International Journal of Current Research in Life Sciences
Vol. 05, Issue. 06, pp. 563-568, June 2016

www.ijcrls.com

Full Length Research Article

MACHINE LEARNING ALGORITHMS FOR CLASSIFYING OLP IMAGES USING HISTOGRAM AND BICC FEATURES WITH SVM AND RBFNN

Dr. Venkatakrishnan, S.

Assistant Professor, Mechanical Engineering, Engg-Wing DDE, Annamalai University

Accepted 17th May 2016; Published Online 10th June 2016

ABSTRACT

Oral lichen planus is a potentially malignant disorder with a considerable risk of malignant transformation. Thus intervention at an earlier level could improve prognosis and also avoid cancer formation. Image analysis and pattern classification methods are already on the rising curve in the field of cancer and precancer research. This research has utilised machine learning algorithms for classifying OLP images using histogram and BICC features with SVM and RBFNN classifiers. The results are evaluated and compared among different methods.

Key words: Oral lichen planus. Cancer , images, classification, SVM, RBFNN.

INTRODUCTION

Oral cancer being one of the major diseases challenging human lives, makes it mandatory to invent newer advanced techniques for an early diagnosis and intervention. Thus health professionals aim at saving the life of patients or at the least, reduce their pain and suffering that could improve their lifestyle. Potentially malignant disorders (PMDs) are certain lesions and conditions that predispose to oral cancer formation at an increased rate as compared to normal oral mucosa (Rajendran and Sivapathasundaram, 2009). Leukoplakia, OSMF (oral submucous fibrosis), Oral lichen planus (OLP), etc are common among them (Rajendran, 2004). But the chances of malignant transformation of OLP is lesser than the other two.

Related works

In 2009, Riikka Mattila, published the paper "Molecular Markers of Oral Lichen Planus". The article explained OLP as a chronic inflammatory mucosal disease and is detected in between 0.5% - 2.2% of the population (Riikka Mattila, 2009). WHO has defined OLP as a potentially precancerous disorder, representing a generalized state associated with a significantly increased risk of cancer. However, only 0.5 – 2.9% of OLP lesions will progress to cancer. Currently, there are no prognostic markers to identify the lesions at increased risk for malignant transformation. The main aim of these studies was to identify cellular and molecular markers in order to understand the pathogenesis of atrophic OLP and its progression towards malignancy. Selected markers for cell proliferation, adhesion, apoptosis, and lymphocytic infiltration were assessed by immunohisto chemistry in addition to static cytometry analyses for DNA content.

Corresponding author:* **Venkatakrishnan, S.,
Assistant Professor, Mechanical Engineering, Annamalai University, Tamilnadu, India

In 2011, Smitha T, Sharada P and Hc Girish, described "Morphometry of the basal cell layer of oral leukoplakia and oral squamous cell carcinoma using computer-aided image analysis". The paper was developed to study and compare the changes in nuclear and cellular size, shape and nuclear–cytoplasmic ratio of the cells in the basal layer of oral leukoplakia and well-differentiated oral squamous cell carcinoma (SCC) with normal buccal mucosa, using computer-aided image analysis in tissue sections (Smitha *et al*., 2011). This was a retrospective study conducted on tissue sections on a total number of 70 cases to determine the various morphometric parameters. The data collected in this study were analyzed statistically by computing descriptive statistics, viz., percentage, mean, standard deviation, standard error of mean, 95% confidence interval for mean. The difference in the control and study groups for various diagnostic variables was compared by means of analysis of variance (ANOVA), Student's t-test for independent samples, wherever applicable. Mann–Whitney U-test and Kruskal–Wallis test were used where the data were found to be asymmetrical and the standard deviations were also different. The results were considered statistically significant whenever $P \leq 0.05$.

Oral Lichen Planus (OLP): Oral Lichen Planus (OLP) is a common chronic inflammatory disorder with a genetic predisposition linked to the T cell-mediated immunological response to an induced antigenic change. It is a muco-cutaneous inflammatory disease which leads to an intense destruction of the basal layer of the epithelium (Carrozzo and Thorpe, 2009). Lichen planus affects about 1% to 2% of the population, being the most frequent dermatological disease that involves oral cavity. The most common clinical presentation of oral lichen planus is the reticular form, and the jugal mucosa and tongue are the most frequent locations of the disease, representing more than 70% of the cases. Oral lichen

planus also shows strong predisposition for occurrence in females, with a proportion of 5:1 in comparison with males. Regardless of gender, most oral lichen planus cases occur between the 4th and 6th decades of life. However, only the clinical presentation of the disease (atypical forms: atrophic, erosive, "in plaque") and the location of the lesions (tongue) were associated with a higher malignant transformation risk.

