Lung Nodule Classification Techniques for Low Dose Computed Tomography (LDCT) Scan Images as Survey

Suganya A.
Student, Department of CSE,
Vivekanandha College of
Engineering for Women,
Tiruchengode, Namakkal,
Tamilnadu, India

Mohanapriya N.
Assistant Professor/ CSE,
Vivekanandha College of
Engineering for Women,
Tiruchengode, Namakkal,
Tamilnadu, India

B. Kalaavathi, PhD
Professor/ CSE, K.S.R Institute
for Engineering and
Technology, Tiruchengode,
Namakkal, Tamilnadu, India

ABSTRACT

Image classification is one of the most complex areas in image processing. It is more complex and difficult to classify if it contain blurry and noisy content. Medical image Classification is an important task within the field of computer vision. Here, perform four type of lung nodules classification , i.e., well-circumscribed, vascularized, juxta-pleural and pleural-tail, in low dose computed tomography (LDCT) scans images. The classification method combine the lung nodule and surrounding anatomical structures through contextual analysis. Here, presents a survey on the various techniques used for classifying image which is mainly based on object in the given image. The proposed method was classified on a publicly available dataset.

Keywords
LDCT, Image, vascularized.

1. INTRODUCTION

Lung cancer, also known as carcinoma of the lung or pulmonary carcinoma, is a malignant lung tumor characterized by uncontrolled cell growth in nodule of the lung. If the left side of lung untreated, this growth can spread beyond the lung by process of metastasis into nearby tissue or other parts of the lung. Most cancers that start in the lung, known as primary lung cancers, are carcinomas that derive from epithelial cells. The main primary types are small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma(NSCLC). The most common symptoms are coughing (including coughing up blood), weight loss, shortness of breath, and chest pains. The vast majority 80–90% of cases of lung cancer are due to long-term exposure to tobacco smoke. About 10–15% of cases occur in people who have never smoked. These cases are often caused by a combination of genetic factors and exposure to radon gas, asbestos, or other forms of air pollution, , including second-hand smoke. At present, the classification from Diciotti et al. [1] is the most popular approach and it divides four types of lung nodule: well-circumscribed (W) with the nodule located centrally in the lung without any connection to vasculature; vascularized (V) with the nodule located centrally in the lung but closely connected to neighboring vessels; juxta pleural (J) with a large portion of the nodule connected to the pleural surface; and pleural-tail (P) with the nodule near the pleural surface connected by a thin tail. Lung cancer may be seen on chest radiographs and computed tomography (CT) scans. CT scan form images from X-rays and it is combination of many X-rays images taken from different angle to produce cross sectional images of specific area of object. In CT scan timing, voltage, and beam current are controlled by computer through control bus.

2. CLASSIFICATION TECHNIQUES

There are several ways to classify the images. Classification techniques are used to extract information from a data set or image and transform it into an understandable structure for further use. Classification is the organization of data in given classes. Also known as supervised classification, the classification uses given class labels to order the objects in the image classification. Classification approaches normally use a training set where all objects are already associated with known class labels.

2.1 K-Nearest Neighbor Algorithm

K-nearest neighbor algorithm introduced by M.Soundarya and R.Balakrishnan[5].Nearest neighbor classifiers are based on learning by analogy. Each sample represents a point in an n-dimensional space. All of the training samples are stored in an n-dimensional pattern space. When given an unknown sample, a k-nearest neighbor classifier searches the pattern space for the k training samples that are closest to the unknown sample. "Closeness" is defined in terms of Euclidean distance, where the Euclidean distance, where the Euclidean distance between two points, $X=(x1,x2,\ldots\ldots,xn)$ and $Y=(y1,y2,\ldots.,yn)$.In the classification , k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

2.2 Decision Tree Classifier (DT)

Decision tree classifier introduced by S.Arunadevi1,Dr. S. Daniel Madan [4] .DT is combination of mathematical and computational technique to description, categorization and Simplification. of set of data. DT classify image by repeat partitioning a dataset into uniform subset. Instances are classified starting at the root node and sorted based on their feature values. Decision tree describe data but not about decision. DT have two main types-*classification tree* used to calculate the outcome of class to which data belongs and *regression tree* calculate outcome when the outcome is considered as real number.

