Volume: 11 Issue: 03 Pages: 4294-4297 (2019) ISSN: 0975-0290

Oral Cancer Detection: Feature Extraction & SVM Classification

Shilpa Harnale
Research Scholar, CSE Dept GNDEC, Bidar
Emailshravi97@rediffmail.com
Dr. Dhananjay Maktedar
HOD, CSE Dept GNDEC, Bidar
Email: dhan_mak@yahoo.com

ABSTRACT

Oral or mouth neoplasm is the type of head & neck cancers. This type of cancer starts in the throat or mouth due to uncontrollable growth of tissues, and it looks like a lump or bump. In the pre- processing step, anisotropic diffusion filter used to filter unwanted distortions from MRI image. Next, the lesion separated from MRI image using a hybrid approach KFCM clustering in segmentation and features extracted using Intensity of Histogram, GLCM & GLRLM. The comparison between these three algorithms is performed to obtain the best feature extraction technique. Next, SVM classifier used to classify the lesion. Classification accuracy obtained for the developed system is 98.04% using GLRLM feature extraction technique.

Keywords**: KFCM Clustering, FO, GLCM, GLRLM, SVM.**

Date of Submission: Nov 02, 2019 Date of Acceptance: Dec 23, 2019

I. INTRODUCTION

Oral neoplasm is a type of carcinoma that occurs in the oral cavity or mouth. It called squamous-cell carcinoma. More than 90 percent of carcinoma occurs in the oral cavity or mouth. Oral lesion is an abnormal growth of squamous-cell. The lesion may be cancerous or non-cancerous. The oral lesion is cancerous or non-cancerous is suggested by the dentist. If the lesion is cancerous, then only go for a biopsy examination for confirmation. It is an invasive and painful process. Many advanced techniques are available today for detecting & diagnosing oral cancers. Radiologist recommends the patients to go for biopsy testing only after identifying an abnormal lesion in MRI image because biopsy is an invasive, expensive, and painful task to the patients. In this approach, a method for classification of normal and abnormal lesions proposed with different feature extraction algorithms and a comparison between these algorithms is carried out to obtain the best feature extraction technique. Detection of oral neoplasm in an earlier stage can be curable. This approach provides the classification of lesions from MRI images.

Input MRI image acquired and preprocessed using anisotropic diffusion filter in image pre- processing. Then, the lesion part is segmented using a hybrid approach of KFCM clustering, and the features of the lesion are extracted using first order & second-order statistics. Later, SVM classifier used to classify abnormal lesion & normal lesion.

This paper is organized as follows: Chapter 2 provided Literature Study. Chapter 3 described the Implementation of proposed approach. The Performance Evaluation discussed in Chapter 4. Finally, Chapter 5 described the Conclusion & Future Work. For this, Matlab R2017a used over the 64-bit operating system on Intel (R) dual-core processor with inbuilt 2 GB RAM and 500GB hard disk.

II. LITERATURE REVIEW

For the early-stage detection and diagnosis of oral lesions, many experts and researchers used various methods.

A morphological mathematical watershed algorithm proposed to detect suspicious lesion at an earlier stage [1]. Various cancer detection and classification methods compared in this study. Also, the comparison between multiple algorithms for the detection, segmentation, feature extraction, and classification of oral cancer explained in detail.

A combination of the segmentation algorithm proposed to achieve high accuracy [2]. In preprocessing, the Gaussian filter used to reduce unwanted distortions & smooth out the dental panoramic images. Next, selection of ROI algorithm used for segmentation. Then features extracted using FO, GLCM, and GLRLM. Finally, SVM classifier used to classify tumors and performance evaluated using receiver operating characteristic curve. In this study, high accuracy achieved by the integration of FO and GLRLM algorithms.

A marker controlled watershed segmentation algorithm proposed [3]. This algorithm is suitable to separate boundaries & edges of an image & to detect the tumor at an earlier stage. Then, GLCM used for features extraction. Finally, tumors classified using SVM classifier. The classification accuracy achieved for the developed system is 92.5%.

Volume: 11 Issue: 03 Pages: 4294-4297 (2019) ISSN: 0975-0290

A marker Controlled Watershed algorithm proposed for segmentation [4]. GLCM, GLRLM, and Intensity Histogram algorithm used for feature extraction and SVM classifier used for classification. In this study, first-order and second-order statistics algorithms compared to achieve high accuracy. The classification accuracy obtained by using GLCM feature extraction technique is 96%, and MCC is 0.92. Performance evaluated by GLCM is efficient than the intensity of histogram & GLRLM feature extraction techniques. Tumors classified using Support Vector Machine classifier.

