SOFTWARE DEVELOPMENT TO DETECT LUNG NODULES IN COMPUTED TOMOGRAPHY SCAN IMAGE USING SUPPORT VECTOR MACHINE

Kustanto[1], Sri Widodo[2], Sri Tomo[3]

[1,3]Department of Informatic Systems, STMIK Sinar Nusantara Surakarta, Central Java, Indonesia
[2]Medical Records and Health Information Academy APIKES CitraMedika Surakarta,   Central Java, Indonesia

---***---

Abstract- *Lung nodules is a circle or oval that is often found on chest radiograph (X-ray) or computed tomography (CT) scan. Lung nodules usually do not cause a symptom, usually found incidentally on chest X-ray or CT. Diagnosis process performed by observing CT Scan. The further process by taking a sample of tissue in the lungs and perform laboratory tests. This procedure is not effective. The aim of this study was develop software to detect lumps on pulmonary CT scan images are nodules or artery. This study consists of five steps. The first step is to build user interface. The second is to develop software for lung segmentation using Active Shape Model. The third is to develop software for candidate pulmonary nodules segmentation using mathematical morphology. The fourth is detection of pulmonary nodules using Support Vector Machine (SVM). The fifth is a 3-D visualization of pulmonary nodules using volume rendering. Results of detection of pulmonary nodules is accuracy 75%, sensitivity 77%, specificity 22%.*

Key Words:ASM, Mathematical Morphology, SVM.

1. INTRODUCTION

Currently, cancer is still a serious problem in the world. World Cancer 2014 International Agency for Research on Cancer (IARC) at the World Health Organization that in 2012 there were approximately 14 million new cases of cancer occur. Lung cancer is the most common cause of death with an estimated 8.2 million deaths. According to IARC, about 70 percent of cancer deaths occur in Africa, Asia, Central and South America [1] [2]. Lung cancer is a disease caused by lung cell division that is uncontrolled. If left untreated, the growth of lung cells to grow and spread beyond the lungs [3].
Lung nodules is a circle or oval that is often found on chest radiograph (X-ray) or computed tomography (CT) scan [4]. At one of the 500 reading chest X-ray can be found a nodule. Benign nodules or benign when found in people who are still aged under 40 years. Causes Tumor and Lung Nodules are still not very clear, but in general there is a problem with inflammation due to infections histoplasmosis, coccidioidomycosis, cryptococcosis, or

aspergillosis. The possibility of inflammation is not caused by an infection, but a result of rheumatoid arthritis, Wegener's granulomatosisSarcoidosis. Among other causes are congenital anomalies (birth defects) as cistik lung and pulmonary malformations [5].

If there is a suspicion of lung cancer, it will be diagnosed in more advanced with chest X-ray and CT Scan. In the X-ray will reveal a picture of a tumor or a liquid, while the CT scan, you will see a tumor, fluid or lymph node enlargement. Furthermore, to ensure nodules or not to be held laboratory tests of samples of sputum and phlegm. It is also taken of lung tissue or gland, for laboratory test. Diagnostic process as above is not effective. Therefore we need a software that can automatically detect lung nodules. Several studies relating to the detection of nodules automatically among others [6] - [11].

The aim of this study is to develop a pulmonary nodule detection software. This study consists of five steps. The first step is to build user interface. The second is to develop software for segmentation of lung using Active Shape Model (ASM). The third is, to develop software for segmentation candidate pulmonary nodules using mathematical morphology. The fourth is detection of pulmonary nodules using Support Vector Machine (SVM). The fifth is a 3-D visualization of pulmonary nodules. The end result is a software that can detect an existing image in CT lung nodule or artery.

2. MATERIAL

The data used in this study is the CT images of 30 patients. Image sliced in axial CT scan. Image size of 505x427 pixels, and a thickness of 0.5 - 10 mm. Ct Scan image were taken from Surakarta Moewardi hospital.
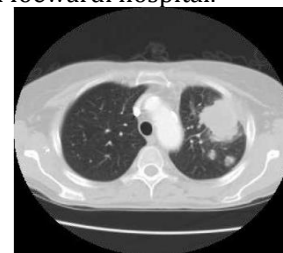

