IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

Machine Learning Approach for Cancer Detection

Anooja Ali[1*], Pooja G[2], Prajeela MP[3], Riddhi Rakesh[4], Tabassum Taj[5]

[1,2,3,4,5] School of Computing & Information Technology, REVA University, Bangalore, India

[*]Corresponding Author: anoojaali@reva.edu.in, Tel.: 8050779104

DOI: https://doi.org/10.26438/ijcse/v7si14.224228 | Available online at: www.ijcseonline.org

Abstract**—** Machine Learning has several applications in Healthcare Domain. It provides more efficient, faster, smarter ways to detect and cure various diseases. Machine learning approaches are widely used for cancer diagnosis. In our approach we classify cancerous and noncancerous Oral Cancer images. We focused on image pre-processing, segmentation using image segmentation app in Matlab to improve the image quality and thereby improving the accuracy of classification and cancer detection. This approach using Support Vector Machines (SVM) obtained an accuracy of 89.2%. This method can be easily adopted for early cancer detection**.**

Keywords**—**Machine Learning,Oral Cancer, Image Segmentation, Support Vector Machines.

I. INTRODUCTION

Healthcare domain is one the most important and largest in the world as it plays a vital role in people's life and directly affects quality of life. The number of health issues are increasing exponentially due to various factors and even though there has been a greater improvement for treatment and diagnosis in the Health sector the need for much better methods and tools are always a necessity [1]. The improving technology and expectations always demands a better and more efficient system. All these factors make this a significant asset if tended to accurately, through approach and practice, with a good data set and algorithm there is a high chance to receive a scope of benefits.

Healthcare investing has to take into consideration the benefits of technology in terms of buying advanced equipment and training professionals on how to use it [2]. This technology mainly includes equipment's for non-invasive surgery, which results in more effective outcome and less suffering for the patient, and software to predict and therefore foresee any outbreaks. Medical advances have effectively eradicated many diseases and disorders that are now treated as minor, like smallpox [3]. People are able to do research and come up with treatments without testing it on animals or humans, all that credit to technology. One of the finest illustrations is the HIV treatment, which is now at a period where the virus does not have to advance into AIDS.

We focused on improving the accuracy of image segmentation and classification. In pre-processing we remove the noise from the image while the image quality is enhanced with Fuzzy C-means. In fuzzy C-means each of the data points are assigned with membership values based on their distance from the centre points. SVM outperform other classifiers in scaling any high dimensional data to binary levels.

The remaining part of the paper is divided as follows. Section 2 indicates literature survey followed by the motivation for this work. Section 4 is methodology and then system design. Section 6 is results and discussion. The paper is concluded with conclusion remarks and scope for further enhancement.

II.RELATED WORK

A challenging task for radiologist is the early detection of cancer. Early diagnostic of cancer helps in quick treatment plan and procedures which increases life expectancy. For this reason we incorporate machine learning techniques for early detection. Most of the existing works are based on automated segmentation of the lesion, clustering by Fuzzy C means, Neuro Fuzzy Inference Systems, Classification by SVM.

The automated image segmentation is based on the axis of symmetry which considers the two sides of the image [4]. The purpose is to detect the dissimilarity in concentration where the tumouroccurs. It attained highly exceptional results including different cyst types. However, the model has limitationwhen the tumour is in the centre of the symmetric axis or on both the side of the side the results are not positive. Anisotropic diffusion filtering is used for pre-processing decreased the unwanted pixels on the CT image meanwhile augmenting the edges of the oral cavity. In addition, the previous study was combined and enhanced withoral cysts classification. The result for96 samples had an accuracy rate of 96.48% which had three different class of oral cysts. However, this study based on the demanding issue of segmentation was not accredited with a large data set due to which the rate of accuracy may be entirely different.

A compound method with the combination of Fuzzy C-Means and neutrosophic algorithm for segmenting tumours in the oral panoramic image was prospected [5]. It reduced the speckle noise by 3 x 3 median filter using speckle contraction, and with the help of neutrosophyreduce the noise in the image. This method paved way for a major improvement in segmentationof oral cavity. The accuracy is the result ofusing the uncertainty approach to cluster the region and determining the cyst. However, if image boundary calculation and calculating the clusters the region with shadows will be falsely detected.

The different types of Support Vector Machine (SVM), like Quadratic SVM, Linear SVM and Cubic SVM were used. The cysts were classified based on optical coherence tomography [6-7]. Under six different classification settings the specificity, sensitivity and accuracy were compared. It was based on biopsy images, and the accuracy is directly proportional to the image quality. Thus, it is obvious that lower image quality would be an issue.

