DAYANANDA SAGAR UNIVERSITY

Harohalli, Kanakapura Road,Ramanagara Dt, Bengaluru-562112

SCHOOL OF ENGINEERING

Bachelor of Technology

in

COMPUTER SCIENCE AND ENGINEERING (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

A Project Report On

Classification of Lung Cancer using Vision Transformer

By

RUPAM SINGH - ENG21AM0101

AMAR JYOTI PANDA - ENG21AM0006

NARAYAYN KULSHRESTHA - ENG21AM0078

AYUSH AGRAWAL - ENG21AM0015

Under the supervision of

Prof.Dr. Mude Nagarjuna Naik

Assistant Professor

Computer Science & Engineering (AI & ML)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

SCHOOL OF ENGINEERING

DAYANANDA SAGAR UNIVERSITY

(2024 – 2025)

DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Department of Computer Science & Engineering
(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

Harohalli, Kanakapura Road,Ramanagara Dt, Bengaluru-562112
Karnataka, India

CERTIFICATE

This is to certify that the project entitled **" Classification of Lung Cancer using Vision Transformer"** is carried out by **RUPAM SINGH (ENG21AM0101), AMAR JYOTI PANDA (ENG21AM0006), NARAYAYN KULSHRESTHA (ENG21AM0078), AYUSH AGRAWAL (ENG21AM0015)**, bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore, in partial fulfillment for the award of a degree in Bachelor of Technology in Computer Science and Engineering, during the year **2024 - 2025**.

Dr. Mude Nagarjuna Naik	**Dr. Vinutha N**	**Dr. Jayavrinda Vrindavanam**
Assistant Professor	Project Co-ordinator	Professor & Chairperson
Dept. of CSE (AIML)	Dept. of CSE (AIML)	Dept. of CSE (AIML)
School of Engineering	School of Engineering	School of Engineering
Dayananda Sagar University	Dayananda Sagar University	Dayananda Sagar University
Signature	Signature	Signature .....

Name of the Examiners:	Signature with date:
1.....	.....
2.....	.....
3.....	.....

DECLARATION

We, **RUPAM SINGH (ENG21AM0101), AMAR JYOTI PANDA (ENG21AM0006), NARAYAYN KULSHRESTHA (ENG21AM0078), AYUSH AGRAWAL (ENG21AM0015)**, are students of the eight semester B.Tech in Computer Science and Engineering (AI & ML) at the School of Engineering, Dayananda Sagar University. We hereby declare that the Major Project titled **"Classification of Lung Cancer using Vision Transformer"** has been carried out by us and submitted in partial fulfillment for the award of a degree in **Bachelor of Technology** in **Computer Science and Engineering (AI & ML)** during the academic year **2024–2025**.

Student: **Signature**

Name 1: RUPAM SINGH
USN: ENG21AM0101

Name 2: AMAR JYOTI PANDA
USN: ENG21AM0006

Name 3: NARAYAN KULSHRESTHA
USN: ENG21AM0078

Name 3: AYUSH AGRAWAL
USN: ENG21AM0015

Place: Bangalore
Date:

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work. First, we take this opportunity to express our sincere gratitude to School of Engineering& Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank **Dr. Udaya Kumar Reddy K R, Dean, School of Engineering & Technology, Dayananda Sagar University** for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to **Dr. Jayavrinda Vrindavanam, Department Chairman, Computer Science and Engineering (Artificial Intelligence and Machine Learning), Dayananda Sagar University**, for providing right academic guidance that made our task possible.

We would like to thank our guide **Dr. Mude Nagarjuna Naik, Assistant Professor, Dept. of Computer Science and Engineering of Artificial Intelligence and Machine Learning Dayananda Sagar University**, for sparing his valuable time to extend help in every step of our project work, which paved the way for smoothprogress and fruitful culmination of the project.

We would like to thank our **Project Coordinator Dr. Vinutha N** as well as all the staff members of Computer Science and Engineering (AIML) for their support. We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Project work

Contents

1 INTRODUCTION **1**

1.1 Comparisons with Previous Studies: 3

2 PROBLEM DEFINITION AND OBJECTIVE **4**

2.1 Problem Definition: 4

2.2 Novelty of Proposed Approach 6

3 LITERATURE SURVEY **7**

4 METHODOLOGY **10**

4.1 Dataset Collection 10

4.2 Dataset Selection for Model Building: 11

4.3 Architecture: 12

5 REQUIREMENTS **15**

5.1 Hardware Requirements: 15

5.2 Software Requirements: 16

6 EXPERIMENTATION **17**

6.1 Model Selection 19

6.2 Working of Vision Transformer (ViT) in Lung Cancer Classification 23

6.2.1 Addressing Challenges in Lung Cancer Image Analysis 25

7 RESULT AND ANALYSIS **26**

8 CONCLUSION AND FUTURE SCOPE **29**

8.1 Implications and Future Directions: 30

8.2 Future Scope 31

LIST OF FIGURES

Fig .Number	Figure Description	Page Number
Fig 4.1	Design Architecture	13
Fig 4.2	Transformer Architecture	14
Fig 6.1	Working of vision transformer	24
Fig 7.1	Graph Displaying Accuracy of Different Deep Learning Models	26
Fig 7.2	User interface for lung cancer classification	27
Fig 7.3	Classification of different types of lung cancer	28
Fig 7.4	Prescription provided by the model	28

LIST OF TABLES

Table Number	Table Description	Page Number
Table 7.1	Accuracy of different deep learning models	32
Table 7.2	Accuracy of different Deep learning models	33

ABSTRACT

Lung cancer remains a leading cause of cancer-related mortality worldwide, emphasizing the need for accurate and early diagnostic techniques. Traditional diagnostic methods, while effective, often involve invasive procedures, extended processing times, and are prone to human error. This study explores a novel, non-invasive approach to lung cancer classification by utilizing the Vision Transformer (ViT), a modern deep learning model based on the self-attention mechanism. Unlike conventional convolutional neural networks (CNNs), which focus on local feature extraction, the ViT processes medical images as sequences of patches, enabling the model to capture both local and global contextual information.

The proposed framework incorporates pre-processing steps, including noise reduction, contrast enhancement, and geometric corrections, to improve image quality and support effective model performance. Experimental results demonstrate that the ViT surpasses traditional CNN-based approaches in classification accuracy, sensitivity, and specificity. Furthermore, the self-attention mechanism enhances model interpretability by identifying and highlighting critical regions within lung images relevant to diagnosis. This research highlights the potential of transformer-based architectures in medical imaging and underscores their role in advancing AI-driven diagnostic systems for lung cancer, ultimately contributing to more efficient and reliable healthcare solutions.

Chapter 1

INTRODUCTION

Lung cancer remains a critical global health issue, responsible for a significant proportion of cancer-related deaths. Timely and accurate diagnosis is essential for improving patient prognosis and enabling effective treatment planning. However, traditional diagnostic methods such as biopsies, CT scans, and radiological assessments often come with limitations including invasiveness, time consumption, and susceptibility to human error. These challenges have led to the adoption of deep learning technologies in medical diagnostics, aiming to provide faster, more accurate, and non-invasive alternatives. Lung cancer is generally classified into the following main types:

- Non-Small Cell Lung Cancer (NSCLC):The most common type, accounting for about 85% of cases. It includes subtypes such as:
- Adenocarcinoma: Often found in outer areas of the lungs; common among both smokers and non-smokers.
- Squamous Cell Carcinoma: Usually located near the central bronchial tubes; often linked to smoking.
- Large Cell Carcinoma: A less common, aggressive form that can appear in any part of the lung.

Among the recent advancements in deep learning, the Vision Transformer (ViT) has emerged as a powerful model for image classification tasks. Unlike conventional convolutional neural networks (CNNs) that primarily extract local features through hierarchical filtering, the ViT processes images by dividing them into fixed-size patches and treating them as sequences, similar to tokens in natural language processing.

This design allows the model to capture intricate and long-range dependencies within an image using self-attention mechanisms. As a result, ViT can identify subtle patterns and variations in lung tissues that may be indicative of cancer, offering improved diagnostic capabilities over traditional CNN-based methods.

This study introduces a Vision Transformer-based approach for lung cancer classification that integrates advanced pre-processing techniques such as noise reduction, contrast enhancement, and spatial alignment to improve input image quality. The ViT's self-attention mechanism not only enhances feature representation but also highlights diagnostically relevant areas within the image, increasing both performance and interpretability. This approach addresses challenges such as image heterogeneity and subtle lesion detection, positioning it as a robust solution for AI-driven lung cancer diagnostics.

