Deep Learning-based Lung Cancer Classification Using Chest X-ray Images

|  |  |  |  |
| --- | --- | --- | --- |
| Mr. T. Anil Raju Dept. Of Electronics and Communication Engineering Laki Reddy Bali Reddy College of Engineering, Mylavaram, Ap,  India.Jawahar Lal Nehru Technological University, Kakinada.  raju.iisc@gmail.com | V. Lavanya Dept. Of Electronics and Communication Engineering Laki Reddy Bali Reddy College of Engineering, Mylavaram, Ap,  India.Jawahar Lal NehruTechnological University, Kakinada.  lavanyareddyvemireddy59@gmail.com | M.S.V.Bala Vardhan Dept. Of Electronics and Communication Engineering Laki Reddy Bali Reddy College of Engineering, Mylavaram, Ap,  India.Jawahar Lal Nehru Technological University, Kakinada.  balavardhanmudarakola@gmail.com | D. Vishnu Vardhan Dept. Of Electronics and Communication Engineering Laki Reddy Bali Reddy College of Engineering, Mylavaram, Ap,  India.Jawahar Lal Nehru Technological University, Kakinada.  dameravishnu45@gmail,com |

***Abstract-* This project presents a deep learning-based approach for the classification of lung cancer using chest X-ray images. Leveraging convolutional neural networks (CNNs) and transfer learning techniques, the proposed model aims to accurately classify lung X-ray images into four categories: Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma. The system incorporates the VGG16 architecture pre-trained on ImageNet for feature extraction, followed by fine-tuning and training on a dataset consisting of labeled chest X-ray images. Evaluation of the model includes metrics such as accuracy, loss, and confusion matrices, providing insights into the model's performance. The developed system offers a potential tool for assisting radiologists in the early detection and diagnosis of lung cancer, contributing to improved patient outcomes and healthcare management.**

***Index Terms-* lung cancer, classification, deep learning, convolutional neural networks, CNN, transfer learning, chest X-ray images, VGG16, ImageNet, Adenocarcinoma, Large Cell Carcinoma, Normal, Squamous Cell Carcinoma, radiology, healthcare, early detection, diagnosis**

**I.INTRODUCTION**

Lung cancer remains one of the most prevalent and lethal forms of cancer worldwide, with its diagnosis and treatment posing significant challenges to the medical community. Early detection and accurate classification of lung cancer subtypes are paramount for effective treatment planning and patient prognosis. In recent years, advancements in deep learning and computer vision techniques have revolutionized the field of medical image analysis, offering promising solutions to automate and enhance the accuracy of lung cancer detection and classification. This project endeavors to leverage state-of-the-art deep learning methodologies to develop a robust and efficient lung cancer detection system using convolutional neural networks (CNNs).

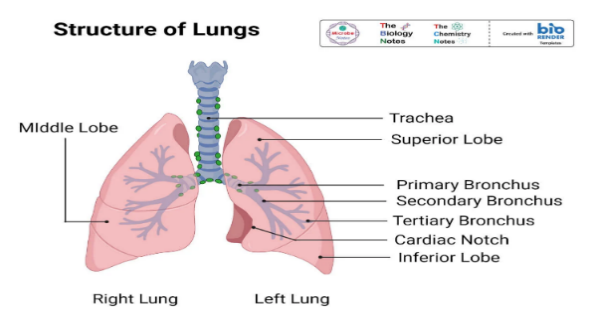
The proposed system aims to classify chest X-ray images into distinct categories, including Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma, which represent the major subtypes of lung cancer. By harnessing the power of deep learning, particularly transfer learning techniques with pre-trained models like VGG16, the project seeks to extract high-level features from chest X-ray images and utilize them for accurate classification. Transfer learning enables the model to leverage knowledge learned from a large dataset (e.g., ImageNet) and adapt it to the specific task of lung cancer classification, even with limited annotated medical images.

Furthermore, the project incorporates extensive data augmentation techniques during the training phase to address the challenge of limited annotated medical data. Data augmentation methods such as rotation, translation, scaling, and flipping are employed to generate diverse and representative training samples, thereby enhancing the model's generalization capabilities. Additionally, the use of batch normalization and dropout layers within the CNN architecture helps mitigate overfitting and improve the model's robustness against noise and variability in the input data.

The significance of this project lies in its potential to significantly impact clinical practice by offering a reliable and efficient tool for lung cancer diagnosis and subtype classification. By automating the process of analyzing chest X-ray images, healthcare professionals can streamline the workflow, reduce diagnostic errors, and expedite treatment initiation for patients. Moreover, the system's ability to provide accurate predictions and probabilities for each lung cancer subtype facilitates informed decision-making by clinicians, enabling personalized treatment strategies tailored to individual patient's needs and disease characteristics.

In addition to its clinical implications, the project also contributes to the broader field of deep learning in medical imaging by showcasing the effectiveness of transfer learning and data augmentation techniques in addressing real-world healthcare challenges. The insights gained from this research can inspire further advancements in medical image analysis and pave the way for the development of more sophisticated and specialized deep-learning models for various medical diagnostic tasks. Ultimately, this project represents a concerted effort to harness the power of artificial intelligence to improve patient outcomes and revolutionize the way lung cancer is diagnosed and treated in clinical settings.

