

An Interpretable Baseline Framework for Binary Classification of Lung CT Images using Logistic Regression

Abdirahman Mohamed Hassan¹ , Ni Haibin (倪海彬)²  Abubakar Abdinur Hersi³  Idris Aweis Hussein³ 

School of Electronics and Communication Engineering, “Faculty of Information Communication Engineering”, Nanjing University of Information Science & Technology, Nanjing, 210044, China;
School of Computer Science, “Faculty of Computer Science”, Nanjing University of Information Science & Technology, Nanjing, 210044, China.

Abstract: - Early diagnosis is essential for improving survival rates in lung cancer, which is one of the leading causes of death worldwide. Computed Tomography (CT) scans are among the most commonly used methods for diagnosing lung cancer; however, their interpretation by radiologists is time-consuming and subject to inter-observer variability. Machine learning (ML) methods have been developed to automate the identification of lung cancer. This study presents a baseline ML framework for classifying lung CT images using logistic regression to distinguish between cancerous and non-cancerous cases. The objective of this study was to evaluate the performance of simple, interpretable models under limited data conditions. Due to the limited sample size and the lack of patient-level identifiers, the reported performance should be interpreted as dataset-specific rather than clinically generalizable.

The original IQ-OTH/NCCD dataset consisted of three categories: Normal, Benign, and Malignant scans. In this study, the dataset was reformulated into a binary classification problem: non-cancerous scans (normal and benign) and cancerous scans (malignant). All CT images were preprocessed by resizing them to a uniform size, converting to grayscale, and normalizing pixel values before extracting features by flattening pixel intensities into one-dimensional vectors. The model was evaluated using multiple metrics: accuracy, precision, recall, F1-score, ROC-AUC, and PR-AUC. The model was compared with several classical ML models and a Convolutional Neural Network (CNN).

Experimental results demonstrated high classification performance; however, further analysis indicates that these results may be influenced by dataset characteristics including limited sample size, potential data leakage, and simplified class structure. The findings also reveal that although simple models can provide high performance under controlled experimental conditions, such performance may not generalize to real-world clinical scenarios. This study contributes by producing a reproducible baseline framework, analyzing the limitations of high-performing classifiers on small datasets, and emphasizing the importance of robust validation strategies in developing classifiers that can be used clinically. The findings contribute to the ongoing discussion on the use of interpretable machine learning models as an initial step toward developing models using larger datasets and more rigorous validation practices.

Keywords: - lung cancer, machine learning, logistic regression, CT image classification, image preprocessing, binary classification.

I. Introduction

In Lung cancer, being one of the prime causes of death from all cancers around the world, is a significant public health issue [1], [12]. Early diagnosis is essential to increasing the chances of survival; by allowing for faster treatment, it provides an opportunity for successful treatment. Medical imaging techniques, particularly CT scans, are frequently employed to reveal abnormalities in the lungs and help detect possible cancers within them [2], [34]. However, manual interpretation of CT images is time-consuming and may be affected by human error, especially when radiologists must review large volumes of scans. With advances in machine learning, automated systems have been created that can provide assistance to medical image analysts, assisting in the analysis of medical images and aiding in the diagnosis, thereby improving the accuracy of diagnosis as well as decreasing the workload on radiologists while also providing consistent results [13], [34]. Machine learning techniques may be used to discover patterns in medical images that may not be clearly seen by the human eye. This study is a development of a machine-learning-based system for detecting lung cancer in CT images. This study proposes a binary classification technique that will classify images into two groups: those with lung cancer (cancerous) and those without lung cancer (non-cancerous). Logistic regression is selected as the primary classification algorithm due to its simplicity, interpretability, and effectiveness in handling binary classification problems, particularly when working with limited datasets [7], [9].

II. Problem statement

Globally, lung cancer continues to be one of the most significant causes of cancer-related death for millions of people [1], [12], mainly because early detection at a more treatable stage remains

challenging. Early detection of lung cancer directly correlates with improved outcomes; however, it continues to be a significant clinical challenge. Computed tomography (CT) imaging is widely used to detect suspicious lung nodules and other abnormalities and provides highly detailed cross-sectional images that support clinical decision-making [2],[34]. Although CT imaging is effective in diagnosis and determining the extent of the disease, the manual interpretation of CT scans is a complex process requiring significant attention to detail. Radiologists must examine a large number of images per scan, and factors such as experience, fatigue, and variability in case presentation can contribute to differences in interpretation of the same CT images. These challenges highlight the potential for inconsistent diagnoses and the need for computerized assistive systems. Computer-aided diagnosis (CAD) systems have shown great promise in their ability to assist clinicians in interpreting medical imaging more accurately and efficiently [13].

III. Related Work

Recent years have witnessed significant growth in the ways that we use advanced computer technology (such as sophisticated data science, machine learning, and deep learning) to assist doctors with diagnosing a great variety of diseases in patients [13],[31],[33]. The new advanced technologies allow for many more possibilities to improve the accuracy with which we are able to diagnose various kinds of diseases [34],[35]. This is especially true for diseases that are very difficult to diagnose, such as lung cancer [1],[12]. Lung cancer is a significant cause of death across the world [1],[12], and because of the often-delayed signs pointing to lung cancer, there are many people who are diagnosed too late for any type of successful treatment. For this reason, physicians and

scientists are continually working on improving automated systems for earlier and more accurately diagnosing lung cancer [13],[18]. The purpose of this chapter is to provide a literature review and an explanation of some of the ways in which computer-based technology and medical imaging can be used to identify lung cancer [2],[34].

This literature review serves to reduce the gap between what has been reported in the literature regarding the use of machine learning techniques and what is currently being done in the practice of medicine today [35]. The literature review will include existing literature and provide examples of where improvements need to be made. This knowledge will be necessary to create a computer-aided diagnosis system that is both functional and user-friendly, particularly to the students and the researcher who are new in the field.

1. Machine learning in medical images

Machine learning (ML) refers to a subset of artificial intelligence (AI) that allows systems to identify patterns and relationships from available data and then output predictions without explicitly programming them to do so [34],[5]. The use of ML technology has been a major factor in improving diagnostic support through medical imaging. This is particularly true for lung cancer detection where ML applications train models to classify CT scans into positive and negative cancer categories.

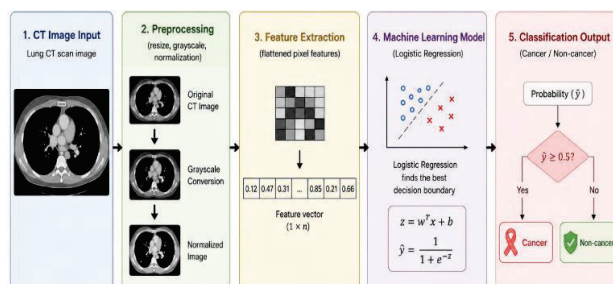


Figure 1: Traditional machine learning pipeline with manual feature extraction and classification

Most traditional types of machine learning applications rely on using features (characteristics) derived from their respective input (CT scans) through feature extraction prior to classifying the input into predefined categories.

For instance, an image classification application that uses ML techniques would often convert an image into a feature vector comprising pixel intensities or other descriptive features before conducting the classification process. These features are then used for classification.

2. Deep learning in medical images

Deep learning (DL) is a subcategory of machine learning that implements multi-layer neural networks to automatically build hierarchical feature representations from raw data [31],[33].

Traditional machine learning methods require the manual extraction of features from data prior to applying the model; however, deep learning models can directly process raw data without manual feature extraction, learning complex patterns within the data through multiple layers of abstraction.

convolutional neural network CNNs are a specific type of deep learning architecture that offers strong performance in image analysis tasks [3]. CNNs consist of a series of layers and use convolutional filters within the network to identify features by extracting spatial relationships from the images (edges, textures, shapes, etc.). CNNs learn spatial hierarchies, making them highly effective for analyzing medical images (e.g., CT scans) for lung diagnosis. Deep learning models have shown excellent performance when applied to the medical imaging, and consequently have demonstrated high accuracy for classification and detection tasks [6],[21],[23].

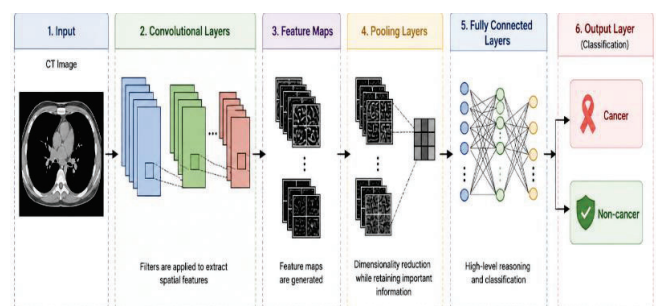


Figure 2: Deep learning pipeline with automatic feature extraction and classification

3. Introduction to Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a deep learning neural network that provides methods for use in 'image and video' classification problems [31],[33].

