INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 03 | MARCH - 2023 IMPACT FACTOR: 7.185 ISSN: 2582-3930

Detection of Lung Cancer Cells Using Inception V3 Model

Swetha Panthagani
Dept of Information Technology
Andhra Loyola Institute of Engineering and Technology
Vijayawada, India

Arthi Surakshitha B
Dept of Information Technology
Andhra Loyola Institute of Engineering and Technology
Vijayawada, India

Rizwana SK
Dept of Information Technology
Andhra Loyola Institute of Engineering and Technology
Vijayawada, India

Dr. D Ratna Kishore
Professor, Department of Information Technology
Andhra Loyola Institute of Engineering and Technology
Vijayawada, India

Abstract— **Lung Cancer is a dangerous disease which is caused by consumption of tobacco and other pollutants it is very hard to predict. But in this era of technology, prediction of lung cancer is possible with the help of machine learning and deep learning algorithms. According to the surveys the survival rate of patients that are affected by cancer is very low, it is because of lack of prediction of cancerous cells in earlier phase. In this Project we use CT(Computed Tomography) scan images of malignant, benign and normal cases are taken in the dataset to classify them as cancerous and non-cancerous images with the help of CNN family Algorithm, i.e, Inception V3 Algorithm. The CT scan images are preprocessed with the help of image preprocessing techniques such as image warping and image cropping. After the preprocessing of the CT images the images are processed with Inception V3 Algorithm, the image dataset is trained under Inception V3 Algorithm of deep learning technique and under the range of 10 epochs the best accuracy the algorithm has provided is 97%. Many Machine Learning Algorithms are used to classify the images such as SVM, CNN, etc. But under deep learning technique the inception V3 algorithm has provided the best accuracy so far.**

Keywords— CT Scan Images, CNN Algorithm, Cancer Nodules, Malignant cases, Benign Cases, Normal Cases, Inception V3 Algorithm, SVM Algorithm, Image Warping and Image Cropping.

I INTRODUCTION

Lung cancer is one of the most ubiquitous and hazardous human diseases. it include in which the cancerous cell grows and spreads abnormal cells with in the human body it can be easily detected and can be well treated if it is diagnosed in initial faces [7] lung cancer is a different type of cancer that begins in the lung tissues according to the statics, nearly 238, 340 familiar cases of this cancer recorded the survival rate for lung cancer is not more than 56 percent only, 14 percent of cancer cases are identified and analyzed in the earlier phases[8]. CT (computed tomography) scan images played a vital role in the COVID-19 pandemic, during this epidemic the laboratorians had very trouble to identify and determine whether the change in the lung is due to the virus or due to the cancer[12]. Machine Learning based models of lung cancer prediction have been proposed in [3] different methods have been implemented to determine and diagnose the cancerous cells in the CT Scan images. In [12] they identified some other risks and side effects of the cancer which includes cancer therapy in the lateral phases. Many researchers have proposed different CAD systems to identify and classify lung cancer within the CT scan images. Lung Cancer is the most staid and spread-out disease that is responsible for the most deaths every year [1]. The cancer cells consist of two types of cells one is Benign and the other is Malignant. Malignant cells are the cells which go to the other cells and spread the cancer throughout the tissues of the lung. Whereas the other Benign tumor typically is not so threatening to life and is regular in shape which does not spread through other cells. Malignant tumor cells spread through the other body cell and have an uneven shape [10]. Correspondingly the treatment of lung cancer has increased with the increase in the technology, the cancer can be detected in the earlier stages and can be diagnosed easily with the help of the technology. Medical science has improved a lot with the help of Machine learning and deep learning algorithms. All the diagnoses of the medical images depend upon the preprocessing of the images and the algorithms that have been used in the processing stage. The therapies that have been taken in the later stages such as chemotherapy and the radiography has also become painless and easy with the help of the technology.[6]