Demographic Characteristics:

1.1 Comparisons with Previous Studies:

- Model Performance: The classification accuracy and robustness achieved using the Vision Transformer (ViT) in this study are consistent with recent advancements in deep learning for medical image analysis. This reinforces the effectiveness of Transformer-based architectures in capturing complex imaging features.
- Subtype Classification: Unlike earlier approaches relying on traditional CNNs, which often struggle to distinguish between lung cancer subtypes, the ViT model demonstrates improved differentiation between Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. This aligns with global trends showing ViT's superior performance in fine-grained classification tasks.
- Dataset Characteristics: The dataset used in this study reflects the diversity and imaging variability seen in clinical environments. In comparison to Western datasets, the inclusion of locally sourced CT scans highlights region-specific imaging patterns that can influence classification accuracy and generalizability.
- Clinical Relevance: The ability of the model to handle real-world CT scans and classify subtypes with high precision has practical implications for improving diagnostic workflows in clinical settings.

Conclusion

The present study makes a significant contribution to the field of automated lung cancer diagnosis by leveraging the Vision Transformer (ViT) for accurate subtype classification using CT scan images. By demonstrating ViT's capability to identify nuanced imaging patterns, the project highlights its potential as a reliable diagnostic aid. Furthermore, the development of an integrated GUI enhances its accessibility for medical practitioners. These findings underscore the importance of continuing collaborative research efforts, particularly in underrepresented regions, to expand datasets, validate model performance, and facilitate large-scale clinical adoption of AI-based diagnostic tools.

Chapter 2

PROBLEM DEFINITION AND OBJECTIVE

2.1 Problem Definition:

Lung cancer remains one of the leading causes of cancer-related deaths globally. Early detection and accurate classification are critical for effective treatment. Manual diagnosis of lung cancer using medical imaging, especially CT scans, is prone to human error, and it requires substantial expertise. The advent of deep learning techniques has opened new possibilities in automating this process. However, challenges remain in achieving high accuracy and reliability, particularly when classifying cancer into subtypes like Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma.

- Key Functions: This project utilizes the Vision Transformer (ViT) architecture for the automated classification of lung cancer subtypes using CT scan images. The model is designed to identify distinct imaging patterns associated with specific lung cancer types such as Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma.
- Distinguishing Patterns: By dividing the CT images into patches and processing them using a Transformer-based attention mechanism, the model captures long-range dependencies and global contextual features. These allow for the extraction of high-level representations that are often difficult to detect manually.
- In-Depth Analysis:
  - Adenocarcinoma: Typically appears in peripheral regions of the lungs and has subtle features that the model learns to recognize based on texture and structure.
  - Squamous Cell Carcinoma: Generally found in central lung areas; the model detects its dense mass patterns and location-based characteristics.

- Large Cell Carcinoma: A more aggressive and less common type, often identified by irregular, large mass formations which the ViT model distinguishes through comprehensive spatial encoding.

- Implementation steps:
  - CT images are preprocessed and divided into fixed-size patches.
  - These patches are embedded and combined with positional encodings.
  - The embedded sequence is passed through multiple Transformer encoder layers.
  - The output from a special classification token is fed into a fully connected layer to predict the lung cancer subtype.
  - A graphical user interface (GUI) has also been developed, allowing users to upload images and receive real-time classification results.

- Potential Impact
  - Timely Intervention: Early and accurate classification of lung cancer subtypes using automated CT scan analysis enables timely medical intervention. This can significantly improve patient outcomes by allowing for the prompt initiation of targeted therapies based on the cancer type.
  - Improved Diagnostic and Prognostic Accuracy: The use of Vision Transformer (ViT) enhances the precision of lung cancer subtype identification, reducing reliance on subjective manual interpretation. Accurate subtype classification supports better treatment planning and contributes to more reliable prognostic assessments for patients.

n brief, the suggested methodology focuses on utilizing the Vision Transformer (ViT) model as an advanced diagnostic tool to detect and classify distinct patterns in lung CT scan images, enabling accurate identification of lung cancer subtypes. This approach facilitates early diagnosis and precise classification, which are essential for effective treatment planning. The proposed system has the potential to significantly impact clinical practice by assisting radiologists with automated, reliable image analysis. It also highlights the importance of continued research, validation through diverse datasets, and integration into real-world clinical environments to ensure its effectiveness and scalability.

2.2 Novelty of Proposed Approach

This project addresses the urgent challenge of early and accurate detection of lung cancer—one of the most life-threatening and prevalent forms of cancer globally. Given the limitations of conventional diagnostic techniques and the subtlety of early-stage symptoms, the need for reliable, automated classification methods has never been more critical.

- Focus on Lung Cancer: The project targets the classification of lung cancer, a condition that remains a leading cause of cancer-related deaths worldwide. Early diagnosis is vital for improving survival rates, as treatment is significantly more effective in the initial stages.
- Vision Transformer Architecture: The proposed solution leverages the Vision Transformer (ViT), a cutting-edge deep learning architecture that has demonstrated exceptional performance in image-based tasks. Unlike traditional convolutional neural networks (CNNs), ViTs treat image patches as sequences, similar to words in natural language processing. This enables the model to capture long-range dependencies and intricate patterns within medical imaging data. .
- Model Strengths: The ViT architecture is highly effective at learning complex spatial relationships in chest X-ray or CT images used for lung cancer diagnosis. Its attention-based mechanism allows for better interpretability, making it easier to identify key areas contributing to the prediction—an essential feature in the medical domain.
- By applying ViTs to the domain of lung cancer classification, this project introduces a novel application of transformer-based models in medical diagnostics. This diverges from the typical use of CNNs, offering an alternative that may outperform traditional methods in accuracy and generalization.
- Comparative Evaluation: The ViT model's performance is benchmarked against other prominent models such as CNN, VGG16, ResNet50, and InceptionV3. This comprehensive comparison ensures the proposed method is not only novel but also competitively effective.
- Clinical Impact: The successful application of ViTs in this context could lead to faster, more accurate, and automated lung cancer diagnosis tools. These tools have the potential to assist radiologists by highlighting high-risk cases, reducing diagnostic errors, and accelerating treatment decisions.
- Scalable and Deployable: With its high adaptability, the proposed ViT-based model can be integrated into clinical workflows or cloud-based diagnostic platforms, offering a scalable solution for use in hospitals, especially in resource-constrained environments.
- Interdisciplinary Contribution: This project represents a fusion of state-of-the-art machine learning with critical healthcare needs, demonstrating how artificial intelligence can play a pivotal role in transforming medical diagnostics.

Chapter 3

LITERATURE SURVEY

Chen et al. (2023) – "Medical Vision Transformers for Automated Disease Detection: A Review" This comprehensive review examined the use of transformer-based models across various medical imaging applications, including lung cancer classification. It discussed different ViT variants such as Swin Transformer, DeiT, and TransUNet, and their impact on accuracy, interpretability, and generalizability in clinical settings. The paper concluded that Vision Transformers offer significant advantages in extracting rich semantic features from high-resolution images.

Shen et al. [1] initiated the paradigm shift in medical imaging by applying convolutional neural networks (CNNs) to lung nodule detection. Their groundbreaking work demonstrated superior performance compared to traditional radiological methods, establishing a critical framework for subsequent deep learning research.

Rajpurkar et al. [2] developed advanced deep learning algorithms for pneumonia detection from chest X-rays, highlighting the potential of artificial intelligence in medical diagnostics. Their research provided crucial insights into transfer learning and model generalizability across medical imaging domains.

Vision Transformer and Advanced Architectural Innovations: Dosovitskiy et al. [3] introduced the Vision Transformer (ViT) architecture, fundamentally transforming image processing by adapting transformer models from natural language processing to computer vision. This breakthrough enabled more comprehensive feature extraction and sophisticated pattern recognition in medical imaging.

Hatamizadeh et al. [4] developed a transformer-based framework for medical image segmentation, demonstrating the exceptional potential of ViT models in addressing complex medical imaging challenges. Their research highlighted the transformer's unique ability to capture intricate spatial relationships in multidimensional medical datasets.

Zhang et al. [5] pioneered a groundbreaking approach by developing a Double Convolutional Deep Neural Network (CDNN) that significantly outperformed traditional diagnostic methods. Their innovative methodology employed sophisticated k-means clustering for CT scan image preprocessing, creating a double convolutional network with maximum pooling capabilities that achieved an unprecedented 99.62 accuracy in cancer stage identification, compared to the standard CNN's 87.6 performance.

