Diagnosis Of Cancer and covid 19 using x-ray images

Ojasvi Yadav

SRM University

Abhigyan Pratap

SRM University

P Murali Sir(Guide and Panel Head)

***Abstract*— With a limited number of COVID-19 test kits available in medical facilities, it is important to develop and implement an automatic detection system as an alternative diagnosis option for COVID-19 detection that can be used on a commercial scale.Specifically, chest X-Ray images can be analyzed to identify the presence of COVID-19 in a patient.**
Due to the high availability of large-scale annotated image datasets, great success has been achieved using a convolutional neural network for image analysis and classification.Input is obtained in the form of chest x- rays images. Output results are acquired instantly in real-time which predicts if the person suffers from it Covid or lung cancer The state-of-the-art methods work on the RCNN algorithm which makes it less accurate and more time-consuming.

Keywords*— *Cancer, X-ray images, deep learning, advanced CNN, COVID-19 & Lung cancer detection

I. INTRODUCTION

A pandemic is a medical outbreak that affects several communities globally. The 20th century saw numerous pandemics across the globe. Pandemics are primarily brought on by flu viruses. Since the behavior of these viruses changes with the seasons, it is necessary to foresee it in order to prevent infection. Most viral forecasts are accurate when made by health specialists. However, certain viruses exhibit unusual behavior and are challenging to predict. Such viruses generate pandemics because people lack the immunity to fight them off. The most recent coronavirus illness, COVID-19, has recently emerged and is spreading quite quickly. The disease has now spread to 199 nations and territories since it was first identified in Wuhan, China, in December 2019. In the literature, there are many methods for diagnosing COVID-19 utilizing machine learning methods on radiological images.

II. NEED OF THIS PROJECT

On CT images, a deep learning-based algorithm is used to look for COVID-19. Additionally, several researchers have created available databases that include the chest X-ray scans of COVID-19 patients. On these open datasets, a technique called COVID-Net is created and used to diagnose COVID-19. Results are good when using deep learning to diagnose chest X-ray pictures. For processing medical imaging data, deep learning models are frequently utilized. The multiresolution analysis feature is made use of by the suggested network. There are many benefits to using deep networks and wavelet transformations together. The network receives the wavelet decomposition as input. It is not a conventional convolutional neural network that is being used (CNN). In this work, a depth-wise separable network is used.

The basic objectives of cancer prediction and prognosis differ from the objectives of detecting and diagnosing cancer. In predicting/prognosis for cancer, three predictive domains are considered: 1) predicting cancer susceptibility (i.e., risk evaluation); 2) predicting cancer recidivism, and 3) predictability of cancer survival. In the former case, an attempt is made to predict the probability of developing a type of cancer before the onset of the disease. In the second case, an attempt is made to predict the probability of cancer rearing as a result of the apparent resolution of the disease. In the third case, an attempt is made to predict a result (life expectancy, survivability, progression, tumor susceptibility) after the disease has been diagnosed. In these last two situations, the success of predictive prognosis obviously depends, in part, on the success or quality of the diagnosis. However, the prognosis of a disease can only be established after a medical diagnosis and a prognostic prediction needs to take into account more than a single diagnosis.

III. APPLICATION

The fast-paced human-to-human transmission of the disease is a matter of great concern for the regulatory authorities globally. The control of COVID-19 largely depends on the diagnosis at the right time. The X-Ray facilities are easily accessible in all parts of the world and the results are also produced at a fast pace. Therefore, chest X-Ray images may be utilized for detecting the presence of COVID-19. The development of an automated method based on chest X-Ray images for support in clinical decision-making will be significant for disease control.

