International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST218236

Lung Cancer Prediction Using Chest X-Rays

Vegi Jeevana , B. Sai Ram Charan
SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

Article Info
Volume 8, Issue 2
Page Number : 236-242

Publication Issue
March-April-2021

Article History
Accepted : 20 March 2021
Published : 25 March 2021

ABSTRACT

Cancer is the most important cause of death for both man and women .The early detection of cancer can be helpful in the curing of the disease completely. So the requirement of techniques to detect the occurrence of cancer in early stage is increasing. To design a model that can be useful in early identification and prediction of a lung diseases have become a necessity in the research, as it can facilitate the subsequent clinical management of patients. Here, we establish a diagnostic tool based on a deep-learning framework for the screening of patients with common treatable blinding retinal diseases. Our framework utilizes a technique, which trains a neural network with a fraction of the data of conventional approaches. For such purpose, we employ the Deep learning techniques and machine learning approaches to train the data and evaluate the efficiency of our proposed research.

Keywords : Deep Learning, Cancer Prediction, ConvNet, Chest Xray, CNN

I. INTRODUCTION

The most common cause of lung cancer is long-term exposure to tobacco smoke, which causes 80-90% of lung cancers. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the centre of the chest. Lung cancer often spreads toward the centre of the chest because the natural flow of lymph out of the lungs is toward the centre of the chest. As for the stages, in general there are four stages of lung cancer; I through IV. One of the major reason for non-accidental death is cancer. It has been proved that lung cancer is the topmost cause of cancer death in men and women worldwide. The death rate can be reduced if people go for early diagnosis so that suitable treatment can be administered by the clinicians within specified time. Cancer is, when a group of cells go irregular growth uncontrollably and lose balance to form malignant tumours which invades surrounding tissues. Cancer can be classified as Non-small cell lung cancer and small cell lung cancer. The various ways to detect lung cancer is by the used of image processing , pattern recognition and artificial neutral network to develop Computer aided diagnosis. In this research we used the techniques and algorithm used in image processing to detect cancer in three types of medical images. In this system first of all the medical images are recorded using a suitable imaging system. The images obtained are taken as input for the system where the image first go through

the various steps of image processing like pre-processing, edge detection, morphological processing ,feature extraction.

Lung cancer which is among the five main types of cancer is a leading one to overall cancer mortality. Cancer is a serious health problem among various kinds of diseases. World Health Organization (WHO) reports that worldwide 7.6 million deaths are caused by cancer each year. Uncontrollable cell development in the tissues of the lung is called as lung cancer. Lung nodule is an abnormality that leads to lung cancer, characterized by a small round or oval shaped growth on the lung which appears as a white shadow in the CT scan. These uncontrollable cells restrict the growth of healthy lung tissues. If not treated, this growth can spread beyond the lung in the nearby tissue called metastasis and, form tumours. In order to preserve the life of the people who are suffered by the lung cancer disease, it should be pre-diagnosed. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. So there is a need of pre-diagnosis system for lung cancer disease which should provide better results.

The main purpose of the research is to predict the cancer patients from Chest X-ray CXR image dataset. Our proposed deep learning algorithm is used to predict the cancer patient more accurately and increase the system performance to gives the result and saves the time from the clinical and Avoid the sparsity problems. We believe the early diagnosis and prognosis of a cancer type have become a necessity in cancer research, as it can facilitate the subsequent clinical management of patients. The main scope of the research is to identify the cancer from chest X-ray CXR image. The deep learning algorithm is used to increase the system performance to predict the cancer accurately.

II. METHODLOGY

The basic methodology of the research were designed on two principles, 1) **Review research** and 2) **Building up a new algorithm**.

1.1. Review research

This study was conducted using four databases: Google Scholar, SAGE, DOAJ and PubMed. Selection of papers was done based on keywords and theme relevant to this review. Further, the published papers from these databases were arranged in systemic order with respect to year of publication for better understanding.

1.2. Building up a new algorithm

After a careful inference from review research, the proposed algorithm was developed. The algorithm was divided into seven modules shown in Fig (1).

