Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 14, 2020

IMPROVED ACCURACY IN DETECTION OF LUNG CANCER USING SELF ORGANIZING MAP

Sri Widodo[1], Ibnu Rosyid[2], Mohammad Faizuddin Bin MD Noor[3], Roslan bin Ismail[4]

[1] Health Science Faculty, Duta Bangsa Surakarta University, Central Java, Indonesia
[2] Radiology Department, Ir. Soekarno General Hospital, Central Java, Indonesia.
[3,4]Unikl Malaysian Institute Of Information Technology, Universiti Kuala Lumpur, 1016 Jalan Sultan Ismail, 50250 Kuala Lumpur
widodosri1972@gmail.com, pastimulia@gmail.com mfaizuddin@unikl.edu.my, drroslan@unikl.edu.my

Received:13.04.2020 *Revised: 18.05.2020* *Accepted: 17.06.2020*

Abstract
Lung cancer is a type of lung disease that is characterized by growth of cells that are not recognized in the lung tissue. Lung cancer cell types are generally divided into two groups, namely: small cell carcinoma lung cancer (SCCLC) and non-small cell carcinoma lung cancer (NCCLC). In the last decade, detection of lung cancer using intelligent systems has been carried out. This paper describes detection of Lung Cancer on CT_Scan Using Self Organizing Map. This work starts from segmentation of lung area using Active Appearance Model (AAM) method, then segmentation of candidates suspected of lung cancer using morphological proses. The next step is feature extraction uses texture and shape feature. The last step is cancer detection using Self Organizing Map (SOM). Results of detection obtained an accuracy rate of 87%.

Keywords: AAM, Ct Scan, Lung Cancer, morphological, SOM.

DOI: http://dx.doi.org/10.31838/jcr.07.14.121

INTRODUCTION

Lung cancer cell types are generally divided into two groups, namely: lung cancer of small cell carcinoma (LCSCC) and non-small cell lung cancer (NSCLC). Types of lung cancer, non-small cell type carcinoma include: adenocarcinoma, squamous cell carcinoma, large cell carcinoma and adenoskuamosa carcinoma[1], [2]. Pulmonologists and radiologists only use naked eye vision to read and diagnose lung cancer on 2-D CT-Scan images. The reading process is done by placing CT-Scan film printed out on the reading lamp. This technique is certainly not effective. In addition, pulmonologists can be different in diagnosing lung cancer, including determining type, shape, size and location of abnormalities in the lung organs. Therefore we need an application that can detect or diagnose lung cancer from CT scan images automatically.

Several studies on detection of lung cancer using intelligent systems have been carried out. These studies include research from Bhagyashri, namely detection of lung cancer cells on CT-Scan using image processing methods [3]. Segmentation method used is thresholding and watershed method. Meanwhile, to predict lung cancer using a binaryization approach and masking process. The watershed segmentation method has an accuracy of 85.27% while segmentation using thresholding has an accuracy of 81.24%. The second study is a research from Vijay A.Gajdhane is detection of lung cancer on CT-Scan images using various image processing techniques [4]. First, preprocessing using filter gabor and segmentation using watershed. For Region of Interest (ROI) extraction using three features, namely area is the number of white pixels in the extracted plane, perimeter is length of Region of Interest (ROI) boundary was extracted, and eccentricity is shape of a round object. Classification using Support Vector Machines (SVM) method.