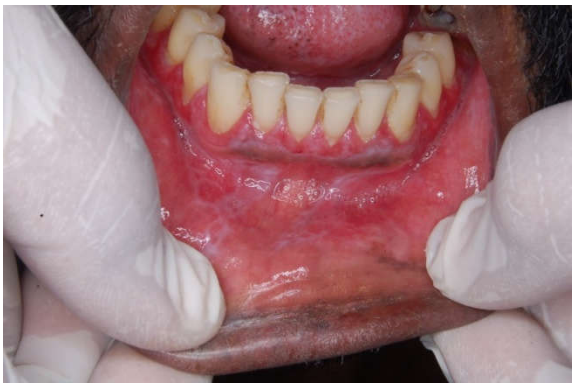

Fig. 1. Oral Lichen Planus

Fig.1 shows the white linear pattern of the OLP lesion in the affected patient's oral mucosa. Oral lichen planus presents as white striations, white papules, white plaques, erythema, erosions or blisters affecting predominantly the buccal mucosa, tongue and gingivae, although other sites are occasionally involved. Lesions are typically bilateral and often appear as a mixture of clinical subtypes. White or grey streaks may form a linear or reticular pattern on an erythematous background. Alternatively, there may be a central area of shallow ulceration (erosion) with a yellowish surface (fibrinous exudate) surrounded by an area of erythema. OLP can present itself as any of the following forms- ulcer, plaque. Patch, erosion, papules, etc. Almost all cases of OLP present with reticular keratotic [white] striae in some area of the oral mucosa.

Microscopic Features

- Basal cell liquefaction.
- Band-like lymphocytic infiltrate at epithelial-stromal junction with obfuscation of basal cell region.
- Normal epithelial maturation pattern
- "Candle-dripping", spindly rete ridges
- Parakeratosis.
- Civatte bodies.
- Ragged separation of epithelium from lamina propria due to basal cell destruction.

Degeneration of the basal keratinocytes and disruption of the anchoring elements of the epithelial basement membrane and basal keratinocytes (eg, hemidesmosomes, filaments, fibrils) weakens the epithelial-connective tissue interface. As a result, histologic clefts (ie, Max- Joseph spaces) may form, and blisters on the oral mucosa (bullous lichen planus) may be seen at clinical examination. B cells and plasma cells are uncommon findings (Rajendran R. and Sivapathasundaram, 2009). World Health Organization has defined OLP as a potentially precancerous disorder, representing a generalized state associated with a significantly increased risk of cancer. However, only 0.5 - 2.9% of OLP lesions will progress to cancer.

