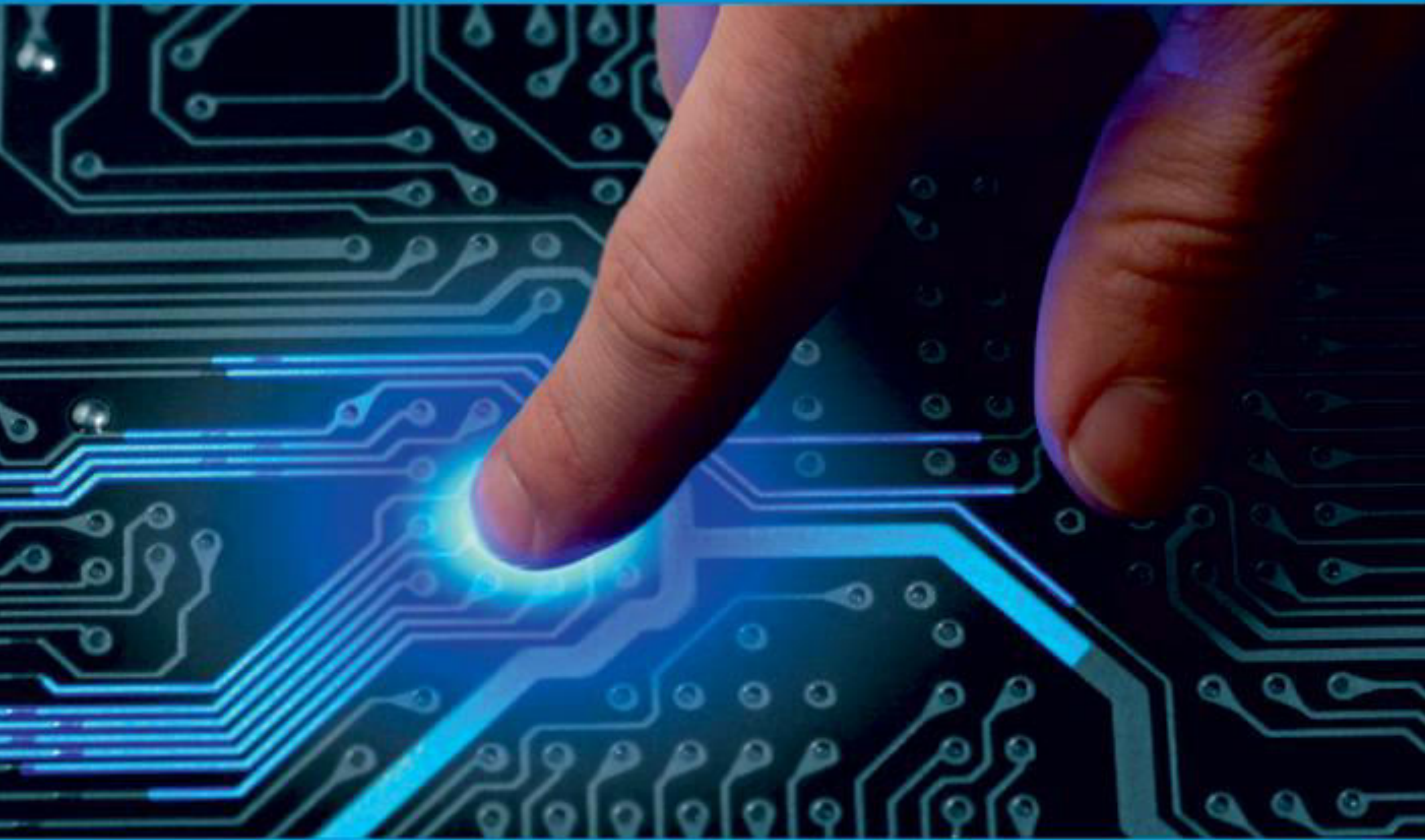
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| DOI: 10.15680/IJIRCCE.2024.1203093 |

Implementation of Lung Tumor Detection Using Machine Learning

Maddirala Amani[1], **Kilari Mounika**[2], **Mekathoti Dharani**[3], **Medikonda Grace Satyavedam**[4], **Narendra Sanam**[5]

UG Students, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Namburu, Andhra Pradesh, India[1,2,3,4]

