e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

MACHINE LEARNING BASED LUNG CANCER DISEASE PREDICTION SYSTEM

Prof. Dr. Pavan Gujjar[*1], Brijesh V[*2], Eshwar Patil[*3], Harshitha Suresh[*4], Nisha CD[*5]

[*1]Professor, Department Of Information Science And Engineering, AMC Engineering College, Bengaluru, Karnataka, India.

[*2,3,4,5]Student, Department Of Information Science And Engineering, AMC Engineering College, Bengaluru, Karnataka, India.

DOI : https://www.doi.org/10.56726/IRJMETS51864

ABSTRACT

Small cell lung cancer (SCLC) is a type of malignant tumour that is characterised by rapid growth and early metastasis spread. Early and accurate SCLC diagnosis is critical for improved survival. Accurate cancer segmentation assists doctors in better understanding the location and size of cancer and making better diagnostic decisions. In this project, we are using the YOLO framework to pinpoint the exact location of a lung tumour that is attached to the border of blood veins, as well as to classify the tumour. The R-CNN techniques demonstrated in Part 1 primarily use regions to localiseobjects within an image. The network does not examine the entire image, but only the portions of the image that are more likely to contain an object. The main advantage of using YOLO is that it is extremely fast and accurate.

Keywords: YOLO, Deep Learning, Lung Cancer, CNN.

I. INTRODUCTION

The lungs are shaped like a pair of sponge cones. The right lung has three lobes, while the left lung has two. The right lung is significantly larger than the left lung. The inhaling process delivers oxygen to the lungs. The lungs' tissue transports oxygen into the bloodstream. Lung cancer is a disease that causes abnormal cells to multiply and grow into tumours. Cancer cells in the blood can be conducted away from the lungs. Because the natural flow of lymph out of the lungs is toward the centre of the chest, lung cancer frequently spreads toward the centre of the chest. Lung cancer is classified into two types: small cell lung cancer and non-small cell lung cancer, which has three subtypes: carcinoma, aden carcinoma, and squalors cell carcinomas. Lung cancer was found to be the second leading cause of death in men and the sixth leading cause of death in women. Image processing has a wide range of applications in medical image processing for diagnosing lung cancer. The second stage employs several image enhancement techniques to achieve thehighest level of quality and clarity. The third stage employs image partition algorithms, which play an important role insubsequent image processing stages, and the fourth stage obtains general features from enhanced partitioned images, which provide indicators of image normality or abnormality.

Objectives:

Deploying CNN Network for Lung cancer Infection Classification:

System: Upload CT scan ,The system should be capable of getting CT scan from Users that will be utilized by the CNN Model.

System: Detection of COVID – The system should be able to detect the Cancer within the CT scan images that users have uploaded.

System: Display Results - The system should be able to give information that our user can appropriately understandand gain insight from it.

Problem Statement:

To implement an image classification model that detects and categorizes the chest X- Ray and CT images into Normal, cancer, using CNN model thereby reducing turnaround time.

Motivation

The motivation behind this project is the rapid growth in cancer incidence and mortality cases worldwide. The reasons are complex but reflect both aging and growth of the population and changes in the prevalence and distribution of prevalence and distribution of the main risk factors for cancer.

DATAFLOW DAIGRAMS

A dataflow outline is a tool for referring to knowledge progression from one module to the next module This graphgives the data of each module's info and yield. The map has no power flow and there are no circles at the same time.

