IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING BASED LUNG CANCER DISEASE PREDICTION SYSTEM

[1]Prof. Dr. Pavan G P, [2]Brijesh V, [3]Eshwar Patil, [4]Harshitha Suresh,[5]Nisha C D

[1]Professor, [2]Student, [3]Student, [4]Student, [5]Student,

[1]Information Science & Engineering,

[1]AMC Engineering College, Bengaluru, Karnataka, India.

Abstract: Small cell lung cancer (SCLC) is a type of malignant tumor that is characterized 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 tumor 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 localizeobjects 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 thelungs 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.

II. LITERATURE SURVEY

According to [1] 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.

According to [2] 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 with the 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.

According to [3] Rapid 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 tumours, 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 difficult for 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.

According to [4] 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.

According to [5] Small cell lung cancer (SCLC) is a type of malignant tumour that is characterized 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 bloodveins, as well as to classify the tumour. The R-CNN techniques demonstrated in Part 1 primarily use regions to localizeobjects 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.

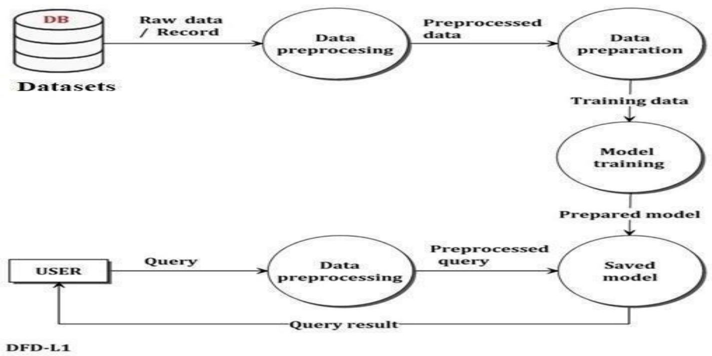
III. EXISTING SYSTEM

A trainable neural network that learns long-range 3D contextual information using a lightweight 3D convolution neural network and fine-grained intra-slice semantic information using a 2D convolution neural network. To deal with the anisotropic dimensions of CT volumes and reduce the computational cost of the 3D convolution neural network, they use spatiotemporalseparable3D (S3D) convolutions. We use dilated convolutions in a 2D convolution neural network to enlarge the receptive field while retaining high resolution to retain a large amount of semantic information about smaller objects. Create a hybrid features fusion module (HFFM) to effectively fuse 2D and 3D features. To address the issue of data imbalance training, they also use the generalized Dice loss function.

IV.PROPOSED SYSTEM

In a proposed system used to extract the features of Lung and it used to increase the resolution and efficiency level of an input. Here, view of Lung density it gives the exact level of a density and it also finds the all supplements in Lung. To find a COVID-19 and fracture it gives high resolution to find level of a COVID-19. It is the combined process of a Lung feature extraction.

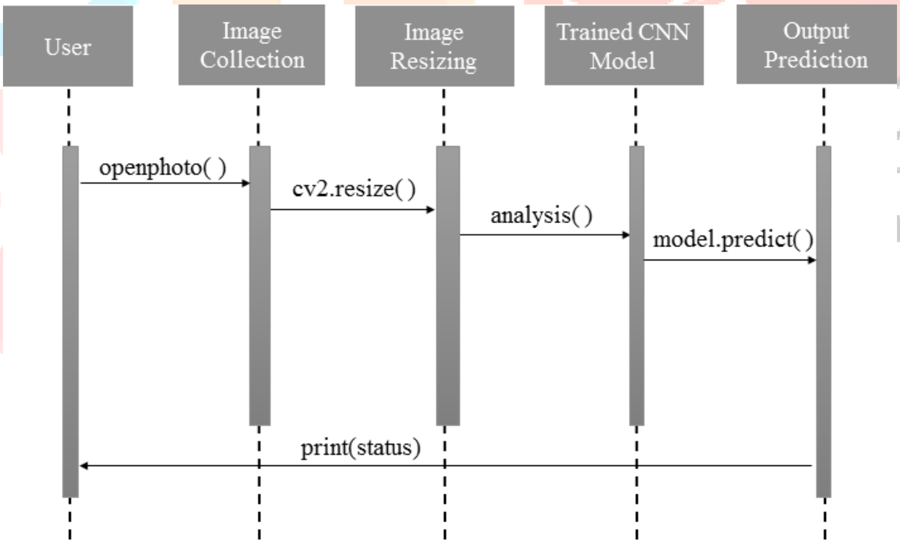
Preprocessing: In preprocessing stage, the median filter is used to restore the image undertest by minimizing the effects of the degradations during acquisition. Various preprocessing and segmentation techniques of lung nodules are discussed further.
Feature Selection: Process of identifying the most relevant and informative features from the dataset. There are various techniques for feature selection.
Feature Extraction: It is the process of transforming raw data into a reduced representation of its most important features. In machine learning and data analysis, feature extraction aims to capture essential information from complex data and represent it in a more concise.
Classification: classification refers to the process of categorizing individuals into different groups based



on their likelihood of having lung cancer.

