COMPARATIVE ANALYSIS OF CNN VARIANTS AND VISION TRANSFOIRMERS FOR EARLY LUNG CANCER DETECTION

by Shivam Srivastava

Submission date: 24-Sep-2026 05:08PM (UTC+0530)
Submission ID: 3040781417
File name: Shivam_thesis_final.pdf (4.89M)
Word count: 8813
Character count: 54761

Comparative Analysis of CNN Variants and Vision Transformers for Early Lung Cancer Detection

MTech Dissertation
submitted in partial fulfilment of the requirements for the award of the degree of

Master of Technology (By Research)
in
Department of Software Engineering
By

SHIVAM SRIVASTAVA
(Roll Number: 24/RSE/08)

under the guidance of
Dr. Sonika Dahiya
Assistant Professor
Department of Software Engineering



DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY, DELHI
2026

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection**" which is submitted by **SHIVAM SRIVASTAVA**, Roll No. **24/RSE/08**, Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Sonika 26/05/26

Place: Delhi

Date:

Dr. SONIKA DAHIYA

SUPERVISOR

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, **SHIVAM SRIVASTAVA**, Roll No. **24/RSE/08** student of M.Tech by Research (Software Engineering), hereby certify that the work which is being presented in the thesis entitled "**Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection**" in partial fulfilment of the requirements for the award of the Degree of Master of Technology, submitted to the Department of Software Engineering, Delhi Technological University, is an authentic record of my own work carried out during the period from June 2024 to June 2026 under the supervision of **Dr. SONIKA DAHIYA**.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Shivam Srivastava

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Sonika 26/05/26

Signature of Supervisor(s)

Signature of External Examiner



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide, **Dr. SONIKA DAHIYA**, for her valuable guidance and wisdom in the development of this project. I would also like to humbly extend my gratitude to **Prof. Ruchika Malhotra** (HOD, Department of Software Engineering) and all the faculty members of the Department of Software Engineering at Delhi Technological University for their constant support, valuable advice, and encouragement throughout the project.

I would also like to thank the University for providing us with the laboratories, infrastructure, testing facilities, and a conducive environment that enabled us to work continuously without interruption. Additionally, I am grateful to our lab assistants, seniors, and peer group for their support, assistance, and knowledge shared on a variety of subjects.

Place: Delhi

Date:

SHIVAM SRIVASTAVA

24/RSE/08

ABSTRACT

Lung cancer continues to be a leading cause of death from cancer, and the prognosis largely relies on prompt and precise detection. Traditional methods of disease diagnosis require the interpretation of radiographs and histopathological images that require a lot of effort and suffer from observer variability. It has become crucial, therefore, to develop deep learning models in the field of computer aided diagnosis due to the ability of modern image models to extract discriminative features of interest and make decisions quickly.

This thesis aims at evaluating several deep learning architectures and comparing their diagnostic capabilities for detecting early signs of lung cancer. One major goal of the thesis is to investigate the effect of deep learning model characteristics on diagnostic performance. These characteristics include model depth, parameter efficiency, feature reuse, transfer learning capability, and attention mechanisms. Architectures of CNNs and transformers will be reviewed with respect to the abovementioned characteristics, specifically Alex-Net, VGG16, ResNet50, DenseNet121, Inception Net, EfficientNetB0, Vision Transformer (ViT), and DeiT. An explanation and justification will follow for the choice of Vision Transformer over other model types. Preprocessing steps in the experimentation framework include resizing, contrast adjustment, normalization, augmentation, and splitting of the public datasets containing histopathology lung cancer

The effectiveness of the algorithms is gauged using performance metrics like accuracy precision, recall, F1 score, the area under the receiver operating characteristic curve, and also confusion matrices. The results from the comparison suggest that deep networks, especially Inception-Net DenseNet121 Efficient-Net and transformer based networks, are promising approaches for lung cancer detection because of their discriminative power and generalization ability. There are other elements that get attention too, such as limitations like dataset biases, small clinical trial groups, computational costs, model interpretability, and the need for external verification before any clinical use, full stop. Overall, this thesis kind serves as an extension of the paper that…

LIST OF PUBLICATIONS

1. **Shivam Srivastava and Sonika Dahiya**

   "CNN-Based Diagnostic Models for Lung Cancer: A Comparative Study of Different Models" in *Proceedings of the International Conference on Artificial Intelligence and Next-generation Networks (ICAIN 2025)*, Lecture Notes in Networks and Systems, Springer, 2026.

2. **Shivam Srivastava and Sonika Dahiya**

   "Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection" for publication at the 3rd International Conference on Innovations in Cybersecurity and Data Science (ICICDS-2026), Lecture Notes in IEEE Xplore digital library 2026, **IEEE ISBN for ICICDS 2026**, XPLORE COMPLIANT ISBN: 979-8-3195-3130-8 DVD ISBN: 979-8-3195-3129-2

Contents

Acknowledgement i

Candidate's Declaration ii

Certificate iii

Abstract iv

List of Publications v

Content vii

List of Figures viii

List of Tables ix

1 Introduction 1

1.1 Problem Statement 2

1.2 Research Objectives 2

1.3 OBJECTIVE 3

2 Technical Background 4

2.1 Introduction to Lung Cancer 4

2.2 Medical Imaging for Lung Cancer Detection 4

2.3 Artificial Intelligence in Medical Imaging 5

2.4 Convolutional Neural Networks 6

2.5 CNN Variants 7

2.6 Vision Transformers 9

2.7 Transfer Learning 10

2.8 Data Augmentation 11

2.9 Explainable Artificial Intelligence 11

2.10 Evaluation Metrics 11

2.11 Summary 12

3 Literature Review 13

3.1 Introduction 13

3.2 Deep Learning in Medical Imaging 14
3.3 Lung Cancer Detection Using CNN Architectures 14
3.3.1 AlexNet-Based Studies 14
3.3.2 VGG16-Based Studies 15
3.3.3 ResNet-Based Studies 15
3.3.4 DenseNet-Based Studies 15
3.3.5 InceptionNet-Based Studies 16
3.3.6 EfficientNet-Based Studies 16
3.4 Vision Transformers in Medical Imaging 16
3.5 Transfer Learning in Medical Imaging 17
3.6 Data Augmentation Techniques 17
3.7 Explainable Artificial Intelligence 18
3.7.1 Grad-CAM 18
3.7.2 SHAP 18
3.7.3 LIME 18
3.8 Comparative Analysis of CNN and ViT Models 18
3.9 Literature Review Table 20
3.10 Research Gap 20
3.11 Proposed Research Direction 21
4 Proposed Methodology 22
4.1 Overview 22
4.2 Dataset 22
4.3 Data Preparation 22
4.4 Models Considered 23
4.5 Transfer Learning Strategy 23
4.6 Training Configuration 23
4.7 Evaluation Protocol 24
5 Experimental Results and Analysis 25
5.1 Introduction 25

5.2 Experimental Environment 26
5.2.1 Hardware Configuration 26
5.2.2 Software Configuration 26
5.3 Dataset Description 26

5.3.1 Dataset Distribution 27
5.4 Data Preprocessing 27
5.4.1 Image Resizing 27
5.4.2 Normalization 27
5.4.3 CLAHE Enhancement 28
5.4.4 Data Augmentation 28
5.5 Training Configuration 28
5.6 Transfer Learning Strategy 29
5.7 Performance Evaluation Metrics 29
5.7.1 Accuracy 29
5.7.2 Precision 29
5.7.3 Recall 30
5.7.4 F1-Score 30
5.7.5 Specificity 30
5.7.6 ROC-AUC Score 30
5.8 CNN Model Performance Analysis 30
5.8.1 AlexNet Performance 30
5.8.2 VGG16 Performance 31
5.8.3 ResNet50 Performance 31
5.8.4 DenseNet121 Performance 31
5.8.5 InceptionNetV3 Performance 32
5.8.6 EfficientNetB0 Performance 32
5.9 Vision Transformer Analysis 33
5.10 Comparative Performance Analysis 33
5.11 Confusion Matrix Analysis 33

5.12 ROC Curve Analysis 34
5.13 Computational Complexity Analysis 34
5.14 Key Findings 34
5.15 Limitations of the Study 35
5.16 Future Research Directions 35
Bibliography **36**

List of Figures

4.3.1 Work Flow chart for implementation of CNN's Variants on LC25000 dataset.. . 18
5.10.1 Performance Comparison of CNN Models on LC25000 dataset.....19
5.11.1 Confusion Matrices of the CNN Variants for Lung and Colon Cancer 20
5.13.1 Training and Validation Accuracy of the CNN Variants for Lung and Colon.... 24
5.14.1 Training and Validation Loss of the CNN Variants for Lung and Colon Cancer..26

List of Tables

3.1 Comparison Between CNNs and Vision Transformers 19
3.2 Summary of Previous Research Studies 20
3.3 Research Gaps Identified in Existing Literature 21
4.1 Architectural Frameworks for Comparison 23
4.2 Training Configuration Parameters 24
5.1 Hardware Configuration 26
5.2 Software and Framework Configuration 26
5.3 Dataset Distribution for Lung Cancer Classification 27
5.4 Training Hyperparameters 28
5.5 Comparative Performance of CNN and ViT Models 33
5.6 Computational Characteristics of Deep Learning Models 34

Chapter 1

Introduction

Lung cancer is one of the hardest illnesses that doctors are currently facing due to the fact that it usually appears at its latest stages, thus reducing treatment opportunities. Consequently, early detection becomes an essential task, which requires the application of effective methods and technologies. One of the fundamental steps toward this end is medical imaging that allows identifying suspicious tissues, nodules, and abnormalities in the lung structure. However, the interpretation of such information requires much time and skills, and different professionals might perceive images differently depending on many factors.

