DAYANANDA SAGAR UNIVERSITY

MINI PROJECT REPORT ON
"LUNG CANCER DETECTION"

Submitted By

ABHAY SHREEKANT SHASTRY (ENG18CS0008)

B. ANTONIO MERVYN (ENG18CS0056)

VARUN MENON (ENG18CS0314)

of

BACHELOR OF TECHNOLOGY

in

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

at

DAYANANDA SAGAR UNIVERSITY
SCHOOL OF ENGINEERING, BANGALORE-560068

7TH SEMESTER
(Course Code: 16CS472)

MACHINE LEARNING FOR HEALTHCARE

UNDER THE SUPERVISION OF

DR JAYITA SAHA

DAYANANDA SAGAR UNIVERSITY

CERTIFICATE

This is to certify that the Networks and USP Mini-Project report entitled **“Lung Cancer Detection”** being submitted by **ABHAY SHREEKANT SHASTRY** USN **ENG18CS0008** to Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, for the 7th semester B.Tech C.S.E of this university during the academic year 2020-2021.

Date:___________

____

Signature of the Faculty in Charge

DAYANANDA SAGAR UNIVERSITY

CERTIFICATE

This is to certify that the Networks and USP Mini-Project report entitled **"Lung Cancer Detection"** being submitted by **B**. **ANTONIO MERVYN** USN **ENG18CS0056** to Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, for the 7th semester B.Tech C.S.E of this university during the academic year 2020-2021.

Date:___________ _____________________________

Signature of the Faculty in Charge

DAYANANDA SAGAR UNIVERSITY

CERTIFICATE

This is to certify that the Networks and USP Mini-Project report entitled **"Lung Cancer Detection"** being submitted by **VARUN MENON** USN **ENG18CS0314** to Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, for the 7th semester B.Tech C.S.E of this university during the academic year 2020-2021.

Date:___________ ____________________________

Signature of the Faculty in Charge

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of a task would be incomplete without the mention of the people who made it possible and whose constant guidance and encouragement crown all the efforts with success.

We are especially thankful to our **Dean, Dr. Srinivas A**, for providing necessary departmental facilities, moral support and encouragement.

We are very much thankful to **Dr Jayita Saha** for providing help and suggestions in completion of this mini project successfully.

We have received a great deal of guidance and co-operation from our friends and we wish to thank all that have directly or indirectly helped us in the successful completion of this project work.

ABHAY SHREEKANT SHASTRY (ENG18CS0008)

ANTONIO MERVYN (ENG18CS0056)

VARUN MENON (ENG18CS0314)

CONTENTS

1. ABSTRACT

Early cancer stage prediction and categorization are required in order to take countermeasures for treatment and diagnosis. The majority of the time, scanned pictures are used to determine the presence of lung small cell carcinoma.

This project proposes a modified convolutional neural network called EfficientNet to diagnose the three forms of lung cancer based on the input scanned images. It implements EfficientNet to diagnose lung cancer efficiently and accurately The proposed model is made up of three sub-paths and a main path. At a low level, the main path extracts tiny features and produces feature maps.To complete the classification process, the sub-paths are responsible for transferring the medium and high level feature maps to fully connected layers.

This project aims to help the early detection of lung cancer as one of the main methods of preventing deaths from lung cancer is to detect its presence at an early, treatable stage, thereby assisting the successful diagnosis of lung cancer.

2. INTRODUCTION

Lung cancer, commonly referred to as bronchial carcinoma, is a malignant lung tumour marked by uncontrolled cell proliferation in lung tissues. Lung carcinomas arise from epithelial cells that have been converted into malignant cells, or from tissues made up of epithelial cells. Other lung cancers, such as the rare sarcomas of the lung, are caused by the malignant transformation of mesenchymal cells' connective tissues (nerve, fat, muscle, and bone). Lung cancer can also be caused by lymphomas and melanomas (cells of the lymphoid and melanocyte lineages).

It is the most common form of cancer and causes more deaths that colon cancer, breast cancer and ovarian cancer combined. Therefore, it is a disease that requires significant attention to treat

Current methods make fibrotic lung diseases difficult to treat, even with access to a chest CT scan. In addition, the wide range of varied prognoses create issues organizing clinical trials. Finally, patients suffer extreme anxiety in addition to fibrosis-related symptoms from the disease’s opaque path of progression. Therefore, early diagnosis is the best outcome to treat lung cancer .

