Title of the Thesis

Comparative Evaluation of Attention-Enhanced Deep Transfer Learning Architectures: Improving Diagnostic Accuracy for Lung Cancer Detection in CT Scans

Submitted By

Ram Proshad Kumar Mohonto
ID: 212-16-571
Department of Computing & Information System
Daffodil International University

Supervised By

Md. Mehedi Hassan
Lecturer (Senior scale),
Department of Computing & Information System
Daffodil International University

Department of Computing and Information System
Daffodil International University.
Dhaka, Bangladesh

Submission Date: 21/10/2025

APPROVAL

This Thesis titled "Comparative Evaluation of Attention – Enhanced Deep Transfer Learning Architectures: Improved Diagnostic Accuracy for lung cancer Detection in CT scans", Submitted by Ram Proshad Kumar Mohonto, ID No: 212-16-571 to the Department of Computing and Information Systems, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computing & Information Systems and approved as to its style and contents. The presentation has been held on 21-10-2025.

BOARD OF EXAMINERS

Md Sarwar Hossain Mollah — **Chairman**
Associate Professor and Head
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Nasimul Kader — **Internal Examiner**
Assistant Professor
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Mehedi Hassan — **Internal Examiner**
Lecturer (Senior Scale)
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Dr. Saifuddin Md. Tareeq — **External Examiner**
Professor
Department of Computer Science and Engineering
University of Dhaka, Dhaka

Declaration

I, **Ram Prosad Kumar Mohonto,** hereby declare that, this thesis titled;"**Comparative Evaluation of Attention-Enhanced Deep Transfer Learning Architectures:Improving Diagnostic Accuracy in Lung Cancer Detection from CT Scan**" has been done by me under supervision of **Md. Mehedi Hassan, Lecturer (Senior Scale),** Department of Computing and Information System (CIS) of Daffodil International University. I am also declaring that this thesis or any part of there has never been submitted anywhere else for the award of any educational degree like, B.Sc., M.Sc., Diploma or other qualifications.

Supervised By

Md. Mehedi Hassan,
Lecturer (Senior Scale),
Department of CIS
Daffodil International University

Submitted By

Prosad

Ram Prosad Kumar Mohanto
ID: 212-16-571
Department of CIS
Daffodil International University

Acknowledgements

I am very thankful to the Almighty Creator, the Most Merciful, for giving me the strength, wisdom, and determination to finish my thesis, "Comparative Evaluation of Attention-Enhanced Deep Transfer Learning Architectures: Improving Diagnostic Accuracy for Lung Cancer Detection in CT Scans."

I am very grateful to my well-known supervisor, Md. Mehedi Hassan, who is a Senior Scale Lecturer in the Department of Computing and Information Systems at Daffodil International University. His constant support, important advice, and insightful feedback were key in shaping the direction and quality of my research.

I want to thank Mr. Sarwar Hossain Mollah, the respected head of the CIS department, for creating an intellectually stimulating environment and providing the necessary resources that inspired me from the start.

I want to thank all of the CIS department's academic staff for their help, advice, and support throughout my research journey. They were very helpful and patient during the hard parts of our project.

The people I mentioned above helped and encouraged me a lot, and I couldn't have finished this thesis without them. I want to thank each of them from the bottom of my heart for helping me with my academic career.

Dedication

I want to thank my parents for always being there for me, loving me, and making sacrifices that have helped me with my academic journey. Without their consistent support and faith in my abilities, I wouldn't have been able to take advantage of this possibility to perform research and study to make a difference in society.

Abstract

Lung cancer continues to be a predominant source of cancer-related mortality, with early identification essential for enhancing patient outcomes. This paper offers a comparative assessment of deep learning architectures, including VGG16, ResNet50, MobileNetV2, and Vision Transformer (ViT), for the identification of lung cancer from CT scans. To improve diagnostic precision, we propose an attention-enhanced VGG16 (AttVGG16) that incorporates the Convolutional Block Attention Module (CBAM) to emphasise prominent problematic areas. Experimental results indicate that baseline VGG16 outperforms conventional architectures, whereas ResNet50, MobileNetV2, and ViT display diminished predictive efficacy due to constrained feature representation or data prerequisites. AttVGG16 surpasses all baseline models, with 97.78% accuracy, 97.80% sensitivity, 97.76% F1 score, and 0.9178 MCC, underscoring the effectiveness of attention processes in accentuating diagnostically pertinent areas and minimising false negatives. The research employed a meticulously curated dataset of lung CT scans, incorporating extensive preprocessing such as normalisation, augmentation, and class balancing to mitigate data scarcity and improve model generalisability. Performance was assessed through many measures to ensure a comprehensive evaluation of categorisation accuracy, sensitivity, and predictive reliability. Moreover, the attention mechanism integrated into AttVGG16 enables the model to emphasise significant areas in CT images, enhancing the network's interpretability for clinical applications. These findings highlight the efficacy of attention-enhanced CNNs for accurate and early lung cancer identification in CT imaging, providing a valuable resource to assist radiologists in diagnostic decision-making. The suggested methodology may aid in diminishing diagnostic inaccuracies, enabling prompt interventions, and eventually enhancing patient management and outcomes. Future endeavours may encompass the expansion of this framework to encompass larger and more heterogeneous datasets, the integration of multimodal imaging, and the creation of real-time clinical decision support systems for the automated identification of lung cancer.

Preface

This bachelor's thesis focusses on forecasting lung cancer using data obtained from a public repository. The main goal is to create a hybrid model that can help us understand the datasets we have analysed. This study seeks to enhance lung cancer research and diagnoses within the realm of CT scan imagery.

This thesis includes 7 chapters which are briefed as follows:

Chapter-1

An overview of lung cancer disease and its effects is given in this chapter. It describes the goals and expected results, highlights the research's contributions, and talks about the reasons for the study.

Chapter-2

The chapter is a review of the current studies on deep learning-based lung cancer detection and computer-aided diagnosis systems. It reviews the previous methodologies and the major findings of different models such as CNNs, attention mechanisms, as well as transformer-based models and also identifies the gaps and limitations in the existing literature that inform the current study.

Chapter-3

The methodology of the study is summarised in this chapter, along with preprocessing steps, a detailed description of the data attributes, and a dataset description.

Chapter-4

Chapter 4 presents the results of implementing diverse baseline and hybrid models on the dataset. It presents a thorough analysis of the outcomes derived from each model, accompanied by an extensive comparison table that elucidates several algorithms in lung cancer pathology.

Chapter-5

Conclusion of the research.

Chapter-6

Future work.

Chapter-7

References

Content

List of Figures

List of Tables

List of Abbreviations

CAD - Computer-Aided Diagnosis
CNN - Convolutional Neural Network
CT - Computed Tomography
LDCT - Low-Dose Computed Tomography
AI - Artificial Intelligence
DL - Deep Learning
VGG16 - Visual Geometry Group 16-layer network
ResNet50 - Residual Network with 50 layers
MobileNetV2 - Mobile Network Version 2
ViT - Vision Transformer
AttVGG16 - Attention-enhanced VGG16
CBAM - Convolutional Block Attention Module
SVM - Support Vector Machine
LSTM - Long Short-Term Memory
GRU - Gated Recurrent Unit
DICOM - Digital Imaging and Communications in Medicine
HU - Hounsfield Units
IRB - Institutional Review Board
ROC - Receiver Operating Characteristic
AUC - Area Under the ROC Curve
MCC - Matthews Correlation Coefficient
ReLU - Rectified Linear Unit
ReLU6 - Clamped ReLU (min(max(0, x), 6))
GELU - Gaussian Error Linear Unit
MLP - Multi-Layer Perceptron
MSA - Multi-Head Self-Attention
LN - Layer Normalization
SGD - Stochastic Gradient Descent
FLOPs - Floating Point Operations
IQ-OTH/NCCD - Iraq-Oncology Teaching Hospital / National Center for Cancer Diseases

Chapter 1

Introduction

1.1 Introduction

Lung cancer is an deadly disease, around the world. I have seen patients with lung cancer. Lung cancer kills people because doctors often find lung cancer only after lung cancer has spread far. Early signs of lung cancer are vague or missing. In the decades doctors have used imaging tools such, as chest X-rays and computed tomography (CT) scans to look for lung problems. Chest X-rays and CT scans help find lung problems early. In my work I see low-dose CT (LDCT) help find disease. I also see the amount of imaging data keep growing. The growing imaging data puts a load on radiologists. That heavy load leads to errors leads to variability and leads to delayed reporting. I notice the inter-observer variability, in assessment makes early detection even harder. The subtle nodular changes can be missed. The missed changes happen especially when the lesions are small or peripheral. I notice developments, in the computer methods, artificial intelligence and machine learning let us run automatic analysis of the medical images. This can give the accuracy the speed and the more repeatable results. I see the impact. Deep learning models, neural networks have had big success in the medical image tasks. Deep learning models can learn features, from the data on their own. Deep learning models pick up edges and textures. Also pick up more complex shape patterns. I see that attention mechanisms and transformer based architectures have recently emerged. I see that attention mechanisms and transformer based architectures give focus on the relevant regions. I see that attention mechanisms and transformer based architectures capture range dependencies, in image data. I think attention mechanisms and transformer based architectures can improve lung cancer detection and lower the workload, on clinicians. I think attention mechanisms and transformer based architectures can improve outcomes.

Lung cancer continues to be the primary cause of cancer-related mortality globally, with over an estimated 1.8 million deaths each year [1]. It is highly lethal for being diagnosed and treated at advanced stages, when treatment efficacy is low; early detection can improve survival from less than 20% to more than 70% over a five-year period[2]. Low dose computed tomography (LDCT) is established as the gold standard for screening leading in a reduction of 20\% of cancer death due to early detection of pulmonary nodules according to National Lung Screening Trial (NLST) trial [3]. However, radiologist expertise is required for manual interpretation of CT scans, which can become a highly cognitive workload for per patient (hundreds slices), causing inter-observer differences, fatigue perception and inconsistent diagnostic results [4].

This has sped up research into Computer-Aided Diagnosis (CAD) systems to make diagnoses more accurate and cut down on mistakes made by people.

Early computer-aided design (CAD) used hand-made radiomic features like texture, shape, and intensity, which were sorted by methods like support vector machines (SVMs) or random forests [5]. However, these methods were limited by the quality of segmentation, their sensitivity to noise, and their ability to express features. Even small changes in how nodules are drawn or artefacts in imaging can make performance much worse, which limits their use in medicine. The introduction of deep learning (DL), especially convolutional neural networks (CNNs), transformed lung image analysis by facilitating automated feature extraction directly from raw pixel data, thereby identifying intricate hierarchical patterns frequently overlooked by conventional techniques [6]. Deep CNNs can model small differences in the shape of nodules, the density of tissue, and the surrounding anatomical structures. This is important for telling the difference between benign and malignant lesions. Transfer learning with ImageNet-pretrained models has improved performance even more in medical fields where there isn't much labelled data [7]. This lets networks use low-level features like edges and textures while changing high-level layers to show disease-specific representations. Architectures like VGG16 [8], ResNet50 [9], and MobileNetV2 [10] have been used a lot for analysing pulmonary nodules, and they have made big improvements over older methods.

But traditional CNNs process images in the same way every time, which isn't good for clinical diagnosis because problems are often only found in small areas. Attention mechanisms get around this problem by dynamically focussing on important features and blocking out unimportant ones [11]. The Convolutional Block Attention Module (CBAM) [12] is a good example of this. It combines channel- and spatial-level recalibration so that the network can focus on small nodules or subtle tissue irregularities that are important for diagnosis. At the same time, Vision Transformers (ViTs) [13] have become a competing model. They use self-attention to model long-range dependencies and capture global context well. However, their usefulness in medical imaging is still limited by a lack of training data and high computational requirements [14]. ViTs can, in theory, improve the detection of subtle or dispersed abnormalities by modelling relationships between distant areas in an image. However, without enough data, overfitting and poor generalisation are still big problems. So, attention-enhanced CNNs, which combine hierarchical convolutional feature extraction with localised attention, seem to be a good compromise because they provide both fine-grained spatial sensitivity and the ability to learn quickly from small datasets.

Even though these improvements have been made, there are still not many rigorous comparisons of attention-enhanced CNNs with canonical CNNs and transformer models on balanced lung CT datasets. A lot of previous research has used small or very unbalanced datasets and only looked at binary classification between benign and malignant nodules. These kinds of studies can be helpful, but they may not apply well to real-life situations where it is important to tell the

difference between normal, benign, and malignant classes. To fill this gap, we put together and improved the public IQ-OTH/NCCD dataset [15] to make sure that the classes were balanced and that the evaluation was strong. We used data augmentation techniques like random rotations, flips, scaling, changing the intensity, and changing the contrast to make the samples more diverse, reduce overfitting, and make the model more general. By adding realistic changes to the way nodules look and how imaging is done, augmentation makes the model less sensitive to common factors that can confuse it, like how the patient is positioned, how they breathe, or differences between scanners. In medical imaging, these strategies are especially important because getting large annotated datasets is often impossible because of privacy issues and the fact that expert labelling takes a lot of time.

We examine the best designs, such as VGG16, ResNet50, MobileNetV2, and ViT. We also suggest an attention-enhanced VGG16 (AttVGG16) that employs CBAM to improve feature representations and focus on clinically essential areas. Many tests suggest that AttVGG16 is better than baseline architectures on a lot of performance parameters, such accuracy, sensitivity, and F1 score. This shows how well attention mechanisms work to reduce false negatives and make things clearer. When clinicians look at attention maps together with visual outputs, they can see which elements of the model's judgements were based on, which makes AI-assisted diagnoses more open and reliable. The channel-wise and spatial attention methods also help the network focus on places where cancer is more likely, such as nodules with uneven boundaries, spiculations, or varying densities. It can also ignore tissues that aren't significant.

These results show how important spatially adaptive processing is for improving the detection of lung cancer and suggest that attention-enhanced networks can be useful decision support tools in clinical workflows. Furthermore, by combining attention maps with a radiologist's review, these systems can boost diagnostic confidence, put high-risk cases at the top of the list, and make it easier to act quickly. Attention-enhanced CNNs can work together with existing CAD pipelines to help radiologists find nodules and score their risk, which makes their work easier and less mentally taxing.

There are still some problems with deep learning-based CAD systems, even though they have shown promise. Many existing models are trained on small or unbalanced datasets, which can cause them to overfit and not work as well on different populations and imaging devices. Also, most studies only look at two types of classification (malignant vs. benign), but making decisions that are important for patients often needs to include normal, benign, and malignant nodules. Interpretability is still a big issue because "black-box" predictions may make it harder for doctors to use them if they need visual or quantitative proof for their diagnostic decisions. To solve these problems, we need to create models that are not only very accurate but also strong, easy to understand, and able to make the most of small amounts of annotated data. The addition of attention mechanisms to CNNs, which is the focus of this study, is a big step towards meeting

these needs because it highlights important areas and gives insight into how the model makes decisions. This makes it easier to trust and use in clinical settings.
The AttVGG16 model that's been suggested has some interesting implications for screening programs and diagnostics on a larger scale. It makes it possible to quickly analyze large CT datasets in an automated way. This can really help with mass screening efforts and make it easier to identify groups that are at higher risk. When this model is used in clinical settings, it can provide real-time alerts for nodules that look suspicious. It also helps prioritize which cases should be reviewed more closely, which might cut down on delays in diagnosis. These efficiency improvements are really important in healthcare systems where there aren't many radiology resources available. They help streamline workflows while still keeping high standards for diagnostics.

Lastly, depending on the application area there is potential for extending the network to multi-modal spaces by incorporating inputs from other modalities (e.g. CT-PET, CT-MR), which will offer additional structural and metabolic information. Hybrid architectures, such as those combining the transformer modules with convolutional backbones, could further improve performance by capturing local fine-grained details and long-term dependencies. Third, light-weight attention models for real-time deployment on edge devices could help make AI-assisted diagnosis available even in low-resource regions, contributing to the global fight against lung cancer death. Taken together, these results suggest an important step forward in AI-aided lung cancer detection are attention-enhanced convolutional neural networks. They also achieve higher diagnostic accuracy, as well as robustness and interpretability over conventional CNNs and state-of-the-art transformer models, suggesting potential applications in clinical tools for screening and decision support. By mitigating the limited data, class imbalance and feature localization challenges, the impact of these methods could be to enhance patient outcome, lower healthcare cost and accelerate AI-based diagnostics in oncology for wider clinical use.

1.2 Research Objective

Despite considerable advances in computer-assisted diagnosis of lung cancer, there are several important research challenges. Conventional machine learning methods suffer from the lack of hand-crafted features, which are susceptible to distortion due to imaging quality, segmentation accuracy and operator knowledge. These models tend to exhibit good performance on curated data, but lack the capability to generalize across heterogeneous populations, imaging protocols or scanners. This is even more relevant in medical tasks where variation is unavoidable and strong generalisation to different settings is a must.

These limitations have partially been addressed by deep learning techniques which permit automated hierarchical feature extraction. CNNs are capable of learning intricate features from raw image data, and can hence remove the need for manual feature engineering. Efficient,

However traditional CNNs treat all regions of the images equally which may not be desirable for clinical tasks, well localized pathologies are frequently detected by utilizing a small part of the image. Conventional CNNs may have difficulty in capturing the subtle nodules, small lesions, and low-grade tumors (early-stage) since these can either be with high similarity between classes or in low contrasted images. However, deeper architectures such as ResNet50 and DenseNet have better feature representation but are vulnerable to overfitting on small datasets and/or might not capture diagnostically relevant areas - limiting their interpretability and trust in clinical practise.

