IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 4, May 2022

# Lung Cancer Detection using CT Scans Images

**Avantika Bhujade[1], Disha Daga[2], Jagruti Bisen[3], Mansi Baraliya[4], Dr. S.B. Patil[5]**

Students, Department of Electronics and telecommunication[1,2,3,4]

Head of Department, Department of Computer Science and Engineeing[5]

Shri Sant Gajanan Maharaj College of Engineering, Shegoan, Maharashtra, India

**Abstract:** *Cancer is the most dangerous disease which causes deaths. Early treatment of such disease is very important, if not it may cause deaths. So, to overcome this complication, advanced and early diagnosis is mandatory. By the help of modern technologies like artificial intelligence we can detect this deadly cancer and could help the patients. Artificial Intelligence technique is the most important convolution neural network that can easily identify the cancer type with high accuracy in very less time. So, in this paper we took many research paper into consideration and found that convolution neural networks are very popular to identify pulmonary cancer than any other algorithms in deep learning.CNN gets over the age compared to its antecedent because it detects the suspected and the non suspected region without any human supervision. The main focus is to maximize the accuracy and specificity. Nevertheless, it will help the radiologists and medical personal to dig into treatments and analyze the proper cure effectively.*

**Keywords:** Lung Cancer

## I. Introduction

Lung cancer is listed among a dangerous disease in the world. The survival and death rates proportions are daunting. If the cancer is detected and treated at early phase, the patient's life can be saved. Further it also helps the radiologist to scan and interpret about the disease easily. Sometimes the nodules are too small to detect by naked eyes and here the CT scans imagining comes in limelight. The screening process properly examines the CT scans and gives results with least failure.

The 5-year survival rate tells you what percent of people live at least 5 years after the cancer is found. Percent means how many out of 100. The 5-year survival rate for all people with all types of lung cancer is 21%. The 5-year survival rate for men is 17%. The 5-year survival rate for women is 24%. The 5-year survival rate for NSCLC is 25%, compared to 7% for small cell lung cancer. In cancer a large group cells mushroomed abnormally and unrestrained. In multiple cases they grow beyond their cell boundaries and seize the neighbouring organs/other parts. Further this process is called metastasizing and it is deadly. The tumours are namely neoplasm and malignant tumour.

According to study, the survival rate is affected by plenty of factors, it includes the habits of the patients, the isoforms of the disease and its stages during treatment. Low dose computed tomography is trusted and deep-rooted technique for lung cancer screening. Despite of such impairment India does not a supervised, well established screening agenda. Multiple parameters like capital generation, logistics, statistical data and high false true ratio influences the implementation.

## II. Literature Review

*Suresh Kallam*, [1] 13 Oct 2021 *Adaptive Diagnosis of Lung Cancer by Deep Learning Classification Using Wilcoxon Gain and Generator* This paper proposes a novel Wilcoxon Signed-Rank Gain Preprocessing combined with Generative Deep Learning called Wilcoxon Signed Generative Deep Learning (WS-GDL) method for lung cancer disease diagnosis.

*Himal Chand Thapa*, [2] October 2020 *Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images* This research work presents lung cancer detection using histopathological images. A convolutional neural network (CNN) was implemented to classify an image of three different categories benign, Adenocarcinoma, and squamous cell carcinoma. The model was able to achieve 96.11% and 97.20% of training and validation accuracy.

*Endalew Simie* [3] *Lung cancer detection using Convolutional Neural Network (CNN)* In this project, used different lung segmentation and nodules segmentation methods. work has consisted of preprocessing, and lung segmentation by using thresholding, and also used the U-net model for detection of the candidate nodules of the patient's lung CT scan and classification methodology

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 4, May 2022

*Hamdalla F. Al-Yasriy* [4]July 2020 *Diagnosis of Lung Cancer Based on CT Scans Using CNN* By using a convolutional neural network technique with AlexNet architecture for helping with the diagnosis of the patient's cases. After the end of the training which done randomly, this model gives after 86 epoch of the 100 epochs of the training, an overall accuracy of 93.548%.

*Debnath Bhattacharyya* [5] 12 Mar 2021 *3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation* In this paper, the lung nodule classification using the improvised 3D AlexNet with lightweight architecture. Conducted the binary classification on computed tomography images from the LUNA 16 database conglomerate and database image resource initiative.

