Identification of Mouth Cancer laceration Using Machine Learning Approach

Prof. Barry Wiling,

Professor, Department of Computer Science, U.S.M.N Oman
barrywiling@usmn.edu

Article History	Abstract
	This Paper describes about Identification of Mouth Cancer laceration Using Machine Learning Approach .The SVM algorithm is used for this purpose. Image segmentation operations are performed using: Resizing an image, Gray scale conversion, Histogram equalization and Classifying the Segmented image using SVM. SVM is used to reduce the complexity faced in the existing system comprising of Texture Segmentation and ANN (Artificial Neural Networks) Algorithm. SVM is a simple Machine Learning algorithm when compared to ANN. The outcome of the paper is to segment and classify the Malignancy from the Non-Malignant region using the classifier SVM. SVM performs the classification based on the dataset that contains the trained images.

I. Introduction

People are dominatingly visual oriented living beings that depend vigorously on our vision to comprehend our general surroundings. Individuals not just glance at things to recognize and order them, yet in addition check for contrasts, and get a general unpleasant inclination for a scene with a snappy look. People have advanced exact visual aptitudes. Natural eye can distinguish a face in a moment, can separate hues, and can process a lot of visual data rapidly. People like their images to be sharp, clear and point by point [1]; machines lean toward their images to be basic and un-clustered [2].

An image is an exhibit, or a grid, of square pixels (picture components) orchestrated in sections and lines. In a (8-piece) gray scale image each image component has a doled out power that ranges from 0 to 255. A Digital image is a two dimensional capacity, where the capacity esteem gives the brilliance of the image at some random point. The estimations of Digital images are discrete in nature. Normally they take on just whole number qualities. The splendour esteems additionally going from 0 (dark) to 255 (white). A computerized image can be considered as an enormous exhibit of discrete dabs, every one of which has a splendour related with it. These spots are called picture components, or all the more just pixels [3].

The pixels encompassing a given pixel establish its neighbourhood. An area can be portrayed by its shape similarly as a network. With the exception of in extremely extraordinary conditions, neighbourhoods have odd quantities of lines and segments; this guarantees the present pixel is in the focal point of the area. To speak to shading images, separate red, green and blue parts must be indicated for every pixel (expecting a RGB shading space)[3] , thus the pixel 'esteem' is really a vector of three numbers. Regularly the three unique parts are put away as three separate 'gray scale' images known as shading planes (one for every one of red, green and blue), [3] which must be recombined when showing or handling [4].

The real gray scale or shading part powers for every pixel may not really be put away unequivocally. Frequently, all that is put away for every pixel is a record into a shading map in which the genuine force or hues can be turned upward. An image surface is a lot of measurements determined in image preparing intended to

evaluate the apparent surface of an image. Image surface gives us data about the spatial course of action of shading or powers in an image or chose locale of an image [5].

II. Existing techniques

Image division is the way toward apportioning an advanced image into numerous portions (sets of pixels, otherwise called super pixels). The objective of division is to improve as well as change the portrayal of an image into something that is increasingly important and simpler to dissect [6][7]. Image division is commonly used to find articles and limits (lines, bends, and so forth.) in images. All the more exactly, image division is the way toward allocating a mark to each pixel in an image to such an extent that pixels with a similar name share certain qualities [8].