II. LITERATURE SURVEY

It is studied that many image processing techniques and machine learning techniques have been implemented to detect cancer and classify the images[3]. In this they have proposed different segmentations and different filters in the pre-processing phase. They have used the Gabor Filter, Median Filter, and added the features such as area, perimeter, also the features like diameter, centroid, also used the pixel mean intensity parameters as a feature, they have used watershed segmentation method to segment the image after the preprocessing phase. In[12] they have used Artificial Intelligence(AI) to determine the degree and the alternations of the CT scan images and identify that the alterations in the lung is caused by the virus or by the cancer. The model has identified and diagnosed the images as the professional doctor has defined. The limits and merits of the virus have been discussed in the paper. Therefore, they have derived that deep learning methodologies and the machine learning techniques showed excellent precision for identifying the images that are cancerous and non-cancerous images. They have used SVM classifier[3] for classifying the cancerous and the normal images which are not cancerous. The images that they have taken is in the form of DICOM format. In many times the DICOM format is very difficult to process, therefore they have converted that DICOM format into the JPEG images for the processing and the pre-processing of the images. In[1] they have proposed a methodology where the image is passed through the CNN algorithm. Here they have used AlexNet architecture which has 25 different layers. The training process is done with the help of the architecture. They have used the CT scan images from the LIDC-IDRI. This datasheet contains three types of images which are Benign cases, Malignant Cases, and the other images which do not have the cancer cells which are normal.

III EXISTING METHODOLOGY

In existing methodology, an image handling procedure has been used to identify the beginning time lung malignant growth in CT scan examine images. The CT filter image is processed early to be pursued by division of the ROI of the lung. Discrete waveform Transform is connected for examined image pressure and highlights are extricated usage of a GLCM. The outcomes are encouraged into a SVM classification algorithm to verify whether the lung CT image is cancerous or not. The SVM classification algorithm is assessed dependent on a LIDC dataset. Also the CNN Algorithm has used to classify the cancerous cells but the accuracy of the dataset is only 96%. The CNN Algorithm architecture has used in the existing methodology to find out the cancerous cells in the CT scan images, train and classify them with the Model that has built based on the CNN Algorithm.

Limitations:

1. CNN Algorithm does not encipher the place and orientation of the image or any object
2. CNN Algorithm does not have to ability to shifting the input signal result in an equally shifted output signal. That means it lacks the ability to be spatially invariant to the input data.
3. CNN Algorithm requires lots of training data in which small dataset may not be sufficient to execute this algorithm.

IV DATASET

We got this data set from online free sources such as Kaggle. The dataset consists of folders such as Malignant cases, Benign cases and Normal cases. The Dataset consists of 1097 images in which Malignant cases are 561, Benign cases are 120 and normal cases are 416 images.