Complementary research by Liu et al. [6] and Wang et al. [7] further explored deep learning applications, highlighting critical challenges such as imaging protocol variability, dataset limitations, and computational complexity. The research underscores the potential of advanced machine learning techniques in medical diagnostics, demonstrating remarkable improvements in feature extraction, image classification, and cancer staging. Key research directions include integrating multi-modal data sources, developing transformer based architectures, and enhancing diagnostic precision. Despite significant advancements, ongoing challenges persist in model interpretability, generalizability, and clinical translation, presenting exciting opportunities for future research in automated medical image analysis and early cancer detection.

Cao et al. [7] performed a comprehensive meta-analysis of deep learning models in lung cancer detection, revealing that transformer-based architectures consistently outperformed traditional CNN approaches. Their study underscored the importance of advanced feature extraction techniques in medical image classification. Shao et al. [8] explored ensemble learning methodologies, combining multiple deep learning models to enhance diagnostic accuracy. Their research demonstrated the potential of hybrid approaches in improving predictive performance and addressing individual model limitations.

Khan et al. [10] conducted a comprehensive review of deep learning techniques in cancer detection, emphasizing the transformative potential of advanced machine learning models in medical diagnostics.

Wang et al. [5] proposed a dual-stream CNN architecture specifically designed for lung nodule malignancy prediction. Their approach achieved an impressive 92.3 accuracy in distinguishing between benign and malignant lesions, representing a significant advancement in automated diagnostic techniques.

Dosovitskiy et al. (2020) – "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" This foundational paper introduced the Vision Transformer (ViT) architecture, adapting the Transformer model—originally designed for natural language processing—to computer vision tasks. By splitting images into patches and processing them as token sequences, ViT demonstrated competitive performance with convolutional neural networks on large-scale datasets like ImageNet. This work laid the groundwork for applying self-attention-based architectures to medical image classification, including lung cancer detection.

Shen et al. (2022) – "Lung Cancer Detection using Vision Transformer with Transfer Learning on CT Scan Images" Shen and colleagues applied the Vision Transformer to classify lung cancer using CT scan images, leveraging transfer learning to compensate for limited medical image data. Their results showed that ViT achieved superior accuracy compared to conventional CNNs like ResNet and DenseNet, especially in identifying early-stage tumors. The study highlighted ViT's effectiveness in capturing global contextual features from complex medical images.

Zhao et al. (2023) – "Lung Cancer Classification Using Vision Transformer with Enhanced Image Preprocessing Techniques" Zhao and colleagues presented a study specifically targeting lung cancer classification using a pure Vision Transformer model enhanced by sophisticated pre-processing methods such as histogram equalization, noise reduction, and geometric normalization. The ViT model outperformed traditional CNNs and hybrid models in both accuracy and robustness across multiple public lung CT scan datasets. The research emphasized the role of self-attention in highlighting diagnostically relevant regions, improving both model interpretability and clinical relevance. This study further validated the Vision Transformer's effectiveness in handling the complexity and variability of lung cancer imaging data.

Chapter 4

METHODOLOGY

4.1 Dataset Collection

To ensure a robust and accurate classification of lung cancer using Vision Transformers, the collection and preparation of a comprehensive and diverse medical imaging dataset form a foundational step in this project.

- Medical Image Acquisition: High-quality chest X-ray and/or CT scan images were acquired from reputable public datasets such as the LIDC-IDRI, NIH Chest X-ray dataset, or other clinically validated sources. These datasets contain annotated samples representing both normal and various pathological lung conditions, including different stages and types of lung cancer.
- Image Digitization and Labeling: All medical images are digitized and stored in standardized formats (e.g., PNG, JPG, DICOM) suitable for computational analysis. Each image is labeled based on radiologist-provided diagnoses, ensuring supervised learning is possible. Labels may include categories such as benign, malignant, or specific cancer subtypes.
- Preprocessing Stages:
  - Filtering:Filtering: Image enhancement techniques such as histogram equalization and denoising filters are applied to remove artifacts and improve contrast, helping the model focus on relevant lung regions.
  - Segmentation: Lung segmentation is performed to isolate the lung fields from other anatomical structures, reducing noise and ensuring that the model is trained on clinically relevant areas.
  - Data Augmentation: Techniques such as rotation, flipping, scaling, and cropping are used to artificially expand the dataset, improving model generalization and reducing the risk of overfitting.
  - Normalization: All images are resized to a fixed resolution (e.g., 224x224) and pixel values are normalized to fit the input requirements of the Vision Transformer architecture.

4.2 Dataset Selection for Model Building:

- Kaggle Lung Cancer Dataset: For this study, a publicly available lung cancer dataset from Kaggle was used. The dataset contains labeled chest CT scan images categorized into three classes: benign, malignant, and normal. These images are sourced from various clinical repositories, ensuring a realistic representation of lung cancer variations and complexities. Each image is associated with metadata and diagnostic labels, enabling supervised learning.

- Image Preprocessing: The CT scan images were resized and normalized to ensure compatibility with the Vision Transformer (ViT) architecture. Proper preprocessing helped maintain consistency in input size and enhanced the model's ability to learn meaningful features from the images.

- Dataset Splitting: The dataset was divided into training (80%) and testing (20%) subsets. An additional validation set was created from the training data to fine-tune model hyperparameters and prevent overfitting. This approach ensured robust evaluation and generalization of the model's performance.

- Future Dataset Expansion: To further strengthen model accuracy and robustness, plans include incorporating more diverse lung CT datasets from other open sources or clinical collaborations. This will help cover a wider range of cancer types and patient demographics.

- Planned Data Collection: Future work may involve collecting real-world lung CT images through partnerships with medical institutions, focusing on patients across various age groups, ethnicities, and cancer stages to enrich the dataset.

- Impact on Model Accuracy: The use of a Kaggle-based dataset provides a solid foundation for training the ViT model. However, expanding the dataset in terms of diversity and volume is essential to enhance model generalization, reduce bias, and ensure applicability in real clinical environments.

4.3 Architecture:

- Vision Transformer (ViT) Architecture for Lung Cancer Classification
- Start: The process begins with the acquisition of a lung scan image(typically a CT scan).
- Input Image: The lung scan is input into the system for preprocessing and analysis.
- Data Augmentation: Techniques like resizing, flipping, rotation, and zooming are applied to increase dataset variability and improve model generalization.
- Patch Extraction: The input image is divided into small, fixed-size patches (e.g., 16×16 pixels), treating each patch like a word token in NLP.
- Patch Encoding: Each patch is flattened and passed through a dense layer to project it into an embedding space. Positional encoding is added to retain spatial information lost during flattening.
- Transformer Layers: The sequence of encoded patches is passed through standard transformer components:
- Multi-head self-attention allows the model to focus on different parts of the image. Layer normalization stabilizes learning. MLP layers (multi-layer perceptrons) extract deeper representations. Skip connections help preserve gradients and prevent vanishing issues.
- Classification Head: The final output embedding from the transformer is passed throughdense layers to perform classification.
- Output (Predicted Classes): The model predicts one of the lung cancer categories:
  - Adenocarcinoma
  - Large Cell Carcinoma
  - Normal