There is another loophole in terms of multi-class classifiers. The majority of studies are based on binary classification, for example malignant vs. benign nodules. In many clinical tasks, however, the classification among these groups of tumours is involved. Very few studies have thoroughly evaluated a variety of architectures like CNNs, attention-enhanced CNNs, and transformer models systematically on balanced multi-class data. The first problem is important because good binary classification performance does not always directly translate to clinically meaningful multi-class situations.

A potential answer to the problem is attention mechanism, which however, has been not well employed in lung cancer detection. Attention modules may help emphasize important features, enhance explainability and alleviate false negative error but few works have properly quantified these benefits in a controlled multi-class test case. Moreover, incorporation of attention mechanisms on top of prevailing symbolic models has to be carefully designed for trade-off between computational complexity and accuracy as well as clinical interpretability. While theoretically being advantageous for global dependency capturing, Transformers are impractical in medical imaging due to excessive data and computational requirements.

Finally, the interpretability and clinical acceptance of AI models remain important challenges: "black-box" models-that is, models that make predictions but do not explain those predictions-are unlikely to see wide adoption in healthcare. Clinicians need systems that perform well on accuracy but also provide some form of visual or quantitative justification. This lack of models with high interpretability, robustness, and generalizability creates a significant barrier to the translation of AI research into clinical practice.

In short, the gap in the research indicates that such models are needed, which are accurate, interpretable, robust to data, can be used to classify multiple classes, and could be used to point out clinically significant features. These gaps need to be addressed to come up with AI systems that become effective to be factored into the usual screening and diagnostic processes of lung cancer. The proposed research will address these gaps by showing an attention-enhanced VGG16 model based on hierarchical feature extraction, adaptive attention, and transfer learning to give accurate, interpretable, and clinically relevant predictions.

1.3 Motivation

The rationale behind this study is the creation and use of sophisticated methods of computation to enhance the early identification and diagnosis of lung cancer using CT images. This is despite the fact that medical imaging and computer-aided diagnosis have made significant progress and still have serious challenges that restrict the efficacy and reliability as well as clinical aptitude of the systems currently in existence. Having lung cancer diagnosed early is one of the most important factors to ensure high survival rates of patients, but due to the presence of manual interpretation, incomprehensive feature representation and insensitivity to smaller or minor changes, the majority of the traditional methods are limited. These shortcomings highlight the need to develop improved approaches that are able to close the gap between unprocessed imaging data and clinical actionable information.

Among the improvement points, the implementation of deep learning architectures with automatic feature extraction can be mentioned. The conventional radiomics and hand-designed feature techniques are both time-consuming due to their high dependence on manual segmentation and pre-defined descriptors, and also subject to inter-operator variance. Conversely, deep convolutional networks like VGG16, ResNet50, and MobileNetV2 have a hierarchy of feature extraction in one capacity that has the ability to extract detailed spatial information in CT images. Through these models, this paper will aim at improving the granularity and accuracy of feature representation, which will allow the identification of small nodules, irregular tissue textures, and other subtle signs of malignancy that can be missed by human eyes or the traditional CAD systems.

The other aspect of improvement is on incorporation of attention systems into such convolutional networks. Standard CNNs treat every part of an image in the same way, which may result in the blurring of the attention to important pathological regions. Such modules like the Convolutional Block Attention Module (CBAM) enable networks to be adaptive by highlighting diagnostically relevant regions both channel-wise and spatially. The ability is especially relevant to the lung cancer detection where nodules may be of different sizes, shapes and densities, and even be in anatomically difficult locations. Attention mechanisms increase the discriminability of the network, false negative, and interpretable attention maps which can be consulted by radiologists, thereby augmenting clinical trust and utility by refining the networks focus.

The area of improvement, also includes data-driven approaches to address the weaknesses related to small and unbalanced datasets. The data of lung cancer usually contains a imbalanced number of malignant, benign and normal cases and this may bias the learning process and deteriorate generalization. This research paper will counter such challenges by using systematic data augmentation, such as geometric transformations, intensity variations, and spatial manipulations to produce a balanced and diverse dataset. This would not only enhance model robustness, but also increase the range of visual features which the model can be trained on and

therefore it is more robust to changes in imaging protocols, patient anatomy and acquisition devices.

Another area of the critical improvement is transfer learning. With the weights of deep networks trained on large scale image data, initialization of the deep networks allows the model to utilize generalized low level features, which include edges, textures, and shapes, and further refine higher levels to the specific features in lung cancer. This method cuts down on requirements on computational resources, prevents overfitting, and does accommodate effective learning on small annotated datasets. It also enables quick implementation in clinical settings where labeled data might be limited, enabling the adoption of AI-based tools in the existing clinical workflow.

In addition to model architecture and data set factors, the sphere of improvement is further expanded to the power of enhancing the interpretability and clinical value of the predictive outputs. The improved system can be used to produce visual explanations, which can be verified by clinicians by producing attention maps and identifying areas with the highest contribution to classification choices. This interpretability is essential to the clinical acceptance because radiologists need to be confident in automated predictions and model reasoning that can be traced. It also has educational and research application, as there is an opportunity to analyze fine imaging patterns which can be added to certain pathological characteristics or stages of the progression of the disease.

Moreover, the improved structure is to be universal to support all the architectures and comparative studies. The study compares attention-enhanced VGG16 with ResNet50, MobileNetV2, and Vision Transformers (ViTs) to give a holistic understanding of the performance of each model, as it shows the benefits and drawbacks of each model over its counterparts. Such comparative assessment will allow identifying the best practices in the model design, attention integration, and deployment strategies that would help make sure that the improvements are not the isolated improvements but rather the improvements that have to be placed in the greater context of the computational lung cancer detection.

Besides such performance measures as accuracy, sensitivity, specificity, F1 score, and AUC, the breadth of improvement covers the provision of better practical utility in the real-life clinical environment. Performing models should be not only statistically good, but also efficient, with both calculably cost and inference processes. This choice of architectures (e.g. MobileNetV2 to be efficient and using attention mechanisms selectively) allows the framework to compromise between performance and operational feasibility, rendering it able to be deployed onto common clinical hardware or, possibly, even portable diagnostic devices.

The improvements also envisage future developments on multimodal and multi-institutional applications. Though this paper concentrates on CT imaging, the principles of hierarchical feature extraction, weighting of attention, and transfer learning can be applied to exploit other

imaging modalities, e.g., PET or MRI, and clinical metadata, e.g., patient demographics and laboratory results. The proposed enhancements have the broad potential, and this multimodal method may enhance the quality of the diagnosis, the risk stratification, and individual treatment planning.

1.4 Rationale of the Study

Lastly, the area of improvement bears in mind the long-term gains of the healthcare systems and patient outcomes. The better framework can help to improve the timeliness of interventions, lower treatment expenses, and increase the survival rates by improving early detection, minimizing diagnostic errors, and offering interpretable AI support. It also decreases the load of radiologists who can devote more time to more complicated cases and risky patients which eventually lead to the improvement of the overall quality and efficiency of lung cancer patients.

To conclude, the areas of improvement in this research are detailed and complex. It discusses fundamental issues of lung cancer detection, such as shortcomings of classic feature based approaches, adaptive attention mechanism, imbalance in datasets, interpretability, computational efficiency, as well as health care integration. The study offers a sound outline of the further development of automated lung cancer detection by building upon the systematic enhancement of the model architecture, training schedules, and the use of data. Such improvements can make a great contribution to clinical practice, patient outcomes, and further research directions, and create a ground of further innovations in medical image analysis.

Chapter 2

Literature Review

2.1 Introduction

In this chapter, I will discuss the current studies in the field of predicting Lung cancer, discussing the methods, results and accurateness of various studies. I will also bring into the limelight my approach and the challenges that I encountered in my study and the lessons that were learnt.

2.2 Related Work

The correct CT scan-based diagnosis of lung cancer has led to the replacement of the classical machine learning with deep learning (DL) computer-aided diagnosis (CAD). The first CAD tools used required significant amounts of manual descriptors, including SIFT, HOG, and wavelet features, and classifiers, like the support vector machines (SVMs) [16]. Although these techniques were a basis of automated analysis of nodules in the lungs, they were limited by the manual limitations of feature engineering. Conventional radiomic methods aimed to identify predetermined quantitative attributes of nodules such as texture, shape, intensity, and edge descriptors that are extracted on segmented nodules. Even though these features might be useful in capturing some morphological features, they were very vulnerable to the quality of segmentation, preprocessing, and operator skill [24][25]. Inter-rater inconsistency due to variability in annotation, scanner procedures, and imaging noise was a common occurrence that reduced the clinical applicability. Further, the handcrafted features were too inexpressive to adequately represent the delicate formations that signified malignancy particularly to small and irregularly shaped nodules.

With the introduction of convolutional neural networks (CNNs), a new approach to the analysis of medical images emerged. The architectures, which include AlexNet [17], VggNet [8], and ResNet [9], achieved the state-of-the-art performance in the natural image recognition task and were later generalized to the medical imaging. CNNs automatically discover hierarchical features using the raw pixel intensities, gradually discovering low-level edges, mid-level textures and high-level semantic patterns. CNNs have been tested to be highly sensitive and specific in the detection of pulmonary nodules in lung CT scans [18][19], with VGG16 specifically used due to its deep and uniform structure, which allows it to extract features effectively on a wide range of data types [20][21]. CNNs have also been used in several other applications that are beyond detection, including multi-class lung cancer, nodule malignancy risk, and segmentation, which underscores their flexibility and clinical importance [22][23]. Moreover, hybrid networks that integrate CNNs with recurrent networks like LSTMs or GRUs have been investigated to support

sequential relationships between CT slices, and they are more successful in the analysis of volumetric data [28]. These hybrid designs combine both spatial and sequential contextual information and overcome the shortcomings of convolutional designs in volumetric data.

Litjens et al. (2017) [16]	Transfer Learning (CNNs)	Showed ImageNet pre-trained models outperform traditional ML; established transfer learning as standard in medical imaging	Avg. gain 10–15% vs. hand-crafted ML
Gulshan et al. (2016) [20]	Transfer Learning (Inception)	Demonstrated effective transfer from natural images to medical images; validated generalizability of CNN features	AUC: 97.5% (retina), adapted in lung CAD
Woo et al. (2018) [12]	CBAM (Attention Module)	Proposed channel + spatial attention to refine CNN features; improved interpretability	+2–4% accuracy in medical imaging tasks
Chen et al. (2021) [24]	TransUNet (Transformer + CNN)	Hybrid model combining CNN encoder with Transformer for medical segmentation; highlighted attention advantages	Outperformed pure CNNs on segmentation benchmarks
Proposed (AttVGG16)	**VGG16 + CBAM**	**Attention-enhanced transfer learning model for multi-class lung cancer classification; enhances localized feature discriminability**	**+3–5% accuracy over baseline VGG16**

Table 2.2.1 : Summary of prior studies on transfer learning and attention mechanisms in computer-aided diagnosis (CAD) systems, specifically emphasizing their roles in enhancing feature representation and model generalization for medical image analysis highlighting Attention based VGG16 achieving notable improvements.

2.3 Significance and Challenges

Transfer learning has become a vital method to enhance performance of medical imaging tasks, annotated datasets of which are often small. With networks trained using large-scale datasets like ImageNet [7], it is possible to use general-purpose low-level feature representations (e.g., edges, corners, textures) and fine-tune higher-level layers to absorb disease-specific patterns. This method will decrease training time, decrease overfitting, and enhance generalization in environments where it is not feasible to label thousands of CT scans. Other researches have indicated large performance improvements in pulmonary nodule classification and lung cancer subtype prediction with transfer learning, especially when data augmentation methods are used to artificially increase dataset diversity [7][24].

Residual learning, introduced in networks such as ResNet50, allows the training of very deep networks, by adding identity shortcut connections, which help to propagate the gradient and alleviate the vanishing gradient problem [22]. Though the deep residual networks can portray the complex hierarchical patterns, they can also fail to perform particularly in situations with little annotated data or when the features of interest are highly localized such as small lung nodules. MobileNetV2 is a computationally-efficient network, which balances accuracy and the number of parameters, which makes it suitable in a real-time or resource-constrained environment [18]. In other architectures such as DenseNet, EfficientNet, and Inception networks, have trade-offs in depth, connectivity and parameter efficiency, and research has found mixed levels of success in detecting and classifying pulmonary nodules [24]. Nonetheless, the traditional CNNs cannot exploit the non-uniform nature of images (regions) because they assume all images are treated as the same, with all image regions treated identically, and thus fail to detect subtle or localized anomalies.

Attention mechanisms offer a remedy to this weakness, whereby feature maps are adaptively weighted to highlight diagnostically significant areas. Convolutional Block Attention Module (CBAM) [12] does refines both channel and spatial features which enable the network to concentrate on salient features and eliminate irrelevant or noisy background data. CBAM and other attention modules have been shown to be more effective in medical imaging tasks including, but not limited to: classifying chest X-rays [23], segmenting brain tumors [30], and nodule detection. Attention mechanisms can improve interpretability, reduce false negatives, and can be made to focus on critical structures, e.g. irregular nodules, spiculations or heterogeneous tissue patterns, to enable clinician trust and adoption. Squeezeandexcitation (SE) block variants, self-attention, and dual attention network variants have been effectively incorporated into CNN pipelines and have continuously achieved better classification and segmentation performance [25][26].

Simultaneously, Vision Transformers (ViTs) [13] have proposed self-attention networks that could be used to capture long-range dependencies among image patches, including global contextual interactions that CNNs may fail to capture. ViTs demonstrate a potential of modeling features in a holistic manner, but their data needs are large, and training on limited samples is sensitive, limiting their use in most medical imaging applications [28]. The hybrid architecture that depends on CNN backbones and transformer blocks have already shown their capabilities to utilize local convolutional feature extraction and global attention to show better results in multi-class classification tasks [26][29]. These methods give an encouraging way of both local fine-grained details as well as long-range dependencies, especially crucial in the detection of subtle and spatially dispersed pathologies.

Nevertheless, relatively few comparative analyses of attention-enhanced CNNs and canonical CNNs as well as transformer models have been performed to classify lung cancer multi-classically. Numerous earlier researches have been based on dichotomous or limited and

skewed datasets, which limit generalization. Also, systematic evaluation of trade-offs in terms of accuracy of classification, complexity of computations, and interpretability of the model is absent. In the effort to fill these gaps, our work compares VGG16, ResNet50, MobileNetV2 and ViT on a curated and balanced dataset and proposes an attention-enhanced VGG16 (AttVGG16) with CBAM. AttVGG16 achieves this by adding a hierarchical feature extraction term with a focus on attention-based feature refinement, which is able to prioritize local diagnostic features, lessen misclassification, and enhance overall robustness. With this method, we have offered a thorough overview of the use of modern deep learning architectures in deep learning to achieve reliable and multi-class lung cancer detection and also indicated the importance of attention mechanisms in improving performance and interpretability.

Chapter 3

Methods and Materials

3.1 Data Collection & Preprocessing

The data used in this study is the publicly accessible dataset of thoracic (computer tomography) CT scans of lung cancer, as provided by the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) [1]. The data sample was filtered within three months of the fall of 2019 in two specialized Cancer centers in Iraq, that is, the Oncology Teaching Hospital and National Cancer Diseases Center. It contains CT scans of 110 participants that were diagnosed with lung cancer at different stages of the disease and healthy control participants. These centers had expert oncologists and radiologists who helped to annotate the data, which meant high-quality clinical labels.

Its original data sample is 1,190 CT slices which are based on 40 malign cases, 15 benign cases and 55 normal cases. Every patient provided between 80 and 200 slices, which were obtained by Siemens SOMATOM scanner under the following imaging parameters: tube voltage was 120 kV, slice thickness was 1 mm, window width was between 350 and 1200 Hounsfield Units (HU), and window center was between 50 and 600 Hounsfield Units (HU). Scans were all done with the subjects holding the breath-hold fully inspired to reduce motion artifact. Digital Imaging and Communications in Medicine (DICOM) was the original format of the raw data that was later de-identified according to ethical and privacy regulations. The approvals of the participating medical centers were taken as the institutional Review Board (IRB) and the oversight review board waivered the written consent.

We balanced the dataset by first upsampling the data with systematic data augmentation until each of the classes had 600 samples, and we obtained a balanced dataset that had 1,800 images. The pipeline of augmentation utilized a complex of geometric and photometric transformations, such as horizontal and vertical flips, left right and top bottom rotations, scaling, random crops, translation and slight manipulation of the intensity. These transformations have been chosen very carefully to maintain the diagnostic consistency of lung structures and enhance intra-class diversity and boost the generalization ability of downstream models [2][3].