An improved MCWS algorithm proposed for the detection of oral cancer [5]. This approach used linear contrast stretching to acquire filtered image. In existing plan, the Watershed algorithm used for segmentation. MCWS used to decrease the problem of over-segmentation. In the proposed method, an improved MCWS algorithm used for improving segmentation accuracy. It also used to overcome the problem of over-segmentation. In this paper, the computation time is estimated before contrast-stretching and after contrast-stretching.

A hybrid approach of FCM and neutrosophic algorithm proposed for segmentation of lesions[6]. This approach used the median filter to eliminate the speckle noise, and the neutrosophy algorithm to reduce image noise of panoramic images. The developed approach is an efficient method to segment the oral lesions.

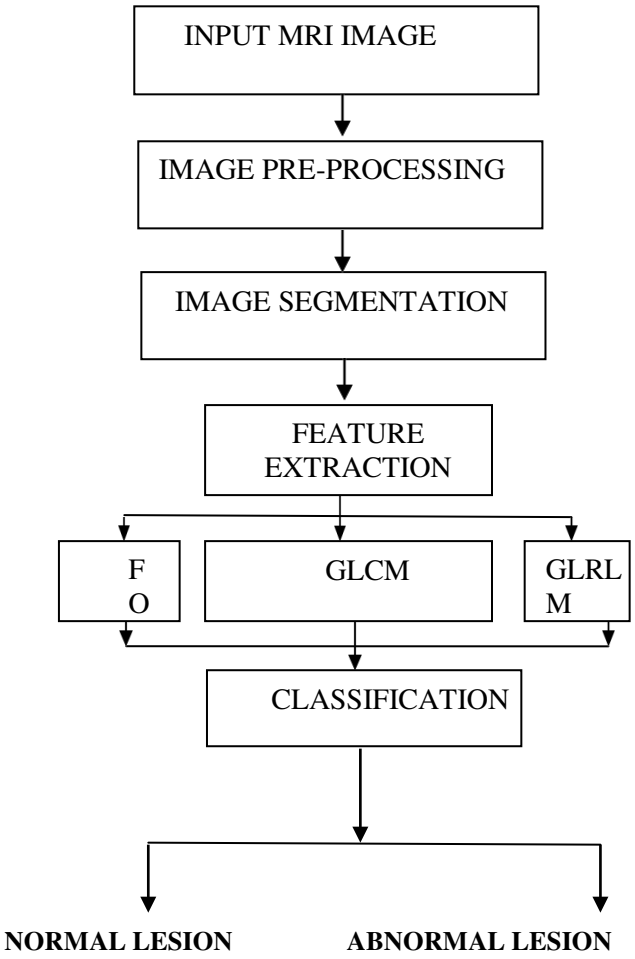
In this, the lesion region is accurately segmented, but computation time is higher than other algorithms. A fuzzy c-means segmentation algorithm for oral cancer detection & the integration of FO & GLRLM algorithms for the feature extraction proposed [7]. Further, normal and abnormal tumors classified using support vector machine classifier.

This approach used the anisotropic diffusion to eliminate the unwanted artifacts and acquire the good quality of CT image in image enhancement. Further, FCM clustering used for segmentation of CT image and feature extraction is performed by the integration of FO and GLRLM. SVM classifier used for the binary classification of lesions. The result of the developed approach achieved 90.11% accuracy, 87.5% specificity, and 92.16% sensitivity.

III. IMPLEMENTATION OF PROPOSED SYSTEM

This approach consists of different stages: 1) Pre-processing of image 2) Segmentation of image 3) Feature Extraction of lesions 4) Classification of lesions.

In the pre-processing step, anisotropic diffusion filter used to remove unwanted distortions and noise from MRI input image. Later, the enhanced image is segmented using a hybrid approach of KFCM to detect lesion. Features extracted from the lesion area using Intensity Histogram, GLCM, and GLRLM algorithms. Next, SVM classifier used to classify lesions.



A. Figure 1. Proposed System

Finally, the comparison made between feature extraction algorithms to find an efficient algorithm to achieve high classification accuracy.