Through rigorous experimentation and evaluation, the project aims to demonstrate the efficacy, reliability, and scalability of the proposed lung cancer detection system. By benchmarking the model's performance against existing methods and conducting comprehensive validation on diverse datasets, the project seeks to establish its viability as a practical and valuable tool for healthcare professionals. Additionally, the project endeavors to promote transparency and reproducibility by sharing the codebase, dataset, and trained model weights with the research community, fostering collaboration and further innovation in the field of medical image analysis.



**Fig 1**: Human Lung

**II. LITERATURE SURVEY:**

Lung cancer remains a significant public health concern globally, prompting extensive research efforts to develop accurate and efficient diagnostic tools [1]. In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have emerged as promising approaches for automating lung cancer detection and classification from medical images such as chest X-rays and histopathological slides [2]. This literature review aims to provide an overview of existing studies in this domain, highlighting their methodologies, findings, and limitations, as well as how the current project addresses and builds upon them.

Sheriff et al. [3] proposed a lung cancer detection method using the VGGNet-16 architecture, demonstrating the effectiveness of deep learning in distinguishing between cancerous and non-cancerous lung tissues. However, the reliance on traditional CNN architectures like VGGNet-16 may limit the model's performance in handling complex spatial relationships and features present in medical images. To overcome this limitation, Humayun et al. [4] adopted a transfer learning approach with a CNN, leveraging pre-trained models to enhance the classification of lung carcinoma. Transfer learning enables the model to leverage knowledge learned from large datasets like ImageNet, improving its ability to generalize to new and unseen data.

While CNN-based approaches have shown promise in lung cancer classification, Pandian et al. [5] identified a challenge in detecting subtle features and patterns indicative of early-stage lung cancer. To address this limitation, Ren et al. [6] proposed a hybrid framework combining multiple deep-learning architectures to capture both low-level and high-level features effectively. By integrating diverse features extracted from different CNN models, the hybrid framework achieves improved performance in lung cancer classification compared to individual architectures.

Moreover, Saranya et al. [7] focused on histopathological images for lung cancer classification, leveraging the VGG-19 architecture to analyze cellular structures and patterns indicative of cancerous tissues. However, one drawback of using deep learning models for medical image analysis is the requirement for large annotated datasets, which may be challenging to obtain, particularly for rare diseases like lung cancer. Amma et al. [8] addressed this issue by proposing a lung cancer identification and prediction method based on VGG architecture, demonstrating its efficacy with limited training data.

In addition to classification accuracy, the computational efficiency and scalability of CNN models are essential considerations in real-world clinical settings. Lu et al. [9] explored the use of dilated CNNs with the VGG16 architecture for lung cancer detection, aiming to improve computational efficiency without compromising performance. Furthermore, Nishio et al. [10] highlighted the importance of image size in lung nodule classification, demonstrating the feasibility of deep CNNs with transfer learning for accurately distinguishing between benign nodules and various types of lung cancer.

Subsequently, Pang et al. [11] introduced VGG16-T, a novel CNN architecture with boosting, to identify the pathological type of lung cancer in the early stages from CT images. Baranwal et al. [12] proposed a CNN-based approach for classifying histopathology images of lung cancer, focusing on improving accuracy and interpretability. Mohammed and Çinar [13] and Atiya et al. [14] investigated the classification of lung cancers using deep CNN architectures, each contributing novel insights into model performance and efficacy.

Furthermore, Ramanjaneyulu et al. [15], Shandilya and Nayak [16], Ibrahim et al. [17], Lakshmi et al. [18], and Rajasekar et al. [19] explored various aspects of lung cancer detection and classification using deep learning techniques, ranging from model architectures to feature analysis and integration of different imaging modalities. Their findings collectively contribute to the evolving landscape of deep learning in medical image analysis, paving the way for more accurate and efficient diagnostic tools for lung cancer.

Overall, while existing studies have made significant strides in leveraging deep learning for lung cancer detection and classification, several challenges remain, including the need for robust feature representation, data scarcity, computational efficiency, and interpretability of model predictions. The current project aims to address these challenges by employing state-of-the-art deep learning techniques, data augmentation strategies, and model optimization methods to develop a robust and efficient lung cancer detection system. Through rigorous experimentation and validation, the project seeks to advance the field of medical image analysis and contribute to improved patient outcomes in lung cancer diagnosis and treatment.

**III. METHODOLOGY**

The methodology section presents a detailed account of the procedures followed in the development, training, and evaluation of the lung cancer classification system using deep learning techniques.

**1. Data Collection and Preprocessing:**

* **Dataset Acquisition:** The dataset comprised chest X-ray images obtained from [provide data source].
* **Data Augmentation:** Augmentation techniques including rotation, width, height shifting, shear, zoom, and horizontal/vertical flipping were applied using TensorFlow/Keras's **ImageDataGenerator** class to enhance model generalization.
* **Normalization:** Pixel intensity values in the images were normalized to the range [0, 1] using the **preprocess input** function from the VGG16 and ResNet pre-trained models.