Assistant Professor, Dept. of CSE, Vasireddy Venkatadri Institute of Technology, Namburu, Andhra Pradesh, India[5]

ABSTRACT: Lung cancer stands out as a pervasive and deadly global disease, playing a significant role in cancer-related deaths. Each year, approximately 2.20 million new cases emerge, with a concern that most patients (75% - 80%) do not survive beyond five years post-diagnosis. The complexities within lung cancer cells, marked by diverse intertumoral heterogeneity and the emergence of drug resistance, create formidable hurdles in developing effective treatment plans. Despite these challenges, advancements in cancer research technologies have fueled extensive collaborative efforts, resulting in vast collections of clinical data, medical images, and sequencing data. These archives are extremely helpful resources that enable researchers to carry out comprehensive studies of lung cancer trends, including diagnosis, treatment modalities, and clinical outcomes. Considerable progress has been achieved in combining various analyses, greatly enhancing our capacity to examine cancer. However, even with methods like matrix and tensor factorizations targeted at dimension reduction, efficiently using such complex, multi-dimensional information for therapeutic applications demands a substantial investment of time and knowledge. Researchers are further challenged by the exponential increase of cancer-related datasets. Therefore, it is imperative to use machine learning (ML) models that can recognize the complex properties included in different kinds of data on their own, therefore supporting doctors in making decisions.

KEYWORDS: Machine Learning, Convolutional Neural Networks (CNN), Tumor and Non-Tumor, Computed Tomography Scan (CT scan), ResNet50

I. INTRODUCTION

Lung cancer stands out as a pervasive and deadly global disease, playing a significant role in cancer-related deaths. Each year, approximately 2.20 million new cases emerge, with a concern that most (75%-80%) patients do not survive beyond five years post-diagnosis. The complexities within lung cancer cells, marked by diverse intertumoral heterogeneity and the emergence of drug resistance, create formidable hurdles in developing effective treatment plans. Despite these challenges, advancements in cancer research technologies have fuelled extensive collaborative efforts, resulting in vast collections of clinical data, medical images, and sequencing data. These archives are extremely helpful resources that enable researchers to carry out comprehensive studies of lung cancer trends, including diagnosis, treatment modalities, and clinical outcomes. Considerable progress has been achieved in combining various analyses, greatly enhancing our capacity to examine cancer. However, even with methods like matrix and tensor factorizations targeted at dimension reduction, efficiently using such complex, multi-dimensional information for therapeutic applications demands a substantial investment of time and knowledge. Researchers are further challenged by the exponential increase of cancer-related datasets.

The effectiveness of lung cancer therapy depends on how quickly the illness can be identified to prevent lung cancer from progressing (in stage) and spreading to other parts of the body. Powerful technology is especially needed to help doctors prevent lung cancer and analyse the disease in its early stages. In particular, image processing, machine learning, and artificial intelligence techniques can process medical field data with the help of engineering solutions to detect and diagnose lung disease. Therefore, it is imperative to use machine learning (ML) models that can recognize the complex properties included in different kinds of data on their own, therefore supporting doctors in making decisions.

CT scans are critical input data for training algorithms in machine learning, particularly Convolutional Neural Networks (CNNs), which are designed to automate the detection and analysis of lung tumours and other abnormalities.

Initially, CT scan pictures are pre-processed to standardize format, resolution, and orientation, assuring consistency throughout the collection. Annotated photos emphasizing the existence and location of abnormalities are utilized to generate training datasets, which serve as the model's ground truth. The CNN architecture, which includes convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification, is designed specifically for analysing CT images. During training, the model learns to correlate extracted features with abnormalities, fine-tuning its performance through validation and testing on previously unknown data. Following optimization, the trained CNN can be used in clinical settings to aid radiologists by automatically recognizing and flagging worrisome spots in CT scans, improving early detection and diagnostic accuracy. Overall, CNNs are an effective method for automating the interpretation of CT scan data, assisting in the identification of lung tumours and other abnormalities.

