LungSightNet: Deep Learning Driven Lung Cancer Prediction Using Compact Convolutional Transformers

by

Md. Al Mamunuzzaman Hredoy
ID: CSE1903018157

Md. Mosfiqur Rahman
ID: CSE2102023045

Md. Efrat Hossain
ID: CSE1901016060

Alamin
ID: CSE2201025016

Supervised by
Imran Hossen

Submitted in partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SONARGAON UNIVERSITY (SU)
January 2026

APPROVAL

The thesis titled "**LungSightNet: Deep Learning Driven Lung Cancer Prediction Using Compact Convolutional Transformers**" submitted by Md. Al Mamunuzzaman Hredoy (CSE1903018157), Md. Mosfiqur Rahman (CSE2102023045), Md. Efrat Hossain (CSE1901016060) and Alamin (CSE2201025016) to the Department of Computer Science and Engineering, Sonargaon University (SU), has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering and approved as to its style and contents.

Board of Examiners

Imran Hossen
Lecturer,
Department of Computer Science and Engineering
Sonargaon University (SU)

Supervisor

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 1

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 2

(Examiner Name and Signature)
Department of Computer Science and Engineering
Sonargaon University (SU)

Examiner 3

DECLARATION

We, hereby, declare that the work presented in this report is the outcome of the investigation performed by us under the supervision of **Imran Hossen, Lecturer,** Department of Computer Science and Engineering, Sonargaon University, Dhaka, Bangladesh. We reaffirm that no part of this thesis has been or is being submitted elsewhere for the award of any degree or diploma.

Countersigned

(Imran Hossen)
Supervisor

Signature

Md. Al Mamunuzzaman Hredoy
ID: CSE1903018157

Md. Mosfiqur Rahman
ID: CSE2102023045

Md. Efrat Hossain
ID: CSE1901016060

Alamin
ID: CSE2201025016

ABSTRACT

Accurate and interpretable medical image classification remains a critical challenge in computer-aided diagnosis, particularly under limited dataset conditions. Deep learning models often struggle to capture both local and global patterns simultaneously, and their black-box nature limits clinical trust. Incorporating attention mechanisms and explainability techniques can enhance both performance and interpretability. This research proposes a modified Self-Attention Compact Convolutional Transformer (SA-CCT) architecture designed to improve feature extraction and classification performance for lung cancer detection. The model integrates enhanced self-attention mechanisms and custom MLP blocks within the transformer framework, coupled with an improved patch-based tokenization strategy that captures richer local and global features from grayscale CT images. To ensure interpretability, Grad-CAM explainability and segmentation-based visualization modules are incorporated that enables spatial localization of discriminative regions and validation of model focus. The proposed SA-CCT model is trained and evaluated on a comprehensive lung cancer dataset that achieves 99% of classification accuracy, along with robust performance across per-class metrics and confusion analysis. These results show that the modified SA-CCT architecture is a very successful and easy to understand way to automatically diagnose lung cancer. It also provides a framework that may be used again and again for future research in medical image analysis and transformer-based categorization.

Keywords: Lung Cancer Classification, CT Image, SA-CCT, CNN-Transformer Hybrid, Self-Attention, Explainable AI, Grad-CAM, Convolutional Tokenization, Custom MLP Block, Sequence Pooling.

ACKNOWLEDGMENT

At the very beginning, we would like to express my deepest gratitude to the Almighty Allah for giving us the ability and the strength to finish the task successfully within the schedule time.

We are auspicious that we had the kind association as well as supervision of **Imran Hossen**, Lecturer, Department of Computer Science and Engineering, Sonargaon University whose hearted and valuable support with best concern and direction acted as necessary recourse to carry out our project.

We are also thankful to all our teachers during our whole education, for exposing us to the beauty of learning.

Finally, our deepest gratitude and love to my parents for their support, encouragement, and endless love.

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
CCT	Compact Convolutional Transformer
CNN	Convolutional Neural Network
CT	Computed Tomography
DeTraC	Decompose, Transfer, and Compose
DSU-Net	Distraction-Sensitive U-Net
FPSO	Fuzzy Particle Swarm Optimization
GGOs	Ground Glass Opacities
Grad-CAM	Gradient-weighted Class Activation Mapping
HoG	Histogram of Oriented Gradients
IQ-OTH/NCCD	Iraq-Oncology Teaching Hospital / National Center for Cancer Diseases
LBP	Local Binary Pattern
MAE	Mean Absolute Error
MHSA	Multi-Head Self-Attention
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
ReLU	Rectified Linear Unit
SA	Self-Attention
SA-CCT	Self-Attention Compact Convolutional Transformer
SGD	Stochastic Gradient Descent
SIFT	Scale-Invariant Feature Transform
UNet	U-shaped Neural Network
ViT	Vision Transformer
WHO	World Health Organization

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION TO LUNGSIGHTNET

1.1 Introduction

Lung cancer occurs when abnormal cells in the lungs grow and multiply uncontrollably, forming a tumor. Lung cancer continues to be a major global health challenge because of how common it is, how deadly it can be, and the urgent need for early detection. According to the World Health Organization (WHO), there were about 2.2 million new cases reported in 2020 [1], making it the most frequently diagnosed cancer in the world. The worrying frequency is increased by the terrible fact that lung cancer is the fore- most cause of cancer-related fatalities, accounting for nearly 2 million lives each year [2] The severity of the issue is compounded by the insidious advancement of lung cancer, frequently reaching advanced stages prior to the emergence of noticeable symptoms [3]. Approximately 70% of patients are diagnosed at an advanced stage [4], constraining treatment alternatives and yielding a discouraging five-year survival rate of only 19% [5]. This alarming number highlights the urgent necessity for advanced diagnostic methods that can identify lung cancer in its early stages, providing potential for enhanced treatment outcomes and reduced mortality rates [6]. Although the link between smoking and lung cancer is well-known, with about 85% of cases caused by tobacco use, it is important to note that non-smokers still represent a notable share of patients, roughly 15% [7], [8]. This detail shows how complicated the disease is, which is why predictive modeling needs to be done in a way that takes into account different patient traits. Lung cancer is more than just common; it's also the top cause of cancer-related deaths, taking the lives of about 2 million people every year [9]. Since lung cancer frequently advances silently and does not exhibit clinical symptoms until its later, more difficult stages, the importance of early identification becomes clear. The complicated development of lung cancer is influenced by genetic factors, environmental exposures, and lifestyle decisions, underscoring the need for sophisticated diagnostic instruments that can treat its various causes [10], [11]. Previous studies on lung cancer diagnosis have largely focused on models such as convolutional neural networks (CNNs) or hybrid architectures that emphasize local feature extraction from medical images. While these approaches effectively capture fine-grained patterns, they often neglect global contextual information, which is crucial for accurately characterizing complex tumor structures. To address the challenges of lung cancer diagnosis, this study investigates the application of the Compact Convolutional Transformer (CCT) model. Unlike conventional convolutional neural networks (CNNs), which excel at extracting local features but struggle with long-range dependencies, and standard transformers, which capture global context but often require large datasets and high computational resources, the CCT architecture offers a balanced approach. By integrating convolutional layers for robust local feature ex-

traction with transformer-based attention mechanisms for global context modeling, CCT effectively captures both fine-grained and holistic patterns in medical images. Utilizing the IQ-OTH/NCCD [12] dataset, the proposed approach achieves a classification accuracy of 99%. These results demonstrate the efficacy of CCT in lung cancer image analysis and underscore its potential to enhance the precision and reliability of diagnostic and prognostic systems.

1.2 Motivation

Deep learning has become the foundation of modern image analysis, but traditional convolutional networks often struggle to capture long-range spatial dependencies that are critical for complex visual reasoning. Transformer-based models overcome these limitations by modeling global relationships through self-attention, yet their high computational cost and data requirements make them difficult to deploy in practical scenarios [13].

The Self-Attention Compact Convolutional Transformer (SA-CCT) [14] provides a more efficient alternative by combining the strengths of convolutional tokenizers and lightweight transformer blocks. However, existing SA-CCT implementations are often rigid, optimized for general classification tasks, and do not fully exploit the potential of hybrid local-global feature extraction. To address these limitations, this research builds a modified SA-CCT architecture that enhances the tokenization stage with additional self- attention, expands the MLP capacity within transformer blocks, and refines the forward pipeline for improved feature learning under limited data conditions [13].

In many real-world applications, high performance alone is not sufficient. Sensitive domains require understandability that practitioners must understand why a model makes a particular prediction and which regions of the image contribute most. To support this requirement, the modified SA-CCT is integrated with Grad-CAM explainability and a segmentation-based visualization approach. This end-to-end system not only performs classification but also provides spatial localization cues, strengthening trust and offering deeper insights into model behavior [15], [16].