Researchers have adopted a compound feature extraction technique and classification algorithm with biomarkers. They came up with five cysts classification methods and tested all of the methods [8]. Adaptive Neuro Fuzzy Inference System (ANFIS), achieved the maximal classification rate of 93.41%. This is obtained by combining biopsy images with the clinical pathology data set. The complete patient's case was included for study. However, if either clinical pathology data set or images alone were considered in segregation, the accuracy of over 90% accuracy rate may not be obtained. In this work it is clear that accuracy depends on the patient's record or information.

The automated oral cancer detection used two models to detect two common types of lesions in the oral cavity [9]. It was estimated to have 92% sensitivity with an average of 0.32 false positives in close border lesions and 85% sensitivity with no false positives in open border lesions. Furthermore, the model also considered improving accuracy to 100% sensitivity with an average of 0.13 false positives in open border lesion algorithm.

The SVM model named RF-SVM [10] obtained the highest (83.58%) accuracy in detecting Inferior Nerve Injury when compared to the other SVM algorithms and by experts. This study usedfilters such as anisotropic diffusion to remove noise. The algorithm provides image with high quality after filtration by removing noise and enhancing the image boundaries. For oral tissue cell classification SVM algorithm was proposed [11]. In this, linear SVM classifier was used to detect biopsy images for abnormal and normal oral cells. It classifies the linearly non-separable and separable data following the classification of cells on 512 × 512 dimension of images.

Every stride to pre-process the medical image and classify the cysts using hybrid algorithms was applied, whilst accomplishing high accuracy (87.18% accuracy rate) by using a number of algorithms, Grey Level Co-occurrence Matrix (GLCM) and Grey Level Run Length Matrix (GLRLM) [12]. It resulted in a high rate of 94.44%.

In our work we followed the series of steps Pre-processing, Segmentation, Feature Extraction and Classification. We combined inbuilt image segmentation in matlabwith SVM and obtained an accuracy of 89.2%.

III.MOTIVATION

Oral cancer ranks in the top three of all cancers. In India, over thirty percent have oral cancers and oral cancer regulation is hastily becoming a global health priority. Rate of oral cancer in India is soaring, that is, 20 per 100,000 population. Figure 1 shows the incidence and mortality rate of oral cancer in India. It is estimated that the percentage would double in the next decade, and a major problem is that ignorance of people leads to severe stage of cancer. Thus it becomes life threatening and also the cost and time taken for the biopsy result is a major concern [13]. Hence to avoid this, by using machine learning approaches, we can detect cancer at a very early stage using a CT scan image and this would be a game changing improvement in hospitals.

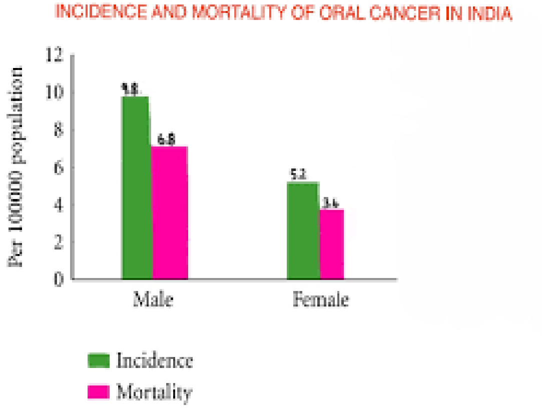


Figure 1. Incidence and Mortality rate of Oral Cancer in India. This figure is adopted from [13].

IV.METHODOLOGY

CT scan images of patients with oral cancer are considered. Figure 2 shows a few sample images. In our work image processing along with Machine Learning to train the system with the data set and classify an image as cancer or not and to give the intensity. The data set contains images of two different categories one being cancerous and the other being non-cancerous. The non-cancerous images have other abnormalities such as birth marks, ulcers and heat sores and perfectly normal oral cavity.

Data sets are obtained from National Cancer website has oral images with cancer, birth mark and ulcers. The reason for choosing this data set is because the model should be able to classify even an abnormal image such as ulcer, heat sores, and birthmarks as non-cancerous. This will help in increasing the performance of the system. The obtained images are resized and segmented using the image segmented app in Matlab. The app displays the thresholder tab and among several thresholding methods a suitable one is chosen. Other methods can also be used to increase the accuracy such as manually drawing regions using active contours.

