Oral Cancer Detection: Hybrid Method of KFCM Clustering

Shilpa Harnale , Dhananjay Maktedar

Abstract: Oral neoplasm is one of the complex diseases in the world. The risk of death is increasing all over the world due to the rapid swelling of abnormal tissues. Early diagnosis of malignancy is necessary to avoid the risk of death. The tumor detected from magnetic resonance imaging (MRI) images is new innovative analysis topic in medical intervention. Normally the internal structure of the mouth can be examined using the MRI scan or CT scan. MRI scan is an advantageous and adequate technique for detection of the oral malignancy. It is non-invasive because it does not use any radiation. In this study, the hybrid approach KFCM is proposed for the segmentation and compared with conventional K-Means & Fuzzy C-Means(FCM).The main objective of merging these two algorithms is to reduce the total iterations generated by initializing an exact cluster to the FCM clustering with less computation time. The developed system concentrated on image enhancement using anisotropic diffusion to improve the quality of image and segmentation technique using KFCM clustering to reduce computation time &improve the segmentation accuracy. It exactly segments the lesion region and evaluates the lesion area.

Keywords: Oral and Maxillofacial Surgery, Pre-processing, KFCM clustering, Morphological operations.

I. INTRODUCTION

Oral neoplasm is the type of head and neck cancers. A lesion is unconstrained growth of tissues in the body. An oral neoplasm can develop in the tissues of the mouth or oral cavity, including the lips, Oropharynx, gums, Nasopharynx, tongue, cheeks, and roof of the mouth. The mortality rate of oral cavity cancer is higher and rising by chewing tobacco. Eighty percent of people die due to this type of Oral cancer. The standard oral precancerous lesions are leukoplakia, erythroplakia, lichen-planus, and so forth. Early detection & diagnosis of these lesions of oral cancer is crucial for successful treatment.MRI scan is more superior & non-invasive compared to CT scan. It provides the detail information of an image. MRI images are analyzed by the radiologists and dentists to detect &diagnose the oral lesions. This approach used anisotropic diffusion filters to remove unwanted distortions, other artifacts & to acquire the good quality of image. This method concentrated on the segmentation accuracy and elapsed time for segmentation.

Image segmentation is a technique to partition an image into multiple segments or non-overlapping regions. The segmentation algorithms based on one of the two basic characteristics of image intensity is similarity and discontinuity. The primary purpose of segmentation is to identify homogeneous regions and related objects. Most of the clustering methods used for diagnosing diseases. We have used clustering method as segmentation techniques for the detection of the lesion and determining the ROI for MRI images. This paper provides an efficient lesion detection method for medical images. In this, a hybrid clustering approach (KFCM) used for segmentation. Also, it reduces the execution time of clustering or segmentation. Finally, the lesion is segmented from the MRI image & calculate the lesion area with the exact location.

The paper represented as follows: chapter 2 provides the previous work, and Chapter 3 describes the methodology of the proposed approach. Then Chapter 4 demonstrates the experimental results and discussion. Chapter 5 gives performance evaluation and conclusion given in Chapter 6. prepared as per journal the template. 3. Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible.

II. PREVIOUS WORK

Different segmentation methods are outlined to detect the oral cancer of medical images in the literature.

A segmentation system of dental x-ray images for cyst or tumor cases presented [1]. In preprocessing, the colored pictures transformed into gray photos and apply Gaussian filter to these photos for smoothing. Then, the active contour models (snake) used in segmentation. This approach used to handle region of interest (ROI), not the entire image. Receiver operating characteristic (ROC) analysis used as a performance measure in this method.

Adv: snakes reduced the need for edge linking compared to conventional edge-based segmentation.

Oral cancer detection using a morphological mathematical watershed algorithm proposed [2]. For lesion segmentation purpose, the marker controlled watershed algorithm used. Then, this segmentation algorithm compared with the watershed segmentation. Linear contrast stretching of input(x-ray) image is carried out in Pre-processing, and then marker-controlled segmentation is used.

It gave better results compared with the watershed segmentation algorithm. Automatic symmetric axis analysis segmentation algorithm proposed[3].In this, the automatic segmentation of the tumor of CBCT images used. Diffusion filtering is used to acquire the good quality of image in preprocessing. An efficient method with new feature points used this automatic detection. Multi-resolution hierarchical FFD used in multi-scale registration with fastness and robustness. More True Positive and lessFalse Positive resultsare obtained using this approach. Performance metrics of this automatic segmentation are Dice's coefficients, Jaccard index, and Hausdorff distance.

