



Prediction Of Lung Cancer In The Ct Image Using Machine Learning Algorithm

¹S. Niranjana, ²DR. D. Chitra Devi, ³S. Nivedha

¹PG Scholar, ²Associate Professor, ³Assistant Professor

¹Department of Computer Science and Engineering,

¹S.A. Engineering College, Chennai, India

Abstract: Lung tumor are responsible for a large number of death fatalities globally emphasizing the urgent need for new CAD tools for early identification and categorization this research automatically detects and categorizes lung cancer subtypes using automated CT scans using the most advanced InceptionV3 model a convolutional neural network to distinguish between adenocarcinoma, squamous and large cell carcinoma the model was trained on a variety of annotated CT scans allowing it to understand and recognize unique imaging features for each subtype. InceptionV3 architecture is designed to be highly precise and computationally efficient allowing lung pictures to be processed quickly by automating the diagnosis process. This research aims to improve patient outcomes and help healthcare providers make better treatment decisions and deployment of this system diagnostics could undergo a revolution if this approach is implemented successfully becoming more dependable and accessible while also relieving the strain on medical practitioners by displaying a better accuracy of 99%.

Keywords: Computed Tomography (CT), Lung cancer, InceptionV3, Deep Learning, Convolutional Neural Network (CNN).

I. INTRODUCTION

A significant percentage of deaths related to cancer worldwide are caused by lung cancer, making it one of the most difficult problems facing modern medicine. Because of its intricate pathophysiology and numerous subtypes, prompt and precise diagnosis is essential to raising patient survival rates and improving their standard of life. The diagnosis process has historically been hampered by the arbitrary interpretation of pathological samples and the use of invasive procedures like biopsies. There is an urgent need for innovative ways to enhance the precision and effectiveness of lung cancer screening [1]. There are many types of cancers that impact the lungs, but the term "lung cancer" typically refers to two primary categories They are classified as non-small cell lung cancer as well as small cell lung cancer [2]. Although it can impact anyone, small cell lung cancer is an uncommon and aggressive type of lung cancer that mainly affects people with a long history of tobacco use. If detected early, some patients can be cured, while others may receive treatment to extend their lives. Quitting smoking is the most effective approach to avoiding small-cell lung cancer. This kind of cancer is less common than non-small cell lung cancer and makes up 15% of all lung cancer cases [3]. Approximately eighty-five percent of lung cancers are non-small cell [4] and it is frequently diagnosed after it has spread to other areas of the body and encompasses three main types. Adenocarcinoma develops from mucus-secreting cells in the outer regions of the lungs and primarily affects current or former smokers, though it can also occur in never-smokers and is often diagnosed at a younger age than other lung cancers. Squamous cell carcinoma, which arises from the flat squamous cells lining the airways, typically forms in the central parts of the lungs as well as is closely connected to a history of smoking. Finally, because of its rapid growth and propensity to spread, big cell carcinoma, sometimes referred to as undifferentiated carcinoma, can develop in any part of the lung and presents treatment issues [5].

Recent advancements in medical imaging, particularly computed tomography (CT), have transformed the diagnostic landscape. CT scans offer cross-sectional, detailed views of the lungs, making it possible to spot minute anomalies that could point to malignancy. However, the interpretation of these images is often complex and requires significant expertise, which can lead to variability in diagnoses. To address these challenges, our project leverages deep learning techniques, specifically the InceptionV3 model, to automate the classification of lung cancer subtypes directly from CT images. The InceptionV3 model stands out for its ability to achieve high accuracy while maintaining computational efficiency, making it an ideal candidate for this task. By training the model on a diverse dataset of annotated CT scans, we aim to enable it to recognize and differentiate between key imaging features associated with adenocarcinoma, squamous carcinoma, and large-cell carcinoma. In addition to minimizing human mistakes, this automatic classification drastically cuts down on diagnosis time, allowing medical practitioners to make well-informed decisions more quickly. This study may be used by doctors as a fallback option for predicting lung cancer, which could lead to better patient outcomes and faster treatment. We want to advance lung cancer identification and treatment by reducing the burden on radiologists and providing reliable and efficient tools [6]. This study employs an ensemble approach to classify lung cancer utilizing the deep learning model InceptionV3. The remaining portions of the manuscript are broken down as follows: Section 2 discusses the research literature review. Section 3 provides a more thorough discussion of the recommended methodology. The results of the study were examined in Section 4, along with a conclusion given in Section 5.

