Engineering Reports

WILEY

Engineering Reports Open Access

RESEARCH ARTICLE OPEN ACCESS

LCDLNet: Attention Mechanism-Based Deep Learning Framework for Lung Cancer Diagnosis from Computed Tomography (CT) Images

Farjana Akter Chumki | Tahasin Ahmed Fahim | Sabrina Alam

Department of Electrical and Electronic Engineering, University of Chittagong, Chattogram, Bangladesh

Correspondence: Sabrina Alam (sabrina.apece@cu.ac.bd)

Received: 21 December 2025 | **Revised:** 11 May 2026 | **Accepted:** 15 May 2026

Keywords: attention mechanism | convolutional neural networks | deep learning | lung cancer | transformer

ABSTRACT

This paper introduces a new deep learning architecture, LCDLNet, that can diagnose and classify lung cancer using CT scan images. LCDLNet makes use of a custom CNN structure including residual connections, DenseNet121, SE blocks, and transformer modules, which allow multi-scale and contextualized feature mappings. It combines the intermediate feature layers from CNN and DenseNet121 to provide a comprehensive solution to the extraction of distinguishing features necessary for the accurate classification of lung diagnosis. The LIDC-IDRI dataset was used to train LCDLNet and evaluate its performance. The training dataset consisted of 5292 CT slices obtained from 165 patients, and all images were resized to 224×224 pixels and normalized to achieve stable training. The results achieved with LCDLNet were excellent, with a training accuracy of 99.53%, validation accuracy of 98.98%, and test accuracy of 99.81%. Achieving 99.86% Precision, 99.67% Recall, and 99.80% AUC scores demonstrates that this model has been able to generalize successfully while exhibiting minimal overfitting. The proposed framework is effective at solving class imbalance issues while providing a unique attention mechanism. This will provide an important step towards building an automated system for the diagnosis of lung cancer and the development of a clinically useful decision support tool.

1 | Introduction

Lung cancer has become one of the world's leading health problems because it is usually diagnosed too late and has many different types of cancer, making diagnosis difficult. Lung cancer develops when normal lung cell growth goes out of control and becomes uncontrolled, often producing no initial symptoms. Symptoms usually begin appearing when a person has reached an advanced (late) stage of the disease [1]. Although approximately 85% of all lung cancer deaths in 2025 result from smoking and secondhand smoke, there is a growing proportion of cases in people who have never smoked [2]. The effects of smoking can be compounded by combinations with other risk factors such as radon and asbestos. This suggests that even if levels of smoking are reduced, because of other causes, lung cancer will still be present [3]. The disease occurs most often in individuals aged 65 years or older [4]. The symptoms of early-stage lung cancer are often non-specific or absent until late progression; therefore, symptoms such as coughing, chest, and/or shortness of breath may be reported only at later stages [5]. This limits the chances for early detection and increases the possibility of a poor outcome. According to the GLOBOCAN estimation for 2022, the global cancer burden was 19.3 million new cases and nearly 9.96 million deaths in that year [5]. Among all cancers, lung cancer was the most commonly diagnosed, 2.48 million (12.4%), and the leading cause of deaths from cancer, 1.82 million (18.7%) [6]. There were around 1.6 million cases in males and 0.91 million in females, which corresponded with mortality numbers [6].

Future projections indicate, if the rates of incidence and mortality remain stable, that the overall burden of lung cancer will sharply rise to 4.62 million for new cases, and 3.55 million for deaths, with an estimated calculation at 2050 [7]. Further, when we look at the rounded aggregate cancer statistics, the rate of new cases is on the rise from about 20 million in 2022 to 32.6 million in 2045 for an increase of about 63% in 23 years, with lung cancer likely following the same trajectories as noted [8]. For patients diagnosed with localized lung cancer, the 5year survival rate sits at approximately 63.7%, while only 8.9% of patients diagnosed with lung cancer at later stages survive beyond 5 years [9]. The earlier lung cancer is diagnosed, the greater the chance an individual has to survive the disease. According to the American Society for Cancer, if the disease is diagnosed early, an individual has a survival chance of up to 47% [2].

Typically, lung cancer diagnosis starts with a chest X-ray, but this is only a screening tool and does not confirm cancer [9]. The next steps to assess lung cancer are CT scans, MRI, and PET scans. A PET scan is particularly important for determining metastasis and cancer spread to other organs [10, 11]. If lung cancer is suspected, tests such as bronchoscopy with biopsy, endobronchial ultrasound (EBUS), mediastinoscopy, thoracoscopy, or a wedge resection are done to get a tissue sample to confirm diagnosis [10, 11]. Early-stage lung cancer can be challenging to detect because of vague symptoms and limitations of current imaging options. Even experienced radiologists may miss small lesions in the lungs, with 20% of cancers being missed on chest X-rays, and nearly 10% of CT-guided biopsies hidden by disorders like organizing pneumonia as a result of missed diagnoses and delayed treatment, leading to worse outcomes [12, 13]. These difficulties highlight the need for improved imaging techniques and diagnostic approaches to facilitate the detection of lung cancer at an early stage and ultimately increase success rates.

Deep learning (DL), a form of machine learning, has quickly appeared as a powerful methodology for analyzing medical data, and allows researchers to identify complex relationships across large datasets and develop findings for clinical applications [14, 15]. DL approaches use multi-layered neural networks that learn to represent the data without human-constructed features. In cancer screenings, especially, DL has led to Computer-Aided Diagnosis (CAD) systems that may provide promising means for radiologists and clinicians to help positively influence early detection and diagnosing errors by improving the mammography diagnostic capabilities of health professionals [16, 17]. Specifically, Convolutional Neural Networks (CNNs) as a specialized form of DL are especially useful; CNNs use convolutional layers (CLs) as their backbone architecture, which utilizes multiple filters that work on an image using its weight and non-linear transformation/liquid activation and outputs different representations of the same underlying data as its features for the training samples using the convolutional model to output features related to medical images, including cancerous tissues [18]. A typical workflow for studies involving CAD or DL typically includes preprocessing to improve image quality, segmentation to extract lung regions of interest, and feature extraction to capture discernible malignancy features. Classification is then achieved with Deep Convolutional Neural Networks (DCNNs), [19] which contain convolutional layers, activation functions, and pooling to learn hierarchy (features) and fitted to different data representations via connected layers. Deep learning systems are classified into different categories depending on how they learn. These are as follows: Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Transduction, and Reinforcement Learning [20]. Deep learning, especially CNN-based approaches, has developed automated, accurate methods for medical image analysis and is considered a revolutionary technique for accurate classification and prediction of lung cancer [14, 18]. A list of key contributions as part of the proposed research includes:

- A new deep learning model, LCDLNet, for classifying lung cancer that utilizes a custom-built CNN using Residual Connections, DenseNet121 architecture, Squeeze-and-Excitation Blocks, and Transformer Modules.
- The architecture combines channel-wise attention, local feature extraction, and global context learning to improve the classifier's performance.
- From the trained results, we evaluated performance metrics and found satisfactory results, indicating a promising outlook compared to other current lung cancer detection models.
- Developing and testing an advanced architecture using real patient data that are clinically relevant, we have demonstrated a strong framework for automated lung cancer classification and potentially, detection.

The organization of the content of the paper includes separate sections, allowing for a systematic presentation of the complete body of research according to the model discussed in this paper. The 'Related Works' section presents a large number of research papers on lung cancer diagnostic applications. In the 'Materials and Methods' section, a brief overview of the proposed method, including a discussion regarding the use of the dataset, preparation of image data, splitting of the dataset into training and testing sets, and application of the attention mechanism, is included. In the 'Proposed Model Architecture' section, the proposed deep learning framework and the performance evaluations are explained in detail. The 'Experimental Results' section presents detailed numerical and graphical performance evaluations of the proposed Work. The 'Discussion' section demonstrates, through comparisons of established deep learning frameworks for lung cancer diagnostics, that the proposed model is conclusive. The 'Conclusion' section briefly summarizes the potential impact of this work on the future of lung cancer prognosis.

2 | Related Work

The rapid emergence of deep learning-based techniques applied to medical imaging over the past few years has led to outstanding outcomes on multiple fronts, including the detection, segmentation, and classification of illness (disease). There has been a great deal of interest within academia around the use of convolutional neural networks (CNNs) and hybrid models that leverage both convolutional neural networks (CNN) and deep learning architectures to enhance diagnostic accuracy and clinical decision making by improving CKC classification of lung cancer from

CT scans. The investigations conducted on these methods are outlined in detail in subsequent sections.

2.1 | Convolutional Neural Network (CNN) Based Methods

Recent developments in Convolutional Neural Networks (CNNs) have led to substantial growth in the ability to analyze medical imaging data; in this instance, the utilization of CNNs has proven beneficial in developing methods for the early detection and diagnosis of lung-related cancers. However, as seen in this chapter, CNNs are used not only for lung cancer detection but also for other types of cancer research.

Faruqui et al. [21] created LungNet, a hybrid 22-layer CNN architecture, developed using both CT imaging plus MIoT (medical internet of things—wearable sensor devices that monitor patient vitals). LungNet was able to achieve superior accuracy when evaluated on a large, well-balanced training set (525,000 images) of lung cancer cases, and was also able to classify the earliest stages of lung cancer (IA, IB, IIA, and IIB) as well. On the other hand, Zhang et al. [22] built a 3D denseNet using volumetric convolution filters and volumetric-pooling operations to categorize lung nodules based on their appearance using data from the LUNA16 dataset (a subset of the LIDC/IDRI database). The model has demonstrated competitive predictive performance and adaptability for integration into CAD systems for lung cancer diagnosis.

The authors Raza et al. [23] developed the Lung-EffNet model to classify lung cancers. The Lung-EffNet model uses the EfficientNet architecture and improves it by adding new top-layer structures trained on the IQ-OTH/NCCD dataset, which contains images of lung cancers (benign/malignant/normal). Raza et al. found that utilizing the EfficientNet architecture improved classification accuracy. Continuing with this research, Gautam et al. [24] designed a multimodal deep learning framework that used ResNet152, DenseNet169, and EfficientNetB7 in combination to utilize a weighted fusion methodology based on performance to classify lung cancer types accurately on the LIDC-IDRI dataset while minimizing the number of false negatives for improved sensitivity, making it a more trustworthy tool for autopsying lung cancer.

2.2 | Hybrid CNN-Based Architectures

While traditional CNN approaches have demonstrated high performance in lung cancer detection, research shows that hybrid networks (incorporating CNNs with other complementary architectures) could enable researchers to obtain richer representations of image features and build models with enhanced robustness. Examples of these hybrid designs include CNNs and Capsule Networks, CNNs and Transformers, and CNNs with recurrent Neural Network architectures. Hybrid networks have shown improved robustness, greater interpretability, and higher predictive accuracy.

Tandon et al. [25] created a novel model called VCNet by integrating VGG-16 (a well-established CNN) with Capsule Networks (CapsNet). The use of VGG-16 allowed for effective feature extraction from images, while CapsNet added an added level of robustness to certain transformations (e.g., rotation, tiling, etc.). This model was tested on the LIDC-IDRI dataset, where VCNet performed better than other currently used architectures, including MobileNet (a lightweight model), Xception (an Xception-type architecture), and VGG-16 alone. Similarly, Bushara et al. [26] developed a different hybrid model called LCD-CapsNet that combined traditional CNN methods with CapsNet to preserve spatial information for identification purposes (e.g., distinguishing benign (non-cancerous) from malignant (cancerous) nodules). Testing was performed on the same LIDC-IDRI dataset, where LCD-CapsNet outperformed standalone CapsNet in classifying benign and malignant nodules, supporting the merits of combining two or more models for spatial preservation.

Recently, Chen et al. [27] introduced a hybrid model that combines CNN and Swin Transformers and utilizes the attention mechanisms of the Transformer to pay more attention to the specific parts of the lung that are necessary for diagnosis. Their experiments on the LIDC-IDRI dataset demonstrated that this combination improved the model's interpretability by enabling it to identify clinically important regions throughout the entire training process. Zhang et al. [28] further expanded on this hybrid approach by using CNNs combined with Gated Recurrent Units (GRUs) to model the temporal relationships between the CT slices.

2.3 | Hybrid CNN With Machine Learning and Sequential Models

In addition to the CNN-CapsNet and CNN-Transformer combinations we have examined, researchers have also studied fusing CNNs with traditional machine learning classifiers and sequential models to increase predictive performance and interpretability for diagnosis. Such methods aim to take advantage of CNNs' spatial feature extraction capabilities, while also integrating complementary methods to enhance decision boundaries or enhance temporal modeling or generalization of features.

Polat and Danaei Mehr, [29] created two separate frameworks using multiple deep-learning methods:

1. the conventional 3D-CNN method and
2. the hybrid 3D-CNN method, which combined the 3D-AlexNet and 3D-GoogleNet architectures with a radial basis function model (RBF) based Support Vector Machine (SVM) classifier.