WORKING OF CT SCANS

CT (Computed Tomography) scans use X-rays to create detailed cross-sectional images of the chest, which aid in the detection of lung tumours. Patients lie on a table that moves through a CT scanner, which emits X-rays and rotates around them to capture multiple cross-sectional images. A computer then reconstructs these images, creating detailed, three-dimensional representations of the chest. In some cases, a contrast dye may be injected into the bloodstream to increase the visibility of specific structures and aid in the differentiation of normal and abnormal tissues. Radiologists examine these images for any abnormal masses or nodules that could indicate the presence of a tumour. The tumour’s size, shape, location, and characteristics are carefully evaluated to determine whether it is malignant. CT scans provide useful information about the characteristics of lung tumours, such as their size, density, and proximity to surrounding structures, which aids in cancer staging and treatment planning. Additionally, CT scans are used to monitor the progression of lung cancer and assess treatment response through repeat scans at regular intervals. Overall, CT scans are critical in detecting, diagnosing, and managing lung tumours, providing detailed images to guide treatment decisions and track therapeutic outcomes.

II. RELATED WORK

In their study titled "An Effective Method for Lung Tumor Screening Using CT Dataset," Islem Daassi, Amine Ben Slama, Sabri Barbaria, Mounir Sayadi, and Hedi Trablesi address the critical issue of lung tumors, known for their high incidence and mortality rates, often diagnosed at advanced stages. Computed tomography (CT) scans serve as crucial tools in distinguishing between various illnesses, prompting the development of computerized systems for disease analysis, particularly in its early phases. The researchers present a fully automated framework for detecting nodules in lung CT images, employing grayscale CT image histograms to automatically separate lung regions from underlying tissue, refined using morphological operators. Subsequently, they extract the internal anatomy of the parenchyma. To differentiate candidate nodules from other structures, they propose a threshold-based technique and extract various statistical and shape-based features to create a node feature vector, classified using a support vector machine. The efficacy of their method is evaluated on a large dataset from the Lung Imaging Database Consortium (LIDC), demonstrating superior results compared to existing methods, with a sensitivity rate of 84.6%.[1]

In their paper titled "Lung tumour Area Recognition and Classification using EK-Mean Clustering and SVM," K. Gopi and Dr. J Selvakumar address the critical issue of lung tumour detection, classification, and area recognition, given the significant impact of lung tumours on mortality rates. Early detection of lung tumours is pivotal in improving survival chances, with Computed Tomography (CT) screening being considered an effective method. The proposed method involves five key steps: pre-processing, feature extraction of lung tumour regions from CT scans, image segmentation of lung tumour regions, additional feature extraction, and classification of lung tumours as malignant or benign. In the proposed approach, pre-processing involves calculating the number of masks through thresholding to remove unwanted image information, followed by the calculation of Suspected Regions of Interest (ROIs). Feature extraction is conducted using the Gray-Level Co-occurrence Matrix (GLCM) on the suspected ROIs. Finally, the extracted features are classified using a Support Vector Machine (SVM) classifier. This methodology offers promise for enhancing lung tumour detection and classification, contributing to advancements in early diagnosis and treatment efficacy.[2]

In their research paper titled "Ensemble Learning Framework with GLCM Texture Extraction for Early Detection of Lung Cancer on CT Images," Sara A. Althubiti, Sanchita Paul, Rajanikanta Mohanty, Sachi Nandan Mohanty, Fayadh Alenezi, and Kemal Polat address the critical issue of lung cancer detection, emphasizing the significance of early diagnosis to mitigate its detrimental impact globally. Leveraging computerized tomography (CT) scan imaging, a widely utilized technique in the medical field, the study aims to enhance cancer detection through computer-assisted diagnosis. Through a comprehensive analysis of 20 CT scan images of lungs, the researchers embark on a preprocessing phase, evaluating various filters such as median, Gaussian, 2D convolution, and mean, determining the median filter's suitability. Subsequently, adaptive histogram equalization is employed to enhance image contrast, followed by optimization using fuzzy c-means and k-means clustering algorithms. Fuzzy c-means demonstrates superior performance with an accuracy of 98%. Feature extraction is conducted using Gray Level Co-occurrence Matrix (GLCM). The study then compares the classification efficacy of bagging, gradient boosting, and ensemble algorithms (SVM, MLPNN, DT, logistic regression, and KNN), with gradient boosting exhibiting the highest accuracy at 90.9%. These findings contribute significantly to the advancement of early lung cancer detection methods, showcasing the potential of ensemble learning frameworks in CT image analysis for improved patient outcomes.[3]