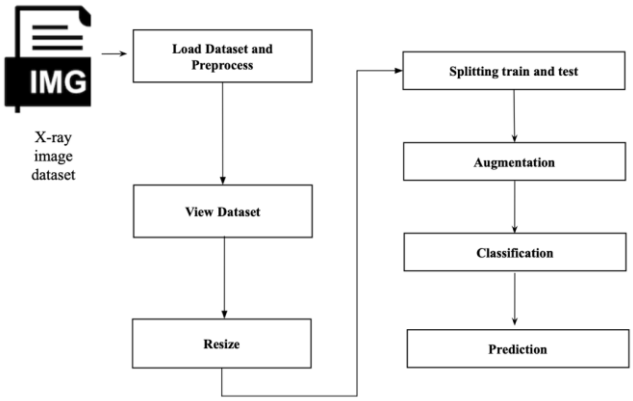


Fig (1). Show the seven modules of algorithm.

III. RESULTS AND DISCUSSION

REVIEW INFERENCE:

a. Understanding of a convolutional neural network

It has been observed that the CNN has an excellent performance in machine learning problems. Specially the applications that deal with image data, such as largest image classification data set (Image Net), computer vision, and in natural language processing (NLP) and the results achieved were very amazing. In

this paper, all the elements and important issues related to CNN, and how these elements work were explained and defined. In addition, it the parameters that effect CNN efficiency were stated. The interest in having deeper hidden layers has recently begun to surpass classical methods performance in different fields; especially in pattern recognition. One of the most popular deep neural networks is the Convolutional Neural Network (CNN).

b. Diagnosis of Liver Disease Using Machine Learning Techniques

Different techniques to see which model has good prediction accuracy were used. The objective of this study was to improve the diagnosis and prediction of the liver disease with the machine learning algorithms. There are different data mining methods like classification, clustering, association, and pattern matching and data visualization. In this paper various data mining techniques development were surveyed through a literature review.

c. Efficient CNN for Lung Cancer Detection: The seven layer combination was chosen to improve the lung cancer detection results. EFFI-CNN was developed based on the experiments performed in ICDSSPLD-CNN and EASPLD-CNN. Normal Lung Image-Cancerous Lung Image showed the Lung cancer data set taken from LIDC-IDRI. It was used to display the difference between normal lung and cancerous lung.

d. A Review of Deep Learning with Recurrent Neural Network: The main aim was to showcase the recurrent Neural Network (RNN) which is a deep learning model that uses the concept of supervised learning. Deep learning belongs to the family of machine learning. It is also called hierarchical learning or deep structured learning. The classic machine learning algorithms are definite, while the deep learning algorithms follow a chain of command. Deep learning has the capability to deal with more complex neural networks and it mainly deals with sequential data. Recurrent networks can process examples one at a time, preserving an element, that reflects over a long period of time.

e. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database: The main aim was to showcase Pneumonia is a bacterial, viral, or fungal infection of one or both sides of the lungs that causes lung alveoli to fill up with fluid or pus, which is usually diagnosed with chest x-rays. This work investigated opportunities for applying machine learning solutions for automated detection and localization of pneumonia on chest x-ray images. It further, proposed an ensemble of two convolutional neural networks, namely Retina Net and Mask R-CNN for pneumonia detection and localization. It was validated by the solution on a recently released dataset of 26,684 images from Kaggle Pneumonia Detection Challenge and it was scored among the top 3% of submitted solutions. With 0.793 recall, we have developed a reliable solution for automated pneumonia diagnosis and have validated it on the largest clinical database publicly available to date.

f. Malignant Lung Nodule Detection using Deep Learning: The main aim was to showcase an approach to detect malignant pulmonary nodules from CT scans using Deep Learning. A pre-processing pipeline was used to mask out the lung regions from the scans. The features were then extracted using a 3D CNN model based on the C3D network architecture. The LIDC-IDRI is the primary dataset that was used along with a few resources from the LUNA16 grand challenge for the reduction of false-positives. The end product is a model that predicts the coordinates of malignant pulmonary nodules and demarcates the corresponding areas from the CT scans. The final

model achieved a sensitivity of 86 percent for detecting malignant Lung Nodules and predicting its malignancy scores.

ALGORITHM CREATION

Data selection and loading: Data selection is the process of selecting the data for cancer chest X-ray image. In this research, Cancer chest X-ray image was used to find cancer and normal patient. The dataset contains information about chest X-ray image of test, train and valid. Three datasets contained cancer and normal patient X-ray as a Gray scale image.

Data processing: Image data pre-processing is the process of getting rescale data from the dataset.

a. Resize image dataset: Rescale the Gray scale chest x-ray image size into 200.
b. Getting data: Categorical data is defined as variables with a finite set of rescaled values. Most deep learning algorithms require array input and output variables.

