ISSN No:-2456-2165

Tongue Pre-Cancer Stages Classification Using Machine Learning

[1]Vidhi B. Chaudhari and [2]Dr.Narendra M. Patel
[1]Assistant Professor, Ganpat University, Mehsana [2]Professor,
VM University, Vidhyanagar

Abstract:- Tongue represents the overall health of a person, so tongue examination is a one of the important tool in medical science. In the modern era, medical imaging refers to techniques used to create and process images of various parts of the human body for diagnostic and treatment purposes within digital health. Medical imaging may include various radiological imaging techniques such as: X-ray diagnostic, MRI, CT SCAN. This work represents the region of interest is found using K-means clustering algorithm. Also used image augmentation method to create dataset/ images. Extracted features namely texture, color, hue and saturation are given to Artificial Neural network (ANN) and Support Vector Machine (SVM). Classification of diagnosis and detection of two pre-cancerous stages named Erythroplakia and Leukoplakia.. By this technology doctor may diagnose and detect tongue cancer in early stages which leads to saves many lives.

Keywords:- *Erythroplakia, Leukoplakia, Pre-Cancer Stage, Medical Diagnosis.*

I. INTRODUCTION

The tongue is the considered as the mirror that reflects the conditions of the body's internal organs. The tongue also reflects the overall digestive, nutritive and metabolic conditions of the entire organism. In India, 20 per 100000 populations are affected by oral cancer, which accounts for about 30% of all types of cancer. Over 5 people in India die every hour every day because of oral cancer[1].Leukoplakia is a white or gray patch that develops on the tongue, the inside of the cheek, or on the floor of the mouth. The risk of developing tongue cancer depends on how different the abnormal cells are in shape, size and appearance compared to normal mouth cells.Erythroplakia is defined as a red patch that cannot be characterized either clinically or pathologically as any other condition. Excess chewing of tobacco/drinking is a major cause. Leukoplakia and erythroplakia are the two most common potentially malignant disorders of the oral cavity. The prognosis and overall survival of a patient with oral cancer is dependent on the early detection of any lesion that might identify a patient with higher risk than normal or with early infiltration before metastatic disease[2].Shape, colour and & texture of the tongue is most important thing for the method [3].

Xiuqin Zhong at li,demonstrated a novel method to segment the tongue image automatically with the mouth location method and active appearance model (AAM).Mouth location has to be fixed and displacement of lips and teeth would lead to wrong segmentation[4].

Wenshu Li at li,described an automatic initialization of contour by the feature of tongue in the HSV colour space. Improved level set method takes not only colour information but also tongue contour shape constraint represented by energy function between the evolving curve and parametric shape model. Colour Space Separation & Comparative of Edge Detection of binary image. It is identify crack in tongue. Only used for shape identification purpose, not used for medical efficacy purpose [5].

Chao Liang at li,explained Combination a feature of tongue shape and the Snakes Correction model. Active contour models offer attractive approach in tackling varieties of problems due to their capability of representing the complex shapes and broad shape variability of anatomical structures .We get regular tongue shape, but accuracy is not high. So necessary to further undergo the shape correction model [6].

Identification of different tongue cancer stages. Here, we are presenting published work on the use of computer technology in the tongue cancer detection as below.

II. LITERATURE REVIEW

ZHAI Xue-ming at li,presented that through the colour space conversion, image pre-processing and then the application of double Snakes algorithm, adjusting the initialization parameters of the internal and lateral contour to improve the accuracy of the segmentation. Experiments show that the accuracy of the single-Snake is 81.63%, and the accuracy of the double Snake is 92.89%. Accuracy is better than traditional

There are two precancerous stages named Leukoplakia and Erythroplakia, Leukoplakia is a white or gray patch that develops on the tongue, the inside of the cheek, or on the floor of the mouth. Erythroplakia is defined as a red patch that cannot be characterized either clinically or pathologically as any other condition[7].In Medical science, tongue cancer is diagnosed by physical examination, observation and clinical sign & symptoms, but it confirmation done after tissue biopsy. Physical examination

and observation is very primary diagnostic tool in the practice in various tongue disorders, in the examination of tongue main focus is tongue shape, colour& texture. If the physician detects cancer in early stage, they can save the life of patient by early treatment. Here computer technology can help the physician to detect the early change or pre cancer stages of analysis of tongue shape, colour & texture.