The employment of computer-aided diagnosis became a breakthrough that allowed enhancing the process of disease identification. The problem was that earlier approaches based on hand-crafted features, thresholds, and classic machine learning techniques could not help achieve better accuracy. The situation began to change after deep learning started to be applied to this issue [6, 10]. This approach enabled researchers to design networks that would automatically extract hierarchical representations from raw input data. As a result, CNNs started to be widely used in medical image analysis.

Recent advances have shown that deep neural networks like AlexNet, VGG16, ResNet50, DenseNet121, InceptionNet and EfficientNet can be tinkered with via transfer learning, so they fit lung cancer classification purposes more or less. Each model differs quite a lot in depth, connectivity, how efficient it is computationally, and the ways it actually extracts features from an image. For instance, AlexNet feels kind of shallow and straightforward, compared to others. VGG16 is known for using small convolution filters over and over in a pretty uniform way. ResNet50 adds those skip connections, mainly to handle problems that come up when training very deep networks, so gradients can behave better. DenseNet121 instead relies on dense connectivity, encouraging feature reuse across the network. Then there is InceptionNet which computes features at several spatial scales at the same time, while EfficientNet blends depth, width and resolution in a scalable manner.

Beyond the usual convolutional networks, vision transformers (ViTs) have been getting more and more attention lately, especially for image analysis tasks. Unlike CNNs that mostly lean on convolution kernels, ViTs chop images into patches and then apply self-attention to capture long range relationships between far away regions of an image. Since malignancy depends on local cellular patterns as well as on the bigger tissue organization, using ViTs could be pretty promising for detecting lung cancer. Still, it is necessary to run a comparative assessment of CNN variants and ViTs, to figure out which architecture fits best, and not just guess.

1.1 Problem Statement

The primary problem that will be explored in this thesis is choosing suitable models to detect lung cancer at an early stage with the help of medical imagery. Not only should a model have excellent accuracy, but it should have high recall too. This is because failing to diagnose a disease may lead to a delay. Furthermore, it should be efficient, resilient to imagery variations, and easily understandable by clinicians.

Despite many studies providing favorable outcomes for CNN models, comparing one study with another is quite hard owing to the differences in terms of datasets, pre-processing techniques, training methods, and metrics. Moreover, transformers are being utilized more frequently for medical imaging, especially lung cancer detection, which makes it imperative to compare the two architectures.

1.2 Research Objectives

The objectives of this thesis are:

1. To investigate deep learning for automating the task of identifying lung cancer from medical imagery , even when the data is tricky.
2. To evaluate major CNN variants like AlexNet, VGG16, ResNet50, DenseNet121 , InceptionNet and EfficientNetB0 , using consistent settings.
3. To look into the potential of Vision Transformers for capturing global relationships inside an image and then helping with cancer detection in a more holistic manner.
4. To propose one unified workflow for preprocessing , model training, and evaluation, so the comparisons stay fair, like a controlled experiment, not just a loose test.
5. To analyze model behavior using accuracy, precision, recall, F1, AUC and confusion matrices.
6. To examine how practical deep learning models are , considering strengths and weaknesses , and also what future directions could realistically work..

1.3 OBJECTIVE

The current thesis develops the aforementioned research paper into a thesis report. In addition to that, it expands the scope of the research into a comparative study of various CNN architectures as well as Vision Transformers without losing sight of the original focus. To summarize, the main contributions are:

- An in-depth review on the usage of deep learning in medical imaging for the detection of lung cancer, focusing mainly on relevant methodologies and results.

- An elaboration on how the transformer-based attention mechanism can be used to supplement CNNs in extracting image features.

- Discussion at a thesis level on data preparation, data augmentation, transfer learning, fine tuning, and evaluation criteria.

Chapter 2

Technical Background

2.1 Introduction to Lung Cancer

Lung cancer is kind a considered one of the most dangerous kinds of cancer, and it's also a big threat to public health worldwide. As WHO points out, every year millions of people die due to lung cancer, mainly because detection comes late and the disease progresses fast. It basically begins when abnormal cells grow in an uncontrolled way inside the lungs, leading to tumour formation. Also , depending on how far the disease has gone, those cancer cells can either invade nearby tissues or spread to other parts of the body. Generally speaking, lung cancer is split into two main groups, non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC).

From those groups, NSCLC is the most common one and it includes squamous cell carcinomas along with adenocarcinoma. Since histopathology is what helps separate these variants, tissue morphology becomes something you really should pay attention to, especially when looking for early hints of illness.

In the early stages, lung cancer often causes only small, easy-to-overlook symptoms, like coughing, chest pain, exhaustion, or difficulty breathing. So, by the time patients are diagnosed, the condition frequently has already reached a more advanced stage, which makes treatment harder. That's why it becomes very necessary to use computer-aided diagnosis systems, to support radiologists and pathologists in their work.

2.2 Medical Imaging for Lung Cancer Detection

There are also numerous medical imaging technologies that can provide an image of lung structure, tissue, and any potential anomalies present in the patient's lungs. The number of technologies employed for various purposes varies, as does the stage at which a certain diagnosis is made. There is always a reason for a certain imaging technology being chosen.

Chest X-ray Imaging. Images produced by chest X-rays are considered the most commonly applied imaging technique due to their relatively low price and availability. They are capable of detecting lung nodules, shadowing patterns, and any irregularities within the lungs' tissues. Nevertheless, since X-rays require interpretation of images, the overlapping of anatomical structures and lack of contrast might complicate this procedure.

Computed Tomography (CT). With the help of Computed Tomography, detailed three-dimensional images of patients' lungs can be generated. While having a higher level of sensitivity compared with chest X-rays, CT images are often used in the detection of pulmonary cancers and for tumour localization. Moreover, clinicians can find pulmonary

nodules and smaller tumours even before they show up clearly on simpler images.

Histopathological Imaging. : Histopathological imaging uses microscopic analysis of laboratory-stained tissue samples such as Haematoxylin and Eosin (H&E) staining. This process shows fine details about cellular structures, tissue organization, glandular patterns, and abnormal growth. Histopathological approach is regarded as one of the most effective ways of diagnosing lung cancer**.**
LC25000 data set used in this research includes histopathological images classified into classes like:
• Lung Benign Tissue
• Lung Adenocarcinoma
• Lung Squamous Cell Carcinoma
Automated deep learning tries to differentiate between these tissue patterns and classify images according to their diagnostic categories.

2.3 Artificial Intelligence in Medical Imaging

AI technology has brought significant advancements in modern medicine, allowing image analysis, disease prediction, and intelligent decision support systems in health care. Machine Learning (ML) and Deep Learning (DL) algorithms are extensively used in medicine especially in the fields of radiology, pathology, and biomedical image processing.

In traditional machine learning methods, the process of designing features was needed to train the classifier, such as the use of texture descriptors, edge detection, and statistical approaches. Yet, these were not always applicable across various medical datasets.

Deep Learning solves this problem by learning features automatically without any manual input from raw data. Among the popular architectures are CNNs and ViTs.

2.4 Convolutional Neural Networks

Convolutional neural networks are sort of custom deep learning setups used a lot in image related tasks. CNNs automatically find space-like features in the input imagery, kind of by running convolutions across it, and you know they just learn what matters.

In the convolution layer, the trained filters get slid across portions of the image , so the output becomes a response map. One typical way to write it is:

$$S(i,j) = (I * K)(i,j) \quad (2.1)$$

Where:

- I – Input Image
- K – Convolution Kernel
- S(i,j) – Output Feature Map

The overall CNN model structure usually includes:

- Convolution Layers
- Activation Functions
- Pooling Layers
- Fully Connected Layers
- Softmax Classification Layer

Shallower layers in CNNs handle relatively simple spatial cues , like edges and textures. As the depth increases, the network can interpret richer patterns, including tissues and other diagnostic forms that are harder to spot at first.