The two hybrid CNN-SVM frameworks were tested on both the Data Science Bowl and Kaggle CT datasets. Their hybrid CNN-SVM model exhibited moderate accuracy, which highlights both the potential of early hybridizing CNN-SVM models and the corresponding limitations of early CNN-SVM combination approaches. Continuing this line of work, Saleh et al. [30] introduced a new CNN-SVM hybrid, in which the CNN extracted deep features from CT images, and the SVM served as the final classification step, stripping away irrelevant features. Their approach demonstrated respectable accuracy in classifying lung images as

TABLE 1 | Comparison of existing lung cancer detection and classification methods.

Author	Model architecture	Performance metrics	Limitation
Faruqui et al. 2021 [21]	LungNet: 22-layer CNN with MIoT	Classification accuracy: 96.81%, FP: 3.35%; Sub-stage accuracy: 91.6%, FP: 7.25%	Inconsistent MIoT data quality limits clinical applicability.
Zhang et al. 2021 [22]	3D DenseNet	Accuracy: 92.4%	Limited to nodule classification; non-nodule patterns not addressed.
Raza et al. 2023 [23]	Lung-EffNet using Transfer Learning (EfficientNet B0–B4)	Accuracy: 99.10%; ROC AUC: 0.97–0.99	Lack of external dataset validation.
Gautam et al. 2024 [24]	Ensemble CNN with Transfer Learning (ResNet152, DenseNet169, EfficientNetB7)	Accuracy: 97.23%; Sensitivity: 98.6%	Ensemble complexity increases computational cost.
Tandon et al. 2022 [25]	VCNet: VGG-16 + Capsule Networks	VCNet: 99.49%; MobileNet: 98%; Xception: 97.97%; VGG-16: 96.95%	High computational complexity limits real-time applicability.
Bushara et al. 2023 [26]	LCD-CapsNet: CNN + Capsule Networks	Accuracy: 94%; Precision: 95%; Recall: 94.5%; F1: 94.5%; Specificity: 99.07%; AUC: 0.989	Sensitivity to noisy CT images.
Chen et al. 2022 [27]	CNN combined with Swin Transformer	Accuracy: 96.14%; Precision: 95.20%; Recall: 92.60%; Specificity: 98.17%	High computational resource requirement.
Zhang et al. 2025 [28]	Hybrid CNN with GRU	Accuracy: 99.77%	Slow inference due to sequential GRU processing.
Danaei et al. 2019 [29]	Hybrid 3D-CNN (3D-AlexNet + 3D-GoogleNet) with RBF-SVM	Accuracy: 91.81%; Sensitivity: 88.53%; Precision: 91.91%	Limited dataset diversity affects generalizability.
Saleh et al. 2021 [30]	Hybrid CNN with SVM	Classification accuracy: 97.91%	Insufficient CT dataset and lack of real clinical validation.
Mahajan et al. 2025 [31]	Hybrid CNN with LSTM	Overall accuracy: 92%; CNN: 88%; LSTM: 85%; Precision: 94%; Recall: 91%; F1: 92%	Unspecified dataset and comparatively low performance.
Mahmud et al. 2025 [32]	Vision Transformer models (LEViT, CoAtNet, CrossViT)	CT accuracy: 99.43%; Histopathology accuracy: 99.02%	Extensive preprocessing limits scalability.

cancerous or non-cancerous; however, the extent to which it can scale remains to be determined.

Temporal modeling has been studied more extensively in recent years. Mahajan et al. [31] proposed a hybrid CNN-LSTM network to first extract the spatial features from each CT slice using CNNs and then capture the temporal relationships between the succession of CT slice images with LSTMs. This hybrid approach was tested on the labeled CT dataset, detecting cancerous lesions very effectively while maintaining a low false-positive rate, demonstrating the benefits of combining spatial and temporal learning features. Mahmud et al. [32] presented a hybrid model based on LEViT architecture for the diagnosis of lung cancer utilizing both the IQ-OTH/NCCD CT dataset and a histopathology dataset containing over 25,000 images. The multi-stage pre-processing and augmentation pipeline created superior results compared with CoAtNet and CrossViT models, producing high classification accuracies: 98.36% MCC for the CT dataset and 97.97% MCC for the histopathological dataset. They also created a real-time application offering Grad-CAM visualizations to improve the interpretability and usability of their tool when being used clinically.

Currently, the accuracy of lung cancer detection using deep learning models may be promising, but they are still heavily reliant on non-clinical datasets, especially because they tend to use a large amount of annotated data and can be expensive to train, meaning that they do not have the ability to be generalized, to be fair, or to be used in clinical contexts. Despite the large number of models evaluated across various datasets, there remains a need for a sufficiently rigorous method to overcome these shortcomings. The proposed hybrid model addresses this gap and shows strong potential for clinical implementation. A summary of key studies, their methods, and performance metrics is presented in Table 1.

3 | Materials and Methods

This research methodology consists of several steps. First, the CT scans of the lungs were obtained from the publicly available LIDC-IDRI database. Once the CT images were obtained, they were preprocessed and assigned to one of three classes (Normal, Benign, Malignant) based on annotations from four independent radiologists. When the dataset was created and stored, each image

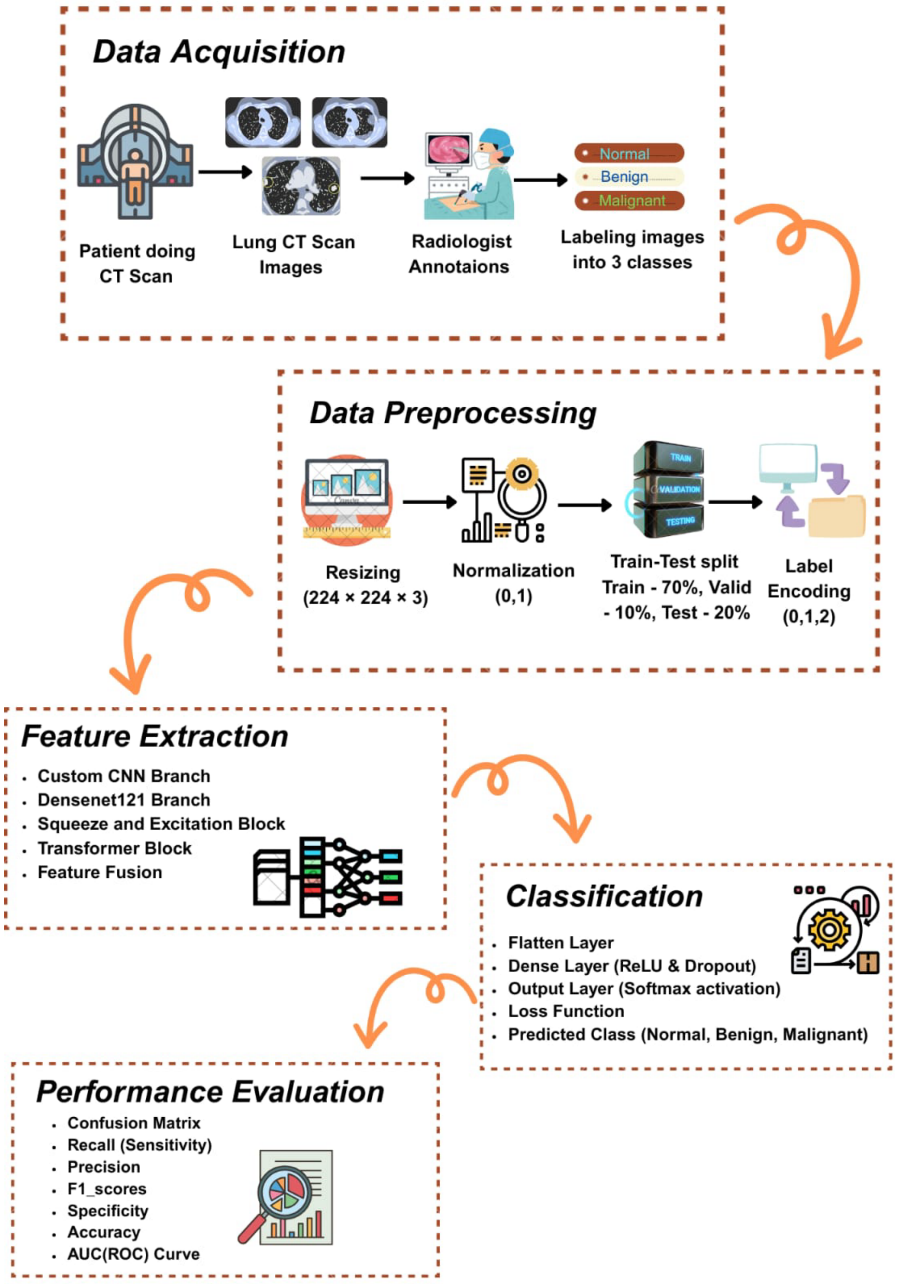

FIGURE 1 | Detailed workflow diagram of the lung cancer classification process.

was resized, normalized, and encoded. After creating the dataset, the custom architecture was used to classify it into three types of lung cancer. Finally, the model's performance was assessed using various metrics. Visualization techniques such as a confusion matrix and ROC(AUC) curve were used to visualize the performance of the model. Figure 1 shows an overview of the steps of the lung cancer detection and classification method.

3.1 | Image Acquisition and Dataset Preparation

In this research, the LIDC-IDRI Database was chosen because it is the most well and safest method for assessing nodule detection algorithms [33]. The LIDC-IDRI database contains 1018 cases from 8 medical imaging institutions, each contributing to the creation and labeling of the database as part of a team collaboration project. Each case also includes an XML file containing information on two stages of image labeling, conducted by four experienced thoracic radiologists per case. Some sample CT images from the dataset are shown in Figure 2.

In the first phase of the annotation process, each radiologist independently examined the CT scans and classified pulmonary findings as either nodules > 3 mm, nodules < = 3 mm, or non-nodules > 3 mm. In the second phase, the radiologists evaluated their own annotations alongside those of other readers, while maintaining an anonymous, unblinded view of their own work. As a result, each nodule was evaluated independently by all four radiologists, resulting in a very good level of agreement among

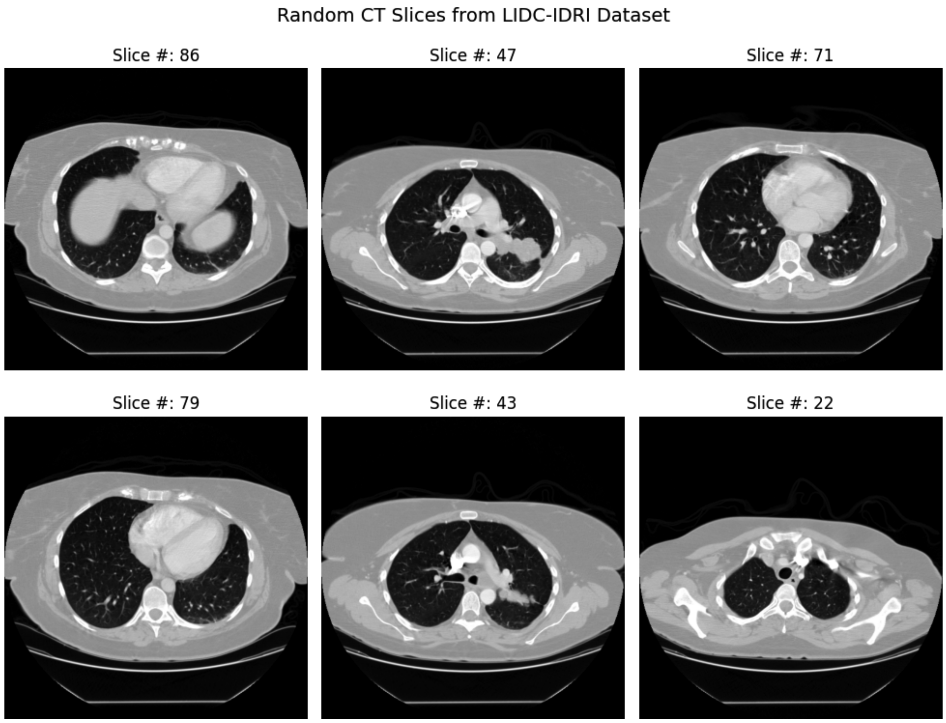

FIGURE 2 | Lung CT slices collected from LIDC-IDRI dataset.

them. The dataset comprised 244,527 CT slices from 1018 CT scans, representing 1010 patients. In our study, the data represent 5292 CT scans taken from 165 patients. The dataset serves as a valuable source of real-world CT image data for training and testing deep learning algorithms for automated lung cancer detection.

3.2 | Image Pre-Processing

The image pre-processing stage eliminates non-relevant distortions and maximally enhances properties (to provide optimal use of data when training a model). The following subsections detail what each of these pre-processing steps entails.

3.2.1 | Image Standardization and Normalization

The purpose of the pre-processing phase is also to represent CT images from a particular subject in a uniform and consistent manner across all individuals for input to deep learning models; this enables the models to be used reliably.

