atan2(Y,X)$, where $X = (2R - G - B), Y = \sqrt{3}(G - B), Saturation = 1 - min(R,G,B)/(R+G+B)$ and the color features were calculated as differences in mean hue values (H_{diff}) and mean saturation values (S_{diff}) between abnormal (lesion) and normal regions. As a shape feature we have chosen area factor, $AF = A/A_R$, where A is lesion's area and A_R is area of its bounding rectangle. The area factor is a measure of object's compactness. In total, the extracted feature vector F consisted of $F = \{H_{diff}, S_{diff}, AF\}$. Due to the low number of training examples we limited our feature vector to three dimensions.

C. Supervised classification

As discrimination functions we have chosen the two competing classifiers: Support Vector Machines (SVMs) and Bayes Point Machines (BPMs). It has been experimentally shown [4] that for the zero training error case the BPMs consistently outperform the SVMs on both artificial and real-world data sets. In the soft-boundary and soft-margin case (allowing classification errors) the improvement over SVMs was reduced. This makes it interesting to compare these classifiers on real-world data and low sample volumes.

The Support Vector Machines (SVMs) [2][11], have the following decision function:

$$f(\mathbf{x}) = sgn\Big(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \Big) \quad (4)$$

where $(\mathbf{x}_i, y_i)$ are training examples, $\mathbf{x}$ is a sample to be classified, l is number of training examples, K is a kernel function, $y_i = \{-1, +1\}$ are class labels, b is bias, α_i are the solutions to the associated quadratic programming problem, $0 \leq \alpha_i \leq C$, and C is a penalty parameter chosen by the user, in a non-separable case. A large C corresponds to a high penalty to classification errors and the SVM will attempt to find in the parameter space a complex surface that separates the data perfectly. The optimization criterion is the width of the margin between classes (maximal margin classifier) or margin distribution (soft margin optimization). In the experiments, we have used the polynomial and radial basis functions (RBF) kernels:

$$K_{Poly}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d, d \in N \quad (5)$$

$$K_{RBF}(\mathbf{x}, \mathbf{y}) = exp\left(-(\|\mathbf{x} - \mathbf{y}\|^2)/2\sigma^2\right) \quad (6)$$

The Bayes Point Machine (BPM) is a Bayesian approach to linear classification. The decision function for the Bayes Point Machine [4] has the same functional form as in (4). However, the α_i coefficients are calculated differently, by averaging all classifiers according to their posterior probabilities. The solution is called the Bayes Point and the corresponding classifier is a Bayes Point Machine. For estimation of the Bayes Point we have used a method called Expectation Propagation (EP), developed by Minka [7]. The EP algorithm uses a multivariate Gaussian approximation to the posterior over the classifier's weights and uses their mean as the estimated Bayes Point. Alternative ways for estimation of the Bayes point are the mean-field and TAP algorithms [8] or different billiard and perceptron learning algorithms [10][4].

In our classification experiments we first estimated (using grid search) the best SVM with respect to kernel type and kernel parameters and then trained the BPM with the same kernel type and the same kernel width.

III. RESULTS

This section summarizes our experiments with SVMs and BPMs on this two-class problem, where the two classes are the precancerous lesions and the harmless lesions. Here, the leukoplakia cases were labelled as $+1$ and lichenoid reactions as -1. To avoid overoptimistic results we present leave-one-out cross-validated error rates and ROC-curves. The ROC-curves have been generated by varying the bias term in (4).

The Table I presents the error rates for the SVM classifiers using different kernels, with different parameter values.

TABLE I

LEAVE-ONE-OUT ERROR RATES FOR SVMS WITH DIFFERENT KERNELS.

$Features = \{H_{diff}, S_{diff}, AF\}, dim(X) = 3, N_1 = 11, N_2 = 23$

Kernel type	$C = 0.1$	$C = 1$	$C = 10$	$C = 10^2$	$C = \infty$
linear	0.32	0.32	0.17	0.20	0.20
poly d=3	0.32	0.20	0.20	0.17	0.29
RBF $\sigma = 0.25$	0.32	0.17	0.15	0.17	0.29
RBF $\sigma = 0.5$	0.32	0.20	0.20	0.20	0.32
RBF $\sigma = 1$	0.32	0.32	0.17	0.20	0.29

All of the kernels attended some minimum in the middle range of the kernel and penalty parameters, with a global minimum for the RBF-kernel with the width $\sigma = 0.25$, $C = 10$ and error rate of 0.15.