IV. EXISTING SYSTEM

The existing systems are also not robust in diagnosing covid-19 and lung cancer from chest X-ray images, as the results are less accurate. The existing models for diagnosis are built on RCNN algorithms which may need a large training time to improve their classification accuracy, the most important aspect of health diagnosis.We also know that the existing systems don't operate well and show inaccurate results when trained with larger datasets. The existing systems are simple
The current system for lung cancer detection employs Recurrent Neural Networks (RNN) as the deep learning technique. RNN is a widely-used technique for sequential data processing, including natural language processing and time-series analysis. However, in the context of lung cancer detection, RNN has some limitations that need to be considered. One of the potential drawbacks of using RNN is that it may not be able to effectively capture the spatial features in lung images. This may lead to lower accuracy rates and misdiagnosis of lung cancer. Additionally, RNN is computationally expensive and requires significant computational resources.

RNN processes sequential data by using memory cells that store the information and pass it to the next cell. However, when dealing with images, RNN processes the image in a sequential manner, which is a suboptimal approach. Moreover, the sequential processing of images by RNN can be time-consuming, especially when dealing with large datasets. These limitations in the existing RNN-based system have led researchers to explore alternative deep-learning techniques for lung cancer detection.

V. PROPOSED METHOD

The objective of this project is the Diagnosis of COVID-19 and Lung cancer from Chest X-Ray Images Using a Wavelet-Based Depth wise Convolution Network The suggested approach is based on spectral pooling using wavelet transforms and a depth-wise separable convolution network. Deep learning and multiresolution analysis are combined to create the network. Due to the high computational cost and the vast number of parameters generated at each layer, classic CNN layers are prone to overfitting enhance by enhancing training convergence, enabling adjustable pooling dimensions, and maintaining accuracies. Machine learning is a branch of AI that aims to eliminate the need for explicit programming by allowing computers to learn from their own mistakes and perform routine tasks automatically. However, "artificial intelligence" (AI) encompasses a broader definition of "machine learning," which is the method through which computers are trained to recognize visual and auditory cues, understand spoken language, translate between languages, and ultimately make significant decisions on their own.

One of the key advantages of CNN is its ability to capture spatial features in images effectively. CNN achieves this by using convolutional layers that learn features from the image, followed by pooling layers that reduce the dimensionality of the features. This process helps CNN to learn the key features of an image, which is crucial for accurate lung cancer detection. Moreover, CNN is computationally efficient and can be run on devices with limited computational resources, making it more accessible for researchers and medical practitioners.

The proposed CNN-based system is a significant improvement over the existing RNN-based system for lung cancer detection. The use of CNN has the potential to
significantly improve the accuracy of lung cancer detection by capturing spatial features in lung images more effectively. Additionally, the proposed system is more computationally efficient than the existing RNN-based system, making it more accessible for researchers and medical practitioners.
In conclusion, while the existing RNN-based system has limitations in capturing spatial features in lung images, the proposed CNN-based system is a promising alternative that offers several advantages. The use of CNN has the potential to significantly improve the accuracy of lung cancer detection and make the system more accessible for researchers and medical practitioners. Therefore, the proposed CNN-based system represents a positive step toward improving the accuracy and efficiency of lung cancer detection.

FEASIBILITYSTUDY

With an eye toward gauging the project's viability and improving server performance, a business proposal defining the project's primary goals and offering some preliminary cost estimates is offered here. Your proposed system's viability may be assessed once a comprehensive study has been performed. It is essential to have a thorough understanding of the core requirements of the system at hand before beginning the feasibility study. The feasibility research includes mostly three lines of thought:

- Economical feasibility
- Technical feasibility
- Operational feasibility
- Social feasibility

3.3.1 ECONOMICAL FEASIBILITY

The study's findings might help upper management estimate the potential cost savings from using this technology. The corporation can only devote so many resources to developing and analyzing the system before running out of money. Every dollar spent must

have a valid reason. As the bulk of the used technologies is open-source and free, the cost of the updated infrastructure came in far cheaper than anticipated. It was really crucial to only buy customizable products.

3.3.2 TECHNICAL FEASIBILITY

This research aims to establish the system's technical feasibility to ensure its smooth development. Adding additional systems shouldn't put too much pressure on the IT staff. Hence, the buyer will experience unnecessary anxiety. Due to the low likelihood of any adjustments being necessary during installation, it is critical that the system be as simple as possible in its design.