Finally, building a unified training, evaluation, and visualization pipeline allows the entire process from input preprocessing to interpretable prediction to be systematic, reproducible, and suitable for research-grade deployment. This motivates the development of a robust, efficient, and transparent modified SA-CCT framework for image classification.

1.3 Objectives

To guide the development and evaluation of the proposed modified SA-CCT architecture, this study sets forth the following key objectives:

- To develop an enhanced Self-Attention Compact Convolutional Transformer (SA-CCT) that strengthens both local and global feature extraction for lung cancer classification.
- To incorporate an improved convolutional tokenizer, additional attention pathways, and expanded custom MLP blocks to achieve richer representation learning in limited-data environments.
- To design a complete transformer-driven classification pipeline with optimized preprocessing, tokenization strategy, encoder configuration, training workflow, and evaluation using accuracy, confusion matrix interpretation, and detailed per-class performance metrics.
- To integrate a comprehensive explainability and visualization framework by combining Grad-CAM applied to attention-related components with a segmentation-assisted module for spatial localization of discriminative regions.

1.4 Research Questions

This research is guided by the following key research questions, formulated to address the challenges of accurate and interpretable lung cancer classification from CT images using deep learning techniques:

1. RQ1:
   How can a compact hybrid CNN–Transformer architecture be designed to effectively capture both local and global features from lung CT images under limited data conditions?
2. RQ2:
   In what ways can convolution-based tokenization and customized MLP blocks enhance feature representation learning within a Self-Attention Compact Convolutional Transformer (SA-CCT) framework?
3. RQ3:
   How does the incorporation of self-attention mechanisms and sequence pooling improve classification performance compared to traditional CNN-based approaches?
4. RQ4:
   How can explainability techniques such as Grad-CAM and segmentation-based visualization be integrated into the model to reduce black-box behavior and improve interpretability for clinical decision support?

5. RQ5:
   To what extent does the proposed SA-CCT framework achieve reliable and balanced performance across different lung cancer classes as evaluated through accuracy, confusion matrix analysis, and per-class metrics?

1.5 Contributions and Impacts

This thesis contributes to the advancement of medical image analysis by addressing the limitations of existing deep learning approaches in terms of feature representation, data efficiency, and interpretability. By integrating convolutional and transformer-based learning paradigms within a compact and explainable framework, the proposed work provides both methodological and practical advancements for lung cancer classification using CT images. The following subsections summarize the key research contributions and the broader impacts of this study.

1.5.1 Research Contributions

The primary research contributions of this thesis are focused on architectural innovation, optimized learning strategies, and enhanced interpretability within transformer-based medical imaging models. Specifically, the contributions are as follows:

1. **A Modified SA-CCT Architecture:**

   A compact and enhanced Self-Attention Compact Convolutional Transformer (SA-CCT) is proposed, specifically tailored for lung cancer classification using grayscale CT images under limited dataset conditions.

2. **Improved Convolutional Tokenization Strategy:**

   An enhanced convolution-based tokenization approach is introduced to preserve local texture and structural information, enabling more effective patch representation compared to fixed patch embedding methods.

3. **Custom MLP Block Design:**

   Customized MLP blocks are integrated within the transformer encoder to improve non-linear feature learning while maintaining model compactness and reducing overfitting.

4. **Sequence Pooling-Based Feature Aggregation:**

   A sequence pooling mechanism is employed instead of a conventional classification token to achieve more robust and informative global feature aggregation.

5. **Integrated Explainability Framework:**

   The model incorporates Grad-CAM and segmentation-assisted visualization techniques to provide spatial interpretability and validate the regions influencing model predictions.

1.5.2 Practical and Research Impacts

Beyond architectural contributions, this research demonstrates meaningful practical and scientific impacts by enhancing the reliability, transparency, and applicability of deep learning models in medical diagnostics. The proposed framework bridges the gap between high-performance classification and clinical interpretability, thereby supporting real-world adoption and future research development.

- **Clinical Impact:**

  The interpretable design enables clinicians to visualize discriminative lung regions, increasing trust and usability in computer-aided diagnosis systems.

- **Technical Impact:**

  The compact SA-CCT architecture demonstrates that transformer-based models can be effectively adapted for medical imaging tasks with limited data availability.

- **Research Impact:**

  The proposed framework serves as a reproducible and extensible baseline for future research in medical image classification, explainable AI, and hybrid CNN–Transformer architectures.

- **Future Applicability:**

  The methodology can be extended to other medical imaging modalities and disease classification tasks, supporting broader adoption in healthcare AI systems.

1.6 Limitations

Despite the promising performance and interpretability of the proposed SA-CCT framework, this study has several limitations that should be acknowledged. These limitations primarily arise from data availability, experimental scope, and architectural constraints inherent to transformer-based medical image analysis.

First, the model was trained and evaluated on a limited-size lung CT dataset, which may restrict its ability to generalize to larger and more diverse patient populations. Although data preprocessing and mild augmentation were applied to mitigate this issue, the performance of the model may vary when applied to datasets collected from different institutions or imaging protocols.

Second, the study focuses exclusively on grayscale CT images resized to $32 \times 32 \times 1$, which, while computationally efficient, may lead to the loss of fine-grained anatomical details present in higher-resolution scans. As a result, subtle pathological patterns might not be fully captured.

Third, the proposed framework addresses classification tasks only and does not perform end-to-end lesion segmentation or precise tumor boundary delineation. While segmentation-assisted visualization was used to enhance interpretability, a fully supervised segmentation model was beyond the scope of this work.

Fourth, the experiments were conducted using a single computational environment (Google Colaboratory) and evaluated using a fixed set of hyperparameters. Further optimization and cross-platform evaluation may improve robustness and performance.

Finally, although explainability mechanisms such as Grad-CAM were integrated to reduce black-box behavior, these visualization techniques provide approximate interpretations and should not be considered definitive clinical evidence without expert validation.

1.7 Organization of Thesis Book

This thesis is organized into several structured chapters that collectively present the development, evaluation, and interpretation of the proposed Self-Attention Compact Convolutional Transformer (SA-CCT) framework for lung cancer detection. Chapter 1 introduces the research background, motivation, problem statement, objectives, research questions, contributions, and overall scope of the study. Chapter 2 reviews relevant literature on lung cancer diagnosis, convolutional neural networks, vision transformers, compact transformer architectures, and explainable artificial intelligence techniques. Chapter 3 describes the proposed methodology in detail, including dataset description, preprocessing pipeline, SA-CCT model architecture, training strategy, and explainability framework. Chapter 4 presents the experimental design and evaluation process, covering implementation details, training configuration, performance metrics, baseline model comparison, and ablation analysis. Chapter 5 discusses the experimental results, analyzes model performance, and evaluates the effectiveness of the proposed approach in comparison with existing methods. Finally, Chapter 6 concludes the thesis by summarizing key findings, highlighting limitations, and outlining potential directions for future research.

CHAPTER 2

RELATED WORKS

To situate our work within the broader scientific landscape, it is essential to examine the research foundations that have shaped the current understanding of lung cancer detection and deep learning driven diagnostic systems.

In the study, A. Asuntha et. al. [2] developed a system that can find cancerous spots (malignant nodules) in lung images and determine how serious the lung cancer is. They used advanced deep learning methods to detect these dangerous nodules and picked some of the best ways to pull out important details from the images, such as Zernike Moments, Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), Wavelet features, and Histogram of Oriented Gradients (HoG). After pulling out different kinds of image details (texture, shape, size, and brightness), the researchers use a smart selection method called Fuzzy Particle Swarm Optimization (FPSO) to pick only the most useful ones. Then, a deep learning system classifies whether the nodule is cancerous and how serious it is. To make the process faster and lighter, they created a new version called FPSOCNN that combines FPSO with a convolutional neural network (CNN). Their final model achieved an accuracy of 94.67%.

Ibrahim M. Nasser et. al. [11] created a smart computer program (an Artificial Neural Network) that can detect lung cancer just by checking everyday symptoms—things like yellow-stained fingers, feeling anxious or worn out all the time, chronic coughing, wheezing, trouble breathing or swallowing, chest pain, and allergies. They combined these clues with some personal details about the person and let the AI figure out if lung cancer is present. Their artificial neural network (ANN) was developed, trained, and validated using the ”survey lung cancer” data set. The model evaluation revealed that the ANN model had a 96.67% accuracy rate in detecting the presence or absence of lung cancer.

