A Survey on Image Classification Methods and Techniques for Improving Accuracy

[1]R.R Darlin Nisha, [2]V.Gowri
[1]PG Scholar, [2]Assistant professor
Department of Computer Science and Engineering
SNS College of Engineering
coimbatore

***Abstract*--Image classification is a composite process that may be precious by many factors. The stress is placed on the summarization of major advanced classification approaches and the techniques used for improving classification accuracy. In accumulation, some important issues affecting classification performance are discussed. This literature evaluation suggests that designing a suitable image-processing procedure is a requirement for a successful classification of a little sensed data into a thematic map. Effective use of multiple features of distantly sensed data and the selection of a suitable classification method are especially significant for improving classification accuracy. Non-parametric classifiers such as decision tree classifier, neural network, SVM classifier and knowledge-based classification have increasingly become important approaches for data classification. More research is needed to identify and reduce uncertainties in the image-processing chain to improve classification accuracy.**

***Keywords*: decision tree, neural network, SVM classifier, k-NN classifier.**

I. INTRODUCTION

Lung cancer is a disease of unusual cells multiplying and growing into a tumor. Cancer cells can be carried away from the lungs in blood, or lymph fluid that backdrop lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes situated in the lungs and in the centre of the chest. Lung cancer repeatedly spreads toward the centre of the chest because the normal flow of lymph out of the lungs is toward the centre of the chest. Metastasis occurs after a cancer cell leaves the site where it began and moves into a lymph node or to an extra part of the body throughout the blood stream [1]. Cancer that starts in the lung is known as primary lung cancer. There are three different types of lung cancer, and divided into two main groups: Non-small cell lung cancer and Small cell lung cancer which has three subtypes: Squamous cell carcinomas, and Adenocarcinoma. The class order of cancers for both males and females among Jordanians in 2008 indicated that there were 356 cases of lung cancer accounting for (7.7 %) of all newly diagnosed cancer cases in 2008. Lung cancer affect 297 (13.1 %) males and 59 (2.5%) females with a male to female share of 5:1 which Lung cancer ranked second between males and 10th among females[2]. The first stage starts with attractive a group of CT images [3]. The second stage applies more than a few techniques of image enhancement, to get most excellent level of quality and clarity. The third phase applies image segmentation algorithms which play a proficient rule in image processing stages, and the last stage obtains the broad features from enhanced segmented image which gives indicators of normality or abnormality of images. Lung cancer is the most unsafe and widespread cancer in the world according to phase of detection of the cancer cells in the lungs, so the method early detection of the disease acting a very important and essential role to avoid serious advanced stages to decrease its percentage of distribution.

II. NAIVE BAYES CLASSIFIER

The naive Bayes classifier is based on a probability representation and assigns the class, which has the greatest estimated subsequent probability, to the feature vector extracted from the ROI. This process is optimal when the attributes are orthogonal. However it performs well without this statement. The effortlessness of the method allows good

ISSN: 2278 – 1323

performance with small training sets. Certainly, by building probabilistic models, it is robust to outliers. In addition, it creates soft decision boundaries, which has the outcome of avoiding overtraining. However, the arbitrary option of the distribution model for estimating the probabilities $P(x)$ along with the lack of flexibility of the decision boundaries fallout in limited performance for complex multiclass configurations.

III. K-NEAREST NEIGHBOR

The k-nearest neighbor classifier cuts out hyperspheres in the space of instances by conveying the majority class of the k-nearest instances according to a defined metric. It is asymptotically optimal and also its implementation allows speedy tests [5]. However, quite a lot of shortcomings are inherent to this method. It is very sensitive to the irritation of the dimensionality [6]. Certainly, increasing the dimensionality has the effect to sparse the feature space, and local homogeneous regions that signify the prototypes of the diverse classes are spread out. The classification performance robustly depends upon the used metric. Moreover, a small value of k results in chaotic boundaries and makes the process very aware to outliers.

IV. J48 DECISION TREES

The J48 decision trees algorithm divides the feature space sequentially by choosing mostly features with the highest information gain. J48 is an achievement of the C4.5 algorithm. In medicine, it is in communication to the approach used by clinicians to create a diagnosis by answering following questions. This is however only partially true when radiologists interpret HRCT images. However, it is aware to the variability of data. The makeup of the tree is possible to change completely when a new instance is added to the teaching set. Another disadvantage is its failure to detect interactions among features, as it treats them separately. This results in decision boundaries that are orthogonal to dimensions, which is not accurate for greatly nonlinear problems.