In their paper titled "Pulmonary Nodule Detection in CT Images Using Optimal Multilevel Thresholds and Rule-based Filtering," Satya Prakash Sahu, Narendra D Londhe, and Shrish Verma address the pressing issue of lung cancer, which ranks as the most common cancer with the highest mortality rate worldwide. The primary challenge lies in the late detection of lung cancer, contributing significantly to its high mortality rate. Early detection of lung nodules is particularly challenging due to various factors such as nodule types, size, unfavourable locations, and attachment with vessels. To tackle this challenge, the researchers propose a method for the early detection of lung nodules in computed tomography (CT) images, aiming to assist radiologists in locating suspected nodules efficiently. The proposed method comprises three main sections: extraction and segmentation of lung parenchyma using clustering algorithms, highlighting regions of interest through optimal multilevel thresholding employing particle swarm optimization, and selecting and applying 2D features to detect initial nodule candidates, followed by iterative false-positive reduction through the analysis of 3D features. The experiment is conducted using 84 CT scans with 301 nodules obtained from the publicly available Lung Image Database Consortium – Image Database Resource Initiative (LIDC-IDRI) dataset. Overall, the proposed method achieves an impressive overall sensitivity of 84.05% with a reduced false-positive rate of 1.93 per CT scan across all types of nodules. Comparative analysis with existing literature methods demonstrates improved false-positive rates alongside comparable sensitivity, highlighting the efficacy of the proposed approach in enhancing lung nodule detection in CT images.[4]

In their paper titled "Lung X-Ray Image Segmentation Using Heuristic Red Fox Optimization Algorithm," Antoni Jaszcz, Dawid Połap, and Robertas Damasˇevicˇius delve into the crucial realm of medical image segmentation, a fundamental step in disease recognition and classification processes. Their research focuses on the utilization of a heuristic red fox optimization algorithm (RFOA) tailored for medical image segmentation. The heuristics of RFOA are adapted to analyse two-dimensional images through equation modification and the introduction of a novel fitness function. The proposed method involves converting selected pixels in the image to either black or white variants based on a threshold value, enabling the automatic selection of segmentation threshold parameters. The study introduces a novel fitness function and adjusts RFOA for image analysis, offering a fresh approach to segmentation techniques. Evaluation is conducted using a publicly available database of lung X-ray images, followed by accuracy analysis and a comprehensive discussion of the method's benefits and limitations. This research significantly contributes to the advancement of image segmentation methods, offering potential applications in medical diagnostics and treatment planning.[5]

III. PROPOSED SYSTEM

The Proposed Machine learning algorithm used for Lung Tumour Detection is Neural Network architecture. Here, supervised learning approaches are utilized on CT scan picture datasets. A Model is built that comprises numerous steps, for the diagnosis of lung tumour techniques including Pre-processing of Images, Feature Extraction, Binarization, Thresholding, Neural Network Classifier (CNN), and Metrics. It is an application that will take input as CT scan images and will predict the possibilities of the tumour or non-tumour classes using Machine Learning. The proposed research work has defaulted to the Kaggle dataset and utilized a profound Convolutional Neural Network (CNN) system for accomplishing high precision.

A. Workflow of System

The proposed system used the following workflow:

1.Data Collection: The system collects the CT scan images of different angles. The data is collected from the Kaggle dataset consists of both tumour and non-tumour CT scan images.

2.Data Processing: The collected images are pre-processed to handle the overfitting and outliers. Normalization and some preprocessing techniques are used.

3.Test-Train Split: Split the data into training and testing sets using the train_test_split() function from sklearn. The features(X) are stored in X_train and X_test, and the target variable (Y) is stored in Y_tain and Y_test.

4.Model Training and Evaluation:

a) Initialize a ResNet50 model with pre-trained weights from the ImageNet dataset. By setting weights="imagenet, the model's weights are initialized to values learned during training on ImageNet, facilitating effective transfer learning.