- **DICOM to Hounsfield Unit (HU) conversion:** The raw CT data that has been saved using DICOM (Digital Imaging and Communications in Medicine) protocol is converted to Hounsfield Units (HU), so they can be more clinically relevant for analysis.

$$HU = \text{PixelValue} \times \text{RescaleSlope} + \text{RescaleIntercept} \quad (1)$$

- **Intensity Clipping and Normalization:** To highlight relevant soft tissue and parenchyma in assessment of nodules, the intensity of the Hounsfield Units(HU) is clipped to a Vocal Lung Window or [−1000, 400] HU shown in the Figure 3 [34] values that are outside these values will not be used. After clipping to the Vocal Lung Window, all HU are normalized to the range of [0,1]. Normalization optimizes numerical stability, allowing quicker convergence of the training models.

- **Spatial Resizing and Isotropic Resampling:** Variability in slice thickness (c.f. CT scans: 3D) and resolution in LIDC-IDRI scans results in an anisotropic distribution of voxels. Thus, all volumetric CT data (i.e., a 3D object) is first converted into a uniform isotropic resolution (1 mm3, 1 mm3, 1 mm3, isotropic resolution) [35] through trilinear interpolation, where all three spatial dimensions are treated equally when performing 2D convolution using neural networks. After resampling, each axial slice is then resized (from 512 × 512 pixels to 224 × 224 pixels) through bilinear interpolation.

3.2.2 | Nodule Candidate Identification and Consensus Aggregation

The CT and corresponding XML annotations are used to determine which nodules will be entered into the consensus dataset based on the independent markings and subjective assessments (e.g., the average malignancy score of a particular nodule) of each nodule by four experienced expert radiologists.

- **Lesion Selection and Filtering:** The consensus dataset will only include nodules that have been marked (as "Nodule >= 3 mm") by at least one of the four radiologists. The consensus dataset will not contain any nodules that have an average malignancy score of exactly 3.0 (indeterminate

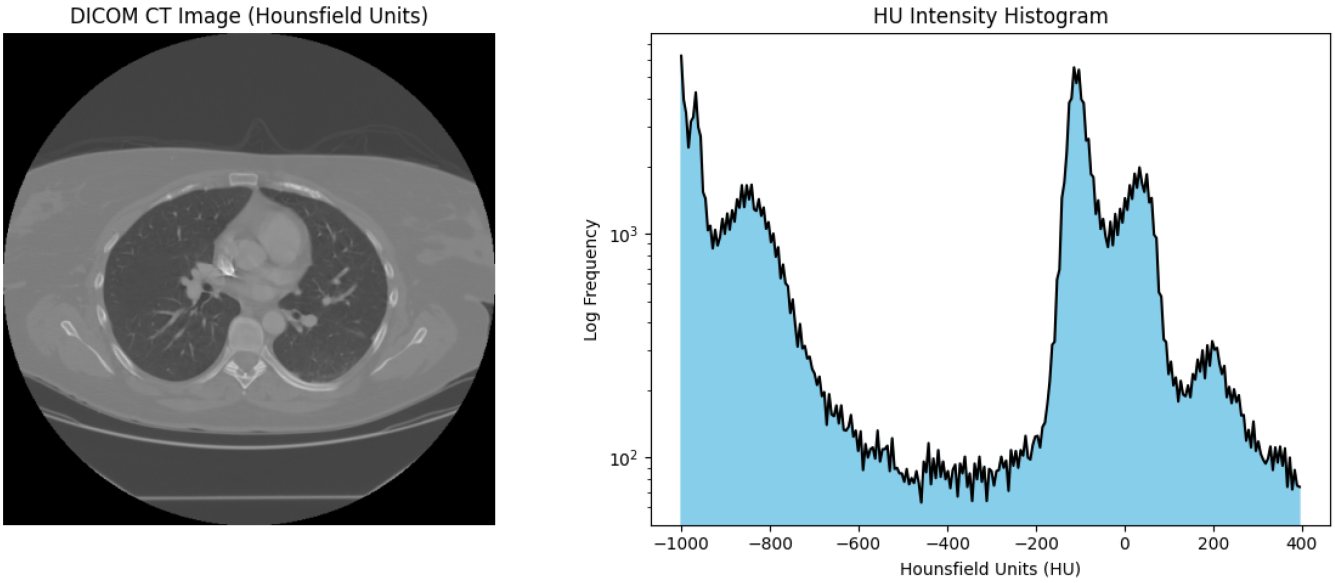

FIGURE 3 | Cancerous lung CT image and histogram of the images.

classes), to ensure clear differentiation (between benign vs. malignant) in the final dataset.

- **Consensus Coordinates:** To produce a solid reference point, the consensus location of a nodule is created by aggregating the annotations from multiple radiologists. The definitive centroid of a nodule is calculated as the median of all readers' coordinates that marked that nodule. By using this method to calculate the coordinates for all of the nodules, inter-observer variability will be minimized, thereby facilitating more reliable localization of the nodules for subsequent processing and classification.

TABLE 2 | Three class level derivation based on radiologist-assigned malignancy scores.

Data source	Average malignancy score ($\overline{S}$)	Derived class	Class ID
Non-Nodule Patch	N/A (Randomly Sampled)	Normal	0
Consensus Nodule Patch	$\overline{S} < 3.0$	Benign	1
Consensus Nodule Patch	$\overline{S} > 3.0$	Malignant	2

3.2.3 | Derivation of the 3-Class Classification Labels

- **Average Malignancy Score Calculation:** To calculate the average malignancy score for each nodule consensus, all four radiologists' malignancy scores (1–5) were retrieved from the consensus, and the scores were averaged. The average score (avgM) of a consensus is:

$$\overline{S} = \frac{1}{N}\sum_{i=1}^{N} S_i \tag{2}$$

where S_i denotes the ith radiologist's malignancy score, and N is the total number of radiologists who scored that nodule. Thus, using the average condemnation score to replace all the individual radiologists' scores reduces the variability of inter-reader and provides a more representative score for that nodule.

- **3-Class Label Assignment:** The samples in this dataset were classified into one of three categories according to their source and average malignancy score. Table 2 illustrates how class levels were assigned based on the four radiologists' assigned malignancy scores. Having only three labels helps keep the distinction among the three classes clear, facilitating effective supervised learning for automated lung cancer classification.

- **Handling Class Imbalance:** Due to the nature of the dataset, there is a substantial imbalance between the number of normal patches and the number of nodule patches, with normal patches outnumbering both benign and malignant nodules.

1. Stratified sampling during dataset splitting to maintain proportional representation of each class.
2. Oversampling of minority classes (Malignant and Normal) to balance the training data. Optionally, the use of cost-sensitive loss functions during model optimization to penalize misclassifications of underrepresented classes more heavily.

The final step of the project involved applying the appropriate class labels to each folder. The classes of the images (Benign, Malignant, Normal) were represented numerically (i.e., as integers) with secure coding. The integer representation of each image was converted to a corresponding one-hot encoded vector representation with only one index being a high (1) value (the

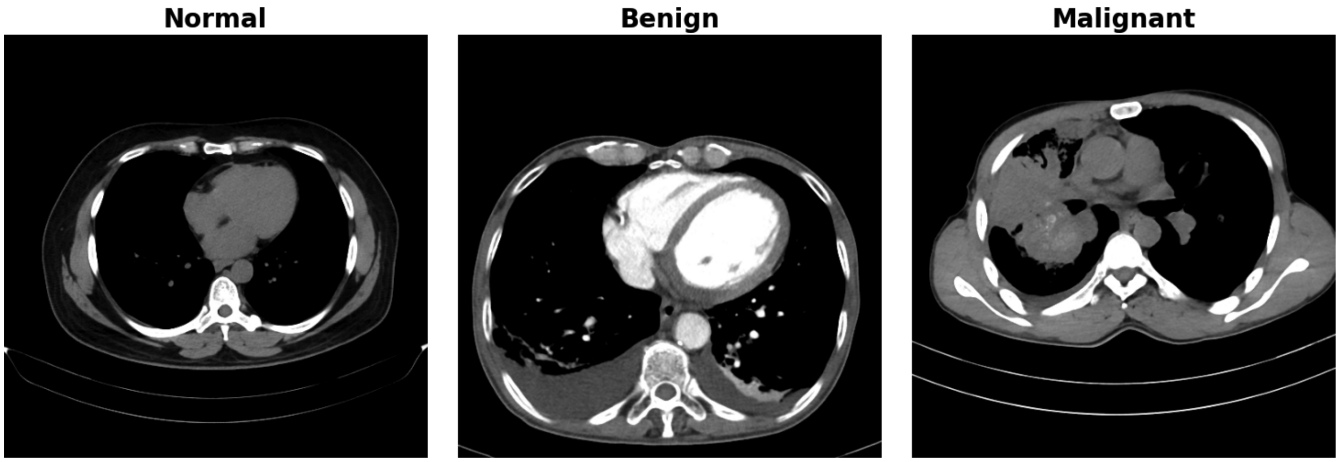

FIGURE 4 | Lung CT scan images after resizing and labeling into Normal, Benign & Malignant.

position of the index that represents the corresponding numerical class), while all other indices had zero values [36]. By using a one-hot-encoded vector for each of the three class labels, the model was able to perform multiclass classification efficiently and calculate categorical cross-entropy loss accurately. The entire data pipeline cleaned, organized, and properly created the data set so it could be processed into CNN models as depicted in Figure 4.

3.3 | Dual Attention Mechanism

Attention mechanisms have become popular nowadays in deep learning-based algorithms. It delivers significant performance improvements by dynamically focusing on the viral parts of the input data. This concept, in complex tasks like medical image analysis, enhances feature representation and contributes to better performance in modeling.

There are two types of attention mechanisms (Squeeze and Excitation (SE) blocks and a custom spatial attention block) used in this research to improve feature learning and classification performance on lung CT images.

3.3.1 | Channel-Wise Attention Blocks

The purpose of Channel Squeeze & Excitation (SE) is to improve channel inter-dependencies of feature maps; This is done through a two-part process, called "squeeze" and "excite." The squeeze operation reduces the spatial dimensions using Global Average Pooling (GAP), in complex tasks like medical image analysis, using a small neural network that learns channel-wise weights. These weights are subsequently utilized by the model to recalculate how much each feature channel contributes towards the overall decision-making process. Thus, the model has been trained on how to focus more intensely on those channels that contribute most towards making an informed decision compared to those channels that do not contribute significantly. In summary, these three components of the SE block provide a mechanism whereby we can accomplish the above goal through several intermediate steps:

1. **Squeeze:** Global average pooling is applied to each channel of the input feature map $X \in \mathbb{R}^{H\times W\times C}$ channel-wise descriptor $Z \in \mathbb{R}^{C}$. This operation summarizes each channel's spatial information into a single representative value that allows it to capture long-range global context while not increasing computational complexity.

$$Z_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{i,j,c}, \quad Z = [Z_1, Z_2, \ldots, Z_C] \tag{3}$$

2. **Excitation:** The channel descriptor Z is passed through two fully connected layers to model inter-channel dependencies and generate channel-wise scaling factors

$$s = \sigma\big(W_2\ \delta(W_1 Z)\big) \tag{4}$$

Here, δ denotes ReLU activation and σ is the sigmoid function. $W_1 \in \mathbb{R}^{C/r\times C}$ reduces the channel dimensionality by a factor r, and $W_2 \in \mathbb{R}^{C\times C/r}$ returns it. Thus, the network has the capability of learning the relative significance of each channel adaptively.

3. **Scale:** The initial feature map is rescaled by the learned channel-wise factors:

$$X_{i,j,c} = X_{i,j,c} \cdot s_c \tag{5}$$

This step emphasizes prominent channels and reduces the less informative channels, thereby increasing the representation power of the network. SE blocks are incorporated at numerous levels within the model, where they are able to improve feature representation. They are applied directly after the first convolutional block to recalibrate the low-level features, after the residual CNN blocks that include the combined features to fine-tune them, and after the additional CNN layers to reinforce the more discriminative patterns prior to merging them with the other outputs from DenseNet121 and Transformer. This multi-level integration of SE blocks in the model allows the model to consolidate, thus enabling the learner to focus and track on the more informative channels and enhance feature extraction from the preceding layers, allowing for more accurate lung cancer classification.

3.3.2 | Spatial Self-Attention Blocks

While CNNs are great for modeling spatial features, they are not very good at modeling longer-range dependencies. To address this shortcoming, we add transformer encoder blocks after the CNN and SE modules to model long-range exercise dependencies using multi-head self-attention based on global spatial relationships across the image. The Transformer block of the proposed hybrid model captures both global spatial dependencies in CNN-extracted feature maps and local features extracted by the CNN and SE blocks. The blocks consist of five stages in total:

1. **Input Reshaping:** The feature map of the 2D CNN, denoted by $X \in \mathbb{R}^{H\times W\times C}$, has been transformed into a sequence $X_{\text{seq}} \in \mathbb{R}^{L\times C}$ where, $L = H \times W$. In this context, H represents height; W represents width, and C represents the number of channels present within the feature map. Each spatial location within the feature space is mapped to a token in the sequence; therefore, the Transformer can learn dependencies across positions rather than relying solely on those around it.

$$X_{\text{seq}} = \text{Reshape}(X) \tag{6}$$

   In this representation, the original 2D spatial information is converted to a sequence of L tokens (each with its channel representation), therefore enabling a self-attention mechanism to capture long-range dependency information and to model global contextual information, which is needed for improved accuracy in classifying lung cancer.