3.3.3 OPERATIONAL FEASIBILITY

An important aspect of our research is hearing from people who have actually used this technology. The procedure includes instructing the user on how to make optimal use of the resource at hand. The user shouldn't feel threatened by the system, but should instead see it as a necessary evil. Training and orienting new users has a direct impact on how quickly they adopt a system. Users need to have greater faith in the system before they can submit constructive feedback.

3.3.4 SOCIAL FEASIBILITY

During the social feasibility analysis, we look at how the project could change the community. This is done to gauge the level of public interest in the endeavor. Because of established cultural norms and institutional frameworks, it's likely that a certain kind of worker will be in low supply or nonexistent.

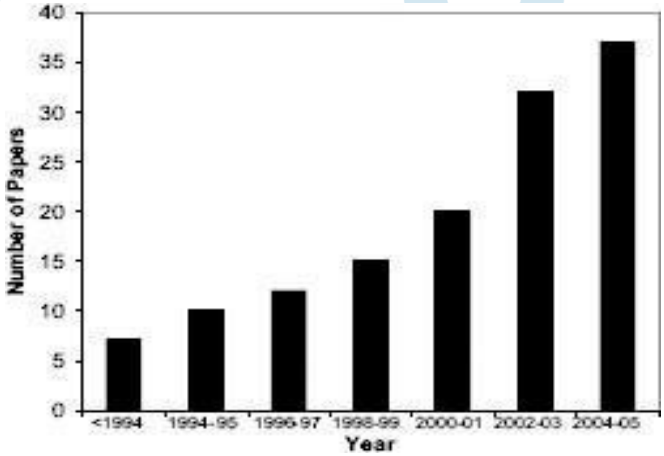


Fig 1.1 Graph - Predict cancer risk, Recurrence and Outcome

VI. TRAINING

Data Gathering

We have gathered the data for our project for the following diseases in the CSV format: Lung and covid x-ray Images: An image dataset

Data Preprocessing

Data preprocessing is a data mining technique used to turn raw data into a format that is both practical and effective. Data preprocessing changes the data into a format that can be processed in data mining, machine learning, and other deep learning tasks more quickly and efficiently. To ensure reliable findings, the techniques are typically applied at the very beginning of the deep learning and AI development pipeline. Although there are many various tools and strategies for preparing data, we have chosen to concentrate on the following measures for our dataset:

Our first step was to collect a dataset of lung cancer images and corresponding labels (e.g., malignant/benign). We used a dataset that is diverse and representative of the population being studied. Once the dataset is collected, it should be preprocessed to ensure that the images are ready for training.
Data Augmentation

Data Augmentation: Data augmentation is a technique used to increase the diversity of the training data by applying various transformations to the images. Data augmentation is especially useful when the dataset is small, as it helps to prevent overfitting and improve the model's generalization ability. Some common data augmentation techniques include random rotations, flips, shifts, and zooms.

The goal of data augmentation is to artificially expand the size of the dataset by generating new variations of the original images. This can help to reduce overfitting and improve the model's ability to generalize to new data. Data augmentation is particularly useful in medical imaging, where the dataset may be limited and the variations in the input data may be small. By applying data augmentation techniques to the input images, the model can be trained on a more diverse set of images, which can improve its ability to detect lung cancer accurately.

Model Training And Validation

Once the base model is set up, we can start training the model using the training data set. We will use a suitable loss function, such as binary cross-entropy for binary classification, and an optimizer, such as Adam optimizer, to train the model. We will monitor the training process to avoid overfitting by using techniques such as early stopping and regularization. During training, we will evaluate the model's performance on the validation set to check if the model is learning correctly and if the hyperparameters are tuned correctly.

Model validation is the process of evaluating the performance of the trained model on a separate set of images to assess its ability to generalize to new data.

During model validation, the performance of the trained model is evaluated on a separate set of images that were not used during training. This helps to assess the ability of the model to generalize to new data and to identify any potential overfitting issues. The performance of the model is typically evaluated using metrics such as accuracy, precision, recall, and F1 score.

