JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

HYBRID APPROACH FOR THE DETECTION OF ERYTHROPLAKIA ORAL CANCER USING SVM AND RNN WITH GLRLM FUSION MODEL

Dr. Pandikumar S, Dr.M.Sengaliappan

Assistant Professor, Department of Computer Applications, Yadava College, TamilNadu, India
Associate Professor , Department of Computer Applications, Nehru College of Management

Abstract - Erythroplakia is an infrequent, quarantined, red, velvety lesion which affects patients mainly in the 6th and 7th decades. This study proposes a hybrid approach for the detection of oral cancer by integrating Support Vector Machine (SVM) and Recurrent Neural Network (RNN) classifiers, utilizing features extracted from Gray Level Run Length Matrix (GLRLM) analysis of medical images. GLRLM, a texture analysis method, captures important textural features that are indicative of malignancy in oral tissues. These features are then fed into SVM and RNN models, which are trained to differentiate between cancerous and non-cancerous lesions. The SVM classifier is employed for its robustness in handling high-dimensional data and separating classes with a maximum margin, while the RNN leverages its capability to learn sequential dependencies, enhancing the model's performance in capturing complex patterns within the image data. Experimental results demonstrate that the combined SVM and RNN approach with GLRLM features significantly improves the accuracy, sensitivity, and specificity of oral cancer detection compared to conventional methods. This hybrid model showcases the potential of advanced machine learning techniques in medical diagnostics, paving the way for more reliable and efficient early detection of oral cancer.

Keywords - Erythroplakia, GLRLM, RNN,SVM, Naive Bayes

1.INTRODUCTION

Erythroplakia is a heterogeneous presentation of combined red and white surface alterations are noted, with an intermingling of these changes characteristic of erythroplakia noted at the lateral aspect of the soft palate and buccal mucosal interface. The tissue diagnosis was squamous cell carcinoma, minimallyinvasive.

However, the erythroplakia is characterized by a smooth, velvety clinical presentation with a homogeneous surface, without ulceration. The tissue diagnosis was oral squamous cell carcinoma in situ.

A biopsy of Erythroplakia involves taking a sample tissue from the afflicted region and sending it to a lab for analysis. To check for aberrant cells that could point to cancer or other disorders, the sample is inspected under a microscope.

During this approach, a way for classification of traditional and abnormal lesions planned with totally different feature extraction algorithms and a comparison between these algorithms administrated to get the simplest feature extraction technique. This approach provides the classification of lesions from MRI images.

Supervised machine learning techniques helps to detect and diagnose oral carcinoma at an early stage and it increases the patient's survival rate.

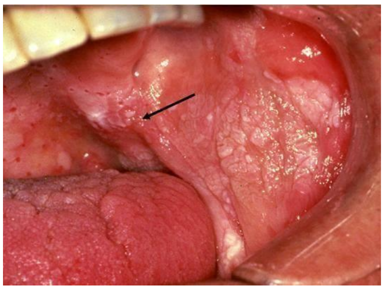

Fig. 1.1 Erythroplakia Image

2.METHODOLOGY

The proposed framework consists offollowing main steps which are

1) Give Oral MRI image of Mouth asinput.
2) Convert it to gray scale image.Apply intensity

Histogram on theimage.

3) Extracting features by using GLRLM
4) Classify with SVM and RNN based on features

The architecture workflow depicts the steps involved and the corresponding input and output. The methodology used for MRI Mouth images is as shown in Fig. 2.1

