http://ijea.jctjournals.com/ ISSN(ONLINE):2320 – 0804

Volume-11, Issue-11, November, 2022
International Journal of Engineering Associates (IJEA)
Page Number: 01-08

A Robust Neural Network for Lung Cancer detection Using Machine Learning Approach

Ketaki Patidar[1], Prof. Sumit Sharma[2]
Research Scholar of CSE[1], Head of Department[2]
[1,2]Department of Computer Science Engineering
[1,2] Vaishnavi Institute of Technology and Science *(VITS), Bhopal, (M.P.), INDIA*
[1]ketkipatidar17@gmail.com

Abstract—Cancer is a disease that is becoming increasingly prevalent around the world. Researchers have conducted numerous studies to determine where on the human body cancer is most common. This research work is proposed to find out the early stage of lung cancer and explore the accuracy levels of various machine learning algorithms. Due to improper handling of DICOM images, there is low accuracy and a high implementation cost. For medical image processing, many different types of images are used, but computer tomography (CT) scans are generally preferred because of less noise. The proposed approaches' performances are evaluated based on their accuracy, sensitivity, specificity, and classification time. Machine learning based lung cancer prediction models have been proposed to assist clinicians in managing incidental or screen detected indeterminate pulmonary nodules. Such systems may be able to reduce variability in nodule classification, improve decision making and ultimately reduce the number of benign nodules that are needlessly followed or worked-up.

Keywords- Pulmonary nodules; lung neplasms; lung; machine learning; decision making, Structural Co-occurrence Matrix (SCM), Classifier, Data Set, ROC curve, Malignant nodule, Benign nodule, etc.

1. INTRODUCTION

1.1 BACKGROUND

Lung cancer is diagnosed in the United States at a rate second only to that of breast cancer. Lung cancer patients have a survival rate of only 15% five years after their diagnosis. Survival analysis is a common topic in medical research. The survival rate of cancer patients can be predicted using a predictor variable that indicates whether or not certain events, such as death or recurrence of a disease, have occurred over a specified time period. A patient's prognosis after a cancer diagnosis must be predicted by the predictor models. You have two sponge-like organs in your chest, the lungs, which are responsible for breathing. He has three lobes in each of his proper lungs. The left lung has two lobes on each side. The left lung is reduced in size to compensate for the increased size of the heart. Air enters your lungs when you inhale through your nose or mouth and travels down your trachea (windpipe). The trachea in the lungs divides into several smaller bronchi.

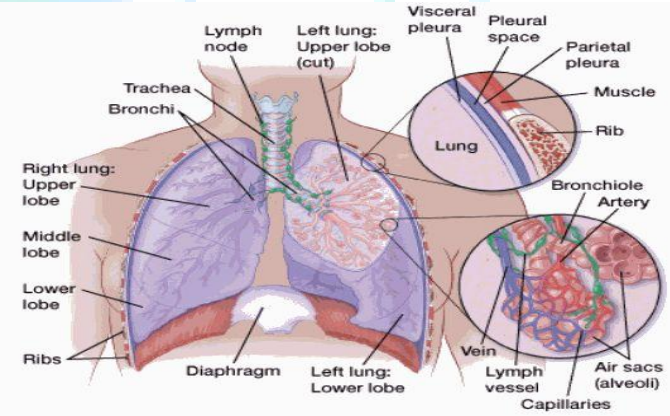

Fig. 1.1 Normal structure and function of the lungs [1]

Smaller branches, known as bronchioles, branch out from the main bronchial tree and are called bronchial branches. The bronchial tubes culminate in little air sacs called alveoli. Carbon dioxide is expelled from the blood as you inhale oxygen via the alveoli. Your lungs are responsible for taking in oxygen and exchanging carbon dioxide. The lining of the

bronchi and other parts of the lung, such as the bronchioles or alveoli, are common places for lung cancer to begin.

The lungs are protected from the outside world by the pleura, a thin layer of lining. The pleura acts as a cushion between human lungs and the chest wall, allowing them to expand as well as contract when you breathe.

This dome-shaped diaphragm divides the upper and lower torsos by forming a barrier between the two areas. When you inhale and exhale, the diaphragm rises and falls, causing the lungs to fill and empty.

