JJBS

Volume 10, Number 4,December 2017
ISSN 1995-6673
Pages 239 - 249

Jordan Journal of Biological Sciences

Lung Cancer Detection Using Multi-Layer Neural Networks with Independent Component Analysis: A Comparative Study of Training Algorithms

Abdelwadood M. Mesleh*

Computer Engineering Department, Faculty of Engineering Technology, Al-Balqa Applied University, Jordan,

Received: May 14, 2017; Revised: June 15, 2017; Accepted: August 10, 2017

Abstract

The present paper presents a Computer-Aided Design (CAD) system that detects lung cancer. Lung cancer detection uses Multi-Layer (ML), Neural Networks (NNs) and Independent Component Analysis (ICA). ICA aims to speed the detection by decreasing the number of features. The ML NNs classifier is trained by Gradient descent algorithm (traingd), Gradient descent with momentum (traingdm), Gradient descent with variable learning rate and momentum (traingdx), Resilient back propagation (trainrp), Fletcher-Reeves Update (traincgf), Polak and Ribiere (traincgp), Powell and Beale Restarts (traincgb), Scaled Conjugate Gradient Algorithm (trainscg), Quasi Newton BFGS (trainbfg), One step secant algorithm (trainoss) and Levenberg–Marquardt (trainlm). The detection algorithm is tuned to determine the existence of cancer in real Computerized Tomography (CT) images and it is validated, trained & tested using 460 CT images, 350 of them belong to lung cancer patients in Jordanian hospitals. The presence of cancer in these images is labeled by experts. The present paper investigates the performance of the ML NN classifier trained by these training algorithms with ICA feature extraction. Results reveal the robustness of the detection algorithm for real CT images. Among the 11 training algorithms, *Levenberg–Marquardt* achieves a classification accuracy of 100% with least number of ICA features.

Keywords: Neural networks, Independent component analysis, Lung cancer detection, Multi-layer neural networks, Training algorithms; Computer aided design, Lung cancer in Jordan.

1. Introduction

Cancer is a disease with an abnormally cell growth, its different types are classified by the type of the initially affected cell, it harms humans when damaged cells divide uncontrollably and it generally forms tumors (Argiris, 2012; Kennedy *et al*., 2000). Tumors grow and interfere with human systems and release hormones that alter body functions. Lung cancer starts in the cells lining the bronchi, bronchioles or alveoli, it starts as areas of precancerous changes that makes the human body cells to grow abnormally, the cell abnormal growth may look a bit abnormal if seen under a microscope, however, it does not form a tumor at this stage and does not cause any symptom. Over time, the abnormal cells progress to true cancer as more genes are changed. Cancer may enable the abnormal cell growth to form a tumor that can be detected on imaging tests. In later stages, cancer cells may spread to other parts of the human body. Most lung cancer cases are not detected until causing symptoms. There are many common symptoms of lung cancer, such as coughing up blood, chest pain, weight loss and loss of appetite, shortness of breath and feeling weak. It is known that tumors (Argiris, 2012; Kennedy *et al*., 2000) are benign or malignant, benign tumors are less than 3 mm, and are curable cancer cases, malignant tumors are greater than 3 mm, and are uncontrollable.

Lung cancer (Miao *et al*., 2016; Howlader *et al*., 2017) is one of the main causes of cancer mortality in many countries, including the United States of America (USA), lung cancer causes more deaths than a combined of breast, prostate, and colorectal cancers, Based on the number of deaths in 2010-2014, lung cancer was the leading cause of death in USA and it was more common for middle and old people, in 2015, 27% of the American deaths were lung cancer patients, lung cancer can be successfully treated at early detected stages, because its symptoms are not noticed until it is at incurable phase, more than half of patients with lung cancer cannot survive more than a year of being diagnosed; moreover, its five-year survival rate is much lower than other cancers. Old attempts, such as chest X-ray that were used to monitor lung cancers cannot provide clinically satisfactory detection of lung cancer. Therefore, a more accurate prescreening method are required (Miao *et al*., 2016). CT imaging (Al Mohammad *et al*., 2017) is among the effective methods to detect lung cancer, as it is able to measure nodule sizes, track the growth of nodules,

* Corresponding author. e-mail: wadood@fet.edu.jo.

support the characterization of morphological lesion and visualize axial sections of chest. The main disadvantage of CT is its radiation that may increase the cause or development of cancer, as a result, it is recommended to use the least radiation dose that assures acceptable quality images. In 2017, the estimated number of new lung cancer cases (Howlader *et al*., 2017) will be 222,500 and the expected number of deaths will be 155,870.

In 2010, the number of cancer cases in Jordan (Tarawneh *et al*., 2010; Al-Sayaideh *et al*., 2012) has increased to 4921, lung cancer cases (males and females) were among the top five cancer cases 380 (7.8%); moreover, its cases were also among the top five cancer cases among Jordanian males 311 (13.3%) and it was a principal cause of death (30.2%) (Tarawneh *et al*., 2010; Al-Sayaideh *et al*., 2012). In 2010, colorectal, lung and Hodgkin lymphoma were the commonest cancer cases for Jordanian males and females, however, colorectal and lung were the most common cancer cases for Jordanian males (age 20-49) and it was the commonest for Jordanian males (age 50+) (Tarawneh *et al*., 2010; Al-Sayaideh *et al*., 2012). In 2010, it is clear that lung cancer was one of the types of cancer cases that most common in males than females, as it affected 311 (13.3%) Jordanian males and 68 (2.7%) Jordanian females (Tarawneh *et al*., 2010; Al-Sayaideh *et al*., 2012).

As mentioned before, it is very important to detect cancer in its early stage to determine the optimal treatment decisions, this treatment may strongly influence the survival opportunity, it is reported that only 15.9% of the lung cancer patients are diagnosed at early stage and their 5-year survival is 55.6% (Howlader *et al*., 2017).

CAD is one of the software technologies to assist radiologists to detect lung nodules efficiently, any CAD cancer detection system (Firmino *et al*., 2015) starts with a preprocessing phase, in which, it converts CT images to grayscale, filters noise if dealing with noisy CT images, converts the grayscale images to binary images using thresholding technique, applies some boundary tracing technique and morphological operation to get rid of unimportant areas of the CT images, and it may apply some smoothing algorithm. In the feature extraction phase, the CAD system minimizes the size of the processed data, and finally, in the classification phase, it classifies / detects lung cancer by implementing data mining or classification methods.

Despite the significant strides to treat cancer cases, lung cancer remains among the unsolved clinical cancer problems. It remains difficult to detect lung cancer in early stage (Argiris, 2012; Kennedy *et al*., 2000). Optimistically, early detection of lung cancer may maximize the treatment opportunity. It is useful to use an invasive biopsy to confirm the detection of lung tumors using imaging techniques, such as CT that are useful in cancer detection. CT is a technique (Herman, 2009; Mayo Clinic Staff, 2017) to better visualize cancer, unfortunately, CT detects lung cancer in late stages, where the chance of survival becomes low. Thus, there is a demand for sophisticated technologies that detect lung cancer in early stages.

