ISSN 2319 – 1953
International Journal of Scientific Research in Computer Science Applications and Management Studies

A Study on Forecasting Mechanism of Lung Cancer Using Artificial Neural Network

Rajathi V [#1], Ponninselvi A [*2]

[1] *Asst. Professor, Department of Computer Science, Jayaraj Annapackiam College for Women (Autonomous), Periyakulam, TamilNadu, India,*

[2] *II. M..Sc(CS& IT), Jayaraj Annapackiam College for Women (Autonomous),Periyakulam, TamilNadu, India,*

[1]rajathiresolute@gmail.com,

***Abstract* - Today Countless humans are being replaced by computers in their working environments. Because they can do work more efficiently at much lower cost to business one thing that gives humans the edge of a computers is that they gives humans and learn according to certain trends but with modern technology computers are being taught to learn as well. Now a day's Machine Learning is going to be a part of our everyday activities. This machine learning and its algorithm is used in healthcare system very effectively. Lots of algorithms and techniques are used in existing paper. Diagnosed cancer in early stage is difficult task. In this paper give summary of techniques which is used in previous. And finally give the best among them. According to this survey Feed forward neural network give the best result. In normally machine learnt from past experience. So In Feed forward neural network predict from stored dataset. Multilayer perceptron was another name of feed forward neural network.**

***Keywords*: Computer Aided Detection, Chemotherapy, Feed Forward Neural Network, Non Small Cell Carcinom, Convolution Neural Network**

I. INTRODUCTION

Cancer is a word given to a group of diseases involving abnormal cellular growth that can potentially spread and threaten other tissues. Cancer is the antisocial behaviour of own body's cells that typically happens when various genetic abnormalities accumulate in its. DNA cell contain a gene specifically meant to prevent cancer occurring it's known as a tumor Suppressor gene. Its job is to stop the cell from growing out of control. And restrict its activity. Cells also contain Proto Oncogenes which do the exact opposite when activated into oncogenes. These gene are switched on and off determines the cell's growth and ultimately whether got cancer. Harmful cancer caused by chemicals into the body known as carcinogens. These can affect the genes inside a cell and change them. Another method is infection as in addition to smoking, risk factors for cancer include heavy alcohol consumption, excess body weight, physical inactivity, and poor nutrition. Other causes of cancer are not preventable. Currently, the most significant unpreventable risk factor is age. This paper presents a forecasting methodology for Lung cancer using artificial neural networks.

II. LUNG CANCER

Lung cancer can start in the windpipe, main airway or lungs. It develops when there is uncontrolled growth of abnormal cells inside one or both lungs. These cells grow to form tumors more than eight out of ten cases of lung cancer caused by smoking. And other causes of lung cancer are Radon gas, Air pollution, Chemicals. In the workspace of cancer that starts in the lung is called a primary lung cancer while a cancer that starts in other part of the body and spread to the lung is called a secondary lung cancer. The stage of cancer refers to its size and how far it is spread. The types of Lung cancer is depicted in Figure 1.

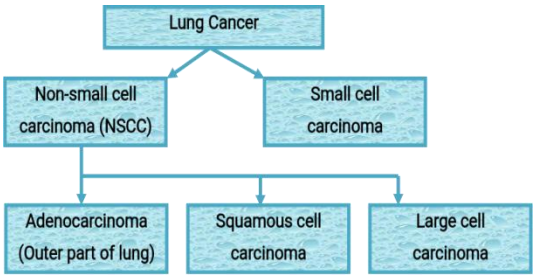

Figure 1. Types of Lung Cancer

According to the American Cancer Society, doctors in the U.S. diagnose 87 percent of cancer cases in people ages 50 years or older. There are two types of lung cancer namely, Non-small cell carcinoma (NSCC) and Small cell carcinoma. And NSCC is further divided into three. That is Adenocarcinoma, Squamous cell carcinoma and large cell carcinoma. Totally it has four stages. In the First stage Cancer is found in the lung, but it has not spread outside the lung. In the Second stage Cancer is found in the lung and nearby lymph nodes. And the Third stage Cancer is in the lung and lymph nodes in the middle of the chest. And it is further divided into two. Stage 3A Cancer is found in lymph nodes, but only on the same side of the chest where cancer first started growing. Stage 3B Cancer has spread to lymph nodes on the opposite side of the chest or to lymph nodes above the collarbone. At the final stage Cancer has spread to both lungs, into the area around the lungs, or to distant organs. Treatment option for cancer includes local therapies such as surgery and

Radiation therapy and systemic therapies such as Chemotherapy and Targeted therapy. It is depending on the type and stage of cancer. Radiation therapy is local treatment that uses high-energy rays to kill or shrink cancer cells to relieve symptoms. Chemotherapy is a systemic treatment that uses drugs to stop the growth of the cancer cells by killing them or inhibiting their cell division once the drugs enter the bloodstream they can travel and reach cancer cells throughout the body. The stages of the Lung cancer are presented in Figure 2.