Surface division, the utilization of image surface can be utilized as a portrayal for areas into fragments. There are two principle kinds of division dependent on image surface [4], district based and limit based. In spite of the fact that image surface is certainly not an ideal measure for division it is utilized alongside different measures, for example, shading, that explains sectioning in image [10]. Locale Based, Attempts to gathering or bunch pixels dependent on surface properties together. Limit Based endeavours to gathering or bunch pixels dependent on edges between pixels that originate from various surface properties. The existing machine learning process algorithm [9] is shown in figure 1.
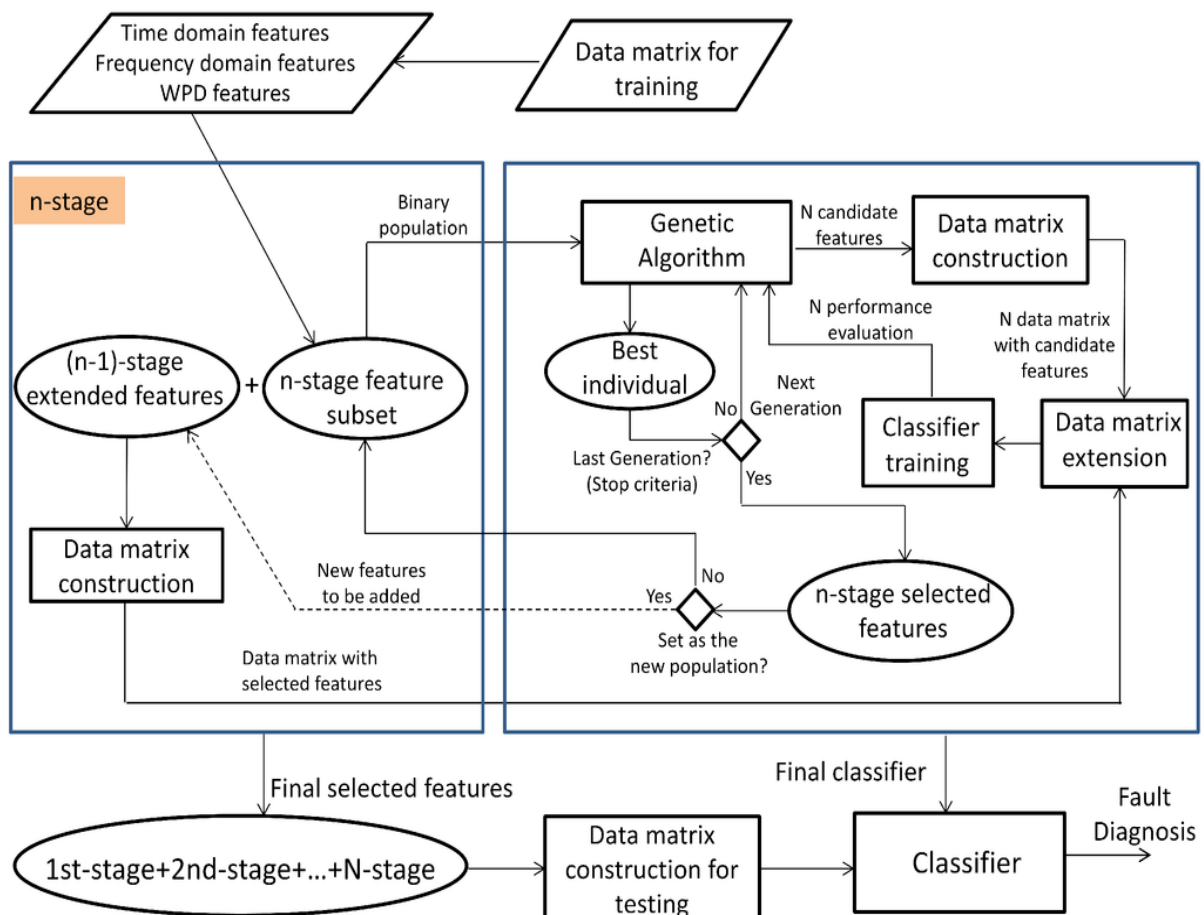

Fig 1: Block Diagram of existing machine learning algorithm

Fundamental strides for division is as per the following Resizing the image, the given info image can be of any size however then it must be changed over to a standard type of size 256*256.once the image has been changed over to a customary size the subsequent stage is to do Histogram Equalization after dim transformation histogram evening out, this technique ordinarily builds the worldwide complexity of numerous images, particularly when the usable information of the image is spoken to by close differentiation esteems. Through this modification, the forces can be better appropriated on the histogram. This takes into account zones of lower nearby difference to increase a higher complexity.

III. Proposed Methodology

Histogram levelling achieves this by adequately spreading out the most incessant power esteems. There are two different ways to consider and execute histogram adjustment, either as image change or as palette change. The activity can be communicated as P (M (I)) where I is the first image, M is histogram balance mapping activity and P is a palette. On the off chance that we characterize another palette as P'=P(M) and leave image I unaltered then histogram evening out is executed as palette change. Then again on the off chance that palette P stays unaltered and image is adjusted to I'=M (I) at that point the execution is by image change. By and large palette change is better as it safeguards the first information.