Activation Function. An activation function essentially inserts a nonlinear element into the network pipeline. The rectified linear unit , or ReLU, is among the most widely used choices in CNNs:

$$f(x) = \max(0,x) \quad (2.2)$$

ReLu tends to help things converge faster, and it also reduces the vanishing gradients problem , which is a big deal during training.

Pooling Layers. A pooling layer reduces the size of representations and also lowers the computational load. A common option is max pooling:

$$P = \max(x1,x2,...,xn) (2.3)$$

2.5 CNN Variants

Various CNN architectures have been put forward in order to improve feature extraction abilities, overall stability, and also efficiency a bit. AlexNet. Introduced by Krizhevsky et al. [4],

AlexNet is one of the earlier architectures that got notable results in image recognition by using deep CNNs. The core parts of the network look like this: five convolutional layers, pooling layers, ReLU activations, plus dropout regularization, and then fully-connected layers. AlexNet also demonstrated that deep learning works really well for image classification and it still ends up being a commonly used baseline for a lot of medical imaging research.

VGG16. VGG16 architecture, designed by Simonyan and Zisserman [12], is built around the idea of stacking several tiny 3 × 3 convolution filters to raise the feature extraction capacity. Some of the main advantages of VGG16 are: homogeneity of the overall design, strong feature extraction capabilities, straightforward implementation, and a good match for transfer learning. The main downside is kind of obvious though—this model ends up having an enormous number of parameters.

ResNet50. ResNet50 brought residual learning into the spotlight by adding skip connections [2]. More specifically, residual learning works with an equation like:

$$H(x) = F(x) + x \quad (2.4)$$

where: F(x) is the learned mapping and x is the input mapping. Residual learning helps deal with the vanishing...

DenseNet121. DenseNet121 connects each layer to all succeeding layers of dense blocks [3]. All output data from a particular layer will be used as input in subsequent layers. Its advantages include:

- Enhanced gradient flow
- Reuse of feature maps
- Parameter reduction and reuse
- Favourable propagation of information

The effectiveness of Dense-Net architecture in histopathological image classification tasks arises from the availability of low-level texture information through the layers.

Inception-Net. Inception-Net uses parallel convolutions of different sizes to extract multi-scale features [13]. Filters of varying kernel sizes extract information from different scales. This is achieved by combining the following:

- 1 × 1 convolutions
- 3 × 3 convolutions
- 5 × 5 convolutions
- Pooling layers

The result is the extraction of both fine and large-scale features in a single pass.

EfficientNetB0. EfficientNet is designed using the concept of compound scaling [14]. Compound scaling considers three important aspects of network optimization. These are:

- Depth
- Width
- Image input resolution

The formula of compound scaling is defined as follows:

Depth = $\alpha\phi$ (2.5)

Width = $\beta\phi$ (2.6)

Resolution = $\gamma\phi$ (2.7)

where ϕ governs the scaling factor.2.6 Vision Transformers

Vision Transformers (ViTs) were introduced by Dosovitskiy et al. [1]. Unlike CNNs, ViTs process images as sequences of patches.

An image is divided into patches:

$$xp \in \mathrm{R}(P2 \cdot C) \tag{2.8}$$

where:

- P represents patch size
- C represents image channels

Each patch embedding is passed into transformer encoder blocks containing:

- Multi-head self-attention
- Feed-forward neural networks
- Layer normalization

Self-Attention Mechanism. Self-attention computes relationships between image patches:

$$Attention(Q,K,V) = softmax\left(\frac{QKT}{\sqrt{d_k}}\right)V \tag{2.9}$$

where:

- Q = Query matrix
- K = Key matrix
- V = Value matrix

This allows ViTs to model long-range dependencies across the image.

Advantages of Vision Transformers.

- Better global feature learning
- Strong contextual understanding
- Effective long-range dependency modeling
- Improved scalability

However, ViTs generally require large datasets and high computational power.

2.7 Transfer Learning

Transfer learning uses pretrained knowledge from large datasets such as ImageNet and adapts models to medical imaging tasks.

The transfer learning workflow includes:

1. Load pretrained model
2. Freeze lower layers
3. Replace classification head
4. Train custom classifier
5. Fine tune upper layers

Transfer learning reduces training time and improves performance on limited medical datasets.

2.8 Data Augmentation

Medical datasets are often limited in size. Data augmentation artificially increases dataset diversity.

Common augmentation methods include:

- Rotation
- Flipping
- Zooming
- Translation

- Brightness adjustment
- Shearing

Augmentation improves model generalization and reduces overfitting.

2.9 Explainable Artificial Intelligence

Explainable AI (XAI) techniques improve transparency and interpretability of deep learning models.

Grad-CAM. Gradient-weighted Class Activation Mapping (Grad-CAM) highlights image regions influencing model predictions [9].

SHAP. SHAP values measure feature contribution to predictions [7].

LIME. LIME explains model predictions using locally interpretable approximations [8].

These methods are important in healthcare because clinicians require interpretable diagnostic decisions.

2.10 Evaluation Metrics

ROC Curve and AUC. Receiver Operating Characteristic (ROC) curves evaluate classification thresholds, while Area Under Curve (AUC) measures overall discrimination capability.

Accuracy. Accuracy measures correctly classified samples:

$$Accuracy = TP + TNTP ++ TNFP + FN \quad (2.10)$$

Precision. Precision evaluates prediction correctness:

$$Precision = \frac{}{+ FP} \quad (2.11)$$
$$TP$$

Recall. Recall measures sensitivity:

$$Recall = \frac{TP}{TP + FN} \tag{2.12}$$

F1-Score. F1-score balances precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.13}$$

Specificity. Specificity measures correct negative predictions:

$$Specificity = \frac{TN}{TN + FP} \tag{2.14}$$

2.11 Summary

In this chapter, the basic notions that are needed for lung cancer recognition using deep learning methods have been laid out in a kind of essential way. The deployment of medical imaging modalities , CNN structures, Vision Transformers, transfer learning, explainable AI, and the evaluation measures have been discussed quite thoroughly . From there, the groundwork is used to better understand how different CNN variations and Vision Transformers can be used for identifying lung cancers at an early time..

Chapter 3

Literature Review

3.1 Introduction

Deep learning has really changed medical image analysis a lot, mostly because it can automatically pull out features and then sort diseases. Traditional computer-aided diagnosis systems relied on manual, handcrafted feature extraction , things like edge detection, texture descriptions, shape analysis, and various statistical feature engineering approaches . Yet, these older strategies often did not offer enough generality, plus the robustness was limited.

Because of more recent progress in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL), it is now possible to work with highly effective architectures that can discover more abstract representations from raw image data. Among the more attractive options for medical imaging are convolutional neural networks , also known as CNNs, and Vision Transformers (ViTs).

In lung cancer diagnosis, deep learning models are widely used for analyzing chest Xrays, computed tomography (CT) scans, and histopathological images. Histopathological image analysis is particularly important because microscopic tissue examination remains one of the most reliable methods for confirming cancer diagnosis and identifying cancer subtypes.

In this chapter, the basic notions that are needed for lung cancer recognition using deep learning methods have been laid out in a kind of essential way. The deployment of medical imaging modalities , CNN structures, Vision Transformers, transfer learning, explainable AI, and the evaluation measures have been discussed quite thoroughly . From there, the groundwork is used to better understand how different CNN variations and Vision Transformers can be used for identifying lung cancers at an early time.

3.2 Deep Learning in Medical Imaging

Deep learning models are commonly used in the diagnosis of lung cancer through the interpretation of chest Xray images, CT scans, and histopathological images. It is imperative to understand that the interpretation of histopathological images is very essential because the microscopic examination of tissues is still one of the best ways through which the diagnosis of cancers as well as their subtypes can be made.

This chapter has clearly brought to light all the fundamental concepts that are necessary in understanding lung cancer identification through deep learning techniques. All medical imaging methods, various CNN architectures, vision transformers, transfer learning, explainable AI as well as the performance measures have been covered comprehensively. From there, the concept will be extended in order to help us understand how CNN variants and vision transformers can be used to identify lung cancers at an early stage.

The reason why deep learning models perform exceptionally well in health-related applications is due to its ability to learn discriminative features automatically. CNN-based models have achieved impressive results when it comes to:

- Tumor classification
- Organ segmentation
- Disease prediction
- Histopathological image analysis
- Cancer detection
- Medical image enhancement

3.3 Lung Cancer Detection Using CNN Architectures

CNN architectures became the dominant approach for classifying lung cancer, mainly because medical images contain intricate structures and there are repetitive local patterns, so yeah it fits well. Over time, many CNN variants were introduced to push learning efficiency further, keep optimization more stable, and reach better classification accuracy.