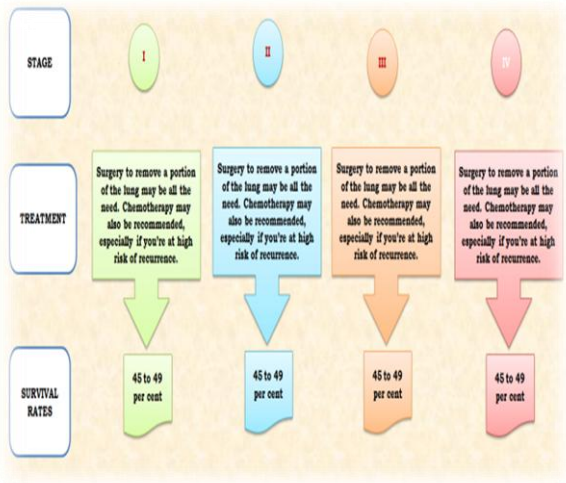

Figure 2. Stages of Lung Cancer

III. LITERATURE REVIEW

Mohamed Shakeel et al., [1] presented automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks. The noise present in the data was eliminated by applying the bin smoothing normalization process. The minimum repetition and Wolf heuristic features were subsequently selected to minimize the dimensionality and complexity of the features. The selected lung features were analyzed using discrete AdaBoost optimized ensemble learning generalized neural networks, which successfully analyzed the biomedical lung data and classified the normal and abnormal features with great effectiveness. In the future, lung cancer prediction can be further improved by capturing and examining the sensor data, using an optimized technique for enhancing the earlier prediction rates.

Xiaojiao Xiao et al., [2] proposed a feature extraction method for lung nodules based on a multichannel principal component analysis network (PCANet). The architecture consists of three seamlessly connected functional layers. RGB multichannel can automatically extract ROI sequence images involving lung parenchyma from the lung imaging sequence. The feature extraction layers, using Principal Component Analysis Network - random binary hash (PCANet-RBH), extract high-level semantic features of the R/G/B channel by cascading PCA and fuse the extracted normal color patterns, and generate multiple binary patterns via RBH to produce richer features with color information. PCANet is a simplified convolutional neural network method that can extract features. It uses the mapping matrix of principal component analysis instead of the convolution kernel of convolutional neural networks.

Furqan Shaukat et al., [3] discussed with artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features. The Computer Aided Detection CAD could be a powerful tool for initial lung nodule detection and preventing the deaths caused by the lung tumor. In this paper, an advanced technique for lung-nodule detection by using a hybrid feature set and artificial neural network is proposed. Initially, the lung volume is segmented from the input Computed Tomography image using optimal thresholding which is followed by image enhancement using with multi scale dot augmentation filtering. Next, lung nodule candidates have been detected from enhanced image and certain features are extracted. The set feature consists of the texture features, shape 2D and 3D and intensity. Finally, lung nodule's classification is attained using two-layer feed forward neural network. Finally, the candidate nodules are detected and segmented instantaneously from an enhanced image by multi scale dot augmentation filter. Feature vectors are formed by using intensity, texture features and shape features from lung nodule candidates and used by ANN classifier to reduce false positives. LIDC dataset has been used to evaluate their proposed system. The achieved sensitivity of their proposed system is 95.5% with 5.72 false positives per scan only.

Maisa Daoud et al., [4] explained a survey of neural network-based cancer prediction models from microarray data. The considered papers were published between 2013-2018 and used gene expression datasets for cancer classification and clustering. This review presented some commonly used architecture, datasets, and the accuracies of each suggested model. This study has summarized most recent approaches and their related neural network architectures. They also highlighted some critical points that have to be considered when building a neural network-based prediction model such as over fitting and class imbalance. More powerful neural network-based approaches can be suggested in future by choosing different network's parameters or combining two or more of the presented approaches.