V. MULTI-LAYER PERCEPTRON

MLPs are inspired by the human nervous system where in sequence is processed during unified neurons. MLP is a feed-forward neural network, which defines that the information propagates as of input to output. The inputs are fed with principles of each feature and the outputs give the class value. Through one layer of neurons, the output is a weighted linear blend of the inputs. This network is branded as linear perceptron. By totaling an extra layer of neurons with nonlinear foundation functions (the hidden layer), a nonlinear mapping among the input and output is prospective. The teaching phase consists of iterative optimization of the weights concerning the neurons by minimizing the mean squared error rate of organization. The learning rate, which controls the adjustments of the weights throughout the teaching phase, must be elected as a trade-off among mistake on the training set and overtraining. An additional critical constraint is the number of units, of the hidden layer. Definitely, the MLP is subject to over fitting and requires an optimal choice of the parameters for regularization. The MLP can generate models with arbitrary difficulty by drawing infinite decision boundaries. It is also strong to noisy features, as these will find a low weight after teaching.

VI. KERNEL SUPPORT VECTOR MACHINES

Kernel SVMs implicitly map input feature vectors to a higher dimensional space by using the kernel function with the width of the Gaussian. In the transformed space, a maximal extrication hyper plane is built considering a two-class problem. Two parallel hyper planes are constructed symmetrically on both side of the hyper plane that separates the data. The aim is to exploit the distance between the two external hyper planes, called the margin. An declaration is made that the enhanced the margin is, the improved the simplification error of the classifier will exist. Indeed, SVMs were residential according to the structural risk minimization attitude which seeks to minimize an upper clear of the generalization error, while most of the classifiers aim at minimizing the empirical risk, the error on the training set. The SVM algorithm aims at finding a decision function, which minimizes the functional.

The SVMs permit training, nonlinear classifiers in high-dimensional spaces using a minute training set. This is enabled during the selection of a division of vectors (called the support vectors) which characterizes the right boundaries between the classes fit.

VII. CLASSIFIERS WORN FOR LUNG TISSUE CLASSIFICATION IN HRCT DATA

A brief review of the recent techniques used for the categorization of lung tissue patterns in HRCT data is given in this section. The system classifies ROIs using decision trees that minimize the entropy over

ISSN: 2278 – 1323

all distributions associated with lung tissue classes and matches the ROI beside suggestion images in JPEG (and not DICOM) that are already indexed in the database. The hierarchical association of the features forced by the structure of decision trees assigns too much consequence to the first selected attributes and is not adapted to integrate information from a set of complementary attributes such as gray-level histogram bins. The small dataset used (nine HRCT image series) leads to a biased classification job, as training and testing using succession from the same patient create instances artificially close together in the feature space.

Nonlinear binning of gray-level morals for co-occurrence matrices is planned in order to be eligible lung tissue fibrosis in HRCT data. A least amount Mahalanobis detachment classifier is used. This extensive naive Bayes classifier relies on the guess that the prospect density functions of the classes are Gaussian, chief to non-flexible decision boundaries.

Two classifiers along with two feature selection techniques are evaluated all through their capability to detect fibrosis in HRCT images using co-occurrence matrices. They are J48 and naive Bayes and decision trees. The attribute selection method showed development of categorization accuracy, whereas two classifier families achieved equivalent performance. Unmoving, the dataset used for testing is fairly small and the classifiers may not be flexible enough for multiclass troubles. Information about the localization of the lung tissue patterns within a lung atlas is incorporated as an additional feature which allows a collection accuracy improvement.

Six lung tissue patterns are classified using an adaptive several texture feature method. Correlated features are in poles apart and a Bayesian classifier is worn. The final might not be accurate for classifying any kind of lung tissue, as it is sensitive to the selection of the probability density occupation of the features.

VIII. CONCLUSION

Classification and quantification of interstitial lung disease is complicated, and even experienced chest radiologists frequently resist with differential diagnoses. Automated schemes that indicate a percentage of affected lungs or the opportunity of a certain disease would certainly be address, but require much more research. Accuracy assessment is an integral part in an image classification procedure.

Accuracy assessment based on error matrix is the most frequently employed advance for evaluating per-pixel classification. The success of an image classification depends on many factors, major one is accuracy. The combination of different classification approaches has shown to be helpful for improvement of classification accuracy.

REFERENCES

[1] Non-Small Cell Lung Cancer, Available at: http://www.katemacintyrefoundation.org/Pdf/non-small-cell.pdf, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, (accessed July 2011).
[2] Tarawneh M., Nimri O., Arqoub K., Zaghal M., Cancer Incidence in Jordan 2008,Available at:http://www.moh.gov.jo/MOH/Files/Publication/Jordan%20Cancer%20Registry_2008%20Report_1.pdf, 2008, (accessed July 2011).
[3] Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/ datac/signon.cgi, (accessed July 2011).
[4] Gonzalez R.C., Woods R.E., Digital Image Processing, Upper Saddle River, NJ Prentice Hall, 2008.
[5] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman, “MDCTbased 3-D texture classification of emphysema and early smoking related lung pathologies,” IEEE Trans. Med. Imag., vol. 25, no. 4, pp. 464–475, Apr. 2006.
[6]L. Sorensen, S. B. Shaker, and M. de Bruijne, “Quantitative analysis of pulmonary emphysema using local binary patterns,” IEEE Trans. Med. Imag., vol. 29, no. 2, pp. 559–569, Feb. 2010.

ISSN: 2278 – 1323