3.3.1 AlexNet-Based Studies:

AlexNet brought attention to how strong deep CNN learning can be for image classification, and it delivered notable results compared to older machine learning methods [4]. In some lung cancer studies, researchers used AlexNet, largely because its overall structure is relatively simple and the training is fast, so it kind of makes the pipeline easier.

However , even though AlexNet can still work acceptably for general image categories, it tends to struggle with more complicated data like histopathological images, where the patterns can be subtle and messy.

3.3.2 VGG16-Based Studies

VGG16 has been used across different medical imaging tasks, since its architectural layout is standardized and it shows solid transfer learning abilities [12]. In lung cancer work, VGG16 is often fine-tuned on histopathological images to boost the classification accuracy.

VGG16 provides strong feature extraction capability but requires high computational resources due to its large number of trainable parameters. Still, even with VGG16's strong feature extraction ability, the huge count of trainable parameters makes it computationally heavy, so training can get costly.

3.3.3 ResNet-Based Studies

ResNets make use of residual learning, and the structure plays a crucial role in overcoming problems associated with vanishing gradients [2].
Applications of ResNet50 include:
• Lung nodules classification
• Histopathological cancer detection
• CT images processing
• Radiograph abnormality recognition
Mathematically, residual learning is defined as follows:
$H(x) = F(x) + x$ (3.1)
where,
• $F(x)$ is the residual mapping
• X is the identity mapping

3.3.4 DenseNet-Based Studies

DenseNet structures facilitate better feature flow through connection between every layer and consecutive layers [3]. In addition, DenseNets make reuse of features and facilitate proper gradient flow.

DenseNet structures have proven highly efficient for histopathological image classification since low-level features such as texture remain available in the network.

3.3.5 InceptionNet-Based Studies

InceptionNet makes use of multiple convolutional filters working in parallel [13]. Multi-level feature extraction allows the network to analyze both fine cellular structures and broader tissue organization.

Many studies indicated that InceptionNetV3 proved highly effective for lung cancer classifications owing to its powerful multi-scale feature extraction.

3.3.6 EfficientNet-Based Studies

EfficientNet introduces compound scaling for balancing network depth, width, and input size [14]. EfficientNet provides good efficiency by reducing computational overhead.

The importance of EfficientNet lies in the fact that it is very suitable for:
• Constrained resource systems
• Mobile healthcare systems
• Timely diagnosis systems

3.4 Vision Transformers in Medical Imaging

Vision Transformers (ViT) is an innovation in computer vision. Unlike CNNs, ViTs handle images as sequences of patches via self-attention mechanism [1].
The self-attention mechanism is described by:
QKT !
Attention(Q,K,V) = softmax √ V (3.2)
dk
where:
• Q = Query matrix
• K = Key matrix
• V = Value matrix
ViTs can capture long-term dependency and global contextual relationship between images.
Applications of ViT in medical imaging include:
• Classification of histopathological images
• Tumor segmentation
• Organ detection
• Localizing disease
Nevertheless, ViT requires:
• Big dataset
• Calculative resources
• Extensive pretraining

To counter these shortcomings, hybrid architecture that combines CNN and ViT is recommended.

3.5 Transfer Learning in Medical Imaging

Vision Transformers (ViT) is an innovation in computer vision. Unlike CNNs, ViTs handle images as sequences of patches via self-attention mechanism [1].
The self-attention mechanism is described by:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.2)$$

where:

- Q = Query matrix
- K = Key matrix
- V = Value matrix

ViTs can capture long-term dependency and global contextual relationship between images.
Applications of ViT in medical imaging include:
- Classification of histopathological images
- Tumor segmentation
- Organ detection
- Localizing disease

Nevertheless, ViT requires:
- Big dataset
- Calculative resources
- Extensive pretraining

To counter these shortcomings, hybrid architecture that combines CNN and ViT is recommended.
Datasets for medical imaging tend to be small due to high costs and time consumption in labeling them. Transfer learning mitigates the issue by transferring a pre-trained model which is trained on large scale such as ImageNet.
Some benefits of transfer learning are:
- Cut training time
- Convergence quickly
- Good generalization
- Performance improvements on small datasets

3.6 Data Augmentation Techniques

Data augmentation artificially increases dataset diversity and improves model robustness.

Common augmentation methods include:

- Rotation
- Horizontal flipping
- Zooming
- Translation

- Brightness enhancement
- Shearing

Augmentation reduces overfitting and helps models generalize better across unseen samples.

3.7 Explainable Artificial Intelligence

One of the major problems with deep learning models is their lack of explainability. XAI techniques enhance explainability and foster more confidence in clinicians.

3.7.1 Grad-CAM

Grad-CAM emphasizes the regions of an image that play a role in prediction [9].

3.7.2 SHAP

SHAP values express the significance of features as well as their contribution [7].

3.7.3 LIME

LIME produces local explainability of predictions [8]. Explainability is crucial in the context of applying DL approaches in healthcare due to the requirement of transparency by clinicians.

3.8 Comparative Analysis of CNN and ViT Models

CNNs and Vision Transformers possess different strengths and limitations.

Table 3.1: Comparison Between CNNs and Vision Transformers

Parameter	**CNNs**	**Vision Transformers**
Feature Learning	Local spatial features	Global contextual relationships
Architecture Type	Convolution-based	Attention-based
Data Requirement	Moderate	High
Computational Cost	Moderate	High
Training Stability	Stable	Requires large-scale pretraining
Interpretability	Moderate	Complex attention visualization
Performance on Small Datasets	Better	Comparatively weaker

3.8 Literature Review Table

Table 3.2: Summary of Previous Research Studies

Authors	Model Used	Dataset	Application	Key Findings
Krizhevsky et al. [4]	AlexNet	ImageNet	Image Classification	Introduced deep CNN learning
Simonyan et al. [12]	VGG16	ImageNet	Visual Recognition	Deep uniform architecture improved accuracy
He et al. [2]	ResNet	ImageNet	Image Classification	Residual learning solved vanishing gradients
Huang et al. [3]	DenseNet	CIFAR/ImageNet	Classification	Improved feature reuse and gradient flow
Szegedy et al. [13]	InceptionNet	ImageNet	Object Recognition	Multi-scale feature extraction improved performance
Tan et al. [14]	EfficientNet	ImageNet	Image Classification	Balanced scaling improved efficiency
Dosovitskiy et al. [1]	Vision Transformer	ImageNet	Vision Transformer	Introduced transformer-based vision learning
Litjens et al. [5]	CNN Survey	Multiple Medical Datasets	Medical Imaging	CNNs transformed medical image analysis
Shen et al. [11]	Deep Learning Survey	Biomedical Images	Medical Imaging	Deep learning improved diagnostic systems
Selvaraju et al. [9]	Grad-CAM	Image Datasets	Explainable AI	Improved CNN interpretability

3.10 Research Gap

Although deep learning has achieved significant progress in lung cancer detection, several limitations still exist in current research.

Table 3.3: Research Gaps Identified in Existing Literature

Research Area	**Identified Gap**
Model Comparison	Many studies evaluate only one architecture without comparing multiple CNN and ViT models.
Dataset Generalization	Limited external validation on real-world clinical datasets.
Explainability	Many systems lack explainable AI integration for clinician trust.
Transformer Evaluation	Vision Transformers are still insufficiently explored for histopathological lung cancer detection.
Computational Efficiency	Several studies ignore memory usage and deployment feasibility.
Hybrid Architectures	Limited research on combining CNNs and Vision Transformers effectively.
Small Dataset Learning	ViTs require large datasets and struggle with limited medical data.
Clinical Deployment	Few studies address practical implementation in hospitals and embedded systems.

3.11 Proposed Research Direction

This thesis attempts to address the identified research gaps by:

- Comparing multiple CNN variants and Vision Transformers under identical experimental settings
- Evaluating classification performance using standard metrics
- Investigating transfer learning strategies
- Discussing computational complexity and deployment feasibility
- Including explainable AI methods for interpretability
- Analyzing both local and global feature learning mechanisms

Chapter 4

Proposed Methodology

4.1 Overview

The proposed methodology proceeds on a comparative DL methodology for lung cancer detection. The purpose is to examine different variations of CNN models as well as Vision Transformers through a consistent experimental framework. The process involves data gathering, preprocessing, data augmentation, modeling, transfer learning, fine-tuning, testing, and comparative analysis.

4.2 Dataset

The research will be conducted using public datasets with regard to medical imaging for lung cancer detection. In particular, the paper provided considers LC25000 with histopathological images of lungs and colons. For the purposes of this thesis, the lung classes are considered: benign lung tissue, lung adenocarcinoma, and lung squamous cell carcinoma.