The third study is a research from Disha Sharma is lung cancer identification using image processing techniques [5]. Study begins with preprocessing stages of image, namely removal of noise, and improvement of image quality using Wiener Filters. Features used to identify nodules are ROI, calcification, shape, nodule size and contrast improvement. Accuracy obtained is 80%. The fourth study is a research from Mahersia, namely detection of pulmonary nodules from CT-Scan images [6] . In this research using three stages to detect pulmonary nodules, namely: preprocessing, lung segmentation and candidate nodule classification. This study emphasizes location of nodules. Method used is classification and clustering, which includes: fuzzy and neural networks, K-nearest neighbors, support vector machines and linear discriminant analysis. The last study was research from Mokhled who conducted research on the detection of lung cancer using image processing techniques [7]. The study began with image processing to improve quality of images using Gabor and Gausian filters. Segmentation uses Binarization and Masking Approach. The features used for classification process are pixels percentage and mask-labeling. Accuracy obtained is 85.7%. From description above it can be seen that all methods used for lung segmentation and nodule candidates use conventional methods that based of contrast values of lung image edge. Disadvantage of this method is if shape of cancer is large and unclear lung border, so if segmentation of cancer image is done, the main focus is not involved in the lung image, so segmentation will fail. In addition, the features used are too few, so classification does not get maximum results.

This paper describes the detection of Lung Cancer on CT-Scan Using Self Organizing Map. This work starts from segmentation of pulmonary area using Active Appearance Model (AAM) method [8], then segmentation of candidates suspected of having cancer using morphological math [9], [10]. The next step is feature extraction [11]. The features used is texture, and the last one is Lung Cancer detection using Self Organizing Map (SOM) [12].

MATERIALS AND METHOD

Data used in this study is CT images of 30 patients in axial slices. Image size of 505x427 pixels, and a thickness of 0.5-10 mm. Ct Scan image were taken from Ir Soekarno Sukoharjo Regional General Hospital, Central Java Indonesia. CT Scan image of Non-Small Cell Lung Cancer Types can be seen at figure 1. Detection proses of non-small cell carcinoma lung cancer on CT-Scan images can be explained as bellow. The first is lung organ

segmentation using Active Appearance Model (AAM) [8]. The second step is segmentation of cancer candidate using morphological math [13], [14]. Third step is feature extraction [15], and then classification of lung cancer using Self Organizing Map (SOM) method [16], and the last is 3-D reconstruction of lung cancer using Volume Rendering [17].

Stages of research on detection of lung cancer on CT-Scan images can be explained as figure 2.

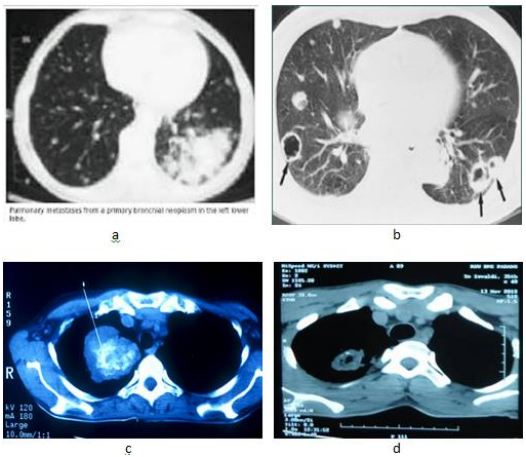

Figure 1. CT Scan of Non-Small Cell Lung Cancer Types (a). Adenocarcinoma, (b) Squamous Cell Carcinoma, (c) Large Cell Carcinoma and (d) Adenosquamous Carcinoma

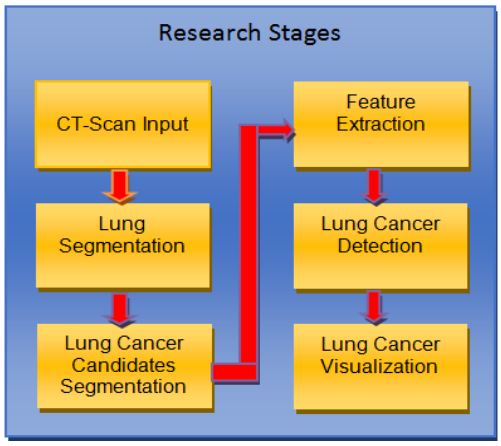

Figure 2. Steps of Lung Cancer Detection

RESULTS AND ANALYSIS

Lung Segmentation Using Active Appearance Model (AAM)