P. Mohamed Shakeel et al. [5] introduced improved image processing and machine learning techniques to predict lung cancer. Using a non-small cell lung cancer CT dataset, they clean the images, then apply an enhanced deep neural network to segment the tumor area and extract features. Similar advanced deep learning methods have also been widely used for automated COVID-19 diagnosis.

S.A. Banday et al. [17] proposed a morphological reconstruction approach to enhance the segmentation and detection of Ground Glass Opacities (GGOs) in CT lung images. Using an Artificial Neural Network (ANN), their method significantly improved the specificity of COVID-19 detection.

Maiello et al. [18] built a 3D UNet that can look at lung scans and tell doctors exactly how much healthy lung is left, where the air is trapped, and how much of the lung might open up with treatment. Their work clearly pointed out that lungs have very complicated structures, so today's segmentation tools still need some fine-tuning to get everything just right.

Junting Zhao et. al. [19] came up with a clever two-stage UNet called DSU-Net (short for Distraction-Sensitive U-Net). It is designed to ignore all the unimportant stuff in lung CT images and focus only on the key regions that doctors need to see.

At the same time, Mohammad Hamid Asnawi et. al. [20] tested several versions of the 3D UNet model e.g., 3D Res-uNet, 3D VGG-uNet, and 3D Dense-uNet to see which one works best for accurately outlining the lungs in CT scans. Their work showed that there are many different ways to build these models, and each has its own strengths. They also pointed out how important it is to carefully choose the right image features so the model performs at its best.

He Ma et. al. [21] built a Mask region-based CNN (Mask R-CNN) specially designed for 3D medical images, and it worked really well for detecting problems in scans. Separately, Asmaa Abbas et. al. [22] created a model called DeTraC (short for Decompose, Transfer, and Compose) that looks at chest X-rays to spot COVID-19. It turned out to be very good at telling COVID-19 apart from other lung diseases, showing a lot of promise for real-world use.

S. Guan et al. [23] proposed the FM-HCF-DLF model, which fuses handcrafted features and deep learning after Gaussian filtering pre-processing, achieving a high accuracy of 94.08% for COVID-19 classification.

These studies demonstrate a wide range of techniques from morphological reconstruction and segmentation models to machine learning and advanced graph-based neural networks for automated lung cancer and COVID-19 diagnosis. Despite significant progress, difficulties in precise segmentation, effective feature extraction, and model interpretability remain, highlighting the need for ongoing research to create more robust and reliable diagnostic tools.

CHAPTER 3

METHODOLOGY

In this research, we developed a modified Self-Attention Compact Convolutional Trans- former (SA-CCT) model for robust lung cancer prediction using CT images. The model integrates enhanced convolutional tokenization, strengthened self-attention pathways, and expanded transformer MLP blocks to improve the extraction of both local and global features under limited-data conditions. To evaluate the effectiveness of the proposed architecture, its performance was compared with several established deep learning models, including InceptionV3, ResNet152, VGG19, and MobileNetV2. The modified SA-CCT demonstrated superior performance across all metrics, achieving 99% classification accuracy on the IQ-OTH/NCCD [12] lung cancer dataset.

3.1 Dataset

The experiments were carried out using the IQ-OTH/NCCD [12] lung cancer dataset. This dataset contains CT images across three classes: benign, malignant, and normal. The dataset consists of grayscale Computed Tomography (CT) scan images representing different lung conditions, including cancerous and non-cancerous cases. CT imaging is widely used in clinical practice due to its ability to provide detailed cross-sectional views of lung tissues, making it suitable for detecting nodules and abnormal structures associated with lung cancer. All images in the dataset are stored in standard digital image file formats JPG, which preserve pixel-level intensity information required for medical image analysis. Since CT images are inherently grayscale, each image contains a single intensity channel, representing tissue density variations within the lung region. Prior to model training, all images are resized to a uniform dimension of $32 \times 32 \times 1$ to ensure consistency in input shape and computational efficiency.

The dataset exhibits class imbalance, a common issue in medical imaging datasets, where certain classes contain fewer samples than others. Overall, the dataset provides a suitable foundation for evaluating the effectiveness of the proposed SA-CCT framework in lung cancer classification. Each image varies in size and intensity, reflecting real clinical diversity. Before training, all samples were resized and normalized for consistency.

Table 3.1: Dataset Summary

Class	Number of Images
Benign	120
Malignant	561
Normal	416

To prepare the images for model training, all samples were resized, normalized, and underwent intensity scaling. Mild data augmentation e.g., rotation, zooming, and contrast variation was applied to enhance generalization and help address class imbalance. The dataset was divided into training, validation, and testing subsets using an 80:10:10 split.

3.2 Architectural Design

The architectural design of the proposed SA-CCT model follows a structured six-phase workflow, as illustrated in Figure 3.1 The process begins with data collection and preprocessing, leading into the core model architecture which combines convolutional tokenizers with transformer encoders to capture both local and global image features. This pipeline concludes with a high-accuracy evaluation and the integration of Grad-CAM heatmaps to provide clear, visual explanations for each clinical prediction.

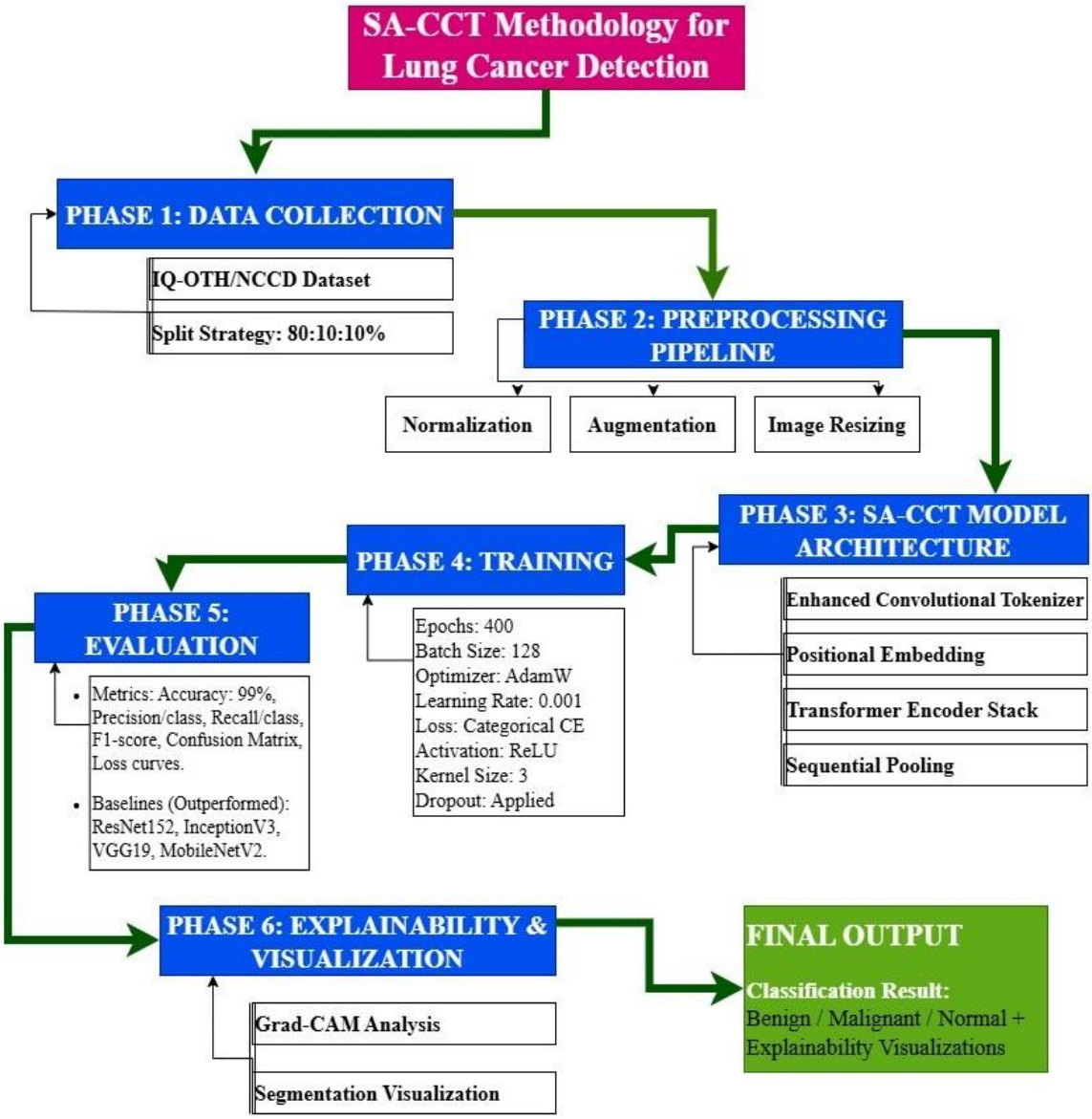

Figure 3.1: SA-CCT Methodology for Lung Cancer Detection

3.3 SA-CCT Model

The Self-Attention Compact Convolutional Transformer (SA-CCT) serves as the core architecture of this research due to its ability to combine convolutional feature extraction with global dependency modeling. Conventional CNNs, which are limited to localized receptive fields, and standard transformers, which require large datasets and fixed patch embedding's, SA-CCT integrates the strengths of both approaches while significantly reducing computational overhead. The model illustrated in Figure 3.2 begins with a convolutional tokenizer that converts spatially rich CT images into