Lakshmanaprabu et al., [5] analysed optimal deep learning model for classification of lung cancer on CT Images. The proposed ODNN with feature reduction demonstrated the better classification in case of lung CT Images compared with others classification techniques. An automatic lung cancer classification approach reduces the manual labelling time and avoids a human mistake. Through machine learning techniques, the researchers planned to achieve better precision and accuracy in recognizing a normal and abnormal lung image. According to the experimental outcomes, the proposed technique is effective for the classification of the human lung images in terms of accuracy, sensitivity, and specificity with its values 94.56%, 96.2%, and 94.2% respectively. The

accuracy level has clearly evident that the proposed algorithm is deeply proficient in recognizing cancer-affected parts in CT images. The classification performances of this investigation demonstrate the advantages of this strategy: it is speedy, simple to operate, non-invasive and cheap. In future work, they will use high dosage CT lung images and optimal feature selection with multi-classifier consisted to cancer detection process.

Ayshath Thabsheera et al., [6] anticipated lung cancer detection using CT scan images a review on various image processing techniques. This paper encapsulates a short review on different techniques used for lung cancer diagnosis using CT images. It is observed from the literature that Gabor filter enhancement gives better performance. Features such as area, eccentricity are calculated and given as input to the classifier, which decides whether the lung nodule in the CT image is cancerous or non-cancerous. As a future work, MRI, X-ray, PET images can be considered for detecting lung cancer. Thus, one can compare and analyze which imaging modality gives better result for the detection of lung cancer. They can also classify the images into several stages of lung cancer using many techniques such as fuzzy or neural network.

Siddharth Bhatia et al., [7] represented lung cancer detection: A deep learning approach. The feature set is fed into multiple classifiers, viz. XGBoost and Random Forest, and the individual predictions are ensemble to predict the likelihood of a CT scan being cancerous. The accuracy achieved is 84% on LIDC-IRDI outperforming previous attempts. In this paper, they propose an approach to lung cancer detection employing feature extraction using deep residual networks. They compare performance of tree-based classifiers like Random Forest and XGBoost. The highest accuracy they get is 84% using ensemble of Random Forest and XGBoost classifier.

Yash Dagli et al., [8] analysed prediction of two year survival among patients of non-small Cell Lung Cancer. In this research, they try to create a prediction model, with the help of machine learning to accurately predict the survival of non-small cell lung cancer patients (NSCLC). Clinical data of 559 patients was taken for training and testing of models. They have developed multilevel perceptron model for survival prediction. Other models developed during this study were compared to measure performance of their model. Attributes that are found to be useful as biomarkers for prediction of survival analysis of NSCLC have also been computed and ranked accordingly for increase in accuracy of prediction model by implementing feature selection method. The final model included T stage, N stage, Modality, World Health Organization Performance status, Cumulative Total Tumor dose, tumor load, Overall treatment time as the variables. Two year survival was chosen as the prediction outcome. Neural Network was found as the best prediction model with area under Curve (AUC) of 0.75. By far to their knowledge Multilevel Neural Network is found to be the best model for predicting two-year survival among patients of non-small cell lung cancer.

Andre Dias et al., [9] offered a solution to a deep learning line to assess patient's lung cancer stages. Goal is to pursue a vision of developing and maintaining a comprehensive and integrated computer model to help physicians plan the most appropriate treatment and anticipate a patient's prospects for the extent of cancer. For example, cancer can be treated at an early stage by surgery or radiation, while chemotherapy may be the care for more advanced stages. In fact, early detection of this type of cancer facilitates its treatment and may raise the patients' prospect of a continued existence. Thus, a formal view of an intelligent system for performing cancer feature extraction and analysis in order to establish the bases that will help physicians plan treatment and predict patient's prognosis is presented. It is based on the Logic Programming Language and draws a line between Deep Learning and Knowledge Representation and Reasoning, and is supported by a Case Based attitude to computing. In fact, despite the fact that each patient's condition is different, treating cancer at the same stage is often similar.

Miao Zhang et al., [10] suggested multi-level CNN for lung nodule classification with gaussian process assisted hyper parameter optimization. This paper investigates lung nodule classification by using deep neural networks (DNNs). DNN has shown its superiority on several medical image processing problems, like medical image segmentation, medical image synthesis, and so on, but their performance highly dependent on the appropriate hyper parameters setting. Hyper parameter optimization in DNNs is a computationally expensive problem, where evaluating a hyper parameter configuration may take several hours or even days. Bayesian optimization has been recently introduced for the automatically searching of optimal hyper parameter configurations of DNNs. It applies probabilistic surrogate models to approximate the validation error function of hyper parameter configurations, such as Gaussian processes, and reduce the computational complexity to a large extent. However, most existing surrogate models adopt stationary covariance functions (kernels) to measure the difference between hyper parameter points based on spatial distance without considering its spatial locations. This distance-based assumption together with the condition of constant smoothness throughout the whole hyper parameter search space clearly violate the property that the points far away from optimal points usually get similarly poor performance even though each two of them have huge spatial distance between them. In this paper, a non-stationary kernel is proposed which allows the surrogate model to adapt to functions whose smoothness varies with the spatial location of inputs, and a multi-level convolution neural network (ML-CNN) is built for lung nodule classification whose hyper parameter configuration is optimized by using the proposed non-stationary kernel based Gaussian surrogate model. Their algorithm searches the surrogate for optimal setting via hyper parameter importance based evolutionary strategy, and the experiments demonstrate Their algorithm outperforms manual tuning and well-established hyper parameter optimization methods such as Random search, Gaussian processes (GP)