II. LITERATURE REVIEW

One of the most promising diagnostic tests is the CT scan, which can detect lung cancer. High-resolution, high-contrast CT scan pictures can be used to evaluate the lungs in three dimensions. This paper presented an automated approach to lung cancer detection and implemented the model on InceptionV3, which is designed to be extremely accurate and computationally efficient, allowing for quick processing of lung pictures. CNN-Based Model was specifically designed for the early identification of lung cancer by CT scan image analysis, and it was proposed in [5] to classify lung cancer using CT scan pictures using DL (deep learning). The outcomes demonstrate our DL model's efficacy and validate its position as an excellent diagnostic tool for early lung cancer diagnosis. With an impressive overall accuracy of 99.20% in predicting cancer types, this model represents a notable advancement over existing approaches. In [6] author proposed Design of DL-based Method to Use CT Scan Images to Predict Lung Cancer for precise and effective model for detecting lung cancer by pre-processing CT images to enhance their quality, followed by extracting features using deep convolutional neural networks. These features are then processed by a classification model to determine the presence of lung cancer. A large dataset of CT images, including both malignant and non-cancerous patients at different stages of malignancy, will be used to assess the model's effectiveness. The study's average classification accuracy for identifying lung cancer from CT scans was 72.41%. The results may provide physicians with a trustworthy instrument for identifying lung cancer early on, enabling timely treatment and favorable outcomes for patients. In [7] proposed CNN algorithm categorizes input data based on its unique features. In the context of analyzing numerous computed tomography (CT) images, early lung cancer detection is crucial for saving lives. The proposed approach focuses on accurately identifying the specific type of cancer and predicting whether it is benign or aggressive. The deployment of this model aims to enhance the system's accuracy. The results indicated an overall detection accuracy of 98.4% for lung cancer and a 98.8% accuracy rate in predicting the cancer type.

III. PROPOSED SYSTEM METHODOLOGY

The suggested approach aims to revolutionize the detection as well as classification of lung carcinoma by efficiently evaluating CT scans while utilizing state-of-the-art deep learning techniques, particularly the InceptionV3 model. This innovative approach seeks to automate the diagnostic procedure by significantly reducing the requirement for manual interpretation. This would facilitate the precise identification of lung cancer subtypes, such as squamous carcinoma, adenocarcinoma, as well as large cell carcinoma, while reducing subjectivity and variability in diagnoses. Effective data processing and evaluation will be made possible by the proposed solution's simple interaction with the healthcare imaging infrastructure now in place. The InceptionV3 model will use a carefully chosen dataset of annotated CT scans to train it to recognize the unique features associated with different types of lung cancer. The automated system will provide healthcare imaging infrastructure. The research area of DL has improved in the medical field which includes detection of the cancers and tumors with the help of CT scan images or MRI Images, many samples of images have been

collected from the different CT scanners or machines, improvising the images and making them into fit a similar format is more likely to build and train a model. In this paper, we have proposed the CNN Algorithm, in which we have used different dense and pooling layers to build a model. The model has been set up with different dimensions and attributes that are related to the Model and the output of the images. This approach classifies as well as detects lung cancer in CT scan images of patients both with & without the initial stages of lung cancer. A precise lung cancer classifier could expedite as well as lower the expenses of lung cancer screening, facilitating broader patient classification. We train CNN as well as InceptionV3 algorithms employing CT Scan Lung Cancer Nodules dataset and determine patient survival rates using both algorithms. Here we will be using 20 epoch iterations so that we will get better accuracy of 99 %.

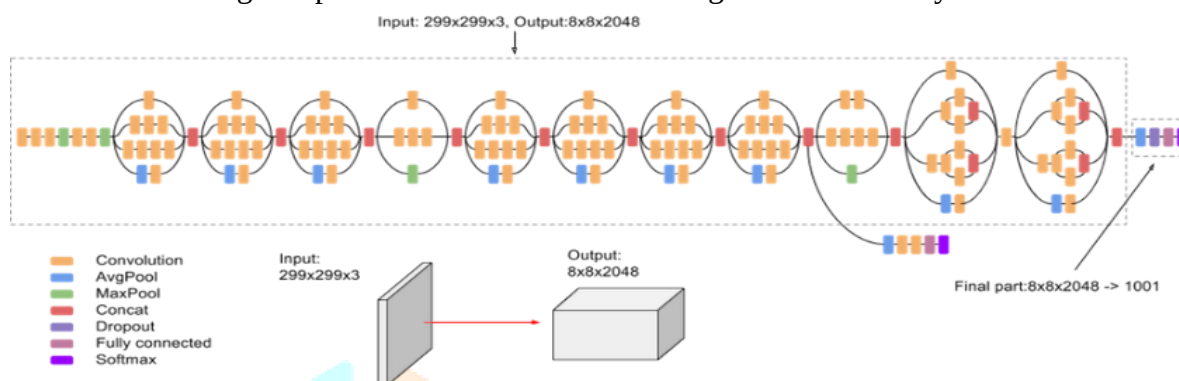


Figure 1: Inception V3 Model

3.1 LUNG CT DATASET

A representative and varied dataset is required to classify different types of lung cancers, specifically adeno, squamous, and big cell carcinoma. To achieve this, we obtained our dataset from the popular Kaggle website. It consists of approximately 1000 CT scans, each carefully selected to capture the distinctive features of each of the three types of lung tumor.