Sample dataset

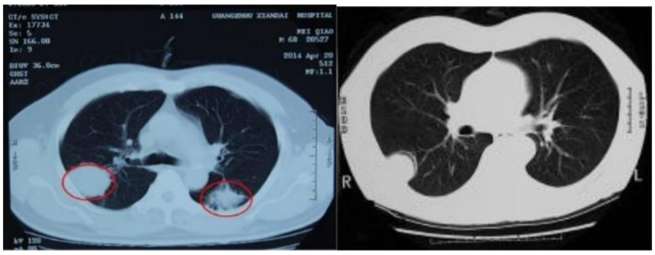

Fig 4.1 Malignant Case CT Image

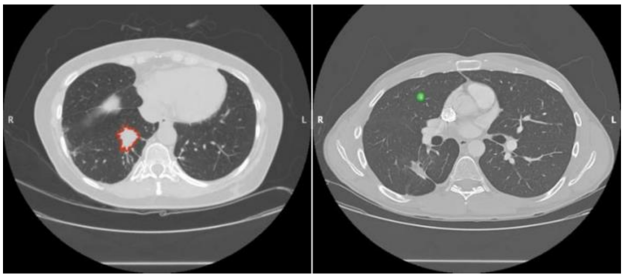

Fig 4.2 Benign Case CT Image

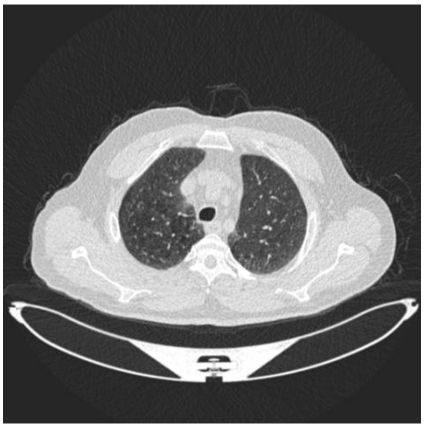

Fig 4.3 Normal Case CT Image

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 07 ISSUE: 03 | MARCH - 2023 IMPACT FACTOR: 7.185 ISSN: 2582-3930

V Preprocessing Techniques

Image Preprocessing:
Image has been preprocessed with the help of different techniques which includes the CT scan images to fit perfectly without overfitting and to use the images without any disturbance. Here we used two techniques:

-Image Warping and
-Image Cropping

Image Warping:
Image warping is the process of technologically manipulating an image such that any shapes describe in the image have been significantly distorted. Warping may be used for correcting image distortion as well as for creative purposes.

While an image can be transformed in various ways, pure warping means that points are mapped to points without changing colors. This can be based mathematically on any part of the plane to the plane. If the function is injective the original can be reconstructed. If the function is bijection any image can be inversely transformed.
Images may be distorted through simulation of optical weirdness. Images may be viewed as if they had been projected onto a mirrored surface. Images can be partitioned into polygons and each polygon distorted. Images can be distorted using morphing.

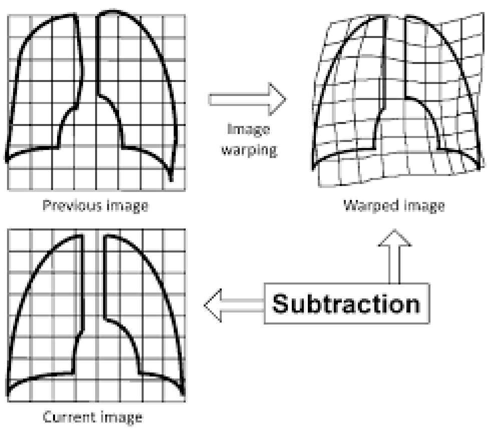

Fig 5.1: Image Warping

Image Cropping:
Cropping is the removal of unwanted outer areas from an illustrated image. The process usually consists of the removal of some of the unimportant areas of a lung image to remove extraneous trash from the lung image, to improve its framing, to change the aspect ratio. A cropping made by trimming off the top and bottom margins of an image.
The information area which contained in the original image should be removed in order to obtain the lung area. The cropping process is done in order to separate the lung areas, which is the region of interest(ROI),from another area of original CT scan image.
Its separate areas of the lungs, to the area outside the lungs.

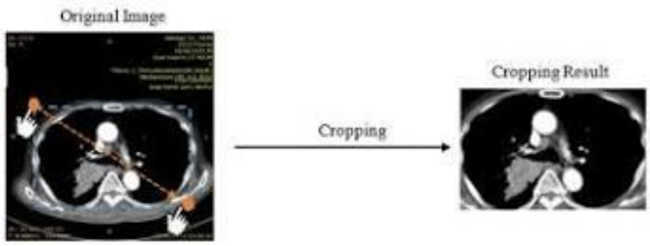

Fig 5.2 Image Cropping

VI. PROPOSED METHODOLOGY

The research area of the deep learning has improved in the medical field which includes detection of the cancers and tumors with the help of CT scan images or MRI Images, many samples of images has been collected from the different CT scanners or machines, improvising the images and making them into fit a similar format is more likely to build and train a model. In this paper we have proposed the CNN Algorithm, in which we have used different dense and pooling layers to build a model. The model has been set up with different dimensions and the attributes that are related to the Model and the output of the images.

