Atom Indonesia

Journal homepage: http://aij.batan.go.id

Early Lung Cancer Detection Using Artificial Neural Network

T. Pandiangan[1*], I. Bali[2] and A.R.J. Silalahi[3]

[1]*Medical Physics Study Programme, Faculty of Science, Technology, Engineering and Mathematics, Matana University, Jl. CBD Barat Kav.1, Tangerang 15810, Indonesia*

[2]*Institute of Research, Science Development and Technology Study, Matana University, Tangerang 15810, Indonesia*

[3]*The Australian National University (Visiting research scholar), Canberra, Australia*

ARTICLE INFO

Article history:
Received 16 May 2018
Received in revised form 30 October 2018
Accepted 5 November 2018

Keywords:
Malignant tumor
Image processing
Enhancement
Neural network
X-ray

ABSTRACT

Lung carcinoma is a malignant lung tumor that is deadly and is characterized by the uncontrolled cell growth in the tissue of lung. Normally the lung cancer detection is done by visual inspection of x-ray image by medical doctor. The purpose of this study is to create a computational tool that can automatically detect early lung cancer from x-ray image. This research has two main steps, with first being to characterize cancer or cancer symptoms based on x-ray images and second step is to develop an artificial neural network (ANN). In first step, particularly it is wanted to lay out a rigorous image processing framework with sequential steps: (i) image noise reduction, (ii) image enhancement, (iii) lung organ segmentation, (iv) object edge detection, and (v) tumor boundary detection. The framework incorporates image processing techniques such as thresholding and morphological detections (erosion and dilation). The framework is expected to reveal the relevant features that define lung cancer or early lung cancer such as area, perimeter, density profile and shape ratio. For the second step, the ANN is built based on machine learning algorithm to study a large set of x-ray images of positively diagnosed lung cancer patients. In addition to learning solely based on the 2D x-ray images, it is also incorporated the previously studied tumor features. The two combined with a large dataset is expected to enable the machine to reach a close to 100 % detection accuracy. Based on the test results of 10 samples obtained the comparative value of the calculated by the ANN with the results of measurement with Matlab program is tends to approach the same. It can be concluded that ANN has been successfully educated so that can identify 10 samples correctly.

INTRODUCTION

In the diagnosis process of conventional x-ray imaging of lungs, radiology physician observation is done visually. This observation becomes more difficult when the existing tissue is wanted to be distinguished from the absence of an early tumor in the lung tissue. In addition to that, noise from x-ray scattering of the environment against image brightness makes it more difficult to interpret the observation. In general, images generated from conventional lung x-ray are considered as the detection tool of early lung tumors, but serious errors in diagnostic could have adverse effects, or even can lead to the death of the patient [1].

There are several studies about early lung cancer detection system. Lingayat and Tarambale [2], Patil and Udupi [3], also Ramaraju and Praveen [4] have identified and characterized lung tumor objects that measure the area, perimeter and shape or irregularity of tumor objects. Kumar and Saini have detected lung cancer using artificial neural networks, however the characterization of cancer object has not been identified [5]. In the present study, the tumor objects are classified and also characterized using artificial neural networks. The goal of the research is to obtain a smart application program or intelligent machine to detect and classify early tumor type using artificial neural network. This correct detection results may help

*Corresponding author.
E-mail address: tumpal@matanauniversity.ac.id
DOI: https://doi.org/10.17146/aij.2019.860

radiologists in analyzing and diagnosing patients' illnesses more accurately, so that early lung tumor test results can help patients in treatment by avoiding risk or no risk. For patients, the results of this study may be useful for identifying early lung tumor types, so the risk for wrong patient treatments can be minimized. Patients who used this conventional lung x-ray modality, in addition to paying cheaper, also received lower radiation dose, compared to alternative methods using CT scan modalities. As for radiology physicians, research results can be useful in assisting radiologists in identifying and classifying patients' tumor types quickly.