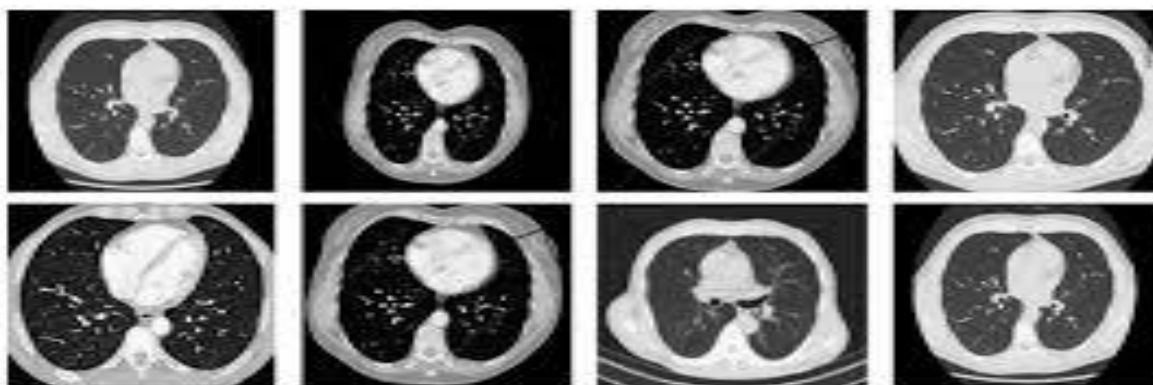


Figure 2: Sample Lung Tumor CT Datasets

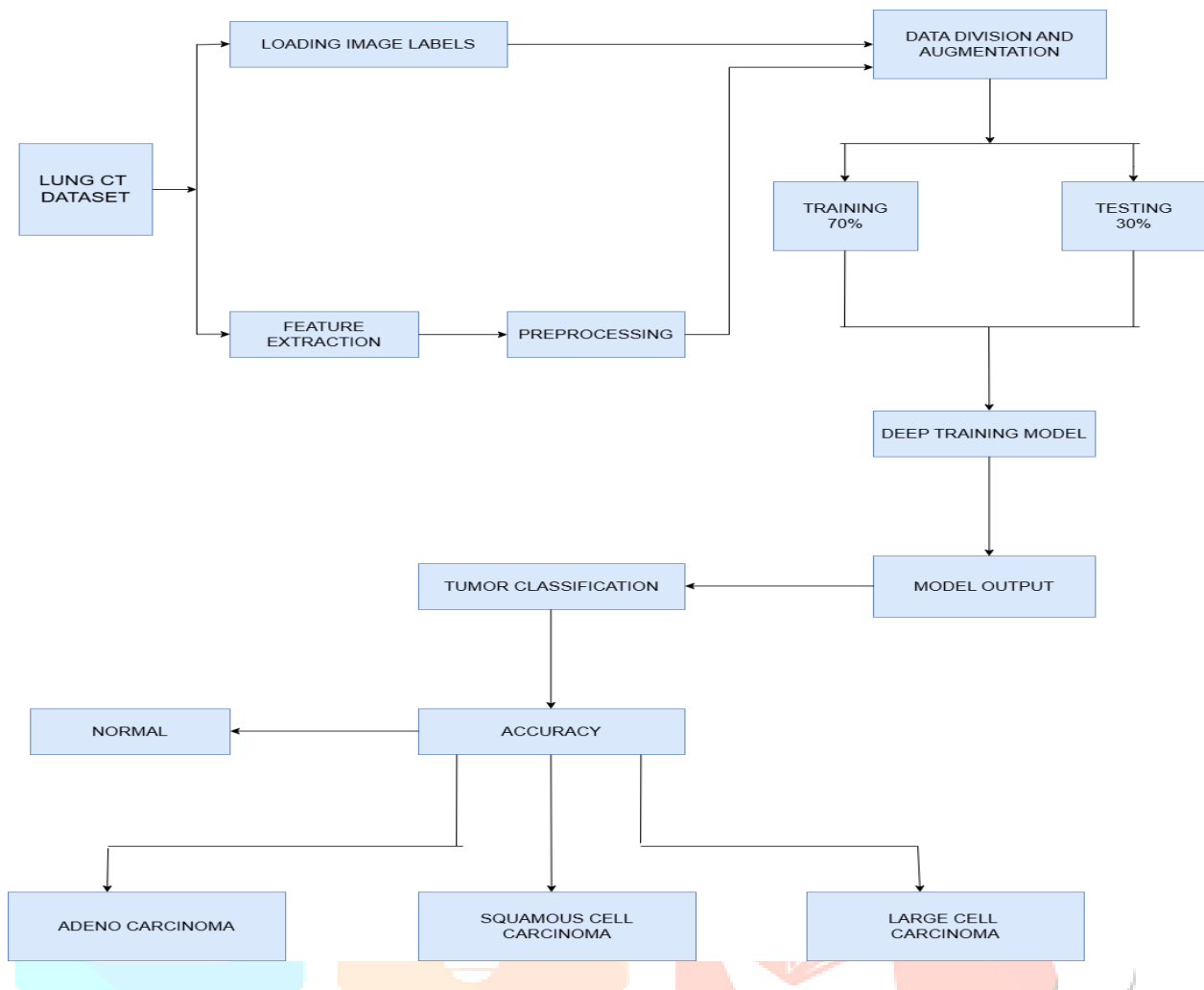


Figure 3: Block Diagram for Detecting and Classifying the Lung Tumor from the CT Images

3.2 FEATURE EXTRACTION

The extraction of patterns from a system image is described as the process that transforms raw data into usable numerical patterns meanwhile preserving the original dataset information. The decrease in our model image size procedure, which breaks down a starting raw dataset collection into small and easier-to-manage components.

3.3 PRE - PROCESSING

This stage is to classify the photographs into a uniform format. To maintain a constant image pixel resolution of 64×64 , the images will be transformed into grayscale. Preprocessing enhances the clarity of input data images and aids in the removal of noisy information.

3.4 DATA DIVISION AND AUGMENTATION

Data augmentations aid in the inclusion of huge amounts of information by adding substantially altered versions of previously stored data or entirely creating synthetic data from existing data. During model training, it functions as a regularizer as well as mitigates overfitting.

3.5 TRAINING AND TESTING

The 3 subfolders contain the raw data set that was downloaded. The initial subset, sometimes referred to as the learning data, is sent to the system so that it may identify and learn patterns from our real dataset. It serves to instruct our system in this manner. You can use the evaluation and verification data in the other subfolder to evaluate and improve our algorithms' training. Test data is usually smaller than train data. The model requires extensive input to identify as well as learn significant sequences.

3.6 INCEPTIONV3 MODEL AND OUTPUT

A major turning point in CNN's growth was the inception network, which classified various image types. It employs numerous strategies to improve performance, both in terms of accuracy and speed. Inception V3 Total having 48 Layer network in this v3 model is trained on ImageNet dataset for 1000 different classes. In the case of Inception, images