In the proposed system our work is to classify the problem to identify and detect the presence of lung cancer in patient CT scan image of lungs with and without early-phase lung cancer. Our objective is to use methods from computer vision and deep learning, particularly 2D and 3D convolutional neural networks, to build an accurate classifier model.
An accurate lung cancer classifier could speed up and reduce costs of lung cancer screening, allowing for more widespread patient classification. In this project we are using CT Scan Lung Cancer Nodules dataset to train CNN and SVM algorithms and then calculate survival rate of patients by using both algorithms. If algorithm predicted 17 records correctly out of 20 records, then survival rate will be (17/20 * 100) = 85%.

For CNN we use multiple filters to filter the dataset for better prediction results. In the first layer CNN uses 64 X 64 image size with 32 filters and in the second layer for 32 X 32 image size also it uses 32 filters and for each filter we will have the best image features and prediction accuracy will be better. Here we will be using 16 epoch iteration so that we will get better accuracy.

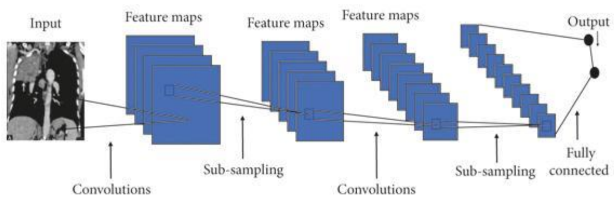

Fig 6.1 CNN Algorithm Architecture

CNN ALGORITHM

A Convolutional Neural Network(CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.
CNNs are trained using a large dataset of labeled images, where the network learns to recognize patterns and features that are associated with specific objects of classes.
Once trained, CNN can be used to classify new images, or extract features for use in other applications such as image segmentation.
Lung cancer detection based on chest CT images using CNN. In the first stage, lung regions are extracted from CT image and in that region each slice are segmented to the tumors the segmented tumor regions are used to train CNN architecture. Then CNN is used to test the patient images. The main objective of this study is to detect whether the cancer is present in a patient's lung is malignant or benign.

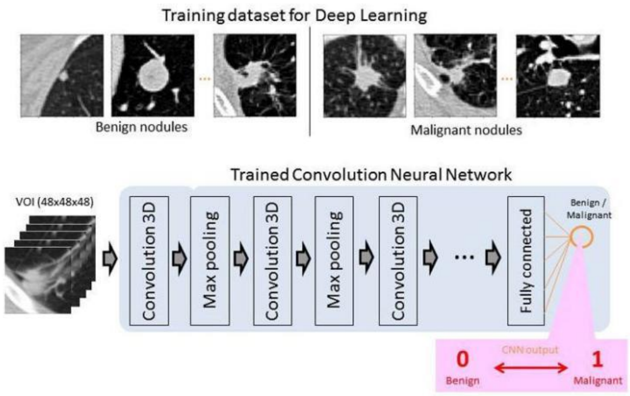

Fig 6.2 **CNN Training Dataset**

Inception V3 Model Architecture:

The inception network was an important milestone in the development of CNN classifying the different types of images . It uses a lot of tricks to push performance; both in terms of speed and accuracy.
Inception V3 Total having 48 Layer network in this v3 model is trained on ImageNet dataset for 1000 different classes. The case of Inception images needs to be 299*299*3 Pixels size.

Inception Layer is a combination of 1*1,3*3 and 585 convolutional layers with their output filter, other models generate some computational complexities and overfitting, therefore we use versions of the inception model to reduce such type of problems.
The authors noted that the auxiliary classifiers did not contribute much until near the end of the training process, when accuracies were nearing saturation. They argued that they function as regularizes, especially if they have Batch Norm or Dropout operations.
The inception v3 model itself is made up of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, concatenations, dropouts and fully connected layers. Batch normalization is used extensively throughout the model and applied to activation inputs; loss is computed using SoftMax.