with stationary kernels, and recently proposed Hyper-parameter Optimization via RBF and Dynamic coordinate search (HORD).

Ryota Shimizu et al., [11] proposed deep learning application trial to lung cancer diagnosis for medical sensor systems. Personal and easy-to-use health checking system is an attractive application of sensor systems. Sensing data analysis for diagnosis is important as well as preparing small and mobile sensor nodes because sensing data include variations and noises reflecting individual difference of people and sensing conditions. Neural means relating to a nerve or to the nervous system. Neural network was interconnection of neural information. Everyone was inspired with other. Similarly, neural network came from biological neuron. So, to implement neural network, first we have to know the functionality and working procedure of biological neuron, but the understanding the biological neuron was complicated task. Character recognition was one of the applications of neural network. It was most interesting researches. The characters are regulated for size, located on a grid, and projections are prepared of the lines through the squares of the grid. These projections form the inputs to a back-propagation network. Deep Neural Network, or Deep Learning, is a well-known method of machine learning and it is effective for feature extraction from pictures. Then, they thought Deep Learning also can extract features from sensing data. In this paper, they tried to build a diagnosis system of lung cancer based on Deep Learning. Input data of the system was generated from human urine by Gas Chromatography Mass Spectrometer (GC-MS) and their system achieved 90% accuracy in judging whether the patient had lung cancer or not. This system will be useful for pre- and personal diagnosis because collecting urine is very easy and not harmful to human body. They are targeting installation of this system not only to gas chromatography systems but also to some combination of multiple sensors for detecting gases of low concentration.

Qian Xiao et al., [12] developed application of kriging models for a drug combination experiment on lung cancer. Combinatorial drugs have been widely applied in disease treatment, especially chemotherapy for cancer, due to its improved efficacy and reduced toxicity compared with individual drugs. Neural Network is made up from the functionality human brain and it is biologically inspired. In this paper character recognition is identified. Normally humans can easily recognize the characters that are digits, numbers, and alphabets. This task to be carried out by computer is involves a high level of complexity for the reason that the computer can only understand binary codes. To recognize the character there are six basic steps are followed. Pre-processing holds noise reduction and binarization techniques. Feature extraction contains zoining method, moment methods. Segmentation is divided into 3 parts, namely line segmentation, word segmentation and character segmentation. Classification includes classifiers like Support Vector Machine and Decision Tree. And finally Post-processing includes Grouping and Error detection and correction. The study of combinatorial drugs requires efficient experimental designs and proper follow-up statistical modelling techniques. Linear and nonlinear models are often used in the response surface modelling for such experiments. They propose the use of kriging models to better depict the response surfaces of combinatorial drugs. They illustrate their method via a drug combination experiment on lung cancer and further show how proper experimental designs can reduce the necessary run size. They demonstrate that only 27 runs are needed to predict all 512 runs in the original experiment and achieve better precision than existing analyses.

Wu et al., [16] investigated the potential utility of artificial neural networks as a decision-making aid to radiologists in the analysis of mammographic data. Three-layer, feed-forward neural networks with a back-propagation algorithm were trained for the interpretation of mammograms on the basis of features extracted from mammograms by experienced radiologists. A network that used 43 image features performed well in distinguishing between benign and malignant lesions, yielding a value of 0.95 for the area under the receiver operating characteristic curve for textbook cases in a test with the round-robin method. With clinical cases, the performance of a neural network in merging 14 radiologist-extracted features of lesions to distinguish between benign and malignant lesions was found to be higher than the average performance of attending and resident radiologists alone (without the aid of a neural network). The authors conclude that such networks may provide a potentially useful tool in the mammographic decision-making task of distinguishing between benign and malignant lesions.