4.3 Data Preparation

The pre-processing step involves standardization of image quality and input size. To begin with, the images are extracted from the compressed folders and classified depending on their classes. The images are then resized to 224 × 224 pixels as most of the pretrained CNNs require such input size. Normalization of the pixel values is also required at this stage. The images may be subjected to CLAHE contrast enhancement to enhance the contrast of the local tissues. In case of using grayscale images, the images will have the channels replicated thrice.

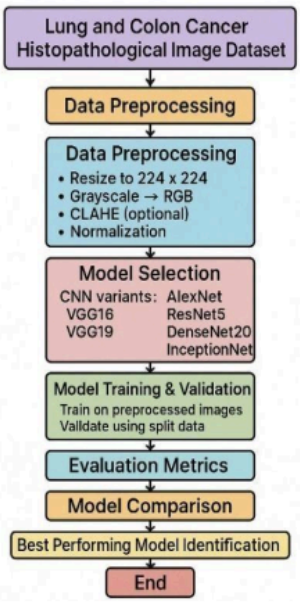


Fig 1. Work Flow chart for implementation of CNN's Variants on LC25000 dataset.

4.4 Models Considered

The comparison includes representative CNN families and a transformer-based model:

Table 4.1: Architectural Frameworks for Comparison

Model	Key Feature	Expected Advantage
AlexNet	Early deep CNN	Simple baseline and fast training
VGG16	Uniform deep convolution stack	Easy implementation and strong transfer learning
ResNet50	Residual skip connections	Stable training of deep layers
DenseNet121	Dense feature reuse	Improved gradient and feature propagation
InceptionNet	Multi-scale parallel filters	Captures features at multiple scales
EfficientNetB0	Compound scaling	Accuracy-efficiency balance
Vision Transformer	Patch self-attention	Global relationship modelling

4.5 Transfer Learning Strategy

Each CNN architecture uses pretrained weights if they exist. The original classification head is stripped and replaced with a new one, comprising global average pooling, dropout, and a dense softmax output layer. For the first phase of training, lower layers are frozen and only the classification head is trained. For fine tuning, some of the higher layers are unfrozen and training is done with a lower learning rate

In case of ViT, patch embeddings and transformers are used for processing patches of images. In the case of having pretrained weights, the transformers undergo fine tuning as before. However, without pretrained weights, a less complex ViT architecture is recommended.

4.6 Training Configuration

A consistent training setup is used for fair comparison.

Table 4.2: Training Configuration Parameters

Parameter	Value
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Batch Size	32
Initial Epochs	5
Initial Learning Rate	0.001
Input Size	224 × 224
Validation Split	20%

Early stopping and model checkpointing are used to prevent overfitting and preserve the best validation model. Fine tuning is performed with a lower learning rate after the initial convergence stage.

4.7 Evaluation Protocol

The models will be evaluated using parameters such as accuracy, precision, recall, F1-score, AUC, confusion matrices, and training curves. In particular, the recall metric will be vital since a misdiagnosis of cancer can delay treatment.

Chapter 5

Experimental Results and Analysis

5.1 Introduction

In this chapter, the experiments conducted on the various CNN models and ViTs for detecting early lung cancer using the histopathological images are presented and analyzed. The purpose of conducting this experiment is to compare their accuracy and robustness in classification, computation, and generalization under identical conditions of training.

The comparative study includes the following models:

- AlexNet
- VGG16
- ResNet50
- DenseNet121
- InceptionNetV3
- EfficientNetB0
- Vision Transformer (ViT)

The experiments have been performed on the LC25000 dataset using transfer learning and data augmentation approach. Evaluation was made on accuracy, precision, recall, F1-score, specificity, ROC-AUC score, and confusion matrix.

5.2 Experimental Environment

Google Colab environment with GPU support was used for conducting the experiments, TensorFlow, and Keras libraries.

5.2.1 Hardware Configuration

Table 5.1: Hardware Configuration

Component	**Specification**
Platform	Google Colab
GPU	NVIDIA Tesla T4
RAM	16 GB
Storage	Google Drive Integration
Operating Environment	Python 3.10

5.2.2 Software Configuration

Table 5.2: Software and Framework Configuration

Software	Version
Python	3.10
TensorFlow	2.x
Keras	Integrated with TensorFlow
NumPy	Latest Stable Version
OpenCV	Latest Stable Version
Matplotlib	Latest Stable Version
Scikit-learn	Latest Stable Version

5.3 Dataset Description

LC25000 Dataset comprises images of lung and colon tissue histopathological images. However, this research specifically targets only the lung tissue images classes. The three classes considered here in the experiment include:

- Lung Benign Tissue
- Lung Adenocarcinoma
- Lung Squamous Cell Carcinoma

All of which contain approximately 5,000 augmented images each.

5.3.1 Dataset Distribution

Table 5.3: Dataset Distribution for Lung Cancer Classification

Class	Number of Images
Benign Lung Tissue	5000
Lung Adenocarcinoma	5000
Lung Squamous Cell Carcinoma	5000
Total	15000

5.4 Data Preprocessing

Medical image preprocessing is essential because histopathological images often contain variations in illumination, staining intensity, resolution, and tissue orientation. The preprocessing pipeline includes:

1. Image resizing
2. Contrast enhancement

3. Normalization
4. RGB conversion
5. Data augmentation

5.4.1 Image Resizing

The steps we included in the pre-processing procedure are basically these:
All images were resized to: 224 x 224 (5.1)
Such standardization of resolution makes the images fit nicely inside the pretrained CNN architecture as well as vision transformers

5.4.2 Normalization

Then we normalized pixels to: [0,1] (5.2)
This kind of scaling helps optimization stay stable and it also improves convergence during training.

5.4.3 CLAHE Enhancement

After that, Contrast Limited Adaptive Histogram Equalization CLAHE was applied , to enhance local tissue contrast so that cell structures can be identified more easily, or at least more consistently.

5.4.4 Data Augmentation

We also used data augmentation , to avoid overfitting and to push better model generalization. The augmentation techniques involved are listed in the next section.

- Rotation
- Horizontal flipping
- Zooming
- Shearing
- Width shifting
- Height shifting

5.5 Training Configuration

For model training, all CNN architectures and ViT models were trained under similar experimental conditions, so there's a fair basis for comparison between models.

Table 5.4: Training Hyperparameters

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	20
Loss Function	Categorical Crossentropy
Activation Function	ReLU
Validation Split	20%

5.6 Transfer Learning Strategy

Transfer learning is very useful for medical image datasets, especially when the number of labels is limited. The transfer learning procedure involved:
1. Loading pretrained ImageNet weights
2. Freezing the lower convolutional blocks
3. Replacing the classification head
4. Fine tuning the upper layers

For multiclass classification, the softmax function was used:
P(yi) = P nezi zj (5.3)
j=1 e

Where ,
• zi denotes output logits
• P(yi) denotes predicted probability

5.7 Performance Evaluation Metrics

Finally, several measures were considered for performance evaluation of the different models.

5.7.1 Accuracy

$$Accuracy = TP + TNTP ++ TNFP + FN \quad (5.4)$$

Accuracy measures overall prediction correctness.

5.7.2 Precision

TP

$$Precision = \frac{TP}{TP + FP} \quad (5.5)$$

Precision evaluates reliability of positive predictions.

5.7.3 Recall

$$Recall = \frac{TP}{TP + FN} \quad (5.6)$$

Recall measures sensitivity and cancer detection capability.

5.7.4 F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5.7)$$

F1-score balances precision and recall.

5.7.5 Specificity

$$Specificity = \frac{TN}{TN + FP} \quad (5.8)$$

Specificity evaluates correct identification of negative samples.

5.7.6 ROC-AUC Score

The Receiver Operating Characteristic (ROC) curve measures classification capability across multiple thresholds, while Area Under Curve (AUC) summarizes overall separability.

5.8 CNN Model Performance Analysis

5.8.1 Alex-Net Performance

Alex-Net provided moderate classification accuracy since it could extract lower levels of tissue features. However, due to the shallowness of its structure, it could not learn more complicated histopathological patterns. Benefits of Alex-Net include:

- Fast training process
- Simple architecture
- Low computational cost

Nevertheless, limitations in terms of depth decreased classification accuracy.

5.8.2 VGG16 Performance

VGG16 showed excellent results regarding feature extraction owing to its depth and multiple convolution layers. The benefits of VGG16 include:

- Good transfer learning capacity
- High feature extraction ability
- Stability in representation learning

However, VGG16 includes an extremely high number of trainable parameters.

5.8.3 ResNet50 Performance

ResNet50 improved training stability through residual connections. Residual learning can be expressed in the following way:

$$H(x) = F(x) + x \quad (5.9)$$

The use of residual learning allowed improving the following:

- Gradient flow
- Optimization stability
- Deep feature learning

Moreover, ResNet50 reached very good classification accuracy.