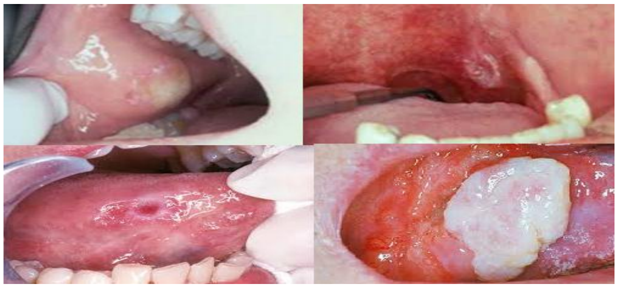


Figure 2. Images of patients with oral cancer.

Setting the iteration numbers minimally can yield good results. The image segmentation app can also be used. The next step is feature extraction. This used to train the system using SVM classifier using the training label of the particular image. The same procedure is carried out with all the images and the training data is extracted and loaded. After training the model input image is given to test the accuracy rate. The output images are segmented to show showing whether it is cancerous or non-cancerous. Clustering is done by using Fuzzy C means or K means clustering.

V. SYSTEM DESIGN

The images are pre-processed and segmented. In pre-processing binary mask are Red-Blue-Green (RGB) mask are created. In feature extraction, features like contrast, homogeneity, energy, mean, standard deviation, entropy, skewness, variance, smoothness are extracted. Figure 3 shows the proposed system. A different approach for segmentation is done using fuzzy C-means segmentation. We compare the accuracy rate of K means and fuzzy C-means. Fuzzy C-means is the most accurate one. Classification is done using SVM classifier. Classifier detects whether the samples are cancerous or not.

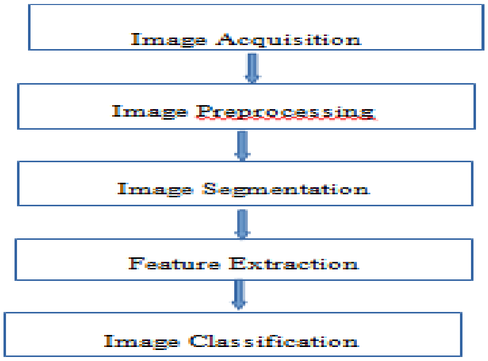
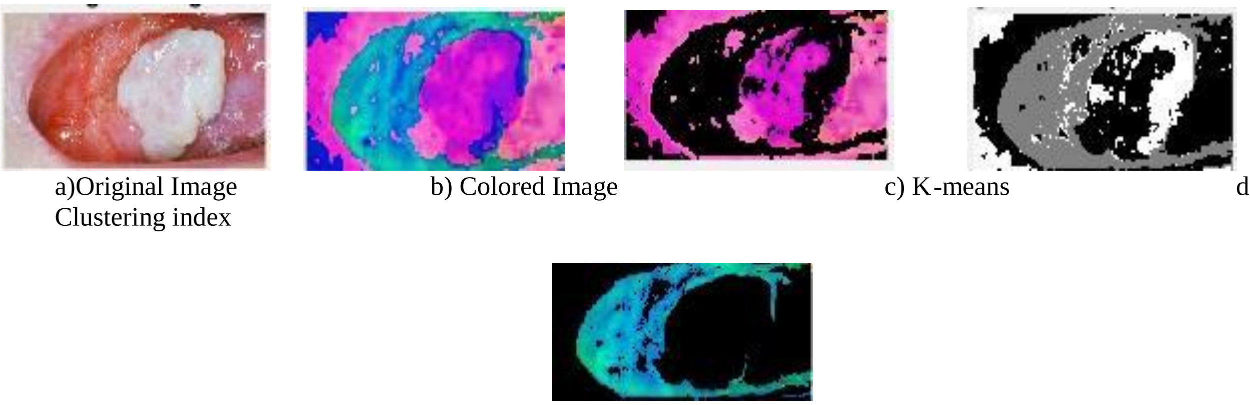


Figure 3. Block diagram of the proposed system

VI. RESULTS AND DISCUSSION

The CT images will have noise. So the pre-processing CT images is very important. The image boundaries of the image are enhanced. Fuzzy C means effectively cluster all the data points. Assigning membership points and then perform grouping gives a better result than any other mean or median clustering algorithm.

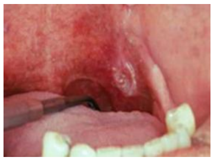
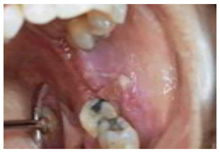
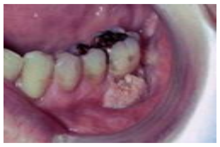
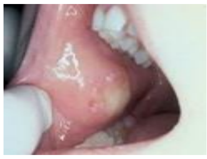
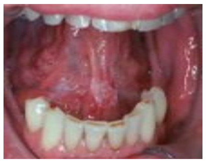


a)Original Image b) Colored Image c) K-means d)
Clustering index

e) Fuzzy C-means

Figure 4: Original image when converted into grey scale for K-means and Fuzzy C-means clustering

Table 1: Run time and accuracy using SVM.