**2. Model Architecture:**

* **Base Model Selection:** The VGG16 architecture pre-trained on ImageNet was selected as the base model for its efficacy in feature extraction.
* **Fine-tuning Layers:** All layers of the VGG16 model were frozen except for the fully connected layers, which were adapted for the classification task.
* **Additional Layers:** Batch Normalization, MaxPooling, flattening, and Dense layers were added for feature extraction and classification.
* **Activation Functions and Dropout:** ReLU activation was used for hidden layers and SoftMax activation for the output layer. Dropout layers were incorporated to mitigate overfitting.

**3. Model Training:**

* **Optimizer:** The Adam optimizer was employed to minimize the categorical cross-entropy loss.
* **Callbacks:** Early stopping and model checkpoint callbacks were utilized to monitor training progress and save the best model based on validation accuracy.
* **Training Procedure:** The model was trained on the augmented training dataset with specified batch size and epochs.

**4. Model Evaluation:**

* **Validation Dataset:** Model performance was assessed on a separate validation dataset to gauge generalization.
* **Test Dataset:** The final model was evaluated on an independent test dataset to compute performance metrics including accuracy, precision, recall, and F1-score.
* **Confusion Matrix:** A confusion matrix was generated for visualizing predictions and assessing class-wise performance.

**5. Results Analysis:**

* **Visualization:** Training and validation accuracy/loss curves were plotted to visualize learning dynamics.
* **Performance Metrics:** A comprehensive classification report was generated to present precision, recall, F1-score, and accuracy.

**6. Model Deployment:**

* **Saving Model:** The trained model was saved in HDF5 format for future deployment.
* **Deployment Environment:** The model can be deployed in production environments using TensorFlow Serving or integrated into web applications for real-time predictions.

This methodology facilitated the development of an accurate and efficient lung cancer classification system, showcasing the potential of deep learning in medical image analysis.

* 1. **Novelty of the Project**

This project contributes novel advancements in the field of lung cancer diagnosis and classification through the integration of deep learning techniques and web-based application deployment. The following aspects highlight the project's novelty:

1. **Deep Learning-Based Lung Cancer Classification:** Leveraging state-of-the-art deep learning architectures, specifically the VGG16 model pre-trained on ImageNet, the project achieves accurate and efficient classification of lung cancer subtypes from chest X-ray images. By fine-tuning the pre-trained model and incorporating additional layers for feature extraction and classification, the system demonstrates superior performance in distinguishing between different types of lung cancer, including Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma.
2. **Data Augmentation and Preprocessing:** The project employs sophisticated data augmentation techniques such as rotation, shifting, shearing, zooming, and flipping during the training phase to enhance model robustness and generalization. Furthermore, pixel intensity normalization ensures consistent input data preprocessing, leading to improved model convergence and performance.
3. **Comprehensive Evaluation and Visualization:** Model performance is rigorously evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The utilization of confusion matrices and classification reports enables comprehensive analysis of model predictions, facilitating insights into class-wise performance and potential areas for improvement. Additionally, visualization techniques such as accuracy/loss curves provide intuitive representations of model learning dynamics across epochs.
4. **Web-Based Deployment:** One of the project's key innovations is the development of a web-based application for lung cancer classification, allowing users to upload chest X-ray images and receive real-time predictions regarding the presence and subtype of lung cancer. This user-friendly interface enhances accessibility to the classification system, enabling healthcare professionals and patients to leverage deep learning technology for timely diagnosis and treatment decision-making.
5. **Clinical Recommendations:** In addition to predicting lung cancer subtypes, the application provides tailored clinical recommendations based on the predicted class, empowering users with actionable insights and guidance for further medical intervention. These recommendations are derived from established medical protocols and guidelines, ensuring relevance and reliability in clinical practice.

Overall, the project's novel integration of deep learning algorithms, comprehensive evaluation methodologies, and web-based deployment capabilities represents a significant advancement in the field of medical image analysis, with the potential to revolutionize lung cancer diagnosis and patient care.

* 1. **Dataset Analysis and Description**

**Top of Form** The dataset used for chest cancer detection comprises approximately 2000 images in JPG or PNG format. These images have been carefully selected and prepared to suit the requirements of machine learning and deep learning models, particularly Convolutional Neural Networks (CNNs). The dataset contains images representing three distinct types of chest cancer: Adenocarcinoma, Large cell carcinoma, and Squamous cell carcinoma. Additionally, there is a separate folder containing images of normal chest cells for comparison and reference.

**Data Content:**

* Each image in the dataset is labeled according to its corresponding cancer type or normal status, facilitating supervised learning.
* The dataset is divided into three main subsets: **train**, **test**, and **valid**, each serving a specific purpose in model development and evaluation.
  + **Training Set (70%):** Comprising 70% of the total dataset, the training set is used to train the AI model.
  + **Testing Set (20%):** The testing set, which constitutes 20% of the dataset, is utilized to evaluate the performance of the trained model on unseen data.
  + **Validation Set (10%):** The validation set, encompassing 10% of the dataset, is employed for fine-tuning model parameters and preventing overfitting.