To investigate this SVM-RBF classifier at other operating points we have generated a Receiver Operating Characteristic (ROC) curve, simply by varying the bias term in (4). For comparison, we have trained a BPM using same kernel (RBF, $\sigma = 0.25$) and the Expectation Propagation as the training method. Fig. 3 presents the ROC-curves for both the SVM and BPM classifiers.

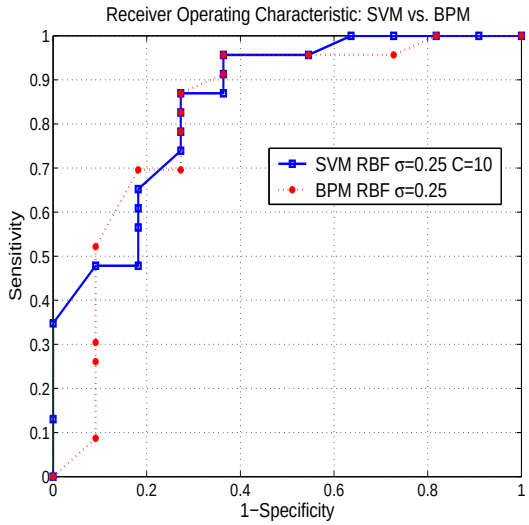

Fig. 3. Receiver Operating Characteristic, Sensitivity vs. specificity for RBF-BPM ($\sigma = 0.25$) and RBF-SVM ($\sigma = 0.25$, $C = 10$), leave-one-out estimation.

Both the ROC-curves have a similar shape. The BPM-ROC has a slightly better performance in the middle range, while the SVM-ROC outperforms at the extreme ranges of sensitivity/specificity. The jaggedness of the ROC-curves might be a result of low number of training examples. Also the applied way of generating ROC-curves might contribute to this effect. A better method for controlling sensitivity vs. specificity could be by using different penalties for different kind of errors, as proposed in [12]. Due to the low number of training examples it is impossible to obtain statistically significant differences between these two classifiers. Both the classifiers attend the best total classification accuracy of 85% and in fact possess the same confusion matrix i.e. the same kind of errors by wrongly classifying one precancerous lesion and four harmless lesions (5 of 34). It is interesting that the SVM faster than the BPM reaches the point of 100% sensitivity (with 34% specificity).

The analysis of individual features showed that the saturation difference, S_{diff}, had a more discrimination power than the hue difference, H_{diff}. This was in contrast with the results for Nordic population [6] for which the hue difference dominated as the discriminatory parameter.

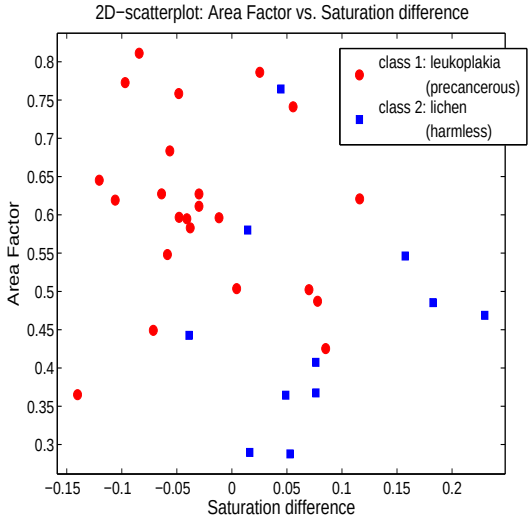

Fig. 4. 2-D scatterplot: Area Factor (AF) vs. Saturation difference (S_{diff}). Class 1 (circles) is leukoplakia, class 2 (squares) are lichenoid reactions.