2.1 Generalized View

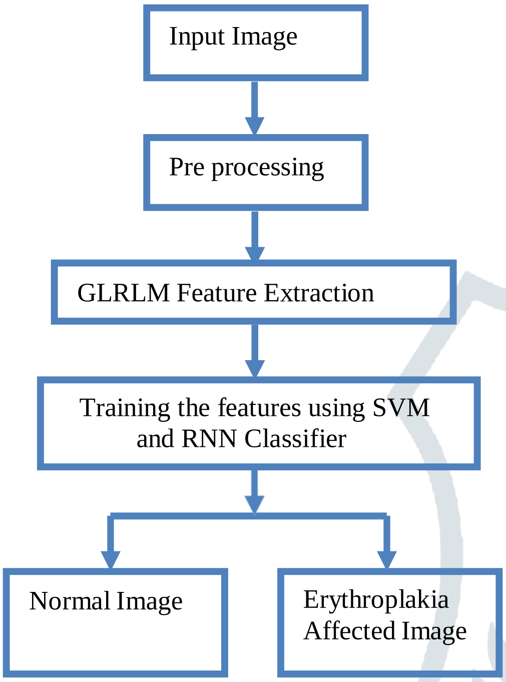

Fig. 2.1 Block diagram of the system

2.2 Image Acquisition

Normal and Erythroplakia affected tissue images were collected from patients of Raja Muthiah Dental College and Hospital (RMDC & H). The research consists of 70 images which belong to two types as Normal and Abnormal. There are 44 normal 26 abnormal (Erythroplakia Affected) images were used for training the SVM and RNN classifier.

2.3. Image Preprocessing

After obtaining digital images, image preprocessing techniques can be further used for analysis of region of interest. Load the image data using the Image Processing Toolbox.The pre-processing is used to read the input image into the MATLAB(R2023 a) and also to remove the noise present in the image. Image preprocessing consists mainly of following steps.

- Resize
- Gray conversion
- Median filter

2.3.1 Intensity Histogram

A frequently used approach for texture analysis is based on statistical properties of Intensity Histogram. A histogram is a statistical graph that allows the intensity distribution of the pixels of an image, i.e. the number of pixels for each luminous intensity, to be represented. By convention, a histogram represents the intensity level using X-coordinates going from the darkest (on the left) to lightest (on the right). Thus, the histogram of an image with 256 levels of grey will be represented by a graph having 256 values on the X-axis and the number of image pixels on the Y-axis. The histogram graph is constructed by counting the number of pixels at each intensity value.

2.4 Feature Extraction

The third step of the proposed work is feature extraction. Transforming the input data into the set of features is called feature extraction. Features are used as inputs to classifiers that assign them to the class that they represent. In this work Gray Level Run Length Matrix (GLRLM) features are extracted.

2.4.1 Gray Level Run Length Matrix

A gray level run –length matrix (GLRLM) method is a way of extracting higher order statistical texture measures. A set of consecutive pixels with the same gray level, collinear in a given direction, constitute the gray level run. The run length is the number of pixels in the run and the run length value is the number of times such a run occurs in an image.

The GLRLM is a two dimensional matrix in which each element p(i, j |θ) gives the total number of occurrences of runs of length “j” at gray level “i” in a given direction θ.

The enhanced image is segmented to detect tumor. Intensity Histogram and GLRLM methods are used to extract features from the image. Next, Support Vector Machine (SVM) and RNN classifier is used to classify the tumor as benign and malignant.

Algorithm

1) *Step 1:* Read and preprocess the Image
2) *Step 2:* Compute GLRLM (Use grayrlmatrix function to compute the GLRLM
3) *Step 3:* Extract GLRLM Features (Use the grayrlprops function to ecxtract statistical properties from the GLRLM).

Sl. No.	Features	Formula
1	Short Run Emphasis	$\dfrac{\sum_i \sum_j p(i,\ j)}{j^2}$
2	Long Run Emphasis	$\dfrac{\sum_i \sum_j j^2 p(i,\ j)}{\sum_i \sum_j p(i,\ j)}$
3	Grey Level Non-uniformity	$\dfrac{\sum_j (\sum_i p(i,\ j))^2}{\sum_i \sum_j p(i,\ j)}$
4	Run-Length Non-uniformity	$\dfrac{\sum_i (\sum_j p(i,\ j))^2}{\sum_i \sum_j p(i,\ j)}$
5	Low Grey Level Run Emphasis	$\dfrac{\sum_i \sum_j \frac{p(i,\ j)}{i^2}}{\sum_i \sum_j p(i,\ j)}$
6	High Grey Level Run Emphasis	$\dfrac{\sum_i \sum_j i^2\ p(i,j)}{\sum_i \sum_j p(i,\ j)}$