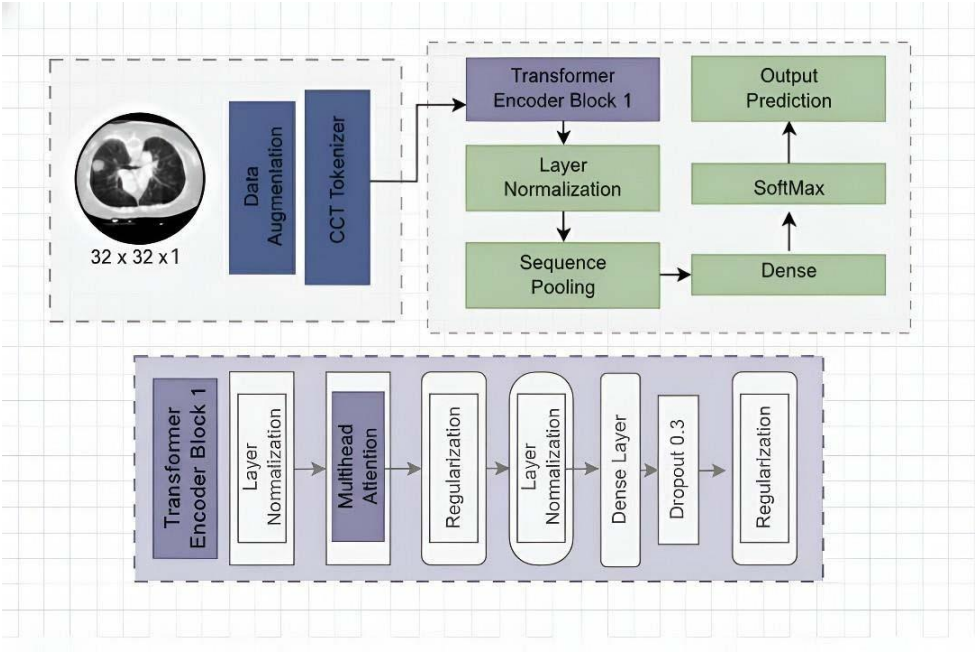

Figure 3.2: SA-CCT Model

flexible latent tokens. These tokens are then processed by self-attention layers that capture wide-range contextual relationships, which is an essential requirement for accurate lung abnormality detection, where benign and malignant regions may differ subtly in shape, density, and boundary textures.

Furthermore, SA-CCT introduces an additional self-attention pathway inside the tokenizer, along with deeper MLP expansion in each transformer block. These enhancements enable richer hierarchical representations, improving the model's capacity to differentiate malignant nodules from benign formations and normal tissue.

Compared with baseline architectures e.g., InceptionV3, ResNet152, VGG19, and Mo- bileNetV2, SA-CCT operates with fewer parameters, faster training cycles, and greater robustness to dataset imbalance. This hybrid mechanism allows the model to effectively process grayscale CT images with variable anatomical structures and subtle lesion pat- terns.

3.4 Proposed Optimized SA-CCT Model

The optimized SA-CCT architecture consists of several interconnected modules designed to enhance spatial encoding, token quality, attention precision, and pooled representation strength. The overall design integrates convolutional tokenization, positional embedding, multi-head self-attention, MLP expansion, sequential pooling, and dense classification.

3.4.1 Convolutional Tokenization

The tokenization block transforms the input image of size $H \times W \times C$ into a compact sequence of length m, enabling the transformer to process spatial information as learnable token embedding. The enhanced tokenizer includes:

- A Conv2D layer with 64 filters,
- A stride of 2 to create overlapped convolutional patches,
- MaxPooling to retain prominent structural features,
- ReLU activation to introduce nonlinearity,
- An additional internal attention pathway to strengthen early-stage global feature propagation.

This design eliminates the rigid patch-splitting step used in ViT and instead employs learned convolutional filters to extract meaningful representations. Such flexibility is particularly valuable in medical imaging, where lesion characteristics do not conform to uniform patch boundaries.

3.4.2 Transformer Encoder with Expanded MLP Blocks

The resulting token sequence is passed into a stack of transformer layers, each consisting of:

- Multi-Head Self-Attention (MHSA), which captures dependencies between distant lung regions,
- Layer Normalization for stable gradients and training consistency,
- ReLU activation to enhance feature smoothing in medical textures,
- Dropout to reduce overfitting in limited medical datasets.

To further improve representational strength, the MLP block within each transformer layer is expanded with wider hidden dimensions, enabling:

- deeper nonlinear transformations,
- stronger channel-wise mixing,
- improved discriminatory capability.

This expansion allows SA-CCT to learn complex variations in radiological density, texture roughness, nodule shape, and inter-lobe differences that factors critical for accurate differentiation between malignant and benign lesions.

3.4.3 Sequential Pooling Mechanism

Instead of relying on a class token, the proposed model incorporates a sequential pooling strategy that dynamically aggregates relevant information across all tokens. This mechanism allows to emphasize the most informative regions of each CT image.

The pooling transformation is defined as:

$$T : \mathbb{R}^{bxnxd} \rightarrow \mathbb{R}^{bxd} .$$

The process begins with the tokenized input:

$$x_L = f(x_0) \in \mathbb{R}^{bxnxc} \quad (3.1)$$

Where x_L represents the output of the final transformer encoder.

Next, a learned projection function $g(\cdot)$ generates token importance weights:

$$x'_L = \text{softmax } (g(x_L)^T) \in \mathbb{R}^{bx1xn} \quad (3.2)$$

Finally, sequential pooling computes a weighted aggregation over all tokens:

$$Z_L = x'_L \times x_L = \text{softmax } (g(x_L)^T)\, x_L \in \mathbb{R}^{bx1xn} \quad (3.3)$$

Where:

- x_0: Tokenized input sequence,
- x_L: Final transformer layer output,
- $g(\cdot)$: Linear projection generating attention weights,
- z: Final pooled embedding used for classification.

Advantages of Sequential Pooling: Sequential pooling provides two primary benefits:

1. It automatically highlights the most discriminative tokens, such as tumour boundaries or regions with abnormal density.
2. It produces a richer global representation compared to average pooling or class- token-based methods.

3.4.4 Classification Layer

The final pooled embedding z is flattened and passed into a fully connected dense layer with a softmax activation function to generate class probabilities for the three categories:

- Benign,
- Malignant,
- Normal.

This classification layer completes the forward pipeline by mapping high-level feature representations into clinically meaningful diagnostic outputs.

3.4.5 Explainability and Visualization

In medical deep learning applications, the ability to interpret how a model arrives at its decision is just as important as achieving high predictive accuracy. Clinicians require transparent systems that not only classify disease patterns correctly but also justify their decisions in a visually meaningful way. To address this need, the proposed SA-CCT frame- work integrates a comprehensive explainability module composed of two complementary visualization techniques: Grad-CAM heatmap generation and a segmentation-based localization mechanism. Together, these methods provide clear and interpretable evidence of the model's internal reasoning.

3.4.6 Grad-CAM Heatmap Visualization

Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to the final self- attention layers of the SA-CCT model to highlight spatial regions that contribute most strongly to the predicted class. By backpropagating gradients from the classification output to the attention-weighted feature maps, Grad-CAM produces a heatmap overlay that visually identifies areas the model considered most relevant.

The generated heatmaps reveal several important interpretative behaviors:

- **High activation on malignant nodules:** Regions containing suspicious masses or high-density abnormalities show significantly elevated activation, confirming the model's ability to detect malignant characteristics.
- **Focused attention on abnormal texture patterns:** The model consistently emphasizes irregular tissue structures, ground-glass opacities, and heterogeneous pixel intensities commonly associated with early cancer indicators.
- **Minimal activation on irrelevant regions:** Non-diagnostic areas such as the rib cage, vacuum regions, or background noise remain largely inactive, demonstrating the model's selective and meaningful feature prioritization.