need to be 299*299*3 Pixels in size. The Inception Layer combines 1*1, 3*3, and 5*5 convolutional layers with their output filter. Since alternative models can lead to overfitting and computational complexity, we employ variants of the Inception model to lessen these kinds of issues. The model is constructed using different hyperparameters and optimization techniques to detect tumor after 30% is testing and 70% is training. When the input is provided and the result matches the expected output, the model's efficiency is demonstrated. A commonly used Python library module, a portable GUI development library is called Tkinter. Tkinter imparts a native appearance to Python-coded graphical user interfaces for Windows, X-Windows, along with Mac OS by offering an object-oriented interface to the open-source and free Tk framework [12]. All major Python distributions include with Tkinter, which is used to create GUIs in Python [11]. This is the only framework in the standard library for Python. Python and Tkinter's interaction makes it quick and simple to create GUI apps. Making a GUI application is made easier with Tkinter. Using Tkinter, we're developing a GUI application that allows users to upload CT lung images. The file manager can be opened instantly to load the image using the fileopenbox approach. The outcome is printed for the given image (figure 5) once the applied algorithm has determined whether or not the image contains a tumor.

IV. RESULTS AND DISCUSSION

This section discusses the training as well as validation results for the suggested CNN classifier model. The model had been constructed utilizing CT-images obtained from Kaggle open-access dataset. There were a number of pre-processing and data-augmentation methods employed to improve the image quality and size. The Keras API was utilized to train the proposed model, which also made use of hyperparameters like batch size and the Adam optimizer. The suggested model had been effectively trained to classify the tumor with an accuracy result of about 99% using 70percent of the training data and 30percent testing data.

