

# CHAPTER 1

## INTRODUCTION

### 1.1. Background

Cancer refers to a collection of diseases marked by the uncontrolled growth and spread of abnormal cells. According to the American Cancer Society, cancer is responsible for approximately 1 in every 6 deaths worldwide, surpassing other diseases such as AIDS, tuberculosis, and malaria combined. It has become one of the most significant health challenges of the 21st century, affecting both developing and developed nations. Cancer not only inflicts physical and mental distress on patients but also places a heavy economic burden on families. In 2020, global cancer statistics revealed an estimated 19.3 million new cases and nearly 10.0 million cancer-related deaths (Sung et al., 2021). The World Cancer Report (2008) projected that the number of cancer-related deaths would rise to 13.1 million annually by 2030. However, the future burden of cancer is expected to be much greater due to the increasing prevalence of risk factors such as smoking, poor diet, and physical inactivity. The World Health Organization (WHO, 2017) estimates that by the age of 75, approximately 21 out of 100 men and 18 out of 100 women globally will be diagnosed with cancer. Over the past four to five decades, the global burden of cancer has increased significantly.

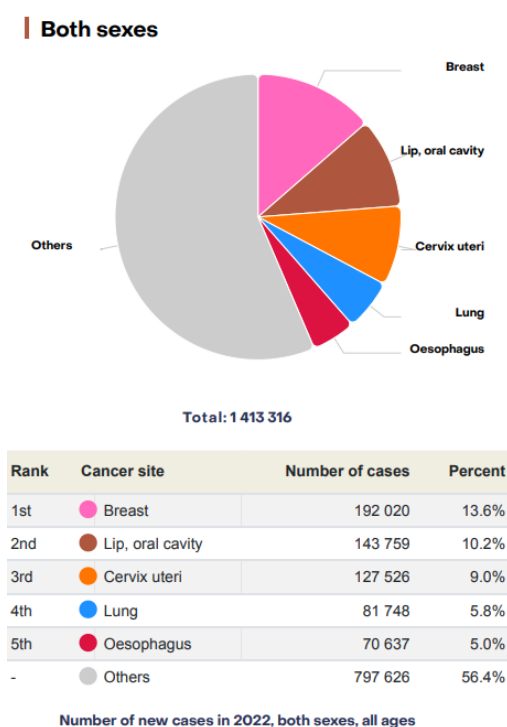


Fig 1.1 Cancer Incidence and Mortality Rates in India(2022)

(Source: <https://gco.iarc.who.int/media/globocan/factsheets/populations/356-india-fact-sheet.pdf>)

According to the GLOBOCAN 2022 report by the WHO, India recorded approximately 1,413,316 new cancer diagnoses and 916,827 cancer-related deaths, with a 5-year prevalence of about 3,258,518 cases.

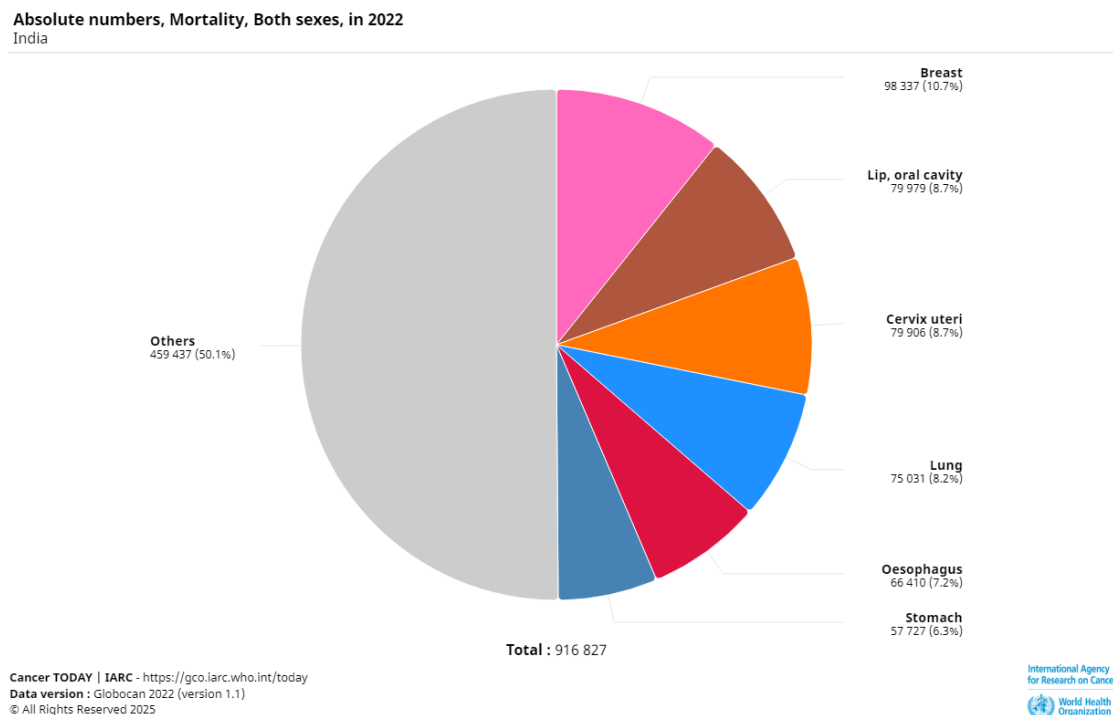


Fig.1.2, Mortality in India (2022).

(Source:[https://gco.iarc.fr/today/en/dataviz/pie?mode=cancer&group\\_populations=1&types=1&populations=356](https://gco.iarc.fr/today/en/dataviz/pie?mode=cancer&group_populations=1&types=1&populations=356))

In 2015, the direct medical cost of cancer in the United States was estimated at \$80.2 billion (Bandera et al., 2016). Similarly, in Europe, cancer-related healthcare expenses reached €83 billion (around \$110 billion) in 2014 (Willett and Hu, 2013). In developing nations like Brazil, Russia, Indonesia, China, and South Africa, the total productivity loss due to premature cancer-related deaths was valued at \$46.3 billion in 2012 (Tan, 2004). Globally, the economic impact of cancer is projected to rise due to the growing number of new cases and the increasing cost of treatments (Onis et al., 2007). Between 2020 and 2050, cancer is expected to cost the global economy over \$25 trillion (Lopes, 2023). The economic burden of cancer includes expenses for primary, outpatient, emergency, and inpatient care, along with medication costs. Additionally, indirect costs such as lost earnings from premature death, cancer-related disabilities, family caregiving, and income loss for individuals who leave employment temporarily or permanently due to illness significantly add to the overall burden (Fernandez et al., 2013).

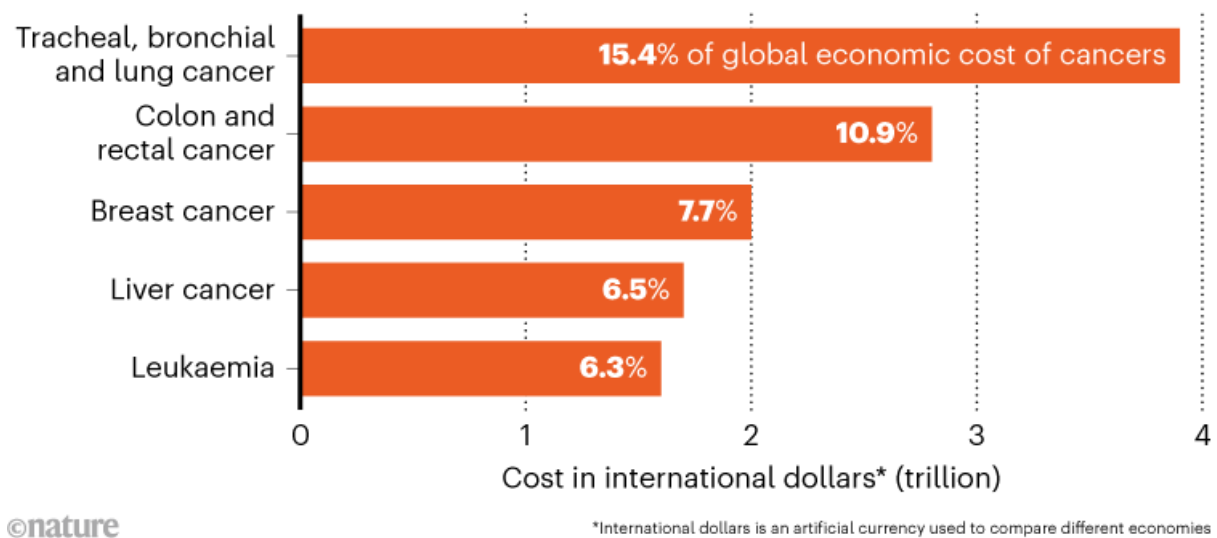


Fig.1.3(a) Estimated Economic burden by 2050 of major cancer types depicting Lung Cancer leading with 15.4% of global economic cost . (Source: <https://canceratlas.cancer.org/> CANCER.ATLAS.ORG)

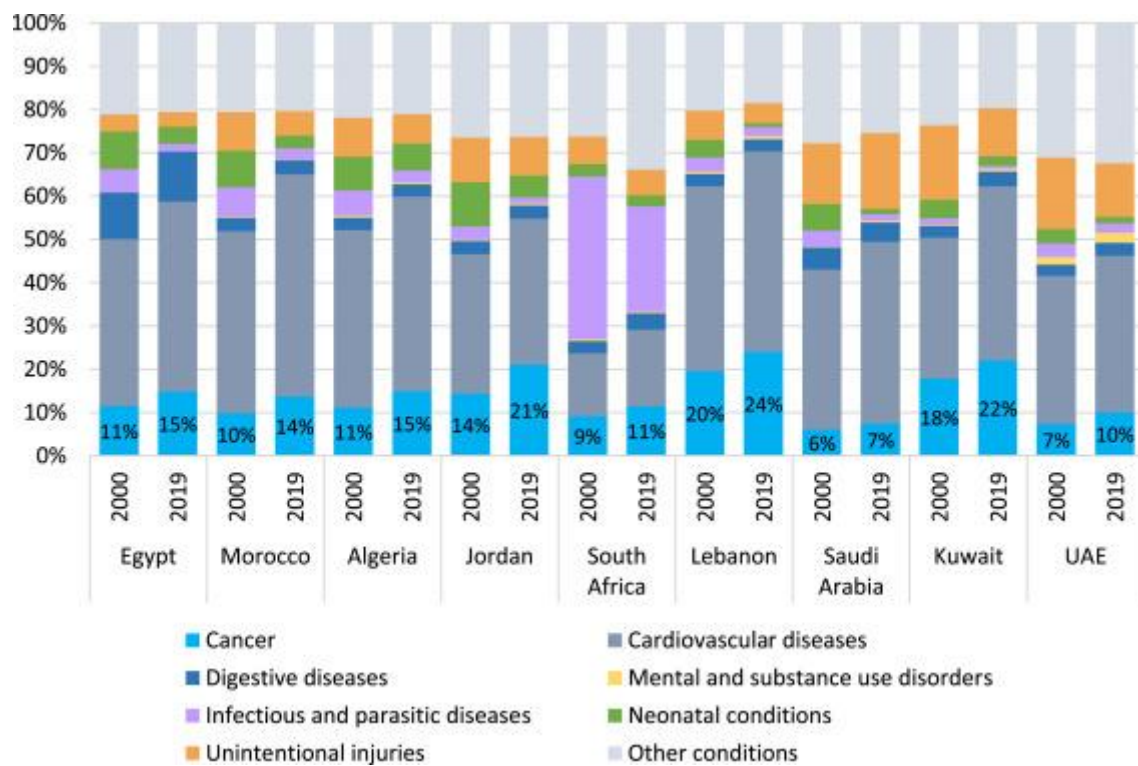


Fig.1.3(b) The Disease and Economic Burden of Cancer in 9 Countries in the Middle East and Africa.