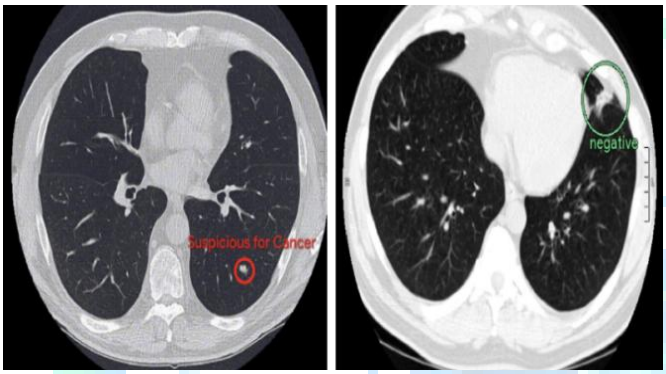

Fig. 1.2 less noisy as compared to MRI lung cancer

Lung cancer is the leading cause of death in the United States. In 2012, there were 1.6 million deaths and 1.8 million new cases reported. Among both men and women, it is the leading cause of cancer-related death in the United States, accounting for more deaths than all other cancers combined. Cancers of the lungs have a 17.8% five-year survival rate, but that number would be much higher if the disease had been discovered earlier. Only 15% of the cases were caught in the early stages, however. This shows that early detection of lung cancer is critical to the success of treatment. Lung cancer is the leading cause of cancer death in smokers. As a result, it has been suggested that cancer may be a result of an individual genetic disposition inherited from family members. In other words, some people are genetically predisposed to developing lung cancer because of genetic mutations or flaws in a gene.

1.2 Classification of Lung Cancer

Epithelial cell tumours, known as carcinomas, account for the majority of lung cancer cases. The size and presence of the malignant tumour can differentiate among non-small cell lung cancer (80%) and small cell lung cancer (20%). cells under the microscope (16.8 percent). In terms of therapy and prognosis, this categorization is quite important.

1.2.1 Non-small cell Lung cancer (NSCLC)

They form a cluster because of their similar prognosis and treatment. The three most common lung cancers are squamous cell carcinoma, adenocarcinoma, and large cell lung carcinoma. Squamous cell carcinoma is the most common form of lung cancer in the central bronchus. The tumor's core is filled with necrosis and cavitations. As a result, well-differentiated squamous cell carcinomas are exceedingly rare in comparison to other types of cancer.

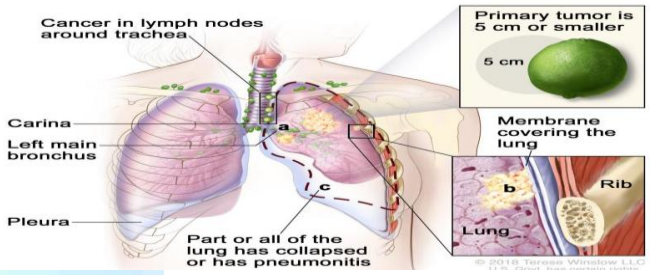

Fig. 1.3 Non-small cell Lung cancer

Non-small cell lung cancer and small cell lung cancer are the two most common forms of lung cancer (NSCLC). Both of these subtypes are given different treatments. Information on NSCLC can be found in this resource. A different resource can provide you with more information on small cell lung cancer. Lung neuroendocrine tumors are also discussed at length on this site.

Tumors, lesions, and nodules—all terms used to describe the abnormal growth of lung cells—are the earliest signs of non-small cell lung cancer. A tumor in the lung can originate in any part of the lungs. A tumor can be either malignant or non-cancerous. As a malignant lung tumor develops, it is conceivable that it may shed cancer cells. Either via the circulation or by a process known as lymphatic exfiltration, these cells may be expelled from your body. Lymph nodes gather lymph after it has travelled via lymphatic pathways.

1.2.2 Types of NSCLC

Epithelial cells are the origin of NSCLC. When it comes to diagnosing lung cancer, clinicians must distinguish between squamous cell carcinoma and other types of the disease. This data is then used to decide the best course of action for a patient.

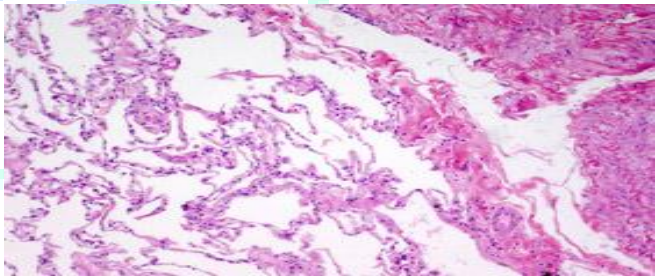

Fig. 1.4 Normal lung tissue [157].

1.3 Prevention

Lung cancer may be prevented or even cured by quitting smoking.