Fig -1: Lung ct scan image with axial sices

3. METHODS

Software development for the detection of pulmonary nodules can be explained by using Data Flow Diagrams (DFD) as in Figure 2. Data Flow Diagrams (DFD) level 1 illustrates the flow of information and transformation which is applied when the data is moved from input to output. At the DFD level 1 detection system of pulmonary nodules present, there is only one entity that is friendly, and there are five processes, namely the process of making the user interface, the process of segmenting the lung, pulmonary nodule segmentation process, the process of detection of pulmonary nodules and pulmonary nodule visualization process.
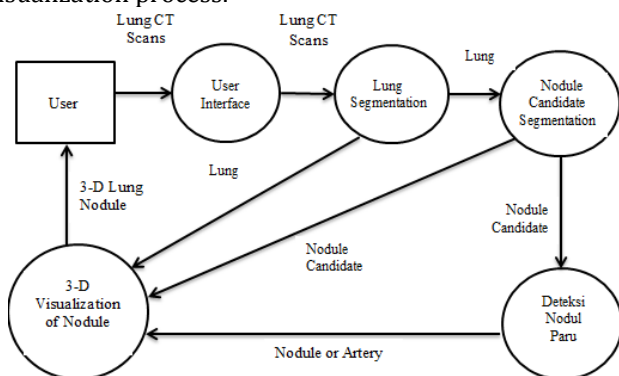

Fig -2:DFD level 1

Step in the development of software for detection of pulmonary nodules can be explained as follows:

3.1. Development of User Interface

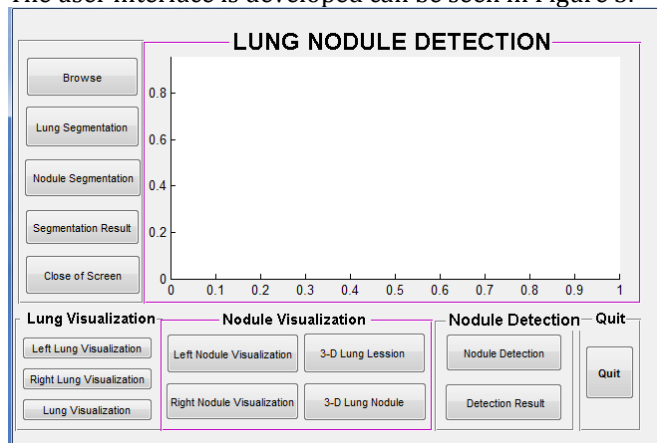The user interface is developed can be seen in Figure 3.

Fig -3: User interface software detection of lung nodules

3.2. Lung Segmentation With ASM

Lung segmentation is done to separate the pulmonary organs with the surrounding tissue. The aim is to simplify the process further. Method used is an Active Shape Model

[12]. In this paper we do not discuss about the segmentation of lung using ASM. This is because ASM has been discussed in another paper [13]. Results of lung segmentation can be seen in Figure 4.
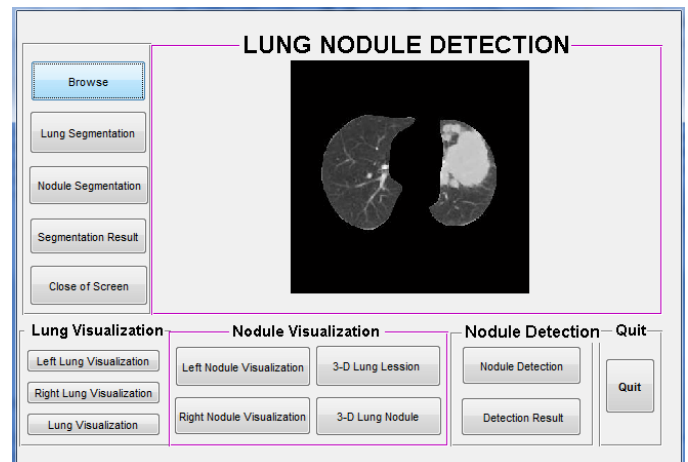

Fig -4: Results of lung segmentation

3.3. Nodule Candidates Segmentation

Method used for segmentation of nodule candidate is mathematical morphology [14] - [16]. Step-by-step process of segmentation candidate pulmonary nodules using mathematical morphology can be explained as follows:

a. Subtraction among lung image with negative lung image

Subtraction is done with the following steps: Image of lung segmentation results with ASM done complement process, which would result in a negative image of lungs. Process of complement in binary image, a value of zero to one and vice versa. Complement of intensity or RGB image, obtained from Subtraction of maximum pixel value (255) minus value of image. Output of complement of an image is if the dark image areas become light, and vice versa. Subtraction of lung image with a negative lung image. Results subtraction will be carried out Mathematical morphological processes.

