Lung Cancer Detection Through Deep Neural Networks Using CT Scan Images

S. Renu Deepti, B. Srivani, Ch. Kamala, and A. Sravani

Abstract Cancer is one of the most disastrous and life-threatening diseases to human beings among which lung cancer affects more than an estimated 2.3 million people around the world each year. Lung cancer is the main cause of mortality worldwide among all types of cancers. The suffering patient's survival rate can be improved by identifying the lung nodules accurately at a fast pace, that is, at an early stage. Nowadays, the field of automated diagnostic systems is becoming popular and thus most used in the diagnosis of any disease. To understand more about this, we can take an example of Image Processing which implements automated diagnostic system especially for the medical diagnosis is one such field where an automated diagnostic system. This will help in reducing the mortality rate and detecting the disease in the initial stage which can be considered as very remarkable in the bioinformatics field. In the majority of cases documented, patients were diagnosed when their disease had progressed to the point where there was no hope of a cure. To examine the presence of any symptoms or signs, screening is used. Hence, the main objective is to design and develop an Inception V3 algorithm to detect lung cancer that improves reliability.

Keywords Convolutional neural network · CT scan · Deep Learning · Flask framework

S. Renu Deepti (✉) · B. Srivani · Ch. Kamala · A. Sravani
IT Department, VNR VJIET, Hyderabad, India
e-mail: renudeepti_s@vnrvjiet.in

B. Srivani
e-mail: srivani_b@vnrvjiet.in

Ch. Kamala
e-mail: kamala_ch@vnrvjiet.in

1 Introduction

Cancerous cells can form tumors, immune system destruction, and other breakdowns that obstruct the body from handling them correctly. Lung cancer is a malignant tumor of the lungs that leads to irrepressible tissue cell proliferation. It is the world's most common disease, with 2,093,876 new cases reported in 2018. From 2005 to 2015, the research found that men's lung cancer rates fell 2.5% per year, while women's rates fell 1.2% per year. Symptoms incorporate a chronic cough, sputum striated with blood, chest pain, alterations in voice, increased conciseness of breath, and cyclical pneumonia or bronchitis, which normally do not appear until the malignancy has progressed [1]. One of the most valuable perilous components of lung cancer disease is still cigarette smoking which is most common in the world.

In the United States, burning weed is still responsible for 80% of accidents that are caused by lung cancer. Lung cancer is thought to be caused by exposure to radon gas, which is emitted through soil and construction materials [2]. For much better picture recognition which leads to accurate predictions, we are employing the InceptionV3 convolutional neural network architecture. Our major goal is to use attributes and information to diagnose the presence of lung cancer cells. The main things that should be considered while diagnosing to check for lung cancer are the attributes. The research checks for the feasibility of using an Artificial Neural Network model to identify lung cancer in an individual's body. The purpose of this research is to identify some significant factors that can cause lung cancer, make a neural network model that can be used for detection of lung cancer and to also know the cancer stage if it is present like malign, benign, or normal.

The main aim to be achieved from cancer screening is to decrease the count of demises caused by lung cancer or to get rid of cancer deaths completely. Screening is being applied to have benefits in clinical practice instead of in a research trial setting. False positives, cost, unintentional results, radiation exposure, and overdiagnosis are all issues that need to be addressed. Despite the truth that CT scan imaging is a substantial image depicting tool in the medical domain, it is difficult for clinicians to interpret and identify cancer from CT scan images, and visual elucidation of these CT scan images may be a fallible task that causes lung cancer detection to be delayed. As a result, clinicians may find computer-assisted diagnostics useful in properly identifying malignant cells.

2 Related Works

Till now, many have tried to develop a neural network frequently aiming at increasing the accuracy of lung cancer diagnosis.