The segmentation of the lung fields is carried out with the aim to facilitate the process of segmenting cancer candidates (18). The method used for segmenting the lung field is Active Appearance Model (AAM)[19]. AAM will produce a deformable model which can change and adjust and estimate the pose of the input image (frame frame video) through the fitting process. The model is built through the training process of a collection of training images so that two model representations are produced, namely shape and appearance which have variations or modes of each. The number of modes of the model determines the model's performance in the fitting process. In addition, the initialization of the placement of the model also affects the performance of the fitting process in achieving convergence [20].

In this paper, we do not discuss AAM segmentation in detail, because the topic has been discussed in other papers [21]. The results of lung segmentation can be seen in Figure 3.

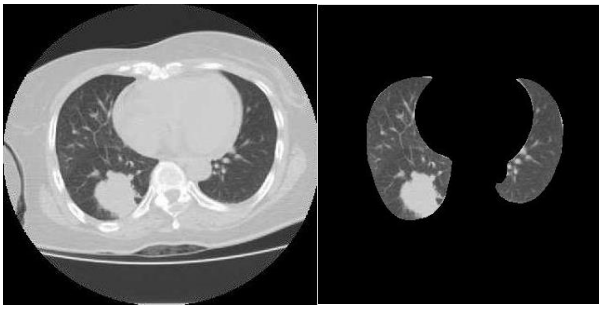

Figure 3. Lung Field Segmentation Results

Segmentation of Lung Cancer Candidates Using Morphological Mathematics

Segmentation of lung cancer candidates with mathematical morphology. Morphological operations are operations that are commonly imposed on binary (black and white) images to change the shape structure of objects contained in an image. The core of morphological operations involves two arrays of pixels. The first array is an image that will be subject to morphological operations, while the second array is called the kernel or structuring element [22]. The process of segmenting lung cancer candidates in detail has been discussed in my previous paper [23]. The results of the lung cancer candidate segmentation are shown in Figure 4.

Figure 4. Cancer Candidate Segmentation Results

Feature Extraction

This research uses texture features. Texture is a mutual relationship between the intensity values of neighboring pixels that repeat over an area larger than the distance of the relationship [24]. The method used to obtain texture features is a statistical method, a method that uses statistical calculations to form features In this study the statistical methods used are first-order statistical methods and second-order statistical methods, or better known as the Gray Level Co-occurrence Matrix (GLCM) method [25]. The simplest method for obtaining textures is based on the histogram The first method discussed is the first order statistical method. The first feature that is calculated statistically is the average intensity. Components of this feature are calculated based on equations

$$m = \sum_{i=0}^{L-1} i \, . p(i) \quad (1)$$

In this case, i is the gray level in the images f and p (i) expressing the probability of the appearance of i and L expressing the highest gray level. The above formula will produce the average brightness of the object. The second feature is a standard deviation. The calculation is as follows:

$$\sigma = \sqrt{\sum_{i=1}^{L-1} (i - m)^2 p(i)} \quad (2)$$

In this case, σ^2 is called a normalized variance or second-order moment because p (i) is a function of opportunity. This feature provides a measure of contrast. The skewness feature is a measure of asymmetry towards the average intensity. Definition:

$$skewness = \sum_{i=1}^{L-1}(i-m)^3 p(i) \quad (3)$$

Skewness is often referred to as a normalized third-order moment. A negative value states that the brightness distribution leans towards the average and a positive value indicates that the brightness distribution leans towards the average. In practice, the skewness value is divided by (L-1) 2 so that it is normalized. Energy descriptors are measurements that express the distribution of pixel intensity over the gray level. The definition is as follows:

$$energi = \sum_{i=0}^{L-1}[p(i)]^2 \quad (4)$$

A uniform image with a gray level will have a maximum energy value, which is equal to 1. In general, images with a little gray level will have a higher energy than those with a lot of gray level values. Energy is often referred to as uniformity. Entropy indicates the complexity of the image. The calculation is as follows:

$$entropi = -\sum_{i=0}^{L-1} p(i)\ \log_2 (p(i)) \quad (5)$$

The higher the value of entropy, the more complex the image. Keep in mind, entropy and energy tend to be the opposite. Entropy also represents the amount of information contained in the data distribution.The second texture feature is the Gray Level Co-occurrence Matrices (GLCM) texture feature, the texture feature in the second order. Measurement of textures in the first order uses statistical calculations based on the pixel value of the original image, such as variance, and does not pay attention to the neighboring pixel relationship. In the second order, the relationship between pairs of two original image pixels is taken into account. For example, f (x, y) is an image of the size of N_x and $N_\cdot$ which has pixels with the possibility to L level and $\vec{r}$ is a vector of spatial offset. $GLCM_{\vec{r}}(i,j)$ is defined as the number of pixels with $j \in 1,..,L$ which occurs at $\vec{r}$ offset of pixels with $i \in 1,..,L$ values, which can be expressed in formulas (Newsam and Kammath, 2005):

$$GLCM_{\vec{r}}(i,j) = \#\{(x_1,y_1),(x_2,y_2) \in (N_x,N_y) \times (N_x,N_y) | f(x_1,y_1) = j^{\vec{}} = \overrightarrow{(x_2-x_1,y_2-y_1)}\} \quad (6)$$

Dalam hal ini, ofset $\vec{r}$ dapat berupa sudut dan/atau jarak. While GLCM features used in this study are 20 features that include first-order features are rotated on the axis 0^0, 45^0, 90^0 and 135^0. The second step is the process of classification using Self Organizing Map (SOM) [26]. In the process of classification using SOM. SOM is used to group (cluster) data based on data characteristics or features.

Lung Cancer Detection Using Self Organizing Map (SOM)

SOM is an artificial neural network method that was introduced around the 1980s by Professor Teuyo Kohonen. SOM is one of the topology forms of the Unsupervised Artificial Neural Network (Unsupervised ANN) where the training process does not require supervision (target output). In this network, a layer containing neurons will arrange themselves based on the input of certain values in a group known as clusters. During process of self-arrangement, cluster that has most weight matches input pattern (has the closest distance) will be selected as winner. Winning neurons and their neighboring neuronneurons will improve their weights [26]. There are m unit groups arranged in architecture of input signals (input) of n. Weight vector for a group unit is provided from the insert patterns joined with the group. During the self-organizing process, group unit that has weight vector that best matches insert pattern (indicated by minimum Euclidean distance) is selected as the winner. The winning unit and its neighbor unit are updated in weight. Each neuron is connected with other neurons that are associated with weights or weight. SOM is used to group (cluster) data based on data characteristics or features.

The unsupervised learning algorithm in Kohonen SOM in grouping data can be explained as follows:

1. Determine number of variables, number of data and number of clusters
2. Initialization. Initialization includes:
   a. The value of w_{ij} weights at random with the value (0-1) with w_{ij} = weight of the connection between i-th input node to the j-th output node, i is the node value at the input layer, and j is node value at output layer.
   b. Initial neighborhood size value N_m (0) with a value large enough but smaller than the number of output nodes. Where m is the winning node's index and N_m (0) is number of neighbors or neighbors of initial winning node
   c. Determine parameters α (t) (learning rate) and σ2 (t) (activation function coefficients) between 0 to 1.
   d. The parameter Ω (epoch), which is number of times a data is entered into network for training process before neighbor size decreases for each iteration.
3. Insert input vector into input layer and calculate distance (d) of this input to the weight w of each node j with euclidean distance equation: d_j

$$d_j = \left|\left|x - w_j\right|\right| = \sqrt{\sum_i^n 1(x_i - w_{ij})^2} \quad (7)$$

Where x_i is the i-th input node, w_{ij} is the weight of the connection between the i-th input node and the n-th output node and n is the number of nodes in input layer. The node with smallest distance is considered winner of m (called the best matching unit).