These heatmaps serve as an intuitive visual explanation for each prediction, allowing clinicians to verify whether the model's focus aligns with established radiological knowledge. They also help identify potential misclassifications by revealing whether the model was distracted by non-relevant features.

3.4.7 Segmentation-Based Localization Module

To further enhance interpretability, a lightweight segmentation-based visualization module is incorporated alongside Grad-CAM. This module outlines and highlights the anatomical structures and suspicious regions within the CT image, providing a more precise spatial interpretation of potential lesions.

3.4.8 Integrated Purpose of Explainability

The integration of Grad-CAM and segmentation-based localization creates a robust interpretability framework. Together, they accomplish the following:

- Strengthen clinician trust by offering transparent, visually grounded justifications for each prediction.
- Provide direct evidence supporting the diagnostic output, allowing cross-verification with radiological expertise.
- Reduce the black-box nature of transformer-based architectures by exposing internal decision pathways.
- Enhance clinical decision-making through interpretable, anatomically relevant visual cues.

Overall, the explainability module ensures that the modified SA-CCT model is not only accurate but also interpretable, making it more suitable for real-world clinical application where transparency and diagnostic reliability are essential.

CHAPTER 4

EXPERIMENTAL DESIGN AND EVALUATION

This chapter describes in detail of the experimental framework adopted to assess the performance, robustness, and interpretability of the proposed Self-Attention Compact Convolutional Transformer model for lung cancer prediction using computed tomography images. The experimental design includes dataset preparation, preprocessing, model training configuration, baseline model comparison, evaluation metrics, and explainability validation. All experiments were conducted with a standardized and reproducible pipeline to ensure consistency and fairness across model comparisons.

4.1 Tools, Frameworks, and Computational Environment

We implemented our model using Python and TensorFlow on Google Colab, utilizing an NVIDIA Tesla T4 GPU, and ensured reproducibility through consistent preprocessing, fixed configurations, and standard open-source libraries. This section presents the software tools, frameworks, and computational resources used to implement, train, and evaluate the proposed SA-CCT model. Clearly documenting these components ensures experimental transparency, reproducibility, and consistency across different research environments.

4.1.1 Hardware and GPU Configuration

To accelerate model training and experimentation, all experiments were conducted using Google Colaboratory, which provides access to high-performance GPU resources. The SA-CCT model was trained using GPU acceleration to reduce training time and handle computationally intensive transformer operations efficiently. Table 4.1 summarizes the computational environment and hardware configuration.

Table 4.1: Computational Environment and Hardware Configuration

Component	Specification
Platform	Google Colaboratory
GPU	NVIDIA Tesla T4
GPU Memory	16 GB
CPU	Intel Xeon (Colab-provided)
RAM	Approximately 12–16 GB
Operating System	Windows (Colab environment)

4.1.2 Software Tools and Frameworks

The proposed SA-CCT architecture was implemented using widely adopted Python-based deep learning libraries that support efficient model development, training, evaluation, and visualization. The selected tools provide flexibility for transformer-based architectures and explainability integration. Table 4.2 summarizes the software tools used in this research.

Table 4.2: Tools and Frameworks Used in This Thesis

Category	Tool / Framework	Purpose
Programming Language	Python	Core language used for model implementation and experimentation
Deep Learning Framework	TensorFlow / Keras	Building, training, and evaluating the SA-CCT model
Numerical Computing	NumPy	Efficient numerical operations and tensor manipulation
Data Handling	Pandas	Dataset handling and experimental result organization
Image Processing	OpenCV / PIL	Image loading, resizing, and preprocessing
Visualization	Matplotlib, Seaborn	Plotting training curves, confusion matrices, and results
Explainability	Grad-CAM	Visual interpretation of model attention and predictions
Development Environment	Google Colaboratory	Cloud-based notebook environment for implementation and training

4.2 Experimental Design

Experiments were done using the IQ-OTH/NCCD lung cancer dataset, which is composed of three classes: benign, malignant, and normal, for a total of 1,097 CT images. Table 4.3 summarizes the characteristics and distribution of the dataset. The dataset was split into the following to allow training and assessment effectively: Training set: 80% Validation set: 10% Testing set: 10%

This split will ensure that the model is evaluated on unseen data, while the validation set remains stable for monitoring convergence.

Table 4.3: IQ-OTH/NCCD Dataset Distribution

Class	Number of Images
Benign	120
Malignant	561
Normal	416
Total	**1,097**

4.3 Preprocessing Pipeline

To ensure consistent data distribution and enhance the generalization capability of the model, all CT images underwent the following pre-processing steps:

- **Image Resizing:** All CT images were resized to 32 × 32 pixels in accordance with the optimized input requirements of the SA-CCT model.
- **Normalization:** Pixel intensities were scaled to the range [0, 1] to stabilize the training dynamics.
- **Data Augmentation:** Mild augmentation techniques were incorporated to mitigate class imbalance and improve robustness, including:
  - Rotation within $\pm 15^{\circ}$
  - Random zooming (5–10%)
  - Slight contrast jitter

These augmentations preserved anatomical integrity while increasing sample diversity during training.

4.4 Training Configuration

The optimized SA-CCT model was trained using the hyper parameters determined in Table 4.4 The final training configuration is summarized as follows:

Table 4.4: Training Configuration for the SA-CCT Model

Parameter	**Value**
Epochs	400
Batch Size	32
Optimizer	AdamW
Learning Rate	1e-4
Loss Function	Categorical Crossentropy
Activation Function	ReLU
Kernel Size	3
Dropout	Applied after transformer attention and MLP layers (0.1)
Sequential Pooling	Used in place of a class token to accumulate token embeddings

Training was conducted in a high-performance GPU environment to ensure computational efficiency and to support the model's transformer-based architecture.

4.5 Baseline Model Comparison

To rigorously assess the effectiveness of the proposed SA-CCT model, several established deep learning architectures were implemented as baselines:

- ResNet-152
- InceptionV3
- VGG19
- MobileNetV2

All baseline models were trained under identical experimental conditions, including the same dataset split, preprocessing pipeline, loss function, and comparable training epochs. This ensured a fair and controlled comparison.

The evaluation focused on demonstrating improvements in:

- Classification accuracy
- Robustness across classes
- Model efficiency with limited training data
- Ability to capture both global and localized CT features

4.6 Evaluation Metrics

A comprehensive set of evaluation metrics was employed to assess model performance from multiple perspectives:

1. **Accuracy:** Measures overall prediction correctness across all classes.
2. **Precision, Recall, and F1-score:** Capture per-class diagnostic performance, particularly essential in medical imaging where minimizing false negatives is critical.
3. **Confusion Matrix:** Provides insights into:
    - Misclassification tendencies
    - Overlap between benign and malignant nodules
    - Model bias toward majority classes
4. **Loss and Accuracy Curves:** Training and validation plots were analysed to evaluate convergence and detect signs of overfitting.

4.7 Explainability

Interpretability is a critical requirement for deploying deep learning models in medical decision-support systems. In clinical settings, it is not sufficient for a model to achieve high classification accuracy; the model must also provide transparent and interpretable explanations that justify its predictions. To address the black-box nature of deep learning models, this study integrates explainability mechanisms that enable visualization of the spatial regions contributing to the classification decisions of the proposed SA-CCT framework. We integrated Grad-CAM and segmentation-assisted visualization to validate that the model focuses on clinically relevant lung regions, reducing black-box behavior. The explainability components were designed to validate whether the model's attention aligns with clinically relevant lung regions and pathological structures observed in CT images. Two complementary visualization techniques Grad-CAM analysis and segmentation-assisted visualization were employed to ensure reliable and anatomically meaningful interpretations.

4.7.1 Grad-CAM Analysis

Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to generate heatmap visualizations from the final attention-related layers of the SA-CCT model. Grad-CAM computes the gradient of the predicted class score with respect to feature maps, allowing the identification of spatial regions that contribute most strongly to the model's decision.

In this study, Grad-CAM heatmaps were overlaid on the original CT images to visualize:

- Lung regions influencing the classification decision
- Nodule boundaries associated with malignant or suspicious areas
- Dense or abnormal tissue patterns relevant to lung pathology

The generated heatmaps consistently highlighted anatomically meaningful regions within the lung fields, demonstrating that the SA-CCT model learned to focus on clinically relevant features rather than irrelevant background information. This confirms that the self-attention mechanism effectively captures global contextual dependencies while preserving important local structures.

4.7.2 Segmentation-Assisted Visualization

To further enhance interpretability and provide anatomical context, a lightweight segmentation-assisted visualization module was incorporated into the framework. Segmentation-assisted visualization provides anatomical boundaries that help validate attention regions and improve clinical interpretability. This module overlays segmentation-based masks on the CT images to support spatial interpretation of attention-guided regions.