This project describes computer aided diagnosis centered on deep learning approaches that improve diagnostics of lung cancer. The early diagnosis of disease’s severity should be possible. Deep Learning will help us tell which spectrum that individual will fall on and it would also become a faster detection process. The Deep Learning method that’s used is a modified Convolutional Neural Network called Efficientnet. This model will not only help accurately predict lung cancer but is also small in size and easy to practically integrate into applications for commercial medical use.

3. PROBLEM STATEMENT AND PROBLEM DESCRIPTION:

Lung cancer is by far one of the most deadly forms of cancer, causing 25% of all cancer deaths. Advancements in treating lung cancer in its later stages have not bore fruit. So while we wait for advancements in cancer treatment, our attention needs to go to early detection which will help offset it before it reaches early stages. CT scans are normally used for early detection of cancer but it is difficult for the human eye to find tumors and carcinogenic lesions.

Therefore Deep learning and computer vision methods can be used for detection of lung cancer from CT scans. The problem with traditional deep learning techniques is the fact that usability is normally a tradeoff for accuracy. A lot of models are extremely accurate with perfect precision and recall values but end up being extremely huge in size which becomes an issue when the model gets exported as a JSON file inorder to be integrated into web or mobile applications.

4. EXPERIMENTAL SETUP:

4.1 Dataset:

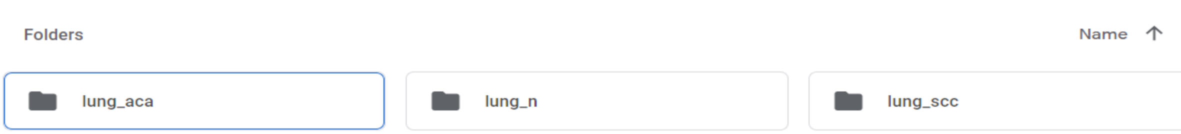
This dataset contains 15,000 histopathological images with 3 classes. All images are 768 x 768 pixels in size and are in jpeg file format.

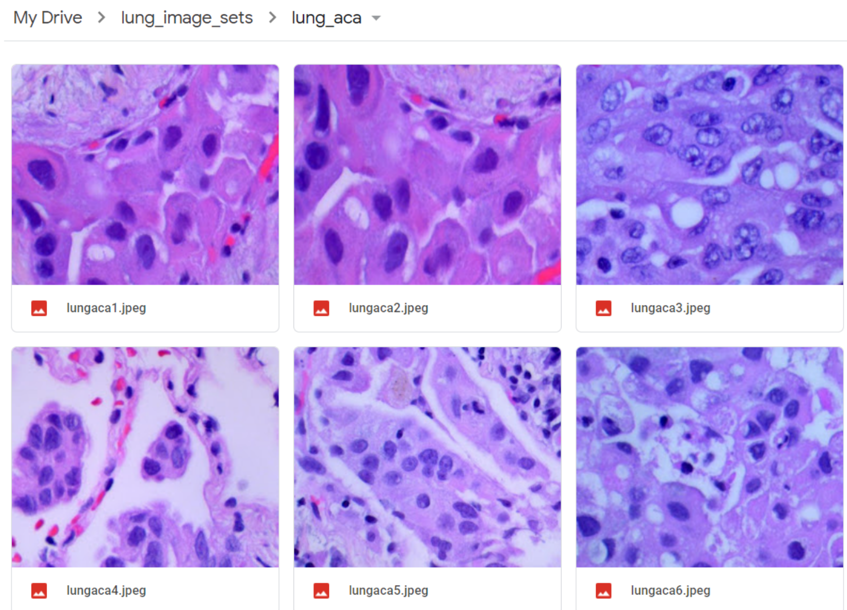
The images were generated from an original sample of HIPAA compliant and validated sources, consisting of 750 total images of lung tissue (250 benign lung tissue, 250 lung adenocarcinomas, and 250 lung squamous cell carcinomas) and augmented to 15,000 using the Augmentor package.

