ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Identifying Lung Cancer Using X-Ray: A Review

Nilakshi Maruti Mule, Prashant Mane, Premkumar Yadu Mohite

M.E. Student, Dept. of Computer Engineering, Zeal Education Society's DCOER, Narhe, Pune, Maharashtra, India

Asst. Professor, Dept. of Computer Engineering, Zeal Education Society's DCOER Narhe Pune, Maharashtra, India

U. G. Student, Dept. of Electronic & Telecommunication, DACOE, Karad, Maharashtra, India

ABSTRACT: Chest X-ray image used for detection of lung cancer. The early detection and treatment of lung nodule are the most challenging clinical tasks performed by radiologists. In this paper, processing the X-ray images using various image processing algorithms like Genetic algorithm, K-nearest neighbour algorithm, Decision tree algorithm, Multilayer perceptron algorithm and various other techniques like MAD(Median Absolute Deviation) .X-ray images of lungs as an input and will give the details of the lung nodule as if it is cancerous or not as the output.

KEYWORDS: Genetic Algorithm, K-NN Classifier, chest X-Ray, Gabor Filter, Feature Extraction, Feature Selection

I. INTRODUCTION

Lung cancer is the uncontrolled growths of tissues in the lung. The vast majorities (80–90%) of cases of lung cancer are due to long-term exposure to tobacco smoke. However, about 10–15% of cases occur in people who have never smoked. Causes of lung cancer are combination of coal factories & exposure to radon gas, asbestos, or other forms of air pollution. In the world, lung cancer cases are approximately reported as 14 million new cases and 8.2 million cancer related deaths in 2012 as per the statistics given by WHO in 2012. In the detection phase of lung cancer if we use the CT scan it will take lot of time and money detect the nodule since it is a 3D approach while the X-ray image is 2D consumes less time and money. CT scan gives extra clarity to the features but applying suitable algorithms the X-ray image can be used to detect and classify the lung nodule. Classifier is used to identify the nodule is cancerous or not. The current computer Tomography scanning is costlier even for middle class family. We gone through some research papers and came to know that using various algorithms and using X-ray images cost can be reduced and efficiency can be improved .This made us to propose this system for classifying the lung nodules.

II. RELATED WORK

In Computerized Detection of Lung Nodules by Means of "Virtual Dual-Energy" Radiography (2013) by Sheng Chen and Kenji Suzuki, Member, IEEE explain that, Major challenges in current computer-aided detection(CADe) schemes for nodule detection in chest radiographs(CXRs) are to detect nodules that overlap with ribs and/or clavicles and to reduce the frequent false positives (FPs) caused by ribs. Detection of such nodules by a CADe scheme is very important, because radiologists are likely to miss such subtle nodules. Our purpose in this study was to develop a CADe scheme with improved sensitivity and specificity by use of "virtual dual energy"(VDE) CXRs where ribs and clavicles are suppressed withassive-training artificial neural networks (MTANNs). The VDE technology suppressed rib and clavicle opacities in CXRs while maintaining soft-tissue opacity by use of the MTANN technique that had been trained with real dual-energy imaging. The original scheme without VDE technology achieved sensitivity of 78.6% (110/140) with 5 (1165/233) FPs per image. By use of the VDE technology, more nodules overlapping with ribs or clavicles were detected and the sensitivity was improved substantially to 85.0% (119/140) at the same FP rate in a leave one-out cross-validation test, whereas the FP rate was reduced to2.5 (583/233) per image at the same sensitivity level as the original CADe scheme obtained(1). In Image feature analysis and computer-aided diagnosis in digital radiography (1998) By S Sanada, KDoi and Heber McMahon. Explain that, In order to aid radiologists in the diagnosis of pneumothorax from chest radiographs, an automated method for detection of subtle pneumothorax is being developed. The computerized method is based on the detection of a fine curved-line pattern, which is a unique feature of

