

Predicting Survival Time of Lung Cancer Patient using Weibull Method

Shanid Malayil¹, Shahad P²

P.G. Student, Department of Computer Engineering, MEA Engineering College, Malappuram, Kerala, India¹

Assistant Professor, Department of Computer Engineering, MEA Engineering College, Malappuram, Kerala, India²

ABSTRACT: Lung cancer seems to most dangerous and most common cause of death among people throughout the world. Early detection of lung cancer can increase the chance of survival. Image processing mechanisms are widely used in several medical areas for early detection and in deciding treatment stages. In this paper the technique presented is for detection of lung cancer and prediction of survival time of lung cancer patient. CT scan image of lungs is given as input the image is enhanced, and the lung nodule is separated by segmentation. In later step the features are extracted from nodule and classify the image into the particular stage of cancer, thus predicting the survival time of patient.

KEYWORDS: CT Image, Weibull Segmentation, Support Vector Machine, Benign, Malignant, Irregularity Index.

I. INTRODUCTION

In medical field, neoplasia term is used for cancer. The term neoplasia means new growth, the new growth produced is called neoplasm or tumor (except new growth of tissues and cells those exist in the process of embryogenesis, regeneration, hormonal stimulations etc.). Satisfactory definition of neoplasm or tumor is a mass of tissues formed as a result of abnormal, excessive, uncoordinated, autonomous and purposeless proliferation of cells. The branch of science dealing with the study of neoplasms or tumors is called Oncology. Neoplasm may be benign when they are slow growing and localized without causing much difficulty to the host or malignant when they proliferate rapidly, spread through out the body and may eventually cause death of the host the abnormal cells then multiply and grow into a tumor. The mortality rate of lung cancer is the highest among all other types of cancer with the smallest survival rate after the diagnosis, with a gradual increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time with higher chance of survival if detected in the early stages. Cancer cells can be carried away from the lungs in blood, or lymph that surrounds lung tissue.

This system could predict the survival time of the lung cancer patient when the CT scan image of the lungs is given as input.

The work in this paper is divided in four stages. 1) Image Enhancement 2) Segmentation 3) Feature Extraction 4) Classification and survival time calculation. The image is enhanced using tools in MATLAB and Gabor filter, after that lung nodule is segmented out, then the features are extracted, finally the classification in to stages and predicting survival time.

Paper is organized as follows. Section II describes image segmentation using weibull distribution method, CT scan image analysis and separation of lung nodule The flow diagram represents the step of the algorithm. After detection of lung nodule, the features of marked region is extracted in Section III. Section IV presents the classification into different stage of cancer and thus predicting survival time based on pre computed dataset, Section VI experimental results showing results of images tested. Finally, Section V presents conclusion.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

II. RELATED WORKS

In computer vision, segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, texture. All image processing operations generally aim at a better recognition of objects of interest, i.e., at suitable local features that can be distinguished from other objects and from the background. The next step is to check each individual pixel to see whether it belongs to an object of interest or not. This operation is called segmentation and produces a binary image. A pixel has the value one if it belongs to the object otherwise it is zero. After segmentation, it is known that which pixel belongs to which object.

The weibull distribution method use weibull distribution which is far more accurate in segmenting the lung nodule region, After applying the segmentation steps, we get a filtered image that contains the marked region.

A complex SAR image pixel represented by $C_{m,n}$. If the SAR amplitude image pixel, $|C_{m,n}|$ is a Weibull distributed random variable with form parameter $\gamma_{m,n}$ and scale parameter $\beta_{m,n}$ its probability density function is given by:

$$P(|C_{m,n}|) = \frac{\gamma_{m,n}}{\beta_{m,n}} \left(\frac{|C_{m,n}|}{\beta_{m,n}} \right)^{\gamma_{m,n}-1} \exp \left[- \left(\frac{|C_{m,n}|}{\beta_{m,n}} \right)^{\gamma_{m,n}} \right]$$

For $\gamma_{m,n} = 2$ the Weibull distribution becomes a Rayleigh distribution and for $\gamma_{m,n} = 1$ it becomes the exponential distribution. It can be shown that the intensity image $|C_{m,n}|^2$ is also Weibull-distributed.