*Margarita Kirienko*[6]30 Oct 2018*Convolutional Neural Networks Promising in Lung Cancer T-Parameter Assessment on Baseline FDG-PET/CT*The algorithm developed and tested in the present work achieved an accuracy of 87%, 69%, and 69% in the training, validation, and test sets, respectively.

*Diego Riquelme*[8]8 January 2020*Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans* While 3D CNN show the best performances overall, 2D CNN techniques also give some very interesting results. Popular techniques such as U-Net, Faster R-CNN, Mask R-CNN, YOLO, VGG, ResNet, and so forth, were used to build the convolutional networks for nodule detection and false positive reduction.

*Lei Cong*[9]*Deep Learning Model as a New Trend in Computer-aided Diagnosis of Tumor Pathology for Lung Cancer*DL-based lung cancer pathology CAD and scientific research has shown good performance and great potential.the authoritative lung cancer pathology database is scarce, which makes it difficult to explain the universality of the training model.

## III. Image Selection

DICOM is an acronym for "Digital Imaging and Communications in Medicine".

DICOM is the degree for communication and management of medical data related with images. DICOM is most ordinarily used for storing and transmitting medical images which enables the combination of medical imaging devices like scanners, servers, printers,

Network, hardware and architecture.

For lung cancer detection the best and the efficient technique for diagnosis of a disease is by using CT-scan images. The CT-scan images helps us to suspect the cancer nodule effectively. Surprisingly even it is a promising tool in medical sciences for diagnosis of disease. Eventually by using the CNN algorithm and architecture with the help of deep learning the detection becomes accurately possible. It has done the work of CNN in biomedical image analysis is extraordinary.

### 3.1 Dataset Collection

The Data set used in the project are collected from The National Cancer Imagining Archive. This data set consist of CT and Pet-CT DICOM images of lung cancer. The images are acquired from patients with suspicion of lung cancer and who underwent standard of care lung biopsy and PET/CT.

We have trained the machine with nearly 6000+ images. The dataset is real and it trains the model effectively. Since starting medical imaging is deeply rooted with x-rays through medical imaging lenses. From further research and development medical sciences have opted digital imaging format. The digital format is called as DICOM and it is in JPEG, png format. DICOM technology offers a wide range of application and advantages. To the fact that it only takes a second to develop the image. On the contrary x rays' images would take hours of processing. This is time efficient and helps in fast curation of patients, increases the efficiency of healthcare workflow.

Regardless to this, DICOM provides zoomed to enhancement to view the physician is observing. Adjusting the brightness, contrast and many other editing features makes it special and extremely useful. Further it helps the physicians to accurately diagnose and to plan a complete and sustainable treatment plan.

## IV. Methodology

In this suggested deep learning architecture for recognition of lung cancer sensing is broadly divided into two sections first contains the filtering model succeeded by CNN network.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 4, May 2022

Impact Factor: 6.252

A filtering paradigm functions the initial stage of the performing activity, it excludes the irrelevant details and choose the most enlighten information for diagnosis. In the next stage CNN model is proposed. The passes through different layers namely max pooling, min-pooling, etc.