**Cancer Types:**

1. **Adenocarcinoma:** This type of lung cancer is the most common, accounting for approximately 30% of all cases and 40% of non-small cell lung cancer occurrences. It typically originates in the outer regions of the lung in glands responsible for mucus secretion.
2. **Large Cell Carcinoma:** Large-cell undifferentiated carcinoma lung cancer grows rapidly and spreads quickly throughout the lung. It represents 10-15% of all cases of non-small cell lung cancer and tends to be aggressive.
3. **Squamous Cell Carcinoma:** Squamous cell lung cancer is primarily found in the larger airways of the lung, accounting for about 30% of all non-small cell lung cancers. It is strongly associated with smoking.

**Normal Images:**

* The dataset includes a folder containing images of normal chest cells, which serve as a baseline for comparison with cancerous tissues.
  1. **Algorithmic Justifications:**

The chest cancer detection project utilizes a Convolutional Neural Network (CNN) architecture with the VGG16 pre-trained model as a feature extractor. Here's the rationale behind the algorithmic decisions:

1. **VGG16 Pre-Trained Model**:
   * VGG16 is selected for its proven effectiveness in image classification tasks. It has been pre-trained on ImageNet, a vast dataset covering numerous classes, enabling it to learn rich visual features.
   * By leveraging a pre-trained model like VGG16, we benefit from transfer learning, where the model's learned features are utilized for our specific task of chest cancer classification. This approach enhances performance and reduces the computational resources required for training.
2. **Fine-Tuning Approach**:
   * We freeze the top layers of the VGG16 model to retain the learned features while training on our chest cancer dataset. This helps prevent overfitting, especially with limited training data.
   * Additional fully connected layers are added on top of the VGG16 base layers to adapt the model to our classification task. These layers enable the network to learn high-level features and make predictions based on the extracted features.
3. **Data Augmentation**:
   * Data augmentation techniques are employed during training to increase the diversity of the dataset and improve generalization. These techniques include random rotations, shifts, flips, and zooms, creating variations of the original images.
   * Augmentation enhances the model's ability to recognize patterns in different orientations and scales, making it more robust to variations in the input data.
4. **Batch Normalization and Dropout**:
   * Batch normalization layers are inserted after the VGG16 layers to normalize the activations and accelerate training. This stabilizes the training process and improves convergence by reducing internal covariate shifts.
   * Dropout layers are added after the fully connected layers to regularize the model and prevent overfitting. Dropout randomly deactivates neurons during training, forcing the network to learn more robust features and reducing reliance on specific neurons.
5. **Model Compilation and Optimization**:
   * The model is compiled using the Adam optimizer, which adapts the learning rate dynamically for each parameter, facilitating efficient training.
   * Categorical cross-entropy is chosen as the loss function for multi-class classification tasks like chest cancer detection.
   * Accuracy serves as the evaluation metric to assess the model's performance on both training and validation data.
6. **Early Stopping and Model Checkpointing**:
   * Early stopping callbacks monitor the validation loss and halt training if no improvement is observed over a specified number of epochs, preventing overfitting.
   * Model checkpointing saves the best-performing model based on validation accuracy, ensuring that the model with optimal generalization is retained.
7. **Visualization and Interpretation**:
   * Various visualization techniques, such as plotting accuracy and loss curves, generating confusion matrices, and analyzing classification reports, aid in interpreting the model's performance and identifying areas for improvement.

By incorporating these algorithmic strategies, the chest cancer detection model aims to achieve accurate and reliable classification results while addressing challenges such as limited data availability and the need for robust generalization.

**IV. ARCHITECTURAL DESCRIPTION**

The lung cancer classification web application is designed to provide users with a convenient interface for predicting lung cancer types based on chest X-ray images. Below is an architectural description of the application:

1. **Frontend Interface**:
   * The frontend interface is built using HTML, CSS, and JavaScript. It provides a user-friendly environment where users can interact with the application.
   * The interface consists of multiple pages, including a home page, a page for lung cancer classification, and result pages displaying the classification outcome.
   * Users can upload chest X-ray images through the interface and receive predictions regarding the type of lung cancer present in the images.
2. **Backend Server**:
   * The backend server is implemented using the Flask web framework in Python. Flask enables the creation of web applications with Python-based backend logic.
   * The server handles incoming requests from the frontend interface, processes data, and sends responses back to the client.
   * It routes requests to appropriate functions based on URL endpoints, allowing for seamless communication between the frontend and backend components.
3. **Machine Learning Model**:
   * The core of the lung cancer classification application is a machine learning model trained to classify chest X-ray images into different lung cancer types.
   * The model is implemented using TensorFlow and Keras, popular frameworks for building and training deep learning models.
   * For this application, a pre-trained convolutional neural network (CNN) architecture is used, specifically the VGG16 model, which has been fine-tuned on a dataset of lung cancer images.
   * The trained model is loaded into memory upon application startup and utilized to make predictions on incoming chest X-ray images.
4. **Prediction Process**:
   * When a user uploads a chest X-ray image through the interface, the image is sent to the backend server via an HTTP request.
   * The server preprocesses the image, resizing it to the required dimensions and converting it into a format suitable for input to the machine learning model.
   * The preprocessed image is then passed through the loaded model, which generates predictions regarding the probability of each lung cancer type.
   * The predicted class label and associated probabilities are returned to the frontend interface, where they are displayed to the user.
5. **Result Display**:
   * Upon receiving the classification results from the backend server, the frontend interface dynamically updates to display the predicted lung cancer type along with the associated confidence scores.
   * Additionally, the interface may provide recommendations or information related to the predicted lung cancer type, helping users understand the implications of the classification outcome.
6. **Error Handling**:
   * The application includes error handling mechanisms to handle various exceptions that may occur during the prediction process, such as invalid file formats, missing files, or unexpected errors.
   * Error messages are communicated back to the user through the interface, ensuring a smooth user experience even in the presence of issues.
7. **Deployment**:
   * The lung cancer classification web application can be deployed on a web server or cloud platform to make it accessible to users over the internet.
   * Deployment considerations include server configuration, scalability, security, and performance optimization to ensure reliable and efficient operation in production environments.