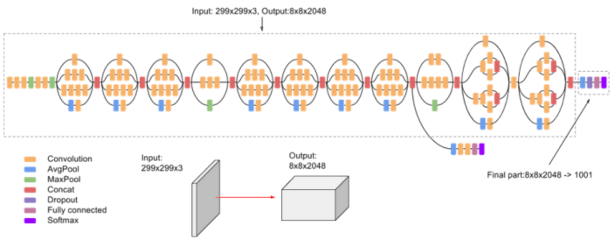

Fig 6.3 **Inception v3 Model Image Classification**

VII.EXPERIMENTAL RESULTS

In this section, the overall results(in terms of accuracy) are represented and discussed using the CNN algorithm, different features have been extracted from the images such as image region, shade region, perimeter, area of the image and the dimensions that have been included in the image.

i) CNN Learning Method

CNN Learning Algorithm has been used in this deep learning method to classify the cancerous images, for the images we have used the CNN algorithm that has specific advantages towards the medical images classification.

ii) Inception V3 Learning Method

Inception V3 Learning Algorithm has been used in this deep learning method to classify the cancerous images, for the images
Accuracy of the model can be derived using the true positive, true negative, false positive, and false negative parameters.
TP- True Positive
TN- True Negative
FP- False Positive
FN- False Negative

 ISSN: 2582-3930

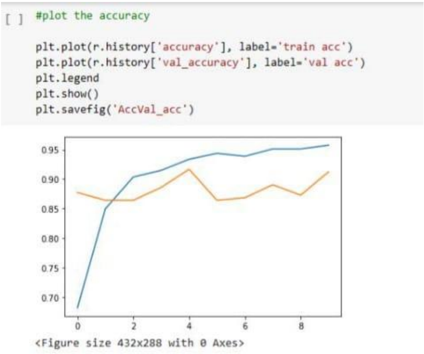

Fig 7.1 Accuracy Graph

Fig 7.1 shows the accuracy graph of the training data and the valid data in which the blue line in the graph shows the accuracy of the training data and the orange line shows the accuracy of the valid data.

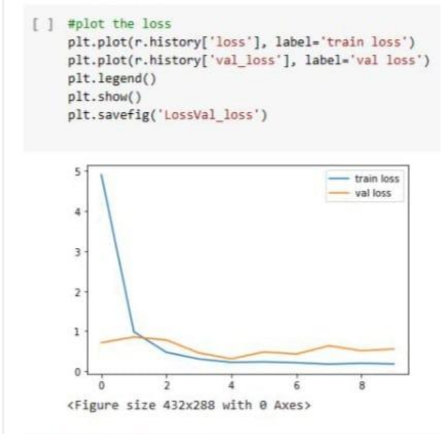

Fig 7.2 Loss Graph

Fig 7.2 shows the loss graph of the training data and the valid data in which the blue line in the graph shows the loss of the training data and the orange line shows the loss of the valid data.

The Accuracy that we got for the CNN Algorithm is 94% and the accuracy that we have got in the Inception V3 Algorithm is 97%.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision tells about how much does the model get it right when it forecasts a good outcome? When the cost of false positive is high, precision helps.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall tells about out of all the values that actually belong to class; how much is predicted as class.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 score is all about to have a combined effect of precision and recall.

$$\text{F1 Score} = 2/(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}})$$

TABLE I. PERFORMANCE MEASURES OF CNN CLASSIFIER AND INCEPTION V3 CLASSIFIER

Algorithm	Precision	Recall	F1-Score	Accuracy
CNN	0.8	0.69	0.74	94%
Inception V3	0.9	0.73	0.69	97%

Fig 7.3 shows that the processed image doesn't have any cancerous cells. The image is set across the dimensions 224* 224 for the processing and after the preprocessing of the image, the image is not having the cancer.