CNN networks are developed for analyzing, utilizing a grid-like data structure, such as the pixels that comprise an image. Input data are analyzed through multiple convolution layers in order to define (extract) the features from the input data [31]. The CNN is comprised of two (2) basic structures: the convolution layer and the pooling layer. A convolution layer uses a mathematical operation called "convolution" to apply a filter on the input data. The result is a feature map, where each filter creates a figure from the input data (i.e., edges and textures). Pooling layers are typically inserted between convolution layers in order to reduce the size of the feature map. This increases the robustness of the model to small variations of the input data and reduces the number of parameters in the model [31].

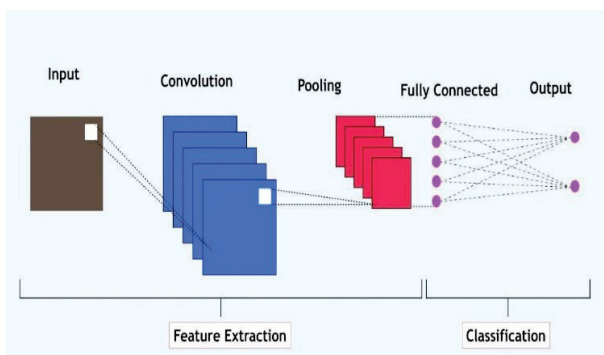


Figure.3: Basic architecture of a Convolutional Neural Network (CNN)

The basic structure of a Convolutional Neural Network (CNN) has been illustrated in Figure 2.3. Its purpose is to classify images. The image will be the input into the Convolutional layer, which involves using a set of filters to find key characteristics of the image (e.g. edges, textures, patterns). after the Convolution is complete and the feature maps extracted from the image are reduced through the use of Pooling operations; these Pooling operations allow the CNN to keep only the most

significant features while reducing the amount of computing needed to process each feature [31]. Once the features have been pooled, they are sent to the fully connected layers, which classify the feature maps for the output of the CNN [31]. The output is the Final Class of the image. As shown in the figure above, CNNs provide a single process for doing feature extraction and classification, which makes them very effective for image-based applications such as Medical Imaging Analysis [31].

4. Medical Image Analysis in Lung Cancer Detection

The evaluation of medical images is important for the diagnosis and early detection of lung cancer [1],[12]. Computed tomography (CT) scans are the most commonly used imaging modality because it provides detailed cross-sectional images of the lungs, allowing for the detection of nodules and abnormal patterns in tissue [2],[34]. High-quality analysis of CT images is critical for the early diagnosis of diseases, as timely diagnosis can significantly improve patient outcomes [12].

In recent years, machine learning and deep learning approaches have begun to be used for the analysis of medical images in order to increase accuracy and efficiency when diagnosing diseases [13],[31],[33]. Machine learning enables automated detection and classification of lung abnormalities while assisting radiologists in interpretation [13],[34]. Automated systems help clinicians identify patterns that are difficult to see, especially for early diagnosis of lung cancer [5],[18].

5. Traditional vs. Modern Methods of Machine Learning

Within the domain of image classification (including medical images), there are two primary approaches for developing image classifiers: traditional machine learning methodology and deep

learning methodology [31],[33]. Traditional machine learning classification relies on feature extraction to identify meaningful patterns with classification potential in the raw data used as input for a classifier [18],[29]. Common algorithms include logistic regression, support vector machines (SVM), and random forest [18],[29]. Traditional machine-learning techniques are often more straightforward, less resource-intensive, and easier to interpret than their modern counterparts [7],[9]. Whereas traditional methods rely on manually engineered features, contemporary methods of machine learning utilize neural networks and learn features automatically (i.e., deep learning) [31],[33]. CNNs extend traditional neural networks by using convolution operations to learn spatial relationships inherent in an image (Convolutional Neural Networks), enabling them to identify complex patterns and hierarchical structures in images. Because of this, CNNs are widely used tools in medical imaging applications, particularly for identifying lung cancer [21],[23]. However, due to the requirement of large numbers of labeled training images (examples) and the high computational resource requirements, their adoption in clinical practice may be limited [6],[7]. The following table summarizes key studies on machine learning and deep learning approaches for lung cancer diagnosis.

Table 1: Literature Review Summary of Related Studies on Deep Learning and Machine Learning Approaches in Lung Cancer Diagnosis

| Ref. No. | Authors | Year | Methodology | Challenges / Limitations | Application Area | Accuracy (%) |
|----------|------------------|------|---|---|------------------------------|--------------|
| [5] | Javed et al. | 2024 | Deep learning models for lung cancer detection | Requires large datasets and high computational cost | Lung cancer detection | ~90–95% |
| [6] | Thanoon et al. | 2023 | Review of deep learning techniques for CT image analysis | Limited generalization and strong dependency on dataset quality | Lung cancer classification | ~88–94% |
| [13] | Li et al. | 2022 | Machine learning methods for diagnosis and prognosis prediction | Model interpretability limitations | Lung cancer analysis | ~85–92% |
| [14] | He et al. | 2016 | Deep Residual Networks (ResNet) for deep image classification | Requires large datasets and high computational resources | Medical image classification | ~94–97% |
| [21] | Bouamrane et al. | 2024 | CNN-based lung cancer diagnosis framework | Sensitive to image quality and limited dataset size | Lung CT diagnosis | ~93–96% |
| [25] | Litjens et al. | 2017 | Survey of deep learning methods in medical image analysis | Limited interpretability and dependence on annotated datasets | Medical image analysis | N/A |

In Table 1:, selected studies employing deep learning and machine learning techniques for lung cancer diagnosis are summarized. Convolutional Neural Networks (CNNs), transformer-based models, and other more advanced methods tend to demonstrate good classification accuracy (> 90%) across these various papers [5, 6, 14, 21, 25]. However, these methods commonly require large annotated data sets and exhibit a relatively high level of computational complexity, thus making their use impractical in many situations where resources are constrained [5, 14, 25].

This study will attempt to address these limitations by developing an interpretable baseline using logistic regression, while also providing a

systematic evaluation of the reliability and limitations associated with achieving high-performance results through the use of small datasets.

6. Classification of Lung CT Images - Challenges

Despite recent developments in the field of machine learning and deep learning for classification of lung Computed Tomography (CT) images have shown great potential; many challenges persist. For example, the limited availability of annotated medical datasets will lead to reduced (generalizability) ability of a model and/or increased risk of overfitting. Other factors that may impact the ability to produce accurate CT image classifications include variability in CT image quality, resolution, and variation in patient anatomy. In addition, deep learning models typically require very large datasets and vast computational resources and their, typically, low interpretability may inhibit trust clinicians have in these models when making treatment decisions. Finally, most published studies do not conduct external validation or patient-level splitting. This omission may lead to overly optimistic estimates of performance and thereby increase the risk of incorrectly classifying a patient's (CT) scan as being "normal" or "abnormal." These limitations highlight the need for the development of interpretable baseline models and application of robust evaluation methodologies.

IV. Methodology

a. Introduction

the methodology used to develop the model is described to classify images of lung cancer based on CT scan data. Our process involves creating a structured pipeline containing dataset preparation, image preprocessing (processing of all images before applying machine learning techniques), feature extraction (the identification of key features that will aid classification), model selection, training, and evaluation. The

methodology's aim is to produce a baseline, interpretable and reproducible model, to classify lung CT scan images as cancerous or non-cancerous. We use logistic regression as our primary machine learning classification algorithm, and we evaluate its performance against various standard classification metrics and by visualizing the information produced through using these tools and methods.

b. Dataset Description

This study utilized the [IQ-OTH/NCCD](#) lung CT scan dataset and consists of three categories of lung images: normal, benign, and malignant.

The dataset was accessed and downloaded in May 2026. The original dataset contains a larger number of CT images across three classes (normal, benign, and malignant). The exact total number of images is not explicitly specified in the available dataset description. In this study, a subset of 300 images was selected for experimental purposes.

To simplify the classification task, the dataset was reformulated into a binary classification problem.

1. Non-cancerous CT scans: including normal and benign cases

2. Cancerous CT scans: containing malignant cases only

A total of 300 CT images were selected for this study based on the following criteria:

- ensuring sufficient image quality and clarity for preprocessing;
- maintaining a balanced representation of the classes;
- excluding corrupted or unreadable images.