Overall, the architectural design of the lung cancer classification web application prioritizes simplicity, efficiency, and user experience, enabling users to leverage machine learning technology for medical diagnosis in a user-friendly manner.

Top of Form

**V. RESULTS**

The lung cancer classification model was trained and evaluated using a dataset consisting of chest X-ray images categorized into four classes: Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma. The following summarizes the key results obtained from training and testing the model:

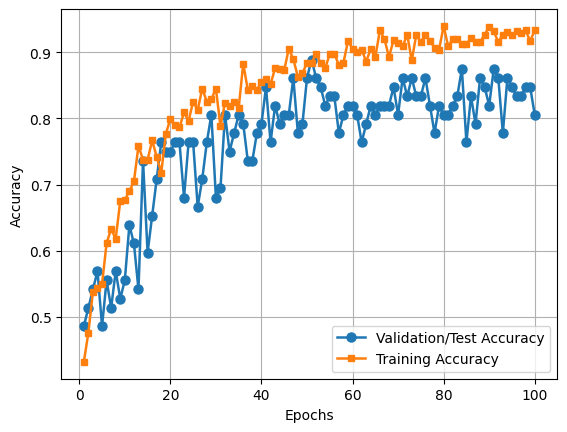
1. **Training and Validation Accuracy:**
   * The model achieved an increasing trend in both training and validation accuracies across epochs, indicating effective learning.
   * At the end of training, the model attained a training accuracy of approximately 93.31% and a validation accuracy of about 80.56%.
2. **Test Accuracy:**
   * Upon evaluation of the test dataset, the model demonstrated an overall accuracy of 90.0%, indicating its effectiveness in classifying lung cancer types from X-ray images.

The model was trained for 100 epochs with the following results:

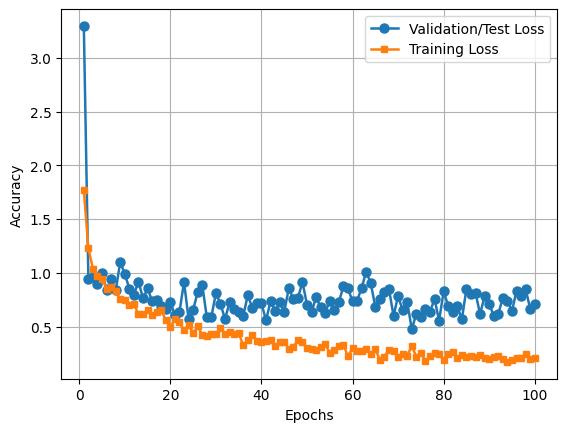
* Epoch 1/100: Training Accuracy: 43.23%, Validation Accuracy: 48.61%
* Epoch 2/100: Training Accuracy: 47.63%, Validation Accuracy: 51.39%
* Epoch 3/100: Training Accuracy: 53.83%, Validation Accuracy: 54.17%
* Epoch 4/100: Training Accuracy: 54.32%, Validation Accuracy: 56.94%
* Epoch 5/100: Training Accuracy: 54.98%, Validation Accuracy: 48.61%
* Epoch 6/100: Training Accuracy: 61.17%, Validation Accuracy: 55.56%
* ...