1. Image Preprocessing:

In this process, the input image is enhanced to acquire a filtered image using one of the filtering methods. The proposed method used the anisotropic diffusion filter to remove unwanted distortions and other artifacts from the MRI input image**.**

2. Image Segmentation:

It is the division of a digital image into a set of inter-related and intra-related pixels. In this study, segmentation is performed by a hybrid approach of KFCM, in which both clustering algorithms work separately. The main aim of merging these two methods is to reduce execution time & segment lesion part accurately.

Initially, K-Means clustering method segmented an enhanced MRI image. Further, the FCM clustering method segmented the output cluster of k-means. Later, morphological operations used to separate lesion appropriately from lesion predicting cluster of fuzzy c-means clustering.

Volume: 11 Issue: 03 Pages: 4294-4297 (2019) ISSN: 0975-0290

3. Feature Extraction:

It is a method of converting input data into a set of features in a proper format. It is a process of obtaining useful & desired information from the image. After segmentation, features of lesion image extracted depending on the detail information of lesions like as color, shape, texture, or context.

In this paper, we used statistical texture-based feature extraction. Two types of texture feature considered in this study, like first-order and second-order statistics. Consider intensity histogram in first-order & GLCM & GLRLM in second-order statistics. In this approach, features are extracted from the lesion using first order (intensity of histogram) & second-order (GLCM, GLRLM).

The first order(Intensity of Histogram)

It provides various statistical features of the intensity of histogram of an image. It depends on individual pixel values and does not relate to neighboring values. The first- order statistics are intensity histogram and intensity features. It included mainly four properties like as mean, standard deviation, skewness & kurtosis.

Second Order(GLCM & GLRLM)

GLCM:

GLCM stands for gray level co-occurrence matrix where the number of rows and columns is equal to the number of distinct gray levels. Generally, it consists of fourteen features. Out of these, only five features used in this study like Energy, Contrast, Entropy, Correlation, and Homogeneity.

GLRLM:

GLRLM stands for Gray Level Run Length Matrix. It is the number of runs with pixels of gray level i & run-length j for a given direction. GLRLM generated for every segment. A gray level run is a set of consecutive pixels with the same gray level. The run length is the number of pixels in a run. GLRLM computed for any orientation. Run length is the length of the particular direction of pixels with the same gray intensity.

This approach used features from GLRLM are SRE, LRE, GLN, RP, RLN, LGRE, HGRE.

Classification

The final step of the proposed system is classification. In this study, a binary classification algorithm used for classification. SVM classifier used to classify the lesion.

SVM Binary Classifier:

It is a supervised learning model with associated learning algorithms that analyze data and recognize patterns used for classification and regression analysis.

SVM is a binary classification method to map input data from the original feature space to higher dimensional feature space. The training phase and the testing phase are the two phases of the classification process. The trained data is known data in the phase of training, and test data is unknown data in the phase of testing.

IV. Performance Evaluation

Performance evaluation of the MRI cases performed by consulting a radiologist working at a JJM Medical college & hospital, davangere.

To test the system, we collected the dataset of 46 MRI images containing normal lesions and abnormal lesions. Out of 46 sample images, 20 normal lesion images, and 20 abnormal lesion images selected randomly to evaluate the effectiveness of the system. The three metrics used for assessing the performance of the developed method, including classification accuracy, sensitivity, and specificity.

Classification Accuracy: It is the total number of correctly predicted MRI images with normal lesions and abnormal lesions over the total number of MRI images.

$$B.\ Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)}$$

Sensitivity: It refers to the True Positive rate, number of correctly predicted MRI images that have normal lesions.

$$C.\ Sensitivity = TP / (TP+FN)$$

Specificity: It refers to the True Negative rate, number of correctly predicted MRI images that have abnormal lesions.

$$D.\ Specificity = TN / (TN + FP)$$

Where,

True Positive (TP) = the number of correctly predicted MRI images that have normal lesions. **False Negative (FN)** = the number of incorrectly predicted MRI images that have abnormal lesions.
True Negative (TN) = the number of correctly predicted **MRI images that have abnormal lesions.**
False Positive (FP) = the number of incorrectly predicted MRI images that have normal lesions.