(Source: <https://www.sciencedirect.com/science/article/pii/S2212109923000547>)

## 1.2 Overview of Lung Cancer and Lung Cancer Variants

Lung cancer is among the most prevalent and fatal cancers globally. Early detection plays a crucial role in improving survival rates and the effectiveness of treatment (Siegel et al., 2020). As reported by the International Agency for Research on Cancer (IARC), lung cancer leads all cancers in incidence and mortality worldwide. Its early stages often remain undetected, resulting in a high mortality rate by the time symptoms develop. While lung cancer can occur at any age, it is more common in individuals over 50, with key risk factors including smoking, radon exposure, air pollution, and occupational hazards (American Cancer Society, 2021). Lung cancer is primarily categorized into two types: non-small cell lung cancer (NSCLC), which accounts for approximately 85% of cases, and small cell lung cancer (SCLC), known for its rapid progression and aggressive nature. According to the GLOBOCAN 2020 report from the World Health Organization, around 2.2 million new lung cancer cases were diagnosed globally, with approximately 1.8 million deaths, reflecting a mortality rate of about 81.8%. These figures highlight the critical need for progress in early detection, treatment innovations, and prevention strategies (Sung et al., 2021).

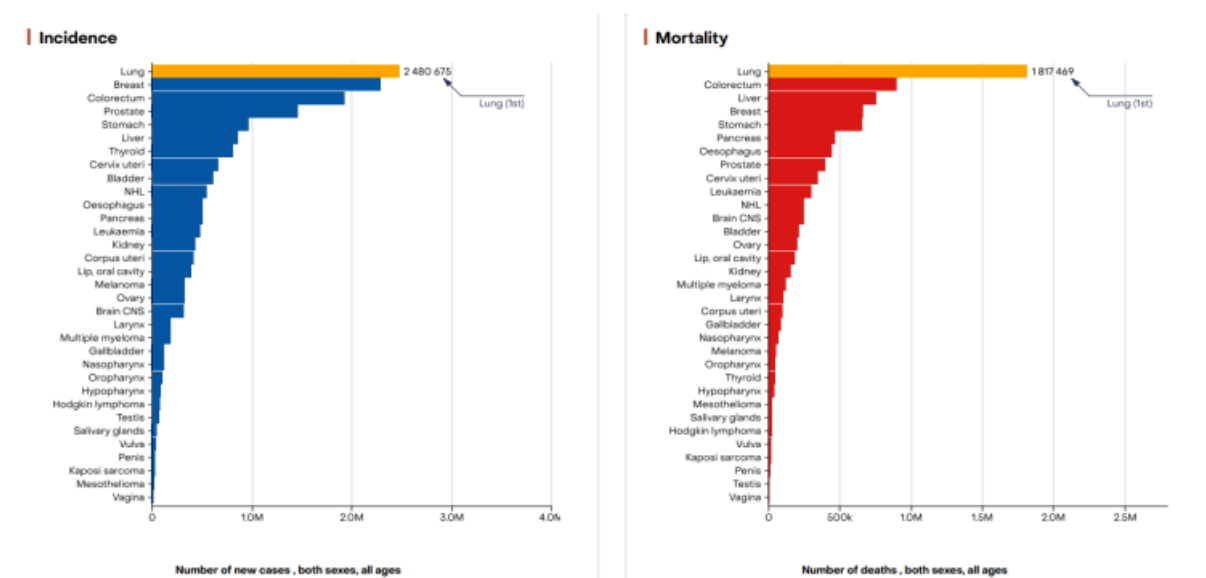


Fig.1.4 Incidence and Mortality of Lung Cancer.

(Source:<https://gco.iarc.who.int/media/globocan/factsheets/cancers/15-trachea-bronchus-and-lung-fact-sheet.pdf> WHO GLOBOCAN 2022 REPORT)

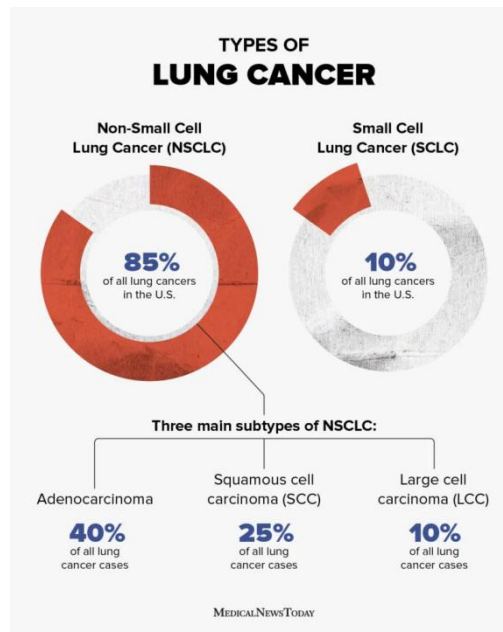


Fig.1.5 Classification of NSCLC

(Source: <https://www.medicalnewstoday.com/articles/prevalence-of-non-small-cell-lung-cancer>)

Lung cancer is generally classified into two major types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), based on the appearance of cancer cells under a microscope. NSCLC makes up about 85% of all lung cancer cases and includes subtypes like adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Molina et al., 2008). Adenocarcinoma, the most common subtype, typically forms in the outer regions of the lungs and is often associated with non-smokers (Thanigaiarasu et al., 2024). Squamous cell carcinoma originates in the central regions of the lungs near the bronchial tubes and is commonly linked to smoking. Large cell carcinoma, though less frequent, is an aggressive form that can appear in any part of the lung. In contrast, SCLC accounts for approximately 15% of lung cancer cases and is noted for its rapid growth and early metastasis, making it harder to treat (Rudin et al., 2021). It is strongly linked to heavy smoking. Other rarer types of lung cancer include mesothelioma, which affects the lung lining, and carcinoid tumors, which grow slowly and originate from neuroendocrine cells. A clear understanding of these cancer types is essential for developing specific diagnostic and treatment strategies.

### 1.3 Identifying Lung Cancer

Lung cancer is a leading cause of cancer-related deaths globally. It is often diagnosed at a later stage due to subtle or absent symptoms in its early phases, limiting treatment options. Early detection plays a crucial role in improving patient outcomes. Diagnosing lung cancer typically involves a combination of clinical evaluation, imaging techniques, biopsy, histopathological

examination, molecular testing, staging and advanced digital pathology tools such as Whole Slide Imaging (WSI).

**Clinical Evaluation:** The diagnostic process starts with a thorough clinical evaluation. Healthcare providers assess the patient's medical history, focusing on factors like smoking habits and exposure to environmental toxins such as radon, asbestos, or second-hand smoke (Samet et al., 2009).

**Imaging Techniques:** Imaging is vital for detecting and evaluating lung abnormalities (Tárnoki et al., 2024). Chest X-rays are commonly used for an initial overview, identifying potential masses or nodules. If any abnormalities are seen, a computed tomography (CT) scan is often performed. CT scans provide detailed cross-sectional images of the lungs, helping to assess tumor size, location, and whether it has spread to nearby structures (Rubin, 2015). Positron Emission Tomography (PET) scans evaluate the tumor's metabolic activity, determining if it is active or has metastasized (Griffeth, 2005). Magnetic Resonance Imaging (MRI) may be employed to obtain detailed images of soft tissues, particularly if the cancer is suspected to involve the brain or spine. These imaging techniques collectively offer essential insights for diagnosis, staging, and guiding further tests or treatments.

**Biopsy and Histopathological Examination:** A biopsy is essential for confirming cancer. During this procedure, tissue samples are taken from the tumor for examination. A pathologist then analyzes the sample under a microscope to confirm the presence of cancer and classify its type (e.g., adenocarcinoma, squamous cell carcinoma, or small cell lung cancer).

**Molecular Testing:** Molecular testing plays a key role in diagnosing lung cancer, especially for identifying tumor types and tailoring treatment. This test examines the tumor's genetic makeup to detect mutations in genes such as EGFR, ALK, ROS1, or KRAS (Lindeman et al., 2013). These mutations assist in determining the type of lung cancer (e.g., non-small cell lung cancer or NSCLC) and help guide targeted therapies. Molecular testing is often performed on biopsy tissue samples, or sometimes on blood samples through liquid biopsy. Identifying genetic changes allows doctors to select more effective treatments, improving outcomes and minimizing side effects.

**Staging:** Staging is the process of assessing the extent of cancer spread in the body (Liam et al., 2020). It helps doctors determine the most appropriate treatment and predict the patient's prognosis. Lung cancer staging is based on tumor size, its spread to nearby lymph nodes, and whether it has metastasized to other parts of the body. Stages range from I to IV, with Stage I indicating early, localized cancer, and Stage IV representing advanced cancer that has spread extensively.

**Whole Slide Imaging (WSI):** Whole Slide Imaging (WSI) plays a transformative role in the histopathological examination of lung cancer. WSI involves digitizing entire microscope slides

at high resolution, allowing pathologists to analyze and share virtual slides remotely. This enhances diagnostic accuracy, facilitates second opinions, and enables the use of artificial intelligence (AI) tools for pattern recognition and tumor classification. WSI also supports integration with molecular data, enabling a more comprehensive understanding of tumor biology. Furthermore, digital slides can be archived and retrieved efficiently for future comparisons or education, making WSI a valuable tool in modern pathology workflows.

#### **1.4 Noise in Histopathological Imaging**

Noise in Whole Slide Images (WSIs) of histopathological samples is a significant challenge that affects image analysis quality and diagnostic precision. This noise can arise from various factors, including inconsistencies in tissue staining, scanner artifacts, and environmental conditions during image capture (Huang et al., 2021). One of the most common types of noise encountered in WSIs is Gaussian noise, which is caused by random electronic fluctuations in imaging devices (Kanmani & Rajivkannan, 2018). Additionally, speckle noise, resulting from coherent signal interference, complicates the detection of subtle morphological features critical for cancer diagnosis (You et al., 2019). Compression artifacts, occurring during storage or transmission, can also introduce impulse noise, leading to distortion in tissue structures (Buades et al., 2005). Tackling these noise types often involves advanced denoising techniques, such as Non-Local Means (NLM) filtering, which helps to reduce noise while preserving essential image details (Buades et al., 2005). Furthermore, deep learning approaches, such as convolutional neural networks (CNNs) and Generative Adversarial Networks (GANs), have shown considerable potential in adaptively reducing noise in WSIs while retaining the integrity of histological features (Jiang et al., 2020). Implementing effective noise reduction methods is vital for improving the accuracy of automated analysis systems and boosting the reliability of pathology workflows (Huang et al., 2021).

#### **1.5 Denoising Method**

Denoising methods are essential for improving the quality of Whole Slide Images (WSI) used in diagnosing lung cancer, as noise can greatly impact the accuracy of image analysis and classification. One commonly used technique is Gaussian filtering, which applies a weighted average to smooth noise while preserving key image features (Kanmani & Rajivkannan, 2018). Another effective method is median filtering, which reduces impulse noise by replacing each pixel's value with the median of its neighboring pixels, maintaining the structural integrity of histopathological images (Kanmani & Rajivkannan, 2018). Advanced techniques, such as Non-Local Means (NLM) filtering, have been applied to WSIs to enhance noise reduction by evaluating the similarity of pixels in a local neighborhood, making it ideal for preserving fine details in medical images (Buades et al., 2005). More recently, deep learning-based denoising approaches, such as autoencoders and convolutional neural networks (CNNs), have gained popularity for their ability to learn complex noise patterns and adaptively remove noise while retaining crucial diagnostic features (Jiang et al., 2020). Additionally, Generative Adversarial

Networks (GANs) have shown promise in denoising WSIs, improving image clarity and reducing artifacts (You et al., 2019). Implementing these denoising techniques can enhance the performance of subsequent tasks, such as cancer classification and segmentation, thereby facilitating more accurate diagnostic results (Huang et al., 2021).

### **1.6. Utilizing WSI Images and AI for Lung Cancer Detection**

Whole Slide Imaging (WSI) has revolutionized lung cancer diagnosis by allowing for in-depth digital analysis of histopathological samples. However, these high-resolution images often contain noise that can obscure important details and affect diagnostic accuracy. Tackling noise in WSI is crucial for enhancing image quality and diagnostic reliability (Zarella et al., 2019).

Recent research has introduced deep learning-based denoising methods specifically designed for WSI. One example is the patch label denoising technique, which improves the classification accuracy of cancerous tissues in WSIs by reducing label noise (Yang et al., 2022). Another method uses convolutional neural networks to minimize noise and variability in tissue images, enabling more consistent and precise analysis (Huang et al., 2021).

Traditional noise reduction techniques, such as Gaussian and median filtering, continue to be important in preprocessing WSI data. Gaussian filtering works by applying a weighted average to smooth noise, while median filtering is effective for preserving edges and removing artifacts (Kanmani & Rajivkannan, 2018).