4. Update the weight vector at the winner node m from its neighbor node with the formula:

$$w_{ij}(t+1) = w_{ij}(t) + c[x_i = w_{ij}(t)] \quad (8)$$

Where w_{ij} (t + 1) is the weight of the connection between the I-nodes input and output at the next iteration.

$$C = \alpha(t)hib(t),$$
$$C = \propto (t)hib(t)$$
$$\propto (t) = \exp\left(\frac{-||\tau t - \tau m||}{\sigma^2(t)}\right) \quad (9)$$

hib (t) is a neighboring function that is a Gaussian function for all nodejs in Nm (t), while Ri - Nm = number of nodes or physical distance between node and winner node m with Euclidean distance.

3. Continue from the third step for Ω epoch, add 1 to t, and reduce the size of neighborhood, α (t) and σ^2 (t).

$$\alpha(t+1) = \alpha(t) *_{N_m(t)}^{N_m(t+1)} \quad (10)$$

Where, α (t + 1) is the learning rate on the next iteration and N_m (t + 1) = neighbor size on the next iteration.

The process of detecting non-small cell carcinoma lung cancer using the SOM algorithm begins with the training data call. While the training data used are CT-Scan training data on types of non-small cell lung cancer which includes Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, Adenosquamosa and Arterial Carcinoma, and then calculated with the initial weight at each iteration. After that, SOM repeats

based on the number of iterations, in this study using 5 iterations. If it has reached 5 iterations, it will issue a cluster or pattern formation. The number of training data used is 18, the number of variables 5, the number of desired clusters is 5, the initial weight is random, learning rate (α): 0.5 and update learning rate (α ($baru$)): 0.5 * α (ama). Clustering results from each input data with a maximum of 10 iterations are expressed with index 0 for cluster 1, index 1 for cluster 2, and index 2 for cluster 3, index 3 for cluster 4 and index 4 for cluster 5. Result of lung cancer detection can be seen on figure 7.

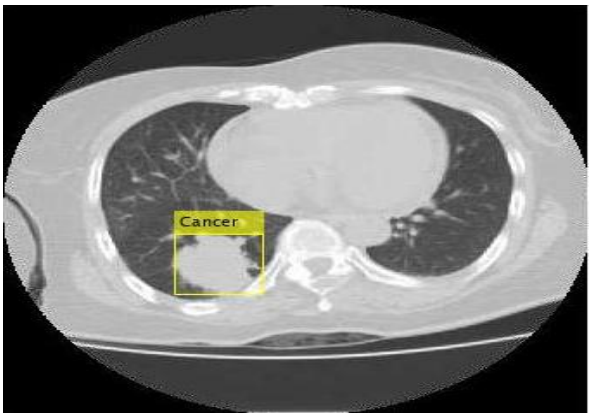

Figure 7. Results of Lung Cancer Detection

Testing process is done after CT Scan image of lung cancer is a segmentation process to separate lung disease image with surrounding tissue, and also to determine Region of Interest (ROI) of lesion object. Furthermore image is normalized to 60x60 pixels with gray degree 255. Testing process is done by 30 sample data and 33 fitur . The number of classes there are 2, artery class, cancer class. Results of test by using data test as much as 75 data can be seen in Figure. 8 below.