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

radiographic findings of pneumothorax. Initially, regions of interest (ROIs) are determined in each upper lung area, where subtle pneumothoraces commonly appear. The pneumothorax pattern is enhanced by the selection of edge gradients within a limited range of orientations. Rib edges included in this edge-enhanced image are removed, based on the locations of posterior ribs that are determined separately (2). In paper Image Processing Technique for Suppression Ribs in chest radiograph by means of MTANN (2006) By Kenji Suzuki, Hiroyuki Abe, Heber McMahon, and Kunio Doi, *Senior Member, IEEE*. Search that, When lung nodules overlap with ribs or clavicles in chest radiographs, it can be difficult for radiologists as well as computer-aided diagnostic (CAD) schemes to detect these nodules. In this paper, we developed an image-processing technique for suppressing the contrast of ribs and clavicles in chest radiographs by means of a multi resolution massive training artificial neural network (MTANN). An MTANN is a highly nonlinear filter that can be trained by use of input chest radiographs and the corresponding "teaching" images. We employed "bone" images obtained by use of a dual-energy subtraction technique as the teaching images. For effective suppression of ribs having various spatial frequencies, we developed a multi resolution MTANN consisting of multi Resolution decomposition/composition techniques and three MTANNs for three different-resolution images (3). In the paper of a fully Automated method for lung nodule detection from Postero-anterior Chest Radiography (2007) By Paola Campadelli, Elena Casiraghi and Diana Artioli, *Member, IEEE*. Search that, in the past decades, a great deal of research work has been devoted to the development of systems that could improve radiologist's accuracy in detecting lung nodules. Despite the great efforts, the problem is still open. In this paper, we present a fully automated system processing digital postero-anterior (PA) chest radiographs that starts by producing an accurate segmentation of the lung field area. The segmented lung area includes even those parts of the lungs hidden behind the heart, the spine, and the diaphragm, which are usually excluded from the methods presented in the literature. The segmented area is processed with a simple multi scale method that enhances the visibility of the nodules, and an extraction scheme is then applied to select potential nodules. Different learning experiments were performed on two different data sets, created by means of feature selection, and employing Gaussian and polynomial SVMs trained with different parameters; the results are reported and compared(4). In Comparison of the Multi-Layer Perceptron and the Nearest Neighbour Classifier for Handwritten Numeral Recognition **By**: K. ROY+, C. CHAUDHURI, M. KUNDU, M. NASIPURI Multi-Layer Perceptron: Multi-Layer Perceptron (MLPs) constitute an important class of feed-forward Artificial Neural Networks (ANNs), developed to replicate *learning* and *generalization* abilities of humans with an attempt to model the functions of biological neural networks. They have many potential applications in the areas of Artificial Intelligence (AI) and Pattern Recognition (PR). This work presents a cost effective design of a 2 layer MLP for recognition of handwritten numerals. It also presents a comparative assessment of the performances of the MLP and a NN classifier for the same. The feature set selected for representing a handwritten numeral includes all the features selected by Weidman *et al.* excepting two, one for the ratio of the energies at the top to the bottom of the loop, if the numeral image contains any, and the other for the number of loops in the numeral image. Design of the NN Classifier: The NN classifier requires a set of sample patterns of known classes, called *reference* set. To classify an unknown pattern X, it first finds the reference pattern closest to X in the feature space and then assigns the class label of the same to X. The k-Nearest Neighbour Classifier: To classify an unknown pattern X, the k-nearest neighbour (k-NN) classifier first finds k of the reference patterns closest to X and then classifies X to a class which is found to have the maximum number of samples among the k reference patterns(5). In the paper of Improving the precision of classification tress: By: WEI-YIN LOH explains that, Besides serving as prediction models, classification trees are useful for finding important predictor variables and identifying interesting subgroups in the data. These functions can be compromised by weak split selection algorithms that have variable selection biases or that fail to search beyond local main effects at each node of the tree. The resulting models may include many irrelevant variables or select too few of the important ones. Either eventuality can lead to erroneous conclusions. Four techniques to improve the precision of the models are proposed and their effectiveness compared with that of other algorithms, including tree ensembles, on real and simulated data sets. Many of the early classification tree algorithms, including THAID, CART and C4.5, search exhaustively for a split of a node by minimizing a measure of node heterogeneity. As a result, if all other things are equal, variables that take more values have a greater chance to be chosen. This selection bias can produce overly large or overly small tree structures that obscure the importance of the variables. Improving the prediction accuracy of a tree and the precision of its splits is a balancing act. On the one hand, we must refrain from searching too greedily for splits, as the resulting selection bias can cause irrelevant variables to be selected. On the other hand, we should search hard enough to avoid overlooking good splits hidden behind interactions or linear relationships. We solve this problem by using three groups of significance tests, with increasing complexity of effects and decreasing order of priority. The first group of tests for main effects is always carried out. The second group, which tests for interactions, is performed only if no main effect