III. FEATURE EXTRACTION

The Image features Extraction stage is very important in our working in image processing techniques which using algorithms and techniques to detect and isolate various desired lung nodule portion of the lung image. After the segmentation is performed on lung region, the features can be obtained from it and the diagnosis rule can be designed to exactly detect the cancer nodules in the lungs. This diagnosis rules can eliminate the false detection of cancer nodules resulted in segmentation and provides better diagnosis. In the literature we found among the features used in the diagnostic indicators. Feature extraction provides certain parameters, on the basis of which computer system takes decision.

This information is very helpful in detecting lung nodule as malignant or non-malignant, the stage of cancer the extracted features can be used as diagnostic indicators. The features extracted from the segmented lung nodule is compared with the standard values prescribed by a physician or by values from training set, below are some features and the equations to compute these values:

1. Area

Area of the segmented tumor is computed by counting the number of pixels which have the value 1 in the image array. The area (A) in the object is the just the count of the ones in the image array. For computing area, binary image is used $A = n [1]$ where, $n [1]$ represents the count of number of the patterns within the parenthesis.

2. Perimeter

Perimeter property is calculating the distance between each adjoining pair of pixels around the border of the region. If the image contains discontinuous regions, then it returns unexpected results. Perimeter is calculated by counting the pixels contained only in the boundary

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

3. Irregularity Index

Lung cancer is characterized partially by the irregularity in its tumor border. For this analysis, the irregularities in the tumor are computed by an index:

$$I = \frac{4\pi}{(\text{Perimeter})^2}$$

The metric value or roundness or circularity index or irregularity index (I) is equal to 1 only for circle and it is < 1 for any other shape. Here it has been assumed that, more circularity of the object, the probability of the object being nodule is high

4. Equivalent Diameter

Scalar that specifies the diameter of a circle with the same area as the region . Computed by formula

$$E_{Diameter} = \sqrt{\frac{4 * Area}{\pi}}$$

5. Convex Area :

The convex hull or convex envelope of a set X of points in the Euclidean plane is the smallest convex set that contains X . For instance, when X is a bounded subset of the plane, the convex hull may be visualized as the shape formed by a rubber band stretched around X . Convex area is the number of pixel in convex image. Convex image is the size of the bounding box of the region. Bounding box is convex hull.

6. Solidity

It gives the proportion of the pixels in the convex hull that are also in the region.

$$Solidity = \frac{Area}{Convex Area}$$

7. Gray Level Co occurrence Matrix Properties

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM and then extracting statistical measures from this matrix.

The number of gray levels in the image determines the size of the GLCM. The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset. Offsets define pixel relationships of varying direction and distance. However, a single GLCM offset might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. For this reason, multiple GLCMs offset for a single input image.

8. Contrast:

It measures the local variations in the gray-level co-occurrence matrix. It returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. Contrast is 0 for a constant image. Formula for contrast is given by :

$$\sum_{i,j} |i - j|^2 P(i, j)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

9. Correlation :

It measures the joint probability occurrence of the specified pixel pairs. Returns a measure of how correlated a pixel is to its neighbor over the whole image. Range is from 1 to -1.

Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is null for a constant image. It is given by formula :

$$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$$

10. Energy

Provide the sum of squared elements in the GLCM. It is also known as uniformity or the angular second moment.

Returns the sum of squared elements in the GLCM. Its range is from 0 to 1. Energy is 1 for a constant image.

$$\sum_{i,j} P(i, j)^2$$

11. Statistical Properties of an Image

It is a single measure of some attribute of a sample. It is calculated by applying a function (statistical algorithm) to the values of the items of the sample, which are known together as a set of data.