VII. TESTING

There are four types of testing:

Unit Testing

The term "unit testing" refers to a specific kind of software testing in which discrete elements of a program are investigated. The purpose of this testing is to ensure that the software operates as expected.

Test Cases

1. Test case for input data pre-processing: This test case should test the pre-processing of input data, such as normalization and resizing. The test should ensure that the input data is correctly pre-processed and ready for the neural network to process.

2. Test case for model creation: This test case should test the creation of the neural network model. The test should ensure that the model architecture is correctly defined and that the layers are connected properly.

3. Test case for layer testing: This test case should test the individual layers of the neural network. The test should ensure that each layer is performing as expected and that the outputs are correct.

Integration Testing

The program is put through its paces in its final form, once all its parts have been combined, during the integration testing phase. At this phase, we look for places where interactions between components might cause problems.

Test Cases

1. Test case for accuracy and precision: This test case should test the accuracy and precision of the lung cancer detection system. The test should ensure that the system is correctly identifying positive and negative cases of lung cancer with a high degree of accuracy.

2. Test case for sensitivity and specificity: This test case should test the sensitivity and specificity of the lung cancer detection system. The test should ensure that the system is correctly identifying true positive and true negative cases of lung cancer.

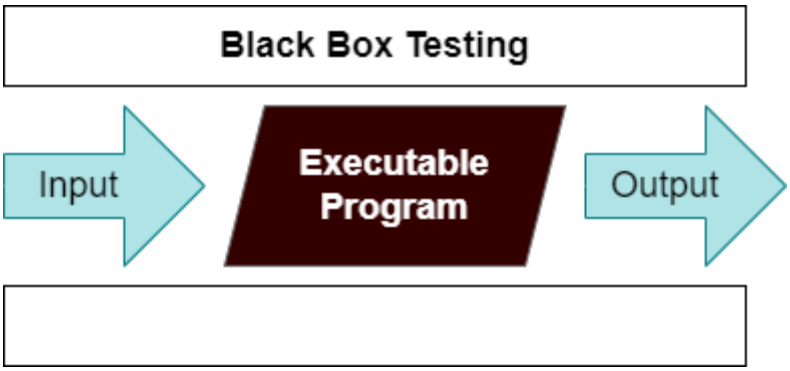
3. Test case for robustness: This test case should test the robustness of the lung cancer detection system. The test should ensure that the system can handle a variety of inputs and scenarios, such as different patient populations or varying image quality.

6.2 TESTING TECHNIQUES

There are many different techniques or methods for testing the software, including the following:

BLACK BOX TESTING

During this kind of testing, the user does not have access to or knowledge of the internal structure or specifics of the data item being tested. In this method, test cases are generated or designed only based on the input and output values, and prior knowledge of either the design or the code is not necessary. The testers are just conscious of knowing about what is thought to be able to do, but they do not know how it is able to do it.



For example, without having any knowledge of the inner workings of the website, we test the web pages by using a browser, then we authorise the input, and last, we test and validate the outputs against the intended result.

Test Cases

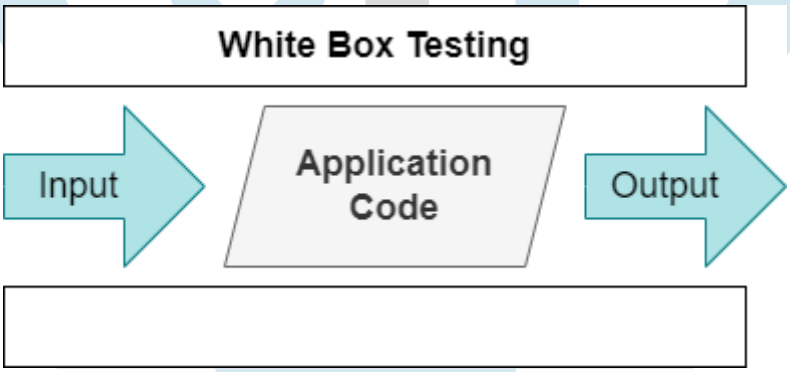
1. Test case for input validation: This test case should ensure that the neural network can handle a wide range of inputs, including inputs that are outside of the expected range or contain errors or anomalies.

