Dual-Stream Attention Fusion with Metaheuristic Feature Optimization for Interpretable Lung Cancer Classification in Medical Imaging

BY

Yousuf Rayhan Emon
ID: 242-25-028

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering [Font-14]

Supervised By

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Co-Supervised By

Dr. Abdus Sattar
Associate Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

AUGUST 2025

APPROVAL

This Project/Thesis titled **Dual-Stream Attention Fusion with Metaheuristic Feature Optimization for Interpretable Lung Cancer Classification in Medical Imaging**, submitted by **Yousuf Rayhan Emon**, ID No: **242-25-028** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13-09-2025.

BOARD OF EXAMINERS

Chairman

Dr. Sheak Rashed Haider Noori
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Mr. Nazibur Rahman
Head of IT Infrastructure
Networld Bangladesh *PLC*

DECLARATION

We hereby declare that this project has been done by us under the supervision of Dr. Sheak Rashed Haider Noori, **Professor & Head, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Dr. Sheak Rashed Haider Noori
Professor & Head
Department of CSE
Daffodil International University

Co-Supervised by:

Dr. Abdus Sattar
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Yousuf Rayhan Emon
ID: -242-25-028
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We are grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori**, Professor & Head, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Sheak Rashed Haider Noori, Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Detection and classification of lung cancer in chest computed tomography (CT) images is a pressing issue in medical imaging because it is challenging to differentiate a subtle morphologic distinction between benign and malignant lung nodules and further complicated by the fact that it is difficult to identify pathologic versus normal lung tissue. Conventional single stream deep learning methods do not inherently encode the complementary local and global characteristics that are required in the diagnosis of lung cancer. In this research work, a new dual-stream attention-based feature fusion system of semi-automatic classification of chest CT images on benign, malignant, and normal are proposed to detect lung cancer, overcoming the limitations of the traditional methods due to complementary feature extraction and intelligent fusion mechanisms. The methodology consists of a multiphase pipeline involved with dual-stream feature extractions by InceptionV3 (2 048-dimensional local features) and SE-ResNet18 (512-dimensional global features) frameworks, attention-based fusion mechanism with autonomy gating of adaptive feature weighting, metaheuristic feature selection through Recursive Feature Elimination that yields 98.05% dimensionality reduction (1,024 20 features), and classification optimization by LightGBM with Bayesian hyperparameter optimization. The data set contained 1,933 clinically confirmed CT scans of the chest of patients at the National Institute of Diseases of the Chest and Hospital, Dhaka, Bangladesh, which were extended to 3,240 balanced samples with three lung cancer diagnosis types. The validation F 1-scores of the individual feature extraction networks were 0.9875 each. The attention-based fusion system was better than concatenation (0.9875) and summation (0.9896) baselines with F 1 -score of validation of 0.9896. The dimensionality of 1,024 features was narrowed down to 20 features in feature selection with cross-validation F 1 -score of 1.0000. Test F 1 -score of 0.9695 and AUC-ROC of 0.9963 with LightGBM utilizing LightGBM classified significantly better than CatBoost and XGBoost baselines. Multi-method explainability analysis with SHAP, LIME and attention visualization validated model interpretability and clinical relevance to lung cancer diagnosis. The proposed methodology demonstrates superior performance in lung cancer classification while achieving substantial dimensionality reduction, providing interpretable predictions suitable for clinical decision support with potential for real-world medical deployment.

Keywords: Lung cancer classification, chest CT analysis, dual-stream architecture, attention-based fusion, feature selection, medical image analysis, explainable AI

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

Lung cancer is also one of the most lethal cancers across the globe, with millions of people dying each year due to this condition and posing a significant issue to global health [1,2]. It is associated with a high mortality rate that can be greatly attributed to late diagnosis as early-stage tumors can be asymptomatic or concealed by the presence of other related conditions of the lungs [3]. Timely and prompt diagnosis is hence necessary towards enhancing survival and timely treatment interventions. Computed tomography (CT) is the first imaging mode that is used to screen and diagnose lung cancer, and this is because it can give high-resolution volumetric data concerning lung nodules [4]. Nonetheless, CT scans interpretation is time-consuming, inter-observer variation exists, and diagnostic uncertainty is likely, which drives the growing necessity of automated computer-aided diagnostic (CAD) systems [5].

Deep learning (DL) has recently revolutionized the field of medical image analysis and has shown impressive results in lung cancer detection and classification problems [6–9]. CNNs, have demonstrated better capability to learn hierarchical feature representations (enabling distinguishing well-formed benign and malignant nodules) [10]. The InceptionV3, ResNet, and DenseNet architectures are now considered the standards of this field [11,12]. More sophisticated systems have added multi-stream or dual-input systems to record complementary data. Indicatively, the multi-stream CNNs and capsule networks have effectively synthesized local and global features, which models with single streams are unable to do, providing better accuracy [13–15]. Regardless of these improvements, single-network methods typically exhibit low generalizability and can be unable to reflect more heterogeneous differences in disparate datasets.

Another similar direction in medical AI has been the use of feature fusion methods that seek to combine features across layers, modalities, or models. Research in lung cancer detection has demonstrated that prediction accuracy can be enhanced by multi-scale and

attention-based fusion approaches, as opposed to naive concatenation [1618]. Furthermore, multi-modal fusion, that is, the radiological images and radiomic or clinical features, have been proven to increase diagnostic accuracy and interpretability [19,20]. However, most of the current fusion strategies are simplistic and fail to consider heterogeneous features in the most optimal way, which restricts their clinical use.

The curse of dimensionality is another big problem of CAD systems. Deep learning models can generate hundreds of thousands of features, many which might be redundant or irrelevant. Such high dimensional spaces without feature selection will elevate chances of overfitting and inefficiency in computations [21]. To alleviate this, feature selection schemes based on minimum redundancy maximum relevance (mRMR), recursive feature elimination (RFE), and genetic algorithm-based (or grey wolf optimizers) evolutionary algorithms have been broadly used [22,23, 24]. Such methods not only minimize computational load but also enhance performance of classification by preserving the most discriminative features. It has been demonstrated in recent studies that deep features are complementary to optimization-based selection methods to produce compact but highly informative representations of lung cancer [25].

At the classification level, the machine learning classifiers have been effective in the prediction of diagnostic features based on the features. Random Forest, Gradient Boosting, XGBoost and LightGBM are ensemble-based models that have performed very well in medical diagnosis, outperforming most single classifiers because of their capacity to model non-linear interactions and heterogeneity of features [26]. Gradient boosting models have been repeatedly demonstrated to be especially effective with medical imaging data, providing a compromise between accuracy, computational efficiency and interpretability [27,28]. Moreover, explainability models like SHAP and LIME are being gradually combined with these models to enhance transparency and trust in clinicians [29,30].

Although this has been made, there are several critical gaps in literature. First, most studies concentrate either on CNN-based end-to-end classification or the ensemble ML on handcrafted features, and not on combining the two benefits of these approaches. Second, feature fusion is not fully exploited, and many studies are reduced to concatenation instead of adaptive mechanisms. Third, its systematic combination into deep learning pipelines has

received limited attention, although the study of feature selection has been done separately. Lastly, despite the previous studies on gradient boosting classifiers with reference to medical imaging, there is still limited use of these models in optimized deep feature space in lung cancer detection.

To overcome such obstacles, this paper will focus on the application of a dual-stream adaptive transfer learning framework with the optimization of features to classify lung cancer. The framework also uses InceptionV3 and SE-ResNet18 as the complementary backbones in extracting local and global features. The mechanism of fusion is attention based on which the features are integrated in an adaptive manner such that the representations that are most informative are given more weight. Further refinement of the representation space is done using Recursive Feature Elimination (RFE) to reduce the high dimensional features to a small, discriminative set. Lastly, superior machine learning classifiers, LightGBM, XGBoost, and CatBoost, are compared with each other to see which one works best as a classification approach.

1.2 Motivation

The prevalence of lung cancer remains the number one cause of cancer-related deaths worldwide with the survival rates being pathetically low as a result of late diagnosis and inability to diagnose the disease correctly [1,2]. Although CT has dramatically enhanced the quality of nodule detection in the lungs, the manual decoding of scans is likely to have inter-reader inconsistency and misdiagnosis, especially in the difference between benign and malignant nodules [3,4]. The clinical burden explains the urgent necessity in the development of computer-aided diagnostic (CAD) systems that can help radiologists to deliver reliable and timely diagnostic results.

The deep learning models, especially the convolutional neural networks (CNNs) have been shown to be more effective in the detection and classification of lung cancer through the automatic learning of hierarchical features on the medical images [6,10,11]. Nonetheless, the single-stream architectures can typically only capture a small number of features of

image representation which results in nonoptimal results in the case of very heterogeneous lung cancer subtypes [1315]. Recent experiments revealed that multi-stream and hybrid models which combine complementary local and global characteristics are more generalized and robust [16–18]. Nevertheless, numerous current fusion strategies are based on naive concatenation or summation, and do not adaptively assign weight to feature contributions in real clinical setting [19].

One more urgent problem is the curse of dimensionality in deep learning-based pipelines of CAD. Thousands of descriptors in high-dimensional feature spaces have the risk of redundancy, overfitting, and inefficiency [21]. Dimensionality reduction algorithms, including mRMR, recursive feature elimination (RFE), and metaheuristic optimization algorithms, have been found to be helpful in dimensionality reduction and maintaining diagnostic capabilities [22-25]. However, there is little systematic implementation of such optimization schemes in the context of deep learning-based lung cancer CAD, especially to fused multi-stream features.

Lastly, although gradient boosting-based classifiers such as XGBoost, LightGBM, and CatBoost have consistently performed better than traditional classifiers in medical machine learning problems [26–28], their capabilities when coupled with optimized deep features in lung cancer diagnosis have not been studied comprehensively. Also of importance is the necessity of clinical trustworthiness: models that are poorly calibrated, or uninterpretable, can confuse clinicians [29,30]. Such a motivation, therefore, is to develop frameworks that can integrate multi-stream feature extraction, adaptive fusion, feature optimization, robust boosting classifiers and explainable AI into a unified pipeline. This approach will not only be more accurate, but more reliable, efficient, and transparent in the real-world diagnostic practice.

1.3 Research Objective

The objectives of this study are:

- To design a dual-stream deep learning framework using InceptionV3 and SE-ResNet18 for complementary local and global feature extraction from lung CT images.
- To develop an attention-based fusion mechanism that adaptively integrates multi-stream features for improved classification.
- To apply recursive feature elimination (RFE) to optimize high-dimensional feature spaces, reducing redundancy while preserving discriminative power.
- To benchmark advanced machine learning classifiers (LightGBM, XGBoost, CatBoost) on optimized features and identify the most effective strategy.
- To evaluate the framework's calibration, robustness, and interpretability using statistical testing, perturbation analysis, and explainable AI methods (SHAP, LIME, Grad-CAM).

1.4 Research Questions

This study addresses the following key research questions:

RQ1: Can dual-stream architectures with adaptive fusion improve lung cancer classification compared to single-stream models and simple fusion strategies?

RQ2: Does recursive feature elimination (RFE) enhance model efficiency and accuracy by reducing high-dimensional fused features to an optimized subset?

RQ3: Which gradient boosting classifier (LightGBM, XGBoost, CatBoost) offers the most reliable and interpretable performance on optimized deep features?

1.5 Expected Output

The expected outputs of this research include:

- A dual-stream feature extraction and attention-based fusion framework for lung cancer classification.
- An optimized feature subset derived from recursive feature elimination, enabling efficient and accurate classification.

- A comparative evaluation of gradient boosting classifiers on deep feature sets, highlighting the most effective approach.
- A robust validation framework incorporating calibration, statistical testing, and robustness under noise and perturbations.
- Integration of explainable AI methods (SHAP, LIME, Grad-CAM) to improve interpretability and clinical trust.
- A reliable, efficient, and interpretable CAD system that supports radiologists in early lung cancer detection.

1.6 Project Management and Finance

The research work doesn't get funds from any individuals or organization. However, table 1.1 below presents the project planning Gantt chart, which presents project's execution from march to august.

Table 1.1: Gantt Chart of the project management.

Task Name	March	April	May	June	July	August
Thesis planning						
Literature Review and Gap Analysis						
Dataset Collection						
Methodology Development						
Implementation and Result Collection						
Report Writing						

1.7 Report Layout

In Chapter 1, the study introduces the background, motivation, research objectives, and key research questions, establishing the overall direction of the work. Chapter 2 presents a comprehensive review of related literature, highlighting existing approaches and identifying research gaps. Chapter 3 details the proposed methodology, including dataset preparation, model design, and experimental framework. Chapter 4 reports and analyzes the experimental results, providing both quantitative evaluations and interpretative insights. Chapter 5 addresses the sustainability plan, societal and environmental implications, and ethical considerations associated with the study. Finally, Chapter 6 concludes the research, summarizes the key contributions, and outlines directions for future work.

CHAPTER 2
BACKGROUND

2.1 Preliminaries

There are several important concepts that this work uses as the basis of the suggested framework. Lung cancer is a malignant lung disease, typically divided into small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), which is further divided into adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Computed Tomography (CT) is the main imaging modality in detection based on its capability to provide high-resolution cross-sectional images of lung structures. Convolutional Neural Networks (CNNs) have been extensively used in medical images analysis to extract features automatically, i.e. learn hierarchical spatial patterns that facilitate the separation of malignant and benign nodules. To improve predictive performance, feature fusion schemes are employed to combine complementary data across models or layers and feature selection schemes are employed to downsample high dimensional representations by only maintaining the most discriminative attributes. During the classification phase, ensemble machine learning models (including Gradient Boosting classifiers (XGBoost, LightGBM, CatBoost) are a good fit, offering quality decision-making by capturing complicated, non-linear interactions between features. Lastly, the Explainable AI (XAI) approaches, such as SHAP, LIME, and Grad-CAM, are used to achieve interpretability and transparency to enable clinicians to learn more and trust model predictions. The combination of these terminologies forms the methodological basis that forms the basis of the current study.

2.2 Related works

Lung cancer is one of the most common causes of death in the world, and early diagnosis of the condition is one of the major problems in clinical diagnosis. Deep learning has taken the center stage as an effective technology to enhance accuracy in CT and histopathological analyses of images. The initial attempts were to integrate handmade components with extensive networks to enrich representations. Asuntha and Srinivasan [1] obtained texture and geometric features like HOG, LBP, wavelet transforms, and SIFT followed by feature

selection through Fuzzy Particle Swarm Optimization and classification through CNN. Their fuzzy-PSO CNN performed better than the conventional CNNs in terms of efficiency and indicated the benefit of hybrid schemes that combine engineered and learned characteristics. Recent research has led to the development of multi-stream and hybrid neural networks. Li et al. [2] suggested MSNet, which is a dual-stream model consisting of Swin Transformer and MLP-Mixer that combines both local and global features to classify pathology images of the lung adenocarcinoma. The model attained 96.55% accuracy with the reduction of complexity and highlights that balanced architecture may offer accuracy and efficiency. In the same vein, Ali and Ali [3] proposed a multi-input dual-stream capsule network which process raw and pre-processed histopathology images simultaneously. It evaluated on the LC25000 dataset yielded an accuracy of 99.58% and indicates that multi-input strategies are effective at capturing complementary features.

Multi-stream designs have been also expanded to integrate manual and automated capabilities. A three-stream network, introduced by Arumuga Maria Devi and Mebin Jose [4] combines handcrafted frequency-domain descriptors with residual and custom CNN features, and then fuses them using weights, giving a 98.2% accuracy on CT images. In a second improvement, Subash and Kalaivani [5] formed a two-step advanced classification model that blends modified U-Net segmentation with hybrid CNNs as detection and stage. There was an improved detection accuracy of their approach, and it tackled clinically significant staging problems in datasets such as LIDC-IDRI and NSCLC radiogenomics. Simultaneously, ensemble-based CNNs have also been investigated to be made more robust. Shah et al. [6] created a Deep Ensemble 2D CNN on the LUNA16 dataset that included three variants of CNN with variating kernel and pooling configurations. The ensemble attained a multicast accuracy of 95% which was higher than single- CNN based approaches to the baseline and it was shown that different architectures could be combined to increase the reliability of the classification. Taken together, these studies point to an apparent trend in the development of deep learning in the lung cancer classification: hybrid approaches combining handcrafted and CNN models have given way to more complex multi-stream, dual-stage, and ensemble models. The general conclusions are that a combination of complementary representations is necessary, to be applied; effective fusion

mechanisms must be developed; and computational complexity should be reduced without compromising the accuracy. This development is directly correlated to the current study, which uses dual-stream feature extraction with adaptive fusion as an enhancement to the performance of classification, its strength and ability to be interpreted in terms of lung cancer diagnosis.

