AI in Lung Health: Advanced Automated Solutions for Lung Cancer Diagnosis and Prognosis Using Multimodality of Medical Data

Goram Alshmrani

School of Computing and Communications

Lancaster University

A thesis submitted in partial fulfillment for the degree of

Doctor of Philosophy

January 2026

"This thesis is devoted to my beloved family. It would not have been feasible to finish my PhD studies without their unending love and encouragement. I adore my family and am grateful for all they have done to support me."

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, in whole or in part, either for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes but excluding the bibliography.

Goram Alshmrani

Abstract

Lung nodules are areas of higher density in the lungs that can happen for a number of reasons, such as smoking or being exposed to airborne pollutants for a long time. It is essential to find and classify tumors on Computed Tomography (CT) scans as soon as possible so that lung diseases can be diagnosed and evaluated, as well as for planning and making treatment plans. For the diagnosis, it is essential to understand the difference between typical lung diseases like Tuberculosis, Pneumonia, and lung cancer, as all the diseases have similar symptoms initially. Initially, all the diseases have respiratory symptoms like cough, difficulty breathing, and chest pain. Pulmonary infiltrates or nodules can be observed in lung cancer, pneumonia, COVID-19, and tuberculosis, posing difficulty distinguishing between the diseases. Thus, this thesis has performed the classification of different types of diseases using X-rays by proposing a novel deep-learning framework for the multi-class classification of lung diseases, including lung cancer. The experimental results show that the Visual Geometry Group Network (VGG) 19 + Convolutional Neural Network (CNN) outperformed other existing work with 96.48% accuracy in the multi-classification of lung diseases.

Moreover, once lung tumor is detected, precise localization enables healthcare practitioners to ascertain the tumor's dimensions, which is crucial for staging and devising treatment strategies. Hence, this research proposes an advanced deep learning model called the Universal Network (U-net) to accurately segment lung tumors utilizing multiple types of imaging data, specifically CT and Positron Emission Tomography (PET) scans. The intricate structures of the suggested models, which incorporate several fusion approaches such as early fusion, late fusion, dense fusion, hyper-dense fusion, and hyper-dense VGG16 U-Net, are discussed in detail. The experimental results, particularly the performance of the hyper-dense VGG16 model, instill confidence in the proposed models, as it outperformed all other analyses, receiving a Dice score of 73%.

Survival analysis for lung cancer patients is a crucial aspect of treatment planning and outcome prediction. Therefore, in-depth stage classification using the TNM (Tumor, Node, metastases) staging system of Lung Cancer is of utmost importance. This thesis suggests an innovative method to classify the overall stage of non-small cell lung cancer (NSCLC) by employing multimodal data, including multi-view CT images and textual clinical information. A comparative analysis of Vision Transformer (ViT) and Convolutional Neural Network (CNN) architectures, evaluating both direct

classification and TNM-based approaches are proposed. The experimental results prove that the ViT-based direct model achieves superior accuracy 98.75%, improving accuracy by 8.75% over the TNM-based ViT model, while also reducing computational complexity by 66.67%. Similarly, the CNN-based direct model achieves 87% accuracy, outperforming the TNM-based CNN model by 7%, with a corresponding reduction in computational demands. The use of the proposed methods in real-time can help practitioners to detect lung cancer and predict the survival of the patient effectively.

Publications and Submissions

Alshmrani G, Ni Q, Jiang R, Pervaiz H, M. Elshennawy N. **A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images**, Alexandria Engineering Journal (AEJ), Elsevier. 2022 Nov 1;64:923-935. doi: 10.1016/j.aej.2022.10.053 (**300 citations till date**)

Alshmrani G, Ni Q, Jiang R. **Hyper-Dense_Lung_Seg: Multimodal-Fusion-Based Modified U-Net for Lung Tumour Segmentation Using Multimodality of CT-PET Scans**, Diagnostics, MDPI. 2023 Nov 20;13(22):3481. doi: 10.3390/diagnostics13223481

Alshmrani G, Ni Q, Jiang R. **Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction Using Vision Transformers**, IEEE Journal of Biomedical and Health Informatics (Submitted)

Acknowledgment

I wish to convey my profound appreciation to my supervisor, Professor Qiang Ni, for his unwavering assistance, forbearance, and mentorship for the entirety of my doctoral odyssey. The important input and support he provided have been crucial in the successful completion of this thesis.

I am deeply appreciative of my co-supervisor, Dr. Richard Jiang, for his exceptional guidance, valuable feedback, and consistent accessibility to engage in discussions regarding my study. The help provided by him has significantly enhanced the quality of my work.

I would like to express my heartfelt gratitude to my colleagues in the School of Computing and Communications (SCC) for their cooperation, enlightening conversations, and the exceptional atmosphere they created. Your companionship and assistance have been crucial in this endeavor.

I express my sincere gratitude to my family for their steadfast support and encouragement. I would like to express my gratitude to my family for their unwavering support, affection, and unwavering belief in my abilities. Dear wife, I express my gratitude for your unwavering patience, comprehension, and for serving as my steadfast support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial intelligence
ASPP	Asterisk Spatial Pyramid Pooling
AUC-ROC	Area under the receiver operating characteristic curve
CNN	Convolutional Neural Networks
CT	Computed Tomography
DL	Deep learning
DRENet	Detailed relation extraction neural network
GeLU	Gaussian Error Linear Unit
GNN	Graph neural network
GSA	Gravitational search
LDA	Linear Discriminate Analysis
LDM	Local Diagonal Masking
LIDC	Lung Image Database Consortium image collection
LN	Layer-Norm
ML	Machine learning
MLP	Multi-layer Perceptron
MRI	Magnetic Resonance Imaging
MSA	Multi-Head Self-Attention
NBIA	National Biomedical Imaging Archive
NSCLC	Non-small-cell lung cancer
ODNN	Optimal Deep Neural Network
PET	Positron emission tomography
ReLU	Rectified Linear Unit
ResNet	Deep residual networks
ROI	Region of interest
RSNA	Radiological Society of North America
RT-PCR	Reverse transcription-polymerase chain reaction
SCLC	Small-cell lung cancer
SIRM	Italian Society of Medical and Interventional Radiology

SVM	Support Vector Machines
TB	Tuberculosis
TCIA	Cancer Imaging Archive
TNM	Tumor, Node, metastases
ViTs	Vision Transformers
WHO	World Health Organization

Chapter 1

1. Introduction

Cancer is a highly fatal and the most challenging disease ever documented in human history. The cure for cancer remains elusive as those afflicted with the condition often become aware of it during advanced stages. Detecting it in its early stages is challenging, and the majority of cancer-related deaths are attributed to lung cancer. Consequently, extensive research has been undertaken to create a system capable of identifying lung cancer from CT scan images [1]. Preventing cancer is difficult due to the manifestation of symptoms at advanced stages, making recovery hard. Additionally, certain lung disorders have symptoms that closely resemble those of lung cancer, leading to the potential misidentification of these diseases as lung cancer in medical images [2].

The American Cancer Statistics 2023 study reveals that lung cancer ranks as the second most prevalent cancer among both males (12%) and females (13%). However, it also has the highest fatality rate of 21% across both genders. Each day, lung cancer claims the lives of around 350 individuals, a statistic that is nearly 2.5 times higher than the number of deaths caused by the second most common cause of cancer-related deaths, which is colon and rectal cancer [3]. There are two primary subcategories of lung cancer: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). Lung cancer typically arises from a combination of variables, such as cigarette use, exposure to dangerous particles in the atmosphere, genetic predisposition, old age, and other unidentified causes. Symptomatic signs of lung cancer include yellowing of fingers, stress, persistent disease, fatigue, reactions to allergens, wheezing, loud breathing, hemoptysis, respiratory difficulties, bone pain, headaches, dysphagia, and chest discomfort [4].

Deep learning (DL) and machine learning (ML) algorithms yield cutting-edge outcomes in various domains, such as object detection, classification, and semantic segmentation [5]. They significantly impact bioinformatics, particularly in cancer detection [6]. Recent advancements in DL technology have facilitated the autonomous identification of graphic components by CAD systems. Consequently, numerous medical image-processing approaches have been effectively implemented [7].

The capacity of deep learning to process intricate, multi-dimensional data and identify significant characteristics has seamlessly aligned with medical imaging. Conventional image analysis techniques frequently depended on features and rule-based algorithms, which restricted their capacity to fully utilize the intricate information in medical images. On the other hand, deep learning acquires the ability to autonomously identify significant characteristics from unprocessed data, enabling more precise and resilient picture analysis. DL has become an exciting approach to diagnosing lung cancer. It can make cancer detection much more accurate and quicker for many types of cancer [8]. The transformation is notably noticeable within medical imaging, wherein pathology and radiology images function as an indispensable diagnostic medium.

1.1. Research Problem

Early recognition is crucial in the treatment of cancer. Early detection of lung cancer frequently enables the utilization of more efficient and less intrusive treatment alternatives, thereby substantially enhancing patient outcomes. On the other hand, a diagnosis that is postponed can result in lung cancers that have progressed to a more advanced stage, making them harder to treat [9]. Incorporating deep learning methods into the analysis of medical images has received significant interest in recent years. Moreover, deep learning can utilize multimodal data to achieve a more thorough evaluation of lung cancer. By combining medical imaging, genetics, and clinical data, a comprehensive understanding of a patient's health can be achieved, enabling more accurate and tailored diagnostic and treatment strategies [10].

Various imaging modalities, including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and PET scans, offer distinct and supplementary insights into lung tissue composition, operation, and metabolic behavior. Integrating various modalities can achieve a more thorough comprehension of the disease. Moreover, including clinical data, encompassing patient history, demographics, and biomarkers data, offers contextual details that can enhance a comprehensive evaluation of a patient's well-being. A thorough understanding of the patient's overall condition is crucial for precise diagnosis and formulation of treatment strategies [11].

Despite advancements in medical imaging and automated systems, several significant challenges remain that hinder the effective identification, localization, and staging of lung cancer. These gaps are evident in current diagnostic practices and automated systems that rely on deep learning and other machine learning methods.

- **Symptom Overlap and Misdiagnosis**: Many lung conditions, such as pneumonia, COVID-19, and tuberculosis, share similar respiratory symptoms with lung cancer, creating significant diagnostic challenges. Chest X-rays, often the first line of investigation due to their widespread availability, are commonly used for preliminary diagnosis. However, traditional automated systems for X-ray analysis usually struggle to distinguish between these conditions. Existing automated systems have proven effective in binary classification tasks (e.g., detecting the presence or absence of pneumonia). However, when handling multi-class classification involving a broader range of diseases with overlapping symptoms, including lung cancer, these systems leave a gap in early lung cancer detection. For instance, specific patterns in X-ray images, such as patchy lung opacities observed in pneumonia or tuberculosis, might look very similar to early-stage lung cancer. This overlap frequently results in misdiagnosis, which leads to delayed or ineffective therapies and has a detrimental influence on patient outcomes.
- **Tumor Localization Challenges:** Currently, automated algorithms for lung tumor localization rely mainly on CT scans, which give rich anatomical data. However, CT alone has significant limitations since it fails to capture critical functional information on tumor metabolism. While MRI offers high-resolution imaging and improved soft tissue contrast, its application in lung tumor imaging is limited due to respiratory motion errors and poor sensitivity to air-filled structures. These difficulties impede precise tumor localization and may result in inadequate assessments. Other modalities, such as ultrasonography, confront similar issues, with low penetration and uneven anatomical information for deep-seated lung cancers, complicating segmentation attempts. PET imaging, when combined with CT, provides critical metabolic insights that improve our understanding of tumor behavior and activity. This combination can improve treatment planning and staging by allowing for a more thorough examination of malignancies. Some existing systems have begun to incorporate multimodal techniques by integrating CT and PET imaging, utilizing their advantages to improve tumor segmentation. However, there is still room for improvement in tumor localization accuracy through more accurate fusion approaches that combine the characteristics of both imaging modalities.

- **Inadequate Staging Classification:** Accurate staging of lung cancer utilizing the Tumor, Node, Metastases (TNM) approach is critical for selecting effective treatment options, estimating patient outcomes, and forming prognoses. Existing automated systems frequently rely primarily on imaging data for TNM categorization, overlooking the vital role that demographic and clinical data might play in staging accuracy. While some techniques may be comparably sensitive in determining tumor size (T), they often overlook other important factors such as lymph node involvement (N) or distant metastases (M). Inaccurate general stage forecasts resulting from the lack of thorough staging models could lead to uneven patient treatment approaches and worse-than-ideal results. The preponderance of the T stage in current models limits their capacity to totally reflect the course of lung cancer, so influencing the prognosis dependability and treatment efficacy.

These inadequacies in current diagnostic procedures highlight the urgent need for novel concepts that use deep learning to improve the accuracy and dependability of lung cancer detection, tumor segmentation, and stage classification. By resolving these problems using multimodal data fusion, the model's diagnosis precision can be improved, enabling more precise tumor localization and establishing solid, data-driven frameworks for lung cancer staging.

1.2. Research Motivation

The most prominent cause of cancer-related fatalities globally is still lung cancer, and proper diagnosis and early discovery significantly affect the survival chances. This research is driven by several motivations in both the critical clinical need for improvement and the technological possibilities presented by contemporary deep learning methods.

- **Enhancing Patient Outcomes by Early Detection:** Early detection is essential for lung cancer survival rates. However, because lung cancer symptoms overlap with those of other respiratory disorders, current diagnostic instruments sometimes miss the early stages of lung cancer. This study is driven primarily by the possibility of significantly enhancing patient outcomes via early detection of lung cancer. We want to create models that can increase the accuracy of early detection by using advanced artificial intelligence technology, therefore enabling quicker interventions and higher survival rates.

- **Use of Multimodality imaging:** Accurate diagnosis and treatment planning in lung cancer depend not only on spotting the disease but also on correctly pinpointing tumors, hence bridging the gap between imaging modalities. Conventional imaging techniques, including CT and PET scans, provide different but insufficient perspectives on lung cancers. PET scans offer vital information regarding metabolic activity, but CT scans concentrate on the anatomical aspects. The possibility of closing the difference between these two imaging modalities primarily drives this research. This work attempts to provide a complete knowledge of lung cancers by proposing improved deep-learning models that combine data from both CT and PET scans, therefore enabling more exact localization and focused treatment.
- **Advancing Precision Medicine in Cancer Staging:** Predicting outcomes for lung cancer patients and choosing therapy courses depend much on the TNM staging system. Though crucial, current automated staging systems may ignore thorough data integration, including demographic information and imaging. Incorporating more data-rich models will help to develop precision medicine, thereby driving this study. Using deep learning approaches, it is possible to offer more complex and personalized staging assessments, therefore facilitating more informed treatment decisions and more accurate prognosis predictions.
- **Harnessing the Power of AI for Clinical Use:** AI and DL are becoming more valuable in healthcare, potentially changing how doctors practice their profession. The therapeutic use of AI has not yet realized its full potential despite its impressive performance in other domains, such as picture recognition and natural language processing. The primary goal of this research is to discover ways to use artificial intelligence's ability to change things to assist doctors with their daily tasks, particularly in the detection and treatment of lung cancer. This study aims to push the boundaries of AI in healthcare by developing models that are robust, scalable, clinically relevant, and simple to integrate into existing processes. The primary motivation for this research is the urgent need to improve lung cancer tumor location, diagnostic accuracy, and staging. With the ultimate goal of improving lung health care and patient outcomes, we are looking for innovative solutions to these problems using advanced deep learning methods.

1.3. Thesis Aims and Objectives

Developing sophisticated deep learning-based models to enhance the diagnosis, localization, and staging of lung cancer is the goal of this work. This work aims to improve the accuracy of early detection, enable more exact tumor localization, and provide solid predictions for TNM-based cancer staging by using multimodal imaging data and integrating clinical demographic information, thus contributing to more efficient patient management and improved outcomes.

The following particular goals help one to reach this aim:

- Using chest X-ray images, create and use a deep learning model to precisely classify six different lung diseases—including pneumonia, TB, COVID-19, and lung cancer. This goal seeks to lower the chance of misdiagnosis by better-automated analysis, therefore addressing the difficulties presented by symptom overlap among these diseases.
- To investigate how PET and CT imaging modalities might be combined to improve tumor localization. This aim will be to create algorithms that combine metabolic information from PET scans with anatomical data from CT scans, enabling a more complete knowledge of tumor traits and behavior. Correct therapy planning and tumor activity evaluation depends on this integration.
- A complete staging classification system based on the TNM (tumor, node, metastases) staging system predicts the general stage of lung cancer by incorporating pertinent demographic and clinical data to increase prognosis accuracy and guide treatment decisions.
- To perform thorough validation of the created models utilizing many datasets and evaluation criteria, including accuracy, sensitivity, specificity, and F1-score. This purpose guarantees dependability and generalizability among various patient populations. Furthermore, the clinical usability and scalability of the suggested models will be evaluated by comparing them with current diagnosis and staging systems.
- By working with medical practitioners, one can evaluate the clinical applicability of the suggested approaches by ensuring that the frameworks created fit clinical procedures and handle practical diagnostic requirements.

1.4. Research Contributions

The significant contributions of the research are given below.

- A deep learning classification model is proposed to identify and differentiate six lung diseases: pneumonia, lung cancer, tuberculosis (TB), lung opacity, COVID-19, and normal cases. Due to overlapping symptoms that result in similar patterns on X-ray images —such as ground-glass opacity and nodular formations—this study utilizes a transfer learning approach with a pre-trained VGG19 model enhanced by three additional convolutional neural network (CNN) layers. Trained on a large and diverse dataset of X-ray images, this model represents one of the first attempts to classify multiple lung diseases effectively using X-ray imaging at the time this investigation took place.
- An innovative architecture is proposed for lung cancer segmentation using multimodal imaging from PET and CT scans. This model leverages both PET and CT scans, providing critical metabolic and anatomical information. It utilizes a modified U-Net architecture with various fusion strategies: early fusion, late fusion, dense fusion, hyper-dense fusion, and hyperdense VGG-16 U-Net. By incorporating dense connections for each modality, the model improves information flow and feature representation over a basic encoder. The hyperdense connections enhance integration between modalities by capturing the complementary details from each modality. Using VGG-16 for deep feature extraction boosts overall segmentation performance. These modifications collectively enhance segmentation performance compared to other strategies.
- An advanced deep learning (DL) approach named Vision Transformer (ViT) is developed for overall stage prediction of lung cancer based on the TNM staging system, integrating both imaging data and demographic information. Unlike existing approaches that classify TNM stages individually before making a final prediction, our model proposes a direct overall stage prediction using ViT. Additionally, we introduce a multi-view approach along with demographic data to enhance prediction accuracy. To validate the effectiveness of our proposed model, we also apply a CNN model for this task and compare the results. To the best of our knowledge, this is the first work applying the ViT model for overall stage prediction of lung cancer based on the TNM system.

Finally, to validate our models across all contributions, we employed various evaluation metrics and loss functions. For the classification model, we assessed metrics such as accuracy, precision, recall, and F1-score. The segmentation model utilized the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) metrics, optimized with multiple loss functions—including Dice loss, binary cross-entropy loss, and focal loss—to compare their performance effectively. In the Vision Transformer model for overall stage prediction, accuracy was used, with cross-entropy loss applied during training. These methods ensured a comprehensive evaluation of our contributions to enhancing lung disease diagnosis and prognosis assessment.

1.5. Thesis Structure

The research conducted in this thesis revolves around three main areas: Multimodal medical data based on an advanced DL framework for multiclass lung disease classification, lung tumor segmentation, and overall stage prediction of lung cancer. The thesis is structured into six distinct chapters, as indicated below.

Chapter 2 comprehensively analyzes the existing literature on identifying and categorizing lung diseases using medical imaging techniques, i.e., Chest X-ray images. The second section of the literature examines the segmentation techniques employed with multi-modality pictures, specifically CT and PET scans. The CT-feature extraction, PET-feature extraction, and feature fusion procedures are thoroughly explained. The third section of the literature review examines research that explicitly investigates lung cancer stage classification and overall stage prediction based on the TNM stage system using both CT imaging and clinical data. The literature discusses many methodologies that utilize machine learning (ML), deep learning (DL), transfer learning, Vision Transformer (ViT) approaches, applied pre-processing processes, dataset robustness, and limitations.

Chapter 3 proposes a novel DL framework for the multi-class classification of Normal, Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity, and the latest addition, COVID-19, from the chest X-ray images. A detailed description of the tremendous datasets from various resources used and the pre-processing steps performed on the dataset are also discussed. The pseudo-code for the proposed algorithm is given along with the architecture of the proposed model. The mathematical notations for the performance metrics used are also explained. Furthermore, the results obtained and the accuracy and loss graphs are presented and discussed

in detail. Finally, the chapter ends with the conclusion and future scope for classifying chest X-ray images into different lung diseases, including lung cancer.

Chapter 4 proposes using an advanced deep learning model called U-net to accurately separate lung tumors utilizing multiple types of imaging data, specifically CT and PET scans. The intricate structures of the suggested models, which incorporate several fusion approaches such as early fusion, late fusion, dense fusion, hyper-dense fusion, and hyper-dense VGG16 U-Net, are discussed. The merits and disadvantages of each model are emphasized. The findings from all the models are compared with the benchmark models. The several loss functions employed for model training are examined, and their mathematical expressions are provided. Each model's anticipated segmented image is compared to the corresponding ground truth.

Chapter 5 suggests an innovative method to predict the overall stage of non-small cell lung cancer (NSCLC) by employing sophisticated deep learning methods, including Vision Transformers, using the multi-input dataset, including radiological and clinical data. The chapter discusses the conventional TNM staging approach, the clinical parameters affecting stage prediction, and the rationale for using Transformers. The thorough comprehension of lung cancer staging, novel strategies, and explanation of the benefits of Vision Transformers in this crucial medical application are discussed. The ViT-based architecture for overall stage prediction is presented and discussed in detail. The findings for the TNM staging classifier are discussed and compared with the overall stage prediction-based approach and the benchmarked models.

Chapter 6 presents the concluding section and summarizes this research by evaluating the effectiveness of the offered techniques and analyzing their practical implications for Lung Cancer Diagnosis and Prognosis assessment based on overall stage prediction using multi-modality imaging and clinical data. It also highlights the limitations and potential future work that enhances patient outcomes, facilitates and expedites the diagnosis process, and saves time for decision-making.

Chapter 2

2. Literature Review

2.1. Introduction

Manually interpreting medical images can be characterized by its time-consuming nature, susceptibility to human error, and susceptibility to intra-observer and inter-observer variability. In recent years, using artificial intelligence (AI) techniques, namely deep learning models, has proven crucial in image processing automation. This development has garnered significant interest within the field of medical imaging. Medical imaging has been significantly transformed by the widespread adoption of CNNs, which have proven to be highly effective at capturing intricate patterns and facilitating the automated identification of diseases and anomalies. Recent research has exhibited noteworthy advancements in lung cancer detection, segmentation, and classification [12, 13].

The literature is divided into six sections. Section 2.2. discusses the background. Section 2.3. provides the literature review of multi-class lung disease classification using DL. Section 2.4. describes the literature for Lung Tumor Segmentation using Multimodality of Computed Tomography (CT) -Positron Emission Tomography (PET) Scans. Section, 2.5. provides detailed literature about Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction Using Vision Transformers. Section 2.6 summarizes the research gaps. Finally, the chapter concludes with a summary presented in section 2.7.

2.2. Background

Numerous studies have discussed the efficacy of computer-aided diagnoses in the medical context, based on collaboration between medical researchers and computer scientists. Certain computer-aided diagnosis systems in medicine may be classified as expert systems since they seek to replicate the decisions of medical professionals. In addition, computer-aided detection systems in medicine can process complicated and large clinical data [14, 15]. Computer-aided detection systems can also assist clinicians to gain new insights into data and apply the knowledge to improve diagnostic accuracy. As a result, the systems are considered intelligent systems since they employ a process of feedback to continuously enhance their performances.

Large clinical data is complicated to analyze. Intelligent Computer-aided diagnosis systems using data mining, artificial intelligence (AI), and deep learning methodologies are beneficial in diagnosing an array of illnesses and medical disorders.

In the last century, researchers have accumulated substantial knowledge regarding human anatomy and physiology. In recent years, chest X-rays (CXR), ultrasounds, and MRI have played vital roles in enhancing the accurate diagnosis of human diseases. Significant improvements in healthcare and medical research have helped people to improve their quality of life as new technologies have facilitated the accurate diagnoses of patients' ailments and diseases. In the last few decades, medical experts have faced challenges in conducting an accurate diagnosis of diseases, which has compounded unnecessary healthcare and malpractice claims for both doctors and patients. Machine learning, deep learning, and statistical analysis are effective tools for computer-aided diagnosis. These tools are used in solving difficult computer vision tasks in medical imaging, such as segmenting lungs, classifying lung diseases, and so on. With recent developments in deep learning, machines can perform equally or better than humans in a wide range of activities. For example, deep learning can be used to calculate treatment outcomes, such as cancer therapy. With huge-labeled datasets and deep learning-based approaches, promising findings are developed in the categorization of thoracic disorders using a CXR modality. In addition, machine learning is the model that can learn and make decisions based on a vast number of input data sets. Artificial intelligence performs activities that require human intellect, such as voice recognition, translation, and the ability to analyze colors and shapes by evaluating incoming data and making predictions. A combination of machine learning algorithms, known as deep learning, has demonstrated remarkable success in various sectors, particularly in the healthcare sector [14, 15]. Deep learning models can accurately predict and categorize numerous diseases, such as tuberculosis (TB), lung cancer, pneumonia, and currently COVID-19, using images without human intervention. As the network becomes larger, data representation becomes deeper, making deep learning to be more effective, contrary to classical machine learning. Consequently, the model automatically collects characteristics and generates more accurate outcomes. Since the models use a combination of non-linear functions rather than linear functions, deep learning algorithms are more accurate than typical machine learning methods.

In late 2019, the coronavirus (COVID-19) pandemic invaded the planet, leading to an alarming scenario. The virus was first formally discovered in Wuhan, China, in December 2019, and the World Health Organization (WHO) designated it as an emergency health problem at the

beginning of 2020. By March 2020, WHO classified it as a pandemic [16]. The Coronavirus causes pneumonia, persistent cough, high fever, and fatigue, among other symptoms. Reverse transcription-polymerase chain reaction (RT-PCR) is employed to identify positive cases of the virus. However, it can take several hours, even days, to generate results using this form of diagnosis. RT-PCRs are both time-consuming and expensive. Subsequently, experts are facing significant challenges in developing alternatives via detection technologies. AI is being used to automate the diagnosis of many diseases today, and AI has been proven to achieve superior performance during automatic image categorization using various machine learning algorithms. The detection is based on the image processing and the classification of the features extracted from the CXR or CT, as shown in Figure 2.1. Furthermore, machine learning specifies models that have the capability of learning and making decisions based on a massive input of data samples.

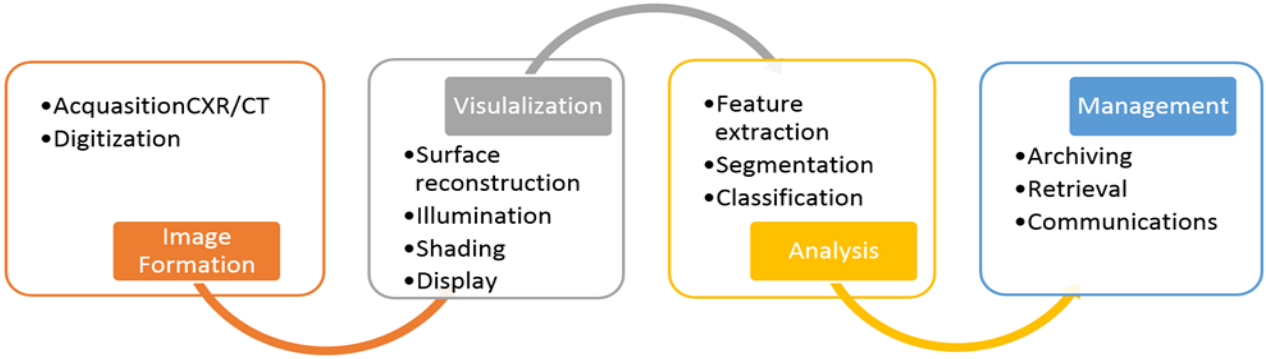

Figure 2.1. Image processing-based classification model

In the context of deep learning, the extraction and classification of features from images is the primary goal. Deep learning has been a huge success in a wide range of industries, including healthcare [15]. In addition, deep learning can develop models that can accurately predict and diagnose illnesses using images. It has been effective in diagnosing TB [17–22], pneumonia [23–30], lung cancer [31–35], and COVID-19 diagnosis without the need for human expertise. Unlike traditional machine learning, the fundamental reason behind using deep learning techniques is its ability to build the model of input as the size of the network deeply grows. Because of this, the model automatically gathers data and generates findings that are more accurate. Deep learning models, in contrast to typical machine learning algorithms, describe features using a sequence of non-linear functions that are incorporated to optimize the accuracy of the utilized model.

In 2020, lung cancer ranked as the second most prevalent cancer, accounting for around 11.4% of all newly identified instances of cancer, with approximately 2.2 million individuals affected. Furthermore, it was the primary cause of cancer-related mortality, responsible for approximately 18.0% of the deaths caused by cancer globally, resulting in approximately 1.8 million fatalities. Loss of appetite, exhaustion, chronic coughing, and chest pain are among the symptoms of lung cancer, which can cause unimaginable anguish for the sufferer [36].

Segmentation of lung tumors, treatment evaluation, and tumor stage classification have become significantly more accessible with the advent of PET/CT scans. Moreover, the molecular characteristics and anatomic aberrations of the target lesion can be observed with PET/CT. PET imaging technique does not involve cutting or surgery. By detecting illness markers earlier, PET allows for earlier diagnosis than imaging modalities like MRI and CT [37]. Their metabolic processes can be analyzed for their physiological function and biochemical features by studying particular organs and tissues. PET can detect molecular and cellular levels of tissue metabolism.

However, multimodality imaging technology, such as PET-CT scanners, has simultaneously made it possible to record functional and anatomical information [38]. It is a rigorous and time-consuming process for oncologists, radiologists, and pulmonologists' to manually segment the lesions and tumors, leading to delays in therapy and decreased survival rates, particularly in clinics with insufficient resources. In addition, specialist knowledge and clinical experience are necessary for high-quality manual localization and segmentation. Because of this, computer-aided diagnostic (CAD) systems [39] were developed to replace radiologists' manual viewing of lung cancer. Combining lung segmentation approaches with radiologists' knowledge can reduce the burden on radiologists and boost their productivity and accuracy. Many recent advancements in image segmentation have allowed for more precise and effective treatment and diagnosis. Thresholding, Atlas, and Region Growing are some examples of classic automatic segmentation methods. These approaches use shallow qualities of an image, such as grayscale, texture, gradient, and many more, to segment the object [40]. However, conventional segmentation techniques have difficulty distinguishing between tumors and surrounding healthy tissue because their intensity distributions are similar. In addition, these tasks typically involve manual processes and are characterized by a significant investment of time. Moreover, they are subject to substantial heterogeneity across operators.

Furthermore, the complexity of the background in CT images consistently provides quite different information when comparing PET and CT scans. As a result of these constraints, deep learning-based algorithms have proven to be superior in au-to-segmenting medical images [41].

Deep learning (DL) models automatically extract features and apply the learned high-dimensional abstractions for performing segmentation. The effectiveness of fully convolutional networks (FCN) for semantic segmentation is promising [42]. In an FCN, the fully connected layer is replaced by a convolutional layer. This comprehensive framework has served as the foundation for subsequent studies of semantic segmentation of medical images. Medical image segmentation commonly uses U-Net [43], built on the FCN architecture. Using skip-connection architecture, each layer's down-sampled features are joined with their up-sampled counterparts. This mechanism is similar to an encoder-decoder, but it is more effective and doesn't require a lot of disk space. FCN-based networks, such as U-Net, have surpassed manual or semi-automatic segmentation methods since the emergence of big data methods.

The U-Net architecture is a convolutional neural network (CNN) primarily used to recognize image patterns. U-Net semantic segmentation relies extensively on the categorization of image pixels. Segmenting lung tumors can be reduced to a foreground/background pixel binary classification problem. The down-sampling and up-sampling module is responsible for the U-Net architecture. The surface layer is where localization information is learned, but the down-sampling procedure, also known as the pooling procedure, may improve the volume of context data the network learns [43].

The VGG model investigates the effect of convolutional network depth on recognition accuracy in a large-scale setting. The main contribution is a thorough evaluation of increasing-depth networks using an architecture with 3x3 convolution filters, which shows a significant improvement. The VGG model, a kind of CNN, was created to improve model performance when more layers are added. The VGG model takes 224x224 color images as its primary input and feeds them via a sequence of convolutional layers with filter sizes of 3x3 and 1x1 with the stride of 1 and valid padding, as well as Max-pooling with 2x2 with the stride of 2. Finally, a three-layer network is developed with a soft-max activation function and 4096 neurons in the first two layers, followed by 1000 neurons in the last layer. VGG [44] presents the two primary models, VGG16 and VGG19. In comparison to the VGG-19 network, which has 19 layers of typical convolutional networks, the VGG-16 network [44] only has 16, each of which has 33

filters and strides of 1. Each of the five blocks is separated from the next by a max-pooling layer. There are three interconnected layers on top of the blocks.

Based on a report published by the American Cancer Society in 2022, it was projected that there would be around 1.9 million newly diagnosed cases of lung cancer and a mortality rate exceeding 600,000 in the United States. Out of the fatalities, lung cancer was responsible for an average daily of 350 deaths [45]. The delayed identification of lung cancer nodules and subsequent management of lung cancer patients is associated with elevated mortality rates. Typically, those diagnosed with lung cancer face a survival rate ranging from 10% to 16% during a five-year period. Nevertheless, the survival rate has the potential to increase to 70% in the event of an early diagnosis of lung cancer [46]. Moreover, people whose lung cancer has progressed to a later stage have a much lower chance of survival. Prioritizing lung cancer in its first stages is imperative to enhance the likelihood of patient survival. In recent decades, medical imaging methods have assumed a progressively significant role in the screening, prognosis, survival estimation, and early identification of lung cancer, ultimately contributing to treatment efficacy and preventative strategies [47]. X-rays are frequently used as a preliminary screening for lung cancer as the X-ray modality can reveal lung abnormalities, but a CT scan is now considered the gold standard. Nevertheless, due to the two-dimensional nature of X-rays, these images cannot precisely determine the specific location of any abnormalities. In contrast, CT images are three-dimensional representations that offer comprehensive insights into cancer, encompassing critical aspects such as tumor position, morphology, and characteristics [48].

Lung cancer is classified into two distinct categories, with differentiation depending on the microscopic characteristics of cancerous cells. There are two primary forms of lung cancer: small cell type lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Among these, NSCLC is the predominant form, representing approximately 80%–85% of lung cancer diagnoses [49]. Both categories are extensions of the TNM system [50]. Medical professionals primarily categorize patients into limited and extensive stages in treatment planning for small cell lung cancer (SCLC). During the initial phase, neoplasms are localized inside a specific thoracic area, such as a solitary pulmonary lesion or a lymph node on the same side. On the other hand, the advanced stage of cancer is characterized by the dissemination of malignant cells to both thoracic areas and multiple other anatomical sites via metastasis. SCLC is characterized by its quick growth and aggressive metastasis, often leading to a fatal outcome within a few weeks. Hence, medical professionals must make critical therapy determinations

expeditiously. Nevertheless, it is not always true that physicians exclusively adhere to a binary classification of small cell lung cancer (SCLC) stages. Occasionally, medical professionals may also opt for TNM staging as an alternative to assessing SCLC. The determination of treatment varies among individual patients. The medical practitioners use to choose the TNM staging instead of evaluating SCLC. However, the selection of treatment options varies with respect to individual patients.

2.3. Multi-Class Lung Disease Classification from Chest X-Ray Using DL

In most countries, chest computed tomography (CT) and X-ray pictures are widely utilized as a feasible option for identification of COVID-19. However, COVID-19 identification is a complex process that requires clinical imaging of patients [51–57]. Lung cancer represents a major source of mortality in humans. The immediate diagnosis could improve human survival [31–34]. Applying machine learning and image processing has presented considerable promise for lung cancer diagnosis. This section discusses an exhaustive evaluation of deep learning models for TB, COVID-19, lung cancer, and pneumonia. Transfer learning methods, such as VGG-16, ResNet-50, and InceptionV3, to clinical pictures of lung illnesses and COVID-19 has offered promising results [18]. It is discovered that pneumonia is among the significant symptoms of COVID-19. Transfer learning helps discover that the same virus causes pneumonia and COVID-19. The following subsections describe the literature for COVID-19, lung cancer, pneumonia, and TB detection.

2.3.1. Covid19 Detection

A study demonstrates that the information obtained by a model trained to detect viral pneumonia may be applied to identify COVID-19 [56]. As a result, Haralick features can be used to facilitate feature extraction. This approach involves statistical analyses that focus on a specific area of COVID-19 diagnosis. In comparison to the traditional classifications, transfer learning has consistently proven to offer statistically significant outcomes [56]. Some studies developed and analyzed a fully automated COVID-19 detection framework utilizing CTX. To diagnose COVID-19, the visual features were extracted from volumetric chest CT images using COVID-19 neural network approach. The outcomes show that the approach has outperformed the existing work. Pre-trained models-based CNN architecture such as Inception-ResNetV2, ResNet152, ResNet50, InceptionV3, and ResNet101 was used in related work to identify COVID-19 pneumonia based on the CXR images. Among the existing models, the ResNet50

exhibited the most accurate classification outcomes [53]. The comparison and modeling were based on CT images of 101 pneumonia, 88 COVID-19 patients, and 86 healthy cases from two areas in China. A detailed relation extraction neural network (DRENet) learning-based CT diagnostic algorithm identified COVID-19 patients. The model correctly distinguished between COVID-19 patients with a recall of 0.93, AUC of 0.99, and accuracy of 0.96. The research showed that deep learning based on CT scans may help to detect COVID-19 patients and automatically identify possible abnormal changes. Another study categorized COVID-19 CXR images by applying modified MobileNet and a ResNet architecture. With this approach, characteristics from multiple CNN layers are dynamically combined to overcome the gradient vanishing problem. The proposed approaches outperform the current methods by 99.3% on the CT image dataset and by 99.6 % on the CXR [54].