THEORY

Artificial neural networks

Artificial neural networks (ANN) are information processing systems that contain many neurons for highly interconnected processing [6,7]. Neural Networks is a form of multiprocessor computer system with high-level interconnects, simple processing elements, adaptive interactions between elements and messages. These neurons work together in a distributed manner: (i) to learn from input information, (ii) to coordinate internal processes, and (iii) to optimize the outcomes. It is like someone who learns from their experience. There are two types of learning methods that are supervised and unattended. Controlled learning is a learning process that needs to be trained first with a lot of data whereas unattended learning does not require a willing response. The ANN plays an important and impressive role to overcome various health problems such as acute illness and even other minor illnesses. The Hopfield Neural Network (HNN) and FCM (Fuzzy C-Means) algorithms for color image segmentation and comparing their classification with their studies show that HNN is better than FCM, and successfully allows the excavation of the nuclear and cytoplasmic regions [8,9]. Based on the image is done detection of true positive nodule candidate in the lung [10]. A lung segmentation technique was proposed by other researchers, to improve the accuracy of segmentation as well as to separate and combat trachea from the lungs. Digital image processing techniques have been used to achieve better quality and accuracy [11], and to illustrate the development of algorithms that detects lung cancer symptoms in x-ray films by simulating the Cellular Neural Network template [12]. Sensitivity is also important in this aspect as a result of the use of the Massive Training ANN filter, the peculiarity and sensitivity of the CAD system increased significantly with the image difference technique achieving the 96 % sensitivity reported by Suzuki [13].

METHODOLOGY

Research methodology is directed to achieve research objectives. The purpose of the research will be achieved by the acquisition of a model and intelligent application program with the artificial neural network method. The process of identification and classification of early tumors is done based on conventional lung x-ray image processing and ANN process. These results can be used to improve visual observation of the object being observed. The steps taken to achieve the goal are as follows [5].

The first step: initiating preliminary process and characterization of object detection

In preliminary process, image processing framework consists of sequential steps: (i) image noise reduction, (ii) image enhancement, (iii) lung organ segmentation, (iv) object edge detection, and (v) tumor boundary detection. All of the samples processing is done with the Matlab GUI program.

For the image noise reduction, with the reason that the process is easier and simpler, the samples of lung x-ray images are converted into gray scale format. Gray scale images carry intensity values only. This image is also known as white (highest intensity) and black (lowest intensity) and among them is gray.

After the object is converted to gray scale image then the image enhancement process is carried out by the spatial domain filtering technique with averaging mean. In general, the image enhancement includes not only limited to intensity and contrast manipulation but also reduces noise, eliminates background and sharpens borders.

Subsequent to the enhancement process, the threshold segmentation process for lung organ segmentation and object edge detection are conducted. The image containing the structure is dark in the bright background, then thresholding can be used to separate the structure [5].

The filtering average results are converted into binary format through the specified threshold value. Separation feature of malignant and non-malignant tumors in x-ray images is based on low contrast, whereas non-malignant tumors do not show light or are too dark. This research will use multi-level threshold value to classify any point in (x, y) in image.

Following observing all x-ray images in the database, the pixels in the malignant tumor are

calculated in the range of values. The pixel value which is less than T_2 and greater than T_1 is the background set to zero. A pixel whose value lies between T_1 and T_2, maintains its pixel value. After this process, then the image is converted into binary format, by setting all the foreground pixel values equal to 255 [5].

In characterization of object detection, the feature is extracted based on the classification process on x-ray binary image [6]. This is a set of connected pixels. It lies on the boundary between the two regions and is detected by the Sobel method because of its accuracy. The extracted features are characterized as area, perimeter and shape. Area, perimeter and shape are scalar values. This feature is used to develop diagnostic rules in detecting malignant tumors.

The second step: perform process identification and classification by artificial neural network with deep structural learning

Artificial neural networks were developed for the classification of tumor types after applying the training and the testing process. Artificial neural networks consist of three main layers, namely the input layer, the hidden layer and the output layer (Fig. 1, Fig. 2 and Fig. 3).