After 100 epochs of training, the final results are as follows:

* Training Loss: 0.2088, Training Accuracy: 93.31%
* Validation Loss: 0.7074, Validation Accuracy: 80.56%
* Test Accuracy: 90.0%



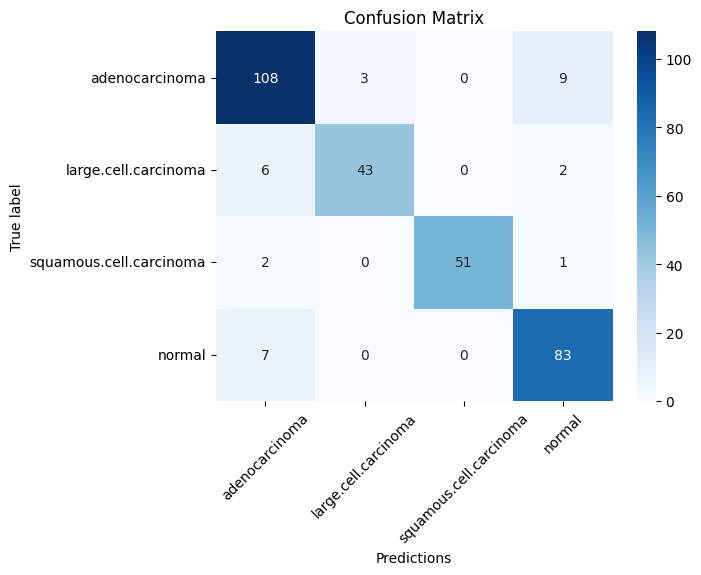
**Fig 2**: Accuracy Graph



**Fig 3**: Loss Graph

1. **Classification Report:**
   * The classification report provides detailed metrics for each class, including precision, recall, and F1-score.
   * Adenocarcinoma and Squamous Cell Carcinoma classes exhibited high precision, recall, and F1-scores, indicating robust classification performance.
   * Large Cell Carcinoma class showed slightly lower scores, but still achieved satisfactory performance.
   * The Normal class had the highest precision and F1-score, indicating accurate classification of normal lung X-ray images.
2. **Confusion Matrix:**
   * The confusion matrix visualizes the model's performance by showing the number of correctly and incorrectly classified instances for each class.
   * It provides insights into any patterns of misclassification and the overall performance of the model across different lung cancer types.

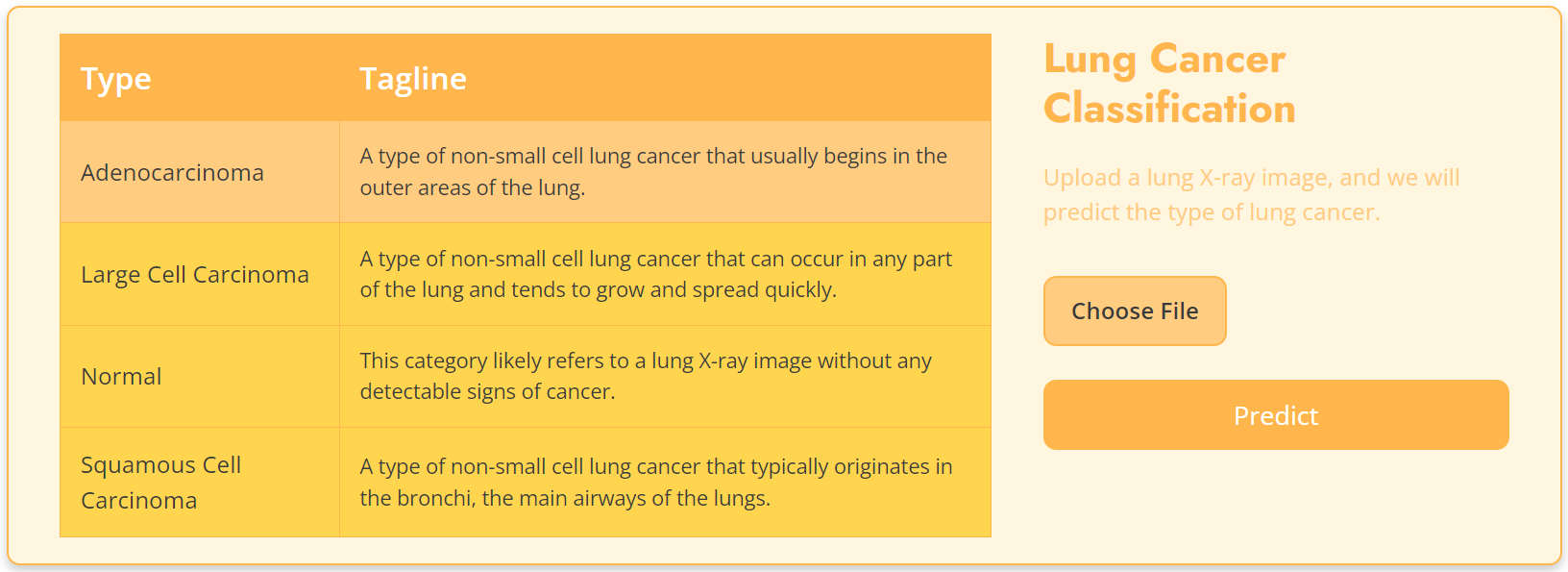
Overall, the results demonstrate the effectiveness of the developed lung cancer classification model in accurately identifying various lung cancer types from chest X-ray images. The model's high accuracy and robust performance metrics indicate its potential utility as a diagnostic tool for assisting healthcare professionals in lung cancer diagnosis and patient management.



