ORIGINAL ARTICLE

Improving Early Detection and Diagnosis of Lung Cancer Using an Enhanced Ensemble Deep Learning Model on CT Images

Subba Rao Dusari *, **Nagendra Panini Challa**

School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India

*Corresponding Author: Subba Rao Dusari

Email: subbaraodusari77@gmail.com

Received: 13 December 2024 / Accepted: 24 May 2025

Abstract

Purpose: Lung cancer is the most common and deadliest type of cancer that is the cause of one million deaths around the world every year. Due to the present level of medical research, identifying lung tumors on chest Computed Tomography (CT) images has become a significant process in modern medicine. Enhancing treatment and reducing lung cancer mortality can be achieved by promptly identifying and accurately diagnosing suspected malignant lung tumors. While many deep learning algorithms have been developed recently for the classification of lung cancer, achieving high accuracy in lung cancer classification is still a challenge. An advanced deep learning technique is designed to boost the effectiveness of early lung cancer diagnosis.

Materials and Methods: In this research, we have proposed an enhanced ensemble deep-learning model for lung cancer classification and segmentation. Initially, we carried out an extensive pre-processing process, including image resizing, noise reduction, and contrast enhancement, to enhance the image quality. The problem of small sample size is addressed by applying conventional data augmentation techniques like flipping, rotating, zooming, and shearing. Next, seven statistical features are retrieved using Improved Empirical Wavelet Transforms (IEWT). After feature extraction, the Enhanced ResNeXt model is used to classify lung cancer into normal, malignant, and benign classes. The interested region of the lung tumor is segmented using the Modified ShuffleNetV2 model. Individuals are classified into normal, malignant, and benign categories based on the presence of lung cancer. The experiments are performed on the benchmark datasets LIDC-IDRI and IQ-OTH/NCCD.

Results: For lung cancer classification, the proposed Enhanced ResNeXt model achieves an exceptional model accuracy of 99.43% for the IQ-OTH/NCCD dataset and 99.37% for the LIDC-IDRI dataset. Furthermore, the proposed Modified ShuffleNetV2 model effectively segments lung tumor regions, achieving an Intersection over Union (IOU) of 98.43% and a Dice Similarity Coefficient (DSC) of 97.24% on the LIDC-IDRI dataset, and an IOU of 97.05% and DSC of 96.23% on the IQ-OTH/NCCD dataset, demonstrating its robustness and accuracy in delineating tumor boundaries. The expected outcomes show that the accuracy and efficiency of our proposed ensemble deep learning model outperform other CNNs.

Conclusion: The proposed models beat existing CNN-based models in terms of speed and number of training parameters, which means that using CT scan images to diagnose lung cancer automatically is a suitable option and a strong selection for extensive use in medical environments.

Keywords: Lung Cancer; Computed Tomography; Data Augmentation; Enhanced ResNeXt; Modified ShuffleNetV2.

DOI:

Frontiers in BIOMEDICAL TECHNOLOGIES

1. Introduction

Globally, lung cancer continues to be one of the deadliest and frequently diagnosed malignancies, responsible for more deaths annually than breast, colon, and prostate cancers combined [1]. It originates in the tissues of the lungs, typically in the cells lining the air passages. Major risk factors include cigarette smoking, exposure to environmental carcinogens such as asbestos and radon, and genetic predisposition [2]. The prognosis for lung cancer is highly dependent on the stage at which it is diagnosed. Early detection is critical yet remains challenging due to the asymptomatic nature of early-stage tumors [3]. As a result, the majority of cases are diagnosed at later stages when treatment options are limited and less effective [4]. Improving early detection strategies is, therefore, a crucial step toward reducing mortality and improving patient quality of life [5].

Medical imaging, particularly Computed Tomography (CT), plays a vital role in lung cancer screening and diagnosis [6]. However, the manual interpretation of CT scans is a time-consuming and error-prone task that heavily relies on the expertise of radiologists [7]. To assist clinicians and reduce diagnostic variability, recent years have seen the rise of Computer-Aided Diagnostic (CAD) systems as useful devices [8]. These systems utilize Artificial Intelligence (AI), especially machine learning and deep learning algorithms, to automatically interpret medical images and support early detection efforts [9]. CAD systems not only reduce the burden on radiologists but also improve the overall accuracy and consistency of lung cancer diagnoses [10].

Among various deep learning techniques, medical imaging tasks have shown remarkable performance from Convolutional Neural Networks (CNNs), including classification, object detection, and segmentation [11]. CNNs can automatically extract hierarchical features from images, making them particularly well-suited for recognizing complex patterns in lung CT scans [12, 13]. However, one of the challenges in training deep CNNs from scratch is the need for extensive annotated data and high computational resources [14]. To overcome these limitations, transfer learning has become an effective solution. It leverages pre-trained models that have learned general visual features from large datasets and applies them to domain-specific tasks, significantly reducing the training time and data requirements [15, 16].

In addition to transfer learning, ensemble learning techniques have gained attention in medical image analysis due to their ability to combine the strengths of multiple models [17]. By aggregating predictions from diverse classifiers, ensemble methods enhance robustness, reduce variance, and improve generalization performance [18]. Techniques such as bagging, boosting, and stacking are commonly used to build ensemble models capable of handling noisy and heterogeneous medical data [19, 20].

Motivated by these insights, this research proposes an enhanced ensemble deep learning framework for lung cancer classification and segmentation using CT images. The proposed methodology integrates a modified ResNeXt model for multi-class classification of lung cancer into normal, malignant, and benign categories and a Modified ShuffleNetV2 model for accurate segmentation of the tumor region. ResNeXt strikes a balance between complexity and accuracy, outperforming traditional CNN models in feature representation, especially in medical imaging scenarios with subtle variations across classes. ShuffleNetV2, on the other hand, is selected for segmentation due to its lightweight design and computational efficiency, which makes it appropriate for tasks involving medical imaging in real-time. Unlike heavier models like U-Net, the Modified ShuffleNetV2 achieves high segmentation accuracy while maintaining low computational overhead, an essential factor for deployment in clinical environments. Experimental evaluation is conducted on two benchmark datasets, including LIDC-IDRI and IQ-OTH/NCCD, to demonstrate the effectiveness and generalizability of the proposed approach.

By combining advanced deep learning architectures, transfer learning, statistical feature extraction, and ensemble learning strategies, this research aims to deliver a high-accuracy, reliable framework for lung cancer detection and segmentation that can assist clinicians in early diagnosis and improve patient care.

1.1. The Contribution of the Proposed Research

- We developed a novel, improved ensemble deep learning model for binary-class and multi-class lung cancer classification.
- The pre-processing approach uses an adaptive median filtering-based noise removal technique, and image contrast is enhanced using contrast-limited adaptive histogram equalization (CLAHE).
- By applying conventional data augmentation techniques, including rotating, zooming, flipping, and shearing, we have resolved the significant imbalance problem to overcome the skewness of the data.
- The seven statistical features, Kurtosis, Standard Deviation, Average Amplitude, Energy, Entropy, Variance, and mean, are extracted based on Improved Empirical Wavelet Transforms (IEWT).
- An Enhanced ResNeXt model is used in the proposed research to classify lung cancer into normal, malignant, and benign classes.
- To segment the cancerous region of the lung, the Modified ShuffleNetV2 model is used, which incorporates the attention mechanism to extract the spatial features of the images.
- The Adam optimizer is employed to optimize model weights during training. The hyperparameters were tuned separately using a grid search strategy to determine the optimal values.

1.2. Organization of the Research

The following is the structure of the research article: The previous literature publications on lung cancer diagnosis and their problem statement are examined in Section 1. The proposed methodology for segmenting and classifying lung cancer is explained in Section 2. The experimental results and a performance comparison of state-of-the-art methods are presented in Section 3. The discussion of the proposed research is explained in Section 4, and Section 5 presents the research's conclusions.

1.3. Related Prior Works

Recent advancements in artificial intelligence have significantly impacted lung cancer diagnosis, with deep learning-based approaches offering promising accuracy in classification and segmentation tasks. Various hybrid and ensemble techniques have been developed to enhance model robustness and adaptability.

Gupta *et al.* [21] proposed a UDCT framework that integrates a Differentiable Architecture Search (DARTS) model for multi-scale feature extraction and a modified U-Net for lung cancer segmentation using CT scans. Architectural search optimization enables efficient gradient-based tuning, while multilevel Otsu thresholding is applied as pre-processing for nodule separation.

Mahesh *et al.* [22] introduced a dual-method approach combining a Markov Random Field-based AHCA algorithm for lung lobe segmentation with a ShuffleNet model trained using an Adaptive Hunger Foraging Optimization (AHFO) algorithm for classification. The AHA mechanism within this framework mimics real-time foraging behavior to refine segmentation and classification results iteratively.

To capture temporal features, Mostafa *et al.* [23] employed a combination of CNN, LSTM, and Bi-LSTM models. CNN was used for spatial feature extraction from CT images, while LSTM and Bi-LSTM models enhanced classification by modelling temporal dependencies in feature sequences.

Patnaik *et al.* [24] developed Mask-EffNet, a hybrid architecture combining Masked Auto-encoder for Distribution Estimation (MADE) with a variant of EfficientNet (EfficientNetB7). This two-stage pipeline leverages MADE for dimensionality reduction and uncertainty estimation, followed by deep transfer learning with EfficientNet for final classification.

Gugulothu *et al.* [25] presented a hybrid deep-learning framework for lung cancer diagnosis using CT images. The CBSO technique was used for segmentation, while the Improved Fish Bee algorithm performed optimized feature selection, an enhanced deep ensemble neural network (HDE-NN) that predicted tumor types based on selected features.

Venkatesan *et al.* [26] proposed WPHT-DLBP, a lung cancer diagnosis framework that incorporates adaptive median filtering for pre-processing, feature extraction using Wavelet Packet Histogram Texture (WPHT) and Discrete Local Binary Pattern (DLBP), and feature selection using the Arithmetic Hunger Herd Optimization (AHHO) technique. The classification was carried out using an optimized Support Vector Machine (OSVM) with parameters fine-tuned by the IW-BS algorithm.

These diverse approaches demonstrate the potential of hybrid architectures and intelligent optimization strategies for lung cancer detection. However, challenges remain in ensuring generalizability, scalability, and interpretability in clinical environments.

1.4. Problem Statement

Despite the significant progress in AI-assisted lung cancer diagnosis, several limitations persist in existing approaches. Many current models are trained on datasets with limited diversity in terms of scanner types, image acquisition parameters, and patient demographics, leading to restricted generalizability across different clinical settings. Furthermore, most models are designed for static analysis and cannot perform longitudinal assessments of lung nodules—an essential aspect of clinical decision-making.

Traditional diagnostic methods remain invasive, time-consuming, and susceptible to human error. While deep learning models have improved performance in image-based diagnostics, they often suffer from scalability issues and lack robustness when applied to real-world datasets with varying conditions. Additionally, some methods rely heavily on handcrafted features or are computationally intensive, limiting their practicality in time-sensitive clinical scenarios.

To address these gaps, the proposed research introduces an enhanced ensemble deep learning framework for the classification and segmentation of lung cancer using CT images. The framework integrates a modified ResNeXt model for robust multi-class classification and a modified ShuffleNetV2 model for efficient and accurate segmentation. By incorporating statistical feature extraction using the IEWT, a diverse data pre-processing strategy, and conventional augmentation techniques, the model is trained on varied datasets capturing different stages of lung cancer. This ensures higher adaptability, early detection accuracy, and support for individualized treatment planning.