Incorporating these advanced noise reduction techniques into WSI enhances the reliability and accuracy of lung cancer diagnoses. Improved image clarity allows pathologists to make more accurate assessments, leading to better patient outcomes (Kothari et al., 2013).

### **1.7 Aims and Objective**

#### **Aim:**

To evaluate and compare the effectiveness of different image denoising techniques in enhancing the quality of histopathological WSIs for accurate lung carcinoma classification using deep learning.

#### **Objectives:**

- a. To collect annotated histopathological image tiles in collaboration with oncopathologists from an ICMR-funded project.
- b. To apply Bilateral, Gaussian, Median, and Wiener filters for denoising the images.
- c. To train a CNN model using denoised images for lung cancer subtype classification.
- d. To compare model performance across denoising methods using standard evaluation metrics.
- e. To identify the most suitable denoising approach for digital pathology workflows



## CHAPTER 2

### LITERATURE REVIEW

In recent years, artificial intelligence has gained significant attention due to its groundbreaking advancements in the field of intelligent medicine (Gu et al., 2021). The progress in AI and deep learning techniques has facilitated notable developments in computer-aided medical diagnosis. These deep learning methods have been extensively utilized for analyzing histopathological images of lung cancer (Davri et al., 2023). Considerable research has focused on developing optimal deep learning models capable of accurately detecting various types of lung cancer using these images. A discussion of recent scholarly contributions in the domain of machine learning is provided below.

R. Golan (2016), in the study “Lung nodule detection in CT images using deep convolutional neural networks,” proposed a framework that utilizes backpropagation to train the CNN weights for detecting lung nodules in CT image sub-volumes. The system demonstrated a sensitivity of 78.9% with 20 false positives per scan and 71.2% with 10 false positives per scan for lung nodules annotated by all four radiologists.

Šarić et al. (2019) developed a convolutional neural network (CNN) trained using the ACDC@LUNGHP dataset, which included annotated images. The dataset produced over 220,000 patches for training and 28,784 patches for testing. Two pretrained models, VGG16 and ResNet50, were evaluated. Image patches (256x256) were classified as "tumor" if 75% or more of the patch area contained tumor cells. The models were trained for 17 epochs with a learning rate of 0.0001, utilizing binary cross-entropy as the loss function. Although ResNet50 demonstrated higher general accuracy on ImageNet, VGG16 achieved slightly better AUC performance in this study. This variation underscores the differing feature requirements between natural image datasets like ImageNet and specialized datasets such as those used in digital pathology.

Hatuwal and Thapa (2020) developed a convolutional neural network (CNN) model. They utilized the LC25000 dataset, which includes histopathological images of three lung tissue types: benign tissue, adenocarcinoma, and squamous cell carcinoma. The proposed convolutional neural network (CNN) achieved a training accuracy of 96.11% and a validation accuracy of 97.20%, demonstrating robust performance in lung cancer classification tasks. The approach offers notable benefits, such as high accuracy, reduced diagnosis time compared to manual methods, and minimized human error through automation. Additionally, the model showed strong class-specific performance, with high precision, recall, and F1-scores, particularly excelling in the benign tissue category. However, the study has limitations, including its narrow focus on three categories, potential overfitting due to dataset variability despite augmentation techniques, dependency on consistent image quality for preprocessing, and significant computational demands requiring GPU resources for model

training. These factors highlight both the potential and constraints of CNN-based approaches for medical image analysis.

Suna W. et al. (2016) in “Computer aided lung cancer diagnosis with deep learning algorithms”, implemented three different deep learning algorithms, Convolutional Neural Network (CNN), Deep Belief Networks (DBNs), Stacked Denoising Auto Encoder (SDAE), and compared them with the traditional image feature based CAD system. The CNN architecture contains eight layers of convolutional and pooling layers, interchangeably. For the traditional algorithm, there were about 35 extracted texture and morphological features. These features were fed to the kernel based support vector machine (SVM) for training and classification. The resulting accuracy for the CNN approach reached 0.7976 which was little higher than the traditional SVM, with 0.7940. They used the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) public databases, with about 1018 lung cases.

The study by Muniasamy et al. (2024) employed the LC25000 dataset, which includes 25,000 histopathological images, of which 15,000 were lung cancer-related images divided equally into three categories: lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue (5,000 each). Using the EfficientNetB7 model with transfer learning, the study achieved a remarkable accuracy of 99.77% for classifying lung cancer histopathology images. The benefits of this study include high diagnostic accuracy, reduced workload for pathologists, and faster detection using advanced CNN models, which enhances early treatment options. Additionally, transfer learning effectively utilizes pretrained knowledge, optimizing resource usage. However, the study has limitations, including reliance on a specific dataset, which may restrict generalizability, and image resizing to 224x224 pixels, potentially losing critical histopathological details. This research highlights the potential of AI in improving lung cancer diagnostics and forms a strong basis for further exploration of generalizable and scalable models.

Khaw et al (2017) proposes a model combining Convolutional Neural Networks (CNN). This study utilized a dataset of 11,000 training and 1,650 testing images with standard benchmarks like Lena and Cameraman, introducing various noise types (Impulse, Gaussian, Speckle, and Poisson) at different levels and combinations. The proposed CNN model integrated with PCA achieved an average accuracy of 99.7% for single noise types, 99.3% for noise types and levels, and 86.3% for mixed noise classification, outperforming human recognition in complex scenarios. Benefits include high accuracy, reduced training time (10 minutes compared to 24 hours with traditional CNNs), and efficient feature extraction via PCA filters. However, limitations involve challenges in recognizing mixed noise types, dependency on appropriate filter sizes, and decreased performance on images with repetitive patterns. These findings demonstrate potential yet highlight areas for refinement in noise type recognition.

Ralla and Burra, 2024 developed the use of deep convolutional neural networks (CNNs) for automated lung cancer detection, utilizing a dataset of 220,025 lung tissue images, with a

split of 130,908 for training and 89,117 for testing. The proposed model, a Deep Belief Network (DBN), achieved impressive performance metrics, including an accuracy of 98.19%, precision of 97.87%, recall of 98.33%, and specificity of 97.83%. These results highlight the model's effectiveness in early cancer detection, which can significantly improve patient outcomes by enabling timely diagnosis. The approach also reduces the need for labor-intensive manual histological analysis, offering a faster and more cost-efficient solution. However, the model demands substantial computational resources and training time, making deployment in clinical settings challenging without advanced infrastructure. Despite these limitations, the study demonstrates the potential of machine learning in enhancing the accuracy and efficiency of medical diagnostics, laying a foundation for further research in this field.

Severeyn et al (2025) developed "AI-Driven Histopathological Analysis for Early Lung Cancer Detection" used the LC25000 dataset, which includes 25,000 histopathological images derived from 1,250 patient samples. Of these, 15,000 images were focused on lung cancer, equally divided into benign tumors, adenocarcinoma, and squamous cell carcinoma. The best-performing model (Model 4) achieved over 97% accuracy across all cancer types, with a perfect 100% accuracy for benign tumors in the test set. This research highlights significant benefits, including high diagnostic accuracy, potential for early detection, automation of the diagnostic process, and a user-friendly interface for clinical integration. However, limitations include dataset homogeneity, potential loss of critical details during preprocessing, reliance on high computational resources, and limited generalizability to diverse patient populations. These findings underscore the promise of AI in enhancing lung cancer diagnostics while identifying areas for further improvement.

Kumar et al (2024) utilized histopathological images of lung tissue for classifying lung cancer (LC) subtypes using the ML3CNet CNN-based model. It achieved a high classification accuracy of 99.72% on the dataset, with a quantized version achieving 98.92% accuracy. The nonlocal mean (NLM) filter was employed for noise reduction, preserving image edges and enhancing quality. Benefits include high diagnostic accuracy, efficient noise reduction, reduced model size via quantization for faster processing, and applicability to other datasets, such as colon cancer images. However, drawbacks include dependency on preprocessing techniques like NLM filtering, the resource-intensive nature of model training, and limited generalizability due to the focus on specific LC subtypes. This study highlights the potential for automated LC diagnosis to aid clinicians in accurate and efficient decision-making.

Chuah et al (2017) developed "Detection of Gaussian Noise and Its Level using Deep Convolutional Neural Network" that proposed a CNN-based model to detect and classify Gaussian noise levels in images. The model was trained using 12,000 standard test images and tested on 3,000 images. The dataset consisted of common test images such as Barbara, Lena, and Mandrill, corrupted with various Gaussian noise levels. The CNN model achieved an overall accuracy of 74.7% in classifying 10 noise levels. This study highlights the application

of CNNs for noise detection and classification, demonstrating their ability to intelligently recognize and classify Gaussian noise levels in images. The model reduces the need for manual preprocessing, streamlining the denoising pipeline. However, the moderate accuracy suggests potential limitations in handling subtle noise variations, and the method is constrained to Gaussian noise without addressing other types. Despite this, the work underscores the growing role of deep learning in enhancing image preprocessing workflows and provides a foundation for developing more advanced denoising methods applicable in fields like medical imaging.

Laine et al (2021) developed "Imaging in Focus: An Introduction to Denoising Bioimages in the Era of Deep Learning" does not provide specific accuracy metrics or patient data since it is a review article discussing various denoising methods for bioimages. Instead, it evaluates deep learning-based techniques like CARE, Noise2Void, and 3D-RCAN, focusing on their ability to improve image clarity in fluorescence microscopy. Here Deep learning-based denoising methods are highlighted as effective tools for enhancing bioimages by removing noise while preserving critical details. These methods, including CARE and Noise2Void, outperform traditional denoising approaches in metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Benefits include improved segmentation accuracy, usability in resource-limited settings, and the potential to train models with minimal or noisy data. However, challenges include reliance on high-quality training datasets, risk of introducing artifacts, and computational intensity. These advancements in image quality improvement can directly aid diagnostic accuracy in medical imaging, including lung cancer histopathology.

Nazir et al.,2024 analysed 104 studies on deep learning-based denoising techniques for medical images, highlighting their significance in enhancing image quality and diagnostic accuracy. Convolutional Neural Networks (CNNs) were the most utilized approach (40%), followed by encoder-decoder models (18%), generative adversarial networks (GANs, 12%), transformers (13%), and other methods. Common noise types included Gaussian noise (35%), speckle noise (16%), and Poisson noise (14%), affecting imaging modalities like MRI, CT, PET, ultrasound, and X-rays. Deep learning models effectively adapt to diverse noise types, improving the clarity and diagnostic value of images while enabling better visualization of anatomical structures. However, challenges remain, including the lack of standardized models across modalities, high computational and data demands, and risks of artifact introduction during denoising. Emerging techniques like GANs and transformers show promise in addressing these issues, underscoring the ongoing advancements and potential for deep learning in medical image denoising.

The above discussed literature on lung cancer detection using Convolutional Neural Networks (CNNs) suggests that the accuracy of the models can vary greatly depending on the size of the dataset and the number of layers used. Some models achieved a high accuracy like Muniasamy et al. (2024) stands out as the best-performing one, achieving an accuracy of 99.77% for classifying lung cancer histopathology images using the EfficientNetB7 model with

transfer learning. This performance is the highest among the studies reviewed, surpassing others in terms of diagnostic accuracy.

While studies like Šarić et al. (2019) and Khaw et al. (2017) demonstrate high accuracy (e.g., 99.7% for Khaw), their focus is more generalized (e.g., noise classification) or limited in scope compared to Muniasamy's targeted lung cancer classification. Kumar et al. (2024) also achieved a high accuracy of 99.72%, but its dependency on specific preprocessing techniques and less generalized applicability reduces its practical utility compared to Muniasamy's approach.