Some studies developed a model to distinguish between critical and severe COVID-19 instances using deep learning characteristics and radionics based on D-Resnet [52, 58]. These authors studied 217 individuals in three Chinese hospitals, 82 with extreme severity and 135 with serious disease. The patients were grouped into two (174 patients) for training and (43 patients) for testing. The authors created a 3-dimensional deep learning network using the clipped segments and multivariable logistic regression to integrate relevant radiomics characteristics and deep learning scores. To test the robustness of their methods, they used stratified analysis, cross-validation, decision curve analysis, and survival analysis. An AUC of 0.909 distinguishes between critical patients in the test and training groups [58]. Another study applied InceptionV3, NASNet, Xception, DenseNet, MobileNet, VGGNet, InceptionResNetV2, and ResNet for classifying the COVID-19, which was tested on the mixed dataset of CXR and CT images. DenseNet121 offered the best performance with an accuracy of 99% [52]

Image segmentation is used to categorize chest CTX into pneumonia, COVID-19, and normal illnesses using four CNN base learners, a modified stack ensemble model, and Naive Bayes as the meta-learner in one research. For COVID-19, pneumonia, and normal classes, the suggested technique beats current techniques by .9867 on standard datasets and 0.98 Kappa on the same datasets [59] based on CT scans. By reducing manually labeled CT images, the suggested technique may accurately detect COVID-19 infections and rule out the case of COVID-19. Based on the positive qualitative and quantitative results, the recommended approach is widely used in large-scale clinical trials [60]. The convolution neural networks are effective in converting 360 X-ray and CT scan pictures into a categorization on a binary class

pneumonia-based translation of decision tree, Inception V2, and VGG-19 models. Compared to decision tree (60%) models and Inception V2 (78%), the fine-tuned version VGG-19 (91%) exhibits the greatest increase in training and validation precision [60].

The GSA-DenseNet121-COVID-19 is a unique mixed CNN architecture that utilizes DenseNet121 and the optimization technique of gravitational search (GSA). The DenseNet121-COVID-19 could identify COVID-19 better than other DenseNet121, which could only diagnose 94% of the test set. The suggested method was contrasted with an Inception-v3 CNN architecture and manual analysis when computing hyperparameter estimates. The GSA-DenseNet121-COVID-19 outclassed the comparison technique, which could only categorize 95% of the test set samples [61].

EfficientNet-based pre-trained models were lowered using kernel principal component analysis. Then, multiple retrieved features were merged using a feature fusion technique. Finally, stacked ensemble meta-classifiers were used to classify the model into two stages. Predictions were made in the first step using a support vector machine (SVM) and a random forest, which were then pooled and fed into the second stage. Next, a logistic regression classifier divides the X-ray and CT data into two classes (COVID and NON-COVID). The model's performance was compared to other CNN-based pre-trained models. The new model outperforms previous approaches and may be used by clinicians for point-of-care diagnosis [62]. In a comparable work, researchers used ResNet32 and the deep transfer learning technique to categorize COVID-19-infected patients, and the results were published. Comparing the COVID-19 classifier to earlier supervised learning models, experimental data demonstrated that it delivered superior outcomes when compared to previous learning models [63].

A cutting-edge attention-based deep learning model with VGG-16 and a fine-tuned classification process was designed using a unique deep learning model that uses a convolution layer of the VGG-16 models for COVID-19. The experimental analysis shows steady and promising performance after comparing the suggested approach to the existing models [64]. The integrated stacking deep convolutional network using pre-trained models like ResNet101 and XceptionV3 was applied for InstaCovNet-19. The accuracy of .99 for three classes (Normal, Pneumonia, COVID-19) and .9953 for two classes (COVID, non-COVID) is achieved. In ternary classification, the suggested model obtained 98% accuracy, whereas binary classification achieved 100% precision and 98% recall [64].

The CNN is used to implement binary and multiclass classification. The model was trained on 3877 CT and X-ray pictures, of which 1917 were of COVID-19-affected people. The binary classifier achieved a 99.64% accuracy and exhibited a 99.58% recall, a 99.56% precision, a 99.59% F1 score, and a 100% ROC. The model was trained with 6077 photographs. A total of 1917 patients were of Covid-19 infected patients, 1960 healthy people, and 2200 pneumonia patients. The suggested technique obtained 98.28% accuracy, 98.25% recall (or sensitivity), 98.22% precision, 98.23% F1-score, and 99.87% ROC for multiclass classification [65].

2.3.2. Lung Cancer Detection

The early detection of lung cancer increases survival chances from 14% to 49%. Although CT approaches are found to deliver more accuracy than X-rays, a conclusive diagnosis relies on many imaging modalities. An artificial DNN can spot lung cancer in CT images. Therefore, studies have proposed an adaptive boosting technique and a DenseNet to classify the lung image as normal or malignant. A total of 201 lung pictures have been included in the training dataset, with 85 percent of them being utilized for training and 15 percent being used for testing and classification. The proposed approach was shown to achieve a 90% accuracy in testing [34]. The MLP classifier offered a higher accuracy of 88.55% than the alternative classifiers, according to the outcome of the analysis of a study [33]. The CNN, DNN, and sparse auto-encoder deep neural networks were employed to identify lung cancer calcification. CT scans of benign and malignant lung nodules were classified using these networks. The Lung Image Database Consortium image collection (LIDC) database examined the networks where accuracy was 84.15%, sensitivity 83.96%, and specificity 84.32% [32]. CNN was the most accurate of the three networks. Another work applied Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA) to evaluate CT lung images that reduce the dimensionality of deep features. The ODNN is used with CT scans and optimized using the Gravitational Search Algorithm to classify lung cancer, thereby offering 96.2% sensitivity, 94.2% specificity, and 94.56% accuracy [31].

2.3.3. Pneumonia Detection

Since medical specialists face challenges in distinguishing between COVID-19 and pneumonia, one study utilized an artificial neural network, ensemble classifier, SVM, and KNN for categorization. However, a RNN with a LSTM has been proposed as a deep learning architecture to identify lung conditions. The outcomes of the experiments demonstrated the resilience and effectiveness of the suggested model [30]. Another work uses an ensemble of InceptionResNet_V2, ResNet50, and MobileNet_V2 for classifications. The outcomes

revealed that the ResNet50, MobileNet_V2, and InceptionResNet_V2 models provide an F1 score of 94.84%, which is higher than other models [29]. In addition, the CNN with pre-trained weights is utilized to categorize COVID-19, pneumonia, and healthy individuals using transfer learning techniques. Those who have active SARS-CoV-2 and pneumonia were accurately categorized in the dataset, which is one of the most important discoveries of that work [25]. Another study examined the potential of using machine learning to delineate and pinpoint pneumonia in CXR using RetinaNet and Mask R-CNN as an ensemble for the identification and localization of pneumonia, thereby achieving a recall of 0.793 for a large dataset [28]. For a variety of lung diseases, the transfer learning approach was used to capture images on CXR and CT. As COVID-19 resembles pneumonic viral lung illness, COVID-19 detection is challenging and relies on a thorough examination of a patient's clinical pictures. The goal is attained using a novel architecture trained to identify virus-related pneumonia for COVID-19 detection. When compared to traditional categories, the findings of transfer learning are strikingly different [27].

One study develops the CNN model from scratch to extract characteristics from an image of pneumonia infected person's chest X-ray and categorize it. This concept might alleviate some of the issues associated with dealing with medical images. It is difficult to obtain a significant number of pneumonia datasets for this classification assignment due to the limited availability of such data. Multiple data augmentation strategies were used to increase the accuracy of the training and validation classification of the proposed model. This has achieved a significant precision of 0.94814 in the validation phase [26]. The transfer learning system automatically differentiates between 3883 CXR pictures classified as exhibiting pneumonia and 1349 that are designated normal. As an initialization, the suggested technique makes use of weights pre-trained on ImageNet using the Xception Network. When compared to current approaches, the model is competitive in obtaining 0.84, 0.91, 0.99, and 0.97 for precision, recall, F1, and ROC, respectively [24]. In a separate study, researchers studied 180 X-ray images of persons who had been infected with COVID-19. The research attempted to employ the most successful systems, such as ResNet50V2 and Xception networks, to detect the virus. Overall, the suggested model achieved a 91.4% accuracy for all classes and a 99.50% accuracy for instances of COVID-19 [23].

2.3.4. Tuberculosis Detection

Using a CXR dataset from the National Library of Medicine Shenzhen No.3 Hospital, researchers developed a DCNN model to detect tuberculosis. This dataset was compared with

a non-TB-specific chest X-ray dataset of a different population. The DCNN offered an AUC of 0.9845 and 0.8502. The AUC of the supervised DCNN model in the CXR dataset, on the other hand, was much lower, at 0.7054, than in the other datasets. A total of 36.51% of aberrant radiographs in the CXR dataset associated with tuberculosis were predicted by the final DCNN model [17].

Another study combined ResNet and depth-ResNet to predict severity scores and analyze TB's likelihood. A depth-ResNet of 92.70% and ResNet-50 of 67.15% were produced for TB detection. The study used the overall severity probability, different likelihoods for high severity (1 to 3 scores), and low severity (4 and 5 scores), where scores of 1 to 5 were converted into the probabilities of 0.9, 0.7, 0.50, 0.30, and 0.2. A 75.88% and 85.29%, respectively, are the averaged accuracies for both approaches [18]. Other studies proposed three standard designs in the ensemble technique, namely AlexNet, GoogleNet, and ResNet. As a result, a new classifier for TB categorization has been developed from scratch. A combined dataset of publicly accessible standard datasets is used to train and test the suggested approach. Accuracy of 88% and the AUC of 0.93%, which is better than most existing approaches, are achieved [19].

The hierarchical feature extraction for abnormality detection method uses two levels of hierarchy to classify characteristics into healthy and unhealthy categories. Two levels of feature extraction are identified: level one is handmade geometrical feature extraction, and level two is typical statistical feature extraction and textural feature extraction from segmented lung fields. They were tested on 800 CXR images derived from two public datasets to verify their performance. AUC = 0.99 0.01 for Shenzhen and 0.95 0.06 for Montgomery, which illustrated that the two TB detection approaches offered a promising performance as compared to the existing techniques, as demonstrated by the obtained findings. Furthermore, Friedman's posthoc multiple comparison methods are demonstrated to statistically validate the suggested method [20]. Latif et al. [66] automate the diagnosis procedure of pneumonia using image processing techniques. It presents a suggested and realized automated method for accurately diagnosing pneumonia utilizing images from the DICOM chest X-ray collection. This research presents a pneumonia diagnosis system with enhanced deep residual networks (ResNet) architectures. The system is evaluated using a dataset of 30,227 DICOM Chest X-rays. Two residual network models, Version 1 and Version 2, were employed. Additionally, the outcomes were compared with three distinct CNN models and methodologies discovered in recent scholarly works. The findings demonstrate that the proposed ResNet (Version 2) technique

attains superior accuracy compared to CNN and other previously suggested approaches. The ResNet model attained an average accuracy of 88.67% after 80 epochs. The reviewed studies about chest disease detection and classification are summarized in Table 2.1.

Table 2.1. Literature summary for multi-class lung diseases classification

Disease	Study	Method	Medical Image	Performance		
				Acc.	Prec.	Sens.
COVID-19	[56]	VGG-16, ResNet-50, InceptionV3	CXR+CT	93	91	90
	[58]	VGG-19+ ResNet-50	CT	94	95	90
	[55]	DRE-Net	CT	86	96	93
	[54]	Modified ResNet	CXR+CT	99.3	99.7	99.1
	[53]	ResNet50	CXR	96.1	76.5	91.8
	[52]	DenseNet121	CXR+CT	98	96	96
	[51]	VGG-16	CXR	98.67	100	98
	[58]	D-Resnet-10 network	CT	81.4	79.8	87.5
	[59]	VGG+CNN	CT	96.2	97.3	94.5
	[60]	VGG-16, InceptionV2, DT	CXR+CT	91	94	97
	[61]	GSA-DenseNet121	CXR	98.38	98.5	98.5
	[62]	Deep learning Meta classifier	CXR+CT	99	99	99
	[63]	ResNet32+DTL	CT	93	95	91
	[67]	VGG-16	CXR	79.58	92	95
	[64]	InstaCovNet-19	CXR	99.08	99	99
	[68]	CNN	CXR+CT	98.28	98.22	98.25
Lung Cancer	[34]	FPSO-CNN	CT	95.62	96.32	97.93
	[33]	Multi-layer Perceptron (MLP)	CT	88.55	86.59	89.84
	[32]	CNN	CT	84.15	84.32	83.96

	[31]	MGSA	CT	94.56	94.2	96.2
Pneumonia	[30]	RNN-LSTM	CXR	95.04	88.89	95.41
	[29]	ResNet50 +MobileNetV2+ InceptionResNetV2	CXR	95.09	95.53	94.43
	[25]	CNN with pre-trained weights on ImageNet	CXR	91	92	87
	[28]	RetinaNet and Mask R-CNN	CXR	83.80	75.8	79.3
	[27]	Transfer learning	CXR+CT	94.9	93	93
	[26]	CNN	CXR	93.73	-	-
	[24]	Xception Network pre-trained weights on ImageNet	CXR	97.3	84.3	99
	[23]	Xception+ResNet 50V2	CXR	99.50	92.69	80.53
Tuberculosis	[17]	DCNN	CXR	98.45	82	72
	[18]	Depth-ResNet	CT	85.29	-	84.16
	[19]	Ensemble (AlexNet, GoogleNet and ResNet)	CXR	88.24	88.0	88.42
	[20]	(SVM+FOSF+GLCM)	CXR	99.40	99.42	99.40
Lung Opacity	[66]	ResNet	CXR	88.67	-	-

2.4. Lung Tumour Segmentation using Multimodality of CT-PET Scans

CT and PET imaging are used in various research papers because of the unique insights they provide into the structure and function of the human body, respectively. Combining the two allows for the early detection of even the tiniest lung tumors. This section provides the detailed literature for lung tumor segmentation using multimodality imaging CT and PET.

Wang et al. [69] advised a DL-based dual-modality approach using CT and PET scans to develop an automated segmenting of lung tumors for radiation therapy planning. Two distinct convolution routes were built into the 3D convolutional neural network for extracting features at different resolutions from the PETs and simulated CTs, and a single deconvolution path was also built into the network. Tumour segmentation via skip connections at each granularity was achieved by aggregating the obtained characteristics from the convolution arms and feeding them into the deconvolution pathway. A panel of oncologists judged the medical effectiveness of the network-generated segmentation strategy. While this work has many promising applications, it does have some caveats. The network may struggle to produce precise segmentations when tumor edges are not precise on CT or PET.

Park et al. [70] presented a two-stage Unet model to boost the segmentation effectiveness of lung tumors by utilizing [18F]FDG PET/CT, as precise segmentation is necessary for determining the functional size of a tumor in this imaging modality. The LifeX program was used to create the tumor volume of interest. In the first step, a 3D PET/CT volume is used to train a global U-net, based on which a 3D binary volume is then retrieved to serve as an initial representation of the tumor's region. In the second stage, the PET/CT slice identified in Stage 1 is sent to the U-net, generating a 2D binary image centered on the eight adjacent slices. The major drawback of the research is the lack of a 3D volume as the final result of the suggested approach. It may cause the coronal and sagittal slices to have gaps between the binary segments.

Xiang et al. [71] recommended a modality-specific segmentation network (MoS-Net) to segment lung tumors. To better understand the differences between PET and CT scans, MoSNet is taught to use modality-specific representations. In contrast, modality-fused representations are employed to convert the typical characteristics of lung tumors in both scan types. The authors suggest an adversarial approach that uses an adversarial purpose concerning a modality

discriminator and a reserved modality common illustration to reduce the modality difference's approximation. As a result, the network's ability to represent data for the segmentation in PET and CT scans is enhanced. By generating a map for each modality, MoSNet can explicitly quantify the weights for the attributes in each modality. However, the limitation of the research is that the proposed approach is developed for 2D thorax PET-CT slices.

Fu et al. [72] proposed a DL system for lung cancer segmentation, i.e., a multi-modal spatial attention module (MSAM). It is trained to highlight tumor-related regions selectively and downplay those physiologically rising from the PET scans. Next, using the created spatial attention maps, a CNN core is trained to focus on areas of a CT image with a higher propensity for tumors. The drawback of the research is that the datasets used only had one observer define the outlines. If numerous observers had been used to reach a consensus segmentation, things would have gone much smoother. Because of the potential vagueness of the related thresholding approach used to create the ground truth for the NSCLC dataset, the segmentation outputs require human adjustment to correct for incorrectly categorized ROIs.

Zhong et al. [73] provided an innovative method for lung tumor segmentation by bringing together a robust FCN-based 3D-Unet and a graph-cut-based co-segmentation model. Initially, high-level discriminative features for PET and CT images are learned by independently training two distinct deep Unets on the data sets. These features then create tumor/non-tumor masks and probability maps. The final tumor segmentation findings are obtained using the PET and CT probability mappings in a graph-cut-based co-segmentation model. Despite fusing their extracted features, the research has given different results for CT and PET.

Hwang et al. [74] recommend a new network architecture called 3C-Net, which uses numerous contexts in three distinct ways. Two decoders in the network are implemented to exploit inter-slice contextual information: a segmentation decoder and a context decoder. The context decoder receives the inter-slice difference features and uses them to predict the segmentation mask's inter-slice difference. Having this 3D background information for each slice helps in attention direction. The prediction results from each decoder stage were used to derive a loss function for network optimization. Since two modalities are used, i.e., PET/CT data, a co-encoder block is implemented to extract mutually reinforcing features from both modalities while simultaneously acquiring contextual knowledge about them. Weights for both CT and PET were modified twice in co-encoder blocks. The co-encoder blocks take in relevant data from both modalities, allowing for interaction while maintaining spatial and structural

coherence. The encoder additionally includes an Asterisk Spatial Pyramid Pooling (ASPP) block in its final step. The ASPP block aids the network in increasing the scope of its observations and avoiding the loss of spatial context, which allows the recording of visual context at various scales.

Kumar et al. [75] improve the multimodality of PET-CT fusion using CNN, which learns to fuse complementary information. The proposed CNN stores modality-specific characteristics before deriving a spatially variable fusion map. It allows quantifying the relevance of each modality's characteristic across various spots. Moreover, multi-plying the fusion maps with the modality-specific feature maps yields representations of the complementary multimodality data at various positions. The recommended CNN is tested on PET-CT scans of lung tumors, where its ability to detect and separate many regions with variable fusion needs is evaluated.

Jemaa et al. [76] demonstrated a comprehensive strategy employing 2D and 3D CNN for rapid tumor classification and metabolic data retrieval from whole-body FDG-PET/CT images. This architecture is relatively economical in terms of tumor load, healthy tissue volume, and the intrinsic heterogeneity of the input images. This is especially important for whole-body scans due to their vast size and high asymmetry.

Zhao et al. [77] developed a novel multimodality segmentation approach that utilizes a 3D FCN and simultaneously includes PET and CT data in tumor segmentation. Initially, the network underwent a multitask training phase, during which two parallel sub-segmentation architectures, each built with a deep CNN, were learned to generate map-like features from both modalities. The PET/CT feature maps' characteristics were re-extracted using a weighted cross-entropy reduction technique, and a feature fusion component was then constructed using cascaded convolutional modules. The softmax function was also used to generate the cancer mask as the network's final output. The research lacks an automatic setting of the weighting parameters of the loss functions, which can affect performance. Also, more effective ways for feature extraction can increase the segmentation's performance.

Using W-net, Zhong et al. [78] evaluate 3D Deep Fully Convolutional Networks (DFCN) for tumor co-segmentation on dual-modality NSCLC and PET-CT images. CT and PET data are combined to understand NSCLC tumors in PET-CT scans better and apply DFCN co-segmentation. The recommended DFCN-based co-segmentation approach uses two connected 3D-UNets with an encoder-decoder to exchange complementing data between PET and CT.

Bi et al. [79] developed a hyper-connected fusion model that uses a CNN-TN fusion encoder and a CNN-TN fusion decoder. With hyper-connections between them, the encoder splits into three forks to independently handle PET, CT, and combined PET-CT scans. The transformer encoders process the encoded image embeddings to learn complimentary characteristics in a long-range dependency between the PET, CT, and concatenated PET-CT images. The transformer decoder combines the learnt embeddings to find characteristics important for segmentation, which are subsequently transformed into a 2D feature map. The segmentation results are then up-sampled using a convolutional neural network. The data came from the soft-tissue sarcoma databases. The data showed that the model's dice had a probability of 66.36%. The summary of the literature research on lung tumor segmentation models is listed in Table 2.2.

Table 2.2. Summary of literature on lung tumor segmentation models using multimodality

Author Year	CT-only Extractor	PET-only Extractor	Feature Fusion	Dataset Description
Wang et al. [69]	3D CNN	3D CNN	3D CNN	Private clinic dataset comprising 290 pairs of CT and PET.
Park et al. [70]	Global Unet	Global Unet	Regional Unet	Private data of 887 individuals with lung cancer.
Xiang et al. [71]	Dual-stream encoder	Dual-stream encoder	Decoder branch	126 PET-CT scans containing NSCLC
Fu et al. [72]	Encoder-decoder backbone CNN	Multimodal spatial attention	CNN architecture containing	Two clinical PET-CT datasets of

Author Year	CT-only Extractor	PET-only Extractor	Feature Fusion	Dataset Description
		module (MSAM).	skip connections.	NSCLC and STS
Zhong et al. [73]	3D-Unet	3D-Unet	graph-cut-based co-segmentation model	PET-CT scans from lung cancer patients
Hwang et al. [74]	Shared co- encoder	Shared co-encoder	Shared co-encoder	F-18-FDG PET/CT scans from a private hospital
Kumar et al. [75]	An encoder using multiscale output	An encoder using multiscale output	Decoder using multiscale multimodal input	Biopsy-proven NSCLC FDG PET-CT scans.
Jemaa et al. [76]	-	-	2D U-Net and selected VNet	Patients with non-lymphoma Hodgkin's and NSCLC, which includes 3664 FDG-PET/CT images from head to toe.
Zhao et al. [77]	VNet	VNet	Voxel-wise addition, along with VNet	Private clinical dataset having 3D PET/CT images.
Zhong et al. [78]	An encoder using multiscale output	An encoder using multiscale output	Decoder using multiscale	NSCLC patients who received stereotactic

Author Year	CT-only Extractor	PET-only Extractor	Feature Fusion	Dataset Description
			multimodal input	radiation treatment
Bi et al. [79]	CNN-TN Encoder	CNN-TN Encoder	TN-CNN decoder	Non-small cell lung cancer (NSCLC) and one soft-tissue sarcoma (STS) dataset.

2.5. Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction Using Vision Transformers

This section is divided into three subsections: one section will be lung cancer detection using vision transformers, and the second section will be based on TNM stage classification.

2.5.1. Lung Cancer Detection Using Vision Transformers

Nevertheless, the emergence of transformers has informed researchers about a significant limitation of CNNs: their inability to capture long-range dependencies effectively. This limitation pertains to the challenges of extracting contextual information and identifying non-local correlations among objects. Malaviya et al. [80] proposed a vision transformer model utilizing CT data. The initial stage involved the classification of CT images from the dataset. To effectively tackle the initial training model's limitations, a segmentation method was utilized to partition the image into smaller patches. The image has been divided into smaller sections to efficiently process it using the transformer encoder. This approach allows the training process to proceed promptly while accounting for the variability in the images. The output of the transformer model has been designated as the multi-layer perceptron head. By employing the recommended model, the accuracy of 91.93% through rigorous training of 100 epochs is attained. The limitation of the recommended technique is its relatively lower level of precision when compared to other established methods. Another limitation is the computational expense associated with the function, which exceeds that of systems constructed using CNNs.

Similarly, Liu et al. [81] introduce a unique architectural framework, Res-trans networks, for classifying CT images for lung cancer. The authors employ various methodologies to investigate the research question. The utilization of local and global blocks was employed to extract features that efficiently maintain the interconnections among pixels. The researchers have devised residual blocks employing convolutional operations to extract local features. Furthermore, the construction of transformer blocks includes the utilization of self-attention processes to capture global properties efficiently. In addition, the Restrans network integrates a sequence fusion block that efficiently merges and extracts the sequence data produced by the transformer. The tenfold cross-validation results on the LIDC-IDRI dataset demonstrate that the suggested method achieves superior performance, with an AUC of 0.9628 and an Accuracy of 0.9292. However, a potential weakness of this study is the utilization of subjective malignant labeling to train the model.

Wang et al. [82] also aim to classify lung nodules on CT images using a CT image-based transformer model, TransPND. The model uses a 2D Panning Sliding Window technique to enrich data, focusing on local features. The encoder component of TransPND can be subdivided into two distinct sections: the Self Attention Encoder and the Directive Class Attention Encoder. The self-attention process in the self-attention encoder resembles the conventional approach, but it integrates Local Diagonal Masking (LDM) as a means to determine the position of attention. The DCA method directs attention towards local features while reducing computational burden. The Weight Learning Diagonal Matrix regulates residual connections in both stages. Extensive tests on the LIDC-IDRI dataset show a precision rate of 93.33 %.

2.5.2. Lung Cancer TNM Stage Classification

A limited number of researchers have devised methodologies for classifying lung cancer stages. Several techniques in this research are derived from emphasizing basic image processing methods, explicitly emphasizing the T descriptor. These techniques involve calculating parameters such as area, perimeter, and eccentricity. These approaches have been previously discussed in references [83–85]. Additional strategies involve utilizing convolutional neural network (CNN) based algorithms specifically emphasizing T or N descriptors. The studies [86]propose a convolutional neural network (CNN) approach in a two-dimensional (2D) framework for the classification of T categories, specifically distinguishing between T1/T2 and T3/T4. The researchers employed FDG PET/CT data and obtained 82.6% average accuracy with cross-validation. The final model yielded a test accuracy of 69%. The current approach is limited to binary categorization and does not account for individual T-class distinctions.

Furthermore, since all three labels are required for accurate TNM staging, the above method omitted the N and M descriptors.

To classify the T-stage of lung cancer, researchers have devised a method based on a double convolution neural network, as described in references [87]. Nevertheless, it may fail to account for certain T-phases and N, M stages. In their study, Paing et al. [88] provide a methodology for the identification and stage classification of lung cancer. This methodology utilizes five distinct methods: the Support Vector Machines, K-nearest neighbor, Neural Networks (NN), decision tree, and ensemble tree. The researchers employed four distinct datasets, and NN attained the highest accuracy of 90.6% for classifying a total of seven T-stages.

The study by Zhao et al. [89] proposes a novel approach that utilizes cross-modal 3D DL techniques to predict lymph node metastasis in patients diagnosed with clinical stage T1 adenocarcinoma. The researchers integrate previous clinical characteristics acquired by combining the clinical data with the image features. The researchers conducted an experiment using a dataset obtained from a privately owned hospital, resulting in an accuracy rate of 87.6%. There is a limited availability of studies that suggest a comprehensive classification system for the TNM staging of lung cancer, with Moitra et al. [90] being the sole identified study that considers all three descriptors (T, N, and M) for this purpose. The researchers utilize an openly accessible dataset known as NSCLC-Radiogenomics [91]. The authors present a 1D CNN as a potential approach to classifying the lung cancer TNM stages and histological grading. The characteristics of the tumor have been derived from the delineation of PET/CT images of the patients.

Tyagi & Talbar [92] aim to provide a new and effective method for categorizing the stages of lung cancer using the TNM criteria. A multi-level 3D deep CNN called Lung Cancer, Stage Classification Network, is recommended. The recommended network architecture has three classifier networks, each designed to classify T, N, and M-labels. Firstly, the data pre-processing stage involves augmenting the CT images and processing the label files to extract the necessary TNM labels. The classification network employs a DCNN incorporating a contemporaneous squeezing and excitation element and asymmetric convolutions to categorize each label separately. The overall stage is determined by combining all three descriptors. The simultaneous squeeze and excitation unit improves the algorithm's classification accuracy by enabling it to concentrate on the crucial information in the image. Asymmetric convolutions

are utilized to reduce the computationally complex nature of the network. The average accuracy for the T-Stage classification was 96.23%, the N-Stage classification was 97.63%, and the M-Stage classification was 96.92%. Furthermore, a classification accuracy of 97% is achieved for the overall stage categorization.

The classification process necessitates some fundamental characteristics, followed by considering a one-dimensional aspect. The implementation of CNN obtains the final classification results. The researchers have attained a mean accuracy rate of 96%, indicating a commendable level of performance. This technique offers several benefits, including a streamlined model, reduced computational requirements, and exceptional precision. Nevertheless, this approach has certain disadvantages, including the need for extensive human pre-processing of the data before its application in a Convolutional Neural Network (CNN).

Initially, the tumor regions undergo segmentation; then, characteristics are extracted from these regions. Subsequently, a manual selection process is employed to identify relevant features for classification. In addition, they used segmentation to isolate the tumor, which prevents examination of adjacent structures for staging purposes. In addition, it is imperative to consider various metastasis forms when assessing M-staging. In the case of brain metastasis, medical professionals choose to utilize a brain magnetic resonance imaging (MRI) scan. However, in the case of other forms of metastasis, such as adrenal metastasis and liver metastasis, a comprehensive CT scan is necessary to examine several organs for metastatic growth. Segmenting lymph nodes or liver tumors that have metastasized from lung cancer poses a significant challenge throughout lung tumor segmentation. Information about the T, N, and M stages cannot be adequately determined based only on the excised tumor region. The primary objective of this study was to rectify the insufficiencies identified in previous research about the categorization of lung cancer staging and forecasting of the overall stage. The constraints observed in previous studies are attempted to address by employing direct overall stage prediction utilizing Vision Transformer architecture.

2.6. Summary Of Research Gaps

The literature review reveals several significant gaps in current research on lung disease diagnosis, tumor segmentation, and staging of lung cancer.

From the literature review, most existing studies focus on binary classification architectures for lung disease diagnosis using X-ray imaging (e.g., [19, 26, 28, 53]). While binary

classification provides effective results, it limits the model's ability to learn complex patterns due to the reduced number of classes. This simplified approach makes the model less capable of capturing diverse patterns, and as the number of classes increases, the model's performance tends to degrade. Additionally, in binary classification, the starting baseline accuracy during testing is inherently 50% due to the limited number of outcomes. While this simplifies the classification task and often leads to seemingly high-performance metrics, it fails to challenge the model to learn the intricate and overlapping patterns that are critical in distinguishing between multiple diseases. This lack of complexity in binary architectures ultimately limits their applicability in scenarios requiring the accurate classification of multiple conditions, where precise differentiation is essential for timely and effective treatment. Although several attempts have been made to develop multi-class architectures for lung disease classification, including lung cancer, using CT scans, no existing work integrates lung cancer diagnosis into multi-class lung disease architectures based on X-ray imaging. This gap is critical, as X-rays are widely used as a first-line diagnostic tool globally. The overlapping characteristic patterns of lung cancer with other lung diseases such as similar shadowing, nodules, or opacity patterns on X-ray images can delay its detection due to misinterpretation or failure to distinguish between conditions. This delay can result in lung cancer being identified at more advanced stages, reducing the effectiveness of treatment and impacting patient outcomes. Thus, there is an urgent need for a comprehensive multi-class classification architecture that includes lung cancer alongside other lung diseases. This approach will allow for diagnosis at an earlier stage and the enhancement of treatment pathways, filling an important gap in the existing literature.

Although the majority of the existing literature on lung tumor segmentation is concerned with primary lung tumors, metastatic tumors, especially soft tissue sarcoma that originate from outside the lung sites and metastasize to the lung, are rarely addressed. The imaging characteristics of primary and secondary lung tumors, are different. Primary tumors usually have borders that are well defined and follow a predictable standard pattern, and secondary metastatic STS, on the other hand, may be irregularly shaped with association of complex boundaries and have features of surrounding lung tissue. These differences make it impossible to use traditional segmentation methods in the accurate identification of metastatic STS. The challenges arise as a result of their heterogeneous appearance and tendency to integrate with adjacent lung tissue, which may result in potential misdiagnosis or segmentation failure. Therefore, specialized segmentation techniques are required that will address the intricacies of metastatic STS tumors to enhance their precise diagnosis and treatment

Utilizing the TNM system (Tumor, Node, Metastasis) for accurate staging of lung cancer is essential for determining appropriate treatment strategies and predicting patient outcomes. However, other research, for instance, [86, 88] have concentrated on tumor size (T), whereas they often ignore other important factors including the involvement of lymph nodes (N) and the presence of metastases in distant sites (M). The studies conducted by [89] which only concentrate on the N stage, do not address the disease in a comprehensive manner, which affects the stage classification accuracy. This can result in different treatment protocols being used that can have an adverse effect on the patients. Additionally, although some studies [90, 92] succeed in implementing TNM-based models for stage prediction, they focus on demographic and clinical variables integration for which are important in order for staging to be accurate. Furthermore, most of the existing models segregate the T, N, and M stages and use individual branches for each stage prediction and then combine the output. This method introduces additional complexities in the model and increased computational time, which compromises the efficiency required for real time applications. In addition, multi-view CT imaging, which provides complete anatomic views improving the accuracy of staging, has not been adequately investigated with respect to TNM-based models. Hence, there is enough justification for developing a unified model that incorporates multimodal imaging including multi-view CT and clinical data to deliver an accurate overall stage prediction, behind only a reasonable computational demand in order to make it practicable in real time.

To sum up, the existing research emphasize key gaps in multi-class classification, tumor segmentation, and TNM-based staging for lung cancer. Addressing these gaps through integrated, efficient, and specialized models will significantly improve diagnosis, treatment, and patient outcomes.

2.7. Chapter Summary

This chapter describes the comprehensive literature review and analysis for lung disease classification and segmentation, along with survival prediction. The chapter is divided into various sections to present a comprehensive overview of the existing research on the classification of lung diseases using deep learning techniques, specifically focusing on multi-class classification. The literature on Lung Tumor Segmentation utilizing a combination of CT and Positron Emission Tomography (PET) Scans is also presented. It also contains an extensive review of the research on Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction using Vision Transformers.

Chapter 3

3. Deep Learning Architecture for Multi-Class Lung Diseases Classification Using Chest X-ray (CXR) Images

3.1. Introduction

In 2019, the world experienced the rapid outbreak of the COVID-19 pandemic, which created an alarming situation worldwide. The virus targets the respiratory system, causing pneumonia with other symptoms such as fatigue, dry cough, and fever, which can be mistakenly diagnosed as pneumonia, lung cancer, or TB. Thus, the early diagnosis of COVID-19 is critical since the disease can provoke patients' mortality. Chest X-ray (CXR) is commonly employed in the healthcare sector, where both quick and precise diagnoses can be made. Deep learning algorithms have proved extraordinary capabilities in terms of lung disease detection and classification. They facilitate and expedite the diagnosis process and save time for the medical practitioners. In this chapter, a deep learning (DL) architecture for multi-class classification of Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity, and most recently COVID-19 is proposed.

3.2. Proposed Methodology

The human respiratory system is attacked by a variety of lung illnesses. These diseases include pneumonia, tuberculosis, lung cancer, and lung opacity, among others. These diseases can cause similar effects on human lungs; therefore, X-ray images are commonly employed for diagnosing these diseases. AI in the form of deep learning algorithms has increasingly played a key role in disease identification and classification. Deep learning facilitates the diagnosis process and saves time for healthcare providers.

The study presents a multiclass deep learning classification model to identify the most common chest diseases. The aim of the research work is to design a deep learning framework and classify multi-classes of Pneumonia, Lung Cancer, TB, Lung Opacity, and most recently, COVID-19. A thorough search of the literature shows that this research is the first attempt to use the single deep learning framework, incorporating and classifying all these six classes at a

time. Figure 3.1. represents the proposed framework in a block diagram. The framework is divided into three phases: pre-processing, feature extraction, and classification. X-ray scacns were used as inputs, and the categorization of the input X-ray image on a disease level was the final output of the model.

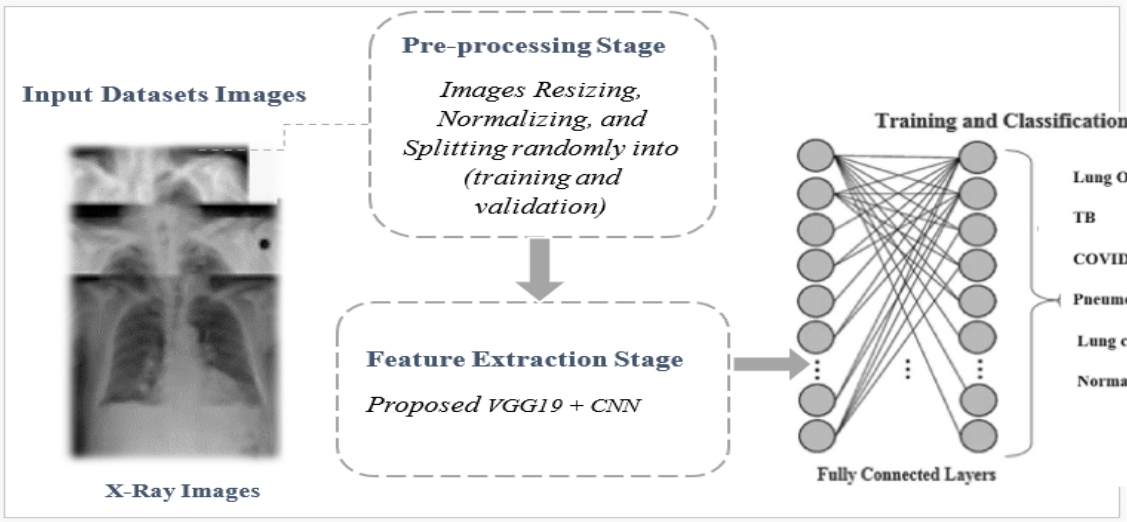

Figure 3.1. The proposed framework for Multi-Class Lung Diseases Classification

During the first phase, the input images undergo pre-processing functions such as normalization, resizing, and data image splitting into 80% training and 20% validation at random. Then, deep learning algorithms are used during the second and third stages. The second phase involves feature extraction, which is performed using VGG19 and CNN techniques. The fully connected network technique is employed during the image categorization step.

3.2.1. DATASET

For the experimental purpose, in addition to healthy cases, tremendous X-ray images of pneumonia, TB, lung cancer, lung opacity, and, most recently, COVID-19 were accessed and collected from reliable sources.

To begin with, for COVID-19, 4189 CXR images [93] were included in this study. Secondly, 7397 CXR images of pneumonia were extracted [93], which are publicly available for research purposes. Furthermore, [93] represents 6,012 CXR images of Lung opacity and 10,192 of Normal samples whilst [94, 95] indicates the dataset resource for a total of 10,000 X-ray images of lung cancer. Ultimately, a total of 4,897 X-ray images for tuberculosis [93] were collected and employed in the research. Over 42,000 specifies the total number of CXR images used in

the experiments. Samples of chest X-ray images for COVID-19, normal, pneumonia, TB, lung opacity, and lung cancer are shown in Figure 3.2. The number of patients for each disease dataset with respect to ages: ages were frequently between 38 and 65 for the COVID-19 dataset, 26 and 62 for the pneumonia dataset, 28 and 58 for the lung cancer dataset, and for normal patients, the ages were between 33 and 58 years.