MATLAB Code

```
% Step 1: Read and preprocess the image
image = imread('path_to_image.jpg');
if size(image, 3) == 3  % Check if the image is RGB
    grayImage = rgb2gray(image);  % Convert to grayscale
else
    grayImage = image;  % Image is already grayscale
end

% Step 2: Compute the GLRLM
numGrayLevels = 8;
offset = [0 1];  % Horizontal direction
glrlm = grayrlmatrix(grayImage, 'NumLevels', numGrayLevels,
'GrayLimits', [], 'Offset', offset);

% Step 3: Extract GLRLM features
properties = {'ShortRunEmphasis', 'LongRunEmphasis',
'GrayLevelNonuniformity', ...
              'RunLengthNonuniformity', 'RunPercentage',
'LowGrayLevelRunEmphasis', ...
              'HighGrayLevelRunEmphasis'};

stats = grayrlprops(glrlm, properties);

disp('GLRLM Features:');
disp(stats);
```

B. References

In 2019, Textures are one of the most important features in computer vision for many applications. Texture Feature Extraction is a method of capturing visual content of images for indexing & retrieval. General features such as extraction of color, texture and shape or domain specific features. GLCM to extract statistical texture features such as Contrast, Correlation, Energy, Entropy and Homogeneity. GLRLM to extract run length features such as SRE, LRE, GLN, RLN, LGRE and HGRE. Constructing combinations of the various extraction methods (GLCM & GLRLM) to get the data with sufficient accuracy. The experimental results demonstrated that texture feature extraction based on the KNN technique achieved a better image recognition, and the accuracy of classification based on this method has been significantly improved and that it requires less computation time efficiently used for real time Object TrackingApplications. [1].

In 2021, Aarti Nayak, Samir Pol, Swaraaj Singh, Prof. Anagha Patil published the paper " Unconscious Oral Cancer Prediction using Supervised Learning" Since the beginning of time, the disease of cancer has been incurable and intimidating. However owing to the tremendous advancement in the field of technology, it is smoothly curable provided detected in the earliest of stage possible. Pre-cisely considering oral cancer, it is the phenomenon of exponential increase in the number of cells which in turn starts damaging the surrounding and neighbouring cells.

In spite of the availability of supremely advanced radiation therapy and chemotherapy, the death rate prevailing is very disappointing and increasing. However an early prediction of the same might help to curb this problem. In order to provide with a substantial solution for the same, we propose to perform a comparative analysis of the supervised learning techniques under the domain of machine learning using the accuracy and time complexity approach to design an effective model using the considered data set of the victim to help predict the unconscious cancer in a user so that he/she can work towards the appropriate line of treatment and also make the suggested lifestyle changes. The aim of this paper is to act as a detailed guide for all to develop a system on similar guideline [8].

2.5 Classification

The goal of the SVM algorithm is to form the best line or decision boundary which will segregate n-dimensional space into categories so we are able to simply place the new information within the correct class in the future. This best decision boundary is named a hyperplane. It utilizes a technique called the kernel trick to switch your information and so supported these changes it identifies an optimum boundary among the potential output.

Model Implementation:

SVM: Use the Statistics and Machine Learning Toolbox to train and validate an SVM classifier on the extracted GLRLM features.

RNN: Use the Deep Learning Toolbox to design, train, and validate an RNN on the image data. Preprocessed image sequences or extracted features can be used as input.

Hybrid Model Development: Combine the predictions from the SVM and RNN models to develop a hybrid model, potentially improving classification performance.

2.5.1 ***Data Description:*** The MRI image data description of the proposed method is shown in table.

Table 2.1: Dataset Description

S.No.	Type of Image	Number of Images
1	Normal Images	44
2	Abnormal Images	26

3.RESULTS AND DISCUSSION

An experiment has been conducted on a MRI oral image data set based on the proposed flow diagram as shown in Fig 3.1.