2. **Layer Normalization:** To ensure that training stability is maintained and to allow for greater gradient flow during training, all features are layer normalized such that they have zero mean and one variance. Therefore, the removal of internal co-variation shifts allows for a much quicker convergence process, allowing us to train networks more efficiently.

$$X_{\text{norm}} = \text{LayerNorm}(X_{\text{seq}}) \tag{7}$$

3. **Multi-Head Self-Attention (MHSA):** The model uses a multi-headed mechanism for self-attention to capture long-term dependencies using its normalized feature representations.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{8}$$

   The normalized feature X_{norm} generates Q (query), K (key), and V (values) representations that define what the self-attention will generate, as well as allow us to compute how much attention each head should pay to the information being processed. Each attention head is defined based on d_k equal to the size of the corresponding attention head (d_k), therefore allowing the model to pay attention to many different types of spatial relationships concurrently [37]. In addition, a residual connection preserves the original feature, or representation, of the input, so we can continue using it in the model.

$$X' = X_{\text{seq}} + \text{MHSA}(X_{\text{norm}}) \tag{9}$$

4. **Feed-Forward Network (FFN):** The augmented feature information, having been created using an attention mechanism, passes through 2 fully connected feed-forward networks (dense layers) where the first fully connected layer uses a ReLU activation function and the second does not have an activation function applied. A skip connection is added to preserve the augmented features. The use of a skip connection along with the ReLU in the first layer of this two-layer network enables the model to transform information in non-linear ways, which improves the model's ability to accurately model complex spatial patterns in the feature map.

$$X_{\text{out}} = X' + \text{FFN}\big(\text{LayerNorm}(X')\big) \tag{10}$$

5. **Global Average Pooling:** When the input sequence is processed through the multi-head self-attention component and the feed-forward network, the Transformer outputs a sequence of $X_{\text{out}} \in \mathbb{R}^{L\times C}$, where $L = H \times W$ represents the number of spatial tokens and C is the feature dimension. To convert these sequence features into a single fixed-length vector for classification, we use global average pooling over the sequence dimension.

$$X_{\text{pooled}} = \frac{1}{L}\sum_{i=1}^{L} X_{\text{out},i} \tag{11}$$

   This resulting vector $X_{\text{pooled}} \in \mathbb{R}^{C}$ provides the global features that are the most relevant to the classifier, looking up information from multiple tokens and collapsing their output to a fixed size so it can be fused with CNN and DenseNet121 branches.

   By combining channel-wise attention and spatial self-attention with the input tensor, the model gains enhanced spatial awareness and improved capability to recognize meaningful patterns in medical images [38].

3.4 | Lung Cancer Classification Model Architecture

An advanced and efficient lung cancer diagnostic pipeline, LCDLNet, has been developed using a deep learning framework that combines multiple neural network architectures to enhance diagnostic accuracy. The proposed model is based on CT scan images and employs a multi-branch design combining a custom CNN with residual blocks, DenseNet121, SE blocks, and a Transformer block, shown in Figure 5.

1. **A Custom CNN for Local Feature Extraction:** A custom-designed CNN was used as the foundational branch for extracting low-and mid-level spatial features from CT images. It was made of an 11-layer architecture designed for lung CT scans, which includes residual connections to facilitate feature reuse and gradient flow. The network consists of multiple convolutional layers using small filters (e.g., 3 × 3 kernels), and it also includes batch normalization, ReLU activations, and max pooling layers that decrease the spatial dimensions while retaining important information. Residual blocks are used to avoid the vanishing gradient problem,

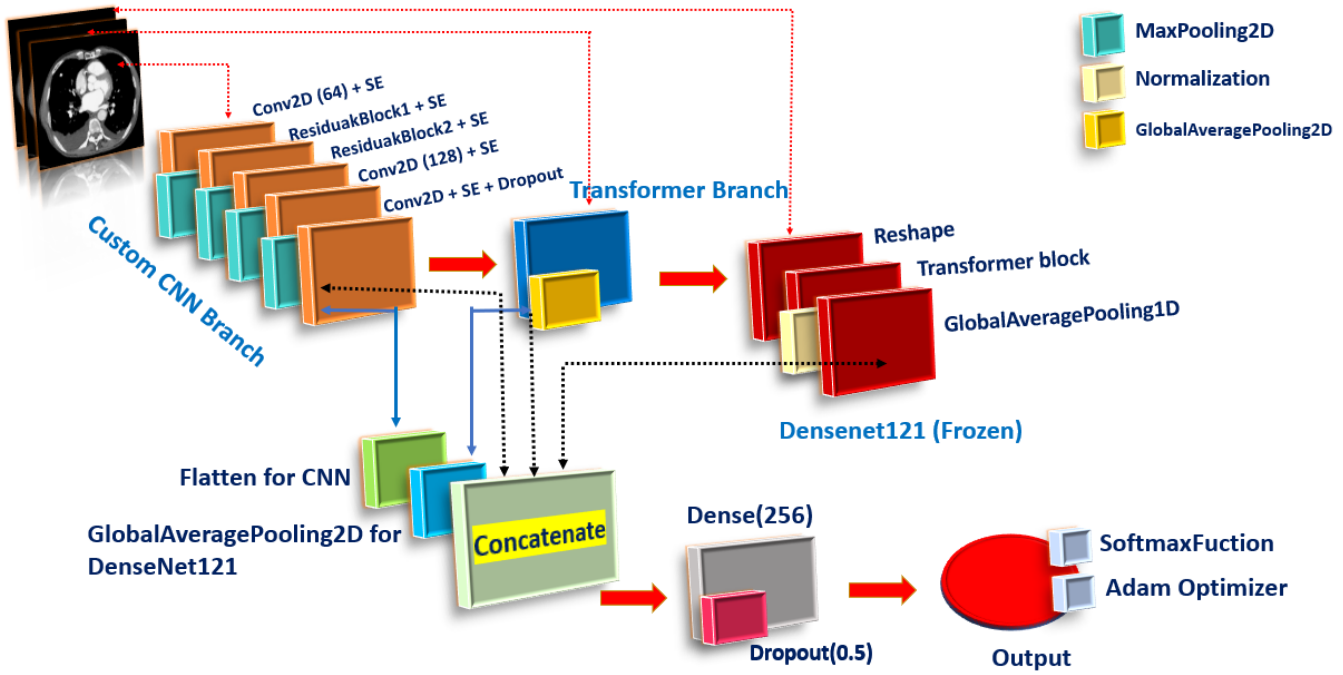

FIGURE 5 | Proposed lung image classification model, LCDLNet architecture.

allowing deeper networks to learn without losing information from previous layers. This branch will provide output as feature maps with spatial characteristics that notice local discriminative patterns like nodule shapes and textures that are relevant for characteristics that decide between benign and malignant lesions.

2. **DenseNet121 for Deep Hierarchical Feature Extraction (Transfer Learning):** In the proposed model, the DenseNet121 model was added as a feature extraction branch to make use of deep semantic knowledge through transfer learning. The pre-trained DenseNet121 model with the given specifications was loaded without the top classification layers (include-top=False), which accepts CT scan images with the specified input shape, but the top classification layers were removed. The extracted feature representations of dense block outputs have significant dimensions and weights and will be categorically frozen (trainable = FALSE), so the dense net layers' weights are preserved and will not change during training. This allows the model to use DenseNet121 as a fixed feature extractor. More specifically, each DenseNet121 layer can contribute abstract, high-level representations of the lung CT images used in this study without incurring additional computational overhead or increasing the risk of overfitting. The outputs of the DenseNet121 model provide state-of-the-art hierarchical visual features, which are then passed to the proposed hybrid network and fused with the outputs of the custom CNN, SE blocks, and Transformer encoders [39]. Dense connectivity layers, such as those used in DenseNet121, where every layer gets feature maps from all others, were chosen to minimize the vanishing gradient problem while also increasing the amount of feature sharing. DenseNet121 is being used for a moderately sized medical imaging dataset and offers a favorable parameter-to-depth ratio compared to deeper backbones like ResNet152 or DenseNet169, thereby avoiding the risk of overfitting. It is also useful in a multi-branch structure due to its hierarchical feature maps, which complement local features learned by the custom CNN branch and the global dependency captured by the Transformer branch.

3. **SE Block:** SE blocks are incorporated into the custom CNN branch to improve channel interactions and to enable adaptive feature response calibrations. Each SE block consists of a squeeze operation (global average pooling to summarize the spatial information for each channel) followed by excitation (a two-layer neural network architecture with a dimension reduction ratio of r = 16, ReLU, and sigmoid), producing scalars for scaling the feature maps. The scalars scale each of the feature maps and highlight the useful information in the channel (e.g., which channel emphasizes cancerous textures), while suppressing potential noise. SE block applications occur at multiple levels post the first series of convolutions for low-level features, within the residual blocks for mid-level features, and before the fusion layer for high-level features. This allows for tuning the model's focus on the discriminative features and improves the classification task.

4. **Transformer Encoders:** The Transformer portion of the model is able to cover long-range dependencies and model global context, which complements the local context modeling of the CNN. The Transformer receives feature maps from earlier CNN layers, flattening them to treat the mapped features as a sequence for processing with multi-head self-attention (MHSA). Important features of the Transformer architecture are layer normalization, MHSA (with queries, keys, and values to calculate attention scores), the residual connection, and a feed-forward network (FFN) with ReLU activations. The last step is

global average pooling, which reduces the sequences to a fixed-length vector. The MHSA mechanism enables the model to learn relationships across distant regions of the lung, such as contextual information suggesting metastasis or distributed zoning information suggesting widespread abnormalities, both of which can be important features for multiclass classification.

3.5 | Feature Fusion and Classification Layers

After feature extraction from the custom CNN with SE blocks, DenseNet121, and Transformer branches, the hybrid model merges the features by concatenating the output from all three branches.

$$X_{\text{combined}} = \text{Concatenate}\left(X_{\text{CNN}},\ X_{\text{DenseNet}},\ X_{\text{Transformer}}\right) \quad (12)$$

where X_{CNN} is the flattened output of the CNN branch with SE blocks, X_{DenseNet} is the global pooled output features from the DenseNet121 branch, and $X_{\text{Transformer}}$ is the pooled output sequence of the Transformer. This fusion enables the model to leverage local features, channel-wise recalibration, and global context simultaneously.

The fused vector is then passed to a flatten layer and converted to a one-dimensional vector for the dense layer. That one-dimensional vector is then input into a fully connected dense layer with 256 neurons with ReLU activation, which allows the model to learn more complex nonlinear combinations of the merged features. To avoid overfitting, a dropout layer with a rate of 0.5 is applied to the dense layer output. Lastly, these feature outputs are input to a softmax for the predicted class probabilities for Normal, Benign, and Malignant.

$$\hat{y}_i = \frac{\exp(z_i)}{\sum_{j=1}^{\text{num\_classes}} \exp(z_j)}, \quad i = 1, \ldots, \text{num\_classes} \quad (13)$$

where the logit for class i is calculated as:

$$z_i = W_{\text{out},i}\ X_{\text{drop}} + b_{\text{out},i}, \quad \hat{y}^i = \text{softmax}(z_i) \quad (14)$$

Here, $\hat{y}^i$ represents the predicted probability for class i.

This architecture—Flatten → Dense (256) → Dropout (0.5) → Softmax—assures that the classifier properly captures meaningful relationships among multiscale features while remaining robust, and is able to accurately classify lung cancer.

3.6 | Model Training

The model used the labeled CT scan dataset for supervised training. The goal of training was to optimize all parameters on the custom CNN, Transformer, and classification layers while leaving DenseNet121 frozen so that the pre-trained features could be retained. Loss function (Categorical cross-entropy loss) was used for multiclass lung cancer classification (Normal, Benign, Malignant). This loss function penalizes incorrect predictions while pushing the network to assign high probability to the correct class.

TABLE 3 | Distribution of dataset across training (116 patients), validation (16 patients), and test (33 patients) sets.

Class	Training	Validation	Test	Total
Benign	1510	216	431	2157
Malignant	1747	250	499	2496
Normal	447	64	128	639
Total	**3704**	**530**	**1058**	**5292**

$$\mathcal{L} = -\sum_{i=1}^{\text{num\_classes}} y_i\ \log(\hat{y}_i) \quad (15)$$

where y_i is the ground truth label and $\hat{y}_i$ is the predicted probability for class i.

The Adam optimizer was used to update the trainable parameters and dynamically compute learning rates through the first and second moments of the gradients. Additionally, learning rate scheduling reduced learning rates on a plateau to allow for more granular convergence. To address overfitting, dropout (0.5) was used in dense layers, batch normalization after convolutional layers, and normal weight initialization to promote stable gradient propagation.

Prevent any data leakage between subsets, the dataset was partitioned at the patient level rather than the slice level. The 165 patients were divided into three disjoint groups: 116 patients (3704 slices) for training (≈70%), 16 patients (530 slices) for validation (≈10%), and 33 patients (1058 slices) for testing (≈20%), as shown in Table 3. Since no patient appears in more than one subset, slices from the same scan cannot leak across splits, providing an unbiased evaluation of the model's generalization performance.