Bhatia et al. in [3] have performed preprocessing on those images to get the areas of the lungs that are sensitive to cancer. Feature extraction was done using UNet and resNet models. Then used a combination of XGBoost and random forest which was

provided with extracted features for predicting cancer using the LIDC-IRDI database, the accuracy obtained was 84%. Makaju et al. in [4] used Watershed segmentation for detection of nodules and SVM for the purpose of classification of nodules into Malignant or benign. He improved the image quality by calculating the weighted mean function. He used the Lung Image database consortium as a database which gave an accuracy of 92%. Faisal et al. in [5] worked with a number of classifiers such as MLP, Neural Network, Decision Trees, Naive Bayes, and SVM with a database obtained from UCI. From all the examinations, he found Gradient-boosted tree works well repository which gave an accuracy of 90%. Alakwaa et al. in [6] used a CT scan dataset from DSB and implemented the CNN algorithm. He predicted sensitive areas using UNet architecture. His model gave an accuracy of 86.6%.

Abdillah et al. in [7, 8] used CT scan Images which are from VIA and ELCAP databases. Region growing, Marker Controlled Watershed, and Marker Controlled Watershed with masking are applied. And among these, one with highest accuracy and robustness was Watershed with masking method. Alam et al. in [9] the preprocessing was done by image enhancement, segmentation, Image scaling, color space transformation, and contrast enhancement. He used a multi-classifier with the lung cancer dataset used for training which was taken from the UCI machine learning database. The precision was 97% for cancer identification and 87% for cancer prediction.

3 Problem Statement

The development of a Neural Network-based Detection Model, capable of identifying tumors of Lung Cancer in CT Scan Images of an individual and thereby categorizing them as benign, malignant, or none, while being completely automated in its functioning and requiring minimal human intervention from the start to finish.

4 Existing System

Currently, there exist the following types of testing methodologies, which are deployed at scale to detect LUNG CANCER:

1. **Computed tomography (CT) scan**—It is not like x-rays; it takes several pictures of the body and the computerized system combines those pictures as a part of the body that will be studied/analyzed. It is used to identify nodes/lumps present in the body
2. **Magnetic resonance imaging (MRI) scan**—MRI scans, like CT scans has radio waves in the place of x-rays.,. These scans are used to analyze lung cancer metastasis to the brain or spinal cord. These were used on a large scale at a point in time [10].

3. **Sputum cytology**—In the lab the mucus you cough up from your lungs is tested to find if it contains cancerous cells. The regular approach to do this is by collecting the samples three times each day for three consecutive days. Squamous cell lung tumors are more likely to be found by this because they start in the lung's primary airways.

4. **Needle biopsy**—A needle is regularly used to collect a sample from a mass. The disadvantage is that they only collect a little amount of tissue, which in a few situations may not be sufficient to do the diagnosis and for additional testing on cancer cells to assist doctors in choosing anticancer medications.

The challenges faced by the traditional methods are:

1. **Long turnaround times**—These results of the model can be obtained on the same day or in 1–2 days in a few situations, although the standard method gives us the result taking time such as 1–2 weeks or so.

2. **False-negative results**—Because there are no exact criteria for negative screening LDCT findings, sensitivity is usually ascertained by classifying new instances of lung cancer and giving the result as false negatives that appear under a year of a screening study. In the six trials that evaluated this characteristic, the sensitivity of LDCT for the identification of lung cancer extended from 80 to 100%.

3. **Radiation Exposures**—For the conventional/standard methods, the person must undergo many radiations. These methods have shown a huge increase in radiation exposure. The diagnosis of the chest/lungs for CT accounted for the major ratio of radiation exposure in cases of screening of lungs. Patients have to undergo/recieve on an average of eight mSv over the treatment that has been undergone for almost 2–3 years, this involves both the detection and diagnostic/treatment assessment.

5 Proposed Method

Speedy recognition of lung cancer has flatter censorious, and image processing and deep learning techniques have made it possible. A lung cancer diagnosis has become a lot easier with deep learning. Lung patient Computer Tomography (CT) scan pictures were beneficial in this investigation to locate and classify lung nodules, as well as to demonstrate their malignancy stage. In this research, we are utilizing a deep learning CNN algorithm to detect lung cancer from CT-SCAN pictures, and to train CNN, we're using the CT-SCAN image dataset. The greater motivation of this research is to see how efficient classification algorithms are at detecting lung cancer early.