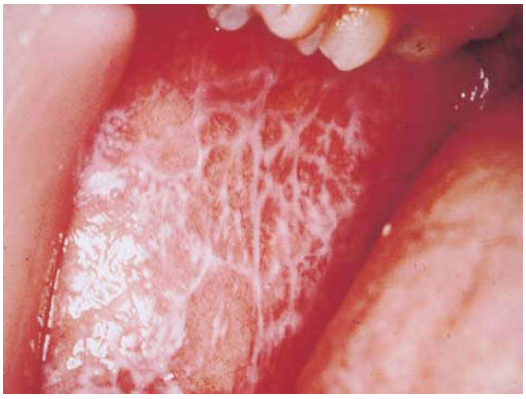

Fig. 2(a). Lichen Planus of cheek mucosa

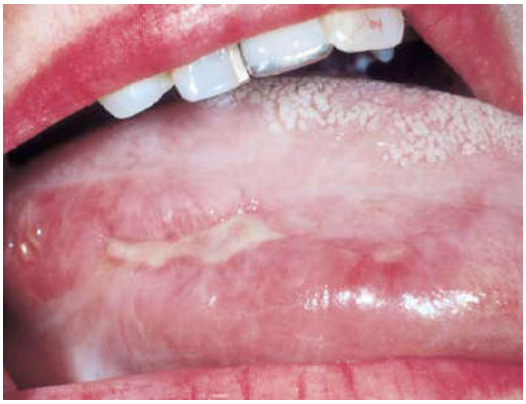

Fig.2(b). Lichen Planus of tongue

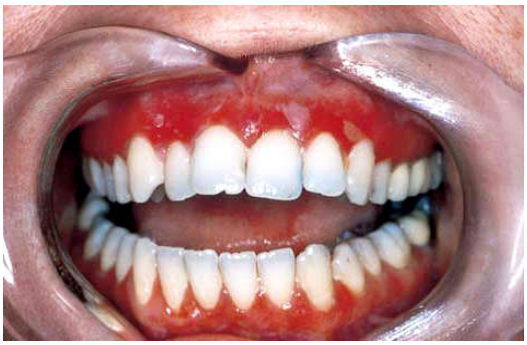

Fig. 2(c). Lichen Planus of Gingiva (gums)

The diagnosis of oral lichen planus demands expertise and experience in clinicians and pathologists. Like any other precancerous lesion, grading of dysplasia and assessment of its malignant potential is highly subjective, requiring more research work which is essential for a more appropriate diagnosis and classification of the disease.

Computer aided diagnosis: The only reliable method to diagnose disease affected oral tissues is microscopic examination of tissue samples. Variations in individual cell morphology, tissue-level changes, cell-nuclei changes and a few other microscopic changes are the features that are used by pathologists to differentiate normal and diseased tissues (Smitha *et al.,* 2011). It was once beyond imagination that a computerized device could scan a tissue section microscopically and make accurate interpretation. However,

the observer can record his findings, according to carefully defined specifications and criteria, without attempting to interpret these findings into a diagnosis. The findings regarding the features observed in the tissue sections can then be coded into a form suitable for computer analysis and various computer techniques can be applied to the problem of pattern recognition and diagnosis (Krishnan *et al.,* 2010). Finally, the results of these computer analyses can be compared with the diagnoses reached on the same cases by conventional methods (Muthu Rama Krishnan *et al.,* 2011). The evolution of computerized image processing and analysis system is a remarkable milestone in the field of computer applications. Research works on computer applications in oral mucosal lesions including cancers and pre-cancers had been done right from 1970s (Crispian Scully *et al.,* 2008; Shekhar Singh, 2012Kramer *et al.,* 1974). Various techniques like segmentation, textural analysis, automated classification, etc, are being continuously applied in oral cancer as well as pre cancers by many computer experts (Paul *et al.,* 2015). With numerous works on OSMF and oral cancer, very few have been done on OLP in this aspect (Rusha Patra *et al.,* 2012; Gao, 1992; Muthu Rama Krishnan *et al.,* 2009). Thus this research work has aimed to come out with a pattern classification system to classify OLP images and normal mucosal images.

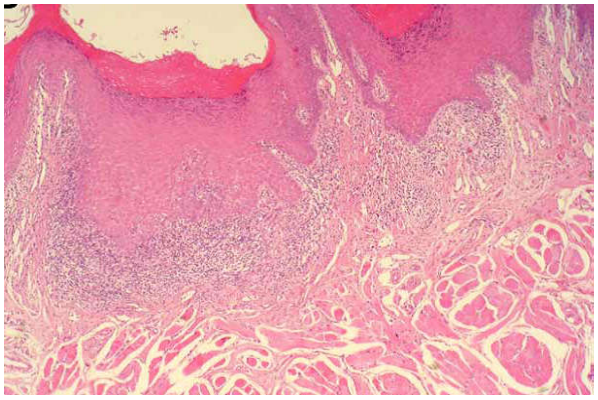

Fig. 3. Microscopic image of OLP (low magnification)

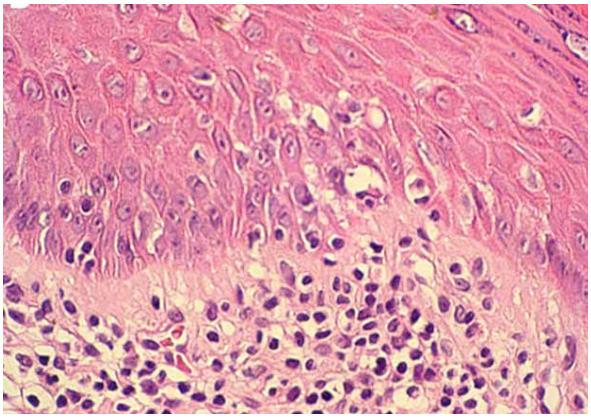

Fig. 4. OLP image under high magnification

Fig. 3. shows the appearance of a Lichen planus lesion under microscope. The epithelium is hyperplastic and the connective tissue shows inflammatory cell accumulation. Fig.4 shows the same lesion under high magnification.