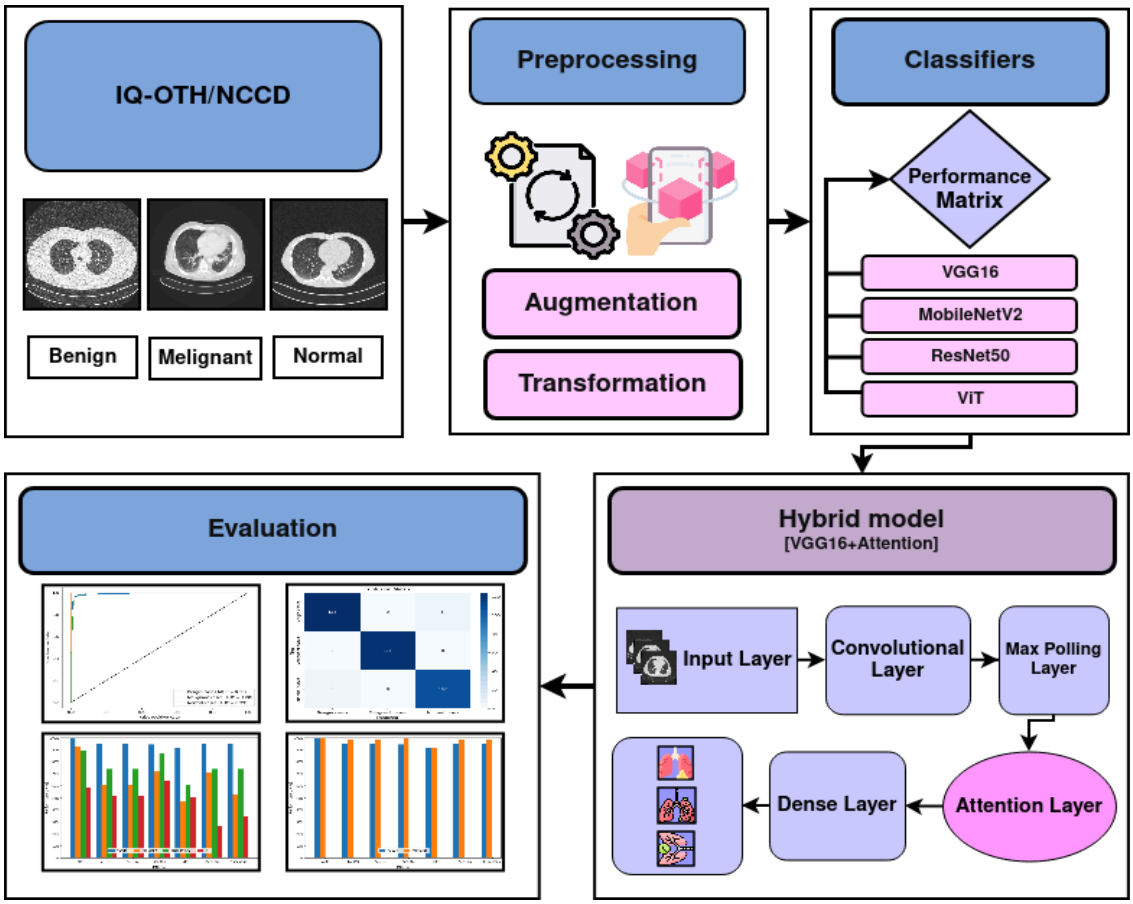

Figure 3.1.1: Detailed workflow of the proposed attention-enhanced AttVGG16 model for automated lung cancer detection from CT scans. The framework integrates transfer learning with the Convolutional Block Attention Module (CBAM) to enhance both channel and spatial feature representations. The process begins with image preprocessing and normalization, followed by feature extraction using a pre-trained VGG16 backbone

3.2 Data Augmentation

Data augmentation is an essential method in contemporary deep learning processes, with notable use in medical imaging processes where large annotated datasets are frequently difficult to obtain because of cost, privacy issues, as well as an expert annotator being required. The inherent aim of data augmentation is to artificially increase the size of the training data, by performing a sequence of controlled transformations on the existing images. The process enables the models to experience a greater range of input patterns during training, thus improving their capability to generalize to unseen data, decreasing overfitting, and increasing their stability to real-world variations.

Considering the analysis of lung CT scans, small differences in image conditions, patient positioning, scanner settings, and anatomical variation can cause a considerable amount of

heterogeneity. Deep learning models can either learn spurious correlation or become too sensitive to certain attributes of the training set, resulting in poor generalization on new patients without having enough variability in the training set. Data augmentation is used to address this issue and generate additional samples which retain the diagnostic features stored within the sample but change appearance, orientation, and scale. This allows the model to acquire features of pathology that are impartial, as opposed to imaging artifacts.

The most used augmentation techniques on medical images are geometric transformations, intensity manipulations, and spatial adaptations. The rotation, flipping, scaling, translation, and cropping geometric transformations provide the simulation of various viewpoints and spatial orientations of the target organ, which is adequate to make sure the network is not overfitted to a particular alignment. The model is used to correct the differences in scanner calibration and patient-specific imaging properties with the aid of intensity adjustments, such as changes in brightness, contrast, and gamma. It can be further enhanced by noise injection, blurring or even sharpening to achieve robustness by replicating the imperfections of real world imaging. Together, these changes form an extensive and diverse dataset which asks the model to infer fixed and discriminative features.

In deep learning pipelines, data augmentation is usually performed online, i.e. one mini-batch of images is transformed in real-time at random. The dynamic method produces an essentially infinite number of variations of the input which can be observed by the network without extra storage and the network can experience a new view of the input data every epoch. This has been found especially useful in medical imaging, where data is usually small, and overfitting may be easily caused by repeated exposure to the same data without variation.

In the case of the proposed attention-enhanced VGG16 model, data augmentation is of primary importance in optimizing the performance of the model. As the attention module also highlights the salient regions in the image, it is important that the augmented images only modify the images with important diagnostic features and not other images. Designing augmentation strategies carefully will make sure that the most important morphological cues, e.g. shape of nodules, their size, texture, and even spatial relations, are preserved. This will ensure that the attention mechanism gets trained to concentrate on clinically significant areas instead of the artifact or other meaningless patterns that augmentation may bring.

Besides, augmentation also adds to the resistance of the model to clinical variability. In practice, CT scans may be configured differently concerning slice thickness, orientation, motion of breathing in the patient, and scanner specific intensity profiles. When the model is exposed to augmented versions of the data which represent such variations, it becomes more resilient and reliable when used in real-life clinical settings. This minimizes chances of misclassification as a result of slight deviation of imaging conditions and this is essential especially in cases of early-stage disease detection where minor differences can be of great diagnostic importance.

An additional advantage of the data augmentation is that it can be used with transfer learning. Although the pre-trained models offer general visual representations, small datasets of medical data can still be overfitted by fine-tuning in case the same examples are presented repeatedly to the network. Augmentation is simply an effective way to increase the effective dataset size giving the model new contexts to use its already learned filters. This not only makes the model more adaptable to the target domain using its high-level feature representations but also makes it use both the general and domain-specific features in a complementary way.

More advanced techniques like elastic deformations, random erasing, mixup, and cutout may be used to further enrich the dataset besides classical methods of augmentation. Elastic deformations add tiny smooth distortions to images that mimic anatomical variability, and mixup and cutout add or remove parts of images to advise the model to pay attention to the global patterns, not a particular localized area. Such strategies are especially beneficial to attention based networks as they put the attention mechanism to the task of identifying salient features in changing conditions.

In general, data augmentation is an essential part of the pipelines of the deep learning in lung CT classification. Augmentation enhances model generalization, model robustness, and clinical reliability by artificially increasing the training set, realistic variability, and diagnostic integrity. In the suggested AttVGG16 structure, the idea of augmentation does not only aid in efficient learning of hierarchical features but also increases the capacity of the attention module to highlight the clinically relevant regions at a consistent rate, leading to the improvement of the sensitivity, accuracy, and overall diagnostic performance.

3.3 Best Classifier Selection

Transfer learning has become one of the most effective approaches of contemporary deep learning, especially where such labeled datasets are scarce or prohibitively costly to acquire, such as medical imaging. The general principle of transfer learning is to make use of knowledge obtained in a large-scale source domain to implement it in a target domain where there are fewer training examples. In image analysis, this usually consists of a process where one uses a model already trained on large amounts of data like ImageNet to extract general visual features which are then refined on a smaller, more task-specific dataset. This strategy is useful to replicate the representational capacity of deep neural networks to a task-specific use case, thereby shortening the training period, alleviating the problem of overfitting and enhancing generalization.

Transfer learning has a number of benefits in the context of detecting lung cancer with help of CT images. CT data tends to be non homogeneous in nature and labeled samples are scarce owing to the complexity and cost involved in medical annotation. Examples on such little data may cause overfitting and unreliable convergence when training a deep neural network

architectures. The network is based on initializing the model with pre-trained weights, and thus the parameters directly encode rich and diverse visual representations. Given parameters can be such basic characteristics as edges, textures and shapes that can be cross-domain transferable. The initial layers of the network that encode these low-level characteristics do not change significantly during fine-tuning, and the later layers are modified to encode high-level characteristics that are important to the morphology and pathology of lung tissues.

Transfer learning is also known to hasten the training convergence. The pre-trained model is already highly optimized, therefore, it takes fewer iterations to fit the new domain. This minimizes computation cost and training time which is very useful when using scarce hardware resources or time. In addition, the pre-trained weights also provide a kind of regularization to the optimization process, helping find a more stabilized and generalized solution, which minimizes the threat of overfitting to the small target dataset.

The other important benefit of transfer learning is that the representational richness is maintained whereas the domain specific nuances are adjusted accordingly. Visual differences among normal and abnormal tissues in the field of medical imaging are usually delicate and complex. Directly training a model might fail to represent these complex patterns particularly in situations where data is limited. Transfer learning offers a base of generalized perception using large natural image datasets, enabling the model to specialise on finer-tuning of domain-specific representations. This selective adaptation assists the model to identify complex features like subtle irregularities in texture, faint nodular borders or differences in tissue density- features, which are imperative in the accurate diagnosis but which can be easily lost by weaker models.

The use of transfer learning was one of the main elements in the training of all the CNN-based architectures (VGG16, ResNet50, MobileNetV2, and the proposed attention-enhanced VGG16) in this study. The convolutional layers of both the models were loaded with ImageNet pre-trained weights, and the latter layers of classification have been substituted and re-trained to suit the binary classification of lung cancer detection. In fine-tuning, a selective training procedure was used: the low-level layers were frozen to preserve the representation of general features, whereas the higher layers were made trainable to learn domain-specific attributes. This strategy offered a good trade-off between the already trained general features and the acquisition of the specific characteristics of the lung CT scans.

The structural features of the natural image and the medical image are both common at the low-level features which are one of the main reasons why transfer learning is effective in this area. As an example, edges, gradients, and textures that constitute the core of the higher pattern recognition can be found in both of them. The network can concentrate on learning the more complicated spatial interactions and morphological variations that characterize pathological areas

by reusing the filters that have already learned to recognize such simple visual patterns. The model therefore not only converges faster, but also generally has better generalization to other imaging conditions, scanners and patients with different demographics.

Moreover, transfer learning is helpful in providing model stability and interpretability. Deep networks can also have unstable gradient propagation due to random initialisation, which results in erratic behaviour during learning and unreliable performance. Pre-trained models, however, offer well-initialized weights, which stabilize the training dynamics, giving a less jagged loss convergence and greater reproducibility. This is stability, which is especially useful in medical use, where reproducibility and reliability are vital to clinical confidence.

Deep models are also made more interpretable using transfer learning. Because the initial layers of pre-trained networks still have the general-purpose visual filters, the activation of the feature in the fine-tuned model can be frequently associated with the intuitive image properties. The visualization (Grad-CAM or saliency mapping) may then be used to identify the areas that arise the most in model decisions enabling clinicians to know why a model considers a given CT scan to be malignant or benign. Such interpretability acts as the interface between the radiological reasoning of humans and deep learning algorithms and contributes to the increased acceptance of AI-assisted diagnostics in clinical settings.

The other significant aspect of transfer learning is that it has a regularizing impact. In small medicine datasets, the overfitting associated with high dimensionality of deep neural networks may cause the model to memorize training examples rather than learn patterns which are generalizable to new scenarios. Transfer learning addresses this problem by restricting the space of learnable parameters to the neighborhood of an existing set of weights, and thus biases the optimization procedure to explore the parameter space in the same direction as strong and transferable representations. Consequently, the fine-tuned model can be used to attain high accuracy and generalization when there is limited training data.

Practically, transfer learning also provides a flexibility of using the pre-trained model to some extent. Researchers may use a combination of one of various strategies according to the size of the dataset and their similarity of the domain: feature extraction, partial fine-tuning, and full fine-tuning. Convolutional backbone is frozen in feature extraction, and the retraining is performed on a new set of data only in the classifier. The method is computationally effective and effective in cases where the target data set is small and is like the source domain. In partial fine-tuning, a subset of deeper layers are unfrozen to fit more specialized features to trade off between generalization and domain adaptation. Full fine-tuning, in its turn, re-trains the entire set of layers of the pre-trained model, which allows the greatest degree of flexibility but is also more susceptible to overfitting in the case of small datasets. Partial fine-tuning was used in our approach to permit the appropriate adjustment whilst maintaining the soundness of the features learned previously.

Its effectiveness is further strengthened by the integration of transfer learning and attention mechanism as applied in the proposed AttVGG16 model. Although transfer learning offers a rich basis of visual knowledge, it is the attention modules that enhance this knowledge, by focusing on features of diagnosis interest. This synergy allows the network to not only inherit the overall visual intelligence of a pre-trained model but also dynamically redistribute its attention to those areas that have the greatest clinical relevance. Transfer learning coupled with attention therefore is a strong paradigm to medical image activities that is both robust and adaptive.

In a more general sense, the effectiveness of transfer learning in this work proves its ability to transform healthcare AI. Medical data is necessarily restricted because of privacy aspects, annotation cost, and patient heterogeneity, and thus large-scale end-to-end training is not feasible. Transfer learning fills this gap and enables models to be global consumers of visual knowledge without clinical specificity. It makes deep learning more accessible to medicine by minimizing data requirements and enabling high-performance models to be available in low-resource research environments.

To conclude, transfer learning was important in this research because it allowed the effective, stable, and accurate training of deep learning models to detect lung cancer. The models were able to converge quicker, generalize more and have higher interpretability than could be achieved by training on fresh data, also through reusing and fine-tuning prior knowledge gained by large datasets. Together with attention mechanisms, the transfer learning allowed not only to boost the feature extraction capacity of the network but also to make it more clinically relevant such that the suggested system may become a consistent and interpretable tool that may be applied to the diagnostic practice in real-world settings.

3.4 Proposed Method

We suggest an attention-enhanced VGG16 (AttVGG16) computation framework to classify lung cancer using CT scans, which intends to use the deep hierarchical feature extraction together with a variable attention approach to enhance the ability of these deep features in classifying cancer. The model has incorporated a sparse attention module into the baseline VGG16 backbone, making it possible to effectively concentrate on discriminative spatial regions of CT slices. This selective focus improves the image of localized pathological characteristics, including nodules or tissue abnormalities which are important to detect more correctly at an early stage. Through this, the network avoids the possibility of incorrectly labeling the low-contrast or low lesion regions, which are normally difficult to distinguish with conventional CNNs.

VGG16 proved to be the best feature extractor to use in this task according to baseline tests with VGG16, ResNet50, MobileNetV2, and Vision Transformer (ViT) showed consistent high performance and stability with a balanced dataset of CT scans. Based on these findings, we

added the attention mechanism to VGG16 to create AttVGG16. Under this architecture, input CT images are first operated using the convolutional layers of VGG16, which generate deep hierarchical feature maps, which have low to high-level representations of pulmonary structures. Such feature maps are then optimally adjusted by the attention module which dynamically recalibrates channel-wise and spatial feature to highlight areas of diagnostics importance and silence background noise.

They are then subject to the refined features that are further classified into three groups out of fully connected layers: normal, benign, and malignant. The attention process enables the network to dynamically place more weight on features at various spatial positions and channels and to maintain fine-grained anatomical structure upon which is needed to differentiate between subtle disease patterns. This hierarchical property of feature extractions and adaptive attention weighting does not only enhance sensitivity to small or ambiguous nodules but also the robustness of the network to inter-patient variation of the CT imaging conditions.

In order to further improve the performance of the models, we also use various training methods such as transfer learning using ImageNet-pretrained weights, data augmentation (rotation, scaling, flipping, and intensive adjustments), and an early-stop to avoid overfitting. We also hyper-optimize factors like learning rate, batch size, and attention module congruency to trade off between classification accuracy and computational efficiency. The interpretability of attention heatmaps produced when observing inference has also been found useful, as clinicians can see where their input affects the model-based predictions and trust automated decision-making.

Generally, the AttVGG16 is capable of merging the advantages of the deep hierarchical feature extraction with the attention-driven spatial prioritization and enhances the ability to detect localized pathological abnormalities and fine pathological patterns. Its capacity to highlight clinically significant regions, as well as strong classification performance, renders AttVGG16 a prospective framework to detect early lung cancer and potentially be integrated into computer-aided diagnostic systems, which will eventually help radiologists in enhancing patient outcomes and minimize diagnostic errors.

3.5 VGG16

3.5.1 Introduction

In our case, to perform the image classification, we used the VGG16 architecture, which is a deep convolutional neural network presented by Simonyan and Zisserman (2014) [8]. The network has 13 convolutional layers and 3 fully connected layers, in five convolutional blocks with each group having a max-pooling operation to gradually decrease the spatial dimension. Each convolutional layer is based on 1-stride kernels and ReLU activation functions are used to introduce non-linearity so that the model can learn complex hierarchical features. Multi-class

classification is based on the final fully connected layer, with the addition of a softmax activation, which results in probabilities of the classes.