Deep neural networks have the benefit of a variable weight-sharing mechanism that improves the algorithm's performance. As a result, we set out to create a reliable diagnostic LUNG CANCER detection algorithm based on CT scan pictures. Our model seeks to use CNN for feature extraction, followed by Deep Learning InceptionV3 Algorithms to accurately characterize a CT scan image as belonging
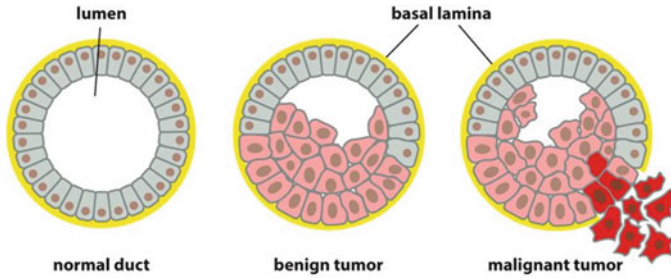
Fig. 1 Normal versus Benign versus Malignant Tumor

to an infected or healthy person. Deep learning is capable of learning a partial deep model on a partition of the overall data [11]. Data flow graphs allow us to understand more easily about a model. These are provided by a popular library called Tensorflow which is quite an intelligent toolset. It is capable of allowing programmers to construct wide-ranging neural networks with several layers.

To preprocess the photos, scale them to a given width, and do data augmentation, use Tensorflow's Keras preprocess function. The different information collectors utilize their own schemata for recording the data and the characteristics of different applications result in various data representations [12]. The Softmax function is utilized in the output layer of neural network models that foresee a multinomial probability distribution as an activation function. Softmax is acquired as the activation function for multi-class categorization that requires class membership on more than two outputs depicted in Fig. 1.

1. Ability to examine the deportment and efficiency of diverse visual models, from inception to Neural Architecture Search (NAS) networks, and then fine-tune them appropriately.
2. We will determine the model's classification ability by measuring the area lower than the graph (AUC) of a recipient operator curve. Visually analyze the behavior of these models by representing class activation maps (CAMs) or heatmaps for all the individual networks (ROC).
3. Ability to correctly diagnose lung cancer cases with an accuracy $\geq 90\%$.
4. Implementation of a fully automated lung cancer diagnosis system.

   The following are the step-by-step procedure to build the model:

1. **Data Collection**: We have downloaded the dataset from a website named Kaggle.of CT Scan images.
2. **Data Preprocessing**: Resizing, flipping, zooming, and rotating are the few preprocessing techniques used.
3. **Building the model**: InceptionV3 is the model we're developing to detect lung cancer. In this case, every layer's output is utilized as an input to the next layers. We have used three techniques of regularization, namely L1, L2, and dropouts for building the best fit model.
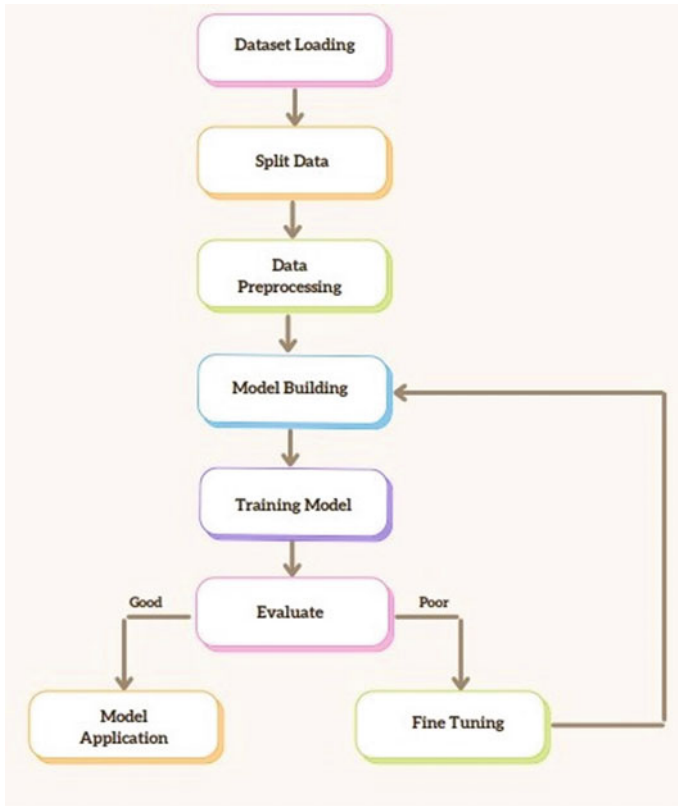
Fig. 2 Model workflow

4. **Evaluating the model**: We fine-tuned the model by changing multiple hyperparameters such as the count of neurons, activation function, optimizer, learning rate, size of the batch, and epochs to get better accuracy and loss curves. The model workflow is represented in Fig. 2.