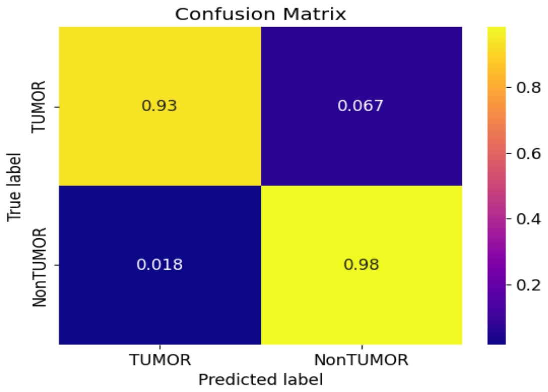
b) Fit the model to the training data using the fit() method.

c) Make predictions on the training and testing data and calculate the accuracy scores.

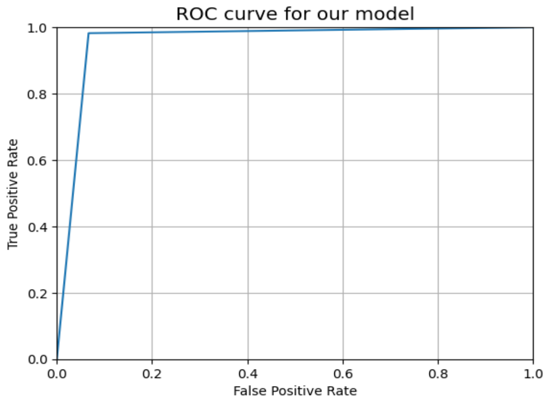
$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

5.Visualization:

Visualize the confusion matrix using Seaborn's heatmap () function with normalization and without normalization.



Receiver Operating Characteristic (ROC) curves is plotted for our model to evaluate the performance of binary classification. This graph depicts the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) at various threshold values.



