International Journal of Research in Engineering, Science and Management
Volume 4, Issue 3, March 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Deep Learning in the Diagnosis of Lung Cancer

Aromal Sen[1*], Shon J. Mathew[2], Ashley Mathew[3]
[1,2,3]*Department of Computer Applications, Saintgits College of Applied Science, Kottayam, India*

***Abstract*: Cancer is one the deadliest disease and is regarded as the second dominant cause of death around the world with an estimated death of more than 9 million cases around the world and about 1 in 6 deaths is due to cancer. Lung Cancer is one the most common cancers that cannot be ignored and is one of the main reasons for the increase of deaths in the world among both women and men with an impressive rate of about five million. In recent few years, so many Computer Aided Diagnosis (CAD) are used for the detection of many diseases. Lung cancer detection in an early stage has been a crucial part for the sustainability of people's life. So many experts and doctors use Deep Learning Techniques to determine the disease in the early stage itself. In this study we use Deep Learning techniques namely Convolutional Neural Network (CNN) for the lung cancer diagnosis and get the details in high efficient and accurate manner. The work done is a CNN network model with the help of LIDC dataset and LSTM network to analyze cancer nodules at early stages.**

***Keywords*: Computer Aided Diagnosis (CAD), Lung Cancer, Deep Learning, Conventional Neural Network (CNN), LIDC, LSTM.**

1. Introduction

Cancer is one of the main reasons for death in the world, with an impressive rate of about 18 million cases in average per year. Approximately, 40% people will be diagnosed with cancer at some point of their lifetime with an overall mortality rate of 171.2% per 1 lakh per year. Lung cancer is one of the most common cancers and the largest cause of cancer related deaths every year due to its aggressiveness and advanced stage when detected. According to World Health Organization, there were 2.09 million cases and 1.76 million lung cancer related deaths in 2018.Lung cancer is regarded as one of the most aggressive, with 5-year survival rate of only 5-10%. However, results are significantly better with a cancer detected at an early stage with 10-year survival rate for stage one cancer up to 75%. It is critical to detect cancer as early as possible since it gives more time for treatment and gives better chances of survival.

Therefore, regular examinations are necessary especially for high risk group. It is where the automated diagnosis can help doctors in their field. This is where computer aided diagnosis with a system that can automate segmentation and evaluation of lung lesion improve healthcare performance. However, to diagnose lung cancer at an early stage, the pulmonologist and cardiologist started using chest computed tomography (CT) images.

In order to support medical decision making, computer-aided diagnosis (CAD) system is an effective medical diagnosis tool and it benefits today's medical image practically. Computer-aided diagnosis has been a major field of research for the past few years. These CAD systems are mainly used by the doctors and medical professionals to provide an extra second opinion in order to obtain accurate and effective diagnosis. Medical Professionals uses these features and Machine Learning (ML) techniques such as Support Vector Machine (SVM) to classify the nodule. Even though many other works use machine learning frameworks, the disadvantage regarding these methods is that, to get best performance it uses many parameters that need to be hand-crafted, thus making it hard to reproduce the exact results. Additionally, this makes these approaches helpless to the variability between different CT scans and different screening parameters. So nowadays medical professionals use deep learning techniques. It has achieved big impact and success in computer vision. Deep Learning is a type of machine learning that teaches the computers to learn from different experience and examples and understand the world in terms different concepts. Since the computers gets the different kinds of knowledge from experience, humans do not need to specify all relayed information to the computer one by one. Computer learn complicated concepts by making them out of simpler ones since the different levels of concepts allows the computer to learn them easily. Deep learning provides the opportunity to increase the accuracy of early detection of diseases.

2. Existing Methodology

Many of the normal image processing systems are deficient in detecting lung carcinomas at early stages of development. The primitive methods which are used in analysis of lung cancer are poor in providing precise or sufficient results which may result in more fatalities every year. The existing designs use simple Computed Tomography images in detection of lung cancers which are deficient as they are just two-dimensional images which makes it difficult for professionals to analyse, detect or calculate cancerous cells in the lungs. The extraction of features from a CT scan image is limited up to a certain extent. This system lacks in a number of features. It is strenuous to detect small malign nodules from a big picture of the whole lungs. So, we propose use of Convolutional Neural Network (CNN) using deep learning in providing a detailed picture of CT scan to improve the accuracy in detection of lung

*Corresponding author: aromalajith07@gmail.com

carcinomas at early stages.

3. Proposed Methodology

In the proposed system figure we use Convolutional Neural Networks along with the help of dataset and RNN to figure out if a person is normal or has lung cancer.

Detection of malignant cancerous nodules needs huge amount of processing on CT scan images and the following procedures are to be done,

1. Input layer
2. Pre-processing
3. Segmentation
4. Extraction of features using CNN
5. Mapping of features using RNN(LSTM)
6. Result

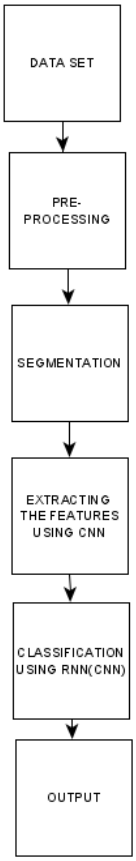

Fig. 1. Flow diagram of the various stages of detection of malignant cancerous nodules

A. Input layer

The DICOM images are processed into the network. For that, we use an LIDC dataset which has a collection database of a broad number of CT scan images.