2. Materials and Methods

The block diagram of the proposed lung cancer diagnosis system is depicted in Figure 1. There are five stages: data collection, image pre-processing, feature extraction, lung cancer classification, and lung cancer segmentation. In the pre-processing stage, adaptive median filtering is used to reduce the noise content successfully. Image improvement with Contrast Limited Adaptive Histogram Equalization (CLAHE) further enhances the clarity of noiseless images. Conventional data augmentation approaches are extensively employed to increase the dataset size and address the issue of class imbalance. The IEWT is used to extract the seven statistical features, including Kurtosis, Standard Deviation, Average Amplitude, energy, entropy, variance, and mean. Then, the extracted features are sent to the Enhanced ResNeXt model to classify the normal, malignant, and benign classes. To separate the cancer tissue from the classified images, segmentation is performed using the Modified ShuffleNetV2 model. The Bottleneck Convolutional Block Attention Module (BCABM) is incorporated here to enhance the segmentation accuracy. The proposed model is evaluated and analyzed according to several performance measures and is contrasted with cutting-edge methods to demonstrate its promise for lung cancer diagnosis.

2.1. Dataset Description

In this research, we have used two popular publicly available datasets for lung cancer classification and segmentation. The LIDC-IDRI dataset images are used for binary classification, including benign and malignant classes. The IQ-OTH/NCCD dataset images are used for multi-class classification, including benign, malignant, and normal.

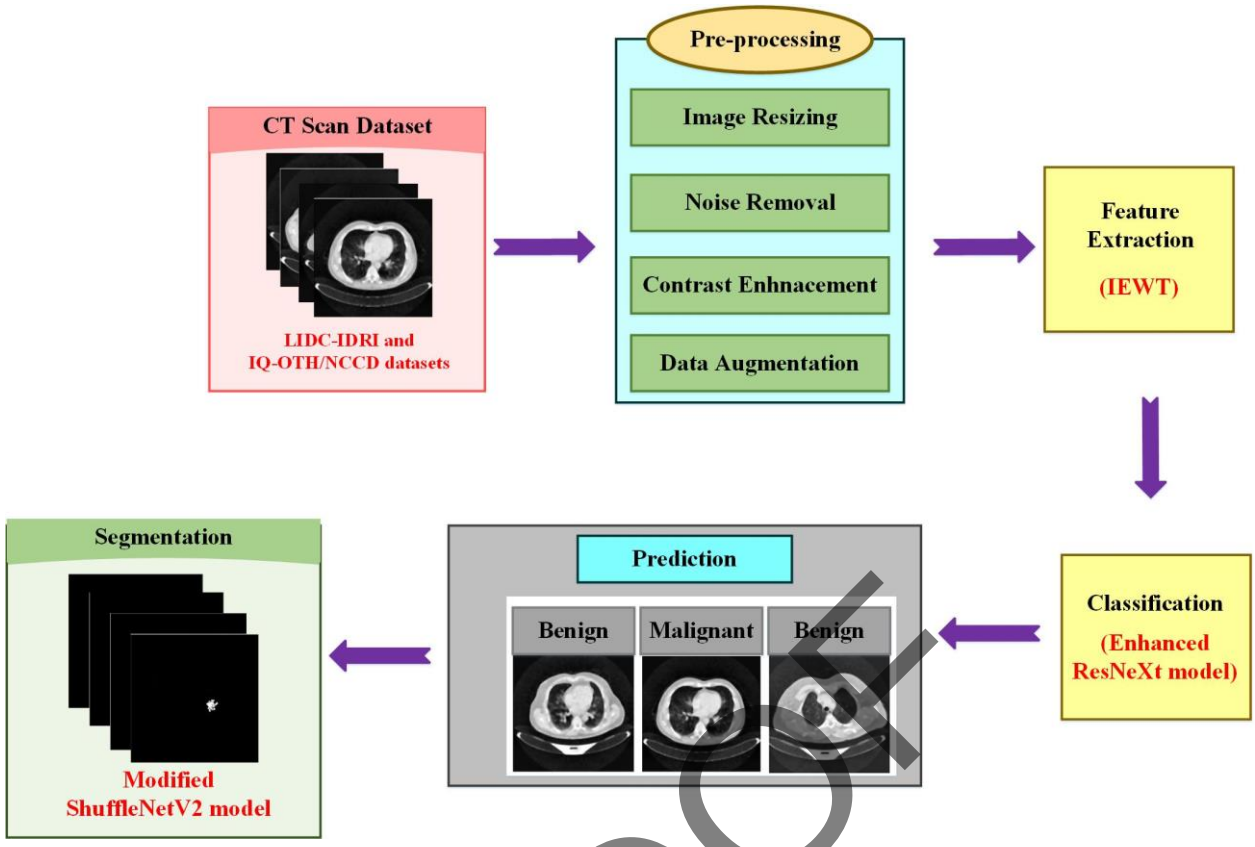

Figure 1. The architecture of the proposed enhanced ensemble deep learning model for lung cancer diagnosis

2.1.1. LIDC-IDRI Dataset

The 1018 cases in total (the CT image data file type is Dicom) with accompanying diagnostic results are included in the LIDC-IDRI dataset, which was gathered at the National Cancer Institute's American effort. Four experienced chest radiologists annotated each case's images. On a range of 1 to 5, four medical professionals assessed each nodule's malignancy; higher scores denoted more malignancy. In total, there are 1051 malignant classes and 1015 benign classes. The dataset images are taken from the Kaggle website (https://www.kaggle.com/datasets/busharakmea/lung-cancer-detection-lidc-idri-subset).

2.1.2. IQ-OTH/NCCD Dataset

The National Center for Cancer Diseases is the source of the dataset used in this research. Three months were spent gathering the data in the fall of 2019. In these two centers, radiologists and oncologists marked IQ-OTH/NCCD slides. In all, 1190 images from 110 cases' CT scan slices are included in the collection. The three groups of these cases are malignant, benign, and normal. 55 of these have been classed as malignant cases, 15 as benign cases, and 40 as normal cases. Initially, the DICOM format was used to gather the CT scans. The Siemens SOMATOM scanner is being utilized. The original CT image format was DICOM, and the scanner used was Siemens SOMATOM. The collection includes 416 images of normal cases, 561 images of malignant patients, and 120 images of benign cases overall. The dataset images are taken from the Kaggle website (https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset).

2.2. Data Pre-Processing

Two primary pre-processing processes are used in this research. Initially, the dataset's images were 512x512 pixels in size; these images were resized to 224x224 pixels. Next, adaptive median filtering is used to eliminate noise from the original images, and the CLAHE algorithm is applied to improve the contrast. Figure 2 shows the sample images after performing pre-processing, such as noise removal and contrast enhancement. After pre-processing, the dataset is balanced by applying conventional data augmentation techniques.

2.2.1. Noise Removing

One noise reduction method that works very well on "impulse noise" or "salt-and-pepper noise" in CT images is adaptive median filtering. Across different image regions, its primary benefit over normal median filtering resides in its responsiveness to the variable noise circumstances. It detects and eliminates noise in the image while maintaining important edge details. Although the borders of traditional median filters may be blurry, the adaptive technique modifies its filtering window according to local conditions. When an appropriate region is located or the noise is separated, the filter dynamically increases the window's (kernel's) size. This helps avoid excessively smoothing out minute details, such as edges, which is significant for CT scans because tiny anatomical information might be critical to a diagnosis. Sharp edges and small structures in the image are intended to be preserved by adaptive median filtering, as traditional noise reduction techniques frequently impact these features.

2.2.2. Contrast Enhancement

Before processing, image-enhancing techniques were applied to the source images to improve their appearance and information value. CLAHE was utilized to increase image visibility. The improved CLAHE component was a part of the Adaptive Histogram Equalization (AHE) approach. Using this method, the transformation function was determined, and the boosting function was applied to all surrounding pixels. CLAHE has less contrast than AHE because of the contrast amplification limit. Rather than using CLAHE on the entire image, CLAHE was applied to discrete data sections termed tiles to improve the contrast of the image. The created adjacent tiles were then successfully fused back together using bilinear interpolation. The "clip-limit" function was used to decrease image noise. After creating it, the histogram was cropped to the grey-level mapping. The average amount of gray pixels in the contextual region was determined by dividing each grey level in the pixel numbers equally:

$$n_{avg} = \frac{n_{CR-x_p} * n_{CR-y_p}}{n_{gray}} \quad (1)$$

Where n_{avg} denotes the average number of pixels. In the contextual region, n_{gray} is used to represent the number of gray levels. In the contextual region's x direction, n_{CR-x_p} is used to indicate the number of pixels. In the contextual region's y direction, n_{CR-y_p} is used to indicate the number of pixels.

The image's resolution can be mainly determined using Equation 1. Images with greater resolutions typically have higher pixel counts, which can be important for segmenting and classifying images.

The clip limit is then determined in the manner shown below (Equation 2),

$$n_{CL} = n_{CLIP} * n_{avg} \quad (2)$$

Providing a method for adaptive contrast enhancement is the goal of the clip limit Equation 2 in image processing. It enables CLAHE to improve an image's local contrast across several regions without exaggerating or amplifying noise.

Because CLAHE excels at improving accessibility to the usually noticeable but ordinarily difficult-to-reach locations. It is an effective instrument for processing biomedical images.

2.3. Data Augmentation

Effective training of deep CNN architecture requires more than only the 1190 samples' original CT scan slices. For this reason, the quantity of dataset samples should be increased by using data augmentation for every class in the original samples. Additionally, data augmentation aids in boosting the model's resilience and preventing over-fitting. This popular method is used to expand the number of datasets that are automatically generated by using various picture transformation methods, including cropping, mirroring, shearing, rotation, translation, and vertical and horizontal flipping. For properly balancing each class's dataset samples, six data augmentation techniques, height shift, width shift, zooming, brightness, rotation, and horizontal flip, are applied to the dataset in this research. Factors 3, 2, and 13 are used to augment the images in the normal, malignant, and benign classifications, respectively. Since the benign class has the fewest images, it has undergone the most augmentation. Each class's total

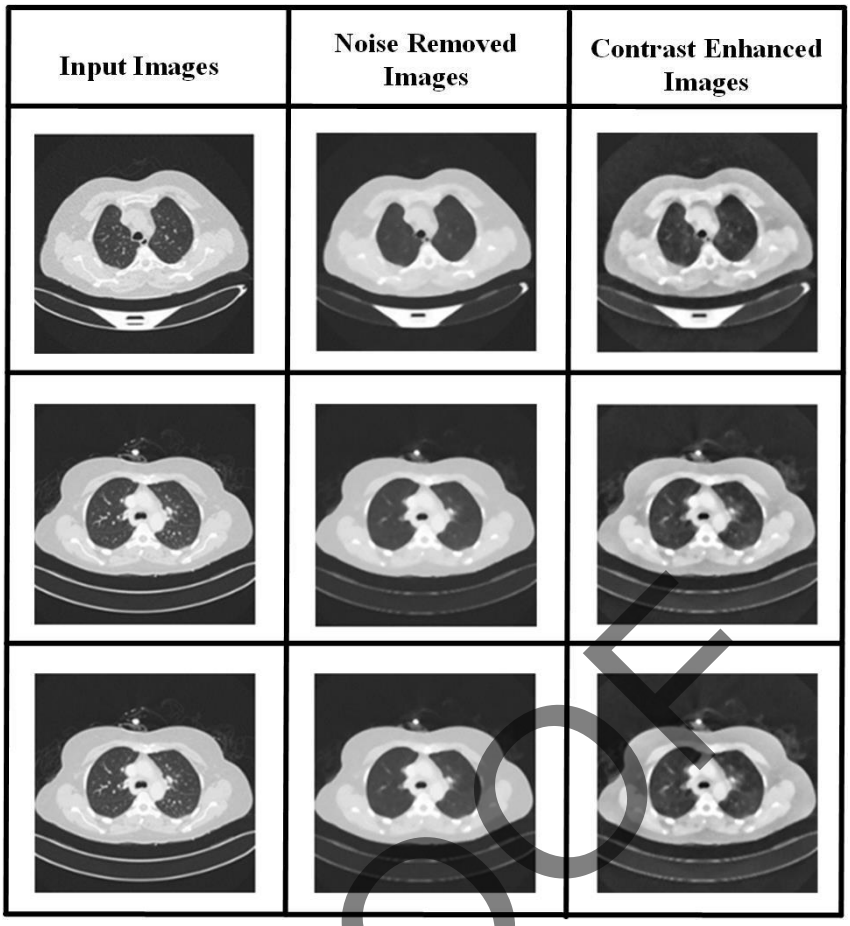

Figure 2. The sample pre-processed images with noise filtering (Adaptive median filtering) and Contrast enhancement (CLAHE)

Table 1. Dataset distribution for before and after data augmentation

Dataset	Class	Before data augmentation	After data augmentation
LIDC-IDRI	Benign	1015	2600
	Malignant	1051	2692
	Total	2066	5292
IQ-OTHNCCD	Benign	120	1680
	Malignant	561	1683
	Normal	416	1664
	Total	1097	5027

number of slices from a CT scan is listed in Table 1, both before and after data augmentation. In this case, a train set's CT scan slices are the only ones that undergo data augmentation. Following the application of the previously indicated data augmentation, Figure 3 displays a few examples of augmented CT scan slices for the IQ-OTH/NCCD dataset and LIDC-IDRI dataset.