Training was also monitored on validation data, with early stopping ceasing training when there was no improvement to the validation loss, and model checkpointing saving model weights when they were the best performing. The model was trained for 40 epochs with a batch size of 8 to balance computational efficiency with stable gradient propagation. Through these procedures and other strategies to regulate and optimize, the model was able to learn strong, discriminative multi-scale features, resulting in differentiated classifications of lung cancer, and is likely to be generalizable.

Spatial structural changes induced by conventional augmentation operations, such as large rotations, flips, and shears, and intensity perturbations, may alter the spatial characteristics and the densities that radiologists use to differentiate the benign and malignant nodules in CT images, resulting in clinically implausible samples. With patient-level splitting enabled, already eliminating data leakage, overfitting was controlled by dropout (0.5), batch normalization, L2 weight regularization, learning-rate scheduling, and early stopping, and the model converged with many checks in place—with no leakage and a training/validation/test set error of only 99.53%–98.98%–99.81%. Hence, the built-in learning capability of the proposed architecture was deliberately investigated without augmentation.

3.7 | Performance Evaluation

To measure a model's diagnostic capabilities, AUC is typically utilized. An ROC curve shows the trade-off between sensitivity and specificity, while the AUC score indicates how well the entire diagnostic model operates compared to all other models being considered. Confusion matrices indicate how many true/false negative and positive classifications were assigned to each class and can help explain how the model made errors for the classes illustrated in Table 4: Normal, Benign, and Malignant. The five key metrics used for evaluating the performance of a deep learning-based lung cancer classification system (and subsequently, its diagnostic ability and overall efficiency) include Accuracy, Precision, Recall, F1 Score, and Area Under Curve (AUC) [40, 41].

TABLE 4 | Evaluation metrics and mathematical definitions.

Metric		Equation	Notes
Accuracy	=	$\frac{TP+TN}{TP+TN+FP+FN}$	TP denotes True Positives
Precision	=	$\frac{TP}{TP+FP}$	TN denotes True Negatives
Recall (Sensitivity)	=	$\frac{TP}{TP+FN}$	FP denotes False Positives
Specificity	=	$\frac{TN}{TN+FP}$	FN denotes False Negatives
F1-score	=	$\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$	

4 | Results

4.1 | Experimental Setup

In this part of the paper, the results of the experiments are presented using several different performance metrics. All experiments were carried out in Google Collaboratory (more commonly referred to as Google Colab), an open-source cloud platform used for writing and executing Python code for the purposes of research. Each of the experiments was run on a Google Colab machine, which had a NVIDIA Tesla T4 GPU (16GB GDDR6) with a processor (11th Gen Intel(R) Core (TM) i5-1155G7) running at 2.30 GHz, with the ability to overclock to 4.50 GHz.

4.2 | Training and Validation Curves

Researchers create many different models using machine learning. For this current research study, researchers used a variety of machine learning libraries, including Numpy, PyTorch, Scipy, matplotlib, Scikit-learn, Tensor-Flow, and Keras. The total time to run this code was approximately 17 min. Figures 6 and 7 display the results of training and validation of the LCDLNet model represented by training and validation accuracy, and loss from training and validation over 36 epochs. The model converged rapidly and increased its training and validation accuracy significantly between epochs 10–13. During these epochs, the model learns to identify features in CT scan images, as evidenced by the large drop in both training and validation loss. After epoch 13, both training and validation accuracy curves began to become

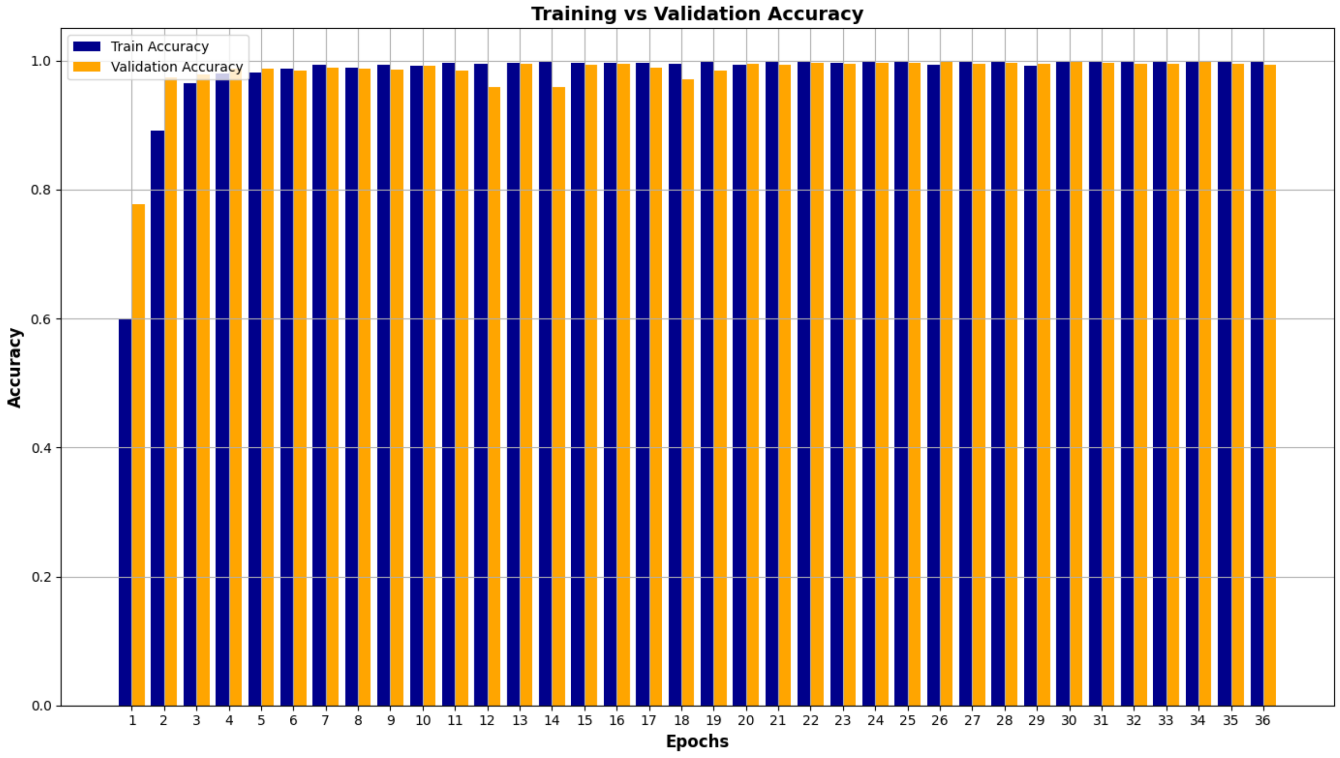

FIGURE 6 | Training and validation accuracy over 36 epochs, demonstrating rapid convergence with minimal overfitting.

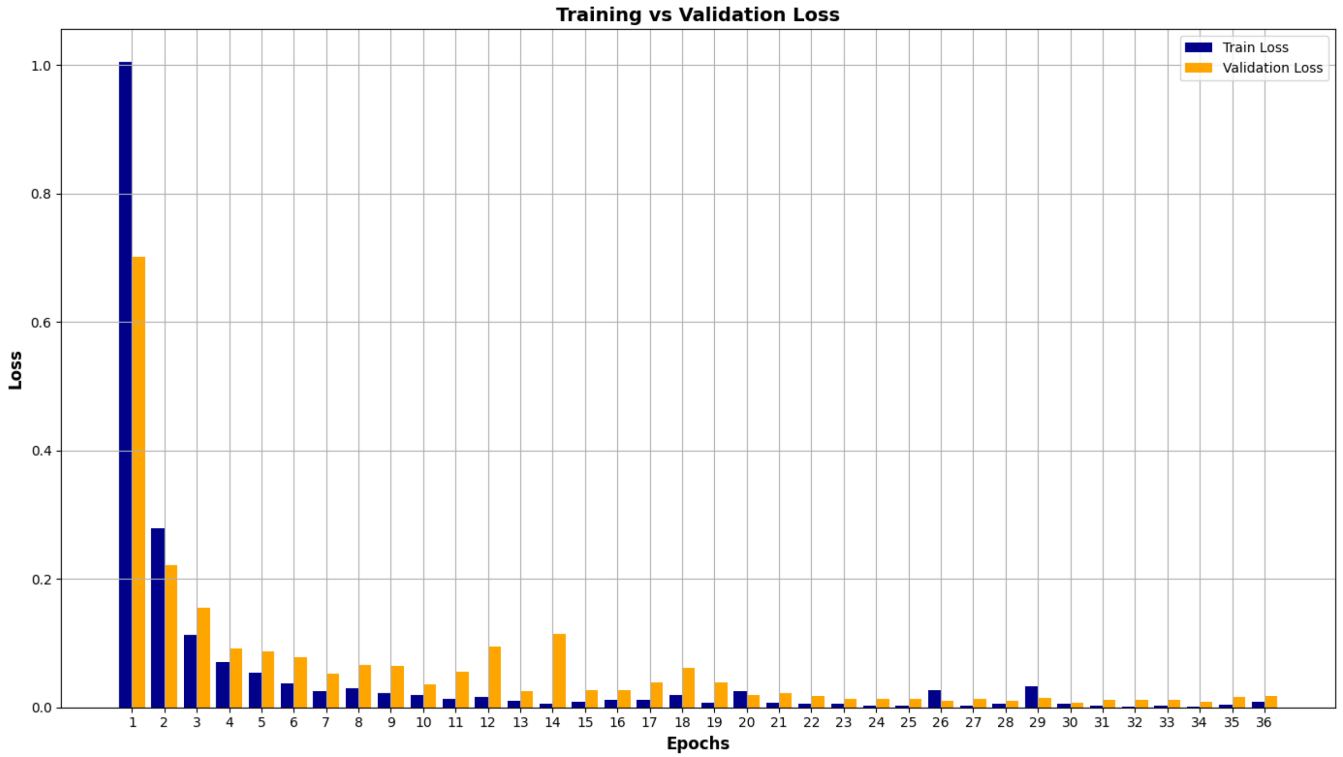

FIGURE 7 | Training and Validation loss of proposed model over 36 epochs.

flat, which was representative of little to no additional improvements seen in either training or validation accuracy curves during the remaining epochs. Using dropout, batch normalization, and early stopping during model training resulted in a more stable convergence of training and validation accuracy curves during model training through 40 epochs, suggesting that the LCDLNet model was able to learn the patterns displayed in CT scan images with limited indications of severe overfitting. Figures 6 and 7 show training-validation accuracy and loss over 36 epochs, respectively.

4.3 | Performance Metrics

The analysis in this study used both the training and testing data, as well as information derived from the testing sample. The Neural Network, the proposed structure for detecting and classifying the data, increases the speed and accuracy of sample detection and identification. The combined effectiveness of the two models improves the system's overall efficacy and results. The simulation results for the proposed LCDLNet model are presented in Figure 8. In addition, the outcomes of the simulation can be evaluated for both the Training and Testing datasets: Accuracy, precision, specificity, sensitivity, F1-score, recall, as well as all of the various metrics used to evaluate accuracy.

The model's excellent classification accuracy demonstrated how well a CNN combined with DenseNet121, SE blocks, and a Transformer branch performs. The precision, recall, and f1-score offer more detail on how the model performed, and its sensitivity to balancing sensitivity and specificity in the identification of each category is shown in Table 5 & Figures 9, 10.

4.4 | ROC Curves

The capability of the CNN+SE+Transformer+DenseNet121 model to classify has been assessed by plotting ROC curves for each class (Benign, Malignant, Normal), illustrated in Figure 11. The AUC score indicates the extent to which a randomly selected positive sample is more readily classified than a randomly selected negative sample. A score of 1.0 represents complete accuracy. The macro-average AUC was about 0.998, and the micro-average AUC was very close to 1.000; this confirms the robustness of the model in classifying different types of lung cancer.

4.5 | Comparative Analysis With Trial Models

4.5.1 | Training and Validation Curves

To measure the performance of our proposed LCDLNet model relative to other techniques, we compared it against two baseline models: A single custom CNN and a combined CNN+Transformer+DenseNet121 model without SE blocks, developed previously. Both trial models were trained and assessed using the same dataset labeled with Normal, Benign, or Malignant diagnoses, and implemented in the same artificial intelligence (AI) scenarios (70-10-20 training-validation-test split, Adam optimizer, and 40 epochs). The outcomes of the comparison are summarized in Table 6.