The segmentation-assisted visualization was used to:

- Highlight suspicious lesions and abnormal lung regions
- Provide anatomical boundaries for clearer spatial understanding
- Align Grad-CAM attention maps with radiologically meaningful structures

By combining segmentation overlays with Grad-CAM heatmaps, the framework enables more intuitive interpretation of the model's predictions. This dual visualization approach bridges the gap between algorithmic attention and clinical anatomy, making the model's behavior more transparent to clinicians and researchers.

CHAPTER 5

RESULT ANALYSIS

5.1 Evaluation Methods

The performance of the proposed SA-CCT model was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These measures were calculated from the confusion matrix, where true positives (TP) represent correctly classified samples for a given class, false positives (FP) represent misclassified negative samples, and false negatives (FN) represent missed positive detections. True negatives (TN) indicate correctly rejected samples. The metrics were computed using the following definitions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5.2 Results and Discussion

The experimental results illustrate the strong diagnostic capability of the optimized SA- CCT architecture. The classification report demonstrates an overall accuracy of 99%, with the malignant class achieving a perfect recall (1.00), the benign class achieving full recall (1.00), and the normal class demonstrating high consistency with a 0.96 recall score. The weighted F1-score of 0.99 indicates balanced performance across all classes.

The confusion matrix (Figure 5.2) further supports this outcome, where malignant and benign cases show no false negatives, and normal cases present only minimal misclassifications. The model reliably distinguishes structural differences in CT images across all three classes.

The training and validation curves show smooth convergence. Training accuracy reached near 100%, while validation accuracy stabilized between 97% and 99%. Training loss dropped steadily, and validation loss remained consistently low, demonstrating limited overfitting and robust generalization.

Explainability modules supported these findings. Grad-CAM (Figure 5.3) heatmaps highlighted radiologically meaningful regions such as nodule centers, high-density tissue zones, and abnormal texture areas. Segmentation overlays confirmed that the model consistently focused on the lung parenchyma while disregarding irrelevant surroundings. This alignment with anatomical structures strengthens clinical interpretability and model trust.

Table 5.1: Classification Report

Class	**Precision**	**Recall**	**F1-score**	**Support**
Normal cases	1.00	0.96	0.98	77
Malignant cases	0.99	1.00	1.00	116
Benign cases	0.93	1.00	0.96	27
Accuracy		0.99		220
Macro avg	0.97	0.99	0.98	220
Weighted avg	0.99	0.99	0.99	220

5.3 Baseline Model Comparison

To show the effectiveness of the proposed optimized SA-CCT model, its performance was compared with several well-known deep learning models that have been widely used in medical image classification. These models include InceptionV3, ResNet152, MobileNetV2, and VGG19, and their results were taken from existing published studies. In this work, these baseline models were not re-trained. Instead, their reported testing accuracies were used as reference values to compare with the performance of the proposed SA-CCT model. Table 5.2 presents this comparison.

Table 5.2: Performance comparison of referenced baseline models and the proposed SA-CCT model

Model	**Image Size**	**Epochs**	**Testing Accuracy (%)**
InceptionV3 [24]	224×224	100	88.66
ResNet152 [25]	224×224	100	91.41
MobileNetV2 [26]	224×224	100	89.84
VGG19 [27]	224×224	100	94.53
Proposed Optimized SA-CCT	32×32	400	**99.09**

The comparison reveals several key findings:

- The proposed SA-CCT model achieved the highest accuracy (99.09%), outperforming all baseline models by a significant margin.
- Despite using a smaller input size (32×32) compared to the baselines (224×224), the SA-CCT model effectively captured rich local and global features through its hybrid convolutional tokenizer and transformer blocks.
- ResNet152 and MobileNetV2 delivered moderate performance; however, both were inferior to VGG19 and SA-CCT.
- VGG19 achieved the highest accuracy among the CNN-based baselines (94.53%), yet it still lacked the sensitivity and discriminative power demonstrated by SA-CCT.

Overall, this comparison highlights that transformer-based architectures particularly enhanced with convolutional tokenizers that offer superior performance in medical imaging tasks where fine-grained structural information must be preserved.

5.4 Confusion Matrix, Accuracy Curve and Loss Curve

The confusion matrix (Figure 5.1) illustrates clean class separation with negligible misclassification across all categories. Both malignant and benign classes exhibit perfect recall, ensuring that no clinically significant cases are overlooked. This high recall is particularly important in medical diagnostics, where missed detections can lead to severe clinical risks.

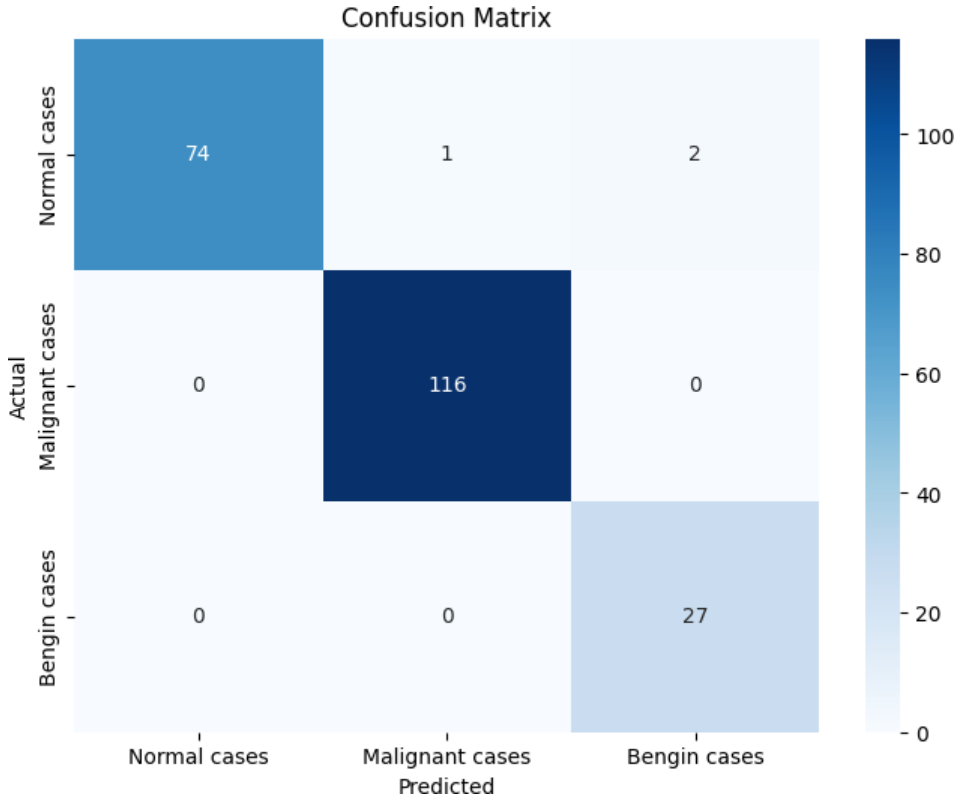

Figure 5.1: Confusion Matrix

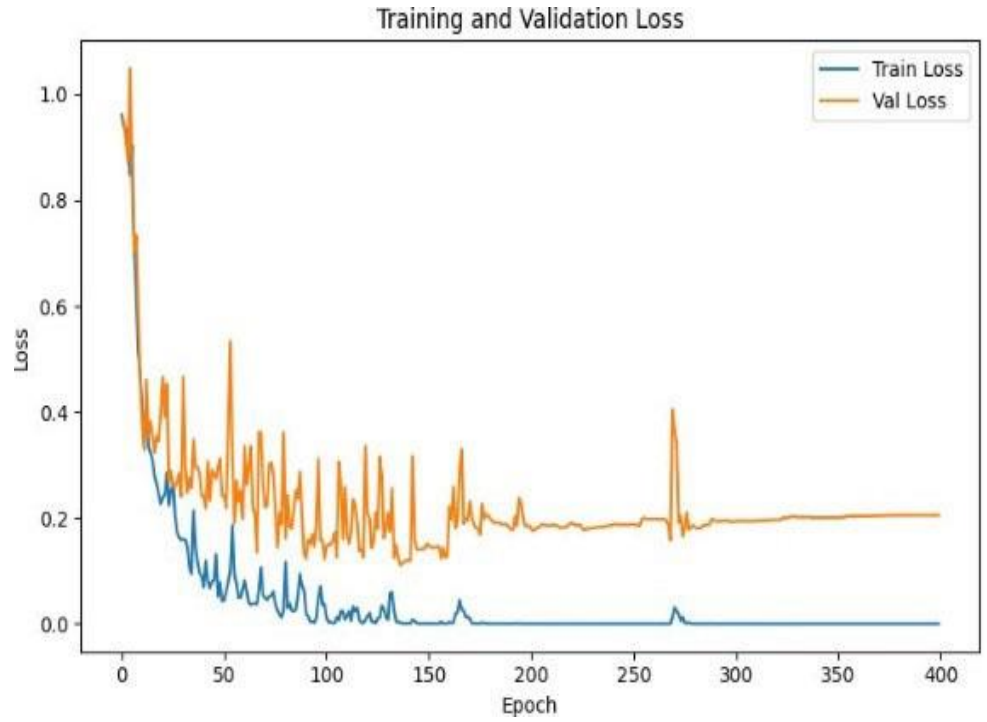

Figure 5.2: Training and Validation Loss

Figure 5.3: Training and Validation Accuracy

The training and validation accuracy curves (Figure 5.2) show smooth and consistent improvement, while the corresponding loss curves demonstrate stable convergence without significant divergence between training and validation losses. These trends confirm that the SA-CCT model learns effectively and maintains strong generalization capability on the testing dataset.