Features for OLP classification

Histogram and BICC feature extraction: Color histogram features as explained in Section 4.2 are extracted from both OLP affected and normal images. BICC features such as discussed in Section 5.2 are extracted from both OLP affected and normal images. The features are combined and normalized using equation (6.1) to produce a feature vector which characterizes the image.

$$y_i = \frac{(x_i - x_{min})}{x_{max} - x_{min}} \qquad \text{.....(1)}$$

where x_{max} and x_{min} are the maximum and the minimum values X_i of the un normalized data using Histogram features were extracted for 16, 32, 64 bins and BICC features were extracted for blocks of size 5 x 5, 10 x 10, 15 x 15 which resulted in 10, 45, 105 dimensional feature vectors respectively.

Fig. 5 shows the OLP image with block 5 x 5 for extracting the BICC features. Fig. 6 shows the histogram features extracted from OLP image.

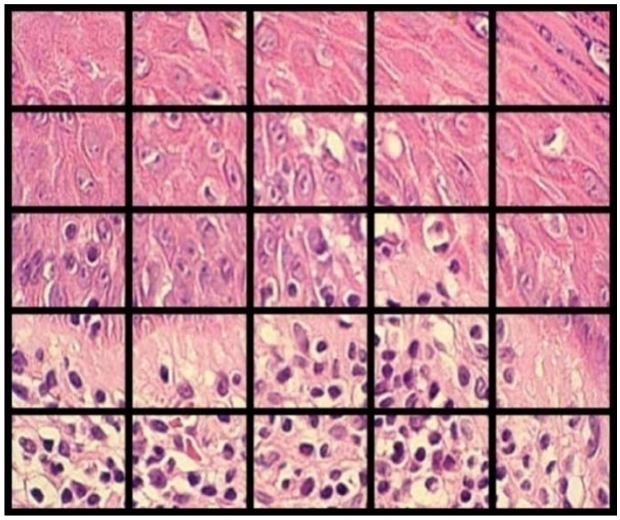

Fig. 5. OLP microscopic image is divided into blocks of size 5 X 5

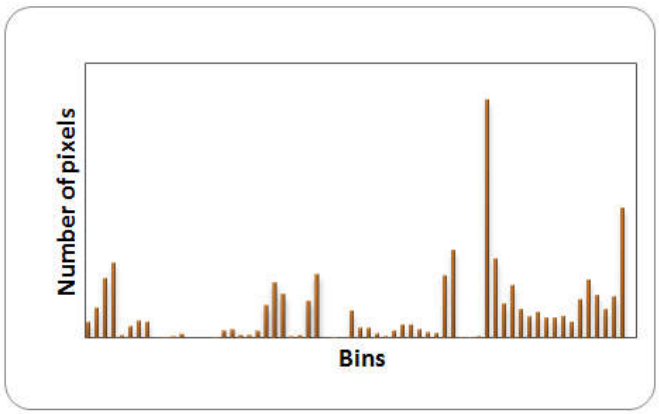

Fig. 6. Histogram of OLP Microscopic Image

EXPERIMENTAL RESULTS

A total of 200 microscopic images which consists of 100 OLP images and 100 normal images are considered. For four fold

cross validation training data gf_i (i=1, 2, 3, 4) consisting of 150 microscopic images [50 images (25 Normal + 25 OLP) + 50 images (25 Normal + 25 OLP) + 50 images (25 Normal + 25 OLP)] are used. For testing, 50 microscopic images (25 Normal and 25 OLP) are used.