Smoking ban

Quitting smoking may help prevent or even cure lung cancer.This is because giving up tobacco will cut down on one's dependence on it.It has become more common in Western countries to implement measures to reduce the number of people who are exposed to secondhand smoke in public settings like restaurants and workplaces.Bhutan has had a no-smoking policy since 2005, whereas India made it illegal

to smoke in public starting in October of the same year.To keep children and teenagers from starting to smoke, the World Health Organization has urged governments to outlaw all forms of tobacco advertising.Tobacco use has dropped by 16% as a result of these regulations, according to the study's authors.

Screening

In order to find cancer, large groups of people who have no symptoms are screened. People at high risk of lung cancer can be detected early with computed tomography (CT) screening, and treatment options can be provided that extend their lives. A 0.3 percent decrease in mortality risk can be attributed to this method of lung cancer screening (a relative percentage of 20 percent). Cigarette smokers who have smoked a pack of cigarettes daily for the past 30 years, including the past 15 years, are considered to be at high risk of developing lung cancer.

II. TECHNICAL BACKGROUND

Machine Learning Artificial Intelligence In Cancer Detection

It is essential to have a solid comprehension of what machine learning is, as well as what it is not, before beginning a detailed analysis of which types of machine learning methods work best for which kinds of scenarios. Specifically, it is important to have an understanding of what machine learning is not. The field of artificial intelligence research known as machine learning employs statistical, probabilistic, and optimization techniques in order to "learn" how to classify new data, discover new patterns, or forecast new trends based on previous examples (Mitchell 1997).

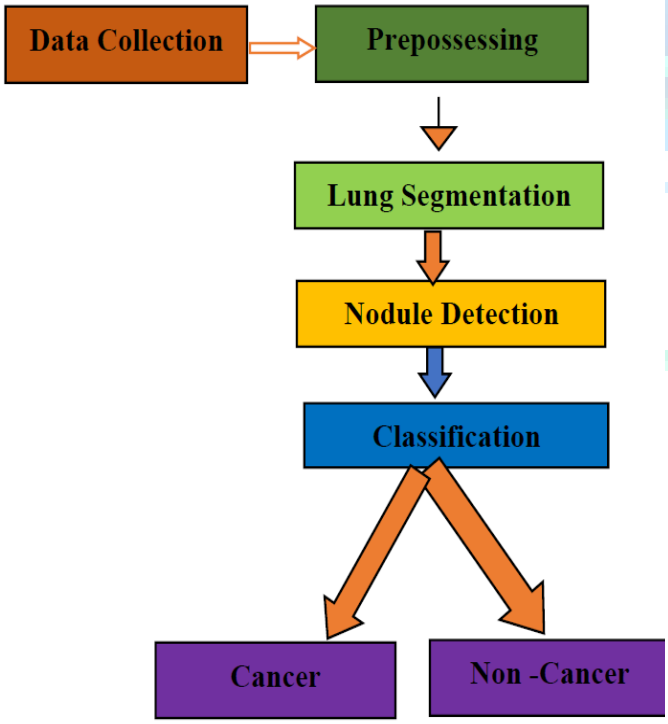

Fig.2.1 Block Diagram for Machine Learning based Lung Cancer Detection

In the beginning, natural language processing methods are used to convert the collected input text into a binary format. However, machine learning methods are then used to analyse this binary data in order to produce the appropriate output and decisions.

- The procedure for identifying variations together in tumours is going to be detailed down below.
- "Images of hearts are categorised based on the shape of the heart itself."
- The method is proposed for use in predicting issues related to the heart.
- Dermatologists would be able to make accurate diagnoses of tumours if they used artificial intelligence.
- A framework for the application of artificial intelligence in the intensive care unit.
- Identifying individuals who are at risk of developing cervical cancer using algorithms is currently being researched and developed.
- It has been shown that AI can be helpful in the diagnosis of breast cancer.

2.2 Machine Learning Steps

Deep learning algorithms, in particular CNN, Fully Connected Convolutional Networks (FCN's), and Deep Belief Networks (DBN's), had swiftly evolved techniques and strategies to study and examine/analyze the imaging in the medical area, such as MRI, X-Ray, and computed tomography (CT) images, etc., in a short amount of time. Image classification, the classification and detection of lesions, organ and lesion segmentation, enhancement, and image generation are all examples of applications for deep learning approaches. These approaches can also be used to combine image data with reports.