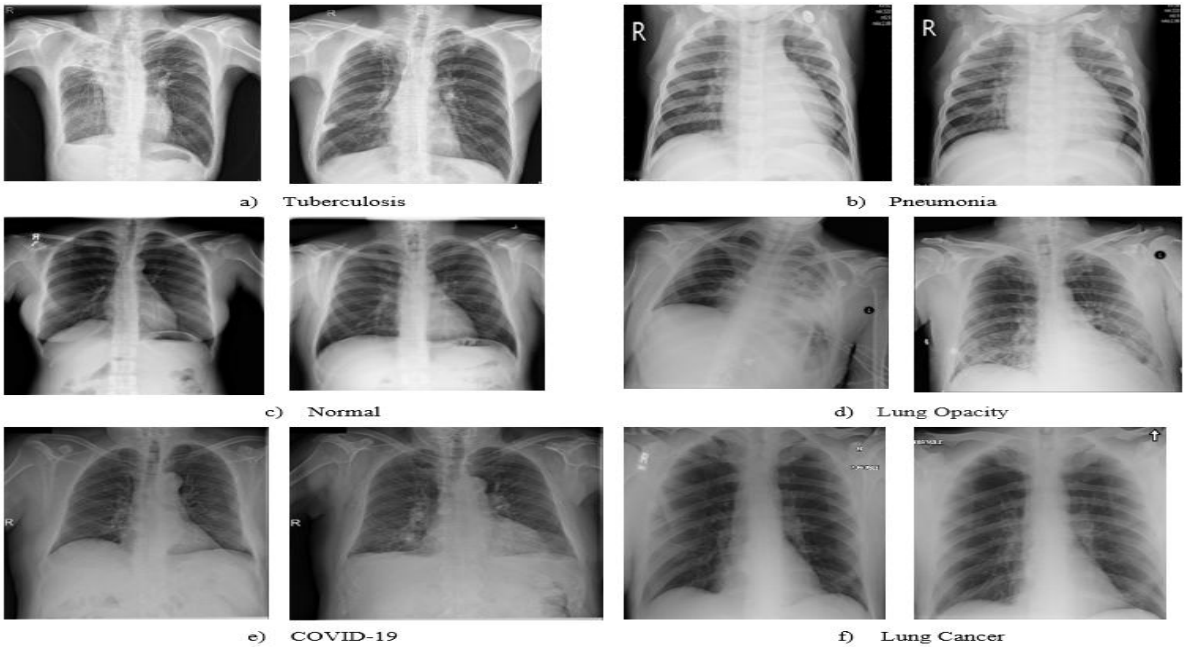

Figure 3.2. Chest X-ray images: (a) Tuberculosis images, (b) Pneumonia images, (c) Normal images, (d) Lung Opacity images, (e) COVID-19 images, (f) Lung cancer images

3.2.2. Dataset Pre-processing

Some pre-processing processes were employed to adjust the input data to meet the requirements of the deep learning model: 1) The images were resized; 2) the images were normalized; 3) the images were converted to an array to be employed as input in the model's next phase. To ensure robust model evaluation and to prevent overfitting, the dataset was randomly divided into training, validation, and testing subsets, corresponding to 70%, 10%, and 20% of the total data, respectively. The test set was kept completely independent and used only for the final evaluation of the proposed model. This setup ensured that the high accuracy reported in Table 3.1. reflects the model's generalization capability on unseen data. To meet the criteria of the framework, all images were scaled to 224*224*3. After normalizing each pixel in the image to the interval [0,1], all images were transformed into an array data representation.

3.2.3. Proposed Deep learning VGG19+CNN Model

This research presents supervised deep learning for multiclass classification of the most common chest diseases. For classification, a pre-trained model, VGG19, is used, and CNN is used as a feature extraction model, which is fully connected.

The choice of VGG19 [96] as the feature extraction backbone in this study was motivated by its proven ability to capture multi-level hierarchical representations through deep convolutional layers with small receptive fields. VGG19 is particularly effective for medical images where subtle intensity changes are critical for identifying disease-specific features. However, to further enhance discrimination among multiple lung diseases, a dedicated CNN block was integrated after the VGG19 feature maps to refine spatial dependencies and strengthen classification sensitivity. This hybrid architecture leverages the transfer learning capability of VGG19 while retaining flexibility for domain-specific adaptation. The combination provided a robust balance between feature generalization, computational efficiency, and classification accuracy across six lung disease classes.

A convolution layer with a ReLU as an activation function is included in each CNN block. Following these three CNN blocks, batch normalization and a max-pooling layer were applied, which were then followed by a dropout layer, as indicated in Figure 3.3.

In the feature extraction step, the output was turned into a one-dimensional data vector, which was then used as an input in the classification stage after being modified through the flattening layer. The remaining components of the categorization step are comprised of three thick layers, each having 512, 256, and 128 neurons. It is a thick layer with six neurons, and the SoftMax activation function generates the final classification output. This layer is responsible for classifying the output image into one of the six chest disease classes: pneumonia, tuberculosis, lung cancer, and lung opacity. A total of 24,622,470 model parameters are span into two categories. First were the trainable parameters (24,622,342), which were revised throughout the training process. The best value for these parameters was required to ensure the training accuracy. The second category was the untrainable parameters (128), which were those that did not change at the time of training. Figure 3.4. illustrates the pseudo-code for the proposed framework.

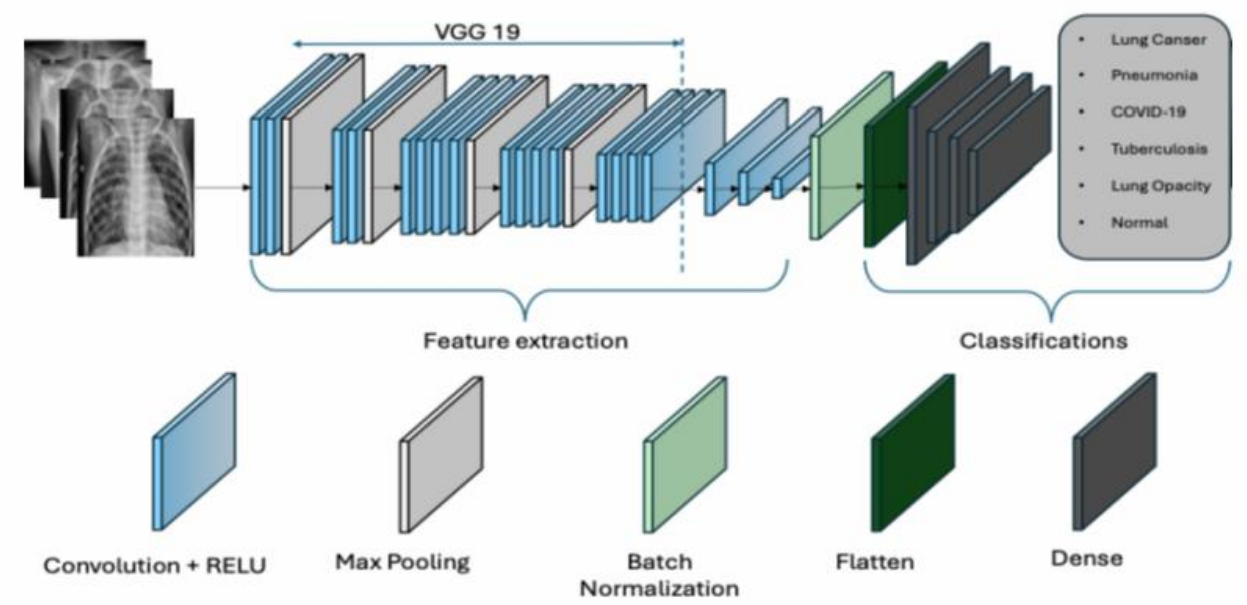

Figure 3.3. Model Architecture

Pseudo-Code of the Proposed Model
1. ***Input: Clinical images of chest diseases dataset***
2. ***Pre-processing: Resize images to 224*224*3 and Normalize images pixel values into interval [0,1]***
3. ***Split train data to (80, 20): 80% training and 20% for validation***
4. ***Extracting features using Vgg19+CNN deep learning approach***
5. ***Classify images by fully connected networks***

Figure 3.4. Pseudo-code for the proposed framework.

3.3. Results

A classification model for chest disease was created using Python 3 and the Keras framework. The model was simulated on a Google Colab Pro edition with 2 TB storage, 25 GB RAM, and CPU-P100. The ImageDataGenerator class in Keras was used during the pre-processing stage, which included picture scaling, normalization, and conversion to an array of data.

The suggested multi-chest illnesses classification deep learning model input was created using the outcome of the pre-processing step. An optimizer and appropriate fit algorithms were used with 5000 epochs to train and validate the model. Eight iterations and 32 batch sizes were

employed in each epoch. With the greatest precision, the performance metrics formulae were entered into the validation data outputs. The Adam [97] optimizer was employed, with a learning rate of 0.000001. This value was determined empirically through parameter tuning experiments to achieve stable convergence and minimize validation loss. The suggested deep learning model's code was published on the GitHub website [98].

Precision, loss, F1-score, accuracy, AUC, and recall were used to evaluate the model's performance. Accuracy was calculated as the proportion of correctly predicted instances to the total number of instances. Precision (positive predictive value) measures the ratio of correctly predicted positive samples to all predicted positives. The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of both metrics. Recall (sensitivity) quantifies the proportion of actual positives correctly identified by the model. These metrics are defined in Equations (3.1)–(3.4) [99]:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{3.1}$$

$$F1-score = \frac{2\,T_p}{2\,T_p + F_p + F_n} \tag{3.2}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{3.3}$$

$$Recall\ (Sensitivity) = \frac{T_p}{T_p + F_n} \tag{3.4}$$

Where the actual positive and negative parameters are denoted by T_p and T_n, respectively. The False positive and false negative values are denoted by F_p and F_n, respectively.

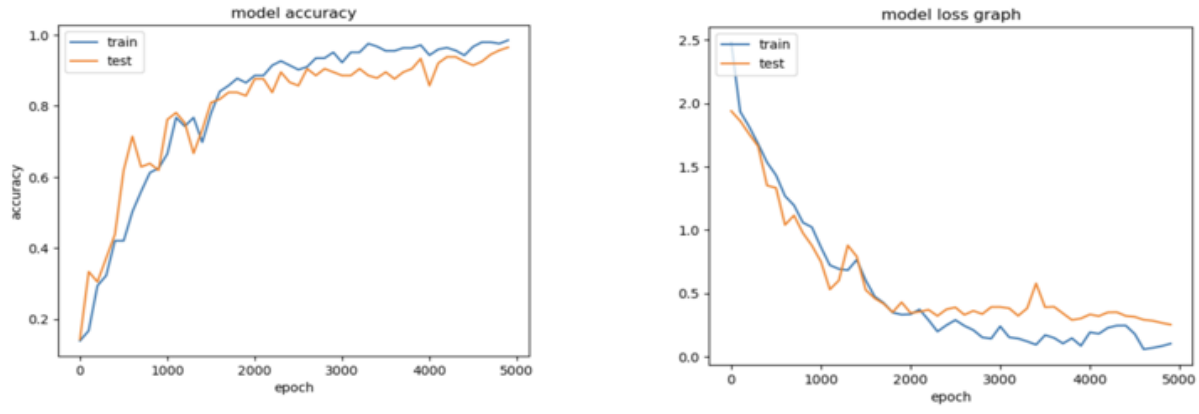

Figure 3.5. Performance metrics change with epochs in the training and validation

In addition, a confusion matrix is computed for the proposed model. Figure 3.5. shows how the performance of the varying epochs during the training and validation phases.

3.3.1. Experimental Results

The proposed VGG19 + CNN model was evaluated using the metrics defined in Equations (3.1)–(3.4). Table 3.1. presents the validation results obtained during the iteration that achieved the highest validation accuracy. The model achieved a loss of 0.1792, an accuracy of 96.48%, a precision of 97.56%, a recall of 93.75%, an F1-score of 95.62%, and an AUC of 99.82%. These results demonstrate that the proposed hybrid model provides a highly reliable and balanced classification performance across all evaluation metrics.

Table 3.1. The performance validation of the VGG19+CNN model

Methods	**Loss**	**Acc**	**Pre**	**AUC**	**F1**	**Recall**
VGG19 + CNN	0.1792	96.48	97.56	99.82	95.62	93.75

3.3.2. Comparative Analysis

To the best of our knowledge, there is no recent existing research has employed a single deep learning model for evaluating and classifying the following chest diseases together: Tuberculosis, Pneumonia, Lung Opacity, Lung cancer, and COVID-19 images. To show the effectiveness of the proposed model, Table 3.2. introduces a comparative analysis of fifteen existing works. The comparative analysis presented in this chapter has been expanded to clarify the fairness and relevance of performance evaluation. The proposed VGG19 + CNN model was trained using a unique dataset combination consisting of the Harvard Dataverse dataset [93] for five lung diseases and the ChestX-ray8 and JSRT datasets [94, 95] for lung cancer samples. To the best of our knowledge, no existing research has included lung cancer alongside five other pulmonary diseases within a single multi-class classification framework. Consequently, previous studies used for comparison were selected on a conceptual and methodological basis rather than as direct dataset replications. Although their datasets differ, these works represent the most relevant state-of-the-art approaches in deep learning-based lung disease classification. Thus, the presented comparisons aim to highlight architectural effectiveness and generalization capability rather than absolute numerical equivalence across datasets.

Table 3.2. The comparison between the proposed model and existing related work

Ref	Number of Classes	Method	Medical Image	Performance		
				Acc.	Prec.	Sens.
[56]	3	VGG-16, ResNet-50, InceptionV3	CXR+CT	93	91	90
[57]	3	VGG-19+ ResNet-50	CT	94	95	90
[55]	3	DRE-Net	CT	86	96	93
[53]	2	ResNet50	CXR	96.1	76.5	91.8
[63]	2	ResNet32+DTL	CT	93	95	91
[58]	2	D-Resnet-10 network	CT	81.4	79.8	87.5
[33]	2	Multi-layer Perceptron (MLP)	CT	88.55	86.59	89.84
[32]	2	CNN	CT	84.15	84.32	83.96
[25]	3	CNN with pre-trained weights on ImageNet	CXR	91	92	87
[28]	2	RetinaNet and Mask R-CNN	CXR	83.80	75.8	79.3
[27]	3	Transfer learning	CXR+CT	94.9	93	93
[26]	2	CNN	CXR	93.73	-	-
[18]	1 Class with 5 Levels of Severity	Depth-ResNet	CT	85.29	-	84.16
[19]	2	Ensemble (AlexNet, GoogleNet and ResNet)	CXR	88.24	88.0	88.42
Proposed	**6**	**VGG19+CNN**	**CXR**	**96.48**	**97.56**	**93.75**

3.3.3. Architecture Performance

Accuracy, precision, and recall (sensitivity) are the major parameters used to measure the performance of the model. The accuracy of the proposed framework produced the highest results, with 96.48, overcoming the rest of the models in Figure 3.6.

As revealed in Figure 3.7., the best precision value was 97.56 with the proposed model. However, Figure 3.8. confirms that the proposed model achieves the highest sensitivity of 93.75 compared to others.

As presented in Figure 3.9., various architectures of individual pre-trained models, transfer learning, and ensemble techniques based on deep learning have been investigated and compared with the multi-class proposed framework. The results show that the proposed VGG19-CNN achieved the best performance. ResNet50 [53] was better than Transfer Learning [27]. However, the Ensemble model [19]records the lowest.

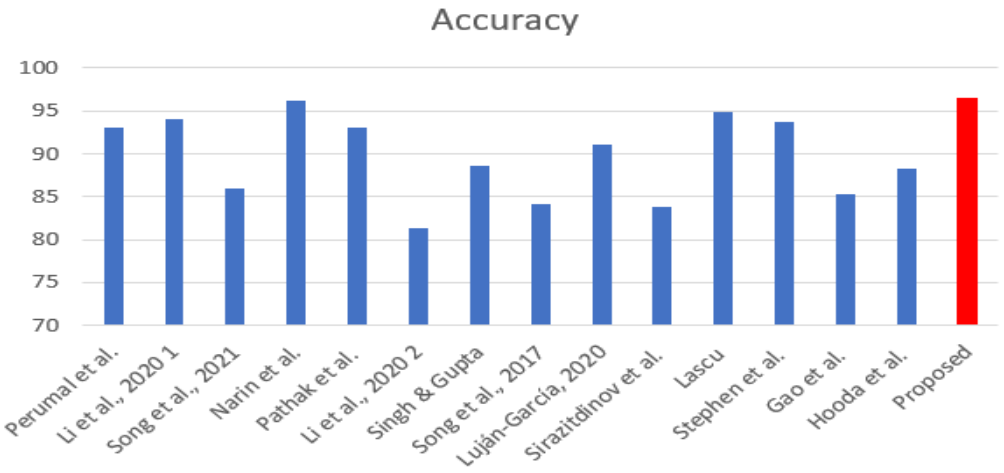

Figure 3.6. Competitive analysis based on Accuracy

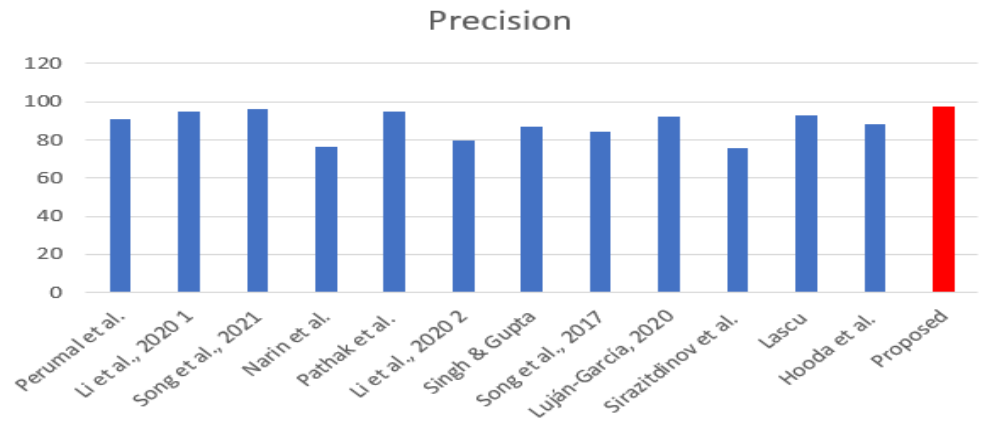

Figure 3.7. Competitive analysis based on Precision

Moreover, Table 3.2. illustrates the proposed multi-class framework used to classify six classes of the most popular chest diseases: tuberculosis, pneumonia, lung opacity, lung cancer, and COVID-19, in addition to normal cases. The model significantly outperformed binary classes presented by [19, 28, 33, 53]. Likewise, the model got over multi-class as observed by [25, 27, 57, 100]. The confusion matrix for the VGG19+CNN proposed model is shown in Figure 3.10, revealing that the VGG19+CNN model can successfully classify the six chest diseases with the highest ratio to COVID-19, starting from lung opacity, normal chest, lung cancer, pneumonia, and lastly the tuberculosis disease.

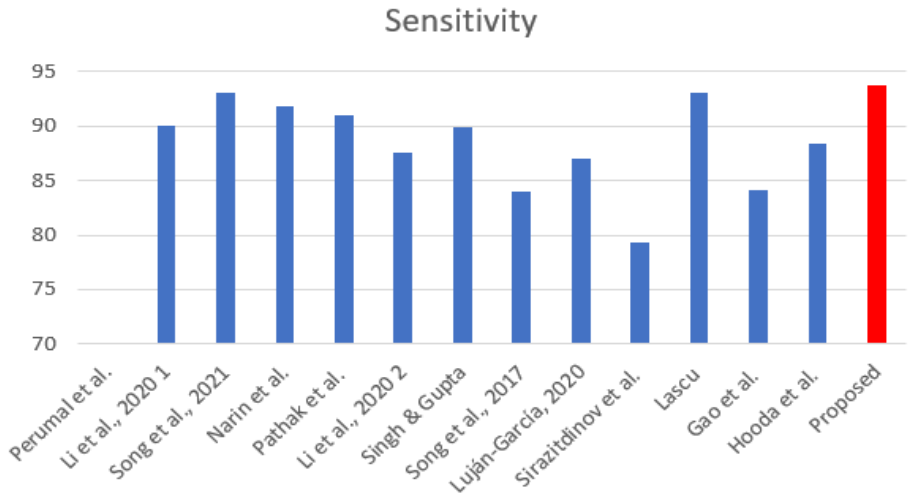

Figure 3.8. Competitive analysis based on Sensitivity

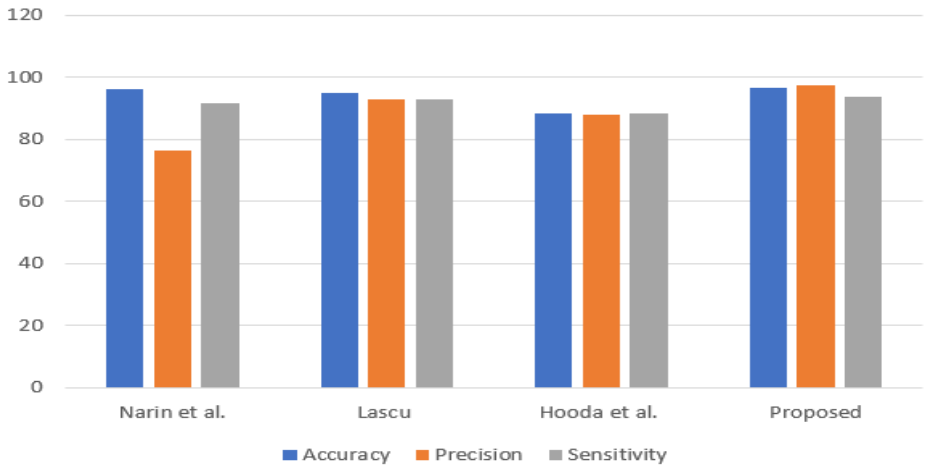

Figure 3.9. Competitive analysis is based on a variety of utilized deep learning approaches.

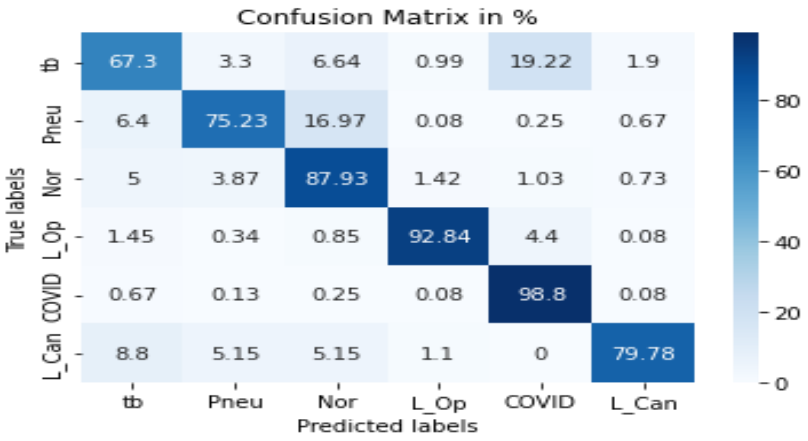

Figure 3.10. The confusion matrix

A multiclass deep learning classification model has been used in this work to incorporate and classify six classes of COVID-19, lung opacity, TB, lung cancer, and pneumonia using the VGG19+CNN approach. The architecture of the model was based on VGG19+CNN for feature extraction and a fully linked network for classification. The recall, accuracy, AUC, F1 score, and precision of the suggested model were all tested. The findings showed that the VGG19+CNN provided satisfactory classification performance with 96.48% accuracy, as shown in Table 3.2.

Based on X-ray images, the VGG19+CNN can identify various chest disorders with 96.48% accuracy, 93.75% recall, 97.56% precision, 95.62% F1 score, and 99.82% AUC. It is expected that the deep learning model will contribute to the development of a model for diagnosing chest disorders from CXR chest pictures, improving patient outcomes, and saving lives.

3.4. Discussion

In this chapter, a multi-class chest disease classification based on a deep learning architecture was developed and evaluated for classifying TB, lung opacity, lung cancer, pneumonia, normal, and COVID-19 using CXR images. In terms of classification, a pre-trained model, VGG19, followed by three blocks of convolutional neural network (CNN) as feature extraction and a fully connected network at the classification stage, was introduced. The experimental results revealed that the proposed VGG19 +CNN outperformed other existing work with 96.48% accuracy, 93.75% recall, 97.56% precision, 95.62% F1 score, and 99.82% area under the curve (AUC).

As Figure 3.5. shows, the training process demonstrated a robust convergence pattern with the train accuracy and validation accuracy scores, indicating that the model's learning trajectory is generally effective. Initially, both training and validation accuracy showed gradual

improvement over the first 1000 epochs, reflecting the model's ability to begin identifying patterns in the data. As training progressed, training accuracy maintained a steady upward trend, consistently achieving slightly higher values than validation accuracy. This discrepancy is expected due to the model's exposure to training data more frequently, which can result in a marginally higher accuracy.

From epoch 2000 onward, training accuracy continued to rise, reaching a plateau near 98.5% at the final epochs. Meanwhile, validation accuracy exhibited minor fluctuations but stabilized around 96.48%, indicating that the model generalized well to unseen data and did not experience significant overfitting.

The fluctuations observed in validation accuracy were minimal, suggesting the model's resilience to overfitting and confirming effective generalization across different classes. The slight disparity between training accuracy and validation accuracy in the final epochs suggests an optimal balance between bias and variance, resulting in accurate and consistent performance across classes.

Additionally, the confusion matrix in Figure 3.10. offers a granular view of the model's performance across the six classes: TB, pneumonia, normal, lung opacity, COVID-19, and lung cancer. Most predictions align closely with true labels, reflecting high specificity and sensitivity across classes. For instance: TB and pneumonia classes had modest misclassification rates, with TB showing some confusion with normal cases and COVID-19 (6.64% and 19.22%, respectively). Normal cases were predicted with 87.93% accuracy, though misclassifications occurred primarily with TB and pneumonia. Lung opacity achieved a high true positive rate at 92.84%, with minimal misclassification, emphasizing the model's capability to distinguish it accurately. COVID-19 was particularly well-classified, with a true positive rate of 98.8%, suggesting the model's vital feature extraction for this class. Due to their similar visual features on chest X-rays, lung cancer had the highest misinterpretation rate, with significant overlap with pneumonia and tuberculosis.

When these results are considered together with the performance measures (96.48 percent accuracy, 93.7 percent recall, 97.5 percent precision, and 95.6 percent F1 score), it is clear that the model is effective in the multi-class categorization of chest disorders. The model's ability to confidently distinguish between classes is emphasized by the AUC score of 99.82%, which is essential for real-world diagnostic applications. Its stability and high accuracy demonstrate the model's practicality for clinical applications during both the training and validation phases.

CT scans can accurately detect aberrant patterns even before symptoms appear. Therefore, employing a combination of CXR and CT images is a potential enhancement parameter for future work. Moreover, the identification of the region of interest (ROI) in conjunction with the classification of severity levels based on a powerful segmentation model is another direction for future work exploration.

3.5. Summary

In this chapter, a deep learning (DL) architecture for multi-class classification of Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity, and most recently COVID-19 is proposed. Tremendous CXR images of 4189 COVID-19, 6012 Lung opacity, 7397 Pneumonia, 10,000 lung cancer, 4897 tuberculosis, and 10,192 normal images were resized, normalized, and randomly split to fit the DL requirements. The proposed model integrated a pre-trained VGG19 backbone with three convolutional neural network (CNN) blocks for feature extraction and a fully connected layer for final classification. Experimental results demonstrated that the proposed VGG19 + CNN framework outperformed existing methods, achieving 96.48% accuracy, 93.75% recall, 97.56% precision, 95.62% F1-score, and 99.82% area under the curve (AUC).

These promising outcomes confirm the potential of deep learning–based diagnostic systems for accurate and automated identification of multiple lung conditions from chest X-ray images. However, while classification models provide effective detection at the image level, they do not offer detailed information about the exact location, shape, or extent of tumours, which are critical for clinical assessment and treatment planning. To address this limitation, the next chapter focuses on lung tumour segmentation using multimodal CT–PET imaging, enabling precise delineation of tumour regions and paving the way for advanced diagnostic and prognostic analysis.

Chapter 4

4. Hyper-Dense -Lung-Seg: Multi-modal fusion based Modified U-Net for Lung Tumour Segmentation using Multimodality of CT-PET Scans

4.1. Introduction

The majority of cancer-related deaths globally are due to lung cancer, which also has the second-highest mortality rate. Segmentation of lung tumors, treatment evaluation, and tumor stage classification have become significantly more accessible with the advent of PET/CT scans. With the advent of PET/CT scans, it is possible to get both functioning and anatomic data during a single examination. However, integrating images from different modalities can indeed be time-consuming for medical professionals and remains a challenging task. This challenge arises from several factors, including differences in image acquisition techniques, image resolutions, and the inherent variations in the spectral and temporal data captured by different imaging modalities. Artificial Intelligence (AI) methodologies have shown potential in the automation of image integration and segmentation. To address these challenges, multi-modal fusion approaches-based U-Net architecture (early fusion, late fusion, dense fusion, hyper-dense fusion, and hyper-dense vgg16 U-net) are proposed for lung tumor segmentation.

4.2. Contribution

The significant contributions of this research are given below.

- The inputs to the proposed architecture are PET and CT scans. Here, dense connections happen along the same pathways that process each modality individually. Last, their features are joined together at a high layer to finish separating them.
- Five deep models based on U-net architecture are suggested for lung cancer segmentation in multimodal image scenarios: Early fusion, Late fusion, Dense fusion, Hyper dense fusion, and Hyper dense VGG-16 U-net.

- The performance of the suggested models was evaluated using three types of loss functions: binary, dice, and focal loss functions.

4.3. Proposed Methodology

The selection of a U-Net-based architecture for tumour segmentation was driven by its strong suitability for medical imaging tasks that require precise localization. U-Net's encoder–decoder structure effectively captures both global context and fine structural boundaries, which are essential for accurate lesion delineation. The proposed Hyper-Dense VGG16 U-Net extends this capability by integrating dense cross-modal connections between PET and CT feature maps, allowing the network to exploit complementary spatial and metabolic information. The use of VGG16 as the encoder improves representational depth and feature reuse while maintaining manageable computational complexity. This design ensures stable gradient flow, improved boundary recovery, and enhanced tumour segmentation consistency across imaging modalities.

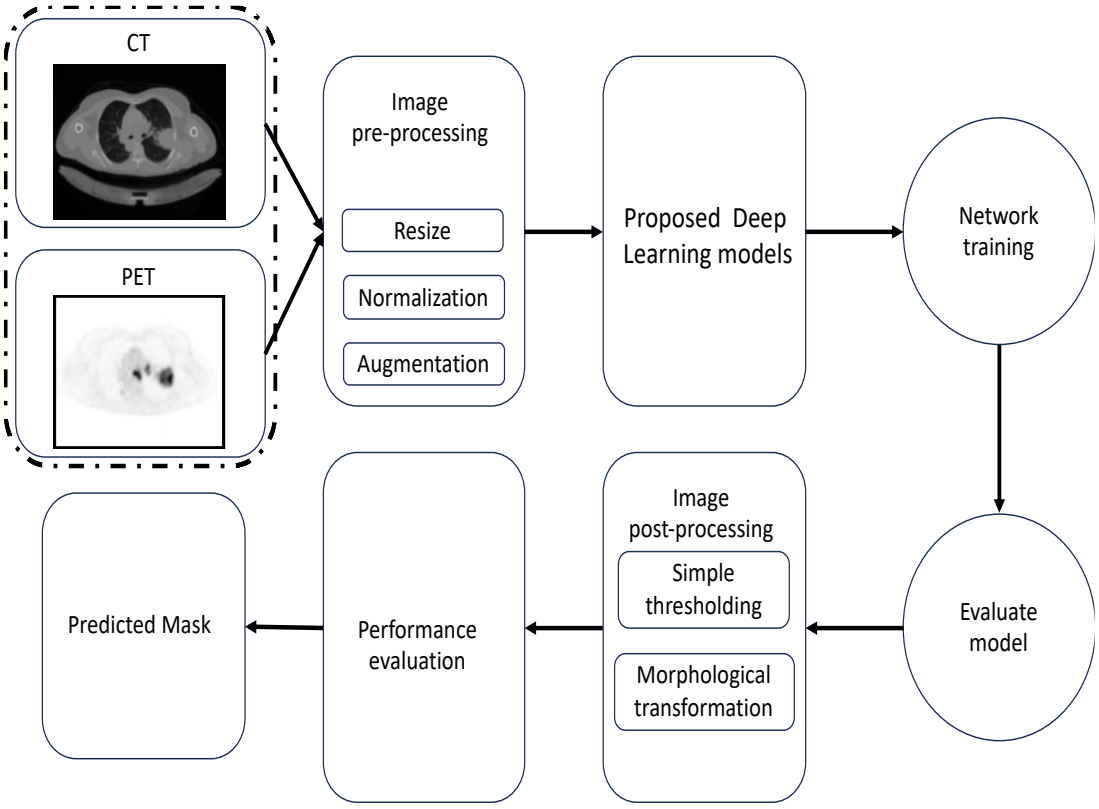

Figure 4.1. Block diagram for the lung cancer segmentation framework

The proposed architecture for lung cancer segmentation is shown in Figure 4.1. It depicts the three main stages of the framework: (1) image pre-processing, (2) multimodality U-net

segmentation, and (3) medical image post-processing. The pre-processing, augmentation, and post-processing methods are discussed in the following subsections.

4.3.1. Images Processing

4.3.1.1. Image Pre-processing

The intensity levels of the image's pixels were normalized to remove any potential for ambiguity. In addition to resizing each image, the pixel scale value was changed from (0 - 255) to (0 -1) to reduce the level of complexity of the images. To simplify model training, the resolution of the CT and PET scans is reduced in the dataset to 256 x 256 pixels. The dataset was divided as follows, at random: 46 examples were used for training, and another five were used for testing.

4.3.1.2. Data Augmentation

The CT-PET images are augmented throughout this phase to prevent overfitting, which helps in enhancing the performance of the model. In addition, the implementation of augmentation techniques, such as random rotations, flips, and cuts, can enhance the model's ability to maintain invariance towards variations in feature position and orientation within the image. This feature proves to be particularly advantageous when working with real-world images that may exhibit variations in object orientation or spatial arrangement.

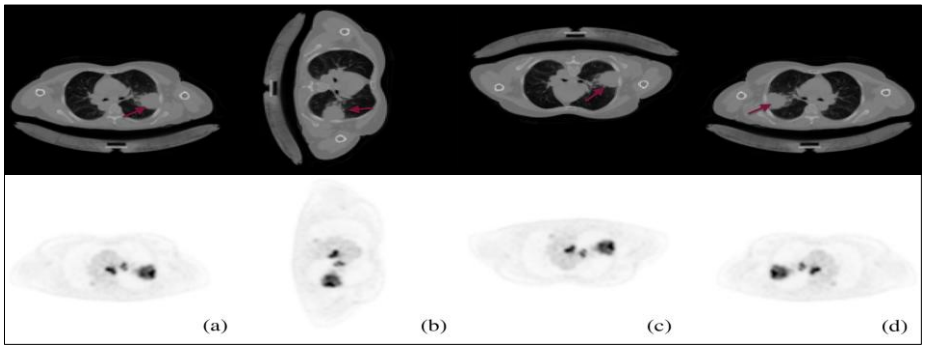

Figure 4.2. Some examples of the augmentation process of CT and PET images for STS: (a) the main CT-PET, (b) rotating the CT-PET by 90 degrees clockwise, (c) flipping the CT-PET upside down, and (d) left-mirroring the CT-PET. Red arrows indicate the tumor region.

Images are augmented in three ways, as shown in Figure 2: rotating the CT-PET by 90 degrees clockwise (2b), flipping the CT-PET upside down (2c), and left-mirroring the CT-PET (2d) as shown in Figure 4.2.

4.3.1.3. Image post-processing

The suggested framework's ultimate stage uses a morphological change and a basic thresholding technique. A morphological gradient accounts for the structure of the input picture to lessen the impact of noise. Its effect is analogous to the difference between expanding and contracting an image.

While Equation (4.1) defines dilation [101] as the process of removing pixels (noises) from object boundaries, Equation (4.2) describes erosion [102] as the process of adding pixels (negative noises) to object boundaries.

$$A \oplus B = \cup_{b \in B} A_b \tag{4.1}$$

$$A \,!\, B = \{z \in E | B \subseteq A\} \tag{4.2}$$

Where A is a set of pixels, and B is a structuring element.

The thresholding technique is defined as:

$$f(x) = \begin{cases} 1, & \text{if } x \geq t \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

where $\boldsymbol{x}$ represents the predicted pixel value and $\boldsymbol{t}$ is the threshold used to separate tumour pixels from the background. A value of $\boldsymbol{t = 0.5}$ was employed, meaning pixels with predicted values equal to or greater than **0.5** are considered tumour regions. This post-processing step ensures crisp binary segmentation boundaries and effectively reduces false positives near object edges.

Figure 4.3. depicts the last stage in processing predicted masks, in which tiny false positive values and blobs at the borders are removed.

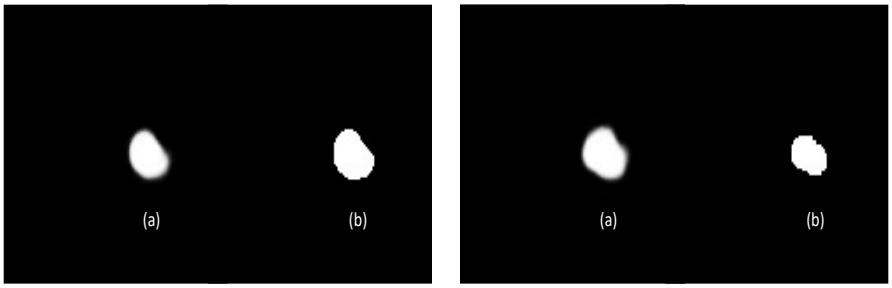

Figure 4.3. Two samples show post-processing effects: (a) predicted mask and (b) image after mask post-processing.

4.3.1.4. Multimodal Feature Fusion

A feature fusion strategy is deployed in medical imaging to generate a higher-quality final image. Professionals in the medical field view fusion processes as a helpful resource. Feature extraction, classification, and making decisions are the three main pillars of any supervised learning-based method. To broaden the types of features recovered and better understand their relationships, the early and late sequences of feature fusion are employed in the encoder portion of the core U-net design. Features from different imaging modalities, like PET and CT, are fused serially to characterize lung tumors better.

4.3.1.5. Early fusion

In early fusion, each medical image scan (CT and PET) has a single input path that contains two CNN layers with 64 units and a Relu activation function. Then, these two paths are concatenated into a single path, which is processed through a unique path in the down-sampling U-net path. This path contains three groups of CNN architecture; each group has three CNN layers with 128, 256, and 512 units, followed by a max-pooling layer. All CNN activation functions are Relu functions. Figure 4.4. shows the Early fusion architecture.

