Comparative Performance Analysis of Naive Bayes and SVM classifier for Oral X-ray images

Karthick.G
Anna University
Chennai, India
karthick.sgs@gmail.com

Harikumar.R
Bannari Amman Institute of Technology
Erode, India
harikumarrajaguru@gmail.com

***Abstract*— This paper presents the development of an automatic system for the classification of tooth wear disease diagnosis. Abnormal detection, disease detection and classification of oral images are substantial in the clinical research. Automatic diagnosis of the oral disease helps the medical practitioner to make decisions easily about the diagnosis process. The diagnosis models can be analyzed with the multiclass classification approach. The multiclass Naive Bayes and Support Vector Machine (SVM) classifier have been used for diagnosing the oral diseases and also their performance has been evaluated and compared. The experimental results are measured in terms of Sensitivity, Specificity, Precision, False Alarm and Accuracy which are considered as the performance parameters of classifier. Furthermore, our analysis proves that the SVM classifier achieves better result.**

***Keywords*— oral images, multiclass, Naive Bayes, SVM and classifier**

I. INTRODUCTION

The rapid development in medical image processing greatly helps in the detection of abnormalities and cancer in human [1]. Different forms of oral diseases are affecting the tooth which leads to the permanent loss of its structure. Over the past several decades, a huge number of measurement criteria have been developed to classify the presence of tooth wear diseases. However, as the perceptive of tooth wear diseases progressed, the clinical criteria systems remained focused on the assessment of disease process and few cases of disease diagnosis difficult. Accurate Identification of the oral problem from the X-ray image is considered to be a challenging job for the dentists. All patients are considered to be at the risk of developing oral problems and the examination should routinely involve looking for the clinical signs. According to the perspective of both the patient and dentist, the assessment on the severity of tooth wear is a subjective one. The stage at where the dentist decides that the tooth wear requires restoration is always a challenging one and it will be depend partly on the dentist's assessment and partly on the patient's wishes. Tooth wear indices or scores have been used to assist this clinical decision.

Many researchers reported diverse techniques for recognition of oral problems. Due to the progression in the medical field and computer technology, the oral problems and cancers can be identified by several techniques. Even though, lots of techniques available, the medical image processing is considered as the best way for diagnosis because of the advancements of soft computing techniques, machine learning and artificial intelligence. The proposed work consist detection for different types of oral diseases diagnosis. Fig. 1 represents Outline of the classification for oral images, X-ray oral images considers for the input (12-Dental caries, 12-Impacted molar, 15-Mandibular angle fracture, 15- Peri apical cyst, 18- Normal images). At first noise are removed by the preprocessing stage using median filter and then segmentation is done by Fuzzy c-means clustering technique. The main goal of segmentation is to representation of an image meaningful and make thing easier to analyze. In order to get the efficient features of images, Fuzzy c-means segmentation algorithm is applied before feature extraction.

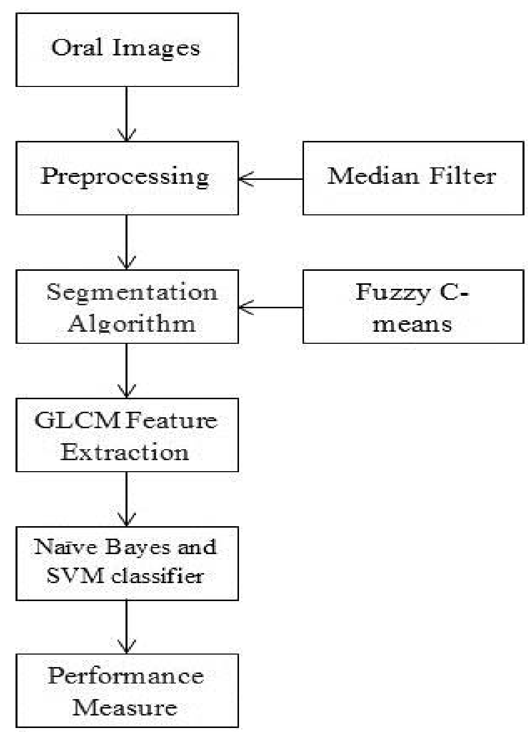

Figure 1 Outline for classification of oral images

Those segmented image features are extracted using GLCM technique. GLCM extracted features are given as input to the classifier. Finally classified images are labeled and performance of the classifier evaluated. This article is organized as follows: Section II discusses the literature

978-1-5090-3355-3/17/$31.00 ©2017 IEEE

reviews and Section III describes the GLCM feature extraction technique. Section IV concerns overview of the Naive Bayes and SVM classifiers for classification of oral X-ray images are discussed. Section V focuses on the performance measures of the classifier. Section VI contains results and conclusions.