VGG16 architecture is a step forward in the development of the convolutional neural network, especially in the classification of images. VGG16, created by Simonyan and Zisserman in 2014, is famous by its simplicity and design uniformity, based on a deep stack of small convolutional filters to obtain an extraordinary representational power. In contrast to the previous CNN architectures that employed large and heterogeneous convolutional kernels, VGG16 is based on the consistent strategy, and it just employs the convolutional filters in all layers and in between layers is max-pooling operation that successively downsamples the image without losing significant information about features. The design is a simple, yet effective, way to model the complex hierarchical characteristics of the network using the number of parameters that are easy to manage, contributing to the efficient learning process and generalization.

VGG16 is a network comprising of 13 convolutional and 3 fully connected layers containing 16 weight layers in total, thus the name. The structure of the network is such that there are five convolutional blocks, separated by max-pooling layers. Convolutional layers are arranged in every block in order to extract more and more abstract features. The initial layers of the network are based on low-level attributes like edges, corners, and textures, which are the essential attributes of visual representation of objects. More complex patterns that include shapes, contours and local motifs are encoded by middle layers whereas high level semantic features that are important to make the difference between one and another category are encoded in deeper layers. This level of hierarchical feature extraction is especially useful in medical imaging: the small changes in tissue texture or nodule morphology can be pathological.

One of the strong sides of VGG16 is that it is able to use transfer learning. Weights obtained by training large-scale datasets like ImageNet give the network a concept of what a generic visual feature should be. When trained on problem-specific data, e.g. lung cancer diagnosis on CT scans, VGG16 can learn domain-specific higher-level representations that capture the anatomical features and pathological patterns of the problem. The method helps to overcome the problems of small labeled medical data, minimizes overfitting risk, and greatly accelerates the convergence during training. VGG16 has become popular in many medical imaging tasks, including tumor detection, organ segmentation and disease classification, due to the transfer learning ability.

The other remarkable aspect of VGG16 is that it uses ReLU functions of activation that add non-linearity to the network and they enable the network to capture more complex relationships between the input data. A convolutional layer is preceded by ReLU, and this allows the network to respond selectively to the relevant features and suppress noise. This is a critical property in medical imaging where variations are usually present in images based on patient anatomy, scan protocols and artifacts of the images. Relu is also used to solve the vanishing gradient problem,

and a stable gradient propagation is achieved regardless of the number of stacked layers, despite the backpropagation of the gradient.

The fully connected layers of VGG16 which are located on the top of the network are used as a classifier through the taxation of the spatially distributed feature maps into a holistic representation. All of these layers are used on all the previous convolutional layers to make the final class predictions. In typical applications, the network ends with a softmax layer that generates the class probability of each class enabling the use of the network in multi-class classification. The deep convolutional feature extraction combined with dense fully connected layers is what gives VGG16 a potent ability to differentiate the indistinct variations of complex image data sets and hence it is most applicable in differentiating normal, benign, and malignant pulmonary nodules.

Nevertheless, VGG16 has certain weaknesses, namely, it is very time-consuming because of depth and a significant number of parameters. As a resource constrained environment, the network cannot be used in many environments with an original form size of more than 138 million parameters, requiring a lot of memory and processing power, which is often scarce. This problem has led to the implementation of tricks like model pruning, knowledge distillation and hybrid networks combining VGG16 with more efficient components, such as attention mechanisms or lightweight convolutional blocks. These improvements enable the model to maintain its capacity to extract its features and minimize its computational burden, becoming more applicable to clinical implementation in cases where fast inference is required.

VGG16 has performed very well in the setting of lung cancer detection as it is able to follow the subtle differences in the CT scan images that can be linked to various stages of malignancy. The richness and consistency of the convolutional layers allow the network to resolve fine variations in nodule shape, margin features and tissue density in the surrounding, which are important markers of early diagnosis. The hierarchical nature of the network is such that the local features are micro-calcifications, represented in the final feature maps and the global contextual features represented in the same feature maps, which is the overall lung morphology. Such multi-scale representation makes the network more sensitive to small nodules that could be missed by visitation in the shallow ones.

Preprocessing and data augmentation are important in the optimization of VGG16 to medical imaging. The network is presented with a wide range of input scenarios by adding variations in rotation, scaling, flipping and intensity thereby enhancing the robustness and generalization to unknown cases. Additional preprocessing methods including normalization, windowing and noise reduction improve the visibility of clinically relevant structures and enable VGG16 to extract clinically significant features of high-resolution CT images. Combined with transfer learning, these measures assure that VGG16 works well even with comparatively small data sets

that is frequently the situation in medical imaging research thanks to the small patient groups and ethical issues.

VGG16 has a high degree of adaptation that can be expanded to include a combination with attention mechanisms, whose additional contribution to the discriminative power is significant. The network can also prioritize the areas of the image that are most appropriate to diagnosing a disease, which may be salient in a feature map, which can be reduced to background noise and irrelevant details, by adding modules that emphasize them. This is especially beneficial in lung CT images where nodules can take up a little fraction of the scan and even the slight change in intensity or texture can have important diagnostic value. The attention-enhanced VGG16 models have demonstrated higher sensitivity, accuracy and the general predictive reliability implying the significance of integrating hierarchical feature extraction and adjustable weighting of the important regions.

Besides classification, VGG16 has also been scaled down to use in segmentation, detection and multi-task learning in medical imaging. Using VGG16 as the backbone and skip connections, researchers have used VGG16 to build U-Net and Mask R-CNN networks to achieve accuracy in localizing pathological structures. This flexibility is a sign of the ability of the network to not only classify but also offer spatially resolved predictions helping to interpret the clinical context and aid in treatment planning. The fact that VGG16-based feature maps are interpretable, particularly with the help of visualization, such as Grad-CAM, gives clinicians an insight into what their model predictions rely on, which builds trust in AI-assisted diagnosis and makes it possible to integrate it into regular clinical practice.

To conclude, VGG16 continues to be a benchmark network in the medical imaging domain of deep learning due to its deep, uniform convolutional, hierarchical feature extractive nature and its ability to use transfer learning and attention. Its capability to develop both low-level and high-level features, as well as its integrability with augmentation strategies and attention modules, makes it especially successful in complex tasks like the lung cancer detection on the CT scans. It is computationally intensive, but still its demonstrated performance, adaptability and interpretability remain the standard by which other advanced CNN-based medical imaging solutions are evaluated and developed. The sustained role of the network in both research and clinical practice supports the lasting utility of the network as a strong and flexible tool of automated image analysis.

The methodology of Sutskever et al. (2013) [17] was used to train the model, which is to use Stochastic Gradient Descent with momentum and categorical cross-entropy as the loss function. To take advantage of transfer learning, the model was started with weights that had been fine-tuned on ImageNet and so converged as fast as possible in addition to performing better on our CT scan dataset despite the limited number of labeled samples. In order to avoid overfitting, we used data augmentation, such as rotation, flipping, scaling and variations of intensities to

make sure that the network performs well on unseen samples. Convolutional layers were optionally followed by batch normalization to make training more stable and speeds up convergence.

3.5.2 Equation

This structure enables VGG16 to capture low-level features, like edges and textures, and high-level features, like complicated anatomical patterns that are pertinent to pulmonary nodules. Having a uniform architecture and high hierarchical representation, it is highly applicable to the process of medical image classification and can be a good baseline when comparing it with more sophisticated engines, such as attention-enhanced and transformer-based networks.

Convolution and pooling are defined as:

$$f_{\text{out}} = \sigma\left(\mathbf{W} * f_{\text{in}} + \mathbf{b}\right), \text{max-pooling: } 2 \times 2 \text{ with stride } 2$$

and the softmax output is:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}.$$

Weights are updated via SGD with momentum μ:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla L(\mathbf{w}_t) + \mu(\mathbf{w}_t - \mathbf{w}_{t-1}),$$

with categorical cross-entropy loss:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i).$$

To avoid overfitting, we used data augmentation methods [17], such as random rotations, horizontal and vertical flips, scaling and intensity variations, which heighten the variety of the training samples and enhance generalization to unseen data. Also, the network was trained using ImageNet weights [25], which has been pre trained, and thus allowing transfer learning to use learned feature representations with the large-scale natural image data, several megawatts, accelerating convergence and boosting performance on the small medical imaging data. The training was also stopped early in case of no further improvement of the validation loss, which allowed the stopping of overfitting and guaranteed the solid performance of the model.

VGG16 is a popular feature extractor, which is able to encode rich hierarchical features such as low-level edge and texture features up to high-level anatomical structures. It has a deep and smooth convolutional structure that improves classification in challenging tasks, including multi-class lung cancer detection of CT scans [26]. Pre-trained weights, data augmentation, and early stopping concur that VGG16 has a high level of accuracy and reduces overfitting, and a solid baseline is established with which other models can be compared.

3.6 ResNet50

3.6.1 Introduction

To classify the images, we used the ResNet50 architecture, which is a deep residual network by He et al. (2015) [9]. In contrast to standard CNNs like VGG16, ResNet50 uses residual connections which enable the network to learn identity mappings, which results in alleviating the degradation issue of very deep networks and allows more easily passing gradients through backpropagation. The architecture has 50 layers, which feature convolutional, batch normalization, and bottleneck blocks, which enables hierarchical feature extraction of low-level edges to high-level semantic features.

To help improve performance and avoid overfitting, we used ImageNet pre-trained weights [25] to initialize ResNet50 and used data augmentation techniques including random rotations, flips, scaling, and adjustments of brightness to augment the diversity of training samples. Validation loss was also used to do early stopping to prevent overfitting to get a generalizable model performance. The residual design of ResNet50 enables the network to learn fine feature representations without causing any instability during training; thus, the network is especially useful when classifying the lung CT images to distinguish between subtle pathological features.

3.6.2 Residual Learning

Each residual block learns a mapping:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathbf{x}$$

Where x is the input, y the output, and $\mathcal{F}$ is the residual function consisting of stacked convolutions with batch normalization and ReLU activation.

3.6.3 Convolution and Activation

Convolutions are defined as:

$$\mathbf{f}_{out} = \sigma(\mathbf{W} * \mathbf{f}_{in} + \mathbf{b})$$

with strided convolutions and global average pooling reducing spatial dimensions [8].

3.6.4 Output Layer and Training

The final layer uses softmax:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

and the model is trained with SGD and momentum, optimizing categorical cross-entropy:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

3.7 MobileNetV2

3.7.1 Introduction

MobileNetV2 was proposed by Sandler et al. (2018), and it is a lightweight but powerful convolutional neural network structure that is optimized to work with mobile and embedded vision tasks. It is based on the principles of its predecessor MobileNetV1, with the additional use of depthwise separable convolutions that lead to a significant reduction in the number of parameters and computational complexity with competitive accuracy. The major innovation of MobileNetV2 is the introduction of and which allow the representation of features efficiently and gradient propagation even in very resource-constrained settings.

Unlike the traditional residual block employed by architectures such as ResNet, where the residual connections join feature maps of increasing size between the input and output, MobileNetV2 has an inverted structure where the residual connections are made between thinner bottleneck layers, and the intermediate layers are temporarily fattened to perform convolutional operations on them. A bottleneck block has three major parts: an expansion convolution to enlarge the dimensionality of the input, a depthwise convolution to filter the spatial information on each channel independently, and a final 1x1projection convolution to reduce the expanded features to a low-dimensional representation. The non-linearity of low-dimensional feature spaces would result in information loss, which is avoided by using a linear activation function (as opposed to the ReLU) following the projection layer.

The capability to balance between the size of the model and its representational capabilities is one of the biggest advantages of MobileNetV2. Its design renders it efficient in running on

devices with limited computational resources thus suitable in large scale implementation and in real time medical imaging application. MobileNetV2 can process high-resolution CT images in seconds without much accuracy loss due to its lightweight architecture, whereas its depthwise separable convolutions can guarantee the capture of both local and global features of the lung nodules. Moreover, the vanishing gradient issues are reduced with the help of the residual connections that guarantee the stability of convergence during training.

On the whole, MobileNetV2 provides a method of fast computation in image-based diagnosis systems. Its high capacity to elicit discriminative characteristics of CT images and combined with its low memory footprint, it is an excellent candidate to be deployed in clinical environments where hardware is limiting. Nevertheless, its representational depth can be shallow compared to more profound architectures, e.g. VGG16 and ResNet50, and thus may result in poor performance on highly complex data. However, due to its scalability and efficiency, it is a useful part of the modern medical deep learning pipeline, particularly, when augmented with attention mechanisms or transfer learning to refine its diagnostic accuracy.

In case of image classification, we used MobileNetV2, which was proposed by Sandler et al. (2018) [10]. In contrast to traditional CNNs like VGG16 [2] and ResNet50 [3], MobileNetV2 is designed to be computationally efficient via depth wise separable convolutions, reverse residual networks, and linear bottlenecks.

3.7.2 Depthwise Separable Convolution:

Standard convolutions are factorized into depthwise and pointwise (1×1) convolutions, reducing computational cost from:

$$\text{Cost}_{\text{standard}} = D_K^2 \cdot M \cdot N \cdot D_F^2$$

to

$$\text{Cost}_{\text{depthwise}} = D_K^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2$$

3.7.3 Inverted Residuals and Linear Bottleneck

The core block expands the input, applies depthwise convolution, and projects back to a low-dimensional bottleneck:

$$\mathbf{y} = \mathbf{W}_p \, \sigma\big(\mathbf{W}_d \, \sigma(\mathbf{W}_e \mathbf{x})\big)$$

with W_p,W_d ,W_e and representing expansion, depthwise, and projection convolutions, respectively, and σ as ReLU6. Shortcut connections are applied when input and output dimensions match:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}$$

3.8 Activation and Output

3.8.1 Introduction

Rectified Linear Unit or ReLU has become an essential part of modern deep neural networks design because it is easy to implement, efficient and highly empirically successful. It is a nonlinear transformation, which decides whether a neuron is to be activated, which enables neural networks to represent more complex relationships in data. ReLU is non-linear, yet without sacrificing computing performance, and thus it is highly applicable in large-scale image classification (such as lung cancer detection with CT scans). Contrary to other traditional activation functions like sigmoid or hyperbolic tangent, which squeeze the value of input into a narrow range, resulting in gradient saturation, ReLU preserves a linear response when the value is positive, and sets any negative values to zero. This acts to prevent the vanishing gradient problem that at one time prevented the training of deep networks as it leads to a reduction in gradient as one backpropagates through numerous layers.

Among the key properties of ReLU, the sparsity of neural activations should be mentioned. Because all negative inputs are zeroed, there are only a limited number of neurons to execute a particular input so the network can therefore allocate computational resources on the best features. This selective activation does not only increase the ability of the network to acquire discriminative features, but also decreases overfitting since neurons get trained to respond to unique meaningful patterns. The sparse representation formed by ReLU has also been associated with an improved ability to generalize to unknown data, which is a positive quality in the context of medical imaging, when minor changes in texture and intensity can be signs of disease development. Moreover, ReLU also hastens training convergence since its derivative is easy and regular at positive values, which enables gradient descent algorithms to update rapidly and steadily.

Nonetheless, ReLU does not lack limitations. The most typical of the problems is the so-called dying ReLU problem, wherein particular neurons become permanently inactive throughout training. It is possible when the weights of a neuron make the neuron generate negative values as the result of all inputs, which is considered to off-turn this neuron as ReLU equals zero in this range. When that occurs, these neurons cease to play a role in the learning process that can decrease the expressiveness of the model as a whole. To overcome this the following extensions have been created: Leaky reLU, Parametric reLU and Exponential Linear Units which permit a small non-zero gradient on negative values to ensure that neurons in the network remain partially active. Irrespective of these issues, the original ReLU fun is one of the most viable and efficient activations that are popularly employed in applications of computer vision and natural language processing.

ReLU has been crucial in medical image analysis based on deep learning to extract and maximize the advantages of relevant structural and textual features of complex imaging data. Consider, in the case of CT scans of lung tissues, ReLU allows the network to extract high-level spatial information, e.g. tumor boundaries, nodule density, and tissue abnormalities, by selectively activating neurons that represent such patterns. Such local activation assists the model to better differentiate between healthy and malignant areas. Its simplicity, sparsity, and strength make ReLU an inseparable element of convolutional neural networks and a major contributor to the success of such architectures as VGG16, ResNet50, and their variations that are applied in medical diagnostics.

On the one hand, the effect of ReLU goes further than the process of activation, as it signifies the transition to the efficient, scalable, and biologically plausible learning processes that have defined the development of modern deep learning. The ReLU remains to be the backbone of models performing well across a wide range of applications, such as object recognition or disease classification, because it enables networks to learn highly hierarchical representations and maintain computational tractability. Use of ReLU in deep medical models like attention-enhanced VGG16 is known to give consistent gradient flow, increased rate of convergence and better results in locating minor abnormalities, which has cemented its status as a fundamental element of deep neural network design.
ReLU6 activation,

$$\sigma(z) = \min(\max(0, z), 6)$$

improves robustness for low-precision computations. The final fully connected layer uses softmax to produce class probabilities:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

3.8.2 Training

MobileNetV2 was trained using SGD with momentum [2], optimizing categorical cross-entropy:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

MobileNetV2 achieves competitive accuracy with significantly reduced parameters and FLOPs, making it suitable for real-time or resource-constrained applications.

3.9 Vision Transformer (ViT)

3.9.1 Introduction

ViT is a computer vision model that is based on transformers, indicating a significant shift in the field as the previous models were convolution-based models, while the new models were natural language processing models. In contrast to classical convolutional neural network, which is based on local receptive field and hierarchical feature extraction, the Vision Transformer learns the global interaction between image regions with a self-attention network. This allows the network to distinguish local texture features along with the long-range spatial features and gives more complete picture of the visual scene. ViT splits an image into a fixed number of patches, which are handled similarly to the word token in language models. These patches are linearly appended to vectors whereby they construct a sequence that undergoes several layers of self attention and feedforward transformations. This design removes the spatial locality bias of convolutional networks and, through this design, learns data-driven, highly flexible representations of the spatial relationship.

One of the main benefits of ViT is that it is capable of learning contextual information on a global scale of the whole image. Whereas CNNs rely on the locality of convolutional kernels and rely on depth to increase their receptive fields, transformers cover all regions concurrently. This implies that ViT is able to figure out the relationship between the far-flung sections of an image which is especially useful in medical imaging tasks in which pathological patterns can occur in far-flung or irregular areas. An example is a lung CT scan, where nodules or lesions can be at different sizes and locations, and by stacking information throughout the image, ViT can be able to see the finer details in the context that may not be easily identified by a convolutional model. This global modelling property causes ViT to be naturally highly suited to the tasks related to the fine-grained pattern recognition, structural analysis, and disease localization.

The lack of reliance on inductive biases in the design of ViT is one of the most notable features. Convolutional networks make heavy assumptions of spatial hierarchy and translation invariance, making them faster to train but being less flexible. ViT, conversely, is trained on the interactions between the data, in which every aspect of spatial relationships is learned by the self-attention process, without depending on any external memory or training. This is a data-driven method of learning that enables ViT to make generalizations across domains, yet it also creates a large dependency on large datasets. Since transformers do not include integrated spatial priors as CNNs, they need large quantities of labeled data to learn useful representations by default. This is of significant challenge in medical imaging in which labeled data may be scarce and costly to acquire. Consequently, a popular strategy to utilize the full potential of ViT is pretraining it on large natural image datasets and fine-tuning it on medical datasets.

Vision Transformers may also provide significant interpretability benefits in practice. The self-attention mechanism also offers explicit attention weights that can show the priority of various areas of the image by the model to allow researchers and clinicians to see which areas a model is focusing to in its decision-making process. This openness comes in especially handy when it comes to the healthcare sector, in which model interpretability is critical to the establishment of trust in automated diagnostic systems. Reviewing the attention maps, clinicians will be able to get a more insight on the anatomical regions that the model correlates with disease indicators, which helps human validation and makes decisions related to the clinical process.

ViT has its limitations, however. Its complexity in terms of computational cost and memory is significantly greater than that of conventional CNNs, and is largely because the attention operation is quadratic in terms of the length of the input. This may render ViT to be inefficient to high-resolution medical images without special care in model optimizations or patch sizes. In addition, the stability of ViT training is also very susceptible to strong regularization methods and data scaling that are necessary to stop overfitting. However, these issues are being overcome with the continued research and the creation of hybrid architectures that combine convolutional and transformer layers.

To conclude, the Vision Transformer is an effective and versatile system of understanding images, based on the use of global self-attention mechanisms, rather than local convolutions. Its capability to model long-term dependencies, some complicated visual semantics and its interpretability enables it to be considered an attractive alternative to the traditional CNNs in terms of medical image analysis. ViT, when used to detect lung cancer, has the capability of detecting subtle structural changes and spatial correlations that may not be easily detected by a convolution-based model. Despite the challenges that demand of data and computational overhead, the conceptual innovation of ViT has established a new direction of learning to represent images and it will be possible to have stronger, explainable and data-efficient medical imaging models in the future.

For image classification, we employed the Vision Transformer (ViT), a transformer-based model adapted from NLP to images (Dosovitskiy et al., 2020) [13]. ViT treats an image $(\mathbf{x} \in \mathbb{R}^{H \times W \times C})$ as a sequence of (N) non-overlapping patches $(\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)})$, each linearly projected into a (D)-dimensional embedding using $(\mathbf{E} \in \mathbb{R}^{(P^2 \cdot C) \times D})$. A learnable [class] token and positional embeddings $(\mathbf{E}_{pos})$ are added to retain global information:

$$\mathbf{z}_0 = [\mathbf{x}_{\text{class}}; \mathbf{x}_{p1}\mathbf{E}; \ldots; \mathbf{x}_{pN}\mathbf{E}] + \mathbf{E}_{pos}.$$

The capability of ViT to capture the global context is especially beneficial in medical imaging, where fine details of abnormalities can cross multiple areas and cannot be accurately represented by local convolutions only. Nevertheless, ViT is sensitive to overfitting on small datasets, which

is typically an issue in the medical field. To overcome this, we used transfer learning using the pre-trained weights of big datasets, and vast data augmentation techniques such as random rotations, flips, changing of intensities and elastic deformations to enhance generalization.

In addition, the patch-based representation is flexible in both addressing different image resolutions, as well as the model can pay attention to both local and non-local features at the same time. ViT was compared to VGG16, ResNet50, and MobileNetV2 in our experiment to compare its performance with these models in classifying multi-class lung cancer (normal, benign, malignant). Although ViT has shown good global contextual modeling, it is sensitive to the size of datasets, which indicates the necessity of adopting hybrid models to incorporate CNN-based feature extraction with transformer-based attention models to be considered in a better diagnostic prediction.

3.9.2 Transformer Encoder

Transformer Encoder is the basis of the modern transformer-based models, like the Vision Transformer (ViT) and BERT. It is made to handle sequential data by estimating dependencies between each element in the sequence using self-attention systems. Unlike the recurrent neural networks, which run in series, the transformer encoder can compute all the elements simultaneously, allowing the computation to be performed effectively and long-range relationships to be learned without the need to consider time. This similarity renders it especially effective in duties in which contextual knowledge about the whole input is required, since it can enable the model to dynamically score the weight of various components of the input in creating feature representations.

Structurally, a transformer encoder consists of several layers of the same material, each having two major components: a multi-head self-attention block and a position-wise feed-forward network. The mechanism of self-attention enables the encoder to assess the relationship between each token (or patch of image in ViT) and all other tokens in the sequence, and calculate contextualized representations that include all token information. With the utilization of several attention heads, the model acquires various feature subspaces, which allows it to learn to capture different kinds of dependencies at the same time. This multi-head architecture adds representational strength to the encoder, with each head specialising in a specific way of the data, e.g. texture, shape, or semantic context in visual tasks.

After the self-attention stage, a feed-forward network performs non-linear mappings on the representation of each token, which further enhances the features acquired by attention. After every sublayer, there are residual connections and layer normalization to stabilize training and enhance gradient flow, such that deep transformer encoders can be trained successfully. The residual paths retain original data and allow the network to acquire complex transformations without sacrificing important contextual information. The philosophy of design, which is a

combination of parallel attention computation, normalization and skip connections has played a key role in aiding the stability, as well as scalability of transformer encoders to large datasets and deep architectures.

Another important innovation in the transformer encoder is how it treats positional information. This is because the architecture does not have an inherent sense of sequence order because it is non-recurrent, so positional encodings are appended to the input embeddings to promote information about the relative or absolute position of every token. Positional encodings can be applied in image analysis positively so that the Vision Transformer can incorporate spatial awareness when handling image patches so that the model may be capable of relating neighbouring regions and keep spatial coherence. Such positioning data is vital to tasks that rely on vision, in which the relative positioning of features carries important structural information.

In medical imaging, the encoder the ability of the transformer to combine information in the whole field of view supports holistic feature extraction compared to convolutional methods. To illustrate, in the lung CT scans, the pathological indicators can be observed in small or spatially separate areas, and the encoder can be effective to associate the patterns into a single diagnostic feature. The receptive field is a global receptive area, which determines contextual relationship between the remote regions of the lung, which is enhanced to classify malignancy and characterization of lesions. To achieve this, as opposed to CNNs where a large stack of convolutional layers is needed to scale the receptive field, the transformer encoder does it with only one operation of attention, and therefore this is inherently efficient in learning global dependencies.

In addition, the interpretability of transformer encoders is also one of the main strengths in clinical use. The self-attention maps give visual clues to the areas that the model finds the most relevant during classification or detection, making the computational decisions consistent with human-intelligible patterns. Such openness is priceless to clinical faith, in that radiologists are able to check that the model can concentrate on anatomically critical characteristics and not on meaningless artifacts.

Transformer encoders, in spite of their advantages, are also associated with problems in computation. The self-attention operation is quadratically dependent on the sequence length, and its implementation requires high memory and processing requirements of high-resolution medical images. In more recent works, however, a number of effective attention variants and hybrid CNN-transformer architectures have been proposed to provide improved mitigation to this problem, only complexity-reducing and not performance-degrading. These hybrid models tend to use convoluted layers to extract local features and use transformer encoders to make global reasoning by synthesizing the best of two worlds.

Overall, transformer encoder is one of the key developmental steps of deep learning that allow models to learn contextual and globally aware representations of visual and sequential data. It is a cornerstone of modern medical image analysis frameworks because of its ability to model complex dependencies, have an interpretable form, and operate on a wide range of data modalities. The transformer encoder, when combined in systems such as the Vision Transformer or attention-enhanced CNNs, forms the basis of more precise, transparent, and generalizable diagnostic models taking the limits of automated disease analysis and detection.

The encoder has *L* layers, each with Multi-Head Self-Attention (MSA) and MLP blocks with LayerNorm (LN) and residual connections [24]:

$$\mathbf{z}'_l = \mathrm{MSA}(\mathrm{LN}(\mathbf{z}_{l-1})) + \mathbf{z}_{l-1},$$
$$\mathbf{z}_l = \mathrm{MLP}(\mathrm{LN}(\mathbf{z}'_l)) + \mathbf{z}'_l.$$

Self-attention for a single head is computed as:

$$\mathrm{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}.$$

The MLP block contains two linear layers with GELU activation [14] :

$$\mathrm{MLP}(\mathbf{z}) = \mathrm{Linear}(\mathrm{GELU}(\mathrm{Linear}(\mathbf{z}))).$$

3.9.3 Classification Head

The last step in a neural network is classification head that transforms the extracted feature representations into the final output of the prediction that needs to be made e.g. class probabilities in image classification. Once the backbone or feature extraction network (such as CNN layers or Transformer encoders) has been run on the input data to produce a small feature map or embedding, it is fed into the classification head to be used to make a decision. It is basically a facilitator between the acquired high-dimensional characters and the target characters, which maps abstract features into interpretive forecasts.

The classification head of a deep learning architecture often includes a single or more fully connected (dense) layers, successively shrinking the dimensionality of the feature vector. The last layer is usually the number of the output classes in the number of the neurons, and either a softmax or a sigmoid activation function (when it is a multi-class problem), or a sigmoid activation function (when it is a binary classification problem). To enable interpretable predictions, the book uses the softmax function to normalize the output to a probability distribution over classes.

The classification head is very important in the optimization and the generalization capability of the model. The design of the model has an influence on the ability of the model to map learned representations to the appropriate class boundaries. As an example, dropout layers or batch normalization to the classification head may also counteract overfitting and make the model more robust in that it is not over-dependent on particular features. Likewise, when activation functions like ReLU or GELU are used, non-linearity is presented and therefore the head can learn complex decision surfaces.

In current architectures, like Vision Transformers (ViT), the classification head is a little bit distinct to the conventional convolutional networks. Once Transformer encoder has achieved input image patches, the resulting sequence of embeddings is extracted and the one that represents the [CLS] token (classification token) is obtained. This token is a summary of all patch images and it is sent to the classification head (typically a linear (fully connected) layer) in order to obtain the final class likelihoods. In this manner, the model can be capable of classifying on a global perception of the input image.

Conversely, CNN-based models like VGG16 or ResNet50 generally reduce the end convolutional feature image to a 1D vector then forward it on to one or more fully connected layers that make up the classification head. These layers decode the spatially aggregated information of the convoluted backbone and assign probabilities of the classes respectively.

The performance of the head of classification can have a great impact on the overall performance of the final model. A model could have a powerful feature extractor, but a poorly designed head could restrict the ability of the model to classify. Conversely, a well-optimized head that is consistent with the structure of the extracted features has the potential of becoming more discriminative and more stable throughout the training.

In other architectures, researchers employ specialized classification heads that are application specific. As an example, multi-task learning models can be used with several heads which are focused to various outputs (e.g., classification, segmentation and localization). Classification heads can also have normalization and projection layers in attention-based or transformer architecture to focus the learned features to prediction.

Altogether, the classification head is one of the crucial elements that complete the model interpretation of the data. It summarizes the finiteness of the learning process of the network and directly defines the capability of the model to make correct and confident predictions. All these properties of its structure, activation functions, and regularization strategies together play a role in making the model more precise, generalizing, and stable in inference.

The final [class] token $(\mathbf{z}_L^0)$ is LayerNormed and passed through a linear layer to produce logits $(\mathbf{z})$. Class probabilities are computed via softmax:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}.$$

3.9.4 Training

The model was trained end-to-end using AdamW [15] and categorical cross-entropy:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i),$$

with data augmentation (RandAugment) and dropout in MLP blocks to mitigate overfitting due to lack of convolutional inductive bias.

3.10 Attention-Based VGG16

3.10.1 Introduction

The Attention-based VGG16 model is an improved model of the traditional VGG16 architecture, which involves an attention mechanism to enhance the capacity of the network to concentrate on the most informative and discriminative parts of an image. Although the original VGG16 architecture is very efficient in hierarchical features extraction by using a set of convolutional and pooling layers, it considers the entire spatial areas equally. This can cause inefficiency in complex vision tasks like medical image classification where important diagnostic information can be localized with small or subtle regions. To eliminate this shortcoming, the attention-based VGG16 architecture incorporates the attention modules to allow the network to prioritize important details while downplaying the background irrelevant information.

In the proposed research, VGG16 architecture is modified to incorporate the Convolutional Block Attention Module (CBAM) and this form of AttVGG16 is developed. CBAM module is a light-weight, plug-and-play module that refines feature maps in two complementary dimensions: channel and spatial. The channel attention scheme assists the model in knowing what to pay attention to by giving the channel features weights of importance, which attaches more weight to the channel features that are more related to the target class. Conversely, the spatial attention mechanism gives the network an instruction of where to pay attention by pointing to the most informative space in the feature map. The two attention processes in combination are sequential in nature so as to increase the representational strength of the model enabling it to acquire finer and meaningful patterns.

The attention-based VGG16 has a similar architecture as the original VGG16 up to convolutional blocks. The CBAM modules are placed after some of the convolutional layers. The feature maps of these modules are refined and sent to the next block with the network gradually learning

attentive to feature representations (spatially and channel-wise). The model is able to apply this integration to dynamically re-calibrate its focus during training to enhance interpretability and efficiency. The last layers of the model are fully connected layers that comprise the classification head, which is used to give the class probabilities through a softmax activation profile.

The greatest benefit of attention-based VGG16 model is that it is more discriminative. In medical imaging studies, including detection of lung cancer, classification of tumours or identification of lesions, the model is trained to focus on diagnostically important details (such as tissue abnormality or texture detail) and ignore noise or non-relevant structures. Selective focus results in improved generalization and accuracy because the model makes better use of its capacity. Also, the attention addition does not raise the computational cost significantly since CBAM is expected to be lightweight and can be easily plugged into current CNN backbones.

Training of the attention-based VGG16 is based on the standard procedures of deep learning, but it is more convergent and more stable. Since attention mechanism offers better gradients due to the salient-based emphasis, the process of optimization becomes more efficient. The attention maps during backpropagation ensure that meaningful parts are reinforced while minimizing the risk of the model overfitting due to the occurrence of irrelevant patterns. The visualization of the learned attention maps also proves that the network actually pays attention to the medically relevant areas that in most cases are not present in deep learning models.

Compared to attention-based VGG16, the vanilla variant and other baseline models, including ResNet50 and MobileNetV2, the attention-based VGG16 has shown superior performance when compared on a variety of performance measures, such as accuracy, precision, recall, F1 score, and AUC. Improve it especially on sensitivity (recall), which means that the model is better suited to detection of positive cases (which is critical of a clinical diagnostic task since a misdiagnosis of a disease case might be lethal).

Finally, the Attention-based VGG16 (AttVGG16) is a hybrid model, which incorporates the strong hierarchical feature extraction of the original VGG16, and the adjustment ability of the CBAM attention mechanism. Such synergy allows the network to not only extract rich visual semantics, but also guide its computational effort to the most informative parts of interest. What comes out is a highly accurate model that is highly interpretable and has immense potential in terms of applications in medical image analysis, especially where fine-grained visual discrimination and a confident diagnostic decision-making process are required.

We have trained attention-enhanced architecture which combined convolutional feature extraction and attention in order to improve feature extraction to identify lung cancer on the basis of CT. Although the hierarchical spatial features were represented by the baseline VGG16 model (Simonyan and Zisserman, 2014) [8], which added the max-pooling operations between the stacked convolutional layers with 3×3 kernels, the attention module was added

following the last convolutional block to enhance the discriminative representation with the re-weighting of both channel and spatial features.

The attention module employed is the Convolutional Block Attention Module (CBAM) [12], which is channel and spatial attention sequentially applied. Channel attention calculates channel-specific weight map using spatial-aggregated global average and max pooling after which a shared multi-layer perceptron (MLP) is used. This enables the network to prioritize channels which are the most significant to malignancy patterns, e.g. the difference in texture or density of pulmonary nodules. Instead, Spatial attention produces a two-dimensional attention map of the rank of important regions of the feature map, allowing the network to detect small nodules and delicate pathological features in the lung parenchyma. The learned feature maps are then subjected to fully connected layers and softmax classifier to generate normal, benign or malignant probabilities.

AttVGG16, a convolutional representation with adaptive attention weighting, is better sensitive to subtle radiological patterns, important in the detection of lung cancer at an early-stage of the disease. The attention block is modular, enabling it to be computationally efficient with very few extra parameters but a significant enhancement in the feature discriminability. Besides, attention visualization may offer interpretability, giving clinicians insight into the parts of the CT image that led most to the classification decision, thus removing the divide between automated diagnosis and clinical reasoning.

3.10.2 Attention Mechanism

Mechanisms of attention have become a revolutionary paradigm of deep learning, first popularized in natural language processing (NLP) and then applied to computer vision problems. On a high level, attention enables neural networks to pay selective attention to the most informative components of the input and repress the rest of the information that might be irrelevant or redundant. This selective attention is mostly useful in medical imaging, where important features, including small nodules in lung CT scans, have a minute portion of the image but a high level of diagnostic importance. Attention mechanisms outperform the predictive performance and interpretability of deep learning models by explicitly modelling the importance of various spatial regions or feature channels.

Attention in convolutional neural networks (CNNs) may be further divided into spatial attention or channel attention (or both). Channel attention considers what features are relevant by learning weights of each map of features in a layer. This is usually done through the combination of the spatial information through global pooling processes, average and max pooling, and then a small feed-forward network which produces channel-wise importance scores. The channels that model important pathological patterns, e.g. irregular tissue density or nodule morphology, are reinforced and channels that are less informative are suppressed. Spatial attention, in contrast,

underlines the location of important features of the image where. It creates 2D attention map on the spatial dimensions of the feature map, which enables the network to target localized areas that are more likely to have abnormalities, including lesions, nodules, or tissue deformations.

The Convolutional Block Attention Module (CBAM) [12] that we use in this paper combines both channel and spatial attention sequentially. The channel attention is applied first to recalibrate the feature maps along the channel axis and then the spatial attention is performed which labels the important spatial locations. This dual-attention mechanism is also light and can be directly inserted in an existing CNN architecture and it improves feature representation without much of a computational burden. Notably, the CBAM is designed in a modular way which makes it interpretable; the attention maps learned can be visualized and this gives information about which regions or features play the most significant role in the decision-making process of the model which makes it a transition between automated diagnosis and clinical thinking.

Attention based image processing has shown significant advances in various areas of medical image analysis such as classification of chest X-rays, retinal disease, classification of brain tumors and segmentation of brain tumors. Attention reaches this objective by allowing networks to focus on diagnostically important regions selectively to reduce the effect of background noise and irrelevant structure, false negativity, and dependence on variations in patient anatomy and imaging. Attention, in respect of the lung CT scans, where the nodules are often tiny, heterogeneous and not easily seen, allows accurate localization and characterization of the pathological aspects, which in the end leads to an early detection and improved accuracy in diagnosis.

In addition to CNNs, the architecture of Vision Transformers (ViTs) is based on an attention mechanism, through which one can model global dependencies between patches of an image using self-attention. In contrast to local convolutions, self-attention identifies long-range interactions, which offer contextual awareness, and that may be crucial in detecting subtle patterns that are spread throughout lung tissue. Although bigger datasets are needed to efficiently train transformers, the general idea of paying more attention to high-priority information instead of insignificant data is applicable to both CNN- and transformer-based models.

To conclude, attention mechanisms are an effective means of deep learning, which allows models to selectively attend to the most informative features within the spatial and channel dimensions. Our proposed AttVGG16, which is a counterpart of VGG16 with the addition of CBAM, capitalizes on the merits of attention to improve the representation of features, the quality of classification, and inherently provide an interpretative ability to detect lung cancer by using CT images.

We have used the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) [12], and channel and spatial attention to intermediate feature maps $(\mathbf{F} \in \mathbb{R}^{H \times W \times C})$. is sequentially applied.

3.10.3 Channel Attention

Channel Attention is a device that focuses on how significant various feature channels of a convolutional neural network (CNN) are. The feature maps or channels in a CNN represent different kinds of image patterns - edges, textures or higher level features. Not every channel is equally significant to the final decision, however, in medical imaging problems such as lung cancer detection, subtle abnormalities may be only represented by a few specific channels. Channel attention tries to solve this by weighting each channel adaptively based on its importance to the task.

The main concept of channel attention is to respond to the following question: what are more information-rich feature maps of the current input? This is usually done by having a mechanism summarising information of space per channel and producing channel attention weights. One of the methods involves global pooling operations, including global average pooling (GAP) and global max pooling (GMP) to reduce each channel to a single scalar, which quantifies the overall importance of that channel over the spatial domain. Such pooled descriptors are passed through a small neural network (usually a two-layer fully connected network with a non-linear activation, e.g. ReLU), and a sigmoid function is applied to produce normalized weights in the range [0,1] [0,1] per channel.

$$M_c(\mathbf{F}) = \sigma\big(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F})) + \mathrm{MLP}(\mathrm{MaxPool}(\mathbf{F}))\big),$$
$$\mathbf{F}' = M_c(\mathbf{F}) \otimes \mathbf{F}$$

3.10.4 Spatial Attention

Spatial Attention is meant to pay attention to the most informative points in a feature map as a complement to channel attention, which pays attention to the importance of the channels of a feature. Compared to traditional CNNs, which focus on all spatial locations, spatial attention enables the network to focus on areas that are more important to the task being undertaken e.g. a tumor lesion or minor nodules in pulmonary CT images. Such is especially useful in medical imaging, where diagnostic characteristics may be critical and can easily take up a very small portion of the image.

Spatial attention is commonly computed in two stages, which are aggregation and weighting. In the first step, the feature map is pooled by channel dimension with operation like average

pooling and max pooling to create a condensed 2D map which summaries the occurrence of features within spatial locations. The two maps are then fused and an operation called convolutional layer is applied to them usually with a 7×7 kernel size to encode spatial dependencies. The spatial attention map is created by a sigmoid activation that has weights ranging between 0 and 1 in every spatial position. Lastly, this attention map is multiplied element-wise with the original feature map, which increases the contribution of information regions and decreases irrelevant or noisy background regions.

Spatial attention is especially efficient in case of subtle and localized abnormalities. To have an example, in lung CT imaging early-stage malignant nodules can be just a few pixels in size but greatly significant in a proper diagnosis. Spatial attention by using these areas improves the sensitivity of the model, false negatives, and overall classification. Spatial attention, when added to channel attention in such architectures as the Convolutional Block Attention Module (CBAM), adds to a holistic feature refinement process that can focus on the significant channels and the locations of these features at the same time, producing more discriminative feature representations to deep learning models including AttVGG16.

Besides, spatial attention is interpretable. The resulting attention maps can be plotted so that it is possible to understand what areas the model perceives as the most important and it helps get insights into model judgments and helps in clinical validation. Such capability of saliency of diagnostically important targets is consistent with the aims of explainable AI, so spatial attention is especially useful in high-stakes fields such as medical imaging.

$$M_s(\mathbf{F}') = \sigma\Big(f^{7\times7}\big([\mathrm{AvgPool}(\mathbf{F}'); [\mathrm{MaxPool}(\mathbf{F}')]\big)\Big),$$
$$\mathbf{F}'' = M_s(\mathbf{F}') \otimes \mathbf{F}'$$

The refined feature map $(\mathbf{F}'')$ is then passed to the fully connected layers for classification.

3.10.5 Training Strategy

ImageNet pre-trained weights were used to get the network started to use previously acquired low-level and mid-level features because that enhances the speed of convergence and minimized the chances of overfitting the small lung CT dataset. Stochastic Gradient Descent (SGD) with momentum was used to fine-tune the model, and is a popular optimistic algorithm that uses gradient data but a proportional part of the prior change to speed up training and stabilize the training process [17]. Momentum aids the optimizer in going around flat areas, as well as,

prevent local minima, especially when utilizing deep networks with many parameters, like VGG16.

Model optimization was achieved by the use of categorical cross-entropy loss function, which is as follows:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Massive data augmentation (random rotations, horizontal flips, intensity normalization) and early stopping were also used to minimize overfitting.

is the predicted probability of similar class. This loss more severely punishes erroneous predictions that are far apart in prediction that is close to the true label and is appropriate in multi-class classification problems.

To enhance generalization and avoid overfitting, there were various regularization techniques that were utilized. The on-the-fly data augmentation methods (random rotations, horizontal and vertical flips, intensity scaling, random cropping) were implemented to the training data to make it more diverse and to allow the network to acquire robust and invariant features [17]. Validation loss was used to early terminate training when the model started to not improve any more after a specified number of epochs, avoiding overfitting as well as unnecessary computing costs.

Also the learning rate scheduling was used and the learning rate of plateaus in the validation loss was decreased so that the weights in the network could be fine-tuned during the later stages of the training. Mini-batch training was conducted using batch sizes that were chosen so as to strike a balance between computational efficiency and stability so that the estimates of the gradient are accurate and do not require a lot of memory. Fully connected layers were also introduction of dropout layers to randomly switch off neurons during training which enhanced model robustness further.

Lastly, the performance of the model was measured on a held-out test set over a variety of metrics such as accuracy, sensitivity, specificity, F1-score and Matthews Correlation Coefficient (MCC) to fully characterize the quality of classification. This cautious approach of pre-training, optimization, regularization and evaluation techniques allowed the network to acquire discriminative representations of lung CT images with a minimum of overfitting and maximum generalization to the unknown data.

The CBAM attention boosts feature representations which makes the model to concentrate on diagnostic regions which are important and enhances discrimination between benign and malignant patterns in CT scans.

Chapter 4

Results and Discussion

4.1 Introduction

This work advanced a set of deep learning frameworks, trained and tested to explore whether they can be used to detect lung cancer and identify it successfully and efficiently with no mistakes on the CT scan image. The compared models are four popular architectures, namely Vgg16, ResNet50, MobileNetV2, and Vision Transformer (ViT) model, and the proposed attention-enhanced VGG16 (AttVGG16) model. The assessment system was thorough, with several quantitative measures such as the accuracy (Acc), the precision (Pre), the sensitivity or recall (Sen), the specificity (Spe), the F1 score, the area under the ROC curve (AUC), and the Matthews correlation coefficient (MCC). All these measures gave a complex insight into the extent to which each model was able to identify malignant and benign lung tissues as well as effectively balance the trade-off between sensitivity and specificity- a critical element in clinical diagnostics.

As the experimental findings showed, VGG16 was the most successful of the baseline models as it showed better performance in most of the evaluation measures. It is characterized by its simple and highly convoluted architecture, with small 3 X 3 convolutional kernels arranged in a pyramidal manner, and it allows a hierarchical spatial feature representation. Such a design enabled the model to capture fine-grained information like texture anomalies, nodular edges and slight density differences in lung CT images. They are the key malignancy predictors, and the hierarchy of VGG16 structured features turned out to be effective in encoding them. This model obtained an impressive accuracy of 94.44 per cent and AUC of 0.9956, which is the strongest baseline in this area. It also had high F1 score indicating preciseness and accuracy in classifying cancerous spots without excessive false positives and thus it had a high degree of accuracy and consistency.

ResNet50, on the other hand, with its more complex and sophisticated layout (residual connections) still exhibited a lower performance level in comparison. The model had an accuracy of 60.56 percent and AUC of 0.9282. Such a fall can be explained by the small size of the data and by the characteristics of CT image data, which do not always contain large-scale variations that would allow utilizing the maximum potential of deep residual networks. Skip connections in ResNet50 are beneficial to avoiding the disappearance of gradients in extremely deep networks, and to a certain degree, they also induce redundant feature propagation when trained on small, homogenous data. As a result, the network might not highlight the importance of local

characteristics related to the early symptoms of cancer, being focused on global networks that are not of great importance when it comes to predicting the malignancy.

MobileNetV2, with its computational efficiency and depthwise separable convolutions, had a moderate trade-off between the speed and accuracy. It obtained an accuracy of 73.89 and the AUC of 0.8873. MobileNetV2 is resource-efficient and lightweight, which allows it to be used in real-time and environments with limited resources, but feature richness can suffer in many cases. Such reduced representational depth can be detrimental to the ability of the network to learn complex visual features in lung CT analysis, where discriminative features can be subtle and locally structured. However, MobileNetV2 has a respectable performance and convergence speed, which contributes to its usefulness as a baseline of deployment-oriented medical imaging systems.

Conversely, the Vision Transformer (ViT) model, though can theoretically model long-range dependencies using the self-attention mechanisms, did not show good results on this dataset, with an accuracy of 51.52% and an AUC of 0.5800. ViT performs so poorly because its reliance on large-scale and diverse training datasets is its first and foremost strength. Transformers tend to perform well in cases associated with enormous volumes of data as they acquire international relations among all image fragments. Nevertheless, in the medical imaging case, where annotated data is scarce and the input images are similar to each other across classes, ViT patch-based tokenization may disrupt spatial continuity that has meaningful information, resulting in underfitting and misclassification. Low precision (26.54) and F1 score (35.03) are also indicative of the fact that the model did not develop strong discriminative boundaries between the benign and malignant areas.

Given these shortcomings, the paper proposed the attention-based VGG16 (AttVGG16) model which incorporates Convolutional Block Attention Module (CBAM) into the default VGG16 model. The addition means that the network can also learn not only hierarchical feature representations but also dynamically pay attention to the most informative spatial and channel-wise regions of the CT scans. Each convolutional block produces the output which is enhanced by channel attention and spatial attention in the CBAM module respectively. Channels attention helps the network to stress feature maps that are strongly related to malignant features and spatial attention identifies the precise pixel locations that represent diagnostically significant structures. Such dual-level feature recalibration greatly improves discriminative capacity of the network, which makes it target more meaningful representations.

The numerical findings indicate that AttVGG16 attained significant performance improvement in all metrics. It received an AUC of 0.9959, accuracy of 97.78, sensitivity of 97.80, specificity of 98.90, MCC of 0.9178 and F1 of 97.76. The sensitivity and the F1 increase increase indicate that the model can better identify the real positive cases with a low false negative rate, which is an important factor in the medical diagnostic field, where a lesion can be crucial in terms of clinical

implications. Moreover, the upward trend in MCC shows that the model remains with balanced performance despite the existence of class imbalance which makes it reliably sound both on the positive cases and negative cases.

The excellent performance of AttVGG16 was also confirmed by the confusion matrices and receiver operating characteristic (ROC) curves analysis. The ROC curve of AttVGG16 exhibited an abrupt rise towards the upper-left corner, which means that its true-positive rate and false-positive rate are high and low, respectively. This trend shows that the attention mechanism was successful in making the model pick subtle textual information in both malignant and benign tissue. Comparatively, both the ROC curves of ResNet50 and ViT had flatter slopes and demonstrated weak discriminative abilities. The confusion matrix analysis confirmed that AttVGG16 was always able to decrease false negatives in comparison to all other models, which is why it has a more practical importance in the early and accurate identification of lung abnormalities.

A qualitative review of activation maps of features supplied additional information regarding the interpretability of the attention-based model. The visualised attention maps, indicated that AttVGG16 network focused on areas of diagnostically importance which include nodular edges, heterogeneous tissue densities, abnormal vascular appearances and inhibited irrelevant background, such as air spaces or normal parenchyma. Besides enhancing the accuracy of classification, this concentrated representation enabled a sensible explanation of model predictions as well, which fit the decisions of a human radiologist. Conversely, regular CNNs such as VGG16 and MobileNetV2 had more extensive and less specific activation areas and in some cases, indicated irrelevant areas that led to misclassifications.

Architecturally, CBAM integration to VGG16 was computationally efficient and it introduced very little overhead in terms of parameters to VGG16 but it produced significant performance gains. The flexibility of the attention mechanism allowed the network to dynamically redistribute its representational capacity during learning, enhancing gradient flow and eliminating overfitting. This feature was especially useful since the size of the dataset was quite small, and it is possible to learn discriminative features without overparameterization. Also, the attention weighting regularization effect encouraged the generalization to unseen samples and the robustness of AttVGG16 was further strengthened.

Another crucial methodological finding made by the results is the combination of hierarchical feature extraction and attention refinement being more advanced than deeper convolutional networks and transformer models on small-to-moderate medical datasets. Although ResNet50 and ViT have some of the best representational abilities, they are complex and need more data which limits them to less annotated data domains. Instead, the AttVGG16 takes advantage of a more basic but more focused methodology that is more efficient at features and less vulnerable to overfitting. Such trade-off between the simplicity of architecture and adaptive concentration

makes it especially beneficial in clinical applications, where the interpretability and reliability are of equal concern as the pure predictive power.

The other interesting finding was that AttVGG16 was robust to advantages of image-level variability, including variations in brightness, contrast, and scanner calibration. The recalibration of the feature responses by the attention mechanism enabled the model to stabilize the responses to varying conditions, resulting in stable performance in cross-validation. This toughness is essential to the reality-world implementation, where the CT images can have various sources of acquisition and vary in the acquisition settings.

All in all, Table 4.3.1 with Figure 4.1.1 supports the findings of the baseline models in that the VGG16 architecture indeed has an advantage over the rest of the baseline classifiers. Further, Table 4.4.1 and Figure 4.4.2 complete the fact that Attention Augmented VGG16 architecture is a better framework to detect lung cancer using CT scans. Its values of high accuracy, sensitivity and MCC values testify to good predictive factors as well as clinical reliability. The success of AttVGG16 is yet another indication of the importance of adding attention mechanisms to convolutional backbones because the latter allow deep networks to learn more like humans, namely, by paying attention to what really matters in the complicated visual information.

To sum up, AttVGG16 model is an important contribution to the sphere of medical image analysis. It is able to balance the structural advantages of VGG16 with the adaptive feature refinement of the attention mechanisms to have both increased performance and interpretability. This makes it a prospective tool of helping radiologists to diagnose lung cancer at its early stages, especially where huge datasets are not present, as it is in the resource-limited settings. Further studies can build on the basis, including the application of multi-scale attention strategies, integrating clinical metadata with imaging data, or even extending the model to multi-modal frameworks, which combine CT with histopathological or genomic characteristics. This may advance the diagnostic accuracy, strength, and translational capability of attention-based deep learning in cancer.

4.2 Performance evaluation metrics

There are a variety of performance measures, each offering a different view of the classification algorithm and each is a measure of success. We have used several measurements in this paper to fully realize the effectiveness of our lung cancer detection framework. The accuracy provides an idea of the reliability of the model with respect to all samples, and it is the measure of the total proportion of the correct predictions. But even the accuracy might not be sufficient to describe performance in the case of class imbalance whereby predicting the majority correctly may exaggerate the measure even when the minority classes are misclassified.

In order to overcome this weakness, the sensitivity or recall or true positive rate was taken into account. The concept of sensitivity establishes how the model identifies positive instances correctly thereby indicating how effective the model is in identifying the real cases of disease. In medical applications, where a detection error may be fatal to the patient, high sensitivity is especially important. Added to sensitivity, specificity evaluates the ability of the model to identify the negative cases, which means that it will not give false alarms and it will not perform unnecessary interventions.

The rate of correct positive predictions of the model divided by the number of positive predictions made by a model is referred to as precision. High precision means that there are minimal false positives produced by the model making positive predictions to be accurate and dependable. Although precision and sensitivity alone can give information regarding various facets of the performance, F1-score is a combination of these two measures and hence a balanced score since it considers the false positive and false negative results.

Lastly, the Matthews Correlation Coefficient (MCC) was added as a powerful indicator that can be used to measure the quality of binary and multi-class classification even in the context of imbalanced datasets. MC is a powerful tool to summarize model performance because it takes into account false and true positives and negatives, and it is especially useful to assess the performance of the classifiers in clinical settings. The overall performance of the models analyzed in all these metrics results in a comprehensive picture of the strengths and weaknesses of the framework, so that we not only aim to achieve a high overall accuracy but also to find critical cases and also reduce the misclassifications.

$$Pre = \frac{TP}{TP + FP}$$

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity(Sen) = \frac{TP}{TP + FN}$$

$$Specificity(Spe) = \frac{TN}{TN + FP}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TN + FP)(TN + FN)(TP + FN)}}$$

$$F1 - Score = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

Where TP, FP, TN, and FN represent true positive, false positive, true negative and false negative respectively.
The receiver operating characteristic (ROC) curve is a graph that shows a comparison between the true positive rate (TPR) and the false positive rate (FPR) using a range of different values of threshold. Sensitivity is TPR, and 1-specificity is FPR.

4.3 Comparative Performance of Baseline Models

VGG16 performed best in the baseline, having a 94.44 percentage of accuracy, 94.44 percentage of precision, 94.44 percentage of recall, 94.35 percentage of specificity, 94.42 F1 score, 0.9956 AUC, and 0.9168 MCC as shown in Table 4.3.1. Conversely, ResNet50 experienced a large drop in predictive accuracy, being 60.56% accurate, which is a 33.88% accuracy drop compared to VGG16. On the same note, MCC of ResNet50 was reduced by 49.00 which means that there was impoverished overall correlation between predicted and actual labels. MobileNetV2 only got moderate performance (73.89-percent accuracy) with a 20.00-percent decrease in accuracy compared to VGG16, because of the lightweight design of the network which constrains the feature extraction capabilities of complex CT images. ViT had the worst performance of 51.52% accuracy, which is 42.92% below VGG16, probably because it had fewer inductive biases to spatial hierarchies with small-to-medium medical image data [13].

Classifier	AUC	Acc(%)	Sen(%)	Spe(%)	MCC	Precision	F1
VGG16	0.9956	94.44	94.44	94.35	0.9168	94.44	94.42
ResNet50	0.9282	60.56	60.56	71.68	0.4672	71.18	52.41
MobileNetV2	0.8873	73.89	73.89	86.94	0.6093	73.76	73.72
Vision Transformer	0.5800	51.52	51.52	63.72	0.5020	26.54	35.03

Table 4.3.1: Performance comparison of CNN-based classifiers for lung cancer classification, highlighting comparative accuracy, sensitivity, and interpretability across multiple evaluation metrics.

The performance decrease of ResNet50, MobileNetV2, and ViT point to the fact that traditional CNN models such as VGG16, which learn hierarchical features through deep convolutional layers, are still more effective in medical image classification tasks with small datasets as shown in Figure 4.1.1. Transformer or lightweight based architectures might need to use bigger datasets or further pretraining to obtain similar performance. Figure 4.3.2 depicted VGG16 accuracy bar plot.

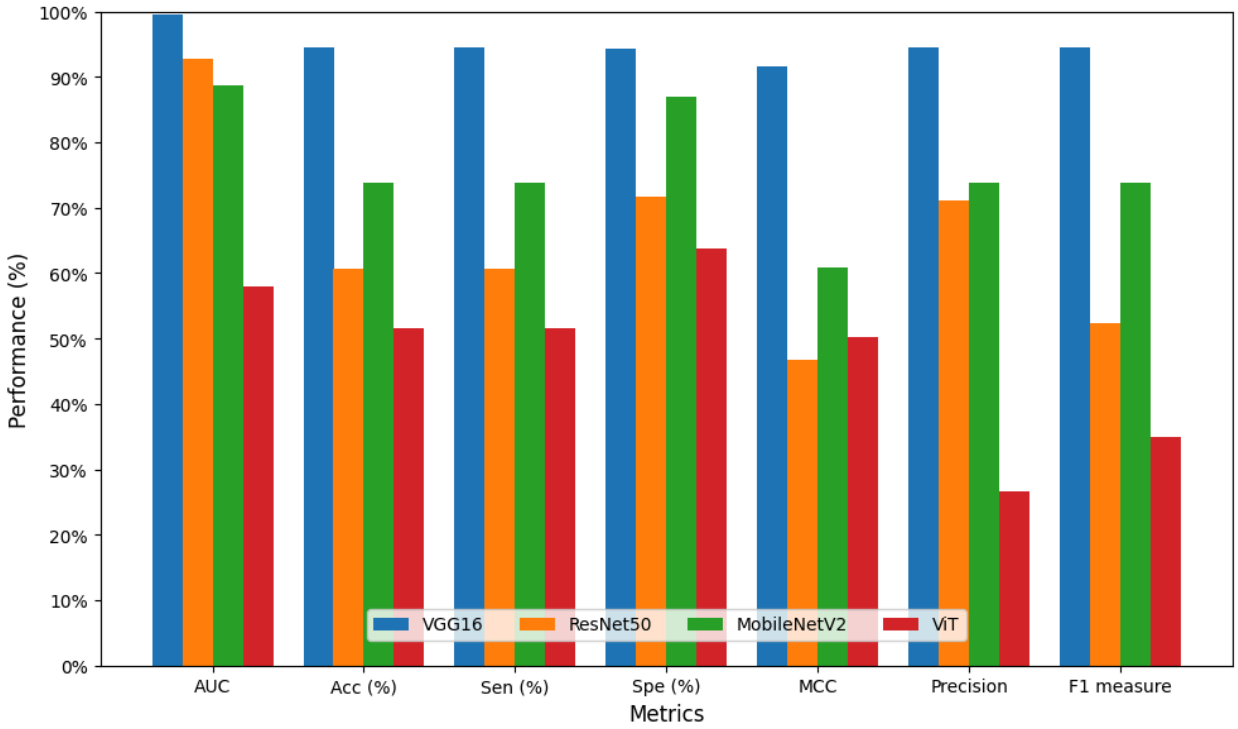

Figure 4.1.1: Comparative performance of baseline CNN and Transformer-based classifiers for lung image classification, illustrating accuracy and robustness differences, with AttVGG16 demonstrating the highest performance among all evaluated models.

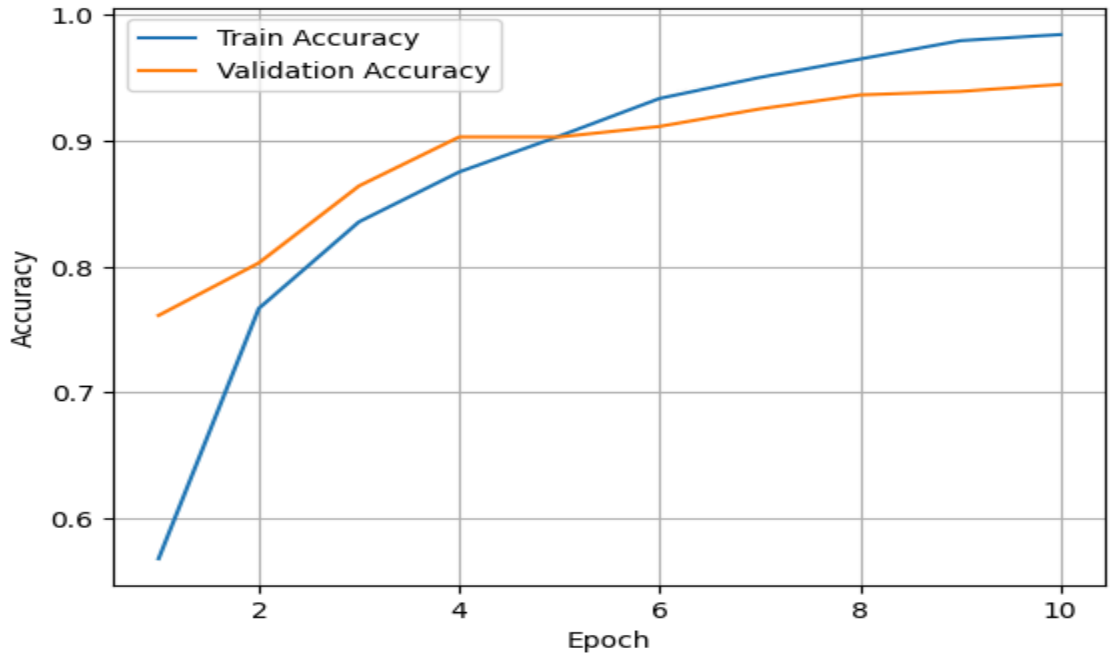

Figure 4.3.2: Illustration of VGG16 accuracy bar plot.

4.4 Improvement Using Attention-based VGG16

By integrating the CBAM attention module, the attention-based VGG16 model improved accuracy to 97.78%, representing a 3.34% increase over the baseline VGG16 (94.44%). Sensitivity improved from 94.44% to 98.18%, a 3.74% gain, highlighting the model's enhanced ability to

Classifier	AUC	Acc(%)	Sen(%)	Spe(%)	MCC	Precision	F1
VGG16	0.9956	94.44	94.44	94.35	0.9168	94.44	94.42
AttVGG16	0.9959	97.78	97.80	98.90	0.9178	97.72	97.76

Table 4.4.1: Performance comparison of baseline VGG16 and the proposed attention-enhanced AttVGG16 model for lung cancer classification, highlighting superior accuracy, sensitivity, and interpretability over conventional architectures across multiple evaluation metrics.

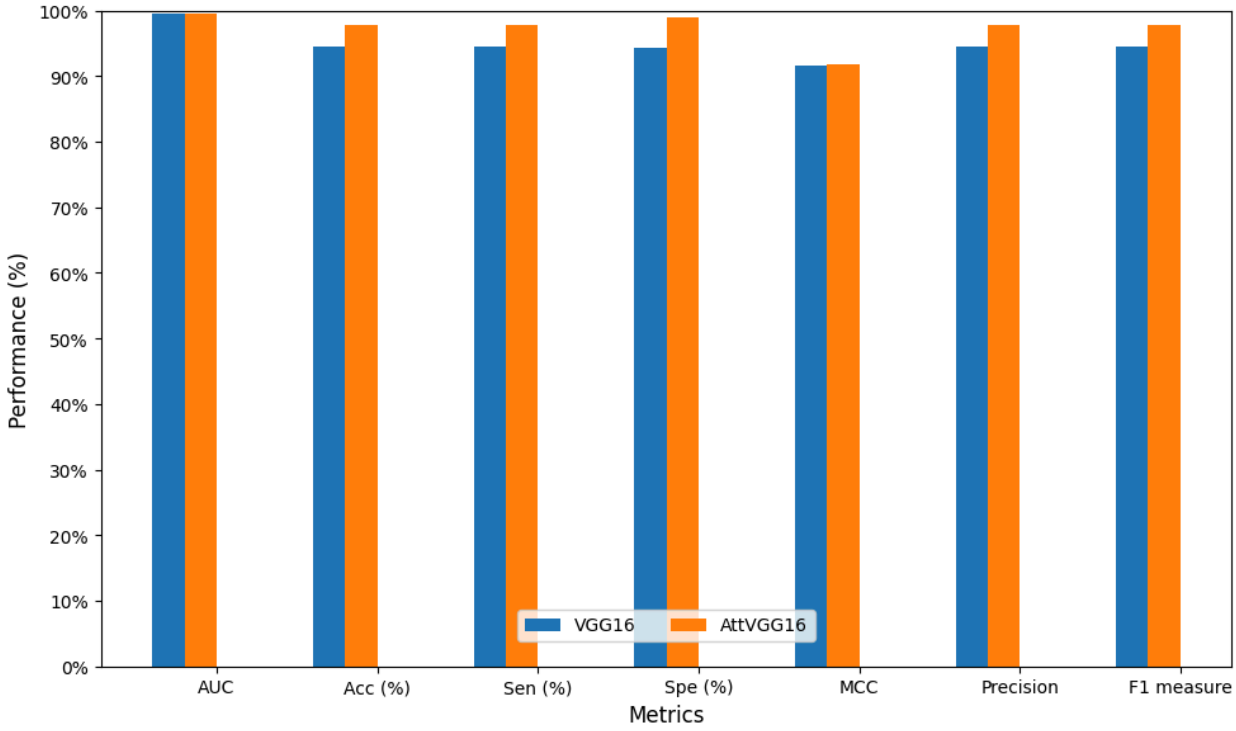

Figure 4.4.2: Comparative illustration showing the superior performance of the proposed attention-enhanced AttVGG16 model over the baseline VGG16 in lung image classification, demonstrating improved accuracy, sensitivity, and overall diagnostic reliability.

correctly detect malignant lung regions. The F1 score increased by 1.58% (from 94.42% to 97.78%), and MCC improved by 0.10 (from 0.9168 to 0.9178), reflecting better overall prediction reliability as depicted in **Figure 4.4.2**.

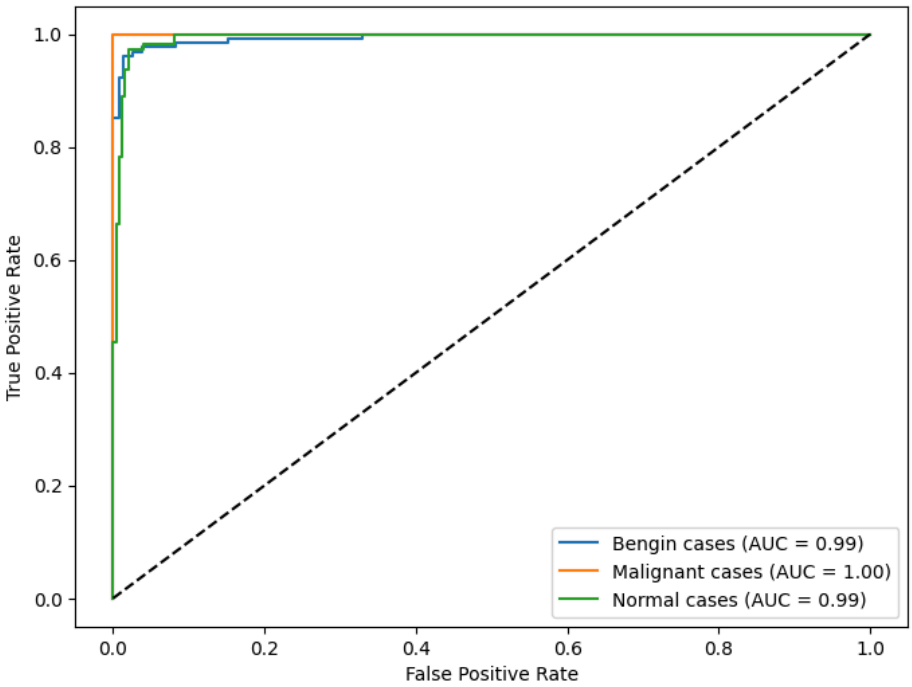

Figure 4.4.3: ROC curve illustrating the superior classification performance of the proposed attention-enhanced AttVGG16 model, showing a high AUC value and strong discriminative capability for accurate lung cancer detection.