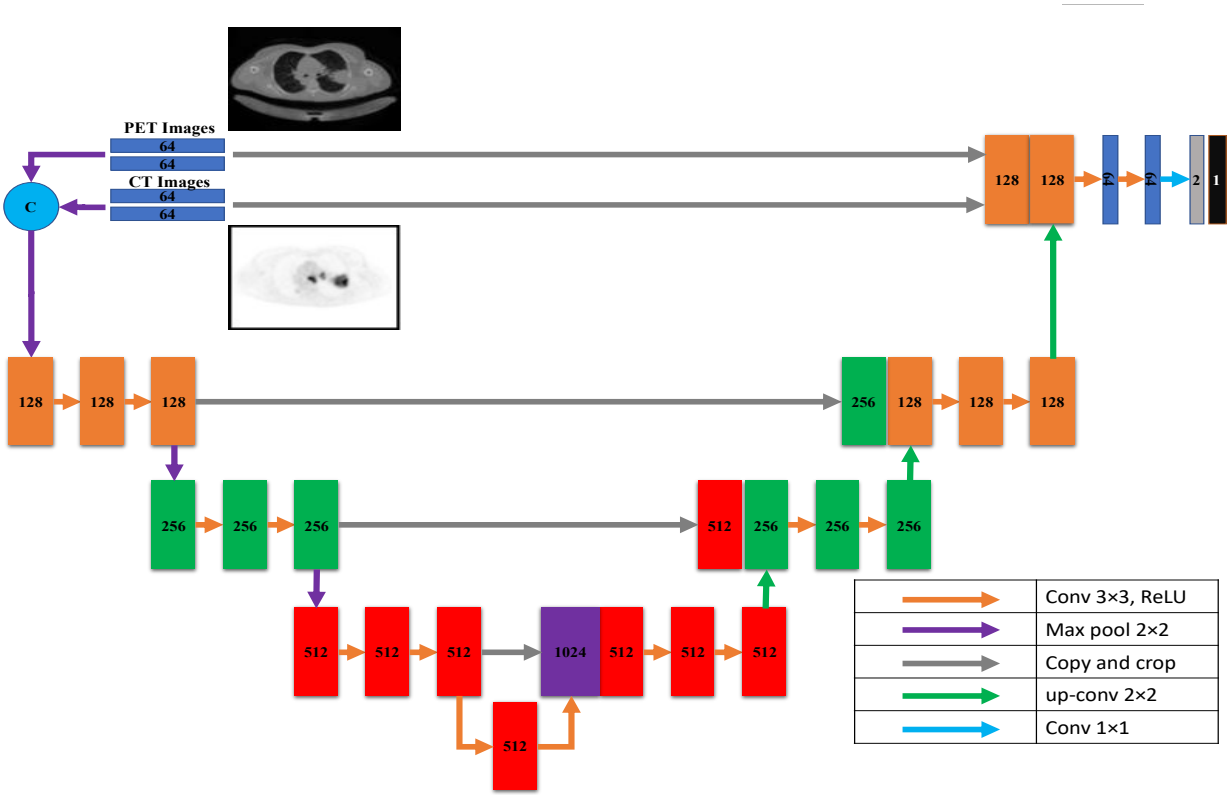

Figure 4.4. Early Fusion Architecture

4.3.1.6. Late fusion

In contrast to most architectures like U-Net, the encoding path is divided into N streams that serve as input for each imaging modality. Each modality learns a unique feature set using images from the other. The two modalities' feature maps are combined at each network's high-level feature layer. This process solves the problem of early fusion strategy. These feature sets are combined into one feature set and then subjected to the last phase of a multimodal classifier's training. The U-net down-sampling path contains four groups of CNN layers. Each group contains three sequential CNNs with several units, 64, 128, 256, and 512 units, respectively, followed by a max-pooling layer. All CNNs have a Relu activation function. At this point, the two paths are concatenated to generate the input of the U-net Up-sampling path. Figure 4.5. shows the late fusion architecture.

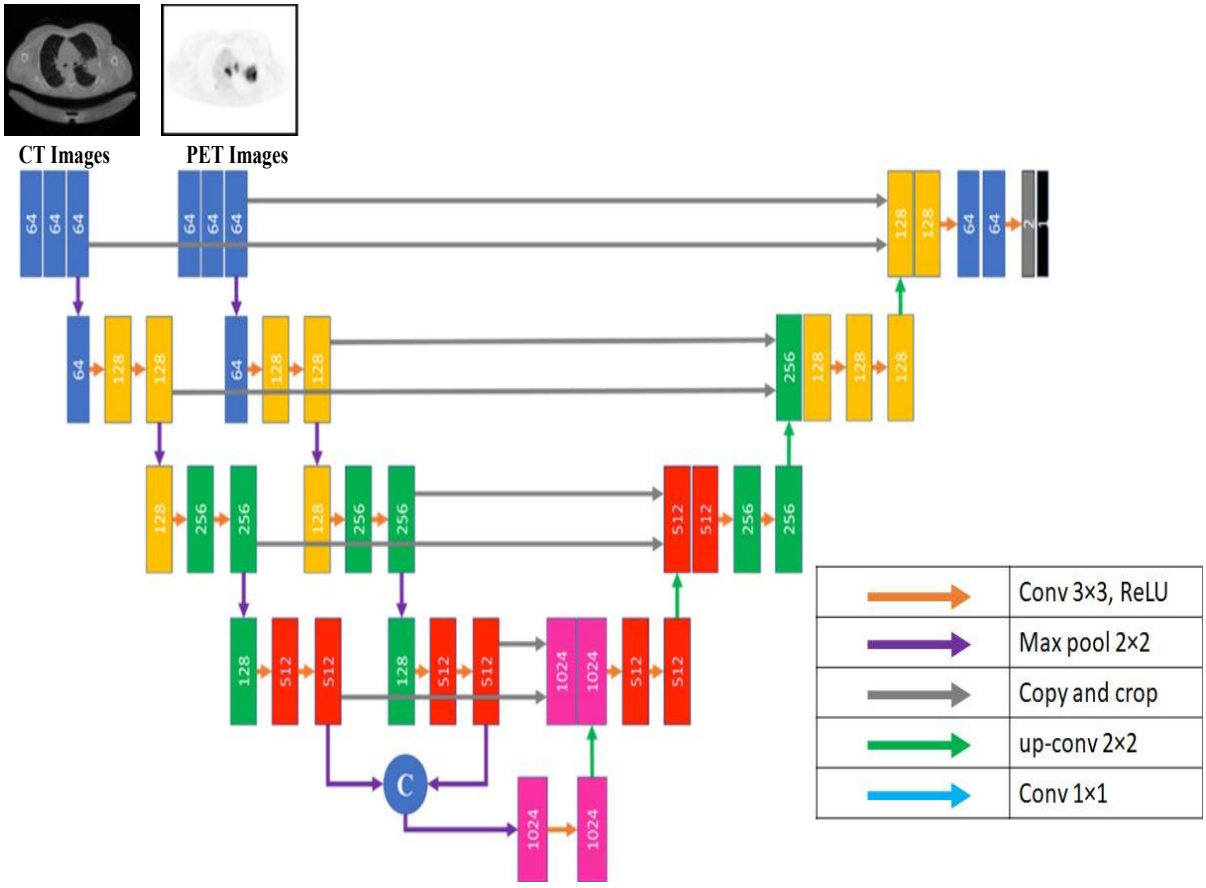

Figure 4.5. Late Fusion Architecture

4.3.1.7. Dense fusion

For lung cancer segmentation, the Dense fusion-based U-net provides two down-sampling routes, one for CT and one for PET images. Eight CNN deep learning building blocks are used along each possible route. All the layers preceding the current layer are inputs to the current

CNN layer. A max-pooling layer follows each pair of consecutive CNN layers. The dimensions of the CNN layer are (in order) 64, 128, 256, and 512. The Relu activation function is standard in all CNNs. The input to the U-net Up-sampling path is generated by concatenating the outputs of the paths following the design described in each path. The dense fusion architecture is shown in Figure 4.6.

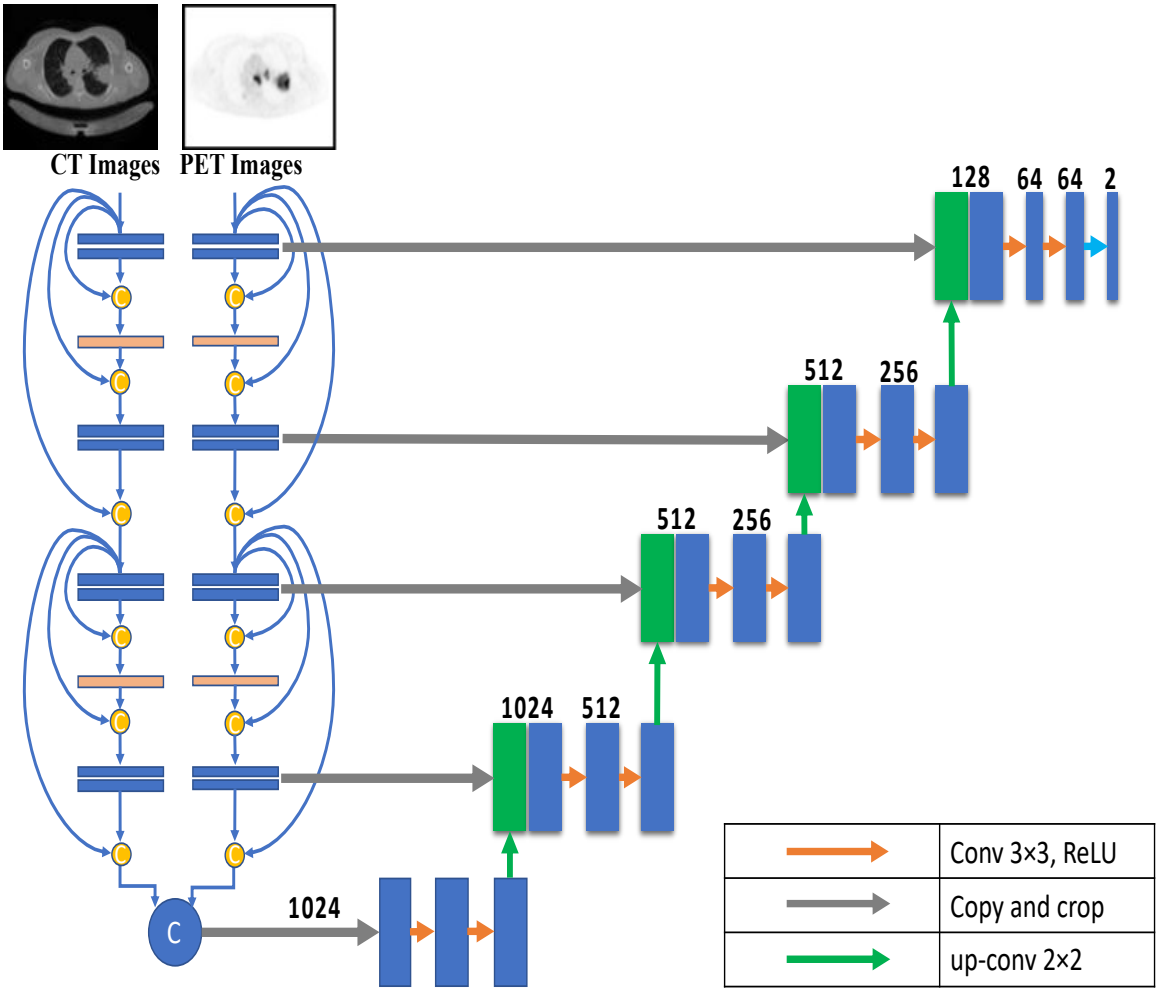

Figure 4.6. Dense Fusion Architecture

4.3.1.8. Hyper dense fusion

Deep learning is essential when an application requires a deep layer to function effectively and efficiently. Reducing the overfitting impact is one of several benefits of using dense architecture for multimodality U-net medical image segmentation. The layers in the same input path provide inputs to all net layers for dense design, which is necessary for U-nets with multiple input paths. Each layer feeds its immediate successor and those in adjacent input channels in hyper-dense fusion. As the network learns the intricate connections between the modalities at each level of abstraction, the hyper-dense connectivity produces a more robust

feature representation than early/late fusion in a multimodal situation. The hyper-dense fusion layout is depicted in Figure 4.7.

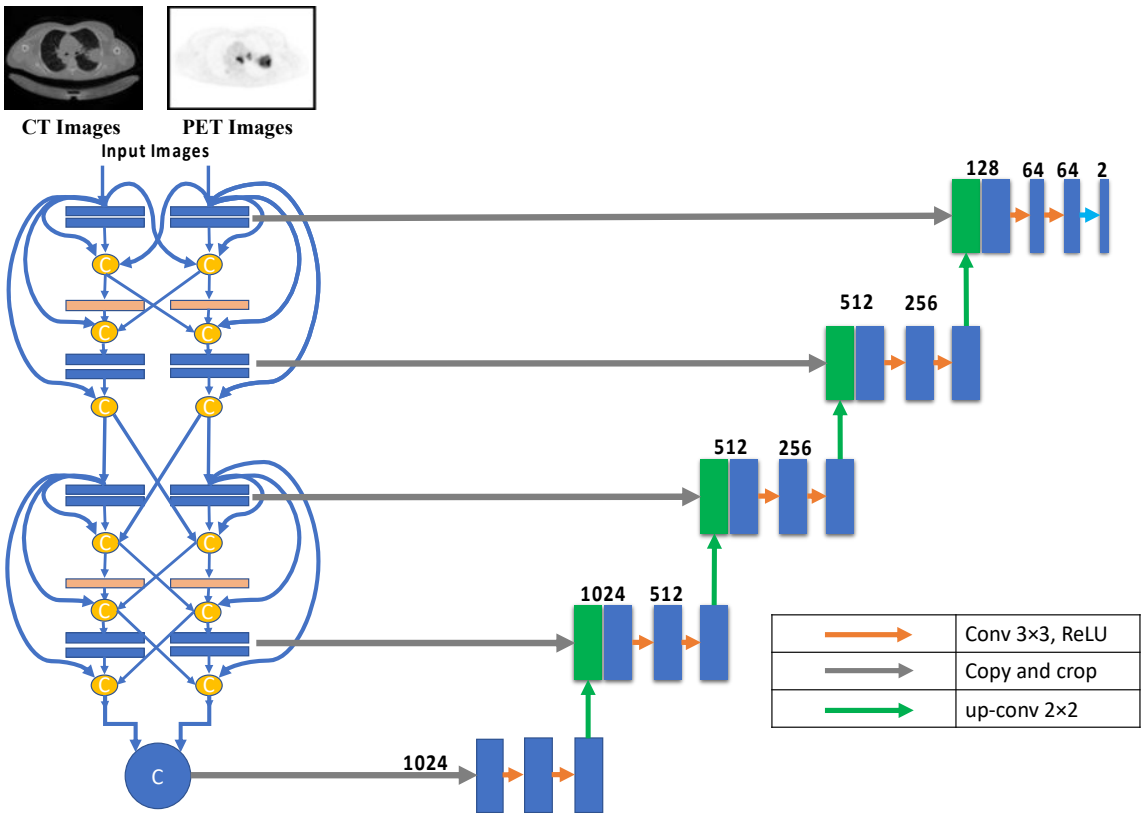

Figure 4.7. Hyper-Dense Fusion Architecture

4.3.2. Loss Functions

In the proposed method, a thorough investigation and comparison of the models using a variety of loss functions is performed. Segmenting an image is essentially a pixel-level classification problem. Each pixel in an image contributes to the overall image, and specific clusters of pixels define particular aspects. Semantic image segmentation is a technique that divides these pixels into their respective components. While designing intricate, deep learning architectures for image segmentation, choosing the loss/objective function is crucial. Loss functions can be broken down into several types based on distribution, region, boundary, and compound. The proposed analysis uses three distinct loss functions, i.e., binary cross-entropy, dice, and focal. The representation as the network discovers the many interconnections between modalities at every level of abstraction, rather than the binary early/late fusion approach.

4.3.2.1. Binary Cross-Entropy

The Binary Cross-Entropy **(BCE)** loss function [103] is widely employed for binary segmentation tasks to measure the difference between predicted probabilities and ground truth labels. It is defined as in Eq. (4.4).

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{4.4}$$

where $\boldsymbol{N}$denotes the total number of training samples, $\boldsymbol{y_i} \in \{\mathbf{0}, \mathbf{1}\}$ represents the true label for each sample, and $\boldsymbol{\hat{y}_i} \in [\mathbf{0}, \mathbf{1}]$ is the predicted probability obtained using a sigmoid activation function:

$$\hat{y}_i = \frac{1}{1 + e^{-\mathrm{w}\cdot \mathrm{x}_i}}$$

The BCE loss penalizes large deviations between predicted probabilities and the corresponding true labels. It is averaged across all samples and backpropagated to update the model weights during training, ensuring optimal discrimination between tumour and non-tumour regions.

4.3.2.2. Focal Loss

The Focal Loss (FL) [104] can be defined as a modification of the Binary Cross-Entropy loss to address class imbalance by focusing more on hard-to-classify samples. The equation is expressed as in Eq. (4.5).

$$FL(p_t) = -\alpha_t(1 - p_t)^{\gamma}\log(p_t) \tag{4.5}$$

where $\boldsymbol{p_t}$ represents the predicted probability of the true class, $\boldsymbol{\alpha_t}$ is a weighting factor that balances the contribution of different classes, and $\boldsymbol{\gamma > 0}$ is the focusing parameter that reduces the loss contribution from well-classified examples. When $\boldsymbol{\gamma = 1}$, the loss reduces to the standard Binary Cross-Entropy loss. In this study, $\boldsymbol{\alpha_t}$ was set according to the inverse class frequency, and $\boldsymbol{\gamma = 2}$ was used following common practice in imbalanced segmentation tasks.

4.3.2.3. Dice Loss

The Dice Loss (DL) is derived from the Dice coefficient, which quantifies the overlap between predicted and ground truth masks [105]. The loss is formulated as in Eq. (4.6).

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i\,\hat{y}_i + \epsilon}{\sum_{i=1}^{N}(y_i + \hat{y}_i) + \epsilon} \tag{4.6}$$

where y_i represents the ground truth label for each pixel, $\hat{y}_i$ denotes the predicted probability, N is the total number of pixels, and ϵ is a small constant added for numerical stability. This formulation penalises low overlap and encourages better alignment between the predicted and true segmentation regions.

Hyper Dense VGG16 U-Net Segmentation Proposed Model

In various computer vision problems, shortcut connections between layers have become increasingly popular since the emergence of residual learning [106]. Unlike in conventional networks, these links back-propagate gradients immediately, which helps prevent gradient-vanishing issues and allows for more complex architectures. The idea of shortcut connections was expanded upon by DenseNet [107], which specified that each layer's inputs should correspond to the outputs of all the layers that came before them. Densely connected convolutional neural networks (CNNs) are built using the feed-forward principle, which entails adding direct connections from any layer to all succeeding layers. Deep networks are more accessible and more accurate to train because of this connectivity. This section proposes independently expanding U-Net to support DenseNet connections within the same multiple N streams of PET and CT modalities. Higher-level layers of the proposed extension will also use the late fusion strategy.

The inspiration for this comes from three separate observations. First, all architectural feature maps are connected by short paths, enabling implicit deep supervision. Second, the network's information and gradients are better able to flow because of the direct connections between all layers. Finally, the regularizing effect of dense connections makes it less likely that training data will be too small for a given task.

Using dense and hyper-dense connections has been demonstrated to have many benefits when segmenting medical images. When the VGG architecture is used for feature extractions, more information can be gleaned from medical images. A multimodality U-net medical image segmentation model is proposed using hyper-dense connections and the VGG16 model.

The primary objective was to refine an existing deep-learning model for lung cancer segmentation. To do this, the U-Net design is modified and used as the starting point. The encoder and the decoder are both CNNs, making up the basic U-Net architecture. The encoder extracts features by first performing convolutional operations and then down-sampling. The usual convolutional processes follow the up-sampling and concatenation layer of the decoder

branch. Connecting feature maps from the encoder network is made possible via a skip link that connects the same-level layers of the decoder and encoder, with the up-sampled feature map conveying coarse global context information. It helps with recovering local characteristics after down-sampling. According to this model, U-net takes data via two distinct input paths, one for each image type. The architecture of both paths is VGG16, with dense and hyper-dense connections between them. This architecture was proposed so that image classification and segmentation tasks may take advantage of VGG, dense, and ultra-dense networks—the suggested VGG16 U-net model's components are given in Figure 4.8.

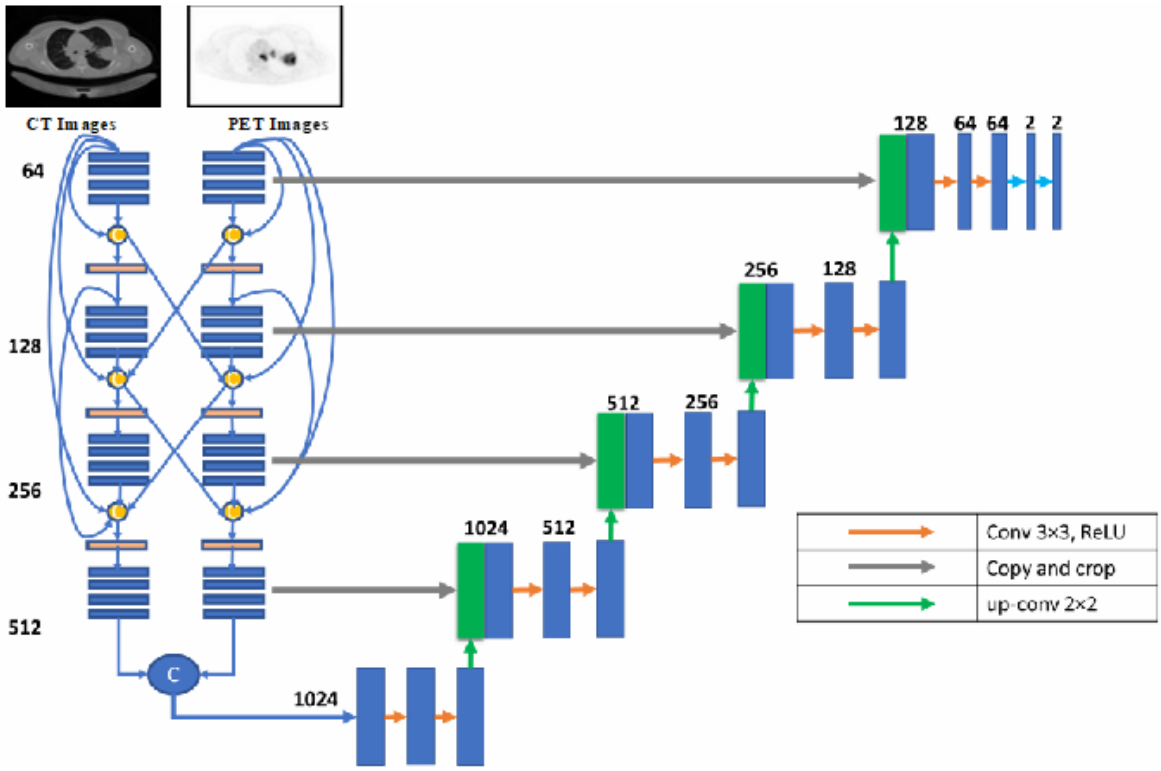

Figure 4.8. The proposed hyper-dense VGG16 U-Net model architecture

Figure 4.8. depicts his proposed Hyper dense VGG16 U-Net model, built upon the U-Net. Both CT and PET images can be fed into the model. The segmented image of lung cancer is the product of the model. In the suggested approach, input images for both CT and PET were 128x128. Each image input type has its dedicated input path, each with 16 CNNs (the number of CNNs in VGG16). Each data set was processed through CNNs of varying sizes (64, 128, 256, and 512). Both input paths are incredibly well-connected, and there are also many connections between the two. All convolutional neural networks used ReLU activation. The decoder's structure comprises four groups of convolutional neural networks (CNNs) of varying sizes (1024, 512, 256, and 128).

4.4. Experiments

The efficiency of the proposed U-Net models for segmenting lung tumors was measured across various performance criteria. The STS dataset was used for both training and testing the models. Experiments compared the newly developed models to benchmarked models widely utilized on the same dataset and other datasets.

4.4.1. Experimental Setup

All experiments were run on servers in the Google Colaboratory environment, and the recommended models for segmenting lung tumors were built using a TensorFlow and Keras backend with an NVIDIA Tesla P100 -PCIE GPU and 32.0 GB RAM. For the training phase, the Adam optimizer is employed with the following settings: learning rate=0.0001, 1=0.9, 2=0.999, and epsilon=1×10^{-8}. One hundred epochs of training were used. The intensity levels of the image's pixels were normalized to remove any potential for ambiguity. The dataset was arbitrarily divided into 70% for training, 20% for validation, and 10% for testing.

4.4.2. Dataset Description

The proposed models are trained on data from a study of soft tissue sarcomas (STSs) [108]. STS includes many types of scans: CT, PET, and MRI, but in this research, CT and PET were used only. In this dataset, a cohort of 51 patients with histologically proven soft-tissue sarcomas (STSs) of the extremities was retrospectively evaluated. With 38,328 images (each patient has around 200-300 images). It included 27 females and 24 males, ranging in age from 16 to 83 years. Also, with various cancer degrees: low, intermediate, and high. The PET slice volumes had a thickness of 3.27 mm and a median in-plane resolution of 5.47 mm x 5.47 mm (range: 3.91–5.47 mm). All images used in the tests were downsized to 128 pixels on the longest dimension.

4.4.3. Performance Metrics

The efficiency of the suggested approach was assessed using the most commonly employed metrics for evaluating segmentation tasks[109]: the Dice score ($Dice$), the most crucial segmentation performance measure. It is defined by Equation (4.5).

$$Dice = \frac{(2 * Tp)}{(2 * Tp + Fp + Fn)} \tag{4.5}$$

In addition, measures of accuracy, sensitivity, and specificity. Equations (4.6) to (4.8) also provide definitions for them.

$$Accuracy = \frac{Tp + Tn}{Tp + Tn + Fn + Fp} \quad (4.6)$$

$$Sensitivity = \frac{Tp}{Tp + Fn} \quad (4.7)$$

$$Specificity = \frac{Tn}{Tn + Fp} \quad (4.8)$$

Where the four primary blocks for computing these metrics were defined as true positive (Tp), true negative (Tn), false positive (Fp), and false negative (Fn) values.

4.5. Results

The effectiveness of the proposed models is discussed in this section. In this section, the results of the model assessments are reported and divided into four categories: loss function comparisons, same-dataset comparisons, cross-dataset comparisons, and cross-model comparisons. Dice, IoU, Accuracy, Spectral Sensitivity, and Area under the Curve (AUC) were utilized as performance measures.

4.5.1. Loss Functions-Based Comparison

Adjustments to the loss functions form the basis for a new comparative evaluation of the models. Focal loss functions, dice, and binary cross entropy are employed in this research. These operations are among the most well-known and often used in deep learning for image segmentation. The outcomes are displayed in Tables 4.1.- 4.3. for Binary, Dice, and Focal loss functions.

Table 4.1. Binary Cross-Entropy

	Dice	IOU	ACC	Sen	Spec
Late	0.67882	0.53651	0.98516	0.73885	0.99068
Early	0.68066	0.54075	0.98397	0.73816	0.99083
Dense	0.69569	0.54016	0.98095	0.68401	0.99225
Hyper	0.71851	0.57818	0.98381	0.72302	0.99284
Hyper+VGG16	0.72532	0.58687	0.98278	0.69209	0.99423

Table 4.2. Dice Loss Function

	Dice	IOU	ACC	Sen	Spec
Late	0.51479	0.69734	0.97806	0.64604	0.99048
Early	0.51465	0.69671	0.97993	0.65452	0.99135
Dense	0.51485	0.51191	0.98112	0.67253	0.99112
Hyper	0.64081	0.69725	0.98295	0.67958	0.99046
Hyper+VGG16	0.66828	0.52222	0.98048	0.69506	0.99102

Table 4.3. Focal Loss Function

	Dice	IOU	ACC	Sen	Spec
Late	0.71217	0.5704	0.98347	0.71046	0.99327
Early	0.66112	0.51011	0.97943	0.6351	0.99232
Dense	0.71554	0.57403	0.98198	0.70131	0.99347
Hyper	0.72713	0.58717	0.98436	0.71786	0.99339
Hyper+VGG16	0.73011	0.55664	0.98103	0.67472	0.99362

Figures 4.9.-4.11. depict the findings using various loss functions, like cross-entropy, focal loss, and dice loss. In contrast, the performance measures for the proposed models using the metrics Dice, IoU, Accuracy, Sensitivity, and Specificity are given in Figures 4.12.-4.16., respectively.

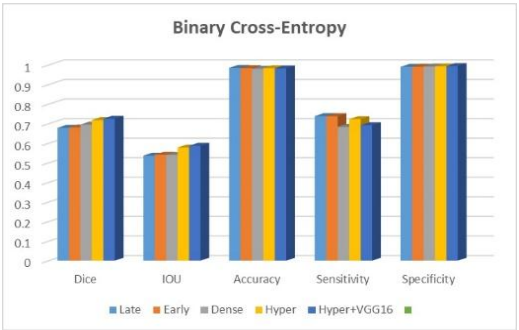

Figure 4.9. Binary Cross-Entropy Function

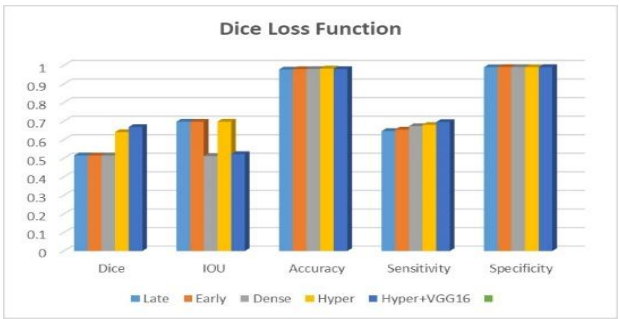

Figure 4.10. Dice Loss Function

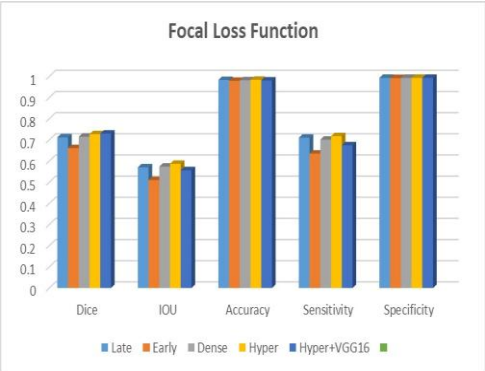

Figure 4.11. Focal loss function

Figure 4.12. Dice Metric

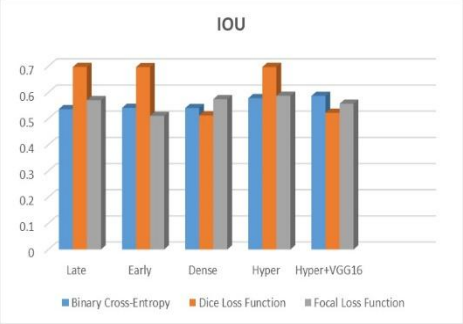

Figure 4.13. IOU Metric

Figure 4.14. Accuracy Metric

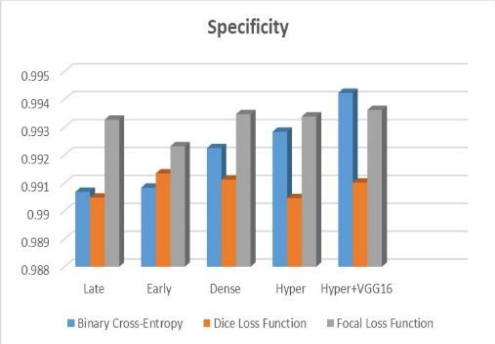

Figure 4.15. Specificity Metric

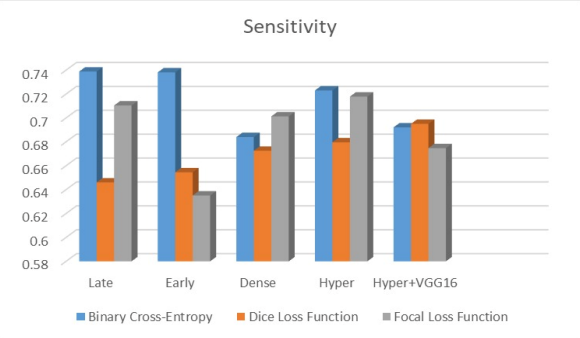

Figure 4.16. Sensitivity Metric

The suggested hyper-dense VGG16 model outperforms the other models in Dice for all types of loss functions, as seen in Tables 4.2.-4.4. The Focal loss function is the only option if you want the best dice performance possible. Figures 4.9.– 4.16. offer graphical representations of the evaluation outcomes.

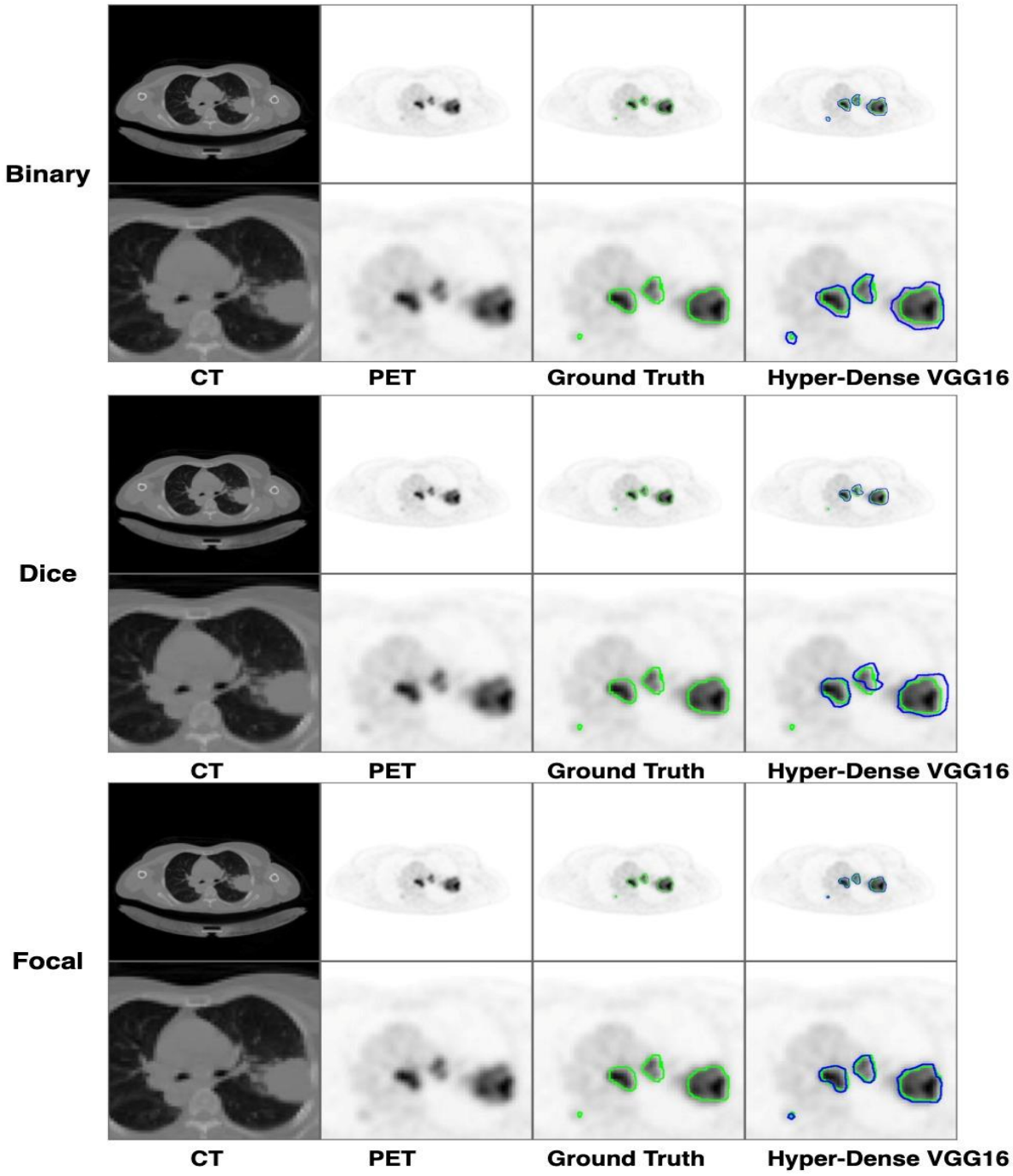

Figure 4.17. The comparison of lung tumor segmentation results, along with the segmentation outcomes for corresponding enlarged tumor regions, using the proposed hyper-dense VGG16 model with various loss functions ("Binary," "Dice," and "Focal"). The green contours outline the "Ground Truth" segmentation, and the blue contours outline the results from the proposed model.

The five presented models are compared in Figures 4.9.–4.12. regarding the binary cross entropy, dice, and focused loss functions used as performance indicators. Figure 4.9. shows that the suggested hyper VGG16 model outperforms the others in terms of dice accuracy (improved by 7%), IOU accuracy (improved by 9%), and specificity accuracy (improved by 0.4%). However, the late fusion model's accuracy and sensitivity are unparalleled. The results of the dice loss function are shown in Figure 4.10., and it is evident that the suggested model outperforms the previously introduced models in terms of dice, specificity, and sensitivity. Finally, the proposed model outperforms the other established models regarding the focused loss function performance, achieving 73% for Dice. Figures 4.10.-4.13. presented visual representations of the performance above metrics about the loss function employed. Figure 4.13. demonstrates that the most outstanding value for the focused loss function is found with the dice metric. The segmentation results of the proposed model for various loss functions are displayed in Figure 4.17. The lung tumor segmentation results generated by hyper-dense VGG16 are compared to the ground truth, employing various loss functions such as binary, Dice, and focal. The observations from Figure 4.17. indicate that the focal loss function yields the most accurate predictions, capturing even the segmentation of small tumor portions and producing a predicted segmentation mask that closely aligns with the ground truth segmentation. Conversely, when utilizing the binary cross-entropy loss function, the segmentation results tend to be slightly larger. The Dice loss function, however, provides the least accurate predictions, as it fails to segment small tumor portions and produces a larger overall segmentation compared to the ground truth.

To comprehensively evaluate the performance of the proposed Hyper-Dense VGG16 architecture, five fusion strategies—Early, Late, Dense, Hyper-Dense, and the proposed Hyper-Dense VGG16—were tested and compared using both the STS dataset and other benchmark datasets. As presented in Table 4.4., all five fusion models achieved superior Dice coefficients compared with the two reference studies (Fu et al. [72] and Bi et al. [79]), which are the only published works that applied their segmentation models to the same STS dataset. The Dice coefficient is recognized as the most significant and widely used metric for medical image segmentation, particularly in tumour delineation, as it measures the spatial overlap between predicted and ground-truth regions. The superior Dice performance of all five fusion methods—including the proposed Hyper-Dense VGG16—demonstrates the effectiveness of multimodal PET–CT feature fusion in improving lesion boundary accuracy.