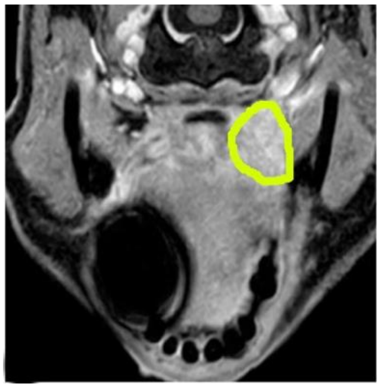

Fig. 3.1 Erythroplakia Affected Image

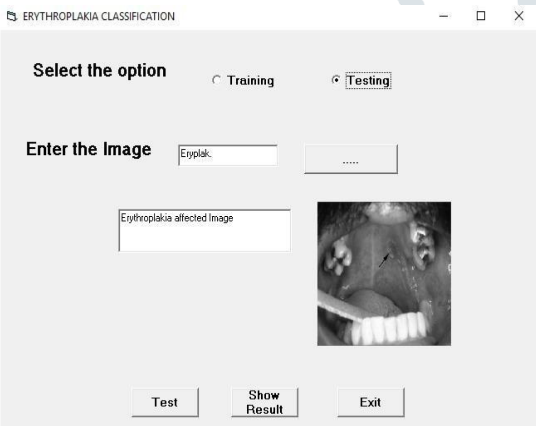

Fig. 3.2 Testing Snapshot of Erythroplakia Erythroplakia Affected Image

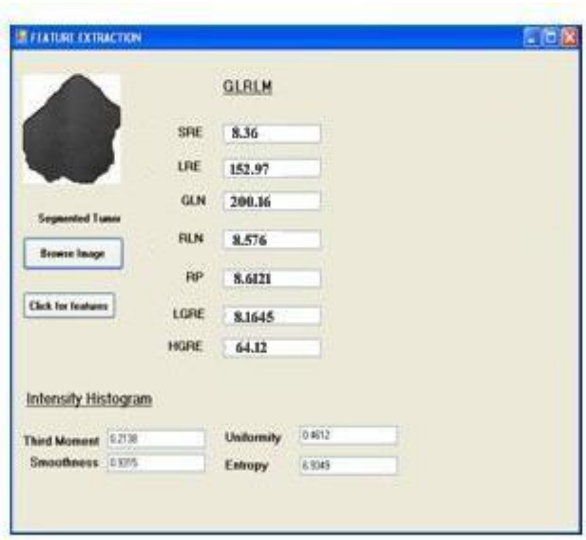

Fig. 3.3 GLRLM Feature extraction of Erythroplakia Affected Image

GLRLM Features	Cyst Lesion		Tumor Lesion	
	Min. Value	Max. Value	**Min. Value**	Max. Value
SRE	0.2193	0.8226	0.3392	0.7997
LRE	1.9790	109.3407	2.8501	182.95
GLN	39.3656	273.6864	37.0880	278.05
RP	0.1431	0.7725	0.1513	0.683
RLN	22.8428	745.6715	54.7440	641.19
LGRE	0.0261	0.3054	0.0264	0.1718
HGRE	9.1975	46.5710	12.5282	63.941

Table 3.1: GLRLM Features of Cyst andTumour lesion

MEASURE	INTENSITY HISTOGRAM	GLRLM
Accuracy	88%	92%
Sensitivity	85%	88%
Specificity	90%	95.45%

Table 3.2: Comparative Results

Confusion Matrix: The confusion matrix is used to measure th performance of two class problem for the given data set. The righ diagonal elements TP (true positive) and TN (true negative) correct classify Instances as well as FP (false positive) and FN (false negativ incorrectly classify Instances. Confusion Matrix Correctly Classifi Instance and TP+TN Incorrectly Classifiy Instance FP+FN.

Value	Intensity Histogram	GLRLM
TP	44	44
FP	2	1
TN	4	3
FN	20	24

Table 3.3: Confusion Matrix for ProposedMethod

Accuracy: It is defined as the ratio of correctly classified instances to total number of instances. The effectiveness of the proposed method has been estimated using accuracy measure.

Accuracy = (TP + TN) / (TP+TN+FP+FN)

where,

True Positive (TP): Number of Abnormal images correctly classified

False Positive (FP): Number of Normal images classified as Abnormal

True Negative (TN): Number of Normal images correctly classified

False Negative ***(FN):*** Number of Abnormal images classified as Normal. The derived features sets are then applied for classifying the new dataset of MRI images as normal and abnormal cases automatically. The result of the prediction provides 92% accuracy in classification.

4. CONCLUSION

In this paper, a hybrid method leverages the strengths of both machine learning and deep learning techniques to enhances the detection performance , showcasing its potential for early diagnosis and intervention in oral cancer. Experimental results provide accuracy of 92%, sensitivity of 88% and specificity of 95.45% is found in classification of oral cancer. This advancement holds significant potential in clinical practice, offering clinicians a valuable tool for prompt identification and management of Erythroplakia, ultimately leading to improved patient outcomes and quality of life.