III. PROPOSED ALGORITHM

For this diagnosis, we suggest a technique as briefed below in the figure. 1. The detail description of the same is explained below it.

1. Acquire tongue images.
2. Convert RGB image into Luminance and Chrominance (LAB) colour space.
3. Select the Region of Interest (ROI).
4. Feature Extraction from image.
5. Classification into two stages named Leukoplakia and Erythroplakia,

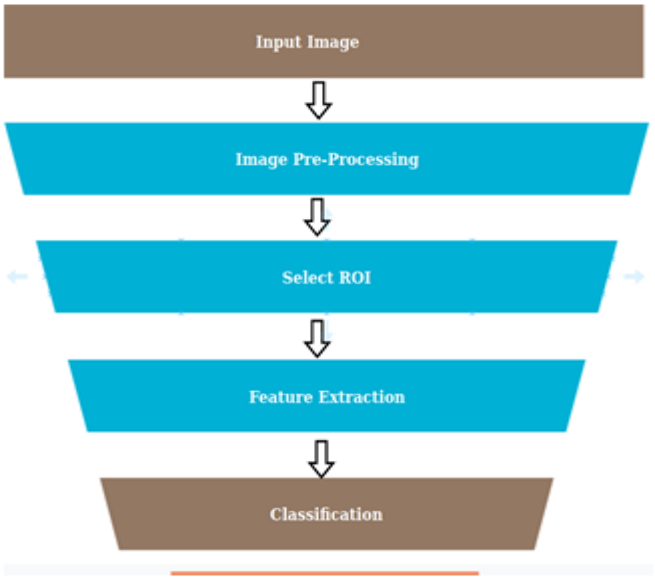

Figure.1 Proposed algorithm

Image Pre-processing

The original tongue image was captured by digital camera under standard light source situation. It usually contained tongue, body, nose, upper lip, partial lower lip and face. In which we need only the area of tongue affected part. Initially, convert RGB color image into LAB color space (The L*a*b* space consists of a luminosity layer 'L', chromaticity-layer ‚a" and „b").Where AB color space has more information so, used for further information. Classify the colors in a*b* color space using K means clustering.

Select region of Interest

For select region of interest used K-means clustering algorithm. K-means clustering is a partitioning method. The function k-means partitions data into *k* mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. The Image has created three clusters on the basis of measure the distance using Euclidean Distance Metric. Here we got three clusters which clusters are nearest to the affected portion. We need to select that cluster. Figure.2 is describing the result of implementation.

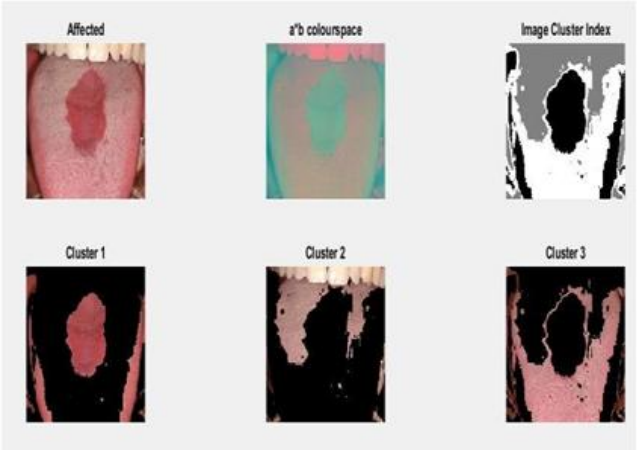

Figure.2 Output screenshot

Feature Extraction

Here we have selected the colour as a feature and find the Mean, Variance, Standard Deviation, Skewness of RGB colour image and also select the Gray Level Co-occurrence Matrix (GLCM) and associated texture feature. The GLCM is a tabulation of how often different combinations of gray level co- occur in an image or image section. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity in the image. Texture variable measures are Contrast, Correlation, Energy, Entropy and Homogeneity. Hue refers to the features that determined by the dominant wavelength of the visible spectrum. It is the attribute that permits colors to be classified as red, yellow, green, blue, or an intermediate color. Saturation as feature defines the strength or purity of the color and represents the amount of gray in proportion to the hue.