Table 4.4. Comparison of The Proposed and Benchmarked Models on The STS Dataset

	Dice	IOU	ACC	Sen	Spec
Fu et al. [72]	0.6226	-	-	0.6474	0.997
Bi et al. [79]	0.6636	-	-	0.6993	0.9969
Late	0.712171	0.5704	0.98347	0.71046	0.99327
Early	0.661116	0.51011	0.97943	0.6351	0.99232
Dense	0.715539	0.57403	0.98198	0.70131	0.99347
Hyper	0.72713	0.58717	0.98436	0.71786	0.99339
Hyper+VGG16	0.730109	0.55664	0.98103	0.67472	0.99362

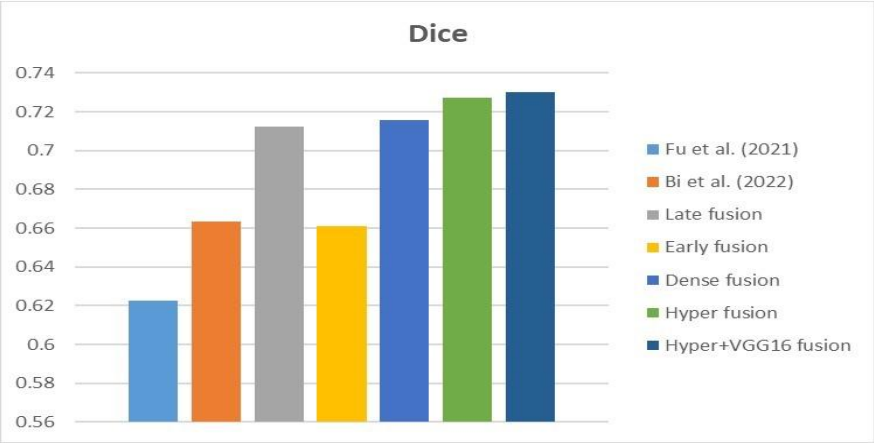

Figure 4.18. Dice

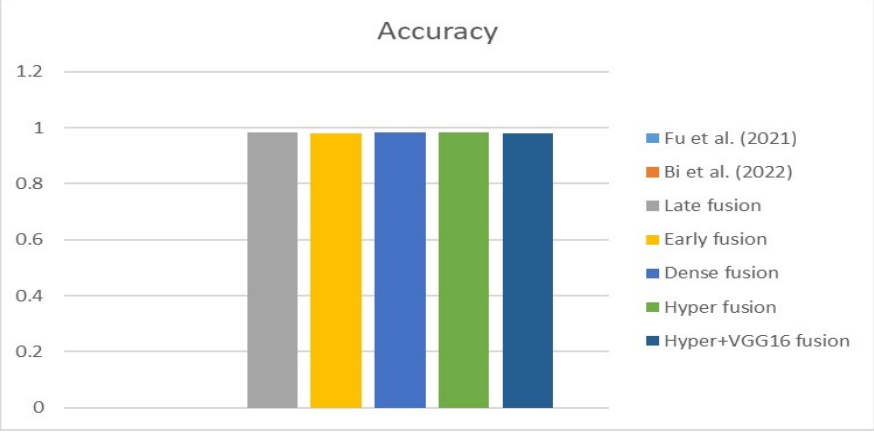

Figure 4.19. Accuracy

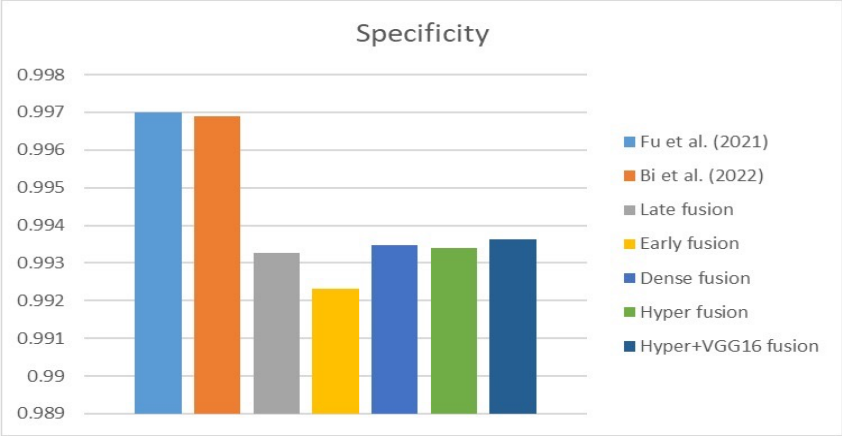

Figure 4.20. Specificity

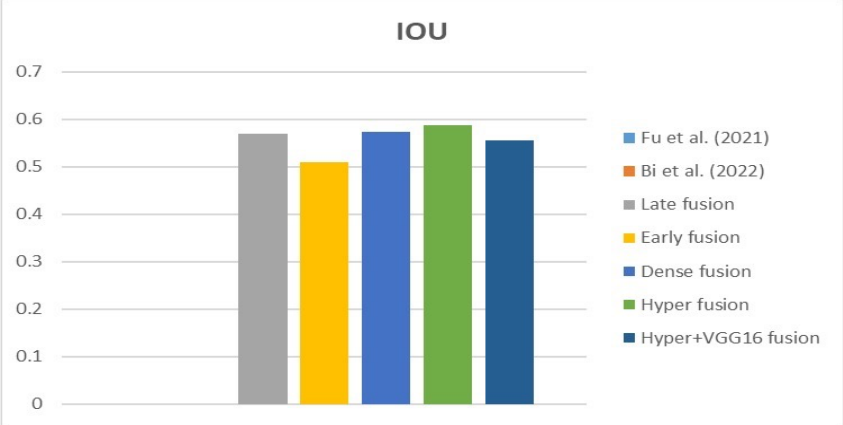

Figure 4.21. IOU

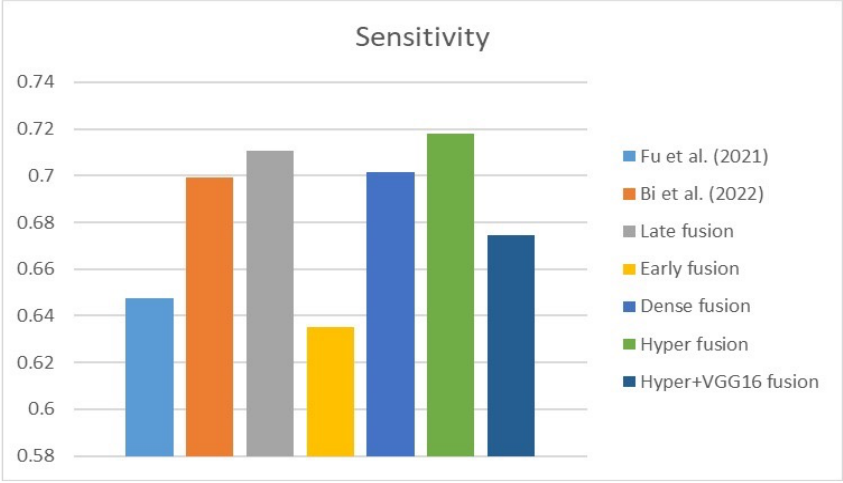

Figure 4.22. Sensitivity

Although the proposed Hyper-Dense VGG16 model achieved the highest Dice (0.7301) and excellent specificity (0.9936), it did not always outperform the other tested strategies across all secondary metrics. For example, the Hyper-Dense model achieved slightly higher sensitivity (0.7178) and IoU (0.5871), while the Late Fusion model achieved marginally better accuracy (0.9834). These differences arise from the trade-off between feature abstraction and pixel-level recall. The integration of the VGG16 encoder in the proposed model deepens spatial representation and strengthens global context learning, resulting in smoother, more consistent segmentation boundaries and fewer false positives. However, this same regularization effect may slightly reduce recall for subtle or irregular tumour edges, lowering sensitivity and IoU. Conversely, shallower configurations such as Hyper-Dense or Late Fusion respond more directly to local intensity variations, improving sensitivity but at the cost of over-segmentation or reduced generalization. Figures 4.18.- 4.22. show visual representations of the performance metrics that were used in the performance evaluation and comparison of the proposed models.

4.5.2. Different Datasets in the State-Of-The-Art

To further validate generalizability, a cross-dataset comparison was conducted using Table 4.5., which evaluates the same five fusion models against Fu et al. [72] and Kumar et al. [75], where Fu et al. [72] used both the STS and an additional dataset, and Kumar et al. [75] applied its method to a completely different PET–CT dataset. These studies were included because they represent the most relevant state-of-the-art multimodal segmentation frameworks, enabling a fair methodological benchmark despite dataset differences. Once again, all five tested models substantially outperformed both reference studies in terms of Dice coefficient, reaffirming that the proposed fusion approaches—especially Hyper-Dense VGG16—are capable of achieving accurate tumour segmentation across varied imaging conditions. Similar to the STS dataset results, the proposed model did not lead in every secondary metric: Hyper-Dense showed slightly higher sensitivity (0.7178), and Late Fusion marginally exceeded it in accuracy (0.9834). These small variations reflect dataset-specific image characteristics such as tumour size, intensity distribution, and PET–CT registration consistency. The proposed model's deeper cross-modal connections provide strong generalization and denoising ability, ensuring stable Dice and specificity scores across datasets, while occasional minor reductions in sensitivity result from its smoother boundary regularization. Figures 4.23.-4.27. show visual representations of the performance metrics used in performance evaluation and comparison of the proposed models.

Overall, the proposed Hyper-Dense VGG16 model consistently achieved the highest Dice coefficient and maintained strong performance across all other metrics on both the STS and different datasets. These results confirm that while the model prioritizes balanced precision–recall trade-offs rather than overfitting to a single metric, it remains the most robust and generalizable architecture for multimodal lung tumour segmentation, outperforming existing state-of-the-art approaches in clinical relevance and stability.

Table 4.5. Comparison of The Proposed and Benchmarked Models on Different Datasets

	Dice	**IOU**	**ACC**	**Sen**	**Spec**
Fu et al. [72]	0.6783	-	-	0.999	0.7616
Kumar et al. [75]	0.6385	-	-	-	-
Late	0.712171	0.570397	0.983471	0.710462	0.993269
Early	0.661116	0.510114	0.979428	0.635095	0.992316
Dense	0.715539	0.57403	0.981976	0.701314	0.993468
Hyper	0.72713	0.587171	0.984362	0.717861	0.993387
Hyper+VGG16	0.730109	0.556635	0.981034	0.674717	0.99362

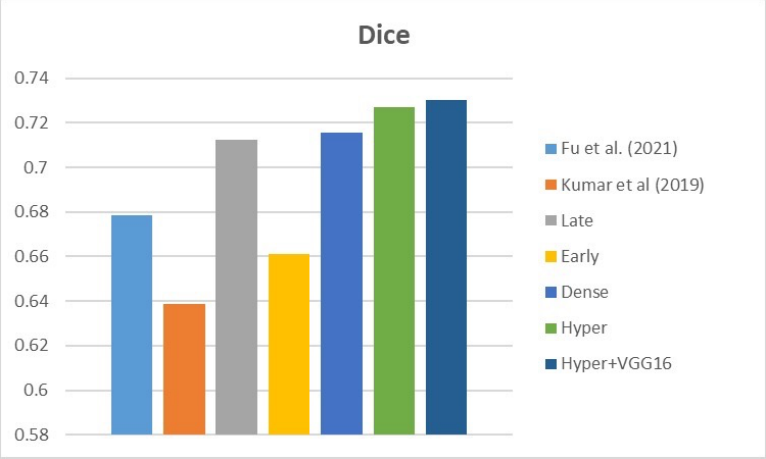

Figure 4.23. Dice

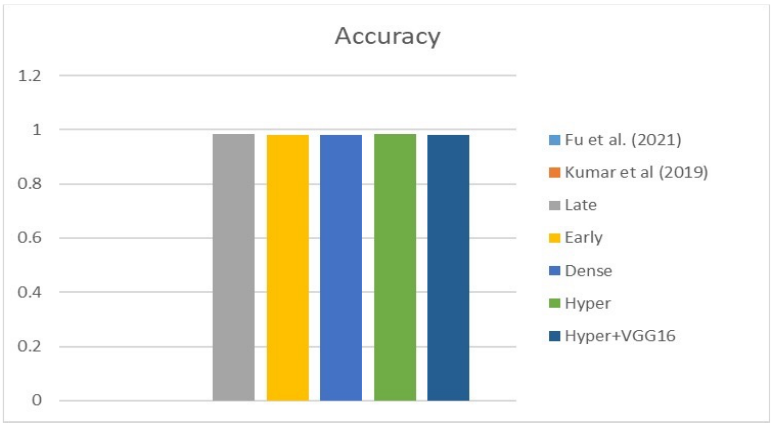

Figure 4.24 Accuracy

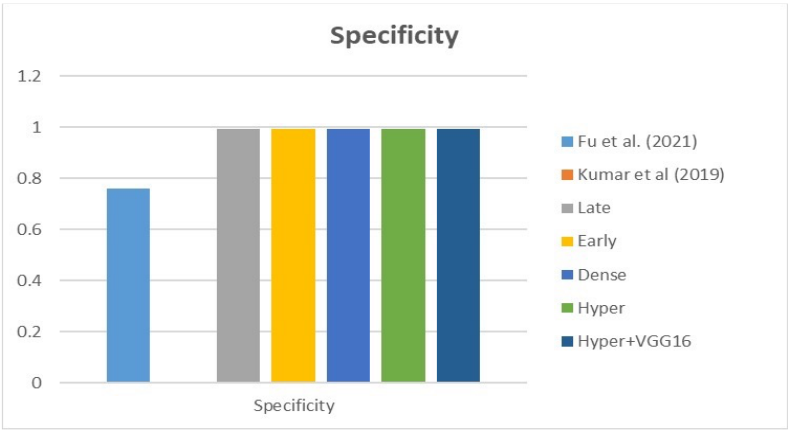

Figure 4.25. Specificity

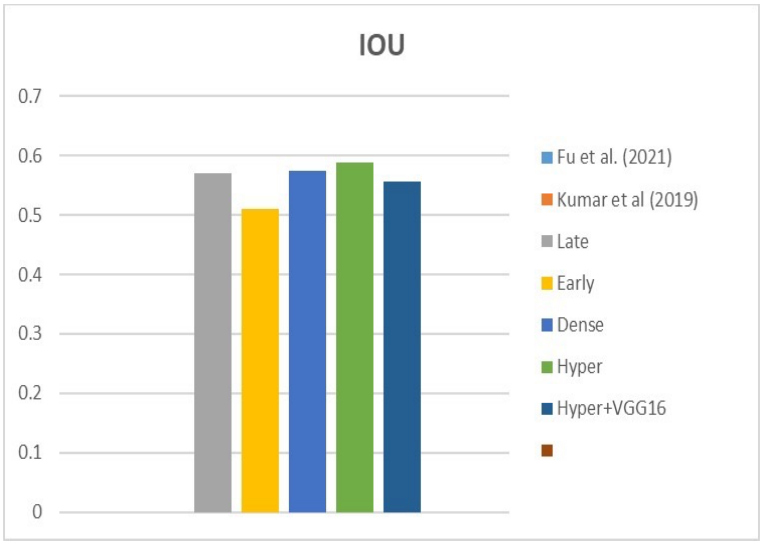

Figure 4.26. IOU

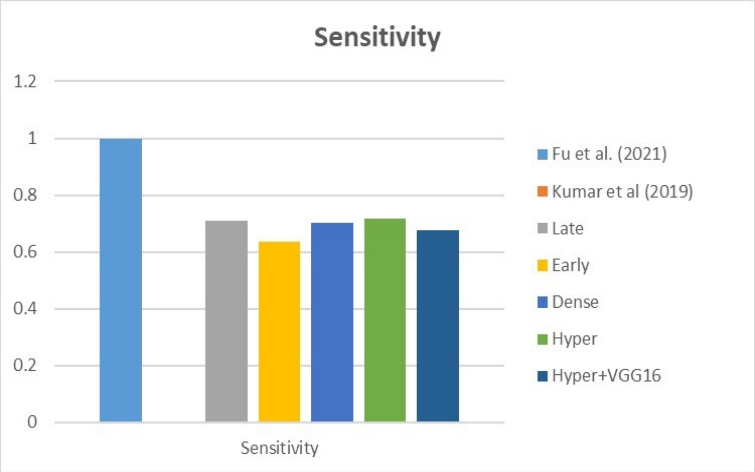

Figure 4.27. Sensitivity

4.6. Discussion

A CT-PET dataset of 51 STS samples was used to assess the five fusion models based on U-Net: Early Fusion, Late Fusion, Dense Fusion, Hyper Dense Fusion, and Hyper Dense VGG-16. A broad category of malignant tumors that start in the body's connective tissues, including muscles, fat, and fibrous tissue, are known as soft tissue sarcomas (STS). When these tumors spread to the lungs, they make imaging-based segmentation more difficult because they have different shapes, fuzzy tissue edges, and a tendency to look like lung tissue around them. Metastatic STS tumors, in contrast to primary lung tumors, necessitate specific methods for accurate delineation since they frequently exhibit distinct radiographic features that make conventional segmentation methods more challenging.

Several augmentation strategies were used to incorporate heterogeneity and improve model resilience, considering the relatively small size of the CT-PET dataset. The images were left-mirrored, rotated 90 degrees, and inverted as part of the augmentations. These additions made the models more accurate at predicting a wider range of tumor types because they mimicked the natural variety that can be found in clinical settings. In order to improve the model's performance and prevent overfitting, these strategies artificially increased the variety of the training set. This was especially helpful while working with a small number of STS samples.

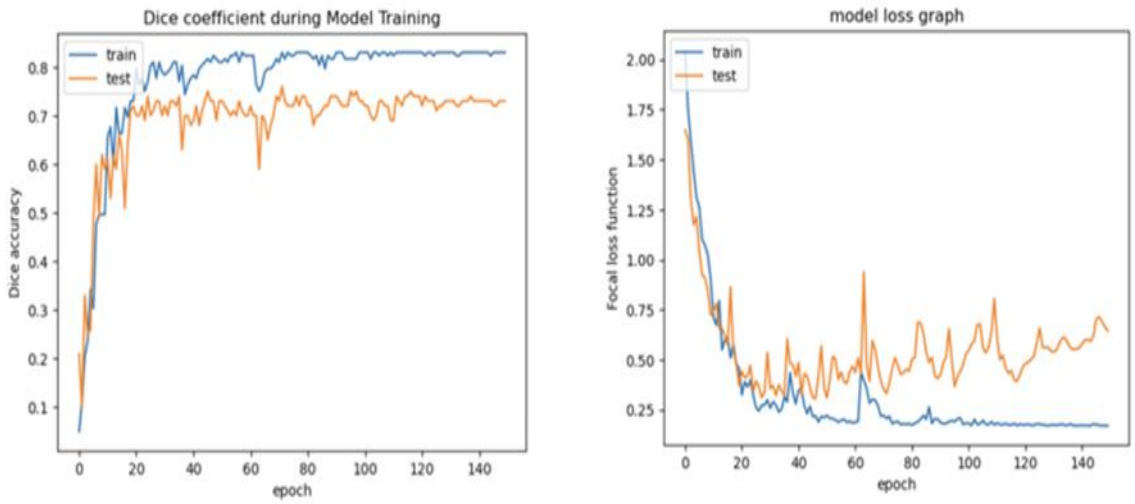

Figure 4.28. Training and Validation Accuracy and Loss Curves for hyper-dense VGG16

The Hyper Dense VGG-16 model demonstrated the best results with a Dice of 0.73 in handling the complexities associated with STS segmentation. Its deep, layered structure, in combination with the VGG-16 backbone, enabled it to more accurately capture the fine details of STS tumor boundaries, which are often difficult to distinguish from surrounding healthy tissue. This model's success underscores the importance of balancing depth with feature retention, a critical aspect hence dealing with the heterogeneous and irregular shapes of metastatic tumors in CT-PET imaging.

As shown in Figure 4.28., the training and validation accuracy and loss curves for the Hyper Dense VGG16 model reveal a steady improvement in performance over time. The Dice coefficient, which is a critical metric for segmentation tasks, demonstrated consistent growth during training. Training accuracy showed a significant learning curve, rising from an initial value of roughly 0.05 to over 0.83. Although progress was initially slow, the model's capacity to stabilize and generalize effectively was demonstrated by the fact that accuracy had reached a plateau at approximately 0.82 by the 50th epoch. On the other hand, there was a more noticeable variation in the validation accuracy, which started at 0.21 and steadily increased to 0.73 by the end of training. Even though there were some changes in the middle epochs, the overall trend showed that the model fit the data excellently, even though there were some problems with generalization. These variations are common among deep learning models, especially when fine-tuning is necessary for hard segmentation tasks. However, the model's ultimate convergence to a greater accuracy highlights how reliable it is at identifying significant

traits. The consistent growth in the Dice coefficient, as well as the stability of both training and validation accuracy, strongly suggest that the Hyper Dense VGG16 model, which combines the Hyper Dense design with the VGG16 backbone, is effective at segmenting lung cancer. With a Dice score of 73%, it clearly performs exceptionally well and could improve segmentation accuracy and lead to new ways for early identification of lung cancer in medical image analysis.

This study examines four distinct fusion strategies—late, early, dense, and hyper-fusion—to enhance lung cancer segmentation. With a Dice score of 0.73, the Hyper Dense VGG-16 fusion model performed better than all of the others, according to the tests and results. This proves that it is capable of effectively managing the complexity of STS segmentation. Because of its deep, layered design and VGG-16 backbone, the model detected small tumor borders that are difficult to distinguish from healthy tissue. The level of detection and preservation of features must be balanced in order to segment heterogeneous and irregular metastatic tumors using CT-PET imaging.

In the studies, we employed three loss functions—Binary Cross-Entropy, Focal Loss, and Dice Loss—to assess their influence on model performance. Focal Loss yielded the most favorable outcomes, particularly in the context of class imbalance and the model's capacity to concentrate on regions that are challenging to classify, which is crucial when managing small or irregularly shaped tumors.

When compared to the other fusion strategies—Early Fusion (Dice = 0.661), Late Fusion (Dice = 0.712), Dense Fusion (Dice = 0.715), and Hyper Fusion (Dice = 0.72)—the Hyper Dense VGG-16 fusion model emerged as the most effective despite its complexity. The Early Fusion model's efficacy was diminished as a result of the increased computational complexity, which was exacerbated by its difficulty with feature alignment. Late Fusion fared well in terms of merging final predictions, but it did not completely benefit from model synergies. Although Dense Fusion outperformed Late Fusion to a small degree, it necessitated more computing power and meticulous regularization to prevent overfitting. However, Hyper Fusion, with sophisticated methods including attention mechanisms, demonstrated excellent results but increased architectural complexity, which could lead to overfitting if not regulated.

Based on these findings, Hyper Dense VGG-16 fusion was selected as the most suitable model due to its superior balance of accuracy, computational efficiency, and model complexity. Critical strategies such as data augmentation, dropout, and regularization mitigated the

overfitting risks associated with more intricate models like Hyper Fusion despite their more intricate architecture. As a result of its performance and generalizability to new data, this method is ideal for STS tumor segmentation in CT-PET scans.

4.7. Summary

This chapter proposes multi-modal fusion approaches based on U-Net architecture (early fusion, late fusion, dense fusion, hyper-dense vgg16 U-net) for lung tumor segmentation. The findings prove that the Dice score of 73% is obtained for the hyper-dense vgg16 U-net, which is superior to the other four proposed models. These results confirm that hyper-dense fusion effectively captures complementary information from both PET and CT modalities, leading to improved tumour boundary delineation.

The proposed segmentation framework provides a crucial step toward automated tumour quantification, supporting radiologists in early detection and treatment planning for lung cancer. However, while segmentation identifies the precise tumour location and shape, it does not by itself provide information about tumour stage or progression, which are critical for clinical decision-making and prognosis. Therefore, the next chapter focuses on TNM classification and overall stage prediction using Vision Transformer (ViT) models, extending the proposed framework from spatial segmentation to disease staging for comprehensive lung cancer assessment.

Chapter 5

5. Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction Using Vision Transformers

5.1. Introduction

This chapter seeks to apply accurate classification of non-small cell lung cancer (NSCLC) stage using deep learning, and in particular, Vision Transformers. Attention is also paid to the historical TNM classification, variables of clinical importance associated with the staging of the disease, and motivation for the use of Transformers for such a task. The aims of this study include a detailed knowledge of lung cancer staging, the development of new approaches, and an emphasis on the benefits of Vision Transformers in this important area of medicine. This chapter describes the main objectives of the study, paying utmost attention to the TNM staging and the prediction of the overall cancer stage. Advanced techniques in deep learning, such as Convolutional Neural Networks (CNNs) and Vision Transformers, have been implemented for precise classification. Particular attention should also be given to addressing class imbalance and improving the performance of the model. The second core objective identifies the importance of the direct pathway in predicting the overall stage of lung cancer patients, incorporating the details of the patients.

5.2. TNM Staging System

The TNM Staging System is used in the field of oncology and refers to the characterization of size related to the tumor, the presence of cancer in the lymph nodes, and whether the disease has metastasized to other organs. The UICC was the first organization to develop it. At the moment it operates through UICC and the American Joint Committee on Cancer [110]. The method of staging for each kind of cancer is universally recognized on a global scale. The foundation of this argument rests upon three fundamental factors

- The size and extent of the primary tumor, as indicated by the T category, refers to the dimensions of the cancer and the degree to which it has infiltrated adjacent tissues.
- The spreading of lymph nodes (N) is a category that details the extent of spread to the lymph nodes in close proximity.
- Metastasis, denoted by the M category, refers to the occurrence of cancer spreading to distant organs or other regions of the body.

The samples of T and N descriptors are given in Figures 5.1. and 5.2. respectively.

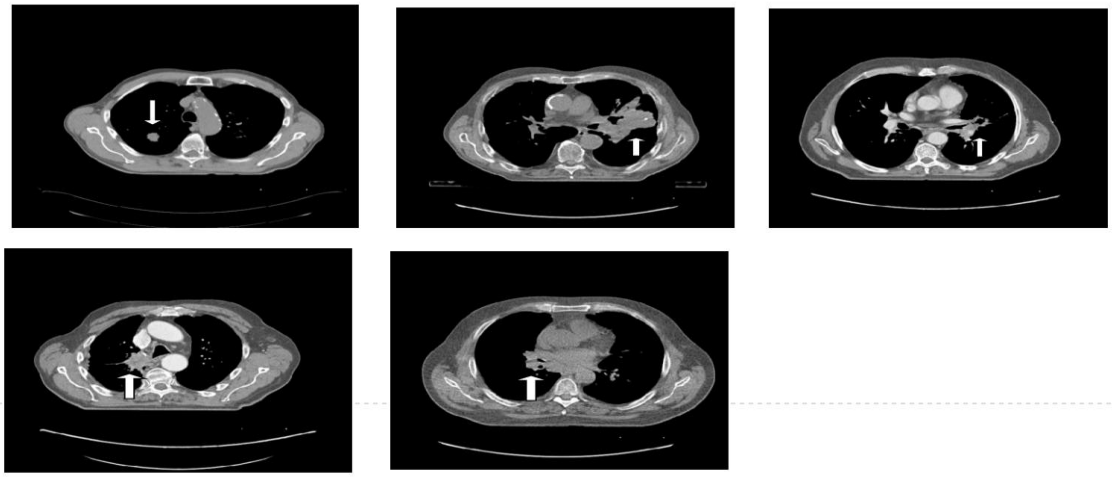

Figure 5.1. T descriptor examples from NSCLC-Radiomics dataset.

Figure 5.2. N descriptor examples from NSCLC-Radiomics dataset.

The TNM classification 8th edition outlines the four stages of lung cancer [111] as given in Table 5.1. Stage 0 represents an additional stage distinguished by the TNM descriptors Tis, N0, and M0. The term "tis" is used to denote a tumor in situ, characterized by its non-malignant nature but with the potential to progress into a malignant state at a later stage. The terms N0 and M0 indicate the absence of metastasis to lymph nodes or distant organs. A diagnostic method frequently utilized for assessing lung cancer stage is CT scanning, sometimes accompanied by a fluorodeoxyglucose (FDG) positron emission tomography (PET) scan.

Table 5.1. Lung cancer staging based on TNM classification 8th edition.

	N0	N1	N2	N3
T1a	IA1	IIB	IIIA	IIIB
T1b	IA2	IIB	IIIA	IIIB
T1c	IA3	IIB	IIIA	IIIB
T2a	IB	IIB	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T3	IIB	IIIA	IIIB	IIIC
T4	IIIA	IIIA	IIIB	IIIC
M1a, M1b	IVA	IVA	IVA	IVA
M1c	IVB	IVB	IVB	IVB

The CT scan images accurately determine the dimensions and spatial coordinates of the tumor within the pulmonary region. Nevertheless, detecting cancer in lymph nodes is a significant challenge for proficient radiologists due to difficulty locating tumors using CT scan pictures. A distinct magnetic resonance imaging (MRI) scan is necessary for brain metastasis, explicitly targeting the brain [48]. The treatment strategy for individuals diagnosed with lung cancer is contingent upon various aspects, including the specific classification of the cancer, namely SCLC or non-small cell lung cancer NSCLC.

5.3. Research Objectives

The primary aim of this study is to delineate the fundamental research objectives, which encompass two pivotal components: TNM stage classification and overall stage prediction. These objectives are driven by the overarching goal of improving the accuracy and effectiveness of lung cancer staging, ultimately contributing to enhanced patient care and treatment outcomes.

Objective 1: TNM Stage Classification

The first core research objective centers on refining and advancing the TNM stage classification process for lung cancer. This objective encompasses several sub-goals:

- Develop and Implement Deep Learning Models: The research aims to develop and implement deep learning models, including Convolutional Neural Networks (CNNs)

[112] and Vision Transformers [113], to accurately classify the TNM stages of lung cancer based on medical imaging data.

- Address Class Imbalance: The methods address class imbalance class imbalance within those in input dataset is reduced to facilitate training so that the models are effective in learning from all TNM stage categories, hence mitigating bias associated to models attaining a potentially 100% accuracy on dominant classes.
- Optimize Model Performance: The goal is to enhance the performance of these models, increasing their accuracy in categorizing lung cancer patients into specific TNM stages. This optimization entails refining model topologies, investigating innovative loss functions, and maximizing classification accuracy.

Objective 2: Direct Overall Stage Prediction

Establishing a straightforward approach for estimating the general stage of lung cancer patients is the second focus of core research. The following main sub-goals comprise this aim:

- Leverage Vision Transformer Architecture: The features of Vision Transformer (ViT) architectures are used to improve the general accuracy of stage prediction even further. ViTs provide a more complete knowledge of the development of the disease since they are quite good in capturing long-range dependencies inside medical images.
- Incorporate Patient-Specific Information: Understanding the importance of patient-specific elements, such age and gender, in lung cancer staging, the aim is to smoothly include this demographic information into the prediction process. This addition guarantees a more exact and customized evaluation of the general performance.

All discussed, the research goals of this work address class imbalance issues, refine TNM stage classification utilizing state-of- the modern deep learning models, and provide a direct overall stage prediction method. By means of these goals, it is aspired to expand the accuracy and clinical utility of lung cancer staging, thereby helping patients by means of therapy recommendations and enhancement of prognostic assessments.

5.4. Research Questions

Keeping with the research goals, this section presents carefully thought-out research questions that will be used as guidelines for the whole study. The goal of these questions is to find important new information and progress in the field of lung cancer staging and aiming to get more critical details.

RQ 1. How can the TNM classification method is applied to accurately predict the overall stages of lung cancer? The question also answers about the most important clinical and pathological factors.

This question investigates the clinical and pathological factors which have significant impact on how well the TNM classification can predict the overall stage.

RQ 2. How can additional imaging and biomarker methods can help the TNM classification system to increase the prediction of the general stages of lung cancer? This questions also identifies the factors which has important role in the staging process.

This study investigates the impact of adding extra imaging and biomarkers methods to the TNM classification system which may improve the general process of predicting the stage of a cancer. The goal is to show the benefits of combining different sources of knowledge.

RQ 3. How does the predictive accuracy of TNM overall stage classification method varies across the different subtypes of lung cancer (e.g., Age/Gender/histology)?

This question focuses on assessing the variability in predictive accuracy when applying the TNM classification system to different subtypes of lung cancer. It aims to uncover how factors such as age, gender, and histological characteristics influence staging outcomes.

RQ4. What is the correlation between lung cancer TNM overall stages and key clinical and demographic factors, including survival rate, patient age, gender, and tumour histology?

This inquiry seeks to establish the correlations between TNM overall stages and critical clinical parameters, including survival rates and subtypes of lung cancer. It aims to provide insights into the prognostic value of TNM staging.

RQ 5. Which specific TNM stage parameters (T, N, and M) have the most impact on overall stage prediction and survival outcomes?

This research question investigates the relative importance of individual TNM stage parameters, namely T (the primary tumor), N (lymph nodes), and M (metastasis), in both overall stage prediction and the prediction of survival outcomes. It aims to identify the most influential factors.

RQ 6. How do different subtypes of lung cancer (e.g., Age/Gender/histology) affect the correlation between overall stages and survival prediction?

Building upon Question 3, this question further explores how different subtypes of lung cancer, categorized by factors like age, gender, and histology, influence the relationship between overall stages and survival prediction. It aims to discern nuanced patterns within specific subpopulations.

RQ 7. What are the potential limitations and challenges associated with predicting overall stages of lung cancer using the TNM classification system, and how can these be addressed?

This final question critically examines the limitations and challenges inherent in predicting lung cancer's overall stages through the TNM classification system. It endeavors to identify potential obstacles and strategies for mitigating them, paving the way for more accurate and reliable staging.

RQ 8. Can the overall stage prediction model effectively address the limitations and challenges associated with the TNM classification system?

Building on the previous question, it is assessed whether the direct overall stage prediction model can effectively circumvent the identified limitations and challenges of the TNM classification system. This indicates that the proposed model has the potential to address these concerns.

RQ 9. What are the possible constraints of direct overall stage prediction models? When examining the direct forecast of the entire stage, this research questions investigates the disadvantages and limitations that may arise in this alternate technique. It is essential to acknowledge these limits in order to conduct a thorough evaluation.

RQ 10. Does a transformer-based design surpass convolutional neural networks in the domain of lung cancer staging?

Transformer-based designs are evaluated against convolutional neural networks (CNNs) in the context of computational techniques for lung cancer staging. This question directs the assessment of the optimal model structure for this crucial medical application.

Ultimately, the research question focuses on exploring the clinical, radiological, and computational components of lung cancer staging, motivated by these specific research themes. The aim is to improve the understanding and precision of lung cancer staging methods by addressing these problems, thereby benefiting both patients and healthcare providers.

5.5. Motivation for Transformers

The application of Transformers, particularly Vision Transformers (ViTs), in the domain of medical image analysis, such as the prediction of various stages of Non-Small Cell Lung Cancer (NSCLC), is driven by their distinctive attributes. This section explores the rationale for choosing Transformers instead of standard Convolutional Neural Networks (CNNs) for this important application.

Key attributes of modeling:

1. Long-Range Dependency: Transformers have an exceptional ability to grasp large interconnections within data. The Multi-Head Self-Attention (MSA) module facilitates the systematic connecting of data patches. This property has similarity to a graph neural network (GNN) [114]enabling Transformers to generate extensive theoretical and efficient receptive fields. In medical imaging, this ability can be very useful because it leads to the comprehension of contextual information and extensive connections that go beyond those exhibited by Convolutional neural networks (CNN).
2. Elaborate Modeling: CNNs generally use pooling and strided convolutions to modify the scales while reducing the feature, whereas Transformer employs MLPs to gradually enhance and adapt embeddings without altering the scale. Because of the well-modeled and learned feature fusion within the Transformer architecture, subtle and semantic details of the images are captured even as deeper levels of the model are accessed. Maintaining intricate information is essential in the field of medical image processing to ensure that correct classifications are made.
3. Inductive Bias: It should be noted that convolutional neural networks (CNN) take into account some strong inductive biases that are closely related to the concept of pixel locality. This means it consistently applies the same set of weights across the entire image. Although this bias can enhance the rate of convergence and the performance attended on small data sets, it also limits adaptability during more challenging scenarios. In contrast, Transformers exhibit a reduced inclination to construct assumptions relying on previous information because of their utilization of global self-attention mechanisms. The primary inductive bias in Vision Transformers (ViTs) is generated from the positional embedding. Transformers experience heightened computational demands and training challenges as a result of their amplified data prerequisites. Nevertheless, they possess the capacity to exhibit more resilience while handling extensive datasets, a crucial aspect to take into account when forecasting the overall stage of NSCLC.

4. Loss Landscape: Transformers generally generate a flatter loss landscape, even when employed with CNN models. This characteristic enhances the efficiency and capacity to apply the model to new data in contrast to Convolutional Neural Networks (CNNs) trained under comparable circumstances. This characteristic can be highly advantageous when working with medical image data that is characterized by noise or variation.
5. Noise Robustness: The application of the Transformer models has exhibited resilience against a more common set of data imperfections and disturbances such as blurring, motion, contrast variations, and noise. Their long-lasting reliability makes them a good fit for medical image processing, a field that often struggles with noisy input data.

Computational factors to consider:

6. Transformers have demonstrated consistent scaling behavior in both Natural Language Processing (NLP) and Computer Vision (CV). Better outcomes are obtained if the scale of processing resources, model size, and dataset volume are raised simultaneously. Because of this scalability, these approaches are well adapted to handling complex problems such as predicting the overall stage of NSCLC, which involves handling large volumes of medical images and meeting considerable accuracy requirements.

In conclusion, the unique architecture and computational aspects of Transformers, particularly the Vision Transformers (ViTs), make these networks suitable for the assessment of medical images. With regard to predicting various stages of NSCLC, describing long-range dependencies, performing holistic modeling, resisting interference, and easy scalability present an opportunity to enhance the performance of lung cancer staging and its application. The properties of these materials are very compatible with the needs of this important medical application and, therefore, very reasonable to include in the study.

5.6. Main Contributions of This Study

This research study makes a contribution to the Non-Small Cell Lung Cancer (NSCLC) staging prediction, including the use of deep learning techniques and novel approaches. The main contributions of the study are as follows:

- Novel Deep Learning Architecture for TNM Stage Classification: This work presents a new deep learning architecture that is constructed with a special focus on increasing the effectiveness and accuracy in the TNM stage classification of non-small cell lung cancer (NSCLC). By utilizing 2D medical images as an input, this architecture shows

tremendous variance in TNM classification with respect to conventional methods. This method offers a more detailed and data-centric approach to determining the stage of a tumor by exploiting the complex patterns and spatial relationships embedded in the images.