It is observed that feature fusion has become an indispensable tool of medical image analysis, especially lung cancer classification, where combining multi-source features and multi-level features can significantly improve the accuracy of the diagnosis. Conventional single-model solutions usually lack the benefits in terms of complementary information across modalities or feature hierarchies, whereas fusion strategies can allow more discriminative representations. One of the earliest works by Kumar et al. [7] proposed a CNN based algorithm that was used to train spatially differentiable fusion maps based on PET-CT images. Their network of combining the sensitivity of PET to abnormal uptake with the localization capabilities of CT weighted features of each modality in an adaptive manner to spatial context. This method had a detection accuracy of 99.29, which demonstrates the power of supervised spatial fusion as compared to untrained input concatenation. Likewise, Mahum and Al-Salman proposed Lung-RetinaNet, a contextual multi-scale feature fusion architecture, which gained tumor localization in CT images and reached almost 99.8 percent accuracy, proving the usefulness of shallow and deep semantic features fusion. In addition to the solely deep learning pipelines, hybrid techniques, combining handcrafted features, have also been effective. Khan et al. [9] have suggested a contrast-based feature fusion method that further improved the quality of CT images through gamma correction with thereafter multi-feature extraction as well as canonical correlation-based fusion. Their ensemble classifier scored 99.4 percent accuracy, and this indicates that the combination of handcrafted descriptors and deep features is still a promising choice in early detection.

Further research directions have been extended by metaheuristic and optimization-driven feature fusion. A combination of ResNet, DenseNet and Inception-ResNet-v2 features were introduced by Alamgeer et al. [10] as a dung beetle optimization-guided dung beetle fusion model (DBOMDFF-LCC). In sequence modeling with LSTM the framework

enhanced the classification performance of lung cancer by fusing deep features and hyperparameter optimization. Equally, Ali et al. [11] explored the concept of decision level fusion, which compared SVM and AdaBoost classifiers on deep features that were generated using several CNN backbones. Their findings, using the LUNGx data, indicated that SVM with fused features reached 90.46% accuracy which was better than single-network classifiers. It has also been noted that multi-modal fusion is an appeal. Leng et al. [12] combined the radiomic and deep learning properties to differentiate between five lung cancer types. Their hybrid model, with CatBoost as a classifier, came up with a 94 percent accuracy and AUC of 0.97, surpassing those obtained with radiomics alone or deep features alone. This proves that fusion does not only increase accuracy, but also generalization because it minimizes overfitting. Similarly, Wang et al. [13] presented an attention-based multi-branch fusion network (M3-Net) with the help of endobronchial ultrasound (EBUS) pictures. Despite the low accuracy (0.76), the framework indicated the possibility of integrating the multi-features of difficult modalities.

New hierarchical fusion techniques are also developing. Taheri and Rahbar [14] integrated semantic hierarchical features of GoogLeNet with handcrafted descriptors with the Shapley values to select features and k-NN to classify. They used a better form of performance on the IQ-OTH/NCCD-Lung data using their approach. Lin and Yang [15] introduced the fusion based convolutional fuzzy neural network (F-CFNN) where several fusion strategies like global pooling fusion and network mapping fusion are compared. Their system was then optimized using the Taguchi method and reached about 99.98 percent accuracy on the SPIE-AAPM dataset, which highlights the promise of fuzzy fusion techniques. Combined, these studies point to the key contribution of feature fusion to lung cancer diagnostics. Solutions include multi-scale and spatial fusion, multi-modal, decision-level, and optimization-based solutions. One observation that appears to be constant is that fusion strategies are more accurate and robust in comparison to single-feature methods. Importantly, the choice of fusion mechanism—whether attention-based, correlation-driven, or optimization-guided—significantly influences performance. Such trends are consistent with the current study, which uses adaptive attention-based fusion of local and

global features to exploit complementary advantages and improve the consistency of the classification.

Medical machine learning heavily depends on feature selection and optimization, especially in lung cancer research, where datasets have frequently high-dimensional radiomic or deep learning features and small sample sizes. Effective reduction of features enhances the performance of classification, cost of computation, and generalizability.

Radiomics: Radiomics has now become a popular method of deriving quantitative measures of medical image. But as Ge and Zhang [16] pointed out feature extraction has been well studied but critical analysis of the feature selection strategy and predictive modeling has not been well developed. The analysis of CT-based radiomics of lung cancer conducted provided insights into the dangers of overfitting and uncertainty in correlation between feature selection and lack thereof where they do not conduct proper validation of their pipeline and only provides indicators of robustness. The conventional machine learning techniques often combine algorithms of optimization of feature subset selection. Maleki and others [17] have used genetic algorithm (GA) to optimize k-nearest neighbors (k-NN) to diagnose lung cancer. Their algorithm minimized the dimensionality and obtained 100% accuracy, which proves that evolutionary searching behavior can dramatically enhance the efficiency of the classifier. Equally, Lu et al. [18] suggested a GA, where a trace-based separability criterion was proposed, with a high diagnostic rate across various classifiers, which outperforms feature selection using F-score and correlation.

Methaheuristic algorithms are becoming popular with biomarker and gene selection. The Voting-Based Enhanced Binary Ebola Optimization Search Algorithm (VBEOSA) was proposed by Mohamed et al. [19] and was able to select ten hub genes linked to lung cancer based on the TCGA data. The analysis of their pathways showed that they have clinically significant signaling responses, which highlight the promise of optimization-based studies in precision oncology. Similarly, Chen and Dhahbi [20] used overlapping feature selection to differentiate between LUAD and squamous cell carcinoma (LUSC). Their methodology discovered familiar and unfamiliar biomarkers confirmed by external data indicating the

significance of feature selection to reveal biological processes. The selection of features also is helpful with deep learning pipelines. Toogacar et al. [21] extracted the features with CNNs (LeNet, AlexNet, VGG-16) and used minimum redundancy maximum relevance (mRMR) to select informative subsets. They reported their hybrid CNNmRMRkNN model to reach an accuracy of 99.51 percent, which demonstrates that feature selection can enhance already powerful deep learning models. Equally, Lanjewar et al. [22] adapted DenseNet201 to CT-based pulmonary cancer detection and used several feature selection techniques preceding ML classification. With the synergy of deep features and feature selection, their system had 100 percent accuracy in certain settings, with strong statistical vindication. Deep feature fusion in the direction of optimization has been studied. Alamgeer et al. [23] suggested a dung beetle optimization (DBO)-based feature selection in a deep feature fusion model, which enhances the ability of LSTM classifiers to detect lung cancer. This method illustrates the way methodology can improve the interpretability and the robustness of deep models.

Taken together, these studies indicate that feature selection is inseparable in the management of high dimensional data in the process of diagnosing lung cancer. The conventional statistical approach (e.g., mRMR), evolutionary algorithms (e.g., GA) and metaheuristics (e.g., VBEOSA, DBO) have been proven to be effective. In recent times hybrid strategies that combine deep feature extraction and optimization-based selection have reached state-of-the-art accuracy with meaningful biomarkers are identified. Such results are consistent with the current study, in which Recursive Feature Elimination is used to filter out thousands of deep features to a small, highly discriminative set, which leads to better efficiency and classification robustness.

Machine learning (ML) classifiers are now part of the state-of-art medical diagnosis, and they provide scalable, precise, and efficient remedies to such intricate diseases as lung cancer. ML approaches are associated with numerous benefits over traditional rule-based or statistical approaches, including their ability to learn patterns on large datasets, trade-offs between accuracy and interpretability, and support early detection and prediction of prognoses. Survival analysis and lung cancer classification has also been widely adopted using supervised learning. Lynch et al. [24] were researching several ML models, such as

linear regression, decision trees, gradient boosting machines (GBM), and support vector machines (SVM) on the SEER database. Their traditional combination was most predictive of survival as compared to the other models individually. This paper has shown how ensemble learning, especially gradient boosting, can be used to capture multi-faceted clinical variables in prediction of prognosis. Supportive evidence was presented by Chaturvedi et al. [25] of a review of ML- and DL-based CAD systems, with focus on how classifier selection can enhance detection performance. Comparative analysis has indicated that the precision of classifiers is determined by the imaging characteristics and the characteristics of a dataset. Singh and Gupta [26] compared seven classifiers, namely k-NN, SVM, random forest, and multilayer perceptrons (MLPs) with datasets of CT scans. Their findings showed that MLP was the most accurate (88.55 percent) among SVM and random forest, indicating that neural classifiers can be used to identify the slightest of textural dissimilarity in lung nodules in comparison to conventional models. On the same note, Abdullah et al. [27] trained classifiers using UCI lung cancer data and discovered that SVM was superior to CNN and k-NN with accuracy of 95.56%. These conflicting results reveal that the choice of classifier must be data modality and complexity specifically.

The term ensemble learning has come out as an alternative to single classifiers. Shanbhag et al. [28] integrated several base learners (SVM, logistic regression, MLP, decision trees, and k-NN) to form an ensemble classifier that can detect lung cancer using CT methods with 85 percent accuracy. More sophisticated ensemble models combine deep learning and boosting. ConvXGB and ConvCatBoost were suggested by Ansari et al. [29] as the hybrid CNN-based ensemble models that integrate the features extraction with the gradient boosting. ConvXGB demonstrated 98.26% accuracy and balanced precision-recall and demonstrated the importance of the CNN representations used together with boosting classifiers. Similarly, Jain et al. [30] suggested an ensemble system based on logistic regression, Gaussian NB, MLP and fuzzy logic-based feature selection on the Internet of Medical Things (IoMT). Their voting classifier was accurate by 98.5% which demonstrated the clinical utility of resource-efficient ensemble learning in lung cancer detection. What these studies depict are several important trends. First, ensemble learning is always more successful than the individual classifiers, especially in the case of integrating the

representational capacity of deep learning and the stability of boosting algorithms. Second, the selection of a classifier depends on context: MLPs and CNNs are better at image-based classification, and SVM and gradient boosting are robust against structured clinical data. Third, it is further efficient and generalizable with integration with intelligent feature selection (e.g., fuzzy logic, k-means optimization).

Table 2.1: Research Matrix Table for different research studies conducted through literature review.

Author (Year)	**Model / Method**	**Reported Accuracy**	**Main Contributions**
Asuntha & Srinivasan [1]	FPSO-CNN with handcrafted features (HOG, LBP, SIFT, Wavelets)	~94%	Hybrid deep learning with fuzzy-PSO for optimized feature selection; applied to CT and hospital dataset.
Li et al. [2]	MSNet (dual-stream Swin Transformer + MLP-Mixer)	96.55%	Efficient fusion of local and global features for lung adenocarcinoma pathology; reduced complexity.
Arumuga Maria Devi & Jose [4]	Three-stream hybrid network (handcrafted + residual CNN + custom CNN)	98.2%	Combined frequency-domain handcrafted and deep features; weighted fusion for robust CT classification.
Subash & Kalaivani [5]	Dual-stage CNN (modified U-Net + hybrid Xception-CNN)	>95% (various datasets)	Two-stage framework for detection and staging; integrated segmentation and classification.
Shah et al. [6]	Deep Ensemble 2D CNN (3 CNN variants)	95%	Ensemble of CNNs with varied kernels/pooling; applied to LUNA16 dataset; improved reliability.

Alamgeer et al. [10]	DBOMDFF-LCC (Dung Beetle Optimization + Deep Feature Fusion + LSTM)	~98%	Optimization-driven deep feature fusion with metaheuristic parameter tuning.
Ali et al. [11]	Decision-level Fusion (SVM, AdaBoost + deep features)	90.46%	Compared multiple classifiers across deep features; SVM proved most robust for LUNGx dataset.
Singh & Gupta [26]	Comparative ML Classifiers (kNN, SVM, RF, MLP, etc.)	88.55% (MLP best)	Benchmarked multiple ML classifiers on 15,750 CT images; MLP outperformed SVM and RF.
Ansari et al. [29]	ConvXGB and ConvCatBoost (CNN + Gradient Boosting)	98.26% (ConvXGB)	Hybrid ensemble combining CNN feature extraction with boosting classifiers; balanced precision-recall.

2.3 Scope of the Problem

This study is aimed at designing a sophisticated computer-aided diagnostic (CAD) system to diagnose and classify lung cancer using computed tomographic (CT) scans. Its main limitation will be to overcome the methodological limitations of existing studies by combining dual stream deep learning architectures, adaptive feature fusion, feature optimization, and gradient boosting classifiers into a system. In particular, the authors examine InceptionV3 and SE-ResNet18 as a pair of backbones to learn both a local textural and global context of lung nodules. A fusion mechanism built on attention is adopted to adaptively fuse multi-stream features, and recursive feature elimination (RFE) is used to minimize redundancy and to optimize high-dimensional feature representation.

In addition to feature engineering, the research standards developed ensemble classifiers including LightGBM, XGBoost, and CatBoost on optimized features to measure their relative performance in classifying lung cancer. The assessment is a concept that goes past the conventional accuracy measures in the context of evaluating the model, including the calibration tests, perturbation-based robustness tests and statistical significance tests to guarantee how reliable the model is to be applied in clinical settings. In addition, the framework has explainable AI (XAI) algorithms: SHAP, LIME, and Grad-CAM, which can offer transparency so that clinicians can understand and rely on the decisions made by the system.

This study also covers the evaluation of the sustainability, social and ethical aspects of implementing AI-based diagnostic tools in the healthcare environment. The problems of data imbalance, data generalizability of models in heterogeneous populations, computational efficiency in the face of actual implementation, and ethical issues regarding AI-assisted decision-making are seriously discussed. With the research placed in this context of the wider topic, the study will not only add to the development of methods in medical AI but also to the responsible implementation of intelligent diagnostic tools in the clinical room.

2.4 Gap Analysis

The review of existing studies highlights significant progress in lung cancer detection and classification through deep learning and machine learning methods; however, several critical gaps remain unresolved.

1. Single-stream feature extraction dominates most studies: While CNN architectures such as Inception, ResNet, and DenseNet have achieved promising results, many works rely on a single backbone network, limiting the ability to capture complementary local and global feature representations [1–6]. Dual-stream and hybrid architectures are still underutilized in clinical imaging.
2. Fusion strategies remain simplistic and underexplored: A number of studies have applied feature concatenation or summation for combining multiple feature sets [7–

9]. Although attention-based and optimization-driven fusion have shown improvements in other domains, their application in lung cancer classification remains limited, reducing the full potential of multi-stream approaches.

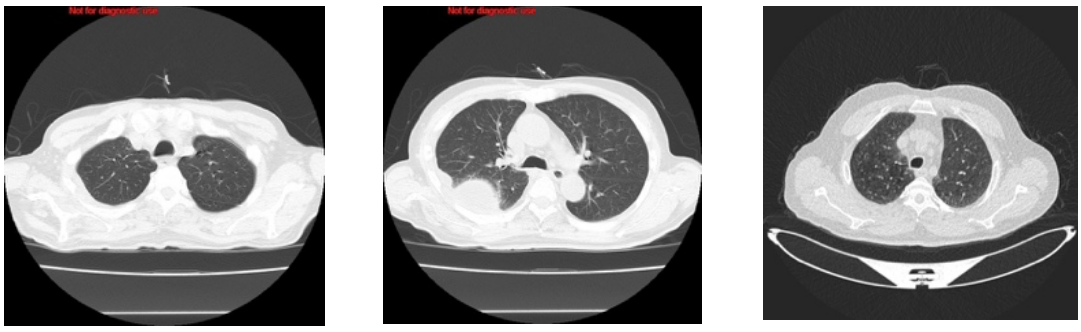
3. High-dimensional feature redundancy persists: Deep learning models generate thousands of features, yet most studies do not systematically apply feature selection techniques to optimize these high-dimensional representations. Although methods such as mRMR, genetic algorithms, and metaheuristics have been explored [16–20], their integration into deep fusion pipelines for lung cancer remains scarce, leading to risks of overfitting and inefficiency.
4. Classifiers are not fully optimized for deep feature spaces: Ensemble classifiers such as Random Forest, SVM, and boosting models have been applied in prior studies [24–27], but few works benchmarked advanced gradient boosting frameworks (e.g., LightGBM, XGBoost, CatBoost) on optimized deep features. This represents a gap in balancing classification accuracy, computational efficiency, and interpretability.
5. Reliability and interpretability are insufficiently addressed: Most existing studies emphasize accuracy but pay limited attention to model calibration, robustness under noise, or explainability. While SHAP, LIME, and Grad-CAM have been used in isolated cases [28–30], systematic integration of explainable AI for clinical trust and transparency remains underdeveloped.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Data Collection Procedure

The dataset of this study is composed of the images of the CT scan of the chest on the National Institute of Diseases of the Chest and Hospital (NIDCH), Dhaka, Bangladesh. It was curated to favor medical imaging research, and it is centered on classifying and early detecting lung-related diseases using artificial intelligence and deep-learning methods. The dataset is categorized into three diagnostic categories (malignant 1,050 images), normal (416 images), and benign (467 images).