II. Fuzzy C-Means (FCM) clustering algorithm

Segmentation plays a vital part in grouping an image into sub regions with respect to a particular presentation. The image might be having certain characteristics like color intensity, Gray Level, texture information, motion based on the calculation. FCM clustering technique has been broadly used for medical image segmentation. As medical images are frequently oxidized by noise and the FCM clustering technique is more sensitive to this noise. The foundation of fuzzy clustering was being proposed by Ruspini [2] and Bellman et al [3], after the development of fuzzy theory by Zadeh in 1965 [4]. In 1974, J.C. Dunn extended the hard means clustering to preliminary concepts fuzzy means [5] and J.C. Bezdek improved the FCM Algorithm by adding the fuzzy factor in 1981[6]. The benefit of FCM is the formation of new clusters from the data points that have close membership values to the existing classes. The procedure is explained below [7].

Step1: Initialize the values for c, cluster partition matrix U(0) and m. Label 'r' represents the number of steps involved.

Step2: The centroid vector C_{ij} is calculated for each step.

$$v_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_{kj}}{\sum_{k=1}^{n} (\mu_{ij})^m} \tag{1}$$

Step3: The Euclidean distance of the each cluster is calculated.

$$D_{ij} = \sqrt{\sum_{j=1}^{m} (x_{kj} - v_{ih})^2} \tag{2}$$

Step4: The partition cluster matrix should be updated for the r^{th} step.

$$\mu_{ij}^{r-1} = \frac{1}{\sum_{j=1}^{c} (\mathrm{d}_{ik}^{r} / \mathrm{d}_{jk}^{r})^{2/m-1}} \tag{3}$$

If $\| U^{k+1} - \mathrm{U}^k \| < \delta$ then the algorithm will be terminated otherwise return to step 3 by updating the cluster centroids.

III. GLCM feature extraction

The purpose of feature extraction is to transform the input image that contains too large information into a reduced representation with set of features. It consists to extract the most relevant features of an image and assign it into a label. In classification, an image is labeled according to its feature contents. In recent years, many feature extraction techniques have been developed and each technique has a strengths and weaknesses. Gray Level Co-occurrence Matrix (GLCM) is the most common way is by using feature extraction technique. GLCM has been introduced by Haralick in the year of 1973 [8], is a second order texture features and contains information about the positions of pixels having similar gray level values. GLCM is recognized as an important texture analytical method and calculates the statistical structures depend on gray level intensities of the image. The GLCM is a measure of different mixtures of pixel brightness values arise in an image.

A. Mean

$$\mu_x = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} iP(\mathrm{i,j}) \tag{4}$$

$$\mu_y = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} jP(\mathrm{i,j}) \tag{5}$$

B. Standard Deviation

$$\sigma_x = \left(\sum_{i=0}^{G-1} P_x(\mathrm{i})\ (\mathrm{i} - \mu_x)^2 \right)^{\frac{1}{2}} \tag{6}$$

$$\sigma_y = \left(\sum_{i=0}^{G-1} P_y(\mathrm{i})\ (\mathrm{i} - \mu_y)^2 \right)^{\frac{1}{2}} \tag{7}$$

C. Contrast

$$F_1 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(\mathrm{i,j})(\mathrm{i-j})^2 \tag{8}$$

D. Correlation

$$F_2 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \mathrm{P(i,j)} \left\lfloor \frac{(i-\mu_x)(\mathrm{j}-\mu_y)}{\sigma_x \sigma_y} \right\rfloor \tag{9}$$

E. Cluster prominence

$$F_3 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^4 P(\mathrm{i,j}) \tag{10}$$

F. Cluster Shade

$$F_4 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i + j - \mu_x - \mu_y\}^3 P(\mathrm{i,j}) \tag{11}$$