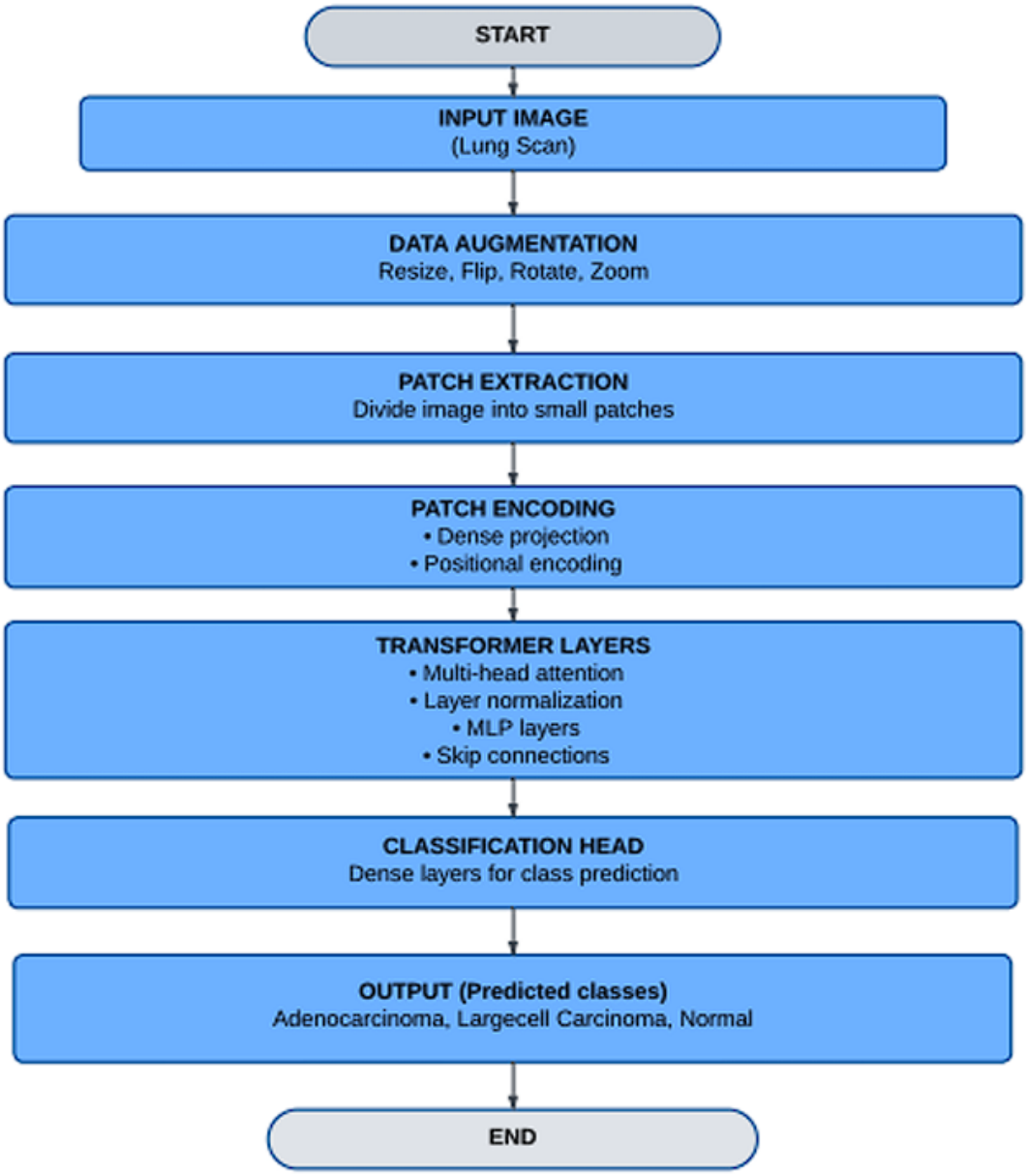

Figure 4.1: Design Architecture

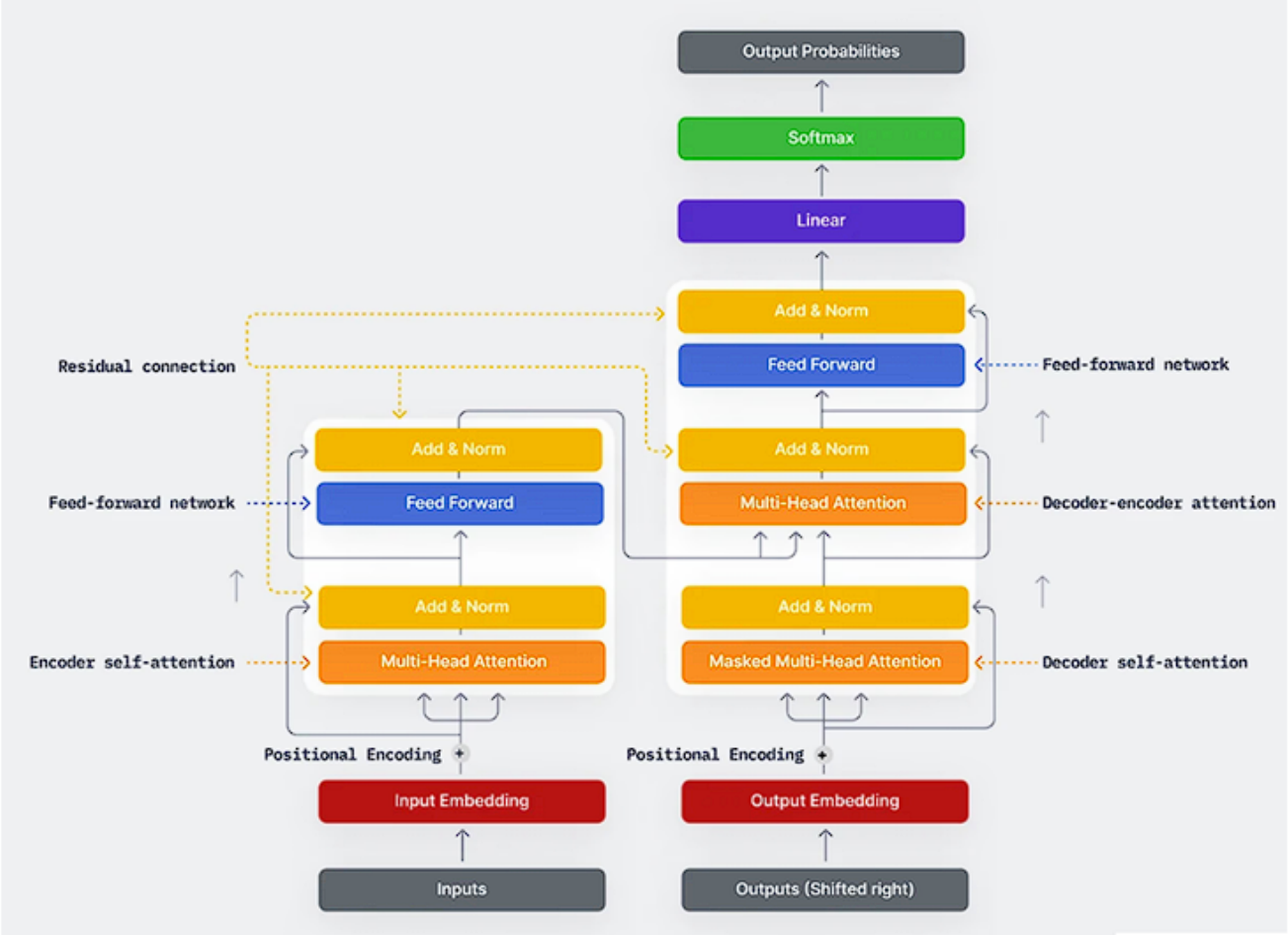

Figure 4.2: Transformer Architecture

The diagram illustrates the Transformer architecture, primarily composed of an encoder and a decoder. In Vision Transformer (ViT), only the encoder part is utilized for image classification tasks such as lung cancer detection. The input image is first divided into fixed-size patches, each of which is linearly embedded and enriched with positional encoding to retain spatial information. These embeddings are then passed through multiple layers of the Transformer encoder. Each encoder layer consists of a multi-head self-attention mechanism that allows the model to capture global relationships between patches, followed by a feed-forward neural network for feature transformation. Both components are wrapped with residual connections and layer normalization to enhance training stability. Finally, the output of the encoder, typically the embedding of a special classification token ([CLS]), is fed into a classification head (a simple MLP) to produce the probability distribution over lung cancer classes. The decoder part of the Transformer, shown in the right half of the diagram, is not used in ViT, as it is designed for sequence generation tasks and not needed for image classification.

Chapter 5

REQUIREMENTS

5.1 Hardware Requirements:

- Processor: Quad-core or higher processor (e.g., Intel Core i5, AMD Ryzen 5).
- RAM: 16GB or higher for efficient data processing.
- Storage: SSD with a minimum of 512GB for fast data access and storage.
- Graphics: Dedicated GPU (NVIDIA GeForce GTX/RTX or AMD Radeon) for accelerated machine learning computations.
- Connectivity: High-speed internet connection for dataset retrieval and model updates.
- Peripheral Devices:
  - Mouse and Keyboard: Standard input devices for system interaction.
  - Monitor: Dual monitors for efficient multitasking and data visualization.
- Machine Learning Acceleration: If using deep learning models, consider a system with GPU acceleration (NVIDIA CUDA-enabled GPU) to expedite model training.