- Vision Transformers for TNM Stage Classification: This study employs transformer technology to harness the advantages offered by Vision Transformers (ViTs) for assessing the TNM stage. ViTs, which are effective in obtaining long-range relationships of distant objects and contextual understanding of particulars in images, are applied to improve the accuracy of precise TNM classification. This novel application of ViTs advances the boundaries of deep learning in the area of medical image analysis by providing a more robust and accurate prediction of the TNM stage.
- Direct Model for Overall Stage Prediction with Multi-Input Structure: This study proposes a direct modeling approach to improve the performance of the overall stage classification task. The system leverages a Vision Transformer architecture, which can accommodate different input structures and features additional information such as the patient's age and gender. This model highlights the importance of demographic attributes in the categorization and, at the same time, integrates them into the prediction task without any effort. By pursuing this aim, the current research contributes to the existing body of knowledge regarding the predictions of the overall stage of NSCLC reviewed so far from a broader perspective of disease advancement.
- To address the class imbalance problem of the dataset, efficient augmentation strategies are employed.
- A comparison examination of the suggested methodology utilizing various state-of-the-art classification networks is conducted for overall stage classification.

5.7. Methodology

5.7.1. Data Collection and Preprocessing

5.7.1.1. Dataset

The NSCLC-Radiomics dataset [115], which can be obtained from both the National Biomedical Imaging Archive (NBIA) and the Cancer Imaging Archive (TCIA), is an excellent tool for performing research regarding Non-Small Cell Lung Cancer (NSCLC) of such type. The aim of this dataset is to offer a comprehensive collection of clinical and imaging-related data for researchers and healthcare providers who are interested in NSCLC. The NSCLC-Radiomics data set discerns 422 records, each of which is comprised of 10 attributes. The

dataset includes a brief explanation of every column. Below is an outline of the size and structure of the NSCLC-Radiomics dataset.

PatientID: Each patient in the dataset is given a unique identification.

Age: This attribute indicates the patient's exact age in terms of the calendar year when the data or information was collected. Age is a significant demographic variable that can impact multiple aspects of cancer diagnosis and treatment.

Clinical.T.Stage: This refers to the clinical stage of the tumor in patients. The T stage offers details regarding the dimensions and scope of the primary lung tumor.

Clinical.N.Stage: Indicates the clinical lymph node (N) stage of the patients. This stage reflects the extent of lymph node involvement by the cancer.

Clinical.M.Stage: Represents the clinical metastasis (M) stage. This stage identifies whether the cancer has spread to distant sites in the body.

Overall.Stage: Reflects the overall cancer stage, which is often determined by combining information from the T, N, and M stages. It provides a comprehensive assessment of the disease's severity.

Histology: Specifies the histological type of the lung cancer. Lung cancers can have different histological subtypes, each with distinct characteristics.

gender: Indicates the gender of the patients, typically categorized as male or female. Gender is another demographic factor that may have relevance in cancer research.

Survival.time: Represents the time (in some specified units, e.g., months) from the initial diagnosis or treatment to a specific event, such as death or the end of the study period. This column is essential for survival analysis.

deadstatus.event: A binary column indicating whether a patient has experienced the event of interest (e.g., death) during the study period. It is commonly used in survival analysis as an outcome variable.

Figure 5.3. shows the clinical data sample. Out of 422 patients, 302 were selected for training, 50 for validation, and 70 for testing purposes.

	PatientID	age	clinical.T.Stage	Clinical.N.Stage	Clinical.M.Stage	Overall.Stage	Histology	gender	Survival.time	deadstatus.event
0	LUNG1-001	78.7515	2.0	3	0	IIIb	large cell	male	2165	1
1	LUNG1-002	83.8001	2.0	0	0	I	squamous cell carcinoma	male	155	1
2	LUNG1-003	68.1807	2.0	3	0	IIIb	large cell	male	256	1
3	LUNG1-004	70.8802	2.0	1	0	II	squamous cell carcinoma	male	141	1
4	LUNG1-005	80.4819	4.0	2	0	IIIb	squamous cell carcinoma	male	353	1
...	...	...	...	...	...	...	...	...	...	...
417	LUNG1-418	53.6712	2.0	0	0	I	adenocarcinoma	male	346	1
418	LUNG1-419	66.5096	4.0	1	0	IIIb	squamous cell carcinoma	male	2772	0
419	LUNG1-420	73.3808	2.0	1	0	II	squamous cell carcinoma	male	2429	1
420	LUNG1-421	61.7041	2.0	2	0	IIIa	squamous cell carcinoma	female	369	1
421	LUNG1-422	68.1260	2.0	0	0	I	NaN	female	1590	1

422 rows x 10 columns

Figure 5.3. Sample of a clinical data CSV file for the NSCLC-Radiomics dataset

5.7.1.2. Pre-processing

The NSCLC-Radiomics dataset used in this study consists of CT and PET-CT scans with sizes 512 × 512. To better computation performance and develop a memory-efficient approach, the data is resized to size 224 × 224. Furthermore, the images are rescaled with the help of the normalization technique. Normalization is required to maintain the general distribution in the dataset and make the convergence of gradient descent faster and smoother.

The data shows a class imbalance problem because the number of patients in each class differs slightly (Figure 5.4.). Data augmentation techniques can be used to tackle this issue. Data augmentation is defined as a technique that is used to create more data samples from existing data. As discussed, many augmentation techniques are available. For the data, scaling and flipping are used to balance the data belonging to each class. In the flipping technique, horizontal flipping and up-scaling are used, and downscaling operations are used.

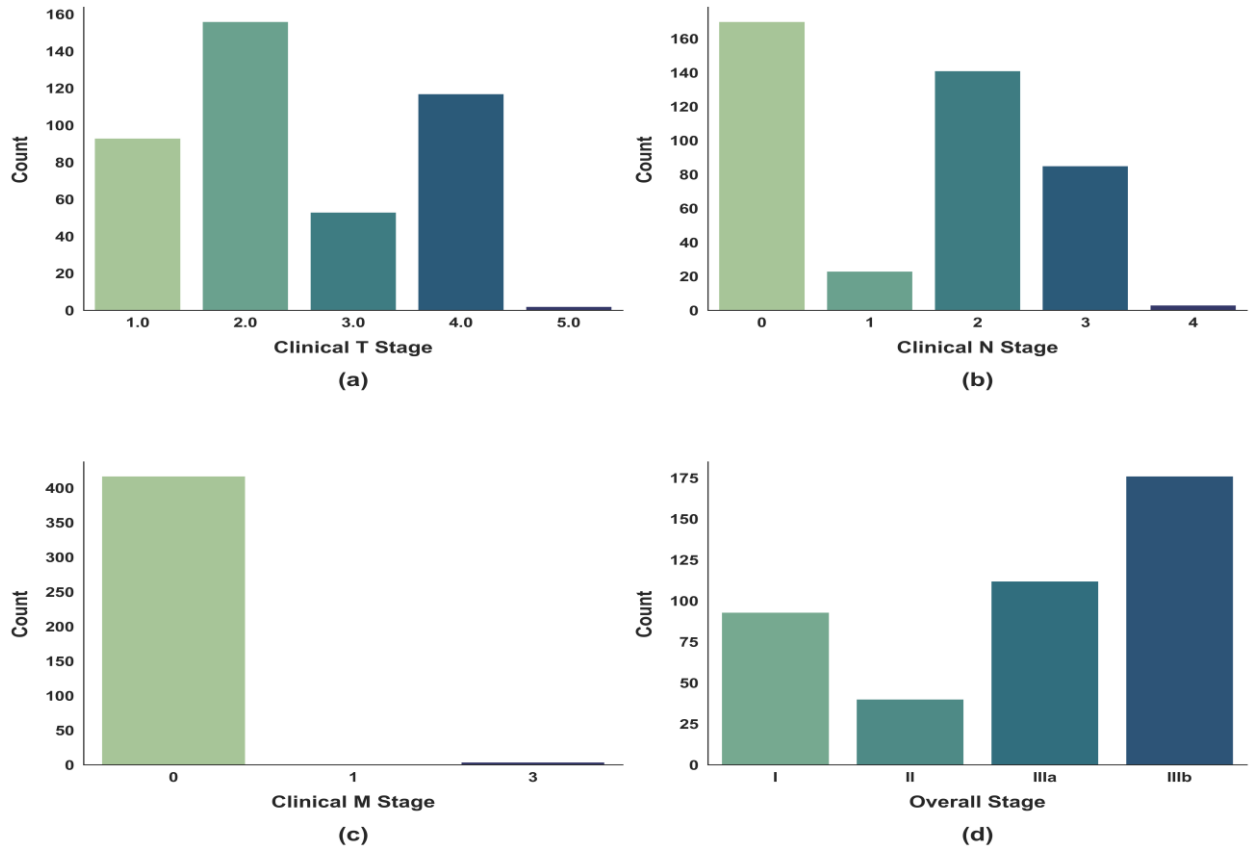

Figure 5.4. Data distribution among different classes.

5.7.1.3. Data analysis

An extensive analysis of the NSCLC-Radiomics dataset is conducted from the Cancer Imaging Archive (TCIA) to gain insights into overall stage prediction for Non-Small Cell Lung Cancer (NSCLC) patients.

In this analysis, key predictors of mortality rates are identified. Age has emerged as a significant risk factor, highlighting its critical role in prognosis. Additionally, it was found that incorporating demographic factors, such as age and gender, along with imaging data has improved the accuracy of the overall stage prediction models. This also stresses the need for incorporating demographic information alongside clinical information to improve the accuracy of the models.

In addition, the study found an interesting pattern in subgroup analysis, as can be seen in Figures 5.5. and 5.6. More specifically, higher model accuracy is noted for individuals aged 65 or older as compared to others. This infers that age is related to predicting the preponderance of the disease in the context of older patients and hence suggests the requirement for improving intervention strategies for this population. The gender analysis indicates that the model was

slightly more accurate in the male group than in the female group. Despite the fact that there have been no apparent gender-related variations in prognosis for NSCLC, as depicted in Figure 5.7., it is, however, essential to delve deeper into understanding the causes that lead to these differences, with the aim of achieving better treatment outcomes.

Additionally, the analysis refers to the effect the clinical N stage has on predicting the overall stage. With this characteristic, the clinical N stage proved to be the most significant factor impacting the predictions. This underlines the importance of clinical N-stage information in understanding the progression, stage, treatment, and outcome of the disease.

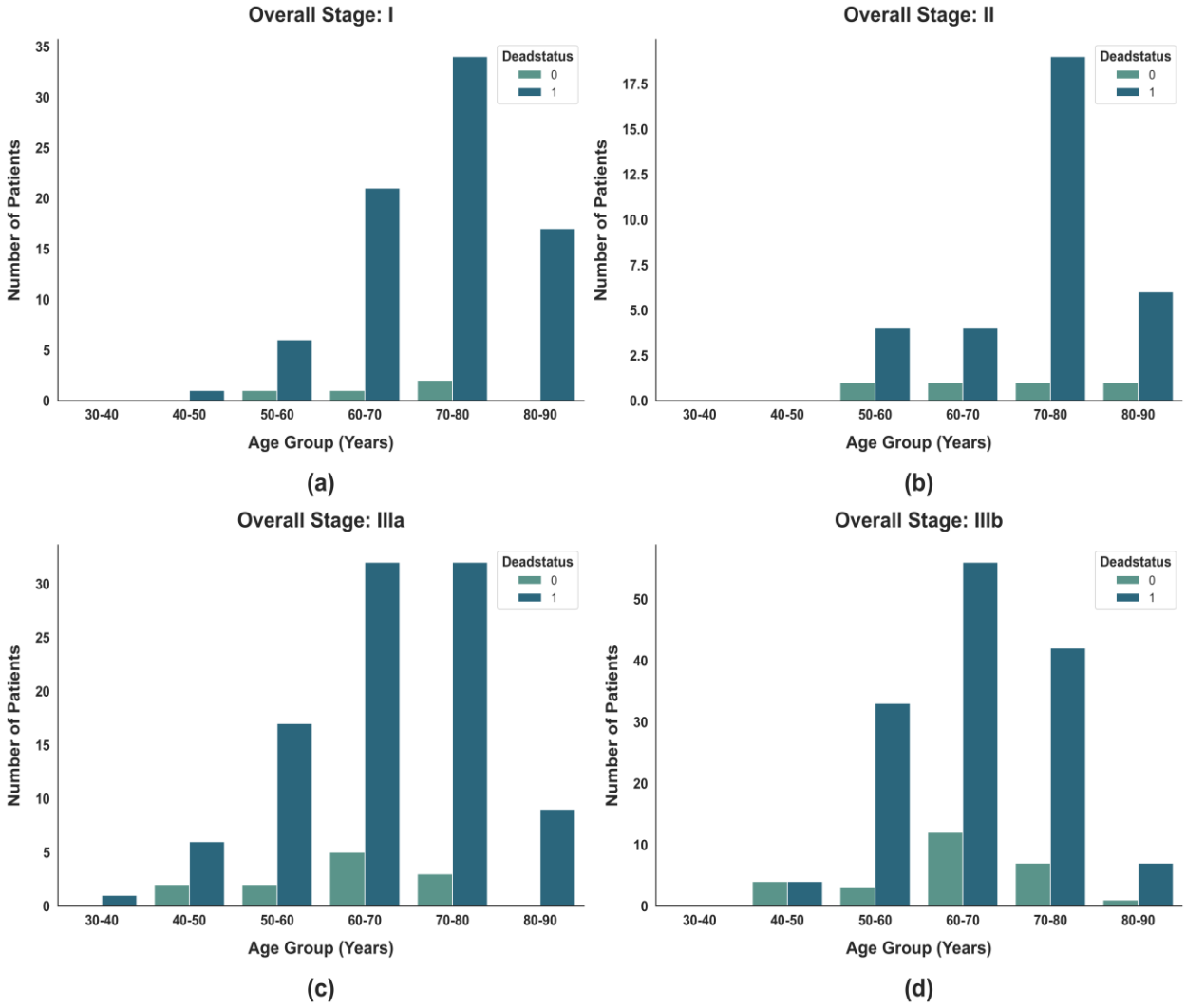

Figure 5.5. Mortality rate distribution within different age groups and overall stages. *Deadstatus* = (1) denotes deceased patients, while *Deadstatus* = (0) represents patients who remained alive.

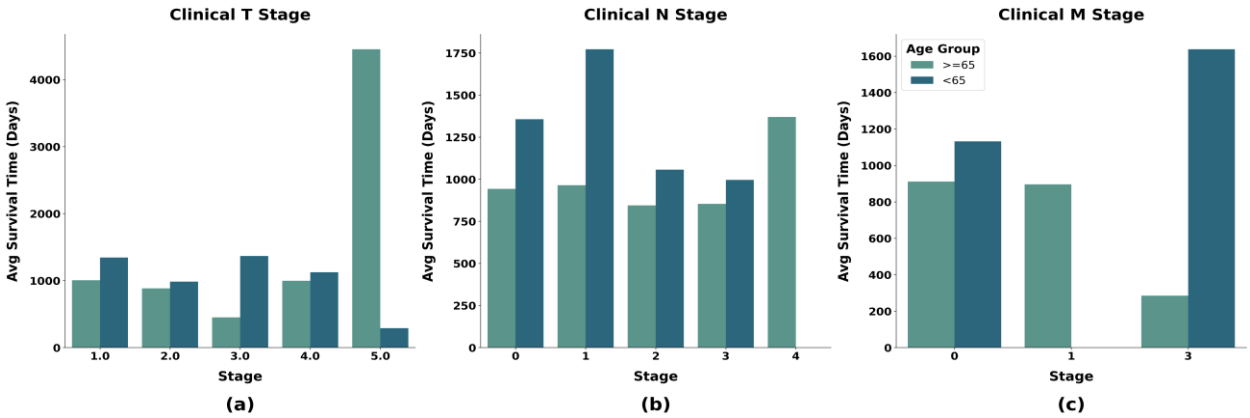

Figure 5.6. Average survival time within different age groups.

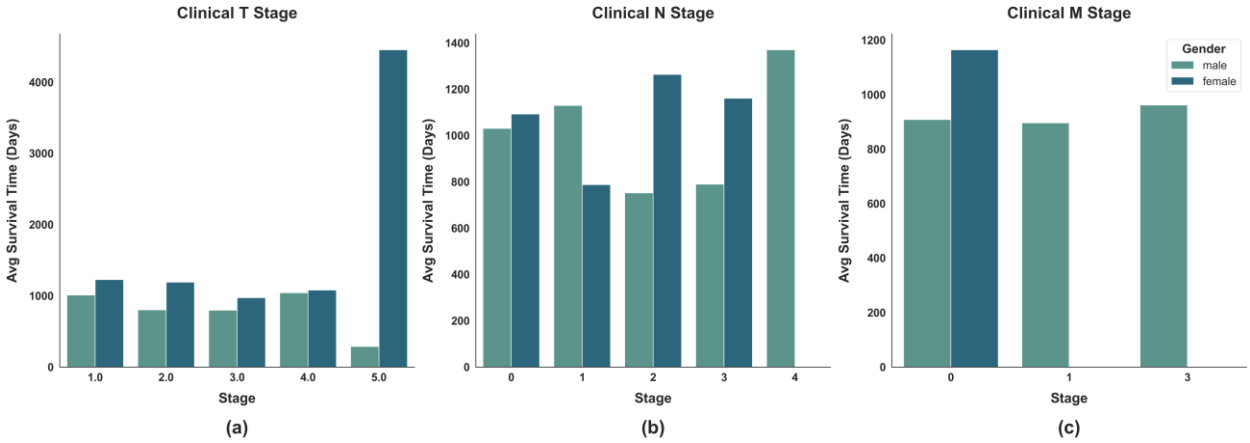

Figure 5.7. Average survival time within male and female groups, including TNM stages.

In summary, age, gender, and histology are essential factors contributing to the overall prediction of the stage and survival of patients with NSCLC. These characteristics are amenable to improvement when machine learning models are applied so that more targeted patient management is achieved. However, these factors, along with clinical, demographic, and other variables, must be taken into account to attain robust predictive models. Furthermore, continuous study is necessary to enhance the comprehension of these connections and enhance the quality of patient care.

5.7.2. TNM Stage Classification

5.7.2.1. Deep Learning Architecture for TNM Stage Classification

The proposed neural network design for TNM (Tumor, Node, Metastasis) stage classification incorporates a novel model that is based on the principles of dense connection [107].The input layer is initialized with a shape of size (224, 224), which corresponds to the typical dimensions of medical photographs. The proposed architecture has a convolutional layer with 164 filters, each possessing a kernel size of 5x5. The Rectified Linear Unit (ReLU) activation function is utilized to introduce non-linearity. Batch normalization is added after each convolutional layer for faster convergence.

Max-pooling with a pool size of 2x2 is used to reduce the feature maps following the initial convolutional layers. To minimize overfitting, dropout with a rate of 25% is implemented. The following convolutional blocks follow this pattern, progressively augmenting the number of filters while preserving the dense connectivity between layers, as depicted in Figure 5.8. and Figure 5.9.

Following the initial convolutional layers, max-pooling with a pool size of 2x2 is applied to downsample the spatial dimensions, and dropout with a rate of 25% is introduced to prevent overfitting. The subsequent convolutional blocks continue this pattern, gradually increasing the number of filters while maintaining the dense connectivity between layers as shown in Figure 5.8 and Figure 5.9.

The architecture incorporates three branches, each dedicated to predicting T, N, and M stages. Each branch follows a similar convolutional block structure but operates independently, allowing the model to capture stage-specific features. The convolutional blocks are interspersed with max-pooling and dropout layers to enhance the network's ability to discern hierarchical features at different scales.

Upon the convolutional blocks, a flattening layer is introduced to transform the multidimensional tensor into a flat feature vector. This vector is then passed through fully connected layers, incorporating ReLU activation, batch normalization, and dropout, fostering non-linearity, stability, and regularization, respectively.

The final layer of each branch employs a SoftMax activation function to generate the probability distribution over the respective TNM classes (Eq. 5.1). The probability that a sample $\boldsymbol{i}$ belongs to class $\boldsymbol{k}$ is computed as:

$$P(i,k) = \frac{e^{z_{i,k}}}{\sum_{j=1}^{K} e^{z_{i,j}}} \tag{5.1}$$

where $\boldsymbol{z_{i,k}}$ represents the model's logit output for class k, and $\boldsymbol{K}$ is the total number of classes. These softmax probabilities are then used to calculate the Categorical Cross-Entropy Loss (Eq. 5.2):

$$L_{CCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log\left(p_{i,k}\right) \tag{5.2}$$

where $\boldsymbol{p_{i,k}} = \boldsymbol{P(i,k)}$, $\boldsymbol{N}$ is the number of training samples, and $\boldsymbol{y_{i,k}}$ denotes the true one-hot encoded class label. This loss function penalises incorrect predictions proportionally to their confidence and serves as the optimization objective for training the TNM classification model.

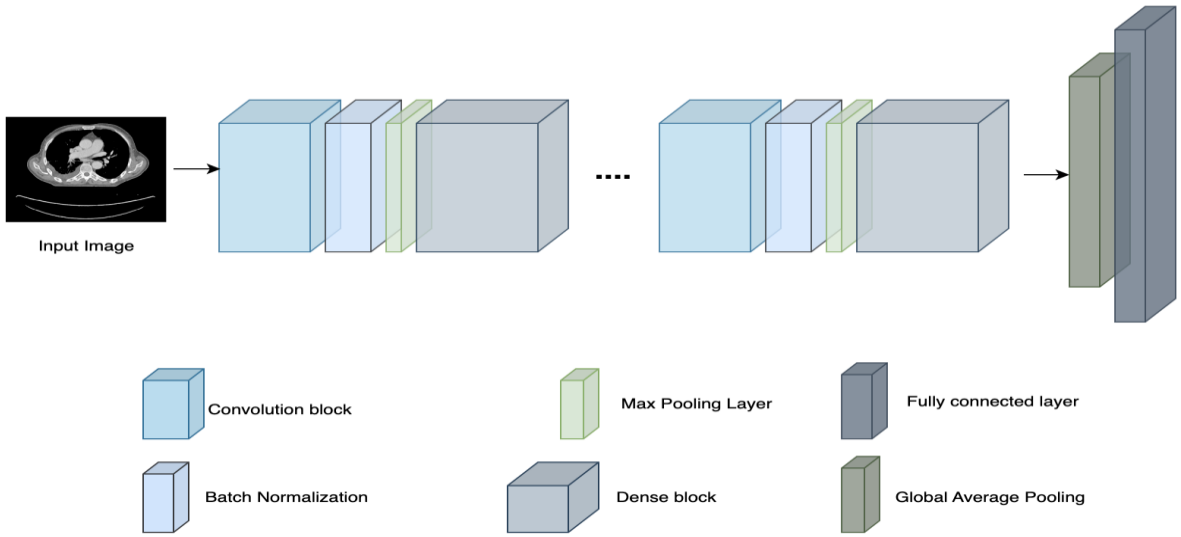

Figure 5.8. Architecture of proposed model for T, N, M stage classification.

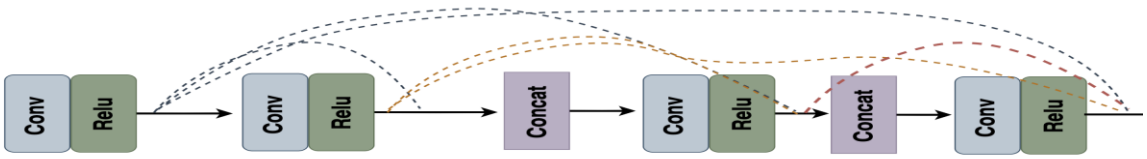

Figure 5.9. Dense Block

Importantly, the architecture adopts a unified decision-making approach, where the predictions from the T, N, and M stages are collectively fed into a decision tree algorithm. This algorithm synthesizes the individual predictions to yield the overall stage classification. The decision tree enhances interpretability and provides a comprehensive strategy for aggregating stage-specific information. The model is trained end-to-end using the Adam optimizer with a learning rate of 0.0001.

5.7.2.2. Enhancements and Extensions of the TNM Stage Classification Architecture:

In the pursuit of refining and extending the TNM stage classification architecture, two key enhancements are introduced: a multi-image approach and the inclusion of demographic features, specifically age and gender. These adaptations are geared towards fortifying the model's robustness, leveraging additional information to improve accuracy and generalize across diverse patient populations.

1. Multi-input Architecture:

The architecture is expanded to accept several image modalities, notably axial and coronal views, to acknowledge the multi-input nature of medical imaging data. This expansion gathers additional information from several imaging planes, which will enhance our understanding of the tumor's geographic distribution and features.

For this augmentation, distinct branches are combined for each image modality. The convolutional blocks within each branch independently process the separate image inputs, enabling the model to distinguish stage-specific features that are present in axial and coronal views. These branches possess the identical convolutional block structure as the original design, but they function on their individual picture inputs. The ensemble approach integrates the final predictions from these branches, leveraging the strengths of both axial and coronal viewpoints to achieve a more comprehensive TNM stage categorization.

2. Inclusion of Demographic Features:

The inclusion of a layer to the design that incorporates gender and age data highlights the importance of demographic characteristics in cancer prognosis (Figure 5.10.). The fully connected layers receive the flattened output from the convolutional blocks, and these demographic features are then added to it. With this update, we hope to give the model a better chance of picking up on gender and age-related subtleties in TNM stage prediction.

Because of its well-documented importance in cancer prediction, including age is very relevant. The ability to learn age-related patterns gives the model a leg up when it comes to identifying how various age groups show unique traits in imaging data.

Gender is ignored in traditional medical imaging models, which introduces a new dimension for analysis. This new dimension allows for the identification of possible gender-based variations in the TNM stage.

Such demographic features are also treated as additional input channels during the model's training, enabling the model to learn from additional demographic data and images. The implemented approach not only improves the model's interpretability but also helps make better TNM stage predictions.

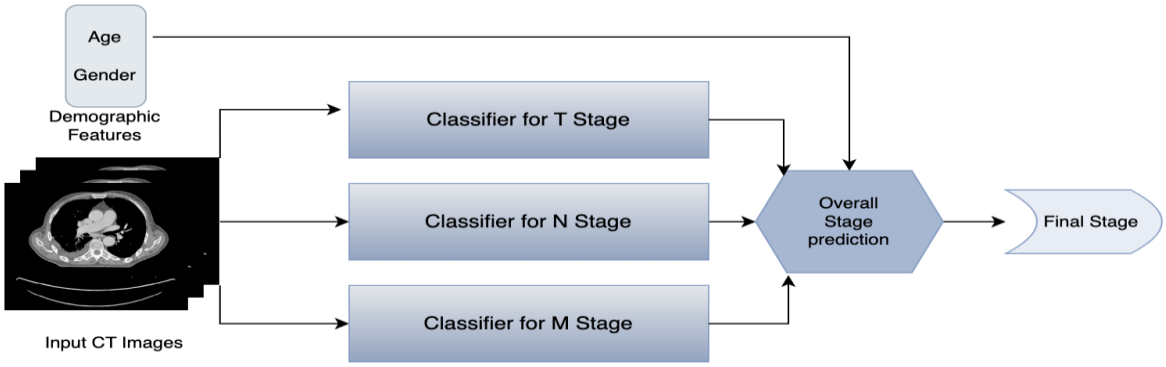

Figure 5.10. Multi-input architecture.

Training and Evaluation:

The extended architectures that incorporate multi-image modalities and demographic features are trained end-to-end using the same categorical cross-entropy, which maintains uniformity in the learning goals. The models are evaluated on comprehensive datasets, assessing their performance across diverse patient cohorts.

These innovations represent progress in the effort towards more precise and refined TNM stage classification. The multi-image approach has the unique structure of using different imaging views at one time, and adding demographic features is a step towards personalizing cancer treatment. These improvements enhance the model's prediction ability and facilitate a comprehensive knowledge of the complex aspects affecting TNM staging in Non-Small Cell Lung Cancer.

1. Advantages of Vision Transformer Integration:

Applying ViT architecture is advantageous in many ways. In the realm of images, ViT excels at identifying intricate structures and contextual cues, hence generating a more comprehensive feature set for TNM stage classification. Large-fine datasets and high-resolution images can be handled with consummate ease. Therefore, it is appropriate for detailed and extensive medical imaging research. Its versatility also allows integration with almost all types of modality and data.

Integrating the Vision Transformer Architecture to TNM stage classification is excellent progress. It provides the model with the expected accuracy and robustness as it combines demographic data with the powerful extraction features of ViT. By improving TNM stage classification, this development visualizes a more focused cancer prognosis, flagging the potential of transformer systems in medical imaging and cancer treatment.

5.7.3. Overall Stage Prediction

5.7.3.1. Deep learning architecture for overall stage prediction.

An effective predictive model is important for predicting the overall stage in the field of Non-Small Cell Lung Cancer prognosis. This section addresses the deep learning architecture that has been custom-developed for this purpose.

Overview of Architecture:

In this stage of architectural development, it is stated that the first step is an input layer. An input layer is one that is intended to receive three-dimensional data, specifically axial and coronal views of medical images. The dimension of the input shape is (224,224). The primary convolutional layer with 64 filters having a kernel of 7x7 serves well as a feature extraction unit with great efficiency in obtaining complex structures from the input data. To facilitate the stability of the model and place the model in a non-linear regime, batch normalization is performed, followed by a rectified linear unit (ReLU) activation. The purpose of the pooling techniques is to decrease the size of spatial dimensions, which enhances the efficiency of the computations and makes it possible to retrieve critical features. The particular feature of the architecture is composed of compact blocks, each made up of several convolution layers with a known expansion rate. Such blocks provide significant opportunities for capturing hierarchical elements that are essential for the recognition of complex patterns from medical images.

Transition blocks are implemented rather seamlessly to attain a desirable balance between the model's accuracy and overfitting to many features. These blocks combine batch normalization, ReLU activation, and convolutional layers to reduce the number of filters effectively. In the transition blocks, the reduction parameter is critical in controlling the flow of information between the layers, thus determining the effectiveness of the model.

As the network expands in size, the global average pool layer aids in dimension reduction and high-level feature extraction. The fully connected layers, augmented by ReLU activation, serve as effective classifiers. The second-to-last dense layer with 1000 units functions as a feature extractor and captures sophisticated features. The softmax-activated last dense layer classifies the data into several discrete classes and thoroughly predicts the general stage of NSCLC.

Exploration of Multiple Inputs:

The research centered on the architecture's capacity to accommodate diverse data sources to enhance predictive performance. The architecture in the model was altered in a manner that allowed extensive use of axial and coronal views of images. This adjustment was due to the fact of the introduction of these various sources of information as illustrated in figure 5.11. The aim of this research is to investigate if the accuracy of prediction could be improved by the use of several images with different views.

Integration of Demographic Features:

Apart from image data, age, and gender description were given as additional input parameters. In their case, the same design principles as the architecture were adopted, and these features were tested to find out whether such information may significantly enhance the model's prediction accuracy. This investigation was premised on the fact that knowledge of the specifics of individual patients can be very important to provide a precise forecast.

In conclusion, the proposed architecture is feasible for the deep learning model that aims to predict the general stage of non-small cell lung cancer (NCLC). It is based on the DenseNet principles and can alter demographic factors to other inputs, indicating that it offers accurate and precise estimations. This comprehensive analysis provides a framework for comprehending the complexities of the architecture, facilitating future enhancement and optimization in the pursuit of a better NSCLC prognosis.

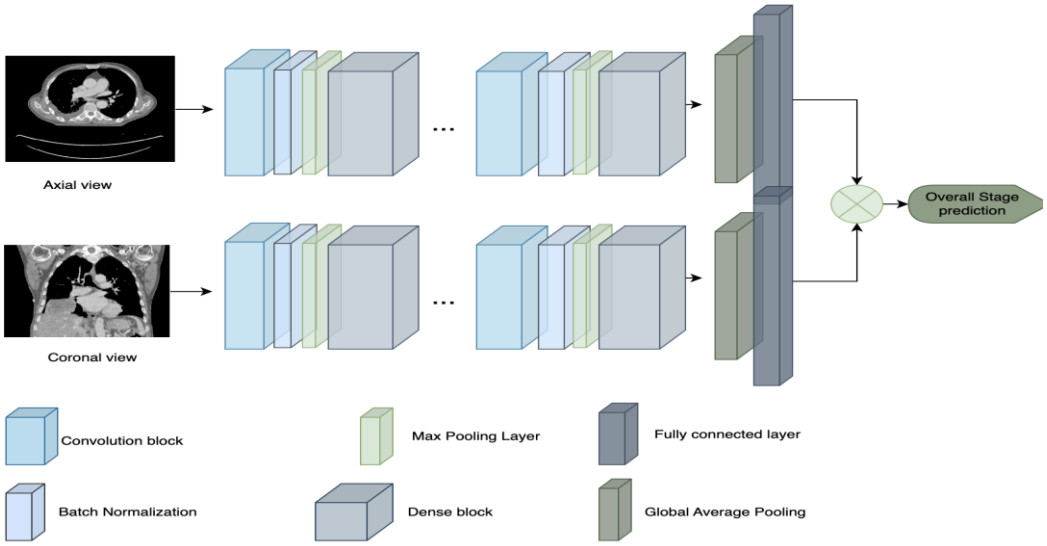

Figure 5.11. Multi-view architecture for overall stage prediction

5.7.3.2. The Vision Transformer-based architecture.

Integration of advanced deep learning architectures becomes essential in the persistent search for better prognostic models for non-small cell lung cancer (NSCLC). This part explores the subtleties of a new method: the multi-input structure based on Vision Transformer (ViT). This innovative design presents a comprehensive framework for general stage prediction, so transforming the area.

Intended initially for general-purpose picture classification, the Vision Transformer [113], has shown amazing adaptability among several computer vision applications. ViT depends on self-attention mechanisms [116] obtained by the Transformer architecture, unlike traditional Convolutional Neural Networks (CNNs). The basic concept considers the input image as a sequence of linearly embedded, fixed-size patches, which are transformed into vectors. Transformer blocks process the input sequence formed by these vectors as well as spatial embeddings.

Self-attention layers enable each Transformer block to capture long-range dependencies inside the sequence. Discerning complex patterns in medical images depends on the model's capacity to simultaneously pay to several areas of the input sequence, hence improving its awareness of spatial linkages. This self-attention mechanism greatly helps the ViT be efficient in feature extraction and representation learning.

ViT Pipeline Overview

The architecture of a Vision Transformer (ViT) typically comprises of a Transformer encoder, and task-specific decoder as shown in Figure 5.12. Taking image processing as an example, the initial step involves dividing the image $X \in \mathrm{R}C\times H\times W$ into a sequence of non-overlapping patches $\{X1, X2, ..., XN\}$, where $Xi \in \mathrm{R}C\times P\times P$, with C denoting the number of channels, $[H, W]$ representing the image size, and $[P, P]$ indicating the resolution of a patch. Subsequently, each patch undergoes vectorization and linear projection into tokens:

$$\mathbf{x}\hat{} = \{X1\mathbf{E}, X2\mathbf{E}, ..., XN\mathbf{E}\},\ \mathbf{E} \in \mathrm{R}CP2\times D \tag{5.3}$$

where ***D*** denotes the embedding dimension, which was set to **768** in this study following the standard ViT-Base configuration. This dimension determines the size of the feature vector representing each patch after linear projection, providing a balanced trade-off between representational richness and computational efficiency.

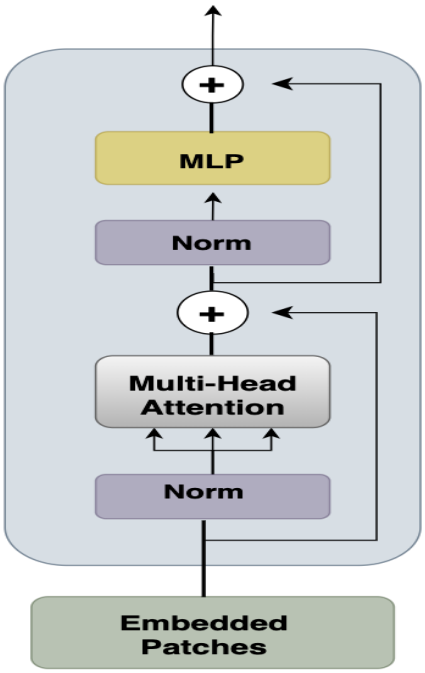

Figure 5.12. Transformer Encoder

Following this, a positional embedding, $\mathbf{E}pos$, is added to preserve the positional information pos of the patches:

$$\mathbf{x}=\mathbf{x}\hat{}+\mathbf{E}pos,\ \mathbf{E}pos \in \mathrm{R}N\times D \tag{5. 4}$$

The resultant tokens are then input into a Transformer encoder, comprising L stacked base blocks. Each base block consists of multi-head self-attention and a multi-layer perceptron (MLP), incorporating Layer-Norm (LN). The feature is expressed as follows:

$$Zl' = \mathrm{MSA}(\mathrm{LN}(Zl{-}1)) + Zl{-}1,\ l \in [1, \ldots, L] \tag{5.5}$$

$$Zl = \mathrm{MLP}(\mathrm{LN}(Zl')) + Zl',\ l \in [1, \ldots, L] \tag{5.6}$$

Generation of Non-Overlapping Patches

In the context of Vision Transformer (ViT) implementation in visual tasks, the generation of patches $\{X1\ , \ldots\ , Xn\ \}$ follows a non-overlapping approach as shown in Figure 5.13. The adoption of this non-overlapping style is aimed at minimizing modifications to the standard Transformer architecture. This choice, however, introduces a partial disruption of the internal structure of an image, as noted by Han et al. (2021a) [117]. To address this challenge, Multi-Head Self-Attention (MSA) blocks are employed to consolidate information from diverse patches, mitigating the impact of the disruption. Simultaneously, the use of non-overlapping patches ensures the absence of computational redundancy when inputting data into the Transformer model.

Positional Embedding Explanation

In the case of Transformers, the processing involves tokenizing and analyzing each patch independently, leading to the unintended consequence of losing positional information concerning the overall image. This is undesirable because understanding the context in the image requires knowledge of the position of each patch. Positional embeddings are proposed to encode such information into every patch so that the positional context is preserved all along the network and helps to solve this problem. Additionally, positional embeddings serve as a manually introduced inductive bias in Transformers. Generally, there are three types of positional embeddings: sinusoidal, learnable, and relative. The first two encode absolute positions ranging from 1 to the number of patches, while the last type encodes relative positions or distances between patches. The subsequent subsections provide a brief overview of each type of positional embedding.