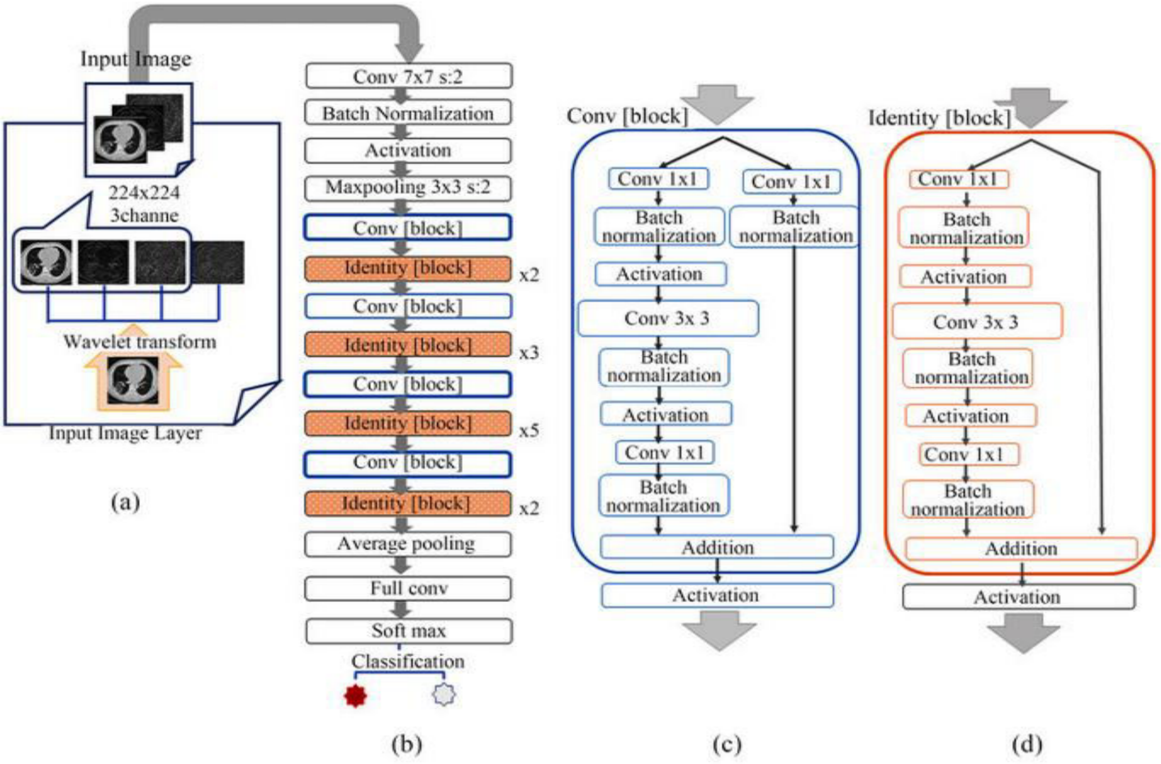
B. ResNet50

A deep convolutional neural network (CNN) architecture called ResNet50 has shown impressive performance in a number of image recognition tasks, including the identification of lung tumors. ResNet50's complex design, which consists of 50 layers, allows it to capture elements and patterns that are crucial for differentiating between tumors and healthy lung tissue in medical imaging.

ResNet50 examines computed tomography (CT) scan pictures of lung tissue in the context of lung tumor identification, identifying possible tumor locations by analyzing minute features within the images. With the help of its deep layers, the network is able to learn representations that are suggestive of the existence of tumors by extracting hierarchical characteristics from the input CT images. ResNet50 solves the problems associated with training deep networks by employing residual blocks with shortcut connections, guaranteeing efficient feature extraction and classification.

Through the use of optimization and backpropagation techniques, the network gains the ability to distinguish between regions that are tumorous and those that are not, and it adjusts its parameters to minimize classification mistakes throughout training. ResNet50 can reliably detect lung tumors in CT images after it has been taught, which can help radiologists and physicians diagnose and treat lung cancer. The accuracy with which it can analyses medical pictures highlights its potential as an effective tool for lung tumor early diagnosis and management, which might eventually lead to better patient outcomes and increase survival rates.

C. Workflow of ResNet50



IV. PREREQUISITES

The following will be necessary for the reader to follow along:

1. Verify that Python is installed on your computer. The official Python website offers downloads and installation instructions for Python.
2. Verify that the necessary libraries(keras,tensarflow,matplotlib,numpy,sklearn,seaborn) are imported and installed.
3. Verify that an image-based dataset is accessible in the local storage.

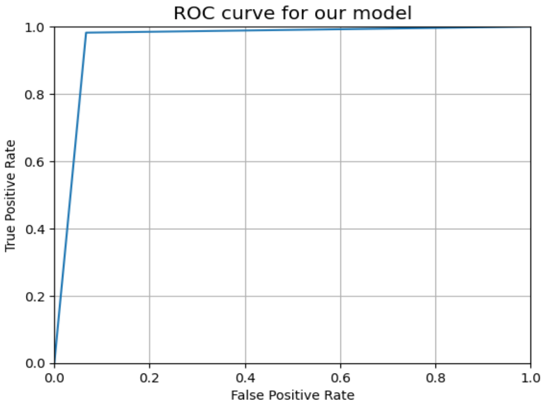
V. RESULTS

When training a Convolution Neural Network (CNN) with a ResNet architecture for the lung tumor procedure, an image processing framework like Python's Keras is used to retrieve pictures and class labels from the file path. A directory holding the picture files may be created and the path assigned to it, along with the Python "cv2" package, to store the image files. change the image's color space from BGR, which OpenCV uses, to RGB, which Keras uses.

Preprocess each training image to 224x224 pixels, as illustrated in the figure. The resized pictures (CT scan) are added to the appropriate lists. Normalization involves converting a list of labels into an array and normalizing it to the interval [0, 1] with a value of 255. To extract the class label, split the file into training and testing sets for images, then combine the sets. Make labels into categories of either 0 or 1.

The pre-trained ResNet50 model consists of a neural network with convolutional, pooling, flattening, and fully connected layers. After training the model for several epochs on CT scan pictures, save the learned model in.h5 format using the 'save' function.

The ROC curve illustrates the trade-off between a binary classification model's true positive rate (TPR) and false positive rate (FPR) at various thresholds. The area under the ROC curve (AUC) measures a binary model's performance for both CT scans.