Evaluation using SVM: A non-linear support vector classifier is used to discriminate the two categories. The N class classification problem can be solved using N SVMs. Each SVM separates a class from the other class. Support vector machine is trained to distinguish Histogram and BICC features of a category from other category. Two SVMs are created for each feature for each category. Microscopic images of 100 normal oral mucosal tissue samples and 100 OLP affected images were collected from Rajah Muthiah Dental College and Hospital. For evaluating the performance of OLP classification system, histogram and BICC features are extracted from the images. A non-linear support vector classifier is used to classify OLP and normal images. The performance is studied for 16, 32 and 64 bins histogram features and blocks of size 5 x 5, 10 x 10 and 15 x 15 resulting in 10, 45 and 105 dimensional BICC feature vectors respectively. Tables 1, 2 and 3 show the performance of normal and OLP classification using SVM for histogram feature, BICC and combined features using different kernel functions respectively.

Table 1. Average performance of normal and OLP classification in terms of accuracy by SVM model using Histogram features

Types of kernel	Accuracy(%)					
	Feature vector dimensions (No. of bins)					
	16		32		64	
	Normal	OLP	Normal	OLP	Normal	OLP
Polynomial	75.0	78.0	78.0	80.0	80.6	82.0
Gaussian	78.0	86.0	82.0	88.0	85.0	89.0
Sigmoidal	83.0	85.0	84.0	87.0	86.0	88.9

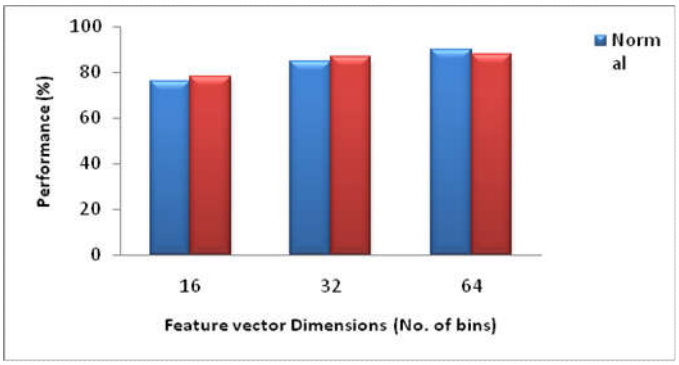

Fig. 7. Performance of normal and OLP classification using Histogram features by SVM Gaussian kernel model

Table 2. Average performance of normal and OLP classification by SVM model using BICC features

Types of kernel	Accuracy (%)					
	Feature vector dimensions (No. of BICC features)					
	10		45		105	
	Normal	OLP	Normal	OLP	Normal	OLP
Polynomial	75.0	79.0	79.0	82.0	80.0	83.4
Gaussian	80.0	83.0	84.0	88.0	86.0	87.0
Sigmoidal	84.0	86.4	89.0	89.0	89.0	90.5

The above three figures show the performance of OLP classification using histogram, BICC and combined feature vectors by SVM Model respectively.

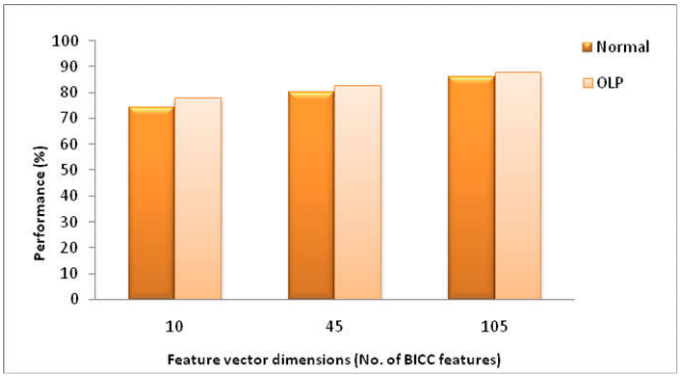

Fig.8. Performance of normal and OLP classification using BICC features by SVM Gaussian kernel model

Table 3. Average performance of normal and OLP classification by SVM model using combined features

Types of kernel	Accuracy (%)					
	No. of combined features (Histogram + BICC)					
	26		77		169	
	Normal	OLP	Normal	OLP	Normal	OLP
Polynomial	79.2	83.0	83.6	85.0	87.0	88.5
Gaussian	86.6	87.9	88.4	91.0	90.1	92.0
Sigmoidal	83.2	86.4	87.4	89.3	89.6	90.2