Following stage is Texture Extraction, An image surface is a lot of measurements determined in image preparing intended to evaluate the apparent surface of an image. Image surface gives the data about the spatial course of action of shading or powers in an image or chose locale of an image. The proposed algorithm is given below:

a) *Entropy channel restores the cluster J, where each yield pixel contains the entropy estimation of the 9-by-9 neighbourhood around the comparing pixel in the info picture I. Any entropy measurement can be:*

$$Entropy = - sum\ (p*log(p))$$

b) *Converts the matrix to the gray scale image.*
c) *Converts the gray scale image to a binary image.*
d) *Expel from a twofold picture every single associated part (protests) that have less than pixels, delivering another parallel picture.*
e) *Perform closing with the structuring element.*
f) *Fill the holes in the binary image.*
g) *Shows raw data*
h) *Convert the matrix to gray scale image.*
i) *Processes a worldwide edge (level) that can be utilized to change over a force picture to a twofold picture with im2bw. level is a standardized power esteem that lies in the range [0, 1].*
j) *Finally the affected region has been segmented from the non-affected region as shown in Fig 3.*
k) *GLCM is given by:*

$$C\Delta x, \Delta y(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} 1, & if\ I(p,q) = i\ and\ I(p + \Delta x, q + \Delta y) = j \\ 0, & otherwise \end{cases}$$

SVM calculations are overseen learning models with related learning [11], [14] estimations that dismember data and see structures, used for portrayal and backslide examination. Given a great deal of planning examples[10],[15]each set apart as having a spot with one of two orders, a SVM getting ready estimation builds a model that assigns new models into one grouping or the other, making it a non-probabilistic combined straight classifier[13]. The square graph of the proposed framework is appeared in figure 2.
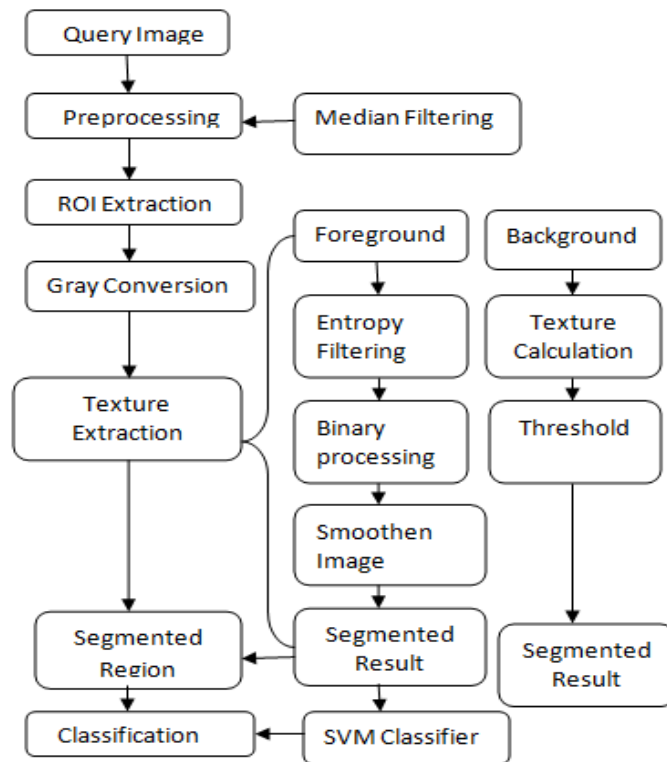
Fig 2: Block Diagram of Proposed System

The SVM classifier involves the following steps

    a)   Load the model stored in the file 'sample. Model'.
    b)   Classify the given dataset.
    c)   Train a new SVM on some new datasheets.
    d)   Change some of the SVM parameters.
    e)   Retrain the SVM with new parameters.
    f)   Perform cross-validation on the training set.
    g)   Save the model file.
    h)   Load the saved model from a file.
    i)   Retrieve the number of classes.
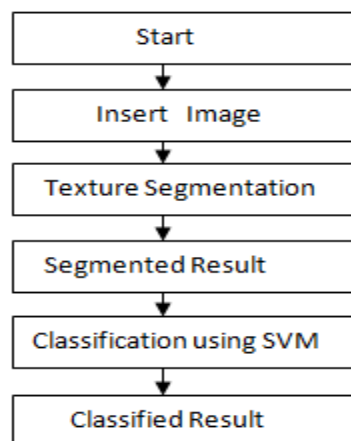   *j)   Display the output.*

Fig 3: Flow chart of the Proposed System

IV. Simulation Results

A SVM model is a depiction of the models as centres in space, mapped with the objective that the occurrences of the various classes are divided by an indisputable gap that is as wide as could sensibly be normal. New models are then mapped into that comparable space and foreseen to have a spot with a characterization reliant on which side of the opening they fall on. Henceforth the influenced district has been sectioned from the non-influenced locale and ordered it to be threatening or non-dangerous utilizing SVM classifier.