2. Test case for output validation: This test case should ensure that the output of the neural network is correctly formatted and contains the expected information, such as the probability of lung cancer or the location of a tumor.

3. Test case for edge cases: This test case should test the neural network's ability to handle edge cases, such as inputs with very low or very high probability of lung cancer or inputs that contain multiple tumors.

WHITE BOX TESTING

During this kind of testing, the user is aware of the internal structure and details of the data item, or they have access to such information. In this process, test cases are constructed by referring to the code. Programming is extremely knowledgeable of the manner in which the application of knowledge is significant. White Box Testing is so called because, as we all know, in the tester's eyes it appears to be a white box, and on the inside, everyone can see clearly. This is how the testing got its name.



As instance, a tester and a developer examine the code that is implemented in each field of a website, determine which inputs are acceptable and which are not, and then check the output to ensure it produces the desired result. In addition, the decision is reached by analyzing the code that is really used.

Test Cases

1. Test case for weight initialization: This test case should ensure that the weights of the neural network are initialized correctly and do not cause numerical instability or other issues.

2. Test case for activation functions: This test case should test that the activation functions used in the neural network are appropriate for the task of lung cancer detection and do not lead to issues such as vanishing gradients.

3. Test case for loss functions: This test case should ensure that the loss function used in the neural network is appropriate for the task of lung cancer detection and is effectively minimizing the error between predicted and actual results.

ADVANTAGES OF USING PYTHON

The following are the advantages of using Python:

Variety of Framework and libraries:

A good programming environment requires libraries and frameworks. Python frameworks and libraries simplify program development. Developers can speed up complex project coding with prewritten code from a library. PyBrain, a modular machine learning toolkit in Python, provides easy-to-use algorithms. Python frameworks and libraries provide a structured and tested environment for the best coding solutions.

Reliability

Most software developers seek simplicity and consistency in Python. Python code is concise and readable, simplifying presentation. Compared to other programming languages, developers can write code quickly. Developers can get community feedback to improve their products or app. Python is simpler than other programming languages, therefore beginners may learn it quickly. Experienced developers may focus on innovation and solving real-world problems with machine learning because they can easily design stable and trustworthy solutions.

Easily Executable

Developers choose Python because it works on many platforms without change. Python runs unmodified on Windows, Linux, and macOS. Python is supported on all these platforms, therefore you don't need a Python expert to comprehend it. Python's great executability allows separate applications. Programming the app requires only Python. Developers benefit from this because some programming languages require others to complete the job. Python's portability cuts project execution time and effort.

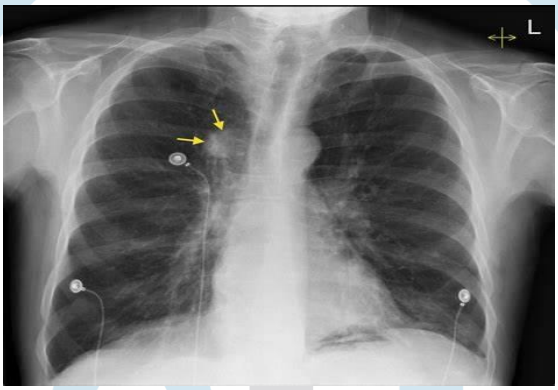
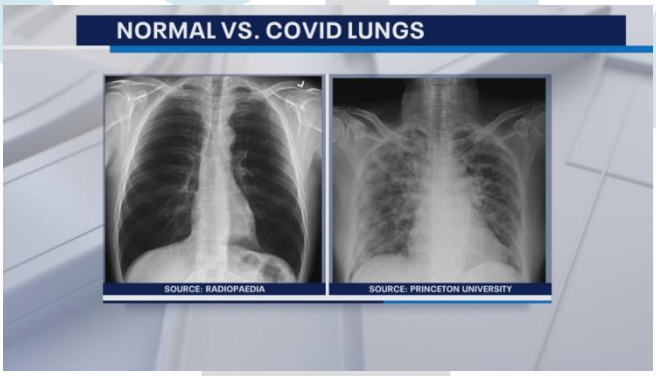