The present paper presents a CAD system that detects lung cancer, determines the existence of cancer in CT images for Jordanian patients, provides a computer based second opinion to assist CT interpretation and uses ML NNs classifier with a ICA feature extraction technique (Guyon *et al*., 2006) to speed the detection process by minimizing the number of features. The present paper compares the performance of 11 training algorithm (Bishop, 1995): traingd, traingdm, traingdx, trainrp, traincgf, traincgp, traincgb, trainscg, trainbfg, trainoss and trainlm, and it fills the gap of investigating the performance of lung cancer detection using ML NNs classifier that is trained using 11 different training methods; moreover, it investigates the performance of these training algorithms with ICA feature extraction method in detecting lung cancer in real CT images. Moreover, the present paper is expected to finally promote an ML NN training algorithm that works best with ICA in detecting lung cancer in CT images.

The rest of the present paper is organized as follows: Section 2 presents some of the related previous work, Section 3 presents the related materials and methods, Section 4 presents the results and the discussion and, finally, the Conclusion is presented in Section 5.

2. Previous Work

In Tao *et al*. (2011), an effective screening method for lung cancer using a radial basis NNs is proposed. In Wu *et al*. (2011), many distinct tumor marker groups are combined using NNs to achieve a 92.8% accuracy in lung cancer detection. In Flores-Fernández *et al*. (2012), NNs and Principal Component Analysis (PCA) are used in lung cancer detection and achieved 90% accuracy by evaluating serum biomarkers levels in lung cancer patients. In Taher *et al*. (2012), lung cancer is diagnosed in early stages using Hopfield NNs and a Fuzzy C-Mean clustering algorithm. In (Abdulla and Shaharum, 2012), NNs classifier is used in detecting lung cancer and achieved 90% accuracy. In Sun *et al*. (2013), many machine learning methods used in diagnosing lung cancer in CT images and support vector machine classifier is recommended for this purpose. In Tariq *et al*. (2013), a neuro fuzzy classifier is proposed to detect lung nodules in CT images, and, lung nodules are classified based on properties, such as area, mean, standard deviation, energy, entropy, and eccentricity. In Chen and Suzuki (2013), virtual dual energy chest radiographs images are incorporated to develop an improved detection scheme for lung cancer using NNs; the sensitivity of the proposed detection scheme is substantially improved especially for subtle nodules. In Kuruvilla and Gunavathi (2014), a new proposed training algorithm is used with back propagation NNs, used some common statistical parameters, such as mean, standard deviation, etc. to detect lung cancer in CT images, and, achieved 93.3% accuracy. In Firmino *et al*. (2016), a detection and a diagnosis system for pulmonary nodules on CT images is proposed using watershed and histogram of oriented gradient techniques to recognize nodules, the diagnosis is based on the likelihood of malignancy. Moreover, the proposed systems deployed support vector machine and rule based classifiers. In Syed and Muhammad (2017), some effectiveness cancer detection algorithms for lung cancer, a survey of nodule detection methods, and, a range of feature extraction, classification, and segmentation algorithms are presented. Furthermore, a set of performance measures are also evaluated, so as to provide as insight into the current advancements in CAD. In Shen *et al.* (2017), the classification of lung nodule malignancy suspiciousness using thoracic CT images is investigated; this investigation used multi-crop convolutional NNs (MC-CNN) in extracting nodule salient information. The proposed classification method achieved accurate classification which was potentially helpful in modeling nodule malignancy. In Tajbakhsh and Suzuki (2017), two dominant classes of end-to-end learning machines are Massive-Training Artificial NNs (MTANNs) and convolutional NNs (CNNs) are compared experimentally and theoretically in the detection of lung nodules and in the distinction between benign and malignant lung nodules in CT images. In addition, it is concluded that MTANNs works better than CNNs especially when dealing with small training datasets. In Manikandan and Bharathi (2017), a hybrid neuro- fuzzy system is proposed to detect lung cancer stages, the algorithm is designed based on the observed symptom values and a classification accuracy of 97.7% is achieved. In Froz *et al*. (2017), a lung nodule classification algorithm, using texture features from CT images, is proposed, the texture features are extracted using artificial crawlers, rose diagram, and, a hybrid model of the artificial crawlers and the rose diagram. In Dimililer *et al*. (2017), an image processing based detection technique for lung cancer is proposed and applied on CT images, these techniques include erosion, filtering, thresholding and feature extraction. Many other NNs classifiers (Ahmad, 2013; Babu *et al*., 2013) are used to detect other cancer cases, such as breast cancer. Moreover, ICA (Nguyen and Dang, 2015) is applied to separate the ribs and other parts in lung images to diagnose lung cancer at early phase resulting in 90% of cases that the ribs are completely and partly suppressed and in 85% of cases increases the nodule visibility.

3. Materials and Methods

3.1. Neural Networks (NNs)

NNs (Bishop, 1995; Haykin, 1999) are efficient processing systems that consist of many neurons. Neurons are highly interconnected processing elements that are connected with each other by links. Links are associated with weights that contain information of input signals. Information signals are used by neuronal networks to solve problems. Neurons have their own activation levels which are the function of the inputs the neurons receive. Gaussian, Linear and Sigmoid are examples of the commonly used activation functions. Learning in NNs is classified into supervised, unsupervised and reinforcement. Consequently, neural network models are specified by interconnections, learning rules to update weights and activation functions. The common neural network architectures are single layer or ML feed forward NNs. Models of NNs differ in architecture, behavior and learning methods; hence they are used to solve different problems, such as character recognition, image compression, pattern recognition and signal processing. Supervised learning is commonly used to train NNs, given input-desired outputs pairs; the NNs process inputs and finally compare the given desired output against the resulted real outputs, errors are propagated back through the network layers are used to adjust weights, this learning process is repeated over and over until achieving high quality results.

In the neural network learning process, it is important to (Haykin, 1999): (i) Pre-process the training examples, and (ii) adjust the architecture of the NNs by tuning the number of layers in the network model, the number of neurons in the input, hidden and output layers, the weights, the transfer function and the training function. In NNs, many training algorithms are used, such as the eleven training methods discussed in the present paper.

3.2. Multi-Layer (ML), Neural Networks (NNs)

ML NNs (Bishop, 1995; Haykin, 1999; Svozil *et al*., 1997), also known as ML NNs (see Figure 1), consist of input layer, output layer and one hidden layer. Noting that each layer has a specific number of neurons. The flow of data inputs starts from input layer, through the hidden layer and finally to output layer. This flow enables the ML to model arbitrary complex functions. Inputs are multiplied by their corresponding weights; then the activation function manipulates the multiplied results and produces the outputs of neurons in the hidden layer. Similarly, the outputs of neurons in the output layer are produced. Sigmoid, hyperbolic tangent, piecewise linear or threshold functions are examples of activation functions in ML networks. An ML operates in training and testing modes: Training aims to minimize the error difference between real outputs of ML and the desired outputs by adjusting weights.