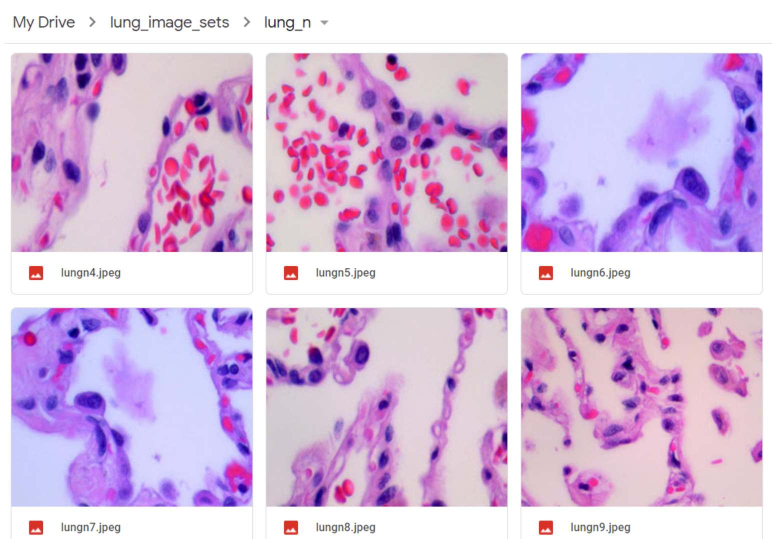
There are 3 classes in the dataset, each with 5,000 images, being:

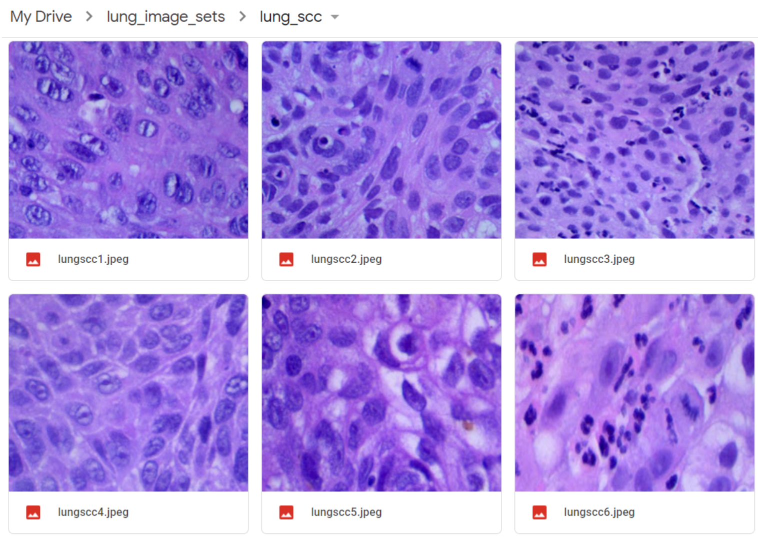
- Lung benign tissue
- Lung adenocarcinoma
- Lung squamous cell carcinoma

The images are in jpeg format.




My Drive > lung_image_sets > lung_aca
lungaca1.jpeg
lungaca2.jpeg
lungaca3.jpeg
lungaca4.jpeg
lungaca5.jpeg
lungaca6.jpeg


My Drive > lung_image_sets > lung_n
lungn4.jpeg
lungn5.jpeg
lungn6.jpeg
lungn7.jpeg
lungn8.jpeg
lungn9.jpeg



4.2 Model Architecture :

The model consists of multiple 3x3 convolution kernels with 3 5 x 5 convolutions sandwiched between every 5 convolution layers. The output shape will have the same dimensions as the input shape. The activation function used is Swish.

4.3 Kernel:

In the context of a convolutional neural network, a convolution is a linear operation that involves the multiplication of a set of weights with the input, much like a traditional neural network. Given that the technique was designed for two-dimensional input, the multiplication is performed between an array of input data and a two-dimensional array of weights, called a filter or a kernel.

The filter is smaller than the input data and the type of multiplication applied between a filter-sized patch of the input and the filter is a dot product. A dot product is the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed, always resulting in a single value. Because it results in a single value, the operation is often referred to as the “scalar product“.