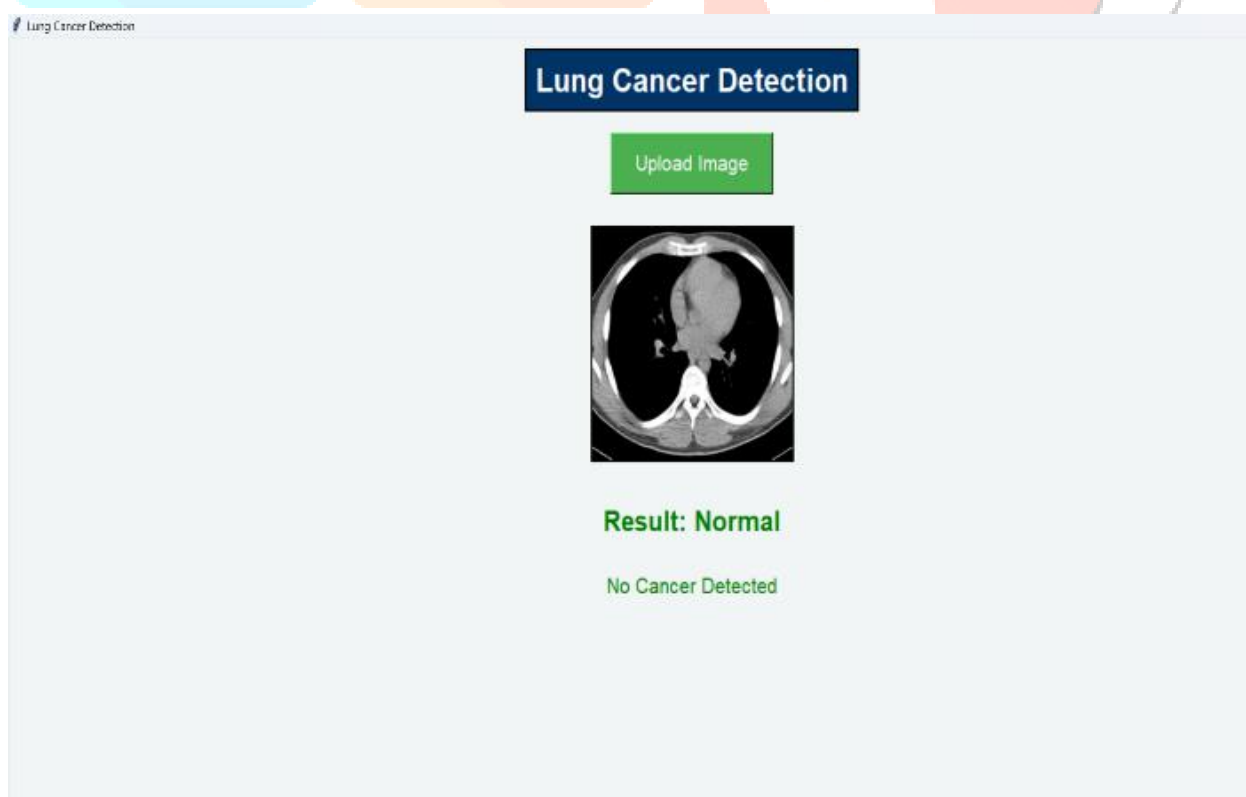


Figure 4: Result of Normal Lung CT Image

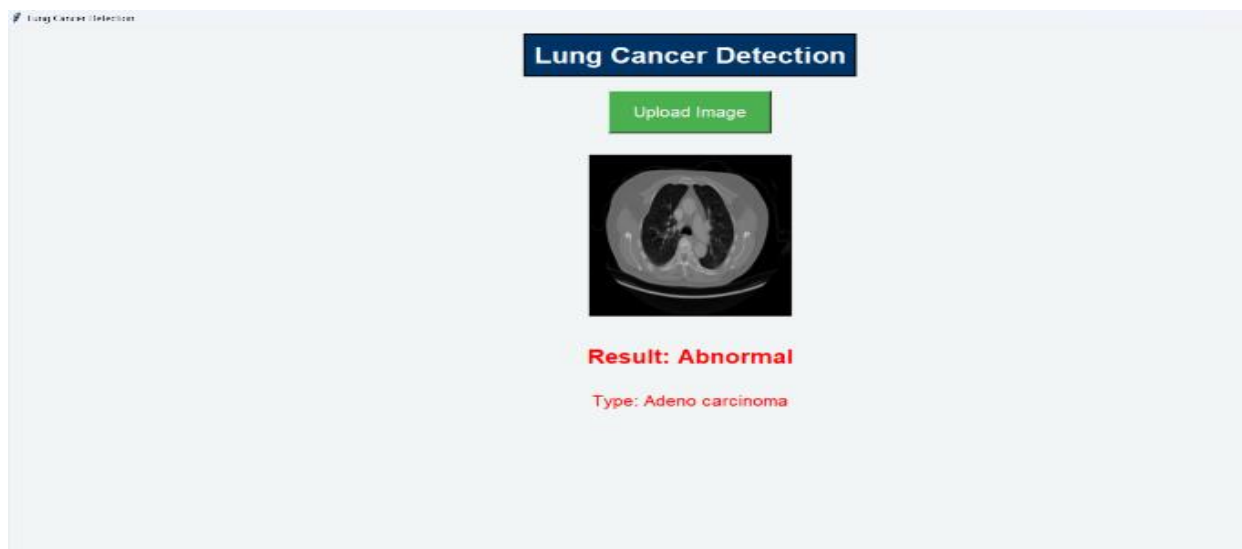


Figure 5: Results of Adeno Carcinoma Lung Image

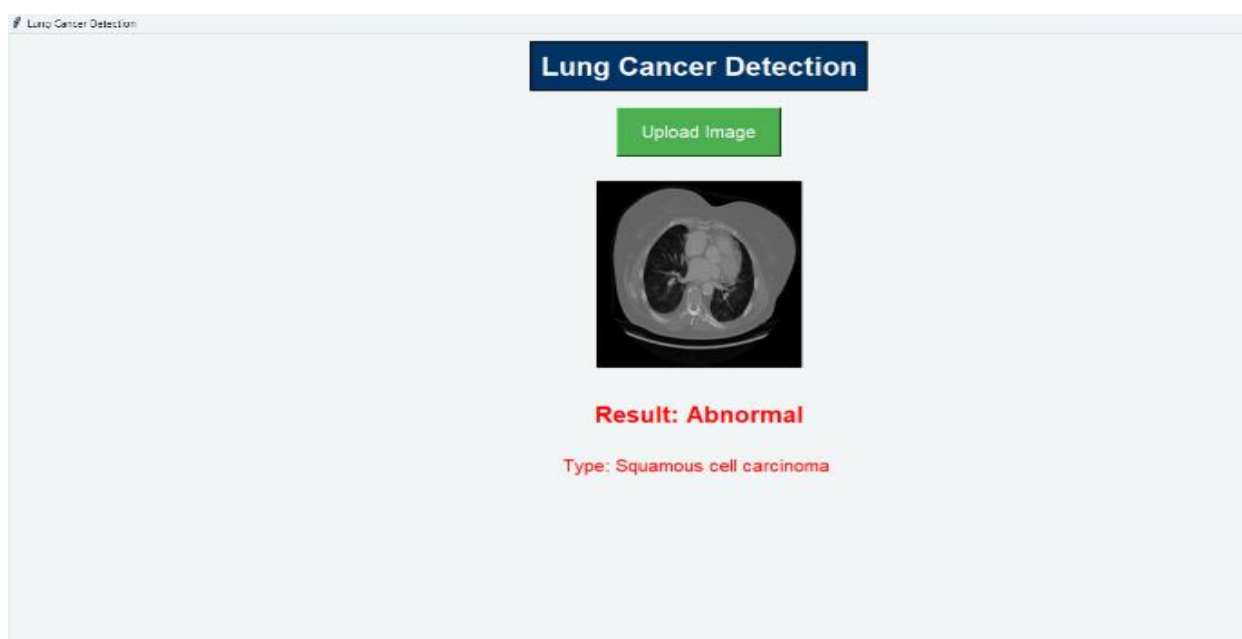


Figure 6: Results of Squamous Cell Carcinoma Image

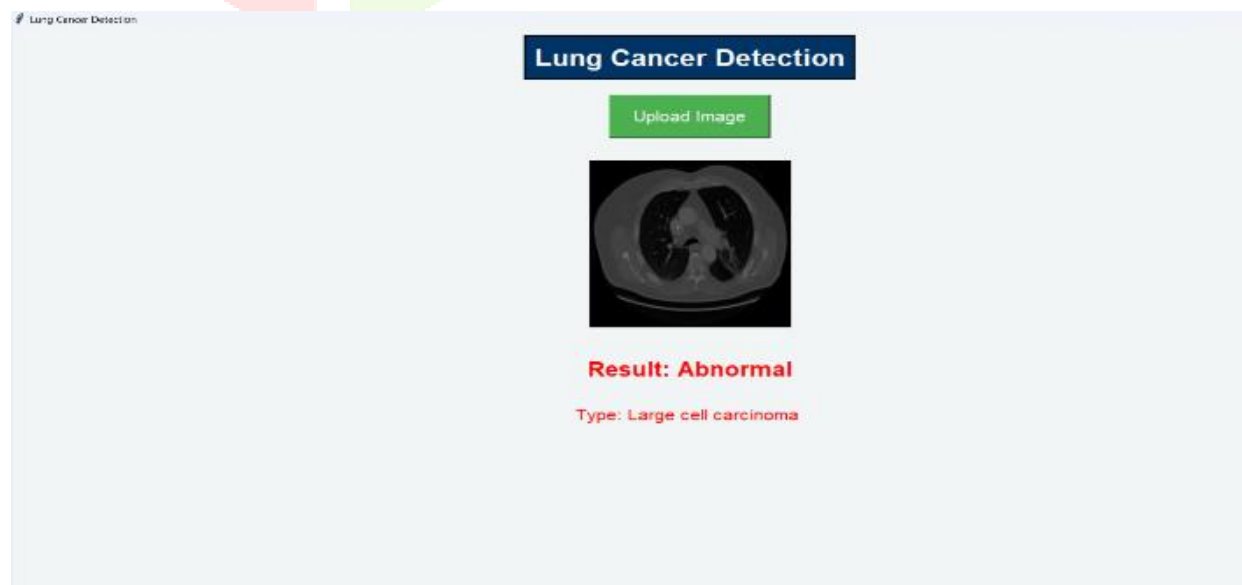


Figure 7: Results of Large Cell Carcinoma Image

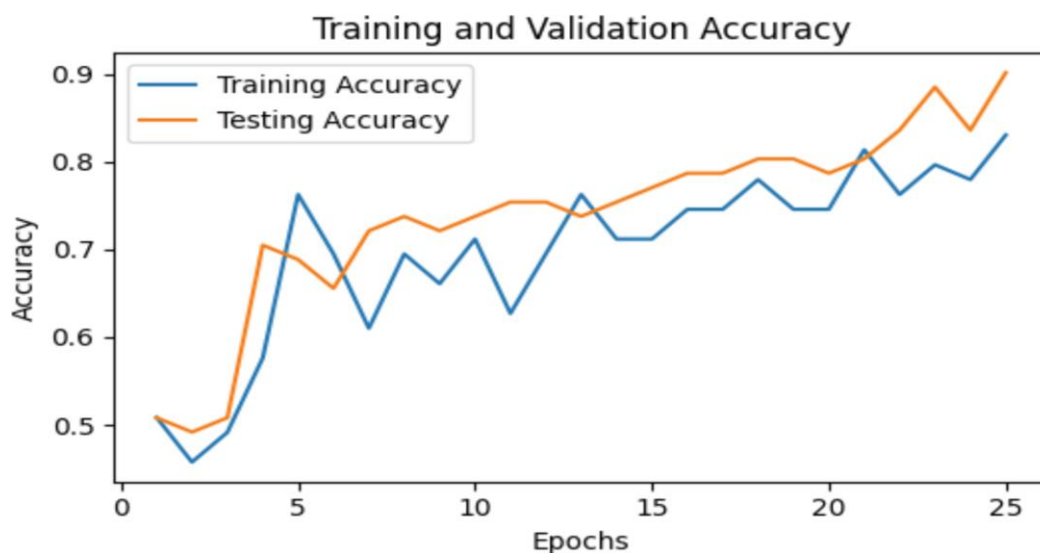


Figure 8: Training and Testing Accuracy curves of the proposed model

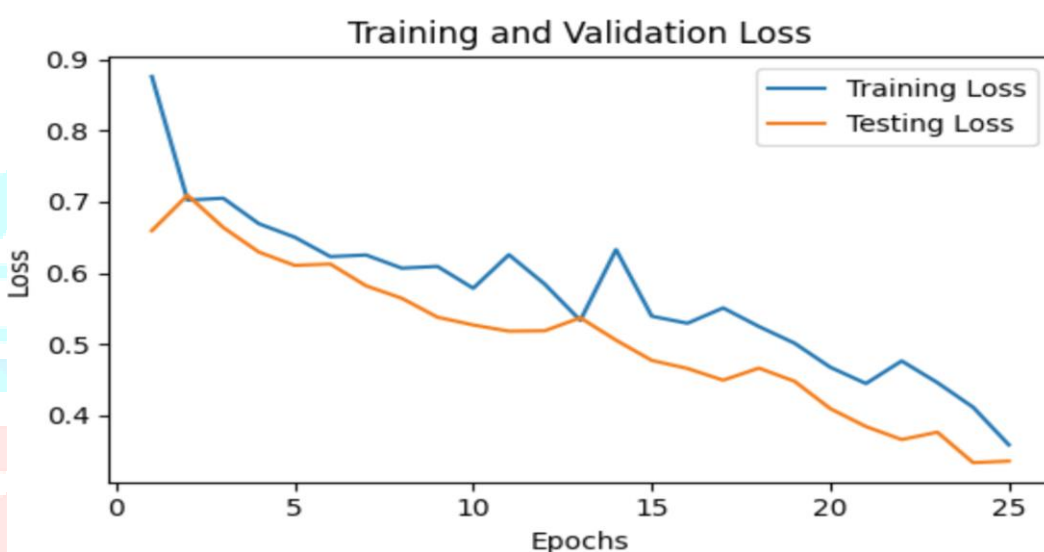


Figure 9: Training and Testing Loss curves of the proposed model

V. CONCLUSION

CT imaging for the identification of lung cancers has grown in popularity as a result of the increasing requirement to guarantee realistic and correct evaluation of vast amounts of health-related data. Lung tumor identification by hand is time-consuming, difficult, and dependent on the expertise of medical specialists. To identify abnormalities in CT images, an automated diagnostic technique will be required. The recommended method showed the best performance in identifying lung cancers with a validation accuracy of 99%.