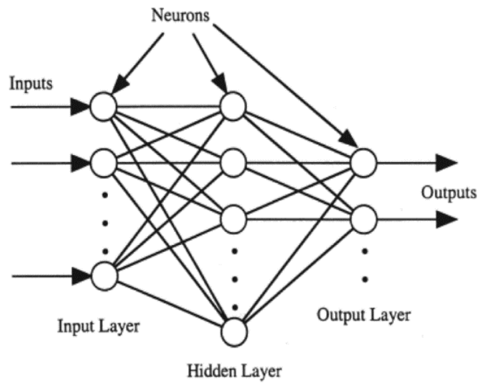

Figure 1. An ML NNs

When the training is completed, the weights of ML NNs are saved in trained states and new (unseen) input data can be presented to the trained ML NNs to determine the appropriate output. It is proved that ML NNs are able to solve larger learning problems; however, more computational efforts are needed to find the correct combination of weights. Back propagation (BP) is a common learning algorithm method that handles many large learning problems, it begins with feeding input data to the neurons in the input layer, passing outputs to neurons in the hidden layer and finally flowing NNs outputs to output layer. BP uses Equation 1 to multiply the neuron's inputs and the corresponding weights for the neurons in hidden and output layers:

$$\mathbf{net}_i = \sum_{j \in A} O_j W_{ji}, \forall\, i: i \in B \qquad (1)$$

Noting that A represents the neurons in a layer, B represents the neurons in the next layer, O_j represents the neuron j in the output layer, W_{ij} is the weight between neurons i and j and finally BP uses sigmoid function (see Equation 2) to calculate the output of all neurons (except input neurons).

$$O_i = f(net_i) = 1/1 + e^{-net_i} \quad (2)$$

After calculating the output of neurons, BP estimates error for output neurons using Equation 3:

$$\delta_i = f(net_i).(T_i - O_i), \forall i: i \in C \quad (3)$$

where C represents the output neurons, T_i represents the target value for neuron i, $f(.)$ estimates derivative of the error function (see Equation 2) and δ_i is the error of neuron i. Similarly, The BP algorithm estimates error for neurons in all hidden layers using Equation 4:

$$\delta_i = \acute{f}(net_i).\sum_{j \in E} \delta_j W_{ij}, \forall i, j: i \in A, j \in D \quad (4)$$

where D represents the neurons of the hidden or output layers and E represents the neurons in the next layer.

Then the BP updates the weights of neurons using Equation 5:

$$W_{ij}(t+1) = W_{ij}(t) + \mu.\Delta W_{ij}, \forall i, j: i \in A, j \in B \quad (5)$$

where $\Delta W_{ii} = O_i \delta_i$ and μ is the learning rate that determines the amount of weight adjustments in each iteration.

3.3. Training Algorithms for ML NNs

In the present paper, eleven training algorithms (Bishop, 1995) (traingd, traingdm, traingdx, trainrp, traincgf, traincgp, traincgb, trainscg, trainbfg, trainoss and trainlm) are used to train ML NNs classifier.

Gradient Descent Algorithm (traingd): In *traingd* (Ramesh *et al*., 2008), the NN training starts with random weights, then iteratively updates weights moving shortly toward the direction of the negative gradient, as shown in equation 6:

$$\Delta w(t) = -\mu \nabla E|_{w(t)} \quad (6)$$

where $\nabla E|_{w(t)}$ is the error minima and μ is the learning rate.

Gradient Descent with Momentum (traingdm): *traingdm* adds a new term known as momentum to the *traingd* training algorithm (Ramesh *et al*., 2008) As a result, it smooths the oscillations of gradient of error minima in the weight space. Accordingly, equations 7 shows the *traingdm*:

$$\Delta w(t) = -\mu \nabla E|_{w(t)} + \propto \Delta w(t-1) \quad (7)$$

where $\propto$ is the momentum term.

Gradient Descent with Variable Learning Rate and Momentum (traingdx): While updating the weights in the NNs, *traingdx* (Hagan et al., 1996) adjusts the training variables according to *traingdm*; however, it changes μ to increase the NNs performance. Initially, it calculates the error of the NNs, then, it calculates the modified weights using the current μ, then it re-calculates error again. Finally, it adjusts the training variables according to the *traingdm*. It is clear that *traingdx* discards the updated weights if the new error is bigger than the previous one.

Resilient Back Propagation (trainrp): trainrp (Riedmiller and Braun, 1993) updates weights and biases of NNs by investigating the sign of the partial derivative, it keeps the same change of weight in case of zero partial derivative, it reduces the change of weight in case of weight oscillation and it increases the weight magnitude in the case of continues change in weights.

Conjugate Gradient Algorithms (CGAs): In CGAs (Charalambous, 1992), a search is accomplished in the direction of conjugate gradient so as to minimize the error minima. Moreover, in CGAs, μ determines the length of weight change where the steepest descent direction (SDD) is searched initially. Noting that P_o is the initial search gradient and g_o is the initial gradient, then the following equation 8 is valid:

$$P_o = -g_o \quad (8)$$

Accordingly, given the current weight X_k, the (k)th iteration computes the current search direction P_k from which a new estimate for the next weight vector X_{k+1} (see equation 9):

$$X_{k+1} = X_k + \mu_k P_k \quad (9)$$

where μ_k is the k-th learning rate. It is clear that the direction of the (k+1)th search is conjugate to the direction of the (k)th search as in equation 10:

$$P_k = -g_k + \beta_k P_{k-1} \quad (10)$$

where P_{k-1} is the earlier search direction and β_k is a positive scalar. It should be noted that computation manner of β_k determines the flavor of CGAs.

Fletcher-Reeves Update (traincgf): traincgf obtains β_k by the following equation 11:

$$\beta_k = g_k^{\prime} g_k / g_{k-t}^{\prime}\, g_{k-1} \quad (11)$$

Polak and Ribiere (traincgp): *traincgp* obtains β_k by the following equation 12:

$$\beta_k = \Delta g_k^{\prime} g_k / g_{k-t}^{\prime}\, g_{k-1} \quad (12)$$

Powell and Beale Restarts (traincgb): In CGAs (Charalambous, 1992), the direction of search is continuously restarted to ensure better performance. *traincgf* and *traincgp* use SDD to restart, *traincgb* is restarted and the direction of search is reset to the negative of the gradient if the following equation 13 is satisfied (Moller, 1993):

$$|g_{k-1}^T g_k| \geq 0.2 \|g_k\|^2 \quad (13)$$

Scaled Conjugate Gradient Algorithm (trainscg): At each iteration, CGAs need line search that requires many calculations related to global error that may raise the complexity of the line search. As a result, *trainscg* combines the model-trust region approach with the conjugate approach to enhance the complexity of the line search.

Quasi Newton BFGS (trainbfg): Quasi Newton BFGS updates weights of NNs according to equation 14:

$$W_{k+1} = W_k - A_k^{-1} g_k \quad (14)$$

where A_k is the Hessian matrix (H) (second derivatives) of the performance index at the current values of the NN weights and biases. Instead of calculating the Hessian matrix, Quasi Newton algorithm updates an approximation of H matrix at each iteration of the algorithm and computes this update as a function of the gradient.

One Step Secant Algorithm (trainoss): *trainoss* (Battiti, 1992) updates the weights of NNs and the related bias values without storing the complete H matrix. *trainoss* assumes that the previous H matrix was the identity matrix

(I) and it calculates the new search direction without computing a matrix inverse.