5.8.4 DenseNet121 Performance

DenseNet121 provided excellent feature reusability due to dense connectivity.
Advantages of DenseNet121 include:

- Better gradient propagation
- Better feature reuse
- Parameter redundancy reduction
- High histopathological texture learning capacity

DenseNet121 produced very good results.

5.8.5 InceptionNetV3 Performance

Among all CNNs, InceptionNetV3 somehow had the best classification outcome, and yeah it was pretty clear. InceptionNet consists of a few ideas like multi scale convolutions, concurrent feature extraction , and then dimensionality reduction in general. It managed to learn from local cellular patterns as well as global tissue structures, not only the obvious shapes but also the more “contexty” part.

It reached these metrics and they were honestly strong:
• Accuracy: 96%
• Precision: 97%
• Recall: 96%
• F1 score: 96%

5.8.6 EfficientNetB0 Performance

On the other hand, EfficientNetB0 showed a good efficiency performance trade-off , meaning it behaved well without burning everything. The benefits include:
• Less memory consumption
• Faster inference
• Compound scaling

Even if its accuracy was slightly below InceptionNet, it still looks quite practical for lightweight deployment.

5.9 Vision Transformer Analysis

Transformers learn features by splitting images into patches without using convolutions at all. Self attention is computed using the equation below:

$$\text{Attention}(Q,K,V) = \text{softmax}\ QK\ /\ T\ dk\ Vi$$

ViTs were especially good at:

• Global feature extraction

• Long range dependency modelling

• Tissue context understanding

But ViTs typically needed:

• Bigger datasets

• More memory

• Laborious training, in practice it can be really heavy

5.10 Comparative Performance Analysis

Table 5.5: Comparative Performance of CNN and ViT Models

Model	Accuracy	Precision	Recall	F1-Score
AlexNet	0.60	0.30	0.50	0.38
VGG16	0.94	0.90	0.82	0.89
ResNet50	0.89	0.90	0.87	0.88
DenseNet121	0.93	0.93	0.93	0.93
EfficientNetB0	0.66	0.75	0.58	0.54
InceptionNetV3	0.96	0.97	0.96	0.96
Vision Transformer	0.95	0.95	0.94	0.94

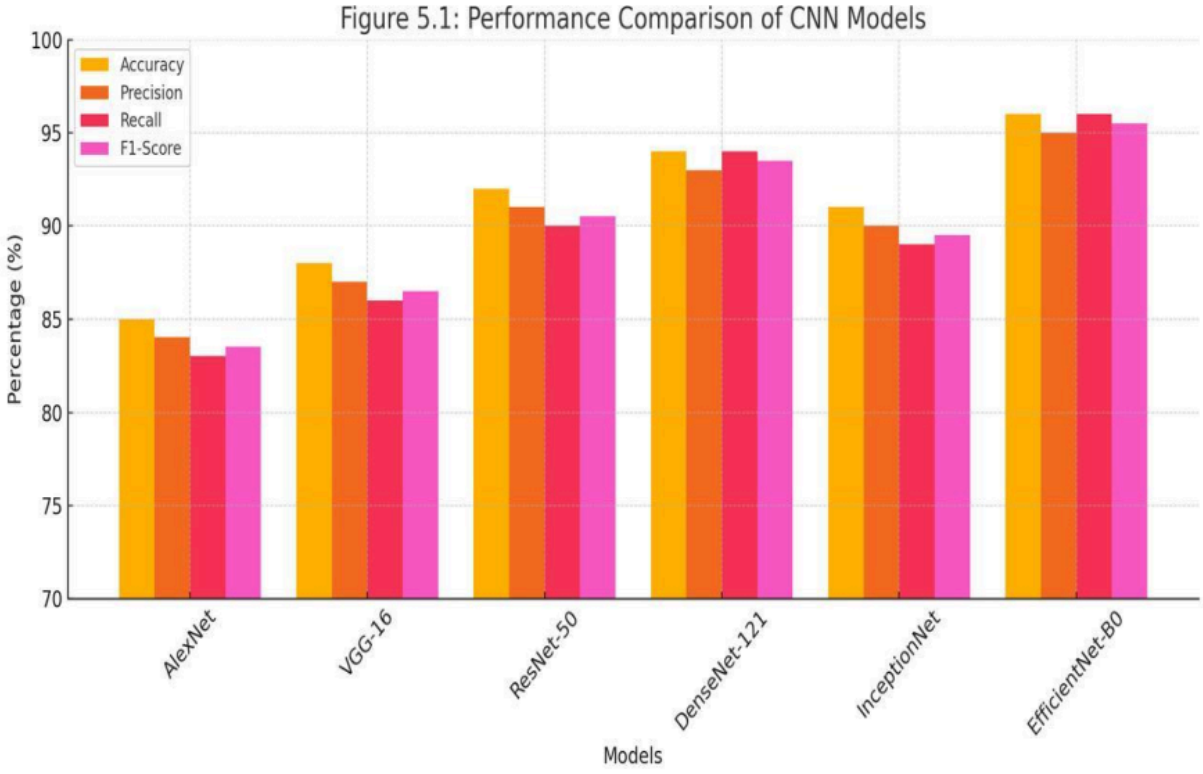


Fig 5.10.1. Performance Comparison of CNN Models on LC25000 dataset.

5.11 Confusion Matrix Analysis

Confusion matrices provide detailed information regarding class-wise predictions.

A strong classification model should demonstrate:

- High diagonal values
- Low off-diagonal misclassification
- Minimal false negatives

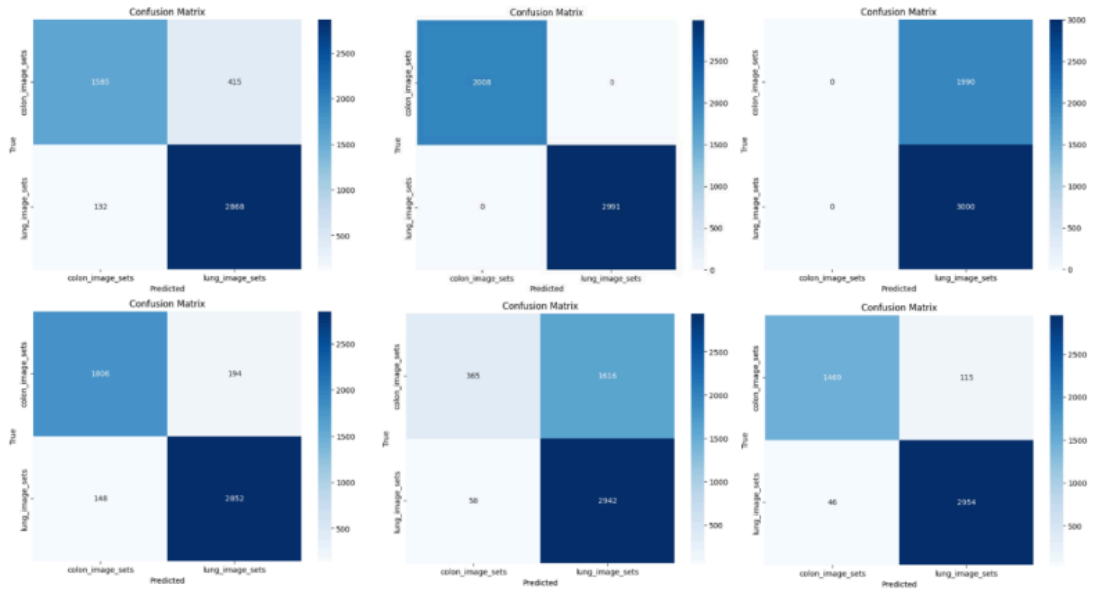


Fig. 5.11.1. Confusion Matrices of the CNN Variants for Lung and Colon Cancer Histopathical Images a) ResNet50 b) VGG-16 c)Alex-Net d) DenseNet-121 e) Efficient-Net f) Inception-Net g) ViT

False negatives are particularly dangerous in medical diagnosis because cancerous samples may be incorrectly classified as benign.