G. Dissimilarity

$$F_5 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i - j| P(i,j) \tag{12}$$

H. Energy

$$F_6 = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (P(\mathrm{i,j}))^2 \tag{13}$$

I. Entropy

$$F_7 = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(\mathrm{i,j}) \log(P(\mathrm{i,j})) \tag{14}$$

J. Homogeneity

$$F_8 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{1}{1+(\mathrm{i}-\mathrm{j})^2}\ \mathrm{P(i,j)} \qquad (15)$$

K. Maximum Probability

$$F_9 = \max i, jP(i,j) \qquad (16)$$

L. Autocorrelation

$$F_{10} = \frac{mn\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} f(i,j)f(i+p,j+q)}{(m-p)(n-q)\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} f^2(i,j)}, \qquad (17)$$

$$f = (\mu_{00}, \sigma_{00}, \ldots \mu_{ij}, \sigma_{ij})$$

Where p and q are the positional difference in the ith jth direction, and m and n are image dimensions [9].

TABLE I .EXTRACTED GLCM FEATURES USING FCM ALGORITHM

Features	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Image1	0.29	0.95	169.21	16.90	0.16	0.25	1.80	0.93	0.40	32.81
Image2	0.37	0.93	145.46	2.68	0.16	0.21	1.84	0.94	0.31	36.61
Image3	0.31	0.92	124.24	-13.73	0.18	0.21	1.90	0.92	0.30	44.42
Image4	0.68	0.88	150.55	-5.90	0.24	0.20	1.93	0.92	0.27	38.67
Image5	0.85	0.84	150.53	8.41	0.30	0.20	2.00	0.90	0.27	35.78
Image6	0.73	0.86	150.26	-11.86	0.29	0.19	2.05	0.90	0.31	42.31
Image7	0.99	0.85	184.11	5.06	0.33	0.22	1.94	0.89	0.36	36.37
Image8	0.36	0.91	78.89	0.53	0.16	0.29	1.61	0.94	0.40	41.92
Image9	1.08	0.76	109.48	5.04	0.43	0.16	2.26	0.84	0.29	32.88

IV. NAIVE BAYES CLASSIFIER

The Naive Bayes Classification represents a supervised learning method as well as a statistical method for classification. It underlies a posterior probabilistic model and it allows us to capture uncertainty about the model in a principled way by defining probabilities of the outcomes. It can explain diagnostic and predictive problems. Naive Bayes classifiers are among the most successful known algorithms for learning to classify images. Naive Bayes algorithm computes the posterior probability belongs to the class c_i depends upon the p_j [10].

$$p(C_i \mid \mathrm{p}_1, p_2, \ldots) = \coprod_j p(C_i \mid \mathrm{p}_j) \qquad (18)$$

According to the Bayes' rule,

$$p(C_i \mid p_j) = p(\mathrm{p}_j \mid \mathrm{C}_i)\frac{p(\mathrm{C}_i)}{p(\mathrm{p}_j)} \qquad (19)$$

using these probabilities in a maximum a posteriori, in which simply take the largest class probability as the classification result and the posterior probability is computed only based on the features. The estimated class $\hat{c}$ is

$$\hat{c} = \arg\max_i \prod_j p(p_j \mid C_i) \qquad (20)$$

V. SUPPORT VECTOR MACHINE CLASSIFIER

From statistical learning theory, SVM was derived by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963. Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik improved it in 1992 [11], by modifying the maximum-margin hyperplanes. SVM is a linear or non-linear classifier, which is a mathematical function that can distinguish two different kinds of images. It is a binary classifier and is used to classify the images and the purpose of classification is to group the items which have similar feature values [12]. It is used in hyper plane to define decision boundaries in addition to separating relevant and irrelevant vectors, sorting out among data points of different classes. In SVM, input data is mapped into higher dimensional space using kernel function for multi-class classification problem. One-Against-All, One-Against-One, Binary Tree and Directed Acyclic Graph are the different ways of solving the multi-class SVM system. One-Against One SVM classifier is the mostly used one. The features of medical images are connected with the class label of m subjects. A set of binary classifiers f¬1, f2,...,fM , is constructed and trained to distinguish one class from the other remaining (M-1) classes and they are combined to get the multi-class classification [13,14].