Levenberg–Marquardt (trainlm): trainlm (Hagan and Menhaj, 1994) updates the weights of the NNs according to a standard nonlinear least squares optimization algorithm which is fast enough. However, its optimization algorithm requires more memory than the other training algorithms. Moreover, *trainlm* approximates H using Jacobian matrix J which includes the derivative of the NNs error e as in $H = J^T J$, The gradient can be computed as in $g = J^T e$. it should be noted that calculation complexity of g is much less than the complexity of computing H. As a result, *trainlm* approximates H as in equation 15:

$$X_{k+1} = X_k - (J^T J + \mu I)^{-1} J^T e \tag{15}$$

3.4. Independent Component Analysis (ICA)

ICA and PCA (Smith, 1992; Cao and Chong, 2002; Hyvarinen, 1999) techniques are used for feature extraction. Originally, ICA is used in blind source separation (BSS) (Cao and Chong, 2002; Hyvarinen, 1999). It recovers the unknown sources of the original input signals from their linear mixtures. Assume that x_t is the original linear mixtures and s_t is the original source input signal, ICA estimates s_t using the following equation 16:

$$s_t = U x_t \tag{16}$$

where U is the $m \times m$ un-mixing matrix. To identify $s_t = U x_t$, it is required that all the components of s_t must be independent and non-Gaussian. There are many approaches for implementing ICA. However, Fixed-Point-Fast ICA is among the best algorithms. In Fast ICA, the Negentropy of s_t is maximized to estimate s_t using the following Contrast Function (CF) equation 17:

$$J_G(u_i) = [E\{G(u_i^T x_t)\} - E\{G(v)\}]^2 \tag{17}$$

where u_i is an m dimensional vector that comprises one of the rows in matrix U. v is a Gaussian variable. G is a non-quadratic function. Equations 18, 19 and 20 show the many used functions for G:

$$G_1(s) = \frac{1}{\gamma_1} \log\cosh(\gamma_1 s) \tag{18}$$

$$G_2(s) = -\frac{1}{\gamma_2} \exp(-\gamma_2 s^2/2) \tag{19}$$

$$G_3(s) = \frac{1}{4} s^4 \tag{20}$$

where γ_1 and γ_2 are parameters with $1 \leq \gamma_1 \leq 2$ and $\gamma_2 \approx 1$. Maximizing $J_G(u_i)$ leads to estimating u_i by equations 21 and 22:

$$u_i^+ = E\{x_t g(u_i^T x_t)\} - E\{\dot{g}(u_i x_i)\} u_i \tag{21}$$

and

$$u_i^* = u_i^+ / \|u_i^+\| \tag{22}$$

where u_i^* is an estimation of u_I, g is the first derivative of G and $\dot{g}$ is the second derivative of G. Based on negentropy, the whole matrix U is computed by maximizing the sum of one-unit CF taking into account the constraint of decorrelation. In Fast ICA (Hyvarinen, 1999), x_t is centered by subtracting its mean. PCA (Smith, 1992) is used to achieve whitening and to reduce the dimension of x_t. this minimizes the number of components of s_t. Centering and whitening preprocessing steps are generally used to make the data suitable for the ICA feature extraction process, to speed the ICA convergence, to have better stability properties for the ICA implementation and to find the orthogonal de-mixing matrix which facilitate the implementation of Fast ICA algorithm. Moreover, these preprocessing steps are beneficial to reduce dimensionality (decrease the complexity) of the processed mixtures and to remove the second-order dependencies between the observed input data images or signals.

It should be noted that PCA and ICA feature extraction methods (Hyvarinen *et al*., 2001) are to reduce the number of features. However, when implementing PCA, the first and second moments of the training data sets are utilized to fit Gaussian distribution, on the other hand, higher moments of the training data sets are employed using ICA to serve a wider range of analysis, especially, when dealing with non-Gaussian noisy data. In the present paper, CT images are not indeed Gaussian.

3.5. Dataset

To create successful CAD detection or diagnosis systems, researchers always need a reference standard dataset that is used for training, and testing their systems. Lung Image Database Consortium (LIDC)[1] (Armato *et al*., 2011; Armato *et al*., 2007) is the standard database for lung cancer, LIDC contains CT images for 1010 patients along with their expert cancer annotations, LIDC provides information about the ratings of cancer nodules for only 157 patients, these tumors are rated as 0 for unknnown,1 for benign, 2 for primary malignant, and, 3 for malignant cancer. These ratings are achieved after performing biopsy, surgical resection, progression and reviewing of the radiological CT images, these ratings show the state of nodules at the patient and the nodule levels, this trusted LIDC dataset is acquired for the mentioned patients over a long period using different scanners.

In the present paper, another real (national) dataset is collected from three different Jordanian hospitals: King Hussein Cancer Center, Prince Hamzah Hospital and Al Karak Hospital. At the time of data collection, the average age of the patients was 54 years, the youngest patient was 32 years, the oldest was 77 years, 85% of them are men and the remaining are women. The sizes of their nodules vary from 3 mm to 7 mm. The CT images are labeled as normal or abnormal by experts in the mentioned hospitals.

In the present paper, a dataset of 460 CT images is used to evaluate the performance of the ML NNs detection classifier trained by 11 training algorithms with ICA:

- Real dataset: 350 of the 460 CT images belong to the local dataset: King Hussein Cancer Center (100 CT images are normal, 70 CT images are abnormal), Prince Hamzah Hospital (50 CT images are normal, 40 CT images are abnormal) and Al Karak Hospital (50 CT images are normal, 40 CT images are abnormal).
- LIDC dataset: 110 of the 460 CT images belong to LIDC lung cancer dataset (60 CT images are normal; 50 CT images are abnormal).

[1] Lung Image Database Consortium Website: http://ncia.nci.nih.gov

It is noted that 260 of the 460 CT images are normal (200 CT images of the real dataset and 60 CT images from the LIDC dataset) while the remaining 200 CT images are infected by cancer (140 CT images of the real dataset and 60 CT images from the LIDC dataset). Figures (2) and (3) show examples of the normal and the abnormal lung images.