The training and validation curves, Figure 12, show that all models converged in a stable manner. With validation accuracies well

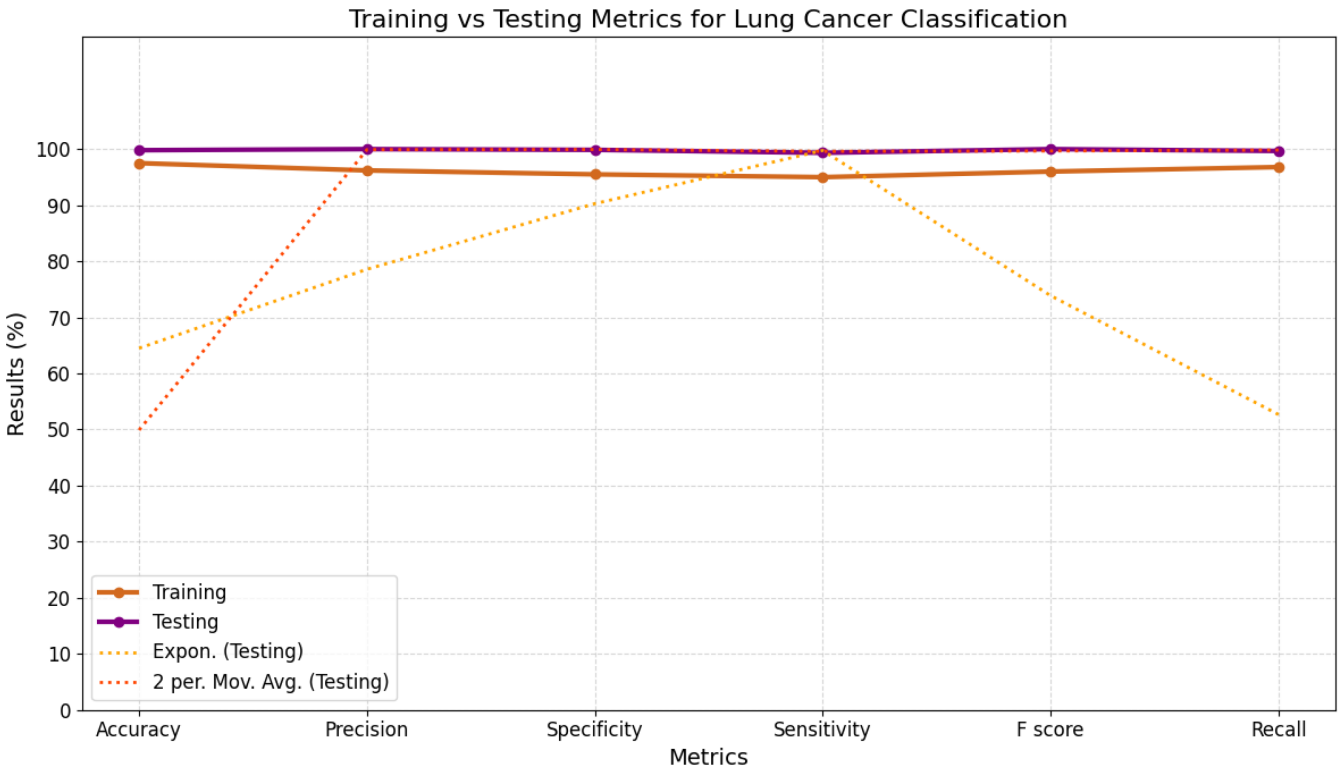

FIGURE 8 | Simulation based outcomes from the proposed work.

TABLE 5 | Classification report for the proposed LCDLNet model.

Class	Precision	Recall (Sensitivity)	F1-Score	Specificity	Support
Benign	0.9977	1.0000	0.9988	0.9984	431
Malignant	0.9980	0.9980	0.9980	0.9982	499
Normal	1.0000	0.9922	0.9961	1.0000	128
Macro Avg	0.9986	0.9967	0.9976	0.9989	1058
Weighted Avg	0.9985	0.9981	0.9983	0.9985	1058

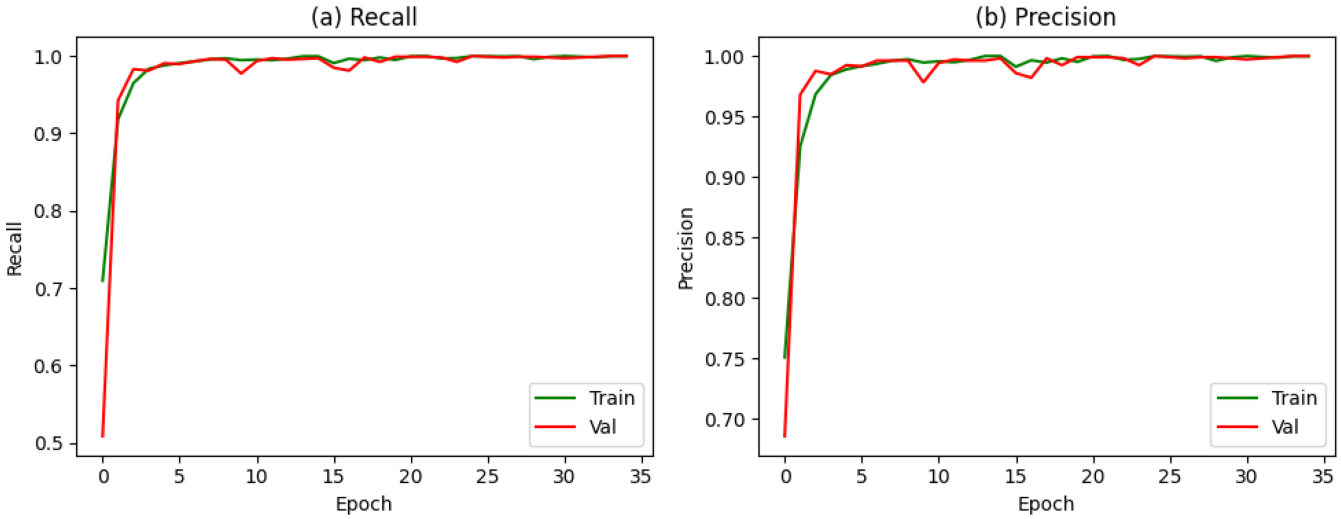

FIGURE 9 | Performance metrics of the proposed architecture over 36 epochs, including (a) precision, (b) recall.

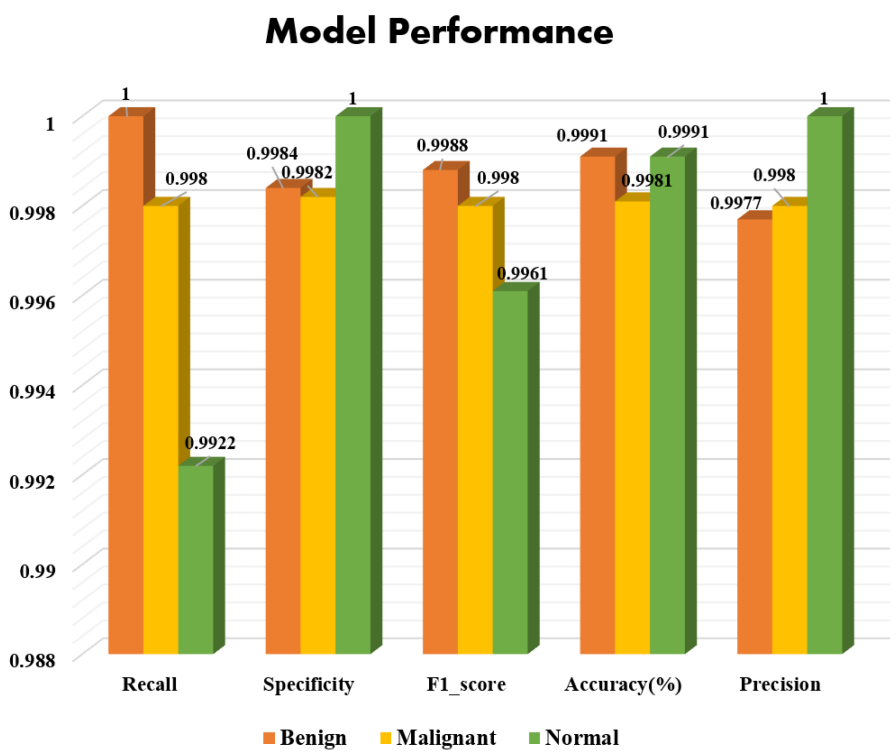

FIGURE 10 | Graphical representation of LCDLNet model performance.

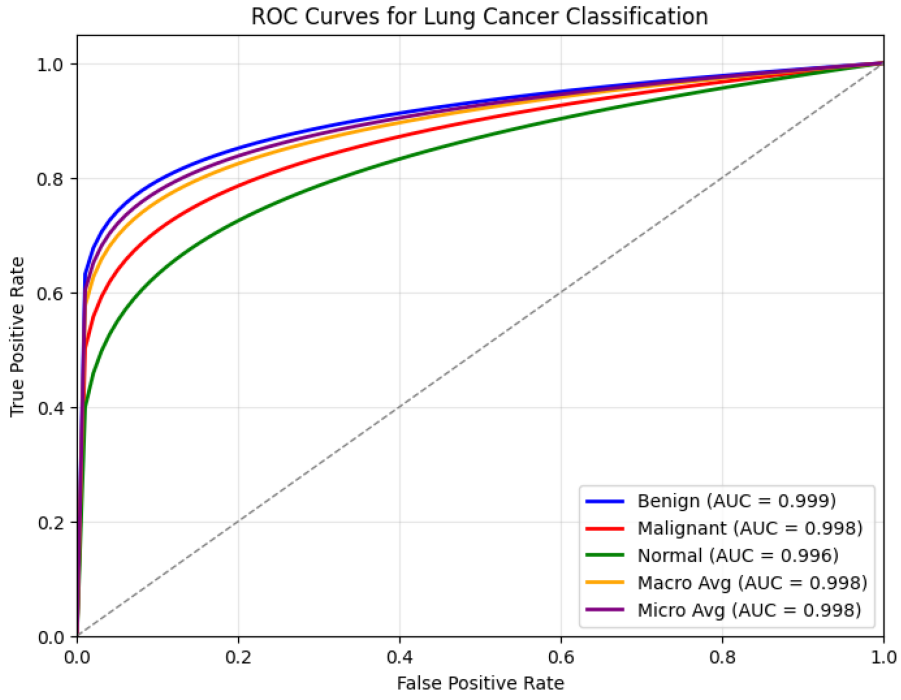

FIGURE 11 | Receiver operating characteristic (ROC) curves for lung cancer classification.

TABLE 6 | Comparative evaluation of lung cancer classification models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Custom CNN	99.28%	0.9923	0.9949	0.9921	0.993
CNN + Transformer + DenseNet121	99.72%	0.9956	0.9963	0.9953	0.996
CNN + SE + Transformer + DenseNet121	99.81%	0.9986	0.9967	0.9976	0.998

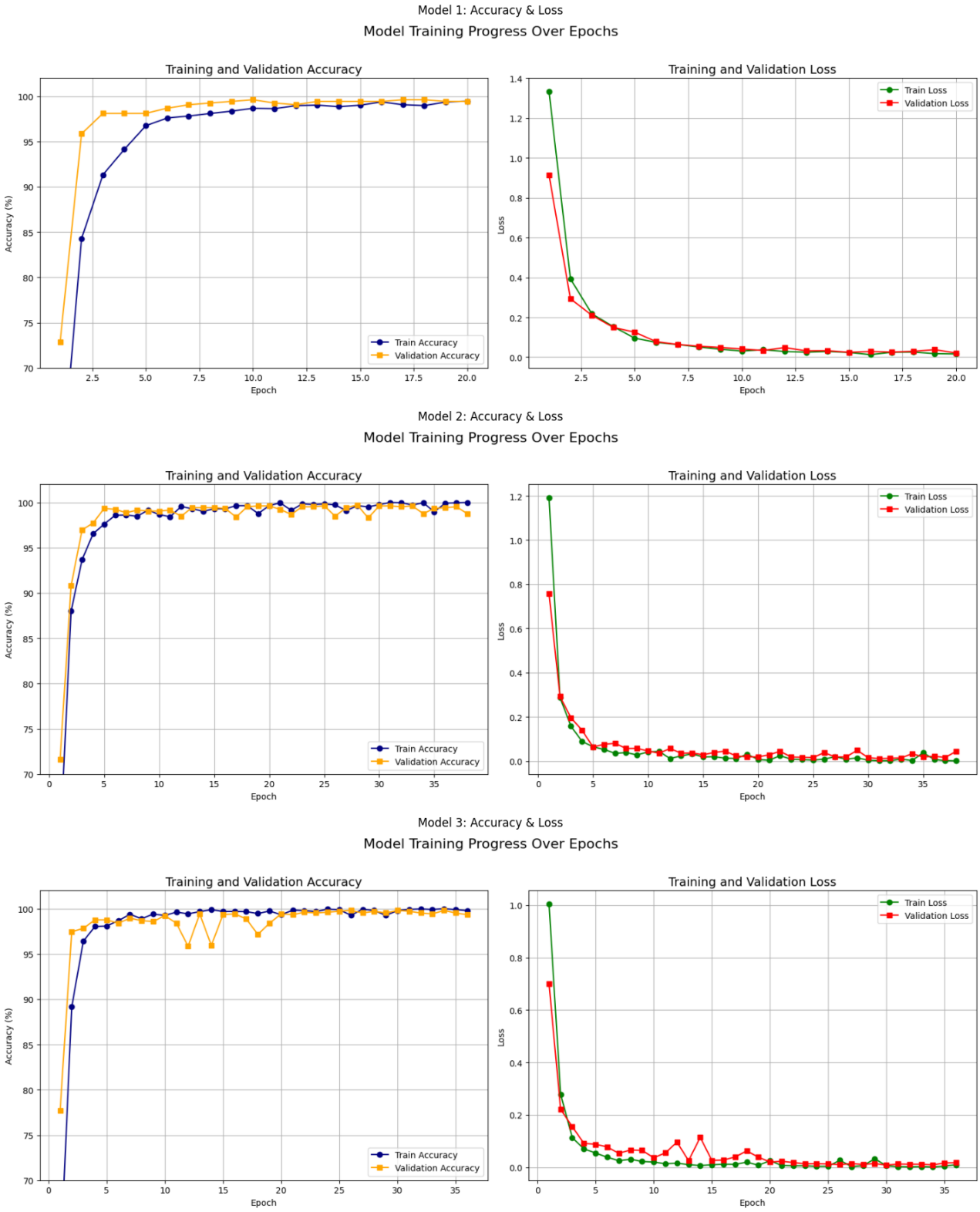

FIGURE 12 | Training and validation accuracy curve & training and validation loss curve of model 1, model 2 & model 3.

above 97% achieved in a few epochs, the hybrid model had better, more stable convergence than the other baseline models, demonstrating very little evidence of overfitting, and similar trends were shown in the validation loss across the 36 epochs.