b. Edge Detection
Edge detection of objects is an operation to detect significant changes in gray level of an image. Changes in the level of intensity is measured by a gradient image. For example, an image f (x, y) is a two-dimensional function, the gradient vector of x and y, respectively, are a first derivative with respect to x and y that can be written in the form of an equation

$$G = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \tag{1}$$

Gradient operator calculates changes in gray level intensity and direction changes. The changes are

calculated by the difference value of the pixel neighbors. In the two-dimensional image, gradient approximated using equation 2.

$$G = \begin{bmatrix} Gx \\ Gy \end{bmatrix} = \begin{bmatrix} f(x+1,y)-(f(x,y) \\ f(x,y+1)-f(x,y) \end{bmatrix} \qquad (2)$$

In the above equation can be written in matrix form

$$G_x = [-1 \ 1] \qquad (3)$$

and

$$G_y = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad (4)$$

While the magnitude of gradient can be calculated using equation models, one of which is shown in equation 5.

$$G[f(x,y)] = \sqrt{G_x^2 + G_y^2} \qquad (5)$$

Edge detection used in this study are: edge detection method prewit convolution. Prewit used size is 3x3 with horizontal elements are in the middle to Gx is equal to 0 and the vertical elements are in the middle to Gy are also equal to 0, as shown in equation 6 and 7 below.

$$G_x = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \qquad (6)$$

and

$$G_y = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad (7)$$

c. Threshold

Thresholding is a process of separation of the pixels based on its degree of gray. Pixels that have a gray level less than the prescribed limit will be given a value of 0, while gray pixels that have a greater degree than the limit will be changed to value 1.

d. Dilation

Dilatation function is added for each pixel on the outskirts of a binary object that is an area that has a value of 1. Where dilatation added 8 pixels interconnected to the surrounding objects. And dilation is a process of combining the dots background (0) become part of the object (1). The use of dilation is put the pivot point S at point A. Give the number 1 for all points (x, y) that is exposed to or affected by the structure of S in that position. Dilation equation shown in Equation 8 below:

D(A,S)=A+S   (8)

e. Fill Images Area

To fill image area used algorithms based on morphological reconstruction. Definition of the image area is an area of dark pixels surrounded by lighter pixels. After the image location is determined, the next operation is to fill location (image area) using four neighboring connection background for the image of input 2-D and 6-connected background neighbors for 3-D input. Output of this process is an area surrounded by brightly colored pixels will have a value of one.

f. Erosion

Erosion function is to eliminate 8 pixels of the binary object associated with the periphery of the object. Erosion is the elimination of the object point (1) become part of the background (0). The use of erosion is place the pivot point S at point A is. If any part of S which are beyond the pivot point A deleted or used as background. Erosion is shown with equation 9 below.

E(A,S)=AxS   (9)

g. Multiplication

Image multiplication performed to obtain the final image of the candidate nodules. Image multiplication can be done by multiplying each pixel with a certain value. Mathematically modeled using equation 10 below.

$$\forall f(x,y); Nf(x,y) = f(x,y) * Th \qquad (10)$$
$$\forall Nf(x,y); if \ Nf(x,y) > 255 \ Then \ Nf(x,y) = 255 \qquad (11)$$

with
Th ≥ 1

F(x,y) is the original image, and Nf (x, y) is image of multiplication of intensity of the original image with a value of Th. In this study, process of multiplication is a multiplication of image of the mask candidate nodules result of erosion process results, with image results in a reduction of complement.Results of candidate nodule segmentation using mathematical morphology can be seen in Figure 5 below:
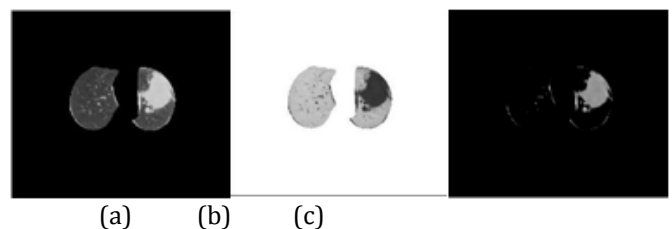

(a)   (b)   (c)

Fig -5: Subtraction process of image
(a) Lung Image, (b) Negative Image, (c) Subtraction Results Between a and b image.