A. Data Prepossessing

The process of machine learning includes the integration of data preprocessing. After being processed, photographs taken at a 20 equivalent magnification (0.5 m/pixel) were used to ensure that both the global perspective and the localised details were maintained. The ASAP technology was developed by the medical professionals in order to meet the requirements of documenting the WSIs in TIFF format using distinctive coloured roughness polygons that identify a particular histological lung tissue type. Deep Neural Networks (DNN)

Our ideal framework would have been a CNN that had both a high level of accuracy and a low amount of tuning time required. The Efficient Net systems received a performance boost, which assisted them in achieving state-of-the-art accuracyperformance levels, as a result of the use of compound scaling in conjunction with auto architecture search. PictureNet has a lower number of operations per second that are performed in floating point (FLOPs).

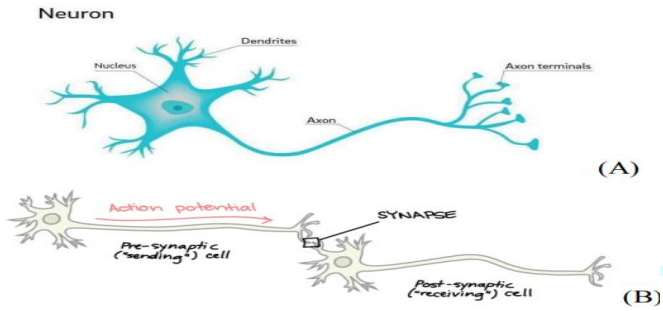

Fig. 2.2 Deep Learning Neural Networks

B. Network Training (NT)

A significant barrier to medical research is the restrictive privacy rules as well as disorganised management systems that apply to most medical materials, particularly labelled ones. Thus, because of our limited training data set, transfer learning methods were used to train the Efficient Net-B5 system.Two stages were involved in the training procedure. To begin, we used default weights from the Image Net data set to create the network, then we froze all but the final fully connected layer before training it with our own data. Second, we thawed out the frozen layers and fine-tuned the entire network to best meet the goal.

III. PROPOSED METHODOLOGYFEATURE SELECTION

3.1 COLOR AUTO CORRELOGRAM

The correlation of a signal with a delayed copy of itself is known as autocorrelation. In the case of discrete time, this type of correlation is referred to as serial correlation. In its most fundamental form, it can be understood as the correlation between observations in terms of the time lags that separate them. Analyzing autocorrelation patterns, for instance, enables one to determine the presence of a periodic signal that is obscured by noise as well as its fundamental frequency. This can be done in either of two ways. The examination of functions and sequences of data, such as time domain signals and temperature readings, is one of its many applications. Autocorrelation can be defined in a number of different ways, and no two of these definitions are identical to one another. The terms "autocorrelation" and "autocovariance" are frequently used interchangeably in the scientific community when talking about autocorrelation. There are many different kinds of autocorrelation processes; some examples include moving average processes, autoregressive processes, trend-stationary processes, and unit root processes.

Auto-correlation of stochastic processes

Real or sophisticated random processes can be described as having an autocorrelation, in which case the Pearson correlation is used to compare the process's values across time rather than as a constant. As examples, a random process and any point in time (whether an integer or a real number in continuous time) are taken as examples. The value (or realisation) that results from a particular process conducted at a given point in time. The mean and variance for each time point in the process are assumed to be constant. The auto-correlation function between times is then definedt_(1) and

t_2〖R〗_XX (t_1,t_2) = E[t_1 t_2] [3.1]

where denotes the expected value operator and the bar is an illustration of complex conjugation. Take into consideration that the expectation might not be clearly defined. The auto-covariance function between times and values is obtained by first subtracting the mean and then multiplying the values.

K_XX (t_1 t_2)=E[(Xt_1 μt_1)]=E[Xt_(1) Xt_2]-μt_1 μt_(2) [3.2]

This equation does not have a definitive answer for all time series or processes because the mean might not exist, the variance might be 0 (for a constant process) or it might be infinite (for a non-constant process), depending on the type of time series or process (for processes with distributions lacking well-behaved moments, such as certain types of power law).

If {X_t} The autocovariance function is independent of the passage of time, and it simply relies on the time elapsed between the two values to calculate the lag. Since the time lag is an even function, we may describe the autocovariance and autocorrelation as an increasing function of timeτ=t_2-t_1 This gives the more familiar forms for the auto-correlation function

R_XX (τ)=E[Xt_(+τ) X] [3.3]

and the auto-covariance function:

(K_XX (τ)=E[(X_(t+τ)-μ(X_(t)-μ)]=E[X_(t+τ) X_t]-μμ[3.4]

Auto-correlation of random vectors

The (potentially time-dependent) auto-correlation matrix (also called second moment) of a (potentially time-dependent) random vector X=(X_1,……X_n)^Tis an n × n matrix contaning as elements the autocorrelation matrix is used in various digital signal processing algorithms.