The Database has been created with the colour images provided by hospitals. The following table shows the parameters calculated for malignant and non-malignant images.

TABLE I DATABASE VALUES

Case	SRE	LRE	LGRE	HGRE	RESULT
1	0.51	78.67	0.12	79.6	Malignant
2	0.43	97.50	0.15	65.8	Non-Malignant

The obtained outputs are displayed in below figure 4 for the malignant cases.
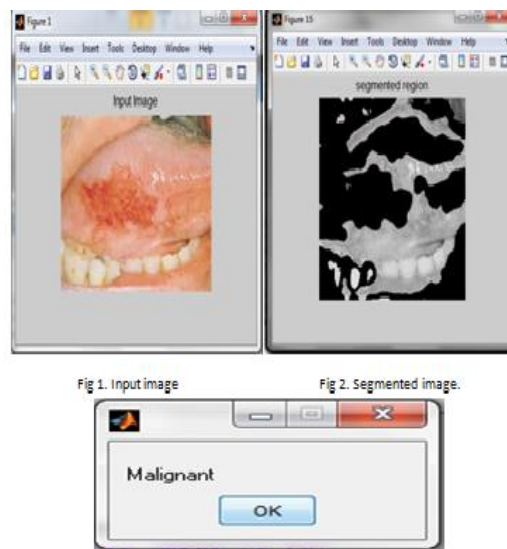

Fig 1. Input image   Fig 2. Segmented image.

Fig 4: Output of Malignant cases
TABLE II DATA SET ANALYSIS FOR ORAL CANCER DETECTION

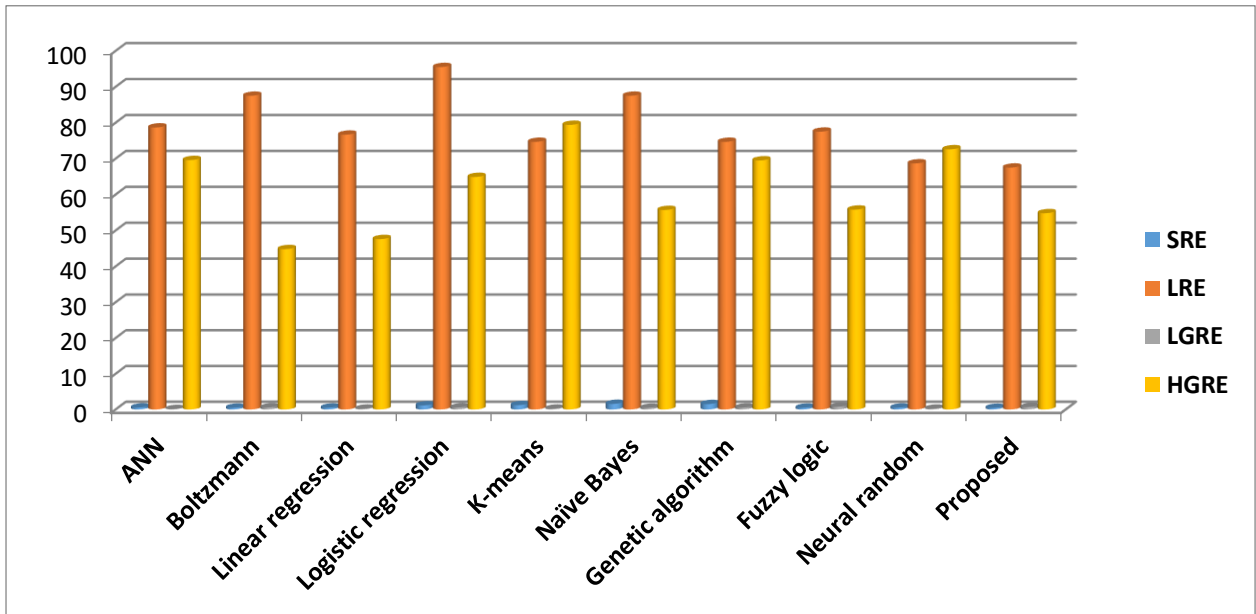| Machine learning algorithms	SRE	LRE	LGRE	HGRE
ANN	0.54	78.67	0.12	69.6
Boltzmann	0.48	87.50	0.55	44.8
Linear regression	0.50	76.67	0.22	47.6
Logistic regression	1.13	95.50	0.56	64.9
K-means	1.21	74.67	0.22	79.4
Naïve Bayes	1.56	87.50	0.45	55.7
Genetic algorithm	1.51	74.67	0.52	69.5
Fuzzy logic	0.48	77.50	0.75	55.8
Neural random	0.51	68.67	0.22	72.6
Proposed	0.43	67.50	0.65	54.8