(a)   (b)   (c)   (d)

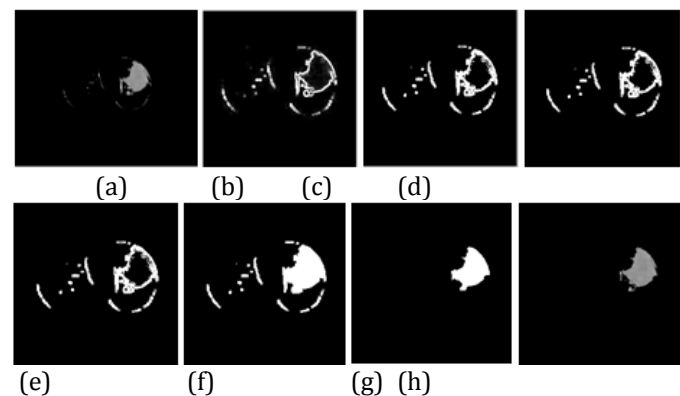(e)   (f)   (g)   (h)

Fig -6: Morphology Process Of Nodules Candidate
(a) image of subtraction results, (b) results of edge detection with convolution prewitt, (c) image of threshold

with value 60, (d) image of erosion process, (e) dilation process results with value: 1.5, (f) image of masking process results, (g) contour of segmentation results, (h) candidate nodules image
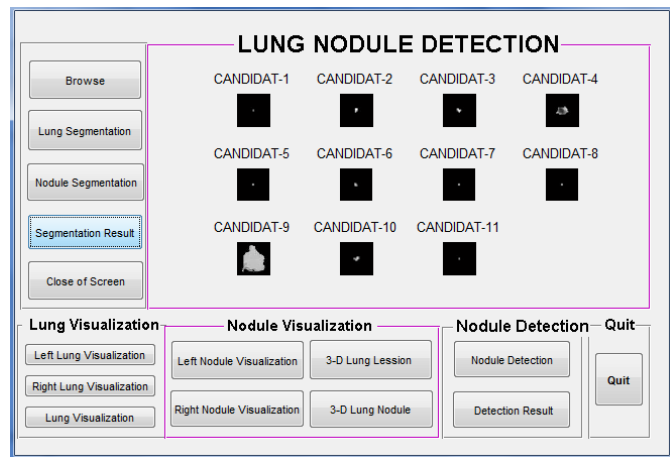

Fig -7: Segmentation results of nodules candidate

3.4. Detection of Lung Nodules Using SVM

Steps in detection of nodules using Support Vector Machine (SVM) can be described in Figure 8.
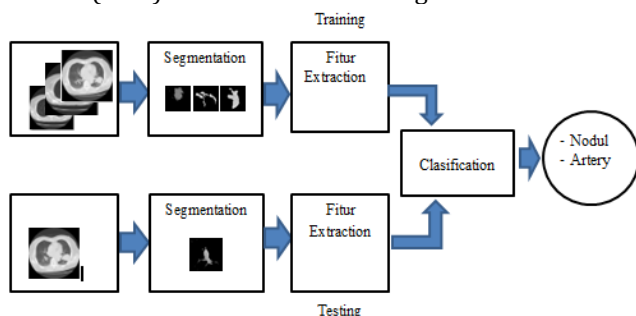

Fig -8: Proposed method for detection of lung nodules

The first step in nodule detection process is feature extraction. Features used is analysis of texture. Texture analysis method used in this study is the second-order statistical feature extraction also called Gray Level Co-occurrence Matrix (GLCM) [17]. Grey-Level Co-occurrence Matrix texture measurements have been the workhorse of image texture since they were proposed by Haralick in the 1970s. The GLCM is a tabulation of how Often different combinations of pixel brightness values (gray levels) occur in an image. First order texture measures are statistics calculated from original image values, like variance, and do not Consider pixel neighbor relationships. Second order measures Consider the relationship between groups of two (usually neighboring) pixels in the original image. Features of the first order include: mean, standard deviation, smoothness, third moment, uniformity, entropy. To calculate the above features are used the following equation.Mean adalah ukuran dari rata-rata intensitas:

$$m = \sum_{i=0}^{L-1} Z_i\ P(Z_i) \tag{12}$$

Standard deviation is a measure of contrast average:

$$\sigma = \sqrt{\sum_{i=0}^{L-1}(Z_i - m)^2 p(Z_i)} \tag{13}$$

Smoothness measure relative smoothness of intensity in the region:

$$R = 1 - \frac{1}{1+\sigma^2} \tag{14}$$

Third moment to measure skewness of histogram:

$$\mu_3 = \sum_{i=0}^{L-1}(Z_i - m)^3\ p(Z_i) \tag{15}$$

Uniformity:

$$U = \sum_{i=0}^{L-1} p^2\ (z_i) \tag{16}$$

Entropy:

$$e = -\sum_{i=0}^{L-1} p(z_i)\ log_2\ p(z_i) \tag{17}$$

Second-order statistics done by creating kookurensi matrix, ie a matrix that represents the adjacency relationship between pixels in the image at various orientation and spatial distance. One technique to obtain the statistical characteristics of second order is to calculate probability of adjacency relationship between two pixels at a certain distance and angular orientation. This approach works by forming a matrix kookurensi of image data, followed by determining the characteristics as a function of the matrix between them. Occurrences means joint events, ie number of occurrences of level of neighboring pixel values with one another level pixel values within (d) and orientation angle (θ) specific. Occurrences matrix formation of an image with 450 in the angular orientation shown in Figure 9.

Fig -9: Occurrences matrix with angle 45⁰

While GLCM features used in this study are 20 features that include first-order features are rotated on the axis 0⁰, 45⁰, 90⁰ and 135⁰. Results of GLCM features can be seen in Figure 10.
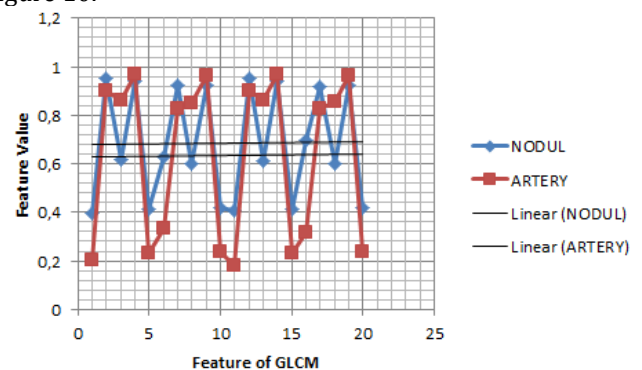

Fig -10: Results of feature extraction using GLCM

The second step is the process of classification using Support Vector Machine (SVM) [18], [19]. In the process of classification using SVM, performed two steps, the first step is the training, while the second phase is the testing step. The amount of training data is 20 per class, while testing is done by using the image data that has not done the training is number 15 CT Scan. The training phase is used to obtain the coordinates of support vector, weight, and distance suppot bias vector, whereas the testing phase is to use data other than the training data to get the results of the classification, so as to know the level of accuracy. Some second-order statistical texture features (GLCM) has been selected as the input system classifier that is considered able to present the nature of the texture of the object.
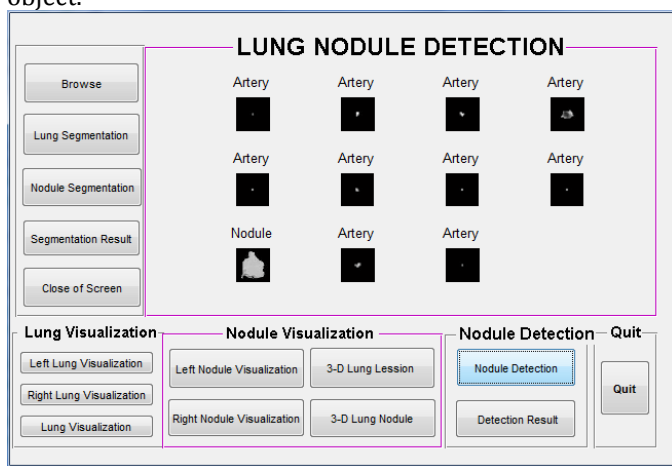