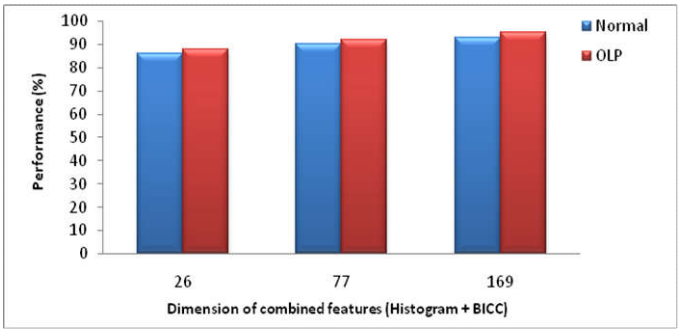

Fig. 9. Performance of normal and OLP classification using combined features by SVM Gaussian kernel model

Among the three kernels, Gaussian kernel gives a maximum performance for Histogram, BICC and combined features. Experiments were conducted for histogram features, BICC features and combined features. Results showed that the combined feature vectors performed well when compared to histogram and BICC features.

Evaluation using RBFNN: For RBFNN training, histogram, BICC and combined features are extracted from the images for each category. These features are given as input to the RBFNN model. In Radial Basis Function Neural Network, the weights in the network are determined using the least squares algorithms for training histogram features for 16, 32 and 64 bins, and BICC features for blocks of size 5 x 5, 10 x 10 and 15 x 15, resulting in 10, 45 and 105 dimensional feature vectors respectively which are extracted from the images. Combined features were also extracted from both the categories of images. These features are given as input to the RBFNN model. The RBF centers are located using k-means algorithm. For each category the value of k is varied from 2 to 8. The system gives optimal performance for k = 8. The weights in the RBFNN network are determined using the least square algorithm. For testing, if the output of the network is greater than the threshold, then it represents a OLP affected

image. The performance results are shown in the following Tables 4 and 5 for histogram, BICC and combined features respectively.

Table 4. Average performance of normal and OLP classification in terms of accuracy by RBFNN model using Histogram features

No. of means	Accuracy(%)					
	Feature vector dimensions (No. of bins)					
	16		32		64	
	Normal	OLP	Normal	OLP	Normal	OLP
2	80.6	82.4	84.3	86.6	86.4	88.9
4	84.9	86.4	87.0	89.5	89.2	90.2
6	85.4	88.9	89.6	92.5	90.4	94.3
8	88.5	90.4	92.3	94.6	94.5	96.5

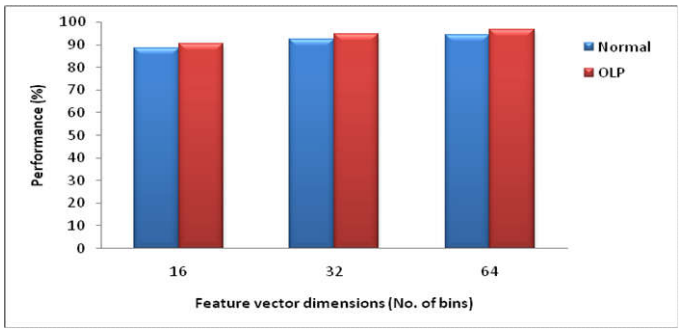

Fig.10. Performance of normal and OLP classification by RBFNN model with k=8 using Histogram features

Table 5. Average performance of normal and OLP Classification by RBFNN model using BICC features

No. of means	Accuracy(%)					
	Feature vector dimensions (no. of BICC features)					
	10		45		105	
	Normal	OLP	Normal	OLP	Normal	OLP
2	83.2	85.0	82.6	86.4	87.8	89.4
4	85.4	88.6	86.6	88.9	88.3	91.0
6	86.9	90.2	87.7	91.6	92.5	95.4
8	88.0	92.0	89.0	93.0	92.0	97.6

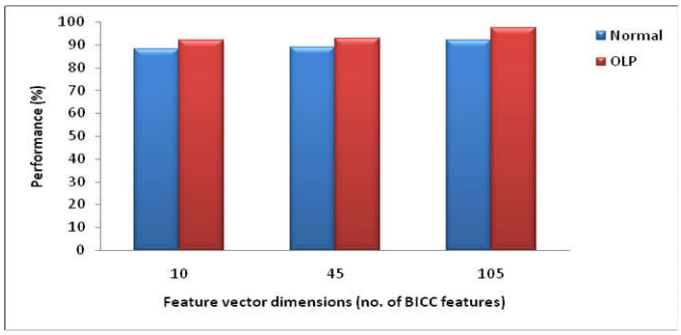

Fig. 11. Performance of normal and OLP classification by RBFNN model with k=8 using BICC features