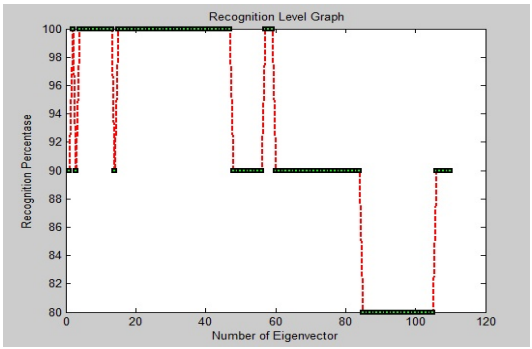

Figure 8. Accuracy of Recognition Level

3-D Visualization of Lung Cancer

Volume rendering is the process of visualizing the characteristics and properties of volumetric data in three-dimensional (3D) objects . Volumetric data often contains two-dimensional images with fixed intervals. then sorted to form a rectangular box, resembling a Rubick Cube structure. Volume rendering has a fundamental difference with surface rendering where polygons make a rendering process using the right surface image while rendering volume represents all data in a large block [27]. Volume rendering includes a set of techniques for rendering volumetric data sets. In medical images, volumetric data can be obtained from various sources such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound or Positron Emision Tomography. Before doing volume rendering, you must first have a volumetric dataset which generally has a set of v samples (x, y, z, v) or commonly known as voxels. Each voxel has location information (x, y, z) and also values of v, and some properties of volumetric data. Voxel values can vary between different types of datasets. Volume rendering is implemented using modules from central mathwork, with vol3d.m name [jatit]. 3-D Visualization of Lung Cancer can be seen in Figure. 9 below.

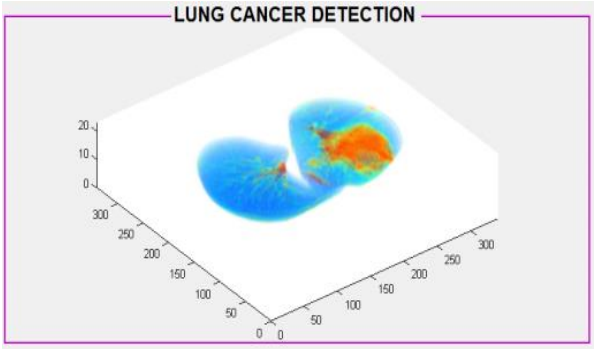

Figure 9. Visualization of Lung Cancer

CONCLUSION

Segmentation with AAM can successfully segment the lung fields that have a large abnormality, resulting in an unclear edge of the lung, this is due to the low contrast value. Segmentation of lung disease candidate with Morphology Mathematics successfully segmented the abnormality well, even abnormalities attached to the artery, which is difficult for some researchers to do. After testing with 75 testing data, by using Self Organizing Map (SOM) methode using 75 input data, showed average accuracy of 78%. Based on the tests, it can be concluded that detection of lung cancer by using Self Organizing Map (SOM) proved capable of being used as a model for the detection of lung cancer.

REFERENCES

1. Lu C, Onn A, Vaporciyan AA et al. (2010), "78: Cancer of the Lung". Holland-Frei Cancer Medicine (8th ed.). People's Medical Publishing House. ISBN 978-1-60795-014-1.
2. J.D. Bronzino, The Biomedical Engineering, Handbook 2nd Edition, vol. 1, CRC Press, Boca Raton, 2000, p.61.
3. Bhagyashri G. Patil ,2014, Cancer Cells Detection Using Digital Image Processing Methods, International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 4 March 2014, pp. 45-49.
4. Mr.Vijay A.Gajdhane, Prof. Deshpande L.M., 2014, Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 28-35
5. Disha Sharma, Gagandeep Jindal,2011, Identifying Lung Cancer Using Image Processing Techniques, International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011).
6. Mahersia H., M. Zaroug, Lung Cancer Detection on CT Scan Images: A Review on the Analysis Techniques, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.4, 2015, 38-45.
7. Mokhled S. Al-Tarawneh, Lung Cancer Detection Using Image Processing Techniques, Leonardo Electronic Journal of Practices and Technologies, Issue 20, January-June 2012, p. 147-158.
8. T. F. Cootes and C. J. Taylor. Active appearance models. In IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 484{498. Springer, 1998.
9. Charles RG, Edward RD (1988), Morphological methods in image and signal processing. Prentice Hall, New Jersey.
10. Yu-qian, Zhao, etc,2005, "Medical Images Edge Detection Based on Mathematical Morphology", Proceedings : IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005.