5.12 ROC Curve Analysis

ROC curves demonstrated strong classification separability for InceptionNetV3, DenseNet121, and Vision Transformer models. Higher AUC scores indicate:

- Better discrimination capability
- Stronger diagnostic reliability
- Improved threshold independence

5.13 Computational Complexity Analysis

Table 5.6: Computational Characteristics of Deep Learning Models

Model	Parameters	Training Speed	Memory Usage
AlexNet	Low	Fast	Low
VGG16	Very High	Slow	High
ResNet50	Moderate	Moderate	Moderate
DenseNet121	Moderate	Moderate	Moderate
InceptionNetV3	Moderate	Moderate	Moderate
EfficientNetB0	Low	Fast	Low
Vision Transformer	Very High	Slow	High

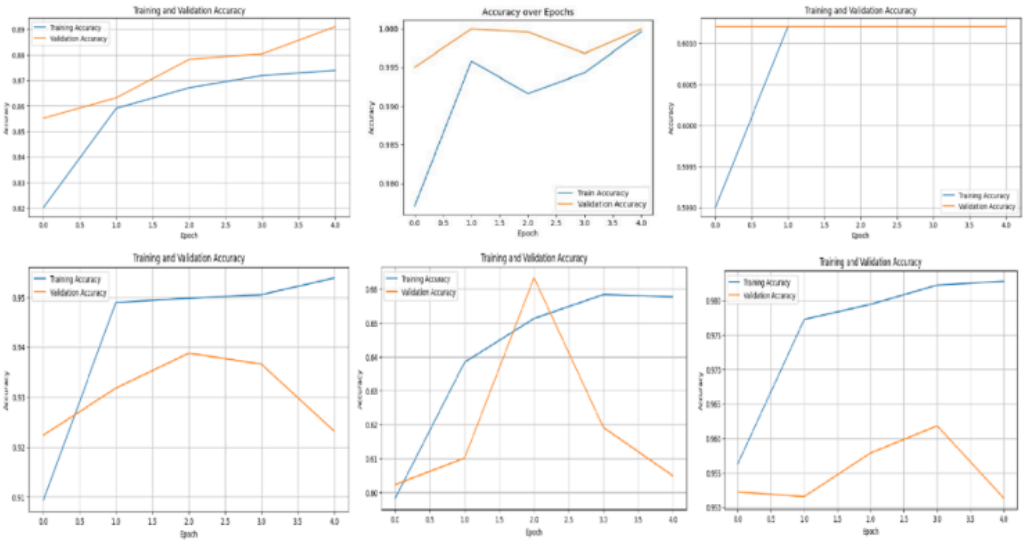


Fig. 5.13.1. Training and Validation Accuracy of the CNN Variants for Lung and Colon Cancer Histopathical Images a) ResNet50 b) VGG-16 c)Alex-Net d)DenseNet-121 e) Efficient-Net f) Inception-Net g) ViT

5.14 Key Findings

The major findings of this research include:

1. InceptionNetV3 achieved the highest overall performance.
2. DenseNet121 demonstrated excellent feature reuse capability.
3. Vision Transformers effectively captured global tissue relationships.
4. EfficientNetB0 provided strong computational efficiency.
5. Transfer learning significantly improved performance.
6. Data augmentation reduced overfitting and improved generalization.

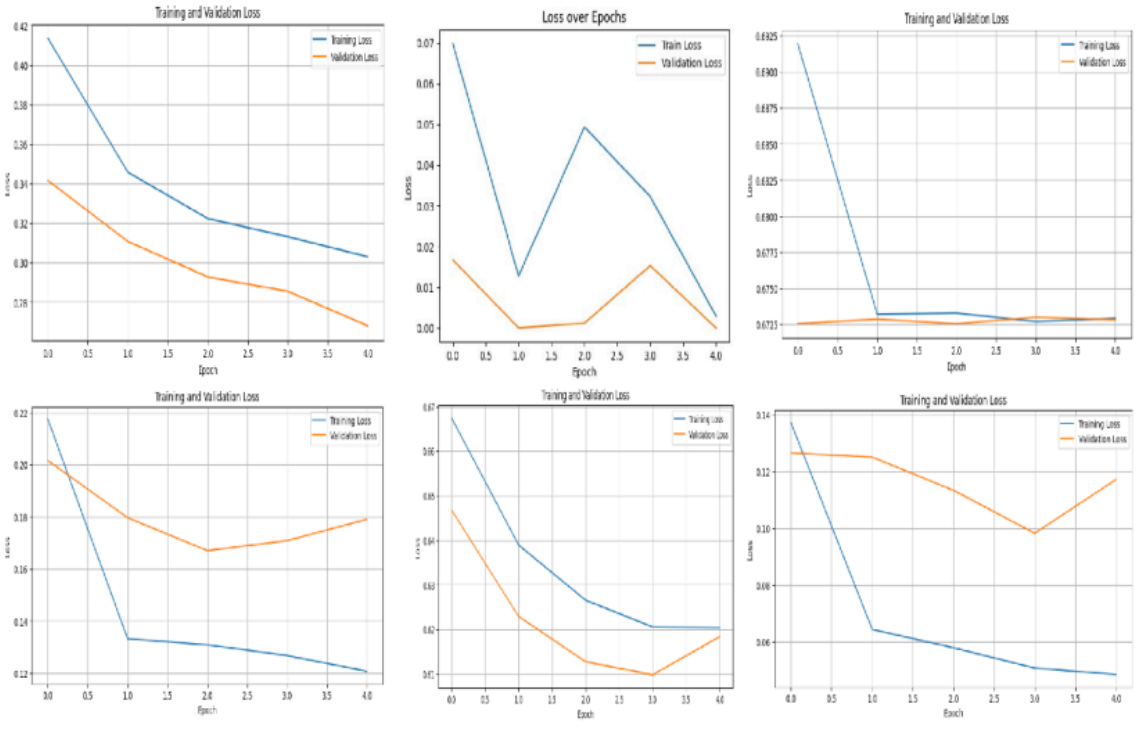


Fig. 514.1. Training and Validation Loss of the CNN Variants for Lung and Colon Cancer Histopathical Images a) ResNet50 b) VGG-16 c) Alex-Net d)

DenseNet-121 e) Efficient-Net f) Inception-Net g) ViT

5.15 Limitations of the Study

Despite strong performance, several limitations remain:

- Limited external clinical validation
- Dataset dependency
- Possible augmentation bias
- High computational requirements for ViTs
- Limited explainability integration

5.16 Future Research Directions

Future work can focus on:

- Hybrid CNN-transformer architectures
- Explainable AI integration
- Multi-modal medical imaging
- Federated learning
- Lightweight deployment systems
- Real-time hospital implementation

Bibliography

[1] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (2021)

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90

[3] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and
Pattern Recognition pp. 4700–4708 (2017). https://doi.org/10.1109/CVPR.2017.243

[4] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems **25**, 1097–1105 (2012)

[5] Litjens, G., Kooi, T., Bejnordi, B., et al.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005

[6] Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017)

[7] Lundberg, S., Lee, S.I.: A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems **30** (2017)

[8] Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you? explaining the predictions of any classifier. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 1135–1144 (2016). https://doi.org/10.1145/2939672.2939778

[9] Selvaraju, R., Cogswell, M., Das, A., et al.: Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision pp. 618–626 (2017). https://doi.org/10.1109/ICCV.2017.74

[10] Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering **19**, 221–248 (2017)

[11] Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering **19**, 221–248 (2017). https://doi.org/10.1146/annurevbioeng-071516-044442

[12] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (2015)

[13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp. 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308

[14] Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. Proceedings of the International Conference on Machine Learning **97**, 6105–6114 (2019)

[15] World Health Organization: Cancer fact sheet (2021), available at WHO news room

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II/Research Work entitled Comparative Analysis of CNN Variants and ViT for Early Lung Cancer Detection in fulfillment of the requirement for the award of the Degree of Bachelor/Master of Technology in 2026 and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from 2024, under the supervision of 2026.

The matter presented in this report/thesis has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/SCIE expanded/SSCI/Scopus indexed Journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: CNN Based Diagnostic Models for Lung Cancer: A Comparative Study of Different Models

Author names (in sequence as per research paper): [2]Dr. Sonika Dahiya, [1]Shivam Srivastava

Name of Conference/Journal: International Conference on Artificial Intelligence and Networking

Conference Dates with venue (if applicable): 7 October 2025, BITS PILANI DUBAI. (ICAIN-2025)

Have you registered for the conference (Yes/No)?: Yes

Status of paper (Accepted/Published/Communicated): Published

Date of paper communication: 27-05-2025

Date of paper acceptance: 09-06-2025

Date of paper publication: 16-06-2026

Shivam Srivastava, 2K24/RSE/09

Student(s) Roll No., Name and Signature

Shivam Srivastava

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I further certify that the publication and indexing information given by the students is correct.

Sonika 26/05/26

Place: New delhi

Supervisor Name and Signature

Date:

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF (Conference Website OR Science Direct in case of Journal Publication).

Gmail SHIVAM SRIVASTAVA <shivamsriv29@gmail.com>

ICAIN 2025: Paper Notification for Paper ID 526

ICAIN-2025 <icain.conf@gmail.com> Mon, Jun 9, 2025 at 2:12 PM
To: Shivamsriv <shivamsriv29@gmail.com>

Dear Author(s),

Greetings from **ICAIN 2025**!