IV. SUMMARY OF PRECEDING METHODOLOGY

TABLE 1. SUMMARY OF PRECEDING METHODOLOGY

PAPER	STAGES	ALGORITHMS/ TECHNIQUES	Data Set
[1, Sha]	Image Enhancement	Bin Smoothing Normalization	ELVIRA
	Feature Learning	Discrete Adaboost Technique	
	Feature Selection	Wolf heuristic	
	Prediction	Ensemble learning	

[2, Xia]	Feature Extraction	Principal Component Analysis Network	LIDC
[3, Sha]	Image Enhancement	Multi scale dot augmentation filter	LIDC
	Classification	Two Layer Feed Forward NN	
	Prediction	Computer Aided Detection (CAD)	
[4, Dao]	Classification	Convolution Neural Network	TCGA
[5, Lak]	Classification	Feed Forward Neural Network	Private
	Feature Extraction	Wavelet	
	Reduction	Optimal Deep Neural Network	
[6, Tha]	Image Enhancement	Spatial and Frequency Domain Gabor Filter	LIDC
	Segmentation	Otusu Thresholding	
	Classification	ANN	
[7, Bha]	Feature Extraction	Deep Residual Network	LIDC – IDRI
	Classification	Random Forest	
[8, Dag]	Feature Selection	ReliefF	-
[10, Zha]	Classification	Deep Neural Network	MNIST

V. CONCLUSION

This paper discussed about Cancer, its types and stages and moreover it gives the summary of existing techniques and algorithms which is used to predict cancer tumor in early stage. So finally it concluded Feed Forward neural network is give best result. Because in normally machines are learn from past experience and feedbacks. So first have to import the knowledge of cancer and its result. Feed forward neural network is analysed it with previous experience to predict the cancer in early stage.

REFERENCE

[1] Shakeel, P. Mohamed, et al. "Automatic detection of lung cancer from biomedical data set using discrete AdaBoost optimized ensemble learning generalized neural networks." *Neural Computing and Applications* (2019): 1-14.

[2] Xiao, Xiaojiao, et al. "A feature extraction method for lung nodules based on a multichannel principal component analysis network (PCANet)." *Multimedia Tools and Applications* (2019): 1-19.

[3] Shaukat, Furqan, et al. "Artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features." *Journal of Ambient Intelligence and Humanized Computing* (2019): 1-15.

[4] Daoud, Maisa, and Michael Mayo. "A Survey Of Neural Network-based Cancer Prediction Models From Microarray Data." *Artificial Intelligence in Medicine* (2019).

[5] Lakshmanaprabu, S. K., et al. "Optimal deep learning model for classification of lung cancer on CT images." *Future Generation Computer Systems* 92 (2019): 374-382.

[6] Thabsheera, AP Ayshath, T. M. Thasleema, and R. Rajesh. "Lung Cancer Detection Using CT Scan Images: A Review on Various Image Processing Techniques." *Data Analytics and Learning*. Springer, Singapore, 2019. 413-419.

[7] Bhatia, Siddharth, Yash Sinha, and Lavika Goel. "Lung Cancer Detection: A Deep Learning Approach." *Soft Computing for Problem Solving*. Springer, Singapore, 2019. 699-705.

[8] Dagli, Yash, Saumya Choksi, and Sudipta Roy. "Prediction of Two Year Survival Among Patients of Non-small Cell Lung Cancer." *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*. Springer, Cham, 2019. 169-177.

[9] Dias, André, et al. "A Deep Learning Line to Assess Patient's Lung Cancer Stages." *Third International Congress on Information and Communication Technology*. Springer, Singapore, 2019.

[10] Zhang, Miao, et al. *"Multi-level CNN for lung nodule classification with Gaussian Process assisted hyperparameter optimization."* arXiv preprint arXiv:1901.00276 (2019).

[11] Shimizu, Ryota, et al. *"Deep learning application trial to lung cancer diagnosis for medical sensor systems."* 2016 International SoC Design Conference (ISOCC). IEEE, 2016.

[12] Xiao, Qian, Lin Wang, and Hongquan Xu. "*Application of kriging models for a drug combination experiment on lung cancer."* Statistics in medicine 38.2 (2019): 236-246.

[13] V.Rajathi, *"Synthesize of Character Recognition in Neural Network: A Survey"*, International Journal of Scientific Research in Computer Science Application and Management Studies (2019).

[14] http://www.ijsrcsams.com/images/stories/Past_Issue_Docs/ijsrcsamsv7i3p126.pdf

[15] http://www.ijsrcsams.com/images/stories/Past_Issue_Docs/ijsrcsamsv7i4p104.pdf

[16] Wu, Yuzheng, et al. *"Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer." Radiology* 187.1 (1993): 81-87.

[17] V.Rajathi, *"A New Architecture For Recognizing Tamil Characters Using Artificial Neural Network"* International Journal of Scientific Research in Computer Science Application and Management Studies (2019).