$$\Phi(\mathrm{w}, \xi) = \frac{1}{2}\sum_{m=1}^{M}(\mathrm{w}_m^T \mathrm{w}_m) + C\sum_{i=1}^{N}\sum_{m \neq y_i}\xi_i^m \qquad (21)$$

Subject to constraints,

$$(W_{y_i}^T \cdot X_i) + b_{y_i} \geq (W_m^T \cdot X)_i + b_r + 2 - \xi_i^m \qquad (22)$$

$$\xi_i^m \geq 0 \text{, for i=1,2...N: m} \in \{1,2\ldots M\} \backslash\{yi\}$$

Where, the coefficients w and b are calculated using the quadratic equation. Here, are slack variables, yi ∈ {1,...,M} are the multiclass labels of the features vectors and m ∈ {1,...,M}\ yi are the multiclass labels except yi . N is the coefficient which is independent of the number of classes and M). Once the training part of the SVM is completed, the K features are extracted from the test images by applying the SVM classifier. By comparing the feature set of test images with the trained images the nearest value and the corresponding class is identified and showed as a result class. The decision boundary is given by,

$$f(\mathrm{x}) = arg\ max_m (W_m^T \cdot x + b_m), \quad (23)$$
$$for\ m = 1,2..m$$

The final classifier will be in the form of,

$$f(\mathrm{x}) = \sum \alpha_i x_i^T x + b_m \quad (24)$$

The number of non-zero α is reduced by directly attempting the regularization.

VI. PERFORMANCE MEASURES

The performance evolution is the method which is used to measure the uses of segmentation in SVM classifier and is carried out using the measures like Sensitivity, Specificity, Precision, False Alarm (FA) and Accuracy. When comparing to other segmentation algorithms, EM algorithm produces the maximum SVM classifier accuracy. The performance of the classifier can be analyzed using the following performance measures [15,16].

True Positive (TP): Abnormal brain correctly identified as abnormal.

True Negative (TN): Normal brain correctly identified as normal.

False Positive (FP): Normal brain incorrectly identified as abnormal.

False Negative (FN): Abnormal brain incorrectly identified as normal.

Sensitivity: To identify the positive results, sensitivity relates to the test's ability.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP+FN})} \quad (25)$$

Specificity: To identify the negative results, specificity relates to the test's ability.

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN+FP})} \quad (26)$$

Precision: Precision which also known as the positive predictive value is the probability that a positive prediction is correct.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \quad (27)$$

False Alarm (FA): FA is also called False Positive Rate (FPR). It is the proportion of negatives cases that were incorrectly classified as positive.

$$\text{False Alarm} = \text{1-Specificity} \quad (28)$$

Accuracy: It is the proportion of the total number of predictions that were correct.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (29)$$

TABLE II. PERFORMANCE ANALYSIS OF NAIVE BAYES CLASSIFIER FOR ORAL IMAGES

Images	Sensitivity (%)	Specificity (%)	Precision (%)	False Alarm (%)	Accuracy (%)
Class 1	50	94.44	66.67	5.56	72.22
Class 2	25	88.89	33.33	11.11	56.94
Class 3	40	76.47	33.33	23.53	58.24
Class 4	20	70.59	16.67	29.41	45.29
Class 5	50	88.89	50	11.11	69.44

TABLE III. PERFORMANCE ANALYSIS OF SVM CLASSIFIER FOR ORAL IMAGES

Images	Sensitivity (%)	Specificity (%)	Precision (%)	False Alarm (%)	Accuracy (%)
Class 1	100	100	100	0	100
Class 2	100	100	100	0	100
Class 3	100	100	100	0	100
Class 4	100	100	100	0	100
Class 5	100	100	100	0	100