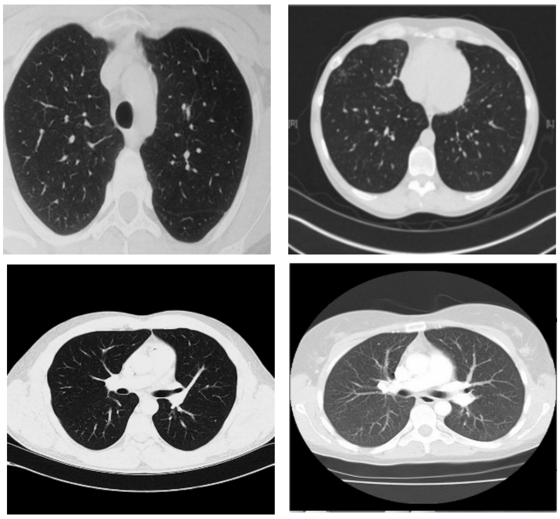

Figure 2. Normal lung CT images

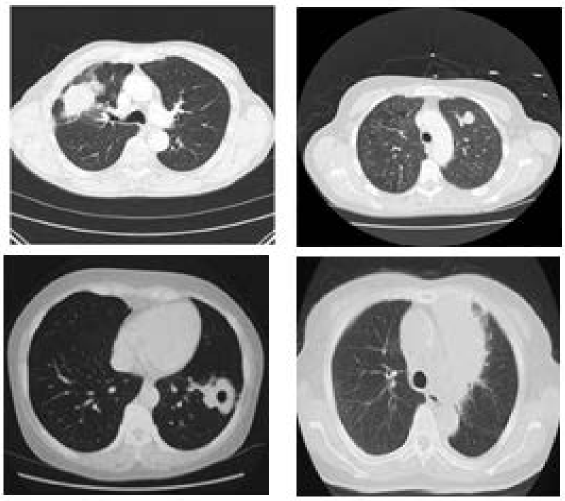

Figure 3. Abnormal lung CT images

3.6. Lung Cancer Detection Algorithm Using ML NNs with ICA

In the medical domain (Suzuki, 2011; Amato *et al*., 2013; Yasmin *et al*., 2013), NNs have been applied in image analysis and interpretation. The proposed lung cancer detection algorithm using ML NNs with ICA is an attempt to solve some of the several challenges that are facing computer aided systems for lung cancer (see El-Baz *et al*., 2013) for challenges discussion).

The proposed algorithm follows the following fundamental steps as shown in Figure (4).

Step 1: Data Collection: in this step, 350 CT lung images were collected for normal and affected lungs from some Jordanian hospitals to train and to test the detection algorithm. Moreover, 110 CT images were selected from the LIDC dataset.

Step 2: Data Preprocessing and Feature Extraction: As a matter of fact, lung images can be directly fed to the input layer of the NNs classifier, but this may negatively affect the complexity of NNs computation. However, to train the classifier more efficiently, data preprocessing procedures are conducted. Data preprocessing is always good since using data with large and small magnitudes may confuse the classification algorithm. In general, when dealing with high dimensional problems, learning algorithms have not been very effective, as a result, the machine learning community has developed many methods (such as ICA) to reduce data dimensions and accordingly to speed the classification process.

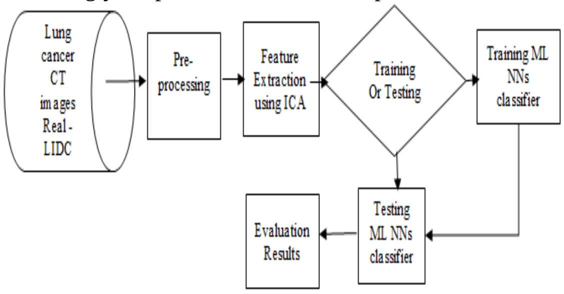

Figure 4. The lung cancer detection algorithm

In the present paper, Otsu method (Otsu, 1979) is used to convert the gray CT images of lungs with a cancerous region to binary images, Otsu method replaced the pixels whose intensities are greater than a threshold level with value ‘1’and replaced the pixels whose intensities are less than the threshold with value ‘0’. Otsu method uses the threshold that minimizes the intra-class variance of the black and white pixels. Morphological opening operation uses aperiodic line structuring element to segment lungs in binary images (Gonzalez and Woods, 2008). In morphological operations, structuring is a shape that is used to draw a conclusion on how this shape fits or misses the shapes in the original CT images. In this work, the structuring element contains 2 × 3 members. After the morphological opening operation, the image is inverted and image structures that are lighter than their surroundings and that are connected to the border of the image are suppressed. This high speed and simple segmentation process segmented the CT images correctly. Extracted features for the CT images are represented in a two-dimensional matrix, the number of rows represents the total number of selected features for each CT image and the number of columns represents the total number of CT images. Finally, ICA is used to reduce data dimensions to speed the classification process and to make the detection algorithm applicable in some of the mentioned Jordanian hospitals.

Step 3: NN Classifier Building: in this step, a ML NNs classifier is used to detect the existence of cancer in CT images for real Jordanian patients. Before deploying NNs for lung cancer detection, a parameter selection is involved to select the parameters of the NNs classifiers, these parameters are often selected by some search algorithm or optimized by a tool. Some of these parameters include those which specify the architecture of the NNs itself and some of them include those which determine the behavior, training and testing of the NNs. The number of layers and the number of neurons in the input and the hidden layers are the most important parameters that affect the overall performance of the presented NN detection algorithm. Researchers in machine learning community (Bishop, 1995; Haykin, 2008) believe that optimizing these parameters may lead to a better performance as more neurons and more layers demand more computations. It is also crucial to determine the parameters that affect the performance of detection algorithms, these parameters include the division of collected CT images into training, validation and testing subsets. As the NNs over-fits data, the validation set error rises. After a predefined number of iterations and when the validation error reaches its minimum value, the training is stopped, and the weights and biases of the epoch, with minimum validation errors are returned as the final NNs structure. In the present paper, a Grid Search (GS) algorithm (Snoek *et al*., 2012) is used for parameter selection. GS is among the well-known parameter selection methods for NNs. GS tries every parameter value over a specified range of values, GS involves an expensive computational cost, fortunately, GS is often parallelized. A user-driven refinement is typically used when searching a large scale of parameter values followed by a fine-tuned or a fine-grained search to achieve better parameter selection. GS parameter selection process used 100 randomly selected CT images in tuning NNs parameters. GS used the Root of Mean Square Error (RMSE) as a performance criterion for the ML NNs detection algorithm. The tuned parameters of the NNs detection algorithm include the number of layers, number of neurons in each layer, the learning rate, momentum term, etc. However, the output layer contains one single neuron.

Step 4: NN Training: in this step, the weights are adjusted so as to minimize the RMSE. Different number of ICA features are evaluated using the above mentioned eleven training algorithms; the number of these features is varied from 1 to 15. At the end of this step, the architecture of the NNs classifier and the other parameters are determined accurately.

Step 5: Testing the Detection Algorithm: in this step, unseen data are exposed to the trained NNs classifier and the performance is evaluated.

4. Results and Discussion

In the present paper, all the experiments are conducted on an Intel Pentium (R) personnel computer with 2.00 GHz Dual-Core CPU and 2 GB RAM, running MS Windows 10 operating system. The lung cancer detection algorithm is programmed using MATLAB version 7.11.0 (R2015a).

The detection algorithm reads the CT images, resizes them to (512 X 512) pixels, converts them to gray images, and, then, converts them to vectors. To prevent overfitting (where the training examples are memorized by the NNs, but the NNs classifier is not able to generalize to new-unseen examples), a validation set of 100 CT images are used to conduct a parameter selection using GS method for the NNs classifier, 50 of them are randomly selected from the real subset and the remaining 50 images are randomly selected from the LIDC subset. This validation set monitors the RMSE of the NNs to terminate the training to avoid overfitting. Finally, best parameter values are selected when achieving best classification accuracy and the least RMSE.