Revised Manuscript Received on January 15, 2020
***** Correspondence Author
Shilpa Harnale , CSE department,BKIT BHALKI-India.
Email: shravi97@rediffmail.com
Dr Dhananjay Maktedar, CSE department,GNDEC Bidar- India
Email: dhan_mak@yahoo.com

Adv: Multi-resolution hierarchical FFD used with fastness and robustness in multi-scale registration.
Disadv : FFD registration computed Inverse FFD.
Limitations: This method is incapable of handling tumors and lesion without specific boundary and cases with significant asymmetric variations.

An efficient method based on artificial neural networks proposed [4]. In this proposed approach, 120 sample images used. For the training process, 120 images used and for the testing and validation process, 30 images used. Then, four statistical features obtained and used for detection of oral lesions. This proposed method is an effective technique for screening and detection of dental and oral diseases. Using this method, diagnosis and prediction of lichen-planus, leukoplakia, and oral squamous cell carcinoma diseases studied.

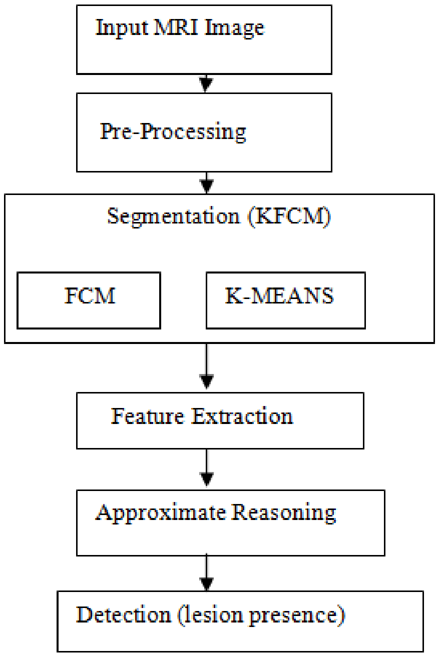
An oral cancer detection system proposed[5]. In preprocessing, Linear Contrast Stretching used to remove noise. In the existing approach, the Watershed algorithm used for segmentation. Marker Controlled Watershed segmentation is also used to decrease the over-segmentation problem of the watershed algorithm. In the developed approach, an Improved Marker Controlled Watershed Segmentation method used to achieve high segmentation accuracy. It used to overcome the problem of over-segmentation. The computation time is estimated before contrast-stretching and after contrast-stretching in this approach.

The hybrid approach of FCM and Neutrosophic segmentation proposed [6].In this, jaw image is segmented to detect the jaw lesion. Both efficiency and accuracy evaluated. To improve the robustness of the developed approach, specificity, sensitivity, and similarity analyses conducted. The developed work compared with the hybrid approach of Firefly Algorithm and the FCM, and another hybrid approach of the Artificial Bee Colony and FCM algorithm. This method produces the most identical lesion region to the manual delineation by the Oral Pathologist and evaluates good results.

Oral cancer detection using fuzzy C-Means Clustering & anisotropic filtering proposed[7]. This approach concentrated on anisotropic diffusion filter to acquire good quality of the CT image in image pre-processing &FCM clustering to achieve high accuracy of tumor detection in segmentation. This approach suppressed unwanted distortions and acquire the good quality of input image by sharpening all edges using diffusion filter. In the next step, the fuzzy c-means clustering focused on improving classification accuracy, specificity and sensitivity.

III. METHODOLOGY

The developed approach consists of different levels: 1) Image Pre-Processing 2) Image Segmentation 3) Feature Extraction4) Approximate Reasoning.



In pre-processing, anisotropic diffusion filter used. Segmentation is performed by a hybrid approach KFCM, in which both algorithms work separately. Morphological operations applied on FCM cluster predicting tumor in Feature extraction and finally, calculate the lesion area and exact position of the lesion in approximate reasoning.

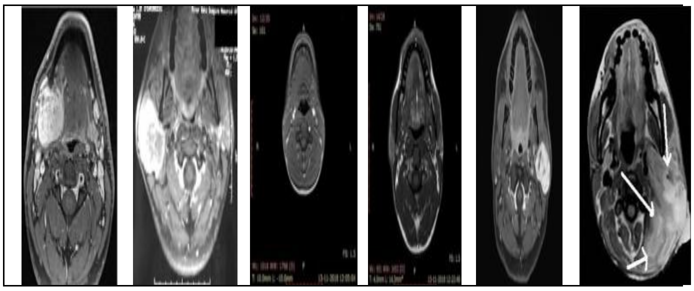


Fig 1.Sample (Normal and Abnormal) images

3.1 Image Preprocessing:

The input MRI image is always prone to various types of unwanted distortions. These distortions can reduce the segmentation accuracy. Image enhancement is an essential process in image processing to enhance the good quality of image. The anisotropic filter is filters which is extensively used for suppressing unwanted distortions, other artifacts in the grayscale images and improve the quality of images by sharpening the edges.