For a random vector X=(X_1,…..X_n)^T containing random random elements whose expected value and variance exist, the auto-correlation matrix is defined

R_XX=E[XX^T [3.5]

Transposition T has the dimensions n n, indicating that it is taking place. Component by component

3.2 Color Moment

As with central moments in probability distributions, colour moments serve to describe an image's colour distribution. If you're looking for photos that look similar based on colour, you can utilisecolour moments in image retrieval applications as features. To locate a similar image, one image is compared to a library of digital photos with pre-calculated attributes. A similarity score is calculated for each image comparison, and the lower the value, the closer the two photos should be Colour moments that are invariant to scale as well as rotation have been discovered.

3.2.1 Mean

It's the main article: Mean

The first color moment can be interpreted as the average color in the image, and it can be calculated by using the following formula

E_i=∑_(j=1)^N▒1/N P_ij [3.6]

where N is the number of pixels in the image and p_{ij} is the value of the j-th pixel of the image at the i-th color channel.

3.2.2 Standard Deviation

In the second colour moment, the standard deviation, the variance of the colour distribution is squared and taken as the standard deviation.

$$\sigma_i = \sqrt{(1/N \sum_{j=1}^{N} [(p_{ij}-R_i)]^3)} \quad [3.7]$$

Where E_i is the mean value, or first color moment for the i-th color channel of the image.

3.2.3 Skewness

The third colour moment is the skewness.As a result, it provides information about the shape of the colour distribution by measuring its asymmetry..

3.2.4 Kurtosis

The kurtosis of a colour distribution provides information on the shape of the distribution of colours, similar to how skewness does. On the other hand, the kurtosis is a measurement of the extremeness of the tails in relation to the normality of the distribution.

3.3 Introduction

In this chapter discuss the proposed method architecture and flow to algorithm to detect cancer tissues.

3.3.1 ANN Approach

ANN initial divides training information into many subsets, exploitation ambiguous clustering techniques. After this, it trains different ANNs by exploitation different subsets. Then it determines the membership grade of those subsets and connects them through a replacement ANN to induce the ultimate result. The whole structure of ANN is shown in figure 5.3. Within the kind of a selected machine learning framework, feed forward ANN covers each the coaching part and therefore the testing part.

Step - I - Featured Selected data training

Step II: Training for training by specific teaching algorithms for training (i = 1, 2, k), ANN model, ANN, (i = 1,2, ... k) for every training set numerous Base ANN model

Step III: to scale back the error for every ANN, we tend to simulate the ANN exploitation the complete training set TR and find the results. Then we tend to use the membership grade, which were generated by the ambiguous cluster module to combine the results.

After this, tend to train another new ANN exploitation combined results. Within the testing part, we tend to input directly the take a look at set information into numerous ANN and receive the output. Supported these outputs, final results is achieved by ultimate ambiguous aggregation module

There are three necessary lawsuits in three phases of the ANN structure-

- Produce totally different training subsets from the initial training dataset TR;
- Produce different base models ANN with different coaching subsets;
- A way to collect numerous results made by numerous base models ANN.

3.4 Forward Neural Network

A directly proportional relationship between input and output occurs in perception, whereas a connection between input and output occurs in FFNNs. The activation function in the hidden layer creates a nonlinear connection. There is a network with a direct connection between the input layer and the output layer that is formed by combining the connection form based on perception with a multilayer network. Cascade Forward Neural Networks (CFNN) are the result of this connection pattern (CFNN). The following are the equations that can be derived from the CFNN model.

$$y = \sum_{i=1}^{n} [f^i(\omega_{t^i}) \chi_i + f^0 (\sum_{j=1}^{n} \omega_j^0 f_j^h (\sum_{i=1}^{n} \omega_{ji}^h \chi_i))] \quad [3.8]$$

Where f^i is the activation function from the input layer to the output layer and ω_{ji}^h is weight from the input layer to the output layer. If a bias is added to the input layer and the activation function of each neuron in the hidden layer is then equation (5.2) becomes.

$$y = \sum_{i=1}^{n} f^i(\omega_i^i \chi_i + f^0 (\omega^b + \sum_{j=1}^{k} \omega_j^0 f^h (\omega_j^b + \sum_{i=1}^{n} \omega_{ji}^h \chi_i))) \quad [3.9]$$

In this research, the CFNN model is applied in time series data. Thereby, the neurons in the input layer are the lags of time series data Xt-1,Xt-2,...,Xt-p, whereas the output is the current data Xt . The architecture of CFNN model in predicting time series is shown at figure 5.4.