2.4. Dataset Splitting

For non-biased training, all images in each class are first shuffled. They are then divided into an 80:20 train-test ratio, with 20% sampled for the testing set and 80% of the total images for the training set. Table 2 shows the number of images in the dataset for training and testing.

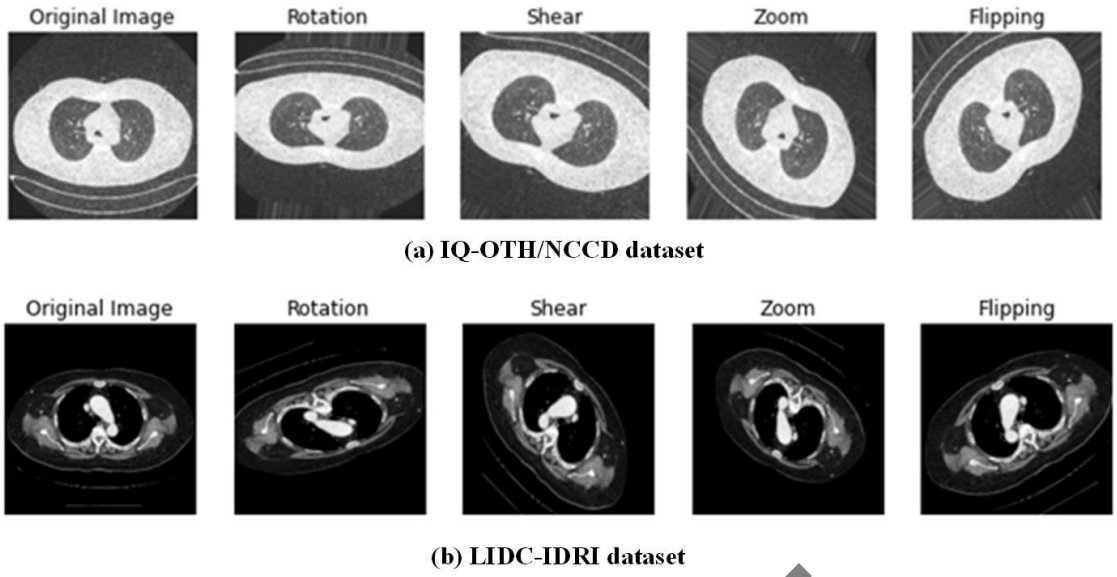

Figure 3. Sample of lung cancer slices from an original CT scan that has been augmented using several data augmentation techniques (a) IQ-OTH/NCCD dataset and (b) LIDC-IDRI dataset

Table 2. The total number of images for training and testing

Dataset Name	Split	Class Name	No. of sample	Total
IQ-OTH/NCCD	Training (80%)	Benign	1344	5027
		Malignant	1346	
		Normal	1331	
	Testing (20%)	Benign	336	
		Malignant	337	
		Normal	333	
LIDC-IDRI	Training (80%)	Benign	2080	5292
		Malignant	2154	
	Testing (20%)	Benign	520	
		Malignant	538	

2.5. Feature Extraction

Empirical Wavelet Transform Improvement helps to clarify the essential features under pre-processing. From the pre-processing outputs, the IEWT is used to extract the essential disease-related features, including Kurtosis, standard deviation, average amplitude, energy, entropy, variance, and mean. Because statistical features offer quantifiable information regarding lung tumor features, they are utilized in the classification of lung cancer. Machine learning methods use statistical features to characterize lung nodules characteristics succinctly, which is then used as input for categorization. These features maintain pertinent information while reducing the complexity of the data. Extracting discriminative features to achieve accurate lung cancer detection is the primary goal of the IEWT feature extraction technique. The following Equation 3 can be used to calculate the image density initially.

$$W_h^g(m,n) = \langle f(t), \psi_m(n) \rangle \quad (3)$$

The image edges are described as m and n, empirical wavelets, and detail coefficients are expressed as $\psi_m\ (t)$ and W_h^g, and $f(t)$ mentions the Fourier transform. The improved empirical wavelet transform is used to segment the boundaries into multiple components. In this way, the wavelet function and empirical scale are used to rebuild each element of the image. Equation 4 is used to calculate the empirical scale function

$$ESF(u,v) = \sum_{n=-\infty}^{+\infty} f(u-n)k(u;v) \quad (4)$$

Where the feature vector is described as k, the complex conjugate functions are defined as u and v. The empirical wavelet function was computed using the subsequent Equation 5.

$$EWF(u,v) = \sum_{n=-N}^{+N} f(u-n)k(u:v) \quad (5)$$

Where the empirical wavelet transform coefficients are represented by u and v.

The retrieved statistical features are Kurtosis, Standard Deviation, Average Amplitude, energy, entropy, variance, and mean. The Mean determines the average image values. Equation 6 is used to express the mean statistical features

$$MEAN(Y) = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (6)$$

Where the pixel value is depicted as i, the number of samples is described as n, and the matrix width is displayed as y.

Equation 7 expresses the variation of the statistical feature

$$V(Y)\frac{1}{n}\sum_{i=1}^{n}(y_i - MEAN(Y))^2 \quad (7)$$

Entropy: This is the most accurate way to determine how much information was contained in the event overall and how unstable the signal disturbance was. Equation 8 is used to express the statistical features' entropy.

$$ENT(Y) = -\sum_{i=1}^{n} \frac{y_i^2}{\sum_{i=1}^{n} y_i^2} log_2 \frac{y_i^2}{\sum_{n=1}^{n} y_i^2} \quad (8)$$

Energy: It is applied while operating a probability framework to describe a data observation. Combined with the Markov Random domain, this is the most extensive a priori assessment. Equation 9 is utilized to express the energy of statistical features.

$$ENE(Y) = \sum_{i=1}^{n} y_i^2 \quad (9)$$

Average Amplitude: Equation 10 represents the statistical feature's Average Amplitude.

$$AA(Y) = \frac{1}{n}\sum_{i=1}^{n} |y_i| \quad (10)$$

Standard Deviation: The average distance between the mean and the pixel value is described by this. In this, the high contrast is defined as a higher standard deviation value. From the mean, the pixel's lesser dispersion is denoted as a lower standard deviation value. Equation 11 shows the statistical structures' standard deviation.

$$SD(Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - MEAN(Y))}^{2} \quad (11)$$

Kurtosis: It is computed as the normal distribution's distribution stability. Equation 12 shows the statistical features of Kurtosis.

$$K(Y) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(|y_i| - MEAN(Y)\right)^4}{SD(Y)^4} \quad (12)$$

The Enhanced ResNeXt model is then used to classify these extracted features.

2.6. Classification

Using CT scans, the Enhanced ResNeXt CNN version was used in the proposed research to predict binary class and multi-class categorization of lung cancer.

ResNeXt was chosen as the backbone classification model in our proposed ensemble framework due to its superior modular architecture and representation capability. The cardinality concept introduced in ResNeXt (the number of parallel paths within a block) offers a more effective trade-off between depth and width, leading to better performance with fewer parameters compared to deeper networks like DenseNet or Inception, enhanced feature representation power, which is essential for

distinguishing complex patterns in lung cancer images (especially subtle differences between benign and malignant cases). While EfficientNet is also parameter-efficient, its performance advantage often depends on transfer learning with large-scale datasets (e.g., ImageNet), whereas our experiments prioritize handcrafted features and custom medical datasets. DenseNet, although effective, can suffer from feature redundancy in dense connectivity when used with statistical features from IEWT.

The ResNeXt model was presented by Xie *et al.* (2017). It is a development of the ResNet model made to handle challenging categorization problems. The cardinality blocks are used by the ResNeXt model, which is made up of several convolutional layer paths, in contrast to ResNet. Each pathway collects different filters, and before moving on to the next layer, their output is merged. The quantity of these paths is called the "cardinality," a configurable hyper-parameter. Through the integration of numerous paths in every block, multiple image classification benchmarks are outperformed by this approach. Through the exploration of many combined feature maps, the wide range of features for increasing the classification performance is enhanced by this proposed method. Particularly, ResNeXt's usefulness extends beyond image categorization; it has demonstrated its efficacy in a variety of tasks, such as object detection and image captioning.

Each with a distinct set of filters, the convolutional layers in the ResNeXt model are partitioned into multiple layers. The different paths are integrated or connected to form a more comprehensive group of features. By acting as several parallel pathways, the cardinality bottleneck modules are connected. Integrating parallel pathways with different filter sizes efficiently controls the relationship between computational costs, depth, and width. Comparing this method to the ResNet design, greater accuracy and efficiency are achieved. The ResNeXt expands upon the residual network (ResNet). Throughout the network, the smooth flow of information is improved by the residual connections in the ResNeXt model.

While the convolutional layers' kernel and channel sizes are greater, which restricts the model's capacity to capture small patterns or local details, the basic ResNeXt model classifies the images with good efficiency. More channels mean more parameters in the network, which raises the possibility of over-fitting, the model's computing cost, and the complexity of the network during training. High-parameter models are more likely to have disappearing or inflating gradients, which makes it more difficult for the model to converge during training. To enhance model effectiveness, the size of the kernels and channels is reduced to improve the structure. Smaller kernels lessen the computational load because they contain fewer parameters. Comparably, fewer channels equal fewer activations, which lowers memory consumption and speeds up model training and inference.

Several improvements are incorporated into the Enhanced ResNeXt model, such as the use of a dense layer followed by a Softmax activation function for classification and grouped convolutions. In the proposed Enhanced ResNeXt model, the standard ReLU activation function is replaced with the Leaky ReLU activation function to address the issue of dying neurons and improve gradient flow during training, thereby enhancing the model's learning capability and overall performance. These improvements are meant to increase the ResNeXt CNN model's efficacy and efficiency. The architectural layout of this Enhanced ResNeXt CNN model is shown in Figure 4.

Grouped convolutions in deep learning architectures are helpful because they allow filters to be replicated across modules, which helps build a more extensive network. By utilizing kernel filters, this replication efficiently extracts feature maps while lowering computational complexity.

This method is advantageous because it makes it easier for the model to recognize and emphasize distinctive features in the data. By using a variety of filters in the grouped convolutions, this method generates a wide range of feature representations, which enables a more sophisticated interpretation of the input data. As a result, the input data is analyzed more thoroughly and effectively, which benefits the model by enhancing its overall effectiveness and precision in classification tasks.