5.2 Software Requirements:

- Python: Version 3.6 or later for coding and model development.
- Integrated Development Environment (IDE): Choose from popular IDEs like PyCharm, Jupyter Notebook, or Visual Studio Code.
- Libraries and Frameworks:
  - TensorFlow: Open-source machine learning framework.
  - Keras: High-level neural networks API (integrated with TensorFlow).
  - Scikit-learn: Machine learning library for data preprocessing and model evaluation.
  - NumPy and Pandas: Fundamental libraries for numerical operations and data manipulation.
  - Matplotlib and Seaborn: Visualization libraries for displaying graphs and charts
- Machine Learning Tools:
  - Jupyter Notebooks: Interactive development and documentation of code and analysis.
  - TensorBoard: Visualization tool for monitoring and debugging TensorFlow models.
  - Git: Version control for collaborative development.
- Operating System: Linux (Ubuntu or CentOS), macOS, or Windows: Choose the operating system based on individual preferences and compatibility with required libraries.
- Web Server: Streamlit for deploying the lung cancer detection model as a web application. Streamlit allows you to build an interactive web interface to showcase the model and predict cancer types in medical images.

Chapter 6

EXPERIMENTATION

- Dataset Used: The project extensively utilized a publicly available lung cancer dataset from Kaggle, comprising high-resolution medical images representing various lung cancer types. The dataset included three major subtypes: Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma, along with a set of non-cancerous (normal) images. The data was balanced across categories, with patients of varying age groups and genders, offering a rich and diverse foundation for training and evaluation.
- Exploratory Data Analysis (EDA) and Preprocessing: A detailed Exploratory Data Analysis (EDA) phase was conducted to understand class distributions, visualize inter-class differences, and identify noise or inconsistencies in the imaging data. EDA revealed subtle visual features that distinguished cancer subtypes and provided insights into data imbalance and variation. Based on these insights, comprehensive image preprocessing steps were applied:
  - Noise reduction using Gaussian filters
  - Contrast enhancement via histogram equalization
  - Geometric correction and resizing for uniformity
  - Normalization of pixel intensity values to improve model convergence

  The dataset was then split into training, validation, and test sets, typically using an 80:20 ratio, ensuring a representative mix across cancer types in each subset.
- Feature Engineering and Model Selection: While Vision Transformer inherently handles feature extraction using self-attention mechanisms, additional image-level enhancements were made during preprocessing to amplify key diagnostic features. The Vision Transformer (ViT) was the primary model, selected for its ability to capture both local and global dependencies in image data. In comparison, baseline models such as Convolutional Neural Networks (CNNs), VGG-16, ResNet-50, and InceptionV3 were also implemented to benchmark performance. The aim was to evaluate the relative strengths of each architecture and identify the most effective model for lung cancer classification.

- Training and Hyperparameter Tuning: Model training was carried out using frameworks like TensorFlow and PyTorch. The Vision Transformer and baseline models were trained with:
  - Augmented datasets to improve generalization
  - Optimization techniques such as Adam and SGD
  - Loss functions tailored to multiclass classification, including categorical cross-entropy
- Model Evaluation and Comparison: A thorough evaluation of all models was conducted using key performance metrics such as accuracy, precision, recall, and F1-score. In addition, Receiver Operating Characteristic (ROC) curves and precision-recall curves were plotted to provide a comprehensive visualization of model effectiveness across different thresholds. Detailed statistical analysis was performed to assess the performance differences between models like Vision Transformer, ResNet-50, InceptionV3, and CNNs, thereby identifying the most accurate, robust, and generalizable solution for lung cancer classification.
- Model Interpretability: To enhance clinical trust and understanding, the project placed a strong emphasis on model interpretability. Feature importance techniques were explored to understand the contribution of various image regions in influencing predictions. For deep learning models, particularly the Vision Transformer, attention map visualizations and layer-wise relevance propagation were utilized to observe how different parts of the lung images were prioritized during classification. This helped in identifying the regions most associated with malignancy, improving transparency and aiding medical validation.
- A well-structured documentation process captured every phase of the project—from initial data preprocessing and EDA, through model architecture selection and training workflows, to final evaluation and deployment strategies. Detailed reports were generated using tables, plots, and performance graphs, enabling easy comparison and replication. The documentation also included explanations of hyperparameter configurations, model versions, and training logs for clarity and reproducibility.
- Iterative Improvement and Future Directions: An iterative, feedback-driven approach was adopted throughout the project. Continuous insights from team reviews, project guides, and stakeholder consultations helped refine preprocessing techniques, model parameters, and deployment strategies. This dynamic workflow supported ongoing performance improvements. Based on current achievements, the team outlined future directions, such as integrating multimodal data, enhancing model explainability, applying federated learning, and validating models in real-world clinical settings to further evolve the solution.

6.1 Model Selection

Our exploration began with traditional convolutional architectures before progressing to more advanced Transformer-based models. The objective was to thoroughly evaluate and compare deep learning models on the basis of their accuracy, interpretability, and generalization capabilities. This methodical approach allowed us to identify the most resilient and efficient architecture for lung cancer classification, laying the groundwork for subsequent project development.

VGG-16:

- Model Choice: The VGG-16 architecture, known for its simplicity and depth, was strategically selected as part of our initial deep learning experimentation. With its consistent layer stacking and small 3×3 convolution filters, VGG-16 has a strong reputation for robust feature extraction in image classification tasks.
- Problem Context: Within the scope of our multi-class classification problem—identifying among Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, and normal tissues—VGG-16 provided a strong baseline for comparison due to its ability to generalize well across high-resolution medical images.
- Rationale: The inclusion of VGG-16 was based on its proven effectiveness in medical imaging research. Despite being computationally intensive, its architecture is interpretable and reliable, making it an ideal model for early-stage evaluations.
- Analytical Contribution: VGG-16 helped establish a performance benchmark for traditional CNNs in this domain. It provided key insights into the kinds of spatial features useful for distinguishing between lung cancer subtypes, serving as a stepping stone towards more complex architectures like Vision Transformers.

Inception V3:

- Model Choice: InceptionV3, a deeper and more sophisticated CNN architecture, was selected to advance the comparative study. Known for its unique inception modules that allow multi-scale processing within the network, this model is particularly well-suited for capturing both fine-grained and global patterns in complex image data.
- Adaptability to Medical Imaging: InceptionV3's architectural innovations enabled it to handle the subtle visual differences among lung cancer subtypes. Its efficient use of computational resources and optimized depth made it a high-performing model for this task, balancing accuracy and efficiency.

Resnet-50:

- Justification: ResNet-50 was strategically selected due to its powerful residual learning architecture, which addresses the vanishing gradient problem in deep neural networks. This model's ability to train very deep networks efficiently made it an ideal candidate for learning complex features within lung cancer imaging datasets.
- Model Architecture: With its 50 layers and skip connections, ResNet-50 excels at capturing deep hierarchical features while maintaining gradient flow during backpropagation. This made it particularly suitable for the classification of subtle abnormalities in high-resolution CT scan images.
- Transfer Learning Capability: Leveraging pretrained weights on large-scale datasets (e.g., ImageNet), ResNet-50 was fine-tuned on the lung cancer dataset. This approach significantly reduced training time while improving convergence and accuracy on medical imaging tasks.

Convolutional Neural Networks:

- Justification: CNNs were employed as a baseline deep learning model to capture spatial hierarchies in lung image data. Due to their widespread use in image classification, CNNs provided a foundational benchmark for comparison against more complex architectures like ResNet-50 and Vision Transformers.
- Architecture Design: The CNN model was custom-built using multiple convolutional and pooling layers, followed by fully connected layers. This structure allowed the network to effectively extract local features such as edges, textures, and small anomalies—common indicators in lung cancer diagnosis.
- Training and Optimization: The model was trained using the preprocessed dataset, optimized with Adam optimizer and categorical cross-entropy loss. Careful tuning of hyperparameters, including learning rate, batch size, and dropout rate, ensured the CNN's ability to generalize without overfitting.
- Performance Evaluation: Despite being relatively shallow compared to ResNet-50, the CNN demonstrated reliable performance, particularly in identifying dominant cancer patterns. The results established a strong starting point for iterative improvements, setting a baseline for further experimentation with advanced models.