When stopping the training process, the weights of the NNs using the different training algorithms with the corresponding parameter settings are saved at the least error of the validation set. The examples of the training subset are randomly divided into four subsets; each subset contains 115 CT images. An evaluation process is conducted to evaluate the performance of the detection algorithm after tuning its architecture and its related parameters using the mentioned validation set. ICA is implemented as a feature extraction method for each CT image (noting that the extracted features for each CT image are represented in a two-dimensional matrix, the number of rows represents the total number of selected features and the number of columns represents the total number of CT images) and the ICA vector size (k) for each CT image is varied from 1 to 15. It should be noted that the proposed algorithm and the related processing are all implemented using a parallel MATLAB code to speed the detection process and to make the proposed algorithm more applicable for lung cancer detection. Accuracy is the primary performance measure that is used to evaluate the performance of the proposed detection algorithm, it is percentage of true results (both True Positive (TP) and True Negative (TN)) in the population as shown in equation 23:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (23)$$

where TP denotes a CT image that is infected and is classified as cancerous, False Positive (FP) denotes a CT image that is not infected and is classified as cancerous, TN denotes a CT image that is not infected and is classified as non-cancerous and False negative (FN) denotes a CT image that is infected and is classified as non-cancer.

Moreover, sensitivity and specificity are used to evaluate the performance of the NNs classifier trained by the eleven learning algorithms at different numbers of ICA features. Sensitivity is the percentage of positives which are correctly classified. On the other hand, specificity is the percentage of negatives which are correctly classified. Practically, higher sensitivity indicates the ability of detecting those individual CT images with cancer, higher specificity indicates the ability of identifying those individual CT images which actually have no cancer.

The classification accuracy, sensitivity and the specificity values for the ML NNs classifier trained by the eleven training algorithms evaluated over different number of ICA features are shown in Tables 1-3.

For each number of features, the highest classification accuracy values are marked in bold typeface. For each training algorithm, the highest classification accuracy values with the minimum number of features are underlined.

Table 1. Accuracy values for the ML NNs classifier trained by the eleven training algorithms at different numbers of ICA features

No. ICA features	Trainlm	Trainbfg	Trainrp	Trainscg	Traincgb	Traincgf	Traincgp	Trainoss	Traingd	Traingdm	Traingdx
1	89	69	90	89	73	70	75	72	87	84	**94**
2	89	91	81	89	86	79	87	**92**	88	88	86
3	**95**	92	88	**95**	86	86	87	74	84	86	80
4	91	88	92	**96**	86	77	**96**	79	84	80	93
5	**96**	92	89	87	81	83	88	78	95	80	**96**
6	**100**	92	96	85	88	82	85	94	86	92	95
7	**100**	91	88	91	92	96	77	85	95	50	94
8	**100**	83	92	96	92	92	92	83	85	86	94
9	**100**	85	96	92	96	**100**	92	85	85	91	96
10	**100**	73	96	**100**	92	**100**	89	89	96	96	92
11	**100**	**100**	96	**100**	**100**	**100**	**100**	92	92	96	96
12	**100**	**100**	96	**100**	**100**	**100**	**100**	**100**	96	96	96
13	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
14	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
15	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

Table 2. Sensitivity values for the ML NNs classifier trained by the eleven training algorithms at different numbers of ICA features

No. ICA features	Trainlm	Trainbfg	Trainrp	Trainscg	Traincgb	Traincgf	Traincgp	Trainoss	Traingd	Traingdm	Traingdx
1	**100**	84	**100**	**100**	61	58	70	**100**	83	**100**	**100**
2	**100**	90	77	88	83	**100**	81	92	**100**	**100**	**100**
3	**100**	84	84	90	83	**100**	76	50	**100**	**100**	**100**
4	**100**	90	84	91	83	57	**100**	66	57	**100**	**100**
5	**100**	**100**	**100**	76	72	90	76	60	80	**100**	**100**
6	**100**	**100**	**100**	69	83	66	69	87	46	**100**	**100**
7	**100**	**100**	**100**	81	92	**100**	53	69	53	**100**	88
8	**100**	69	**100**	91	92	83	84	69	80	**100**	91
9	**100**	69	0.83	84	**100**	**100**	84	69	61	0.90	**100**
10	**100**	46	**100**	100	84	**100**	76	76	76	**100**	**100**
11	**100**	**100**	**100**	92	**100**	**100**	**100**	84	84	**100**	**100**
12	**100**	**100**	**100**	**100**	**100**	**100**	**100**	92	92	**100**	92
13	**100**	92	**100**	**100**	**100**	**100**	**100**	**100**	83	**100**	**100**
14	**100**	**100**	**100**	**100**	**100**	58	70	**100**	**100**	**100**	**100**
15	**100**	**100**	**100**	**100**	**100**	**100**	81	92	**100**	**100**	**100**

In Table 1, it is clear that increasing the number of ICA features does positively influence the classification accuracy, best accuracy values are achieved by Trainlm training algorithm with 6 features, by *Traincgf* training algorithm with 9 features, by *Trainscg* training algorithm with 10 features, by *Trainbfg*, *Traincgb* and *Traincgp* training algorithms with 11 features, by *Trainoss* training algorithm with 12 features and by *Trainrp*, *Traingd*, *Traingdm* and *Traingdx* training algorithms with 13 features. It is also clear that increasing the features more than 13 does not positively increase the accuracy (see Table 1).

It is known that the sensitivity for cancer detection (Brown, *et al*., 2003) decreases when the size of cancer nodules decreases, for example, cancer detection sensitivity of 91-100% can be achieved when the nodules are bigger than 3 mm in diameter, whereas the sensitivity of detection dropped to 70% when size of nodules are less than 3 mm. A similar detection sensitivity of 91% is achieved for nodules bigger than 3 mm and a sensitivity of 86% for nodules less than 3 mm (Ko *et al*., 2001). Similarly, sensitivity of 94.1% was achieved for solid nodules bigger than 10 mm (Setio *et al*., 2015). Conversely, a good performance is achieved over different sizes of cancer nodules (Marten *et al*., 2005). In 2010, a comparative study of six different CAD systems using the same data set, five of them achieved better detection sensitivity for smaller nodules, the outcome of the present study contradicts the performance achieved in (Ko *et al*., 2001; Brown *et al*., 2003; Marten *et al*., 2005; Setio *et al*., 2015).

In the present work, the best sensitivity values (see Table 2) are achieved by trainlm and traingdm training algorithms. However, *trainoss, traincgp* and *traingd* training algorithms have not achieved good sensitivity values. Other training algorithms achieved in between sensitivity values. Moreover, *trainlm, trainrp, trainsgf, trainoss, traingdm* and *traingdx* training algorithms achieved best sensitivity values with minimum number of features (with a single feature).

In the present work, the best specificity values (see Table 3) are achieved by *traingdm* and *traingdx* training algorithms; moreover, *traingd*, *trainlm* and *trainrp* work well. However, *traincgb* does not achieve good specificity values. Other training algorithms achieved in between values. Moreover, *traingd*, *traingdm* and *traingdx* training algorithms achieved best specificity values with minimum number of features. After the evaluation process of the detection algorithm and the above reported results, it becomes ready for deployment in real environment, i.e., in one of the mentioned Jordanian hospitals.