Table 6. Average performance of normal and OLP classification by RBFNN model using combined features

No. of means	Accuracy(%)					
	No. of combined features (Histogram + BICC)					
	26		77		169	
	Normal	OLP	Normal	OLP	Normal	OLP
2	81.1	83.0	80.6	84.0	85.8	87.4
4	83.0	86.6	84.6	86.9	86.2	89.0
6	84.9	88.2	85.7	89.6	90.5	93.4
8	87.0	88.0	85.0	93.0	94.0	95.0

Table 6 shows the average performance of classifying OLP images by RBFNN model using combined features (Histogram + BICC).

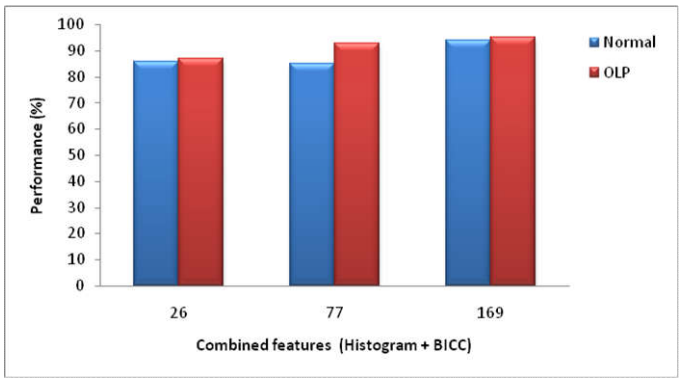

Fig. 12. Performance of normal and OLP classification by RBFNN model with k=8 using combined features

Figs.10, 11 and 12 show the performance of OLP classification using histogram, BICC and combined features respectively by RBFNN model. In RBF network, the weights are determined using least square algorithm where k=8. Fig. 13 and 14 show the snapshot of OLP classification system for training and testing using combined features.

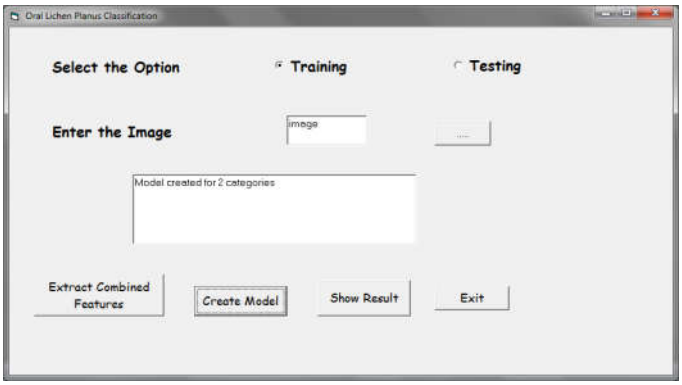

Fig. 13. Snapshot of normal and OLP classification system (Training)

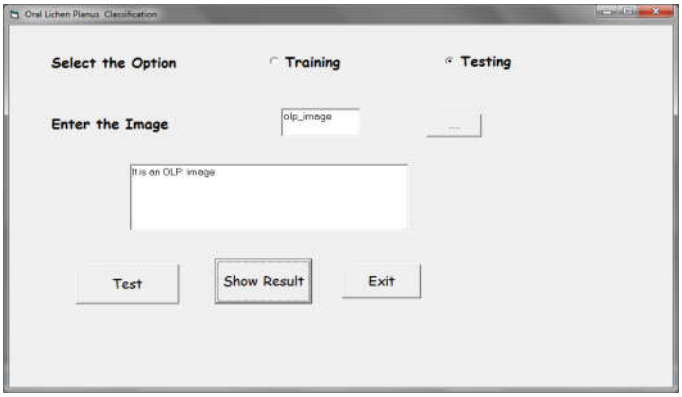

Fig. 14. Snapshot of normal and OLP classification system (Testing)

DISCUSSION

Oral lichen planus, although has a lesser risk of malignant transformation as compared to ther PMDs, needs accurate diagnosis to rule other similar lesions of oral mucosa and also

to pln for a proper treated. As an alternative to histopathological investigation, this eearch work has made an attempt to classify OLP images and normal mucosal images with the help of two pattern classifiers namely SVM and RBFNN by using histogram and BICC features extracted from the images used in the research work. An attempt was made to apply combined features with the same techniques. Experimental results show that the proposed OLP classification scheme is very effective with an accuracy rate of nearly and above 90.0 for all models. The performance was compared among SVM and RBFNN classifiers. RBFNN was found to provide the maximum accuracy for combined features, compared to SVM.