Table 2: Distribution of Cancerous and Non-Cancerous CT Images

| Cancerous | Non-cancerous |
|--|---|
| 120 CT scans with a malignant (cancerous) lung conditions | 180 CT scans with normal or benign (non-cancerous) lung conditions |

The dataset consists of 300 CT images, including 120 malignant (cancerous) and 180 non-cancerous (normal

and benign) cases. This results in a moderate class imbalance toward the non-cancerous class. No explicit duplicate or near-duplicate image removal process was performed. Additionally, due to the absence of patient-level identifiers, it is possible that multiple slices from the same patient are included in the dataset. Although this imbalance is not severe, it may still influence model learning and evaluation. Therefore, evaluation metrics such as precision, recall, and F1-score were used to provide a more reliable assessment of model performance.

c. Image Preprocessing

Preprocessing of the images collected is crucial to preparing CT image data for machine learning. When collecting the CT images, there are differences in image size, pixel intensity, clarity, and quality, and these inconsistencies can significantly reduce the performance of a model unless they have had some form of preprocessing done to standardize them before being input into a machine learning classification model.

We applied the following preprocessing techniques:

- **Resize:**

Each of the images in the dataset has been resized to a consistent fixed dimensional size (64×64 pixels) so that each of the images will have a consistent input size once in the model.

- **Convert RGB to Grayscale:**

All images are converted from full-color with three channels (RGB color images) to single channels (grayscale) to reduce computational complexity while preserving important structural elements present in each image.

Pixel values were normalized by scaling the pixel intensity values to the range $[0,1]$ using min-max normalization (dividing by 255), ensuring consistent numerical input for the classification models.

- **Normalizing Pixel Values:**

Normalizing the pixel intensity to a range of $[0$ to $1]$ allows for better numerical stability and should allow for faster convergence during training.

Collectively, these preprocessing techniques achieve consistency and uniformity across the dataset, which contributes to more accurate results by the classification model.

d. Feature Extraction

To create numerical data for machine learning models, we need to turn image data into feature vectors. After pre-processing, every 64×64 grayscale image is flattened to become a one-dimensional feature vector resulting in a 4096-dimensional feature vector. Flattening converts the two-dimensional Structure of the image into a one-dimensional Array of pixel intensity values—this pixel representation will be the input features for our classifier models.

This method of flattening the image removes any spatial relationship that may exist between pixels and, therefore, is not the ideal way to represent each pixel in the Image. However, because flat pixel representations can be processed quickly and are a simple way to represent pixel data, we chose to represent images in this manner given the small size of our dataset. Future work should explore more advanced feature extraction techniques such as texture-based features (e.g., Gray-Level Co-occurrence Matrix), Local Binary Patterns, or radiomic features to capture structural information and improve classification performance.

e. Handcrafted Feature Extraction

The aim of this study is to analyze three handcrafted feature extraction methods: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gray-level Co-occurrence Matrix (GLCM). These feature extraction methods were evaluated to determine whether texture- and structure-based representations provide better classification performance than raw flattened pixel features. HOG is a feature extraction technique that captures edge orientation and gradient

information in an image. HOG is commonly used for representing structural patterns in images. LBP is a feature extraction technique that characterizes local texture patterns by comparing each pixel with its neighboring pixels and encoding the results into binary patterns. GLCM is a feature extraction technique that analyzes the spatial relationships between pixel intensity values. GLCM provides texture-related measurements, such as contrast, correlation, energy, and homogeneity. All feature extraction processes were applied to all images after the same preprocessing process (conversion to grayscale, resizing to 64×64 pixels, normalization, feature extraction). Logistic Regression was then applied to the feature representation under the identical experimental conditions to ensure a fair comparison. The experiment is not intended to create an HFE framework but rather to investigate whether more descriptive handcrafted features provide improved discriminative capability compared with flattened pixel features.

f. Model Selection

Logistic Regression was chosen as our primary classification Model due to its relative simplicity, ease of interpretability and overall accuracy when applied to Binary Classification tasks. Logistic Regression is regularly used in the medical and academic research communities as a baseline model.

In order to enhance our experimental analysis, we also considered using the following additional classifier models for Comparison Purposes:

1. Support Vector Machine (SVM)
2. Random Forest (RF)
3. k-Nearest Neighbor (k-NN)

In addition, Principal Component Analysis (PCA) was used as a technique for Dimensionality Reduction to see how this affects Classification Performance.

For comparative analysis, the additional classifiers were implemented using standard configurations. The Support Vector Machine (SVM) used a radial basis function

(RBF) kernel, the Random Forest (RF) model used 100 decision trees, and the k-Nearest Neighbors (k-NN) model used $k = 5$ with Euclidean distance. These settings were selected to provide a fair baseline comparison without extensive hyperparameter tuning.

This study intentionally adopts a simple and interpretable approach using logistic regression as a baseline model. The objective is not to achieve state-of-the-art performance, but to establish a transparent, reproducible benchmark for lung CT image classification under limited data conditions. This baseline can serve as a reference for future studies employing more complex models.

g. Model Training

The dataset was randomly split into training (240 samples) and test (60 samples) subsets using an 80:20 stratified split. The proportions of the two classes were maintained in both the training and test subsets.

Table 3: Dataset Splitting

| Set | Total | Cancer | Non-Cancer |
|----------|-------|--------|------------|
| Training | 240 | 96 | 144 |
| Testing | 60 | 24 | 36 |

The dataset was divided at the image level into an 80:20 ratio of training to test images using stratified splits. These splits do not take patient-level identifiers into account. Therefore, images from the same patient may appear in both training and test datasets, creating potential data leakage. This can lead to over-inflation of the model's performance because the model may have learned patient-based features, rather than more general or common features. Therefore, the model's final evaluation metrics should be seen as very specific to the dataset and should not be interpreted as generally applicable in a clinical setting. Future experiments should split at the patient level or use grouped cross-validation methods in order to provide a more accurate evaluation of model performance.

The logistic regression model was built with Scikit-learn using the following parameters:

- Solver = LBFGS
- Penalty = L2
- Regularization strength C = 1.0
- Maximum iterations = 1000
- Class weight = None
- Random seed = 42
- Classification threshold = 0.5

All models were implemented using Scikit-learn with standard configurations. Logistic regression was configured with the LBFGS solver and a maximum of 1000 iterations. Other classifiers, including Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbors (k-NN), were applied using default parameters unless otherwise specified. Feature scaling was applied where necessary to ensure fair comparison between models.

This model was trained on the training dataset by learning the relationship between the input feature vectors and class labels. After training, the trained model was validated against the test dataset to determine its generalizability.

The remaining models (SVM, Random Forest, and k-NN) were built and validated on the same training/testing split to allow for a fair comparison between the performance of the different models. In addition to creating train-test split datasets, 5-fold stratified cross-validation was utilized to calculate model performance. The original dataset was subdivided into five groups while still retaining the distribution of classes. Each group was trained and validated once and then used as the validation dataset during the remaining four training runs to generate the final evaluations. The final evaluation results were reported as the mean and standard deviation of the various performance metrics.

h. Dimensional Reduction (Using PCA)

Before classification, PCA was applied to the dataset for dimensional reduction (to reduce the dimensionality of the feature space).

PCA was set up to retain 95% of the variance within the features of the dataset. The benefit of PCA is that it reduces redundancy in the feature space while keeping the most relevant information (for each sample).

The performance of the models was evaluated on the test dataset both pre-and post-dimension reduction, to determine whether PCA adds any value. Principal Component Analysis (PCA) was fitted only on the training dataset and subsequently applied to the test dataset to prevent data leakage.

i. Proposed Methodology

The proposed approach to lung cancer classification uses a structured machine learning pipeline to classify CT images into two categories: "cancerous" or "non-cancerous." The diagram illustrating how the proposed system works is shown in Figure 3.1. Finally, the classification performance of the model is evaluated using conventional metrics (e.g., accuracy, precision, recall and F1-Score) and visualization methods (e.g. confusion matrix and/or ROC curves).

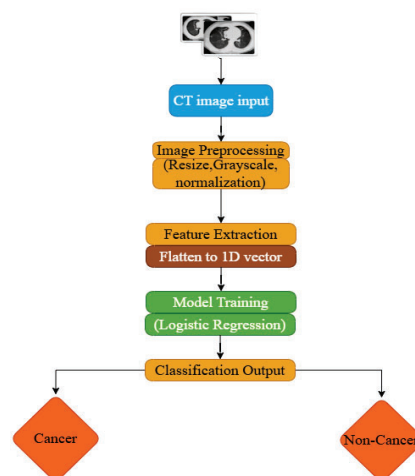


Figure 4: Proposed workflow of the lung cancer classification system

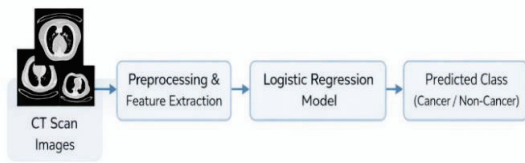


Figure 5: Logistic regression model pipeline for binary classification of CT images

In summary, the proposed methodology provides an effective and reproducible framework for CT image classification using traditional machine learning techniques.

j. Convolutional Neural Network (CNN) Model

The CNN model is included as an exploratory extension to compare traditional machine learning with deep learning approaches under the same dataset constraints. Due to the limited dataset size, the CNN is not expected to fully demonstrate its advantages over simpler models.