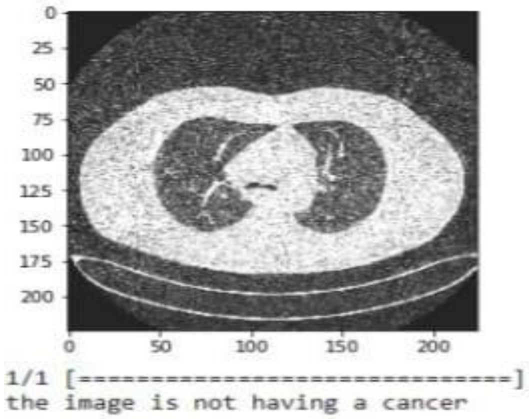

FIG 7.3 Non-Cancerous Image

Fig 7.4 shows that the processed image does have cancerous cells. The image is set across the dimensions 224* 224 for the processing and after the preprocessing of the image, the image is said to be having the cancer.

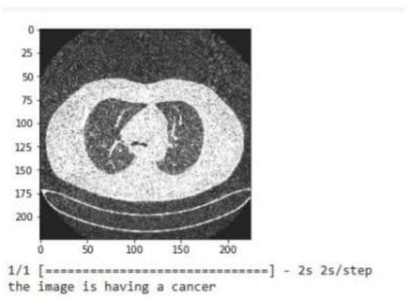

FIG 7.4 CANCER IMAGE

VIII CONCLUSION

Diagnosing the cancerous cells from the CT scan images is a hard task but diagnosing the CT scan images with the help of deep learning algorithms such as CNN Algorithm and Inception V3 Algorithm makes it easier to identify if the patient has cancer or not. The model that we have built using the CNN Algorithm has given 94% . TABLE 1 shows the accuracy of the CNN Algorithm that the model has built and in which the images are trained. The results presented showed 97% accuracy, with the model build by Inception V3 algorithm.

IX FUTURE ENHANCEMENTS

For the future enhancements, we can use some filters like Gabor Filter, HOG Filter, etc to preprocess the image correctly and the Images can use different preprocessing techniques and many more features can be extracted.

REFERENCES

[1] Sushma Garud, Dr. Sudhir Dhage "Lung Cancer Detection using CT images and CNN Algorithm." IEEE 2021 International Conference on Advances in Computing, Communication and Control (ICAC3) doi:10.1109/ICAC353642.2021.9697158 Epub: 7 February 2022, INSPEC Accession Number: 21563065.2021 International Conference on Advances in Computing, Communication and control(ICAC3).

[2]Suren Makajua, P.W.C. Prasad*a, Abeer Alsadoona, A. K. Singhb, A. Elchouemic "Lung Cancer Detection using CT Scan Images." 2018 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017.

[3]Timor Kadir, Fergus Gleeson "Lung Cancer prediction using machine learning and advanced imaging techniques"affiliated by Optellum Ltd Oxford UK, doi:10.21037/tlcr.2018.05.15 Epub: 15 May 2018, PMID:30050768.

[4]Diego Riquelme , Moulay A. Akhloufi "Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans"

[5]R Pandian, V.vedanarayanan, D.N.S Ravi Kumar, R.Rajkumar "Detection and classification of lung cancer using CNN and Google net" doi:10.1016/j.measen.2022.100588 (2022)

[6]Vikul J. Pawar1 , Kailash D. Kharat2 , Suraj R. Pardeshi3 , Prashant D. Pathak4 Lung Cancer Detection System Using Image Processing and Machine Learning Techniques.

[7]Eali Stephen Neal Joshua1* An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study Journal Homepage: http://iieta.org/journals/ria.

[8]Dakhaz Mustafa Abdullah & Nawzat Sadiq Ahmed A Review of most Recent Lung Cancer Journal homepage: ijsab.com/ijsb

[9]Smita Raut1 LUNG CANCER DETECTION USING MACHINE || Volume 6 || Issue 1 || January 2021 || ISSN (Online) 2456-0774.

[10]S.Sasikala, M.Bharathi, B.R.Sowmiya "Lung Cancer Detection and Classification using Deep CNN" International Journal of Innovative Technology and Exploring Engineering(IJITEE) ISSN:2278-3075 (2018)