It is concluded that Levenberg–Marquardt works best as it achieves the highest detection accuracy of 100% (with only 6 features), 100% specificity and 100% sensitivity. Therefore, results recommend using Levenberg–Marquardt with 6 ICA features as it achieved the best classification accuracy. It should be noted that the elapsed CPU time of training is few seconds and it is noted that the memory requirements is also small for the different training algorithms.

Table 3. Specificity values for the ML NNs classifier trained by the eleven training algorithms at different numbers of ICA features

No. ICA features	Trainlm	Trainbfg	Trainrp	Trainscg	Traincgb	Traincgf	Traincgp	Trainoss	Traingd	Traingdm	Traingdx
1	92	53	88	80	84	81	80	50	**100**	**100**	**100**
2	**100**	90	**100**	88	90	66	0.91	90	**100**	**100**	**100**
3	**100**	**100**	**100**	**100**	90	76	**100**	**100**	90	**100**	**100**
4	**100**	84	**100**	**100**	90	**100**	92	91	**100**	**100**	**100**
5	**100**	84	**100**	**100**	90	76	**100**	**100**	**100**	**100**	**100**
6	**100**	84	**100**	**100**	92	**100**	**100**	**100**	**100**	**100**	**100**
7	**100**	84	**100**	**100**	92	91	**100**	**100**	**100**	**100**	**100**
8	**100**	**100**	**100**	**100**	91	**100**	**100**	**100**	**100**	**100**	**100**
9	**100**	**100**	**100**	**100**	92	**100**	**100**	**100**	**100**	**100**	**100**
10	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
11	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
12	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
13	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
14	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
15	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

5. Conclusion

It is noted that lung cancer is one of the causes of cancer death because of its severity and its late stage when detected. In the present paper, a CAD cancer detection algorithm is implemented to predict the existence of lung cancer in CT images in early stages, the detection system is based on ML NNs classifier with ICA. ICA is to speed the detection algorithm, so as to make it applicable in one of the mentioned Jordanian hospitals, the ML NN classifier is trained by 11 training algorithms. Evaluation used two datasets: one of them is the well-known LIDC dataset, the other dataset is locally collected from three Jordanian hospitals. The paper compares the performance of the 11 training algorithms: *traingd*, *traingdm*, *traingdx*, *trainrp*, *traincgf*, *traincgp*, *traincgb*, *trainscg*, *trainbfg*, *trainoss* and *trainlm*. The present paper investigates the performance of these training algorithms with ICA feature extraction. Among the 11 training algorithms, the *trainlm* training algorithm achieves the highest detection accuracy of 100% with minimum number of features, with specificity of 100% and sensitivity of 100%. i.e., using training algorithm with ICA in ML NN classifier is the best choice to detect lung cancer in CT images collected from the mentioned Jordanian hospitals. The present paper concluded that the ICA is beneficial in cancer classification using NNs especially when dealing with real CT images as it enhances the accuracy and the speed of detection. The proposed detection system is applicable to detect lung cancer at any of the mentioned Jordanian hospitals, this detection is expected to decrease the mortality rate of lung cancer in Jordan.

Acknowledgement

The author likes to thank Al-Balqa Applied University for the financial support to conduct the present work that has been carried out during sabbatical leave granted to the author during the academic year 2015/2016 at Al-Ahliyya Amman University, Jordan. The author likes to thank the Computer Engineering Department at Al-Ahliyya Amman University for providing Plagiarism checker software that was helpful to have a high caliber paper.

References

Abdel-Razeq H, Attiga F and Mansour A. 2015. Cancer care in Jordan. *Hematology/Oncology and Stem Cell Therapy*. **8(2)**: 64-70.

Abdulla A and Shaharum S. 2012. Lung cancer cell classification method using artificial neural network. *Information Eng Let*., **2**: 50-58.

Ahmad F, Isa N, Noor M and Hussain Z. 2013. Intelligent Breast Cancer Diagnosis Using Hybrid GA-ANN. Proceedings of the 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks. 9-12.

Al Mohammad B, Brennan P and Mello-Thoms C. 2017. A review of lung cancer screening and the role of computer-aided detection. *Clin Radiol*. **72(6)**: 433-442.

Al Qadire M. 2014. Jordanian cancer patients' information needs and information-seeking behaviour: A descriptive study. *Europ J of Oncology Nursing*. **18(1)**: 46-51.

Al-Sayaideh A, Nimri O, Arqoub K, Al- Zaghal M and Halasa W. 2012. **Annual statistical book**. Directory of Information & Research, Jordanian Ministry of Health.

Amato F, López A, Peña-Méndez E, Vaňhara P, Hampl A and Havel J. 2013. Artificial neural networks in medical diagnosis. *J Appl Biomed*., **11**: 47-58.

Argiris, A. 2012. **Lung Cancer (Emerging Cancer Therapeutics)**. First ed., Demos Medical, New York.

Armato S, McLennan G, Bidaut L, McNitt-Gray M, Meyer C, Reeves A.*et al*., 2011. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. *Medical Phys.*, **38(2)**: 915-931.

Armato S, McNitt-Gray M, Reeves A, Meyer C, McLennan G, Aberle D. *et al*,, 2007. The Lung Image Database Consortium (LIDC): An Evaluation of Radiologist Variability in the Identification of Lung Nodules on CT Scans. *Academic Radiol*., **14(11)**: 1409-1421.

Babu G, Bhukya S and Kumar R. 2013. Feed forward network with back propagation algorithm for detection of breast cancer. Proceedings of the 2013 8th International Conference on Computer Science & Education. 181-185.

Battiti R. 1992. First and second order methods for learning: Between steepest descent and Newton’s method. *Neural Computation*. **4(2)**: 141-166.

Bishop C. 1995. **Neural networks for pattern recognition**. Oxford University Press, New York.

Brown M, Goldin J, Suh R, McNitt-Gray M, Sayre J and Aberle D. 2003. Lung micronodules: automated method for detection at thin-section CT-initial experience. *Radiol*., **226(1)**: 256-262.

Cao L and Chong W. 2002. Feature extraction in support vector machine: a comparison of PCA, KPCA and ICA, Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'OZ), 1001-1005.

Charalambous C. 1992. Conjugate gradient algorithm for efficient training of artificial neural networks. *IEEE Proceedings*. **139(3)**: 301-310.

Chen S, Suzuki K. 2013. Computerized Detection of Lung Nodules by Means of "Virtual Dual-Energy" Radiography. *IEEE Transactions on Biomed Eng*., **60(2)**: 369-378.

Dimililer K, Ugur B and Ever Y. 2017. Tumor detection on CT lung images using image enhancement. *The Online J Sci Technol*., **7(1)**: 133-138.

El-Baz A, Beache G, Gimel'farb G, Suzuki K, Okada K, Elnakib A, Soliman A and Abdollahi B. 2013. Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies. *Inter J Biomed Imaging*, Hindawi Publishing Corporation, **2013**:1-46.

Firmino M, Angelo G, Morais H, Dantas M and Valentim R. 2016. Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. *Biomed Eng*., **15(1)**: 1-17.