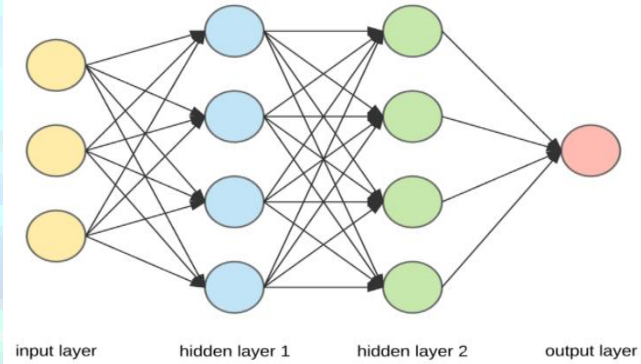

Fig 3.1 Different layers of back proportion method

3.5 Training Back Propagation

Back-propagation is a technique employed in artificial neural networks to calculate a gradient that's required within the calculation of the weights to be utilized in the network. it's ordinarily accustomed train deep neural networks, a term bearing on neural networks with quite one hidden layer. There are two main kind of neural network feed forward and back forward network. There are three main layers in neural network.

- Input Nodes – The Input nodes give data from the surface world to the network and are along noted because the “Input Layer”.
- Hidden Nodes – The hidden nodes don't have any direct association to the surface of the planet (hence the name is "hidden"). They calculate and transfer information from the input nodes within the output nodes. A group of hidden nodes creates a "hidden layer". Whereas a feed forward network can exclusively have one input layer and one output layer, it'll have zero or multiple Hidden layer.
- Output Nodes – The Output nodes are together noted because the “Output Layer” and are

accountable for computations and transferring data from the network to theoutside world.

IV. SIMULATION AND RESULTS

4.1 Introduction

The simulation model and its results will be discussed in this chapter along with the proposed algorithm. Matrix laboratory should be utilised for the process of implementing the proposed algorithm. The Matrix Laboratory is a well-known piece of software that can be used for various calculations related to the implementation of algorithms for data analysis. The number of data analysis tools available in MATLAB is quite extensive.

4.2 Simulation outcomes and Results

In the above discuss the GUI design for simulation of proposed method shows discuss the simulation outcomes of proposed method. The proposed method simulation divided into three parts. In the first part discuss the simulation outcomes on 70% training data and remaining 30% data for testing as well as validation. Similar that apply training with 80% data for testing use remaining 20% data.

Training testing outcomes on 70% data

In the training of proposed method use GUI interface with modified feed forward neural network, number of epochs ,15 and other training parameters as an input. For initialization of training process follow these steps,

1. Using browse button to select the target input folder data.
2. After that in the next enter training ratio in the editable text window. Give the file name in the for training data.

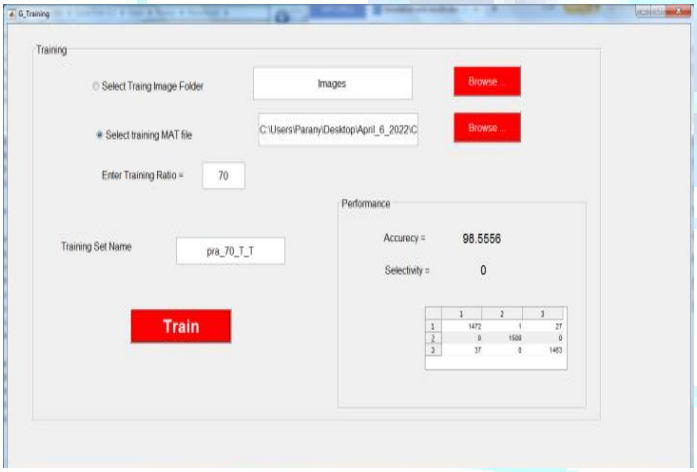

Fig. 4.1 GUI of training data with 70%

3. Now click on train button.
4. Start training, after completion of training, the outcomes of proposed method in 70% training accuracy is 98.5556% .
5. The confusion matrix of proposed method also shown in the GUI.

In the above figure 6.7 shows the training outcome of proposed method.