Fig 5: Comparison for 8bit length

The obtained outputs are displayed in below figure 6 for the non-malignant cases.

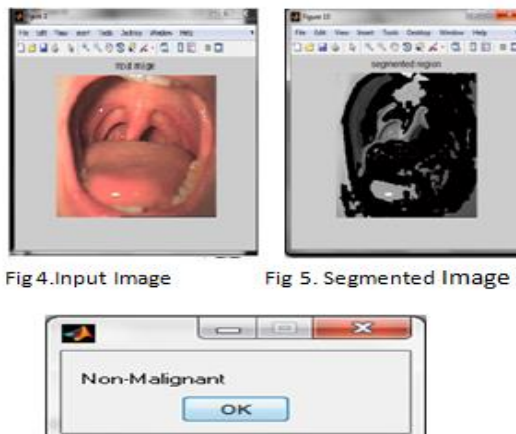Fig 4.Input Image   Fig 5. Segmented Image

Fig 6: Output of non-malignant cases

V. Conclusion

In this paper, Segmentation based on texture analysis and classification using SVM has been applied on cancer Images. Investigation on Textures works similarly, aside from the highs and lows are splendour esteems. GLCM surface thinks about the connection between two pixels one after another, called the reference and the neighbour pixel. In the proposed structure the neighbour pixel is picked to be the one toward the east (right) of each reference pixel. This can similarly be conveyed as a (1,0) association: 1 pixel in the x heading, 0 pixels in the y bearing. Each pixel inside the window transforms into the reference pixel in this way, starting in the upper left corner and proceeding to the lower right. Pixels along the right edge have no right hand neighbour, so they are not used for this check. Thus the proposed work uses the segmented region to classify into two categories such as Normal and Affected. For this classification the proposed system utilizes SVM classifier for classification purpose. This provides better classification accuracy while comparing with conventional approaches.

References

[1] T.F. Chan, L.A. Vese, "Active Contours without Edges," IEEE Trans. image processing, vol. 10, no.2, pp. 266-277, Feb 2001.

[2] C. Feng, S. Wu and N. Liu, "A user-centric machine learning framework for cyber security operations center," 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, 2017, pp. 173-175, doi: 10.1109/ISI.2017.8004902.

[3] A.Chodorowski, U. Mattsson, T.Gustavsson, "Oral Lesion classification using true color images," in 1999 Proc .of SPIE Conf.vol.3661, pp. 1127 –1138.

[4] P. Shimpi, S. Shah, M. Shroff and A. Godbole, "A machine learning approach for the classification of cardiac arrhythmia," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, 2017, pp. 603-607, doi: 10.1109/ICCMC.2017.8282537.

[5] B.V.Dasarathy, E.Holder, "Image characterizations based on joint gray level -run length distributions", Pattern Recogntion Letters, vol. 12, no.8, pp. 497-502, Aug 1991.

[6] S. Athmaja, M. Hanumanthappa and V. Kavitha, "A survey of machine learning algorithms for big data analytics," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8276028.

[7] Fritz Albregtsen "Statistical Texture Measures Computed from Gray Level Co-ocurrence Matrices", University of Oslo ,Tech. Rep., Nov 2008.

[8] M.M. Galloway, "Texture Analysis Using Gray Level Run Lengths", Computer Graphics and Image Processing, vol. 4, pp. 172-179, 1975.

[9] Cerrada, Mariela, et al. "Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal." Sensors 15.9 (2015): 23903-23926.

[10] C. Arunkumar, M. P. Sooraj and S. M. P. Ramakrishnan, "Finding expressed genes using genetic algorithm and extreme learning machines," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 2017, pp. 1-4, doi: 10.1109/ICACCS.2017.8014609.