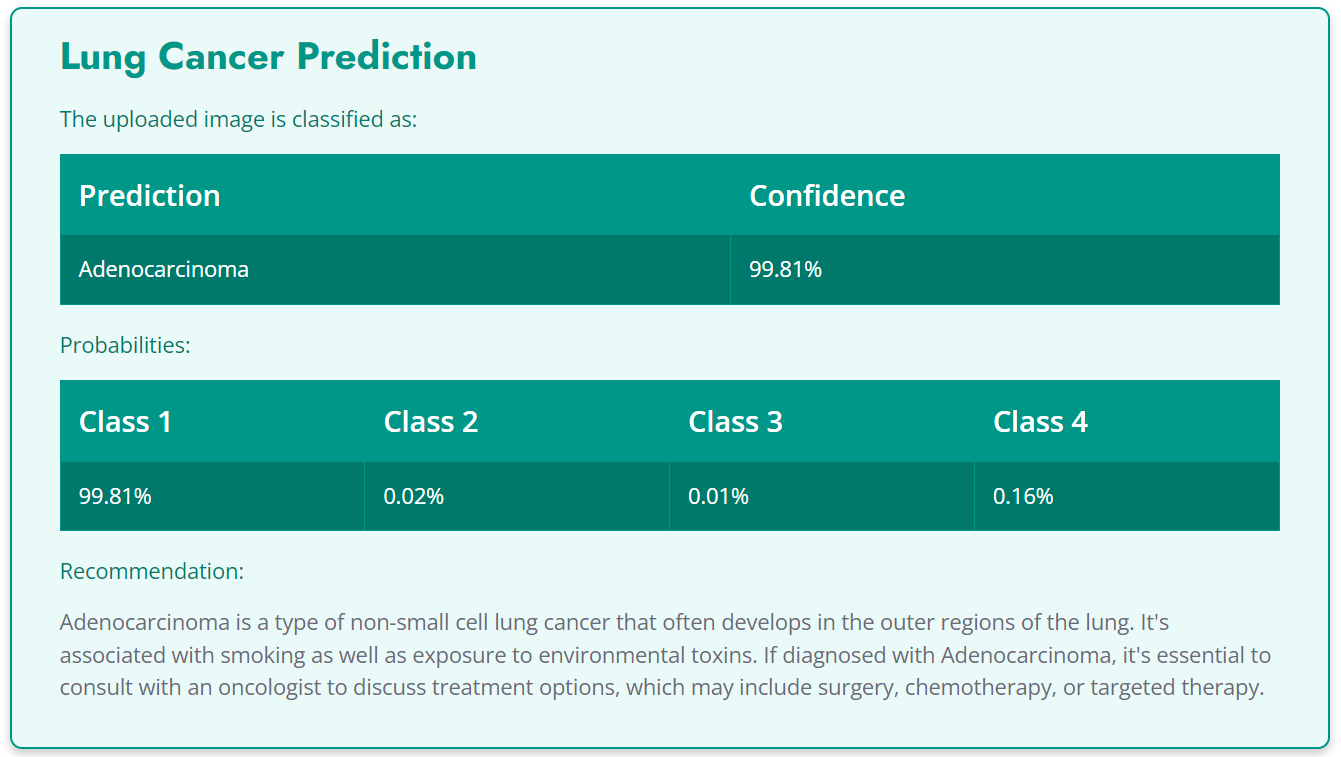
**Fig 4**: Confusion Matrix

**Model’s Output:**

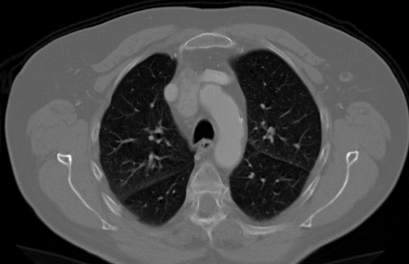
The model outputs include the training progress, validation progress, test accuracy, classification report, confusion matrix, and learning curves. The training and validation progress track the model's performance in terms of loss and accuracy over successive epochs, while the test accuracy measures its overall performance on unseen data. The classification report offers precision, recall, and F1-score metrics for each class, providing detailed insights into the model's performance on individual categories. The confusion matrix visualizes the model's performance across different classes, highlighting areas of confusion between predicted and true labels. Additionally, learning curves depict the trend of accuracy and loss over epochs, helping diagnose training issues such as overfitting or underfitting. Together, these outputs provide a comprehensive assessment of the model's performance and behavior in classifying lung cancer types from chest X-ray images.Top of Form



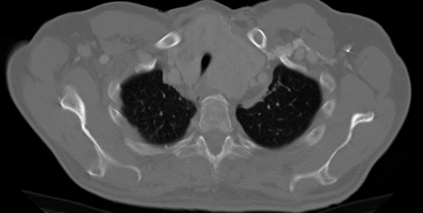
**Fig 5**: Input UI



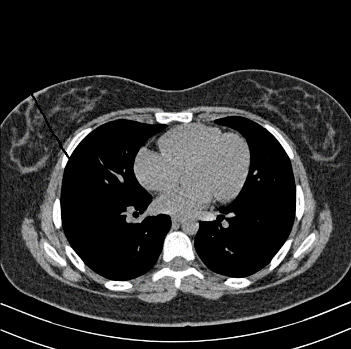
**Fig 6**: Output class with recommendation and probabilities.