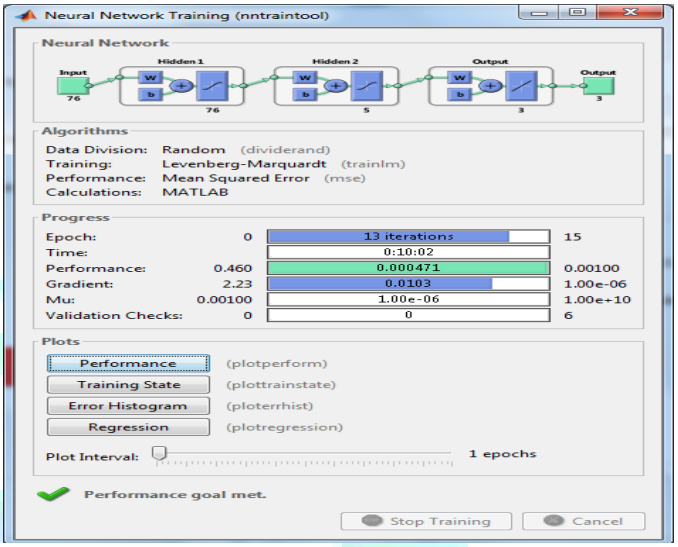

Fig. 4.2 Neural Network and Input Parameters

In the above figure 4.2 shows the neural network input parameters. The proposed method is based on feed forward neural network.

- There are 76 input parameter are selected as an input parameters.
- There are two input hidden layers are apply.
- Data division is in random process.
- For training process use levenbergmarquardt (LBM).

These are major input parameters which are apply in the input of the proposed method. Now discuss the testing outcome plots. There are three major plot for the performance analysis of proposed method. First is performance validation, second one is training state, third one is error histogram and last one is regression plot.

shows best validation performance output **13** epoch and the mean square error **0.0085855**.

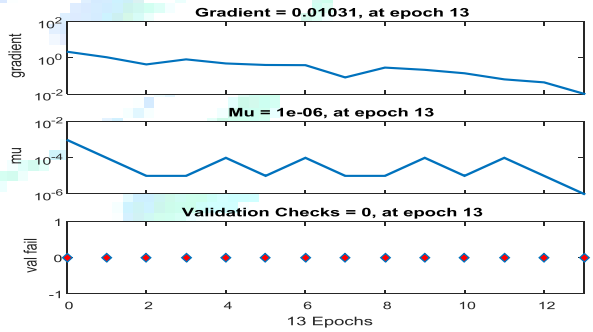

Fig. 4.4 Shows the proposed method gradient, Mu, and validation

In the above figure 4.4 shows the proposed outcomes in terms of gradient, mu, and validation checks. In the above divided into four subplots., in the first subplots shows the gradient outcome plot of proposed work, in second subplot shows the Mu plot with optimum error value in the last shows validation checks. In the X axis shows the number of epochs in the simulation and Y axis in the first plot denote gradient

value, second plot denote mu value and third plot shows the validation fail.

V. CONCLUSION

The Cancer is a disease that is becoming increasingly prevalent around the world. Numerous researchers have conducted a variety of studies in an effort to determine the areas of the human body that are most commonly affected by cancer. The results of one study like this one motivated us to carry out this research in the field of detecting lung cancer. The most common cause of death attributable to cancer in both men and women is lung cancer. The likelihood of a favourable prognosis for lung cancer patients who undergo early detection is increased. The utilisation of image processing systems makes it possible to detect and diagnose abnormalities earlier and more quickly than is possible with the use of other screening tests. When developing a method for the early diagnosis and treatment of disease, taking into account the passage of time is one of the considerations that goes into the process.

REFERENCES

[1]. Bharti, Meenakshi, Jaytrilok Choudhary, and Dhirendra Pratap Singh. "Detection and Classification of Pulmonary Lung Nodules in CT Images Using 3D Convolutional Neural Networks." In *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, pp. 1319-1324. IEEE, 2022.

[2]. Halder, Amitava, Saptarshi Chatterjee, and Debangshu Dey. "Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification." *Biomedical Signal Processing and Control* 72 (2022): 103347.

[3]. Gnanavel, S. "Identification and Classification of Lung Nodules Using Neural Networks." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.6 (2021): 1956-1961.https://turcomat.org/index.php/turkbilmat/article/view/4799

[4]. Abdullah, Dakhaz Mustafa, Adnan Mohsin Abdulazeez, and Amira Bibo Sallow. "Lung cancer prediction and classification based on correlation selection method using machine learning techniques."Qubahan Academic Journal1.2 (2021): 141-149.https://journal.qubahan.com/index.php/qaj/article/view/58

[5]. Sujitha, R., and V. Seenivasagam. "Classification of lung cancer stages with machine learning over big data healthcare framework." Journal of Ambient Intelligence and Humanized Computing 12 (2021): 5639-5649.https://link.springer.com/article/10.1007%2Fs12652-020-02071-2

[6]. Shanthi, S., and N. Rajkumar. "Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods." Neural Processing Letters 53.4 (2021): 2617-2630.