The equation for general anisotropic filter is:

$$\frac{\partial I}{\partial t} = \mathrm{div}(c(x,y,t)\nabla I) = \nabla c.\nabla I + c(x,y,t)\Delta I$$

x, y, t: Represents the control rate of the diffusion coefficient. It represents an image gradient which used for sharpening the boundaries in the image.

∇: Represents the gradient

Div (): Represents divergence operator.

Δ: Represents the Laplacian

I: denotes grayscale image matrix.

3.2. Image Segmentation

Image segmentation is a familiar process of separating foreground from background in image analysis. Segmentation based on the characteristics of the pixels in the image. The segmentation algorithm is having one of the two characteristics of intensity values, namely discontinuity and similarity. The first property is to divide an image based on abrupt changes in the intensity level of images, such as edges in an image, Second property is to divide an image into multiple segments that are similar according to predefined criteria.

3.2.1. K-Means clustering

K-Means is most straightforward, iterative, unsupervised learning algorithms. It is used to classify given dataset to many clusters. Number of clusters (groups) K is defined. Later, randomly select K centers for each cluster. The distance between data point to each cluster centers determined using Euclidean distance algorithm. Each data point is compared to all cluster centers. The data point moved to a particular cluster which has the shortest distance among all. Then, the centroid is re-calculated. Again each data point is compared to all centroids. The process continues until the center converges. The clusters formed by minimizing the Euclidean distance between the data and the corresponding cluster centroid. Thus, the k-means clustering is used to cluster the data.

Algorithm:

a) Initially, Number of clusters (groups) K is defined.
b) Select the k cluster centers randomly
c) Estimate the mean or center of the cluster
d) Determine the distance between each data point to every cluster center
e) When the distance is near to the center of one cluster, then move to that concerned cluster.
f) Otherwise, go to the next cluster(group).
g) Re-calculate the center.
h) Continue the process until the center doesn't proceed.

3.2.2. Fuzzy C-Means clustering

FCM is an unsupervised clustering algorithm. In this method, one piece of data belongs to more than one cluster. The partial membership value is given to each pixel of the image in the fuzzy logic. The membership value of the fuzzy set ranges from 0 to 1. Data are bound to each cluster using a membership function, which represents the fuzzy behavior of this algorithm. Fuzzy partition is carried out by an iterative optimization of an objective function, with the update of the membership function and cluster center. Fuzzy clustering is a multi-valued logic derived from fuzzy set theory. The clusters formed according to the distance between data points and cluster centers for each cluster. Fuzzy C-Means (FCM) is a form of clustering method in which each data point can be a member of more than one cluster with different degrees of membership.

FCM used to minimize the following objective function.

$$J_m = \sum_{i=1}^{D}\sum_{j=1}^{N} u_{ij}^{m} ||x_i - c_j||^2$$

Where,

D is the number of data points.

N is the number of clusters.

M is fuzzy partition matrix exponent for controlling the degree of fuzzy overlap, with $m > 1$.

Fuzzy overlap refers to how indistinct the boundaries between clusters.

Xi is the ith data point.

Cj is the center of the cluster j.

μij is the degree of association of xi in the jth cluster(group).

For a given data point, xi, the sum of the membership values for all clusters is one.

Algorithm:

Algorithmic steps of FCM clustering is given below:

1. Randomly initialize the cluster membership values, $\mu_{i.}$
2. Determine the cluster centers:

$$c_j = \frac{\sum_{i=1}^{D} u_{ij}^{m} x_i}{\sum_{i=1}^{D} u_{ij}^{m}}$$

3. Update μ_{ij} according to the following:

$$u_{ij} = \frac{1}{\sum_{k=1}^{N}\left(\frac{||x_i - c_j||}{||x_i - c_k||}\right)^{\frac{2}{m-1}}}$$

4. Calculate the objective function, J_m.
5. Repeat steps 2–4 until J_m improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

3.2. Feature Extraction

The feature extraction is segmenting the cluster, which shows the predicted tumor at the output cluster of FCM (Fuzzy C-Means). Apply morphological operations to segment tumor from this output cluster.