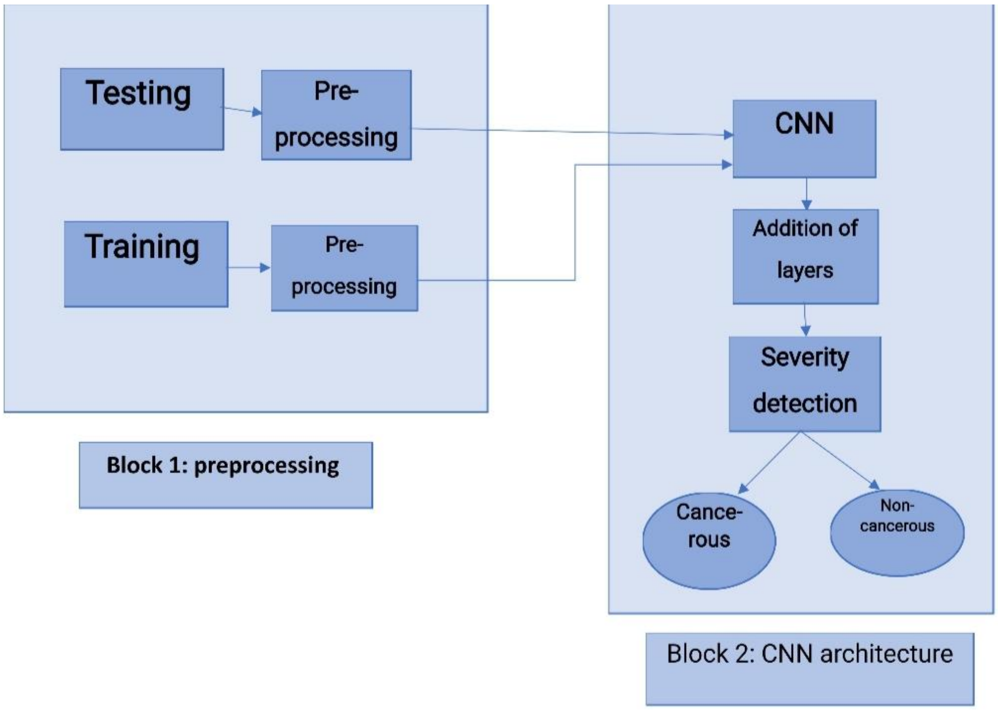



**Figure 1:** Proposed methodology

**Step 1: Filtering Model**

A Computed Tomography (CT) scan images does not only contain the lung but also it is circumferential with other substances like tissues, water, bones, air. The existence of these substances is pointless. It negatively affects the performance of the proposed prototype. It increases the unnecessary noises in an image. Thus reduces the accuracy. Hence elimination of these mass increases the percentage of accuracy.

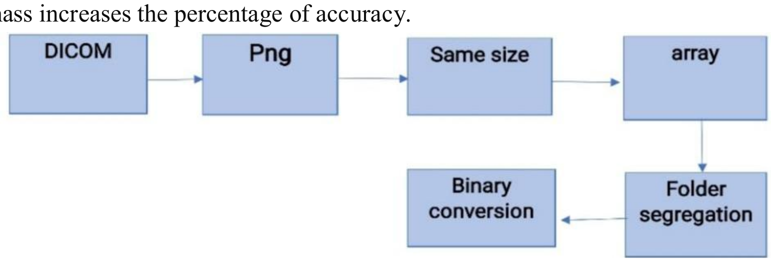



**Figure 2:** Filtration model

DICOM images are superior in its own type. It does not need a long pre-processing. Features can be easily extracted and evaluation of lung nodule is uncomplicated. Further the extracted features(details) are fed to the training model and analysis is corroborated.

The DICOM images has been converted to PNG format. The strength of the method includes

- Higher accuracy in lung nodule detection
- Removes the unnecessary noise that causes false detection
- Data visualization is improved
- It speeds up the training process
- The risk of over fitting is reduced.

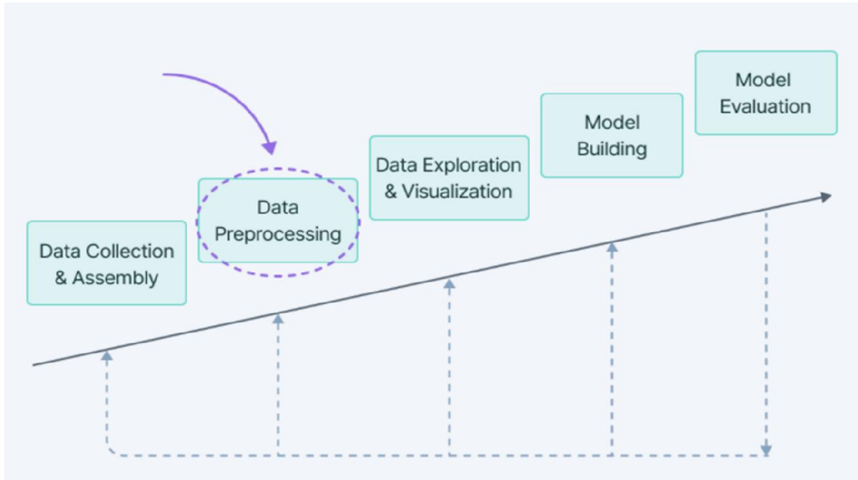



**Figure 3:** Data processing

**Step 2: CNN Development analysis**

The remarkable benefit of Deep learning it does not need any human assistance, after training and testing the machine gets trained and extract the features. During training, the network learns to extract features. All you have to do now is send the image to the network (pixel values). Convolutional neural networks are built with the goal of reducing the number of parameters in the network's architecture for image classification. Convolutional neural networks are made up of a sequence of layers that are grouped together based on their features and functions. The architecture of a ConvNet is similar to that of the human brain.