Future work will focus on expanding the dataset, incorporating multi-modal data, and improving model interpretability to facilitate wider adoption in clinical practice.

ACKNOWLEDGMENT

I would like to thank with overwhelmed gratitude, the enormous support and guidance rendered to us by Dr. Napoleon, Associate Professor, Dept. of Computer Science, Bharathiar University, Tamil Nadu,India.

REFERENCES

[1] Prof. Mrs. K. ChandraPrabha, C. Bharathi, Texture Analysis using GLCM & GLRLM Feature Extraction Methods, International Journalfor Research in Applied Science & Engineering Technology (IJRASET**),** Volume 7 Issue V, (pp. 2059-2064), 2019.

[2] K. Anuradha, K. Sankaranarayanan, "Identification of Suspicious Regions to Detect Oral Cancer at Earlier Stage-A Literature Survey," International Journal of Advances in Engineering & Technology, ISSN: 2231-1963, Vol 3,Issue 1, (pp. 84-91), 2012.

[3] Pandikumar S, Sureshkumar C, Venkatakrishnan S , Classification of Leukoplakia Oral Disease using ANFIS Classifier, Design Engineering, ISSN: 0011- 9342, Issue: 8 | (pp.15548-15557), 2021.

[4] K Thongprasom, Oral lichen planus: Challenge and management, John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.(pp.172-173), 2018.

[5] Pandikumar S, Sureshkumar C, Venkata krishnan S, Oral Cancer Classification Using SVM Classifier, International Journal of Research and Analytical Reviews (IJRAR),Volume 6,Issue 2, (pp.665-669), 2019.

[6] Bahzad Taha Jijo, Adnan Mohsin Abdulazeez ,Classification based on Decision Tree Algorithm for Machine learning, Journal of Applied Science and Technology Trends,Vol 02, No. 01, (pp.20-28), 2021.

[7] Gonzalez, R.C. Richard, E.W; “Digital Image Processing,”, II Indian Edition, Pearson Education, New Delhi, India.2004.

[8] Aarti Nayak, Samir Pol, Swaraaj Singh, Anagha Patil, Unconscious Oral Cancer Prediction using Supervised Learning, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 10 Issue 03,(pp.562-568), 2021.

[9] Pandikumar S, Sureshkumar C, Venkatakrishnan S, Classification of Oral Lichen Planus using Decision Tree Classifier, Proceedings of International Conference on Research Practices of Cutting Edge Technologies in Computing-ICRPCETC 2021, ISBN 978-93-91977-06-1, (pg. 5-10), 2021.

[10] K.Anuradha, Dr.K.Sanakaranarayanan, Comparison of Feature Extraction Techniques to classify Oral Cancers using Image Processing, International Journal of Application or Innovation in Engineering & Management (IJAIEM),Volume 2, Issue 6,(pp.456- 462)2013.

[11] Kassnera and R. E. Thornhilla , “Texture analysis: a review of neurologic MR imaging applications,” American Journal of Neuroradiology, vol. 31, no. 5,(pp.809–816), 2010.

[12] R. M. Haralick, K. Shanmugam, and I. Dinstein , “Textural features for image classification,” IEEE Transactions on Systems, Man and Cybernetics, vol. 3, no. 6,(pp. 610–621), 1973.

[13] L. Liu and P. Fieguth, “Texture classification from random features,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 3, (pp. 574– 586), 2012.

[14] S.Saraswathi, “Detection of lung cancer in CT Images and classification of pulmonary nodules using an efficient three dimensional texture feature and machine learning algorithms” Ph.D. diss. Dept. of Computer Science, Bharathiar University, 2018.

[15] J. M. Guerrero, R. Martínez-Tomás, M. Rincón, and H. Peraita, “Diagnosis of Cognitive Impairment Compatible with Early Diagnosis of Alzheimer's Disease: A Bayesian Network Model based on the Analysis of Oral Definitions of Semantic Categories,” Methods of Information in Medicine, vol. 55, (pp. 42-49), 2016.