3.3. Approximate Reasoning

The lesion area is estimated using the binarization method in the approximate reasoning step. Each transform coefficient compared with a threshold. It is zero when it is smaller than the threshold value. Otherwise, it is one. An image has only two possible values for each pixel, either 0 or 1(black or white). Here maximum image size is 256x256 jpeg. Two colors are used to represent the binary image.

Image,

$$I = \sum_{W=0}^{255}\sum_{H=0}^{255}[F(0) + F(1)]$$

Pixels= Width (W) X Height (H) =256 X256

F (0) = white pixel (digit 0)

F (l) = black pixel (digit 1)

No_ of_ white pixel

$$P = \sum_{W=0}^{255}\sum_{H=0}^{255}[F(0)]$$ Where,

P = number of white pixels (width*height)

1 Pixel = 0.264 mm

The area calculation formula is

Size_ of_ tumor, S = [(√□)* 0.264] mm2

Where,

P= no-of white pixels, W=width, H=height

Algorithm:

A step of oral lesion detection is given below.

1) Initialize the process.
2) Get the MRI input image in JPEG format.
3) Verify whether the input image is in the required format then go to step 4.
4) If MRI image is in rgb format, convert that image into grayscale else go to next level.
5) Detect the edge of the grayscale image.
6) Determine the number of white pixels in the image.
7) Estimate the lesion area using the formula.
8) Display the lesion area.
9) Compare the lesion area acquired with 5mm2. If the lesion area is more significant than 5mm2, then MRI image has a severe abnormal lesion.
10) End the process.

Initially, this algorithm uses the RGB or grayscale image. Next, binarization technique used to convert grayscale image into binary image. After binarization, the edge of tumor pixels in the binary image detected. Later, this algorithm estimates the area of the tumor by determining the number of white pixels (digit 0) in binary representation.

In last step, the predicted tumor area estimated. Here consider that it detects as a tumor when the size of the tumor area is more extensive than 5.00mm^2.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this approach, detection of the lesion in MRI images is done using hybrid segmentation of KFCM clustering techniques. Main objective of merging two clustering methods to minimize the total iterations generated by initializing an exact output cluster to the FCM clustering.

Input MRI image preprocessed to suppress the unwanted distortions and acquire good quality of MRI image. The noise-free MRI is the input of the image segmentation process. The hybrid approach of KFCM used to get the segmented image. K-Means clustering divides the preprocessed image into the clusters having the same or nearby intensity value [8]. Then total iterations of FCM are reduced to get an exact output cluster in FCM clustering. Tumor predicting cluster (group) is the output of FCM. Region of interest is detected from this cluster and extracted using morphological operations to separate the lesion.

The developed method tested on the dataset of 46 MRI images. Out of which the 10 MRI images arbitrarily selected & the output of these images discussed. The table entities are as follows: the image name, Area of the Lesion, elapsed time, and the detection decision (lesion present and lesion absent). Elapsed time: It is the amount of time taken for the segmentation of an input image. This method calculated by Tic and Toc methods in Matlab.

Table1: Table of Lesion Area

Image name	Area of the Lesion(mm^2)		
	KFCM	FCM	K-Means
ab7	13.58	13.78	14.85
ab11	14.5	14.4	15.7
ab15	12.4	12.9	13.74
ab23	8.51	8.43	9.50
ab25	7.67	7.49	8.65
ab37	0.64	0.69	3.08
ab41	1.69	1.69	5.65
ab43	1.37	1.49	4.49
ab44	0.95	0.95	5.09
ab45	1.73	1.77	5.40

Table2: Table of Elapsed Time

Image Name	KFCM		Lesion Detection
	Area of the Lesion(mm^2)	Elapsed Tim	
ab7	13.58	2.64	Present
ab11	14.5	2.56	Present
ab15	12.4	2.76	Present
ab23	8.51	2.54	Present
ab25	7.67	2.55	Present
ab37	0.64	2.67	Absent
ab41	1.69	2.62	Absent
ab43	1.37	3.01`	Absent
ab44	0.95	2.62	Absent
ab45	1.73	2.70	Absent

Table3: Table of Lesion Area, Elapsed Time &Lesion Detection of KFCM

Image Name	KFCM		Lesion Detection
	Area of the Lesion(mm^2)	Elapsed Tim	
ab7	13.58	2.64	Present
ab11	14.5	2.56	Present
ab15	12.4	2.76	Present
ab23	8.51	2.54	Present
ab25	7.67	2.55	Present
ab37	0.64	2.67	Absent
ab41	1.69	2.62	Absent
ab43	1.37	3.01`	Absent
ab44	0.95	2.62	Absent
ab45	1.73	2.70	Absent

Table1& Table 2 show the area of the lesion and elapsed time comparison respectively of three algorithms. Table 3gives an idea of exact detection of the normal and abnormal image of the proposed method by displaying a message of Lesion present or absent.