2.3 Fuzzy Measure

Fuzzy Measure introduced by M .Soundarya and R.Balakrishnan [6]. Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

Propositional fuzzy logics:

The most important propositional fuzzy logics are given by below:

1) Monoidal t-norm-based propositional fuzzy logic MTL is an axiomatization of logic where conjunction is defined by a left continuous t-norm, and implication is defined as the residuum of the t-norm. Its models correspond to MTL-algebras that are pre-linear commutative bounded integral resituated lattices.

2) Product fuzzy logic is the extension of basic fuzzy logic BL where conjunction is product t-norm. It has the axioms of BL plus another axiom for cancel activity of conjunction, and its models are called product algebras.

2.4 Artificial Neural Network Based Classification (ANN)

ANN is introduced by Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal [2]. ANN is a computational model inspired by the biological neural network. It could be considered as a weighted directed graph in which nodes are neurons and edges with weights are connection among the neurons. Each artificial neuron computes a weighted sum of its input signals and generates an output, based on certain activation functions, such as piecewise linear, sigmoid, Gaussian, etc. It consists of one input layer, one output layer, and depending on the application it may or may not have hidden layers. The number of nodes at the output layer is equal to the number of information classes, whereas the number of nodes at the input is equal to the dimensionality of each pixel. Feed-forward ANN with the back propagation learning algorithm is most commonly used in ANN literature. In the learning phase, the network must learn the connection weights iteratively from a set of training samples. The network gives an output, corresponding to each input. The generated output is compared to the desired output. The error between these two is used to modify the weights of the ANN. The training procedure ends when the error becomes less than a predefined threshold. Then, all the testing data are fed into the classifier to perform the classification.

2.5 Support Vector Machine

Support Vector Machine is introduced by S.Archana, Dr. K.Elangovan [3]. The support vector machine usually deals with pattern classification that means this algorithm is used mostly for classifying the different types of patterns. Now, there is different type of patterns i.e. Linear and non-linear. Linear patterns are patterns that are easily distinguishable or can be easily separated in low dimension whereas non-linear patterns are patterns that are not easily distinguishable or cannot be easily separated and hence these type of patterns need to be further manipulated so that they can be easily separated. Basically, the main idea behind SVM is the construction of an optimal hyper plane, which can be used for classification, for linearly separable patterns. The optimal hyper plane is a hyper plane selected from the set of hyper planes for classifying patterns that maximizes the margin of the hyper plane i.e. the distance from the hyper plane to the nearest point of each patterns. The main objective of SVM is to maximize the margin so that it can correctly classify the given patterns i.e. larger the margin size more correctly it classifies the patterns.

2.6 Bayesian Classifier

Bayesian classifier was introduced by Jipsa Kurian, V.Karunakaran[7]. The goal of classification is to classify examples or samples to a class based on the value of several attributes. It approximate the joint probability of the class and attributes. Bayesian classification use chain rule to decompose the joint distribution. Chain rule is a formula for computing composition of two or more function.

$$P_r(C,A_1,...A_k)=P_r(C)\ P_r(A_1,...A_k/C)$$

The first term on the right side is previous probability of the class label. This can directly estimated from the training data. Second term on Right hand is distribution of attribute value given in label. Once estimate $P_r(C)$ and $P_r(A_1,...A_k/C)$, use bayes rule to get conditional probability of class,

$$P_r(A_1,...A_k/C) = \alpha\ P_r(C)\ P_r(A_1,...A_k/C)$$

Table 1: Advantages And Disadvantages Of Classification Techniques

S.NO	CLASSIFIER	ADVANTAGES	DISADVANTAGES
1.	k-nearest neighbor Algorithm	• It is an easy to understand and easy to implement classification technique. • Training is very fast. • Robust to noisy training data. • 4.It is particularly well suited for multimodal classes	• It is sensitive to the local structure of the data. • Memory limitation. • Being supervised learning lazy Algorithm i.e., runs slowly.
2.	Decision Tree classifier	• Nonlinear relationships between parameters do not affect tree performance • Easy to interpret and explain	• Complexity • Possibility of duplication with the same sub-tree on different paths
3.	Fuzzy Measure	• Efficiently handles uncertainty. • Properties are describe by identifying various stochastic relationships.	• Without priori knowledge output is not good • Precise solutions depends upon direction of decision