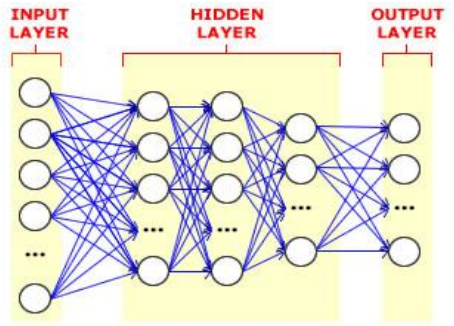

Fig. 1. The process of identification and classification by neural networks with deep structural learning.

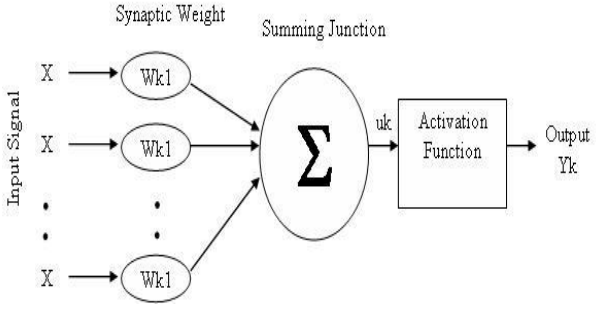

Fig. 2. The process of identification and classification by artificial neural networks.

$$uk = \sum_{j=1}^{p} Wkj.Xj \Longrightarrow Yk = \frac{1}{1+\exp(-uk)}$$

Delta rule $W_{new}=W_{old} + \beta EX/IXI$

Fig. 3. Artificial neural network learning rules.

This research uses back propagation algorithm [14-16]. The usefulness of this algorithm is to reduce errors generated by the difference between the actual output and the expected results accurately. Artificial neural networks using Matlab NNTOOL give results, which show the performance of ANN with graphics [5].

Research path identification and classification of tumor types

Neuron modeling has *Xp* input of *p*, derived from another cell or from external input (not from neuron). Furthermore, each input given weighted *Wk*. Each *Xp* input will be multiplied by the corresponding weights of *Wk*. For all the multiplication result will be summed up as in equation 1 as below.

$$uk = \sum_{j=1}^{p} Wkj.Xj \quad (1)$$

where X_j is the input signal, W_{kj} represents the weight of nerve synapses *k*, *p* is the number of input signal, and *uk* is a linear combiner output depending on the input signal. The result of the equation will be the input for the activation function to obtain the degree of output signal in the neuron, where there are various types of activation functions. For the type of sigmoid function can be described by equation 2, where *Yk* is the output of activation function. The graph of sigmoid function is shown in Fig. 4.

$$Yk = \frac{1}{1+\exp(-uk)} \quad (2)$$

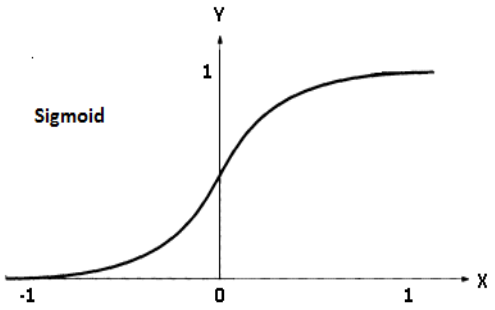

Fig. 4. Graph of sigmoid function.

The activation function of each neuron is different, this is because of the difference in the weight values received per different neuron. The next steps are ANN output analysis, output identification and output classification, respectively.

RESULTS AND DISCUSSION

Image processing results

The image processing methods (enhancement, segmentation and feature separation) [1,5] were performed for each of 50 standard lung x-ray samples of parents, of which 25 are lung nodule samples (LN) and the remaining 25 are non-lung nodule samples (NLN). The example of image processing results for lung nodule samples are given in Fig. 5. In Fig. 5, each of the white dots represents the lung nodule that has been separated from the x-ray image of the lung. The size and the shape of lung nodule objects varies from one image to another. To find out the character of each object, a further process is required in determining the value of area and perimeter. The area and perimeter values dictate the value of shape variable.