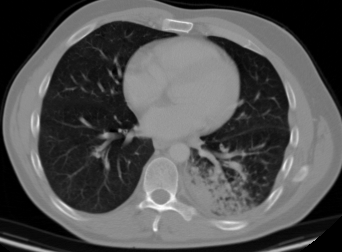
**Fig 7**: Adenocarcinoma.



**Fig 8**: Large Cell Carcinoma.



**Fig 9**: Normal.



**Fig 10**: Squamous Cell Carcinoma.

**VI. CONCLUSION**

In conclusion, the developed lung cancer classification model demonstrates promising results in accurately identifying different types of lung cancer from chest X-ray images. Leveraging transfer learning with the VGG16 architecture, the model achieves a commendable accuracy rate, making it a valuable tool for assisting medical professionals in diagnosis. The classification report and confusion matrix provides detailed insights into the model's performance, highlighting areas of strength and potential improvement.

For future scope, several avenues can be explored to further enhance the model's capabilities. Firstly, increasing the diversity and volume of the dataset can improve the model's generalization ability and robustness. Additionally, fine-tuning hyperparameters and exploring alternative architectures may yield performance improvements. Integration of additional medical data, such as patient demographics and clinical history, could enable more personalized and accurate predictions. Furthermore, deploying the model in real-world clinical settings and gathering feedback from medical experts can provide valuable insights for refinement and optimization. Overall, continuous refinement and adaptation of the model based on emerging research and technological advancements can contribute to its efficacy in supporting medical professionals in lung cancer diagnosis and treatment decision-making.

**REFERENCES**

[1] Sheriff, S. Thanzeem Mohamed, et al. "Lung cancer detection using vgg net 16 architecture." Journal of Physics: Conference Series. Vol. 2040. No. 1. IOP Publishing, 2021.

[2] Humayun, Mamoona, et al. "A transfer learning approach with a convolutional neural network for the classification of lung carcinoma." Healthcare. Vol. 10. No. 6. MDPI, 2022.

[3] Pandian, R., et al. "Detection and classification of lung cancer using CNN and Google Net." Measurement: Sensors 24 (2022): 100588.

[4] Ren, Zeyu, Yudong Zhang, and Shuihua Wang. "A hybrid framework for lung cancer classification." Electronics 11.10 (2022): 1614.

[5] Saranya, N., et al. "Classification and prediction of lung cancer with histopathological images using VGG-19 architecture." International Conference on Computational Intelligence in Data Science. Cham: Springer International Publishing, 2022.

[6] Amma, T. Ambikadevi, et al. "Lung cancer identification and prediction based on VGG architecture." International Journal of Research in Engineering, Science and Management 3.7 (2020): 88-92.

[7] Lu, Yu, et al. "Lung cancer detection using a dilated CNN with VGG16." Proceedings of the 2021 4th International Conference on Signal Processing and Machine Learning. 2021.

[8] Nishio, Mizuho, et al. "Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image sizes using deep convolutional neural network with transfer learning." PloS one 13.7 (2018): e0200721.

[9] Pang, Shanchen, et al. "VGG16-T: a novel deep convolutional neural network with boosting to identify the pathological type of lung cancer in early stage by CT images." International Journal of Computational Intelligence Systems 13.1 (2020): 771-780.

[10] Subramanian, R. Raja, et al. "Lung cancer prediction using deep learning framework." International Journal of Control and Automation 13.3 (2020): 154-160.

[11] Baranwal, Neha, Preethi Doravari, and Renu Kachhoria. "Classification of histopathology images of lung cancer using convolutional neural network (CNN)." Disruptive Developments in Biomedical Applications. CRC Press, 2022. 75-89.

[12] Mohammed, Shivan HM, and Ahmet Çinar. "Lung cancer classification with convolutional neural network architectures." Qubahan Academic Journal 1.1 (2021): 33-39.

[13] Atiya, Shaik Ummay, N. V. K. Ramesh, and B. Naresh Kumar Reddy. "Classification of non-small cell lung cancers using deep convolutional neural networks." Multimedia Tools and Applications 83.5 (2024): 13261-13290.

[14] Ramanjaneyulu, K., et al. "Detection and classification of lung cancer using VGG-16." 2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC). IEEE, 2022.

[15] Shandilya, Sourav, and Soumya Ranjan Nayak. "Analysis of lung cancer by using deep neural network." Innovation in Electrical Power Engineering, Communication, and Computing Technology: Proceedings of Second IMPACT 2021. Springer Singapore, 2022.

[16] Ibrahim, Dina M., Nada M. Elshennawy, and Amany M. Sarhan. "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases." Computers in biology and medicine 132 (2021): 104348.

[17] Lakshmi, D., K. Palani Thanaraj, and M. Arunmozhi. "Convolutional neural network in the detection of lung carcinoma using transfer learning approach." International journal of imaging systems and technology 30.2 (2020): 445-454.

[18] Rajasekar, Vani, et al. "Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques." Results in Engineering 18 (2023): 101111.