Image of a person diagnosed with lung cancer



Healthy lungs vs lungs diagnosed with covid

VIII. CONCLUSIONS

Lung cancer and covid in recent times is a leading cause of death worldwide, and early detection is crucial for improving patient outcomes. Convolutional neural networks (CNNs) have emerged as a powerful tool for computer-aided diagnosis of lung cancer, providing accurate and efficient detection of suspicious nodules and lesions in medical images. CNNs are a type of deep learning algorithm that is able to automatically learn and extract features from images, allowing them to accurately classify the images based on those features. In the case of lung cancer detection, CNNs can be trained on large datasets of lung CT scans to learn the relevant features of malignant and benign nodules and lesions. Several studies have demonstrated the effectiveness of CNNs for lung cancer detection, with high levels of sensitivity and specificity. These models have also been shown to outperform traditional computer-aided detection (CAD) systems and human radiologists in certain cases. However, there are still some challenges and limitations associated with CNNs for lung cancer detection. One major limitation is the need for large amounts of high-quality training data, which can be difficult to obtain and may not be representative of diverse patient populations. Additionally, CNNs may struggle with identifying subtle nodules or lesions that are not easily distinguishable from surrounding tissue. Overall, CNNs have the potential to greatly improve the accuracy and efficiency of lung cancer detection, particularly when used in conjunction with other diagnostic tools and expert medical interpretation. Further research and development of these models may lead to even more accurate and effective methods for the early detection and treatment of lung cancer and covid. Developers can get community

feedback to improve their products or app. Python is simpler than other programming languages, therefore beginners may learn it quickly.

REFERENCES

[1] Jieli Zhou, Baoyu Jing, Zeya Wang, Hongyi Xin, Hanghang Tong, "SODA: Detecting COVID-19 in Chest X-rays with Semi-supervised Open Set Domain Adaptation", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021

[2] Iosif Mporas, Prasitthichai Naronglerdrit, "COVID-19 Identification from Chest X-Rays", International Conference on Biomedical Innovations and Applications (BIA), 2020

[3] Talha Anwar, Seemab Zakir, "Deep learning based diagnosis of COVID-19 using chest CT-scan images", IEEE 23rd International Multitopic Conference (INMIC), 2021

[4] Kaoutar Ben Ahmed, Gregory M. Goldgof, Rahul Paul, Dmitry B. Goldgof, Lawrence O. Hall, "Discovery of a Generalization Gap of Convolutional Neural Networks on COVID-19 X-Rays Classification", IEEE Access (Volume: 9), 2021

[5] Surbhi Anand, Nilesh Nishant, Tanyam Singal, "Automated detection of COVID-19 from chest X-rays using CNN", 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022

[6] D. Haritha, M. Krishna Pranathi, M. Reethika, "COVID Detection from Chest X-rays with DeepLearning: CheXNet", 5th International Conference on Computing, Communication and Security (ICCCS), 2020

[7] D. Haritha, N. Swaroop, M. Mounika, "Prediction of COVID-19 Cases Using CNN with X-rays", 5th International Conference on Computing, Communication and Security (ICCCS), 2020

[8] Pir Masoom Shah, Faizan Ullah, Dilawar Shah, Abdullah Gani, Carsten Maple, Yulin Wang, Shahid, Mohammad Abrar, Saif Ul Islam, "Deep GRU-CNN Model for COVID-19 Detection From Chest X-Rays Data", 2021

[9] Vaibhavi C. Shinde, Pradnya S. Kulkarni, "Automatic COVID-19 Detection from Chest X-Rays using Deep Learning Techniques", International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022