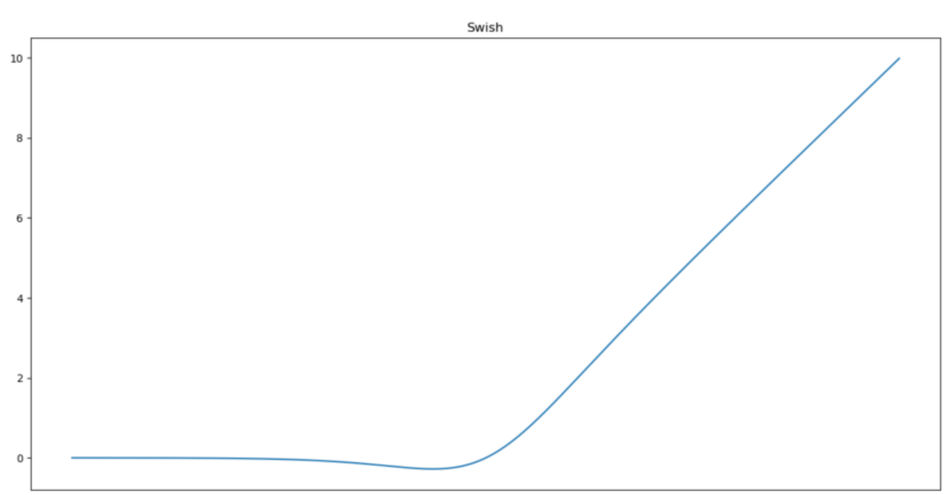
Using a filter smaller than the input is intentional as it allows the same filter (set of weights) to be multiplied by the input array multiple times at different points on the input. Specifically, the filter is applied systematically to each overlapping part or filter-sized patch of the input data, left to right, top to bottom.

4.4 Softmax Layer :

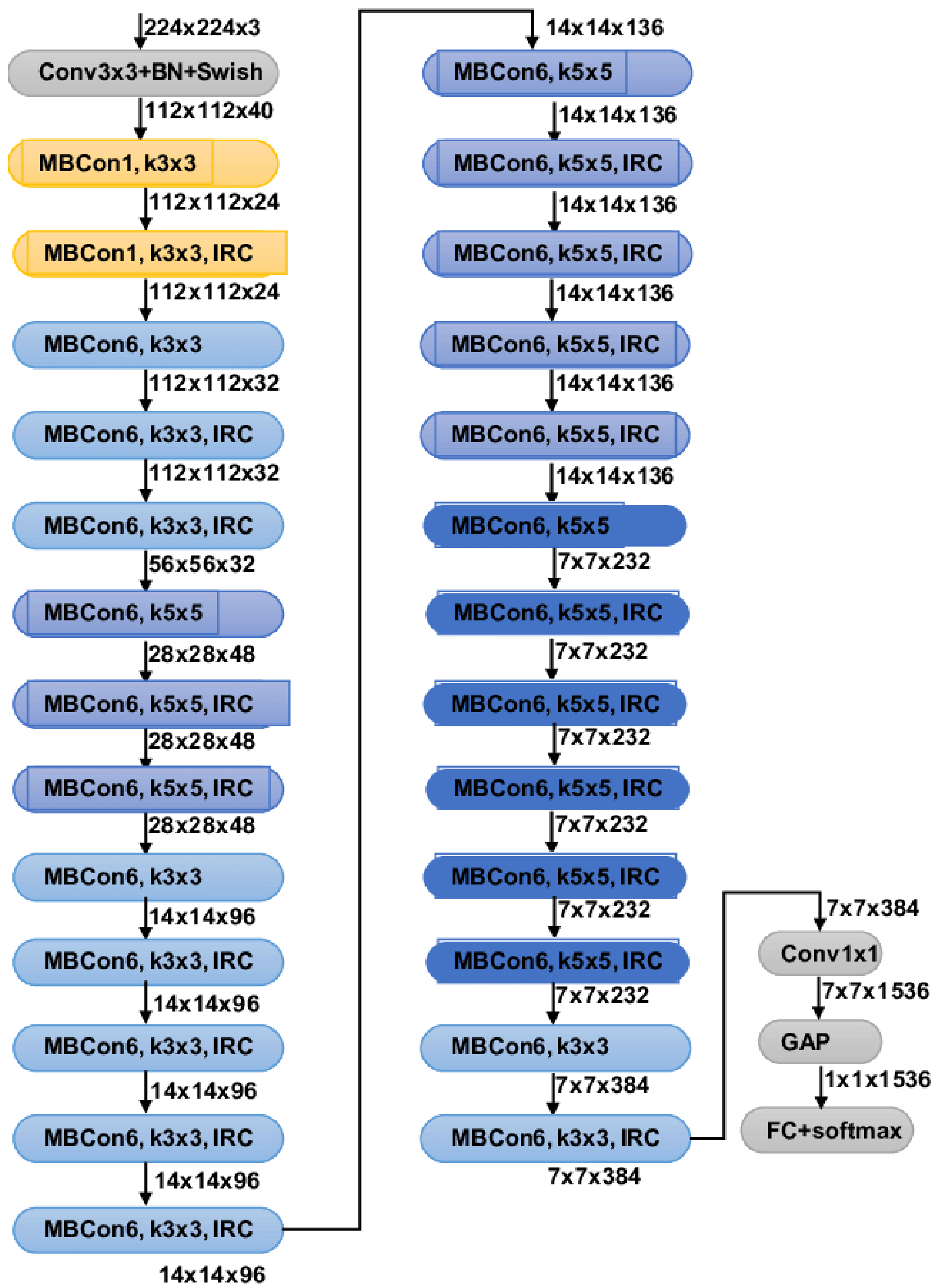
Softmax is a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in the vector.
The most common use of the softmax function in applied machine learning is in its use as an activation function in a neural network model. Specifically, the network is configured to output N values, one for each class in the classification task, and the softmax function is used to normalize the outputs, converting them from weighted sum values into probabilities that sum to one. Each value in the output of the softmax function is interpreted as the probability of membership for each class.