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

test is significant. The third group, which tests for linear structure, is performed only if no test in the first two groups is significant. A Bonferronicorrection controls the significance level of each group. In addition, if an interaction test is significant, the split is found by a two-level search on the pair of interacting variables (6). An introduction to genetic algorithm By: MELANIE MITCHLLE search that, Genetic algorithms can be a powerful tool for solving problems and for simulating natural systems in a wide variety of scientific fields. In examining the accomplishments of these algorithms, we have also seen that many unanswered questions remain. It is now time to summarize what the field of genetic algorithms has achieved, and what are the most interesting and important directions for future research. From the case studies of projects in problem−solving, scientific modelling, and theory we can draw the following conclusions: GAs is promising methods for solving difficult technological problems, and for machine learning. More generally, GAs is part of a new movement in computer science that is exploring biologically inspired approaches to computation. Advocates of this movement believe that in order to create the kinds of computing systems we need—systems that are adaptable, massively parallel, able to deal with complexity, able to learn, and even creative—we should copy natural systems with these qualities. Natural evolution is a particularly appealing source of inspiration. Genetic algorithms are also promising approaches for modelling the natural systems that inspired their design. Most models using GAs are meant to be "Gedanken experiments" or "idea models" (Rough garden et al. 1996) rather than precise simulations attempting to match real−world data. The purposes of these idea models are to make ideas precise and to test their plausibility by implementing them as computer programs (e.g., Hinton and Nowlan's model of the Baldwin effect), to understand and predict general tendencies of natural systems (e.g., Echo), and to see how these tendencies are affected by changes in details of the model (7).

III. METHODOLOGY

In the lung cancer detection process, there are two phases first to detect the lung nodule which will be differentiated from other things that get stuck into the lung. Next phase is to classify the Lung cancer using various classifier techniques such as decision tree algorithm, K- nearest neighbour. Fig.1 shows the block diagram of lung cancer detection. These phases and techniques are briefly described below.

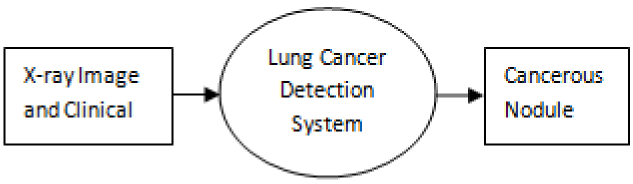

Fig.1 Block Diagram

IV. PRE-PROCESSING

In pre-processing, Median filter and morphological smoothening filter techniques are applied to remove the noise from the images and enhance the image. The median filter will remove the noises and produces the enhanced image. The growth and decrease process is done for morphologically smoothening of the images. Decrease involves the alteration (removal) of pixels at the edges of regions, i.e., exchanging binary 1 value to 0, while growth is the reverse process with regions growing out from their boundaries. The decrease followed by growth is known as Opening operation which suppresses the bright details smaller where growth followed by decrease is closing operation which suppresses the dark details are computed. Here in this work the darker details are suppressed where the growth is performed followed by decrease. The pre-processed image is obtained by applying the median filter and morphological smoothening.

V. FEATURES EXTRACTION

The feature extraction refers to creating a set of new appearance by combination of the existing appearance [12]. Each sub-image is taken from top left corner of the original image and the texture appearance and pixel coefficient values are

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

extracted [12]. The analysis and characterization of textures present in the x-ray images can be done by using feature extraction method. Texture analysis is a computable method that can be used to quantify and indentify structural abnormalities [12].

VI. FEATURES SELECTION

Feature Selection (FS) algorithm is a process of choosing a reduced related features that improves classification by identifying for the feature subset, from the database of original features according to a given classification accuracy[12]. It removes unrelated or unnecessary features dealing with the dimensionality cruse, thus leading to reduce the computational and memory cost [12]. Genetic algorithm was used to identify the optimal solutions. The features were selected and their ability is identified with various classifiers [12].

VII. NEAREST NEIGHBOUR RULE

The k-nearest neighbour rule is among the simplest statistical learning tools in density estimation, classification, and regression. During runtime, where we denote runtime values with subscript, the adaptation of -NN determines the high-resolution image patch from a single low-resolution image patch [8]. For mathematical reasons, it is easier to represent image patches and as vectors instead of square patches. Therefore, in subsequent derivations and are both vectors that have been rearranged from image blocks into a single column. The typical -NN estimate for regression at the test point is given as where depending on whether or not is among the nearest neighbours. In fact, there are several common weighting schemes, ranging from posterior probability like expressions to iteratively determine convex solutions, all functions of distances or weights that can used to minimize some criterion as in. An extensive study on error rates for regression-based -NN estimates was analysed in, where it was conceded that for jointly normal, under the squared-error loss case, the unconditional, large sample risk as of the -NN estimate satisfies Here, is the Bayes risk (minimum expected loss), and parameters and are variance parameters in probability distribution functions (PDF's) and. The trade-off in would like to keep term to limit erroneous reconstruction. While simultaneously favouring a larger to keep the final term in small. This type of trade off is common in -NN problems, and provides much need for cross validation as will be seen even for adaptable values. Of course, it is not always the case that is large and for these problems, the risk of nearest neighbour (where) is usually smaller than the risk of -NN, but overall, when is large, -NN is invariably the rule of choice. Actually, when is large and the dimensionality of, is small, -NN is almost always preferable or at least competitive among other estimation techniques such as SVR, where SVR performs well when is small and is large. This is fairly intuitive because we would like to blanket the entire domain with samples, possible only with large, and easier if were sufficiently small. It is this last scenario that tends to be the case with the potential of today's computing power. As memory for computing tasks increases and computing time for higher complexity routines decreases, motivation for various -NN problems is justified due to the capability of supporting large values [8].