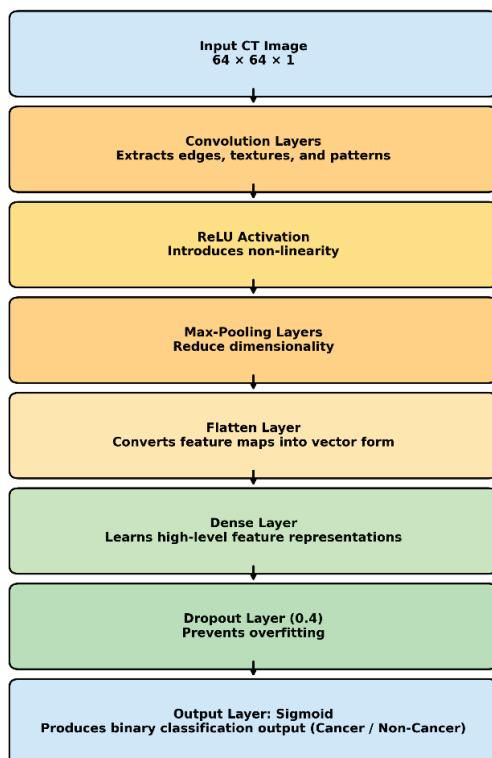


Figure 6: CNN architecture for lung CT image classification

• Architecture Description

In total, there are multiple layers of the CNN model. Each layer has a specific function in the extraction of spatial features from the images of the CT scanner to classifying the images into either cancerous or non-cancerous images. The most significant of these five components are as follows:

1. convolutional layers

The convolutional layers are used to extract the important features of the input images. In this case, the convolutional layers apply multiple filters (or kernels) to the input data to find the desired patterns within the input images. Depending on the number of filters applied to the input images will determine the number of different features detected in the process. As the depth of the CNN model increases, the more complicated features that will be extracted that allow for the identification of whether the CT image contains a cancerous or non-cancerous image within the dataset of images produced from images of CT scanners.

2. ReLU Activation Function

After each application of the convolutional operation of the CNN model, ReLU (Rectified Linear Unit) will be applied. ReLU helps to introduce a level of non-linearity to the CNN model, thus providing the CNN model the ability to learn and extract complex patterns within the input images.

ReLU is defined as follows:

$$f(x) = \max(0, x) \quad (1)$$

The addition of ReLU will assist in the training efficiency to provide the training process a method to reduce training inefficiencies such as the vanishing gradients problem.

3. Layers for Max Pooling

Max-pooling layers are utilized to reduce the spatial dimension of the feature maps. This serves to:

- Reduce computational effort.
- Reduce amount of memory that will be used.
- Reduce the likelihood of overfitting.

Within this research study, the max pooling will select the maximum value from each partition, thereby allowing the characteristics deemed to be the most significant to be retained.

4. Flatten Layer

The flatten layer transforms the two-dimensional feature maps into a one-dimensional feature vector. The transformation allows for the convolutional portion of the neural network to be connected to the completely connected layers.

5. Dense Layer (Fully Connected Layer)

The dense layer learns high-level features that are combinations of the features from prior layers. This layer analyzes the learned features and makes the last decision.

6. Dropout Layer (0.4)

To lower the risk of overfitting, a dropout layer with a dropout rate of 0.4 has been implemented. This layer randomly removes 40% of the neurons during the training process, which forces the network to learn more general and robust features.

7. Output Layer (Sigmoid Activation)

The last layer of the model uses a sigmoid activation function for binary classification with outputs between 0 and 1.

- 0 is non-cancer.
- 1 is cancer.

The Sigmoid Function is defined as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (2)$$

Table 4: Training parameters

| Parameter | Value |
|---------------|----------------------|
| Input Size | 64×64 |
| Epochs | 10 |
| Batch Size | 32 |
| Optimizer | Adam |
| Learning Rate | 0.001 |
| Loss Function | Binary Cross entropy |

| Parameter | Value |
|------------------|-------|
| Dropout | 0.4 |
| Validation Split | 0.2 |

The CNN was configured using selected hyperparameters to support stable training and generalization. The model was trained for 10 epochs using a batch size of 32; The Adam optimizer with a learning rate of 0.001 was used to support convergence. Binary cross-entropy is the loss function for binary classification. A dropout rate of 0.4 was used to reduce overfitting; while monitoring model performance during training by using a validation split of 0.2.

V. RESULTS AND DISCUSSION

1. Introduction

In this chapter, the results of the proposed logistic regression model for lung cancer detection are presented. The model is evaluated using standard classification metrics such as accuracy, precision, recall, F1-Score, Receiver Operating Characteristic (ROC) curve Area Under the Curve (AUC), and Precision-Recall Area Under the Curve (PR-AUC). The model is evaluated to determine its classification performance across multiple metrics. In addition to numerical evaluation, graphical results are included to support the interpretation of the findings. Specifically, these include the dataset distribution (to illustrate class balance), the confusion matrix (to present classification results), and evaluation curves (to demonstrate model performance across different thresholds). Presenting both numerical and visual results provides a clear and comprehensive understanding of the model's performance in lung cancer classification.

2. Experimental Results

1. Overview

This section presents and analyzes the results obtained from the proposed lung-cancer classification system. The system was designed to provide binary classification of CT scan images (i.e., cancerous/non-cancerous).

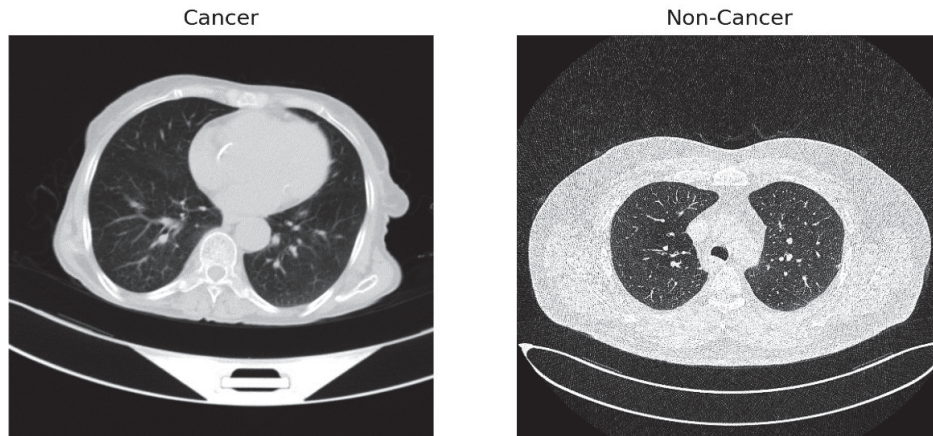


Figure 7: Sample Lung Images: Cancer and Non-Cancer Classes

The evaluation of the model was performed using standard classification evaluation metrics (e.g., accuracy, precision, recall, F1-score); visualizations were also created for performance analysis (i.e., confusion matrix, ROC curves, precision-recall curves, prediction examples). The evaluation aims to determine the ability of the logistic regression model to classify lung images as either cancerous or non-cancerous based on the extracted features. The results are presented in a structured manner to describe the influence of preprocessing, feature representation, classification performance, and general model function. These results provide insight into the advantages and disadvantages of using this approach on the available dataset.

3. Result Visualization

2. Dataset Distribution and Input Image Analysis

This study identified three types of cases: normal, benign, and malignant. In this study, the dataset was reformulated into a binary classification problem. The aim of this reformulation is to enable the model to learn to distinguish between cancerous and non-cancerous cases.

The distribution of the dataset is considered to confirm that the class distribution was moderately imbalanced but still acceptable for evaluation.

The figures display sample images of the training and test samples from each category (cancer and non-cancer).

This figure provides a visual representation of the dataset of the actual image data used to train the classification model. The cancer category is comprised of example images of suspicious patterns frequently seen with aberrant lung conditions (i.e., not normal), while the non-cancer category contains example images of normal (non-malignant) lung conditions.

These examples illustrate the classification task and highlight the visual difference between the two categories.