Sinusoidal Positional Embedding

In the context of encoding the position of each patch, a straightforward approach might involve assigning an index value between 1 and the total number of patches to each patch. However, a notable challenge emerges when dealing with a large number of patches, as this may lead to a substantial disparity in index values, adversely affecting network training. The pivotal concept here is to represent distinct positions using sinusoids with varying wavelengths. For a given patch position n, the sinusoidal positional embedding is defined as per the formulation introduced by Vaswani et al. (2017) [116]:

$$E_{sin}(n, 2d) = sin(\frac{n}{10000^{2d/D}}) \tag{5.7}$$

$$E_{sin}(n, 2d+1) = cos(\frac{n}{10000^{2d/D}})$$

where $\boldsymbol{d = 1, \dots, D_2}$, and $\boldsymbol{D_2 = D/2 = 384}$ in this study. The constant 10,000 acts as a wavelength scaling factor to ensure smooth variation of sinusoidal functions across embedding dimensions. Although this constant was initially introduced for sequence modelling in natural language processing (Vaswani et al., 2017) [116], it has been widely adopted in Vision Transformer (ViT) architectures for image-based tasks (Dosovitskiy et al., 2020) [113]. The formulation remains effective for encoding spatial positions of image patches, as it provides a stable numerical range and preserves relative positional relationships across tokens in the image sequence.

Learnable Positional Embedding

Rather than encoding precise positional information directly onto the patches, a more direct approach involves the use of a learnable matrix denoted as $\mathbf{E}lrn$. In this method, the network is tasked with learning the positional information autonomously. This is commonly referred to as learnable positional embedding.

Relative Positional Embedding

In contrast to utilizing a fixed embedding for each location, as seen in sinusoidal and learnable positional embeddings, relative positional embedding captures the relative information based on the offset between elements in Q and K being compared within the self-

attention mechanism [118]. Various approaches to relative positional embedding have been developed, and it remains an active area of research. Nonetheless, the fundamental principle remains consistent, wherein they encode information about the relative position of Q, K, and V through a learnable or hard-coded additive bias during the self-attention computation.

Multi-Layer Perceptrons

In the conventional Transformer architecture, such as in the original Vision Transformer (ViT) by Dosovitskiy et al. (2020) [113] and the Transformer model proposed by Vaswani et al. (2017) [116], the Multi-Layer Perceptron (MLP) follows each self-attention module.

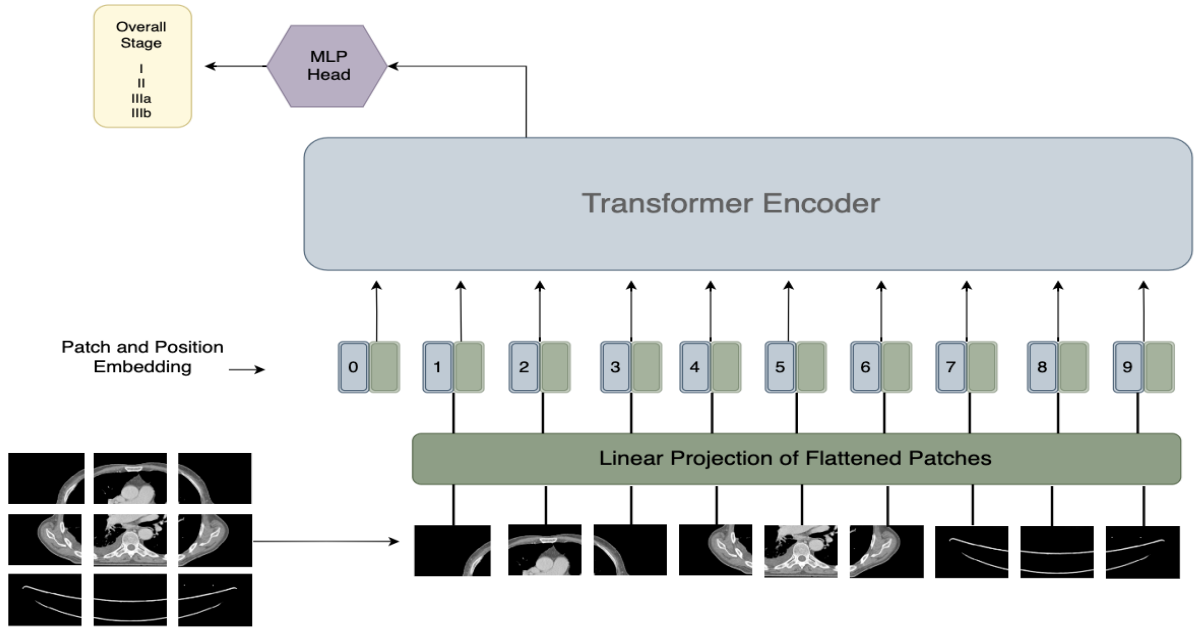

Figure 5.13. ViT-based Architecture for Overall Stage Prediction

The MLP plays a crucial role by introducing inductive bias into the Transformer, addressing the absence of inductive bias in the self-attention operation. This distinction arises from the fact that the MLP is both local and translation-equivariant, while self-attention computation is a global operation. The structure of the MLP consists of two feed-forward networks with an activation function (typically a Gaussian Error Linear Unit, GeLU) in between:

$$MLP(x) = \phi(xW_1 + b_1)W_2 + b_2 \tag{5.8}$$

Here, x represents the input, and W and b denote the weight matrix and bias of the corresponding linear layer, respectively. The dimensions of the weight matrices, W1 and W2, are typically set as D x 4D and 4D x D. As the input is a matrix of flattened and tokenized

patches, applying W to x is akin to employing a convolutional layer with a kernel size of 1x1. Consequently, the MLPs in the Transformer exhibit high localization and equivariance to translation.

Integration of Multi-Input Structure:

The multimodal architecture represents a dynamic synergy between ViTs and textual data, blending medical imaging and patient-specific information as shown in Figure 5.14. This novel method takes advantage of ViTs' ability to record complex visual patterns and long-range dependencies inside images.

The architecture of the proposed model consists of below elements:

1. Axial and Coronal View Imaging Data: Two-dimensional axial and coronal view medical imaging data is accepted as input by the model. These pictures provide a complete picture of the internal components of the lung, which helps the ViT to identify minute visual signals related with cancer development.
2. Backbone Vision Transformer (ViT): The ViT model is the foundation of the design; it has shown remarkable ability in managing medical imaging duties. ViTs use multi-head self-attention systems to help to represent complex interactions among picture patches. These ViT models are made to recognise and encode the spatial aspects of the axial and coronal view images.
3. Gated Fusion: A gated fusion system is included to harmonise the insights obtained from axial and coronal views. By combining the information from the two views, this fusion approach improves the general interpretative power of the model. Gated fusion guarantees appropriate integration of the subtleties from every view, therefore producing a more accurate prediction.
4. Textual Data Inclusion: Acknowledging the value of patient-specific data, age, and gender, we now present textual elements. These features are included into the model to improve personalisation and identify differences in lung cancer development depending on demographic elements. In the framework of lung cancer, age and gender are crucial factors; their inclusion enhances the prediction ability of the model.
5. Concatenation of Visual and Textual Data: The gated fusion output—which shows a harmonic blending of axial and coronal view data—is concatenated with the textual data (age and gender). Combining visual and demographic qualities, this composite feature vector captures the whole patient's condition.

The multimodal architecture enables highly informed forecasts regarding the general stage of lung cancer by means of the complicated interaction between the Vision Transformer's comprehension of complex spatial patterns in medical images and the incorporation of patient-specific knowledge. Combining various modalities gives the model a sophisticated knowledge of the illness and its development, which eventually helps to produce more exact and individualised stage prediction.

This adaptive technique is expected to advance the field of lung cancer overall stage prediction and usher in a new era of precision and personalising in the field of oncology. The multimodal architecture is a significant advancement toward improving patient treatment planning and care, influenced by the Vision Transformers.

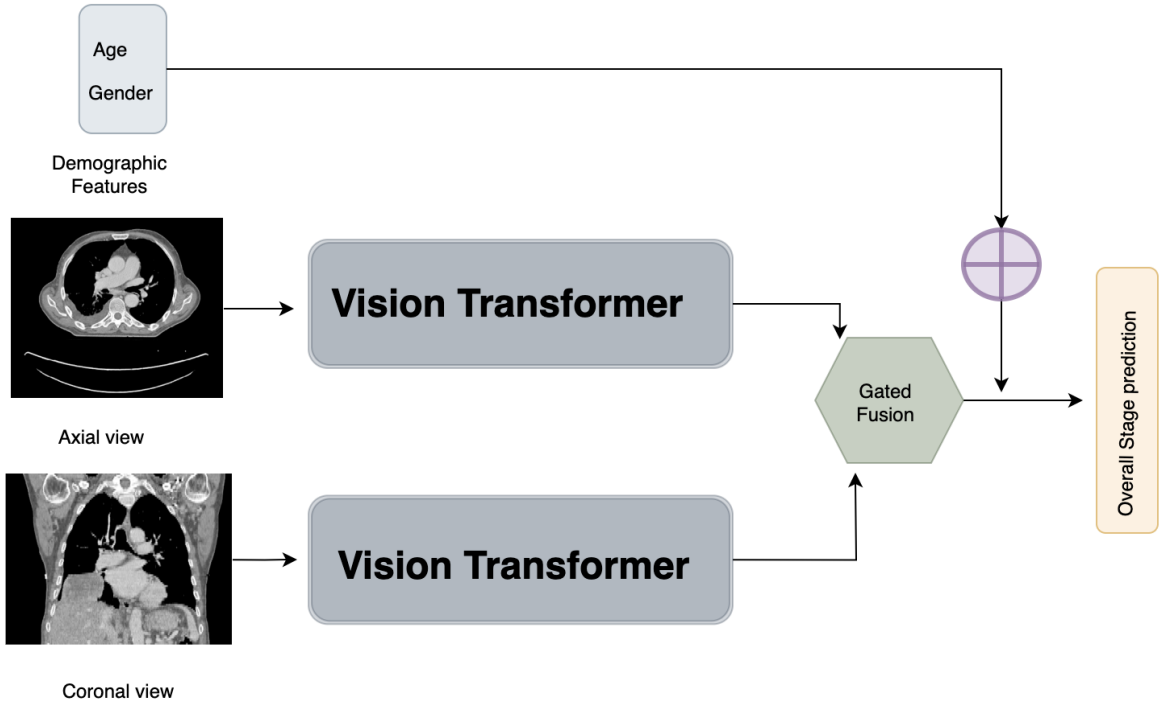

Figure 5.14. Multi-input ViT architecture for overall stage prediction

5.8. Experimental Results

5.8.1. Experimental Setup

This section offers a full description of the experimental setup including information regarding the data splitting procedure, evaluation measures and the training process. During the experiments, Google Colab was utilized for the purpose of accelerating the training times, particularly through the use of its strong T4 GPU.

5.8.1.1. Model Training

The training processes for the CNN and ViT models demonstrate notable differences in methodology and performance, particularly in the context of both TNM stage and overall stage classification models.

In the CNN-based architecture, independent models are trained for the TNM classification and the overall stage classification. Each model is optimized using the Adam optimizer [97] with a learning rate of 0.0001, chosen for its ability to balance effective optimization and convergence stability. The categorical cross-entropy loss function is used in all models during the training and provides uniformity in the optimization for classification tasks. While CNNs can be trained relatively easily, their architecture does not reach high accuracies when trained on complex multi-modal datasets. This shortcoming is observed acutely wherein the trained architectures fail to outperform their ViT counterparts even when the training process is stable.

The ViT models go a step further and utilize an advanced training strategy due to their transformer-based architecture, which helps them in working with multi-modality data. They apply a learning rate of 2e-5 to fine-tune the pre-trained weights in order to enjoy stable convergence during training. In order to accelerate the training on a T4 GPU, mixed-precision training with fp16 is applied while both high accuracy and low computational overhead are hoped to be achieved. Given its design, it is clear that the ViT model inherently is better in regard with learning complex dependencies in data and therefore performs well in TNM as well as overall stage classification. Training and validation accuracy improves for both models steadily across epochs with less divergence and overfitting. These features emphasize the strength of ViT in solving complex classification challenges such as stage of lung cancer.

5.8.1.2. Loss Function

The categorical cross-entropy loss function is used for model optimization for multi-class classification problems. The categorical cross-entropy loss quantifies the dissimilarity between the predicted probability distribution and the true distribution of class labels. Its formula is given by:

$$L(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{5.9}$$

Here (y) is the true class distribution; (y^) is the predicted probability distribution, and (C) is the number of classes.

5.8.1.3. Evaluation Metrics

The main benchmark used to evaluate the models' performance is accuracy. The accuracy of classification is termed as the ratio of events predictively matching the actual observation over a total number of instances and is expressed mathematically as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{5.10}$$

In addition, the confusion matrix and classification reports are employed to gain an understanding of the model's performance. Such classification reports include performance metrics of precision, recall, and F1 score for all classes, thereby aiding an understanding of the model's discrimination of the various levels of cancer. A confusion matrix, which is illustrated in Figure 5.15., is used to assess the efficiency of a developed classification technique.

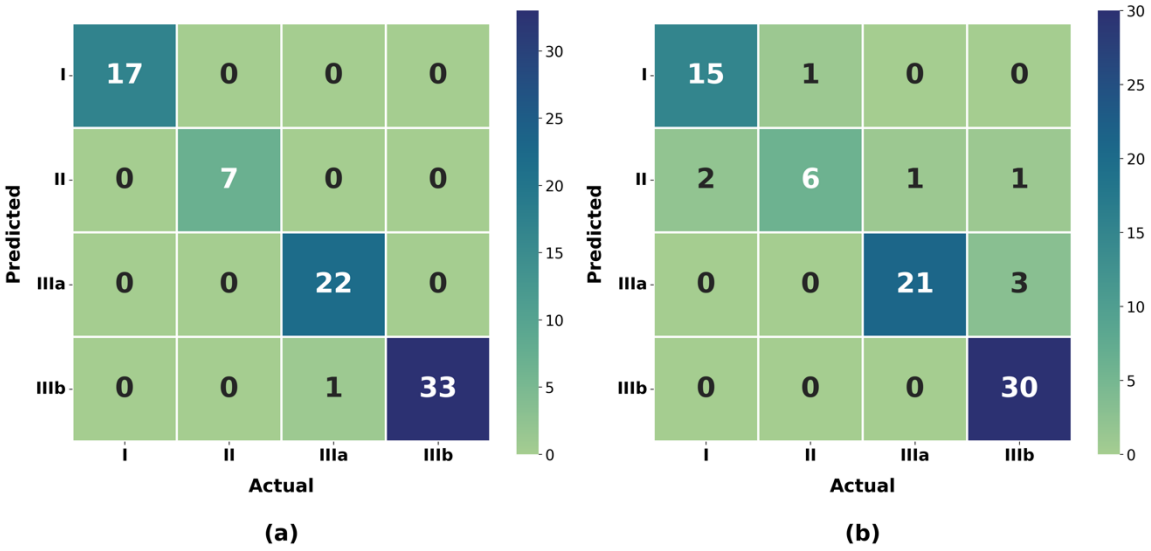

Figure 5.15. Confusion Matrix (a) for testing data using multi-input ViT model for Overall stage and (b) multi-input ViT TNM model.

5.8.1.4. Data Splitting and Cross-Field Validation

We use stratified data-splitting technique to guarantee the dependability and fairness of the model evaluations. This approach kept the distribution of several general phases across the test, validation, and training sets, therefore avoiding a distorted representation that would have biased the model.

Moreover, cross-field validation is used to firmly evaluate the generalization capacity of the suggested models. This included splitting the dataset into several folds and iteratively training and assessing the model on many combinations of training and validation sets. More thorough

evaluation of the performance of the model over several data subsets is offered by cross-field validation. Figure 5.16. shows the five-fold cross-validation; Table 5.2. shows its accuracy at every split.

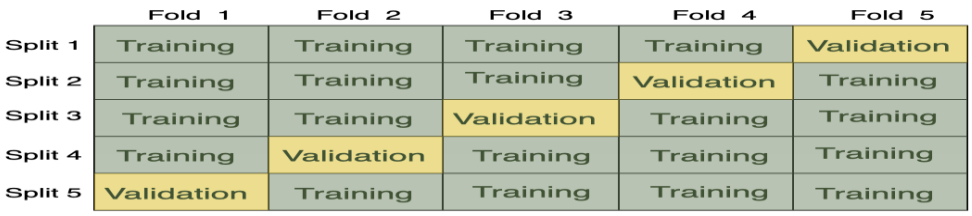

Figure 5.16. Dataset split for cross-validation analysis.

Table 5.2. Validation accuracy using multi-input ViT model for overall stage prediction.

CV Split	Validation accuracy using overall stage ViT model
Split 1	97.88
Split 2	98.26
Split 3	98.65
Split 4	98.75
Split 5	97.76
Average	98.28

To summarize, the experimental design was meticulously designed to ensure the models' integrity and effectiveness for NSCLC staging. Together with Google Colab's T4 GPU's processing capability, innovative designs, suitable loss functions, and comprehensive evaluation metrics establish a strong basis for the subsequent study of the data.

5.8.2. Results for TNM Stage Classification

Results of the TNM stage classification demonstrate the performance of two different architectures: a vision transformer and a deep learning model. Various input setups ranging from single-view images to including multiple views (axial, coronal and sagittal) and extra demographic data were used to assess these structures.

For the deep learning model, the accuracy scores varied across different input scenarios. Notably, the model achieved a 78% accuracy when trained on single-view images. Introducing multi-view data (axial and coronal) led to an improvement, resulting in an accuracy score of

81%. Further enhancements were observed with the inclusion of demographic features, where the model achieved an accuracy score of 83% for the combination of axial view with age and gender. The highest accuracy of 85% was attained when incorporating both axial and coronal views alongside age and gender information.

For the deep learning model, the accuracy scores varied across different input scenarios. Notably, the model achieved 78% accuracy when trained on single-axial-view images. Introducing multi-view data (axial and coronal) led to an improvement, resulting in an accuracy score of 81%. Further enhancements were observed with the inclusion of demographic features, where the model achieved an accuracy score of 83% for the combination of axial view with age and gender. The highest accuracy, which is 85% was attained when incorporating both axial and coronal views alongside age and gender information.

Table 5.3. Accuracy scores for TNM stage models

Model	Axial	Coronal	Sagittal	Multiview	Axial + Demo	Axial + Coronal + Demo
CNN	78	72	69	81	83	85
Vision Transformer	83	75	71	85	86	90

As seen in Table 5.3., the vision transformer regularly outperforms the deep learning model across all conditions. The vision transformer obtained an accuracy score of 83% upon given single-axial-view images. With multi-view data and an accuracy score of 85%, the vision transformer's advantages become clearer. Including demographic characteristics kept improving performance; axial view combined with age and gender had an accuracy score of 86%. Including axial and coronal views with age and gender information produced the most notable improvement, yielding an amazing accuracy score of 90%.

Especially in using multi-modal input data, these results highlight the relative strengths of the vision transformer design. The findings show how much it might improve TNM stage classification accuracy. Furthermore, underlined in the study is the need of incorporating several input configurations, including demographic data, to improve cancer staging predictive models.

5.8.3. Overall Stage Prediction Results

Two different architectures, Deep learning, a CNN model, and a vision transformer, are evaluated to demonstrate the performance of the stage prediction model. The accuracy scores

are presented for several input configurations, ranging from single-view images to multiple views, i.e., axial and coronal, along with other demographic data.

Depending on the input data provided, the deep learning model had varying levels of accuracy. While the model was trained solely with images taken from an axial view, accuracy was achieved at the level of 79%. The use of multi-view images (axial and coronal images) greatly improved the accuracy and achieved a score of 83%. The addition of demographic data achieved further improvement in the model with an accuracy score of 86% for combined axial view with age and gender. When both axial and coronal views, along with age and gender, are all combined, an accuracy of 87% is achieved, as shown in Table 5.4.

Table 5.4. Accuracy scores for Overall stage prediction

Model	Axial	Coronal	Sagittal	Multiview	Axial +Demo	Axial + Coronal + Demo
CNN	79	73	70.5	83	86	87
Vision Transformer	98.65	81.5	78	97.92	97.55	98.75

Among all scenarios, the outcome of predicting the overall stages was the best for the vision transformer as compared to the deep learning, i.e., the CNN model. For single-axial-view images, an accuracy score of 98.65% was achieved for the vision transformer. As the number of views increases in using multi-view images, the accuracy score drops to 97.92% for the vision transformer implementation. Gender and age features were included along with both the axial and coronal views, which resulted in an astonishing accuracy of 98.75%.

These results demonstrate the effectiveness of the vision transformer architecture in overall stage prediction which shows the superior accuracy compared to the deep learning model. It emphasizes the promise of the vision transformer when it comes to using multi-modal inputs and demographic information for more accurate predictions of the overall stage of cancer in patients.

5.8.4. Comparison with Existing Methods

5.8.4.1. TNM Classifier vs. Overall Stage Classifier

Several notable remarks signify how efficient and comprehensive the workings of the TNM stage classifier and the general stage classifier are when their findings are compared. The TNM stage classifier, integrated into a unified decision-making process through a decision tree algorithm, comprises distinct deep learning models for the categorization of T, N, and M stages.

This method seeks to improve interpretability by combining predictions tailored for different stages. The models obtained variable accuracy ratings; their performance varied significantly depending on the input arrangement.

Using both visual transformer architectures and deep learning, the general stage classifier concentrated on holistically predicting the cancer state. When multi-view data and demographic information were included, the deep learning model showed small but consistent accuracy gains. Emphasizing its robustness in using multi-modal data and demographic information, as shown in Figure 5.17., the vision transformer routinely exceeded the deep learning model across all scenarios.

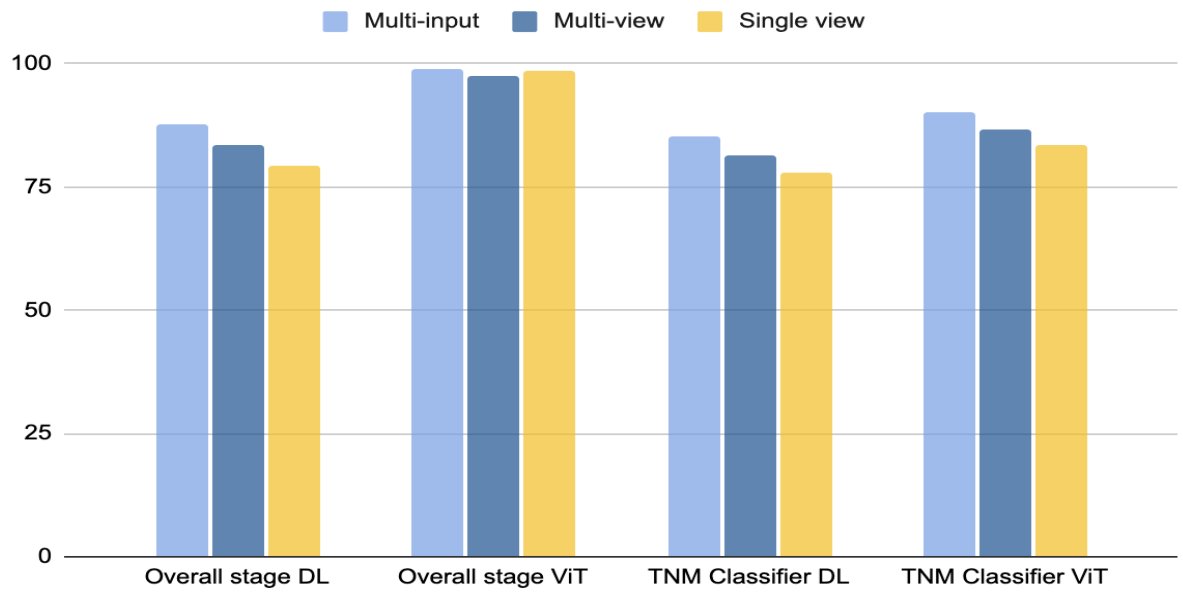

Figure 5.17. Comparative analysis of TNM classifier with Overall stage classifier

Although the TNM stage classifier explores the minutiae of tumor, node, and metastatic classifications, generally the stage classifier offers a more complete picture considering the whole cancer staging. Especially in general stage prediction, the vision transformer constantly outperforms the deep learning model in terms of flexibility to multi-modal input and demographic information.

5.8.4.2. Competitive analysis of Overall stage and TNM stage classifier based on CNN architecture

The deep learning-based models for both the Overall stage and the TNM stage performed differently depending on how the inputs were set up. To figure out the overall stage, the accuracy scores were: 79% for a single axial view, 73% for a single coronal view, 70.5%

for a single sagittal view, 83% for multi-view, 86% for an axial view plus demographic data, and 87% for axial and coronal views plus demographic data (87%). These results show that estimates are much more accurate when different points of view and demographic data are used. The better accuracy shows that looking at the imaging data from more than one angle seems to help us understand the tumor's traits better. This proves that using more than one view makes the model work better. In the same way, adding demographic information helps the model learn trends that are unique to each age group and gender. This makes it even better for prediction.

The TNM stage model did much better than the Overall stage model after multi-view and demographic data were added, but still it needs improvement There were seven different types of accuracy scores: 78% for a single axial view, 72% for a single coronal view, 69% for a single sagittal view, 81% for a multi-view, 83% for an axial view with demographic data, and 85% for axial and coronal views with demographic data. The experimental findings show the importance of multi-modal inputs and demographic data together for improving the deep learning model's accuracy in identifying different stages of tumors, lymph nodes, and metastases. Even it is hard to capture the details of each stage automatically, the TNM stage model gains from the extra data provided by inputs.

Table 5.5. Comparative Performance and Computational Requirements of CNN Architectures for Overall Stage and TNM Stage Classification

CNN	Overall stage	TNM
Number of layers	39	117
Number of parameters	102 814 661	308 443 983
Training time	6 hr	18hr
Testing time	30 ms	90 ms
Accuracy	87	80
Hardware requirements	CPU	CPU

As shown in Table 5.5., the Overall stage model clearly does better than the TNM stage model across all measures when comparing the two CNN classifiers. With multi-view and

demographic data added, the Overall stage predictor is much more accurate having accuracy of 87% than 80%, and its performance is greatly improved. The Overall stage model is also simpler, with 39 layers instead of 117 and 102,814,661 parameters instead of 308,443,983. This means that it takes less time to train, i.e., 6 hours as compared to 18 hours and compute (30 ms vs. 90 ms). The Overall stage model is more accurate and efficient, even though it has fewer parameters. This makes it a better choice for real-world uses. For both models, training and testing were done using CPU hardware, which made them easy to use.

5.8.4.3. Competitive analysis of Overall stage and TNM stage classifier based on ViT architecture

The Vision Transformer (ViT) architecture exhibited greater performance in comparison to CNN-based models, highlighting its increased capacity to handle intricate medical imaging data. The performance validity of the ViT model was quite impressive when it came to forecasting the overall stage. These scores were as follows: the Accuracy was 98.65 for the single axial view, 81.5 for the single coronal view, 78% for the single sagittal view, 97.92% for the multi-view approach, 97.55% for the combination of axial view with demographic data and 98.75% for the combination of axial and coronal view with demographic data. The enhanced accuracy attained through ViT architecture can be attributed to its capability to gather comprehensive context and detailed patterns within the entire image. The use of a multi-view approach enhances the performance of the model by combining different views, which is beneficial. Enhancement of the ViT model is made possible by the addition of demographic data, as it allows the ViT to learn age and gender-oriented features during the classification stages.

By adding the ViT structure into the TNM stage model, the classification accuracy scores have significantly risen: 83% for single axial, 75% for single coronal, 71% for single sagittal views, 85% for multiple views, 86% for axial and demographic data and 90% for axial, coronal and demographic data. The outcomes depict the efficiency of ViT architecture in handling the multi-modal input and demographic data with resilience and flexibility. The ViT consistently outperforms deep learning models in all scenarios due to the large datasets and high-resolution images, as well as its adaptability to include various data sources. It also performs exceptionally well, demonstrating its potential to provide more accurate and customized cancer staging, which will ultimately improve patient outcomes in the oncology domain.

In every criterion, the overall stage model outperforms the TNM stage model when compared to the ViT classifiers. The stage classifier shows a considerably higher accuracy of 98.75% as compared to 90%, indicating the exceptional performance of the ViT in this field. Table 5.6. shows that the Overall stage model requires a shorter training time of 10 hours as compared to 30 hours of computing time of 60 ms as compared to 180 ms, although having fewer layers of 32 as compared to 96) and parameters of 171,605,002 as compared to 514,815,006. The complex design and high processing demand of the ViT make GPU hardware necessary. However, its outstanding accuracy and efficacy in handling multi-modal inputs and demographic information make it the preferred choice for cancer staging applications.

Table 5.6. Comparison of ViT Classifiers for Overall Stage and TNM Stage Prediction

ViT	Overall stage	TNM
Number of layers	32	96
Number of parameters	171 605 002	514 815 006
Training time	10hr	30hr
Testing time	60 ms	180 ms
Accuracy	98.75	90
Hardware requirements	GPU	GPU

5.8.5. Compare the Proposed Approach with Existing Methods in the Literature.

The comparative results presented in this chapter have been refined to ensure transparency and fairness in evaluation. The proposed TNM and overall stage prediction models were developed using the NSCLC-Radiomics dataset [115], which was selected because it provides a comprehensive set of CT images and detailed clinical information, including patient age, gender, and tumour characteristics. Incorporating these clinical features significantly improved the predictive accuracy of the proposed model, achieving 98.75% accuracy and outperforming existing approaches, as shown in Table 5.7.

It is important to note that only a limited number of prior studies have attempted TNM stage prediction using similar data sources. Paing et al. [88] utilized the same NSCLC-Radiomics dataset [115], but their work focused solely on T-stage classification, without addressing the complete TNM staging system. In contrast, Moitra et al. [90] and Tyagi et al.

[92] applied their models using the NSCLC-Radiogenomics dataset [91], which differs from the dataset used in this study and contains limited demographic and clinical variables. These two studies were therefore included as comparative references to maintain a fair evaluation of the complete TNM classification framework, since they are among the few works addressing all TNM components.

Table 5.7. Comparison with other TNM classification approaches.

Method	Dataset	Classification task	Accuracy (%)
Krienko et al. [86]	Private	T-stage as T1/T2 and T3/T4	**82.6**
Paing et al. [88]	LIDC-IDRI [119] NSCLC-Radiomics [115] NSCLC-Radiomics-Genomics [120] NSCLC Radiogenomics [91]	T-stage as 7-stage classification	**90.6**
Zhao et al. [89]	Private	N-stage	**87.6**
Moitra et al. [90]	NSCLC Radiogenomics [91]	TNM stage	**96**
Tyagi, et al. [92]	NSCLC Radiogenomics [91]	TNM stage	**96.6**
The proposed work	NSCLC-Radiomics [115]	TNM stage	**98.75**

Consequently, the comparisons presented in Table 5.7. are based on methodological relevance rather than direct dataset equivalence. The superior results achieved by the proposed model demonstrate that integrating both imaging and clinical features from the NSCLC-Radiomics dataset provides more comprehensive and clinically aligned predictions for TNM and overall stage assessment.

Notably, Tyagi et al. [92] achieved 96.6% accuracy in overall stage classification; however, their method required unique TNM stage classifications and extensive preprocessing due to highly imbalanced T, N, and M data distributions. Additionally, their approach relied on computationally expensive 3D CT scans. In contrast, the proposed model leverages Vision Transformers to efficiently learn discriminative features from 2D CT images captured in multiple views. This design enhances feature representation while significantly reducing computational cost and processing time, leading to improved classification accuracy.

5.9. Discussion

The adoption of the Vision Transformer (ViT) and CNN-based architectures for TNM stage prediction was guided by the need to effectively capture both local imaging cues and global contextual relationships. Traditional CNNs are powerful for extracting spatially localised tumour characteristics from axial, coronal, and sagittal views, whereas ViTs enable long-range dependency modelling through self-attention mechanisms. This combination facilitates holistic feature representation, allowing the network to integrate morphological and clinical information (e.g., tumour size, lymph node spread, and metastasis indicators). The ViT-based design thus provides superior interpretability and scalability for multimodal, multi-view TNM classification and overall stage prediction. This chapter uses CNN and ViT architectures to compare and implement TNM stage classification with direct overall stage classification to improve lung cancer staging. The outcomes, as shown in Tables 5.3. and 5.4., offer a thorough analysis of the model's performance in various input configurations.

In terms of TNM stage and total stage predictions, ViT architecture continuously outperforms the CNN models. The results indicate that for TNM stage classification as given in Table 5.3., the ViT performed better than the CNN in all input configurations, with the greatest accuracy of 90% reached when using combined axial and coronal views with demographic data. Similarly, using the identical input configuration, the ViT achieved an astounding 98.75% accuracy for total stage prediction, as given in Table 5.4, substantially surpassing the CNN, which only achieved 87%.

Tables 5.5. and 5.6. provide a detailed analysis of the CNN and ViT model comparisons for the two staging approaches. The ViT models were more sophisticated and required GPU hardware, even though they had fewer layers and parameters than their CNN equivalents. This was demonstrated by the longer training times and higher computational resources needed.

Nonetheless, the significant improvements in accuracy and the capacity to efficiently handle intricate, multi-modal input data justify this computational expenditure.

There were some important differences in how the CNN and ViT models were trained for overall and TNM stage classifications. Figures 5.18a and 5.18b show the train and validation accuracy and loss curves for each model, which show these changes.

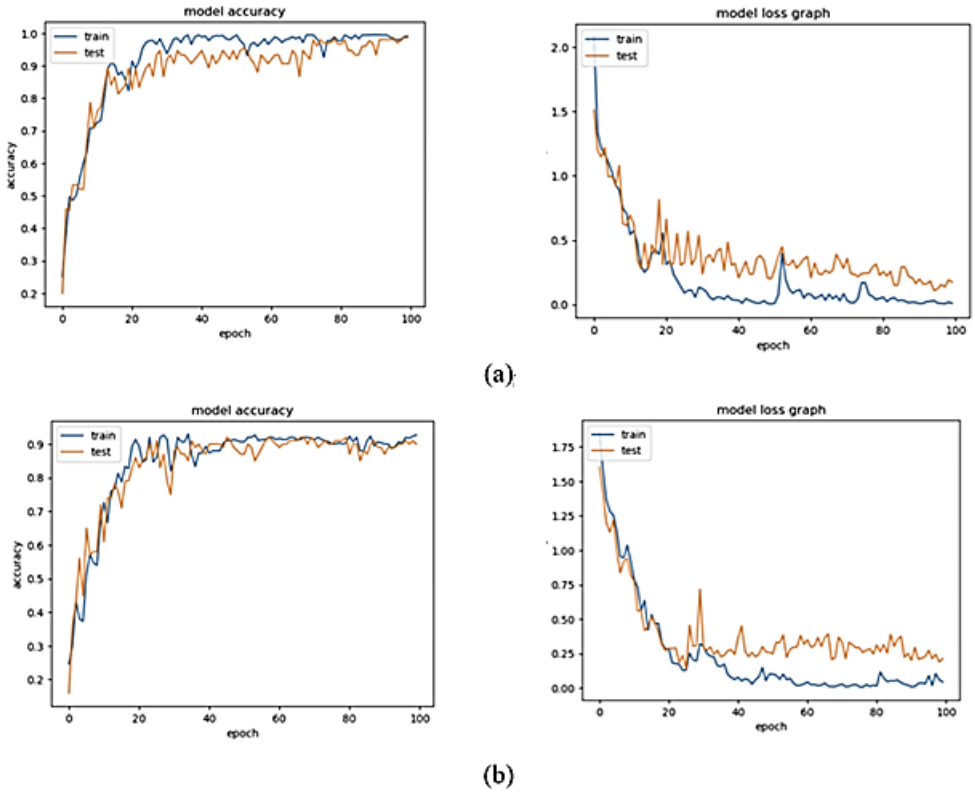

Figure 5.18. Training and Validation Accuracy and Loss Curves for Overall Stage (a) and TNM Stage Models (b) Using ViT Architecture

The training accuracy of the Overall Stage Model begins at about 25% and steadily improves, reaching more than 99% by the end of training. The validation accuracy, on the other hand, begins at 20% and rises gradually, with notable spikes around the mid-epochs, reaching a peak of about 98.7%. This pattern suggests that the model is gradually acquiring features that are pertinent to the overall classification of lung cancer stages, with periods of accelerated performance improvement. The train and validation loss curves show the same trend. The initial high training loss of about 2.07 drops slowly to under 0.05, showing that the model is learning well and convergent. At the completion of training, validation loss lowers and levels off at 0.02. This model is well-suited for clinical deployment where it must remain stable over

unseen data because the train and validation metrics are quite similar, indicating effective regularization and robustness with minimal overfitting.

The ViT model demonstrates strong performance in the TNM Stage Model. Consistent with the general stage model, both training and validation accuracy steadily rise across epochs; however, validation accuracy does experience small swings before settling near 90-92% at the conclusion. The training process generalizes well to the validation set because the validation loss is very close to the training loss across all epochs. This demonstrates that the ViT model is capable of rapidly processing complex multi-modal data, such as fluctuations in TNM stages, without experiencing significant overfitting.

In general, the effectiveness of ViT for both TNM and general lung cancer staging tasks is underscored by the models' consistent and seamless convergence during training. When it comes to lung cancer, the ViT model is an invaluable resource for precise and personalized treatment planning due to its high reliability and accuracy in staging the disease. Despite the fact that it requires a lot of processing power, this is demonstrated by the high accuracy of the validation and low loss at the end of the training process.

Figure 5.15. depicts the confusion matrices for the Overall and TNM stage models adjacent to each other. This study enables us to evaluate each model's classification accuracy and stage-specific discrimination capabilities.

The confusion matrix shows that all of the stage labels in panel (a), which shows the Overall stage model, are very accurate. In particular, 100% accuracy was achieved in the classification of stages I, II, and IIIa. Stage IIIb had a 97% classification accuracy, with only one example misclassified from stage IIIa, demonstrating the model's superior performance in detecting later stages.

Panel (b) shows the TNM stage model, which is also accurate across all levels. Stage I was correctly identified 94% of the time; only once was it wrongly identified as stage II. Stage II attained a classification accuracy of 86%, with misclassifications occurring in stages I, IIIa, and IIIb. On the other hand, stages IIIa and IIIb had high recognition rates of 88% and 100%, respectively.

Overall, the comparison of the confusion matrices shows that the Overall stage model does a good job of classifying things with few mistakes. The results show that the multi-input ViT model can categorize stages well, which suggests it could be helpful in clinical settings where it is important to identify stages precisely.

The choice of suitable treatment strategies for patients depends much on the correct classification of lung cancer stages. Using the strength of learning at several stages, the

suggestion of a multi-input Vision Transformer network has demonstrated excellent promise in enhancing performance. The creation of an autonomous lung cancer overall stage classification system has great potential to help medical professionals in developing more exact and customized treatment strategies for lung cancer patients.

This work represents significant progress in the application of the most advanced machine learning approaches, including Vision Transformers, particularly in lung cancer stage diagnosis. Besides stressing the importance of proper stage identification, the proposed work also accentuates the importance of automating the process. This would greatly enhance the ability of the medical doctor to make decisions and, hence, improve patient care.