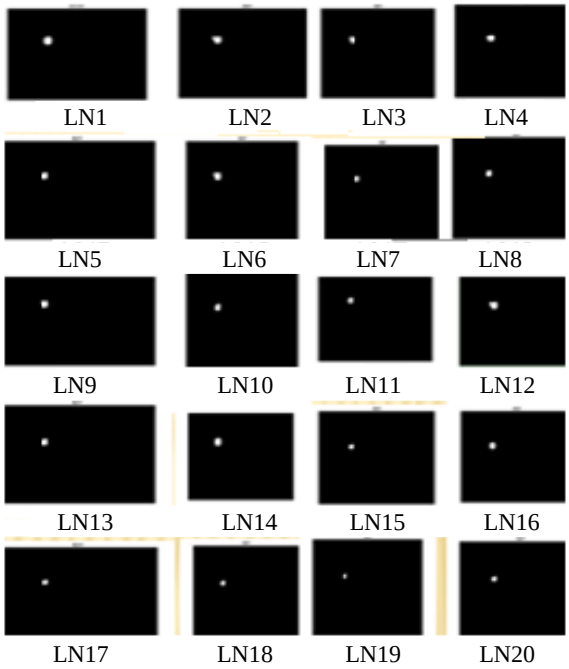

Fig. 5. Lung nodule image results from enhancement, segmentation and feature separation process.

Of the 25 lung nodule samples, 20 samples were used for training and 5 samples were used to test the neural network. In each lung nodule, the perimeter value, the area and the size of the shape of the object were calculated using Matlab program [2,4]. The calculation results are presented in Table 1. The calculated values of perimeter, area and shape size are shown in Table 2.

Table 1. Data of area, perimeter and shape of lung nodule (LN).

Code	Area	Perimeter	Shape
LN1	1100	130.089	0.816396575
LN2	731	99.515	0.927107131
LN3	467	79.964	0.917312896
LN4	920	114.831	0.816396575
LN5	608	88.755	0.969410149
LN6	476	79.212	0.952828271
LN7	1100	130.089	0.816396575
LN8	733	98.863	0.941946082
LN9	470	79.311	0.938470544
LN10	622	89.144	0.983095683
LN11	469	79.311	0.936473798
LN12	731	99.515	0.927107131
LN13	467	79.464	0.928892968
LN14	1128	133.273	0.797653713
LN15	920	114.831	0.876312836
LN16	608	88.755	0.969410149
LN17	851	105.361	0.962851767
LN18	1154	123.241	0.954300214
LN19	1193	123.308	0.985479457
LN20	1154	123.241	0.954300214

The shape values of all objects in Table 1 are close to one, which corresponds to circular object. As noted, the higher the shape value, the higher is the probability of the object being a lung nodule [3]. Conversely, if the value of the shape of the object is much less than one or close to zero then the probability of the object as lung nodule is low or non-lung nodule [3]. The shape values of the test samples are also close to one, which means that the probability level of the object as the lung nodule is quite high. The calculation results of object shape value are shown in Table 2 at column 4.

Table 2. Data of area, perimeter and shape for test sample of lung nodule (LNu).

Code	Area	Perimeter	Shape
LNu1	95	35.526	0.945410936
LNu2	412	73.016	0.970623452
LNu3	89	34.418	0.94364444
LNu4	92	35.972	0.892993499
LNu5	133	40.894	0.998900188

The next 25 samples are non-lung nodule samples, in which twenty samples were used for training and five samples were used for testing of neural network. In each non-lung object, the size of the perimeter, the area and the shape of the object are calculated using Matlab program [1], which are shown in Table 3. Table 4 presents the calculated values from the 5 standard samples of non-lung nodule, which are designated as test samples of artificial neural network.

Table 3. Data of area, perimeter and shape for non-lung nodule (NLN).

Code	Area	Perimeter	Shape
NLN1	250	76.246	0.540126541
NLN2	218	58.768	0.792801222
NLN3	267	68.95	0.705394972
NLN4	678	109.39	0.711646124
NLN5	790	119.132	0.699133102
NLN6	485	87.976	0.787051143
NLN7	600	120.266	0.521020914
NLN8	710	129.311	0.53330671
NLN9	426	85.3	0.735361987
NLN10	426	85.3	0.735361987
NLN11	557	103.948	0.647459428
NLN12	897	132.886	0.638004328
NLN13	423	79.864	0.832967181
NLN14	276	67.242	0.766685901
NLN15	494	88.382	0.794308003
NLN16	311	69.784	0.802118087
NLN17	459	99.284	0.584849056
NLN18	250	62.538	0.802863421
NLN19	325	72.056	0.786199387
NLN20	627	115.512	0.590205121

The shape values in Table 3 are significantly lower than 1, meaning the objects are not classified as having circular shapes [2,4]. This also means that the probability of the objects being lung nodules are low or the objects are more likely non-lung nodules.