We'll look at more powerful and important deep CNN models in the future to classify and divide lung cancers with less time complexity. We will also expand the dataset utilized for this study to include more CT scans in order to increase the accuracy of the suggested model. To create the foundation for future studies, the suggested methodology will also be expanded to include other medical images, such as ultrasound, MRI (magnetic resonance imaging), as well as X-rays.

REFERENCES

- [1] Hosseinzadeh, Ali & Shahin, Mohammad & Maghanaki, Mazdak & Mehrzadi, Hamed & Chen, F. (2024). Minimizing waste via novel fuzzy hybrid stacked ensemble of vision transformers and CNNs to detect defects in metal surfaces. *The International Journal of Advanced Manufacturing Technology*. 135. 5115-5140. 10.1007/s00170-024-14741.
- [2] Lung Cancer: Types, Stages, Symptoms, Diagnosis and Treatment. (2022). <https://my.clevelandclinic.org/health/diseases/4375-lung-cancer>
- [3] American Society of Clinical Oncology. (2022). [Online]. Available: <https://my.clevelandclinic.org/health/diseases/6202-small-cell-lung-cancer>
- [4] Non-Small Cell Lung Cancer – Wikipedia. October (2023). https://en.wikipedia.org/wiki/Non-small-cell_lung_cancer
- American Cancer Society. Treatment Choices for Non-Small Cell Lung Cancer, by Stage. (2022). [Online]. Available: <https://my.clevelandclinic.org/health/diseases/6203-non-small-cell-lung-cancer>.
- [5] Bherje, Akanksha and Jidge, Avantika and Roy, Chandan and Hulke, Abhishek and Aswathy, M. A and Yadav, Vipin and Veenamol, K. V. (2024). "Design of Deep Learning-based Approach to Predict Lung Cancer on CT Scan Images," *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Kottayam, India, 2024, pp. 1-5, doi: 10.1109/ICITIIT61487.2024.10580370.
- [6] N. K. Karthikeyan, S. S. Ali and V. S. R, "Lung Cancer Classification Using CT Scan Images Through Deep Learning And CNN Based Model," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 01-05, doi: 10.1109/ADICS58448.2024.10533528.
- [7] Bherje, Akanksha & Jidge, Avantika & Roy, Chandan & Hulke, Abhishek & Aswathy, M. & Yadav, Vipin & Veenamol, K., (2024). Design of Deep Learning-based Approach to Predict Lung Cancer on CT Scan Images. 1-5. 10.1109/ICITIIT61487.2024.10580370.
- [8] Chandraiah, Kavitha & Bhoganna, Naveen. (2024). An optimal model for detection of lung cancer using convolutional neural network. *Indonesian Journal of Electrical Engineering and Computer Science*. 34. 134. 10.11591/ijeecs.v34.i1.pp134-143.
- Nasser, Ibrahim M. and Abu-Naser, Samy S., Lung Cancer Detection Using Artificial Neural Network (March 2019). *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3), 17-23, March 2019, Available at SSRN: <https://ssrn.com/abstract=3369062>
- [9] Neal Joshua ES, Bhattacharyya D, Chakkravarthy M, Byun YC. 3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation. *J Healthc Eng*. 2021 Mar 11;2021:6695518. doi: 10.1155/2021/6695518. PMID: 33777347; PMCID: PMC7979307
- [10] Inceptionv3 Model Image. [Online]. Available: <https://paperswithcode.com/method/inception-v3>
- [11] Available online: <https://www.activestate.com/resources/quick-reads/what-is-tkinter-used-for-and-how-to-install-it>
- [12] Available online: <https://www.oreilly.com/library/view/python-pocket-reference>