	Existing system(FCM)			Proposed system(KFCM)		
Methods	FO+ SVM	GLCM +SVM	GLRLM+ SVM	FO+ SVM	GLCM +SVM	GLRLM +SVM
Accuracy	78%	87.50%	91.20%	83.29%	96.29%	98.04%
Sensitivity	75%	90%	95%	80%	95%	100%
Specificity	80%	85%	90%	80%	95%	95%

E.Table1. Experimental Results

Volume: 11 Issue: 03 Pages: 4294-4297 (2019) ISSN: 0975-0290

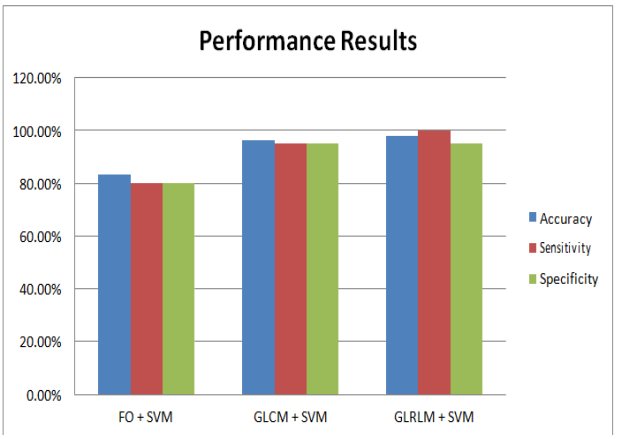


Figure.2 Comparative Results of Feature Extraction Techniques

The classification result of the existing and the proposed methods from 40 cases shown in Table 1. Using the same features extraction and classification algorithms, results of previous and the developed methods compared. According to the evaluated result of performance, the developed approach achieved high accuracy, specificity, and sensitivity. GLRLM + SVM achieved an excellent classification accuracy compared to other feature extraction techniques.

V. CONCLUSION AND FUTURE WORK

A hybrid approach of KFCM clustering segmentation is an effective algorithm to detect lesions at the early stages of disease & achieve high accuracy. In this, classification accuracy improved by comparing with the existing system.

In this developed study, the MRI images acquired, and multiple operations are carried out to classify the lesion as normal or abnormal. In the pre-processing step, the developed approach used anisotropic diffusion to filter the unwanted distortions, artifacts and acquire a good quality of MRI image. Then hybrid approach KFCM clustering used to segment the MRI images in segmentation. Later, feature extraction performed using first order, second-order statistics then SVM classifier used to classify the normal lesion and abnormal lesion. The developed approach achieved 98.04% accuracy using GLRLM feature extraction technique. In future, different classifiers can also be compared to find the best classifier to improve the performance of classification.

REFERENCES

[1]. 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, March 2012.

[2]. I., Nurtanio, E. R., Astuti, I. K. E., Purnama, M., Hariadi, & M. H., Purnomo, " Classifying cyst and tumor lesion using support vector machine based on dental panoramic images texture features," IAENG International Journal of Computer Science, 40(1), 29-37, 2013.

[3]. K. Anuradha, K. Sankaranarayanan," Statistical Feature Extraction to Classify Oral Cancers. Transform". Journal of Global Research in Computer Science, Volume 4, No. 2, February 2013.

[4].K. Anuradha, K. Sankaranarayanan "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, June 2013.

[5]. Anuradha K, Sankaranarayanan K. Oral cancer detection using an improved segmentation algorithm. Int J Adv Res Comput Sci Softw Eng 2015;5(1):451–6.

[6].M. K. Alsmadi, "A hybrid Fuzzy C-Means and Neutrosophic for jaw lesions segmentation," [Press release], 2016. Retrieved from https://doi.org/10.1016/j.asej.2016.03.016.

[7]. Zhalong Hu, Abeer Alsadoon, Paul Manoranjan , P.W.C. Prasad, Salih Ali, A. Elchouemic, "Early Stage Oral Cavity Cancer Detection: Anisotropic Pre-Processing and Fuzzy C-Means Segmentation, "IEEE 978-1-5386-4649-6/18 2018.

[8]. K. Anuradha,'' Efficient oral cancer classification using glcm feature extraction and fuzzy cognitive map from dental radiographs,'' International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018.

[9]. M. M., Galloway, "Texture analysis using gray level run lengths," Computer graphics and image processing, 4(2), 172-179, 1975.

[10].Shivshanker. P. Biradar, Dr. T.S. Vishwanath, Network Lifetime Maximization of Sensor Network Based on Energy Aware Source Tree Routing, Int. J. Advanced Networking and Applications, 10(2), 2018, 3788-3793.