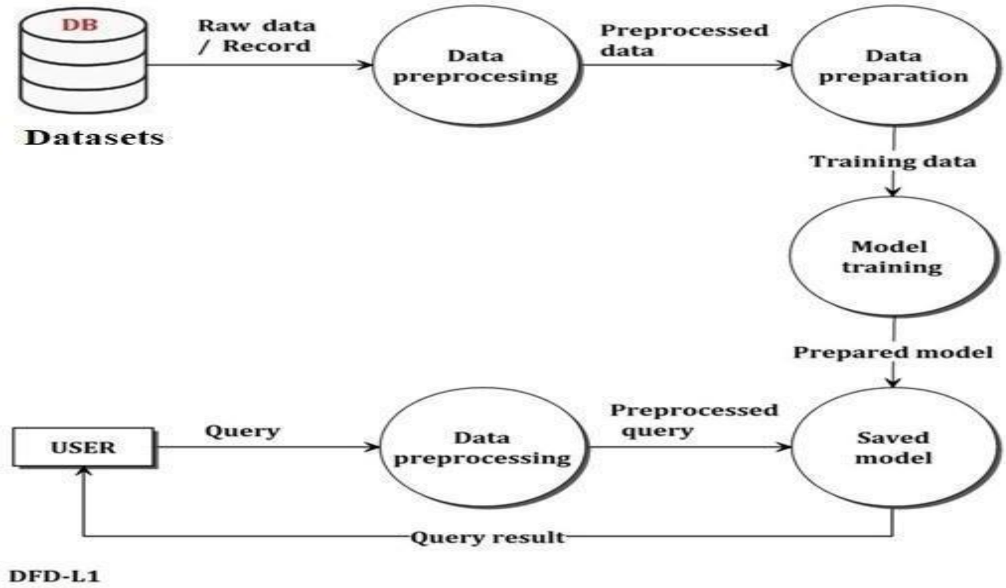
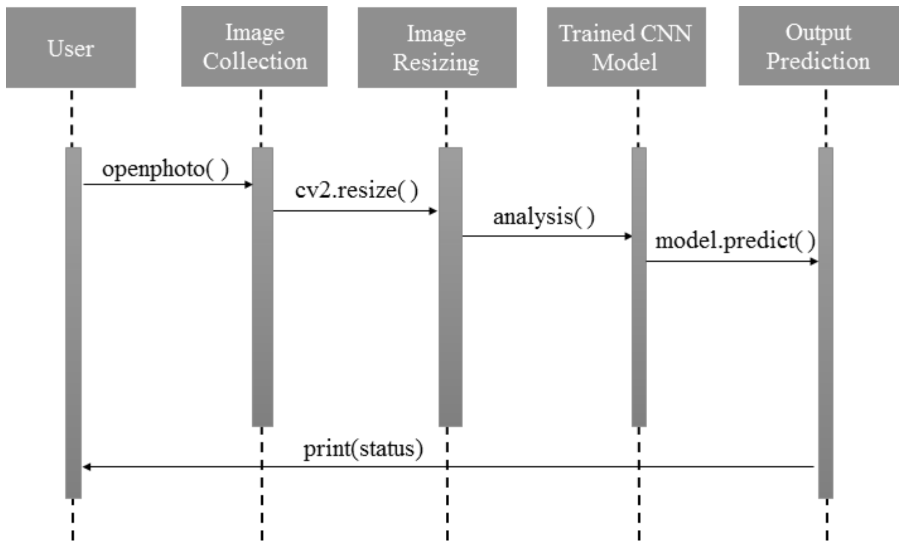


Figure 1. Data Flow Diagram

Sequence Diagrams

A sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which theseinteractions take place as shown in Fig,



II. LITERATURE SURVEY

- Efficient Identification of Lung Cancer on Computed Tomography Images by using Methodology Classification based on Deep Learning, Dr. K Gurnadha Gupta, Dr. T Kumaresan, Dr. Sasikala Dhamodaran 2021.

Deep learning is associate AI feature that mimics the human brain's operations within the process of knowledge for object detection, speech recognition, language translation, and higher cognitive process. The prediction of cancer at earlier stages in recent years is mandatory to maximize the probability of the sufferer's survival. Lung cancer, which isknown as one of the most prevalent diseases in humans worldwide, is the most dreadful type.

- Lung Cancer Detection Using Deep Learning Network: A Comparative Analysis Susmita Das, Swanirbhar Majumder2020Deep learning is an emergent and influential method which is used for feature learning and pattern recognition. We provide a comparison between Computer Aided Diagnosis scheme using Deep Learning Technique andtraditional Computer Aided Diagnosis scheme in our paper. In this paper, we have compared several deep neural networks for recognition of pulmonary cancer.