Conclusion

Few other pattern classifiers such as AANN, GMM and others could also be applied in the same manner to classify precancer and cancer images. This could bring out a better technique for oral lesion classification in the future.

REFERENCES

Rajendran R. and Sivapathasundaram B, "Shafer's Textbook of Oral Pathology", 6th Edition, Elsevier India, 2009.

R. Rajendran, "Oral leukoplakia (leukokeratosis): Compilation of facts and Figs", *Journal of Oral Maxillofac Pathol*, vol. 8, pp. 58-68, 2004.

Martorell-Calatayud A, Botella Estrada, Bayan Sebastian J. V, Sanmartin and Jimeney O, "Oral Leukoplakia: Clinical, Histopathologic and Molecular Features and Therapeutic Approach", *Actas Dermosifiliogr,* vol. 100, pp. 669-684, October 2009.

Carrozzo M and Thorpe R, "Oral lichen planus: a review", Minerva Stomatol, vol. 58, no. 10, pp. 519-537, October 2009.

Riikka Mattila, "Molecular Markers of Oral Lichen Planus", Ph.D. thesis, Medica – Odontologica, 2009.

Crispian Scully, Jose V. Bapan, Colin Hopper and Joel B. Epstein, "Oral cancer: Current and future diagnostic techniques", *American Journal of Dentistry,* vol. 21, no. 4, pp. 199-209, August 2008.

Paul R. R, Mukharjee A, Dutta P. K, Banarjee S, Pal M, Chatterjee J, Chaudhuri K and Mukherjee K, "A novel wavelet neural network based pathological stage detection technique for an oral precancerous condition", *Journal of Clinical Pathology,* vol. 58, pp. 932-938, September 2005.

Rusha Patra, Chandan Chakraborty and Jyotirmoy Chatterjee, "Textural Analysis of Spinous Layer for Grading Oral Submucous Fibrosis", *International Journal of Computer Applications,* vol. 47, pp. 0975-0987, June 2012.

Shekhar Singh, "Cancer Cells Detection and Classification in Biopsy Image", *International Journal of Computer Applications,* vol. 38, no. 3, pp. 15-21, January 2012.

Gao S, "Cell morphometric analysis in oral submucous fibrosis, leukoplakia and squamous cell carcinoma", Zhonghua Kou Qiang Yi Xue Za Zhi, vol. 27, no. 3, pp. 145-147, May 1992.

Muthu Rama Krishnan M, Mousumi Pal, Suneel K. Bomminayuni, Chandan Chakraborty, Rajan Rashmi Paul, Jyotirmoy Chatterjee and AjoyK. Ray, "Automated classification of cells in sub-epithelial connectivetissue of oral sub-mucous fibrosis – An SVM based approach", *Computers in Biology and Medicine,* vol.39, no. 12, pp. 1096-1104, December 2009.

Kramer I. R. H, El-Labban N. G and Sonkodi S, "Further Studies On Lesions Of The Oral Mucosa using Computer-Aided Analyses Of Histological Features", *British Journal of Cancer,* vol. 29, p. 223, 1974.

Muthu Rama Krishnan M, Acharya U. R, Chakraborty C and Ray A. K, "Automated Diagnosis of Oral Cancer Using Higher Order Spectra Features and Local Binary Pattern: A Comparative Study", Twin Cities Radiation Therapists, vol. 10, no. 5, pp. 391-504, October 2011.

Krishnan M. M. R, Shah P, Ghosh M and Pal M, "Automated characterization of sub-epithelial connective tissue cells of normal oral mucosa: Bayesian approach", *Proceedings of IEEE Student Technology Symposium,* pp. 44-48, April 2010.

Smitha T, Sharada P and Hc Girish, "Morphometry of the basal cell layer of oral leukoplakia and oral squamous cell carcinoma using computer-aided image analysis", *Journal of Oral Maxillofacial Pathology*, vol. 15, no. 1, pp. 26-33, January-April 2011.

****