Furthermore, Grad-CAM (Figure 5.4) visualizations and segmentation overlays highlight the model's ability to focus on clinically meaningful lung regions, such as tumor boundaries, abnormal tissue densities, and structural irregularities. These visual explanations demonstrate that the model avoids irrelevant artifacts or background structures, reinforcing its interpretability and suitability for clinical applications.

5.5 Explainability and Segmentation Analysis

Explainability is essential in medical diagnosis, where the transparency of model behavior directly influences clinical trust and adoption. To meet this requirement, the proposed SA- CCT model incorporates a two-stage interpretability framework consisting of: (1) Grad- CAM–based visual explanation and (2) segmentation-assisted visualization. Together, these methods enable direct inspection of the model's decision-making process and allow clinicians to verify that predictions correspond to meaningful anatomical and radiological features.

5.5.1 Grad-CAM based Visual Explaination

Grad-CAM heatmaps were generated from the final attention layers of the SA-CCT model to highlight spatial regions that most strongly influenced each classification output. The visual analysis revealed several key patterns:

- **Malignant samples**: High-intensity activations concentrated around dense, irregular structures typical of malignant tumours.
- **Benign samples**: Moderate activation localized near smooth, low-density abnormalities consistent with non-aggressive lesion morphology.
- **Normal samples**: Minimal and diffuse activation, indicating that the model does not incorrectly identify pathological regions.

Across all classes, the Grad-CAM heatmaps were anatomically coherent rather than random or noisy. The highlighted regions consistently mapped to clinically relevant radiological cues. These observations confirm that the SA-CCT architecture effectively integrates local and global spatial information through its tokenization and attention mechanisms. As a result, the model produces interpretable attention maps that correspond to meaningful radiological features.

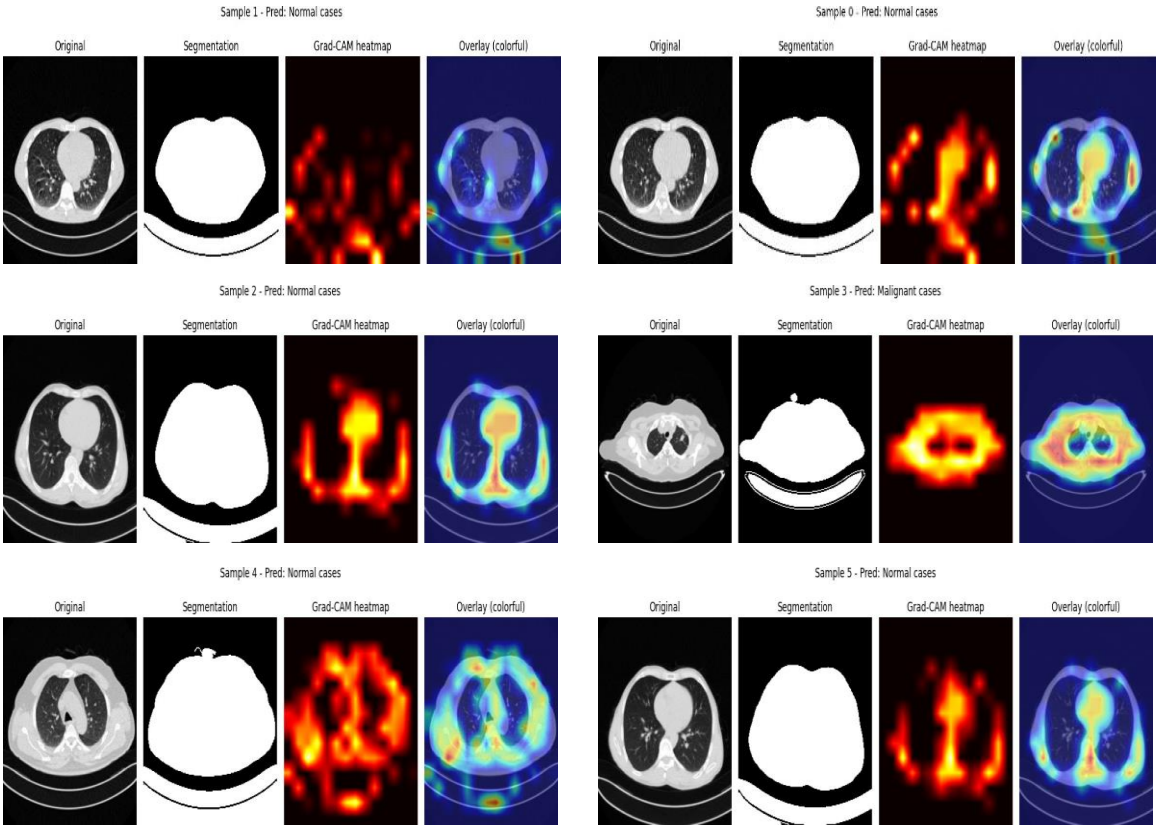

Figure 5.4: Six images arranged in three rows with two images per row

5.5.2 Segmentation-Assisted Visualization

To further enhance interpretability, segmentation masks were incorporated into the visualization pipeline. These masks isolate the lung fields from surrounding tissues before applying Grad-CAM.

When combined with Grad-CAM, segmentation achieves the following:

- Restricts heatmap activation to lung tissue only,
- Eliminates artifacts and irrelevant regions such as bone, muscle, and background,
- Clarifies anatomical boundaries including lobes and pleural surfaces,
- Improves overall ease of interpretation for clinicians.

By constraining visual explanations to clinically relevant zones, segmentation reduces noise and increases confidence that the model evaluates appropriate anatomical structures.

5.6 Ablation Study

A series of ablation experiments was conducted to determine the optimal structural and training configurations for the proposed SA-CCT architecture. Five key components were evaluated: pooling strategy, batch size, loss function, optimizer, and learning rate. The results are summarized in Table 5.3.

Table 5.3: Ablation study results for the SA-CCT model.

Ablation Factor	Configuration	Accuracy (%)
Pooling Layer	Flatten	**93.26**
	Global Max Pooling	93.22
	Global Average Pooling	92.88
Batch Size	128	94.87
	64	94.26
	32	**97.77**
Loss Function	Categorical Cross-Entropy	**95.51**
	MAE	93.23
	MSE	94.87
Optimizer	AdamW	**97.28**
	Adam	96.88
	SGD	96.51
Learning Rate	$1e^{-}4$(*optimal*)	**99.09**

Overall, these ablation experiments guided the refinement of the SA-CCT architecture, ensuring a well-balanced and high-performing configuration for medical image classification.

CAPTER 6

CONCLUSION AND FUTURE WORKS

6.1 Conclusion

In this work, we proposed an optimized Self-Attention Compact Convolutional Transformer (SA-CCT) model for lung cancer prediction using CT images from the IQ-OTH/NCCD dataset. The model combines convolutional tokenization, enhanced multi-head self-attention, widened MLP blocks, and sequential pooling to effectively capture both local and global features under limited data conditions.

Experimental results demonstrate that the optimized SA-CCT achieves a testing accuracy of 99.09% and delivers balanced performance across benign, malignant, and normal classes. The confusion matrix shows almost perfect class separation, with particularly strong recall for malignant and benign cases.

Compared with standard CNN-based models such as InceptionV3, ResNet152, VGG19, and MobileNetV2, the proposed SA-CCT attains higher accuracy while using a smaller input resolution (32×32) and fewer parameters. This confirms that transformer-based architectures with convolutional tokenizers are highly suitable for medical image analysis tasks.