- Lung Cancer Detection Using Convolutional Neural Network Dunke Siddhi1, Tarade Swapna2, Waghule Pratiksha3JAN 2022 Lung cancer is commonly cause of cancer death in the world. detection of lung cancer will help tosave the patient. The CNN, a method that good describe a deep learning models featuring filter that can be trained withthe local pooling operations being incorporated on input CT images in an alternating manner and create an array of hierarchical of complex features. This paper presents an approach which uses a Convolutional Neural Network (CNNs)to classify cancer or normal seen in lung cancer screening computed tomography scans as malignant or benign.
- A Novel Approach For Detecting Lung Cancer Using Image Processing N.V.Chaudhari1, Dr. A.V.Malviya2 July 2021Rapid identification is a difficult issue for researchers since noise signals are mixed in with original signals duringthe picture capture process, causing the cancer picture quality to deteriorate and resulting in poor efficiency. So, as to avoid this image processing techniques become popular in a variety of medical fields for image analysis in earlier diagnosis and treatment stages, especially in cancer tumors, where time is essential for detecting abnormality issues in target images. Despite computed tomography is the most widely used imaging technique in medicine; it can be difficultfor doctors to precisely identify and diagnose cancer from CT pictures. As a result, computer-aided diagnosis can be beneficial to doctors in accurately identifying cancer cells. The methodology is examined by applying image processingtechniques and machine learning classification.
- Lung cancer Detection Techniques: A Review Namita Awasthi1 ,Vimal Kumar Awasthi2,Sudhriti Sen Gupta3 June 2021 Modern 3D medical imaging offers the potential for major advances in science and medicine as higher fidelity images are produced. Due to advances in computer-aided diagnosis and continuous advancements in the field of computerized medical image visualization, there is a need to develop one of the most important fields of scientific imaging. There are many types of cancer, of which lung cancer is one of the most common cancers. Machine learning techniques are widely used for lung cancer screening. This article compares different machine learning techniques for lung cancer detection.
- Lung Cancer Detection using Convolutional Neural Network Spoorthi Hiremath1, Dr. Dhananjaya. V2 May 2022 Small cell lung cancer (SCLC) is a type of malignant tumour that is characterised by rapid growth and early metastasisspread. Early and accurate SCLC diagnosis is critical for improved survival. Accurate cancer segmentation assists doctors in better understanding the location and size of cancer and making better diagnostic decisions. In this project, we are using the YOLO framework to pinpoint the exact location of a lung tumour that is attached to the border of bloodveins, as well as to classify the tumour. The R-CNN techniques demonstrated in Part 1 primarily use regions to localise objects within an image. The network does not examine the entire image, but only the portions of the image that are more likely to contain an object. The main advantage of using YOLO is that it is extremely fast and accurate.

III. METHODOLOGY

1. Dataset Collection:

The first step in our methodology involves the acquisition of a diverse and representative dataset for training and testing our deep learning model. This dataset should include a variety of lung images encompassing both healthy and diseased conditions. Sources may include medical imaging repositories, hospitals, or research institutions specializing in pulmonary images.

When it comes to collecting a dataset for lung disease classification using a CNN model, there arespecific considerationsand steps to ensure the dataset's relevance and effectiveness.

2. Data Preprocessing:

Prior to model training, the collected dataset undergoes preprocessing to ensure consistency and quality. Preprocessing isa crucial step in preparing data for a CNN model for lung disease classification. The goal is to enhance the quality of the input data and facilitate the learning process for the model.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

3. Segmentation:

Segmentation is a process of dividing an image into meaningful regions or segments, often to identify and isolate specific structures or abnormalities. In the context of a CNN model for lung disease classification, segmentation can be valuable for highlighting and extracting relevant regions of the lung that may aid in the classification task.

4. Feature Extraction:

In the context of a CNN model for lung disease classification, feature extraction is a crucial step where the model learns to identify important patterns and features from the input data (in this case, likely medical images of lungs). Feature extraction involves identifying and extracting meaningful features from the segmented lung images. This step aids in capturing essential patterns and characteristics that distinguish between healthy and diseased conditions. Convolutional Neural Networks (CNNs) are often employed in this stage to automatically learn discriminative features.

5. Classification:

The core of our methodology is the classification stage, where a deep learning model is trained to differentiate between healthy and diseased lungs based on the extracted features. The classification step in a CNN model for lung disease involves taking the features extracted in the previous step and using them to make predictions about the presence or absence of specific diseases. This is typically done in the fully connected layers of the network. After the feature extraction through convolutional and pooling layers, the high-level features are flattened into a vector and fedinto a series of fully connected layers. These layers are responsible for learning the relationships between the extractedfeatures and the target classes (e.g., different lung diseases).

6. Analysis and Result:

Once a CNN model for lung disease classification has been trained, the analysis and result method involves evaluatingits performance on new, unseen data.

IV. RESULTS

Various parameters achieved in our CNN Model:

1. Accuracy: 0.9901

2. Loss: 0.08496

3. val_loss: 0.00902

4. val_acc: 1.0000

Comparison based on features :

Accuracy refers to an instrument's capacity to measure an exact value. In other terms, itis the measure's resemblance to a standard or real value.

Formula: Accuracy = (TP+TN) / (TP+TN+FP+FN)

In the context of lung disease detection using CNN (Convolutional Neural Networks), recall and precisionare two commonly used evaluation metrics to measure the performance of the model.