11. Mark S. Nixon A and Alberto S. Aguado, (2008), "Feature Extraction And Image Processing", Second Edition, AcademicPress is an imprint of Elsevier.
12. Riries Rulaningtyas, Andriyan B Suksmono, Tati L R Mengko and Putri Saptawati, (2012), Color segmentation of multi variants tuberculosis sputum images using self organizing map, International Conference on Physical Instrumentation and Advanced Materials, IOP Conf. Series: Journal of Physics: Conf. Series 853 (2017) 012012, pp. 1-6.
13. Angenent, S., Eric Pichon, and Allen Tannenbaum (2000), Mathematical Methods in Medical Image Processing, Buletin of the American mathematical society.
14. Jonker, P. P. (2000). Morphological operations on 3D and 4D images: From shape primitive detection to skeletonization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1953 LNCS*(March 2001), 371–391. https://doi.org/10.1007/3-540-44438-6_31
15. Mark S. Nixon A and Alberto S. Aguado, (2008), "Feature Extraction And Image Processing", Second Edition, AcademicPress is an imprint of Elsevier.
16. Setiawan, Kuswara, The Intelligent System Paradigm, Artificial Intellegence, Bayumedia Publishing, 2003.
17. Widodo, S., & Wijiyanto. (2014). Software development for three dimensional visualization of lung on computed tomography scans using active shape model and volume rendering. *Journal of Theoretical and Applied Information Technology*, *65*(1), 154–160.
18. Kainulainen, Jukka, Clustering Algorithms: Basics and Visualization, Finland: Helsinky University of Technology, 2002.
19. Chu, E.C.P., Wong, J.T.H.Subsiding of dependent oedema following chiropractic adjustment for discogenic sciatica(2018) European Journal of Moleculr and Clinical Medicine, 5, pp. 12-15. DOI: 10.5334/ejmcm.250
20. T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, 1994, The use of active shape models for locating structures in medical images, Image Vis. Computing, vol. 12, no. 6, pp. 355–366, 1994.
21. Widodo, S., Rohmah, R. N., & Handaga, B. (2018). Classification of lung nodules and arteries in computed tomography scan image using principle component analysis. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, *2018-Janua*, 153–158. https://doi.org/10.1109/ICITISEE.2017.8285485
22. Shih, F.Y. 2009. *Image Processing and Mathematical Morphology*. New York: CRC Press.
23. Widodo, S., Rohmah, R. N., Handaga, B., & Arini, L. D. D. (2019). Lung Diseases Detection Caused By Smoking Using Support Vector Machine. *Telkomnika (Telecommunication Computing Electronics and Control)*, *17(3)*, 1256-1266. https://doi.org/10.12928/TELKOMNIKA.V17I3.9799
24. Kulkarni, A. D. 1994. *Artificial Neural Networks for Image Understanding*. New York: Van Nostrand Reinhold.
25. Widodo, S., Rosyid, I., Faizuddin, M., Roslan Bin Ismail. (2020). Texture Feature Extraction To Improve Accuracy Of Malignant And Benign Cancer Detection On Ct-Scan Images. *International Journal of Psychosocial Rehabilitation,* 24(9), 3540-3554.
26. Ely, Hamzah, Rofiqoh, Local Soybean Classification Based on Physical Characteristics Using Artificial Neural Network Kohonen Self Organizing Maps Algorithm, Jurnal Informatika. Universitas Sebelas Maret. Surakarta, (2014).
27. Noon C.J. 2012. *A Volume Rendering Engine for Desktops, Laptops, Mobile Devices and Immersive Virtual Reality Systems using GPU-Based Volume Raycasting*. Iowa State University. United State of America.