International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Lung Cancer Detection using Deep Learning

Prof. Prerana Kulkarni[1], Mrunal Sontakke[2], Isha Jaiswal[3], Soham Bhujbal[4], Hitesh Poonia[5]

[1-5]Department of Computer Engineering & Pillai College of Engineering(Autonomous), New Panvel, India

---***---

Abstract - Higher resolution images created by modern 3D medical imaging techniques have the potential to lead to significant advancements in research and medicine. There is a need to create one of the most significant disciplines of scientific imaging because of improvements in computer-aided diagnosis and ongoing developments in the area of computerized medical image visualization. Lung cancer is one of the most prevalent malignancies among the many different types of cancer. Lung cancer detection commonly makes use of computed tomography (CT) images. Techniques for machine learning are frequently employed for lung cancer detection.CNN is a technique for describing deep learning models that have filters that may be learned through local pooling operations that are alternately implemented on input CT images to produce a variety of hierarchical or complex features. The method used in this project classifies lung cancer or normal tissue seen during computed tomography images for screening as Lung Cancer or Normal using a convolutional neural network (CNN).

Key Words: Lung Cancer Detection, Deep Learning, Image processing, Feature extraction, Convolutional Neural Network (CNN), CT Images.

1. INTRODUCTION

Lung cancer, commonly referred to as lung carcinoma, is a malignant tumour that is characterised by unchecked cell proliferation in lung tissues. To prevent this tumour from metastasizing to other body parts, it must be treated. Finding lung cancer early gives many people the best chance of recovering. The Computed Tomography (CT scan) pictures are utilized to gather the characteristics that are needed for the detection of lung cancer.

A fresh approach called deep learning enables us to improve the accuracy of the outcome. In order to guarantee that the model would be trained using the proper architecture, deep learning research has also been carried out. The pre-processing of CT scans, feature extraction, and classification are the first three steps in the diagnosis technique. The calculation that divides images into categories based on their similarity includes a crucial step called classification. The majority of the cells in the structure of cancer cells overlap one another. The goal of early cancer detection is therefore more difficult.

2. LITERATURE SURVEY

This article presents the various advantages and limitations of various techniques that have been developed and used extensively in the literature.

1. **Lung Cancer Detection Using Convolutional Neural Network.** (2019) [1]

CAD systems have shown to be quite beneficial in the medical field. A technology called computer-aided design (CAD) evaluates the components in medical images that show the unique disease process and emphasises the area of concern that needs to be addressed. If left untreated, a lung nodule is an aberrant cell-tissue formation that has the potential to spread throughout the body. These forming cells may be benign (non-cancerous) or malignant (cancerous). These cells may grow and proliferate at a dizzying rate. The only method to lessen the number of fatalities brought on by this dangerous condition is early detection.In this paper, the topic of CNN is discussed.With a 98.08% accuracy rate, 96.6% specificity rate, and only 0.985 false positives per scan, this strategy enhanced the CAD system's overall performance. The system spends 0.054 seconds processing each image.

2. **Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network.**(2020) [2]

In this study, we used a straightforward 3D CNN classifier to determine if a lung CT image was cancerous or not. Before applying the 3D CNN, we preprocessed the CT image using a thresholding method. With the LUNA 16 dataset, we carried out a thorough experiment. We have a detection accuracy of 80%, which is better than earlier methods. stage. For this investigation, we used 1500 people's CT scan images. However, we focused on the 100 patients' CT images, which each had over 120 DICOM 3D images. Among them, 20 patient images are used for testing while 80 patient images are used for training. Our approach works well for lung cancer early detection.

3. **Lung Cancer Nodules Classification and Detection Using SVM and CNN Classifiers**(2019) [3]

The diagnosis of lung cancer is the main topic of this paper. It suggested a strategy that makes use of both SVM and CNN. Support Vector Machine is a machine learning technique that is used to categorise the system. In a very high dimensional space, SVM creates a hyperplane that can be used for classification, regression, or outlier detection. For the purpose of pre-processing CT scans, a CNN employs 2D convolutional layers. Very minimal preprocessing is used in this procedure. The SVM classifier is used to obtain output. The CNN classifier is used to determine the nodule's malignant or non-cancerous status.