## CHAPTER 3

# MATERIALS AND METHODOLOGY

### 3.1 MATERIALS

#### 3.1.1 Data Set

The dataset comprises 11,580 high resolution histopathological images obtained from an ICMR-funded project (Grant No: EM/SG/Dev. Res/120/0847-2023) involving onco-pathology consultants. These images, are categorized into three classes: Lung Adenocarcinoma (LUAD), Lung Squamous Cell Carcinoma (LSCC), and Non-Malignant (NM). These histopathological images were derived from 300 Whole Slide Images (WSIs) of lung cancer tissues, with Regions of Interest (ROIs) manually annotated by a medical expert. The WSIs were obtained and downloaded from The Cancer Imaging Archive (TCIA). Each image in the dataset was extracted at a resolution of 512×512 pixels. The images for each class were split into training, validation, and test sets in an 8:1:1 ratio. Below is a sample image from the dataset:

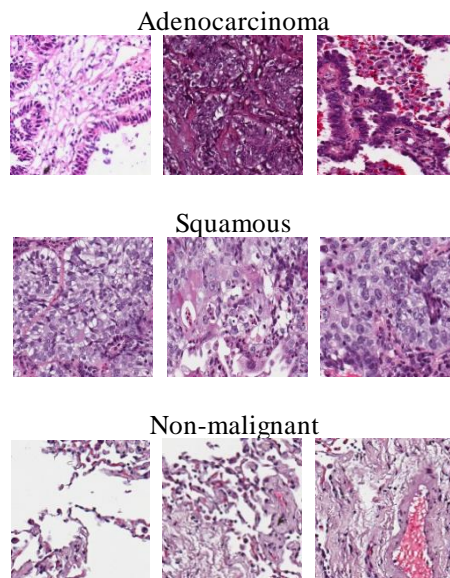
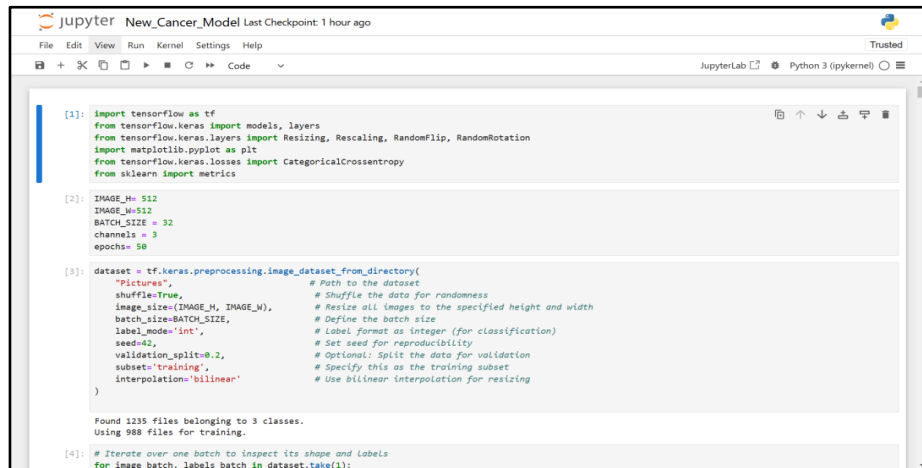


Fig. 3.1 Sample from the used dataset adenocarcinoma, squamous cell carcinoma and non-malignant.

### 3.1.2 Jupyter Notebook



```
[1]: import tensorflow as tf
from tensorflow.keras import models, layers
from tensorflow.keras.layers import Resizing, Rescaling, RandomFlip, RandomRotation
import matplotlib.pyplot as plt
from tensorflow.keras.losses import CategoricalCrossentropy
from sklearn import metrics

[2]: IMAGE_H= 512
IMAGE_W=512
BATCH_SIZE = 32
channels = 3
epochs= 50

[3]: dataset = tf.keras.preprocessing.image_dataset_from_directory(
    "Pictures", # Path to the dataset
    shuffle=True, # Shuffle the data for randomness
    image_size=(IMAGE_H, IMAGE_W), # Resize all images to the specified height and width
    batch_size=BATCH_SIZE, # Define the batch size
    label_mode='int', # Label format as integer (for classification)
    seed=42, # Set seed for reproducibility
    validation_split=0.2, # Optional: Split the data for validation
    subset='training', # Specify this as the training subset
    interpolation='bilinear' # Use bilinear interpolation for resizing
)

Found 1235 files belonging to 3 classes.
Using 988 files for training.

[4]: # Iterate over one batch to inspect its shape and labels
for image_batch, labels_batch in dataset.take(1):
```

Fig 3.2 Jupyter Notebook

Jupyter Notebook is a web-based interactive computing platform that allows users to create and share documents containing live code, equations, visuals, and text. It provides a powerful and user-friendly environment for teaching, scientific computing, and data analysis.

Widely used by data scientists, machine learning engineers, and researchers in academia and industry, Jupyter Notebook supports multiple programming languages such as Python, R, Julia, and more. It organizes work into notebooks, which are divided into cells that can hold code, markdown text, images, and other media types.

One of its key strengths is the ability to produce a comprehensive, executable record of a computational process, making it easy to share and reproduce results. Additionally, Jupyter Notebook supports collaboration by allowing multiple users to work on the same notebook simultaneously.

### 3.1.3 ImageScope

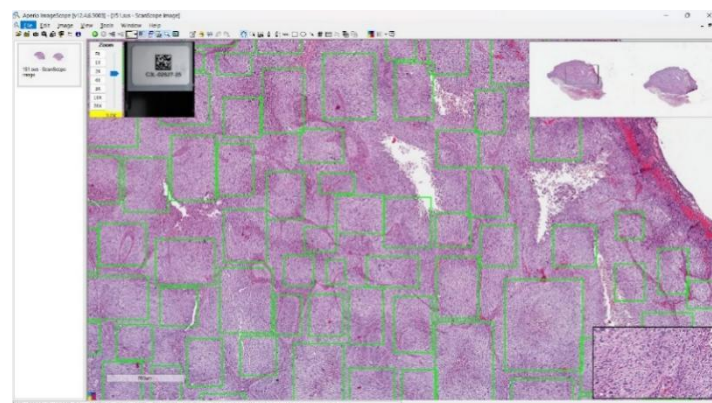


Fig 3.3 ImageScope

ImageScope is a powerful digital pathology viewer developed by Leica Biosystems, primarily used to view and analyse whole slide images (WSI). Designed to support researchers and pathologists, ImageScope allows users to zoom, pan, and annotate high-resolution pathology images with precision. It provides tools for measuring tissue structures, comparing different regions side-by-side, and running image analysis algorithms. With its user-friendly interface and support for multiple file formats like SVS, ImageScope is widely used in clinical and academic settings for diagnostic review, education, and research.

### 3.1.4 Anaconda Prompt

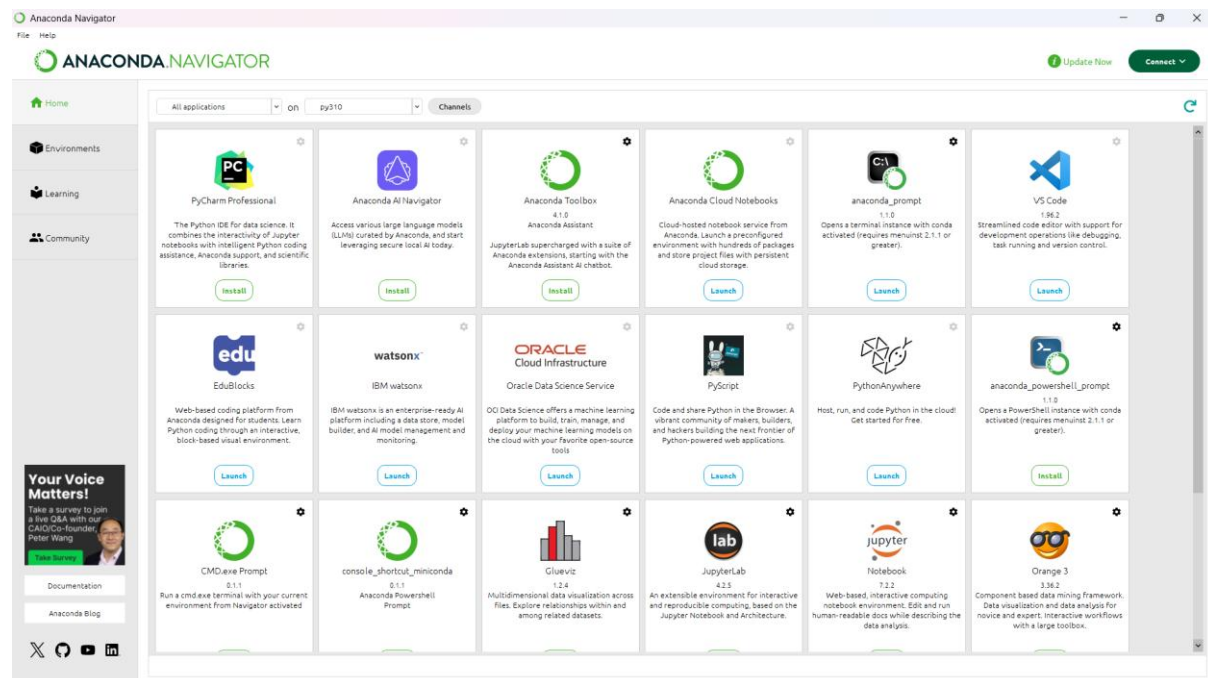


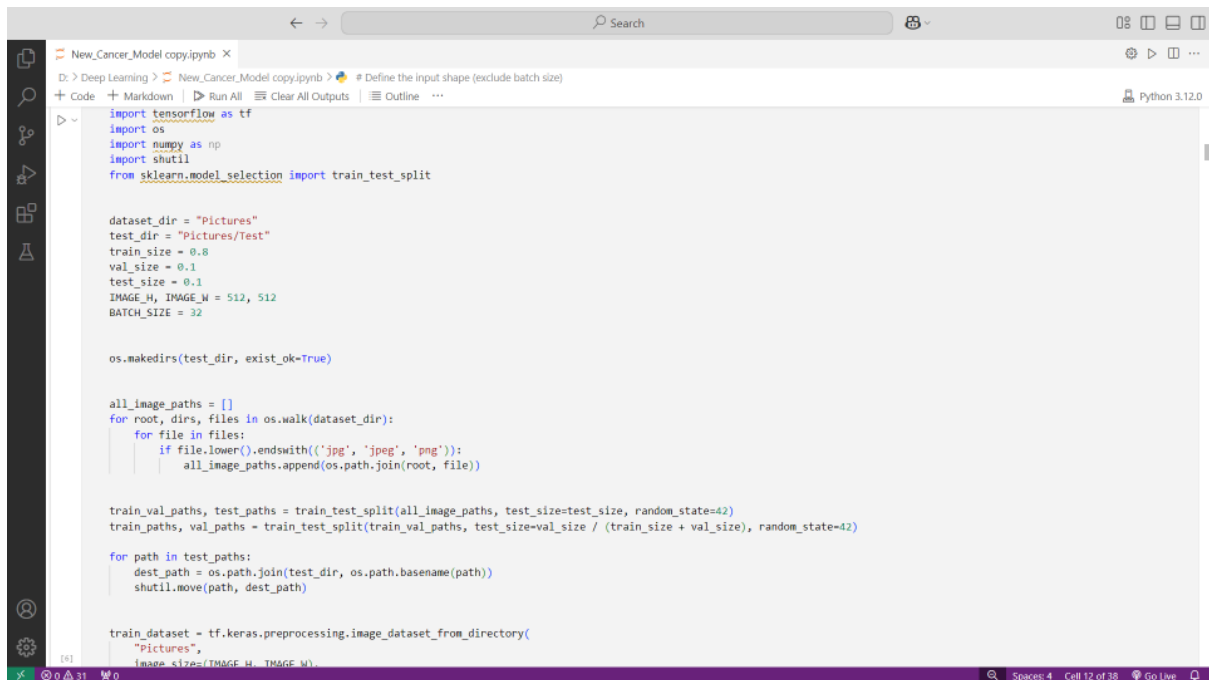
Fig 3.4 Anaconda Prompt

The Anaconda Prompt is a command-line interface (CLI) included with Anaconda, a widely used distribution of Python and R designed for scientific computing, data science, and machine learning. It provides an easy-to-use interface for managing packages, creating virtual environments, running Python scripts, and working with Jupyter notebooks.

The Anaconda Prompt ensures a consistent user experience across different operating systems, allowing users to access their Anaconda environment from any terminal or command prompt. It is an essential tool for data scientists and machine learning developers to effectively manage and control their development environments.