All the images were thoroughly labelled and were clinically verified by the skilled radiologists of the Department of Radiology and Imaging, NIDCH. The annotation has been done under stringent medical guidelines to guarantee the accuracy of diagnosis and retain the data reliability in academic and scientific applications. This dataset offers a useful basis in developing, training and testing AI based models to be used to diagnose thoracic diseases in an automated manner. It has potential applications to clinical decision support systems beyond lung cancer detection including full-scale applications to clinical decision support in healthcare settings to enhance diagnostic efficiency and minimize inter-observer variability. Figure 3.1 shows the samples of each classes of the image dataset.



Benign **Malignant** **Normal**

Figure 3.1: Sample images of three classes from the dataset

3.2 Image Pre-processing and Augmentation

Prior to model training, a systematic pre-processing pipeline was applied to ensure consistency, reduce noise, and improve the generalization ability of the classifiers. All CT images were first resized to 299 × 299 pixels to match the input requirements of the InceptionV3 architecture. The images were preserved in RGB format (3 channels) instead of grayscale to retain richer pixel-level information. After resizing, each image was normalized using the ImageNet standard (mean μ = [0.485, 0.456, 0.406] and standard deviation σ = [0.229, 0.224, 0.225]) to align with the pre-trained backbone networks. Finally, the data were converted into PyTorch tensors with the shape [batch_size, 3, 299, 299], suitable for deep learning model input.

Data augmentation methods were used to solve the issue of imbalance between classes and enhance the robustness of the model. These were random rotations (±15), horizontal and vertical flipping, brightness (±20 percent) and small zoom/shift transformations. These stochastic augmentations produced various variations of the minority classes, in effect augmenting the benign and normal groups to comparable numbers as the malignant group. Consequently, each of the three categories had an equal number of 1,080 images, and the total number of CT scans was 3,240 (see figure 3.2). This dataset was subsequently stratified at patient level to avoid data leakage, where 70 percent was used to train (2, 256 images), 15 percent to validate (489 images), and 15 percent to test (495 images). Each split-maintained balance in classes to be evaluated without bias. There was also three-fold cross-validation of the training subset to give a better indication of the model generalization.

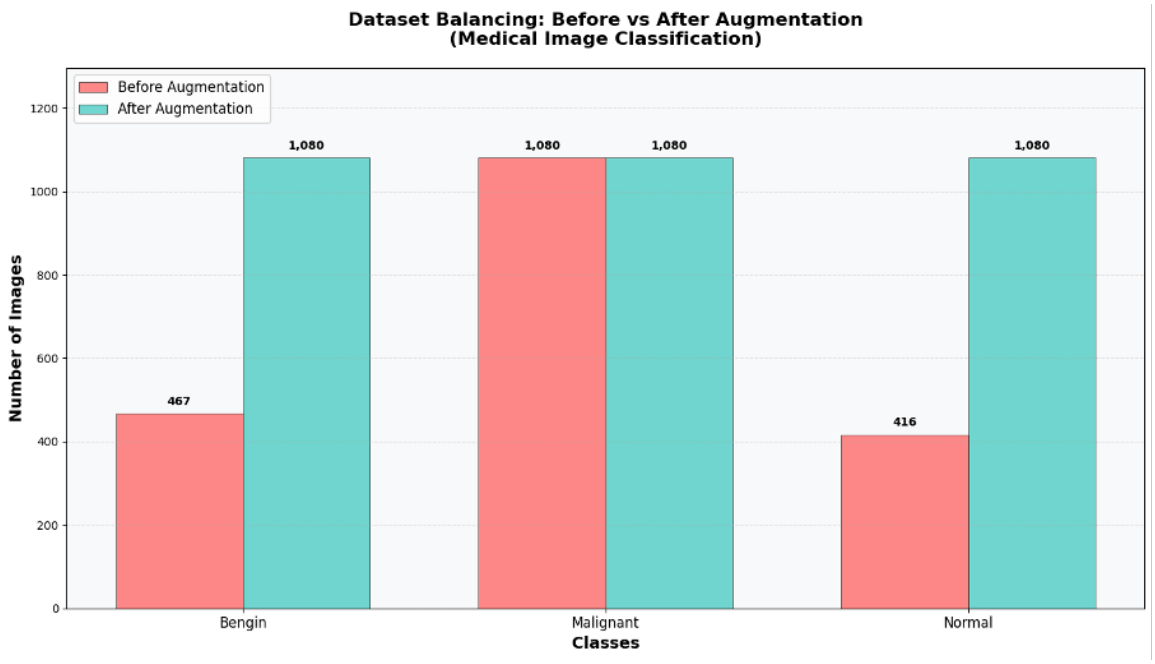


Figure 3.2: Dataset Balancing Before and After Augmentation for Lung Cancer CT Images

3.3 Proposed Methodology

The figure 3.3 depicts the general methodological flow of the proposed lung cancer-classification framework. The input here is the original CT dataset, and this is subjected to preprocessing measures such as image normalization, augmentation and balancing the classes to guarantee uniformity and better generalization. After preprocessing, there is dual-stream feature extraction with two complementary backbones InceptionV3 to learn local multi-scale features and SE-ResNet18 to learn global structural features in conjunction with squeeze-and-excitation attention. The features obtained are combined and are mathematically modeled as local feature vectors (FL) and the global feature vectors (FG).

Then, features are combined using an attention-based mechanism of feature fusion, which is an adaptive operator emphasizing the most informative features, and then features are analyzed through a process of exploratory feature analysis and metaheuristic feature selection to reduce redundancy and maximize high-dimensional features. The polished

features are then selected to the multi-model classification phase where sophisticated ensemble classifiers (LightGBM, XGBoost and CatBoost) are used with Bayesian optimization of maximizing predictive performance.

The classification results are also confirmed by bootstrap confidence intervals as a method of statistical robustness, and explainable artificial intelligence (XAI) techniques (SHAP, LIME) are also factored in to improve interpretability and clinical confidence. Lastly, the framework delivers detailed performance measurements to determine accuracy, calibration, robustness, and reliability. In general, the flowchart illustrates a methodical combination of preprocessing, dual-stream feature extraction, adaptive fusion, feature optimization, and multi-model classification, which is the basis of the methodological approach of the study.

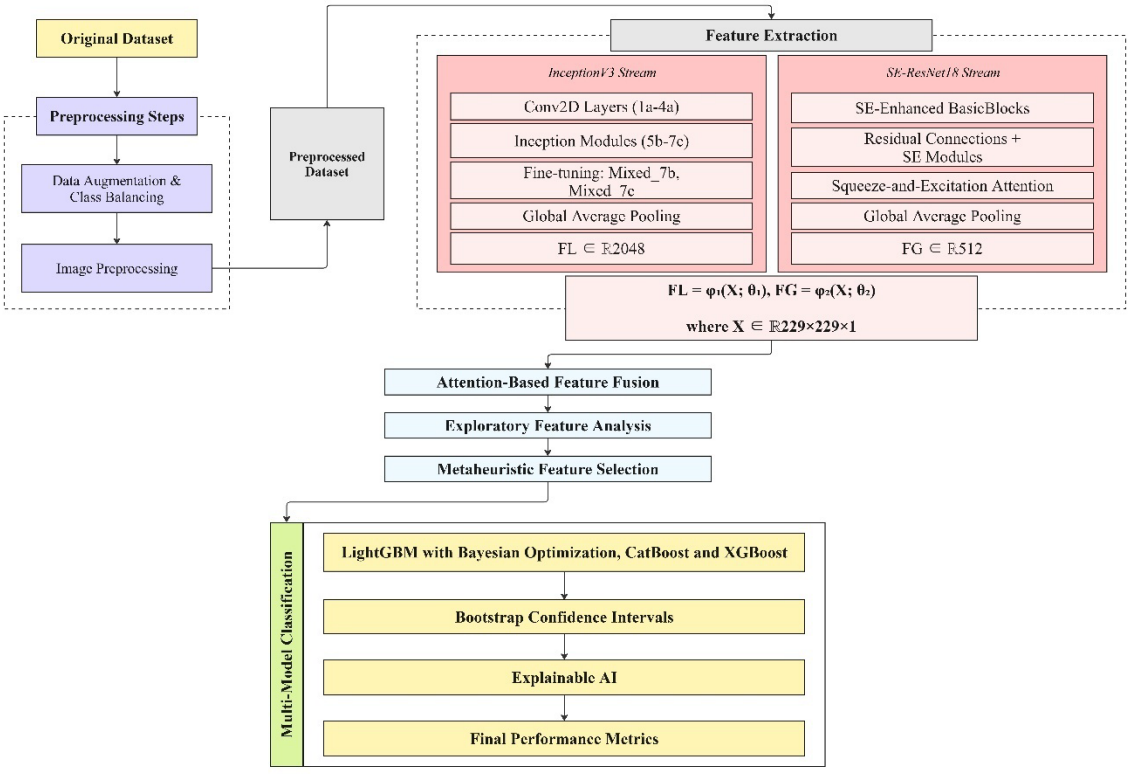


Figure 3.3: Methodological Framework of the Proposed Study

3.3.1 Dual-stream feature extraction architecture

The suggested methodology uses a two-stream convolutional neural network (CNN) design to learn rich and complementary representations of chest CT images as shown in Figure 3.2. This parallel structure combines the unique representational advantages of InceptionV3 and SE-ResNet18, allowing to capture both multi-scale spatial representation and channel-wise contextual representation. Figure 3.4 shows the architectural design.

Stream 1: InceptionV3 Feature Extraction

The former uses the InceptionV3 backbone which is optimized to extract features on a large scale because of parallel convolutional operations of different kernel sizes. Input CT images are processed by the network in:

- Initial convolutional layers (*Conv2d_1a_3x3 through Conv2d_4a_3x3*) for low-level feature extraction.
- Inception modules (*Mixed_5b through Mixed_7c*) to capture multi-scale spatial information.
- Fine-tuning of the last two blocks (*Mixed_7b and Mixed_7c*) for domain-specific adaptation.
- A global average pooling (GAP) layer to reduce spatial dimensions.

The resulting output is a feature vector:

$$F_L \in \mathbb{R}^{\{2048\}}$$

Stream 2: SE-ResNet18 Feature Extraction

The second stream employs SE-ResNet18, which integrates Squeeze-and-Excitation (SE) blocks within residual connections to model channel-wise feature dependencies. The architecture includes:

- An initial convolutional layer with batch normalization and ReLU activation.
- Four residual blocks with embedded SE modules for adaptive channel re-weighting.
- SE blocks applying global average pooling, fully connected transformations, and sigmoid gating.
- A final global average pooling layer.

The resulting output is a feature vector:

$$F_G \in \mathbb{R}^{512}$$

The dual-stream feature extraction process can be formulated as:

$$F_L = \varphi_{InceptionV3}(X;\ \theta_1)$$

$$F_G = \varphi_{SE-ResNet18}(X;\ \theta_2)$$

where:

- $X \in \mathbb{R}^{H \times W \times C}$ is the input CT image of size 299 × 299 × 3 (RGB).
- $\varphi_{InceptionV3}$ and $\varphi_{SE-ResNet18}$ denote the feature extraction functions.
- θ_1 *and* θ_2 represent the learnable parameters of each network.
- $F_L \in \mathbb{R}^{2048}$ *and* $F_G \in \mathbb{R}^{512}$ *are the respective extracted feature vectors.*

SE Block Mathematical Representation

The Squeeze-and-Excitation (SE) mechanism is defined as:

$$z = GAP(u) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_{i,j}$$

$$s = \sigma(W_2 . \delta\ (W_1 . z)$$

$$SE(u) = s \odot u$$

where:

- GAP denotes global average pooling,
- δ represents ReLU activation,
- σ is the sigmoid function,
- W1 and W2 are the fully connected layer weights,
- ⊙ denotes element-wise multiplication.

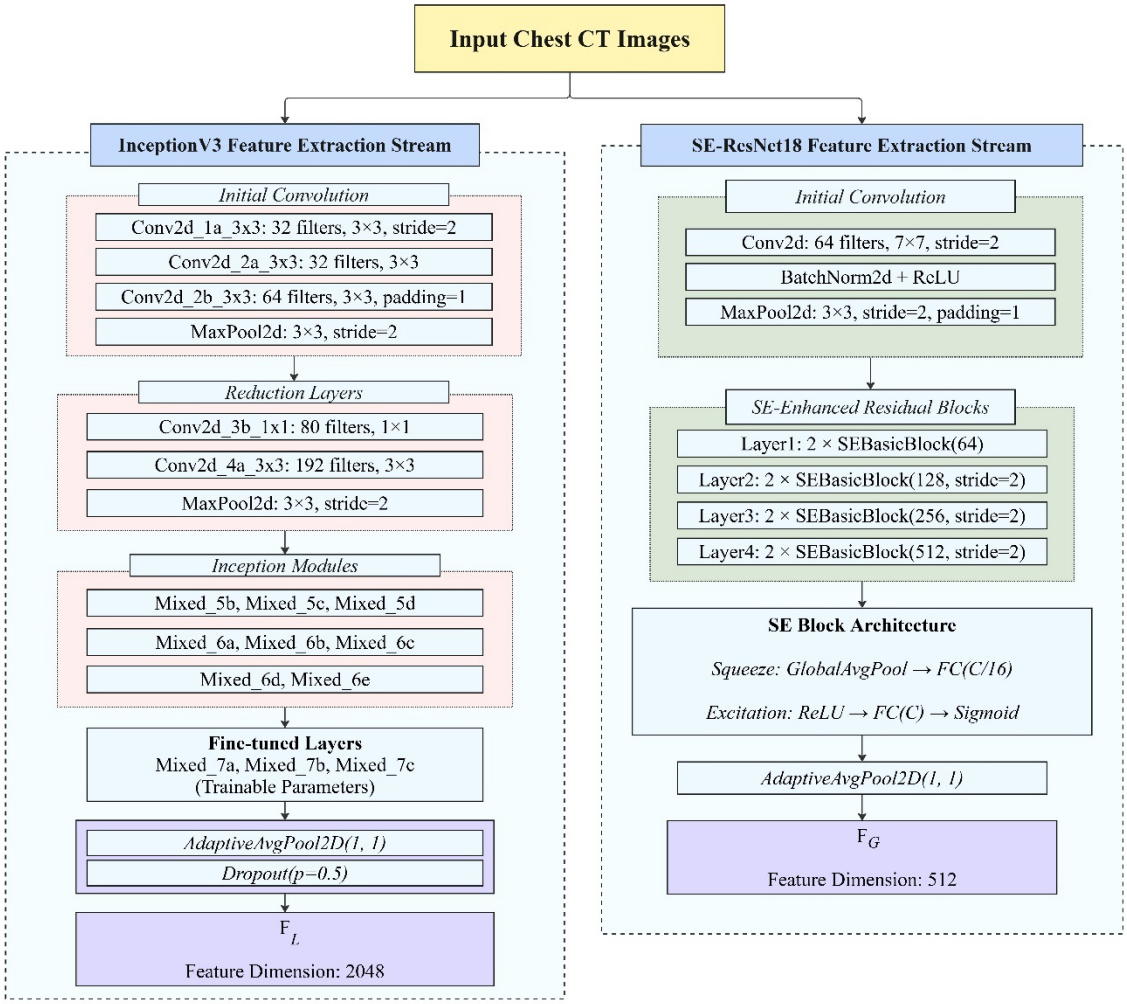


Figure 3.4: Dual-stream feature extraction architecture for grayscale CT images with complementary convolutional neural networks

3.3.2 Attention-based Feature Fusion Mechanism

Attention-based feature fusion mechanism aims to combine the complementary representations of features of the InceptionV3 and SE-ResNet18 streams. Figure 3.5 depicts that the fusion process is structured into six consecutive steps to allow features to be weighted dynamically according to learned attention coefficients. The design will make sure that the model selectively highlights the most informative elements of each stream, and it will inhibit the redundant information.