The confusion matrix summarizes the performance of binary classification models. It displays the number of true positives, false positives, true negatives, and false negatives in each class

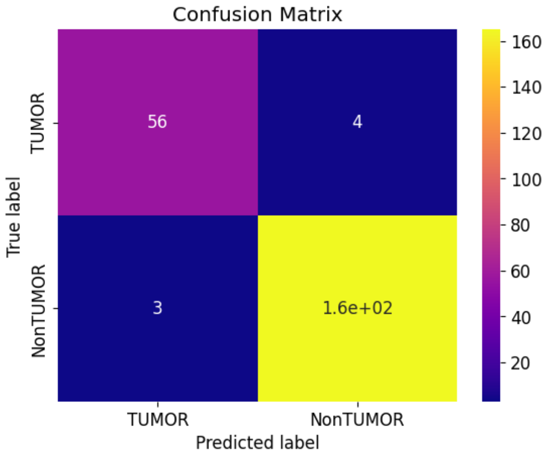


Fig-1 Confusion Matrix without Normalization

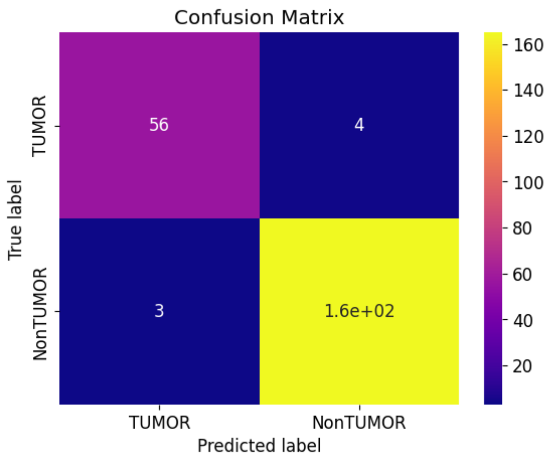
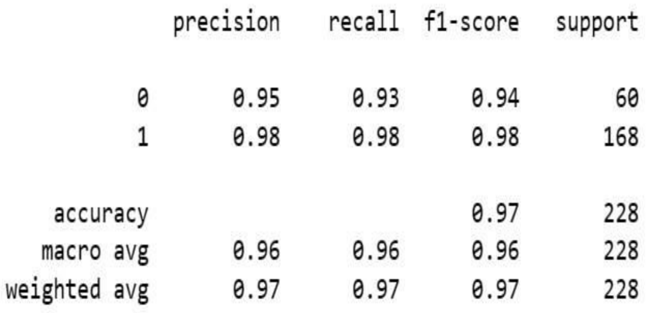


Fig-2 Confusion Matrix with Normalization

A binary classification model's performance is assessed and opportunities for enhancement are pinpointed through the use of the classification report. We compute F1-score, precision, recall, and support.

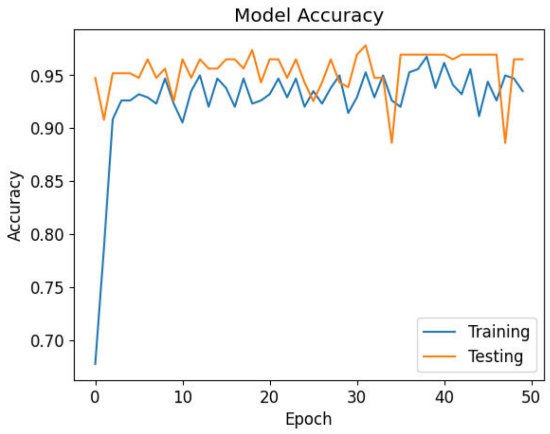


```
              precision    recall  f1-score   support

           0       0.95      0.93      0.94        60
           1       0.98      0.98      0.98       168

    accuracy                           0.97       228
   macro avg       0.96      0.96      0.96       228
weighted avg       0.97      0.97      0.97       228
```

The accuracy plot classifies training and validation data during training. To avoid overfitting, training accuracy should be increased while validation accuracy remains consistent. Out of all the cases in the dataset, the model's accuracy quantifies the percentage of properly identified occurrences.



The loss plot illustrates the model's ability to minimize its loss function during training, aiming to decrease its training loss while simultaneously keeping validation loss low to avoid overfitting. Model loss serves to measure the disparity between predicted outputs and actual labels.