12. Entropy

Measures the randomness of a gray-level distribution. The Entropy is expected to be high if the gray levels are distributed randomly throughout the image. It is given by

$$E = - \sum_i^m \sum_j^n P[i, j] \log P[i, j]$$

13. Mean

It provides the mean of the gray levels in the image. The Mean is expected to be large if the sum of the gray levels of the image is high. It is given by

$$\mu = \frac{1}{N \times M} \sum_{i=0}^M \sum_{j=0}^N P(i, j)$$

14. Variance

Variance tells us, how spread out the distribution of gray levels is. The Variance is expected to be large if the gray levels of the image are spread out greatly. It is given by

$$\sigma^2 = \frac{1}{N \times M} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (P(i, j) - \mu)^2$$

15. Standard Deviation

Standard Deviation shows much ,variation or dispersion exists from the average.

$$Std = \sqrt{\sigma^2}$$

After the segmentation is performed on lung region, the features can be obtained from it and the diagnosis rule can be designed to detect nodules in the lung. The entire feature which are calculated from the image, convey some information regarding lung nodule.

IV. CLASSIFICATION

The classifiers were selected to test a range of techniques, and to determine those that provide the best predictive accuracy. Support vector machines classifier used in this system, they are based on statistical learning theory developed by Cortes and Vapnik and have been shown by Kramer, among others, to obtain high accuracy on a diverse range of application domains such as the letter, page, pendigit, satimage, and waveform data sets. SVMs map the input data to a higher dimensional feature space and construct a hyperplane to maximize the margin between classes. A linear decision surface is constructed in this feature space. The hyperplane construction can be reduced to a quadratic optimization problem; subsets of training patterns that lie on the margin were termed support vectors by Cortes and Vapnik. The formulation we used allows for errors to be on the wrong side of the decision border during training. A cost parameter C is multiplied by the distance the errant example is from the decision boundary. The larger the value of C the larger the penalty applied in the learning process. Parameter tuning of the cost parameter was conducted on training data using a grid search after feature selection. Different kernels, such as a linear kernel, radial basis function kernel, and sigmoid kernel, can be chosen for SVMs. We used the linear kernel with a degree of 3. The svm type was set to C-SVC (classification) and the default termination criteria were used. Dehmeshki used support vector machines effectively on CT-scan image data of the lungs in a Computer-Assisted Detection (CAD) system for automated pulmonary nodule detection in thoracic CT-scan images. We used the support vector machine libSVM by Chang and Lin.

V. EXPERIMENTAL RESULTS

Figures shows the results of lung cancer detected from an image and marked by using segmented marker algorithm. Figs. 2, 3, 4 (a) shows the input image. (b) Shows the image obtained by applying first level of segmentation. (c) shows second level of segmentation (d) Pattern matching to detect lung nodule (e) Lung Nodule marked (f) Shows the final result with survival time of patient

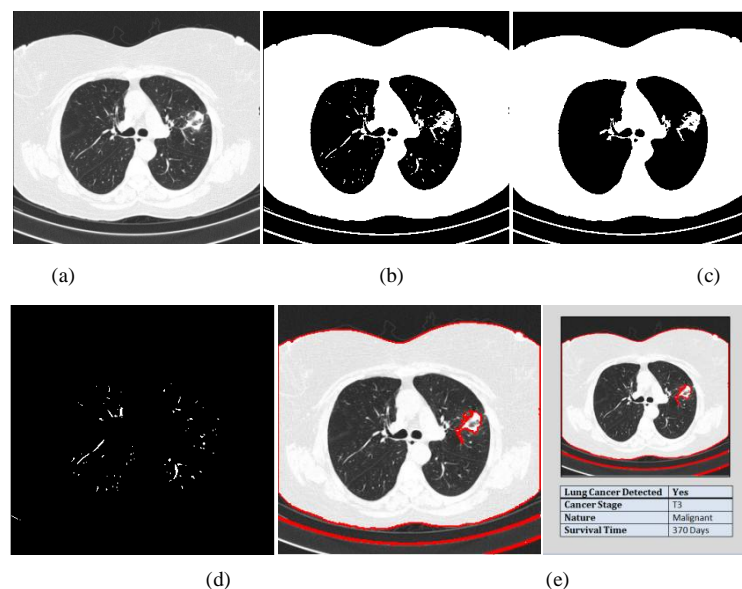


Fig. 1. Lung Cancer Detection (a) Original image (b) Shows the image obtained by applying first level of segmentation. (c) shows second level of segmentation (d) Pattern matching to detect lung nodule (e) Lung Nodule marked (f) Shows the final result with survival time of patient