The proposed model uses the Leaky ReLU function following batch normalization. The dot product is one possible way to perform a combining transformation:

$$\sum_{j=1}^{M} W_j\, n_j \qquad (13)$$

Where a neuron's M-channel input vector is represented by the symbol "n", a filter's weight for the j^{th} neuron is depicted as $'w_j'$. The basic transformation $w_j n_j$ is replaced with possibly a network. The following equation defines the aggregate transformation:

$$T(n) = \sum_{j=1}^{c} S_j\,(n) \qquad (14)$$

To emulate the actions of a basic neuron, a low-dimensional embedding of 'y' is produced by the arbitrary function $'S_j(n)'$.

$$m = n + \sum_{i=1}^{c} S_j\,(n) \qquad (15)$$

Where the model's cardinality is depicted as 'c', an arbitrary function is described as S_j, the neuron's input vector is represented as 'n', and the model output is described as 'm'

Within the last dense layer, the Softmax optimizer transforms logits into the probability of changing to a specific disease cluster. The probability value is determined by using input biases and weights. Next, the value acquired is converted to the value associated with the disease class. The Softmax optimizer's value is determined using Equations 16 and 17.

$$P\big(n = j|\varphi(i)\big) = e^{\varphi(i)}{}_{\sum_{j=0}^{k} e^{\theta_K}(I)} \qquad (16)$$

$$\varphi = w_0 m_0 w_1 m_1 + \ldots\ldots\ldots + w_K m_K \qquad (17)$$

The input vector is described as $'\varphi'$, the $'j^{th}'$class bias is depicted as $'\,{}_0 m_0\,'$, and the number of classes is described as 'K'. The index connected to the representation of the corresponding classes is indicated by the variable "j", which has a range of 0 to 9.

In the proposed Enhanced ResNext model (Figure 4), different components in the architecture are color-coded to enhance visual clarity: pink blocks for convolution layers (1×1, 3×3), blue circular blocks for addition operations (+), which denote the residual connections, green and blue blocks for classification section (fully connected layers and Softmax). This helps distinguish different operations visually and improves readability. The model consists of an initial 5×5 convolutional layer followed by grouped convolutional blocks repeated X3, X4, X6, and X3 times, respectively. The numerical values inside the pink blocks (like $1\times1,64,3\times3,128$) reflect the

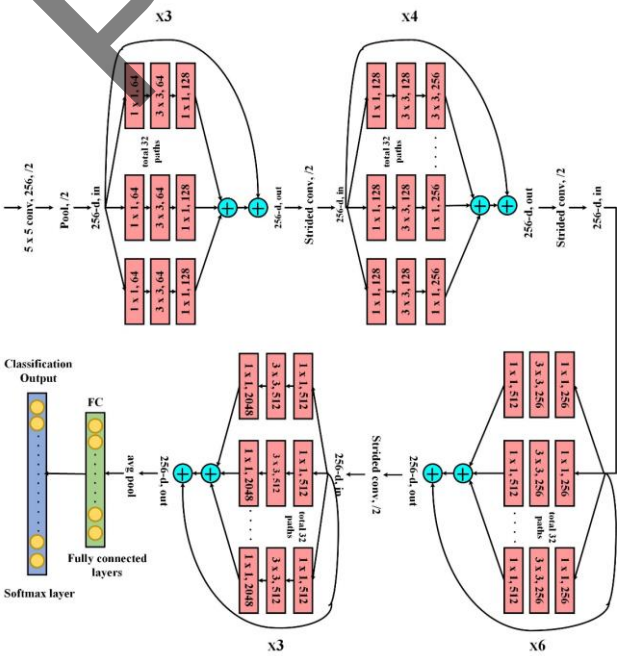

Figure 4. The structure diagram of the Enhanced ResNeXt

actual convolution operations and channel depths used at each stage. These values correspond to the convolution kernel size and output channel dimensions. A total of 32 paths have been added under the grouped convolution blocks. X3, X4, and X6 denote the number of times each grouped residual block is repeated in the corresponding stage. The final stage includes a global average pooling layer, fully connected layers, and a Softmax output layer.

2.7. Segmentation

The Modified ShuffleNetV2 was chosen for lung tumor segmentation due to its lightweight architecture with reduced computational overhead, making it suitable for real-time and resource-constrained environments (e.g., clinical settings), efficient channel split and shuffle operations, which promote better feature fusion and preserve spatial details crucial for accurate tumor boundary detection, capability to outperform heavier models in terms of speed while still achieving competitive segmentation performance. Moreover, our modified version includes adjustments that improve localization sensitivity specifically for tumor regions.

While UNet and its variants (UNet++, Attention-UNet, etc.) have been widely successful in medical image segmentation, they tend to be heavier and more memory-intensive, which may not be optimal for low-resource deployments, require large amounts of annotated data to prevent over-fitting; the segmentation task in our datasets includes challenges such as sparse and irregular tumor shapes, which benefit from the flexibility and lightweight nature of ShuffleNetV2. Our experiments showed that the Modified ShuffleNetV2 achieves a favorable trade-off between performance, efficiency, and robustness when compared to conventional UNet-based models.

The Modified ShuffleNetV2 model accurately segments the lung tumor's interested lesion region. A lightweight model that expands on the ShuffleNetV1 architecture is the ShuffleNetV2 network. The following four essential factors can significantly impact the ShuffleNetV2 network's speed. (1) The convolutional layer's equal input and output channels, the model executes as fast as possible with the least amount of memory access time. (2) A model's speed can be slowed down by performing too many convolution processes. (3) The speed increases with the number of branches in the model network. (4) The model can slow down by adding point-by-point processes, and the number of additions can be reduced.

There are two primary parts to ShuffleNetV2: the base unit and the down-sampling unit. Using a channel split operation, the input image features in the basic unit are separated evenly into two groups. The left branch is left unprocessed while the right branch iteratively runs through a 3 × 3 depth-wise convolution and a 1 × 1 convolution. Afterwards, information sharing is improved by using channel shuffling amongst various groups. The image features have direct access to both branches in the down-sampling unit. The right and left branches are then combined, and channel switching improves the flow of information among various groups.

In lung tumor segmentation models, a lightweight tumor identification model was presented to address the existing problems of low accuracy levels and complicated structures of lung cancers. The first step toward preventing neuronal death was the introduction of a parameterized linear rectification function. Secondly, an improved mixed attention mechanism was incorporated into the model to enhance its attention to tumor characteristics. Later, to simplify the architecture, the original model's network structure was modified. Lastly, a Modified ShuffleNetV2 model was created using transfer learning methods. The final fully connected layers in this model have been exchanged for a transposed convolutional layer, which allows us to up-sample the feature maps to the original image size.

The location, texture, shape, and size of the various forms of lung tumors varied significantly. Because tumors can have fuzzy boundaries, odd shapes, or small sizes, it can be challenging for typical models to represent them adequately. As a result, ShuffleNetV2 was chosen as the backbone network and modified appropriately. Figure 5 shows the design of the Modified ShuffleNetV2. The channel shuffling is represented as Channel Shuffle, the max pooling operation is performed using MaxPool, the channel concatenation is indicated as Concat, the channel separation is accomplished by channel split, and the depth-wise convolutions are represented as DWConv. The convolution operations are represented by Conv, Conv1, and Conv5. The down-sampling units are

layered together in Stages 2, 3, and 4. The enhanced version of the CBAM module is known as the Bidirectional Convolutional Block Attention Mechanism (BCBAM). The activation function used in the model is PReLU, the transposed convolutional layer is depicted as TConv, and the global average pooling process is performed using GlobalPool.

In Figure 5, the network begins with a convolutional layer (Conv1), followed by a PReLU activation and max-pooling operation. The core of the model is organized into four main stages (Stage 2 to Stage 5), each consisting of repeated modular units. Each stage is composed of two types of blocks: The basic unit (left part of each stage) is responsible for maintaining the spatial resolution and enhancing feature representations. It includes channel split, depth-wise separable convolutions, BCBAM attention, concatenation, and channel shuffle. The downsampling unit (right part of each stage) reduces the spatial resolution while increasing channel depth. It also integrates depth-wise convolutions and BCBAM modules, with batch normalization (BN) and PReLU used to preserve gradient flow and mitigate over-fitting. The model ends with a 1×1 convolution, global average pooling, and a final 1×1 convolution (TConv) to segment the tumor regions. The novel integration of PReLU and the proposed BCBAM module enhances non-linearity and attention, leading to improved feature learning and segmentation accuracy.

The classified input image size is 224×224×3, and it passes through a Conv1 layer followed by PReLU activation and produces a 112×112×24 feature map. Then MaxPooling reduces it to 56×56×24. Stages 2 to 4 are the main feature extraction stages. Each stage contains a basic unit (left block) and a down-sampling unit (right block). The basic unit performs feature transformation with 1×1 point-wise convolution → 3×3 depth-wise convolution (DWconv) → 1×1 convolution. Followed by the proposed BCBAM attention module (combining channel and spatial attention in parallel). Output features are concatenated with the bypass connection and passed through channel shuffle to mix features. The down-sampling unit has a similar structure but is applied in parallel to the basic unit to reduce feature map size while doubling the channel depth. It includes two 3×3 DWconv branches, one with residual, and BCBAM follows both, then concatenated and shuffled. Stage 2 of the proposed Modified ShuffleNetV2 model produces feature maps of size 28 × 28 × 116, where the spatial resolution is moderately reduced while enhancing channel-wise discriminative power for spatial localization of tumor regions. Stage 3 reduces the resolution to 14 × 14 and increases the depth to 232, allowing the network to capture more abstract features essential for identifying complex boundaries of lung tumors. Stage 4 outputs 7 × 7 × 300 feature maps, enabling the network to extract highly abstracted semantic information relevant to the structure and extent of tumor regions. Stage 5 (Conv5) further deepens the network to produce 7 × 7 × 1024 features, enriching the feature representation needed for accurate delineation of tumor areas. A Global Pooling layer condenses spatial information while retaining critical contextual cues for tumor boundary understanding. Finally, the 1 × 1 × 7 convolutional layer acts as a decoder head that outputs the segmentation mask with pixel-wise predictions, effectively highlighting the segmented tumor region in the lung.

2.7.1. Parametric Linear Rectification Function

The ShuffleNet structure's convolutional layers used ReLU as their activation function because of two main advantages: it can solve gradient vanishing problems and speed up training. However, the ReLU has several drawbacks. When large gradients or negative values flow through the ReLU neurons during training, they may stay dormant, which stops the related parameters from being updated. The ReLU's shortcomings were addressed by using a parametric linear rectifier function (PReLU). The learnable parameter α was developed by the PReLU to serve as the ReLU's foundation. As a result, the output multiplies the input value by α rather than being zero when the input is negative. With this modification, the dead neuron issue related to ReLU is lessened because even negative inputs result in nonzero outputs that have a slope that can be controlled. The model's performance could be improved by learning this slope during training. E represents the PReLU activation function.

$$P\,Re\,L\,U(X) = \begin{cases} X & ,X > 0 \\ \alpha_i X & ,X \leq 0 \end{cases} \qquad (18)$$

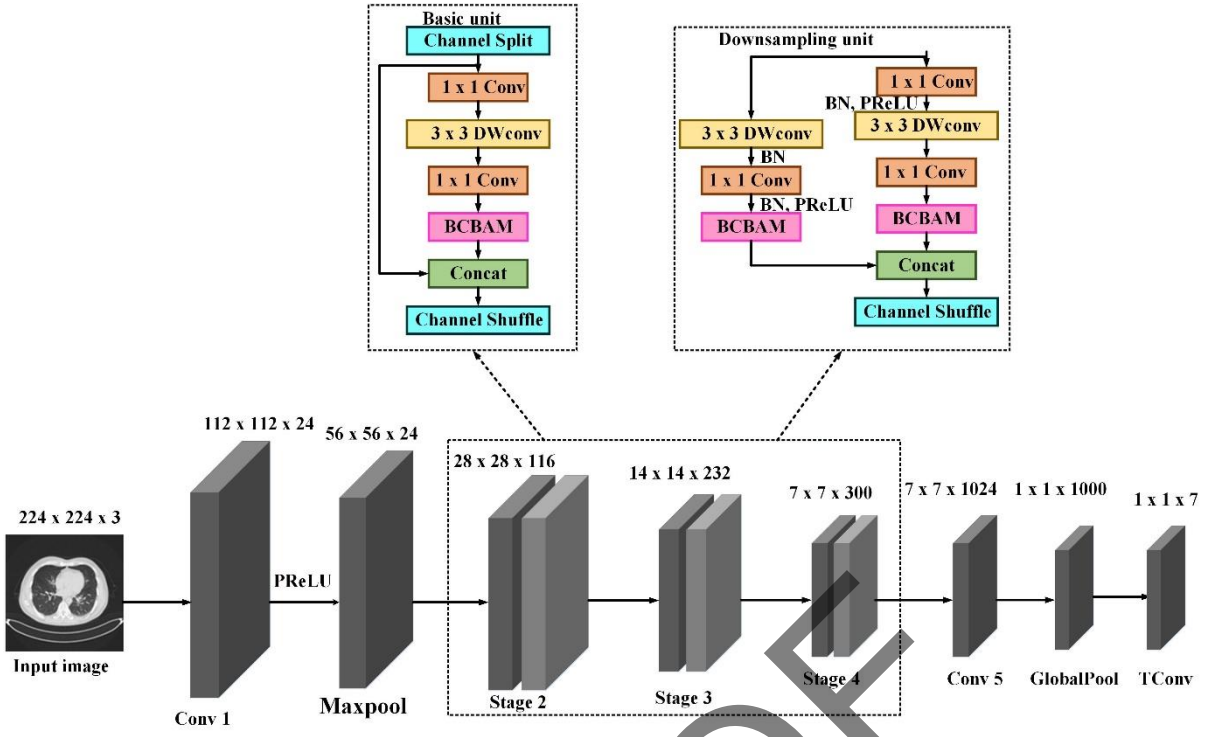

Figure 5. The structure of the Modified ShuffleNetV2 model

With a normal distribution, a random number is represented as α that ranges from 0 to 1, and a specific channel is indicated as i. $P\,Re\,L\,U(X)$ function becomes a ReLU function when $\alpha_i = 0$. The $P\,Re\,L\,U(X)$ function turns leaky at$\alpha_i > 0$. When $\alpha_i$is a variable parameter, $P\,Re\,L\,U(X)$ it becomes a parametric ReLU function.