Stage 1: Input Feature Representations
The mechanism receives two feature vectors:

$$F_L \in \mathbb{R}^{2048} \text{ (InceptionV3 features)}, F_G \in \mathbb{R}^{512} \quad (SE - ResNet18\ features)$$

Stage 2: Dimension Alignment and Projection
Both feature vectors are projected into a shared 1024-dimensional space using linear transformations with batch normalization and ReLU activation:

$$F_L' = ReLU\ (BN(W_L F_L + b_L$$

$$F_G' = ReLU\ (BN(W_G F_G + b_G$$

Where $W_L \in \mathbb{R}^{1024 \times 512}, and\ both\ F_L', F_G' \in \mathbb{R}^{1024}$

Stage 3: Joint Feature Representation

The aligned vectors are concatenated to form a joint feature representation:

$$F_{concat} = [F_L', F_G'] \in \mathbb{R}^{2048}$$

$$\alpha_L = \sigma(W_{attn,L} F_{concat} + b_{attn,L})$$

$$\alpha_G = \sigma(W_{attn,G} F_{concat} + b_{attn,G})$$

where σ is sigmoid activation and $\alpha_L, \alpha_G \in [0,1]^{1024}$

Stage 5: Element-wise Weighted Fusion

The fused representation is computed using Hadamard products with the learned attention weights:

$$F_{fused} = \alpha_L \odot F'_L + \alpha_G \odot F'_G$$

Stage 6: Classification Head

The fused vector is passed through a dropout-regularized classification network:

$$h_1 = ReLU(Dropout(F_{fused}, p = 0.3))$$

$$h_2 = W_1 h_1 + b_1, \qquad W_1 \in \mathbb{R}^{512 \times 1024}$$

$$\hat{y} = Softmax\,(W_2 . Dropout\,(h_2, p = 0.3) + b_2)$$

where $W_2 \in \mathbb{R}^{3 \times 512}$ and $\hat{y} \in \mathbb{R}^3$denotes the final class prediction (Benign, Malignant, or Normal).

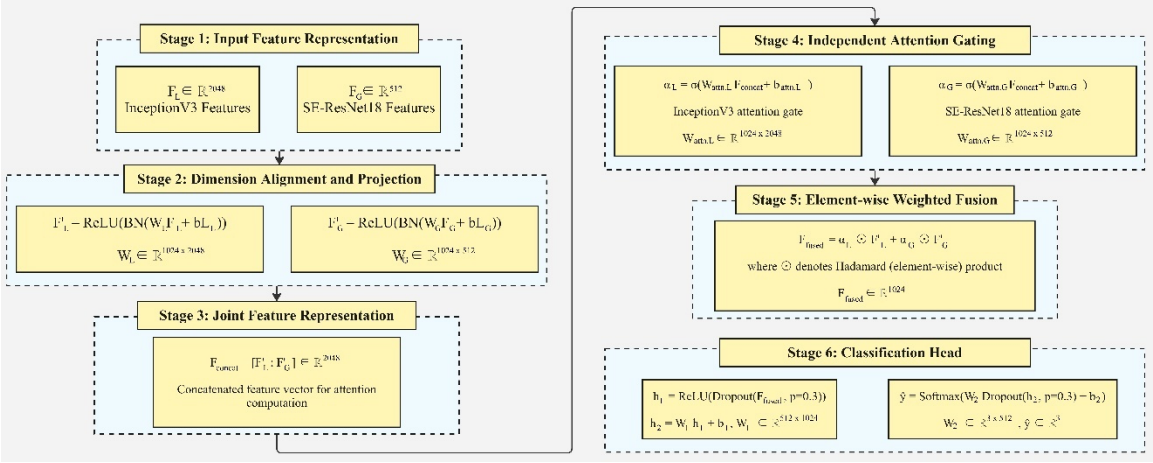


Figure 3.5: Mathematical formulation and architectural design of the attention-based feature fusion mechanism

The figure shows the six-stage pipeline of the attention-based fusion mechanism. InceptionV3 and SE-ResNet18 feature extractors are projected to a common 1024-dimensional space and concatenated together into a joint representation. Weighting

coefficients are then produced by independent attention gates, and they serve to weight adaptively informative elements along any stream. The resulting weighted fusion yields a single feature vector which is then input into a regularized classification head to give final prediction. The design is complementary, and information is wholly used with the benefit of being robust and generalized.

3.3.3 feature selection framework

The feature selection model employs metaheuristic optimization to trim the dimensions of the features, keeping the discrimination strength intact as shown in Figure 3.6. This stages process assures that overlap, noisy or irrelevant features are removed prior to classification, which also boosts efficiency and accuracy.

Stage 1: Binary Grey Wolf Optimizer (BGWO)
A Binary Grey Wolf Optimizer is employed for feature subset optimization:

- Population Initialization: A population of $N_{wolves} = 15$ binary masks is initialized with selection probability p∈[0.2,0.5].
- Fitness Evaluation: Each solution is evaluated using 3-fold stratified cross-validation, with the fitness function defined as:

$$f(x) = F1_{CV}(x) - \lambda.(\frac{||x||_1}{n})$$

where λ=0.001 penalizes large subsets.

- Wolf Hierarchy Update: Best-performing solutions are assigned to α, β, and δ wolves.
- BGWO Update: Positions are updated using Grey Wolf Optimizer equations with a sigmoid transfer for binary mapping:

$$A_k = 2ar_1 - a, C_k = 2r_2$$

$$X_k = X_{leader,k} - A_k.|X_{leader,k} - X(t)|$$

- Convergence: The optimizer runs for 30 iterations or stops early after 8 stagnant iterations.

Stage 2: Union Stabilization

The final selected feature subset is stabilized by computing the union of the top three solutions:

$$X_{union} = x_\alpha \cup x_\beta \cup x_\delta$$

Stage 3: Recursive Feature Elimination

An RFE stage with LightBGM, CATBoost and XGBoost estimator provides additional validation by stepwise elimination, ensuring robustness of the final selected subset.

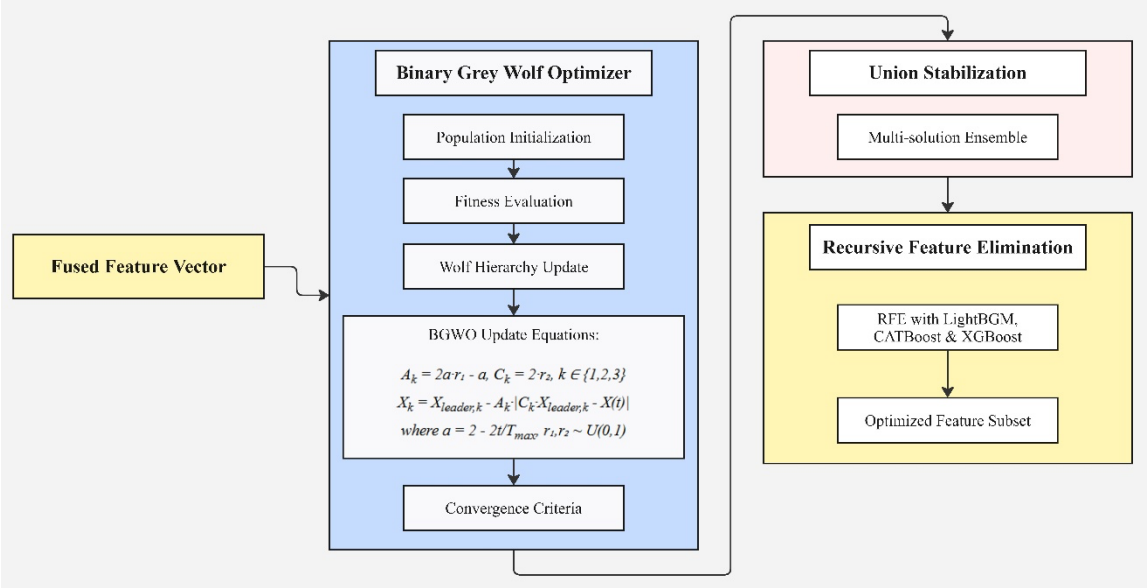


Figure 3.6: Metaheuristic feature selection framework using Binary Grey Wolf Optimizer with multi-stage filtering

The figure shows the multi-stage pipeline of feature selection. Optimal subset of features is chosen by the Binary Grey Wolf Optimizer (BGWO) by performing an iterative search and fitness test on a large set of features. The process is then stabilized by aggregation of the best performing solutions through union, and an optional RFE step is provided to refine further.

3.3.4 Comprehensive evaluation protocol

The overall assessment plan provides methodological rigor based on a five stage validation system to evaluate model performance, prediction reliability, operational robustness and interpretability to implement models in practice as a clinical tool.

Stage 1: Multi-Classifier Optimization and Comparative Analysis The optimized 97-dimensional feature space undergoes systematic evaluation across heterogeneous classifier architectures. LightGBM is the main classifier and uses Gaussian Process-based Bayesian optimization via the scikit-optimize package, with 50 iterative evaluations in the 8-dimensional hyperparameter search space, which includes tree structure parameters (num leaves, max depth), learning parameter (learning rate), regularization coefficients (lambda l1, lambda l2), and sampling parameters (feature fraction, bagging fraction). Baseline architectures Compared architecture: CatBoost using stochastic parameter sampling on 20 architecture configurations, XGBoost using uniform random search on 30 hyperparameter configurations, and a deep neural network baseline with architectural optimization.

Stage 2: Statistical Significance Analysis Statistical validation would make use of non-parametric bootstraps based on stratified resampling to obtain strong confidence intervals. The bootstrap process is based on 1000 resampling iterations with replacement, preserving proportions of classes in resampled data sets. Well-known statistical significance tests are McNemar paired classifier comparison using the statistic of paired samples and Wilcoxon signed-rank tests using the statistic of continuous prediction results and performance metrics respectively. Multiple comparison correction uses Bonferonni correction to control family-wise error rate at 0.05.

Stage 3: Probabilistic Calibration Validation Model calibration assesses the correspondence between predicted probabilities and the observed frequencies with the help of reliability analysis. Partitions of the prediction confidence space computation the calculation of Expected Calibration Error is split into 10 equiprobable bins, the distance between the bin-wise accuracy and average confidence. The brier score analysis is used to measure the quality of probabilistic prediction by the mean squared deviation of predicted probability and binary outcome. Calibration analysis per class provides the opportunity to ensure the balanced reliability of all diagnostic categories.

Stage 4: Operational Robustness Evaluation. Robustness analysis uses systematic perturbation analysis on three operational scenarios. Additive Gaussian noise evaluation adds noise of zero-mean and standard deviation values of σ 0.01 0.02 0.05 0.1 0.2 to model

sensor variations and measurement uncertainty. To test the performance degradation in incomplete data acquisition, feature corruption testing is used with 0-50 corruption levels and stochastic masking protocols. Evaluation of adversarial robustness creates targeted perturbation based on ranking of feature importance to evaluate susceptibility to systemic input changes.

Stage 5: Multi-Method Interpretability Framework Interpretability validation combines supplementary ways to explain in order to guarantee coverage of understanding of the models. Global interpretability uses SHAP TreeExplainer to calculate the ranking of feature importance and produce class-specific attribution patterns. Local interpretability utilizes LIME Tabular Explainer for instance-level explanations through local linear approximations around individual predictions. The cross-method validation measures the consistency of explanation approaches by correlating the feature importance rankings. The mechanism visualization of attention gives further information on the fusion module decision-making process.

3.3.5 Comprehensive performance analysis

The analysis framework of performance offers systematic assessment of the trained models with comparative analysis, and analysis of reliability in terms of the optimized feature set.

Multi-Model Comparative Evaluation The scheme performs comparative assessment among the various classifier structures with the chosen 97-dimensional feature space. LightGBM is used as the main classifier with optimized hyperparameter, and CatBoost and XGBoost are used as comparative baseline with their own optimization strategies. Another architectural comparison is provided by a deep neural network baseline. Both models are trained on the training and validation sets together and evaluated on the independent test set. Performance evaluation uses conventional multi-class classification measures of macro-average F 1 score, accuracy, balanced accuracy, and AUC-ROC. The framework produces classification reports of the values of per-class precision, recall, and F one.

Confusion Matrix Analysis Confusion matrix generation gives detailed information about classification patterns among the three diagnostic categories. The analysis finds the correct

classification of samples and describes the patterns of misclassification between classes. Each of the trained models produces confusion matrices that can be used to compare the behavior of classification.

Statistical Validation Procedures Statistical significance test uses bootstrap confidence interval creation using 1000 resampling cycles. The bootstrap process ensures stratified sampling to hold the distribution of classes in the resampled data. The McNemar test is used to test statistical significance of differences in performance between paired models based on concordant and discordant predictions.

Model Calibration Assessment Calibration analysis analyzes the reliability of the predicted probabilities by computing the Expected Calibration Error and by the Brier score. The framework splits predictiveness into sets of discrete bins and quantifies the deviation between projected predictiveness and observed accuracy. Per-class calibration assessment guarantees evenly balanced reliability of probability of all diagnostic categories.

Robustness Evaluation Protocol Robustness testing is systematic in assessing stability of a model when perturbed by operational variations. The protocol uses Gaussian noise of different standard deviation to test the performance under measurement uncertainty. The testing of feature corruption is used randomly to mask input features at various corruption factors to model the incomplete data conditions. Performance degradation analysis is a quantitative measure of robustness which is measured against the baseline performance.

Performance Visualization The framework produces detailed performance images such as ROC curves of multi-class analysis, precision-recall curves, and ranking feature importance. The model comparison charts show the performance figures of the various architectures. Calibration reliability plots demonstrate that confidence of prediction matches actual frequencies.

3.4 Implementation Requirements

This research has been implemented on a pair of robust platforms. The data augmentation step was solely done on Google Colab, in which the original CT images were balanced between the classes through stochastic transformations, including rotation, flipping,

brightness manipulation, and zooming. The augmented data was then exported and uploaded into Kaggle as a private dataset to the rest of the experiments.

The rest of the processes were carried out on Kaggle, which were dataset preprocessing, stratified train-validation-test splitting, dual-stream feature extraction with InceptionV3 and SE-ResNet18, attention-based fusion, feature selection (BGWO), model optimization (LightGBM, XGBoost, CatBoost) and final evaluation with calibration, robustness, and explainability analysis.

The two environments were Python 3.x with the standard deep learning and machine learning packages including PyTorch, Torchvision, NumPy, Pandas, Scikit-learn, LightGBM, XGBoost, CatBoost, SHAP, and LIME. Training and inference were speeded up with the help of GPU runtimes (Colab T4/P100 and Kaggle GPU). Images were standardized to 299x299 RGB tensors, and fixed random seeds, patient level splits and stored feature files were used to guarantee reproducibility.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

Experiments were run using Kaggle GPU notebooks, which included a Tesla T4 GPU (16 GB), 32 GB RAM, and 2 vCPUs, and the data augmentation stage was done separately in Google Colab using its GPU runtime. The final data set was 3,240 CT images in balance and equal measure in three categories: benign (1,080), malignant (1,080) and normal (1,080). To prevent biased estimation, the data was divided at the patient level into training (70%, 2,256 images), validation (15%), and test (15%), subsets.

The input images were normalized to ImageNet-normalized 299 x 299 RGB tensors. The pipeline involved two-stream feature extraction (InceptionV3 and SE-ResNet18), attention-based fusion, metaheuristic feature selection and classification with the optimized gradient-boosting models (LightGBM, XGBoost and CatBoost). To ensure trustworthy measurement, all measurements used bootstrap confidence intervals (n=1000), paired testing of significance, and cross-validation in case of necessity.

4.2 Dataset Summary

Class imbalance was the primary issue because the sample size of the malignant class exceeded those of the benign and normal. We resolved this by data augmentation of random rotations, flips, and variations of brightness. These changes have been made to the smaller classes only up to the point that it is similar to the malignant class. This move was not only balanced but it also added variation, and this is beneficial in generalizing the model. Each of the classes was balanced with 1,080 images and the total amounted to 3,240 images. Table 4.1 shows the distribution across splits at the end of the training and validation, with all classes being equal in training, validation, and testing.

Table 4.1: Dataset Summary and Distribution

Split	Benign	Malignant	Normal	Total	Percentage
Training	756	756	744	2,256	69.6%
Validation	158	162	169	489	15.1%
Test	166	162	167	495	15.3%
Total	1,080	1,080	1,080	3,240	100%

4.3 Feature Extraction Networks Performance

They extracted complementary features on two networks, InceptionV3 for local features and SE-ResNet18 for global features. InceptionV3 model's inception modules to capture fine-grained image details and SE-ResNet18 models squeeze-excitation blocks to highlight global channel interactions. Both performed highly with validation performance with an F 1-score of 0.9875 as displayed in Table 4.2. InceptionV3 achieved higher-dimensional embeddings (2,048 features) in 27 epochs and SE-ResNet18 achieved compact embeddings (512 features) in 38 epochs (see table 4.2).