### 3.1.5 Python



```
# Define the input shape (exclude batch size)
import tensorflow as tf
import os
import numpy as np
import shutil
from sklearn.model_selection import train_test_split

dataset_dir = "Pictures"
test_dir = "Pictures/Test"
train_size = 0.8
val_size = 0.1
test_size = 0.1
IMAGE_H, IMAGE_W = 512, 512
BATCH_SIZE = 32

os.makedirs(test_dir, exist_ok=True)

all_image_paths = []
for root, dirs, files in os.walk(dataset_dir):
    for file in files:
        if file.lower().endswith(('.jpg', '.jpeg', '.png')):
            all_image_paths.append(os.path.join(root, file))

train_val_paths, test_paths = train_test_split(all_image_paths, test_size=test_size, random_state=42)
train_paths, val_paths = train_test_split(train_val_paths, test_size=val_size / (train_size + val_size), random_state=42)

for path in test_paths:
    dest_path = os.path.join(test_dir, os.path.basename(path))
    shutil.move(path, dest_path)

train_dataset = tf.keras.preprocessing.image_dataset_from_directory(
    "Pictures",
    image_size=(IMAGE_H, IMAGE_W),
```

Fig 3.5 Python

Python is a widely-used, high-level programming language known for its versatility and popularity. It is commonly employed in web development, machine learning, and various advanced technologies within the software industry. Compared to many other programming languages, Python is considered one of the easiest to learn and highly practical for a broad range of applications.

## 3.2 METHODOLOGY

### 3.2.1 Image Preprocessing

In this study, Whole Slide Images (WSIs) provided by the hospital under an ICMR-funded project were pre-processed by extracting smaller, high-resolution patches to enable efficient analysis and model training. WSIs, which are typically very large, contain lung adenocarcinoma and lung squamous cell carcinoma. To manage this, the images were divided into smaller, standardized patches of fixed dimensions (e.g., 512x512 pixels). Resizing these patches to uniform dimensions simplifies the processing workflow and accelerates the training process. The patch extraction specifically targeted regions of interest (ROIs) identified by expert pathologists, ensuring that each patch captured critical histopathological patterns essential for accurate classification.

Normalization refers to adjusting the pixel values of an image to a specific range, usually between 0 and 1 (Krizhevsky et al., 2012). This is done by dividing each pixel value by the maximum possible value, such as 255 for an 8-bit image (LeCun et al., 1998). This process

standardizes the input data, ensuring that all pixels contribute equally to the learning process, thereby enhancing the performance and convergence speed of machine learning models (Goodfellow et al., 2016).

In deep learning, normalization is a critical step as it minimizes the impact of large variations in input values that could disrupt the training of neural networks (Ioffe et al., 2015). By bringing pixel values into a uniform range, it also helps prevent numerical instability and improves the model's robustness (Santurkar et al., 2018). This preprocessing is particularly vital in fields like medical imaging, where precise and efficient model training is essential for tasks such as tumour detection and disease classification (Litjens et al., 2017).

These strategies, including horizontal flipping and brightness adjustments, significantly expand the size and variability of training datasets, which is essential for improving CNN performance in image classification tasks (Iqbal, 2024). Using the ImageDataGenerator module in TensorFlow/Keras enabled real-time data augmentation while preserving the structural integrity of medical images through transformations like horizontal flips and brightness variations (Rosebrock, 2019).

### **3.2.2 Type of noise in a histopathological image**

Images can contain noise due to various factors, often related to the methods used for capturing or processing them. Identifying the type of noise present is crucial for effective image preprocessing and analysis (Zhang, 2015). Gaussian noise, characterized by random variations in pixel intensity, is commonly caused by sensor imperfections during image capture or scanning (Islam et al., 2018).

Among the common types of noise that affect images are Gaussian noise, salt-and-pepper noise, and Poisson noise. In this study, we primarily focus on Gaussian noise, which can be effectively reduced using techniques like Gaussian blur or other filtering methods. Additionally, advanced machine learning approaches, such as convolutional neural networks (CNNs), can be trained to detect and classify different types of noise, enhancing image quality for further analysis or model training (Hami & JameBozorg, 2024).

### **3.2.3 Measurement of Level of Noise in WSI Image (in DB)**

Signal-to-Noise Ratio (SNR) is a key metric that measures the ratio of meaningful information to unwanted noise in an image, making it an essential indicator of image quality (Peksen, 2007).

It is calculated as:

$$SNR = 10 \log_{10} \left( \frac{S}{N} \right)$$

where  $S$  : Power of the meaningful signal in the image or data.

$N$  : Power of the unwanted noise in the image or data (Proakis & Salehi, 2008).

A high SNR indicates that the image has more signal (clear information) compared to noise, resulting in better image quality, while a low SNR means that noise overwhelms the signal, diminishing the clarity and usefulness of the image. In image processing, a higher SNR is associated with better denoising performance, as the goal is to strengthen the signal and reduce noise, thus improving the effectiveness of denoising techniques (Gnanasambandam & Chan, 2021).

The Structural Similarity Index Measure (SSIM) is a perceptual metric used to evaluate the similarity between two images by assessing structural content, luminance, and contrast to gauge image quality (Wang et al., 2004). SSIM measures the preservation of structural information after processing, with values ranging from 0 (no similarity) to 1 (perfect similarity), where higher values indicate better retention of image details and structure (Wang & Bovik, 2009).

The formula for SSIM is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Where  $\mu_x, \mu_y$  : Means of images  $x$  and  $y$

$\sigma_x^2, \sigma_y^2$  : Variances of image  $x$  and  $y$

$\sigma_{xy}$  : Covariance between  $x$  and  $y$

$C_1, C_2$  : Constants to stabilize the division , (Wang et al., 2004)

It measures image quality by evaluating how well the image's structure is maintained after processing (such as denoising or compression). Higher SSIM values suggest that the processed image is more similar to the original, effectively preserving its details and structure.

### 3.2.4 Remove noise from image with appropriate denoising method

Denoising is essential for improving image quality and making key details more visible and accessible for analysis (Zhang et al., 2017). In this study, we applied the Non-Local Means (NLM) algorithm for denoising, which reduces random noise while preserving the overall structure of the image (Buades et al., 2005). The Non-Local Means algorithm is an advanced technique that enhances image quality by effectively removing noise while maintaining important details and textures (Buades et al., 2005). It works by comparing image patches to assess similarity, assigning weights to pixels based on their proximity and similarity to other regions in the image (Buades et al., 2005). The NLM algorithm takes a global approach,

searching for similar patches even if they are far apart in the image (Buades et al., 2005). This method is further optimized to reduce computational time while maintaining the denoising effectiveness (Rousson & Paragios, 2006). In this study, the technique was applied to Whole Slide Images (WSIs) of lung carcinoma to enhance the quality of the input data prior to training the CNN model (Wang et al., 2020). This technique is particularly effective in removing high-frequency noise, such as Gaussian, salt-and-pepper, or speckle noise, and is especially useful for histopathological images, where clarity and structural integrity are crucial for diagnosis (Chawla & Tiwari, 2020).

### 3.3 Model training and development

The CNN model was developed using a convolutional neural network with 1,373 histopathological images obtained from an ICMR-funded project (Grant No: EM/SG/Dev. Res/120/0847-2023) involving onco-pathology consultants. These images, annotated by Oncopathologists affiliated with Apollo Kolkata under the project, are categorized into three classes: lung adenocarcinoma, lung squamous cell carcinoma, and non-malignant.

In this work, we designed an architecture for the model that includes 6 convolutional layers, 3 layers for max pooling, 2 dense layer, and 2 dropout layers. The model's configuration is represented as follows: [(2 Conv x 1 MaxPool x 1 Batch Normalization) \* 3] + 1 Flatten + 1 Fully Connected (FC) + 1 Drop Out + 1 Fully Connected (FC) + 1 Drop Out + 1 Output].

Table 3.1. Model Summary

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 150, 150, 64)	1792
batch_normalization (BatchNormalization)	(BatchNormalization (None, 150, 150, 64)	256
conv2d_1 (Conv2D)	(None, 150, 150, 64)	36928
batch_normalization_1	(None, 150, 150, 64)	256
max_pooling2d (MaxPooling2D)	(None, 75, 75, 64)	0
conv2d_2 (Conv2D)	(None, 75, 75, 128)	73856
batch_normalization_2 (BatchNormalization)	(None, 75, 75, 128)	512
conv2d_3 (Conv2D)	(None, 75, 75, 128)	147584
batch_normalization_3 (BatchNormalization)	(None, 75, 75, 128)	512
max_pooling2d_1 (MaxPooling 2D)	(None, 37, 37, 128)	0
conv2d_4 (Conv2D)	(None, 37, 37, 256)	295168
batch_normalization_4 (BatchNormalization)	(None, 37, 37, 256)	1024
conv2d_5 (Conv2D)	(None, 37, 37, 256)	590080

batch_normalization_5 (BatchNormalization)	(None, 37, 37, 256)	1024
max_pooling2d_2 (MaxPooling 2D)	(None, 18, 18, 256)	0
flatten (Flatten)	(None, 82944)	0
dense (Dense)	(None, 512)	42467840
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 3)	771

Total params: 43,748,931

Trainable params: 43,747,139

Non-trainable params: 1,792

A kernel size of 3x3 was used for the convolution process, which involved applying element-wise multiplication to the input images. Pooling layers were added subsequent to the convolutional layers to minimize the feature map dimensions and enhance memory efficiency. After several convolutional and pooling stages, we integrated three fully connected layers to reduce the dependence on manual intervention. These layers connect neurons across various levels using designated weights and biases. Dropout layer were added after each dense layer to mitigate overfitting. The flattened version of the input image is processed within these fully connected layers, where fundamental mathematical calculations are carried out to commence the classification process. In the end, the output layer utilized SoftMax activation function to identify which inputs should be activated or remain inactive during the forward propagation.

### 3.4 Proposed CNN model architecture

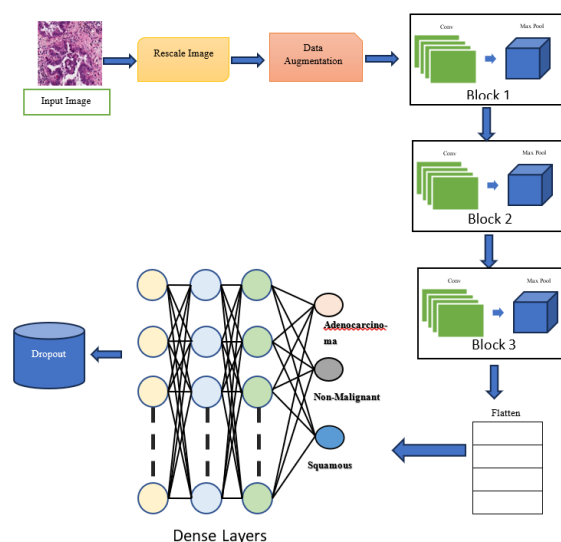


Fig 3.6 Diagram of the Model Architecture.

### 3.5 Evaluation and cross-validation of our proposed model

To thoroughly assess the model's effectiveness, multiple evaluation metrics were utilized. These metrics offered an overall idea of the model's performance. Additionally, overall accuracy was computed as a primary indicator of the model's effectiveness. The formulas for these metrics are outlined in equations (6) to (10).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (9)$$

$$\text{F1-Score} = \frac{2TP}{2TP+FN+FP} \quad (10)$$

Where, Tp =True Positives, TN = True Negatives, FP = False Positives, and FN =False Negatives