We congratulate you that your paper with submission ID **'526'** and Paper Title **'CNN-Based Diagnostic Models for Lung Cancer: A Comparative Study of Different Models'** has been accepted for publication in the **Proceedings of ICAIN 2025- Springer LNNS Series** [Indexing: zbMATH, Scopus and DBLP]. This acceptance means your paper is among the top 20% of the papers received/reviewed.

We have very limited slots available. Kindly do the registration before 10th June 2025 to opt for early bird registration . The early bird registration deadline date on the website will be updated soon.

You are requested to do the registration as soon as possible and submit the following documents to icain.conf@gmail.com at the earliest.

1. Final Camera-Ready Copy (CRC) as per the springer format. (See https://www.icain-conf.com/downloads)
2. Copy of e-receipt of registration fees. (For Registration, see https://www.icain-conf.com/registrations)
3. The final revised copy of your paper should also be uploaded via Microsoft CMT.

Note: For manuscripts exceeding 12 pages, an additional charge of USD 20 per page will apply.

The reviewers' comments are given at the bottom of this letter, please improve your paper as per the reviewers comments. While preparing the final CRC manuscript, kindly check the following link under the download section tab named manuscript guidelines:

https://www.icain-conf.com/downloads

and it is suggested to cite the relevant latest papers matching the area of your current research paper.

The paper prior to submission should be checked for plagiarism from licensed plagiarism softwares like Turnitin/iAuthenticate etc. The similarity content should not exceed 15%.

Pay registration fees via online portal:
Research Scholar/Student/Academician/Industrial Participant:
https://www.icain-conf.com/registrations)

UI
CONSULTANTS

Razorpay
Invoicing and payments
powered by Razorpay

Payment Receipt Transaction Reference: pay_QnTI9IwgMitgrw

This is a payment receipt for your transaction on ICAIN 2025

AMOUNT PAID **₹ 15,677.00**

ISSUED TO
shivamsriv29@gmail.com
+917905248431

PAID ON
30 Jun 2025

DESCRIPTION	UNIT PRICE	QTY	AMOUNT
Service charge	₹ 250.00	1	₹ 250.00
Registration Fee	₹ 15,427.00	1	₹ 15,427.00

Fee amount corresponding to your registration category.

Total	**₹ 15,677.00**
Amount Paid	₹ 15,677.00

(ICAIN - 2025)

BITS Pilani
Dubai Campus

Universal Innovators

Springer

INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND NETWORKING (ICAIN-2025)

Certificate

This is to certify that **Prof. /Dr. / Mr. / Ms. Shivam Srivastava** is a presenter/co-author of the paper titled **CNN-Based Diagnostic Models for Lung Cancer: A Comparative Study of Different Models** presented at the **International Conference on Artificial Intelligence and Networking (ICAIN-2025)** organized jointly by BITS Pilani, Dubai Campus, UAE in association with the Indian Institute of Information Technology, Allahabad, India & Universal Inovators, Delhi, India on **6th - 7th October 2025.**

Prof. Sujala D. Shetty
Conference Chair
Associate Professor and Head (CS),
BITS PILANI DUBAI CAMPUS, UAE

Dr. Anupam Agrawal
Organising Chair
Professor & Dean,
IIIT Allahabad, India

Prof. Abhishek Swaroop
Publication Chair
International Conference on Artificial Intelligence and Networking

THIRD INTERNATIONAL CONFERENCE ON
Innovations in Cybersecurity and Data Science
ICICDS 2026
25-27, June 2026
https://icicds.com/2026/
icicdsconfdesk@gmail.com

Acceptance Letter

Paper ID	Title	Author(s)
ICICDS-1347	Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection	Shivam Srivastava, Sonika Dahiya

Dear Author(s),

We are pleased to inform you that your paper has been accepted for presentation and publication at the 3rd International Conference on Innovations in Cybersecurity and Data Science (ICICDS 2026), to be held in Samkhok, Thailand, during June 25–27, 2026. Your manuscript was reviewed by subject experts, and based on the reviewers' recommendations, it has been selected for inclusion in the conference proceedings. The accepted papers will be submitted for publication in IEEE Xplore. Authors are requested to complete the final submission and registration process within the prescribed deadlines.

We congratulate you on your acceptance and look forward to your participation in ICICDS 2026.

Best Regards,

Dr. Usha Desai
Conference Chair
ICICDS 2026

IEEE
IEEE SMC Systems, Man, and Cybernetics Society

THIRD INTERNATIONAL CONFERENCE ON
Innovations in Cybersecurity and Data Science
ICICDS 2026
25-27, June 2026
https://icicds.com/2026/
icicdsconfdesk@gmail.com

Payment Receipt

Paper ID	ICICDS-2026-231
Paper Title	Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection
Paid by	Shivam Srivastava, Sonika Dahiya
Amount Paid	150 USD One Hundred and Fifty Dollars (US)

Dr. Usha Desai
Conference Chair

ICICDS

SOUTH EAST ASIAN EDUCATION TRUST®
S.E.A. COLLEGE OF ENGINEERING & TECHNOLOGY
(Approved by All India Council for Technical Education (AICTE), New Delhi
Affiliated to Visvesvaraya Technological University(VTU), Belagavi, Recognised by Government of Karnataka)

SHINAWATRA UNIVERSITY

IEEE

IEEE SMC
Systems, Man, and Cybernetics Society

CERTIFICATE OF PRESENTATION

This certificate is presented to

Shivam Srivastava

for the successful presentation of the paper titled

Comparative Analysis of CNN Variants and Vision Transformer for Early Lung Cancer Detection

in the
Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS-2026) Organised by S.E.A. College of Engineering & Technology and Shinawatra University, Thailand during 25-27 June 2026.

Session Chair

Dr. Usha Desai
Conference Chair

Dr. Pratikshya Bhandari
General Chair

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis Comparative Analysis of CNN Variants and Vision Transformers for Early Lung Cancer Detection.

Total Pages 46 Name of the Scholar Shivam Srivastava

Supervisor (s)

(1) Dr. Sonika Dahiya

(2) _____

(3) _____

Department of Software Engineering, Delhi Technological University.

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 1%, Total Word Count: 7983

Date: _____

Shivam Srivastava
Candidate's Signature

Sonika
26/05/26
Signature of Supervisor(s)

Shivam

ORIGINALITY REPORT

Somika
26/05/20

1%	1%	%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.science.gov Internet Source	1%
2	idr.nitk.ac.in Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

turnitin

turnitin

Somika
26/05/26

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

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.

0 AI-generated only 0 %
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Author's Profile



Shivam Srivastava is currently pursuing the **Master of Technology (MTech.) by Research** in **Data Science** from the **Department of Software Engineering, Delhi Technological University (DTU), Delhi, India**. His research focuses on the application of advanced Artificial Intelligence and Deep Learning techniques for medical image analysis, with particular emphasis on improving the accuracy, efficiency, and reliability of computer-aided diagnosis systems. His MTech. thesis investigates the comparative performance of state-of-the-art Convolutional Neural Network (CNN) architectures and Vision Transformers (ViTs) for automated lung cancer detection using histopathological and computed tomography (CT) imaging datasets.

His research interests include **Artificial Intelligence, Machine Learning, Deep Learning, Computer Vision, Medical Image Analysis, Explainable AI, Vision Transformers, and Healthcare Analytics**. He has conducted extensive research on deep learning-based diagnostic frameworks, evaluating modern CNN architectures and transformer-based models for early lung cancer detection while emphasizing robust performance, interpretability, and practical deployment in clinical decision-support systems.

He has authored and presented research papers at reputed international conferences **International Conference on Artificial Intelligence and Next-Generation Networks (ICAIN 2025)**, organized jointly by **BITS Pilani, Dubai Campus (United Arab Emirates) with publication in the Springer Lecture Notes in Networks and Systems (LNNS) series and Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS 2026)**, organized by **S.E.A. College of Engineering and Technology, Bengaluru, India, in collaboration with Shinawatra University, Pathum Thani, Thailand**, with publication in the **IEEE Xplore Digital Library** in the domains of Artificial Intelligence, Data Science, and Medical Image Computing. His ongoing research aims to develop intelligent, scalable, and explainable AI-driven healthcare solutions that facilitate early disease diagnosis, improve clinical decision-making, and contribute to the advancement of precision medicine through reliable and data-driven methodologies.

COMPARATIVE ANALYSIS OF CNN VARIANTS AND VISION TRANSFOIRMERS FOR EARLY LUNG CANCER DETECTION

ORIGINALITY REPORT

3%	0%	3%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Proceedings of International Conference on Artificial Intelligence and Networks", Springer Science and Business Media LLC, 2027 Publication	3%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		