4. **Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images.**(2020) [4]

In order to classify benign, adenocarcinoma, and squamous cell carcinomas, this research report has examined employing convolutional neural network (CNN) architecture. An image of three distinct categories—benign, adenocarcinoma, and squamous cell carcinoma—was classified using a convolutional neural network (CNN). The model's accuracy during training and validation was 96.11% and 97.20%, respectively. To evaluate the effectiveness of the model, the precision, f1-score, recall, and a confusion matrix plot were produced.

5. **Lung Cancer Classification and Prediction Using Machine Learning and Image Processing.**(2022) [5]

According to this study, utilizing machine learning and image processing technology, lung cancer can be classified and predicted with accuracy.The dataset for the experimental study consisted of 83 CT scans from 70 different patients.After that, the photos are segmented using the K-means method.Some of the machine learning methods that were employed for the classification include ANN, KNN, and RF. It is discovered that the ANN model produces more precise outcomes for predicting lung cancer.

3. EXISTING SYSTEM ARCHITECTURE

The image of lung cancer detection obtained by a CT scan is what this technique now uses. The algorithm reads the output from all the ways mentioned above and establishes the best possible production settings, like those for a detection system. The data sets now employ the parameters we were able to collect to extract and present the results. The main calculation used in our program is SVM, a parallel order technique that creates a model document for describing another unnamed piece of information and classifying it into one of two categories using mark information from two classes as input. When preparing an SVM, known information is fed to the SVM together with dataset values, hence defining a constrained preparation set. A SVM can recognise cloud data with reasonable success when given the planning set. In SVM, the input data is translated onto a higher-dimensional space using pieces, leading to two class game plan difficulties. The concept of the choice plane, which defines the choice limit, is essential to the SVM.

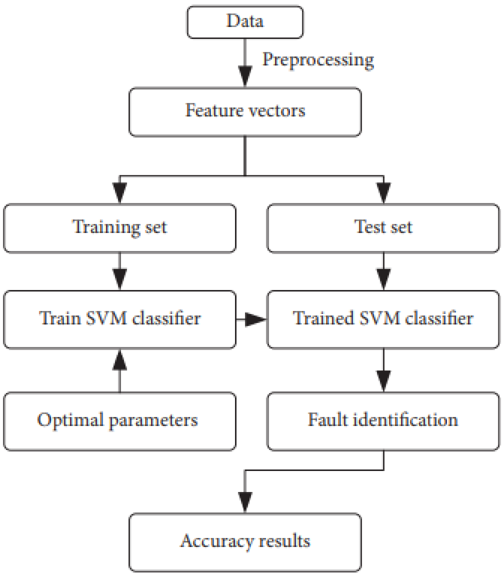

Fig -1: Existing System Architecture

4. PROPOSED SYSTEM ARCHITECTURE

For classifying an image dataset, CNN performs best. Transfer learning and employing a particular convolutional network are the two methods that CNN can be applied, respectively. There are numerous pre-trained architectures that employ CNN-based transfer learning, including VGG16, Inception V3, Resnet Inception-ResNet, Xception, and others. The ImageNet dataset was used to train each of these designs. The two methods for using pre-trained architectures are feature extraction and fine tuning.VGG16 and Inception V3 are utilised for feature extraction in this work. Later, these collected features are used as input by classifiers such as random forest and neural networks. The effectiveness of these models is contrasted with the suggested CNN design.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

The following list of procedures is employed in the proposed work:
(1) Read the digital image that was scanned
(2) use image processing techniques to get rid of noise and other restrictions
(3) From the labelled database, use a classifier to extract the various image attributes.
(4) Train the machine learning algorithm using 80% of the sample database. With the remaining 20% of samples, test the network.
(5)Apply the trained network to the scanned images and check the results.
(6)Compare the effectiveness of the proposed algorithm to other lung cancer detection techniques.
(7)Visualize the results.