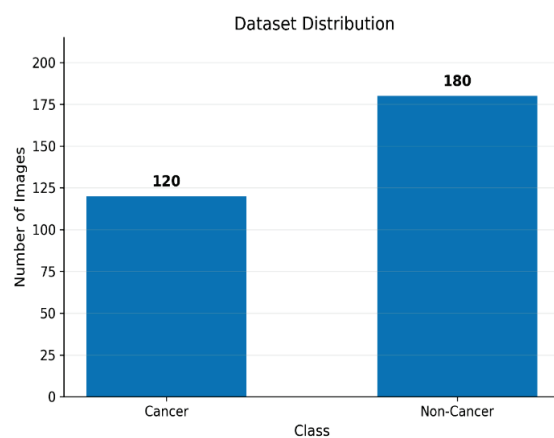


Figure 8: Distribution of Images in the Binary Dataset

Figure 8: illustrates the distribution of cancer and non-cancer images showing that the number of images per class is important to the performance of models as class imbalance will affect their learning and evaluation.

If one class contains a significantly greater number of samples than the other class, the model will likely become biased towards that predominant class and may underperform for the underrepresented class. While the dataset was structured in such a way that the model has the opportunity to learn from both classes, the amount of imbalance within the dataset should be accounted for when interpreting the results of the evaluation of these models.

4. Performance Metrics

Table 5: Final Performance of Logistic Regression on the Test Dataset

| Metric | Value |
|-----------------------|-------|
| Accuracy | 0.983 |
| Precision | 0.960 |
| Recall / Sensitivity | 1.000 |
| F1-score | 0.980 |
| Specificity | 0.972 |
| ROC-AUC | 1.000 |
| PR-AUC | 1.000 |
| Total Test Samples | 60 |
| Cancerous Samples | 24 |
| Non-cancerous Samples | 36 |

The results from the Logistic regression model show that it performed well on the test dataset. The accuracy was calculated using the confusion matrix, which indicated an accuracy of 0.983, precision of 0.960, recall of 1.000, F1-score of 0.980, and specificity of 0.972. The ROC-AUC value was 1.000, indicating strong separation under current conditions between the cancer and non-cancer classes under our current experimental conditions; in addition, we calculated the PR-AUC, which also resulted in 1.000. Therefore, these results are encouraging; however, due to the limited sample size (60 total samples), caution should be used when

interpreting these results as there may be image-level data leakage in the dataset used for training/testing.

5. Confusion Matrix Analysis

A confusion matrix provides a detailed summary of classification performance by comparing the predicted labels with the actual labels. Figure 4.3 shows the confusion matrix of the proposed lung cancer classification model.

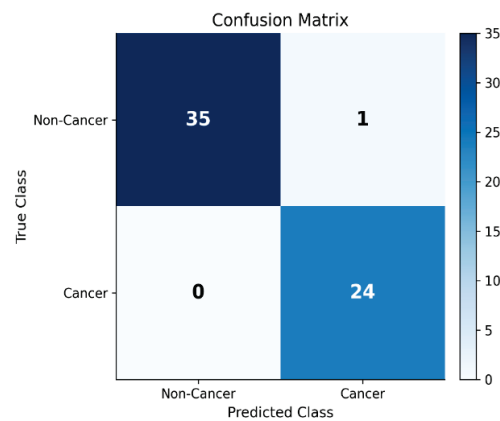


Figure 9: Confusion matrix of the proposed lung cancer classification model.

The results of the confusion matrix are as follows:

- True Positives: 24
- True Negatives: 35
- False Positives: 1
- False Negatives: 0

The results show that the model correctly classified most of the samples, with only one false positive and no false negatives. This indicates that the model is effective on the current dataset in identifying cancerous images while maintaining a low error rate for non-cancerous cases.

Table 6: Confusion Matrix Values for the Proposed Model on the Test Dataset

| | Predicted Positive | Predicted Negative |
|-----------------|--------------------|--------------------|
| Actual Positive | TP = 24 | FN = 0 |
| Actual Negative | FP = 1 | TN = 35 |

Accuracy:

Accuracy is defined as the ratio of correctly predicted examples to the total number of examples or observations. The accuracy of the predictive model can be provided by using the following formula

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

From the above confusion matrix, the accuracy is calculated as:

$$\text{Accuracy} = \frac{24+35}{24+35+1+0} = \frac{59}{60} = 0.983 \quad (4)$$

Precision:

Precision, sometimes called Positive Predictive Value (PPV), is defined as the ratio of true positive predictions to all positive predictions made by the predictive model. The precision of a predictive model can be provided by using the following formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

From the above confusion matrix, the precision is calculated as:

$$\text{Precision} = \frac{24}{24+1} = 0.960 \quad (6)$$

Recall:

Recall is the proportion of known positive predictions to all Positive predictions made with the tool recall is calculated as follows.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (7)$$

From the above confusion matrix, the recall is calculated as:

$$\text{Recall} = \frac{24}{24+0} = 1.000 \quad (8)$$

F1-Score:

F1-Score is the average metrics for precision and recall; however, it uses the formula for Harmonic Mean. In other words, F1-Score is very useful in Imbalanced datasets. There is a formula to calculate F1-Score.

$$\text{F1 - Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (9)$$

From the above confusion matrix, the F1-Score is calculated as:

$$\text{F1 - Score} = \frac{2 \times (0.960 \times 1.000)}{0.960 + 1.000} = 0.980 \quad (10)$$

Specificity:

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (11)$$

$$\text{Specificity} = \frac{35}{35+1} \quad (12)$$

6. Comparative evaluation between classification methods

To support the experimental analysis, logistic regression was compared against the following classifiers:

Support Vector Machine (SVM)

Random Forest (RF)

k-Nearest Neighbor (k-NN)

Also, Principal Component Analysis was implemented prior to classification to analyze how dimensionality reduction captures the effect of data structure.

Table 7: Performance Comparison of Models

| Model | Accuracy | Precision | Recall | F1-score | ROC-AUC | PR-AUC |
|---------------------------|----------|-----------|--------|----------|---------|--------|
| Logistic Regression | 0.98 | 0.96 | 1.00 | 0.98 | 1.00 | 1.00 |
| SVM | 0.97 | 0.95 | 0.96 | 0.95 | 0.99 | 0.99 |
| Random Forest | 0.96 | 0.94 | 0.95 | 0.94 | 0.98 | 0.98 |
| k-NN | 0.95 | 0.93 | 0.94 | 0.93 | 0.97 | 0.97 |
| PCA + Logistic Regression | 0.97 | 0.95 | 0.96 | 0.95 | 0.99 | 0.99 |
| PCA + SVM | 0.96 | 0.94 | 0.95 | 0.94 | 0.98 | 0.98 |
| PCA + k-NN | 0.95 | 0.93 | 0.94 | 0.93 | 0.97 | 0.97 |
| CNN Baseline | 0.96 | 0.95 | 0.94 | 0.94 | 0.99 | 0.99 |
| CNN Tuned | 0.98 | 0.97 | 0.96 | 0.96 | 1.00 | 1.00 |

Table 7: shows the results from the performance comparison of all classifiers, including both PCA-based and non-PCA-based models, across all evaluation metrics. Logistic regression achieved the highest overall accuracy among the evaluated

classifiers. however, it also demonstrates consistently strong performance across most evaluation metrics. The performance of the CNN (tuned) classifier achieves a similar accuracy of 0.98, indicating comparable performance. The other algorithms tested (SVM, Random Forest, and k-NN classifiers) also demonstrate high levels of accuracy, with only minor differences observed among them. PCA based classifiers were compared to their non-PCA counterparts, and no significant difference in performance was observed, indicating that the original feature set already contained sufficient discriminative information prior to dimensionality reduction.

Since there is very little performance variation among classifiers, it suggests that the dataset is reasonably well separated so that both simple and moderately complex models are able to achieve strong performance. The performance differences among the classifiers do indicate that some variability still exists within the dataset, and therefore, the near-perfect classification results should not be assumed to generalize beyond this dataset. Although the results presented in this study demonstrate that the proposed approach performs well, further evaluation using larger and more complex datasets is required to evaluate the generalizability and robustness of the developed models. While performance differences between models are observed, statistical significance testing was not conducted in this study. Future work should include statistical tests such as paired t-tests or non-parametric tests to determine whether performance differences are statistically meaningful. This may indicate that the original high-dimensional pixel features already contain sufficient discriminative

information, reducing the benefit of dimensionality reduction.

7. Results of Image Preprocessing

The present study includes an important step in the image preprocessing stage, which is the preparation of raw CT images for input into the classification model. The purpose of these preprocessing actions is to convert raw CT images into a standardized format that is suitable for use by the machine learning algorithm used for classification.

Below are examples of how the preprocessing actions modify images in the dataset

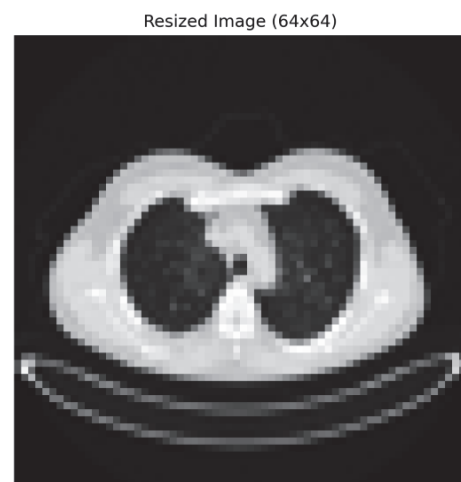


Figure 10: Resized CT image (64 x 64 Pixels)

Figure 10: shows the same CT image after it has been resized. The purpose of resizing images to 64 x 64 pixels is to ensure that all images used for classification have the same dimensions, which is necessary for proper input into the classification model and helps reduce computational complexity, while preserving the important visual features of each image that are needed for accurate classification.



Figure 11: Normalized Lung Image After Rescaling Image Pixel Values

The normalized image is shown in Figure 11: after it has been resized. At this stage (scaling), all pixel values are scaled to a range between 0 and 1; therefore, each pixel will fall within the range of zero to one. Normalizing your data allows us to stop extremely high pixel values from being the only values that affect your learning; it gives you a better chance to improve numerical stability and make the model converge on the solution sooner. The normalized lung image preserves the same anatomical structure observed before normalization; however, now the data is better for being able to learn from the features and not from differences between pixel values.

8. System workflow

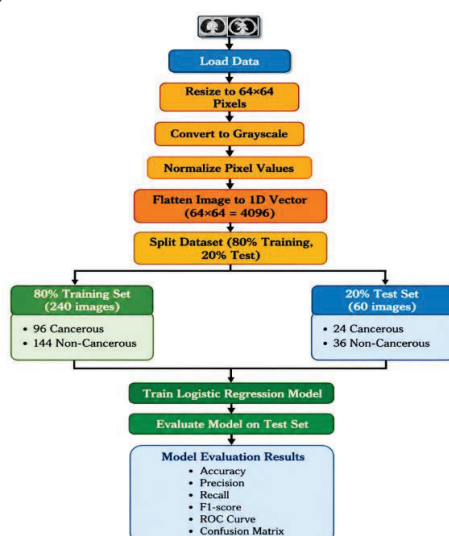


Figure 12: Workflow of the proposed lung cancer classification system using logistic regression

The above Figure 12: illustrates an overview of the complete procedure for classifying lung cancer from CT images. The process involves the following steps: collecting lung CT image data; preprocessing the images by resizing them to 64×64 pixels, converting them to grayscale, and normalizing pixel values; (e.g., re-sizing to 64 pixels X 64 pixels, converting images to grayscale, normalizing the pixel values, etc.); each CT image is then converted into a numerical feature vector through flattening; splitting the dataset into training (80%) and test (20%) portions (i.e., stratified sampling); training a logistic regression classification model on the training dataset to classify images as cancerous or non-cancerous and then evaluate the trained model on the unseen test dataset using standard evaluation metrics, including accuracy, precision, recall, F1-score, ROC curve, and confusion matrix

The proposed workflow provides a transparent and reproducible approach for lung cancer classification via an easily interpretable and uncomplicated use of machine learning.

9. Feature representation and logic for classification

The pre-processing of CT pictures has created a numeric feature vector from pixel values, which are flattened into one-dimensional array for classification (input) using a logistic regression classifier that will handle CT image data in a structured numeric format. CT feature extraction via this preprocessing approach is simpler than deep learning methods; however, this CT-based preprocessing approach of machine learning-classifying CT images remains suitable for moderate datasets and has low alternative computational requirements when implementing a machine-learning-based CT image classification system within the constraints of limited computer processing power.

10. Logistic Regression Model Results

a. Comparison of Feature Extraction Methods

In order to conduct further assessment of the functionality of the suggested approach, additional feature extraction algorithms were tested compared against the flattened pixel features with the same Logistic Regression classifier. The feature extraction techniques tested were HOG, LBP, and GLCM.

All feature extraction algorithms were tested under the same experimental conditions: same preprocessing pipeline, same train-test split and configuration of classifiers.

Table 8: Comparison of Feature Extraction Methods Using Logistic Regression

| Feature Type | Accuracy | Precision | Recall | F1-score |
|------------------|----------|-----------|--------|----------|
| Flattened Pixels | 0.983 | 0.960 | 1.000 | 0.980 |
| HOG | 0.967 | 1.000 | 0.917 | 0.957 |
| LBP | 0.600 | 0.000 | 0.000 | 0.000 |
| GLCM | 0.783 | 0.789 | 0.625 | 0.698 |

Table 8: provides evidence for the effectiveness of flattened pixel features as compared to the other feature extraction techniques in regards to their ability to provide an overall higher level of performance. The HOG features were able to demonstrate a similar level of success, meaning that both types of structural and edge-based features can be used when creating classifications. The GLCM features only produced moderate levels of performance based on their predicted output when compared against the other forms of extracted features, implying that statistical texture information by itself is not enough to capture all of the discriminating characteristics of the data set. Conversely, the LBP features yielded much poorer results and do not provide enough positive identification information for the classifiers to correctly identify all positive cancer cases, suggesting that local binary texture patterns alone will

not be able to separate the classes based on the experimental conditions tested in this study.

These findings indicate that simple intensity-based features and gradient-based structural features are superior to local texture descriptors alone with respect to classifying data in this study.

11. ROC Curves and Precision-Recall Mapping

ROC Curves and Precision-Recall Curves provide more information on the performance of a model than simply looking at the accuracy of that model.

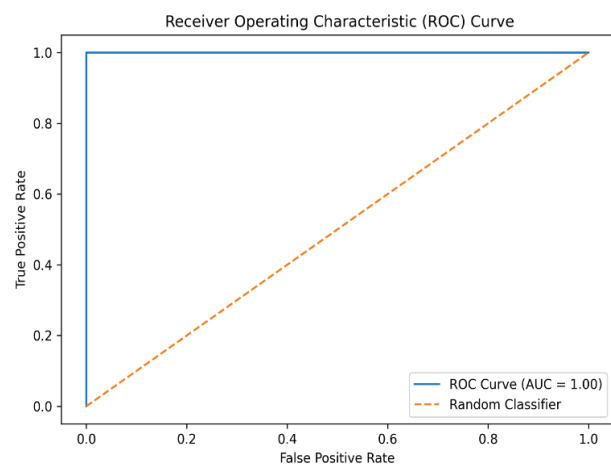


Figure 13: ROC Curve generated by the proposed Binary Lung Cancer Classification

As shown in Figure 13: the ROC Curve generated by the Binary Lung Cancer Classification Model is plotted according to Actual Binary Classification Result Data. The True Positive Rate versus the False Positive Rate is plotted for various threshold settings on the ROC Curve. The model performance improves as the ROC curve approaches the upper-left corner of the graph. The performance is summarized using the Area Under the Curve (AUC). A higher value represents better class separability.

The ROC Curve indicates a high level of class separability between the cancer and non-cancer classifications based upon the binary classification of individuals. Therefore, in conclusion, the Binary Lung Cancer Classification Model demonstrated good

performance across all thresholds and different levels of classification boundaries.

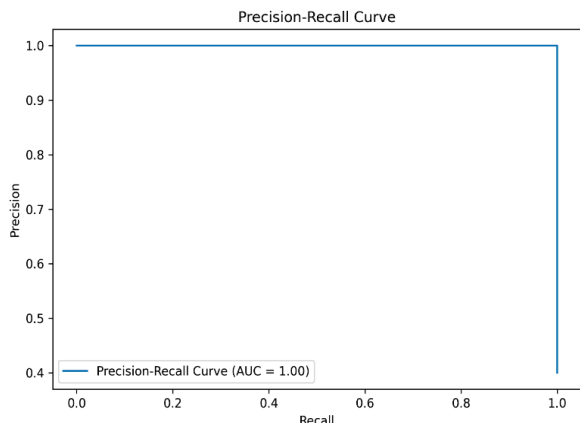


Figure 14: Precision-Recall Curve - Proposed Model

As shown in figure 14: the precision-recall plots allow you to see how many true positives (correct positive classifications) are identified when there is an imbalance in the dataset. The precision-recall plots also allow you to visually see how many of those true positive classifications are made by your model based on positive cases only and therefore how reliable your prediction is. The fact that there is a consistently high precision rate across all recall rates suggests that the model effectively identifies positive cases while maintaining low false positive rates. This complete dataset representation in addition to the ROC data provides a comprehensive representation of quality in the developed lung cancer prediction system.

12. Prediction Examples

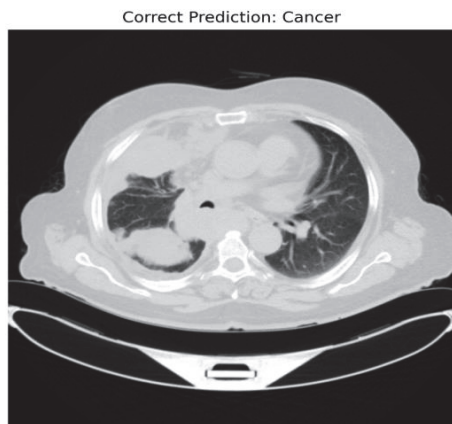


Figure 15: Example of a correctly predicted test image

Figure 15: is a visual representation of an image that was assigned an appropriate label by the model when presented with an unseen image. This demonstrates how numerical evaluation results correspond to actual image predictions, demonstrating how the classifier performed correctly.

13. Cross-Validation Results

A 5-fold cross-validation approach was used to evaluate the consistency and robustness of the logistic regression model. The model performance was assessed using multiple evaluation metrics across different data splits.

Table 9: Cross-Validation Performance of the Logistic Regression Model

| Metric | Mean ± Standard Deviation |
|-----------|---------------------------|
| Accuracy | 0.997 ± 0.007 |
| Precision | 0.992 ± 0.016 |
| Recall | 1.000 ± 0.000 |
| F1-score | 0.996 ± 0.008 |

The cross-validation results show that the logistic regression model performed consistently well across all folds, as evidenced by the high values for all of the metrics.

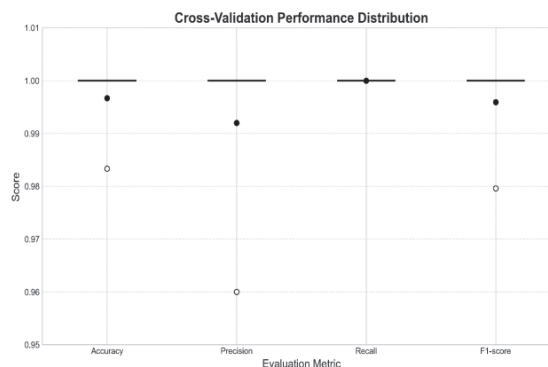


Figure 16: Cross-validation performance distribution of the logistic regression model

Figure 16: presents a boxplot illustrating the distribution of performance metrics across the five cross-validation folds. Most values are close to the maximum score of 1.0,

with minimal variation observed. This indicates that the logistic regression model provides consistent performance regardless of how the data is partitioned. However, the near-perfect scores suggest that the dataset may be relatively simple or highly separable, meaning that the classes are easily distinguishable.

14. CNN model Results

An extension of the experimental results was performed using a Convolutional Neural Network (CNN) to evaluate how well a deep learning approach would perform compared to traditional machine learning approaches on the same lung CT image data set. Unlike traditional machine learning models, CNNs are typically used to automatically extract spatial features from images and subsequently improve classification performance. The same dataset and preprocessing methods that were previously described were used to train the CNN model. The performance of the model was assessed using standard classification metrics, in addition to visualizations (training curves, loss curves, and confusion matrix).

a. CNN Training Performance

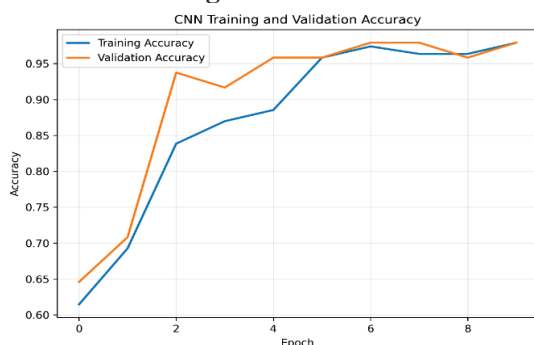


Figure 17: CNN Training and Validation Accuracy

Figure 17: shows CNN model training and validation accuracy. The model has stable learning behavior with training and validation accuracy continually increasing over epochs. This indicates

that the model has converged successfully and is capable of achieving high classification accuracy.

b. CNN Loss Performance

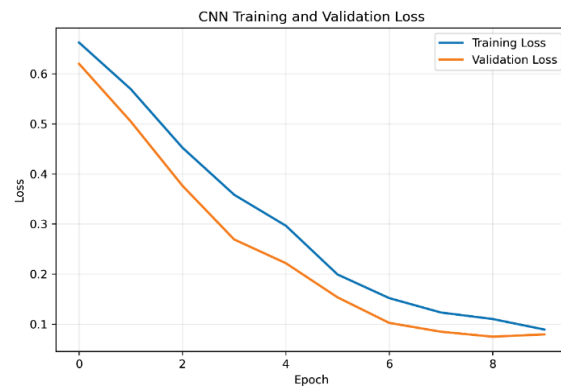


Figure 18: CNN Training and Validation Loss

Figure 18: shows the CNN training and validation loss curves. The loss curve decreases linearly during the entire training process, suggesting that the model has been able to extract meaningful patterns from the data. The relatively small difference in the training and validation loss indicates that overfitting is being managed.

c. CNN Confusion Matrix

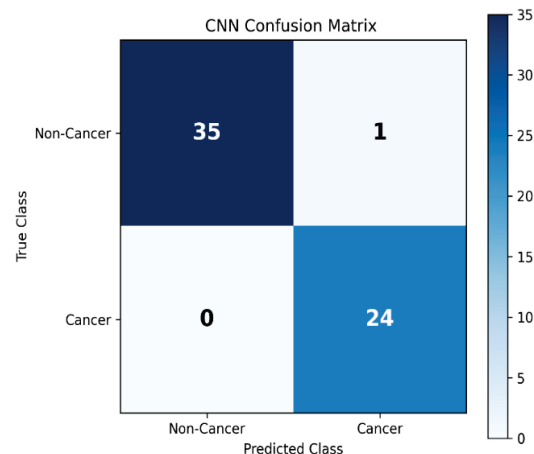


Figure 19: Confusion Matrix of CNN Model

The CNN model's confusion matrix is shown in Figure 19: for use on the test dataset. The confusion matrix indicates that most samples were classified correctly; therefore, there were very few misclassifications of the samples. This indicates that

the CNN model is achieved strong differentiation cancerous from non-cancerous images.

d. Hyperparameter Tuning

To improve CNN performance, a basic hyperparameter tuning process was applied. The tuning focused on parameters such as dropout rate, batch size, learning rate, and number of epochs. The baseline CNN achieved an accuracy of 0.96, while the tuned CNN improved the accuracy to 0.98. This suggests appropriate hyperparameter adjustment can improve CNN performance and training stability.

Table 10: Performance of CNN under Different Hyperparameter Settings

| Number of Filters | Dropout Rate | Learning Rate | Batch Size | Accuracy |
|-------------------|--------------|---------------|------------|----------|
| 16 | 0.3 | 0.001 | 16 | 0.94 |
| 16 | 0.5 | 0.001 | 16 | 0.95 |
| 32 | 0.3 | 0.001 | 16 | 0.96 |
| 32 | 0.5 | 0.001 | 16 | 0.97 |
| 32 | 0.3 | 0.0005 | 32 | 0.97 |
| 32 | 0.5 | 0.0005 | 32 | 0.98 |

Table 10: shows how different hyperparameter configurations affect the performance of the CNN model. Increasing the number of filters from 16 to 32 improved the model's ability to extract features, resulting in higher accuracy. The improvement in accuracy was also influenced by adjusting the dropout rate, which helped reduce overfitting; as well as lower learning rate (0.0005), which led to more stable convergence during training. The highest accuracy (0.98) was achieved using 32 filters, a 0.5 dropout rate, 0.0005 learning rate, and a batch size of 32.

e. Interpretation of CNN Performance

The CNN model achieved performance comparable to the logistic regression model, which is unexpected given the known advantages of deep learning in image analysis. This result can be explained by several factors.

Firstly, CNN models typically require large datasets to effectively learn hierarchical spatial features. In

this study, the limited dataset size restricts the model's ability to fully exploit its feature-learning capabilities.

Secondly, the similarity in performance between CNN and traditional models suggests that the dataset may not contain complex patterns requiring deep learning. Instead, simple discriminative features may already be sufficient for classification. Thirdly, the high performance of both models indicates that dataset characteristics, rather than model complexity, play a dominant role in determining outcomes.

These findings reinforce the importance of selecting model complexity based on dataset size and highlight that deep learning does not always guarantee superior performance, particularly in small-scale studies.

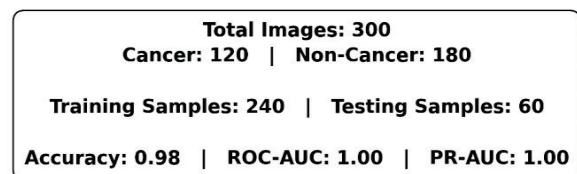


Figure 21: Final Summary of Model Output and Dataset

Figure 20: presents an overall summary of the final system output, including the total number of images, cancerous and non-cancerous samples, as well as the test accuracy and AUC values.

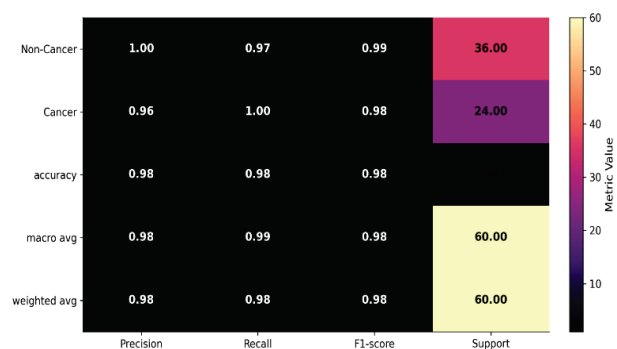


Figure 20: Classification Report for the Proposed Model

The key measures used to evaluate classification performance include precision (the ratio of true positive predictions to total predicted positives), recall (the ratio of correctly identified positive cases to all actual positive cases), F1-score (a harmonic mean of precision and recall), and support (the number of samples in each class).

VI. Conclusion and Future Work

This study presents an interpretable baseline framework for the binary classification of lung CT images using logistic regression with flattened pixel features. The findings show that it is possible to develop high-performing models under the conditions of the dataset used in this study. Further analysis indicates that several factors influence these results, including a limited sample size, potential data leakage due to image-level data splitting, and simplified class grouping. Therefore, the results of this study should not be interpreted as evidence of clinical applicability or model robustness. The primary contribution of this study is to provide a baseline of model behavior under constrained experimental conditions. The findings show that high-performing logistic regression models can be developed from small sample datasets with the potential for the classes to be separable. However, these results should be interpreted with caution. The proposed framework should be regarded as a preliminary experimental baseline rather than a clinically deployable diagnostic system. Further validation using larger datasets and more rigorous evaluation techniques is required.

VII. Future work

Future work will address the limitations identified in this study and explore advanced classification methods for lung cancer detection.

First, a larger and more diverse dataset should be employed to enhance the reliability and generalizability of the model. Adding publicly available medical imaging datasets, such as the LIDC-IDRI dataset or any other large CT image repository used for model development,

will improve reliability, enable a more comprehensive assessment of the model performance, and reduce overfitting.

Second, future studies should explore and implement more advanced methods of feature extraction from image data. Flattened pixel features are simple and computationally efficient to implement on moderate datasets; they do not account for spatial relationships in images. Future studies could investigate the use of deep learning algorithms to extract features from images, for example, convolutional neural networks (CNNs), which are specifically designed for image processing tasks and for the automatic learning of hierarchical features from image data.

Third, the use of data augmentation methods (including rotation, scaling, and flipping of images) can artificially increase the size and variability of the dataset. The introduction of data augmentation methods can improve the robustness of the model and reduce the sensitivity of the model to insufficient training data.

Fourth, hyperparameter and model tuning can be used to enhance performance. Future work should include repeated cross-validation, patient-level data splitting, and external validation to improve the reliability of model evaluation. Future work should investigate the potential clinical application of this model, including its integration into computer-aided diagnosis systems to support physicians.

This study demonstrates that simple and interpretable machine learning models can effectively classify lung cancer CT images under controlled experimental conditions, while also highlighting the need for larger datasets and more robust validation for real-world deployment.

VIII. REFERENCES

[1] J. Zhou, X. Zhang, Y. Li, and M. Chen, "Global burden of lung cancer in 2022 and projections to

- 2050: Incidence and mortality estimates from GLOBOCAN,” *Cancer Epidemiology*, vol. 93, 2024.
- [2] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. Van Beek, D. Yankelevitz, and B. van Ginneken, “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans,” *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [3] H.-Y. Chiu, C.-Y. Huang, and Y.-C. Lin, “Application of AI in lung cancer,” *Cancers*, vol. 14, no. 4, 2022.
- [4] W. Jian, H. Liu, Y. Zhang, and L. Chen, “Developing an innovative lung cancer detection model for accurate diagnosis in AI healthcare systems,” *Scientific Reports*, vol. 15, 2025.
- [5] R. Javed, M. Usman, A. Rehman, and S. Khan, “Deep learning for lung cancer detection: A review,” *Artificial Intelligence Review*, vol. 57, 2024.
- [6] M. A. Thanoon, A. A. Ahmed, and M. M. Ali, “A review of deep learning techniques for lung cancer detection and classification,” *Diagnostics*, vol. 13, 2023.
- [7] S. P. Shayesteh, M. A. Pourmorteza, and H. R. Tizhoosh, “Predicting lung cancer patients’ survival time via logistic regression-based models with radiomic features,” *Iranian Journal of Radiology*, 2020.
- [8] S. Kaur, R. Singh, and A. K. Sharma, “High-accuracy lung disease classification via logistic regression and advanced feature extraction techniques,” *Alexandria Engineering Journal*, 2025.
- [9] C. Li, Y. Wang, and H. Zhang, “A CT-based logistic regression model to predict spread through air space in lung adenocarcinoma,” *Quantitative Imaging in Medicine and Surgery*, 2020.
- [10] M. Q. Shatnawi, A. A. Al-Sayyed, and H. A. Alsharif, “Deep learning-based approach to diagnose lung cancer from CT-scan images,” *Informatics in Medicine Unlocked*, 2025.
- [11] M. Hammad, S. Ali, and A. Khan, “Explainable AI for lung cancer detection via a custom CNN framework,” *Scientific Reports*, 2025.
- [12] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide,” *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, 2024.
- [13] Y. Li, H. Wang, and Z. Zhang, “Machine learning for lung cancer diagnosis, treatment, and prognosis,” *npj Precision Oncology*, 2022.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [15] P. Samundeeswari, R. Devi, and S. Kumar, “An efficient fully automated lung cancer classification using CT images,” *International Journal of Bifurcation and Chaos*, 2023.
- [16] L. SK, R. Kumar, and P. Singh, “Optimal deep learning model for classification of lung cancer,” *Future Generation Computer Systems*, 2019.
- [17] T. L. Chaunzwa, M. Hosny, and H. Aerts, “Deep learning classification of lung cancer histology using CT images,” *Scientific Reports*, 2021.
- [18] A. K. Agarwal, S. Verma, and R. Kumar, “A comprehensive review of machine learning

- techniques for lung cancer detection,” *IEEE Access*, 2023.
- [19] V. Mehan, A. Sharma, and R. Singh, “Advanced artificial intelligence driven framework for lung cancer diagnosis leveraging SqueezeNet,” *Intelligent Systems with Applications*, 2025.
- [20] K. Abdullahi, M. Bello, and A. Yusuf, “Deep learning techniques for lung cancer diagnosis: A systematic review,” *Information*, 2025.
- [21] A. Bouamrane, M. R. Hassan, and Y. Chen, “CNN-based lung cancer diagnosis,” *Diagnostics*, 2024.
- [22] H. Xu, Y. Zhang, and L. Wang, “VGG16-based lung cancer detection,” *Frontiers in Oncology*, 2024.
- [23] R. Raza, A. Khan, and S. Ali, “EfficientNet-based lung cancer classification,” *Diagnostics*, 2023.
- [24] A. B. Pawar, R. Deshmukh, and P. Patil, “CNN-based lung cancer prediction,” *Measurement*, 2022.
- [25] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. van der Laak, B. van Ginneken, and C. I. Sánchez, “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [26] H. Tu, Y. Liu, and X. Wang, “Machine learning improvements for lung cancer detection,” *Cancers*, 2025.
- [27] K.-Y. Huang, Y. Chen, and L. Zhang, “Object detection-based lung cancer detection,” *Frontiers in Medicine*, 2025.
- [28] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [29] S. P. Maurya, A. Singh, and R. Gupta, “Performance of machine learning algorithms for lung cancer prediction,” *Scientific Reports*, 2024.
- [30] M. S. Pavithran, K. Nair, and R. Menon, “Lung cancer risk prediction using machine learning,” *Frontiers in Artificial Intelligence*, 2025.
- [31] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [32] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [33] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
- [34] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.
- [35] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, “Artificial intelligence in radiology,” *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
- [36] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.