4.	Artificial Neural Network Based Classification	• Efficiently handles noisy inputs. • Computational rate is high	• It is semantically poor. • Difficult in choosing the type network architecture.
5.	Support Vector Machine	• Produce very accurate classifiers. • Less over fitting, robust to noise. • Especially popular in text classification problems where very high-dimensional spaces are the norm. • Memory-intensive.	• SVM is a binary classifier. To do a multi-class classification, pair-wise classifications can be used (one class against all others, for all classes). • Computationally expensive, thus runs slow.
6.	Bayesian Classifier	• To improves the classification performance by removing the irrelevant features. • Good performance. • It is short computational time	• **Information theoretically infeasible** • **Computationally infeasible**

Table 2: Comparisons Of Classification Techniques Based On Performance And Accuracy

CLASSIFIERS	PERFORMANCE	ACCURACY
K-Nearest Neighbor Algorithm	92.78%	90.54%
Decision Tree Classifier(DT)	92.40%	88.32%
Fuzzy Measure	90.66%	90.00%
Artificial Neural Network Based Classification(ANN)	92.30%	91.50%
Support Vector Machine	94.80%	95.30%
Bayesian Classifier	89.05%	91.70%

3. CONCLUSION

This paper deals with various classification techniques used in medical image processing and a study on each of them. Image Processing is a wide area that integrates techniques from various fields including machine learning, artificial intelligence, statistics and

pattern recognition, for the analysis of large volumes of dataset. These classification algorithms can be implemented on different types of publically available data sets. Many techniques were used In this paper, SVM provides accurate results when compared to other classifier

4. REFERENCES

[1] S. Diciotti, G. Picozzi, M. Falchini, M. Mascalchi, N. Villari, and G. Valli, "3-D segmentation algorithm of small lung nodules in spiral CT images," IEEE Trans. Information Technology in Biomedicine, vol. 12, no. 1, pp. 7–19, 2008.

[2] Pooja Kamavisdar, Sonam Saluja, Sonu Agrawal, " A Survey on Image Classification Approaches and Techniques ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, 2013.

[3] S.Archana, Dr. K.Elangovan ," Survey of Classification Techniques in Data Mining " International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 2, pg. 65-71,2014.

[4] S.Arunadevi1,Dr. S. Daniel Madan Raja, " A Survey on Image Classification Algorithm Based on Per-pixel ", International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November 2014.

[5] M.Soundarya, R.Balakrishnan," Survey on Classification Techniques in Data mining ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.

[6] Ravivarman.R, Sasirekha.N," Survey On Lung Nodule Classifications ", International Journal of Research in Engineering and Technology ,Volume: 04 Issue: 01 ,2015.

[7] Jipsa Kurian, V.Karunakaran," A Survey on Image Classification Methods ", International Journal of

Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 1, Issue 4, October 2012.

[8] Ameet Joshi, Shweta Bapna, Sravanya Chunduri, " Comparison Study of Different Pattern Classifiers", Mulder server at MSU,2012.

[9] Fan Zhang ,Yang Song,Weidong Cai, Min-Zhao Lee, Yun Zhou, Heng Huang, Shimin Shan, Fulham M.J., Feng D.D.,"Lung Nodule Classification With Multilevel Patch BasedContext Analysis", IEEE transaction on

Biomedical Engineering,volume:61,2014 , Page(s): 1155 – 1166.

[10] M. Gomathi and Dr. P. Thangaraj," Lung Nodule Detection using a Neural Classifier", IACSIT International Journal of Engineering and Technology, Vol.2, No.3, June 2010.

[11] P Sathya, L Malathi, "Classification and segmentation in satellite imagery using back propagation algorithm of ANN and k-means algorithm", Int J Mach Learn Comput, Volume 1,Issue 4 Pages 422-426   2011.