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

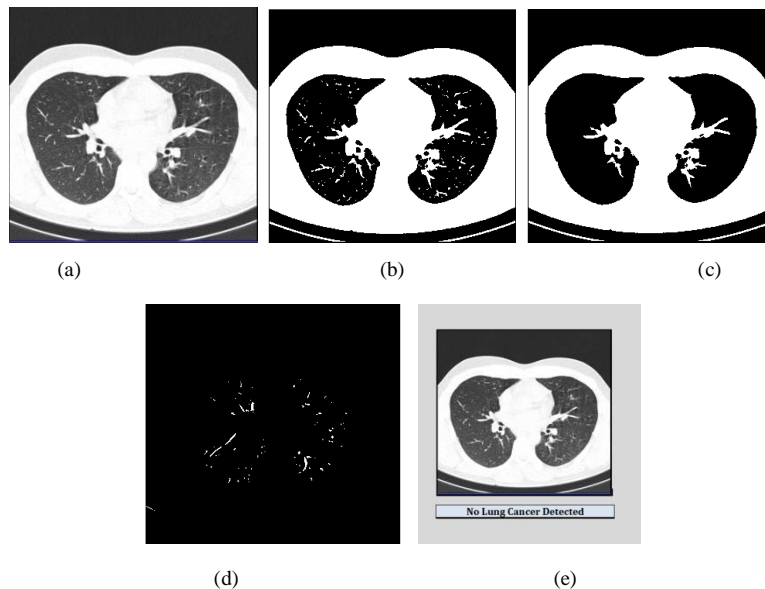


Fig. 3 Lung Cancer Detection (a) Original image (b) Shows the image obtained by applying first level of segmentation.(c)shows second level of segmentation (d) Pattern matching to detect lung nodule (e) Shows the final result with no Lung cancer detected

VII. CONCLUSION

We have implemented an automatic Lung cancer detection and survival time prediction technique from a CT scan lung image. Our algorithm successfully detects the lung cancer nodule from the image using weibull distribution which consists of complex regions. The features extracted from the lung nodule using parametric equations are computed primarily and classified as different the stage of lung cancer, thus calculating and predicting the survival time of patient. We have applied our algorithms on many CT scan lung images and found that it successfully detect the lung cancer nodule and predict survival 81% accurate .

REFERENCES

- [1] V. Kumar, "Radiomics: The process and the challenges," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1234_1248, 2012.
- [2] B. Ganeshan, S. Abaleke, R. C. D. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imag.*, vol. 10, no. 1, pp. 137_143, Jul. 2010.
- [3] R. Samala, W. Moreno, Y. You, and W. Qian, "A novel approach to nodule feature optimization on thin section thoracic CT," *Acad. Radiol.*, vol. 16, no. 4, pp. 418_427, 2009.
- [4] T.W.Way, "Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours," *Med. Phys.*, vol. 33, no. 7, pp. 2323_2337, 2006.
- [5] O. S. Al-Kadi and D. Watson, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 7, pp. 1822_1830, Jul. 2008.
- [6] Y. Zhu, Y. Tan, Y. Hua, M. Wang, G. Zhang, and J. Zhang, "Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography," *J. Digit. Imag.*, vol. 23, no. 1, pp. 51_65, 2010.
- [7] M. C. Lee *et al.*, "Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction," *Artif. Intell. Med.*, vol. 50, no. 1, pp. 43_53, 2010.
- [8] Huang Q., Gao W., Cai W., *Thresholding technique with adaptive window selection for uneven lighting image*, Pattern Recognition Letters, Elsevier, 2004, 26, p. 801-808.
- [9] Anita chaudhary, Sonit Sukhraj Singh, "Lung Cancer Detection on CT Images Using Image Processing ", International transaction on Computing Sciences, VOL 4, 2012
- [10] B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey", IEEE, transactions on medical imaging, vol. 20, NO.12,DEC-2001
- [11] Y. Gu, "Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach," *Pattern Recognit.*, vol. 46, no. 3, pp. 692_702, Mar. 2013.