4.5 Swish Activation Function :

Swish is a smooth, non-monotonic function that consistently matches or outperforms ReLU on deep networks applied to a variety of challenging domains such as Image classification and Machine translation. It is unbounded above and bounded below & it is the non-monotonic attribute that actually creates the difference. With self-gating, it requires just a scalar input whereas in multi-gating scenario, it would require multiple two-scalar input. It has been inspired by the use of Sigmoid function in LSTM and Highway networks where 'self-gated' means that the gate is actually the 'sigmoid' of activation itself.Swish is as computationally efficient as ReLU and shows better performance than ReLU on deeper models. The values for swish ranges from 0 to infinity. The function is defined as – f(x) = x*sigmoid(x) (or) f(x) = x/(1-e^-x).



5. ARCHITECTURE :

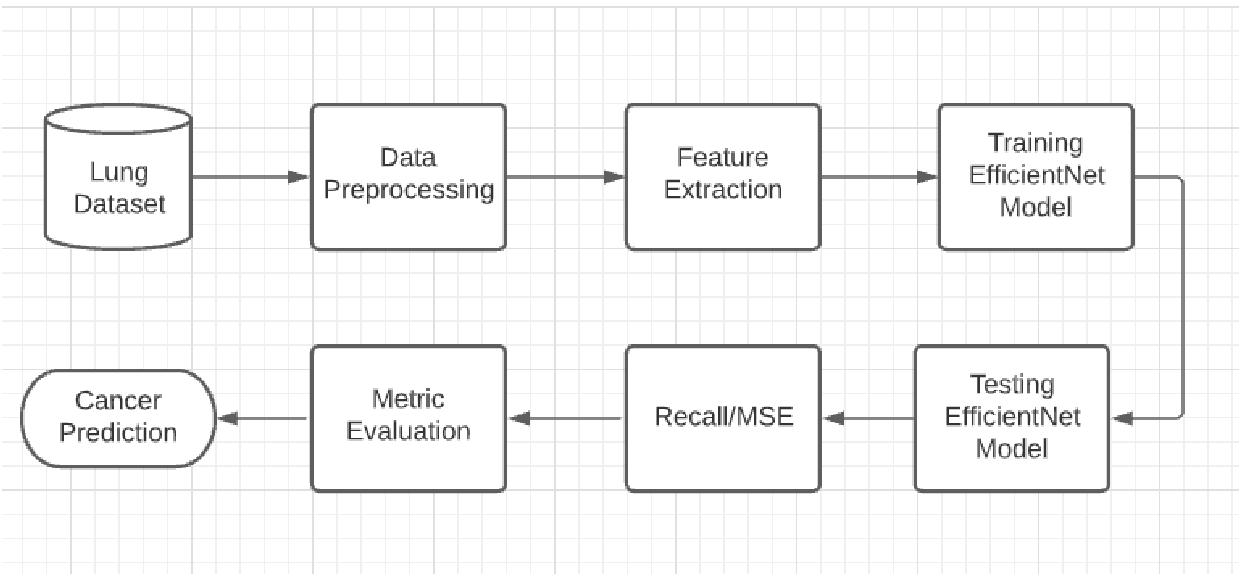


6. LITERATURE SURVEY:

Sl. no	Title	Authors	Description	Year
1.	Deep learning for lung Cancer detection and classification	A. Asuntha, Andy Srinivasan	This study employs unique Deep learning approaches to detect the location of malignant lung nodules. The feature extraction techniques used in this study, includes the Histogram of Oriented Gradients (HoG), wavelet transform-based features, Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT), and Zernike Moment.The Fuzzy Particle Swarm Optimization (FPSO) technique is used to select the optimal feature after extracting texture, geometric, volumetric, and intensity information. Finally, Deep learning is used to classify these features.	2019
2.	Lung cancer prediction using machine learning and advanced imaging techniques	Timor Kadir, Fergus Gleeson	Machine learning-based lung cancer prediction models have been developed to aid doctors in the management of ambiguous pulmonary nodules discovered by accident or on a screen.This paper presents an overview of the primary lung cancer prediction algorithms proposed to date, as well as some of their relative strengths and drawbacks. This article, goes over some of the difficulties in developing and validating such procedures, as well as the steps to clinical acceptance.	2018
3.	Lung Cancer Detection Using Artificial Neural Network	Ibrahim M. Nasser, Samy S. Abu-Naser	An Artificial Neural Network (ANN) in this research is used to detect the presence or absence of lung cancer in the human body. Symptoms such as Yellow fingers, Anxiety, Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing Difficulty, and Chest pain	2019

			were used to diagnose lung cancer. They were employed as input variables for the ANN, along with additional information about the person. The ANN was created, trained, and verified on a data set called "survey lung cancer." The ANN model can detect the absence or presence of lung cancer with 96.67 percent accuracy, according to model evaluation.	
4.	Lung Cancer Detection Using Image Processing Techniques	Mokhled Altarawneh	The fundamental aspects of this study are image quality and accuracy; image quality assessment and improvement are dependent on the enhancement stage, where low-pre-processing approaches based on Gabor filters inside Gaussian rules are applied. Following the segmentation principles, an improved region of the object of interest is created, which is then used as the basis for feature extraction. A normalcy comparison is done based on general characteristics. Pixel percentage and mask-labelling were discovered to be the most important features for reliable picture comparison in this study.	2012
5.	Lung Cancer Detection using CT Scan Images	Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A.Elchouemi	The major goal of this study is to assess several computer-aided approaches, analyse the existing best technique and determine its limitations and shortcomings, and then propose a new model that improves on the present best model. The method employed was to sort and list lung cancer detection techniques based on their detection accuracy. The procedures were examined step by step, and general limitations and flaws were identified. Some have low accuracy, while others have higher accuracy but are not close to 100 percent. As a result, the goal of our research is to improve accuracy to 100 percent.	2018