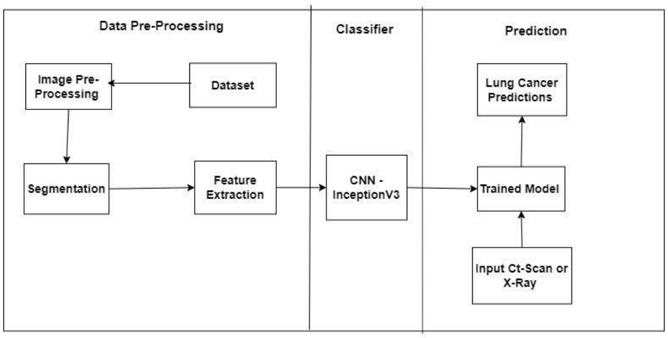

Fig -2: Proposed System Architecture

5. IMPLEMENTATION DETAILS

On our website, users can upload CT Scans using a POST request, the user uploads two files to the back end system: a raw file and a metadata file. The system uses the metadata to open the raw image files that contain the photographs. Using OpenCV and Numpy, the system then captures the picture data and saves it as image files (.png) and image data arrays (.npy). Users can then view CT Scans, the system displays the image files that it has retrieved from the back end in the front end. The system carries out this action by making GET requests for every image. Using a carousel picture, the photos on the front end can be shown. Trained models then make Predictions: when a user chooses an image from which they wish to draw conclusions. The system uses the filenames entered by the user during selection to refer to a numpy file. Before being input to the deep learning model, this numpy file is preprocessed. Following that, the model produces an image . Users can view predictions: The system uses the related mask and the original CT scan reference picture to apply an image contour to the original image.The user can view if he has lung cancer or not. A bar graph showing the chances of which type of cancer he/she might have. It classifies the lung cancer into 3 main types, mainly Adenocarcinoma","Large Carcinoma" and "Squamous Cell lung cancer

6. ALGORITHMS AND TECHNIQUES

1. Convolutional Neural Network (CNN)

The Deep Learning neural network design known as a Convolutional Neural Network (CNN) is frequently employed in computer vision. A picture or other visual data can be understood and interpreted by a computer thanks to the field of artificial intelligence known as computer vision.

There are three different kinds of layers in a typical neural network:

Input layers It is the layer into which we feed data for our model. The entire number of features in our data (or, in the case of a picture, the number of pixels) is equal to the number of neurons in this layer.

Hidden layer -The input from the input layer is subsequently fed into the hidden layer. Depending on our model and the volume of the data, there may be numerous hidden layers. The output from each layer is calculated by matrix multiplying the output of the preceding layer with the learnable weights of that layer, adding the learnable biases, and then computing the activation function, which makes the network nonlinear.

Output Layer: After being passed into a logistic function like sigmoid or softmax, the output from the hidden layer is transformed into the probability score for each class.

2. Support Vector Machine

One of the most well-liked algorithms for supervised learning is called the Support Vector Machine (SVM), and it is used to solve both classification and regression issues. It is largely utilised in Machine Learning Classification issues, though.The SVM algorithm's objective is to establish the best decision boundary or line that can divide n-dimensional space into classes so that subsequent data points can be quickly assigned to the appropriate category. The term "hyperplane" refers to this optimal decision boundary.In order to create the hyperplane, SVM selects the extreme points and vectors. Support vectors are used to describe these extreme circumstances, which is how the method got its name, Support Vector Machine.

3. Resnet50

For picture classification tasks like object recognition or scene comprehension, ResNet50, which has 50 layers, is frequently employed. Convolutional layers, batch normalization, and rectified linear activation functions (ReLU), which help to expedite training and increase accuracy, make up the network architecture. The usage of residual connections, the main innovation of ResNet50, enables information to be transmitted directly from one layer to another, skipping over multiple layers in the process. This enables the network to learn richer and more complex characteristics while reducing the issue of disappearing gradients.

4. Inception V3

For picture classification and object recognition applications, the deep convolutional neural network architecture Inception v3 was developed. The foundation of the Inception v3 network is the concept of "Inception modules," which are a collection of parallel convolutional layers with various filter sizes. For the network to reliably recognise objects in photos, it is critical that it be able to capture information at various scales and resolutions. The use of residual connections, which enable information to be transmitted directly from one layer to another without passing through multiple layers in between, is the main innovation of ResNet50. As a result, the issue of vanishing gradients is lessened and the network is able to learn richer and more intricate features.

7. USE CASE DIAGRAM

Actor graphs, use cases with system boundaries, and communication linkages between the actors and use cases make up use case diagrams. The use case diagram explains how a system interacts with external actors, and each use case represents a feature that a system provides to its users. An actor is represented as a stick figure with their name underneath, while a use case is represented by a circle with their name inside. During the analysis phase of a project, the system functionality is divided and classified using the use cases. The system is segmented using actors and use cases. Users of the system take on the roles portrayed by actors. The use case diagram for the system depicts both the direct and indirect connections between the user and the system.