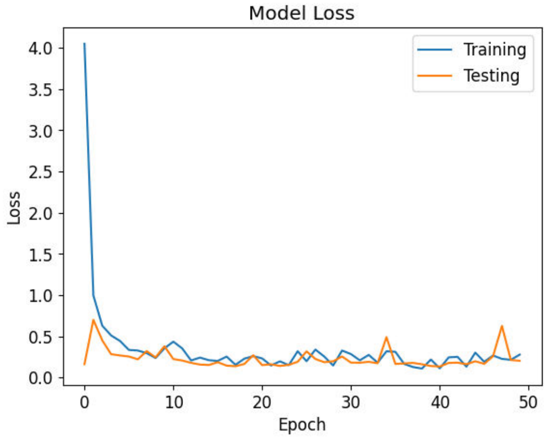


Table-1: Accuracy for various number of epochs.

No of Epochs	Accuracy
25	0.95
30	0.98
50	0.97
75	0.97
100	0.97

Table-2: Accuracy with change in test and train ratio.

No of Epochs	Accuracy
25	0.98
30	0.98
50	0.98
75	0.95
100	0.95

Table-3: Accuracy with change in optimizer.

No of Epochs	Accuracy
25	0.98
30	0.98
50	0.98
75	0.98
100	0.98

VI. CONCLUSION

In conclusion, the application of convolutional neural networks (CNNs) in lung tumor detection represents a notable advancement in medical imaging and cancer diagnosis. This study demonstrates the potential of CNNs, particularly ResNet50, to enhance the precision and effectiveness of identifying lung tumors from computed tomography (CT) scans. Through the adoption of sophisticated machine learning techniques, such as CNNs, the study significantly improves the accuracy of tumor detection, offering prospects for early diagnosis and improved patient outcomes. Moreover, by integrating cutting-edge technologies into healthcare practices, there is a recognition of the importance of addressing the complexities associated with diseases like lung cancer. Overall, this research provides valuable insights and methodologies that hold promise for revolutionizing the landscape of lung tumor detection and diagnosis, ultimately benefiting both patients and healthcare providers.

REFERENCES

[1] An Effective Method for Lung Tumor Screening Using CT Dataset - Islem Daassi, AMine Ben Slama, Sabri Barbaria, Mounir Sayadi, Hedi Trabelsi - 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)

[2] Lung tumor Area Recognition and Classification using K-Mean Clustering and SVM - K.Gopi, Dr J.Selvakumar - 2017 International Conference on Nextgen Electronic Technologies.

[3] Ensemble Learning Framework with GLCM Texture Extraction for Early Detection of Lung Cancer on CT images - Sara A. Althubiti, Sanchita Paul, Rajanikanta Mohanty ,Sachi Nandan Mohanty ,Fayadh Alenezi, and Kemal Polat - 2022 Hindawi Computational and Mathematical Methods in Medicine Volume 2022, Article ID 2733965, 14 pages https://doi.org/10.1155/2022/2733965

[4] Pulmonary Nodule Detection in CT Images Using Optimal Multilevel Thresholds and Rule-based Filtering - Satya Prakash Sahu, Narendra D Londhe & Shrish Verma - 2019 IETE Journal of Research.

[5] Lung X-Ray Image Segmentation Using Heuristic Red Fox Optimization Algorihum - Antoni Jaszcz,Dawid Połap ,and Robertas Damasˇevicˇius - 2022 Hindawi Scientific Programming ,Volume 2022, Article ID 4494139, 8 pages ,https://doi.org/10.1155/2022/4494139

[6] Computer Systems, [6] Janee Alam, S., & Hossan, A. Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier. 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)

[7] Vaishnavi. D, Arya. K. S, Devi Abirami. T, M. N. Kavitha B.E-CSE, Builders Engineering College, Kangayam, Tirupur, Tamil Nadu, “Lung Cancer Detection using Machine Learning”.

[8] S. Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, “Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks”, International Conference on Electronics, Communication and Aerospace Technology, ICECA, 100-103, 2017.

[9] Chunran, Yang, Wang Yuanvuan, and Guo Yi. "Automatic Detection and Segmentation of Lung Nodule on CT Images." 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2018.

[10] A.Amutha, Dr.R.S.D.Wahidabanu,” Lung Tumor Detection and Diagnosis in CT scan Images,”International conference on Communication and Signal Processing, April 3-5, IEEE, 2013.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING