A Hybrid Feature Learning Model with Residual CNN Blocks and Vision Transformer for Improving Lung Cancer Detection from Medical CT Images

Abdullah Al Jubyer

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

APPROVAL

This thesis titled on "**A Hybrid Feature Learning Model with Residual CNN Blocks and Vision Transformer for Improving Lung Cancer Detection from Medical CT Images**", submitted by **Abdullah Al Jubyer (ID: 221-35-860)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. A. H. M. Saifullah Sadi
Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

---

Internal Examiner 1

Dr. Rubaiyat Islam
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Dr. Md. Abdul Kader
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Nuruzzaman Faruqui
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Md. Mostafiz Khan
Managing Director
Tecognize Solutions Limited

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Abdullah Al Jubyer
Date of Birth	:	30/06/2001
Title	:	A Hybrid Feature Learning Model with Residual CNN Blocks and Vision Transformer for Improving Lung Cancer Detection from Medical CT Images
Academic Session	:	2022-2025

I declare that this thesis is classified as:

☒	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☒	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

221-35-860

Student ID
Date:12/24/2025

(Supervisor's Signature)

Dr. Md Abdul Kader

Name of Supervisor
Date: 12/24/2025

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

_____

(Supervisor's Signature)

Full Name : Dr. Md Abdul Kader

Position : Associate Professor

Date : 12/24/2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

_____

(Student's Signature)

Full Name : Abdullah Al Jubyer

ID Number : 221-35-860

Date : 27 November 2025

A Hybrid Feature Learning Model with Residual CNN Blocks and Vision Transformer for Improving Lung Cancer Detection From Medical CT Images

Abdullah Al Jubyer

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Software Engineering)

DAFFODIL INTERNATIONAL UNIVERSITY

NOVEMBER 2025

ACKNOWLEDGEMENTS

I would like to say a big thank you to Almighty Allah who provided me with strength, patience, and determination to complete this research. His advice has never left me demotivated and has made me remain clear-headed throughout this entire process.

I really like my supervisor because of their constant assistance, constructive feedback, and valuable guidance. Their support and considerate recommendations played a significant role in the way this thesis was to be

I am extremely grateful to the teachers at the Department of Software Engineering at Daffodil International University who have been of great academic support and encouragement to me as an undergraduate student

My friends and fellow researchers also helped me a lot. Their encouragement, collaboration, and constructive discussions have gotten me through several difficult experiences in this project

Lastly, I feel deeply grateful to my parents and family for their endless love, support, and prayers. Their belief in me has always been the main reason I was able to achieve everything I have.

DEDICATION

This thesis is dedicated to my beloved parents, whose unconditional love, numerous assistances, and constant prayers had been the driving force of all my success. Their sacrifices, support, and unchanging vision in me have helped me through every activity and accomplishment in my life.

I also dedicate these paintings to my personal group of family members and well-wishers, whose inspiration and sincere benefits have constantly challenged me to achieve perfection

Lastly, I dedicate this thesis to whoever believes and believes in me in my path, and I will not forget that perseverance and religion can turn dreams into reality.

ABSTRACT

Lung cancer is a major cancer death that should be diagnosed at an early stage. Even though deep learning has resulted in slight advance in the diagnosis of CT, certain models have limited data sets, restricted in their global features, which ought to be too costly to compute CNNs are too local in their focus, and ViTs require large data sets and resources. These gaps are filled in the current paper, which proposes a computationally efficient and lightweight hybrid deep learning network that consists of custom residual CNN blocks and a Vision Transformer block. The CNN component adds to the gradient flow and is learned on more detailed spatial representations of the lung CT slices and the ViT component is learned on the global contextual relationship by multi-head self-attention. In order to train this model, the IQ-OTH/NCCD dataset (scaled to 15,000 images) was employed to ensure that this model is powerful and that it is free of overfitting. The specified architecture reached an accuracy of 0.98, a macro-average precision of 0.98, a recall of 0.98, and an F1-score of 0.98, with class-wise AUC scores of 1.00. The model also possesses very less parameters that stand at 2.46 million and is performing well and this saves on a lot of cost in calculating the model relative to the traditional transformer-based approach. The interpretability of the model is also supported by the explainable AI techniques such as Grad-CAM and LIME because of the presentation of the clinically relevant features as nodule boundaries. On the whole, the findings propose that the proposed hybrid CNN-ViT is a valid, interpretable, and computationally efficient model of automatic multi-class lung cancer discovery..

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

VIT	Vision Transformer
CT	Computed Tomography
CNN	Convolutional Neural Network
CAD	Computer-Aided Detection
LIDC-IDRI	Lung Image Database Consortium and Image Database Resource Initiative
DBN	Deep Belief Network
SDAE	Stacked Denoising Autoencoder
GGO	Ground Glass Opacity
VGG	Visual Geometry Group
MSViT	Multi-Scale Vision Transformer
IQ-OTH	Image Quality Other (class/category)

CHAPTER 1

INTRODUCTION

1.1 Overview

Lung cancer remains the major cause of cancer-related deaths all over the world and one of the most urgent issues of the 21st century in the field of public health. The global cancer statistics show lung cancer is the leading cause of cancer with their reported cancer cases amounting to 2.2million annually, which is about 11.4 percent of all cancer cases and the deaths numbering above 1.8million/year, which are due to lung cancer, making it the deadliest type of cancer in the world [1]. The fatality rate is very high simply because of late detection- patients tend to be highly developed as a result of the disease since lung cancer in its early stages is usually asymptomatic. It has been shown that with the identification of lung cancer at an early stage, chances of survival within a span of five years are more than 55-60 years as opposed to less than 20 years with late detection [2]. The best technique in the early detection of pulmonary nodules is low dose computed tomography (LDCT); although the manual review of CT scan is labor intensive and subject to concealed variance. In a single CT scan, hundreds of slices can be present, and thus even professional radiologists can be challenged in terms of misinterpretation, fatigue, false positives, and the inability to identify thin lesions or ground-glass opacities (GGOs). The challenges have led to the realization that automated Computer-Aided Diagnosis (CAD) systems have become a necessity to enhance the accuracy, speed, and consistency of early lung cancer detection [3].

Convolutional neural networks (CNNs) have become one of the most important tools in the medical image analysis field, and deep learning has become a potent tool to extract meaningful representations out of imaging data. CNNs are predator filters, pooling layers and hierarchical feature learning, which enable them to identify patterns in images automatically and are therefore most distinguished to tumour detection, classification and segmentation [4]. CNNs are effective in lung cancer detection, where

local texture variation, nodule outline, and shape are the important features that can be used to differentiate between benign and malignant lesions. Experiments have proved that CNNs are effective to detect lung nodules with high sensitivity and specificity and are much more effective than the conventional handcrafted feature-based CAD systems [5]. Nevertheless, CNNs continue to have a problem with the modelling of long-range dependencies in 3D CT volumes, which prevents them from capturing global contextual information crucial in the proper characterization of complex nodules.

Vision Transformers (ViTs) have been a widely studied topic over the last few years because of their capability to model the global feature interactions by self-attention mechanisms. In contrast to CNNs, which learn on local receptive fields, ViTs subdivide an image into constant-size patches and learn each patch as a token, which allows the model to learn the connections between spatially separated components [6]. This is because global attention is able to detect diffuse or irregular nodules and pick up subtle patterns that are found in several slices. ViTs have proven themselves successful in a number of medical imaging tasks, such as multi-class disease detection, segmentation tasks, and volumetric analysis [7]. Nevertheless, they tend to be limited by the requirement to use large-scale annotated datasets, and cannot process small datasets because their inductive bias is relatively low.

With the realization of the complementary skill of CNNs and ViTs, current studies have centred on hybrid CNN-ViT models, combining the strong local feature extraction of CNNs with the global contextual learning power of transformers. Such hybrid models have excelled in various medical imaging tasks because they can be used to capture the fine-grain anatomy, as well as the long-range dependencies. In addition, hybrid architectures lower the computational cost of their pure transformer models and increase the stability during training [8]. In lung cancer detection in particular, these hybrid systems can be promising in terms of accuracy, robustness and generalization, which is why it is usable in actual clinical settings. Because of these benefits, this thesis provides a CNN-ViT hybrid model that has the potential to successfully analyze CT scan data to correctly predict lung cancer, which is able to provide increased diagnostic accuracy and higher computation efficiency.

With the growing complexity of deep learning models, the black-box nature of such systems is one of the greatest barriers in clinical adoption. To have confidence in automated predictions, clinicians need the ability to trust them and this relies on transparency and interpretability, particularly in high stakes areas such as lung cancer diagnosis. Explainable Artificial Intelligence (XAI) satisfies this requirement by providing visual and analytical utilities that help to understand how a model comes to its conclusion [9]. Image-based XAI algorithms that include Grad-CAM, LIME, saliency maps, and attention heatmaps mark the parts of the image that contributed to the model prediction most of all. XAI can be used in CT-based lung cancer detection to localize suspicious nodules, visualize malignancy-related patterns, and confirm that the model focuses on the radiologists as expected in the CT-based lung cancer [10]. This does not only increase the model transparency but also the clinical validation, error analysis and trustworthiness of AI-driven decisions.

Since early detection is clinically critical and current solutions have drawbacks, this study suggests a computationally efficient hybrid CNN-ViT model that is directly trained to classify multi-class lung cancer based on CT images. Added integration of Explainable AI (XAI) boosts even more transparency given that it identifies areas of decision-making, which fosters trust between clinicians..

1.2 Problem Statement

The problem of lung cancer remains to be one of the main global health issues because of the lateness of its diagnosis and the insidiousness of radiological symptoms. CT imaging can have high sensitivity in the detection of lung nodules, but the interpretation is very difficult to be performed manually. In a patient, radiologists have to scrutinize hundreds of slices, and very often the nodules are rather small, of very low contrast, and look like harmless anatomic features. It is a lengthy process which is subject to human constraints in the form of fatigue, inter-observer reliability as well as uneven diagnostic opinion. This in turn leads to early lesions being missed or improperly diagnosed which has a negative impact on the treatment.

Although computer-aided detection (CAD) systems based on deep learning have demonstrated their potential in the automated process of lung cancer detection, the current

models are associated with a range of limitations. Most CNN-based networks are good at obtaining the local spatial features but not global contextual information needed to discern intricate malignant patterns. Models based on transformers cover this problem but can be expensive to scale to real-time and resources-constrained clinical settings, and may demand huge datasets. What is more, the majority of the current models are implementable as black-box systems, which provides a limited interpretation ability and decreases the confidence of clinical providers in automated decision-making. The main developmental problem, therefore, is to create a model of lightweight, computationally sparse, and interpretable deep learning that can correctly classify lung CT images as normal, benign, and malignant in order to overcome the shortcomings of the existing models.

1.3 Aim and Objectives

The role of the proposed study is to create and build an efficient and faster hybrid deep learning system that is capable of classifying lung cancer using CT images and at the same time make the models understandable using interpretability methods. The goal is to bring together the benefits of convolutional neural network (CNN), able to extract local features, and the Vision Transformers (ViTs), able to model the global context and consequently build a powerful and clinically relevant diagnostic solution. The key objectives for this thesis are:

i. To investigate explore how local features with detailed convolutional networks used with global context understanding transformers can be better than either one to identify lung cancer in CT images.

ii. In order to develop a systematic pipeline taking into account a hybrid CNN-ViT architecture and a pre-processing environment with noise elimination, lung regions segmentation, intensity corrections, contrast amplification and a variety of data augmentations, etc.

iii. For the construction and training of the proposed system on the data segregation of the patient level, and also to check the work of the proposed system with the best CNN models, transformer models, and hybrid models by various metrics such as accuracy, F1, Area under the curve (AUC), sensitivity, etc..

iv. To find examination of the way in which the model makes decisions using the interpretability tools like Grad-CAM as well as attention based visualizations and make the results understandable and clinically meaningful.

1.4 Research Scope

The research paper proposed is restricted to the idea, evaluation and examination of the hybrid deep learning architecture to categorize the lung cancer into the distinct classes with the assistance of 2D slice images of CT scans. It is analyzed and is the IQ-OTH/NCCD data (110 cases of patients and 1190 marked CT images). Only 2D mode is used to analyze axial sections, multi-views synthesis or 3D volume reconstructions are beyond the scope of this paper.

The research covers several key areas:

i. The neural network is implemented on the CNN-ViT hybrid network that is efficient in computation and optimized to process CT slices (2D/3D) with minimum cost of computation, maximum value and stability.

ii. New models can be reconfigured, including optimized fusion schemes between CNN and transformer modules that ought to enable a more fruitful feature synergy as well as less dependent on large datasets by the transformer.

iii. Enhancement of an enduring preprocessing and training process, separation at patient level, improved the quality of data and improved the generalizability in the case of heterogeneous imaging was observed.

iv. Greater model clarification utilizing integrative XAI techniques including a variety of layers of elucidation by integrating CNN feature activations and transformer attention maps to encourage clinical implementation and trust.

This study lacks other elements of clinical trials, integration of patient demographics, engineering radiomics features, optimization of hardware and deployment as a cloud-based system or their adoption as a hospital-ready system. Instead, the study is limited by the algorithmic design only, while the computational experimentation and interpretation assessment in the controlled research situation.

1.5 Research Contribution

The following way this thesis contributes to the sphere of medical image analysis is:

i. A novel, computationally-efficient CNN-ViT hybrid network, specifically trained to examine 2D/3D CT slices, which reduces the computational requirements but maximizes the accuracy and consistency.

ii. The state-of-the-art model customization, including optimized fusion strategy between CNN and transformer modules, allowing more interaction between features and reducing the dependence of the transformer on massive datasets.

iii. An elaborate preprocessing and training pipeline, which ensures patient-level separation, improved quality of data and improvement of generalization in a variety of imaging scenarios.

iv. Better model interpretability using multi-level explanations, i.e. combined XAI, to enable clinical deployment and trust, CNN feature activations as well as transformer attention maps.

1.6 Research Activities

The research was carried out in a series of systematic mutually supportive steps to achieve the methodological correctness, scientific validity and reproducibility. These phases will involve the preliminary literature search, the design, and the construction of the proposed hybrid deep learning model, and the comparison of the results through the assistance of the current evaluation measures. Everything was critical in the formulation, testing, and explanation of the suggested model of lung cancer classification.

1.6.1 Review of the Literature

The paper began with an elaborate literature review to have a glimpse of the state of the art in computer-aided lung cancer detection. It analyzed a huge number of past works that had applied convolutional neural networks (CNNs), Vision Transformers (ViTs), and hybrid designs and traditional radiomics-driven methods. The review has reviewed the primary shortcomings as the lack of generalization of datasets, excessive reliance on huge volumes of training data, inability to interpret the black-box deep learning models and also the computational ineffectiveness of the transformer networks as applied to the medical imaging data.

Attention studies, and multi-class classification structures were also mentioned to see what architectural elements may have been possibly refined to enhance the performance. On the basis of this analysis, there were some gaps that were found, i.e.,

that lightweight yet high-performing models were required, that the features are represented better depending on the local and global dependencies, and that the explainability methods should be more helpful to facilitate clinical adoption. These gaps also led to the motivation and design of the proposed hybrid CNN-ViT model since they were highlighted in various research papers [4-6].

1.6.2 Design and Implementation

Based on the acquired knowledge after the literature analysis, a hybrid architecture has been created as a customized one to exploit the strengths of CNNs and Transformers that are complementary to each other. The architecture begins with a sequence of residual CNN blocks that are useful in extracting local spatial feature of CT images without destabilizing the gradient through shortcut connections. These blocks are inputted into transformer module in a specific way that prioritizes on long range contextual relationships that are normally significant in the process of differentiating benign nodules and malignant lesions. The multi-head self-attention enables the model to attend to the fine and spatially distant regions of the lung fields.

Implementation process consisted of a number of steps:

i. The preprocessing of the data including the resizing process, normalization of the intensity and denoising to improve the visual quality and consistency of the results in the samples.

ii. Geometric transformations, flips, translations and noise injection to enrich data in order to minimize overfitting and enhance generalizability.

iii. The model is trained, with the assistance of TensorFlow/Kera's, and hyperparameter optimization, such as the learning rate, batch size, dropout rate and early stopping to avoid overfitting.

iv. The process of optimization and tuning, whereby such parameters as the number of filters, attention heads, and feed-forward dimensions were systematically varied to trade-off between performance and computational complexity.

An iterative experimentation to maximise the accuracy of the model, its stability, and convergence was done on the augmented dataset.

1.6.3 Benchmarking and Analysis

The final move was to carry out a vigorous benchmarking of the proposed hybrid model against the current base line and state of the art architectures. In order to quantify the quality of classification between normal, benign, and malignant it was tested on standard measures (accuracy, precision, recall, F1-score and macro-averaged statistical measures) to determine the quality of the performance in the three categories.

In addition to the numerical indicators, a set of analytical tools were applied:

i. Validation line and training to monitor the learning behavior and detect issues such as underfitting or overfitting.

ii. Confusion matrices to follow the performance of the classes, the tendencies of misclassification, and sensitivity to the lesions at an initial stage.

iii. Explainable AI (XAI) visualizations, including Grad-CAM heatmaps and attention maps of the transformer module, were applied to be aware of the decision-making process of the model and identify clinically significant areas in the CT images.

This benchmarking step did not only ensure that the proposed architecture was effective, but also demonstrated that it was understandable and applicable to the actual diagnostic setting in the real-world. The general discussion established the reliability, power, and advantages of the model among other competitive strategies.

1.7 Structure of the Thesis

The remaining part of this thesis will be structured in five chapters that will be a major milestone of the research.

Chapter 1 introduces the study is the introduction of the study which includes the background and motivation, research problem, objectives, and scope and key contributions.

Chapter 2 conducts a literature review of CNN-based models, Vision Transformers and hybrid deep learning models in lung cancer detection. It also touches on the modern limitations in the field, including those on the size of the data (and the external validity of the models), and the necessity of open and clinically reliable systems.

Chapter 3 defines the study methodology. It also includes the description of IQ-OTH/NCCD dataset, preprocessing and augmentation process, the development of the proposed CNN-ViT hybrid model, the training process, and the interpretation tools, such as Grad-CAM and LIME.

Chapter 4 It records the performance metrics, the confusion matrices, ROC-AUC scores and interpretability reports. The chapter also compares the proposed model with the existing methods and identifies the strengths and weaknesses which are evident.

Chapter 5 is the last section of the work that summarizes the main findings, remarks on their scientific and clinical usage, and offers the future research perspectives, including the usage of 3D CT data, greater multi-centers data use, and better deployment schemes.

CHAPTER 2

LITERATURE REVIEW

The below is an overview of the latest advancements in the deep learning techniques of lung cancer detection and classification. The papers have highlighted the implementation of the various models and methods such as CNN, DCNN, support vector machines (SVM), and hybrid models that demonstrated the flexibility and effectiveness of the models and methods to improve the diagnostic accuracy. The reviewed research talks about different methods, which include the first detection, image segmentation, and model optimization, and it is important to mention that the first possibility of deep learning is to enhance the possibility of diagnosing and treating lung cancer better.

2.1 CNN-Based Approaches

Convolutional neural networks (CNNs) have become quite a popular selection of the working architecture of the lung cancer detection and analysis through CT scans. Initial deep learning systems relied primarily on 2D CNNs and were trained on patches that were excised out of CT slices, and could only learn simple texture and intensity but not contextual information at a volume. Among the first systematic comparisons of CNNs with DBN and SDAE structures on the LIDC-IDRI data set is that of Sun et al. [6], who discovered that with patch-based 2D learning, the accuracy was approximately equal to 79-81% a significant improvement over classical CAD pipelines, but again limited by the constraints of patch-based 2D learning. Classical segmentation-plus-classification pipelines which combined image processing tasks (e.g. watershed segmentation) with shallow machine learning methods were also reviewed by Makaju et al. [12]. Though these systems have been demonstrated to be accurate over the 84-94% range in past studies cited in its survey, the systems are highly dependent on manual feature engineering and thus cannot easily adapt to new scanning conditions and nodule morphologies.

The advent of large publicly available datasets such as LIDC-IDRI and LUNA16 also saw increased interest in 3D CNNs that are capable of learning volumetric structures. The architecture presented by Crasta et al. [1] is 3D-VNet and 3D-ResNet, which has two steps of nodule segmentation and malignancy classification with near-perfect DSC (99.34), and classification accuracy (99.2). These are amazing results that are subject to overfitting because external multi-center testing was not done. A more clinically inspired source was [21], as its authors developed a multi-resolution 3D CNN that both identified nodules and their malignancy. Their model was sensitive to experts (84.4%) and even did better on a set of 25 radiologists, which clearly indicates that 3D CNNs can be a good choice to be applied to practice. However, the result was poorer in case of small nodules or ground-glass nodules (GGO) that indicates the persistence of the inability of modeling fine-scale malignant patterns.

To complement these entirely volumetric methods, several studies examined the transfer learning and ensemble CNNs to categorize subtypes of lung cancer. VGG-16, ResNet-50, MobileNet, and Xception were tested on the LIDC-IDRI by Chiet et al. [9], and VGG-16 had the highest accuracy of approximately 86.7% although some of the ML classifiers were reported to have a score of nearly 100 that is suspicious of leakage of data due to inappropriate splitting of slices. Similarly, Bhuiyan et al. [10] and the authors [20] conducted their studies with CNNs and classical ML but with high accuracy (up to 100 percent in [20]) but very small datasets and mixed-modalities, which lowered the validity of the articles. Other systems that were enhanced with segmentation-oriented functionality were the nested 3D fully connected CNN by the authors in [22], with a DSC of 0.845, which outperformed U-Net variants. Nevertheless, the single center data and the limited study scope on adenocarcinoma cause a limitation to generalizability. Overall, the CNN-based approaches formed the basis of automated lung cancer processing, even though they continue to suffer the issue of scales, generalization, overfitting, and limited modeling of the long-range contextual information.

2.2 Hybrid, Transformer-Based, and Vision Transformer (ViT) Approaches

Since CNNs were found to be poor at long-range dependencies and global contextual patterns, some hybrid architectures, a combination of several deep learning components,

began to be developed. Alsheikhy et al. [2] proposed one of the first hybrid pipelines, which integrates VGG19 with an LSTM block and PCA to downsample the number of dimensions to a dimensionality and achieved an 99.42 percent accuracy on a mixed dataset of 850 images. It is problematic because their findings suggest the benefits of the application of spatial and sequential feature extractors, but the small size of the data set and the lack of patient-level splits are issues related to overfitting and overly high performance rates. The patch extraction, clustering-based segmentation, and CNN-based classification is another hybrid system developed by Venkatesh et al. [8], which claimed to have an enhanced accuracy and statistical image measures. Patch-based approach however is computationally expensive and requires manual or semi-automatic selection of ROI and hence not as scalable to clinical use. Subsequent hybrid networks, including ensemble model with ResNet-50/101 + EfficientNet-B3 proposed by the authors in [22]) have demonstrated high precision and multi-class classification performance on LIDC-IDRI, but again on small sample sizes and not externally validated.

The rapid evolution of transformer architectures revolutionized the field and presented self-attention mechanisms to capture the global relations in the volumetric CT data. Said et al. [15] conducted high segmentation accuracy (97.83) and high classification performance (98.77) with UNETR, transformer-based encoder with CNN decoder on the Medical Decathlon dataset. Although these results indicate the usefulness of transformers in the medical image segmentation procedure, the uniformity of the Decathlon data raises the question of whether the process will be useful when using heterogeneous CT data in the real world. Literature Transformer-based models like MSViT and pyramid attention networks were observed to be sensitive to up to 97.8% on LUNA16 [3], [7], [16], [17], and [14]. However, these reviews also pointed at the lack of head-to-head comparison of CNNs and transformers under the conditions of standardized evaluation, and it is difficult to possess common benchmarks.

Despite the transformative advances, there are still several limitations of current transformer-based, as well as hybrid methods. Few works could develop coherent CNN-transformer models, which would compromise the CNNs local receptive field of finer features with the receptive field of the transformers. Most of the existing models, according to Tan et al., are not only tested on small subsets of LIDC-IDRI or LUNA16

but almost never on a prospective or multi-centre scope [11]. Even mixed methods such as the ones in [22], do not often apply full 3D processing or cross-dataset testing thereby generating generalization errors. In addition, explainability tools, robustness analysis with variations of noise, and misclassification analysis are not well intertwined in the literature. These gaps suggest that the high-quality hybrid CNN-ViT systems, which are efficient and well-tested, are urgently needed in order to be able to extrapolate and generalize to the heterogeneous CT datasets without causing the high interpretability and clinical reliability to be compromised.

2.3 Clinical Deployment, Real-World Screening Systems, and Multi-Modal AI

The other significant area of research is directed at the implementation of AI systems in the clinical environment where they are required to be robust, interpretable, and large-scale validated. The article by Wu et al. [4] reviewed the AI systems that have been applied in screening of lung cancer with large LDCT screenings such as NLST and NELSON, and revealed that the systems have improved sensitivity, reduced false-positive rates and detected nodules earlier. However, they also pointed out that many algorithms do not work due to inconsistency of parameters of CT acquisitions, scanner heterogeneity, and lack of standardized preprocessing pipelines. Other reviews such as those of Thanoon et al. [3] and Usharani et al. [14] also indicated that even though the models have high internal performance, external validation is not available in most of them and therefore they cannot be implemented at any given time in the clinical practice. General summaries like Tian et al. [19] have emphasized the importance of the integration of AI systems into the real-time clinical workflow, interpretability problems, and regulatory suitability issues, but it is not used in non-research environments.

One of the most significant contributions to the clinical field is the LCP-CNN research by Heuvelmans et al. [13], as it is well-validated in a multicenter study of screening cohorts in Europe. Having an AUC of 0.945 and a safe rule-out value of 22.1 percent of benign nodules at 99 percent sensitivity, the present study points to the opportunities of deep learning systems in practice. However even these high performing systems were found to have a limit such as the incidences of missed carcinoid and potential calibration issues in the various institutions with varying disease prevalence.

Zhang et al. [18] enhanced clinical relevance since they focused on prognostic modelling using histopathology images. Their self-supervised phenotype clustering model (PathoSig) could stratify survival data in several cohorts, suggesting that deep learning could be applied more widely than as a detection tool. However, they depend on the quality of pathology slides, as well as, they are not dedicated to the problem of early detection using CT.

Despite this being encouraging, there is still a huge clinical deployment. The surveys by Javed et al. [7], Tan et al. [11] and Gayap and Akhloufi [17] have mentioned the fact that the major obstacles were lack of explain ability, absence of large-scale prospective trials, inconsistent reporting of evaluation metrics and poor management of the heterogeneity of multi-centers. Even high-performance hybrid or transformer-based systems would often fail to report calibration plots, decision-curve analysis or nodule-level error analysis-key components of clinical integration. Taken together, the clinical research demonstrates that despite a considerable potential of AI to supplement the radiological processes, the science lacks solid, generalizable, and interpretable models that can harmoniously unite local morphological knowledge and global volumetric logic. This also justifies the motivation to develop a CNN-ViT hybrid architecture that is able to achieve a decent predictive performance and a faithful clinical application.

2.4 Research Gaps and Critical Analysis

Through critical review of the available literature in the detection of lung cancer, it is evident that there are still several gaps to be sealed in order to come up with superior and clinically viable deep learning models. Firstly, the majority of studies still use small datasets and unsuitable splits on patient level that lead to data leakage and inflated performance measures. The works such as Alsheikhy et al. [2], Sun et al. [6], Venkatesh et al. [8], Chiet et al. [9], and the ensemble-based works in [20] and [22] commonly use small or heterogeneous datasets, where slices of the same patient are commonly mixed in training and testing datasets. This undermines the generalization and removes the realistic estimation of the model reliability. Second, there are also several high-performing models such as the models proposed by Crasta et al. [1], Said et al. [15], and the accuracy rates of their own models are extremely high (typically over 99%), and they

are not rigorously tested. Such near perfect results are a good sign of overfitting, especially when evaluated on individual datasets, or on small internal splits, and the issue of whether they can be generalized to the actual clinical setting should be doubted.

Third, despite the fact that CNN-based methods have been improved significantly, the majority of them do not have volumetric knowledge. The 2D CNNs that have been used conventionally in the literature of [6], [9], [12] and [20] lack the ability to identify 3D structural relationships within CT volumes which is fundamental in identifying small, irregular or diffuse nodules such as GGOs. Transformer-based models, including UNETR and others described in [11], [14], [15], and [17], in their turn, do not only have the exceptional global feature modeling capabilities but are highly data-sensitive and unstable when trained on small medical datasets. Untrained models based on pure ViT are trained on large scale and fail to learn small CT-based datasets, leading to poor convergence and worsening performance. This explains why there is an utter need to possess hybrid architectures capable of trading the locality of CNNs in terms of feature extraction with the global reasoning of transformers.

Finally, an unacceptable weakness in the literature is the inadmissibility of explanation, external validation and clinical preparedness. The fact that most of the AI systems remain confined to the laboratory is indicated by clinical appraisals and deployment-focused investigations such as Wu et al. [4], Tan et al. [11], Heuvelmans et al. [13], Usharani et al. [14], Tian et al. [19] and expert-level execution studies such as [21]. Few of them apply explainability tools, misclassification analysis, calibration curves, and decision-curve assessment, which are critical criteria of clinical adoption. The majority of the models do not extrapolate the results between multi-center datasets of other scanner types, imaging protocols, and prevalence distributions, limiting its use to real screening program.

Table 2.1: Key Research Gaps Identified in Existing Literature

Reference	Summery of the work	Observation
[2], [6], [8], [9], [20], [22]	The vast majority of the studies are founded on small datasets or slice-level splitting that leads to data leakage and overstated accuracy scores. Models are also able to memorise slices rather than develop actual nodule properties.	Minimal Data Set and Unsuitable Splitting on Patient Level.
[1], [2], [15], [20], [22]	Some of the studies have almost perfect accuracy (>99%), and lack external validation, which suggests that they are overfitting to small samples or are not heterogenous cohorts.	Overfitting & Unrealistically High Performance Claims.
[6], [9], [12], [20]	2D CNNs lack the ability to learn the morphology of 3D nodules and long distance spatial interactions between CT slices and thus are not as good at detecting small nodules, irregular nodules, or diffuse nodules.	Weak Volumetric Cues of CNN-based models.

[11], [14], [15], [17]	Not just effective, pure transformer models (including UNETR) are sensitive to data, and thus may result in unstable training or overfitting when it comes to the exorbitant demand of large training samples.	Transformers are Memory Intensive and that is the reason why they are not stable on small CT Data.
[4], [11], [13], [14], [19], [21]	The overwhelming majority of models are inexplainable (XAI), is not a misclassification analyzer, and fails to work in a multi-center or prospective deployment, this is why they are not applicable in clinical practice.	Lack of Explainability, External validation and clinical preparedness.

2.5 Chapter Summary

This chapter reviewed the literature on the lung cancer detection methods, and in particular deep learning methodologies. The literature on convolutional neural networks (CNNs) revealed their benefits in the local spatial information and also revealed the limitations with regard to long-range relationship between CT scans. Multi-architecture hybrid methods were proved to be more performance-efficient, yet not necessarily interpretable or computationally efficient. It was mentioned in the literature on Vision Transformers (ViT) that they can be more applicable since they are very efficient in encompassing relationships in the world and typically, they require large datasets. The clinical implementation studies indicated the significance of reliable, transparent, and comprehensible CAD systems to support radiologists. The gaps in the studies were found in all 25 of the reviewed papers, as they comprised small datasets, lack of explainability, unstable generalization, and the lack of hybrid CNN-ViT solutions. The existing gaps led to the development of the given hybrid architecture, that will cover the existing gaps, including CNN-based local feature extraction, ViT-based global attention, and XAI-based interpretability.

CHAPTER 3

METHODOLOGY

The article is well-structured and it consists of dataset preparation, image preprocessing, image augmentation, hybrid model development, model training, model evaluation, and model explainability. The original investigation and analysis is performed by use of IQ-OTH/NCCD CT data of the lung cancer and after that, some preliminary processing steps such as image resizing, image normalization, contrast amplification and noise suppression are undertaken to give the same quality of image. He has used the technique of balancing the data and number of classes through enormous geometric and photometric boosting doubling the number of samples to 15,000 which are balanced.

A CNN-Vision image transformer model is then developed that is lightweight. CNN backbone This is a convolutional block of data with a residual convolutional block which finds the local spatial and textural data and the Vision Transformer block finds both a long-range and global correlation between lung regions. This process is repeated to get the features which are then sent to a small classification head to be eventually predicted. Model training The model may be trained end to end using categorical cross-entropy loss and Adam and trained at the best learning rate using such techniques as learning rate scheduling and early stopping.

The performance of the model, F1-score and the confusion matrices are measured using the accuracy, the precision and the recall of the model. Finally, the Explainable AI techniques, including Grad-CAM and LIME are applied to interpret the model predictions and ensure that the network only focuses on clinically significant features. These two pipelines are strong, robust and understandable in the sense that the combination of the two pipelines produces a strong structure of lung cancer on the CT images.

3.1 Dataset Description

The IQ-OTH /NCCD Lung Cancer Dataset that is taken in this paper is also free in Kaggle. The gathered material was carried out within three months of the fall of 2019, the Teaching Hospital of Oncology in Iraq and the National Cancer Diseases Center. It consists of 1190 CT scan sections that had been done on 110 different cases of patients that were classified into three clinically relevant subsets which include normal, benign and malignant. These 110 cases comprise 40 cases of malignant, 15 cases of benign and 55 cases of normal which is a real time scenario of the presentation of lung cancer in a clinical practice.

Each CT scan had original scans which were in the DICOM format and had been scanned using a Siemens SOMATOM CT scanner. The scanning protocol was 120 k V, slice thickness 1 mm and window level parameter that was suitable to the visualization of the lungs where the window widths are 350 1200 Hounsfield Units and window center is 50 600 Hounsfield Units. The scanning was performed fully and with inspiration breath hold and hence it was more appropriate in giving more lung structures and less motion artifacts. The information consists of 80-200 slices of both scans that are the volume of anatomical views and thickness of the part of the thoracic cavity. Ethical consent was obtained by institutional review boards of the respective institutions, and the scans were all made anonymous.

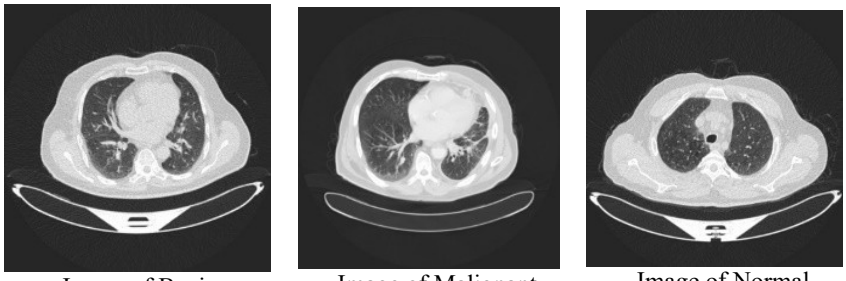

Image of Benign Image of Malignant Image of Normal

Figure 3.1: Samples of images used in the study

3.2 Data Preprocessing

To generate congruent and effective training of the CNN-ViT hybrid model, a collection of preprocesses have been applied to the CT images in order to normalize the images in addition to optimizing the images. Since the raw IQ-OTH/NCCD data will be stuffed with the CT slices in various formats, resolutions and intensity levels, preprocessing will hold great significance to decrease the level of noise, fix the locations of the anomaly and optimise the exposure of the weak lung structures such as nodules, lesions and opacities. The original DICOM slices were first converted into the standard format of image of PNG or JPEG all the images were resized to a standard spatial resolution of 248 x 248 pixels and three channels to fit into the input requirement of the proposed hybrid network. Such resizing will ensure that the size of the inputs of the dataset is not lost, the task to be computed is simplified, and the processing of the batches is efficient when training the models.

The spatial normalization was the next followed by the intensity normalization to offset the natural variation of the CT pixel values due to the difference in the scanners, different variables of acquisition and patient disparities. The pixel values were rescaled to [0,1] with the help of the min-max normalization and helps the training to stabilize and evenly distribute the values across the slices. Further, lung-window normalization that was based on similar ranges of Hounsfield Unit (HU) was performed to emphasize the relevant lung-field anatomical features so that the soft-tissue features could be boosted to the learning model. The noise reduction methods were then employed in a way that would not interfere with the quality of the images but would improve the diagnostic content. The filter was used to remove noise artifact such as using filters such as the Gaussian smoothing or medium filter since the CT images are usually full of noisy images, the low dose scans have a lot of noise. The benefits of the filters are that they remove the random variation and have the ability to learn meaningful structures of the structure in question, in such a way that what the model learns are meaningful features, rather than the noise patterns.

Each CT slice was also subjected to Contrast Limited Adaptive Histogram Equalization (CLAHE) so as to enhance visual quality. CLAHE uses redistribution of the

intensity value within the small surrounding windows to increase contrast of images in the areas with slight noise over-enhancement. The given intervention measure improves the disclosure of small abnormalities of the lungs comprising ground-glass opacities, small nodules, and interrupted tissue demarcation. As a result, more local features and transformer block can be enabled by CNN layers to obtain more global contextual embedding. On the whole, all images pipetting into the hybrid CNN-ViT model are of standardized, noise-reduced, contrast-enhanced and deep-feature-optimal images, which is a good basis of adequate lung cancer recognition.

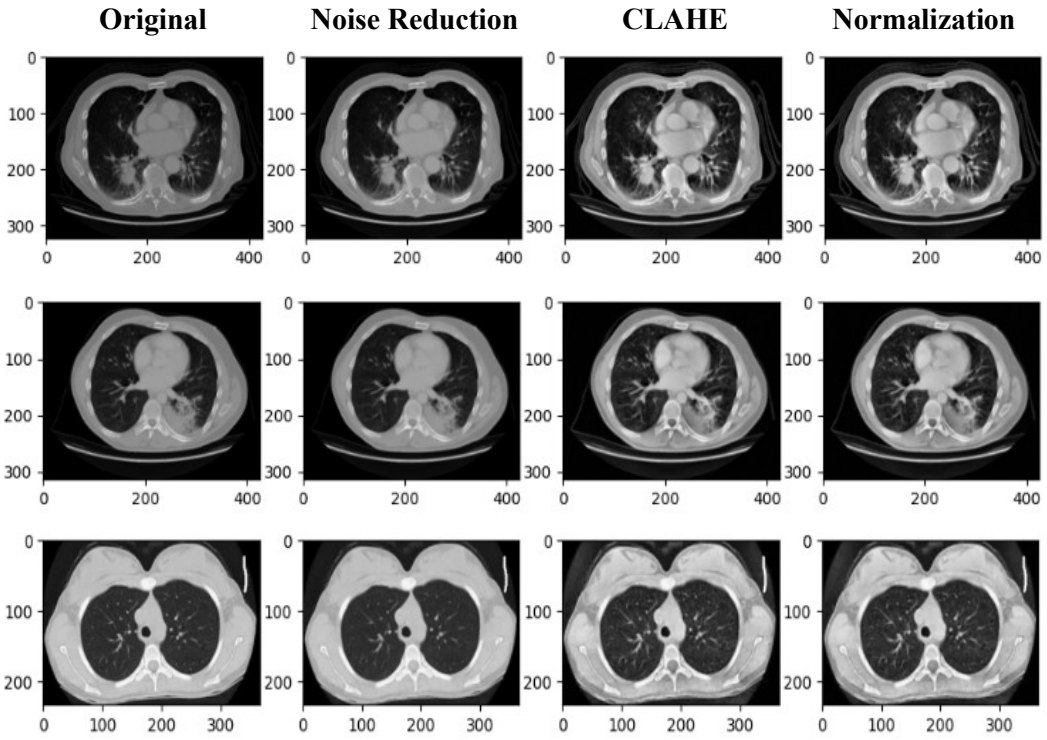

Figure 3.2: Samples of pre-processed images

3.3 Data Augmentation

The original IQ-OTH/NCCD data set is extremely small, and the three groups of the data (normal, benign, and malignant) are disproportional hence data augmentation was

brought to radically increase the size of the training samples and the generalizability of the proposed CNN-ViT hybrid model. Specifically, augmentation will be vital in the field of medical imaging in which large amounts of data are time-consuming, costly and need clinical experience to be annotated. The augmentation process aids the model to be less frail in the imaging mode of the real world, scanner fluctuation, dissimilarity in the location of the patient and minimal anatomical variations.

Such augmentation method involved exploiting the geometry, photometric and intensity based transformations. Geometric augmentation had image rotations, horizontal and vertical flipping, shifting, zooming, optimized range and range cropping and cropping. These are patient positioning and angle simulations which are applied in obtaining CT scans. Photometric controls, such as the brightness and contrast, gamma correction and histogram equalization (with CLAHE) were modified based on the simulated change in the exposure levels, viewability of the tissue density, and change in the scanner calibration. Besides, local clarity was controlled using a selective application of the Gaussian blur or sharpening filters that have no diagnostic qualities. All the methods of augmentation were chosen due to the fact that they do not eliminate the clinical meaning of the lung structures and offer sufficient amount of variability of data.

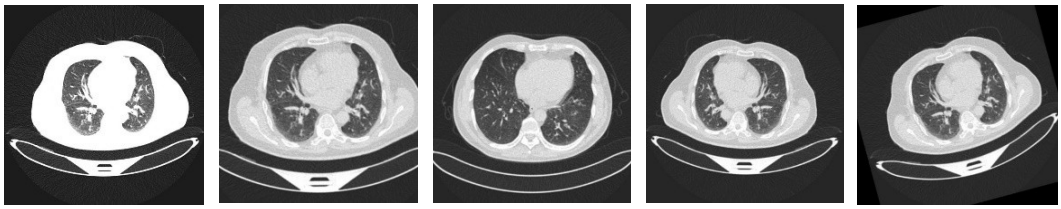

Figure 3.3: Samples after data augmentations.

The augmentation techniques led to the expansion of the dataset. The training set that had previously had a few slices in each of the classes had been increased to over 10000 augmented samples. The validation set and the test set were also grew proportionally but at a more conservative rate so as to preserve integrity of evaluation. The augmented data set was also able to stabilize the procedure of optimization of the model training by providing the appropriate ratio of three classes. Table 1 is the summary of the original and augmented samples.

Table:3.1 Class-wise Image Distribution Before and After Augmentation

Class	Before Augmentation	After Augmentation
Benign	120	5,000
Malignant	561	5,000
Normal	416	5,000
Total	1,097	15,000

Table 3.2 Class Distribution in Training, Testing, and validation.

Class	Training Set	Validation Set	Test Set
Benign	3499	750	751
Malignant	3499	750	751
Normal	3499	750	751

3.4 Model Architecture

It is mentioned that one can implement CNN-Vision Transformer model which is based on a combination of convolutional feature and global self-attention which labels the lung CT images as normal and malignant or even benign. Immediately after, the second sub-section of this paper contains some mathematical equations as well as the description of each of the architectural elements.

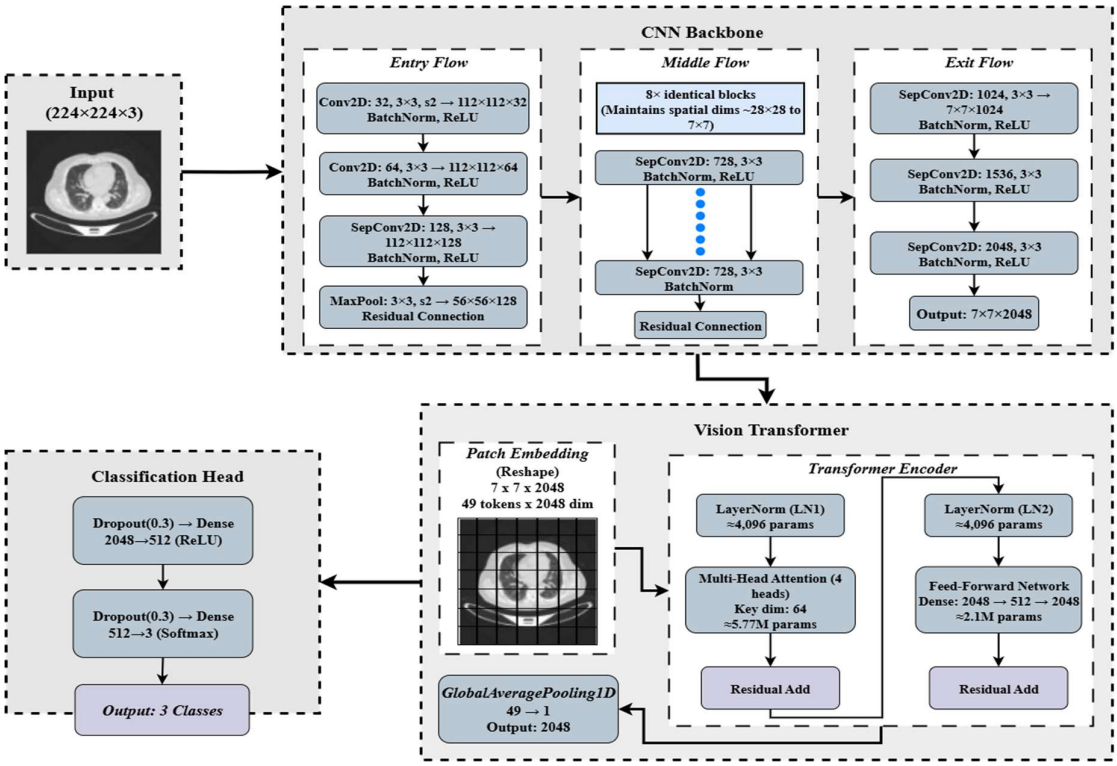

Figure 3.4: Model Architecture of the proposed CNN-ViT model

3.4.1 Input Layer and Convolutional Stem

It starts with a CT image $X \in \mathbb{R}^{248\times248\times3}$ which is first convoluted using 32 filters and a 3x3 convolution and then a batch normalization and ReLU are applied to it:

$$F_0 = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3,32}(X)\right)\right) \tag{3.1}$$

Conv₂D: This will calculate 32 features maps and this is achieved by sliding 3x3 kernel kernels across the image.

The calculation of all the output feature maps is:

$$F_0^{(k)} = X * W^{(k)} + b^{(k)} \tag{3.2}$$

Batch Normalization (BN): Normalizes intermediate activations to stabilize training.

ReLU: Introduces the non-linearity in order to allow learning more complex patterns. The reason is that the spatial dimensions are down sampled with max pooling of 2x 2:

$$F_1 = \text{MaxPool}_{2\times2}(F_0) \quad (3.3)$$

The strongest activation in every 2x2 window is chosen using the Max pooling and it has been shown to be effective in eliminating noise and retaining strong features.

3.4.2 Residual Blocks

The remaining blocks include two convictions and a break in between. The main convolutional path is:

$$H_1 = \text{BN}(\text{ReLU}(\text{Conv}_{3\times3}(F_i))) \quad (3.4)$$

$$H_2 = \text{BN}(\text{ReLU}(\text{Conv}_{3\times3}(H_1))) \quad (3.5)$$

- The outcome of the two 3x3 convolutions is of higher order spatial values.
- Normalization Batch normalization is the method of providing certain stability to the gradients.
- ReLU activation favors thick and squashed representations.
- The convolution is made 1x1 in such a way that addition can be done as the depth of the feature (e.g. 64-128 filters) is varied.
- In case of a similarity in dimensions, then identity shortcut is taken and data stored.

The rest of the output shall be calculated as:

$$R_i = H_2 + S \quad (3.6)$$

The element wise addition will also admit the residual learning, and can be applied to use the gradients flowing through the shortcut and not the original element and eliminate the vanishing gradients. The down sampling is performed at the conclusion of each block:

$$F_{i+1} = \text{MaxPool}_{2\times2}(R_i) \quad (3.7)$$

3.4.3 Additional Convolutional Block

Before the transformation of transformer, the result of the last convolutional layer (256 filters) is:

$$F_4 = \text{MaxPool}_{2\times2}(\text{BN}(\text{ReLU}(\text{Conv}_{3\times3,256}(F_3)))) \quad (3.8)$$

- Combines the features of high-level objects (e.g. nodule boundaries, textures).
- Efforts are put to down sample the feature map into 7x7x256 that is small representation that can be processed by attention mechanisms.

3.4.4 Transformer Block (Self-Attention)

Flattening to Token Sequence: The 49 token sequence of F4is was flattened:

$$T = \text{Reshape}(F_4) \in \mathbb{R}^{(HW)\times C} \quad (3.9)$$

- The grid on the spatial plane (7 x 7) is translated into an order of 49 tokens..
- Each of the tokens is 256-dimensional.

Multi-Head Self-Attention (MHSA): For each head:

$$Q = TW_Q, K = TW_K, V = TW_V \quad (3.10)$$

- **Q (Query):** Provides the query search query that the token is searching.
- **K (Key):** is the value of sum of tokens.
- **V (Value):** This is where the real data is.

  Calculations of self-attention will be made:.

  Self-attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \quad (3.11)$$

- The dot product QK^Tmeasures similarity between tokens.
- Division by $\sqrt{d_k}$stabilizes gradients.
- The similarities are converted into weights under the help of the softmax.

- Each coordination of the tokens is multiplied with vaggregates information.
- It could allow the model to include the long-range interaction, e.g. the correlation of the further parts of the lungs.

Feed-Forward Network (FFN)

$$\text{FFN}(x) = W_2(\text{ReLU}(W_1 x)) \quad (3.12)$$

- The dimensional expansion was acquired to W1 where expressive representation was realized.
- ReLU provides non-linearity.
- The W2projects are returned to the original dimension.

Residual and Normalization Layers

$$T_1 = \text{LayerNorm}(T + \text{MHSA}(T)) \quad (3.13)$$

$$T_2 = \text{LayerNorm}(T_1 + \text{FFN}(T_1)) \quad (3.14)$$

- A residual connection stabilizes the sub-layers of the transformer to make the gradient flow steady.
- Normalization of the layer causes the algorithm to be numerically stable and reach a faster convergent rate. Finall:

$$F_T = \text{Reshape}(T_2) \quad (3.15)$$

This brings back the token sequence to the form of spatial features maps.

3.4.5 Classification Head

Global Average Pooling

$$G = \text{GAP}(F_T) \quad (3.16)$$

This means that it averages all the channels of features and produces a smaller dimension vector.

Dense Layers for Classification

$$Z = \text{ReLU}(W_z G + b_z) \quad (3.18)$$

$$\hat{y} = \text{Softmax}(W_o Z + b_o) \quad (3.19)$$

- Dense(512) acquires high level abstractions.
- No chance disabling of the neurons will overfit dropout(0.5).
- The Softmax-based probability distribution of the three classes was the following:

$$\hat{y} = [P(Benign), P(Malignant), P(Normal)]$$

3.4.6 Summary of the Architecture

Table: 3.3 Layer-wise Output Shapes and Parameter Summary of the Proposed CNN–ViT Hybrid Model

Layer / Block	Output Shape	Params
Input Layer	(248, 248, 3)	0
Initial Conv + BN + MaxPool	(124, 124, 32)	1,024
Residual Block 1 (64 filters)	(62, 62, 64)	57,888
Residual Block 2 (128 filters)	(31, 31, 128)	230,912
Residual Block 3 (256 filters)	(15, 15, 256)	885,760
Extra Conv Block (256 filters)	(7, 7, 256)	591,104
Transformer Block (4 heads)	(7, 7, 256)	527,104
Global Average Pooling	(256,)	0
Dense Layer (512 units)	(512,)	131,584
Dropout	(512,)	0
Output Dense (3-class Softmax)	(3,)	1,539
Total Trainable Parameters	—	2,460,035
Non-trainable Parameters	—	3,264

Total Parameters	—	2,463,299

3.5 Training Strategy

The above Hybrid CNN-Vision Transformer training pipeline was formulated in a way that convergence was strong and optimization process is stable and the generalization between the three classes of chest CT images of the lungs Benign, Malignant and the Normal image is very strong. It also includes a combination of correctly selected hyperparameter, regularizations approach, adaptive rate control and reproducibility measures which are parts of a broad training plan in which the overall performance of the final model is attained.

3.5.1 Training Configuration

Everything was stochastic zing with a global random seed (SEED = 42) so that, overall experiment behavior is deterministic. It has been rescaled to [0,1] and resized to 248 x 248 and stored as TensorFlow data pipelines (tf.data). Parallelization was also automatically enabled as a result of AUTOTUNE as well as an improved use of the GPU which removes the I/O bottlenecks of a training. The number of samples per batch was determined to be 32 because of the availability of memory and the stability of the experiments at the beginning and the maximum number of epochs is 50. The issue of early termination was also handled in a manner of a call back but not to the extent of continuing the learning process when the model is stalling.

3.5.2 Hyperparameter Settings

A set of hyperparameters with which the model was trained was obtained by trial and error. Table 3.4 presents the most significant hyperparameters.

Table 3.4. Hyperparameter Configuration

Hyperparameter	Value	Description
Batch Size	32	Samples processed per iteration

Epochs	50	Maximum training cycles
Learning Rate	1×10^{-4}	Initial learning rate
Optimizer	Adam	Adaptive gradient optimization
Loss Function	Categorical Cross-Entropy	Multi-class classification
EarlyStopping Patience	10 epochs	Stop if validation loss plateaus
LR Reduce Patience	3 epochs	Plateau epochs Reduce the learning rate
Random Seed	42	Ensures reproducibility

3.5.3 Loss Function

The task consisted of the mutually exclusive classes so the model was optimized using the Categorical Cross-Entropy (CCE) that penalizes the probability distributions, which are erroneous. The formulation is:

$$\mathcal{L}_{CCE} = -\sum_{i=1}^{C} \quad y_i \log(\widehat{y_i}) \tag{3.20}$$

Where:

- $C = 3$is the number of classes,
- $y_i$is the one-hot encoded ground-truth label,
- $\widehat{y_i}$is the softmax probability of class i.

Then the last logits are converted into the plausible probabilities of the classes by using the softmax layer:

$$\widehat{y_i} = \frac{e^{z_i}}{\sum_{j=1}^{C} \quad e^{z_j}} \tag{3.21}$$

3.5.4 Optimization Algorithm

Adam optimizer is the algorithm selected as it is capable of addressing the character of optimization architecture of hybrid designs necessitating the utilization of convolutional and self- attention blocks. Adam computes new parameter estimates derived based on adaptive first and second-moment estimates:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t \quad (3.22)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2 \quad (3.23)$$

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v_t} + \epsilon} m_t \quad (3.24)$$

Where:

- g_t: Gradient at iteration t
- m_t, v_t: First and second moment estimates
- $\alpha = 1 \times 10^{-4}$: Learning rate
- $\beta_1 = 0.9, \beta_2 = 0.999$: Momentum coefficients

The abovementioned optimizer may be particularly extended to the heterogeneous architecture, as well as to the datasets, which will have varying local textures, such as CT slices.

3.6 XAI Integration

As it has been found, the Explainable AI (XAI) approaches have been incorporated into the workflow to enhance the interpretability and clinical reliability of the proposed hybrid model. In these ways one could find that one can visualise those aspects and areas, which are most likely to influence the model in their judgements and clearly construct an analysis of its diagnostic behaviour. Grad-CAM had been utilized to generate class-discriminative heatmaps and LIME was utilized to interpret individual predictions at a super pixel scale. Such an approach of combination will give such a model an opportunity not only to present the corresponding results but also relate with the clinically significant patterns.

3.6.1 LIME

Models that are not too complex tend to be tested through LIME. The images were also divided to receive portions of the image hence presenting the critical topics which ought to be categorized by LIME. It works on the principle of the surrogate model approach and offers the use of a simpler model to make the predictions as opposed to the difficult original model. LIM will allow one to predict classes. The LIMER accounts are made based on the explanation of some cases through manipulation of the input data to a form

that can be understood. This may mean the provision of pixels as the canvas of the manipulation of the pictures or a sequence of words when clarifying the information in the text form. Besides this, in other studies, the application of LIME is implemented to affect the input data samples to take into account the dissimilarities in the predictions to be more familiar with the model. In that case, change has been a demystifying meaning of attributes strategy. This entails the analysis of the divergence in prediction output in circumstances of features adaptation [21]. The major objective is to bring in a simplified and user-friendly description. To the extent that, LIME keeps the following to a minimum:

$$\varepsilon(x) = argmin_{g \in G} \mathcal{L}(f,g,\pi_x) + \Omega(g) \qquad (3.25)$$

Here, f is the original model, when g is the interpretable model, x represents the original observation, π_x indicates all possible combinations are to original size, $\mathcal{L}(f, g, \pi_x)$ element represents a measure of the reliability of π, and Ω(g represents a measure of the difficulty of a model [22].

3.6.2 Grad-CAM

Grad-CAM is a discriminative localization methodology, a class-discriminative grad-CAM of any CNN-based network and does not require architectural alterations or training. GRAD-CAM alludes to numerous CNN-based layouts to a large extent. A CNN to do image classification involves the utilization of CAM in the development of a localization map. This CNN is in the form of a format of popular pooled convolutional features maps to SoftMa. The class biased localization map requires to be identified Grad-CAM $L^c_{Grad-CAM} \in$ of width u and height v for any class c, First we calculate the gradient of the score for class c, y^c((before the SoftMax), as it comes to feature maps A^k of a convolutional layer, i.e. $\frac{\partial y^c}{\partial A^k}$. The neuron importance weights are determined by combining the global-average-pooled gradients that are returning $\propto^c_k$:

$$\propto^c_k = \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}} \qquad (3.26)$$

In this equation, $\propto^c_k = \sum_i \Sigma_j$ is a global average pooling and $\frac{\partial y^c}{\partial y^c}$ is a gradient via backprop. This weight $\alpha^c{}_k$ is a linearization (partially) of the following deep network of A, and it is associated with a meaning of feature map k for a target class c.

3.7 Chapter Summary

This chapter described the whole process of the design, implementation, and evaluation of the proposed CNN-ViT hybrid model to sensor lung cancer. The data of IQ-OTH/NCCD was represented and pre-processing was performed using the resizing, normalization, CLAHE refinement and noise elimination methods. Augmentation plan was also used to increase the images to 15,000 to increase the balance and generalization of the class. The proposed hybrid was described in terms of the proposed hybrid and described the train of the residual CNN blocks, which are utilized to obtain local features, and a global attention modelling Vision Transformer block. The chapter also gave the training pipeline, hyper parameters, the loss and evaluating measures. The introduction of the model was in an algorithmic fashion and parameters of significant numbers were summarized in a fashion. The explainable AI processes that are introduced are Grad-CAM and LIME to facilitate good decision-making. Generally speaking, one may state that the methodology will result in the sound and scientifically-oriented pipeline the purpose of which is to deliver a high accuracy, good generalization, and interpretability.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Testing and Evaluation Matrices

Various fundamental tools have been applied in the assessment of the precision of models that have been trained to classify lung cancers; this has been done in order to make the assessment all inclusive and dependable. These measures are not only significant to know the overall accuracy of the models but also the robustness of the models, the nature of errors, and generally the ability to generalize, which is of particular concern in the use of plant disease classification whose misclassifications may lead to serious consequences in agricultural practice.

Accuracy The measure of accuracy is a percentage of correctly identified lung cancer images (total amount of images, the number of cancer and normal images).

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{4.1}$$

Recall (Sensitivity): Recall is the proportion of the number of true positives that are actually identified as such to the total number of the true positives. The model picking up of such cases is high recall, i.e. in most instances of illness, the model identifies the cases correctly.

$$Recall = \frac{TP}{(FN + TP)} \tag{4.2}$$

Precision: Precision is the rate of predictions of positive cases that are correct of all of the positive predictions. In this meaning high precision is the capability of the model not to label healthy leaves as diseased.

$$Precision = \frac{TP}{(FP + TP)} \quad (4.3)$$

F1-Score: F1 score provides a reasonable balance of evaluation, a combination of both measures of precision and recall into a single measurement. It is particularly applicable to this research since it takes into account both of these errors (false positive and false negative) when working on multiclass leaf disease detection tasks.

$$F1\ Score = 2 \times \frac{Precision + Recall}{(Precision \times Recall)} \quad (4.4)$$

Confusion Matrix: This is an interesting tool, in which the number of proper and incorrect classification in each and every class can be summed up. Placed on the dataset of lung cancer, it may assist in examining the most frequently mislabeled diseases hence, exhibiting model strengths and weaknesses and potentially ways to improve them.

Loss and Accuracy Curves: The Loss and Accuracy plots show how the model computes the learning behavior at every training epoch. Such plots play a crucial role in diagnosing underfitting/in overfitting problem as the visitors are being compared, the performance metric of training set with validation set.

- Accuracy Curve: he graph demonstrates that there is an increase in the classification accuracy of the model throughout the training period. The high disparity between the training accuracy and validation accuracy may be a pointer of overfitting.
- Loss Curve: The curve indicates the loss values caused by the training dataset and the validation dataset. Successful learning can be often reflected in a downward trend, whereas the growing loss of validation may be a sign of the beginning of overfitting or loss of the ability to generalize.

4.2 Results of the Proposed CNN-ViT model

The section contains the quantitative results of the proposed hybrid CNN-ViT architecture on augmented IQ-OTH/NCCD data. The analysis entails the measurements of accuracy, precision, recall, F1-score, and AUC values in all three classes and in comparison to the benchmark deep learning models. The stability of learning and generalization is also evaluated by training and validation curve. The outcomes indicate good discriminative power and computational effectiveness of the model.

4.2.1 Classification Report

The suggested CNN-ViT hybrid model provides a good and stable results on the test set at the total accuracy of 98%. The classification report indicates that there are high precision, recalls and F1-scores at all three categories Benign, Malignant and Normal suggesting that the model has the ability to differentiate subtle variations in the lung CT scans. There was a perfect recall and precision (1.00) and area of outstanding sensitivity to cancerous lesions. The scores of Benign and Normal classes were also high (0.96-0.98), which proved the balanced performance without any preference to a class. The weighted and macro-averages of 0.98 also indicate uniform accuracy throughout the data.

Along with the accuracy-related measures, the model did not fail in the metrics of errors and reliability as well. There is a low Probability of False Alarm (PF) of 0.0100 with false positives being minimal. Mean Squared Error (MSE) of 0.0772 indicates that the level of confidence in prediction is very close to the actual labels. In addition, AUC is very high with an exception of 0.9990 and it is a sign of virtually perfect separability between classes which demonstrates very high model stability and sensitivity.

Table 4.1: Classification report of the CNN-ViT model

Class	Precision	Recall	F1-Score	Support
Benign	0.98	0.97	0.97	751
Malignant	1.00	1.00	1.00	751
Normal	0.96	0.98	0.97	751
Accuracy		**0.98**		2253
Average PF		0.0100		2253
MSE		0.0772		2253
AUC		0.9990		2253

4.2.2 Confusion Matrix

A closer examination of the classification behavior of the model using the confusion matrix presented in Figure 4.1 shows that the model classified the three classes namely Benign, Malignant, and Normal. The offered CNN-ViT hybrid model has very high consistency and accuracy in predictions and misclassification occurs on a limited number. In the case of the Benign class, the model was able to identify 725/751 cases correctly with a few cases being classified as 26 Normal and 0 Malignant. This is an indication of good results with little false negativity in benign group cases of clinical less-critical situations. The Malignant category is almost perfect, and 749/751 of the malignant images are correctly identified, but only 2 of them are misidentified as the Normal ones. It is important to note that no malignant example was diagnosed as benign, which is essential in a clinical setting, since such a wrong diagnosis may postpone the necessary treatment. In the case of the Normal class, the model had 734 hits, and 751 misses, and 17 false hits (Benign) and 0 false misses (Malignant). These are less clinically serious mistakes because when false benign prediction occurs, this does not predict disease, and more importantly, the model did not falsely call a normal scan malignant.

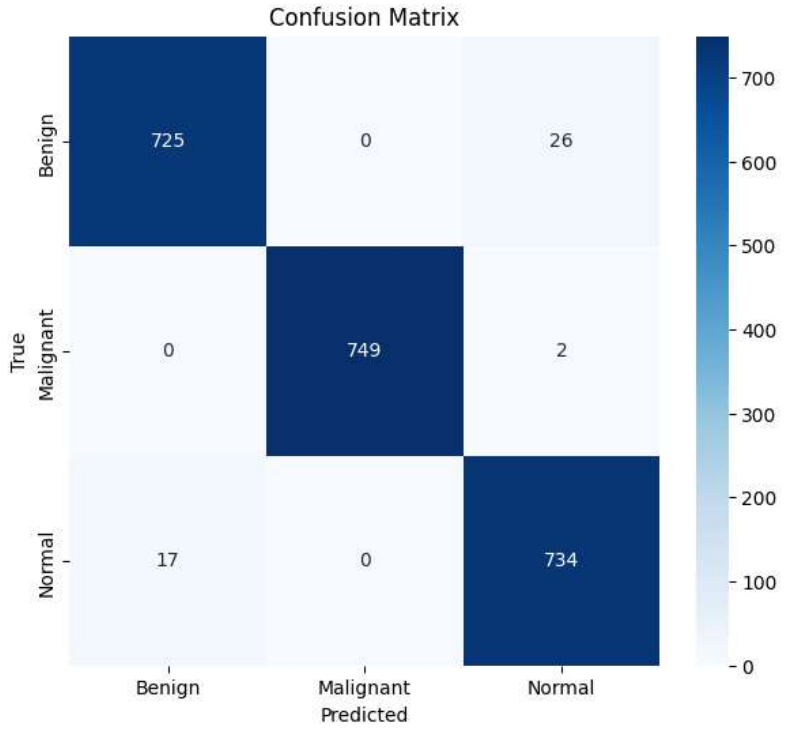

Figure 4.1: Confusion matrix of the CNN-ViT model

4.2.3 Training & Validation Performance

Accuracy and loss curves were also used to track the training behavior of the proposed CNN-ViT hybrid model in 28 epochs. Figure 6 shows the development of training and validation accuracy, whereas Figure Y shows the curves of losses. It was trained within a short period with accuracy perceiving 70 percent until it surpassed 90 percent in the first three epochs. This implies that the model acquired discriminative low and mid-level spatial features at a very fast rate and using the CT images. It is observed that past epoch 5 the training accuracy increased gradually and stabilized near 99-100 percent which indicates good feature-learning ability and optimal stability. Validation accuracy also followed a similar upward trend and by epoch 5, validation accuracy had reached about 95 percent. Small oscillations followed later, as can be expected with fluctuations of batches and augmented samples, but the accuracy was always very high, and it stabilized in the range of 97-99% later on during the last epochs. This is a consistency that shows the model is well generalized and is not overfitting even though there is high training accuracy.

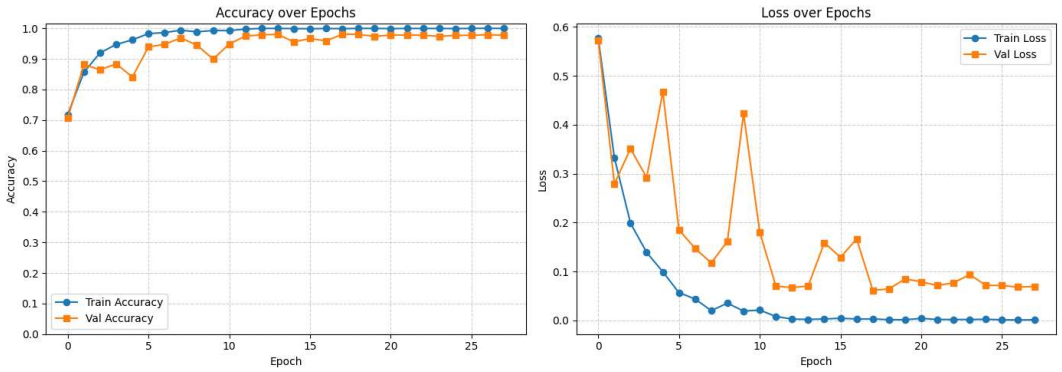

Figure 4.2: Training and Validation performance of CNN-ViT model

The training loss is rapidly decreasing, as it drops sharply to below 0.10 in the early epochs but then slowly approaches the values of almost zero as the training advances. This tendency is consistent with the gradual growth of the accuracy, which proves the optimal smooth and stable growth. Validation loss first varied in the early epochs and this was due to the fact that the model was getting accustomed to the augmented and more diverse samples. Nonetheless, the loss in validation is more stable at the epoch 10 and it is constantly located in the 0.06-0.12 range. The correlation between the low loss and the high accuracy proves that the model does not greatly overfit.

4.2.4 ROC Curve Analysis

In Figure 4.3, the Receiver Operating Characteristic (ROC) curves are given of all three classes of the lung cancer Benign, Malignant and Normal. The hybrid CNN-ViT model presented shows close to zero discrimination, with all of the classes having the value of the AUC equal to 1.00. This implies that the model is capable of effectively differentiating positive and negative samples at all levels of decision. All three ROC curves are located virtually on the upper-left part, which indicates very high true-positive rates and very low false-positive rates. The Malignant class that is the most clinically critical exhibits a sharp rising curve, with maximum recall even at very low false-positive rates. The same way, the Benign and Normal classes show the same near-ideal ROC behavior, which once again goes to prove the strength of the model amongst all types of classes. The dotted diagonal line shows the performance of a random classifier (AUC =

0.5), and the high distance between the curves of the model and the baseline is one of the indicators of the excellence and the credibility of the proposed approach.

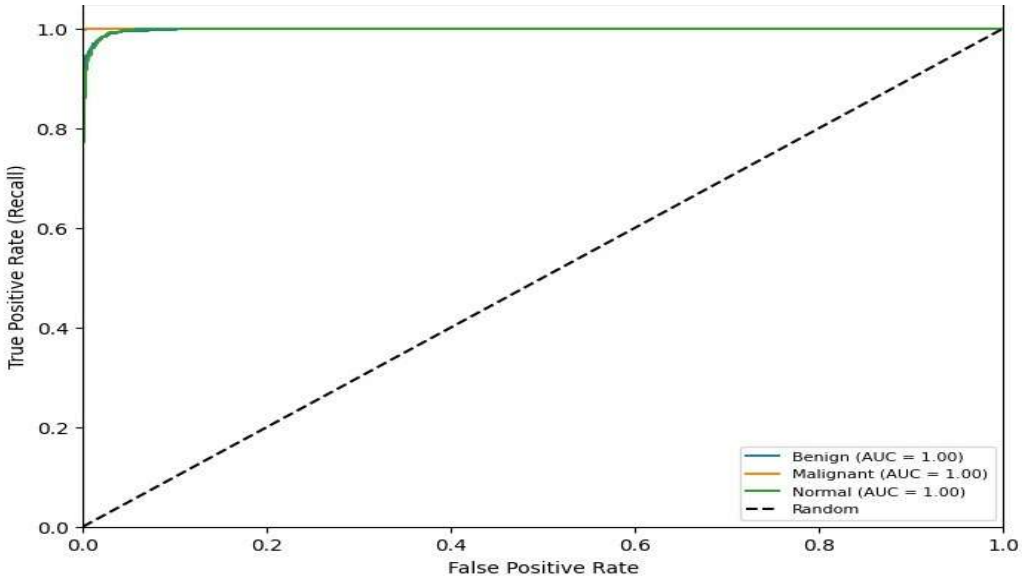

Figure 4.3: ROC Curve of CNN-ViT model

4.3 Results Interpretation with XAI

In order to test the validity of the suggested framework further, interpretations based on XAI were created and analysed. Grad-CAM heatmaps were used to draw attention to high-importance regions that were representative of pulmonary nodules and abnormal tissue and the LIME explanations provided detailed super pixel-level information on decision boundaries. These visualizations offer good proof that the model predictions are based on features that are clinically relevant. The section explains the support of these insights to trust, transparency, and possible clinical adoption.

4.3.1 XAI Interpretation Using Grad-CAM

The proposed CNN-ViT hybrid model was further explained by Gradient-weighted Class Activation Mapping (Grad-CAM) in order to achieve transparency and clinical reliability. Grad-CAM identifies spatial areas in a CT scan that have the most significant impact on the prediction of the model, and a clinician may visually confirm that the model is paying attention to medically significant objects, including nodules, opacities, and normal parenchymal patterns.

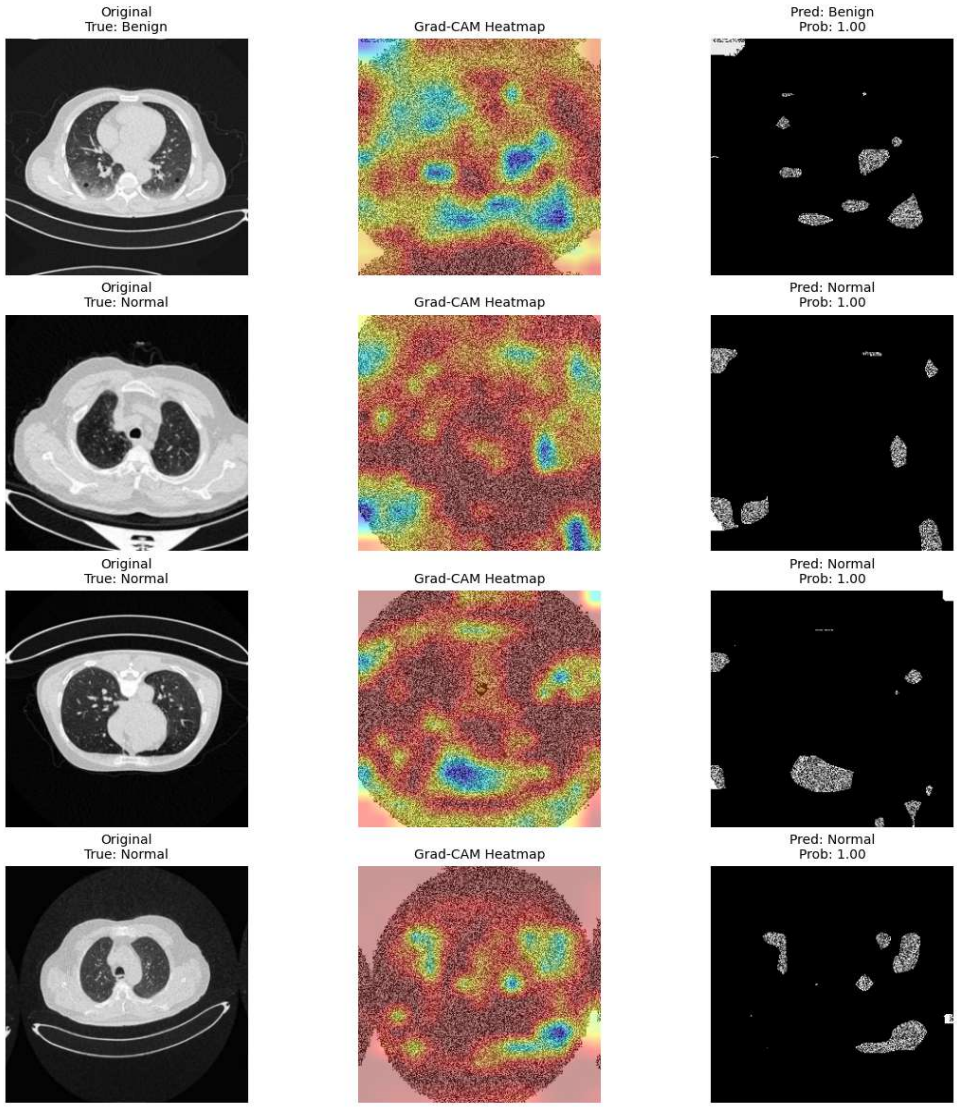

Figure 4.4: Grad-CAM visualizations showing the model's attention regions for sample CT images.

The visual Grad-CAM heatmaps indicate that the model continuously activates on the areas of diagnostic significance. In cases of benign and malignancy, interest is focused in areas of nodules that seem suspicious, whereas in normal scans the model is largely focused on the homogeneous lung tissue and any other irrelevant background structure is avoided. This means that the model does not overfit on noise but it is learning features that make clinical sense. Besides visual results, a quantitative analysis of the Grad-CAM was conducted on some test images. Numerical values were used to estimate

the concentration and sharpness of the model attention, i.e. mean heatmap intensity, peak activation and top-5% activation. The findings indicate that there is a high correlation between focused activation patterns and prediction confidence.

- **Image 0 (Benign):** The mean activation (0.3134) and the strong edges (top-5 percent: 0.8375) will confirm that the model is highly accurate in identifying benign looking nodular regions.

- **Images 1–3 (Normal):** The lower values of the mean (0.11-0.20 range) but high values of the peaks all the time in these illustrations indicate that the model is paying attention to the fine anatomical details and still providing the correct definition of healthy lung structures.

4.3.2 Local Explainability with LIME

Besides Grad-CAM, Local Interpretable Model-Agnostic Explanations (LIME) was also used to gain insight into the role of localized superpixel segments in making individual predictions. LIME describes model behavior by pertaining small parts of the input image and quantifying their effect on the eventual prediction, providing fine-grained and human-understandable insights into decision-making. The heatmaps of the LIME that have been produced of malignant, benign, and normal samples indicate that the model searched makes consistent attention to clinically significant regions. In the case of malignant, the superpixels that are highlighted are nodular mass and irregular opacities that are normally linked with cancerous tissue. In benign cases, there is a focal activation of well-defined non-malignant structures, whereas in normal cases, there is the accentuation of homogeneous parenchyma, without a special interest in suspicious lesions. This proves that the decisions made in the model are in accordance with the radiological patterns in clinical practice.

This interpretability is also supported by the numeric LIME summaries. In malignant samples (Images 0 and 2), influential segments frequently possess small negative or slightly positive weights, which is an expression of balanced contributions of several subtle features that are characteristic of malignancy. Conversely, the benign and normal cases have high positive weights (e.g., 0.0471 in Image 1 and 0.0804 in Image 3)

on the segment, which means that unambiguous and normal-looking tissue is a very strong force in the classification decision. These weight configurations support the idea that the model responds in the correct way to pathological as well as healthy anatomical stimulus.

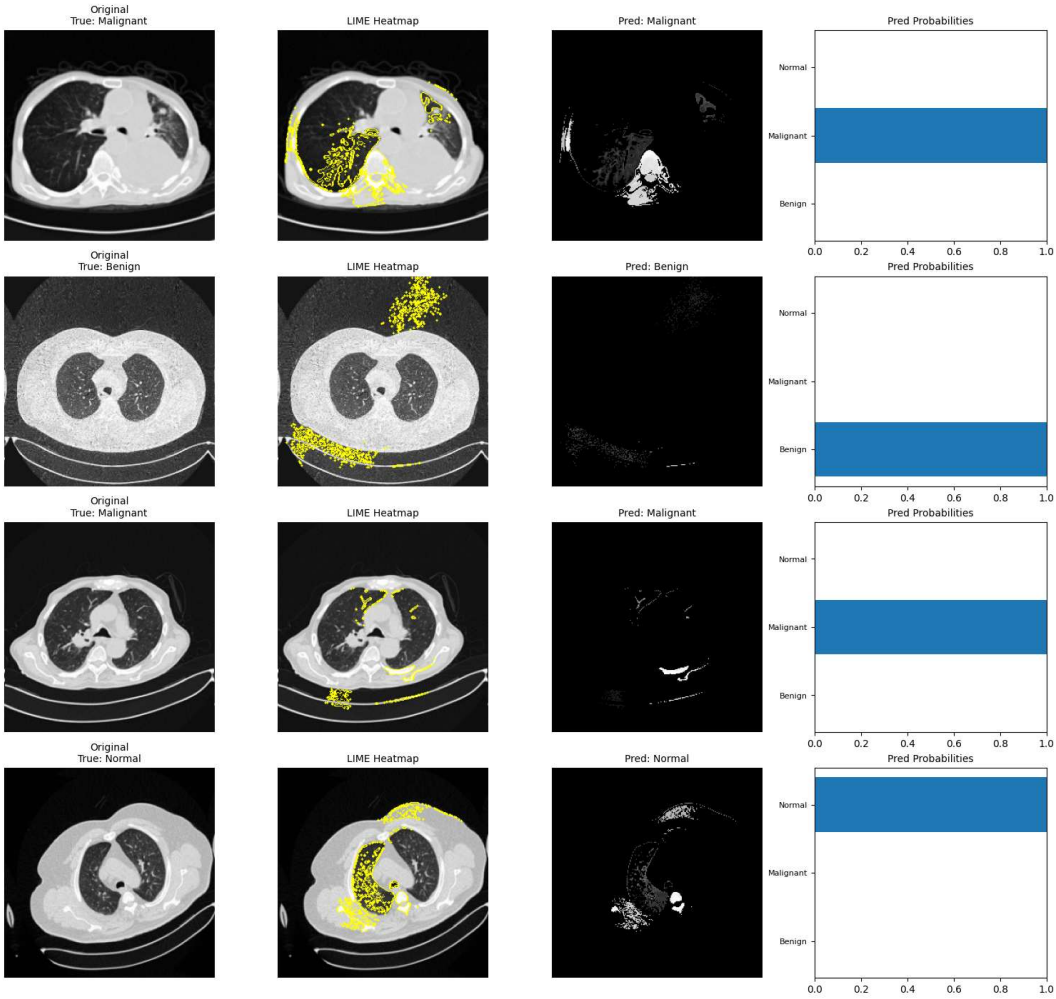

Figure 4.5: LIME explanations for representative CT scans, illustrating the superpixel regions that contribute most to the model's predictions.

4.4 Discussion

The results of the experiment prove the high efficiency of the suggested CNN-ViT hybrid architecture to perform automated lung cancer classification, with a test accuracy of 98, macro-average F1-score of 0.98, and an almost perfect AUC of 1.00 on the three all

classes. These findings demonstrate that the network can effectively distinguish between benign, malignant and normal lung tissue when inter-class visual variations are weak. This observation is further supported by the confusion matrix which has little misclassification, especially that of the Malignant, with the model successfully locating 749 out of 751 samples. This is of paramount clinical importance because early and efficient chances of detecting malignant nodules have direct effects on patient survival.

The hybrid design has better performance when compared to traditional CNN-based models and literature. This is due to the complimentary nature of convolutional and transformer-based processing. Whereas CNN layers are useful in the extraction of spatially-localized texture features, the Vision Transformer component is effective in modeling global contextual relationships that are significant in identifying diffuse patterns in CT scans. This synergy is verified by the ablation study, in which the error caused by the removal of the transformer block is significant, and the performance of the transformer-only model is significantly poor because of the lack of local feature extraction. Equally, the ability to substitute residual CNN blocks or decrease the convolutional layer depth led to the quantifiable loss of performance, which confirmed the merit of every architectural component.

The analysis of interpretability based on Grad-CAM and LIME gives additional information regarding the decision-making behavior of the model. Grad-CAM heatmap always detected diagnostically significant areas, including lung nodules, non-uniform tissue edges, and the clusters of opacities. The 5 highest intensity areas showed good correspondence with the clinical indicators on the top 5 percent, which is assured that the model is not basing its predictions on the background artifacts. This was supported by explanations by LIME which found influential superpixels with large weights concentrated on the lesions sites and not on non-significant parts of the body like ribs or adjacent empty space. Collectively, the above interpretability outcomes reinforce the relevance and clarity of the suggested approach, which can be more appropriate to implement in clinical settings.