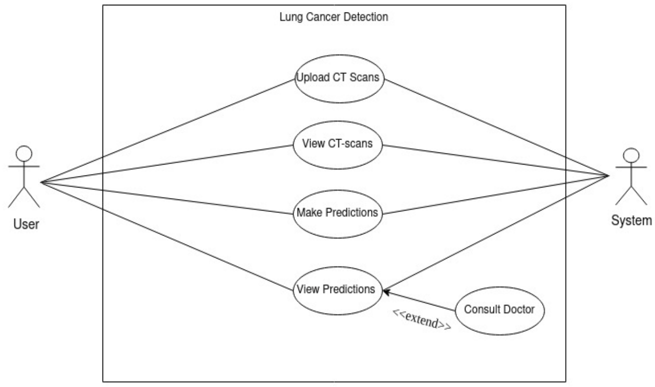

Fig -3: System Use Case Diagram

Additionally, the system displays both the explicit and implicit uses of the input parameters.

8. PERFORMANCE EVALUATION

Based on the input, or training, data, machine learning model accuracy is the measurement used to discover which model is best at recognizing correlations and patterns between variables in a dataset.

Confusion Matrix:

The prediction summary is shown as a confusion matrix. It displays the number of accurate and wrong predictions made for each class. It aids in clarifying the classes that models mistakes for other classes.

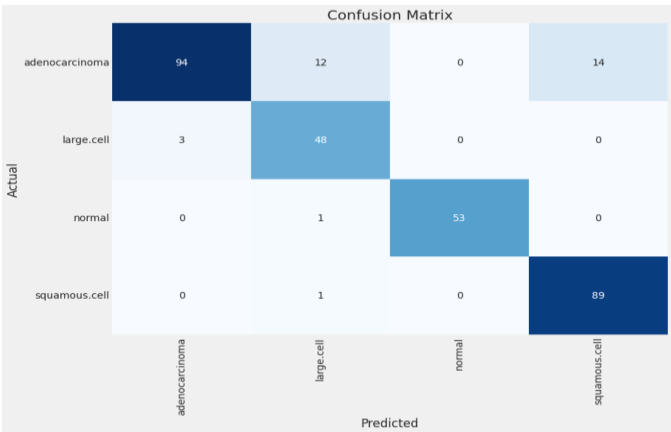

Fig -4: Confusion Matrix

Classification Report:

A classification-based machine learning model uses it as one of its performance evaluation measures. It presents the accuracy, recall, F1 score, and support of your model. It helps to clarify how well our trained model has performed overall. You must be familiar with all of the metrics shown in the machine learning model's categorization report in order to comprehend it. To help you comprehend the categorization report of your machine learning model, I have clearly explained each measure below.

```
Classification Report:
----------------------
                precision    recall  f1-score   support

adenocarcinoma     0.9691    0.7833    0.8664       120
    large.cell     0.7742    0.9412    0.8496        51
        normal     1.0000    0.9815    0.9907        54
 squamous.cell     0.8641    0.9889    0.9223        90

      accuracy                         0.9016       315
     macro avg     0.9018    0.9237    0.9072       315
  weighted avg     0.9128    0.9016    0.9009       315
```

Fig -5: Heat Map of Input Features

9. Dataset Used

The following is a "Chest CT-Scan images Dataset" that was taken from Kaggle. Three different forms of chest cancer—adenocarcinoma, large cell carcinoma, and squamous cell carcinoma—as well as one folder for normal cells are included in the data. Testing set size is 20%, training set size is 70%, and validation set size is 10%. In this dataset, there are 1000 total photos.

Fig -6: Dataset Image

10. Result And Discussion

The experimental results of the proposed system are provided in this given section. This presents the Home Page of our Web application and depicts the features provided on our website.