### 3.6 Flow Chart of The Methodology

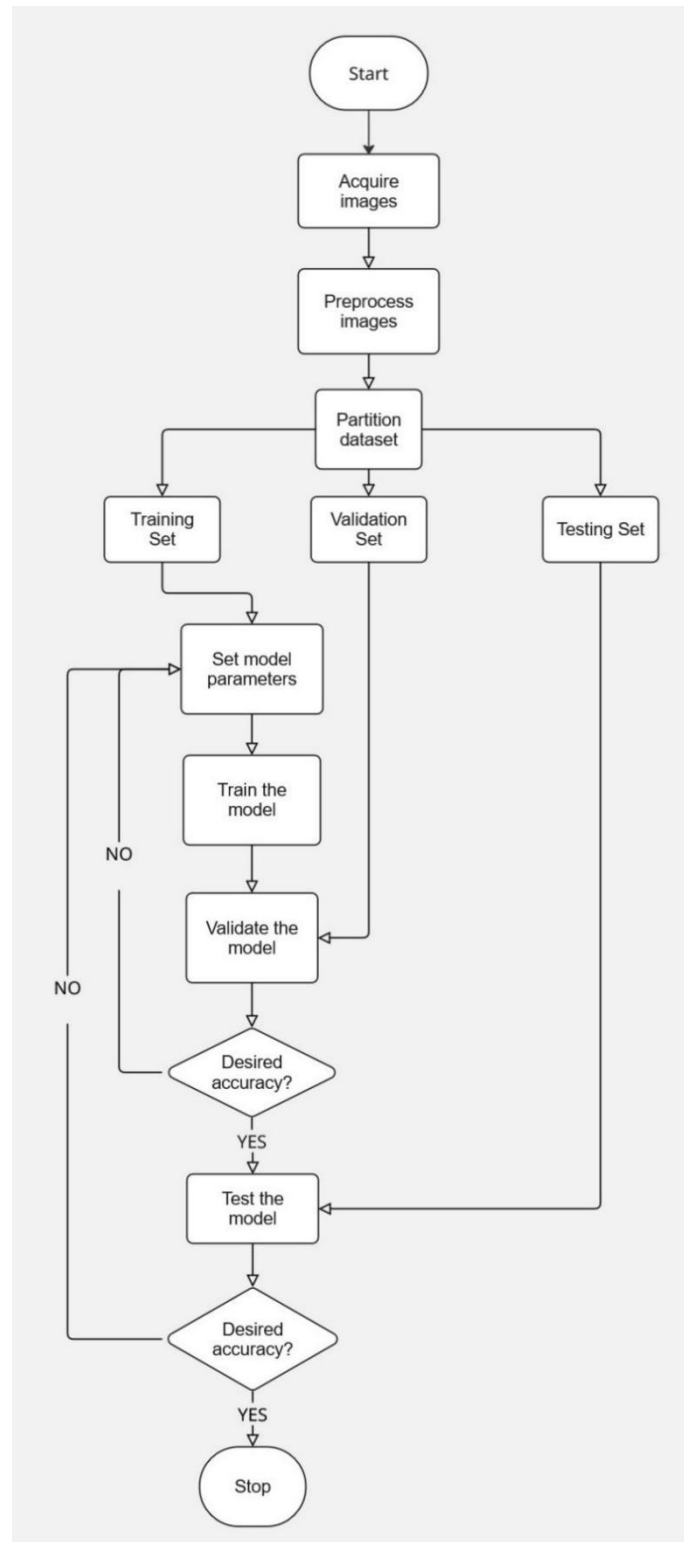


Fig 3.7: Diagram of the flowchart of methodology.

## CHAPTER 4

### RESULTS AND DECISION

#### 4.1 Model Performance

##### 4.1.1 Classification curves of the proposed models

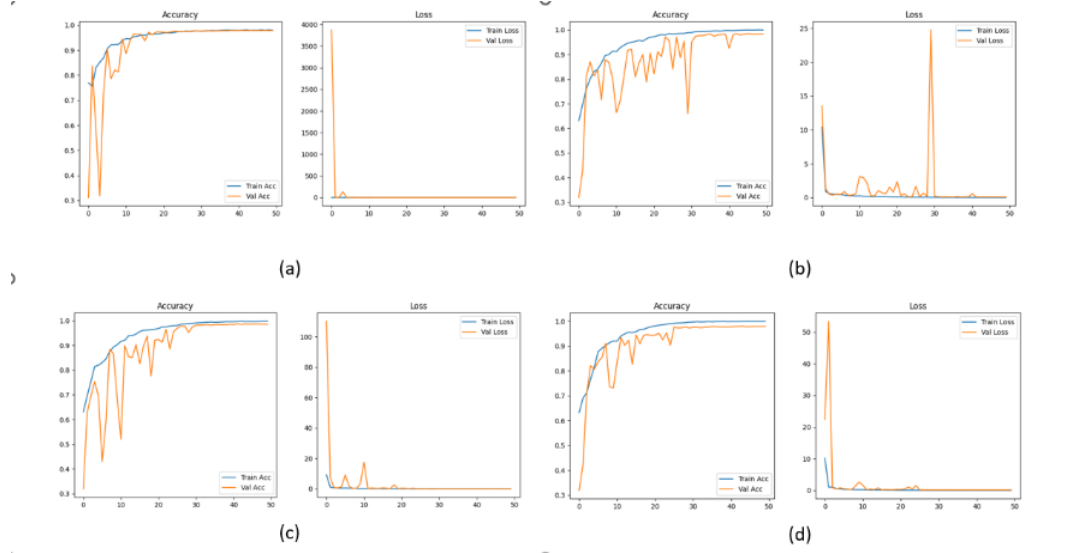


Fig. 4.1 Visualization of (a) Training and Validation accuracy and loss for the Bilateral model (b) Training and Validation accuracy and loss for the Gaussian model (c) Training and Validation accuracy and loss for the Median model (d) Training and Validation accuracy and loss for the Wiener model.

The training and validation curves for all four denoising techniques—Bilateral, Gaussian, Median, and Wiener—show that each method enhances model performance, though with varying stability. Bilateral filtering achieves fast convergence and high accuracy, indicating strong edge preservation. Gaussian filtering leads to good accuracy but with noticeable fluctuations in validation loss, likely due to over-smoothing. Median filtering ensures smooth and consistent training with minimal loss, effectively balancing noise removal and feature retention. Wiener filtering also performs well, showing stable and aligned accuracy and loss curves, thanks to its adaptive nature, which preserves essential image details for classification.



### 4.1.2 Confusion Matrix

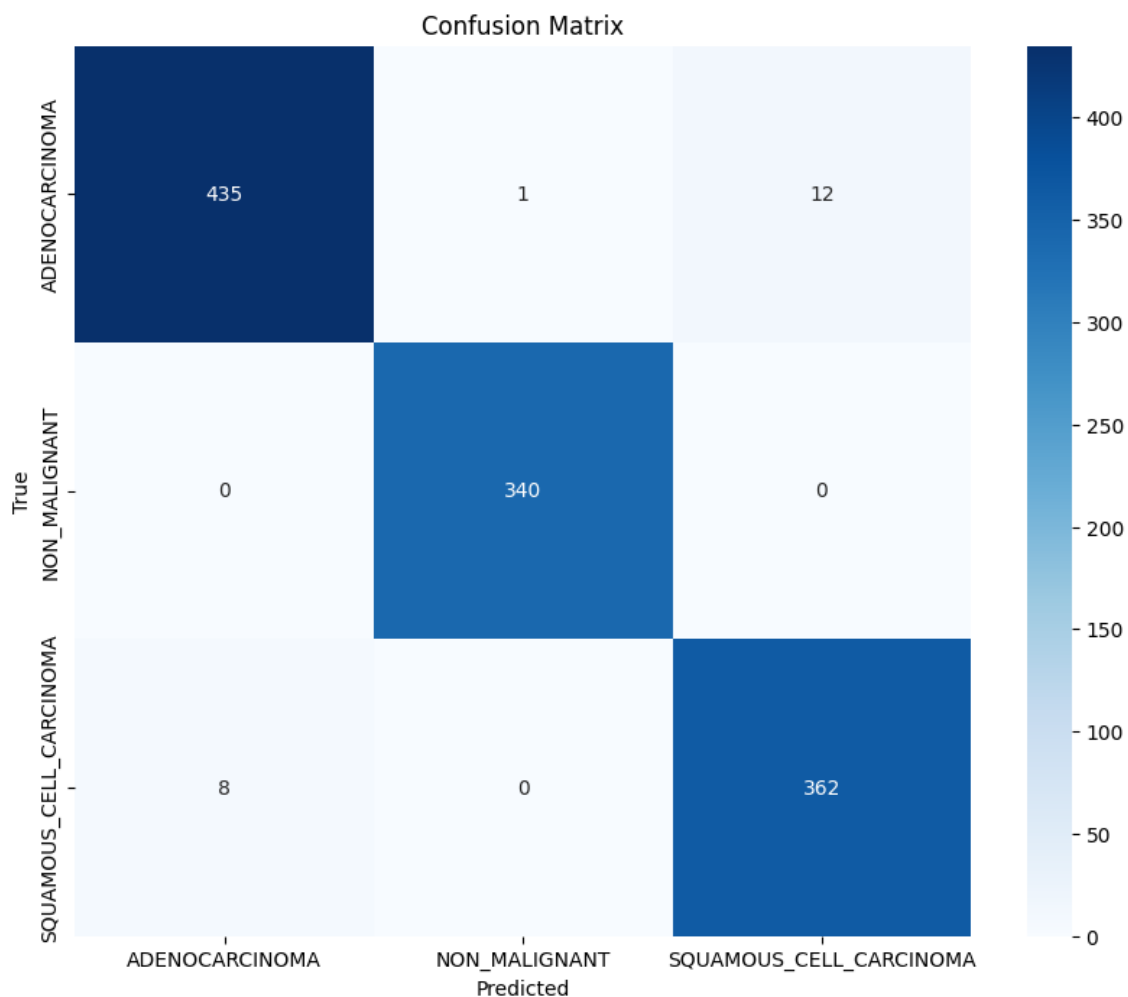


Fig 4.2 Confusion matrix of the proposed model before denoising.

The confusion matrix illustrates the performance of a multiclass classification model trained to identify lung carcinoma subtypes from histopathological images. The model demonstrates high accuracy across all three classes: Adenocarcinoma, Non-Malignant, and Squamous Cell Carcinoma. Out of 448 true Adenocarcinoma cases, the model correctly predicted 435, misclassifying 12 as Squamous Cell Carcinoma and 1 as Non-Malignant. For the 340 Non-Malignant cases, the model achieved perfect classification with no misclassifications. In the Squamous Cell Carcinoma class, out of 370 cases, the model correctly identified 362, with 8 instances misclassified as Adenocarcinoma. These results suggest the model is highly effective, especially in identifying Non-Malignant cases, while minor confusion occurs between the malignant subtypes, which is common due to overlapping histological features.

Table 4.1 Model performance on original datasets Before Denoising

Class	ACC	PRE	SEN	SPE	F1-SCORE
Adenocarcinoma	0.98	0.91	1.00	0.97	0.95
Squamous	0.98	1.00	0.97	1.00	0.99
Non-Malignant	1.00	1.00	1.00	1.00	1.00

The table presents the performance of a CNN model in classifying lung histopathology images into Adenocarcinoma, Squamous Cell Carcinoma, and Non-Malignant categories before applying any denoising techniques. The model shows strong performance across all classes, with an accuracy of 98% for both Adenocarcinoma and Squamous, and perfect accuracy (100%) for Non-Malignant cases. Adenocarcinoma achieved a precision of 0.91 and perfect recall (1.00), indicating some false positives but no missed cases. Squamous had perfect precision (1.00) and a slightly lower recall of 0.97, suggesting a few missed instances. The Non-Malignant class was predicted flawlessly, with perfect scores across all metrics. Overall, the model demonstrates high effectiveness in distinguishing between the three classes, especially in correctly identifying Non-Malignant cases.

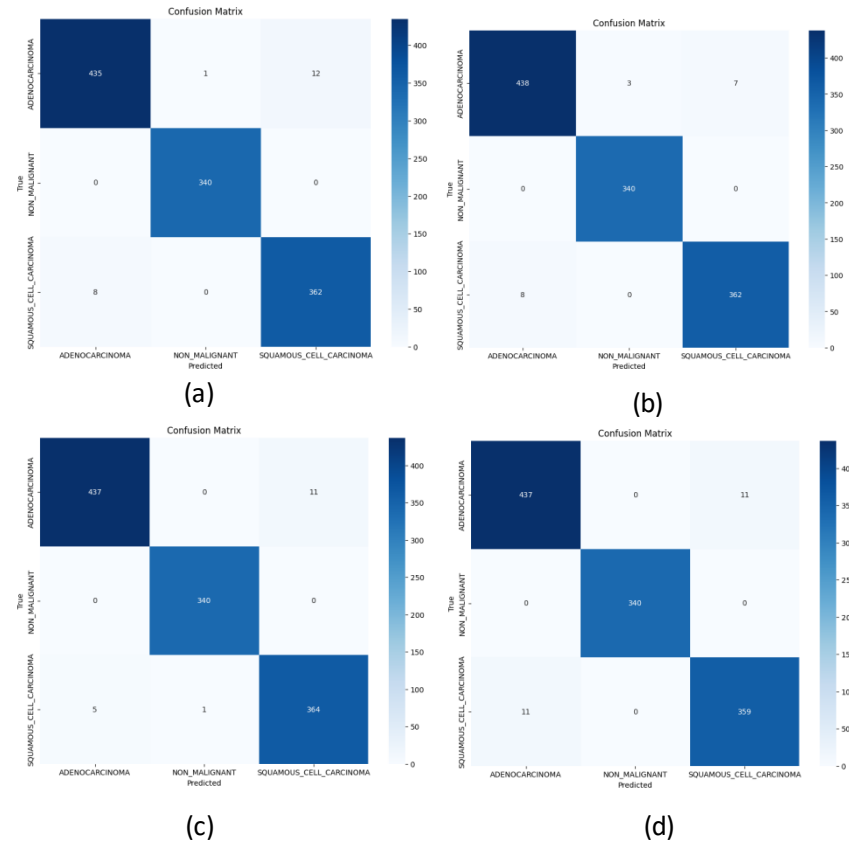


Fig 4.3 Confusion matrix of the proposed model after denoising.