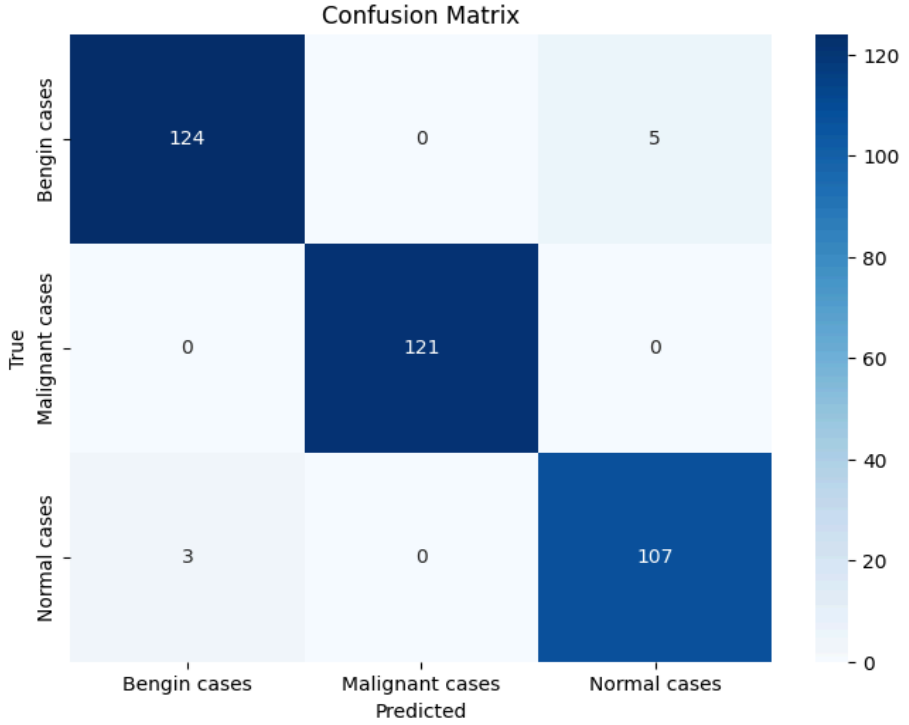

Figure 4.4.4: Confusion matrix demonstrating the classification accuracy of the attention-enhanced AttVGG16 model, indicating reliable differentiation between lung cancer classes with minimal misclassification.

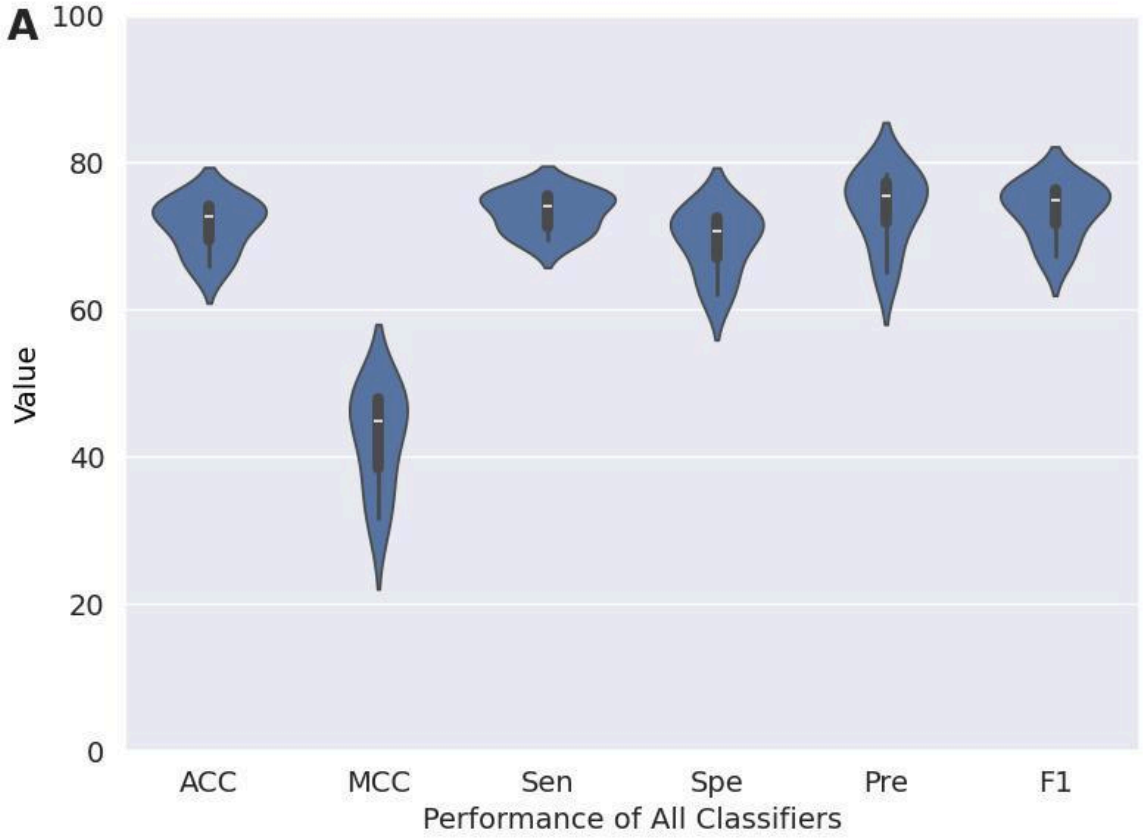

Figure 4.4.5: Violin plot demonstrating the data distribution of performance matrix of all applied classifiers.

Attention mechanism enables the network to selectively attend to the most indicative parts of CT images which are most likely to represent malignancy, and blocks background information. This specific attention is a reason why sensitivity and F1 score are better, which are essential in the clinical practice where false negative should be minimized [12][9]. **Figure 4.4.3** and **Figure 4.4.4** contain ROC curves and confusion matrices of the more successful Attention Based VGG16 (AttVGG16).

These findings verify that, in the detection of lung cancer on CT images, attention-enhanced CNNs are superior to traditional CNNs, lightweight CNNs, or transformers. Selective enhancement of discriminative areas, false negative reduction and global enhancement of model robustness are attained by the addition of attention modules [9][12][16] and are critically important in medical imaging tasks. **Figure 4.4.5** shows a violin plot that shows the performance metrics of each of the five classifiers.

Chapter 5

Conclusion

As this paper will illustrate, attention-enhanced convolutional neural networks (CNNs) have assumed an important leap in automated lung cancer detection of CT scans. We carried out a comparative analysis of four most popular networks: VGG16, ResNet50, MobileNetV2, and Vision Transformer (ViT) through extensive experiments on the balanced IQ-OTH/NCCD data. VGG16 was the most effective baseline model among them, as it has the most effective hierarchical feature extraction capacity to enable the network to distinguish small morphological differences with malignant, benign, and normal tissue patterns.

On this solid foundation we have incorporated the Convolutional Block Attention Module (CBAM) in VGG16 to make the proposed framework AttVGG16. This attention mechanism adapts to highlight the diagnostically important areas whilst inhibiting irrelevant background information and thus the network selectively highlights the areas with the most information on the disease. This focus selectivity led to better classification performance with a higher sensitivity, F1 score, and Matthews correlation coefficient (MCC) and decreased false negatives. The attention-enhanced model was also found to be robust in addressing the problem of inter-class similarity which is a significant problem in pulmonary cancer detection as benign and cancerous nodules can look similar.

Along with the direct performance improvement, the findings highlight the clinical implications of attention processes in deep learning networks. Furthermore, there is scalability and flexibility of the framework with the attention module being applicable to other CNN backbones or even hybrid architectures to other imaging modalities.

In the future, it can be suggested that future studies can be narrowed in this way to multimodal imaging data, using a combination of CT with either PET or MRI data to use the complementary diagnostic data. It could also be considered to use hybrid transformer-CNN architectures to capture both local feature and long-range dependencies in contexts, which might also enhance predictive accuracy and robustness. Also, explainable AI methods and attention could be integrated to gain a better understanding of model decision-making, which will help enhance trust and implementation in clinical care.

To sum up, the suggested AttVGG16 model is a step to a correct comprehensive and immediate diagnosis of lung cancer. Its high performance, interpretability and versatility introduce the transformative capabilities of attention-based CNNs in medical imaging, which opens the way to more useful computer-aided diagnostic tools that eventually can lead to better patient outcomes.

Chapter 6

Future Work

Despite the promising results obtained with the proposed AttVGG16 model, it still has a number of significant directions, in which future research can focus on. To assess the model and re-train it on Bangladesh-specific lung cancer CT datasets would first offer the opportunity to assess the model and improve the localized diagnostic accuracy in relation to demographic, environmental, and clinical particulars of the Bangladeshi patient population. Also, it is possible to further develop the work by using more advanced Explainable AI (XAI) algorithms, including Grad-CAM++, integrated gradients, and layer-wise relevance propagation, in order to make the visual explanations more understandable and more closely related to clinical interpretation; therefore, radiologists will be more willing to trust AI-assisted diagnoses. In addition, the next-generation research needs to be directed towards real-time clinical implementation and optimization, such as model compression, quantization, and edge-device implementation, to provide fast and resource-efficient inference accessible in hospitals and diagnostic centers throughout Bangladesh, with many lacking computational support. It is hoped that the integration of these advancements would make the proposed system much more clinical-ready, reliable, and have a tremendous impact on the real world.

Chapter 7

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022.

[2] M. Gaga *et al.*, "Early lung cancer diagnosis and treatment," *European Respiratory Review*, vol. 31, no. 163, 2022.

[3] D. R. Aberle *et al.*, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[4] S. G. Armato III *et al.*, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.

[5] M. N. Gurcan *et al.*, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," *Medical Physics*, vol. 29, no. 11, pp. 2552–2558, 2002.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *International Conference on Learning Representations (ICLR)*, 2015.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[11] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical Transformer: Gated Axial-Attention for Medical Image Segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 36–46.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[13] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *International Conference on Learning Representations (ICLR)*, 2021.

[14] A. H. Khan *et al.*, "Transformers in Vision: A Survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 10s, pp. 1–41, 2022.

[15] O. S. Al-Karawi *et al.*, "IQ-OTH/NCCD: A New Lung Cancer Computed Tomography Scans Dataset for Image Analysis and Machine Learning Applications," *Data in Brief*, vol. 33, 2020.

[16] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[18] A. A. A. Setio *et al.*, "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.

[19] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019.

[20] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[21] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[23] X. Wang *et al.*, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[24] J. Chen *et al.*, "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[25] J. Donahue *et al.*, "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," *arXiv:1310.1531*, 2014.

[26] S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," *Measurement: Sensors*, vol. 24, 100506, 2022.

[27] U. Haziq *et al.*, "Improving lung cancer detection with enhanced convolutional sequential networks," *Scientific Reports*, vol. 15, no. 1, 32099, 2025. doi:10.1038/s41598-025-06653-y.

[28] S. Aburass *et al.*, “Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases,” *Journal of Imaging Informatics in Medicine*, 2025. doi:10.1007/s10278-025-01481-y.

[29] J. Yang, H. Wan, and Z. Shang, “Enhanced hybrid CNN and transformer network for remote sensing image change detection,” *Scientific Reports*, vol. 15, 10161, 2025. https://doi.org/10.1038/s41598-025-94544-7

[30] A. A. Abe *et al.*, “A robust deep learning algorithm for lung cancer detection from computed tomography images,” *Intelligence-Based Medicine*, vol. 11, 100203, 2025.

212-16-571

ORIGINALITY REPORT

14%	11%	10%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	arxiv.org Internet Source	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
4	ebin.pub Internet Source	<1%
5	public-pages-files-2025.frontiersin.org Internet Source	<1%
6	www.mdpi.com Internet Source	<1%
7	export.arxiv.org Internet Source	<1%
8	link.springer.com Internet Source	<1%
9	www.nature.com Internet Source	<1%
10	hal.science Internet Source	<1%
11	Alireza Rahi. "Ensemble Deep Learning for Histopathological Breast Cancer Detection", Cold Spring Harbor Laboratory, 2025 Publication	<1%
12	"Advanced Computing and Intelligent Technologies", Springer Science and Business	<1%

	Media LLC, 2026 Publication	
13	Submitted to University of Greenwich Student Paper	<1%
14	pmc.ncbi.nlm.nih.gov Internet Source	<1%
15	ceur-ws.org Internet Source	<1%
16	dr.ntu.edu.sg Internet Source	<1%
17	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
18	pure.rug.nl Internet Source	<1%
19	Said Si Kaddour, Larbi Boubchir, Boubaker Daachi. "Multi-Task Deep Learning for Multimodal Biometric Recognition", 2022 Ninth International Conference on Software Defined Systems (SDS), 2022 Publication	<1%
20	iranarze.ir Internet Source	<1%
21	Satyanarayana Murthy Nimmagadda, Gunnam Suryanarayana, Padarti Vijaya Kumar, Goli Sai Vamsi et al. "Early Detection of Lung Cancer Using Deep Learning Techniques: A Comprehensive Review", Archives of Computational Methods in Engineering, 2025 Publication	<1%

22	Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computational Methods in Science and Technology - Volume 2", CRC Press, 2026 Publication	<1%
23	"Quantitative Phase Imaging and Artificial Intelligence: A Review", IEEE Journal of Selected Topics in Quantum Electronics, 2018 Publication	<1%
24	Rejwan Bin Sulaiman, Usman Javed Butt, Yassine Maleh, Mohammad Aljaidi, Md. Simul Hasan Talukder, Musarrat Saberin Nipun. "Securing Health - The Convergence of AI and Cybersecurity in Healthcare", CRC Press, 2025 Publication	<1%
25	Submitted to Addis Ababa University Student Paper	<1%
26	Tobias, Zubin Mario. "Domain-Specific Customization for Improving Speech to Text", Rochester Institute of Technology Publication	<1%
27	web-backend.simula.no Internet Source	<1%
28	Ravikumar, Deepak. "Towards Trustworthy AI: Understanding Memorization, Privacy, and Security in Deep Learning", Purdue University, 2025 Publication	<1%
29	dergipark.org.tr Internet Source	<1%
30	library.iugaza.edu.ps Internet Source	<1%
31	www.hindawi.com Internet Source	<1%

32	era.ed.ac.uk Internet Source	<1%
33	dokumen.pub Internet Source	<1%
34	Anjan Bandyopadhyay, Tanvir Habib Sardar, Saurav Mallik, Ruhul Amin Hazarika, Mahendra Kumar Gourisaria. "AI and Data Engineering for Healthcare - Real-World Applications and Case Studies", CRC Press, 2025 Publication	<1%
35	T. Ananth Kumar, R. Rajmohan, M. Niranjanamurthy, G. Sambasivam. "Deep Learning Models towards Health Informatics Management - Foundations, Challenges and Opportunities", CRC Press, 2026 Publication	<1%
36	ro.ecu.edu.au Internet Source	<1%
37	Submitted to Victorian Institute of Technology Student Paper	<1%
38	bmccancer.biomedcentral.com Internet Source	<1%
39	opus.lib.uts.edu.au Internet Source	<1%
40	Frank Y. Shih. "AI Deep Learning in Image Processing", CRC Press, 2025 Publication	<1%
41	Manoj Kumar, Tanweer Ali, Jaume Anguera, Suman Lata Tripathi. "Emerging Technologies in AI, Computation, Communication, and Cybersecurity - Proceedings of the First International Conference on Artificial Intelligence, Computation, Communication	<1%

	and Network Security (AICCoNS 2025)", CRC Press, 2026 Publication	
42	1login.easychair.org Internet Source	<1%
43	cdn.techscience.cn Internet Source	<1%
44	jsts.org Internet Source	<1%
45	people.dmi.uns.ac.rs Internet Source	<1%
46	www.coursehero.com Internet Source	<1%
47	de Oliveira, Ana Catarina Fontes. "Segmentation of Lungs on CT: Tools to aid Radiotherapy Planning", Universidade de Coimbra (Portugal), 2024 Publication	<1%
48	Alshagathrh, Fahad Muflih. "Advancing Non-Alcoholic Fatty Liver Disease Diagnosis: A Deep Learning Framework for Detection and Staging in Ultrasound Imaging.", Hamad Bin Khalifa University (Qatar) Publication	<1%
49	Brij B. Gupta, Michael Sheng. "Machine Learning for Computer and Cyber Security - Principles, Algorithms, and Practices", CRC Press, 2019 Publication	<1%
50	Polyak, Adam, and Lior Wolf. "Channel-level Acceleration of Deep Face Representations", IEEE Access, 2015. Publication	<1%
51	bmcmedimaging.biomedcentral.com Internet Source	