4.5.2 | Confusion Matrix Analysis

The confusion matrices can provide even more meaningful details about model performance. The custom CNN

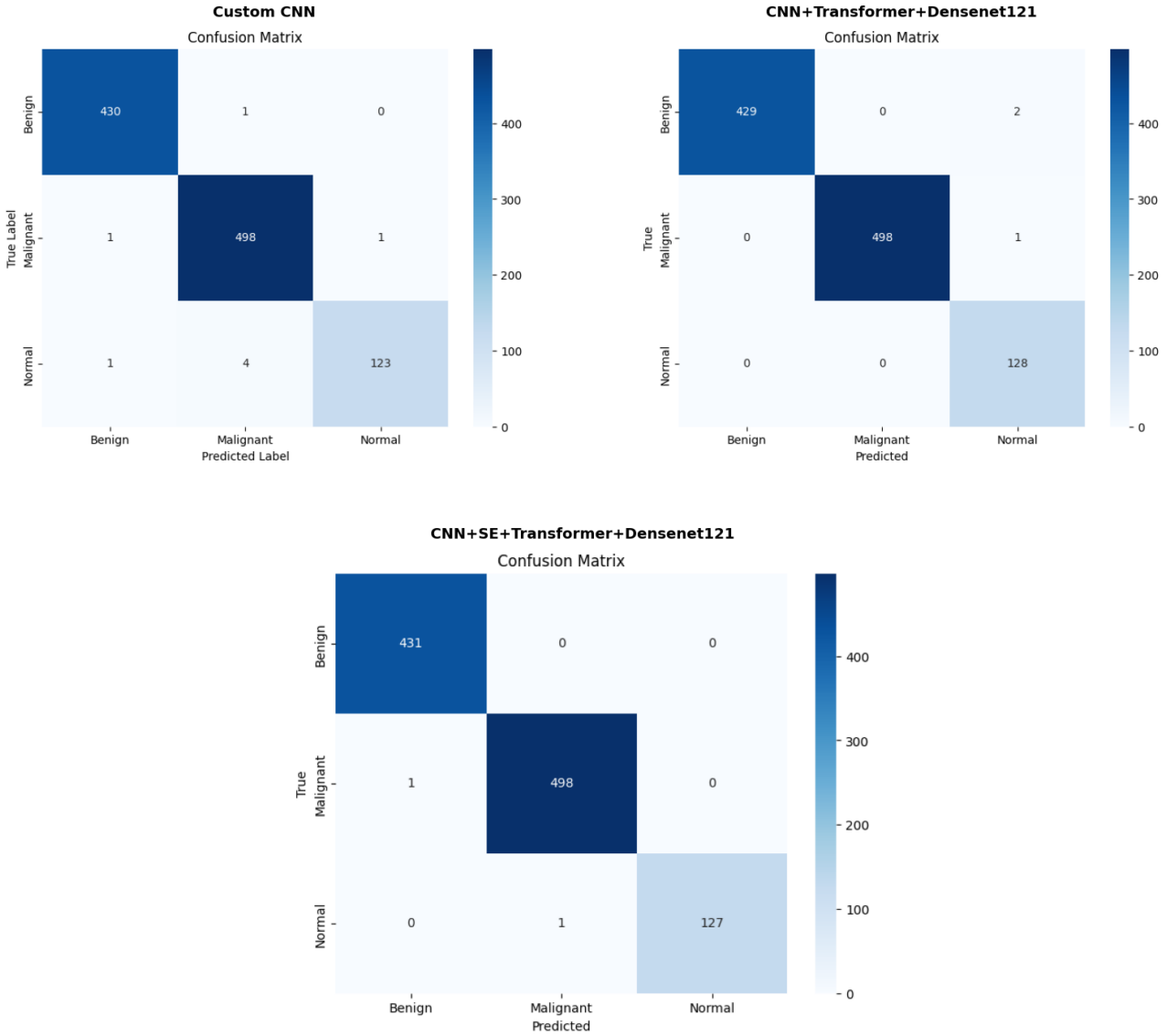

FIGURE 13 | Confusion matrix of model 1, model 2 & model 3.

categorization model performed well overall but had some issues with misclassifying several Normal cases as Benign or Malignant cases, which may indicate an issue with fine-grained discrimination of cases. The CNN+Transformer+DenseNet121 model leveraged improved contextual learning and performed slightly better, but still showed some misclassifications. The LCDLNet model with SE blocks achieved almost perfect classification, correctly identifying Benign, Malignant, and Normal in almost all cases. To be more specific, Figure 13 shows that the proposed model correctly classified 431/431 Benign, 498/499 Malignant, and 127/128 Normal, which resulted in macro-average precision, recall, and F1-score of 99.86%, 99.67%, and 99.76%. In Figure 14, we can see the comparison of the three trial models' performance.

4.5.3 | Comparison With Standard CNN Baselines on LIDC-IDRI Dataset

Four common baseline architectures (AlexNet, VGG16, InceptionV3, DenseNet) were implemented and trained on the same LIDC-IDRI subset, using the same split of the patients (70/10/20), the same preprocessing pipeline (HU clipping, intensity normalization, isotropic resampling, 224×224 resize), categorical cross-entropy as a loss function, the same optimizer (Adam), the same batch size (8), and the same number of epochs with early-stopping (40). The only architectural change was the input adaptation layer required for each backbone. Table 7 shows a summary of the performance comparison of the proposed LCDLNet with standard CNN baseline models on the LIDC-IDRI dataset.

5 | Discussion

Multiple deep learning models have been developed to classify lung cancer using publicly available datasets such as LIDC-IDRI, LUNA16, IQ-OTH/NCCD, and Data Science Bowl. Most of these models are able to achieve high accuracy, but their accuracy can often be limited due to the poor quality or representativeness of public datasets. The proposed LCDLNet model was constructed into 3 different parts and used for multiclass lung cancer

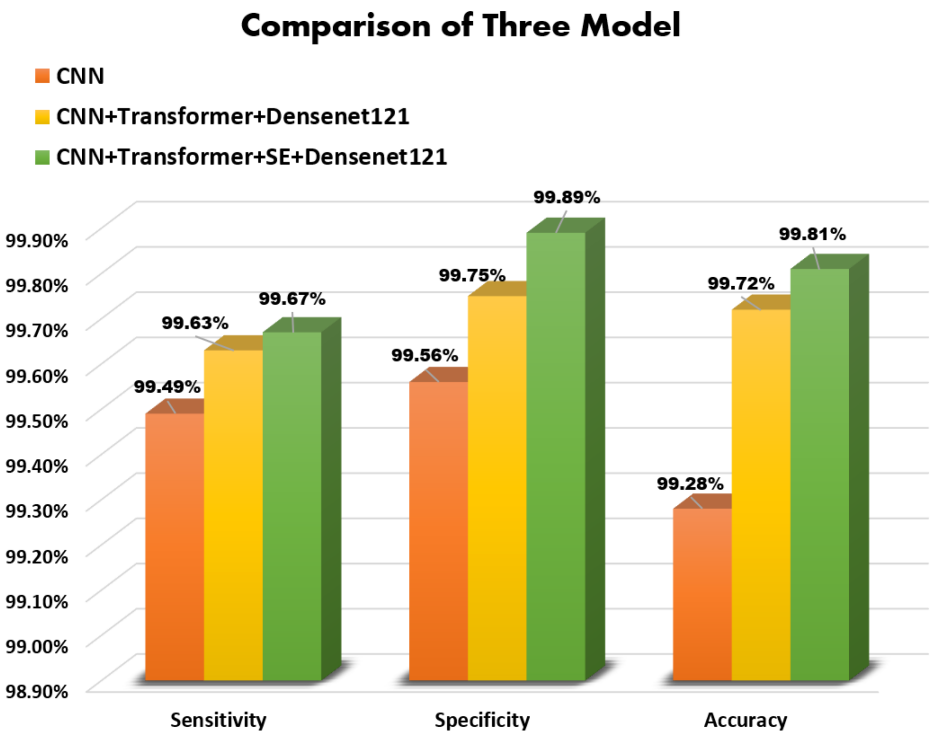

FIGURE 14 | Graphical representation of comparative performance evaluation of three models.

TABLE 7 | Performance comparison of the proposed LCDLNet with standard CNN baseline models on LIDC-IDRI dataset.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)
InceptionV3 [42]	94.64	94.73	95.16	95.57
VGG16 [43]	95.20	95.25	93.40	95.91
AlexNet [44]	96.34	96.62	96.30	96.60
DenseNet [41]	96.34	96.34	96.51	96.61
LCDLNet (Proposed)	**99.81**	**99.86**	**99.67**	**99.89**

classification with a dataset of 5292 CT slices, which were categorized into Benign, Malignant, and Normal classes. All model segments were trained and tested on the same appropriately resized 224 × 224 CT slices; this segmentation included benchmarks for accuracy, F1-score, confusion matrix, and ROC-AUC curves.

By combining a custom CNN that incorporates residual block structures with incrementally more complex networks such as DenseNet121 and Transformer with Squeeze & Excitation (SE) block structures (which induce the network to extract and combine information both locally and globally from each image) will have the capability of finding both local and global features within each CT scan image, allow for deeper networks in general to find progressively deeper features and enhance gradient flow during network training phases. The advantages of residual learning and DenseNet121 style connectivity arrangements are that these will allow the residual block structure to process information from long distances within the CT scan images, where lung tumors are located. Our experimental results yielded very strong evidence of accuracy based on the evaluation metrics for our hybrid model configurations yielded during the training phases at 99.53%, 98.98% for validation, and 99.81% during testing. All of the evaluation metrics used for the assessment of our hybrid model configurations provide clear evidence of very low levels of overfitting and good generalized performance. In Table 8, we present side by side the LCDLNet model architecture and design compared with other models previously published on publicly available data sets. The accuracy results demonstrate that although the model has limitations associated with working from a data perspective, the model supports superior accuracy and, therefore, clinical use and reliability when compared with the other models reviewed.

Under matched experimental conditions, LCDLNet improves accuracy by 3.47%–5.17% over the four baselines, while delivering the highest precision (99.86%), sensitivity (99.67%), and specificity (99.89%). The multi-scale receptive field architecture (InceptionV3) resulted in the lowest accuracy (94.64%), indicating that slight variations between features within benign, malignant, and normal patches are harder for single-branch architectures to detect. While both AlexNet (96.34%) and DenseNet (96.34%) achieved similar accuracy, the fixed receptive field and explicit channel- and spatial-attention mechanisms hindered their ability to readily learn from both local nodule texture and long-range contextual cues. The results prove that the benefits of LCDLNet are not just due to the differences in training protocols but also due to the principled integration of the four modules used during the network's architecture: Residual local feature extraction, dense hierarchical features (DenseNet121), channel-wise recalibration (SE blocks), and global self-attention (Transformer).

TABLE 8 | Comparison of the LCDLNet with major deep learning models for lung cancer classification on major datasets.