Original image	Type	Pre-processing time (in sec)	Segmentation TIME (in sec)	Tumor detected	Accuracy
	Squamous cell carcinoma	1.07	1.521	Yes	True
	Squamous cell carcinoma	.823	1.456	Yes	True
	Squamous cell carcinoma	.876	1.508	Yes	True
	Mouth Ulcer	.963	1.452	No	True
	Squamous cell carcinoma	1.02	1.501	Yes	False

Fuzzy C-means performs well than k-means clustering. Figure 4 indicates this. K-means wrongly detect even the small ulcer present in the image as cancer. A few of the sample images used for our work are represented in table. Squamous cell carcinoma are cancerous. Accuracy true indicates the diagnosis is true. Patient has cancer and it is detected correctly. Accuracy false indicates the cancerous image is detected as noncancerous or a noncancerous image is detected as cancerous.

VII. CONCLUSION

This work automates the entire process of cancer detection and prognosis. It resolves many problems that we face in the current cancer diagnostic scenario. The proposed segmented classifier have an upper side compared to the traditional cancer detecting techniques such as biopsy, which require more than 24 hours to obtain the result.. The oral cavity image of the patient when loaded it into the system we can classify cancerous and noncancerous. Hence with this system the waiting time and the expenses would be less.

The work can be enhanced by using any other ensemble classifier so that a higher accuracy may be obtained. Considering the dataset Support Vector Machine was the best algorithm for this model. Convolution Neural Network could be better option if there are lakhs of images. The work can be expanded to detect several other variants of cancer like as breast cancer, brain cancer, lung cancer etc. This work can even be enhanced to detect the intensity of any cancer at each location so that a better treatment can be given to the patient. The present system work with an accuracy of 89.2%. The fuzzy C-means segmentation along with the binary classifier, SVM can be strongly recommended for cancer detection.

REFERENCES

[1] Van Dyck, &, W Stremersch S, “Marketing of the LifeSciences: A New Framework and Research Agenda for a Nascent Field”. Journal of Marketing, 73(4), 430, 2009

[2] Rowlands, S., Coverdale, S., & Callen, J, “Documentation of clinical care in hospital patients’ medical records: A qualitative study of medical students’ perspectives on clinical documentation education”. Health Information Management Journal, 45(3), 99–106, 2016

[3] Lenzer, Jeanne, “Claim that smallpox vaccine protects against HIV is premature, say critics.” BMJ (Clinical research ed.) vol. 327,7417: 699. doi:10.1136 /bmj.327.7417.699, 2003

[4] F., Abdolali, R. A., Zoroofi, Y., Otake, &Y., Sato, “Automated classification of maxillofacial cysts in cone beam CT images using contour less transformation and Spherical Harmonics,” Computer Methods and Programs in Biomedicine, 139, 197-207, 2017.

[5] Early stage oral cavity cancer detection: Anisotropic pre-processing and fuzzy C-means segmentation 2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, Volumes 2018-January, 2018

.[6] T, Karthikeyan, “Unified RF-SVM model based digital radiography classification for Inferior Alveolar Nerve Injury (IANA) identification,” Biomedical Research, 27(4), 2016.

[7] Amin, Javeria, et al. "A distinctive approach in brain tumor detection and classification using MRI." Pattern Recognition Letters ,2017.

[8] 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.

[9] S., Galib, F., Islam, M., Abir& H.-K. , Lee, “Computer aided detection of oral lesions on CT images,” Journal of Instrumentation, 10(12), 2015.

[10] M. M. R., Krishnan, P., Shah, C. Chakraborty, & A. K., Ray, “Statistical analysis of textual features for improved classification of oral histopathological images,” Journal of medical systems, 36(2) , 2012.

[11] 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.

[12] K., Nguyen, A. K., Jain, & R. L., Allen,” Automated Gland Segmentation and Classification for Gleason Grading of Prostate Tissue Images,” 20th International Conference on Pattern Recognition, 23-26, 2010.

[13] N. P., Malek, S., Schmidt, P., Huber, M. P., Manns,T.F.Greten, “The diagnosis and treatment of hepatocellular carcinoma,” Alcohol, 20, 2014.