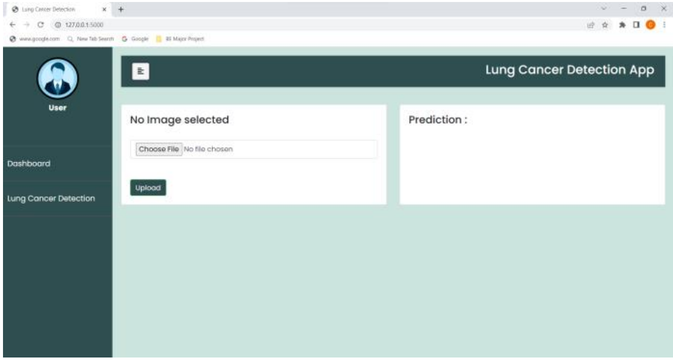

Fig -7: Web application Home Page

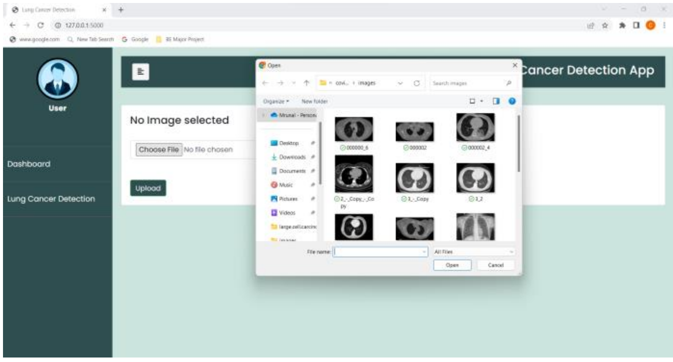

Fig -8: Input Image Selection

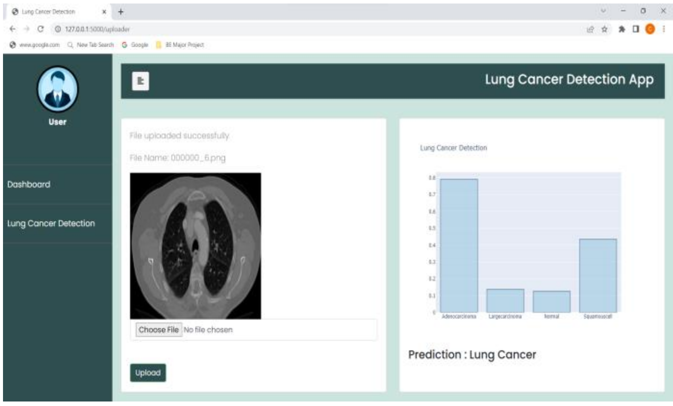

Fig -9: Diagnosed Results Displayed

11. CONCLUSIONS

Two pre-trained CNN-based architectures, Resnet50 and Inception V3, are used in this paper. Images from CT scans and X-rays with four classes can be found in the dataset. Following a comparison, it was found that the suggested CNN model outperforms pretrained architectures that employ CNN-based transfer learning. By creating confusion matrices and measuring several performance metrics, including accuracy, precision, recall, and F1-score, the performance of the suggested architecture has been examined. Specifically, all of these produce satisfactory results. Other pre-trained architectures, such as Xception, VGG19, MobileNet, etc., can be tested on the dataset as an expansion of this project in the future. Additionally, the malignant patch can be indicated on the lung tissue histopathology slide image, which will aid in pinpointing the precise locations.

REFERENCES

[1]Sayana Sharma , Mandeep Kaur , Deepak Saini,"Lung cancer detection using Convolutional Neural Network", International Journal of Advance Research, Ideas and Innovations in Technology,2019

[2]Swati Mukherjee , S.U. Bohra ,"Lung Cancer Disease Diagnosis Using Machine Learning Approach", 3rd International Conference on Intelligent Sustainable Systems (ICISS),2020

[3]Amjad Khana , Zahid Ansarib,"Identification of Lung Cancer Using Convolutional Neural Networks Based Classification", Turkish Journal of Computer and Mathematics Education ,Vol.12 No.10 (2021), 192-203, 28 April 2021

[4]Yashaswini S.L , K.V Prasad,"Lung Cancer Nodules Classification and Detection Using SVM and CNN Classifiers", International Research Journal of Engineering and Technology,Volume-06 Issue- 07, July 2019

[5]Bijaya Hatuwal , Himal Thapa,"Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images", International Journal of Computer Trends and Technology, October 2020

[6]Sharmila Nageswaran, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso,," Lung Cancer Classification and Prediction Using Machine Learning and Image Processing", BioMed Research International, August 2021

[7] Link: https://app.diagrams.net/

[8] Dataset Link: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images