6 Experimental Results

6.1 Dataset Description

A total of 1093 CT scans were included in the data set, of which 120 belonged to benign cases, 557 belonged to malignant cases, and 416 to normal cases. The current Model is trained in a way that uses 764 images as training data, 164 images as validation data, and 165 images as testing data. Uniformity between all cases is maintained in all three datasets. All these images were gathered from Open source
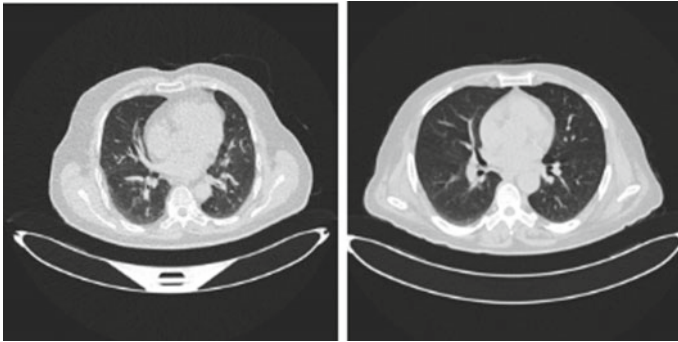
Fig. 3 CT scans of benign and malign cases

and are given due credit. The CT scans of selected benign and malignant cases and their counterparts with probable illness are shown in Fig. 3.

There is a few parameters and hyper parameters that need to be considered to build the best fit model. In the process, we have experimented with different optimizers, learning rates (0.1, 0.01, 0.2), and activation functions (relu, sigmoid, softmax, swish). The model has been trained on different optimizers to capture the suitable optimizer for our model. The table represents the accuracies obtained with different optimizers shown in Fig. 4.

So, the final model was built using an SGD optimizer with 0.001 learning rate, relu, and softmax activation functions. The model was run for 50 epochs over training and validation datasets resulting in the following accuracy and loss curves are represented in Fig. 5.

Figure 6 shows the classification result. Hence, the overall accuracy is 95.75, 97% for training, and 95.12% for the testing set. We have built a webpage using Flask which is a lightweight Python web framework. The main aim of building it is that everyone can use it without any difficulty and predict cancer easily and quickly.

This works just by uploading an individual's CT scan image of anyone to predict whether that person has a tumor or not. The three expected results are benign (non-cancerous tumor), malignant (cancerous tumor), and normal (no tumor). The output is obtained in seconds which helps the doctors or the individual who is using. This is very user-friendly, and everyone can easily understand it.

Optimizer	Training Accuracy	Validation Accuracy
Adam	86%	74.63%
RMSprop	77%	71%
SGD (0.001 LR)	97%	95.21%

Fig. 4 Optimizer function variations
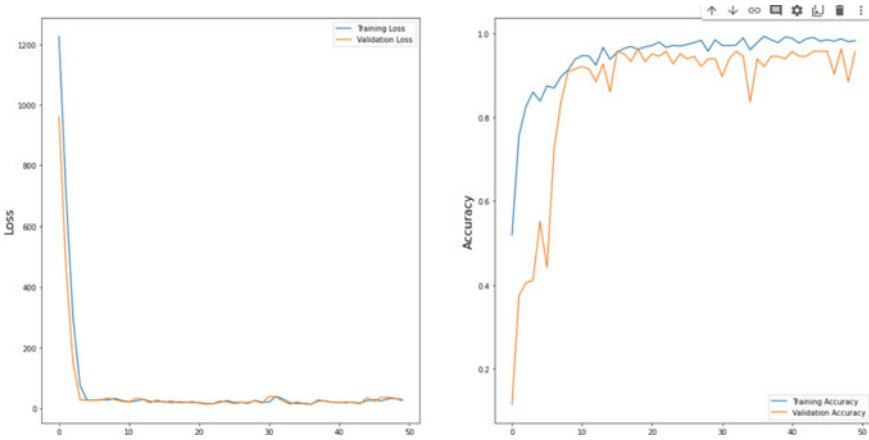
Fig. 5 Loss and accuracy curves