Splitting dataset into train and test data: Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

Classification: Classification involves predicting which class an item belongs to. Some classifiers are binary, resulting in a yes/no decision. Others are multi-class, able to categorize an item into one of several categories. Classification is a very common use case of machine learning—classification algorithms are used to solve problems like email spam filtering, document categorization, speech recognition, image recognition, and handwriting recognition.

In this context, a neural network is one of several machine learning algorithms that can help solve classification problems. Its unique strength is its ability to dynamically create complex prediction functions, and emulate human thinking, in a way that no other algorithm can. There are many classification problems for which neural networks have yielded the best results.

Prediction: It's a process of predicting the Pneumonia from the dataset. This project has effectively predicted the data from dataset by enhancing the performance of the overall prediction results.

Algorithm: CNN in deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analysing visual imagery. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing, brain-computer interfaces, and financial time series. CNNs are regularized versions of multilayer perceptions. Multilayer perceptions usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to over fitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme. Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual

cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

A. Existing Model

Cancer chest X-ray Gray scale image based on test, train and valid was used. Detection of pneumonia patients and Normal patients can only be done with this model and X-ray image was tough to predict. The existing system was used to predict the low level accuracy based on image processing. The existing system didn't effectively classify and predict the pneumonia chest X-ray image.

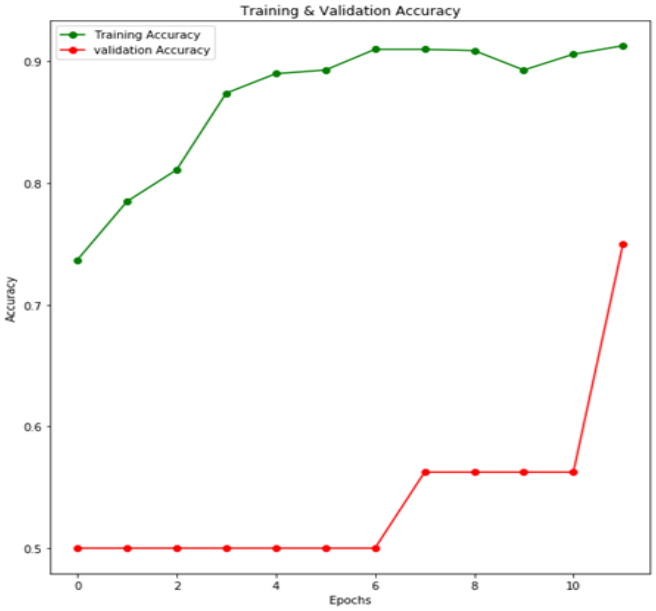
B. Proposed System

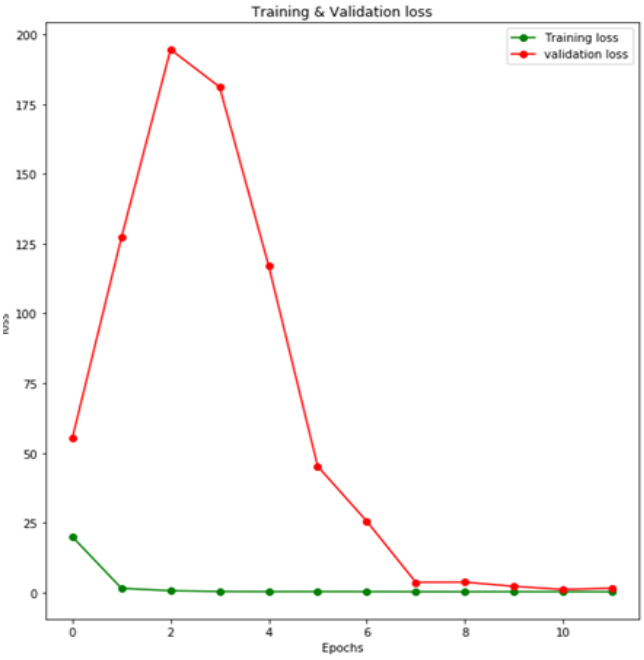
The proposed model was introduced to overcome all the disadvantages that was raised in the existing system. This system will increase the accuracy of the neural network results by classifying the chest X-ray Gray scale image dataset using Deep learning algorithm. It has enhanced the performance of the overall classification results. It also predicts the Gray scale image of chest x-ray to find the accuracy in a more reliable way.

Prediction: It's a process of predicting the Pneumonia from the dataset. This project has effectively predicted the data from dataset by enhancing the performance of the overall prediction results.