B. Pre-Processing

The DICOM images are difficult to understand and needs further efficient processing in order to extract sufficient details from it. Therefore, computer tomography scan images are further developed into .png images or .jpeg. The images are made noise free and processed into grey scale images.

C. Segmentation

Segmentation is being used in this work to develop the greyscale images into binary image. This is done by using the threshold method. The binary images are further converted back to binary images suing addition methods.

D. Extraction of features using CNN

Using CNN, the dataset is extracted. It is used to extract more information from the CT scan image. It consists of multiple layers. The convolutional layer in CNN is used to form featured maps from the CT scan images. Inputted scan images are filtered and it provides featured maps. The use of max-pooling layer is to reduce the feature maps.

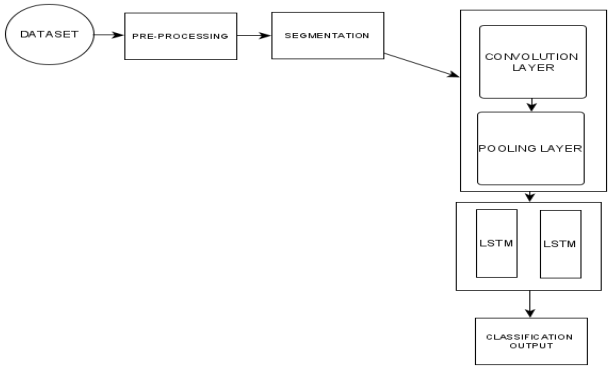

Fig. 2. Classification using RNN (LSTM)

The RNN(LSTM) networks have a pivotal role in mapping the features. The features are mapped to the suited layer. It Is the final layer of CNN-LSTM network.

4. Data Set

The dataset being used in this work is LIDC-IDRI dataset. It includes 1018 cases and was created jointly by medical companies and centers. The Lung Database Consortium dataset consist of CT scans having modules:

1. >or=3mm
2. <3mm
3. Non-nodule>or=3mm

5. Literature Survey

In [1], Ahmed Cinar proposed a method in which tensor flow is used in order to perform the diagnosis of lungs.

This is highly favoured of its adaptability and flexibility. This method uses 3D network architecture and dat from data sets such as LIDC-IDRI dataset. It helps in diagnosis of cancer at early stages. It is proven successful in most scenarios and can be further improved with larger dataset and more efficient architecture.

In [2], Amrit sreekar, Karthika, Sneha, Ganesh, and Jyothika used to deep learning in order to identify malignant pulmonary nodules from CT scans. The DICOM images are pre-processed and converted to Hounsefield units and segmentation was done to make it better. The C3D architecture was used in extracting more features about the images. The final result is used to have a precise location of the malignant nodules and estimate the

percentage of malignancy.

In [3], Riksha Mharke, Kannan, Ruchta propose a way of using the computer tomography images in order to detect or analyse lung cancer at early stages with the help of deep learning techniques. They use CAD system for the purpose. The CT scan images are further processed and segmented and with the help of CNN networks, more information is extracted from it. By using recurrent neural network, it is used to process the images and show if the images are normal or whether lung cancer is detected. It is claimed to be the most accurate procedure with 97% accuracy.

In [4], Naji Khosravan and Ulas Bagci set out to test the hypothesis that joint learning of false positive (FP) nodule reduction and nodule segmentation can improve the computer aided diagnosis system performance. To support this study, they propose a 3D deep multi-task CNN to tackle these problems. To overcome the limitation of lack of labled information for the 3D segmentation task, Semi-supervised approach can be used.

In [5], Shuo Wang, Zhenyu Liu, Xi Chen, Yongbei Zhu, Hongyu Zhou proposed an unsupervised deep learning method to take credit of the unlabeled data for survival analysis, and demonstrated a far better performance than using hand-crafted methods. They recommended a residual convolutional auto encoder and trained the model using pictures from 274 patients without survival time. The analysis results showed that unsupervised deep learning feature achieved better performance with a C-Index of 0.70 than using hand-crafted features with a C-Index 0f 0.62.

6. Conclusion

The work done is a CNN network model with the help of LIDC dataset and LSTM network to analyze cancer nodules at early stages. This is done by using a CT scan image of the lungs and further processing it using deep learning features of CNN. The Developed system is precise in nature and aims to provide correct results. It improves the detection rate of the systems by a huge margin with an accuracy of more than 95% which is a great achievement towards diagnosis of cancerous nodules at early stages.

References

[1] Emine CENGİL "A Deep Learning Based Approach to Lung Cancer Identification", 2017.

[2] Ahmet Kadir Arslan, Şeyma Yaşar, Cemil Çolak, "An Intelligent System for the Classification of Lung Cancer Based on Deep Learning Strategy."

[3] Pujan Kandel, Sarfaraz Hussein, Michael B. Wallace, Candice W. Bolan, and Ulas Bagci: Lung Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches, 2018.

[4] Ryota Shimizu, Shusuke Yanagawa, Yasutaka Monde, Hiroki Yamagishi, Mototsugu Hamada, Toru Shimizu, and Tadahiro Kuroda, "Deep Learning Application Trial to Lung Cancer Diagnosis for Medical Sensor Systems", 2016.

[5] Amrit Sreekumar, Karthika Rajan Nair, Sneha Sudheer, Ganesh Nayar H. and Jyothisha J. Nair: "Malignant Lung Nodule Detection using Deep Learning", 2020.