Fig. 4 presents a scatterplot for two of the features: Area Factor (AF) and Saturation difference (S_{diff}), to get an impression of features' distribution. The area factor is larger for leukoplakia than for lichenoiod reactions, and this coincides with the human perception. The leukoplakia lesions are usually solid and compact, giving a high area factor, while lichenoid reactions may have very irregular borders which results in a low area factor. The second feature, the saturation difference is much harder to detect with a naked eye. This feature is a difference between lesion saturation and normal tissue saturation. The results suggest that leukoplakias, as potentially precancerous lesions, has a lower saturation than the tissues regarded as normal. The lichenoid reactions, usually harmless lesions, seem to be more saturated than normal tissues.

IV. CONCLUSION

We have presented development steps toward a computer aided analysis and diagnosis system for oral mucosal lesions. Two common classifiers were evaluated on real-world data obtained from an odontological domain.

The obtained overall classification accuracy around 85% may be compared with the human performance. The diagnostic performance of the human specialist is estimated at 90%, while the diagnostic ability of dentists in general practice is estimated at 75%, when compared to a specialist [5].

Our current oral lesion database contains some cases of additional common lesions in Indian population, such as erytroplakia and oral submucous fibrosis, which have also potential to develop into oral cancer. These lesions will be analysed as the number of cases will grow. The larger number of examples will give more accurate estimates of the classification performance.

The advantage of standard color imaging is its simplicity, non-invasive character and the low cost, which is important for spreading the technique to the South Asian countries. However, due to the overlapping between the classes, to improve the system's classification accuracy we probably need to incorporate additional imaging data from frequencies beyond the visual spectrum of the light. Another possibility would be to extract additional high-discriminatory features.

The developed method will be implemented as a lesion analysis tool and decision support system to be used in the South-Asian countries at odontological clinics and institutes.

REFERENCES

[1] W. Barrett and E. Mortensen, Interactive live-wire boundary extraction, *Medical Image Analysis* 1(4), pp. 331-341, 1997

[2] C. Cortes, V. N. Vapnik, "Support vector networks", *Machine Learning* 20 (1995) 273-297

[3] P. C. Gupta and A. Nandakumar, Oral cancer scene in India, *Oral Diseases*, 1999, No 1: 1-2.

[4] R. Herbrich, T. Graepel and C. Campbell, "Bayes point machines", *J. Mach. Learn. Res.* 1, pp.245-279, 2001

[5] J. A. Jullien, M. C. Downer, J. M. Zakrzewska and P. M. Speight, "Evaluation of the screening test for the early detection of oral cancer and precancer", *Community Dent Health* 12, pp.3-7, 1995

[6] U. Mattsson, A. Chodorowski, T. Gustavsson, M. Jontell, F. Bergqvist, "Use of computer-assisted image analysis for noninvasive evaluation of oral lichenoid reactions and oral leukoplakia." *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 1995 Feb;79(2):199-206

[7] T. P. Minka, *A family of algorithms for approximate Bayesian interference*. Doctoral dissertation, Massachusetts Institute of Technology, 2001

[8] M. Opper and O. Winther, Gaussian processes for classification: Mean field algorithms. *Neural Computation* 12, pp. 2655-2684, 2000

[9] J. Reibel, Prognosis of oral pre-malignant lesions: Significance of clinical, histopathological and molecular characteristics. *Crit Rev Oral Bio Med* 14(1) (2003) 47-62

[10] P. Rujan, "Playing billiard in version space", *Neural Computation*, vol. 9, no. 1, pp.99-122, 1997

[11] V. N. Vapnik, "The Nature of Statistical Learning Theory", 1st ed., Springer-Verlag, New York (1995)

[12] K. Veropoulos, C. Campbell and N. Cristianini. "Controlling the Sensitivity of Support Vector Machines.", In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI99)*, Workshop ML3, pp. 55-60. Stockholm, Sweden, 1999

[13] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio and V. Vapnik "Feature selection for SVMs", in *Advances in Neural Information Processing Systems*, Vol, 13, 2000

[14] N. Yamamoto, T. Kuroiwa, A. Kutakura, T. Shibahara and C. Choudhury "Loss of Heterozygosity (LOH) on Chromosomes 2q, 3p and 21q in Indian Oral Squamous Cell Carcinoma", *Bull. Tokyo Dent Coll*, 48(3):109-117, 2007