This is because, as evidenced by this study, lung cancer staging is a developing field, and more advancements and improvements can still be made. With regard to the model expansion, clinical validation and the inclusion of additional sources and modalities seem to be the priority directions for development. The cause for the further improvement of the more effective, more accurate staging of lung cancer is a never-ending pursuit, and the expectation work presents progress toward that aim.

5.9.1. Key Findings

The exploration of deep learning, i.e., CNN and visual transformer architectures, has revealed interesting results concerning the classification of TNM stages and the prognosis of patient stages. These findings are quite basis for understanding and treating this type of cancer.

Granularity vs. Holistic Perspective:

The TNM stage classifier provides a deeper understanding or analysis of the progress made in the progression of cancer by categorizing the T, N, and M stages into their corresponding subdivisions, thus creating new means of viewing various aspects of the disease. An understanding of the whole is crucial, especially for the customization of treatment and for assessing the peculiarities of the disease effects in the region. Conversely, the comprehensive stage prediction model, especially utilizing the vision transformer, evaluates the entire cancer staging from a holistic perspective. This comprehensive approach enhances the global understanding of the pattern of evolution of diseases over a time frame.

Model Performance Disparities:

Within the domain of deep learning, the performance of the vision transformer was comparatively better than the conventional deep learning model with varied input configurations. As such, the vision transformer's strong points include its ability to incorporate multi-modal data and demographic factors, which proves useful in predicting the cancer stage.

Multi-Modal Advantage:

Axial, coronal, and demographic information has improved the accuracy of both TNM and overall stage classifiers. This means that using additional information enhances the accuracy of the estimation of the cancer stage. However, the vision transformer performed better than the traditional deep learning model because it was designed to comprehend multimodal data.

Clinical Implications:

The current study findings highlight the need for personalized treatment approaches that consider specific nuances derived from TNM staging, particularly for localized therapies. At the same time, overall stage prediction bolstered with a vision transformer architectural approach remains an effective approach to comprehensively evaluating the burden of NSCLC and planning appropriate management.

Path Forward:

The implementation of vision transformer architectures within cancer staging models indicates the possibility of revolutionary progress in medical imaging. Further research could focus on enhancing the predictive models by incorporating additional clinical and molecular information to improve our understanding of the disease and allow treatment tailoring.

In conclusion, the principal findings reveal the interplay of details and holistic perspectives in cancer staging. The vision transformer, with its astounding performance, provides avenues for future studies in the use of transformer-based deep learning models for improved medical image analysis and cancer prediction.

5.9.2. Research Questions Revisited

During the evaluation, several critical factors were identified that have an impact on the mortality rates of patients with NSCLC. Age appears to be one of the factors that is important, which seems to be critical in the diagnosis of lung cancer. Furthermore, the inclusion of demographic factors such as age and gender combined with imaging data significantly

improved the accuracy of the overall stage diagnosis. This underlines the need to integrate clinical data with demographic observations to improve the accuracy of predictive models.

The subgroup studies we conducted revealed interesting patterns, particularly indicating a higher level of model accuracy in individuals who are 65 years or older, with an accuracy score of 99.3%. This underscores the significance of age as a prognostic indicator, especially among the senior cohort, necessitating tailored treatment and support strategies for this specific group. Among individuals under the age of 64, a significant accuracy score of 98.2% has been reported. With respect to gender, there was a slight-fine variation in model accuracy between the males and females, with values of 98.2% and 98.85%, respectively.

The study revealed no notable differences in the prediction accuracies among the histological diagnoses. It would appear that in spite of the diversity that exists in histological types such as large cell, adenocarcinoma, and squamous cell carcinoma, the accuracy measures were similar, around about 98.5%. Such an observation will imply that the presented predictive model is effective in predicting regardless of the histological subtypes employed and thus is likely to be useful to improve overall stage prediction outcomes.

Moreover, the investigation included an evaluation of how the clinical N stage impacts overall stage prediction. In predictive model building, the clinical N stage turned out to be the most dominant feature, with a significant impact on accuracy. It is important to note that the clinical N stage is critical in assessing how the disease progresses, what therapeutic options should be taken, and the prognosis of patients with non-small cell lung cancer (NSCLC).

In conclusion, it should be stressed that the forthcoming research concerning the stage of non-small cell lung cancer will be of significant importance in age and gender definition factors affecting patients' stage perceptions and survival rates. Machine learning models are able to utilize these variables to increase the accuracy of prediction and personalized patient treatment. Diagnostic imaging studies and treatment with these factors necessitate the need to combine their use with other clinical and demographic factors to build holistic and robust predictive models. Relative to this last factor, more work needs to be done with regard to understanding these relationships accurately and enhancing patient management.

5.10. Summary

In conclusion, this chapter is a pioneering initiative in the prediction of NSCLC stages. This study substantially enhances NSCLC staging techniques by presenting a novel deep learning

architecture for TNM stage classification, employing Vision Transformers to improve TNM classification, and creating a direct model for overall stage prediction with a multi-input framework. The integration of these advanced techniques significantly improves the accuracy and interpretability of lung cancer staging, offering a more precise and personalised framework for clinical decision-making. These contributions demonstrate how deep learning can bridge the gap between tumour segmentation and clinical staging, establishing a foundation for automated, end-to-end cancer assessment.

Building upon these results, the next chapter provides the overall conclusion and future research directions, summarizing the key outcomes from all three technical chapters and outlining how the proposed AI-based solutions can be further developed for real-world clinical integration.

Chapter 6

6. Conclusion

6.1.1. Thesis Summary

The primary objective of this thesis is to classify Chest X-ray images based on lung problems, segment lung tumors using CT and PET images, and predict the overall stage of lung cancer. The primary contribution of this study is the implementation of an efficient deep-learning framework for the precise classification of lung diseases, including pneumonia, lung cancer, tuberculosis, lung capacity, and COVID-19. The design employs a pre-trained VGG16 model and three convolutional neural network blocks for classification. The U-Net-based deep models are employed to segment lung cancer in PET and CT images, effectively identifying and separating lung cancer across various data types. The structures have profound consequences for comprehending and managing NSCLC. The study also examines the classification of the TNM stage and predicts the overall stage, evaluating the suggested segmentation and prediction models using well-established performance measures. The discoveries have substantial ramifications for comprehending and managing NSCLC.

The thesis is organized into six chapters, with Chapter 2 offering a thorough analysis of existing studies on lung diseases utilizing medical imaging techniques, primarily focusing on Chest X-ray pictures. The literature's second portion looks at segmentation methods used with multi-modality images, specifically CT and PET scans. The third section emphasizes on studies for classifying lung cancer stages using clinical and imaging data. Several methods have been studied in the literature that make use of applied pre-processing, transfer learning, deep learning, and ML

Chapter 3 presents a novel deep-learning system that utilizes chest X-ray images to classify Pneumonia, Lung Cancer, tuberculosis (TB), Lung Opacity, and COVID-19. There is also a thorough explanation of the pre-processing procedures that were done on the dataset and the open-source dataset that was used. The architecture of the proposed model is provided for the suggested method. There is also an explanation of the mathematical notations used for the performance indicators. Moreover, a thorough presentation and discussion of the accuracy and loss graphs, together with the results, are provided. The chapter concludes with a discussion of

future directions and possibilities for categorizing chest X-ray images into various lung disorders, including lung cancer.

In Chapter 4, the recommendation is to utilize a powerful deep-learning architecture known as U-net for accurate segmentation of lung cancer. This approach utilizes several types of imaging data, mostly CT and PET scans, to obtain accurate outcomes. The proposed models exhibit intricate structures that incorporate various fusion approaches, such as early, late, dense, hyper-dense, and hyper-dense VGG16 U-Net. Each model's benefits and drawbacks are highlighted. All model's results are compared with those of the benchmark models. The experiments with various loss functions are performed in model training, and their performance is compared. The predicted segmented image of each model is compared with the matching ground truth, and a performance evaluation is performed using standard performance metrics.

Chapter 5 introduces an innovative method for classifying the overall stage of non-small cell lung cancer (NSCLC) by employing advanced deep-learning algorithms, including Vision Transformers. This approach entails the examination of a dataset that has many inputs, such as radiological and clinical data. This considers the conventional TNM approach in staging, investigates the stage determining clinical variables, and justifies the application of Transformers in the medical sector. The stage predictions performed using the TNM staging classifier are compared with the overall stage classifier and the benchmarked models.

6.1.2. Limitations and Future Work

6.1.2.1. Multi-class Lung Disease Classification

The research emphasized the categorization of different types of chest diseases using CXR images and achieved remarkable results. However, there are some drawbacks and some prospects for further research.

The primary limitations pertain to the quality and representativeness of the training data. Biases within the dataset may potentially affect the model’s generalizability, particularly with respect to classes that are underrepresented. Furthermore, the model's performance may differ among various populations or imaging protocols not included in the training data.

Future studies may focus on integrating multi-modal imaging data (e.g., combining CXR and CT scans) with additional clinical and demographic information such as patient age, gender, histopathology, and genetic markers. Such integration can enhance diagnostic accuracy and improve predictive modelling for lung disease progression and treatment outcomes.

Moreover, the development of more sophisticated segmentation models aimed at identifying specific ROIs in the chest images may aid the classification model. Nailing specific structures or pathological abnormalities inside an image may enhance algorithm performance.

6.1.2.2. Lung Tumor Segmentation Using Multimodality of CT-PET Scans

The deep learning techniques for lung tumor segmentation, as demonstrated in this study, have been successfully utilized. There are some limitations and possibilities for future scope to be considered. All trained systems depend on the availability and size of annotated datasets, which in turn defines the maximal performance of the segmentation model. In such conditions, the small data size may induce overfitting and restrict the ability of the model to generalize. Furthermore, inconsistency in the imaging protocols and the annotating standards among different datasets may also bring challenges during the training and validation of the model.

Future research could apply multiple data sources to overcome these challenges and make further progress in the investigation. The segmentation model would improve heterogeneous lung tumor treatment and delineation using complementary imaging datasets. Moreover, segmentation performance could be enhanced by optimizing the loss functions, for example, by examining the combination of binary and Dice loss functions. Using the loss functions to address class imbalances or specific properties of lung tumor variants may help alleviate some complications arising from data variability and noise. In conclusion, systematic hyperparameter tuning experiments could improve segmentation performance even further. Investigating different hyperparameter values and testing them on different datasets would achieve optimum performance in a variety of imaging conditions.

6.1.2.3. Non-Small Cell Lung Cancer TNM Classification and Overall Stage Prediction Using Vision Transformers

To enhance lung cancer staging using the aforementioned Vision Transformer-based multimodal architecture, it is essential to appreciate the current shortcomings and indicate avenues for future exploration. The study notes that firstly, it relies on one data source, which, although useful, is likely to be insufficient to achieve the full variability of data necessary for an in-depth model. The representation of a few stages in the dataset may still be inadequate, as class imbalance has been addressed systematically with the use of data augmentation. More studies, therefore, need to be done to understand possible ways in which this challenge can be addressed.

In addition, although Vision Transformers are effective in identifying visual patterns, the challenge of understanding their workings persists. Future research should prioritize enhancing the model's transparency in its decision-making procedures. Additionally, the present model utilizes age and gender solely as textual features. Future work should be geared towards increasing specificity and prognostic accuracy by accommodating a wider variety of data concerning patients.

Further studies can focus on conserving more nuanced approaches that incorporate clinical records, genetic information, and other similar attributes to aid in understanding the progression of lung cancer. The performance of the model should also be assessed in clinical settings based on a larger and more diversified data set than the one used to develop the model. Such prerequisite validation will be designed to involve medical institutions and healthcare practitioners to prove the applicability of the model in practice. It also recommends developing approaches to explainable AI to ease the model's acceptance in clinical settings by explicating the model's decision. Future work should concentrate on improving the computational cost of the Vision Transformer models to expand their usage in more health care settings. One important aspect is the design of a real-time support system for decision-making for oncologists. This needs study into streamlining the model for rapid diagnosis and treatment planning.

In conclusion, the limitations of the study include factors such as the lack of diversity in the datasets and poor generalizability across populations and imaging options. Small annotated datasets and class imbalance add more benefits but also encompass threats to the model performance, hence posing an overfitting risk. Moreover, the interpretability of deep learning models is a major barrier to their clinical implementation because conditions of a non-transparent “black box” lead to considerable discomfort among clinicians regarding the reliability of predictive capabilities. Future clinical models should concentrate on integrating multi-modal data, encompassing clinical and imaging information alongside patient-specific data, such as genomics, to enhance predictive accuracy. Also, advanced explainable artificial intelligence (XAI) diagnostics derived from these models can be incorporated along with accelerating processing speed to ensure the clinical applicability of the diagnoses. These advancements will improve both the performance and efficiency of AI in lung cancer detection and its practical deployment.

References

1. Rahane, W., Dalvi, H., Magar, Y., Kalane, A., Jondhale, S.: Lung cancer detection using image processing and machine learning healthcare. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT). pp. 1–5. IEEE (2018). https://doi.org/10.1109/icctct.2018.8551008.

2. Bhatt, M.L.B., Kant, S., Bhaskar, R.: Pulmonary tuberculosis as differential diagnosis of lung cancer. South Asian J Cancer. 1, 36–42 (2012). https://doi.org/10.4103/2278-330x.96507.

3. Siegel, R.L., Miller, K.D., Wagle, N.S., & Jemal, A. (2023). Cancer statistics, 2023. CA: A Cancer Journal for Clinicians, 73(1), 17–48. https://doi.org/10.3322/caac.21763

4. Krishnaiah, V., Narsimha, G., & Chandra, N.S. (2013). Diagnosis of lung cancer prediction system using data mining classification techniques. International Journal of Computer Science and Information Technologies, 4, 39–45.

5. Emek Soylu, B., Guzel, M.S., Bostanci, G.E., Ekinci, F., Asuroglu, T., Acici, K.: Deep-learning-based approaches for semantic segmentation of natural scene images: A review. Electronics (Basel). 12, 2730 (2023). https://doi.org/10.3390/electronics12122730.

6. Shah, A.A., Alturise, F., Alkhalifah, T., & Khan, Y.D. (2022). Evaluation of deep learning techniques for identification of sarcoma-causing carcinogenic mutations. Digital Health, 8, 20552076221133704. https://doi.org/10.1177/20552076221133703

7. Amin, S.U., Alsulaiman, M., Muhammad, G., Mekhtiche, M.A., Hossain, M.S.: Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Future Generation computer systems. 101, 542–554 (2019). https://doi.org/10.1016/j.future.2019.06.027.

8. Immanuel, R.R., Sangeetha, S.K.B.: Analysis of EEG Signal with Feature and Feature Extraction Techniques for Emotion Recognition Using Deep Learning Techniques. In: International Conference on Computational Intelligence and Data Engineering. pp. 141–154. Springer (2022). https://doi.org/10.1007/978-981-99-0609-3_10.

9. Ning, J., Ge, T., Jiang, M., Jia, K., Wang, L., Li, W., Chen, B., Liu, Y., Wang, H., Zhao, S.: Early diagnosis of lung cancer: which is the optimal choice? Aging (Albany NY). 13, 6214 (2021). https://doi.org/10.18632/aging.202504.

10. Wankhade, S., Vigneshwari, S.: A novel hybrid deep learning method for early detection of lung cancer using neural networks. Healthcare Analytics. 3, 100195 (2023). https://doi.org/10.1016/j.health.2023.100195.

11. Pei, X., Zuo, K., Li, Y., Pang, Z.: A review of the application of multi-modal deep learning in medicine: Bibliometrics and future directions. International Journal of Computational Intelligence Systems. 16, 44 (2023). https://doi.org/10.1007/s44196-023-00225-6.

12. Alshmrani, G.M.M., Ni, Q., Jiang, R., Pervaiz, H., Elshennawy, N.M.: A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images.

Alexandria Engineering Journal. 64, 923–935 (2023). https://doi.org/10.1016/j.aej.2022.10.053.

13. Alshmrani, G.M., Ni, Q., Jiang, R., Muhammed, N.: Hyper-Dense_Lung_Seg: Multimodal-Fusion-Based Modified U-Net for Lung Tumour Segmentation Using Multimodality of CT-PET Scans. Diagnostics. 13, 3481 (2023). https://doi.org/10.3390/diagnostics13223481.

14. Khan, M.A.: An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier. IEEE Access. 8, 34717–34727 (2020). https://doi.org/10.1109/ACCESS.2020.2974687.

15. Khan, M.A., Quasim, M.T., Alghamdi, N.S., Khan, M.Y.: A Secure Framework for Authentication and Encryption Using Improved ECC for IoT-Based Medical Sensor Data. IEEE Access. 8, 52018–52027 (2020). https://doi.org/10.1109/ACCESS.2020.2980739.

16. Albahli, S.: Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. Int J Med Sci. 17, 1439–1448 (2020). https://doi.org/10.7150/ijms.46684.

17. Sathitratanacheewin, S., Sunanta, P., Pongpirul, K.: Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability. Heliyon. 6, e04614 (2020). https://doi.org/10.1016/j.heliyon.2020.e04614.

18. Gao, X.W., James-Reynolds, C., Currie, E.: Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. Neurocomputing. 392, 233–244 (2020). https://doi.org/10.1016/j.neucom.2018.12.086.

19. Hooda, R., Mittal, A., Sofat, S.: Automated TB classification using ensemble of deep architectures. Multimed Tools Appl. 78, 31515–31532 (2019). https://doi.org/10.1007/s11042-019-07984-5.

20. Chandra, T.B., Verma, K., Singh, B.K., Jain, D., Netam, S.S.: Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. Expert Syst Appl. 158, 113514 (2020). https://doi.org/10.1016/j.eswa.2020.113514.

21. Kumar, J.S., Balamurugan, S.A. alias, Sasikala, S.: Analysis of Deep Learning Techniques for Tuberculosis Disease. SN Comput Sci. 2, 302 (2021). https://doi.org/10.1007/s42979-021-00680-y.

22. Lopes, U.K., Valiati, J.F.: Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput Biol Med. 89, 135–143 (2017). https://doi.org/10.1016/j.compbiomed.2017.08.001.

23. Rahimzadeh, M., Attar, A.: A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inform Med Unlocked. 19, 100360 (2020). https://doi.org/10.1016/j.imu.2020.100360.

24. Luján-García, J., Yáñez-Márquez, C., Villuendas-Rey, Y., Camacho-Nieto, O.: A Transfer Learning Method for Pneumonia Classification and Visualization. Applied Sciences. 10, 2908 (2020). https://doi.org/10.3390/app10082908.

25. Luján-García, J.E., Moreno-Ibarra, M.A., Villuendas-Rey, Y., Yáñez-Márquez, C.: Fast COVID-19 and Pneumonia Classification Using Chest X-ray Images. Mathematics. 8, 1423 (2020). https://doi.org/10.3390/math8091423.

26. Stephen, O., Sain, M., Maduh, U.J., Jeong, D.-U.: An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. J Healthc Eng. 2019, 1–7 (2019). https://doi.org/10.1155/2019/4180949.

27. Lascu, M.-R.: Deep Learning in Classification of Covid-19 Coronavirus, Pneumonia and Healthy Lungs on CXR and CT Images. J Med Biol Eng. 41, 514–522 (2021). https://doi.org/10.1007/s40846-021-00630-2.

28. Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., Ibragimov, B.: Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Computers & Electrical Engineering. 78, 388–399 (2019). https://doi.org/10.1016/j.compeleceng.2019.08.004.

29. El Asnaoui, K.: Design ensemble deep learning model for pneumonia disease classification. Int J Multimed Inf Retr. 10, 55–68 (2021). https://doi.org/10.1007/s13735-021-00204-7.

30. Goyal, S., Singh, R.: Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. J Ambient Intell Humaniz Comput. 14, 3239–3259 (2023). https://doi.org/10.1007/s12652-021-03464-7.

31. S.K., L., Mohanty, S.N., K., S., N., A., Ramirez, G.: Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 92, 374–382 (2019). https://doi.org/10.1016/j.future.2018.10.009.

32. Song, Q., Zhao, L., Luo, X., Dou, X.: Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. J Healthc Eng. 2017, 1–7 (2017). https://doi.org/10.1155/2017/8314740.

33. Singh, G.A.P., Gupta, P.K.: Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Comput Appl. 31, 6863–6877 (2019). https://doi.org/10.1007/s00521-018-3518-x.

34. Kalaivani, N., Manimaran, N., Sophia, Dr.S., D Devi, D.: Deep Learning Based Lung Cancer Detection and Classification. IOP Conf Ser Mater Sci Eng. 994, 012026 (2020). https://doi.org/10.1088/1757-899X/994/1/012026.

35. ALzubi, J.A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A., Thaventhiran, C.: Boosted neural network ensemble classification for lung cancer disease diagnosis. Appl Soft Comput. 80, 579–591 (2019). https://doi.org/10.1016/j.asoc.2019.04.031.

36. Li, J., Li, J., Wu, J., Song, Y., Hu, S., Hong, J., Wang, W.: Change in symptom clusters perioperatively in patients with lung cancer. European Journal of Oncology Nursing. 55, 102046 (2021). https://doi.org/10.1016/j.ejon.2021.102046.

37. Ju, W., Xiang, D., Zhang, B., Wang, L., Kopriva, I., Chen, X.: Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images. IEEE Transactions on Image Processing. 24, 5854–5867 (2015). https://doi.org/10.1109/TIP.2015.2488902.

38. Greco, C., Rosenzweig, K., Cascini, G.L., Tamburrini, O.: Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). Lung Cancer. 57, 125–134 (2007). https://doi.org/10.1016/j.lungcan.2007.03.020.

39. Cellina, M., Cè, M., Irmici, G., Ascenti, V., Khenkina, N., Toto-Brocchi, M., Martinenghi, C., Papa, S., Carrafiello, G.: Artificial Intelligence in Lung Cancer Imaging: Unfolding the Future. Diagnostics. 12, 2644 (2022). https://doi.org/10.3390/diagnostics12112644.

40. Cellina, M., Cacioppa, L.M., Cè, M., Chiarpenello, V., Costa, M., Vincenzo, Z., Pais, D., Bausano, M.V., Rossini, N., Bruno, A., Floridi, C.: Artificial Intelligence in Lung Cancer Screening: The Future Is Now. Cancers (Basel). 15, 4344 (2023). https://doi.org/10.3390/cancers15174344.

41. Hosny, K.M., Elshoura, D., Mohamed, E.R., Vrochidou, E., Papakostas, G.A.: Deep Learning and Optimization-Based Methods for Skin Lesions Segmentation: A Review. IEEE Access. 11, 85467–85488 (2023). https://doi.org/10.1109/ACCESS.2023.3303961.

42. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015).

43. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. Presented at the (2015). https://doi.org/10.1007/978-3-319-24574-4_28.

44. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. (2014).

45. Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A.: Cancer statistics, 2022. CA Cancer J Clin. 72, 7–33 (2022). https://doi.org/10.3322/caac.21708.

46. Valente, I.R.S., Cortez, P.C., Neto, E.C., Soares, J.M., de Albuquerque, V.H.C., Tavares, J.M.R.S.: Automatic 3D pulmonary nodule detection in CT images: A survey. Comput Methods Programs Biomed. 124, 91–107 (2016). https://doi.org/10.1016/j.cmpb.2015.10.006.

47. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med Image Anal. 42, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005.

48. Purandare, N.C., Rangarajan, V.: Imaging of lung cancer: Implications on staging and management. Indian Journal of Radiology and Imaging. 25, 109–120 (2015). https://doi.org/10.4103/0971-3026.155831.

49. Cancer, About Lung Cancer, https://www.cancer.org/cancer/lungcancer/ about/what-is.html., last accessed 2024/03/31.

50. Detterbeck, F.C., Boffa, D.J., Kim, A.W., Tanoue, L.T.: The eighth edition lung cancer stage classification. Chest. 151, 193–203 (2017). https://doi.org/10.1016/j.chest.2016.10.010.

51. Singh, R.K., Pandey, R., Babu, R.N.: COVIDScreen: explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays. Neural Comput Appl. 33, 8871–8892 (2021). https://doi.org/10.1007/s00521-020-05636-6.

52. Kassania, S.H., Kassanib, P.H., Wesolowskic, M.J., Schneidera, K.A., Detersa, R.: Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach. Biocybern Biomed Eng. 41, 867–879 (2021). https://doi.org/10.1016/j.bbe.2021.05.013.

53. Narin, A., Kaya, C., Pamuk, Z.: Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Analysis and Applications. 24, 1207–1220 (2021). https://doi.org/10.1007/s10044-021-00984-y.

54. Jia, G., Lam, H.-K., Xu, Y.: Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. Comput Biol Med. 134, 104425 (2021). https://doi.org/10.1016/j.compbiomed.2021.104425.

55. Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., Chen, J., Wang, R., Zhao, H., Chong, Y., Shen, J., Zha, Y., Yang, Y.: Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) With CT Images. IEEE/ACM Trans Comput Biol Bioinform. 18, 2775–2780 (2021). https://doi.org/10.1109/TCBB.2021.3065361.

56. Perumal, V., Narayanan, V., Rajasekar, S.J.S.: Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. Applied Intelligence. 51, 341–358 (2021). https://doi.org/10.1007/s10489-020-01831-z.

57. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., Xia, J.: Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology. 296, E65–E71 (2020). https://doi.org/10.1148/radiol.2020200905.

58. Li, C., Dong, D., Li, L., Gong, W., Li, X., Bai, Y., Wang, M., Hu, Z., Zha, Y., Tian, J.: Classification of Severe and Critical Covid-19 Using Deep Learning and Radiomics. IEEE J Biomed Health Inform. 24, 3585–3594 (2020). https://doi.org/10.1109/JBHI.2020.3036722.

59. Shi, J., Yuan, X., Elhoseny, M., Yuan, X.: Weakly Supervised Deep Learning for Objects Detection from Images. Presented at the (2020). https://doi.org/10.1007/978-3-030-45099-1_18.

60. Dansana, D., Kumar, R., Bhattacharjee, A., Hemanth, D.J., Gupta, D., Khanna, A., Castillo, O.: Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm. Soft comput. 27, 2635–2643 (2023). https://doi.org/10.1007/s00500-020-05275-y.

61. Ezzat, D., Hassanien, A.E., Ella, H.A.: An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. Appl Soft Comput. 98, 106742 (2021). https://doi.org/10.1016/j.asoc.2020.106742.

62. Ravi, V., Narasimhan, H., Chakraborty, C., Pham, T.D.: Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images. Multimed Syst. 28, 1401–1415 (2022). https://doi.org/10.1007/s00530-021-00826-1.

63. Pathak, Y., Shukla, P.K., Tiwari, A., Stalin, S., Singh, S., Shukla, P.K.: Deep Transfer Learning Based Classification Model for COVID-19 Disease. IRBM. 43, 87–92 (2022). https://doi.org/10.1016/j.irbm.2020.05.003.

64. Gupta, A., Anjum, Gupta, S., Katarya, R.: InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using Chest X-ray. Appl Soft Comput. 99, 106859 (2021). https://doi.org/10.1016/j.asoc.2020.106859.

65. Monowar, K.F., Hasan, Md.A.M., Shin, J.: Lung Opacity Classification With Convolutional Neural Networks Using Chest X-rays. In: 2020 11th International Conference on Electrical and Computer Engineering (ICECE). pp. 169–172. IEEE (2020). https://doi.org/10.1109/ICECE51571.2020.9393135.

66. Latif, G., Al Anezi, F.Y., Sibai, F.N., Alghazo, J.: Lung Opacity Pneumonia Detection with Improved Residual Networks. J Med Biol Eng. (2021). https://doi.org/10.1007/s40846-021-00656-6.

67. Sitaula, C., Hossain, M.B.: Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Applied Intelligence. 51, 2850–2863 (2021). https://doi.org/10.1007/s10489-020-02055-x.

68. Thakur, S., Kumar, A.: X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN). Biomed Signal Process Control. 69, 102920 (2021). https://doi.org/10.1016/j.bspc.2021.102920.

69. Wang, S., Mahon, R., Weiss, E., Jan, N., Taylor, R.J., McDonagh, P.R., Quinn, B., Yuan, L.: Automated Lung Cancer Segmentation Using a PET and CT Dual-Modality Deep Learning Neural Network. International Journal of Radiation Oncology*Biology*Physics. 115, 529–539 (2023). https://doi.org/10.1016/j.ijrobp.2022.07.2312.

70. Park, J., Kang, S.K., Hwang, D., Choi, H., Ha, S., Seo, J.M., Eo, J.S., Lee, J.S.: Automatic Lung Cancer Segmentation in [18F]FDG PET/CT Using a Two-Stage Deep Learning Approach. Nucl Med Mol Imaging. 57, 86–93 (2023). https://doi.org/10.1007/s13139-022-00745-7.

71. Xiang, D., Zhang, B., Lu, Y., Deng, S.: Modality-Specific Segmentation Network for Lung Tumor Segmentation in PET-CT Images. IEEE J Biomed Health Inform. 27, 1237–1248 (2023). https://doi.org/10.1109/JBHI.2022.3186275.

72. Fu, X., Bi, L., Kumar, A., Fulham, M., Kim, J.: Multimodal Spatial Attention Module for Targeting Multimodal PET-CT Lung Tumor Segmentation. IEEE J Biomed Health Inform. 25, 3507–3516 (2021). https://doi.org/10.1109/JBHI.2021.3059453.

73. Zhong, Z., Kim, Y., Zhou, L., Plichta, K., Allen, B., Buatti, J., Wu, X.: 3D fully convolutional networks for co-segmentation of tumors on PET-CT images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 228–231. IEEE (2018). https://doi.org/10.1109/ISBI.2018.8363561.

74. Hwang, S.E., Hwang, D., Kang, S.K., Choi, H., Ha, S., Eo, J.S., Lee, J.S.: 3C-Net: Deep Learning-based Lung Cancer Segmentation Using Multi-Context Information on FDG PET/CT Images Journal of Nuclear Medicine. 63, 3349 (2022).

75. Kumar, A., Fulham, M., Feng, D., Kim, J.: Co-Learning Feature Fusion Maps From PET-CT Images of Lung Cancer. IEEE Trans Med Imaging. 39, 204–217 (2020). https://doi.org/10.1109/TMI.2019.2923601.

76. Jemaa, S., Fredrickson, J., Carano, R.A.D., Nielsen, T., de Crespigny, A., Bengtsson, T.: Tumor Segmentation and Feature Extraction from Whole-Body FDG-PET/CT Using Cascaded 2D and 3D Convolutional Neural Networks. J Digit Imaging. 33, 888–894 (2020). https://doi.org/10.1007/s10278-020-00341-1.

77. Zhao, X., Li, L., Lu, W., Tan, S.: Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. Phys Med Biol. 64, 015011 (2018). https://doi.org/10.1088/1361-6560/aaf44b.

78. Zhong, Z., Kim, Y., Plichta, K., Allen, B.G., Zhou, L., Buatti, J., Wu, X.: Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks. Med Phys. 46, 619–633 (2019). https://doi.org/10.1002/mp.13331.

79. Bi, L., Fu, X., Liu, Q., Song, S., Feng, D.D., Fulham, M., Kim, J.: Hyper-connected transformer network for co-learning multi-modality pet-ct features. ArXiv. (2022).

80. Malaviya, N., Rahevar, M., Virani, A., Ganatra, A., Bhuva, K.: LViT: Vision Transformer for Lung cancer Detection. In: 2023 International Conference on Artificial Intelligence and Smart Communication (AISC). pp. 93–98. IEEE (2023). https://doi.org/10.1109/AISC56616.2023.10085230.

81. Liu, D., Liu, F., Tie, Y., Qi, L., Wang, F.: Res-trans networks for lung nodule classification. Int J Comput Assist Radiol Surg. 17, 1059–1068 (2022). https://doi.org/10.1007/s11548-022-02576-5.

82. Wang, R., Zhang, Y., Yang, J.: TransPND: A Transformer Based Pulmonary Nodule Diagnosis Method on CT Image. Presented at the (2022). https://doi.org/10.1007/978-3-031-18910-4_29.

83. Kulkarni, A., Panditrao, A.: Classification of lung cancer stages on CT scan images using image processing. In: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. pp. 1384–1388. IEEE (2014). https://doi.org/10.1109/ICACCCT.2014.7019327.

84. Ignatious, S., Joseph, R., John, J., Prahladan, A.: Computer Aided Lung Cancer Detection and Tumor Staging in CT image using Image Processing. Int J Comput Appl. 128, 29–33 (2015). https://doi.org/10.5120/ijca2015906607.

85. Firdaus Abdullah, M., Noraini Sulaiman, S., Khusairi Osman, M., Karim, N.K.A., Lutfi Shuaib, I., Danial Irfan Alhamdu, M.: Classification of Lung Cancer Stages from CT Scan Images Using Image Processing and k-Nearest Neighbours. In: 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). pp. 68–72. IEEE (2020). https://doi.org/10.1109/ICSGRC49013.2020.9232492.

86. Kirienko, M., Sollini, M., Silvestri, G., Mognetti, S., Voulaz, E., Antunovic, L., Rossi, A., Antiga, L., Chiti, A.: Convolutional Neural Networks Promising in Lung Cancer T-Parameter Assessment on Baseline FDG-PET/CT. Contrast Media Mol Imaging. 2018, 1–6 (2018). https://doi.org/10.1155/2018/1382309.

87. Jakimovski, G., Davcev, D.: Using Double Convolution Neural Network for Lung Cancer Stage Detection. Applied Sciences. 9, 427 (2019). https://doi.org/10.3390/app9030427.

88. Paing, M.P., Hamamoto, K., Tungjitkusolmun, S., Pintavirooj, C.: Automatic Detection and Staging of Lung Tumors using Locational Features and Double-Staged Classifications. Applied Sciences. 9, 2329 (2019). https://doi.org/10.3390/app9112329.

89. Zhao, X., Wang, X., Xia, W., Li, Q., Zhou, L., Li, Q., Zhang, R., Cai, J., Jian, J., Fan, L., Wang, W., Bai, H., Li, Z., Xiao, Y., Tang, Y., Gao, X., Liu, S.: A cross-modal 3D deep learning for accurate lymph node metastasis prediction in clinical stage T1 lung adenocarcinoma. Lung Cancer. 145, 10–17 (2020). https://doi.org/10.1016/j.lungcan.2020.04.014.

90. Moitra, D., Kr. Mandal, R.: Classification of non-small cell lung cancer using one-dimensional convolutional neural network. Expert Syst Appl. 159, 113564 (2020). https://doi.org/10.1016/j.eswa.2020.113564.

91. NSCLC Radiogenomics - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki, https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics.

92. Tyagi, S., Talbar, S.N.: LCSCNet: A multi-level approach for lung cancer stage classification using 3D dense convolutional neural networks with concurrent squeeze-and-excitation module. Biomed Signal Process Control. 80, 104391 (2023). https://doi.org/10.1016/j.bspc.2022.104391.

93. Basu, A., Das, S., Ghosh, S., Mullick, S., Gupta, A., Das, S.: Chest X-Ray Dataset for Respiratory Disease Classification. Harvard Dataverse, V5. . (2021). https://doi.org/https://doi.org/10.7910/DVN/WNQ3GI.

94. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule. American Journal of Roentgenology. 174, 71–74 (2000). https://doi.org/10.2214/ajr.174.1.1740071.

95. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2097–2106 (2017).

96. Bisong, E.: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA (2019). https://doi.org/10.1007/978-1-4842-4470-8.

97. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. (2014).

98. A Deep Learning Architecture for Multi-Class Lung Diseases Classification using Chest X-ray Images, last accessed 2024/03/21.

99. Khan, M.A., Algarni, F.: A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS. IEEE Access. 8, 122259–122269 (2020). https://doi.org/10.1109/ACCESS.2020.3006424.

100. BIMCV-COVID19 – BIMCV, https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/.

101. Gil, J.Y., Kimmel, R.: Efficient dilation, erosion, opening, and closing algorithms. IEEE Trans Pattern Anal Mach Intell. 24, 1606–1617 (2002). https://doi.org/10.1109/TPAMI.2002.1114852.

102. Sreedhar, K.: Enhancement of Images Using Morphological Transformations. International Journal of Computer Science and Information Technology. 4, 33–50 (2012). https://doi.org/10.5121/ijcsit.2012.4103.

103. Yi-de Ma, Qing Liu, Zhi-bai Quan: Automated image segmentation using improved PCNN model based on cross-entropy. In: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004. pp. 743–746. IEEE. https://doi.org/10.1109/ISIMP.2004.1434171.

104. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal Loss for Dense Object Detection. IEEE Trans Pattern Anal Mach Intell. 42, 318–327 (2020). https://doi.org/10.1109/TPAMI.2018.2858826.

105. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M.J. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In Proceedings of the International Workshop on Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA 2017), Lecture Notes in Computer Science (Vol. 10553, pp. 240–248). Cham: Springer. https://doi.org/10.1007/978-3-319-67558-9_28

106. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (2016). https://doi.org/10.1109/CVPR.2016.90.

107. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2261–2269. IEEE (2017). https://doi.org/10.1109/CVPR.2017.243.

108. Vallières, M., Freeman, C.R., Skamene, S.R., El Naqa, I.: A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol. 60, 5471–5496 (2015). https://doi.org/10.1088/0031-9155/60/14/5471.

109. Müller, D., Soto-Rey, I., Kramer, F.: Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes. 15, 210 (2022). https://doi.org/10.1186/s13104-022-06096-y.

110. TNM, Cancer Staging System, https://www.uicc.org/resources/tnm, last accessed 2023/03/31.

111. Detterbeck, F.C., Boffa, D.J., Kim, A.W., Tanoue, L.T.: The Eighth Edition Lung Cancer Stage Classification. Chest. 151, 193–203 (2017). https://doi.org/10.1016/j.chest.2016.10.010.

112. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE. 86, 2278–2324 (1998). https://doi.org/10.1109/5.726791.

113. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. (2020).

114. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. (2016).

115. Aerts, H., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D.: Data from NSCLC-radiomics. The cancer imaging archive. (2015).

116. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv Neural Inf Process Syst. 30, (2017).

117. Han, K., Xiao, A., Wu, E., Guo, J., XU, C., Wang, Y.: Transformer in Transformer. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., and Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems. pp. 15908–15919. Curran Associates, Inc. (2021).

118. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research. 21, 1–67 (2020).

119. Armato, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., MacMahon, H., van Beek, E.J.R., Yankelevitz, D., Biancardi, A.M., Bland, P.H., Brown, M.S., Engelmann, R.M., Laderach, G.E., Max, D., Pais, R.C., Qing, D.P. -Y., Roberts, R.Y., Smith, A.R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G.W., Jude, C.M., Munden, R.F., Petkovska, I., Quint, L.E., Schwartz, L.H., Sundaram, B., Dodd, L.E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Vande Casteele, A., Gupte, S., Sallam, M., Heath, M.D., Kuhn, M.H., Dharaiya, E., Burns, R., Fryd, D.S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B.Y., Clarke, L.P.: The Lung

Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans. Med Phys. 38, 915–931 (2011). https://doi.org/10.1118/1.3528204.

120. Data From NSCLC-Radiomics-Genomics.