7. DATA FLOW DIAGRAM:



8. RESULT:

1	ACA	N	SCC
2	9.93E-01	1.85E-03	5.37E-03
3	9.91E-01	8.75E-03	6.24E-04
4	9.97E-01	1.04E-03	2.43E-03
5	9.13E-01	4.39E-03	8.31E-02
6	9.92E-01	1.43E-03	6.63E-03
7	9.54E-01	1.21E-02	3.36E-02
8	7.82E-01	3.87E-03	2.14E-01
9	9.88E-01	4.63E-03	7.56E-03
10	8.60E-01	2.16E-03	1.38E-01
11	9.90E-01	3.77E-03	5.97E-03
12	9.92E-01	1.48E-03	6.93E-03
13	9.89E-01	1.77E-03	8.89E-03
14	9.99E-01	1.10E-03	2.53E-04
15	9.98E-01	1.63E-03	4.84E-04
16	9.99E-01	9.25E-04	4.07E-04
17	9.87E-01	9.12E-03	3.48E-03
18	9.75E-01	1.10E-02	1.36E-02
19	9.93E-01	5.79E-03	1.44E-03
20	9.88E-01	6.17E-03	5.85E-03
21	9.86E-01	3.34E-03	1.10E-02
22	2.18E-01	4.64E-04	7.82E-01
23	9.91E-01	4.04E-03	4.99E-03
24	9.98E-01	9.30E-04	8.74E-04
25	9.90E-01	1.36E-03	8.17E-03

The output of the test set is in the form of a csv file with 3 columns dedicated to ACA (adenocarcinomas), N(Benign Lung Tissue) and SCC(lung squamous cell carcinomas). The values associated with these columns are probability values of cancer. The csv file is then augmented to round the values as shown below.

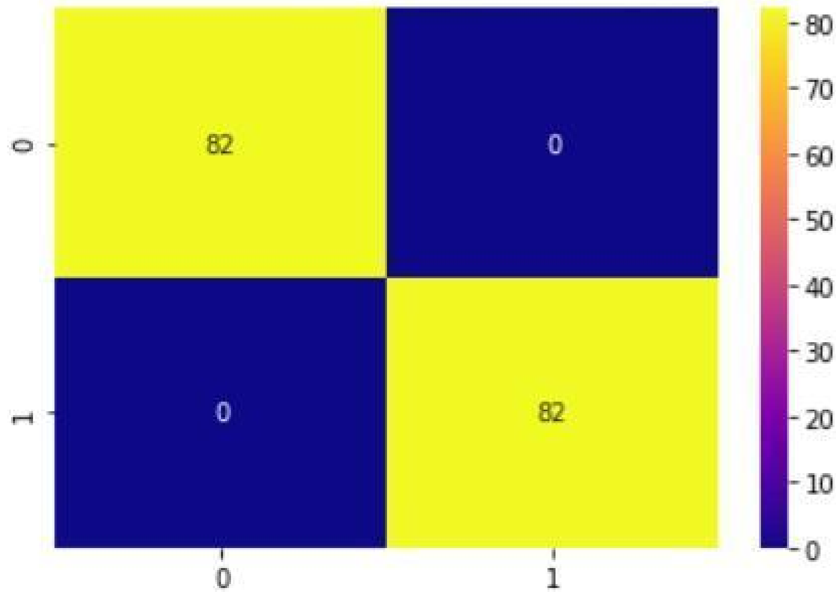
	1	0	0.1
0	1.0	0.0	0.0
1	1.0	0.0	0.0
2	1.0	0.0	0.0
3	1.0	0.0	0.0
4	1.0	0.0	0.0
...	...	...	...
1494	0.0	0.0	1.0
1495	0.0	0.0	1.0
1496	0.0	0.0	1.0
1497	0.0	0.0	1.0
1498	0.0	0.0	1.0

9. ANALYSIS:

	precision	recall	f1-score	support
0	0.95	0.98	0.96	499
1	0.99	1.00	1.00	500
2	0.99	0.94	0.97	500
micro avg	0.98	0.98	0.98	1499
macro avg	0.98	0.98	0.98	1499
weighted avg	0.98	0.98	0.98	1499
samples avg	0.98	0.98	0.98	1499

```
[ ] sns.heatmap(cm, cmap='plasma',annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd278c7cd68>
```




```
[ ] #0->normal, 1->pf
    # TP, FN
    # FP,TN
```

10. CONCLUSION:

One of the most important ways to avoid lung cancer mortality is to discover it at an early, treatable stage. We propose a novel method for automating lung cancer detection with a model employing Deep Learning techniques.This project detects cancerous lung nodules in a given lung image and to classify lung cancer and its severity. This project employs unique Deep learning approaches to detect the location of malignant lung nodules and is also used to classify features in this case. Our main goal is to create a model that can not only diagnose lung cancer with high accuracy but also be used commercially. Our model not only has near perfect precision and recall but also has a model size of only 46.9 megabytes.