[7]. https://link.springer.com/article/10.1007/s11063-020-10192-0

[8]. Doppalapudi, Shreyesh, Robin G. Qiu, and YouakimBadr. "Lung cancer survival period prediction and understanding: Deep learning approaches." International Journal of Medical Informatics 148 (2021): 104371.

[9]. Naik, Amrita, and Damodar Reddy Edla. "Lung nodule classification on computed tomography images using deep learning." Wireless Personal Communications 116.1 (2021): 655-690.

[10]. Patel, Divek, Connor Cowan, and Prateek Prasanna. "Predicting Mutation Status and Recurrence Free Survival in Non-Small Cell Lung Cancer: A Hierarchical ct Radiomics–Deep Learning Approach." 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021.

[11]. Gao, Yang, et al. "Improving the Subtype Classification of Non-small Cell Lung Cancer by Elastic Deformation Based Machine Learning." Journal of Digital Imaging (2021): 1-13.

[12]. Shakeel, P. Mohamed, M. A. Burhanuddin, and Mohammad Ishak Desa. "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier." Neural Computing and Applications (2020): 1-14.https://link.springer.com/article/10.1007/s00521-020-04842-6

[13]. Toğaçar, Mesut, Burhan Ergen, and Zafer Cömert. "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks." Biocybernetics and Biomedical Engineering 40.1 (2020): 23-39.https://www.sciencedirect.com/science/article/abs/pii/S0208521619304759

[14]. Yuan, Fei, Lin Lu, and Quan Zou. "Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms." Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease 1866.8 (2020): 165822.

[15]. Valluru, Dinesh, and I. Jeya. "IoT with cloud based lung cancer diagnosis model using optimal support vector machine." Health care management science 23.4 (2020): 670-679.

[16]. Günaydin, Özge, MelikeGünay, and ÖznurŞengel. "Comparison of lung cancer detection algorithms." 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). IEEE, 2019.https://ieeexplore.ieee.org/abstract/document/8741826

[17]. Luna, José Marcio, et al. "Predicting radiation pneumonitis in locally advanced stage II–III non-small cell lung cancer using machine learning." Radiotherapy and Oncology 133 (2019): 106-112.https://www.sciencedirect.com/science/article/abs/pii/S0167814019300076

[18]. Huo, Donglai, Mark Kiehn, and Ann Scherzinger. "Investigation of Low-Dose CT Lung Cancer Screening Scan "Over-Range" Issue Using Machine Learning

Methods." Journal of digital imaging 32.6 (2019): 931-938.https://link.springer.com/article/10.1007/s10278-019-00233-z

[19]. Skourt, Brahim Ait, Abdelhamid El Hassani, and AichaMajda. "Lung CT image segmentation using deep neural networks." Procedia Computer Science 127 (2018): 109-113.

[20]. Kumar, N. Komal, et al. "Predicting non-small cell lung cancer: a machine learning paradigm." *Journal of Computational and Theoretical Nanoscience* 15.6-7 (2018): 2055-2058.

[21]. Li, Hongming, et al. "Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy." *Radiotherapy and Oncology* 129.2 (2018): 218-226.

[22]. Pradeep, K. R., and N. C. Naveen. "Lung cancer survivability prediction based on performance using classification techniques of support vector machines, C4. 5 and Naive Bayes algorithms for healthcare analytics." *Procedia computer science* 132 (2018): 412-420.

[23]. Lustberg, Tim, et al. "Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer." *Radiotherapy and Oncology* 126.2 (2018): 312-317.

[24]. Lin, Tong, and Yucheng Lin. "Markerless tumor gating and tracking for lung cancer radiotherapy based on machine learning techniques." *Artificial Intelligence in Decision Support Systems for Diagnosis in Medical Imaging*. Springer, Cham, 2018. 337-359.

[25]. Sharma, Karan, HarshilSoni, and Kushika Agarwal. "Lung Cancer Detection in CT Scans of Patients Using Image Processing and Machine Learning Technique." *Advanced Computational and Communication Paradigms*. Springer, Singapore, 2018. 336-344.

[26]. Kohad, Rashmee, and Vijaya Ahire. "Application of machine learning techniques for the diagnosis of lung cancer with ANT colony optimization." *International Journal of Computer Applications* 113.18 (2015): 34-41.