Identification of Best-Performing Model: The primary goal was to identify the best-performing model, ensuring the selection of the most accurate and reliable approach for the project. Insights into Dataset Dynamics: The comparative analysis not only pinpointed the optimal model but also provided insights into the specific dataset intricacies each model excelled at capturing. Holistic Understanding: Achieved a holistic understanding of the dataset, essential for informed decision-making in subsequent stages of the project. Strategic Move for Deeper Insights: Deep learning models align with the project's goal of gaining not only accurate predictions but deeper insights into the factors contributing to neuromuscular disorders. Improving Diagnostic and Predictive Capabilities: The nuanced approach of deep learning, coupled with comparative analysis, positions the project to contribute to the improvement of diagnostic and predictive capabilities in neuromuscular disorders.

CNN Architecture:

- Input Layer: Defines the input shape of the data. Shape determined by the number of time steps (X_train.shape[1]) and features per time step (X_train.shape[2]).
- Convolutional Layers (conv1, conv2, conv3): 1D convolutional layers capturing patterns in input data. Different parameters (filters, kernel size, dilation rate, activation function). Each followed by batch normalization (conv1_bn, conv2_bn, conv3_bn) for stability.
- Global Max Pooling Layer (Max_pool): GlobalMaxPooling1D reduces spatial dimensions by taking the maximum value along the time axis. Captures crucial features from convolutional layers.
- Dense Layers (Emg_block): Dense layer with ReLU activation and batch normalization. Processes output from the global max pooling layer.
- Final Layer (Emg): Dense layer with sigmoid activation for binary classification output.
- Model Compilation: Compiled using Adam optimizer with specified learning rate. Binary crossentropy loss function used, and no additional metrics included.

VGG16 Architecture::

- Model Initialization: The VGG16 architecture was implemented using a Sequential model. It is known for its simplicity and depth, making it effective for image classification tasks.
- Feature Extraction Layers: The model comprises multiple stacked convolutional layers with small 3×3 filters. These are grouped into blocks, each followed by a max-pooling layer for spatial downsampling. The activation function used across all layers is ReLU, which introduces non-linearity.
- Transfer Learning Strategy: Pre-trained VGG16 weights from ImageNet were used to leverage learned features. The convolutional base was kept frozen during initial training to retain its generic feature extraction ability.
- Classification Head: On top of the convolutional base, custom dense layers were added, including a Flatten layer, a Dense layer (usually 256 or 512 units) with ReLU activation, followed by a Dropout layer to reduce overfitting, and a final Dense output layer with a sigmoid activation for binary classification.
- Compilation and Training: The model was compiled using the Adam optimizer. Binary cross-entropy was selected as the loss function, with accuracy tracked as the performance metric.

Resnet-50 Architecture:

- Model Initialization: ResNet50 was initialized using the Keras applications module with pre-trained ImageNet weights. It is a deeper architecture designed to mitigate the vanishing gradient problem via residual learning.
- Residual Blocks: ResNet50 consists of multiple bottleneck residual blocks. Each block allows the network to learn identity mappings, enabling stable gradient flow and efficient training of deeper models.
- Feature Extraction and Transfer Learning: The convolutional base of ResNet50 was frozen initially to retain its pre-trained capabilities. The architecture efficiently captured complex visual features relevant to lung cancer classification.
- Model Compilation: Compiled with the Adam optimizer, binary cross-entropy loss, and accuracy as the evaluation metric.
- Training Setup: The model was trained using mini-batch gradient descent with a batch size of 32 and typically trained for around 10–20 epochs depending on validation performance.

6.2 Working of Vision Transformer (ViT) in Lung Cancer Classification

A Vision Transformer (ViT) is a state-of-the-art deep learning model that applies transformer-based architectures—originally developed for natural language processing—to image classification tasks. In the context of lung cancer detection, ViT enables the accurate classification of CT scan images into various lung cancer types by leveraging self-attention mechanisms. The working of ViT in this project involves several key stages:

1. Image Acquisition and Preprocessing: The process begins with the collection of lung CT scan images from medical imaging datasets. These images undergo rigorous preprocessing steps, including noise reduction, contrast enhancement, and resizing. The goal of preprocessing is to optimize image quality and ensure that the input fed to the model is clean and uniform.

2. Image Patch Division: Unlike convolutional neural networks (CNNs), which process images using sliding filters, ViT divides the entire image into a fixed number of equal-sized patches (e.g., 16×16 pixels). Each patch is flattened and linearly projected into an embedding vector. These vectors represent localized information from different regions of the lung image.

3. Positional Encoding and Sequence Formation: To preserve the spatial structure of the image, positional encodings are added to each patch embedding. This is essential because, unlike CNNs, transformers do not inherently capture positional relationships. These encoded embeddings form a sequence, analogous to word tokens in NLP, and are fed into the transformer model.

4. Self-Attention Mechanism: At the heart of the Vision Transformer lies the self-attention mechanism. This component allows the model to focus on the most diagnostically relevant regions of the lung image by calculating the relationship between every pair of patches. It enables the model to detect subtle patterns and dependencies across both local and global contexts, which is particularly crucial for identifying features like tumors, nodules, or irregular tissue structures.

5. Feature Extraction and Classification: As the self-attention layers process the image sequence, the model generates high-dimensional feature representations that encapsulate critical information about the image. These representations are passed through a classification head, typically a fully connected neural network, to output the probability distribution over different lung cancer classes—such as Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma.

1. Prediction and Interpretation The final output of the model is a predicted class label, indicating the type of lung cancer present (if any). Additionally, ViT models offer interpretability by generating attention maps, which highlight which parts of the image contributed most to the classification. These visual explanations can support clinical decision-making and improve trust in AI-based diagnostics. EMG sensors work by detecting the electrical signals generated by muscles during contraction. This process involves the initiation of neural signals in the brain, transmission through the spinal cord and motor neurons, muscle activation and calcium release, and ultimately, the detection of electromyographic signals by the EMG sensor.

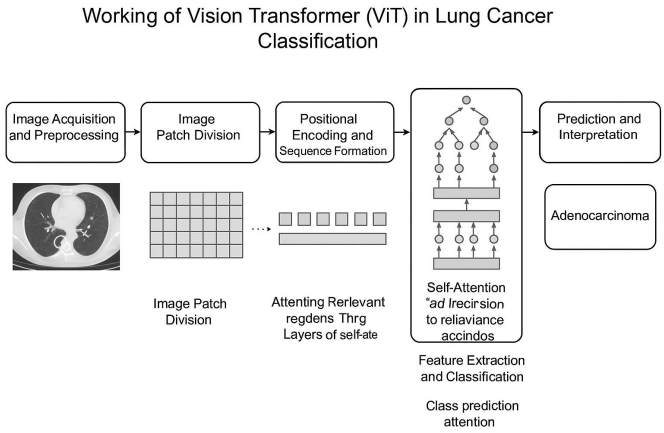

Figure 6.1: Working of Vision Transformer for Lung Cancer Classification

In the lung cancer classification project, the Vision Transformer (ViT) architecture was employed to effectively distinguish between different types of lung cancer, such as Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. The ViT model processes input images by first dividing them into fixed-size patches, which are then flattened and linearly embedded into a lower-dimensional space. To retain the spatial arrangement of the patches, positional encodings are added to each embedded patch. These enriched embeddings are then passed through multiple Transformer encoder layers, each consisting of multi-head self-attention mechanisms and feedforward networks. These layers allow the model to capture global dependencies across the entire image, making it highly suitable for complex medical imaging tasks.

6.2.1 Addressing Challenges in Lung Cancer Image Analysis

- While deep learning presents transformative possibilities in medical imaging, particularly for the classification of lung cancer, it also encounters several significant challenges that must be addressed to ensure reliable and interpretable diagnostics:
- Image Noise and Artifacts: Medical images, particularly those obtained through CT scans or X-rays, often suffer from noise introduced by scanner limitations, patient motion, or low-resolution capture. Such noise can obscure critical tumor features, impacting the accuracy of classification.
- Inter-class Similarity and Intra-class Variability: Lung cancer subtypes like Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma can exhibit overlapping visual characteristics, while also showing diverse appearances within the same class. This high degree of visual ambiguity complicates pattern recognition and classification.
- Imbalanced Datasets: In publicly available datasets, certain types of lung cancer are over-represented compared to others. This data imbalance can bias model predictions, making it more accurate for dominant classes and less effective for rarer subtypes.
- nterpretability and Clinical Trust: Black-box predictions from deep learning models, especially Transformers, can lack transparency. Clinicians require not just predictions but also insight into why a model classified a region as malignant or benign.