For testing of artificial neural network, there are 5 standard samples of non-lung nodule as shown in Table 4.

Table 4. Data of area, perimeter and shape for test sample of non-lung nodule (NLNu).

Code	Area	Perimeter	Shape
NLNu1	436	91.030	0.660856313
NLNu2	297	94.049	0.421733135
NLNu3	389	85.778	0.664029554
NLNu4	259	64.830	0.773993583
NLNu5	415	79.014	0.521052288

Results of artificial neural network training

The artificial neural network (ANN) training is done using Matlab application program, in which it utilizes 1 input layer, 1 output layer and 10 hidden layers. The input layer uses 3 type of parameters (perimeter, area, and shape value) and the output layer has two possible values: nodule and non-nodule. The program learns using 40 samples consisting of 20 lung nodule samples and 20 non-lung nodule samples were analyzed in the training. While for testing, ANN utilizes 5 samples of lung nodule and 5 samples of non-lung nodule. The targets used in the training were 40 samples, in which the outputs are 1 for lung nodule type and -1 for non-lung nodule sample, respectively. Figure 6 shows the input mapping with output on ANN [6].

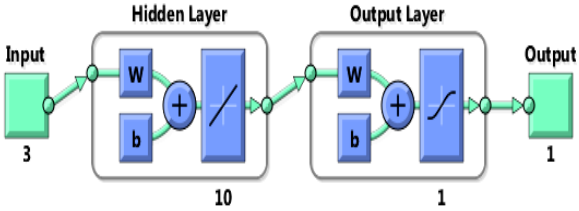

Fig. 6. Artificial neural network form.

As shown in Fig. 7, the value of regress is 0.99992, which means that the output value of the ANN process is almost the same as the planned target value. After the target value is equal to the output value of the ANN process, as shown in Fig. 7, then the training process of the sample is done.

Results of artificial neural network testing

The graph in Fig. 7, shows the value of regress, h = 0.99992, which means that strong predictive power of this value is also shown in Fig. 7 in which the output value coincides with the target value where the target graph and the output are almost the same. After the target value is almost equal to the output value of the ANN process, then the sample testing process is done. ANN learns the samples based on known parameters such as perimeter, area, and shape values. The value of the test results by ANN on test samples of lung nodule (LNu) is shown in Table 5 at line 1, whereas the value of the calculation using Matlab is at line 3. The value on line 1 is almost equal to value in line 3. This test results indicate that ANN has been successfully educated/trained to recognize sample. The sample has been identified with a known shape value as lung nodule.

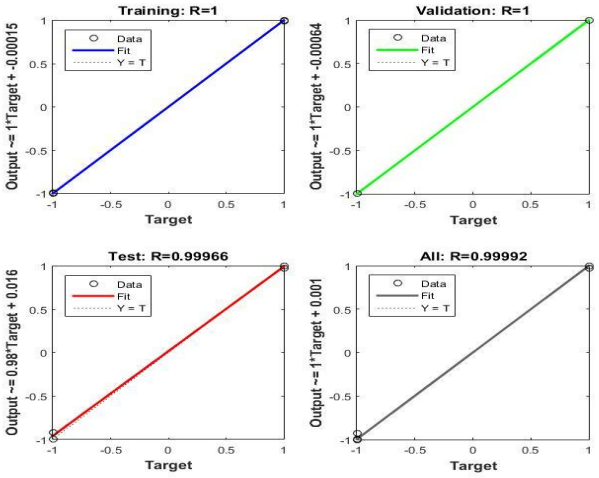

Fig. 7. The training output for the process output approaches the target value.

Table 5. Results of ANN test on LNu test samples.