Fig -11: Results of nodules detection

Classification results will be made a comparison with groundtruth, so that would be obtained four values, each of which is a true positive, false negative, false positive and true negative. True positive (TP) shows the image of nodules identified precisely in its class. False positive (FP) is the image of nodules identified another class. True negative (TN) is an image not nodules, identified the right not nodules. False negative (FN) shows the image that should not nodules, identified as nodules. Based on the four values obtained true positive rate (TPR), which is known as sensitivity. Sensitivity formula is as follows:

$$TPR = \frac{TP}{TP+FN} \qquad (18)$$

The false positive rate (FPR) or specificity is a value that indicates the level of error in identification obtained by the following equation:

$$FPR = \frac{FP}{FP+TN} \qquad (19)$$

While the value that indicates the accuracy of identification is obtained from the following equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} x100\% \qquad (20)$$

The greater the degree of accuracy, sensitivity and specificity declare the results more accurate prediction system.

3.5. 3-D visualization of Lung Nodules

To assist the user in detecting pulmonary nodules, the developed software comes with modules for the visualization of the lung and pulmonary nodules. This is because the medical image is generally a collection of 2D image slices are difficult to interpret, to make it more resembles its original form, subsequent 2D medical images is transformed into a 3D image. Visualization is done by using volume rendering. Reconstruction begins with the process of reading a collection of 2D medical images to be set based on the position spatialnya. The next process, applying the technique of image rendering using volume rendering (VR). Actually, there are other rendering techniques used for 3-D reconstruction that surface rendering and volume rendering. Surface rendering technique shows the surface of the object regardless of the object inside information. It also visualize an object as a set of surface called iso-surface. While volume rendering technique to display the overall volume of the object. Volume rendering is implemented using modules from central mathwork, with vol3d.m name. Image rendering is performed lung from the smoothing, in a way to render the texture field of x, y, z simultaneously. The aspect ratio of the data [1 1 0.6], which means that one unit on the x is equal to the length of one unit in y and 0.6 units in z. The next is to create a linear alphamap by increasing opacity. Results of lung volume rendering is displayed using the command view (3), the purpose of which is to display 3D shape with az = -37.5 el = 30.
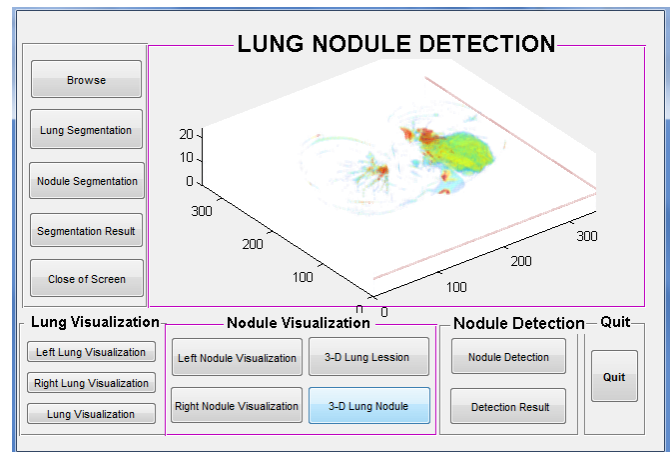

Fig -12: Visualization of lung nodules

4. EXPERIMENTAND RESULTS

The test is conducted using two stages. In the first scenario, the system was tested using data that is used for training data. This scenario is used to see the effect of selection on feature detection model performance. Results from the first trial showed a 100% accuracy rate. The second trial was conducted using data input in the form of

lung CT scan images were not used in the training data. Data used were 15 CT Scan. System performance is then measured by indicators of the value of accuracy, sensitivity and specificity. From the test results showed an accuracy of 75%, sensitivity 77%, specificity 22%.

5. CONCLUSION

Based on the tests, it can be concluded that software detection of pulmonary nodules by using Support Vector Machine (SVM) proved capable of being used as a model for the detection of lung nodules. This is demonstrated by the average value of an accuracy of 75%, a sensitivity of 77% and specificity of 22%.

ACKNOWLEDGEMENT

REFERENCES

[1] U.S. Cancer Statistics Working Group, United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2015.

[2] Horn, L; Pao W; Johnson DH, Chapter 89, In Longo, DL; Kasper, DL; Jameson, JL; Fauci, AS; Hauser, SL; Loscalzo, J. Harrison's Principles of Internal Medicine (18th ed.). McGraw-Hill. ISBN 0-07-174889-X, 2012.