VIII. CONCLUSION

The detection of lung cancer using X-Ray is a complex task that consists of several subtasks. In identifying the boundary of lung and the boundary of nodule is important task because it gives the nodule size which is one of the main tasks. Nodule size is less than 1 cm. After identifying the boundary or it determine shape of these tumors even in the neighbourhood of sharply defined rib contours, it identify that present nodule is cancerous or non cancerous. This is useful mainly for the doctors so as to help them in their predictions in detection of lung cancer and also to those industries that are at higher risk of cancer like textiles, coal mines and fiber industries.

REFERENCES

[1] Sheng Chenand Kenji Suzuki, "Computerized Detection of Lung Nodules by Mean of "Virtual Dual-Energy" Radiography" IEEE Transactions on Biomedical Engineering, vol. 60, No. 2, Feb 2013.

[2] M. L. Giger, K. Doi, and H. McMahon, "Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peri Mar.–Apr. 1988pheral lung fields," *Med. Phys.*, vol. 15, pp. 158–166.

[3] Kenji Suzuki, Hiroyuki Abe, Heber McMahon, and KunioDoi, *Senior Member, IEEE*. "Image Processing Technique for Suppression Ribs in chest radiograph by means of MTANN" (2006).

[4] Paola Campadelli, Elena Casiraghi and Diana Artioli, *Member, IEEE* "A fully automated method for lung detection from Postero-anterior Chest Radiography "(2007).
[5] K. ROY+, C. CHAUDHURI, M. KUNDU, M. NASIPURI "Comparison of the Multi-Layer Perceptron and the Nearest Neighbour Classifier for Handwritten Numeral Recognition**".**
[6] WEI-YIN LOH Annal Appl Stat 2009; 3(4):1710–37. "Improving the precision of classification tress"
[7] MELANIE MITCHLLE "An introduction to genetic algorithm".
[8] Karl S. Ni, *Member, IEEE*, and Truong Q. Nguyen, *Fellow, IEEE* "An Adaptable k-Nearest Neighbours Algorithm for MMSE Image Interpolation" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 18, NO. 9, SEPTEMBER 2009.
[9] M. L. Giger, K. Doi, and H. McMahon, "Image feature analysis and computer-aided diagnosis in digital radiography automated detection of nodules in peripheral lung fields" Med. Phys, vol. 15, pp. 158–166.
[10] A. M. Schilham, B. van Ginneken, and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database," *Med. Image Anal.*, vol. 10, pp. 247–258, Apr. 2006.
[11] M. Loog, B. van Ginneken, and A. M. Schilham, "Filter learning: Application to suppression of bony structures from chest radiographs," *Med Image Anal.*, vol. 10, pp. 826–840, Dec. 2006.
[12] C. Bhuvaneswari , P. Aruna , D. Loganathan www.elsevier.com/locate/eij "A new fusion model for classification of the lung diseases using genetic algorithm"
[13] Goldberg DE, Hassan MR, Hossain MM, Bailey J, Ramamohanarao K."Genetic algorithms in search, optimization and machine learning" Newyork: Addison Wesley; 1989.
[14] Hassan MR, Hossain MM, Bailey J, Ramamohanarao K. in databases LNAI. Berlin Heidelberg: Springer; 2008. p 489–504. "Improving k-nearest neighbour classification with distance functions based on receiver operating Characteristics. In: Machine learning and knowledge discovery"
[15] Gil-Pita, Yao R. Evolving edited k-neighbor classifier. Int J Neural Syst Dec 2008:18(6):459-67
[16] Lin D, Yan C. "Lung nodules identification rules extraction with neural fuzzy network. Neural Inform Process" IEEE 2002; 4:2049–53.
[17] Graham R, Perriss R, Scarsbrook A. Dicom demystified: "a review of digital file formats and their use in Radiological practice" ClinRadiol 2005; 60:1133–40.