Results	Sample				
	LNu1	LNu2	LNu3	LNu4	LNu5
Testing results	0.99999	0.99998	0.99999	0.96432	0.99930
Educating Target	1	1	1	1	1
Calculated results	0.94541	0.97062	0.94364	0.89299	0.99890

Table 6. Results of ANN test on NLNu test samples.

Results	Sample				
	NLNu1	NLNu2	NLNu3	NLNu4	NLNu5
Testing results	-1	-1	-1	-1	-1
Educating Target	-1	-1	-1	-1	-1
Calculated results	0.66086	0.42173	0.66403	0.77399	0.52105

Table 6 presents test result of 5 standard test samples of non-lung nodule (NLNu) which is assumed as the test sample. Previously, ANN had been educated to recognize the NLNu sample with the target value being minus one (-1). After that, ANN is tested with 5 test samples. The perimeter, area and shape values of each of the 5 test samples are made into the process inputs on ANN. The result of the 5 test samples is shown at line 1, which are minus one (-1). This value is equal to the target or destination value for NLNu sample identification. The similarity between the target value and the test results proves that ANN identified well the 5 test samples are NLNu. In line 3, there is a value of each shape of 5 test samples calculated based on image processing with Matlab.

CONCLUSION

Artificial neural network (ANN) training program using Matlab application, utilizes ten hidden layers of neural networks, one-layer output, and three types of input. After the neural network training with 40 samples the program proceeds with predicting output of 10 mixed samples. The ANN calculation results show 100 % agreement with the reference data, therefore it can be concluded that ANN has successfully identified all 10 test images that consists of 5 samples of early lung nodule (cancer) and 5 samples of non-lung nodule.

ACKNOWLEDGMENT

The authors would like to thank the Matana University for the financial support and providing the facilities to this study.

REFERENCES

1. T. Pandiangan and I. Bali, Journal of Medical Physics and Biophysics **5** (2018) 222. (in Indonesia)
2. N.S. Lingayat and M.R. Tarambale, International Journal of Biochemistry and Bioinformatics **3** (2013) 624.
3. S.A. Patil and V.R. Udupi, Journal of Scientific and Industrial Research **69** (2010) 271.
4. P.V. Ramaraju and S. Praveen, International Journal for Research in Applied Science & Engineering Technology **3** (2014) 1.
5. V. Kumar and A. Saini, International Journal of Enhanced Research in Management & Computer Application **2** (2013) 40.
6. A.A. Abdullah and S.M. Shaharum, Information Engineering Letters **2** (2012) 49.
7. G. Bhat, V.G. Bridar and H.S. Nalini, Computer Engineering and Intelligent System **3** (2012) 116.
8. F. Taher and R. Sammouda, *Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods*, IEEE GCC Conference and exhibition (2011) 295.
9. F. Taher, N. Werghi, H.A. Ahmad *et al*., American Journal of Biomedical Engineering **2** (2012) 136.
10. J. Tong, D.Z. Zhao, W. Ying *et al*., *Computer-Aided Lung Nodule Detection Based on CT Images*, IEEE- International Conference on Complex Medical Engineering (2007) 816.
11. A. Chaudhary and S.S. Singh, International Journal of Research in Engineering and Applied Sciences **(**IJREAS) **2** (2012) 1351.
12. A. Abdullah and H. Mohamaddiah, *Development of Cellular Neural Network Algorithm for Detecting Lung Cancer Symptoms*, IEEE EMBS Conference on Biomedical Engineering & Sciences (2010) 138.
13. K. Suzuki, Z. Shi and J. Zhang, *Supervised Enhancement of Lung Nodules by use of a Massive-Training Artificial Neural Network (MTANN) in Computer-aided diagnosis*, IEEE, 19th International Conference on Pattern Recognition (2008) 1.
14. P.K. Nancy, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) **4** (2015) 341.

15. P. Naresh and R. Shettar, International Journal of Engineering Research and Applications **4** (2014) 78.

16. R.S. Shriwas and A.D. Dikondawar, International Journal of Electronics & Communication (IIJEC) **3** (2015) 17.