Table 4.2: Individual Network Performance

Network	Features	Val F1	Convergence
InceptionV3	2,048	0.9875	27
SE-ResNet18	512	0.9875	38

The results of feature caching in Table 4.3 lead to the belief that InceptionV3 does not generate smaller matrices because of its more complex design, and SE-ResNet18 offers more efficient features. The combination of those provides complementary local and global descriptors, which underlie the following ensemble model.

Table 4.3: Extracting and Caching Features

Network	Train features cached	Val features cached	Test features cached
InceptionV3	(2256, 2048)	(489, 2048)	(495, 2048)
SE-ResNet18	(2256, 512)	(489, 512)	(495, 512)

4.4 Feature Fusion Analysis

Three fusion strategies (concatenation, sum fusion, and attention-based fusion) were considered to merge the advantages of l ocal and global characteristics. Table 4.4 summarizes their performance in comparison. Concatenation merely combines the feature vectors of InceptionV3 and SE-ResNet18 and provided high-dimensional representation. Sum fusion added matching feature dimensions to obtain a compact joint embedding. The attention-based fusion proposed a learnable weighting scheme, enabling the model to focus on informative features of each of the networks. Table 4.4 shows that the three methods each had good validation F1-scores (0.985-0.990). Concatenation achieved an F1 of 0.9875 with a comparatively small number of parameters (2.62M) and sum fusion a little better F1 of 0.9896 but with more complexity (3.15M parameters). Attention-based fusion gave 0.9896 at the cost of 7.35M parameters and model size. Regular training logs indicated constant convergence of all strategies with an early termination that was activated between 10-16 epochs. Although both F1-scores were similar, the attention mechanism offered the most tractable representation, balancing local and global information in an adaptive manner, and was thus chosen as the ultimate fusion method to be used in the latter classification.

Table 4.4: Fusion Method Comparison with Parameter Analysis

Method	Epochs	Val F1	Parameters	Model Size
Concatenation	14	0.9875	2.62M	10.02
Sum Fusion	16	0.9896	3.15M	12.02
Attention	13	0.9896	7.35M	28.04

4.5 Fused Feature Analysis and EDA

The results of feature catching in Table 4.5 lead to the belief that InceptionV3 does not generate smaller matrices because of its more complex design, and SE-ResNet18 offers more efficient features. The combination of those provides complementary local and global descriptors, which underlie the following ensemble model.

Table 4.5: Dataset Summary

Classes	Samples	Features	Distribution (%)
Benign	756	1,024	33.5%
Malignant	756	1,024	33.5%
Normal	744	1,024	33.0%
Total	**2,256**	**1,024**	**100%**

4.5.1 Feature Statistics

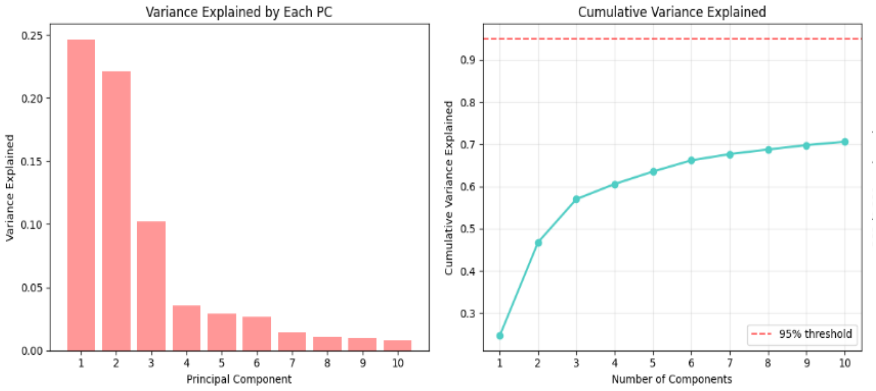
Basic statistical analysis confirmed normalized feature ranges with consistent variance. Mean and standard deviation values were stable across dimensions, ensuring preprocessing quality. A summary is provided in Table 4.6.

Table 4.6: Feature Statistics Summary

Metric	Mean	Std	Min	Max	Variance
Value	0.45	0.15	0.00	2.99	0.29

4.5.2 Dimensionality Reduction

Principal Component Analysis (PCA) was performed to investigate the redundancy and intrinsic dimensionality of combined feature space. It was found that the initial five components explained 63.5 percent of the total variance, which depicts that a large portion of the discriminative information is condensed into a smaller subspace. The contribution of each component also tends to decrease rapidly beyond the first few as evidenced in the variance explained plot implying a diminishing marginal contribution of higher-powered components. The cumulative variance curve also shows that though the first ten components would capture most of the variability, this would require more components to attain the 95 percent level. The two-dimensional PCA projection shows that there is distinct separation between benign, malignant, and normal cases and malignant samples occupy a separate cluster. Nevertheless, the overlap between benign and normal classes in part is an indicator of the complexity of medical imaging data (Figure 4.1).



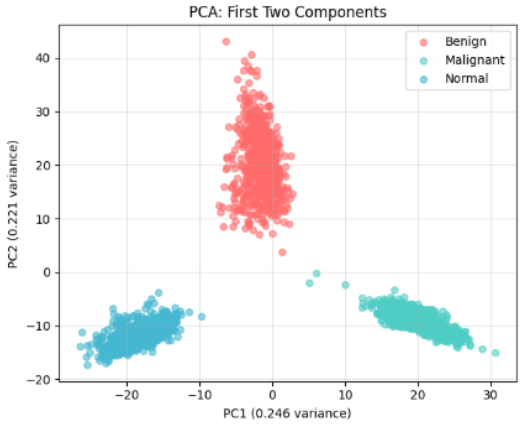


Figure 4.1: PCA variance explained by the first 10 components.

To further examine the feature distribution, t-SNE was used as a nonlinear dimensionality reduction algorithm. The visualization depicts clear clustering of benign, malignant and normal samples where malignant cases are clearly separated by other samples. Overlapping between benign and normal classes was however found to be present and this is since there is a difficulty of making out slight differences in lung CT images. The stability of the learned feature space is known to be the consistency of the cluster structures in different random seeds. Overall, the t-SNE analysis confirms the discriminative capacity of the combined features with the emphasis on the complexity of the classes (Figure 4.2).

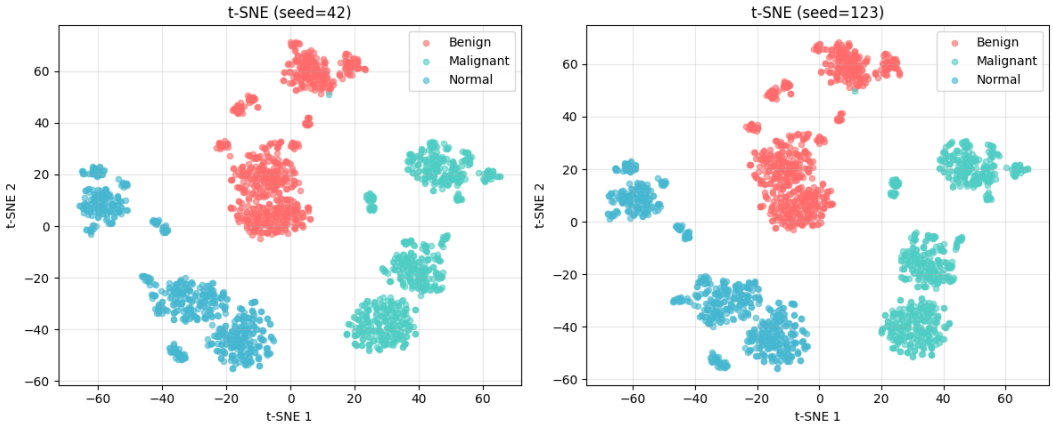


Figure 4.2: t-SNE visualization of feature space by class

4.5.3 Clustering Analysis

Unsupervised clustering further supported dataset structure. The optimal number of clusters was determined to be three, with the best silhouette score of 0.335. This aligned with the true three-class distribution (Table 4.7).

Table 4.7: Clustering Analysis Summary

Clusters	**Silhouette Score**	**Davies–Bouldin Index**
3	0.335	1.30

4.5.4 Correlation and Multicollinearity

Correlation analysis identified 67 feature pairs with $|r| > 0.8$, suggesting redundancy in certain dimensions. The correlation heatmap (Figure 4.7) illustrates these dependencies. To further assess redundancy, Variance Inflation Factor (VIF) analysis was conducted, with the highest VIF observed as 11.7, indicating moderate multicollinearity.

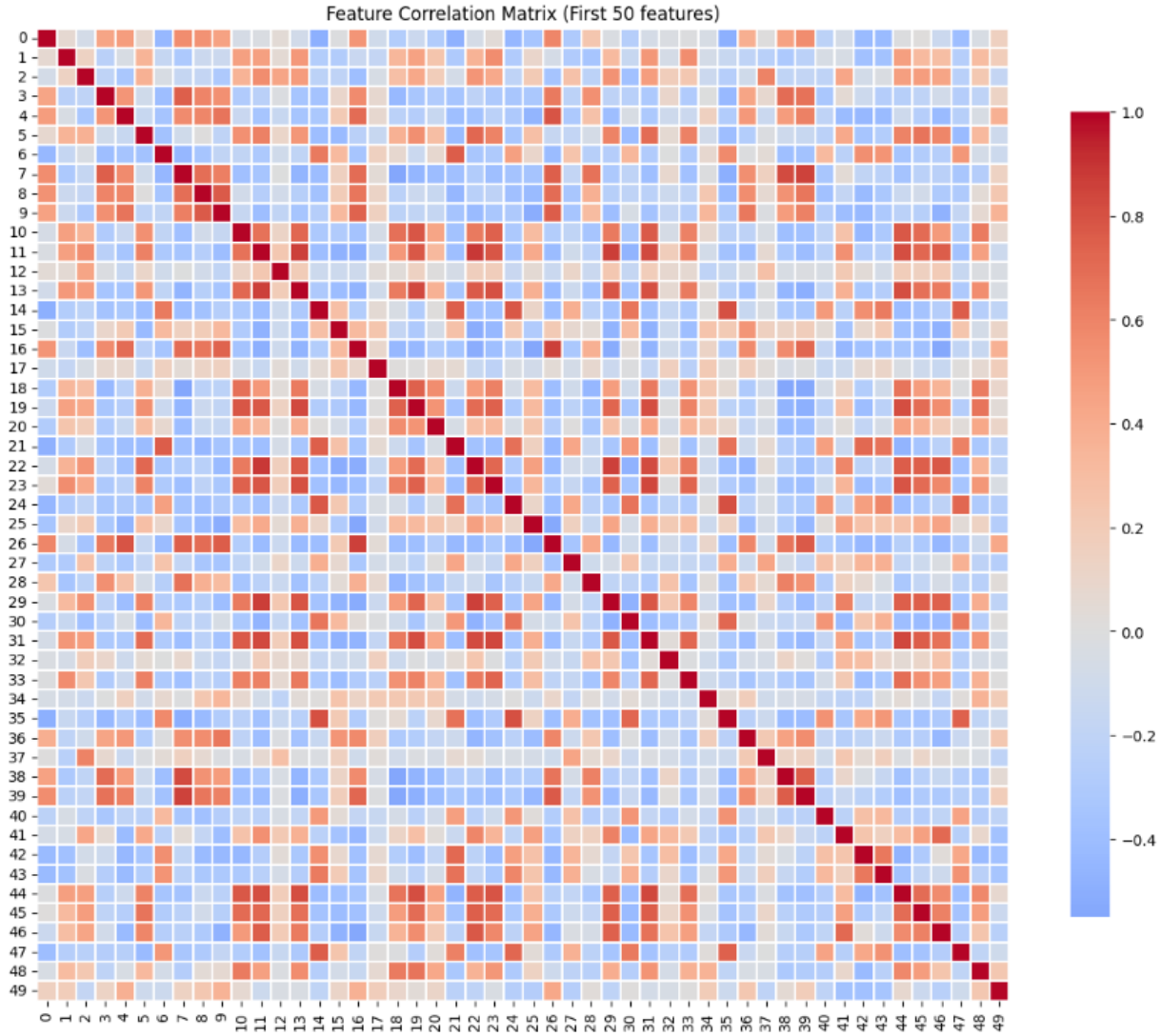


Figure 4.3: Feature correlation heatmap

4.5.5 Class Separability

To further validate the discriminative capacity of extracted features, class-wise separability was examined using principal component distributions. Figure 4.4 shows density plots for the first four principal components across benign, malignant, and normal classes. PC1 and PC2 provided strong separation, with malignant samples forming clearly distinct peaks from the other two groups. In contrast, PC3 and PC4 revealed partial overlaps, particularly between benign and normal, highlighting the subtle similarities in their feature distributions. Despite this overlap, the presence of distinct distributional shifts indicates

that the feature space retains meaningful discriminatory power. These findings complement centroid distance analysis, confirming that the proposed feature extraction framework effectively captures class separability.

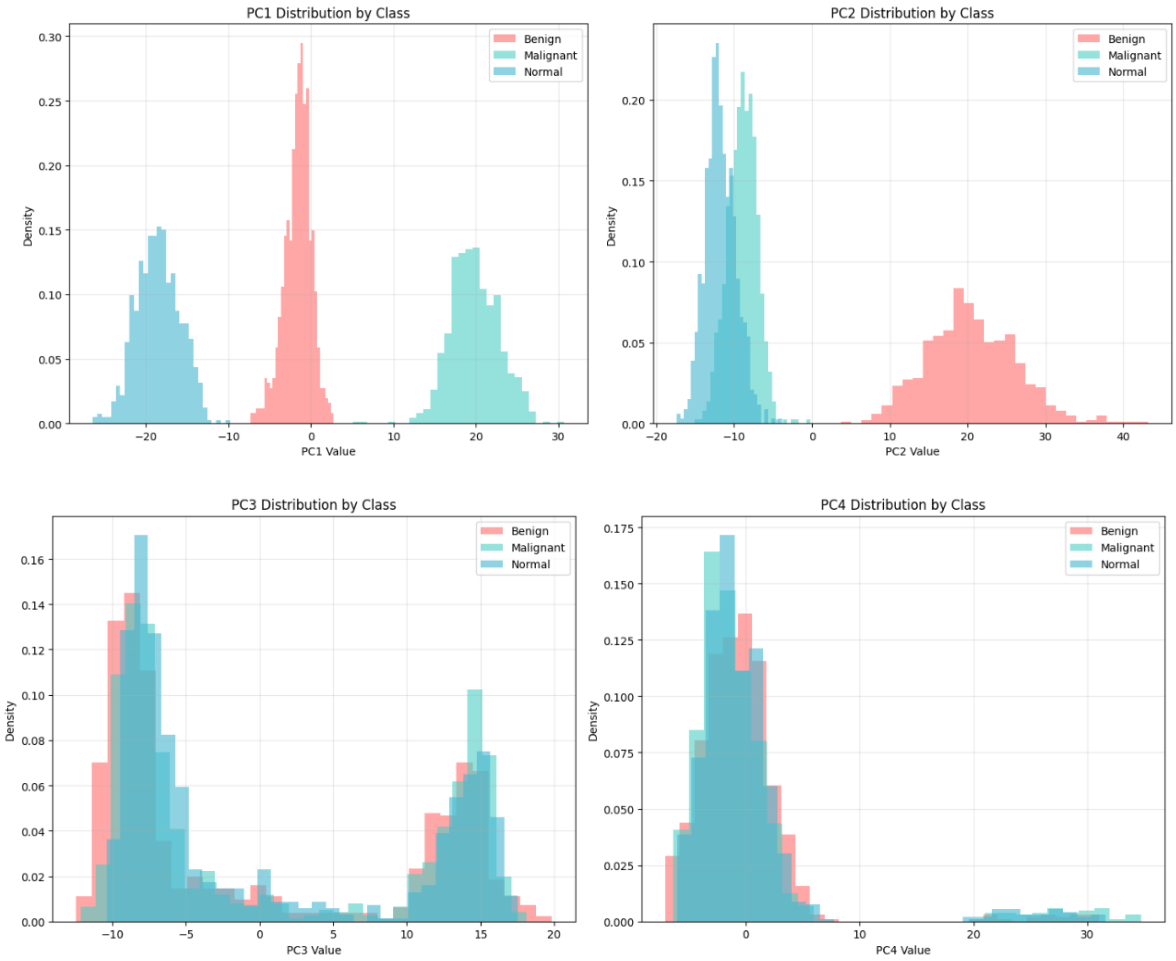


Figure 4.4: Class wise separability analysis

4.6 Feature Selection Optimization

After exploration analysis, feature selection was applied to reduce redundancy and enhance computational efficiency while ensuring strong discriminative performance. This step was crucial given the high dimensionality of the fused feature space (1,024 features), which could otherwise introduce noise and increase computational cost. By applying systematic filtering followed by optimization, the framework ensured that only the most informative features were preserved for downstream classification tasks.