Energy feature

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2$$

……..1)

Entropy feature

$$Entropy = \sum_{i,j=0}^{N-1} -\ln\left(P_{ij}\right) P_{ij}$$

……..4)

Contrast feature

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2$$

……..2)

Homogeneity feature

$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$$

……..5)

.Correlation feature

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

……..3)

Where;
Pij = Elements i,j of the normalized symmetrical GLCM
N= Number of gray level in the image
M=GLCM Mean

ISSN No:-2456-2165

$$\mu = \sum_{i,j=0}^{N-1} iP_{ij}$$

Classification

SVM (support vector machine) method used for classification. SVM classify function classifies each row of the data in Sample, a matrix of data, using the information in a support vector machine classifier structure SVMStruct, created using the svmtrain function. Like the training data used to create SVMStruct, Sample is a matrix where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable. Therefore, data must have the same number of columns as the training data. This is because the number of columns defines the number of features. Group indicates the group to which each row of Sample has been assigned. This plot appears only when the data is two-dimensional. That classifies into two classes named lukoplakia and erytroplakia.Figure.3 is describing the classification pop-up box using svm model.

Another way of classifying the pre-cancer stage is to use neural network model. The Artificial neural network (ANN) model consists of input data, hidden layer neuron and classified output as its components. We have to perform the analysis on data for different iterations and generate the results. We performed the analysis on pre-cancer dataset using neural network models. In the following section we show you the results in a table.1.

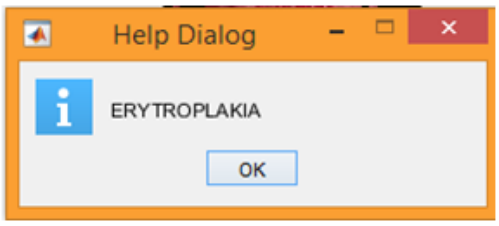

Figure.3 Classification pop-up box(Output)

Image augmentation

Image augmentation method used to Increase dataset. I convert 164 images into 540 images, by using following operation, Rotate 90, Rotate 60, Flip, Scaling.

IV. RESULT AND DICUSSION

Sr.no	Hidden layer	Iteration	Epoch	Accuracy
1.	3	10	500	88.89%
2.	5	10	100	86.95%
3.	5	50	500	91.54%
4.	10	50	1000	88.00%
5.	50	100	500	88.49%
6.	100	100	500	83.53%
7.	200	200	500	80.01%

Table.1 Neural network result

Feature	Classification Method	Accura cy
Color	Support Vector Machine(SVM)	75.60%
Color,Texture	Support Vector Machine(SVM)	88.89%
Color,Texture	Artificial Neural Network (ANN)	91.99%
Color,Texture,HS	Support Vector Machine(SVM)	91.54%
Color,Texture,HS	Artificial Neural Network (ANN)	94.99%
Color,Texture,HS	Support Vector Machine(SVM) 540 Augmented images	93.21%
Color,Texture,HS	Artificial Neural Network (ANN) 540 Augmented images	96.17%

Table.2 Result comparison

V. CONCLUSION

This work presents an detail work of applying k-means clustering for automatic select region of interest. Color, texture and HS (Hue, saturation) feature are used to classify tongue pre-cancer stages Leukoplakia and Erythropakia. We have used two classifier names SVM and ANN. We got 91.54% accuracy in SVM And 94.99% in ANN .Also used augmentation method to increase dataset images, after used of 540 augmented images we got 93.21% accuracy in SVM and 96.17% Accuracy in ANN. For this dataset and mentioned methods Artificial neural network (ANN) is more accurate than Support vector machine (SVM).

The way of diagnosing the disease is using SVM (support vector machine) method. The initially requirement of data collection and classification need to be accomplished. By Using three features color, texture, HS (Hue,Saturation) Around 164 images with 16 feature vector was collected for pre cancer dataset. Output values 1 for Lukoplakia 2 for Erytroplakia. Among 164 images, where 70% data used for training data and 30% for testing. Result comparisons are mentioned in a table.2.