Retrieval Number: E5797018520/2020©BEIESP
DOI:10.35940/ijrte.E5797.018520

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

V. PERFORMANCE EVALUATION

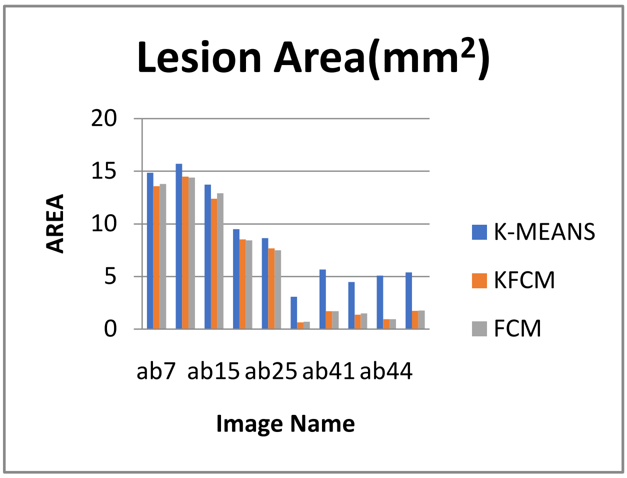


Fig.1: **A graph of Lesion Area Comparison**

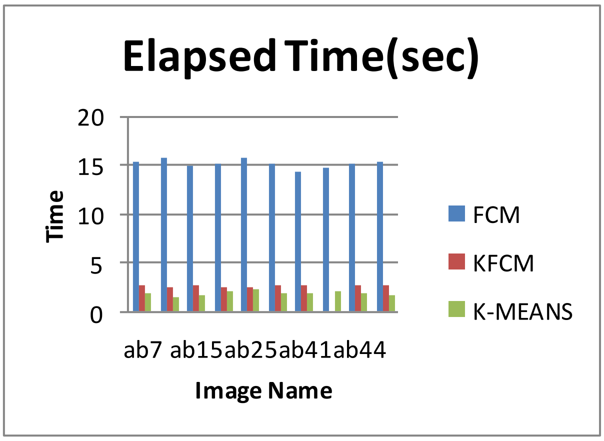


Fig. 2: **A graph of Elapsed Time Comparison**

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

The primary motivation of merging FCM & K-Means algorithms is 1) FCM clustering is used to reduce the total iterations and accurately segment lesion area. 2) K-Means clustering is used to reduce execution time.

This chapter clearly presents that the lesion area of KFCM method is more precise and near to FCM lesion area compared to K-means lesion area as shown in fig.1. When compared with FCM, KFCM produced most identical lesion region to the manual delineation by the oral dentist & radiologist. However, the execution time required by KFCM is lesser than FCM and near to K-Means as shown in fig.2. The developed system segment exact lesion area from MRI image using KFCM segmentation,and morphological operations. To test the system, we collected the dataset of 46 MRI images containing lesions and others with no lesions. Out of 46 sample images, 36 oral lesion images and 10 with no lesion images used to evaluate the performance of the developed approach using three measures. These are the segmentation accuracy, sensitivity, and specificity.

Segmentation Accuracy: It is the total number of correctly segmented MRI images with lesions and without lesions over the given number of MRI images.

Sensitivity: It describes True Positive rate, number of correctly segmented MRI images that have oral lesions.

Specificity: It describes True Negative rate, number of correctly segmented MRI images that have no oral lesions.

Table 1. Segmentation Accuracy of the Proposed System

	Number Of MRI Images	Correctly Segmented MRI Images		Segmentation Accuracy	
		FCM	KFCM	FCM	KFCN
No Lesion	10	10	10	100%	100%
Lesion	36	33	34	91.6%	94.44
Total	46	43	44	93.47%	95.65

Table 2. Experimental Results

Method	Number Of MRI Images	Accuracy	Sensitivit	Specificity
FCM	46	94.44%	100%	76.92%
KFCM	46	95.65%	100%	83.33%

Table 1 & 2 show the experimental results of the developed system, where the segmentation accuracy obtained is 95.65% after testing the proposed system by lesion and non-lesion MRI images. The sensitivity and specificity of the developed approach is 100 % and 83.33% respectively. The proposed approach demonstrated an excellent efficiency in segmenting lesions of an MRI image using KFCM segmentation technique.