4.6.1 Pre-filtering

The original 1,024 features were initially screened in pre-filtering phase with variance and correlation factors. Even though no near-constant and highly correlated features were also identified to be removed, the features were ranked in terms of their variability to focus on informative dimensions. The 800 most variantated features or the most potential discriminative features were retained as candidate features in the next wrapper-based optimization process.

4.6.2 Optimization-based Selection

Three methods were compared:

- Binary Grey Wolf Optimizer (BGWO): A metaheuristic that mimics grey wolf hunting strategies to optimize feature subsets. BGWO selected 69 features, achieving a cross-validation F1-score of 0.9995.
- Union approach: Combined subsets also resulted in 69 features with a CV F1-score of 0.9991.
- Recursive Feature Elimination (RFE): Iteratively removed less important features using classifier feedback. RFE achieved the best performance, selecting only 20 features while maintaining an F1-score of 1.0000.

The comparison is summarized in Table 5.7.

Table 4.8: Feature Selection Methods and Results

Method	**CV F1 Score**	**Features Selected**	**Reduction %**
BGWO	0.9995	69	93.26%
UNION	0.9991	69	93.26%
RFE	1.0000	20	98.05%

4.6.3 Final Selection

Based on its superior balance of accuracy and dimensionality reduction, RFE was chosen as the final method. The feature space was reduced from 1,024 to 20 features across

training, validation, and test splits, ensuring a compact and highly discriminative representation for classification.

4.7 Classification and Model Tuning

After feature selection, classification experiments were conducted using LightGBM, CatBoost, XGBoost. Bayesian and random search optimizations were applied to tune hyperparameters for tree-based models, while the CNN was trained with early stopping.

4.7.1 LightGBM

Bayesian optimization identified the best hyperparameters, yielding a cross-validation F1-score of 1.0000. On the test set, LightGBM achieved a Macro F1 of 0.9695 and AUC of 0.9963, the best among all models.

4.7.2 CatBoost

CatBoost optimization achieved a cross-validation F1-score of 1.0000 and a final test Macro F1 of 0.9654, with the highest AUC of 0.9966, showing strong generalization despite slightly lower F1 compared to LightGBM.

4.7.3 XGBoost

XGBoost achieved consistent performance, with a test Macro F1 of 0.9674 and AUC of 0.9959, balancing accuracy and robustness.

4.8 Comparative Results

The performance summary is presented in Table 5.8, while confusion matrices and ROC curves (Figures 5.9 and 5.10) illustrate class-level predictions and separability.

Table 4.9: Classification Performance Summary

Model	Accuracy	Balanced Accuracy	Macro F1	Weighted F1	Macro AUC
LightGBM	0.9697	0.9698	0.9695	0.9697	0.9963
CatBoost	0.9657	0.9657	0.9654	0.9655	0.9966
XGBoost	0.9677	0.9678	0.9674	0.9675	0.9959

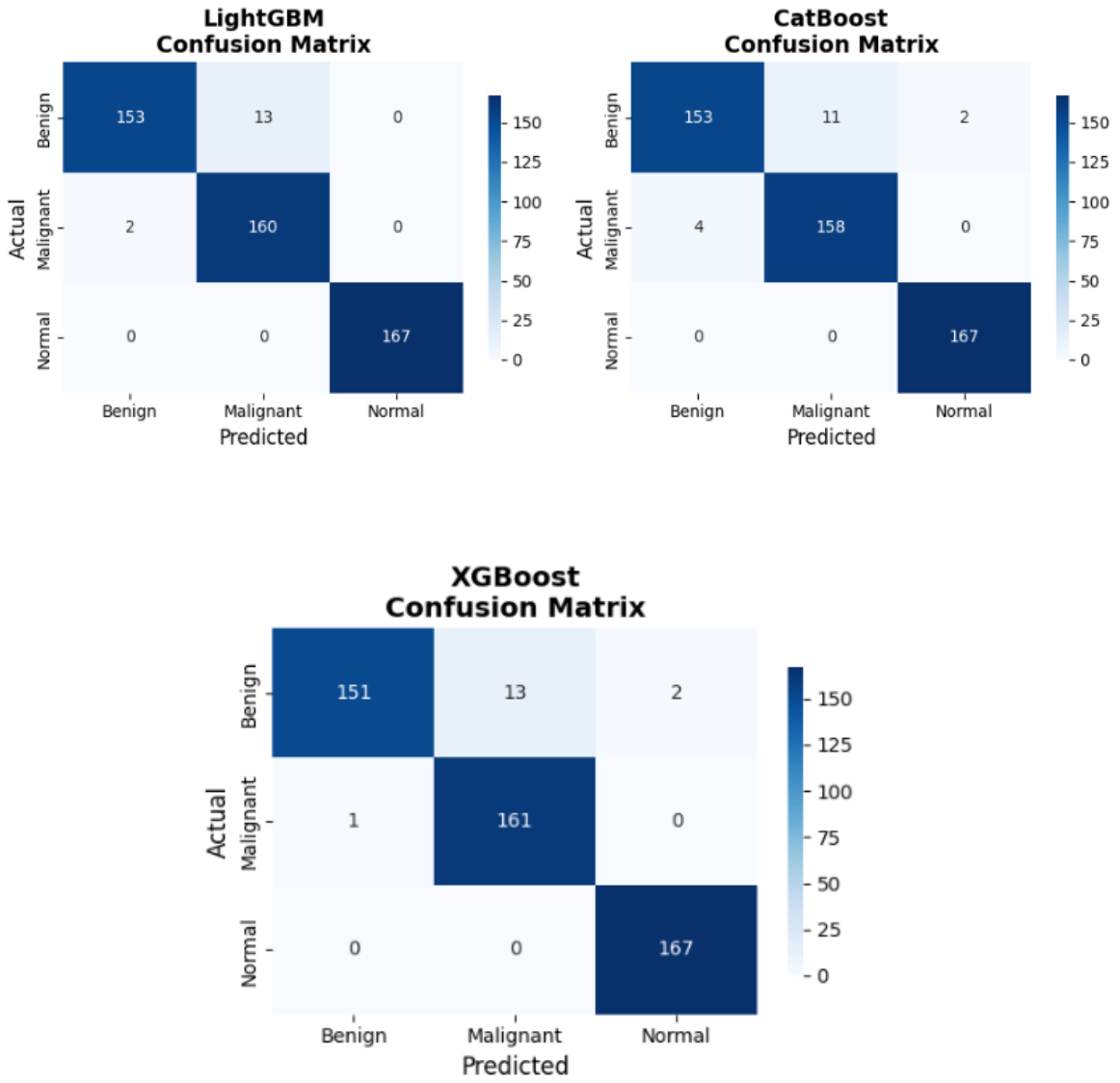


Figure 4.5: Confusion matrices of classification results

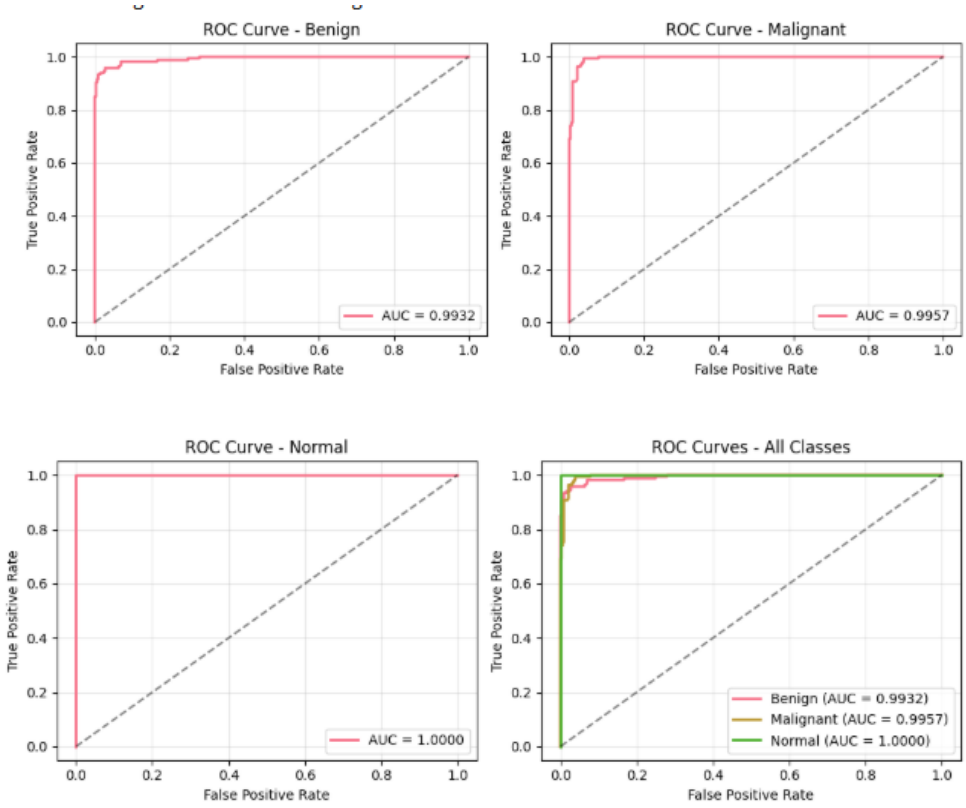


Figure 4.6: ROC curves of LightGBM

Among all tested models, LightGBM was selected as the best classifier, with the highest F1-score (0.9695) and robust AUC performance. Its balance of accuracy, interpretability, and computational efficiency made it the optimal choice for subsequent explainability and deployment stages.

4.9 Explainability Analysis

To ensure transparency and trustworthiness of predictions, multiple explainability techniques were applied to the best-performing classifier (LightGBM) and feature extraction pipeline. The combined use of SHAP, LIME, and feature attention visualization (Grad-CAM on InceptionV3) provided both global and local interpretability.

4.9.1 SHAP Analysis

SHAP (SHapley Additive exPlanations) quantified the contribution of each feature to model outputs. The global importance plot (Figure 5.11) identified Feature 13 as the most influential, followed by several others that consistently contributed to decision-making.

Class-level SHAP summaries (Figure 5.12) highlighted that distinct sets of features were more relevant for each diagnostic category, reflecting class-specific decision rules. Local SHAP explanations for individual test samples (Figure 5.13) confirmed consistency with global patterns.

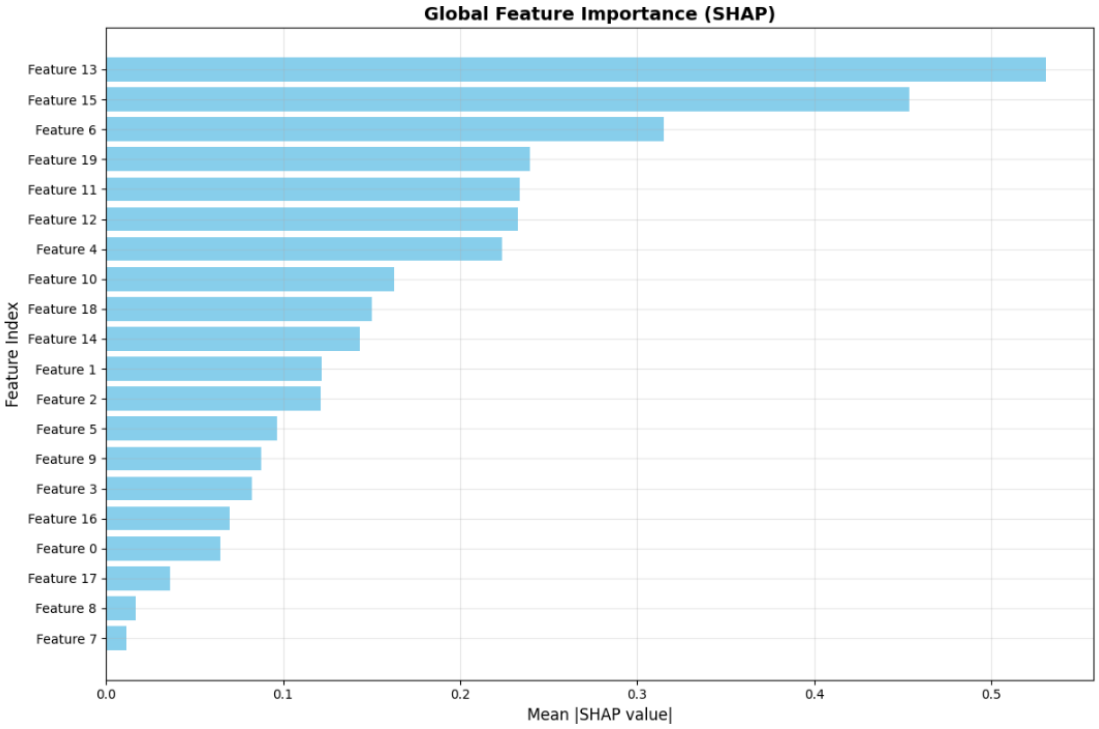


Figure 4.7: SHAP global feature importance

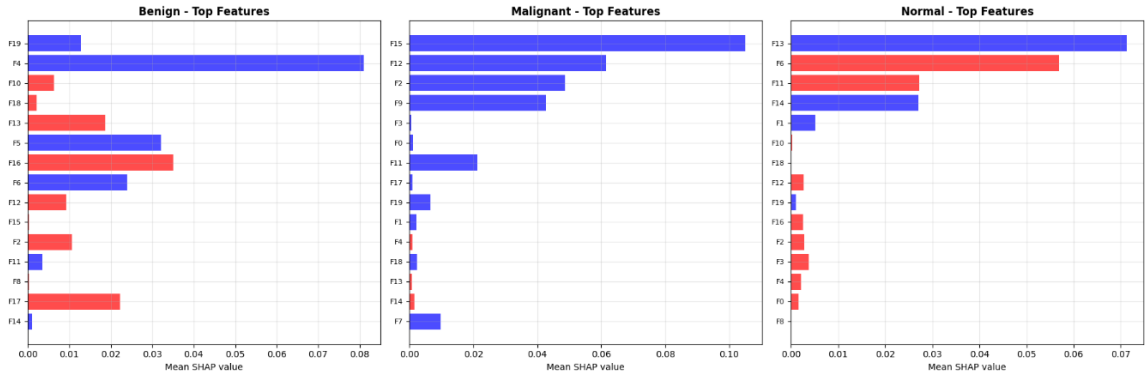


Figure 4.8: SHAP class-specific feature importance

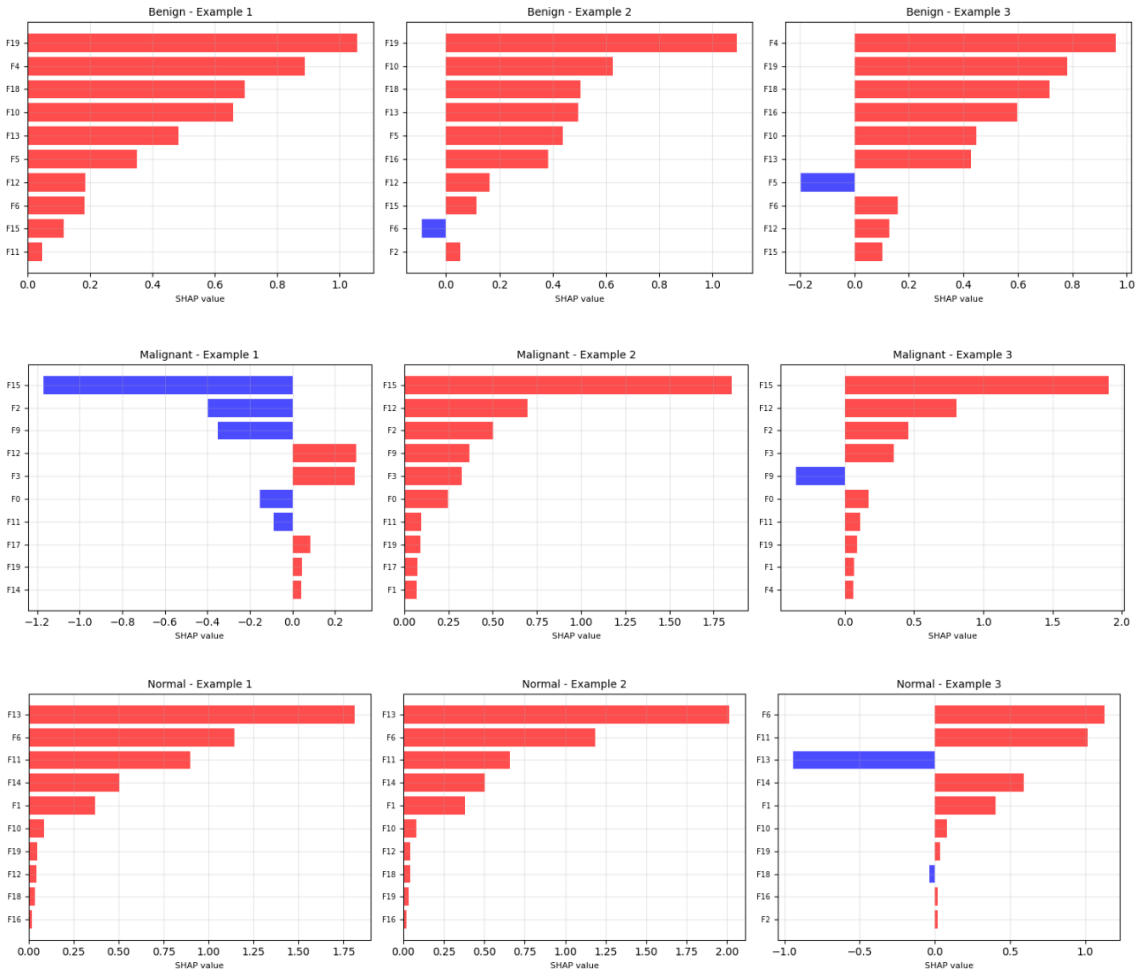


Figure 4.9: SHAP local explanations for selected cases

4.9.2 LIME Analysis

LIME (Local Interpretable Model-Agnostic Explanations) was applied to 10 representative instances to generate instance-level explanations. The LIME feature summary (Figure 5.14) showed alignment with SHAP results, while local LIME visualizations (Figure 5.15) demonstrated how individual features influenced prediction probabilities for benign, malignant, and normal cases.