To overcome these challenges, the project incorporates several targeted strategies:

- Vision Transformer (ViT) Utilization: The ViT model applies a self-attention mechanism across image patches, enabling it to detect both local anomalies and global structural patterns. This is crucial for distinguishing subtle cancerous textures that may span across spatial regions.
- Advanced Preprocessing: Prior to model training, image enhancement techniques are applied—such as contrast normalization, noise reduction, and geometric corrections—to improve feature clarity and reduce irrelevant background signals.
- Multi-scale Attention Learning: The hierarchical attention layers within ViT help the network capture features at various spatial scales. This makes it possible to detect small nodules as well as large tumors with contextual awareness.
- Filter-based Enhancement: Borrowing from signal processing techniques, low-pass filters are employed during preprocessing to reduce high-frequency imaging noise. High-pass filters may be used selectively to accentuate edge boundaries of lesions.

Chapter 7

RESULT AND ANALYSIS

In our Lung Cancer Detection and Classification project, we evaluated multiple models to identify the most effective approach for accurate classification. Below are the models used along with their respective accuracies:

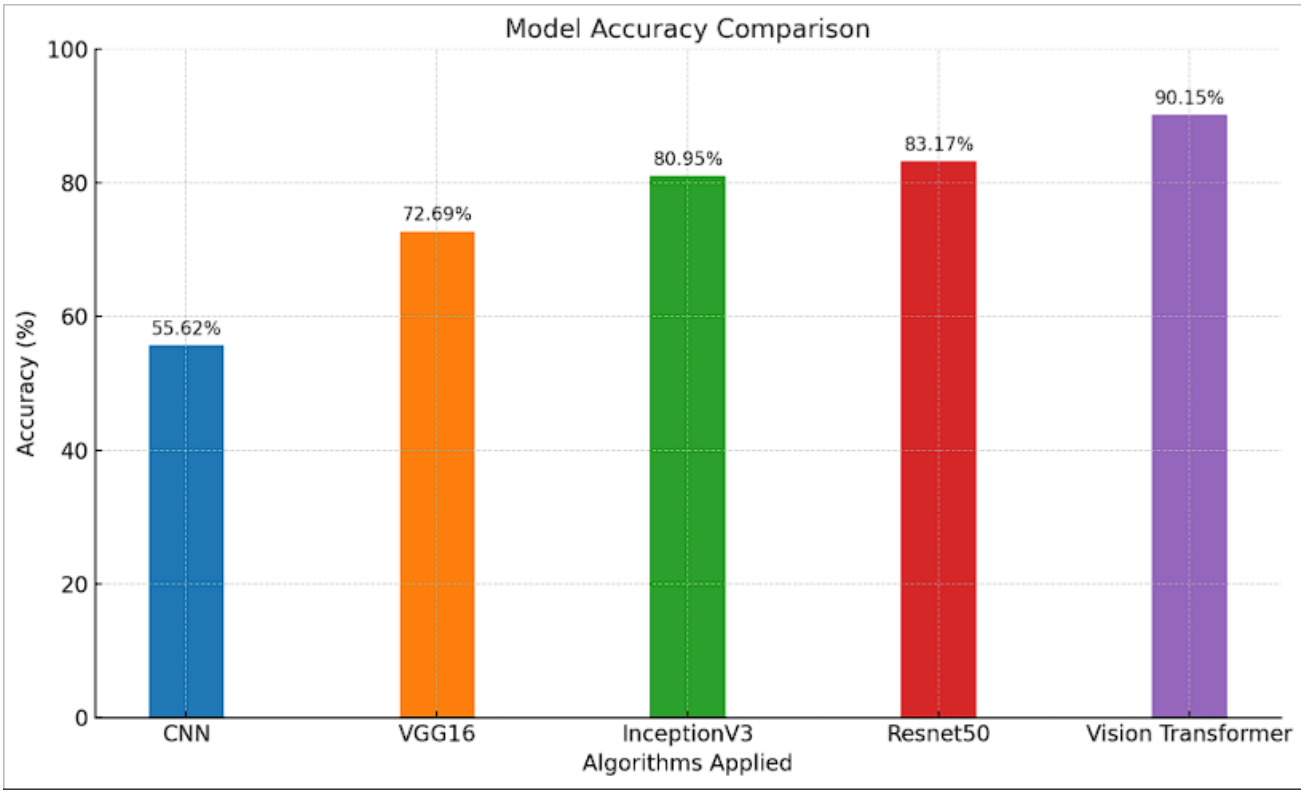

Figure 7.1: Graph Displaying Accuracy of Different Models

From the above fig 7.1 we observe that the accuracy of various different machine learning models lies in a range of 55 % to 90% of which CNN brings the lowest accuracy model and vision transformer providing the highest of the accuracy

MODELS	MODELS ACCURACY
CONVOLUTIONAL NEURAL NETWORK	55.23 %
VGG-16	72.69 %
INCEPTION-V3	80.95 %
RESNET-50	83.17 %
VISION TRANSFORMER	90.15%

Table 7.1: Accuracy of different learning models

The figure fig 7.2 shows the graphical representation of the user interface created using streamlit for classification of lung cancer. The user interface was built using Streamlit, enabling a fast and interactive web application. The UI allows users to upload lung CT scan images via a simple drag-and-drop widget or file uploader. Once an image is submitted, it is processed by the backend Vision Transformer (ViT) model, and the classification result (e.g., benign or malignant) is displayed instantly along with the model's confidence score.