The process of increasing the quantity and complexity of data is known as data augmentation. Rather than transforming existing data, we will acquire new information. Data augmentation is a phase in deep learning. Because deep learning requires large amounts of data and it's not always practical to gather thousands or even millions of photos, data augmentation comes into play. It enables us to make the most of our datasets. Furthermore, patients can log in to their page at any moment for future references to be simply inferred.

CNN are a class of neural networks designed to learn, during training, convolution parameters from a set of available data. In general, they are comprised of different layers such as convolutional layers, deconvolutional layers, pooling layers and so forth. Different architectures were proposed in recent years to improve the performance and overcome some limitations of the standard CNN. [7]

**Operation on Proposed Architecturere:**

The working is divided into two parts i.e., training and testing. We had 6000+ sample images. The results of the training and testing phases are fed into CNN layers, which classify the images and display the result.

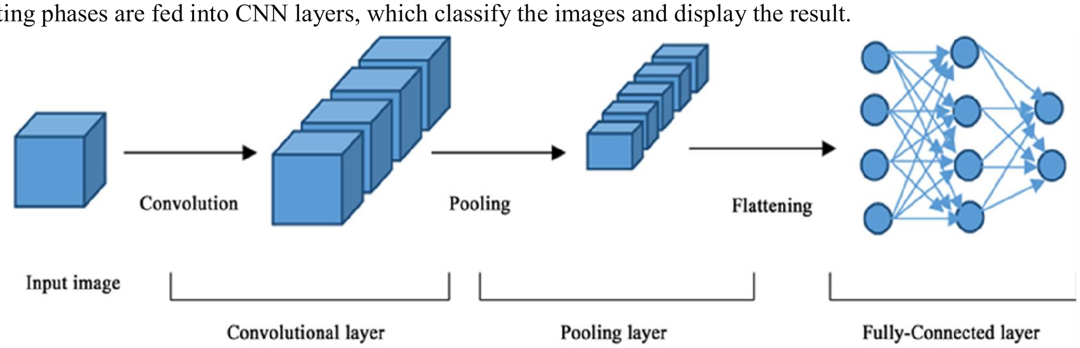



**Figure 4:** CNN architecture

- A CNN structure i.e. the convolutional neural network works on spatial correlation of the input data to its advantage. CNN was created to process input pictures precisely. Their architecture is more definite after then, consisting of two major components. Because this form of neural network serves as a feature extractor, the first block establishes its uniqueness.

- The first layer uses many convolution kernels to filter the picture and produce "feature maps," further they are normalised (using an activation function). This method can be repeated multiple times. Finally, the last feature map's values are concatenated into a vector. This vector specifies the first block's output and the second block's input.
- The second block is not unique to CNNs; in fact, it appears at the conclusion of all categorization neural network. The input values are transformed to return a new vector to the output. The chance that the picture corresponds to class I is represented by element I of this last vector, which has as many elements as there are classes. As a result, each element has a value between 0 and 1, and the total value is 1. The probabilities are recorded and calculated by the last layer, this last layer of second block uses the logistics function.

Kernel traverses across the input data and further it performs a dot product on the subregion of data. This matrix produces the output as a dot matrix Certain characteristics in a picture can be detected using convolution neural layers. A convolution product is another name for it.

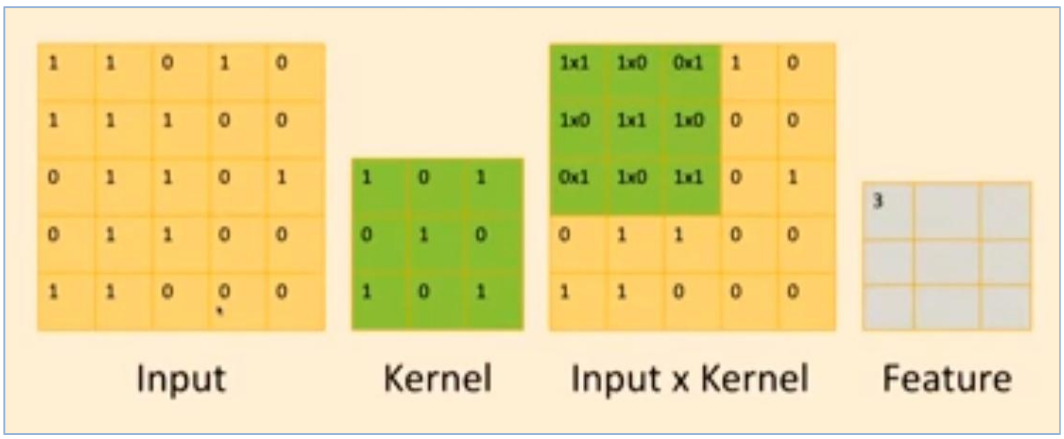



**Figure 5:** CNN block 1

**Max Pooling Layer:**

- Helps in **reduction** of the dimensions of the datasets.
- Reduces parameters, reduces overfitting
- Uses the max value from a cluster of neurons at the prior layer
- Uses as best extractor