Result generation: The final result was generated based on the overall classification and prediction. The performance of this proposed approach was evaluated using some measures like 1)**Accuracy**, 2) **Precision**, 3) **Recall**, 4) **F-measure**, 5) **Confusion matrix**

IV. RESULTS BY ALGORITHM





	precision	recall	f1-score	support
cancer 0	0.77	0.87	0.82	150
Normal 1	0.85	0.74	0.79	150
accuracy			0.80	300
macro avg	0.81	0.80	0.80	300
weighted avg	0.81	0.80	0.80	300

The ***Fig 2*** show that the percentage of accuracy increases as the number of epochs increases .The training accuracy and validation accuracy are plotted in the graphs where the accuracy increases with the epochs and another graph of training loss with validation loss are plotted which decreases with increasing epochs.

The main aim of this chapter is to introduce how lung cancer is an incredibly complex problem and how important is th reliability and validity of the proposed system.

V. HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are base of this research.

VI. AVAILABILITY OF DATA AND MATERIALS

The author confirms that the data supporting the findings of this research are available within the article.

VII. FUNDING ACKNOWLEDGEMENT AND CONFLICT OF INTEREST

The authors whose names are listed immediately above certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

VIII. REFERENCES

[1]. Schmidhuber, J. Deep Learning in neural networks: An overview. Neural Networks vol. 61 (2015).

[2]. Cheng, T. Y. D. et al. The international epidemiology of lung cancer: Latest trends, disparities, and tumor characteristics. J. Thorac. Oncol. 11, (2016).

[3]. Christopherp, T. & Banup, J. J. Study of Classification Algorithm for Lung Cancer Prediction. IJISET-International J. Innov. Sci. Eng. Technol. 3, (2016).

[4]. Fu, L., Ma, J., Ren, Y., Han, Y. S. & Zhao, J. Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features. in Medical Imaging 2017: Computer-Aided Diagnosis vol. 10134 (2017).

[5]. Affonso, C., Rossi, A. L. D., Vieira, F. H. A. & de Leon Ferreira de Carvalho, A. C. P. Deep learning for biological image classification. Expert Syst. Appl. 85, (2017).

[6]. Bhatt, J., Joshi, M. & Sharma, M. Early detection of lung cancer from CT images: nodule segmentation and classification using deep learning. in (2018). doi:10.1117/12.2309530.

[7]. Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. Computational Intelligence and Neuroscience vol. 2018 (2018).

[8]. Krishna, M. M., Neelima, M., Harshali, M. & Rao, M. V. G. Image classification using Deep learning. Int. J. Eng. Technol. 7, (2018).

[9]. Ponnada, V. T. & Srinivasu, S. V. N. End to end system for Pneumonia and lung cancer detection using deep learning. Int. J. Eng. Adv. Technol. 8, (2019).

[10]. Gao, J., Jiang, Q., Zhou, B. & Chen, D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. Mathematical Biosciences and Engineering vol. 16 (2019).

[11]. Ponnada, V. T. & Srinivasu, S. V. N. Efficient CNN for lung cancer detection. Int. J. Recent Technol. Eng. 8, (2019).

[12]. Sirazitdinov, I. et al. Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Comput. Electr. Eng. 78, (2019).

[13]. Bhatia, S., Sinha, Y. & Goel, L. Lung cancer detection: A deep learning approach. Advances in Intelligent Systems and Computing vol. 817 (2019).
[14]. Xu, Y. et al. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin. Cancer Res. 25, (2019).
[15]. Cao, W., Wu, R., Cao, G. & He, Z. A Comprehensive Review of Computer-Aided Diagnosis of Pulmonary Nodules Based on Computed Tomography Scans. IEEE Access vol. 8 (2020).
[16]. Xu, Y. M. et al. Deep learning in ct images: Automated pulmonary nodule detection for subsequent management using convolutional neural network. Cancer Manag. Res. 12, (2020).
[17]. Shak, K. et al. A new semi-supervised self-training method for lung cancer prediction. arXiv (2020).
[18]. Asuntha, A. & Srinivasan, A. Deep learning for lung Cancer detection and classification. Multimed. Tools Appl. 79, (2020).
[19]. Algan, G. & Ulusoy, I. Image classification with deep learning in the presence of noisy labels: A survey. Knowledge-Based Syst. 215, (2021).

Cite this article as :

Vegi Jeevana , B Sai Ram Charan , Dr . S. Usha Kiruthika, "Lung Cancer Prediction Using Chest X-Rays", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 2, pp. 236-242, March-April 2021. Available at doi : https://doi.org/10.32628/IJSRST218236
Journal URL : https://ijsrst.com/IJSRST218236