	Precision	**Recall**	**F1-Score**	**Support**
Bengin	0.76	0.89	0.82	18
Malignant	1.00	1.00	1.00	84
Normal	0.97	0.92	0.94	63
Accuracy			0.96	165
Macro avg	0.91	0.94	0.92	165
Weighted avg	0.96	0.96	0.96	165
Total Accuracy is 0.95757575757575				

Fig. 6 Classification report

7 Conclusion and Future Work

We used InceptionV3 architecture for accurate image recognition and obtained an accuracy of 95%. Also, a model is implemented which can be accessed by anyone to easily predict lung cancer. In conclusion, our work is at a stage of minute improvements but a stabilized and working state with maximum accuracy. There are a few improvements like increasing the accuracy of the model by using ensembling algorithms, creating an application for lung cancer detection, and showing the size of the tumor and degree (stage) of cancer.

References

1. Deepti SR, Srivani B (2016) Efficient algorithm for mining high utility itemsets from large datasets using vertical approach. IOSR J Comput Eng 18(4):68–74

2. Siegel KD, Miller JA (2017) Cancer statistics. Cancer Journal for Clinicians 67(1):7–30
3. Bhatia S, Sinha Y, Goe L (2019) Lung cancer detection: a deep learning approach. Advances in Intelligent Systems and Computing 817
4. Makajua S, Prasad PWC, Alsadoona A, Singh AK, Elchouemi A (2018) 6th international conference on smart computing and communications, ICSCC 2017, 125, 107–114
5. Faisal MI, Bashir S, Khan ZS, Khan FH (2018) An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer. In: 3rd international conference on emerging trends in engineering, sciences and technology
6. Alakwaa W, Nassef M, Badr A (2017) Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). International Journal of Advanced Computer Science and Applications 8(8)
7. Abdillah B, Bustamam A, Sarwinda D (2017) Image processing based detection of lung cancer on CT scan images. J Phys: Conf Ser 893(1):012063
8. Sharma D, Jindal G (2011) Computer aided diagnosis system for detection of lung cancer in CT scan images. International Journal of Computer and Electrical Engineering 3(5):714–718
9. Alam J, Alam S, Hossan A (2018) Multistage lung cancer detection and prediction using multi-class SVM classifier. In: International conference on computer, communication, chemical, material and electronic engineering
10. Bhatnagar D, Tiwari AK, Vijayarajan V, Krishnamoorthy A (2017) Classification of normal and abnormal images of lung cancer. IOP Conference Series: Materials Science and Engineering 263
11. Srivani B, Sandhya N, Padmaja Rani B (2020) Literature review and analysis on big DataStream classification techniques. Int J Knowledge Based Intelligent Eng Syst 24(3):205–215
12. Srivani B, Sandhya N, Padmaja Rani B (2021) An effective model for handling the big data streams based on the optimization enabled spark framework. Intelligent system design. Springer, Singapore, pp 673–696
13. Fan DP, Zhou T, Ji GP, Yi Z, Chen G et al (2020) Inf-net: automatic lung infection segmentation from CT images. IEEE Trans Med Imaging 39(8):2626–2637
14. Rahimzadeh M, Attar A (2020) A modified deep convolutional neural network for detecting pneumonia from chest X-ray images based on the concatenation of exception and resnet50v2. Informatics in Medicine Unlocked 19:100360
15. Li L, Qin L, Xu Z, Yin Y, Wang X et al (2020) Artificial intelligence distinguishes lung cancer from community-acquired pneumonia on chest CT. Radiology 296(2):E65–E71
16. Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F (2014) International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. Lung Cancer 84(1):13–22