2.7.2. Improved Mixed Attention Mechanism

Equation 19 illustrates how the CBAM, also referred to as a mixed attention mechanism, was added to the base block above. The convolutional layer's output image feature map F is used as its input, getting F_2 by repeatedly accessing the channel and spatial attention modules to obtain F_3. In particular, whether there is a deep network or a large feature map, the feature map multiplication operations increase the computational complexity of the model and require more processing resources. Since each module in the sequential CBAM connection calculates the weights on its own, there is an increase in memory utilization and computational expenses. Furthermore, the preceding module's features are only allowed by this module, and the subsequent module's features are ignored.

$$\begin{Bmatrix} F_2 = M_C(F) \oplus F \\ F_3 = M_S(F_2) \otimes F_2 \end{Bmatrix} \quad (19)$$

The output that is obtained from F experiencing the channel attention module is represented by $M_C(F)$, and after F_2 passing through the spatial attention module, the output weight is depicted as $M_S(F_2)$. The output that is generated when F moves via the channel attention module is $M_C(F)$. After the spatial attention module's F_2 traversal, the $M_S(F_2)$ is the output weight. The intermediate variable is depicted as $F_2$and $\otimes$ serves as the operator for the feature map's weighted multiplication.

A parallel connection was established instead of a hybrid attention mechanism's serial connection. Furthermore, the feature maps were transformed from multiplication to summation before output, which improved information integration and made it easier to extract features with stronger representational strength. The proposed parallel attention mechanism, summing feature maps, facilitates more direct gradient flow during back-propagation. This approach reduces the risk of gradient diminishment across layers, thereby enhancing the training stability of deep neural networks. By enabling multiple modules to compute weights simultaneously and integrating their outputs through summation, the network maintains stronger gradient signals, which improves feature learning and overall model performance. Various modules can calculate weights simultaneously in parallel connections, and the addition of feature maps helped

to reduce gradient vanishing, which enhanced deep neural network training outcomes. Figure 6 depicts the proposed BCBAM attention mechanism, and Equation 20 provides its functional formulation.

$$\left\{\begin{array}{c} F_4 = M_C(F) \\ F_5 = M_S(F) \\ F_6 = F \oplus F_4 \oplus F_5 \end{array}\right\} \quad (20)$$

The output weight is represented as the$M_S(F)$. The intermediate variables are depicted asF_4 and F_5. The operator for the feature map's weighted addition is defined by $\oplus$.

2.7.3. Adjusting Network Architecture

To improve performance, the ShuffleNetV2 network used the PReLU activation function in conjunction with the BCBAM. Nevertheless, the addition made the model more difficult. This research modified the Conv5 structural layer to implement dilated convolution by adding a parameter with an interval of 2 and changing the stride from 1 to 2. This made it possible to achieve a lightweight construction with performance. The main goal of dilated convolution was to increase the receptive field to enhance network performance. Furthermore, Stage 4's output channels were downsized from 464 to 300.

2.7.4. Transfer Learning

Applying features to one's network model after they have been pre-trained is known as transfer learning. This reduces the number of training steps, shortens the training period, and lessens over-fitting. When it comes to solving the problems caused by images in the datasets that are either too small or too high quality, transfer learning works well. In particular, specific dataset categories had few images with complex backgrounds, requiring as many as 200 training iterations and thus increasing the training time. As a result, transfer learning became an essential component of our proposed research.

3. Results

For lung cancer detection, this section presents the comprehensive findings and the accompanying evaluation of the proposed model. There is also a discussion of the results' implications. Furthermore, we provide a comparative analysis to demonstrate that the proposed method outperforms competing models.

3.1. Experimental Setup

Jupyter is used to perform all the experiments in the proposed research, which Google Colab has provided using an NVIDIA Tesla T4 GPU with 12 GB. The implementation makes use of the Python 3 environment and open-source libraries Pandas, Numpy, Scikit, Tensorflow, Matplotlib, and Keras.

3.2. Hyper-Parameter Settings

To identify the optimal base learner combination for our proposed ensemble technique, we first conduct a thorough experimental process with various combinations of CNN models. Table 3 lists the hyperparameters chosen for this experiment.

The optimization of hyperparameters in our proposed research was conducted using a grid search strategy. We performed systematic experiments across multiple combinations of key hyperparameters, including the number of epochs, batch size, and

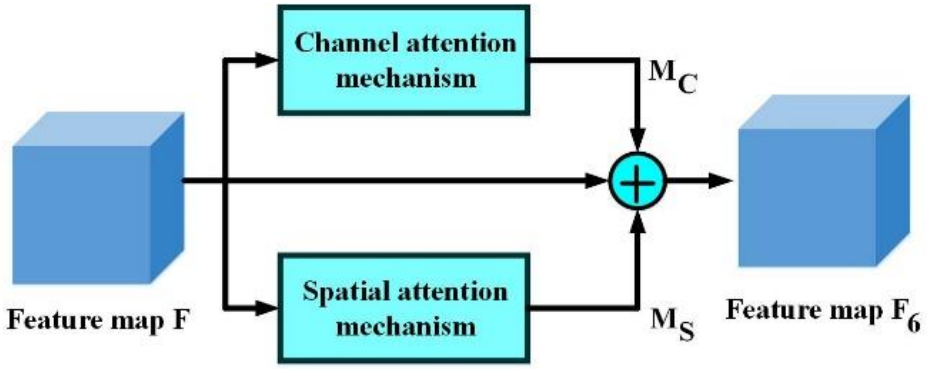

Figure 6. The structure of BCBAM

learning rate. The Adam optimizer was selected due to its adaptive learning rate property and stable convergence across various deep-learning tasks. Among the tested configurations, a learning rate of 0.001, batch size of 20, and 100 training epochs provided the most optimal trade-off between model convergence and generalization based on validation accuracy and loss. Other hyperparameters, like momentum (0.9) and the categorical cross-entropy loss function, were selected based on their empirical success in multi-class classification tasks involving Softmax outputs. While more complex tuning methods, such as Bayesian optimization, were considered, grid search was chosen due to its transparency and interpretability in our experimental setup.

Table 3. Optimized hyperparameters and their respective values

Hyper-parameter	Value/Range (Enhanced ResNeXt)	Value/Range (Modified ShuffleNetV2 model)
Optimizer	Adam	Adam
Momentum	0.9	0.9
Number of Epochs	100	100
Batch Size	20	20
Learning rate	0.001	0.001
Loss function	Categorical Cross-entropy loss function	Dice loss function

3.3. Performance Measures

The most often used metrics for classifying lung cancer were F1-score, recall, precision, AUC, and accuracy, according to the analysis of related publications. The authors of this paper will additionally make use of the following measures to make it feasible and easier to compare the results with the literature

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (21)$$

$$Pr\,e\,cision = \frac{TP}{TP + FP} \quad (22)$$

$$Re\,c\,all = \frac{TP}{TP + FN} \quad (23)$$

$$F1 - score = \frac{2TP}{FN + FP + 2TP} \quad (24)$$

$$AUC = 1 - \frac{1}{2}\left(\frac{FP}{FP + TN} + \frac{FN}{FN + TP}\right) \quad (25)$$

Where the quantity of accurate positives is known as TP (True Positives); the quantity of precise negatives is indicated by TN (True Negatives); The number of incorrect positives is indicated by FP (False Positives); The quantity of incorrect negatives is represented by FN (False Negatives).

There are two standard measures used to analyze segmentation performance. Two popular metrics for evaluation are the DSC and the IoU. These metrics aid in evaluating how well the ground truth and the actual region of interest in medical images, such as the tumor location, match the predicted segmentation.

IoU is a statistic that's used to assess how well the predicted segmentation and the ground truth segmentation overlap. The combined area of both regions calculates the degree of overlap between the expected and actual regions.

$$IoU = \frac{TP}{TP + FP + FN} \quad (26)$$

Another metric for comparing the overlap of the segmentation regions in the ground truth and prediction is the Dice Similarity Coefficient. It is frequently utilized in medical image segmentation tasks because of its exceptional sensitivity to small areas

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (27)$$

3.4. Classification Results

This paper presents enhanced ensemble deep learning models, such as the Enhanced ResNeXt model and Modified ShuffleNetV2 model, with modified network layers, a transfer learning approach, and an advanced attention mechanism that works on CT images for lung cancer diagnosis.

Excellent results are achieved on benchmark datasets for image classification using the proposed Enhanced ResNeXt model. For multi-classification lung cancer prediction, the Enhanced ResNeXt model

is a viable method based on the results of the proposed approach. The pre-processing techniques used on the lung pictures and the enhanced ResNeXt model's fine-tuning allowed the model to achieve exceptional outcomes when performing image classification tasks. Due to a limited dataset, this research uses transfer learning and data augmentation techniques. As seen in Figures 7 and 8, the class-wise results are provided in the classification report for the IQ-OTH/NCCD dataset and the LIDC-IDRI dataset. The model's high classification accuracy suggests a high level of reliability. An overall 98.33% precision and 99.43% accuracy in multi-class classification using the IQ-OTH/NCCD dataset are especially notable, providing evidence that the proposed model can accurately identify lung images that are normal, malignant, and benign. Further proof that the proposed model can accurately identify normal lung images derives from the 99.66% recall and 99.66% F1-score for the IQ-OTH/NCCD dataset. The proposed classification model achieves an overall 99% F1-score, 99.50% recall, 99.50% precision, and 99.37% accuracy for the LIDC-IDRI dataset. The classification results of the proposed model for the IQ-OTH/NCCD dataset and LIDC-IDRI dataset are visually shown in Figure 9 and Figure 10.

3.5. Segmentation Results

The highly accurate segmentation results are achieved by the proposed Modified ShuffleNetV2 model with ground truth and expected masks. Segmented images contain simply tumor segments that are easier to classify based on their morphology compared to those in other models. Ideally, both predicted and ground truth masks have been used to classify normal classes since they do not include any tumor feature. In the existing segmentation papers, several noises and artifacts are found, leading to incorrect categorization in the CT image normal class.

The segmentation results for the lung tumor are shown in Figure 11. The proposed model segments the IQ-OTH/NCCD dataset images with 97.05% IoU and 96.23% DSC metrics. The proposed model segments the LIDC-IDRI dataset images with 98.43% IoU and


```
Accuracy: 0.9943
28/28 ━━━━━━━━━━━━━━━━━━━━ 0s 7ms/step
Classification Report:
              precision    recall  f1-score   support

      Bengin       0.95      1.00      0.97        91
   Malignant       1.00      1.00      1.00       452
      Normal       1.00      0.99      0.99       334

    accuracy                           0.99       877
   macro avg       0.98      1.00      0.99       877
weighted avg       0.99      0.99      0.99       877
```

Figure 7. The classification report for multi-class classification using the IQ-OTH/NCCD dataset

```
Accuracy: 0.9937
149/149 ━━━━━━━━━━━━━━━━━━━━ 1s 3ms/step
Classification Report:
              precision    recall  f1-score   support

      benign       1.00      0.99      0.99      2330
   malignant       0.99      1.00      0.99      2432

    accuracy                           0.99      4762
   macro avg       0.99      0.99      0.99      4762
weighted avg       0.99      0.99      0.99      4762
```

Figure 8. The classification report for binary classification using the LIDC-IDRI dataset

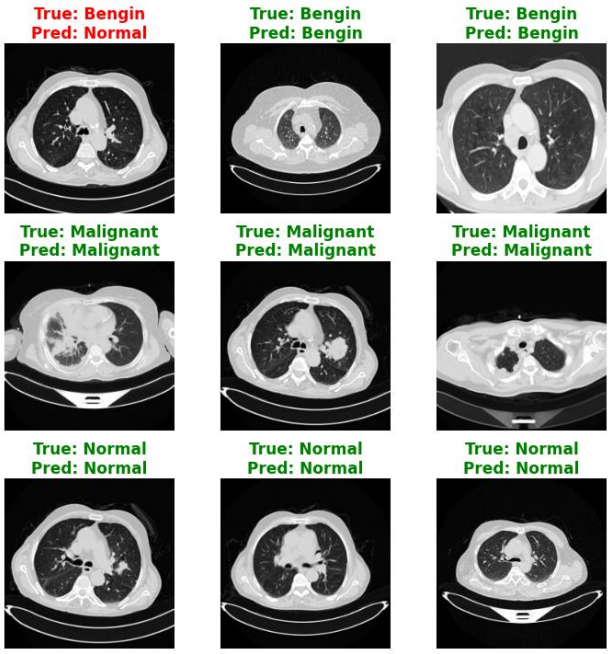

Figure 9. The classified results of the proposed model using the IQ-OTH/NCCD dataset

97.24% DSC measures. Through accurate segmentation, the proposed methodology aids in the differentiation of benign and malignant lung tumors. This could lead to an initial and more precise

diagnosis. Accurate segmentation can assist in identifying patient-specific disease patterns, allowing for more personalized treatments.

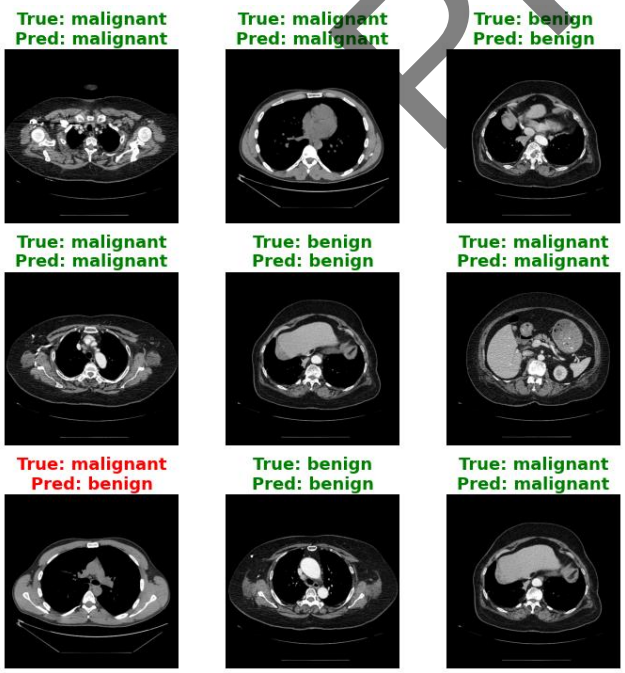

Figure 10. The classified results of the proposed model using the LIDC-IDRI dataset

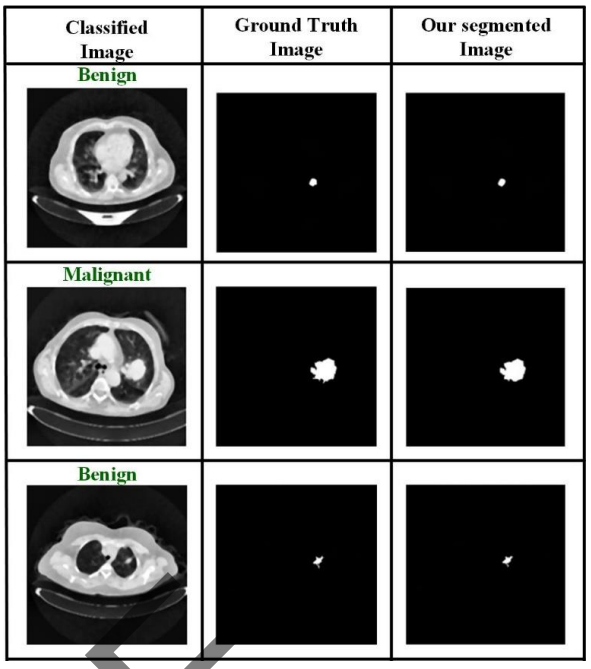

Figure 11. The segmentation results of the proposed model (Modified ShuffleNet)

3.6. Confusion Matrix

The confusion matrix used to predict lung cancer using multi-classification (IQ-OTH/NCCD dataset) and binary classification (LIDC-IDRI dataset) is displayed in Figure 12. In the form of a matrix plot, it is utilized to obtain the results of the classifications or misclassifications report presented between the true label and the predicted label. The proposed approach accurately detects 538 images as malignant and 520 images as benign in binary classification; however, it mistakenly detects 3 images as malignant and 3 images as benign.

The proposed classification model correctly labels 334 images as benign, 335 as malignant, and 331 as normal in multi-class classification. Still, it incorrectly labels two benign case images as malignant, two normal case images as benign, and two malignant case images as both normal and benign. Improved overall model accuracy is directly correlated with a high percentage of accurate classifications. The proposed approach yields fewer false positives, false negatives, and false positives in both binary and multi-classification.

3.7. Accuracy and Loss Curves

Line graphs plots measuring training and testing results for multi-class classification on the IQ-OTH/NCCD dataset experiment and binary classification on the LIDC-IDRI dataset experiment are displayed in Figures 13 and 14. In this manner, the Y-axis represents the accuracy rate attained for the training and testing set as indicated in the plot, and the X-axis represents the chosen training and testing accuracy as per epoch, which spans from 0 to 100. In the case of the IQ-OTH/NCCD dataset, the accuracy rate was 86% at the first epoch and quickly rose to over 99% by the 200th epoch; nevertheless, the testing accuracy is approximately 95%. Initially, it dropped progressively to roughly 85% by the 25th epoch. There are a lot more possibilities, and the stability is approximately 99%. The accuracy rate of the LIDC-IDRI dataset started at 95% at the beginning of the epoch and quickly improved to about 99.5% by the 200th epoch. The proposed model (Enhanced ResNeXt) achieves better classification accuracy for lung cancer diagnosis with fewer training and testing losses.

3.8. ROC Curve

Regarding this dataset, the final ROC curve indicates a notably low false positive rate. This suggests that the proposed classification model is highly effective for both binary classification and multi-class classification problems. Figure 15 displays the area under the curve (AUC) for the IQ-OTH/NCCD and LIDC-IDRI datasets. Here, we can see that for the IQ-OTH/NCCD dataset, Benign is at 99%, Malignant is at 100%, and Normal is at 97%. For the LIDC-IDRI dataset, Benign is at 99% and Malignant is at 99%. The proposed method's validity is further supported by the ROC curve of 0.99 to 0.1. Excellent AUC values for both benign and malignant cases enhance the model's capacity to identify abnormalities early on, potentially leading to life-preserving early intervention.

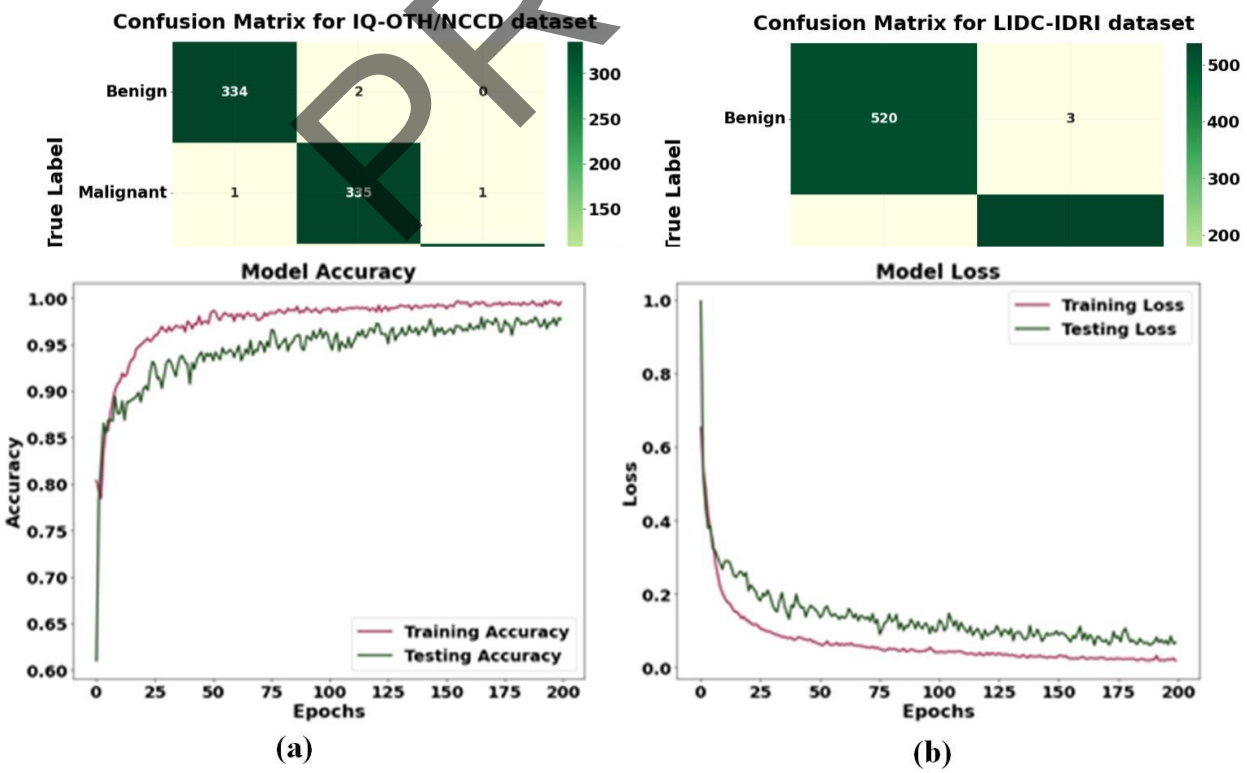

Figure 13. Accuracy and loss score graph for the training and testing sets for the IQ-OTH/NCCD dataset

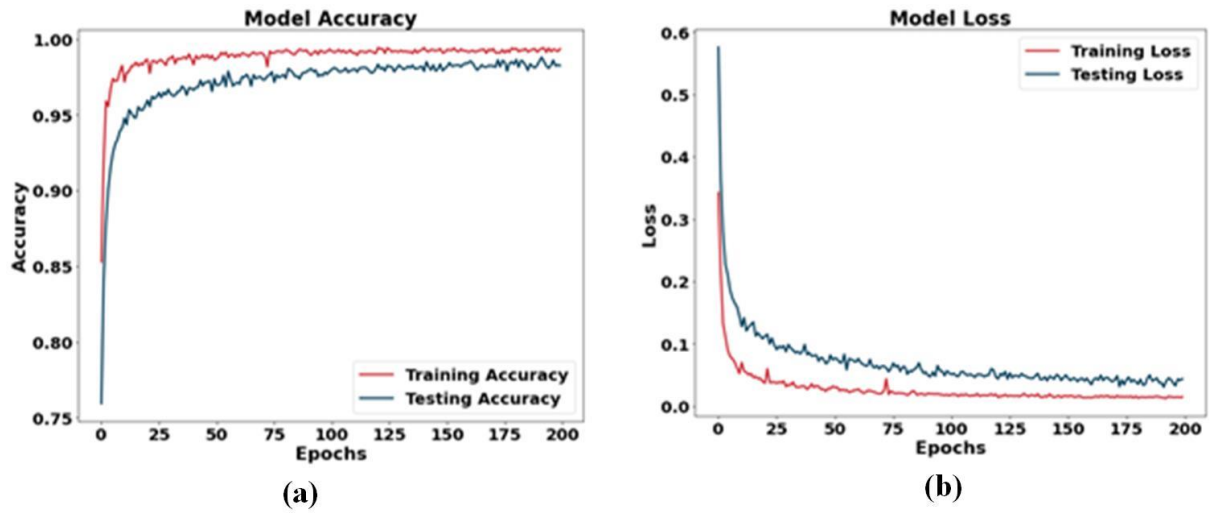

Figure 14. Accuracy and loss score graph for the training and testing sets for the LIDC-IDRI dataset

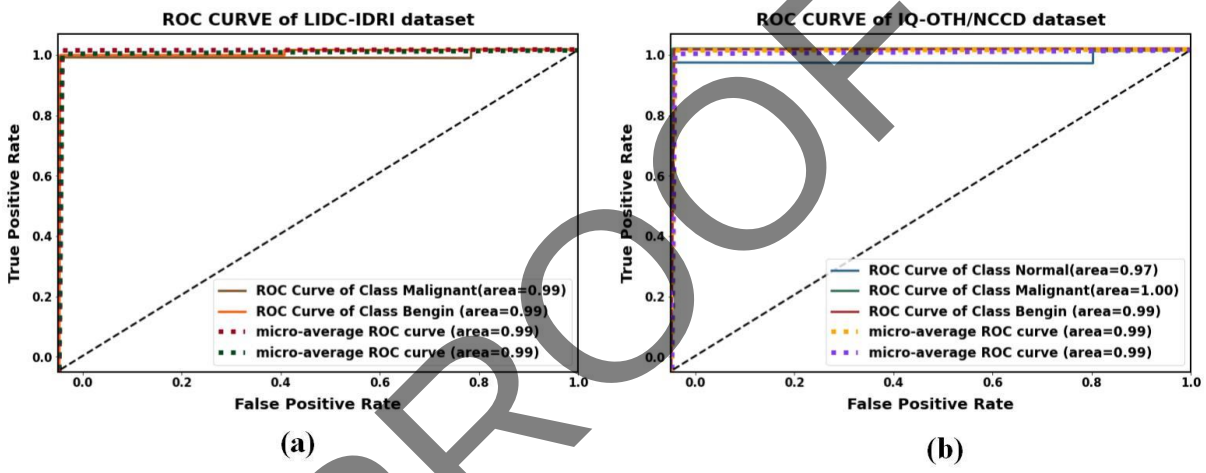

Figure 15. ROC curve for lung cancer classification using (a) LIDC-IDRI dataset and (b) IQ-OTH/NCCD dataset

3.9. Performance Evaluation

It has been demonstrated that the proposed classification model, Enhanced ResNeXt, performs better on lung cancer classification tasks than other existing models. From the input images, the enhanced ResNeXt model's ability to extract features is one of the factors contributing to its superior performance. From the images, the features are extracted by combining the convolutional and pooling layers. This makes it possible for the model to gather intricate visual representations, which are essential for precise categorization. It has been shown that the proposed classification model (Enhanced ResNeXt) is far more effective and versatile than previous deep learning-based CNN designs such as Deep ShirmpNet [28], ShuffleNet [22], MENet [29], CNN-BiLSTM [23], CNN [30], and LCCNet [31], and traditional machine learning methods such as SVM and XGBoost [27]. The accuracy of the proposed approach is compared to the most recent models, and a fair comparison was guaranteed by maintaining the same train-test split ratio and dataset. We provided a thorough examination of the Enhanced ResNext model using cutting-edge methods for classifying lung cancer, as shown in Table 4.

Concerning Table 4, the proposed approach, which delivers a 1.5% error rate and has accuracies of 99.43% and 99.37% for the IQ-OTH/NCCD and LIDC-IDRI datasets, provides the best confirmation with the real data. Additionally, the F1-score of 98.66% and 99% of the proposed approach are the highest when compared to the other existing methods, showing a higher rate of true negatives and the

system's ability to diagnose non-tumor pixels. With a 99.66% and 99.5% recall rate, the proposed model is significantly more adept at accurately diagnosing equivalent pixels or the unaffected positive rate. The advanced similarity of the diagnostic images is demonstrated by the proposed method's precision of 98.33% and 99.5%. These results are significant not only because of the high metric scores but also because the model can accurately distinguish between benign and malignant instances.

Also, a detailed comparison was conducted against the baseline ResNeXt model on two benchmark datasets: IQ-OTH/NCCD and LIDC-IDRI to evaluate the effectiveness of the proposed Enhanced ResNeXt model. The proposed model introduces several enhancements to the original ResNeXt architecture, such as the integration of residual blocks with concatenated convolutions for better feature representation, the use of ReLU activation functions to improve non-linearity and model learning capacity, and the addition of a dense layer before the Softmax activation function to refine the final decision boundaries. These architectural modifications have led to notable performance gains across all evaluation metrics. On the IQ-OTH/NCCD dataset, the proposed model achieved an accuracy improvement of 2.16% (from 97.27% to 99.43%), precision improvement of 2.88%, recall improvement of 2.65% and F1-score improvement of 2.13%. Similarly, on the LIDC-IDRI dataset, the proposed model achieved an accuracy improvement of 2.03% (from 97.34% to 99.37%), precision improvement of 2.81%, recall improvement of 1.67% and F1-score improvement of 1.14%. These consistent gains across datasets clearly demonstrate the superiority of the proposed Enhanced ResNeXt model in lung cancer classification tasks. The results confirm that the enhanced architectural features contribute significantly to more accurate and robust decision-making in medical image analysis.

The majority of lung cancer diagnosis models now in use focus exclusively on binary categorization, frequently overlooking the subtle differences between instances that are benign and those that are malignant. The multi-class classification capabilities of our proposed model in this research set a new benchmark and offer a more comprehensive diagnostic tool. The results show that our model has more advanced detection capabilities than existing methods, which may lead to a more insightful and detailed diagnostic image than those offered by the currently available instruments.

Table 5 describes the comparison methodologies and the results of the proposed Modified ShuffleNetV2 model for lung tumor segmentation. This Table makes it clear that, in terms of segmentation results, the proposed Modified ShuffleNetV2 produces higher results than the existing segmentation methods. The results for the LIDC-IDRI dataset (0.9843 IOU, 0.9724 DSC) and the IQ-OTH/NCCD dataset (0.9705 IOU, 0.9623 DSC) were obtained by the proposed model. The existing approaches, such as EfficientNet-B3 [32], DeepLabV3+ [33], and improved UNet [34], employ an exhaustive search method for region segmentation, which is computationally intensive and takes a long time to compute. To segment the images, a few parameters and a smaller training time are required by the proposed Modified ShuffleNetV2 model. The model requires 43 minutes to train. Thus, the proposed model is a rapid and accurate segmentation method.

3.10. Computational Complexity

A thorough examination is conducted to ascertain the computational complexity of the proposed improved ResNeXt model and Modified ShuffleNetV2. The evaluation takes into account several variables, such as the size of the network, the number of parameters, the training and inference times, and the number of floating-point operations (FLOPS). Table 6 summarizes the computational cost of training and evaluating our proposed deep learning models using CT image datasets using the same hyperparameter settings. Compared to other models, the 4.01 million parameters presented in the Enhanced ResNeXt are a relatively simpler architecture. In the direction of segmentation models like Modified ShuffleNetV2, we find that there are 6.51 million parameters.

4. Discussion

To lower the number of lung cancer-related deaths, this research focused on creating an ensemble lung cancer stage classifier and diagnosis model that includes detection modules and stage classifiers. In

Table 4. Performance comparison for lung cancer classification

Reference	Year	Methods	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Qadir *et al* [27]	2024	XGBoost	IQ-OTH/NCCD	98.54	98.63	96.35	97.36
Deepa *et al* [28]	2023	Deep ShirmpNet	IQ-OTH/NCCD	99.26	99.38	99.34	99.21
Mahesh *et al* [22]	2024	ShuffleNet	IQ-OTH/NCCD	90.71	90.39	91.37	-
Majumder *et al* [29]	2024	MENet	LIDC-IDRI	95.75	92.36	92.95	
Mostafa *et al* [23]	2024	CNN-BiLSTM	IQ-OTH/NCCD	99.2	97	95	99
Parveen *et al* [30]	2023	CNN	IQ-OTH/NCCD	97	98	97	97
Khaliq *et al* [31]	2023	LCCNet	IQ-OTH/NCCD	99	99	99	99
Base Model		ResNext	IQ-OTH/NCCD	97.27	95.45	97.01	96.53
			LIDC-IDRI	97.34	96.69	97.83	97.86
Proposed Research		Enhanced ResNeXt	IQ-OTH/NCCD	99.43	98.33	99.66	98.66
			LIDC-IDRI	99.37	99.5	99.5	99

Table 5. Performance comparison for lung tumor segmentation

Reference	Year	Models	Dataset name	IOU (%)	DSC (%)
Suji *et al* [32]	2024	EfficientNet-B3	LIDC-IDRI	59	-
Alshayeji *et al* [33]	2023	DeepLabV3+	LIDC-IDRI	71.5	-
Naseer *et al* [34]	2024	improved U-Net	LIDC-IDRI	-	85
Base Model		ShuffleNetV2	LIDC-IDRI	95.82	95.02
			IQ-OTH/NCCD	93.63	95.54
Proposed Research		Modified ShuffleNetV2	LIDC-IDRI	98.43	97.24
			IQ-OTH/NCCD	97.05	96.23

Table 6. The proposed models' computational complexity

Models	Model Size (MB)	No of parameters	Flops (G)	Testing time (hh:mm:ss)	Training time (hh:mm:ss)
Enhanced ResNeXt	47.2	4,011,391	0.96	0.00.02	0.03.20
Modified ShuffleNetV2	76.4	6,517,027	1.42	0.00.01	0.46.40

the lung tissue, the primary cause of lung cancer is the unrestricted distribution of cells, which results in lung cancer as a deadly disease. Since early detection may increase patient survival, lung cancer detection is essential. The Enhanced ResNeXt architecture and the Modified ShuffleNetV2 architecture are two enhanced architectures that are combined to create a unique model for lung cancer classification and segmentation in the present research. There are two phases to the proposed model. The Enhanced ResNeXt architecture is used in the first stage to classify lung cancer. Once removed from the lung lobes, the interesting tumor region is segmented using a Modified ShuffleNetV2 architecture. The proposed Enhanced ResNeXt model achieves 99.43% accuracy, 98.33% precision, 99.66% recall, and 98.66% F1-score for the IQ-OTH/NCCD dataset and 99.47% accuracy, 99.50% precision, 99.50% recall, and 99% F1-score for the LIDC-IDRI dataset in the first stage of lung cancer classification. For the segmentation of the tumor region, the proposed Modified ShuffleNet model shows a DSC of 96.23% and an IOU of 97.05% for the IQ-OTH/NCCD dataset and a DSC of 97.24% and an IOU of 98.43% for the LIDC-IDRI dataset.

To segment lung cancers, our research used the BCBAM attention mechanism in the Modified ShuffleNetV2 model. Segmentation performance can be significantly increased by adding the BCBAM attention mechanism to the ShuffleNetV2 model. This can be achieved by reducing redundant features, improving the handling of complicated structures, and improving emphasis on essential areas and channels. The proposed research uses the IEWT to extract important features. The model's computational complexity is decreased by the proposed method by eliminating extraneous features from the images, which leads to faster training periods and predictions. Lung cancer is categorized into several classes using the Enhanced ResNeXt model. In the proposed research, the modified ResNeXt model, faster training times, and greater computing efficiency are made possible by the grouped convolutions, and adequate

gradient flow of Leaky ReLU in ResNeXt training is more efficient and stable when a leaky ReLU is used, as it ensures predictable gradients. The replication learning of parameters is reduced, model performance is increased, and the over-fitting problem is reduced using the grouped convolutions in the ResNeXt model. An Enhanced ResNeXt model with an enhanced layer structure and Leaky ReLU activation function offers significant advantages in training stability, generalization, feature representation, and overall performance for several deep learning applications. The Adam optimizer is employed to optimize model weights during training, which results in increased processing efficiency and quicker convergence.

5. Conclusion

By automating the diagnosis process, CAD systems have been shown to improve diagnostic precision and reduce the possibility of human error. To enhance the model's accuracy for lung cancer prediction, we have created an ensemble model in the present research called the Enhanced ResNext model for classification and the Modified ShuffleNetV2 model for segmentation. This model incorporates transfer learning and attention mechanisms. A few pre-processing approaches were employed before deploying the primary diagnosis system, including adaptive median filtering for noise removal, contrast enhancement with CLAHE for better image quality, and traditional image augmentation techniques for artificially increasing the number of dataset images. The IEWT approach is used to extract the essential tumour-related features following pre-processing to enhance the performance of classification and segmentation. For classification, the retrieved features are fed into the Enhanced ResNext model. The Modified ShuffleNet model is then used to identify the precise lung tumor region using the classified images. The proposed approach has been tested on the IQ-OTHNCCD and LIDC-IDRI lung CT scan datasets, and the results are superior to many of the recently proposed approaches. The results of the performance evaluation demonstrate that the proposed model's accuracy can reach up to 99.43% for the IQ-OTHNCCD dataset and 99.37% for the LIDC-IDRI dataset; these findings illustrate the utility of this hybrid model for identifying lung cancer and its applicability in real-life scenarios.

However, there are false negatives and some false positives, which pose serious problems for medical professionals since they have a direct effect on patient care. Therefore, we must decrease these errors in the future. In order to achieve this, the basic CNN models are enhanced by adding advanced attention mechanisms. By concentrating on the most important regions, these mechanisms may aid in creating better feature maps, which may then lead to the creation of a better prediction model. To make the whole model relevant in real-world scenarios, we would like to investigate some lightweight CNN models in the future. Furthermore, we understand the significance of carefully examining and addressing any possible issues related to our proposed approach. Future research endeavors will encompass a thorough analysis of obstacles, encompassing matters such as comprehensibility and flexibility problems. Creating solutions for these issues will increase the proposed ensemble models' resilience and dependability in a variety of healthcare environments. Future research can assess the model's flexibility in responding to new data and possible conceptual drift, given the dynamic nature of medical datasets.

Acknowledgment

We declare that this manuscript is original, has not been published before, and is not currently being considered for publication elsewhere.

References

1- C. S. Parvathy and J. P. Jayan, "Automatic Lung Cancer Detection Using Computed Tomography Based on Chan Vese Segmentation and SENET", *Optical Memory and Neural Networks*, vol. 33(3), pp.339-354, (2024).

2- M. Coumarane, M. Balasubramanian and S. Sathiya," Deep Learning Techniques for Discerning Various Phases in Lung Cancer Disease using Chest X-Ray Images", *Journal of Electrical Systems*, vol. 20(3), pp.2826-2834, (2024).

3- R. Paramasivam, S. N.Patil, S. Konda and K. L.Hemalatha, "Lung cancer computed tomography image classification using Attention-based Capsule Network with dispersed dynamic routing", *Expert Systems,* vol. 41(9), p.e13607, (2024).

4- A. Feroui, M. Saim, M. E. A. Lazouni, S. A. Lazzouni, Z. A. Elaouaber and M. Messadi, "Improved diagnosis of lung cancer classification based on deep learning method", *International Journal of Biomedical Engineering and Technology*, vol. 46(2), pp.138-159, (2024).

5- S. Jain and P. Jaidka, "Lung Cancer Classification Using Deep Learning Hybrid Model", *In Future of AI in Medical Imaging* pp. 207-223, (2024) IGI Global.

6- X. Cheng, J. Li, M. Mi, H. Wang, J. Wang and P. Su, "Accuracy Study on Deep Learning-Based CT Image Analysis for Lung Nodule Detection and Classification", *Traitement du Signal*, vol. 41(2), (2024).

7- L. K.Gaddala, V. K. Reddy Radha, S. R.Buraga, V. L.Narla, K. R. Kodepogu and S. Yalamanchili, "Machine Learning-Based Lung Cancer Classification and Enhanced Accuracy on CT Images", *Traitement du Signal*, vol. 41(2), (2024).

8- T. Lan, F. Zeng, Z. Yi, X. Xu, and M. Zhu, "ICNoduleNet: Enhancing Pulmonary Nodule Detection Performance on Sharp Kernel CT Imaging", *IEEE Journal of Biomedical and Health Informatics*, (2024).

9- B. Mostafa, A. Keshk and M. Sakr, "A Deep-Learning Model Based on Transfer-Learning Technique for Detecting and Classifying Anomalies in Lungs Images", *IJCI. International Journal of Computers and Information*, vol. 10(3), pp.63-72, (2023).

10- D. Wang, C. Han, Z. Zhang, T. Zhai, H. Lin, B. Yang, Y. Cui, Y. Lin, Z. Zhao, L. Zhao and C. "Liang, FedDUS: Lung tumor segmentation on CT images through federated semi-supervised with dynamic update strategy", *Computer Methods and Programs in Biomedicine*, vol. 249, p.108141, (2024).

11- J. Subash and S. Kalaivani, "Dual-stage classification for lung cancer detection and staging using hybrid deep learning techniques", *Neural Computing and Applications*, vol. 36(14), pp.8141-8161, (2024).

12- V. Gampala, V. Ramya, B. Maram and S. R.Pappu, "Identification of lung cancer using archimedes flow regime optimization enabled deep belief network", *Multimedia Tools and Applications*, pp.1-30, (2024).

13- M. Jha, R. Gupta and R. Saxena, "Convolutional neural network based detection of lung adenocarcinoma by amalgamating hybrid features", *International Journal of Advanced Technology and Engineering Exploration*, vol. 11(111), p.160, (2024).

14- Z. Gao, Y. Guo, G. Wang, X. Chen, X. Cao, C. Zhang, S. An and F. Xu, "Robust deep learning from incomplete annotation for accurate lung nodule detection", *Computers in Biology and Medicine*, vol. 173, p.108361, (2024).

15- C. S. Parvathy and J. P.Jayan, "Lung cancer prediction in chest CT using an active contour-based segmentation and 3DCNN", *Multimedia Tools and Applications*, vol. 83(21), pp.60493-60517, (2024).

16- P. Ramasamy, E. Alabdulkreem, N. Alruwais and V. P.Gladis Pushparathi, "An attention-based neural network for lung cancer classification and gradient in MRI", *Automatika*, vol. 65(4), pp.1379-1390, (2024).

17- A. Hassan, H. Khan, A. Ali, A. Sajid, M. Husain, M. Ali, A. Naz and H. Fakhar, "An Enhanced Lung Cancer Identification and Classification Based on Advanced Deep Learning and Convolutional Neural Network", *Bulletin of Business and Economics* (BBE), vol. 13(2), pp.136-141, (2024).

18- S. A. Agnes, A. A. Solomon, and K. Karthick, "Wavelet U-Net++ for accurate lung nodule segmentation in CT scans: Improving early detection and diagnosis of lung cancer", *Biomedical Signal Processing and Control*, vol. 87, p.105509, (2024).

19- K. Priyadarshini, M. Alagarsamy, K. Sangeetha and D. Thangaraju, "Hybrid RNN-FFBPNN Optimized with Glowworm Swarm Algorithm for Lung Cancer Prediction", *IETE Journal of Research*, vol.70(5), pp.4453-4468, (2024).

20- P. Saravanaprasad and S.A. Karuppusamy, "Advanced lung tumor diagnosis using a 3D deep neural network based CAD system", *Biomedical Signal Processing and Control*, vol.89, p.105650, (2024).

21- A. Gupta, A. Kumar and K. Rautela, "UDCT: lung Cancer detection and classification using U-net and DARTS for medical CT images", *Multimedia Tools and Applications*, pp.1-21, (2024).

22- S. Mahesh, "Hybrid optimized MRF-based lung lobe segmentation and lung cancer classification using Shufflenet", *Multimedia Tools and Applications,* 83(17), pp.52335-52364, (2024).

23- B. Mostafa, M. Sakr and A. Keshk, "Employing the Capabilities of LSTM and Bi-LSTM for Lung Cancer Detection and Classification", *International Journal of Intelligent Engineering & Systems,* 17(5), (2024).

24- R. Patnaik, P.S. Rath, S. Padhy and S. Dash, "Enhancing Lung Cancer Diagnosis through Advanced CT Scan Image Analysis: A Novel Approach using Mask-EffNet", *Nanotechnology Perceptions*, pp.1-26, (2024).

25- V.K. Gugulothu and S. Balaji, "An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques", *Multimedia Tools and Applications*, vol. 83(1), pp.1041-1061, (2024).

26- N. Venkatesan, S. Pasupathy and B. Gobinathan, "An efficient lung cancer detection using optimal SVM and improved weight-based beetle swarm optimization", *Biomedical Signal Processing and Control,* vol.88, p.105373, (2024).

27- Qadir, Abdalbasit Mohammed, Peshraw Ahmed Abdalla, and Dana Faiq Abd. "A Hybrid Lung Cancer Model for Diagnosis and Stage Classification from Computed Tomography Images", *Iraqi Journal for Electrical and Electronic Engineering,*

28- V. Deepa and P.M. Fathimal, "Deep-Shrimp Net fostered lung cancer classification from CT images", *Int.*

J. Image Graph. Signal Process. Vol. 15(4), pp.59-68, (2023).

29- S. Majumder, N. Gautam, A. Basu, A. Sau, Z.W. Geem and R. Sarkar, "MENet: A Mitscherlich function-based ensemble of CNN models to classify lung cancer using CT scans", *Plos one*, vol. 19(3), p.e0298527, (2024).

30- R. Parveen, U. Saleem, K. Abid and N. Aslam, "Identification of Lung Cancer by using Watershed Machine Learning Algorithm", *VFAST Transactions on Software Engineering*, vol. 11(2), pp.70-79, (2023).

31- K. Khaliq, A. Naeem, N. Aslam, A. Malik and K. Abid, "LCCNet: a deep learning based Method for the identification of lung Cancer using CT scans", *VFAST Transactions on Software Engineering*, vol. 11(2), pp.80-93, (2023).

32- R.J. Suji, W.W. Godfrey and J. Dhar, "Exploring pretrained encoders for lung nodule segmentation task using LIDC-IDRI dataset", *Multimedia Tools and Applications*, vol. 83(4), pp.9685-9708, (2024).

33- M.H. Alshayeji and S.E. Abed, "Lung cancer classification and identification framework with automatic nodule segmentation screening using machine learning", *Applied Intelligence*, vol. 53(16), pp.19724-19741, (2023).

34- I. Naseer, T. Masood, S. Akram, Z. Ali, A. Ahmad, S.U. Rehman and A. Jaffar, "Empowering diagnosis: cutting-edge segmentation and classification in lung cancer analysis", *Computers, Materials and Continua*, vol. 79(3), pp.4963-4977, (2024).