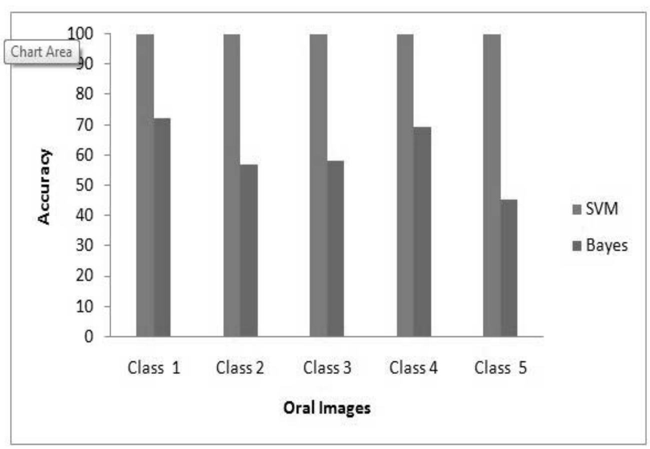

Figure 2 Accuracy comparison of classifier

VII. RESULTS AND CONCLUSION

In this session, we present the experimental results of applying GLCM feature extraction method in the Naive Bayes and SVM classifier. All the experimentation where conducted on the 72 X-ray oral images. Images are categorized by five classes. Class 1 to class 5 represented by the Dental caries, Impacted molar, Mandibular angle fracture, Peri apical cyst and Normal oral images respectively. As a first step, each images had been preprocessed and segmentation done by

Fuzzy C-Means algorithm. The Table I represent the GLCM features extracted data using Fuzzy C-Means segmentation algorithms. The most common 10 GLCM textured features are extracted from the oral X-ray images. In each classifier system split the images into two parts: training images and testing images. The performance measures used to evaluate and analyze the results are: Sensitivity, Specificity, Precision, False Alarm and Accuracy. The Table II and Table III describe the numerical result of Naive Bayes and SVM classification obtained by GLCM feature extraction used in multiclass classification. Figure 2 shows that accuracy comparison of Naive Bayes and SVM Classifier The analysis of the result shows that SVM classifier reached high classification accuracy rate compared to Naive Bayes classifier.

References

[1] T. Acharya and A.K. Ray, Image processing: Principles and applications, Wiley, Hoboken, 2005.

[2] Ruspini, E.H.: A new approach to clustering. Inf. Control 15(1), 22–32 (1969).

[3] Bellman, R.E., Kalaba, R.A., Zadeh, L.A.: Abstraction and pattern classification. J. Math. Anal. Appl. 13, 1–7 (1966).

[4] Zadeh, L.A.: Fuzzy sets. Inf. Control 8(3), 338–353 (1965).

[5] Dunn, J.C.: A fuzzy relative ISODATA process and its use in detecting compact wellseparated clusters. J. Cybern. 3(3), 32–57 (1974)

[6] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981).

[7] J. Nayak, B. Naik, H. Behera, Fuzzy c-means (FCM) clustering algorithm: A decade review from 2000 to 2014, in: 557 Computational Intelligence in Data Mining-Volume 2, Springer, pp. 133–149, 2015.

[8] K. Shanmugam, R.M. Haralick, and I.H. Dinstein, Textural features for image classification, IEEE Trans Syst Man Cybernet 3 (1973), 610–621.

[9] S. Selvarajah and S.R. Kodituwakku, Analysis and comparison of texture features for content based image retrieval, Int J Latest Trends Computing 2(2011).

[10] Boser, B.E., Guyon, I., & Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop of Computational Learning Theory, 5, 144-152, Pittsburgh, ACM.

[11] B.K. Elfarra and I.S.I. Abuhaiba, New feature extraction method for mammogram computer aided diagnosis, Int J Signal Process Image Process Pattern Reorganization 6 (2013), 13–36.

[12] F.F. Chamasemani and Y. P. Singh, Multi-class support vector machine classifiers—An application in hypothyroid detection and classification, Sixth International Conference on Bio-Inspired Computing: Theories and applications, 2011.

[13] V. Vapnik, Statistical learning theory, Wiley, New York, 1998.

[14] E. Costa, A. Lorena, A. Carvalho, and A. Freitas, A review of performance evaluation measures for hierarchical classifiers, in Evaluation Methods for Machine Learning II: Papers from the 2007 AAAI Workshop. AAAI Press, pp. 1–6, 2007.

[15] J.M. Sokolova and G. Lapalme, A systematic analysis of performance measures for classification tasks, Inform Process Manage 45 (2009), 427–437.

[16] Harikumar Rajaguru, Karthick Ganesan, Vinoth Kumar Bojan, Earlier detection of cancer regions from MR image features and SVM classifiers, International Journal of Imaging Systems and Technology, vol:26(3), pp: 196-208, 2016.