Furthermore, Grad-CAM and segmentation-assisted visualization demonstrate that the model focuses on radiologically meaningful regions, such as nodules, abnormal densities, and structural irregularities in the lung parenchyma. This improves interpretability and supports clinical trust in the automated diagnostic system.

6.2 Future Works

Although the proposed optimized Self-Attention Compact Convolutional Transformer (SA-CCT) demonstrates excellent performance and strong interpretability for lung cancer prediction, several promising research directions remain open for future exploration.

First, the current study is conducted on a single publicly available dataset (IQ-OTH/NCCD). Future work can focus on validating the proposed model on multiple large-scale and multi-center datasets, including datasets collected from different hospitals and imaging devices. Such cross-dataset evaluation would help assess the generalization capability and clinical robustness of the SA-CCT model under real-world conditions.

Second, while the present work focuses on three-class classification (benign, malignant, and normal), future research may extend the framework to multi-stage lung cancer grading and subtype classification. Incorporating clinical staging information (e.g., tumor size, lymph node involvement, and metastasis indicators) could enable more fine-grained diagnostic and prognostic predictions.

Third, the explainability framework can be further enhanced by integrating additional interpretability techniques such as attention roll-out, transformer attribution maps, or counterfactual explanations. Combining multiple explainability methods may provide deeper insight into how transformer-based architectures reason across different anatomical regions and improve clinician confidence in automated predictions.

Fourth, future work may explore the integration of multimodal data sources. In addition to CT images, incorporating clinical metadata such as patient age, smoking history, genetic markers, and laboratory findings could significantly improve predictive accuracy and clinical relevance. A multimodal transformer-based architecture may better capture the complex relationships between imaging and non-imaging features.

Fifth, although the optimized SA-CCT achieves high accuracy with a relatively small input resolution, further optimization for real-time and edge deployment can be investigated. Model compression techniques such as pruning, quantization, and knowledge distillation may enable deployment on low-resource clinical systems and portable diagnostic devices.

Finally, future research could extend the proposed framework to related medical imaging tasks such as lung tumor segmentation, treatment response prediction, and survival analysis. Adapting the SA-CCT architecture for longitudinal CT analysis may help track disease progression over time and support personalized treatment planning.

Overall, these future directions aim to enhance the clinical applicability, scalability, and interpretability of the proposed SA-CCT framework, paving the way toward more reliable and intelligent computer-aided lung cancer diagnostic systems.

REFERENCES

[1] D. Radhakrishnan *et al.*, "The emergence of nanoporous materials in lung cancer therapy," 2022. doi: 10.1080/14686996.2022.2052181.

[2] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification," *Multimed Tools Appl*, vol. 79, no. 11–12, 2020, doi: 10.1007/s11042-019-08394-3.

[3] D. Bouchaffra, F. Ykhlef, and S. Benbelkacem, "Advancement in Lung Cancer Diagnosis: A Comprehensive Review of Deep Learning Approaches," 2024. doi: 10.1007/16833_2024_302.

[4] D. Mathios *et al.*, "Detection and characterization of lung cancer using cell-free DNA fragmentomes," *Nat Commun*, vol. 12, no. 1, 2021, doi: 10.1038/s41467-021-24994-w.

[5] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier," *Neural Comput Appl*, vol. 34, no. 12, 2022, doi: 10.1007/s00521-020-04842-6.

[6] E. Dama *et al.*, "Biomarkers and lung cancer early detection: State of the art," 2021. doi: 10.3390/cancers13153919.

[7] T. Zhang *et al.*, "Genomic and evolutionary classification of lung cancer in never smokers," *Nat Genet*, vol. 53, no. 9, 2021, doi: 10.1038/s41588-021-00920-0.

[8] T. L. Chaunzwa *et al.*, "Deep learning classification of lung cancer histology using CT images," *Sci Rep*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-84630-x.

[9] D. Riquelme and M. A. Akhloufi, "Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans," 2020. doi: 10.3390/ai1010003.

[10] O. Rodak, M. D. Peris-Díaz, M. Olbromski, M. Podhorska-Okołów, and P. Dzięgiel, "Current landscape of non-small cell lung cancer: Epidemiology, histological classification, targeted therapies, and immunotherapy," *Cancers (Basel)*, vol. 13, no. 18, 2021, doi: 10.3390/cancers13184705.

[11] I. M. Nasser and S. S. Abu-Naser, "Lung Cancer Detection Using Artificial Neural Network," 2019. [Online]. Available: www.ijeais.org

[12] "IQ-OTH_NCCD lung cancer dataset".

[13] W. Sun, Y. Pang, and G. Zhang, "CCT: Lightweight compact convolutional transformer for lung disease CT image classification," *Front Physiol*, vol. 13, 2022, doi: 10.3389/fphys.2022.1066999.

[14] A. I. Jajja *et al.*, "Compact Convolutional Transformer (CCT)-Based Approach for Whitefly Attack Detection in Cotton Crops," *Agriculture (Switzerland)*, vol. 12, no. 10, 2022, doi: 10.3390/agriculture12101529.

[15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int J Comput Vis*, vol. 128, no. 2, 2020, doi: 10.1007/s11263-019-01228-7.

[16] T. He *et al.*, "MediMLP: Using Grad-CAM to Extract Crucial Variables for Lung Cancer Postoperative Complication Prediction," *IEEE J Biomed Health Inform*, vol. 24, no. 6, 2020, doi: 10.1109/JBHI.2019.2949601.

[17] S. A. Banday, R. Nahvi, A. H. Mir, S. Khan, A. S. AlGhamdi, and S. S. Alshamrani, "Ground glass opacity detection and segmentation using CT images: an image statistics framework," *IET Image Process*, vol. 16, no. 9, 2022, doi: 10.1049/ipr2.12498.

[18] L. Maiello *et al.*, "Automatic Lung Segmentation and Quantification of Aeration in Computed Tomography of the Chest Using 3D Transfer Learning," *Front Physiol*, vol. 12, 2022, doi: 10.3389/fphys.2021.725865.

[19] J. Zhao, M. Dang, Z. Chen, and L. Wan, "DSU-Net: Distraction-Sensitive U-Net for 3D lung tumor segmentation," *Eng Appl Artif Intell*, vol. 109, 2022, doi: 10.1016/j.engappai.2021.104649.

[20] M. H. Asnawi *et al.*, "Lung and Infection CT-Scan-Based Segmentation with 3D UNet Architecture and Its Modification," *Healthcare (Switzerland)*, vol. 11, no. 2, 2023, doi: 10.3390/healthcare11020213.

[21] H. Ma *et al.*, "Automatic pulmonary ground-glass opacity nodules detection and classification based on 3D neural network," *Med Phys*, vol. 49, no. 4, 2022, doi: 10.1002/mp.15501.

[22] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Applied Intelligence*, vol. 51, no. 2, 2021, doi: 10.1007/s10489-020-01829-7.

[23] C. S. Guan *et al.*, "Imaging Features of Coronavirus disease 2019 (COVID-19): Evaluation on Thin-Section CT," *Acad Radiol*, vol. 27, no. 5, 2020, doi: 10.1016/j.acra.2020.03.002.

[24] H. R, S. T, and V. S, "A Comprehensive Analysis of Implementing Convolutional Neural Networks (CNN) and InceptionV3 for Early-Lung Cancer Detection," in *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*, 2024, pp. 157–165. doi: 10.1109/ICMCSI61536.2024.00030.

[25] A. and M. R. M. and L. Y. and A. S. and A. S.-S. M. and B. M. and J. M. B. and A. M. A. Xuan Darren Soong Kai and P.P. Abdul Majeed, "Leveraging Transfer Learning as Feature Extractors for Lung Cancer Classification: Insights from VGG19 and ResNet152 Pipelines," in *Selected Proceedings from the 2nd International Conference on Intelligent Manufacturing and Robotics, ICIMR 2024, 22-23 August, Suzhou, China*, A. and P. T. A. H. and Z. F. and Y. Y. and L. Y. and

H. L. and L. C. and Z. Y. Chen Wei and PP Abdul Majeed, Ed., Singapore: Springer Nature Singapore, 2025, pp. 750–756.

[26] A. Souid, N. Sakli, and H. Sakli, “Classification and Predictions of Lung Diseases from Chest X-rays Using MobileNet V2,” *Applied Sciences*, vol. 11, no. 6, 2021, doi: 10.3390/app11062751.

[27] N. and B. S. and B. S. and K. raja R. Saranya N. and Kanthimathi, “Classification and Prediction of Lung Cancer with Histopathological Images Using VGG-19 Architecture,” in *Computational Intelligence in Data Science*, P. and K. M. and S. M. Kalinathan Lekshmi and R., Ed., Cham: Springer International Publishing, 2022, pp. 152–161.