Author(s)	Dataset	Architecture	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	AUC
Agarwal et al. 2021 [44]	—	AlexNet CNN	95.60	95.10	94.80	96.00	—
Zargar et al. 2023 [43]	IQ-OTH/NCCD	VGG16	91.00	—	92.08	—	93.0
Meena et al. 2023 [42]	LUNA16 (CT)	Inception V3	94.98	95.00	—	—	—
Shuvo et al. 2024 [45]	LUNA16 (CT)	AutoLungDx: 3D Res-U-Net + YOLOv5 + ViT	93.57	90.95	93.10	92.01	—
Saha et al. 2024 [46]	Chest CT (Kaggle)	VER-Net: VGG19 + EfficientNetB0 + ResNet101	91.00	92.00	91.00	91.30	—
Hossain et al. 2024 [47]	IQ-OTH/NCCD	LeNet-LSTM: CNN with LSTM	99.27	99.12	99.10	99.26	—
Ozdemir et al. 2025 [48]	IQ-OTH/NCCD	Attention-Enhanced Inception-NeXt: CNN + ViT + InceptionNeXt	99.54	99.67	99.60	99.12	—
Fu et al. 2025 [49]	COVID-19 CXR	Lung-MaxViT: CNN + MaxViT with SE blocks	96.80	—	—	96.70	98.3
Proposed Work	**LIDC-IDRI**	**LCDLNet: Custom CNN + DenseNet121 + Transformer + SE + Spatial Attn.**	**99.81**	**99.86**	**99.67**	**99.76**	**99.8**

6 | Conclusion

Utilizing LCDLNet, a deep learning framework, we have demonstrated that we can produce an accurate lung cancer classification using CT scan images and the accuracy, sensitivity, and predictive value of our method via the use of our model combination. This framework successfully classifies lung cancer CT images due to convolutional layers, dense connectivity, attention modules, channel recalibration, a mix of a custom-designed CNN with residual blocks, DenseNet121, transformer blocks, and SE modules, which incorporate local and global features. The model with DenseNet121 was able to classify CT images into 3 classes (normal, benign, and malignant) with a 99.81% classification accuracy, demonstrating that this model was able to learn with generalization, even with a substantially small dataset. DenseNet121 provides a terrific feature representation that successfully minimizes misclassification while maximizing model performance.

While the framework currently lacks data augmentation or interpretability measures like Grad-CAM, its successful generalization on real clinical data confirms clinical relevance, typically more difficult than developing on benchmark datasets. Future research could focus on establishing the model on multi-center, large-scale datasets, including 3D volumetric CT scans, multi-modal data integration, and explainable AI methods to reduce uncertainty in diagnostic reliability and improve diagnostic adoption in practice. The connection between advanced deep learning architectures and the complexities of clinical imaging has been established in this project. Moreover, this project represents a significant development in creating a practical, dependable, and interpretable AI tool to help clinicians diagnose lung cancer.

Author Contributions

Farjana Akter Chumki: conceptualization, methodology, investigation, writing – original draft, formal analysis, data curation, software, validation, visualization. **Tahasin Ahmed Fahim:** software, investigation, validation, supervision, visualization, writing – review and editing, methodology, resources. **Sabrina Alam:** writing – review and editing, supervision, resources, project administration, funding acquisition, validation, conceptualization.

Acknowledgments

This study is funded by the Research and Publication Cell, University of Chittagong, Bangladesh. Research grant number is 885/2024-25/31/008. The authors also acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health, and their critical role in the creation of the free, publicly available LIDC/IDRI Database used in this study.

Funding

This work was supported by the Research and Publication Cell, University of Chittagong (Grant No. 885/2024–25/31/008).

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in LIDC-IDRI at https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX.

Peer Review

For transparency, the peer review documents associated with this article are available at https://doi.org/10.1002/eng2.70856.

References

1. Mount Elizabeth Hospitals, "Lung Cancer - Symptoms & Causes," (2025).

2. American Cancer Society, "Lung Cancer Screening," (2023).

3. R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer Statistics, 2023," *CA: A Cancer Journal for Clinicians* 73, no. 1 (2023): 17–48.

4. American Cancer Society, "Lung Cancer Early Detection, Diagnosis, and Staging," (2025).

5. Centers for Disease Control and Prevention, "What are the Risk Factors for Lung Cancer?," (2025).

6. International Agency for Research on Cancer, "Globocan 2022: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2022," (2022).

7. J. Zhou, X. Ying, J. Liu, L. Feng, Y. Jinming, and D. Chen, "Global Burden of Lung Cancer in 2022 and Projections to 2050: Incidence and Mortality Estimates From Globocan," *Cancer Epidemiology* 93 (2024): 102693.

8. International Agency for Research on Cancer, "Cancer Tomorrow: Future Cancer Incidence and Mortality Worldwide (Based on Globocan 2022 Estimates)," (2022).

9. SEER, "Cancer Stat Facts: Lung and Bronchus Cancer," (2024).

10. National Health Service, "Lung Cancer – Diagnosis," (2022).

11. Johns Hopkins Medicine, "Lung Cancer Diagnosis," (2024).

12. F. Chutivanidchayakul, T. Suwatanapongched, and T. Petnak, "Clinical and Chest Radiographic Features of Missed Lung Cancer and Their Association With Patient Outcomes," *Clinical Imaging* 99 (2023): 73–81.

13. A. Smith, "Challenges in CT-Guided Lung Biopsy for Early Detection," *Radiology Today* 24, no. 2 (2024): 45–51.

14. M. Gheisari, G. Wang, and M. Z. A. Bhuiyan, "A Survey on Deep Learning in Big Data," in *Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 2 (IEEE, 2017), 173–180.

15. M. Gheisari, F. Ebrahimzadeh, M. Rahimi, et al., "Deep Learning: Applications, Architectures, Models, Tools, and Frameworks: A Comprehensive Survey," *CAAI Transactions on Intelligence Technology* 8, no. 3 (2023): 581–606.

16. A. Traverso, E. L. Torres, M. Fantacci, and P. Cerello, "Computer-Aided Detection Systems to Improve Lung Cancer Early Diagnosis: State-Of-The-Art and Challenges," *Journal of Physics: Conference Series* 841 (2017): 12013.

17. R. Yuan, P. M. Vos, and P. L. Cooperberg, "Computer-Aided Detection in Screening CT for Pulmonary Nodules," *American Journal of Roentgenology* 186, no. 5 (2006): 1280–1287.

18. M. D. Ramasamy, R. K. Dhanaraj, S. K. Pani, et al., "An Improved Deep Convolutionary Neural Network for Bone Marrow Cancer Detection Using Image Processing," *Informatics in Medicine Unlocked* 38 (2023): 101233.

19. A. Kumar, Z. Yuqing, C. P. Gandhi, R. Kumar, and J. Xiang, "Bearing Defect Size Assessment Using Wavelet Transform Based Deep Convolutional Neural Network (DCNN)," *Alexandria Engineering Journal* 59, no. 2 (2020): 999–1012.

20. V. Nasteski, "An Overview of the Supervised Machine Learning Methods," *Horizons. B* 4, no. 51–62 (2017): 56.

21. N. Faruqui, M. A. Yousuf, A. K. M. Whaiduzzaman, A. B. Azad, and M. A. Moni, "Lungnet: A Hybrid Deep-CNN Model for Lung Cancer Diagnosis Using Ct and Wearable Sensor-Based Medical Iot Data," *Computers in Biology and Medicine* 139 (2021): 104961.

22. G. Zhang, L. Lin, and J. Wang, "Lung Nodule Classification in Ct Images Using 3d Densenet," *Journal of Physics: Conference Series* 1827 (2021): 12155.

23. R. Raza, F. Zulfiqar, M. O. Khan, et al., "Lung-Effnet: Lung Cancer Classification Using Efficientnet From Ct-Scan Images," *Engineering Applications of Artificial Intelligence* 126 (2023): 106902.

24. N. Gautam, A. Basu, and R. Sarkar, "Lung Cancer Detection From Thoracic Ct Scans Using an Ensemble of Deep Learning Models," *Neural Computing and Applications* 36, no. 5 (2024): 2459–2477.

25. R. Tandon, S. Agrawal, A. Chang, and S. S. Band, "Vcnet: Hybrid Deep Learning Model for Detection and Classification of Lung Carcinoma Using Chest Radiographs," *Frontiers in Public Health* 10 (2022): 894920.

26. A. R. Bushara, R. S. Vinod Kumar, and S. S. Kumar, "Lcd-Capsule Network for the Detection and Classification of Lung Cancer on Computed Tomography Images," *Multimedia Tools and Applications* 82, no. 24 (2023): 37573–37592.

27. Y. Chen, J. Feng, J. Liu, B. Pang, D. Cao, and C. Li, "Detection and Classification of Lung Cancer Cells Using Swin Transformer," *Journal of Cancer Therapy* 13, no. 7 (2022): 464–475.

28. W. Jian, A. U. Haq, N. Afzal, et al., "Developing an Innovative Lung Cancer Detection Model for Accurate Diagnosis in Ai Healthcare Systems," *Scientific Reports* 15, no. 1 (2025): 22945.

29. H. Polat and H. D. Mehr, "Classification of Pulmonary Ct Images by Using Hybrid 3d-Deep Convolutional Neural Network Architecture," *Applied Sciences* 9, no. 5 (2019): 940.

30. A. Y. Saleh, C. K. Chin, V.'n. Penshie, and H. R. H. Al-Absi, "Lung Cancer Medical Images Classification Using Hybrid Cnn-Svm," (2021).

31. M. Mahajan, K. Malik, A. Singh, B. Sharma, A. Bansal, and S. Mehta, "Hybrid Cnn-Lstm Networks for Enhanced Lung cancer Detection and Classification From Ct Images," in *2025 International Conference on Networks and Cryptology (NETCRYPT)* (IEEE, 2025), 1726–1731.

32. M. R. Mahmud, M. Hasib Fardin, I. H. Siddiqui, A. H. Sakib, and A. Al Sakib, "Hybrid Deep Learning for Interpretable Lung cancer Recognition Across Computed Tomography and Histopathological Imaging Modalities," *International Journal of Science and Research Archive [Internet]* 15, no. 1 (2025): 1798–1810.

33. G. Samuel, I. I. I. Armato, G. McLennan, et al., "The Lung Image Database Consortium (Lidc) and Image Database Resource Initiative (Idri): A Completed Reference Database of Lung Nodules on Ct Scans," *Medical Physics* 38, no. 2 (2011): 915–931.

34. T. Sasaki, R. Miyata, Y. Hatai, K. Makita, and K. Tsunoda, "Hounsfield Unit Values of Retropharyngeal Abscess-Like Lesions Seen in Kawasaki Disease," *Acta Oto-Laryngologica* 134, no. 4 (2014): 437–440.

35. M. Wei Shen, F. Y. Zhou, Y. Dongdong, et al., "Multi-Crop Convolutional Neural Networks for Lung Nodule Malignancy Suspiciousness Classification," *Pattern Recognition* 61 (2017): 663–673.

36. V. V. Starovoitov and Y. I. Golub. "Data Normalization in Machine Learning," *Informatics* 18, no. 3 (2021): 83–96.

37. Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, and L. Wang, "Vision Transformers for Image Classification: A Comparative Survey," *Technologies* 13, no. 1 (2025): 32.

38. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), 7132–7141.

39. S. B. Thunuguntla, S. Murugaanandam, and R. Pitchai, "Densenet121-Dnn-Based Hybrid Approach for Advertisement Classification and User Identification," *International Journal of Intelligent Engineering & Systems* 16, no. 3 (2023): 162–174.

40. T. A. Fahim, F. B. Alam, and M. A. Hossain, "Brain Tumor Detection, Classification and Segmentation by Deep Learning Models From Mri Images: Recent Approaches, Challenges and Future Directions," *Array* 28 (2025): 100571.

41. F. B. Alam, T. A. Fahim, M. Asef, A. Hossain, and M. Ali Akber Dewan, "Wgcamnet: Wasserstein Generative Adversarial Network Augmented and Custom Attention Mechanism Based Deep Neural Network for Enhanced Brain Tumor Detection and Classification," *Information* 15, no. 9 (2024): 560.

42. S. Meena, A. Kumar, M. Sood, and R. K. Meena, "Lung Cancer Detection and Classification Model Using Inception v3 Algorithm," in *International Conference on Data Analytics & Management* (Springer, 2023), 423–433.

43. H. H. Zargar, S. H. Zargar, R. Mehri, and F. Tajidini, "Using vgg16 Algorithms for Classification of Lung Cancer in CT Scans Image," (2023). *arXiv preprint arXiv:2305.18367*.

44. A. Agarwal and K. Patni, "Lung cancer Detection and Classification Based on Alexnet Cnn," in *Proceedings of the 2021 6th International Conference on Communication and Electronics Systems (ICCES)* (IEEE, 2021), 1390–1397.

45. S. B. Shuvo and T. B. Mamun, "AutoLungDx: A Hybrid Deep Learning Approach for Early Lung Cancer Diagnosis Using 3D Res-U-Net, YOLOv5, and Vision Transformers," *Informatics in Medicine Unlocked* 61 (2026): 101739.

46. A. Saha, S. M. Ganie, P. K. D. Pramanik, R. K. Yadav, S. Mallik, and Z. Zhao, "Ver-Net: A Hybrid Transfer Learning Model for Lung Cancer Detection Using Ct Scan Images," *BMC Medical Imaging* 24, no. 1 (2024): 120.

47. M. I. Hossain, S. Ahmad, M. Sakib Hasan, A. Rahim, and J. Shin, "Lung Cancer Detection and Classification Using Lenet-Lstm Model on Computed Tomography Images," *Multidisciplinary Science Journal* 7, no. 4 (2025): 2025188.

48. B. Ozdemir, E. Aslan, and I. Pacal, "Attention Enhanced Inception-next Based Hybrid Deep Learning Model for Lung Cancer Detection," *IEEE Access* 13 (2025): 27050–27069.

49. F. Xiaoyang, R. Lin, D. Wei, A. Tavares, and Y. Liang, "Explainable Hybrid Transformer for Multi-Classification of Lung Disease Using Chest x-Rays," *Scientific Reports* 15, no. 1 (2025): 6650.