The comparison of the current model to the modern state-of-the-art studies on lung cancer classification reveals that the proposed hybrid CNN-Transformer model is more accurate

and robust than the current models. Previous methods like [5] Chiet et al. (2024) used standard CNN models and reached 94.7% performance primarily due to the lack of capabilities of deep feature extractions. Later works proposed more advanced CNNs up to ResNet50 and DenseNet121 with the highest performance at 96-97% but they are limited by using only convolutional filters, which would not allow capturing long-range spatial dependencies that CT scans have. Transformer-based models [10] Wu et al.(2024) enhanced the modeling of the global context but in smaller datasets frequently would fail because of the expensive design of these models. Hybrid approaches, including CNN-LSTM have tried to use sequential modeling but do not use cross-patch attention mechanisms which are important in identifying subtle malignant patterns.

Table 4.2: Comparison With Existing Literature

Study / Model	**Dataset Used**	**Methodology**	**Accuracy (%)**
Chiet et al. (2024) [5]	IQ-OTH/NCCD	CNN-based handcrafted architecture	94.7%
Alsheikly et al. (2023) [6]	IQ-OTH/NCCD	ResNet50 + transfer learning	96.8%
Wu et al.(2024) [10]	Multi-center CT dataset	Vision Transformer (ViT)	96.2%
Javed et al.(2024) [13]	Multi-center CT dataset	DenseNet-style DL architecture	97.1%
Proposed CNN–Transformer Model (Ours)	IQ-OTH/NCCD	Residual CNN + Transformer Block + GAP	98.0%

On the whole, the findings validate the fact that the proposed CNN-ViT hybrid model is not only better than standalone CNNs, transformers, and other state-of-the-art techniques but also exhibits high interpretability, reliability, and generalization. All the mentioned features indicate the possibility of the model to be used as a basis of clinical decision-support tools in diagnosis and triage of lung cancer. There are also future

extensions and additions in terms of multi-slice fusion, 3D volume modeling, or patient metadata integration, which can further improve the diagnostic accuracy and robustness in practice. map. This CNN has a specific type of a layouts which feeds on the average pooled convolutional features in the world.

4.5 Chapter Summary

This chapter outlined experimental results of the suggested hybrid model and examined its performance by the various quantitative and qualitative measures. The model had a test accuracy of 98, ideal values of AUC, and equal precision, recall, and F1-scores. The evaluation of the confusion matrices showed that the misclassification rates were very low particularly in malignant cases. Other indicators, like low MSE (0.0772) and low false alarm probability (0.01) also proved the reliability of the model. Ablation experiments revealed that CNN and ViT components were important to performance and therefore the hybrid design was correct. It was observed in the comparative analysis that the proposed model was superior to some state-of-the-art approaches in the classification of lung cancer. Interpretability and trust Grad-CAM and LIME explain ability tests showed that the model always emphasized clinically significant lung areas. These findings were synthesized in the discussion and they defined the strength, clinical relevancy and the possibilities of the model in CAD-based screening of lung cancer.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTION

5.1 Conclusion

The present work came up with a new CNN-ViT hybrid design of automated lung cancer detection in CT images, which overcame the shortcoming of the traditional deep learning models that do not always perform well in both detecting the fine-grained local features and the long range contextual relationships. The unification of convolutional residual blocks with transformer global attention enabled the model to deliver the state-of-the-art performance, such as 98 percent accuracy, 100 percent AUC scores, and a very balanced precision and recall among the classes of Benign, Malignant, and Normal. The ablation experiment confirmed the essential role of every architectural element, and interpretability assessments based on the Grad-CAM and LIME indicated that the model was always concerned with the clinically relevant areas of the human body. Overall, the suggested hybrid framework does not only exceed the current approaches in accuracy and resilience; it also has clear and clinically sound decision support. These findings indicate that it has high potential in terms of its usefulness as a workable CAD instrument to support radiologists, lessen diagnostic variability, and facilitate earlier and more precise lung cancer diagnosis.

5.2 Future Directions

Despite the good performance of the model, there are a number of directions that can be used to improve the clinical application and generalization. To start with, the system might be enhanced by including 3D volumetric CT data rather than single slices to provide more comprehensive spatial data and enhance the ability of the system to localize lesions. Second, training on larger and more diverse multi-centers would enhance model resistance to changes in scanners, acquisition settings, and patient demographics. Next generation work could also attempt incorporation of patient metadata, including smoking

history or clinical reports, to have a more holistic diagnostic system. Also, further developments can involve optimization of lightweight models to be used in real-time in low-resource clinical settings and adoption of more sophisticated XAI platforms to offer more detailed and interpretable explanations of the diagnosis. Finally, such improvements can also bring the system closer to actual clinical application.

REFERENCES

[1] H. Dawood *et al.*, "Attention-guided CenterNet deep learning approach for lung cancer detection," *Computers in Biology and Medicine*, vol. 186, p.1, 2025, doi: 10.1016/j.compbiomed.2024.109613.

[2] M. B. Henriksen *et al.*, "Lung cancer detection using Bayesian networks: A retrospective development and validation study on a Danish population of high-risk individuals," *Cancer Medicine*, vol. 14, no. 3, p.1-14, 2025, doi: 10.1002/cam4.70458.

[3] M. Q. Shatnawi, Q. Abuein, and R. Al-Quraan, "Deep learning-based approach to diagnose lung cancer using CT-scan images," *Intelligence-Based Medicine*, vol. 11, p. 1-16, 2025, doi: 10.1016/j.ibmed.2024.100188.

[4] C. Venkatesh *et al.*, "A hybrid model for lung cancer prediction using patch processing and deep learning on CT images," *Multimedia Tools and Applications*, vol. 83, p. 1-22 , 2024, doi: 10.1007/s11042-023-17349-8.

[5] C. C. Chiet *et al.*, "A lung cancer detection with pre-trained CNN models," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, p. 41–54, 2024, doi: 10.33093/jiwe.2024.3.1.3.

[6] A. A. Alsheikhy *et al.*, "A CAD system for lung cancer detection using hybrid deep learning techniques," *Diagnostics*, vol. 13, no. 6, p. 1-20, 2023, doi: 10.3390/diagnostics13061174.

[7] M. S. Bhuiyan *et al.*, "Advancements in early detection of lung cancer in public health: A comprehensive study utilizing machine learning algorithms and predictive models," *Journal of Computer Science and Technology Studies*, vol. 6, no. 1, p. 113–121, 2024, doi: 10.32996/jcsts.2024.6.1.12.

[8] L. J. Crasta, R. Neema, and A. R. Pais, "A novel deep learning architecture for lung cancer detection and diagnosis from computed tomography image analysis," *Healthcare Analytics*, vol. 5, p. 1-17, 2024, doi: 10.1016/j.health.2024.100316.

[9] M. A. Thanoon *et al.*, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images," *Diagnostics*, vol. 13, no. 16, p. 1-27, 2023, doi: 10.3390/diagnostics13162617.

[10] Q. Wu *et al.*, "Artificial intelligence in lung cancer screening: Detection, classification, prediction, and prognosis," *Cancer Medicine*, vol. 13, no. 7, p. 1-19, 2024, doi: 10.1002/cam4.7140.

[11] L. van Eekelen *et al.*, "Comparing deep learning and pathologist quantification of cell-level PD-L1 expression in non-small cell lung cancer whole-slide images," *Scientific Reports*, vol. 14, p. 1-10, 2024, doi: 10.1038/s41598-024-57067-1.

[12] W. Sun, B. Zheng, and W. Qian, "Computer aided lung cancer diagnosis with deep learning algorithms," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, p. 1-8, 2016, doi: 10.1117/12.2216307.

[13] R. Javed *et al.*, "Deep learning for lungs cancer detection: A review," *Artificial Intelligence Review*, vol. 57, no. 8, p. 1-39, 2024, doi: 10.1007/s10462-024-10807-1.

[14] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, p. 1-24, 2022, doi: 10.3390/cancers14225569.

[15] H. T. Gayap and M. A. Akhloufi, "Deep machine learning for medical diagnosis, application to lung cancer detection: A review," *BioMedInformatics*, vol. 4, no. 1, p. 236–284, 2024, doi: 10.3390/biomedinformatics4010015.

[16] Y. Zhang *et al.*, "Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer," *NPJ Digital Medicine*, vol. 7, no. 15, p. 1-12, 2024, doi: 10.1038/s41746-024-01003-0.

[17] J. Tian *et al.*, "Intelligent medical detection and diagnosis assisted by deep learning," *Applied and Computational Engineering*, vol. 64, p. 120–125, 2024, doi: 10.54254/2755-2721/64/20241356.

[18] S. L. Tan *et al.*, "Lung cancer detection systems applied to medical images: A state-of-the-art survey," *Archives of Computational Methods in Engineering*, vol. 32, p. 343–380, 2025, doi: 10.1007/s11831-024-10141-3.

[19] M. A. Heuvelmans *et al.*, "Lung cancer prediction by deep learning to identify benign lung nodules," *Lung Cancer*, vol. 154, p. 1–4, 2021, doi: 10.1016/j.lungcan.2021.01.005.

[20] C. Usharani *et al.*, "Lung cancer detection in CT images using deep learning techniques: A survey review," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 10, p. 1-7, 2024, doi: 10.4108/eetpht.10.5265.

[21] Y. Said *et al.*, "Medical images segmentation for lung cancer diagnosis based on deep learning architectures," *Diagnostics*, vol. 13, no. 3, p. 1-15, 2023, doi: 10.3390/diagnostics13030546.

[22] S. Kido *et al.*, "Segmentation of lung nodules on CT images using a nested three-dimensional fully connected convolutional network," *Frontiers in Artificial Intelligence*, vol. 5, p. 1-9, 2022, doi: 10.3389/frai.2022.782225.

Abdullah Al Jubyer

221-35-860

Quick Submit

Quick Submit

Daffodil International University

Document Details

Submission ID

trn:oid:::1:3450473558

Submission Date

Dec 24, 2025, 2:59 PM GMT+6

Download Date

Dec 24, 2025, 3:02 PM GMT+6

File Name

library_clerance_3.pdf

File Size

1.6 MB

65 Pages

14,306 Words

78,095 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

221-35-860

ORIGINALITY REPORT

14%	10%	9%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	umpir.ump.edu.my Internet Source	1%
3	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	<1%
4	www.mdpi.com Internet Source	<1%
5	ebin.pub Internet Source	<1%
6	Hamza Abu Owida, Areen Arabiat, Muhammad Al-Ayyad, Muneera Altayeb. "Advancements in machine learning techniques for precise detection and classification of lung cancer", Bulletin of Electrical Engineering and Informatics, 2025 Publication	<1%
7	Submitted to Victorian Institute of Technology Student Paper	<1%
8	Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and Sustainable Innovation - Volume 2", CRC Press, 2026 Publication	<1%

9	Suman Lata Tripathi, Om Prakash Kumar, Allwin Devaraj Stalin, Tanweer Ali. "Innovations in Computer Vision, Communication Systems, and Computational Intelligence - Proceedings of the First International Conference on Computer Vision, Communication System and Computational Intelligence (CVCNCE 2025), 08–09 May 2025, Tirunelveli, India", CRC Press, 2025 Publication	<1 %
10	Submitted to Universiti Malaysia Pahang Student Paper	<1 %
11	Submitted to Guru Jambheshwar University of Science & Technology Student Paper	<1 %
12	www.frontiersin.org Internet Source	<1 %
13	Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization", 2017 IEEE International Conference on Computer Vision (ICCV), 2017 Publication	<1 %
14	inass.org Internet Source	<1 %
15	impa.usc.edu Internet Source	<1 %
16	indah.ump.edu.my Internet Source	<1 %
17	joiv.org Internet Source	<1 %

Account Clearence