Firmino M, Morais A, Mendoça R, Dantas M, Hekis H and Valentim R. 2014. Computer-aided detection system for lung cancer in computed tomography scans: Review and future prospects. *Biomed Eng OnLine*. **13(41)**: 1-16.

Flores-Fernández J, Herrera-López E, Sánchez-Llamas F, Rojas-Calvillo A, Cabrera-Galeana P, Leal-Pacheco G, González-Palomar M, Femat R and Martínez-Velázquez M. 2012. Development of an optimized multi-biomarker panel for the detection of lung cancer based on principal component analysis and artificial neural network modeling. *Expert Systems with Applications*. **39(12)**: 10851-10856.

Froz B, Filho A, Silva A, Paiva A, Nunes R and Gattass M. 2017. Lung nodule classification using artificial crawlers, directional texture and support vector machine. *Expert Systems with Applications*. **69(1)**: 176-188.

Gonzalez R and Woods R. 2008. **Digital Image Processing**. 3rd Ed. Prentice Hall, New York.

Guyon I, Gunn S, Nikravesh M and Zadeh L. 2006. **Feature Extraction Foundations and Applications**. Springer-Verlag, Berlin Heidelberg.

Hagan M, Demuth H and Beale M. 1996. **Neural Network Design**. 2nd Ed., PWS Publishing, Boston, MA, USA.

Hagan M and Menhaj M. 1994. Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*. **5(6)**: 989-993.

Haykin S. 1999. **Neural Networks: A Comprehensive Foundation**. 2nd Ed., Prentice Hall, New York.

Haykin S. 2008. **Neural Networks and Learning Machines**. 3rd Ed. Pearson Education, New York.

Herman G. 2009. **Fundamentals of computerized tomography: Image reconstruction from projection**. 2nd Ed., Springer, London.

Howlader N, Noone A, Krapcho M, Miller D, Bishop K, Kosary C, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis D, Chen H, Feuer E and Cronin K. 2017. SEER Cancer Statistics Review 1975-2014. National Cancer Institute. Bethesda, MD: https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission.

Hyvarinen A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*. **10(3)**: 626-634.

Hyvarinen A, Karhunen J and Oja E. 2001. **Independent Component Analysis**. John Wiley & Sons, INC., New York.

Kennedy T, Miller Y and Prindiville S. 2000. Screening for lung cancer revisited and the role of sputum cytology and fluorescence bronchoscopy in a high-risk group. *Chest J.*, **117(4)**: 72-79.

Ko J and Betke M. 2001. Chest CT: automated nodule detection and assessment of change over timed preliminary experience. *Radiol*., **218(1)**: 267-273.

Kuruvilla J and Gunavathi K. 2014. Lung cancer classification using neural networks for CT images. *Computer Methods and Programs in Biomed.*, **113(1)**: 202-209.

Manikandan T and Bharathi N. 2017. Hybrid neuro-fuzzy system for prediction of stages of lung cancer based on the observed symptom values. *Biomed Res.*, **28(2)**: 588-593.

Marten K, Engelke C, Seyfarth T, Grillhösl A, Obenauer S and Rummeny E. 2005. Computer-aided detection of pulmonary nodules: influence of nodule characteristics on detection performance. *Clin Radiol.*, **60(2)**: 196-206.

Mayo Clinic Staff. 2017. CT scan. Mayo Clinic: http://mayoclinic.org, retrieved on May 2017.

Miao Q, Derbas J, Eid A, Subramanian H and Backman V. 2016. Automated Cell Selection Using Support Vector Machine for Application to Spectral Nanocytology. *Biomed Res Inter.*, **2016**: 1-10.

Moller M. 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*. **6**: 525-533.

Nguyen H and Dang T. 2015. Ribs Suppression in Chest X-Ray Images by Using ICA Method. IFMBE Proceedings of the 5th International Conference on Biomedical Engineering in Vietnam. **46**: 194-197.

Otsu N. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. **9(1)**: 62-66.

Ramesh J, Vanathi P and Gunavathi K. 2008. Fault classification in phase-locked loops using back -propagation neural networks. *ETRI J.*, **30**: 546-553.

Riedmiller M and Braun H. 1993. A direct adaptive method for faster back propagation learning: the RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks. San Francisco, CA, USA.

Setio A, Jacobs C, Gelderblom J and Ginneken B. 2015. Automatic detection of large pulmonary solid nodules in thoracic CT images. *Medical Phys.*, **42(10)**: 5642-5653.

Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y and Tian J. 2017. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*. **61**: 663-673.

Smith L. 2002. A tutorial on Principal Components Analysis. Downloaded from:

http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

Snoek J, Larochelle H and Adams R. 2012. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*. **25**: 2960-2968.

Sun T, Wang J, Li X, Lv P, Liu F, Luo Y, Gao Q, Zhu H and Guo X. 2013. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Computer Methods and Programs in Biomed.*, **111(2)**: 519-524.

Suzuki, K. 2011. **Artificial Neural Networks Methodological Advances and Biomedical Applications**. Published by InTech Janeza, Croatia

Svozil D, Kvasnicka V and Pospichal J. 1997. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Syst.,* **39**: 43-62.

Syed M and Muhammad S. 2017. Recent Developments in Computer Aided Diagnosis for Lung Nodule Detection from CT images: A Review. *Current Med Imaging Rev.,* **13(1)**: 3-19.

Taher F, Werghi N, Al-Ahmad H and Sammouda R. 2012. Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods. *Am J Biomed Eng.,* **2(3)**: 136-142.

Tajbakhsh N and Suzuki K. 2017. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. *Pattern Recognition*. **63**: 476-486.

Tao W, Jianping L and Bingxin L. 2011. Research of Lung Cancer Screening Algorithm Based on RBF Neural Network. Proceedings of the 2011 International Conference on Computer and Management (CAMAN), Wuhan, China, 1-4.

Tarawneh M, Nimri O, Arkoob K and AL Zaghal M. 2010. **Cancer Incidence in Jordan 2010**. The fifteenth report, Jordanian Ministry of Health.

Tariq A, Akram M and Javed M. 2013. Lung nodule detection in CT images using neuro fuzzy classifier. Proceedings of the 2013 IEEE Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI). 49-53.

Wu Y, Wu Y, Wang J, Yan Z, Qu L, Xiang B and Zhang Y. 2011. An optimal tumor marker group-coupled artificial neural network for diagnosis of lung cancer. *Expert Systems with Applications.* **38(9)**: 11329-11334.

Van Ginneken B, Armato S, de Hoop B, de Vorst S, Duindam T, Niemeijer M., Murphy K, Schilham A, Retico A, Fantacci M, Camarlinghi N, Bagagli F, Gori I, Hara T, Fujita H, Gargano G, Bellotti R, Tangaro S, Bolaños L, De Carlo F, Cerello P, Cheran S, Torres E and Prokop M. 2010. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Med Image Anal.*, **14(6)**: 707-722.

Yasmin M, Sharif M and Mohsin S. 2013. Neural Networks in Medical Imaging Applications: A Survey. *World Appl Sci J.,* **22(1)**: 85-96.