VI. CONCLUSION AND FUTURE WORK

There are varieties of lesions. Lesion may be a soft tissue lesion or hard tissue lesion in the mouth. If it is a soft tissue, then it is extracted by K- Means algorithm easily from the oral cells. If noise or other artifacts are present in the MRI images, it removed in pre-processing. The noise-free MRI image given as an input to the K-Means. In this research, a K-Means algorithm performs the initial clustering of the images. After an initial segmentation, an appropriate cluster selected depending on the criteria of an updated membership set. Next, a cluster predicting tumor is obtained using FCM. Then, the tumor is segmented from the image using morphological operations and finally, tumor area calculated in approximate reasoning. The experimental results clarify that the suggested research gives better segmentation result with minimum execution time& accurate lesion area. Furthermore, results proved that even small size lesions also detected.

Segmentation accuracy of KFCM is compared with FCM Clustering in the experiment, as shown in table 1 & 2. These results show higher efficiency in segmenting lesions of MRI images using KFCM segmentation technique by achieving high accuracy of 95.65% .In future, advanced algorithm with more features can implement to achieve more segmentation accuracy and less execution time.

REFERENCES

1. Nurtanio I, Purnama IKE, Hariadi M, Purnomo MH. Cyst and tumor lesion segmentation on dental panoramic images using active contour models. IPTEK J Technol Sci 2011;22(3).
2. Anuradha K, Sankaranarayanan K. Detection of oral tumor based on the marker-controlled watershed algorithm. Int J Comput Appl 2012;52(2).
3. F., Abdolali, R. A., Zoroofi, Y., Otake, & Y., Sato, " Automated classification of maxillofacial cysts in cone-beam CT images," Computers in biology &medicine, 72 108-119, 2016.
4. Maghsoudi R, Bagheri A, Maghsoudi MT. Diagnosis prediction of lichen planus, leukoplakia, and oral squamous cell carcinoma by using an intelligent system based on artificial neural networks. Dentomaxillofacial Radiol, Pathol Surg 2013;2(2):1–8.
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 1, 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. Nikhita Biradar, Prakash H. Unki,’’Brain TumorDetection Using Clustering Algorithms in MRI Images,’’International Research Journal of Engineering and Technology 1587-91, 2017;
9. Eman Abdel Maksoud, Mohammed Elmoji, Rashid Al-Awadi, "Brain Tumor Segmentation Based On Hybrid Clustering Technique, " Elsevier B. V. on behalf of Faculty of Computers & Information, Cairo University.
10. Eman A. Abdel Maksoud, Mohammed Elmogy, and Rashid Mokhtar Al-Awadi,” MRI Brain Tumor Segmentation System Based on Hybrid Clustering Techniques,” Springer International Publishing Switzerland 2014, CCIS 488, pp. 401–412, 2014.
11. S., Galib, F., Islam, M., Abir& H.-K. , Lee, “Computer-aided detection of oral lesions on CT images," Journal of Instrumentation, 10(12), 2015.
12. S., Banerjee, S., Chatterjee, A., Anura, J., Chakrabarty, M., Pal, B., Ghosh,J. ,Chatterjee, “ Global spectral and local molecular connects for optical coherence tomography features to classify oral lesions towards unravelling quantitative imaging biomarkers,” RSC Advances, 6(9), 7511-7520, 2016.
13. S.-W., Chang, S., Abdul-Kareem, A. F., Merican, & R. B. , Zain, “ Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods,” BMC Bioinformatics, 14(1), 2013. T., Karthikeyan, “ Unified RF-SVM model based digital radiography classification for Inferior Alveolar Nerve Injury (IANI) identification,” Biomedical Research, 27(4), 2016.
14. M. M. R., Krishnan, P., Shah, C., Chakraborty, & A. K., Ray, “ Statistical analysis of textural features for improved classification of oral histopathological images,” Journal of medical systems, 36(2), 2012.

AUTHORS PROFILE

Shilpa Harnale currently working as a Assistant Professor of CSE Dept At BKIT Bhalki Dist- Bidar.

Dr. Dhananjay Maktedar currently working as a Professor and HOD of CSE Dept At GNDEC BIDAR.