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,



IV. RESULTS AND DISCUSSION

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:

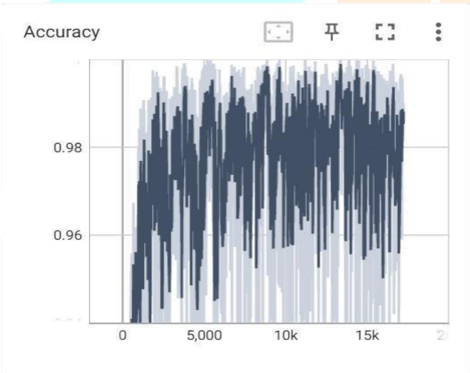
Recall = True Positives / (True Positives + False Negatives)

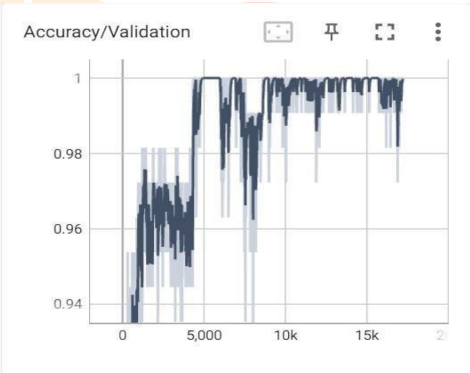
where True Positives (TP) are the cases that the model correctly identified as positive. And False Negatives(FN) are the cases that the model incorrectly identified as negative.

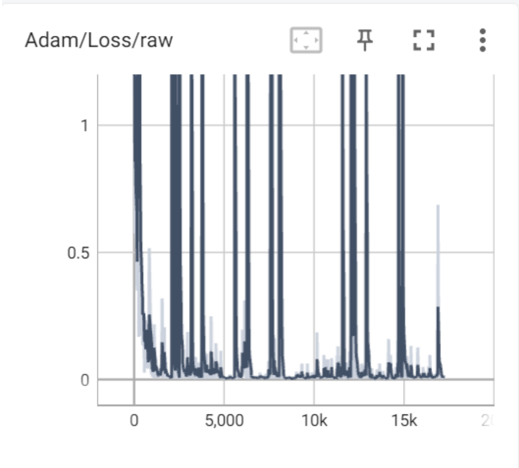
Precision is the percentage of cases that the model correctly identified as positive (lung cancer patients) out of all the cases that the model predicted as positive. Precision can be calculated using the following equation:

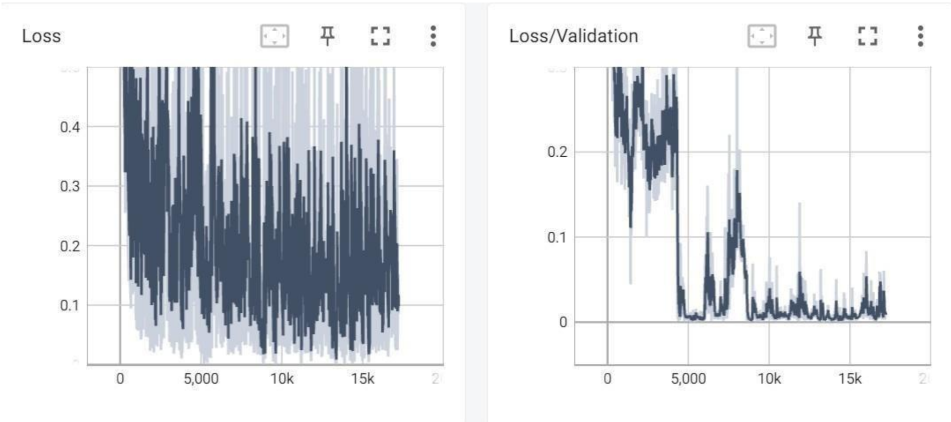
Precision = True Positives/ (True Positives + False Positives)

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









CONCLUSION

Here the survey papers on lung nodule classification are discussed. The detection and classification process which is to be adopted for lung nodule classification are discussed based on the available survey. This paper gives a broader description of the types of lung nodules present in our body and the detection procedures in order to save the human life in early stages.

ACKNOWLEDGMENT

This is a great pleasure and immense satisfaction to express my deepest sense of gratitude and thanks to everyone who has directly or indirectly helped me in completing my paper work successfully. I express my gratitude towards guide Dr. Pavan G P who guided and encouraged me in completing the work in scheduled time. I am very thankful to acknowledgement to Prof. Dr. Amutha R (HOD, Information Science and Engineering). I present my sincere thanks to Prof. Dr. K Kumar, Principal for moral support and providing excellent infrastructure in carrying out the paper work. No words are sufficient to express my gratitude to my family for their unwavering encouragement. I also thank all friends for being a constant source of my support.

REFERENCES

[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. 2015 Aug 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 , Dantong Yu, Mallesham Dasari , Fei Hou , and Hong Qin, "A Lightweight Multi-Section CNN for Lung Nodule Classification