Recall also known as sensitivity, is the percentage of actual positive cases (lung cancer patients) that themodel correctly identified as positive. Recall can be calculated using the following equation:

Recall = True Positives / (True Positives + False Negatives)

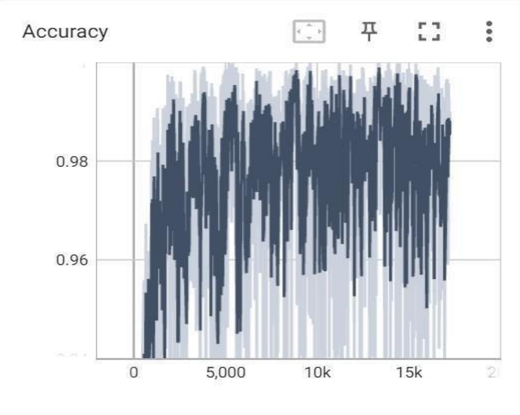
where True Positives (TP) are the cases that the model correctly identified as positive. And False Negatives (FN) are the cases that the model incorrectly identified as negative.

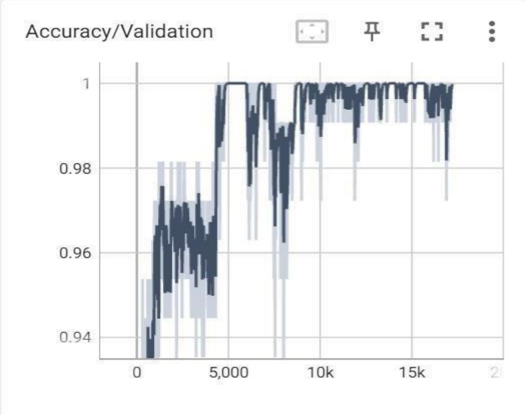
Precision is the percentage of cases that the model correctly identified as positive (lung cancer patients) out of all the cases that the model predicted as positive. Precision can be calculated using the following equation:

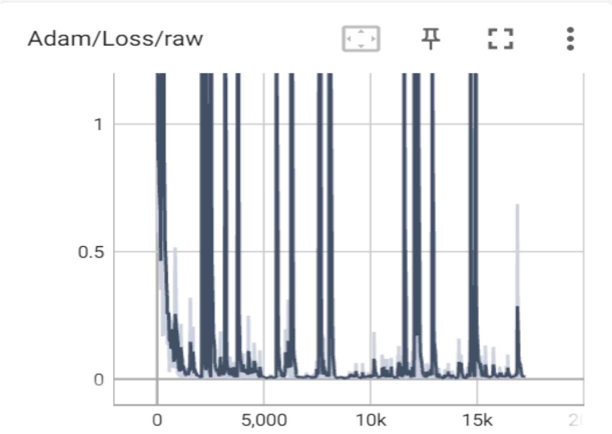
Precision = True Positives/ (True Positives + False Positives)

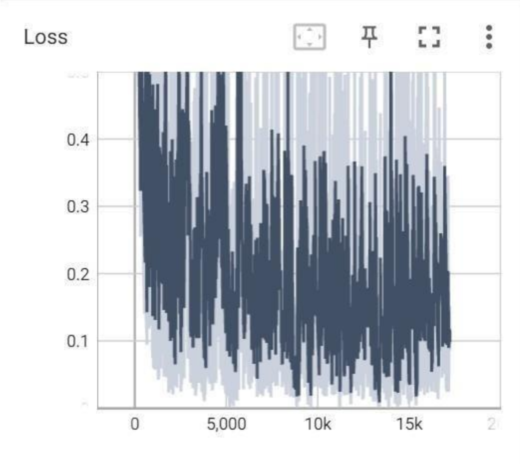
where False Positives (FP) are the cases that the model incorrectly identified as positive.

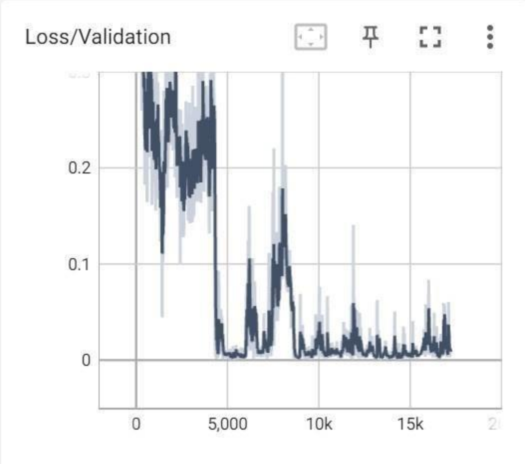
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com