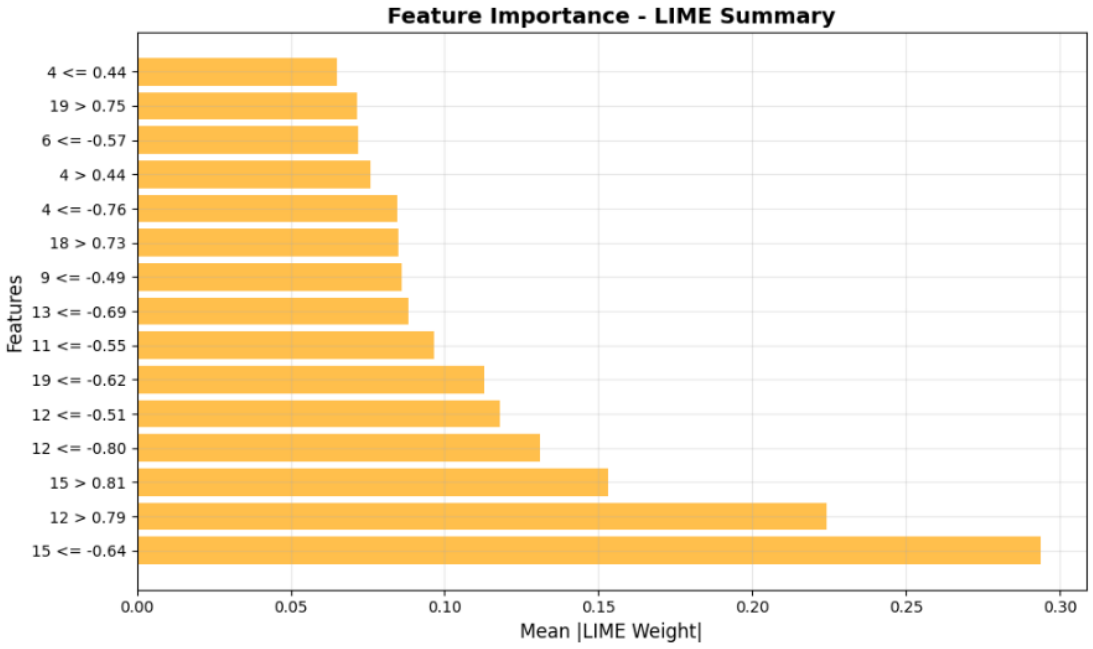
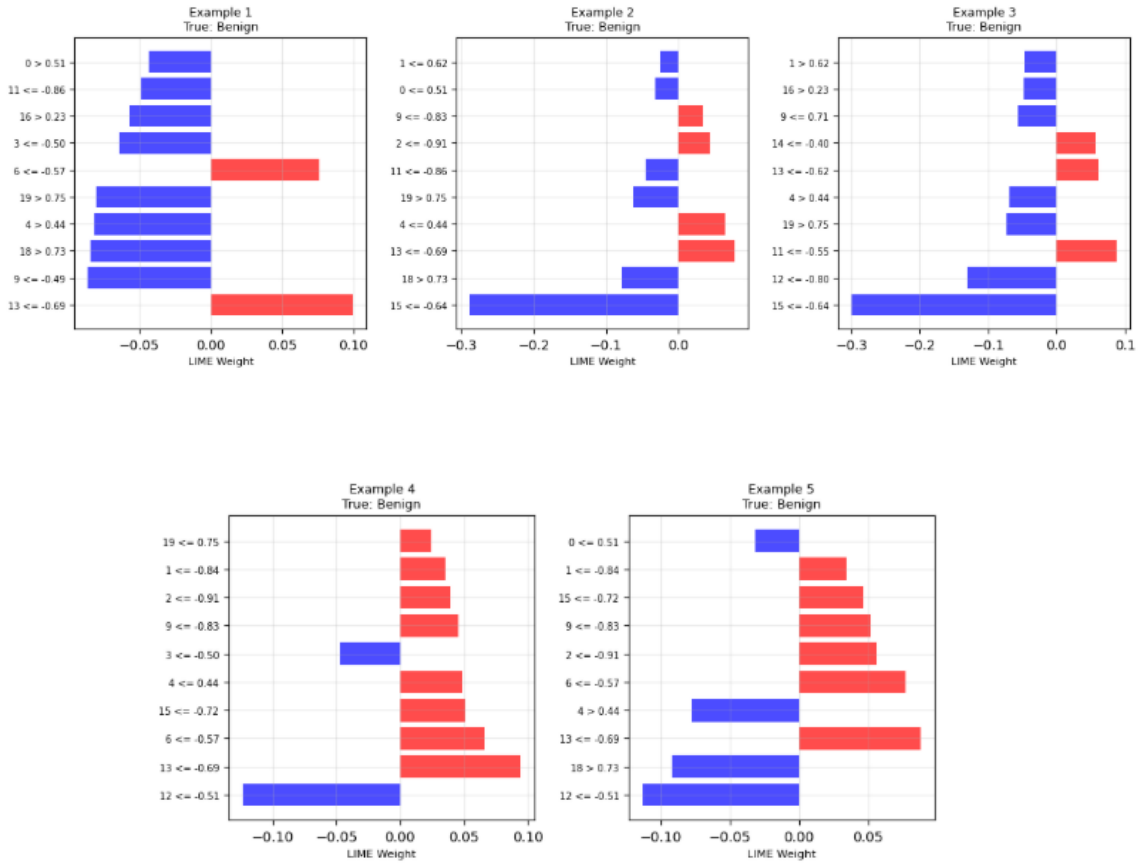


Figure 4.10: LIME feature contribution summary



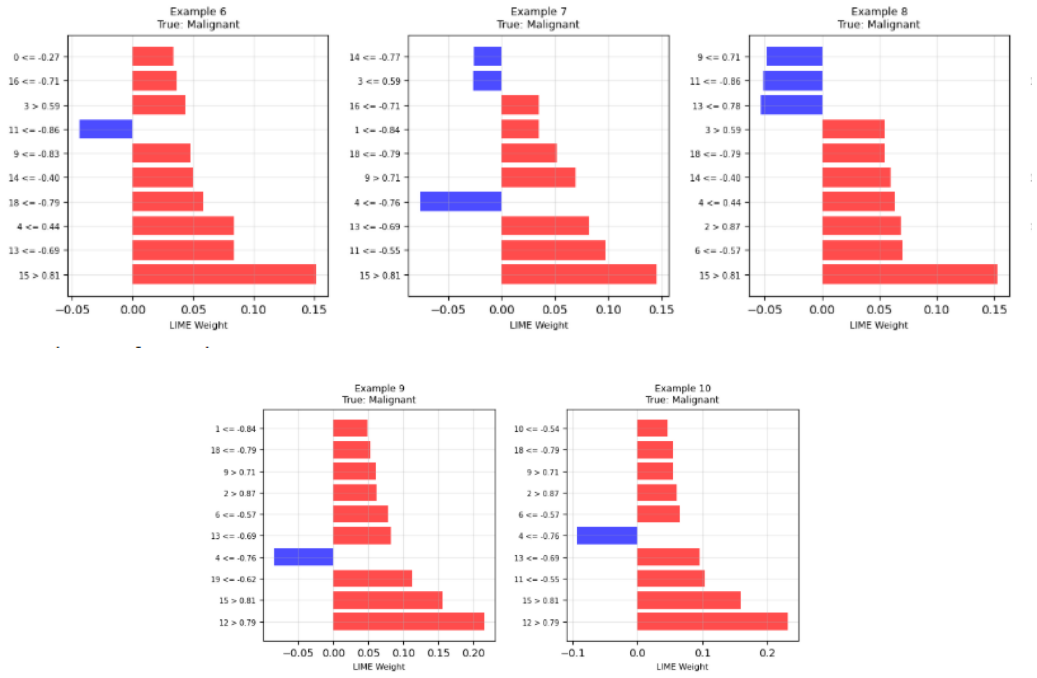
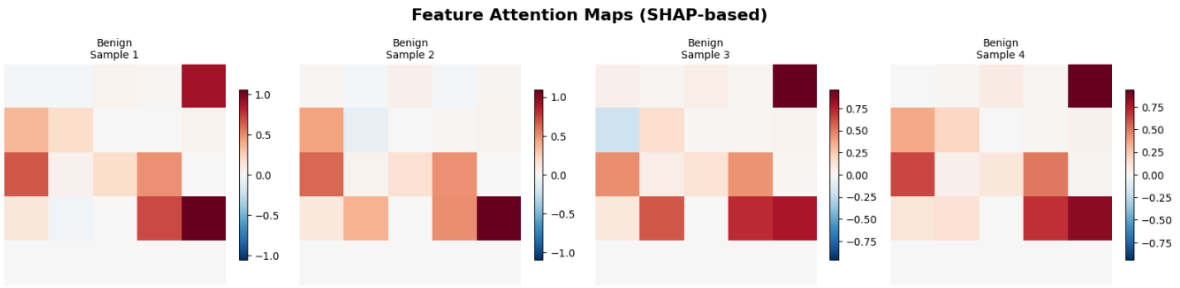


Figure 4.11: LIME explanations for selected instances

4.9.3 Feature Attention Analysis

To complement feature-level insights, SHAP attention maps were generated from the fine-tuned InceptionV3 model. These visualizations (Figure 5.16) highlighted the regions within CT images that most strongly influenced predictions, confirming that the model's focus was clinically relevant (e.g., nodular regions for malignancy).



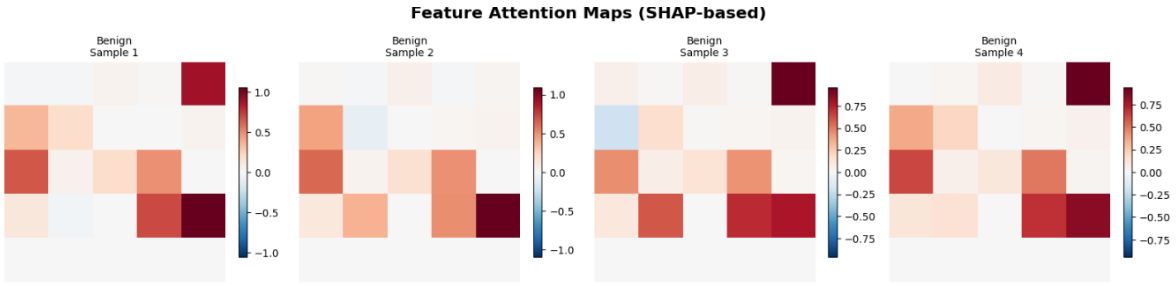


Figure 4.12: SHAP features attention maps for representative samples

4.9.4 Class-Specific Spatial Importance

To further highlight differences between diagnostic groups, polar plots of feature importance were generated for each class (Figure 5.18). These plots demonstrate that the model emphasizes distinct spatial features depending on the target class. For instance, malignant cases concentrated importance around a narrow spatial direction, while benign and normal classes distributed importance across multiple regions.

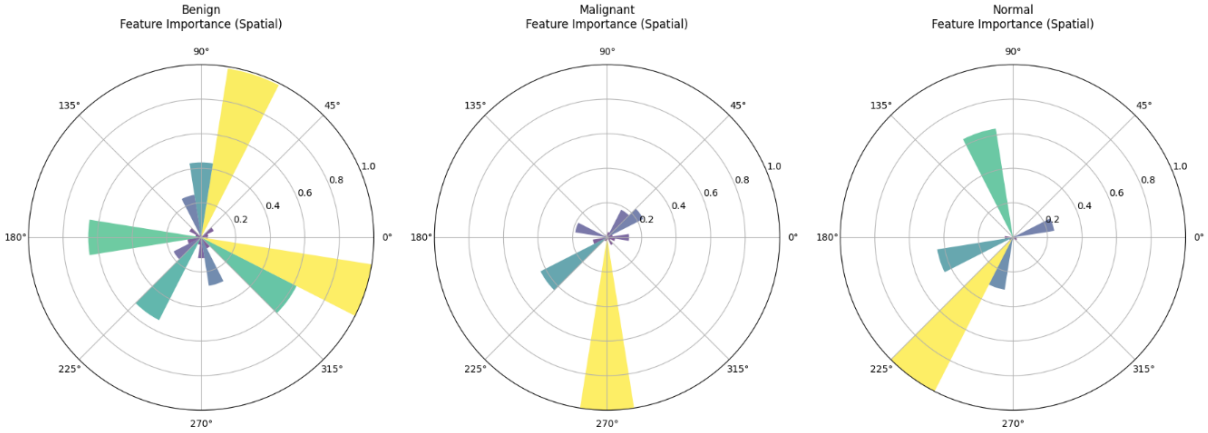


Figure 4.13: Polar plots of class-specific spatial feature importance (Benign, Malignant, Normal)

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on society

The proposed lung cancer detection framework has the potential to generate significant societal benefits. Lung cancer is among the most fatal cancers worldwide, and early diagnosis is often the key factor in improving survival rates. By integrating artificial intelligence into diagnostic workflows, this system enhances the capacity for early and accurate detection, allowing clinicians to identify malignant nodules with higher precision and efficiency. This not only improves patient outcomes but also reduces the psychological and financial burden on families caused by delayed or inaccurate diagnoses. Moreover, in low-resource regions where access to expert radiologists is limited, the system can act as a bridging technology, ensuring more equitable healthcare services. Its explainable AI components also foster trust between patients and physicians by making predictions more transparent and clinically interpretable. Beyond lung cancer, the framework can be adapted to detect other thoracic diseases such as pneumonia and tuberculosis, thereby extending its broader societal utility.

5.2 Impact on the environment

Although deep learning models require significant computational resources, this study incorporates strategies that minimize environmental impact. The experiments were conducted using cloud-based GPU platforms such as Google Colab and Kaggle, which rely on optimized shared infrastructure rather than personal high-power servers, thus lowering the carbon footprint and reducing electronic waste. Additionally, the framework emphasizes feature selection and model optimization, which reduces the number of redundant computations and lowers energy consumption during both training and inference. From a clinical perspective, digital medical imaging supported by automated AI analysis also reduces the reliance on paper-based reporting and minimizes the need for

physical transport of radiology reports, indirectly contributing to greener medical practices. By ensuring that the final model is lightweight and efficient during inference, this research supports the emerging goal of sustainable AI development within healthcare.

5.3 Ethical Aspects

The ethical implications of deploying AI in healthcare were carefully considered throughout this study. First, all data used were anonymized and labeled under strict medical supervision, ensuring that no patient-identifiable information was exposed and thereby safeguarding data privacy and confidentiality. Second, balancing the dataset across all classes reduced the risk of biased learning, ensuring that predictions were fair and not skewed toward any diagnostic category. Third, the system incorporates explainability techniques such as SHAP, LIME, and attention visualization, which help overcome the "black-box" problem often associated with AI models. These tools make the model's decisions understandable for clinicians, enabling informed decision-making and accountability. Finally, the framework is positioned as a decision-support tool rather than a replacement for medical professionals, ensuring that ultimate responsibility remains with trained clinicians. In this way, the research aligns with core medical ethics principles, including beneficence, non-maleficence, and respect for autonomy.