**Applied CNN model:**

In our model we have implemented 6 layers. They follow the pattern as:

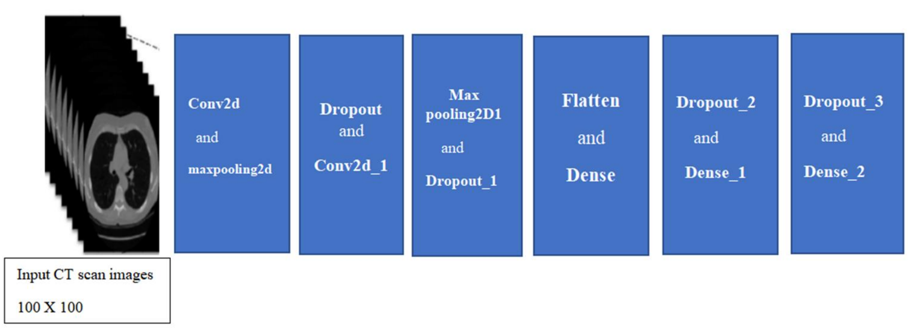



**Figure 6:** Convolutional layers

IJARSCT
ISSN (Online) 2581-9429
IJARSCT
Impact Factor: 6.252
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022

```
Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 96, 96, 16)        1216
 max_pooling2d (MaxPooling2D  (None, 48, 48, 16)       0)

 dropout (Dropout)           (None, 48, 48, 16)        0
 conv2d_1 (Conv2D)           (None, 46, 46, 32)        4640

 max_pooling2d_1 (MaxPooling  (None, 23, 23, 32)       02D)
 dropout_1 (Dropout)         (None, 23, 23, 32)        0

 flatten (Flatten)           (None, 16928)             0
 dense (Dense)               (None, 1024)              17335296

 dropout_2 (Dropout)         (None, 1024)              0
 dense_1 (Dense)             (None, 512)               524800

 dropout_3 (Dropout)         (None, 512)               0
 dense_2 (Dense)             (None, 2)                 1026

=================================================================
Total params: 17,866,978
Trainable params: 17,866,978
Non-trainable params: 0
```

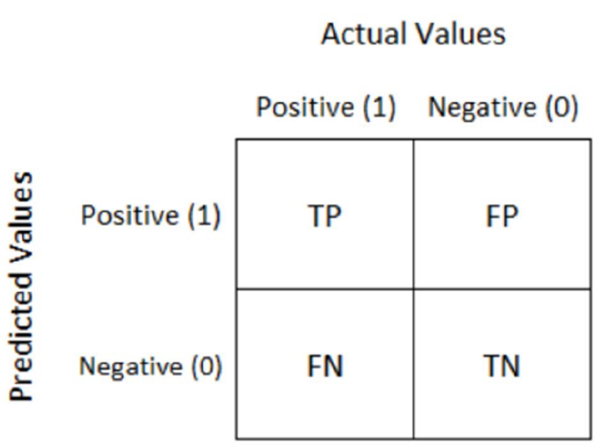
```
Epoch 1/5
148/148 [================] - 78s 522ms/step -
loss: 0.4122 - accuracy: 0.8931

Epoch 2/5
148/148 [================] - 71s 481ms/step -
loss: 0.0033 - accuracy: 0.9998

Epoch 3/5
148/148 [================] - 73s 493ms/step -
loss: 0.0029 - accuracy: 0.9994

Epoch 4/5
148/148 [================] - 68s 459ms/step -
loss: 0.0036 - accuracy: 0.9998

Epoch 5/5
148/148 [================] - 73s 493ms/step -
loss: 0.0049 - accuracy: 0.9994
```

The stacking of the Convolutional layers allows a hierarchical decomposition of the input and the convolutional layers are used to help the computer to determine the features that could be missed in simply flattening an image into its pixel values.