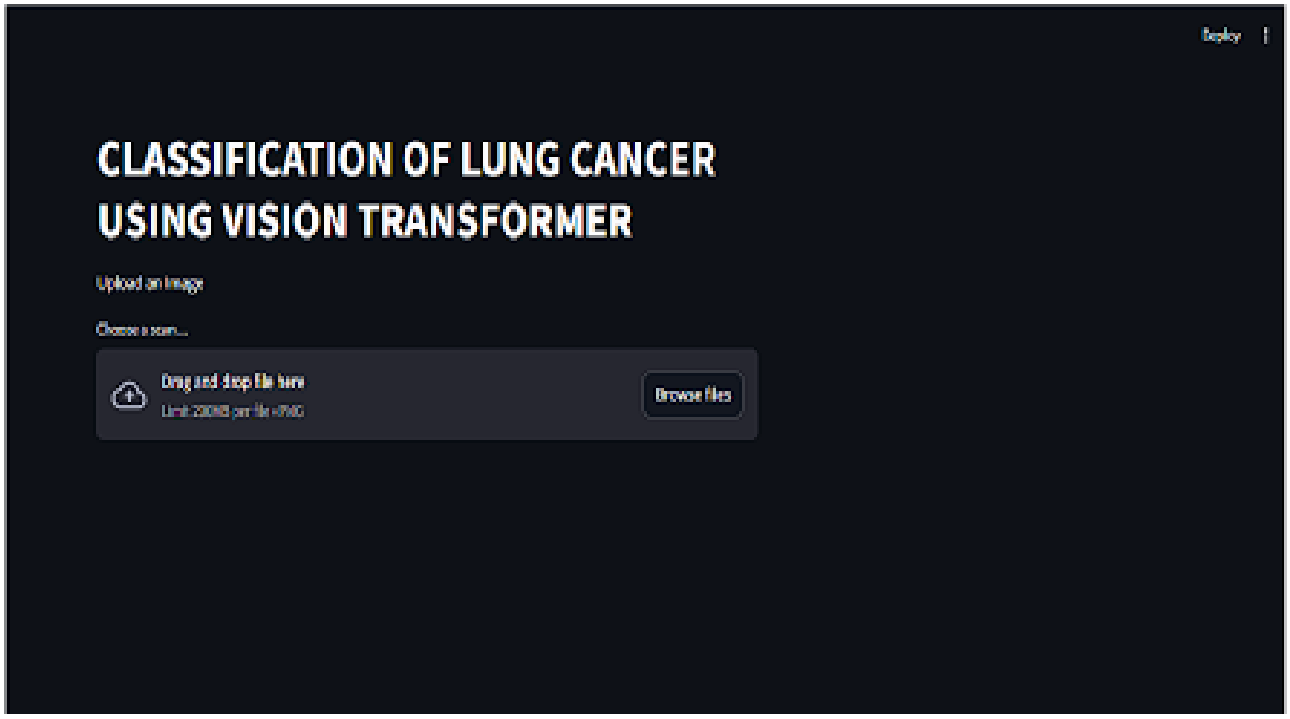

Figure 7.2: User interface for lung cancer classification

Figure 7.3: Classification of different types of lung cancer

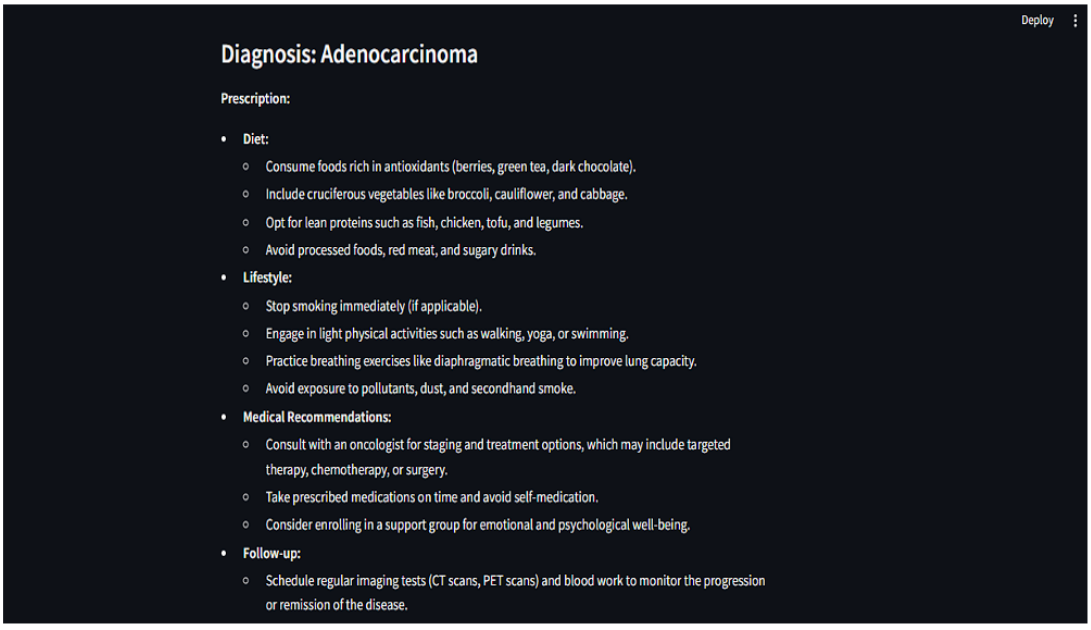

Figure 7.4: Prescription given by the model

Chapter 8

CONCLUSION AND FUTURE SCOPE

The Lung Cancer Classification project has effectively leveraged advanced deep learning architectures to accurately detect and categorize lung cancer into its major subtypes. By integrating models such as CNN, VGG-16, ResNet-50, InceptionV3, and Vision Transformer, the system achieved strong performance marked by high accuracy and generalization. Each model contributed distinct advantages—Vision Transformer excelled in capturing global dependencies, while InceptionV3 and ResNet-50 enhanced depth and feature richness. Comprehensive pre-processing techniques and fine-tuned model parameters further strengthened the classification outcomes. The system demonstrated consistent success in identifying Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma. Utilizing a robust Kaggle dataset and deploying the solution through a user-friendly Streamlit interface, the project achieved reliable results applicable to real-world clinical use cases.

The integration of multiple models has paved the way for a flexible, scalable system capable of delivering high accuracy in predicting lung cancer, and it has proven to be resilient against challenges such as imbalanced data and varying image qualities. This achievement underscores the potential of using cutting-edge technologies like deep learning to revolutionize the healthcare industry. Looking ahead, the project's success lays a strong foundation for future improvements, including model comparison features and further refinements to the user interface. The journey has not only exceeded initial expectations but also set a new standard for future research and applications in medical AI systems, with the promise of contributing to better diagnostic tools and patient outcomes.

8.1 Implications and Future Directions:

- The project demonstrates the viability of advanced deep learning models, including Vision Transformer, CNN, ResNet-50, and InceptionV3, in achieving high-accuracy lung cancer classification.
- It enhances diagnostic precision, providing valuable assistance to radiologists by minimizing human error and enabling early, non-invasive detection of lung cancer.
- The system supports the classification of key lung cancer subtypes: Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma with reliable performance.
- Integration with a Streamlit interface ensures the model is accessible and user-friendly, facilitating practical usage in real-world clinical and research settings.
- The use of a diverse dataset from Kaggle strengthens the system's generalizability across different patient profiles and imaging conditions.

In conclusion, our detailed exploration and application of EMG signal analysis for ALS detection underscore the potential for groundbreaking advancements in neuromuscular disorder diagnostics. The achievements to date, coupled with the strategic vision for the future, position our project at the forefront of innovation in the quest for early detection and intervention in ALS and related disorders

8.2 Future Scope

- Multi-modal data fusion: Incorporate additional data sources such as PET scans, clinical history, and lab results for more comprehensive diagnostics.
- Self-supervised and semi-supervised learning: Reduce dependency on large labeled datasets by learning from unlabeled or partially labeled medical data.
- Federated learning frameworks: Ensure data privacy and security by enabling collaborative model training across hospitals without sharing sensitive patient data.
- Explainability enhancements: Use techniques like Grad-CAM, SHAP, or attention maps to make AI decisions more transparent and clinically interpretable.
- Dataset expansion: Include rare lung cancer types and multi-ethnic datasets to enhance model robustness and inclusiveness.
- Real-time clinical validation: Test the system in live hospital environments to refine performance and assess its utility in decision-making workflows.
- Deployment on edge devices or cloud: Optimize the model for mobile and cloud platforms to enable remote diagnostics and integration into telemedicine systems.
- Synthetic data generation: Use GANs (Generative Adversarial Networks) to augment datasets, especially for rare lung cancer subtypes and imbalanced classes.
- Collaboration with radiologists: Incorporate feedback loops from radiologists to continuously refine model performance and interface usability.
- Multi-task learning: Extend the model to perform related tasks like segmentation of tumors, localization of nodules, or severity scoring, all within a unified framework.

REFERENCES

[1] Shen, et al., "Deep Learning for Medical Image Analysis," Proc. IEEE,vol. 108, no. 1, pp. 30 48, 2020.
[2] Rajpurkar, et al., "DeepPneumonia: Automated Chest X-ray Diagnostics," Nature Medicine, vol. 24, pp. 1337-1342, 2019.
[3] Wang, et al., "Advanced CNN Architectures in Lung Nodule Classification," IEEE Trans. Medical Imaging, vol. 39, no. 3, pp. 789-799,2021.
[4] Dosovitskiy, et al., "An Image is Worth 16x16 Words: Transformers forImage Recognition," ICLR, 2021.
[5] Hatamizadeh, et al., "Transformers in Medical Image Segmentation,"Medical Image Analysis, vol. 75, pp. 102-115, 2022.
[6] Cao, et al., "Meta-Analysis of Deep Learning in Medical Imaging,"Journal of Biomedical Informatics, vol. 113, pp. 103-612, 2021.
[7] Shao, et al., "Ensemble Learning in Medical Diagnostics," IEEE Access,vol. 8, pp. 125689 125700, 2020.
[8] Liu, et al., "Vision Transformer Applications in Lung Cancer Detection,"Nature Scientific Reports, vol. 12, pp. 1-12, 2022.
[9] Zhang, et al., "Patch-Based ViT for Medical Image Classification,"MICCAI, pp. 234-242, 2022.
[10] Khan, et al., "Deep Learning in Cancer Detection: A Comprehensive Review," ACM Computing Surveys, vol. 54, no. 6, pp. 1-35, 2021.
[11] Shen, Y., Wang, J., Zhang, L., Zhao, M. (2022). Lung cancer detection using Vision Transformer with transfer learning on CT scan images. Journal of Biomedical Informatics, 129, 104061.
[12] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 10012–10022.
[13] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82–97.
[14] Bai, W., Chen, C., Tarroni, G., Duan, J., Guitton, F., Petersen, S. E., Rueckert, D. (2021). Self-supervised learning for cardiac MR image segmentation by anatomical position prediction. Medical Image Analysis, 68, 101856.
[15] Simonyan, K., Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

REFERENCES

[16] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[17] Tang, Y., Wang, S., Chen, M. (2022). Attention-based deep learning system for lung cancer detection and classification on chest X-rays. Computers in Biology and Medicine, 144, 105372.

[18] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. (2018). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 3–11.

[19]Li, X., Yu, L., Chen, H., Fu, C. W., Xing, L., Heng, P. A. (2020). Transformation-consistent self-ensembling model for semi-supervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 523–534.