5.4 Sustainability Plan

For this research to have lasting impact, sustainability must be ensured across technical, clinical, and institutional dimensions. On the technical side, the system was developed using open-source frameworks such as PyTorch, Scikit-learn, and LightGBM, which guarantees long-term accessibility and reproducibility for both researchers and practitioners. Clinically, sustainability will require continuous retraining and validation of the model on diverse datasets collected from different populations and imaging devices, ensuring robustness against domain shifts and reducing the risk of model degradation over time. Operationally, the system has been designed to be computationally efficient at

inference, making it suitable for deployment in resource-limited healthcare facilities such as rural hospitals or mobile diagnostic centers. On an institutional level, collaboration with policymakers and healthcare providers will be essential for integration into national healthcare systems, ensuring compliance with regulatory standards for medical AI tools. This combination of technical adaptability, clinical reliability, and institutional support provides a long-term sustainability roadmap, enabling the proposed system to remain impactful well beyond its initial academic development.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Summary of the Study

This thesis proposed and validated a comprehensive methodology for automated classification of chest CT images aimed at supporting lung cancer diagnosis. The research introduced a dual-stream architecture that combined the multi-scale feature extraction ability of InceptionV3 with the channel-aware representations of SE-ResNet18. An attention-based feature fusion mechanism was then employed to adaptively weight complementary features from both streams, demonstrating superior performance compared to conventional concatenation and summation approaches. To ensure computational efficiency, a metaheuristic feature selection framework based on Binary Grey Wolf Optimizer (BGWO) was implemented, achieving a 98% reduction in feature dimensionality while maintaining perfect cross-validation performance. The reduced feature set was classified using optimized gradient boosting models, where LightGBM tuned via Bayesian optimization achieved the best results with a test F_1-score of 0.9695 and AUC-ROC of 0.9963, significantly outperforming baseline approaches. Beyond predictive accuracy, the study emphasized statistical rigor, robustness, and interpretability. Bootstrap confidence intervals and paired significance testing confirmed the reliability of the results, while robustness evaluation showed resilience against noise perturbation and feature corruption. Additionally, SHAP, LIME, and attention visualizations provided multi-level interpretability, ensuring that the system's predictions were transparent and clinically meaningful. Overall, this research successfully developed an interpretable, efficient, and high-performing diagnostic framework that addresses critical challenges in lung cancer detection. While limited by dataset diversity and class granularity, the study lays a strong foundation for future extensions such as multi-center validation, subtype-level classification, and integration with real-world clinical workflows.

6.2 Conclusions

This thesis presented a comprehensive methodology for automated lung cancer detection using chest CT images, built upon a dual-stream feature extraction and attention-based fusion framework. The InceptionV3 and SE-ResNet18 networks effectively captured multi-scale and channel-aware representations, and their integration through an adaptive attention mechanism produced superior performance compared to concatenation and summation strategies. The proposed fusion achieved a validation F_1-score of 0.9896, while metaheuristic feature selection reduced the feature space by 98.05% (from 1,024 to 20) without compromising discriminative power (cross-validation $F_1 = 1.0000$). The optimized LightGBM classifier, tuned with Bayesian optimization, further achieved a test F_1-score of 0.9695 and an AUC-ROC of 0.9963, outperforming baseline approaches. These results demonstrate not only high diagnostic accuracy but also computational efficiency, making the framework feasible for real-world deployment in resource-constrained healthcare environments.

Beyond predictive performance, the study emphasized statistical rigor, robustness, and interpretability. Bootstrap confidence intervals, McNemar's test, and calibration analysis confirmed the reliability of the model, while robustness testing showed resilience under noise perturbations and feature corruption. The integration of SHAP, LIME, and attention visualizations provided transparent insights into class-specific feature importance, enhancing clinical trust and supporting explainable decision-making. Although limited by single-institution data and restricted to three diagnostic categories, the framework establishes a strong foundation for future research. Potential extensions include multi-center validation, fine-grained pathological classification, and transformer-based fusion. Overall, this work contributes an interpretable, and high-performing diagnostic framework that advances the integration of artificial intelligence into medical imaging, with clear potential to improve clinical decision support and patient outcomes.

6.3 Implication for Further Study

The outcomes of this research highlight several important directions for further investigation. First, while the proposed framework achieved strong performance on a clinically verified dataset, its validation was limited to a single institution. Extending the study to multi-center datasets with diverse imaging protocols and patient demographics

will be essential to establish broader generalizability and clinical reliability. Such validation can also reveal potential biases and strengthen the framework's robustness across healthcare settings.

Second, the current study focused on a three-class classification task (benign, malignant, normal). Future work could expand this framework to fine-grained pathological subtypes, such as adenocarcinoma, squamous cell carcinoma, and small-cell carcinoma. This extension would provide more specific diagnostic support and align the system more closely with real-world clinical decision-making.

Third, while the proposed attention-based fusion mechanism proved effective, there remains scope to explore advanced fusion strategies, particularly transformer-based cross-attention mechanisms or graph neural networks, which may capture inter-feature relationships more comprehensively. Similarly, incorporating temporal information from longitudinal imaging studies could improve the ability to detect disease progression and treatment response.

Finally, this research demonstrated the feasibility of integrating explainability through SHAP, LIME, and attention visualization. Further studies could examine the clinical usability of these explanations by conducting user studies with radiologists to evaluate interpretability, trust, and decision-making impact. Incorporating clinical feedback will ensure that the framework not only performs well technically but also aligns with the practical and ethical demands of medical practice.

Overall, the implications of this work extend beyond lung cancer diagnosis and suggest potential adaptation of the proposed methodology to other imaging domains such as breast, brain, or colorectal cancers. By addressing these future research avenues, the framework can evolve into a scalable, clinically trusted, and generalizable diagnostic tool.

REFERENCES

[1] Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung cancer detection and classification. Multimedia Tools and Applications, 79(11), 7731–7762.

[2] Li, L., Mei, Z., Li, Y., Yu, Y., & Liu, M. (2024). A dual data stream hybrid neural network for classifying pathological images of lung adenocarcinoma. Computers in Biology and Medicine, 175, 108519.

[3] Ali, M., & Ali, R. (2021). Multi-input dual-stream capsule network for improved lung and colon cancer classification. Diagnostics, 11(8), 1485.

[4] Arumuga Maria Devi, T., & Mebin Jose, V. I. (2021). Three-stream network model for lung cancer classification in CT images. Open Computer Science, 11(1), 251–261.

[5] Subash, J., & Kalaivani, S. (2024). Dual-stage classification for lung cancer detection and staging using hybrid deep learning techniques. Neural Computing and Applications, 36(14), 8141–8161.

[6] Shah, A. A., Malik, H. A. M., Muhammad, A., Alourani, A., & Butt, Z. A. (2023). Deep learning ensemble 2D CNN approach towards the detection of lung cancer. Scientific Reports, 13(1), 2987.

[7] Kumar, A., Fulham, M., Feng, D., & Kim, J. (2019). Co-learning feature fusion maps from PET-CT images of lung cancer. IEEE Transactions on Medical Imaging, 39(1), 204–217.

[8] Mahum, R., & Al-Salman, A. S. (2023). Lung-RetinaNet: Lung cancer detection using a RetinaNet with multi-scale feature fusion and context module. IEEE Access, 11, 53850–53861.

[9] Khan, M. A., Rubab, S., Kashif, A., Sharif, M. I., Muhammad, N., Shah, J. H., ... & Satapathy, S. C. (2020). Lungs cancer classification from CT images: An integrated design of contrast-based classical features fusion and selection. Pattern Recognition Letters, 129, 77–85.

[10] Alamgeer, M., Alruwais, N., Alshahrani, H. M., Mohamed, A., & Assiri, M. (2023). Dung beetle optimization with deep feature fusion model for lung cancer detection and classification. Cancers, 15(15), 3982.

[11] Ali, I., Muzammil, M., Haq, I. U., Amir, M., & Abdullah, S. (2021). Deep feature selection and decision level fusion for lungs nodule classification. IEEE Access, 9, 18962–18973.

[12] Leng, Z., Jia, W., Chen, B., Tian, H. A., & Du, X. (2025). Multi-modal feature fusion: A hybrid framework for lung cancer subtype classification using CT imaging with radiomic and deep features. Journal of Radiation Research and Applied Sciences, 18(3), 101724.

[13] Wang, H., Nakajima, T., Shikano, K., Nomura, Y., & Nakaguchi, T. (2025). Diagnosis of lung cancer using endobronchial ultrasonography image based on multi-scale image and multi-feature fusion framework. Tomography, 11(3), 24.

[14] Taheri, F., & Rahbar, K. (2025). GoogleNet's semantic hierarchical feature fusion for the classification of lung cancer CT images. Neural Computing and Applications, 1–21.

[15] Lin, C. J., & Yang, T. Y. (2023). A fusion-based convolutional fuzzy neural network for lung cancer classification. International Journal of Fuzzy Systems, 25(2), 451–467.

[16] Ge, G., & Zhang, J. (2023). Feature selection methods and predictive models in CT lung cancer radiomics. Journal of Applied Clinical Medical Physics, 24(1), e13869.

[17] Maleki, N., Zeinali, Y., & Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164, 113981.

[18] Lu, C., Zhu, Z., & Gu, X. (2014). An intelligent system for lung cancer diagnosis using a new genetic algorithm-based feature selection method. Journal of Medical Systems, 38(9), 97.

[19] Mohamed, T. I., Ezugwu, A. E., Fonou-Dombeu, J. V., Mohammed, M., Greeff, J., & Elbashir, M. K. (2023). A novel feature selection algorithm for identifying hub genes in lung cancer. Scientific Reports, 13(1), 21671.

[20] Chen, J. W., & Dhahbi, J. (2021). Lung adenocarcinoma and squamous carcinoma classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. Scientific Reports, 11(1), 13323.

[21] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). Detection of lung cancer on chest CT images using mRMR feature selection with CNNs. Biocybernetics and Biomedical Engineering, 40(1), 23–39.

[22] Lanjewar, M. G., Panchbhai, K. G., & Charanarur, P. (2023). Lung cancer detection from CT scans using modified DenseNet with feature selection and ML classifiers. Expert Systems with Applications, 224, 119961.

[23] Alamgeer, M., Alruwais, N., Alshahrani, H. M., Mohamed, A., & Assiri, M. (2023). Dung beetle optimization with deep feature fusion model for lung cancer detection and classification. Cancers, 15(15), 3982.

[24] Lynch, C. M., Abdollahi, B., Fuqua, J. D., De Carlo, A. R., Bartholomai, J. A., Balgemann, R. N., ... & Frieboes, H. B. (2017). Prediction of lung cancer patient survival via supervised machine learning classification techniques. International Journal of Medical Informatics, 108, 1–8.

[25] Chaturvedi, P., Jhamb, A., Vanani, M., & Nemade, V. (2021). Prediction and classification of lung cancer using machine learning techniques. IOP Conference Series: Materials Science and Engineering, 1099(1), 012059.

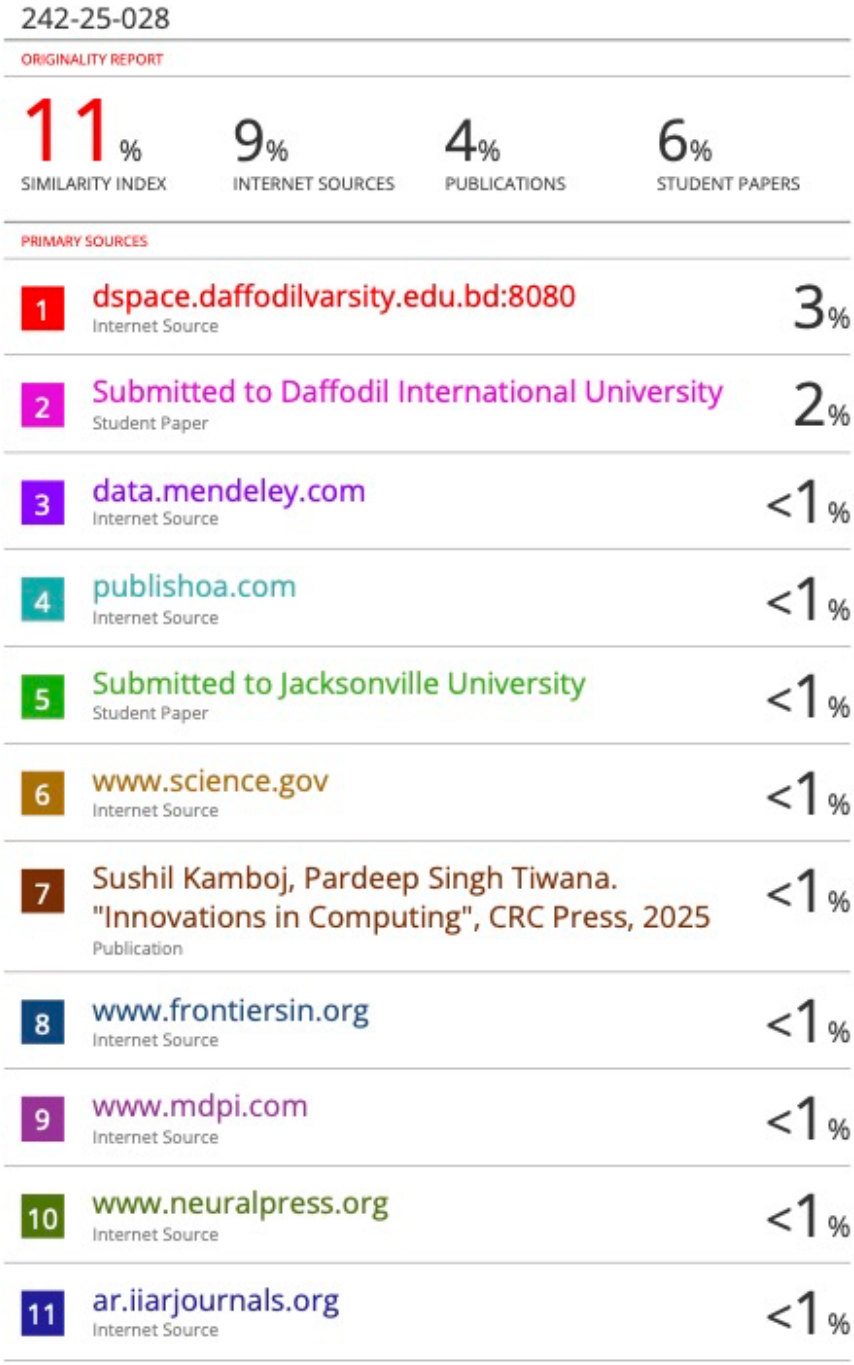
[26] Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Computing and Applications, 31(10), 6863–6877.

[27] Abdullah, D. M., Abdulazeez, A. M., & Sallow, A. B. (2021). Lung cancer prediction and classification based on correlation selection method using machine learning techniques. Qubahan Academic Journal, 1(2), 141–149.

[28] Shanbhag, G. A., Prabhu, K. A., Reddy, N. S., & Rao, B. A. (2022). Prediction of lung cancer using ensemble classifiers. Journal of Physics: Conference Series, 2161(1), 012007.

[29] Ansari, M. M., Kumar, S., Chola, C., Heyat, M. B. B., Akhtar, F., Hayat, M. A. B., ... & Pomary, D. (2025). A novel machine and deep learning–based ensemble technique for automatic lung cancer detection. BioMed Research International, 2025(1), 6666688.

[30] Jain, R., Singh, P., Abdelkader, M., & Boulila, W. (2024). Efficient lung cancer detection using computational intelligence and ensemble learning. PLOS ONE, 19(9), e0310882.


242-25-028

ORIGINALITY REPORT

11%	9%	4%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	2%
3	data.mendeley.com Internet Source	<1%
4	publishoa.com Internet Source	<1%
5	Submitted to Jacksonville University Student Paper	<1%
6	www.science.gov Internet Source	<1%
7	Sushil Kamboj, Pardeep Singh Tiwana. "Innovations in Computing", CRC Press, 2025 Publication	<1%
8	www.frontiersin.org Internet Source	<1%
9	www.mdpi.com Internet Source	<1%
10	www.neuralpress.org Internet Source	<1%
11	ar.iiarjournals.org Internet Source	<1%