ACKNOWLEDGMENT

The authors are highly indebted and thankful to **Dr. Darshak Chaudhari (Eva Dental Clinic & Implant Center, F- 154, Shrijidada complex, Nr. marketyard, Bayad, Arvalli)** for his massive supports and suggestions. We are really grateful to **Dr. Darshak Chaudhari** for sharing the personal information of various patients for the purpose of our research.

ISSN No:-2456-2165

REFERENCES

[1]. Varshitha . A ,"Prevalence of Oral Cancer in India",journal of pharmaceutical science and research,2015.
[2]. https://www.ncbi.nlm.nih.gov/pubmed/23957095
[3]. ZHAI Xue-ming, LU Hang-dong , ZHANG Li-zhong "Application of Image Segmentation Technique in Tongue Diagnosis" International Forum on Information Technology and Applications, 2009.
[4]. Xiuqin Zhong and Hongguang Fu "Automatic Segmentation in Tongue Image by Mouth Location and Active Appearance Model" , Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.
[5]. Wenshu Li, Shenning Hu,Shuai Wang, Su Xu "Towards the Objectification of Tongue Diagnosis: Automatic Segmentation of Tongue Image" Industrial Electronics,35th annual conference of IEEE, 2009.
[6]. Chao Liang, Dongcheng Shi "A Prior Knowledge-based Algorithm for Tongue Body Segmentation" International Conference on Computer Science and Electronics Engineering, 2012
[7]. http://www.cancer.ca/en/cancer-information/cancer-type/oral/oral-cancer/precancerous-conditions/?region=on
[8]. E A Patel , kavindra jain, Nirav P. Desai"Optimal edge detection technique for diagnosis & treatment of Erythroplakia in Oral Mucosa" , International Conference On Innovations In Electrical And Electronics Engineering (ICIEEE),2016.
[9]. E A Patel , kavindra jain, Nirav P. Desai, "Optimal Edge Detection Technique for diagnosis & treatment of Lukoplakia in Oral Mucosa" Conference on Computing, Communication, Electrical, Electronics, Devices and Signal Processing (CCEEDS), Andhra Pradesh (To be published in IEEE/ Springer Journal),2015.
[10]. Du Jian-qiang, Lu Yan-sheng1, Zhu Ming-feng, Zhang Kang, Ding Cheng-hua "A Novel Algorithm of Color Tongue Image Segmentation Based on HSI" International Conference on BioMedical Engineering and Informatics, 2008 .
[11]. Shengyang Yu, Jie Yang, Yonggang Wang, Yan Zhang "Color Active Contour Models Based Tongue Segmentation in Traditional Chinese Medicine" Bioinformatics & Biomedical Engineering, ICBBE, 2007
[12]. L. Shen, B. Wei, Y. Cai, X. Zhang And Y. Wang, "Image Analysis For Tongue Characterization", Chinese Journals Of Electronics, 2003, 12(3):317-323.
[13]. B. Pang, D. Zhang, And K. Wang, "Tongue Image Analysis Forappendicitis Diagnosis," Transactions on Information Sciences, Vol. 175, No. 3, Pp: 169-176, 2005.
[14]. Zhou Yue, Shen Li, YangJie, "Feature Analysis Method of Tongue Image by Chinese Medical Siagnosis Based On Image Processing", Infrared and Laser Engineering, 2002, Vol. 31 No. 6, 490-494.
[15]. Jia Wu, Yonghong Zhang, Jing Bai, "Tongue Area Extraction in Tongue Diagnosis of Traditional Chinese Medicine," IEngineering in Medicine and Biology Society, IEEE-EMBS 2005. 27th Annual International Conference of the Volume , Issue , 2005 pp. 4955 - 4957
[16]. Jijun,Ren,"Color Tongue Image Segmentation Algorithm Based On Grey Space Auto-Thresholding Selection," Journal of Shaanxi University of Science and Technology(Natural Science Edition) vol.23,1995, pp.73 - 77
[17]. Bo Pang, Kuanquan Wang, David Zhang, Fengmiao Zhang, "On Automated Tongue Image Segmentation in Chinese
[18]. Medicine," icpr, p. 10616, 16th International Conference on Pattern Recognition (ICPR'02) Volume 1, 2002[i]