[3] Lung Carcinoma: Tumors of the Lungs, Merck Manual Professional Edition, Online edition. Retrieved 15 August 2007.

[4] Lu C, Onn A, Vaporciyan AA et al., "78: Cancer of the Lung". Holland-Frei Cancer Medicine (8th ed.). People's Medical Publishing House. ISBN 978-1-60795-014-1, 2010.

[5] Collins, LG; Haines C; Perkel R; Enck RE (January, Lung cancer: diagnosis and management, American Family Physician (American Academy of Family Physicians) 75 (1): 56–63. 2007.

[6] Golosio B, Masala GL, Piccioli A, Oliva P, Carpinelli M, A novel multithreshold method for nodule detection in lung CT. Med Phys 36(8): 3607–3618, 2009.

[7] Messay T, Hardie R, Rogers S, A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 14(3): 390–406, 2010.

[8] Riccardi A, Petkov TS, Ferri G, Masotti M, Campanini R, Computer-aided detection of lung nodules via 3D fast radial transform, scale space representation, and Zernike MIP classification. Med Phys 38(4): 1962–1971, 2011.

[9] Camarlinghi N, Gori I, Retico A, Bellotti R et al, Combination of computer-aided detection algorithms for automatic lung nodule identification. Int J CARS 7(3): 455–464, 2011.

[10] Tan M, Deklerck R, Jansen B et al., A novel computer-aided lung nodule detection system for CT images. Med Phys 38(10): 5630–5645, 2011.

[11] Hardie RC, Rogers SK, Wilson T, Rogers A., Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med Image Anal 12(3): 240–258, 2008.

[12] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, The use of active shape models for locating structures in medical images, Image Vis. Computing, vol. 12, no. 6, pp. 355–366, 1994.

[13] Sri Widodo, Wijiyanto, Lung Field Segmentation On Computed Tomography Image Using Active Shape Model, Jurnal Kursor, Universitas Trunojoyo, Volume 7, Nomor 2, Juli 2013, pp 99-108, 2013.

[14] Yu-qian, Zhao, etc, "Medical Images Edge Detection Based on Mathematical Morphology", Proceedings : IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005.

[15] Angenent, S., Eric Pichon, and Allen Tannenbaum, Mathematical Methods in Medical Image Processing, Buletin of the American mathematical society, 2000.

[16] Charles RG, Edward RD, Morphological methods in image and signal processing. Prentice Hall, New Jersey, 1988.

[17] Mark S. Nixon A and Alberto S. Aguado, "Feature Extraction And Image Processing", Second Edition, AcademicPress is an imprint of Elsevier, 2008.

[18] Cristianini N, Shawe-Taylor J., An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge, 2000.

[19] Burges CJC, Tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2): 121–167, 1998.

BIOGRAPHIES

Kustanto

He was born in Sragen on March 10, 1973. He graduated Bachelor in University of Muhammadiyah Surakarta on majoring in Electrical Engineering in 2000. post-graduate at University of Gajah Mada in Information systems in 2011. Field which concentrates on currently is image processing, neural networks, artificial intelligence, and MATLAB. At this time author actively working as a computer lecturer in Department of Informatics Engineering, STMIK Sinar Nusantara Surakarta, Cenral Java, Indonesia.

Sri Widodo

He was born in Sragen on October 28, 1972. He graduated Bachelor inSekolahTinggiManajemendanInformatikaKomputer 'Akakom' Yogyakarta on majoring in Computer Science in 1998, post-graduate at the University of Muhammadiyah Surakarta in the management department in 2005. Field which concentrates on currently is image processing, neural networks, artificial intelligence, biomedical, and MATLAB. At this time author actively working as a computer lecturer in Department of Medical Records and Health Informatics, APIKES Citra Medika Surakarta, Cenral Java, Indonesia.

Sri Tomo

He was born in Sukoharjo on September 7, 1966. He graduated Bachelor in University of Gajah Mada Yogyakarta on majoring in Electrical Engineering in 1994, post-graduate at University of Dian NuswantaraSeamarang in Information systems in 2012. Field which concentrates on currently is image processing, neural networks, artificial intelligence, and MATLAB. At this time author actively working as a computer lecturer in Department of Informatics Engineering, STMIK Sinar Nusantara Surakarta, Cenral Java, Indonesia.