Figure 4.3 displays the confusion matrices show that all four noise removal filters—Bilateral, Gaussian, Median, and Wiener—enable the CNN model to classify lung cancer types with high accuracy. Non-Malignant cases are consistently classified perfectly across all filters. The Median and Gaussian filters show slightly better performance in reducing misclassifications for Adenocarcinoma and Squamous Cell Carcinoma, suggesting they preserve critical features more effectively. Bilateral and Wiener filters also perform well but show slightly higher confusion between malignant classes. Overall, Median and Gaussian offer the best balance between noise reduction and feature retention.

Table 4.2 Proposed Model performance on Lung cancer dataset. Adam optimizer with a learning rate of 0.0001.

Model	Noise Removal	Class	Assessment Metrics				
			ACC	PRE	SEN	SPE	F1-Score
CNN model	Bilateral	Adenocarcinoma	0.98	0.98	0.98	0.99	0.98
		Non-Malignant	1.00	1.00	1.00	1.00	1.00
		Squamous Cell Carcinoma	0.98	0.97	0.97	0.98	0.97
	Gaussian	Adenocarcinoma	0.98	0.98	0.98	0.99	0.98
		Non-Malignant	1.00	0.99	1.00	1.00	1.00
		Squamous Cell Carcinoma	0.99	0.98	0.980	0.99	0.98
	Median	Adenocarcinoma	0.99	0.99	0.98	0.99	0.98
		Non-Malignant	1.00	1.00	1.00	1.00	1.00
		Squamous Cell Carcinoma	0.99	0.97	0.98	0.99	0.98
	Wiener	Adenocarcinoma	0.98	0.98	0.98	0.98	0.98
		Non-Malignant	1.00	1.00	1.00	1.00	1.00
		Squamous Cell Carcinoma	0.98	0.97	0.97	0.99	0.97

This table presents a comparison of CNN model performance using four different noise removal filters—Bilateral, Gaussian, Median, and Wiener—across three lung cancer classes: Adenocarcinoma, Non-Malignant, and Squamous Cell Carcinoma. Overall, all filters demonstrate high accuracy, precision, sensitivity, specificity, and F1-scores, indicating effective classification.

The Non-Malignant class consistently achieves perfect scores (1.00) across all filters, showing that the model reliably identifies this category regardless of preprocessing technique. Median filtering slightly outperforms others overall, offering the highest or tied-highest scores for Adenocarcinoma and Squamous Cell Carcinoma, with balanced sensitivity and precision. Gaussian filtering yields strong performance as well, though slightly less consistent for Squamous Cell Carcinoma. Wiener and Bilateral filters also perform well, with Bilateral showing marginally lower scores for Squamous Cell Carcinoma.

In summary, while all filters support strong classification, Median filtering provides the best balance of noise reduction and feature preservation, leading to the most reliable performance across all cancer classes.

### 4.1.3 Classification: Precision-Recall Curve

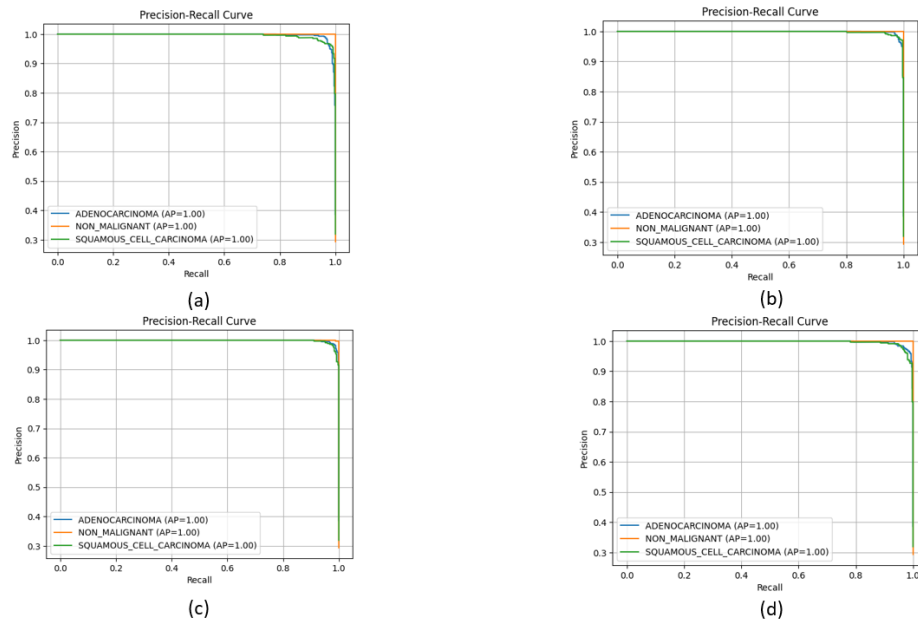


Fig 4.4 Precision-Recall curve for (a) bilateral model (b) Gaussian model (c) Median model (d) Wiener model.

The Precision-Recall (PR) curves for all four denoising methods—Bilateral, Gaussian, Median, and Wiener—demonstrate consistently high performance across all three classes: Adenocarcinoma, Non-Malignant, and Squamous Cell Carcinoma. Each model achieves an average precision (AP) of 1.00 for all classes, indicating excellent classification capability regardless of the filtering method applied.

The Bilateral and Wiener filtered images yield PR curves that are tightly clustered and nearly ideal, with high precision maintained even as recall approaches 1. This suggests these filters are especially effective in preserving critical class-specific features, leading to confident and accurate predictions. The Gaussian filter, while still achieving perfect AP, exhibits slightly more deviation near the high-recall end, implying occasional uncertainty in class separation likely due to over-smoothing. The Median filter offers a clean and stable PR curve, similar to the Wiener filter, reflecting its ability to maintain structural integrity while suppressing noise. Overall, all four filters enable robust classification, with Bilateral and Wiener showing particularly reliable and stable predictive performance.

## **CHAPTER 5**

### **CONCLUSION**

This study demonstrates the crucial role of image denoising in improving the classification accuracy of NSCLC using Whole Slide Images. By conducting a comparative analysis of four widely-used denoising techniques—Bilateral, Gaussian, Median, and Wiener—we found that Median filtering consistently outperformed others in balancing noise reduction and the preservation of essential histopathological details. Gaussian and Wiener filters also showed strong performance, particularly in enhancing classification precision and recall.

Our deep learning model, trained on denoised image patches, achieved high classification accuracy across all cancer classes, with perfect scores for the Non-Malignant category. These results confirm that appropriate preprocessing, especially noise removal, is a key factor in optimizing model performance for medical image analysis.

This project reinforces the value of combining expert-annotated datasets, classical image processing methods, and deep learning models for advancing automated lung cancer diagnosis. Future research can extend this work by integrating deep learning-based denoising techniques or exploring real-time WSI processing systems, ultimately contributing to faster and more accurate cancer detection in clinical settings.

## REFERENCES

1. Bandera, E. V., Fay, S. H., Giovannucci, E., et al. (2016). The use and interpretation of anthropometric measures in cancer epidemiology: A perspective from the World Cancer Research Fund International Continuous Update Project. *International Journal of Cancer*, 139(11), 2391–2397.
2. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263. <https://doi.org/10.3322/caac.21834>
3. Brinkman, J. E., Toro, F., & Sharma, S. (2023). Physiology, respiratory drive. In *StatPearls*. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK482414/>
4. Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 60–65). IEEE. <https://doi.org/10.1109/CVPR.2005.38>
5. Chatterjee, P., & Milanfar, P. (2010). Is denoising dead? *IEEE Transactions on Image Processing*, 19(4), 895–911. <https://doi.org/10.1109/TIP.2009.2037087>
6. Chawla, R., & Tiwari, R. (2020). Image denoising in histopathology using non-local means: A review. *Journal of Pathology Informatics*, 11, 31. [https://doi.org/10.4103/jpi.jpi\\_45\\_20](https://doi.org/10.4103/jpi.jpi_45_20)
7. Cheng, E. S., Weber, M., Steinberg, J., & Yu, X. Q. (2021). Lung cancer risk in never-smokers: An overview of environmental and genetic factors. *Chinese Journal of Cancer Research*, 33(5), 548–562. <https://doi.org/10.21147/j.issn.1000-9604.2021.05.02>
8. Chuah, J. H., Khaw, H. Y., Soon, F. C., & Chow, C. O. (2017). Detection of Gaussian noise and its level using deep convolutional neural network. In M. F. A. Fauzi (Ed.), *Proceedings of the 2017 IEEE Region 10 Conference (TENCON)* (pp. 2447–2450). IEEE. <https://doi.org/10.1109/TENCON.2017.8228254>
9. Ciofiac, C. M., Mămuleanu, M., Florescu, L. M., & Gheonea, I. A. (2024). CT imaging patterns in major histological types of lung cancer. *Life*, 14(4), 462. <https://doi.org/10.3390/life14040462>
10. Cui, J. W., Li, W., Han, F. J., & Liu, Y. D. (2015). Screening for lung cancer using low-dose computed tomography: Concerns about the application in low-risk individuals. *Translational Lung Cancer Research*, 4(3), 275–286. <https://doi.org/10.3978/j.issn.2218-6751.2015.02.05>
11. Davri, A., Birbas, E., Kanavos, T., Ntritsos, G., Giannakeas, N., Tzallas, A. T., & Batistatou, A. (2023). Deep learning for lung cancer diagnosis, prognosis and prediction using histological and cytological images: A systematic review. *Cancers*, 15(15), 3981. <https://doi.org/10.3390/cancers15153981>

12. Dela Cruz, C. S., Tanoue, L. T., & Matthay, R. A. (2011). Lung cancer: Epidemiology, etiology, and prevention. *Clinics in Chest Medicine*, 32(4), 605–644. <https://doi.org/10.1016/j.ccm.2011.09.001>
13. Doyle, J., & Cooper, J. S. (2023, July 4). *Physiology, carbon dioxide transport*. In StatPearls. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK532988/>
14. Gandhi, Z., Gurram, P., Amgai, B., Lekkala, S. P., Lokhandwala, A., Manne, S., Mohammed, A., Koshiya, H., Dewaswala, N., Desai, R., Bhopalwala, H., Ganti, S., & Surani, S. (2023). Artificial intelligence and lung cancer: Impact on improving patient outcomes. *Cancers*, 15(21), 5236. <https://doi.org/10.3390/cancers15215236>
15. Gnanasambandam, A., & Chan, S. H. (2021). Exposure-referred signal-to-noise ratio for digital image sensors. *arXiv preprint arXiv:2112.05817*. <https://doi.org/10.48550/arXiv.2112.05817>
16. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
17. Grewal, R. K., Rosenzweig, K. E., & Ginsberg, M. S. (2010). Thorax. In R. T. Hoppe, T. L. Phillips, & M. Roach (Eds.), *Leibel and Phillips Textbook of Radiation Oncology* (3rd ed., pp. 362–377). W. B. Saunders. <https://doi.org/10.1016/B978-1-4160-5897-7.00019-6>
18. Griffeth, L. K. (2005). Use of PET/CT scanning in cancer patients: Technical and practical considerations. *Proceedings (Baylor University Medical Center)*, 18(4), 321–330. <https://doi.org/10.1080/08998280.2005.11928089>
19. Gu, X., Shen, Z., Xue, J., Fan, Y., & Ni, T. (2021). Brain tumor MR image classification using convolutional dictionary learning with local constraint. *Frontiers in Neuroscience*, 15, 679847. <https://doi.org/10.3389/fnins.2021.679847>
20. Hami, M., & JameBozorg, M. (2024). Assessing the impact of CNN autoencoder-based image denoising on image classification tasks. *arXiv preprint arXiv:2404.10664*. <https://doi.org/10.48550/arXiv.2404.10664>
21. Hatuwal, B. K., & Thapa, H. C. (2020). Lung cancer detection using convolutional neural network on histopathological images. *International Journal of Computer Trends and Technology*, 68(10), 21–24. <https://doi.org/10.14445/22312803/IJCTT-V68I10P104>
22. Herbst, R. S., Morgensztern, D., & Boshoff, C. (2018). The biology and management of non-small cell lung cancer. *Nature*, 553(7689), 446–454. <https://doi.org/10.1038/nature25183>
23. Huang, L., Li, Y., Yang, Y., & Zhang, J. (2021). A deep learning approach for noise reduction in whole slide imaging for lung cancer detection. *Scientific Reports*, 11, 90444. <https://doi.org/10.1038/s41598-021-89552-4>
24. IEEE. (2017). *IEEE Region 10 Annual International Conference, Proceedings/TENCON* (Vol. 2017-December). <https://doi.org/10.1109/TENCON.2017.8228272>
25. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 448–456). <https://proceedings.mlr.press/v37/ioffe15.html>



26. Iqbal, P. A. (2024). Easily image data preparation with ImageDataGenerator in TensorFlow. *Medium*.
27. Islam, M. T., Saha, D., Rahman, S. M. M., Ahmad, M. O., & Swamy, M. N. S. (2018). A variational step for reduction of mixed Gaussian-impulse noise from images. *arXiv preprint arXiv:1811.00244*. <https://doi.org/10.48550/arXiv.1811.00244>
28. Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. *Cancers*, 15(14), 3608. <https://doi.org/10.3390/cancers15143608>
29. Kalemkerian, G. P. (2012). Staging and imaging of small cell lung cancer. *Cancer Imaging*, 11(1), 253–258. <https://doi.org/10.1102/1470-7330.2011.0036>
30. Kanmani, S., & Rajivkannan, R. (2018). Performance analysis of noise filters using tissue images. *Semantic Scholar*.
31. Khan, Y. S., & Lynch, D. T. (2023). Histology, lung. In *StatPearls*. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK534789/>
32. Khaw, H. Y., Soon, F. C., Chuah, J. H., & Chow, C. (2017). Image noise types recognition using convolutional neural network with principal components analysis. *IET Image Processing*, 11(12), 1238–1245. <https://doi.org/10.1049/iet-ipr.2017.0374>
33. Kothari, S., Sharma, R., & Goyal, M. (2013). A study of image processing techniques for the analysis of histopathological images. *Computer Methods and Programs in Biomedicine*, 111(2), 213–221. <https://doi.org/10.1016/j.compmedimag.2013.01.003>
34. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 25, 1097–1105. [https://papers.nips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://papers.nips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)
35. Kumar, A., Vishwakarma, A., & Bajaj, V. (2024). ML3CNet: Non-local means-assisted automatic framework for lung cancer subtypes classification using histopathological images. *Computer Methods and Programs in Biomedicine*, 251, 108207. <https://doi.org/10.1016/j.cmpb.2024.108207>
36. Kumar, T., Mileo, A., Brennan, R., & Bendeche, M. (2023). Image data augmentation approaches: A comprehensive survey and future directions. *arXiv preprint arXiv:2301.02830*. <https://doi.org/10.48550/arXiv.2301.02830>
37. Kürkcü, R. A., Baylan, İ. Y. Y., Attar, E., Selçuk, B., & Şerif, T. (2024). AI-enhanced endometrial cancer diagnosis system. In *2024 9th International Conference on Computer Science and Engineering (UBMK)* (pp. 1–5). Antalya, Turkey. <https://doi.org/10.1109/UBMK63289.2024.10773504>
38. Laine, R. F., Jacquemet, G., & Krull, A. (2021). Imaging in focus: An introduction to denoising bioimages in the era of deep learning. *The International Journal of Biochemistry & Cell Biology*, 140, 106077. <https://doi.org/10.1016/j.biocel.2021.106077>

39. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
40. Liam, C. K., Liam, Y. S., Poh, M. E., & Wong, C. K. (2020). Accuracy of lung cancer staging in the multidisciplinary team setting. *Translational Lung Cancer Research*, 9(4), 1654–1666. <https://doi.org/10.21037/tlcr.2019.11.28>
41. Lindeman, N. I., Cagle, P. T., Beasley, M. B., Chitale, D. A., Dacic, S., Giaccone, G., Jenkins, R. B., Kwiatkowski, D. J., Saldivar, J. S., Squire, J., Thunnissen, E., & Ladanyi, M. (2013). Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *Journal of Thoracic Oncology*, 8(7), 823–859. <https://doi.org/10.1097/JTO.0b013e318290868f>
42. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., & Ghafoorian, M. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
43. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. A. W. M. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
44. Liu, B., Zhou, H., Tan, L., et al. (2024). Exploring treatment options in cancer: Tumor treatment strategies. *Signal Transduction and Targeted Therapy*, 9, 175. <https://doi.org/10.1038/s41392-024-01856-7>
45. Lopes, G. (2023). The global economic cost of cancer—Estimating it is just the first step! *JAMA Oncology*. <https://doi.org/10.1001/jamaoncol.2022.7133>
46. Luengo-Fernandez, R., Leal, J., Gray, A., & Sullivan, R. (2013). Economic burden of cancer across the European Union: A population-based cost analysis. *The Lancet Oncology*, 14(12), 1165–1174.
47. Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (2008). Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. *Mayo Clinic Proceedings*, 83(5), 584–594. <https://doi.org/10.4065/83.5.584>
48. Muniasamy, A., Alquhtani, S. A. S., Bilfaqih, S. M., Balaji, P., & Karunakaran, G. (2024). Lung cancer histopathology image classification using transfer learning with convolution neural network model. *Technology and Health Care*, 32(2), 1199–1210. <https://doi.org/10.3233/THC-231029>
49. Nazir, N., Sarwar, A., & Saini, B. S. (2024). Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges. *Micron*, 180, 103615. <https://doi.org/10.1016/j.micron.2024.103615>
50. de Onis, M., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., & Siekmann, J. (2007). Development of a WHO growth reference for school-aged children and adolescents.

- Bulletin of the World Health Organization*, 85(9), 660–667.  
<https://doi.org/10.2471/BLT.07.043497>
51. Peksen, Y. (2007). *Performance analysis of signal-to-noise ratio (SNR) estimates in AWGN and time-selective fading channels* (Master's thesis, Texas A&M University).
  52. Proakis, J. G., & Salehi, M. (2008). *Digital communications* (5th ed.). McGraw-Hill.
  53. Rajdev, K., Siddiqui, A. H., Ibrahim, U., Patibandla, P., Khan, T., & El-Sayegh, D. (2018). An unusually aggressive large cell carcinoma of the lung: Undiagnosed until autopsy. *Cureus*, 10(2), e2202. <https://doi.org/10.7759/cureus.2202>
  54. Ralla, S., & Burra, A. (2024). Enhanced lung cancer detection using deep convolutional neural networks. *International Journal of Research and Analytical Reviews*, 11(1), 947–952.
  55. Rosebrock, A. (2019). *Keras ImageDataGenerator and data augmentation*. PyImageSearch. <https://pyimagesearch.com>
  56. Rousson, M., & Paragios, N. (2006). Non-local denoising. *Journal of Mathematical Imaging and Vision*, 25(3), 319–335.
  57. Rubin G.. D.. (2015). Lung nodule and cancer detection in computed tomography screening. *Journal of thoracic imaging*, 30(2), 130–138.  
<https://doi.org/10.1097/RTI.0000000000000140>
  58. Sabbula, B. R., Gasalberti, D. P., Mukkamalla, S. K. R., & Anjum, F. (2024). Squamous cell lung cancer. In *StatPearls*. StatPearls Publishing.  
<https://www.ncbi.nlm.nih.gov/books/NBK441933/>
  59. Samet, J. M., Avila-Tang, E., Boffetta, P., Hannan, L. M., Olivo-Marston, S., Thun, M. J., & Rudin, C. M. (2009). Lung cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clinical Cancer Research*, 15(18), 5626–5645.  
<https://doi.org/10.1158/1078-0432.CCR-09-0376>
  60. Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How does batch normalization help optimization? In *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 1–10.
  61. Šarić, M., Russo, M., Stella, M., & Sikora, M. (2019). CNN-based method for lung cancer detection in whole slide histopathology images. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)* (pp. 1–4). IEEE.  
<https://doi.org/10.23919/SpliTech.2019.8783041>
  62. Sathishkumar, K., Chaturvedi, M., Das, P., Stephen, S., & Mathur, P. (2022). Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme, India. *The Indian Journal of Medical Research*, 156(4–5), 598–607. [https://doi.org/10.4103/ijmr.IJMR\\_1821\\_22](https://doi.org/10.4103/ijmr.IJMR_1821_22)
  63. Severeyn, E., Rangel, C., La Cruz, A., & Velásquez, J. (2025). AI-driven histopathological analysis for early lung cancer detection. *Journal Name, Volume(Issue)*, Pages.
  64. Siegel, R. L., Miller, K. D., & Fuchs, H. E. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17–48. <https://doi.org/10.3322/caac.21763>

65. Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67, 101813. <https://doi.org/10.1016/j.media.2020.101813>
66. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
67. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. Advance online publication. <https://doi.org/10.3322/caac.21660>
68. Tárnoki, Á. D., Tárnoki, D. L., Dąbrowska, M., Knetki-Wróblewska, M., Frille, A., Stubbs, H., Blyth, K. G., & Juul, A. D. (2024). New developments in the imaging of lung cancer. *Breathe (Sheffield, England)*, 20(1), 230176. <https://doi.org/10.1183/20734735.0176-2023>
69. Thanigaiarasu, N., Kailash, A., Verma, G., Revendran, J., & Kishan, R. (2024). A case series on lung adenocarcinoma with varying clinical presentation. *Indian Journal of Applied Research*. <https://doi.org/10.36106/ijar/5004804>
70. Travis, W. D., Brambilla, E., Nicholson, A. G., & Yatabe, Y. (2015). The 2015 World Health Organization classification of lung tumors. *Journal of Thoracic Oncology*, 10(9), 1243–1260. <https://doi.org/10.1097/JTO.0000000000000630>
71. Wang, Q., Gümüş, Z. H., colarossi, c., memeo, l., wang, x., kong, c. Y., & Boffetta, P. (2023). SCLC: Epidemiology, risk factors, genetic susceptibility, molecular pathology, screening, and early detection. *Journal of Thoracic Oncology*, 18(1), 31–46. <https://doi.org/10.1016/j.jtho.2022.10.002>
72. Wang, X., Zhang, Y., & Xu, C. (2020). Non-local means denoising for whole slide image preprocessing in digital pathology. *Journal of Digital Imaging*, 33(6), 1493–1504.
73. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
74. Warren, G. W., & Cummings, K. M. (2013). Tobacco and lung cancer: Risks, trends, and outcomes in patients with cancer. *American Society of Clinical Oncology Educational Book*, (33), 359–364. [https://doi.org/10.14694/EdBook\\_AM.2013.33.359](https://doi.org/10.14694/EdBook_AM.2013.33.359)
75. World Health Organization. (2017). Obesity and overweight fact sheet. Retrieved from <http://www.who.int/mediacentre/factsheets/fs311/en/>
76. Willett, W., & Hu, F. (2013). Anthropometric measures and body composition. In *Nutritional Epidemiology* (3rd ed., Vol. 15, pp. 257–270). Oxford University Press.
77. World Health Organization. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*, 363, 157–163.

78. Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights into Imaging*, 9(4), 611–629. <https://doi.org/10.1007/s13244-018-0639-9>
79. Yang, Y., Zhang, T., & Liu, Z. (2022). Patch label denoising for improved classification in histopathological whole slide images. *PMC*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8791954/>
80. Zarella, M. D., Bui, A. A. T., & Feldman, M. D. (2019). Whole slide imaging in diagnostic pathology: A review. *Archives of Pathology & Laboratory Medicine*, 143(4), 421–431. <https://doi.org/10.5858/arpa.2018-0147-RA>
81. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7), 3142–3155. <https://doi.org/10.1109/TIP.2017.2662206>
82. Zhang, L., Zhang, L., & Wu, X. (2017). A review of image denoising algorithms, with a new one. *Journal of Imaging Science and Technology*, 61(3), 030501-1–030501-12.
83. Zhang, Z. (2015). Image noise: Detection, measurement, and removal techniques. *University of Tennessee*.