**A Confusion Matrix** is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

There are two possible predicted classes in confusion matrix and they are, "yes" and "no". If we were predicting the presence of a disease, for example, "yes", it means that they actually have the disease, and "no" would mean they don't have the disease. The most basic terms which are used in the confusion matrix are as follows:

- True Positives (TP): In this case, we estimated yes i.e they have the disease, and actually they do have the disease.
- True Negatives (TN): We estimated no, and they don't have the disease.
- False Positives (FP): We estimated yes, but they don't actually have the disease. False positives are also known as Type 1 error.
- False Negatives (FN): We estimated no, but they actually do have the disease. False negatives are also known as Type 2 error.




Accuracy = TP+TN/TP+FP+TN+FN

**Figure 7:** Confusion metrics

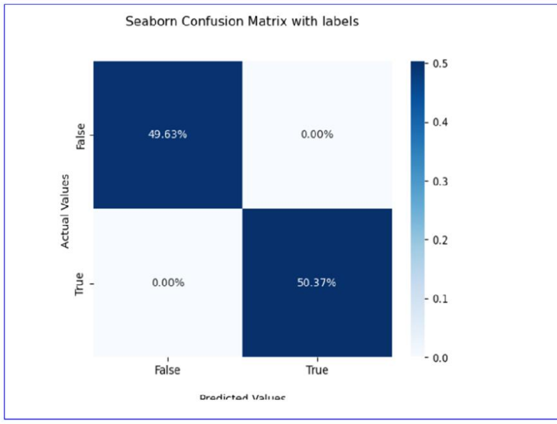
IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 4, May 2022



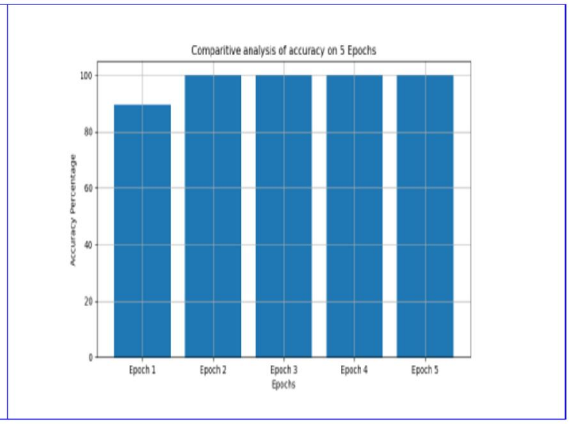




**Figure 8:** Visualization of model

| Epoch | Time in seconds | Error | Accuracy |
|---|---|---|---|
| 1 | 78 Seconds | 0.4122 | 0.8931 |
| 2 | 71 Seconds | 0.0033 | 0.9998 |
| 3 | 73 Seconds | 0.0029 | 0.9994 |
| 4 | 68 Seconds | 0.0036 | 0.9998 |
| 5 | 73 Seconds | 0.0049 | 0.9994 |

**Figure 9:** Analysis of model

## V. Future Work

Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods, or simply curiosity. Capsule neural network is immensely powerful, it uses capsules rather than neurons. They have the ability to keep a track on direction of feature.  It may increase the scope of the project. By using deeper neural hacks, we can also justify the level of cancer patient is suffering. Adding more to it we can collect data regarding severity percentage of heart (i.e the percentage affected).

## VI. Conclusion

Deep learning algorithm has its own benefits over machine learning as its ability in implementing feature engineering on its own. After training machine examines the data and search for features and work to provide a quicker learning. This project has used the DICOM ct scans images of lung nodule. The training and testing of images are done in pre-processing and followed by feature extraction of DICOM images are done. After the successful training and testing, CNN algorithm classifies the input lung nodule as either cancerous or noncancerous. The following output will be displayed accordingly. CNN has efficient accuracy; deep CNN networks is used for better analysis and configurations of lung nodule and to classify the lung images as abnormal and normal images. (add accuracy and data obtained after testing )

ISSN (Online) 2581-9429

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 4, May 2022

## REFERENCES

**[1].** Adaptive Diagnosis of Lung Cancer by Deep Learning Classification Using Wilcoxon Gain and Generator

**[2].** Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images

**[3].** Lung cancer detection using Convolutional Neural Network (CNN) Hamdalla F. Al-Yasriy

**[4].** Diagnosis of Lung Cancer Based on CT Scans Using CNN

**[5].** 3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-

**[6].** Convolutional Neural Networks Promising in Lung Cancer T-Parameter Assessment on Baseline FDG-PET/CT 8 January 2020

**[7].** Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans

**[8].** Deep Learning Model as a New Trend in Computer-aided Diagnosis of Tumor Pathology for Lung Cancer