V. CONCLUSION

In conclusion, this report has detailed the implementation of a sophisticated Convolutional Neural Network (CNN) using the ResNet architecture for the crucial task of diabetic retinopathy detection in retinal images. Diabetic retinopathy, a common complication of diabetes, necessitates accurate and timely diagnosis for effective intervention. The implemented model, leveraging the power of deep learning and ResNet's distinctive residual blocks, offers a promising solution for automated classification of retinal images into distinct stages of diabetic retinopathy. The journey from retinal images. The utilization of callbacks, such as Early Stopping and through the code showcased meticulous data preprocessing, employing real-time augmentation to enhance the model's ability to generalize. The ResNet model, with its depthand skip connections, has demonstrated its effectiveness in learning intricate hierarchical features Model Checkpoint, during training ensures optimal model performance and efficient convergence. The model's evaluation on a separate test set has shown promising results, with accuracy serving as a foundational metric for assessing its classification prowess. Additionally, performance metrics such as confusion matrices and classification reports provide a nuanced understanding of the model's strengths and areas for improvement. The inclusion of image prediction and visualization components enhances the interpretability of the model's outputs, allowing for a qualitative assessment of its performance on sample retinal images. The incorporation of diverse performance metrics and visualization techniques contributes to a comprehensive evaluation framework, essential for deploying the model in real-world clinical scenarios. In summary, the implementation presented in this report signifies a significant stride toward the development of a reliable and efficient tool for diabetic retinopathy detection. As we continue to advance in the field of medical image analysis, this model holds the potential to augment existing diagnostic processes, providing a valuable resource for healthcare practitioners in their efforts to combat diabetic retinopathy and improve patient outcomes. Future work may involve fine-tuning the model, exploring larger datasets, and collaborating with healthcare professionals for practical integration into clinical workflows.

VI. REFERNCES

[1] Faridoddin Shariaty, Mojtaba Mousavi, “Application of CAD systems for the automatic detection of lung nodules”,Informatics in Medicine Unlocked Volume 15, 2019, 100173, https://doi.org/10.1016/j.imu.2019.100173.

[2] Hongsheng Jin, Zongyao Li, Ruofeng Tong, Lanfen Lin, “A Deep 3D Residual CNN for False Positive Reduction in Pulmonary Nodule Detection”, Med Phys. 2018. May; 45(5):2097-2107. doi: 10.1002/mp.12846. Epub 2018 Mar 25.

[3] Kai-Lung Hua, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, Yu-Jen Chen3, “Computer-aided classification of lung nodules on computed tomography images via deep learning technique”, Onco Targets Ther. 2015Aug 4;8:2015-22. doi: 10.2147/OTT.S80733. eCollection 2015.

[4] Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, Jie Tian, “Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation”, Medical Image Analysis 40 (2017)172- 183, http://dx.doi.org/10.1016/j.media.2017.06.014.

[5] Pranjal Sahu , DantongYu, Mallesham Dasari , Fei Hou , and Hong Qin, “A Lightweight Multi-Section CNN for Lung Nodule Classification.

[6] Jue Jiang, Yu-chi Hu1, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy, Gig Mageras and Harini Veeraraghavan, “Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images,” 23 July 2018, volume 38, doi : 10.1109/TMI.2018.2857800, IEEE

[7] Ramkumar M, Ganesh Babu C, Abdul Wahhab A R, Abinaya k, Abinesh Balaji B, Aniruth Chakravarthy N, “Detection and Diagnosis of Lung Cancer using Machine Learning Convolutional Neural Network Technique,” doi : 10.1109/STCR55312.2022.10009607

[8] Shahid Mehmood, Taher M. Ghazal, Muhammad Adnan Khan, Muhammad Zubair, Muhammad Tahir Naseem, Tauqeer Faiz, Munir Ahmad, “Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing,” 10 February 2022, volume 10, doi : 10.1109/ACCESS.2022.3150924

[9] Somya Srivastav, Kalpna Guleria, Shagun Sharma, “Lung Cancer Detection using Deep Learning-based Convolutional Neural Networks,” doi : 10.1109/ASIANCON58793.2023.10269839

[10] Ahmed Saihood, Hossein Karshenas and Ahmad Reza Naghsh-Nilchi, "Multi- Orientation Local Texture Features for Guided Attention-Based Fusion in Lung Nodule Classification," 16 February 2023, volume 11, doi : 10.1109/ACCESS.2023.3243104