HybridMVNet: A Hybrid MobileNet–VGG19 Ensemble Framework for Accurate Lung Nodule Classification

Md. Rashed[1], Muhammad Jamil[2], Adnan Kavak[2], Md. Imran Hossain[1], A F M Zainul Abadin[1], and Md Zahidul Islam[3]

[1]**Department of Information and Communication Engineering, Pabna University of Science and Technology, Pabna-6600, Bangladesh**
[2]**Kocaeli University, Department of Computer Engineering, Izmit, 41380, Turkey**
[3]**Department of Information and Communication Technology, Islamic University, Kushtia 7003, Bangladesh**

Corresponding author:
Md. Imran Hossain[1]

Email address: imran05ice@pust.ac.bd

ABSTRACT

Lung cancer continues to be one of the primary causes of death from cancer around the world, and getting a correct diagnosis as soon as possible is very important for improving patient outcomes. The purpose of this study is to use variants of convolutional neural networks (CNNs) and ensemble methods to improve lung nodule classification. We have implemented state-of-the-art deep learning models such as MobileNet, MobileNetV2, and VGG19, and developed a unique hybrid model, Hybrid MobileNet (V1+V2), that combines features of both MobileNet and MobileNetV2. Then, we incorporated the VGG19 architecture into this hybrid model and created the final ensemble model. We evaluated these models on the LIDC-IDRI dataset and reported detailed evaluation metrics, including F1 score, recall, accuracy, and precision, to demonstrate that our method works. The proposed ensemble model achieves an accuracy of **96.13%**, precision of **97.10%**, recall of **98.57%**, specificity of **98.6%**, F1-score of **96.16%**, and AUC of **0.989** with the lowest training time of 1347.89 s and memory footprint of 16.6 MB, ensuring its effectiveness for accurate lung cancer detection. This approach provides a promising tool for early detection and diagnosis of lung cancer, with potential applications in clinical decision-making and patient care.

INTRODUCTION

Lung cancer is still the number one cause of cancer-related deaths worldwide and represents a major part of the global cancer burden. Getting the right diagnosis early is key to saving lives and improving outcomes. Lung and bronchus cancer is diagnosed in 49.0 per 100,000 men and women each year and has a mortality rate of 32.4 per 100,000 individuals, according to recent statistics. These numbers are based on cases from 2017 to 2021 and deaths from 2018 to 2022 (12). Lung cancer is the second most commonly diagnosed cancer globally, with an estimated 2.2 million new cases reported in 2020 (4). Lung cancer is predicted to become the leading cause of the doubling of the global cancer burden by 2050 (26). It kills more people than breast cancer and is estimated to cause 2.45 million deaths globally by 2030. Of these deaths, nearly 80% are due to tobacco use alone, highlighting the critical importance of preventative and diagnostic measures (34).

Despite advances in diagnostic imaging, the differential diagnosis of benign and malignant lung nodules remains challenging. This complexity is due to the variability of lung nodules, which requires precise analysis for clinical decision-making. Conventional diagnostic methods are highly specialized and subject to inconsistencies. Recent advances in machine learning and deep learning have revolutionized medical imaging, providing tools to increase accuracy and efficiency (30). In particular, convolutional neural networks (CNNs) have demonstrated promising results in automating image analysis tasks and

have shown significant potential to improve the accuracy of lung nodule classification. Ensemble learning is a promising approach to improving classification performance by leveraging the strengths of multiple models. Individual models can be combined to overcome their individual weaknesses and to generate more robust predictions. In this paper, a new hybrid ensemble model for lung nodule classification is proposed. This approach combines features from the MobileNet, MobileNetV2 and VGG19 architectures and leverages the strengths of each. MobileNet and MobileNetV2 are merged to form a hybrid model called Hybrid MobileNet (V1+V2). MobileNet is lightweight, and MobileNetV2 is efficient. This hybrid model was further extended by introducing VGG19 into an ensemble framework, which provided complementary features to improve classification accuracy. The ensemble model was designed to distinguish benign and malignant lung nodules with high accuracy and robustness. Training large models requires high computational resources. We optimized the training process with multithreading. This method allows tasks to be parallelized, making effective use of multi-core processors and aiding in managing complex datasets (14).

Also, to improve model interpretability, Grad-CAM (Gradient-weighted Class Activation Mapping) was used to generate heatmaps showing the regions of lung CT scans that have the greatest impact on classification. These visual aids increase confidence in the model's clinical use by providing transparency and helping physicians understand its decision-making process (27). This strategy parallelizes the tasks, enabling efficient use of multi-core processors and processing of complex datasets (19). In addition, model interpretability was assessed using Grad-CAM (Gradient-weighted Class Activation Mapping) to generate heatmaps highlighting the most relevant regions in lung CT scans for classification. Such visualizations provide transparency and help clinicians better understand the model's decision-making process, fostering trust in its clinical use (23).

The proposed ensemble model was thoroughly evaluated on the LIDC-IDRI dataset (10) and achieved excellent performance metrics, with an accuracy of 96.13%, a precision of 97.10%, a recall of 98.57%, and an F1-score of 96.16%. These results demonstrate the potential of deep learning and ensemble approaches to address key challenges in lung cancer diagnosis.

The problem of lung nodule classification can be formulated as a supervised binary image classification problem. Given a CT image $x_i \in \mathbb{R}^{H\times W\times C}$ and its corresponding diagnostic label $y_i \in \{0,1\}$, where 0 and 1 denote the benign and malignant nodule, respectively, the task is to learn a discriminative mapping function $f_\theta(\cdot)$ to estimate the malignancy probability of each input image. The problem is formally stated as:

$$\hat{y}_i = f_\theta(x_i), \qquad f_\theta : \mathbb{R}^{H\times W\times C} \to [0,1], \tag{1}$$

where $\hat{y}_i$ represents the predicted probability of malignancy and θ denotes the learnable parameters of the proposed ensemble model. The final class decision is obtained by applying a decision threshold τ to the predicted probability:

$$\tilde{y}_i = \begin{cases} 1, & \hat{y}_i \geq \tau, \\ 0, & \hat{y}_i < \tau. \end{cases} \tag{2}$$

In this study, τ is set to 0.5 for binary classification.

Despite the promising advances in deep learning-based lung nodule classification, several critical challenges remain insufficiently addressed in the existing literature. These include the trade-off between computational efficiency and classification accuracy, the limited interpretability of black-box deep learning models in clinical settings, and the lack of lightweight architectures suitable for resource-constrained healthcare environments. To address these gaps, this study is guided by the following research questions: (i) Can a hybrid fusion of lightweight MobileNetV1 and MobileNetV2 architectures produce more discriminative feature representations than either model alone for benign and malignant lung nodule classification? (ii) Can the classification performance of the Hybrid MobileNet (V1+V2) model be further improved by incorporating VGG19 as a complementary deep feature extractor in an ensemble framework? (iii) Can the proposed ensemble model achieve comparable or better classification performance than existing deep learning methods with much lower computational cost in terms of training time and memory footprint? (iv) How much can Grad-CAM visualization provide clinically interpretable insights about the model's classification decisions of benign and malignant lung nodules?

To address these research questions, this study defines the following objectives:

(1) To introduce a novel hybrid model, **HybridMVNet**, by combining MobileNetV1 and MobileNetV2 with a weighted feature-level hybrid strategy, exploiting the complementary advantages of two lightweight architectures for efficient lung nodule feature extraction. (2) To build a final ensemble model by combining the Hybrid MobileNet (V1+V2) branch and VGG19 with an equal-weight averaging strategy, which allows the model to learn lightweight discriminative features and deep hierarchical spatial representations from lung CT images. (3) To enhance the computational efficiency of the proposed model with multithreaded training, reducing the training time and memory footprint while maintaining the high classification performance suitable for real-world clinical deployment. (4) To perform a comprehensive evaluation of the proposed ensemble model on the LIDC-IDRI dataset using standard performance metrics such as accuracy, precision, recall, specificity, F1-score, MCC, AUC, and the confusion matrix and compare the performance with existing deep learning approaches. (5) To improve the clinical interpretability of the proposed model by using Grad-CAM visualization, generating class-discriminative heatmaps to localize the most diagnostically relevant regions in lung CT images for benign and malignant classification decisions.

The contribution of this research is as follows:

A. Developing an innovative hybrid model, Hybrid MobileNet (V1+V2), that combines the advantages of MobileNet and MobileNetV2 for classification of lung nodules.

B. Incorporation of VGG19 in an ensemble framework with Hybrid MobileNet (V1+V2) for further improvement of classification performance.

C. Multithreading optimization of training efficiency, reducing computational time and improving scalability for real-world applications.

D. Incorporate Grad-CAM for model explainability to gain insight into the classification decisions and ensure reliability for clinical practice.

E. Extensive evaluation of the proposed method on the LIDC-IDRI dataset demonstrates its robustness and applicability to lung cancer detection.

The remainder of this paper is organized as follows. Section 2 reviews related work on deep learning-based lung nodule classification. Section 3 describes the proposed methodology, including data preprocessing, augmentation, and the HybridMVNet architecture. Section 4 presents the experimental results and comparative analysis. Section 5 provides a discussion of the findings, followed by conclusions and directions for future work in Section 6.

RELATED WORKS

Deep learning and artificial intelligence techniques have led to significant improvements in lung cancer classification, particularly in differentiating small cell lung cancer (SCLC) from non-small cell lung cancer (NSCLC) using imaging modalities such as PET/CT and CT. Early deep learning approaches used image processing techniques such as adaptive bilateral filtering, noise reduction, top-hat transformation, and segmentation using the marker-based watershed algorithm. Thereafter, the features were provided to machine learning classifiers for the segmentation and classification tasks using these methods (36). This study proposed an inflammatory-immune-nutritional integrated scoring system (IINS) for predicting survival in NSCLC patients. The PLR-, SIRI- and PNI-based model was superior to the traditional TNM staging system in prognostic prediction. However, the study had limitations inherent to a retrospective design and a small external validation cohort (16). In 2021, a transfer learning model was developed for the IQ-OTH/NCCD lung cancer dataset based on GoogleNet, a convolutional neural network (CNN) based on the Inception architecture. The model reached a remarkable validation accuracy of 94.38% (7).

He-Xuan Hu et al. proposed a hybrid attention mechanism and combined it with DenseNet for lung tumor segmentation, achieving 94% accuracy and setting a new benchmark in segmentation performance (15). Another approach is the DGMM-RBCNN method, which achieved an accuracy of 87.79% by extracting features like area, perimeter and entropy (17). A combination of AlexNet and a CNN was also used to classify lung cancer as benign or malignant, but the model had limitations in its diagnostic and treatment scope (6). MobileNet V2, a modified transfer learning model, was employed for lung pathology

classification from frontal thoracic X-rays. This model demonstrated an average accuracy of over 90% and an AUC of 0.811 (21). A hybrid deep learning approach, using a particle swarm optimizer for feature enhancement and a CNN for classification, achieved 95.8% accuracy on a dataset of 3010 images (33).

Gamma correction was applied to CNNs for histology classification, with the model achieving 87.16% accuracy in identifying adenocarcinoma and squamous cell carcinoma. Similarly, deep CNNs that utilize multi-scale features and dilated units predicted patient survival with 88.58% accuracy (2). Comparisons of classifiers such as KNN, SVM, Naïve Bayes, and NNN showed that SVM achieved the highest accuracy, 92.6%, outperforming the other methods (8). Ensemble learning methods such as XGBoost, LightGBM, and AdaBoost, coupled with SMOTE-based oversampling, achieved 94.42% accuracy and 95.66% precision on a lung cancer dataset (22).

The VGG19 model has also been used for lung cancer image classification (3). At the same time, InceptionV3 showed exceptional performance in classifying histopathological images of NSCLC subtypes, such as adenocarcinoma and squamous cell carcinoma (20). An ensemble-based approach combining pre-trained networks such as VGG, Xception, and ResNet was proposed to improve classification accuracy on lung cancer images from CT scans. This method demonstrated improved performance with better generalization across datasets (9). A deep ensemble 2D CNN model integrating multiple CNNs was reported to yield a combined accuracy of 95% for lung cancer classification (35). Other studies used transfer learning and pre-trained models such as VGG16, ResNet50V2, and DenseNet201 to build an ensemble model. This ensemble achieved an accuracy of 91%, outperforming each model (29). Another study used ResNet101 and InceptionV3 in combination to classify lung cancer subtypes, achieving a peak accuracy of 93.7% (25).

A specialized algorithm, combined with FPSO, achieved 94.97% accuracy in identifying malignant lung nodules (11). Hybrid intelligent methods using a generalized rough set and ensemble classifiers reached an overall accuracy of 96%, though false positives remained a challenge (31). In 2023, a Le-Net deep learning model was used for lung cancer classification, achieving impressive results with sensitivity of 93.14%, specificity of 95.91%, and accuracy of 97.88% (5). Another method, LCD-CapsNet, which uses CNNs and capsule networks, distinguished between benign and malignant conditions with 94% accuracy on the LIDC dataset, achieving 95% precision and 95% recall (24). Finally, a comparison of ResNet-50 and Inception V3 models on CT images revealed that ResNet-50 outperformed Inception V3 with an accuracy of 93.09% compared to 69.94% (18).

PROPOSED METHODOLOGY

We frame the problem of lung nodule diagnosis as a supervised binary classification task where each CT image is labeled as benign or malignant. The proposed methodology follows a well-defined computational pipeline comprising image preprocessing, data augmentation, feature extraction via pre-trained convolutional neural networks, hybrid feature-level fusion, ensemble prediction and final performance evaluation. Let the dataset be denoted by:

$$\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{N}, \tag{3}$$

where x_i is the i-th CT images, $y_i \in \{0,1\}$ is the corresponding class label and N is the total number of images. The proposed model tries to learn a robust classification function which minimizes the difference between the predicted label $\hat{y}_i$ and the ground-truth label y_i.

Experimental Setup

The proposed ensemble model was trained and evaluated using a carefully designed experimental setup to ensure robust, reproducible and computationally efficient learning. The hyperparameters are all listed in Table 1. The input CT images were resized to $224 \times 224 \times 3$ pixels to fit the input size of the pre-trained MobileNetV1, MobileNetV2, and VGG19 architectures. The model was trained with a batch size of 16, which is a good trade-off between memory consumption and gradient estimation stability. The **Adam optimizer** was used with a fixed learning rate of **0.001** and its adaptive moment estimation capabilities to ensure stable and efficient convergence throughout the optimization process. The model was trained for **200 epochs**, with a maximum of **83 iterations per epoch**, providing adequate exposure to the training data for the model to learn discriminative patterns associated with benign and malignant lung nodules. The training objective is the **binary cross-entropy** loss function, defined as:

$$\mathscr{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right], \tag{4}$$

where N is the total number of training samples, y_i is the ground-truth class label and $\hat{y}_i$ is the predicted malignancy probability. The model is trained to minimize this objective function, thereby reducing the discrepancy between its predictions and the benign-malignant labels. **ReLU** activation was used in all hidden dense layers to introduce non-linearity. In contrast, **sigmoid** activation function was used in the output layer to get a malignancy probability in the range of $[0,1]$. The binary classification used a decision threshold $\tau = 0.5$. The ensemble fusion weight β and the Hybrid MobileNet fusion weight α were both set to **0.5** so that all the constituent models contributed equally. Each branch has a fully connected dense layer of **128 neurons**, which provides sufficient representational power without overcomplicating the model.

All pre-trained backbone weights were initialized from **ImageNet** pre-training, and the convolutional bases were kept frozen during feature extraction to preserve learned visual representations and reduce the number of trainable parameters. To speed up training and enable fast model evaluation, we trained all experiments on the **Google Colab** platform using a **Tesla T4 GPU with 25 GB of RAM**. Further reductions in training time and improved scalability for the LIDC-IDRI dataset evaluation were achieved by parallelizing computationally intensive operations using multithreading.

Dataset Description

We used a large set of thoracic CT scans from the LIDC-IDRI dataset for this study, which is intended for lung tumor detection and analysis. It contains two types of lung nodules, benign and malignant. The dataset is divided into three parts: training, validation and testing. The training subset has 1,323 images, which were used to train the model and help it learn the patterns and features of lung nodules. The validation set of 330 images was used to optimize the model hyperparameters and assess the model performance during training. The model's final performance was evaluated on the test set, which consisted of 413 images. The dataset contains 2,066 images and is a useful resource for developing accurate lung cancer detection models. This dataset is publicly available on Kaggle (unk) and is an important resource for advancing research on lung cancer detection.

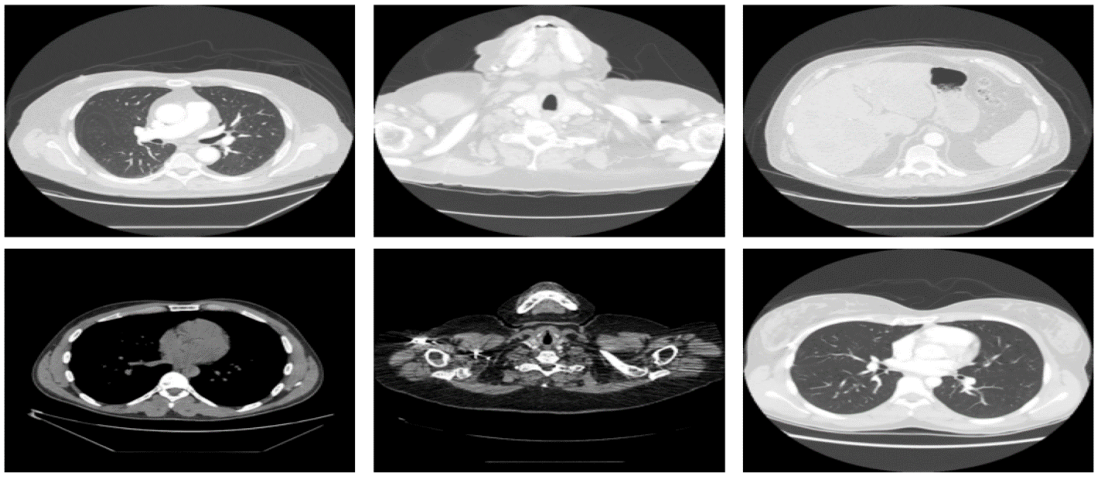


Figure 1. Original Images from the LIDC-IDRI Dataset

Data Preprocessing

In this study, the publically available LIDC-IDRI dataset was used for lung cancer classification. It contains thoracic CT scans with annotated lesions labeled as benign or malignant. The dataset is divided into train, validation and test sets. All images were resized to 224×224 pixels to ensure compatibility. This is the normal input size required by the model architecture. Pixel values were normalized by scaling

to the range of [0,1] for uniformity across all images. For each input image x_i, we resize and normalize the image to obtain a standardized image representation $\bar{x}_i$. The normalization is defined as:

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{5}$$

where $x_{\min}$ and $x_{\max}$ represent the minimum and maximum pixel intensity values, respectively. This transformation maps the pixel values into a normalized range and reduces intensity variation across CT images. Additionally, mean subtraction was performed using pre-set ImageNet mean values to standardize the data and facilitate the model's learning process.

Hyperparameter	**Value**
Input Image Size	$224 \times 224 \times 3$
Batch Size	16
Learning Rate	0.001
Optimizer	Adam
Number of Epochs	200
Loss Function	Binary Cross-Entropy
Activation Function	ReLU (hidden), Sigmoid (output)
Dense Layer Neurons	128
Ensemble Fusion Weight (β)	0.5
Hybrid MobileNet Weight (α)	0.5
Hardware	Tesla T4 GPU (25 GB RAM)

Table 1. Hyperparameter settings used for training the proposed ensemble model.

Table 2 summarizes the distribution of the LIDC-IDRI dataset across training, validation, and testing subsets. The dataset comprises a total of 2,066 CT images, split into 1,323 training, 330 validation, and 413 testing samples, with approximately equal class distribution between benign and malignant nodules in each subset.

Class	**Training**	**Validation**	**Testing**	**Total**
Benign	662	165	206	1033
Malignant	661	165	207	1033
Total	**1323**	**330**	**413**	**2066**

Table 2. Dataset split of the LIDC-IDRI dataset used for model training, validation, and testing.

Data Augmentation Techniques

To improve the dataset's robustness and diversity, several data augmentation approaches were used. Improving model generalization and preventing overfitting, particularly in medical imaging tasks where the dataset is often small. The augmentation pipeline included vertical and horizontal flipping to introduce spatial variations, ensuring the model remains invariant to the scan orientation. Random rotations within ±20 degrees were applied to simulate various patient and imaging orientations. Brightness and contrast adjustments were made to mimic variations in scanner settings and lighting conditions. Small shifts, scaling of 10%, and rotations within ±15 degrees were also applied to simulate common clinical imaging errors, such as patient positioning. Additionally, elastic transformations were added to simulate natural anatomical and soft tissue variations. This augmentation process generated four augmented versions of

each input image, expanding the dataset significantly and enhancing the model's generalization to new, untested data. The augmentation process can be expressed as a transformation function applied to each preprocessed image:

$$x_i^{(a)} = \mathscr{T}_a(\bar{x}_i), \tag{6}$$

where $\mathscr{T}_a(\cdot)$ denotes an augmentation operation such as flipping, rotation, shifting, scaling, contrast adjustment, or elastic transformation, and $x_i^{(a)}$ represents the augmented image. These transformations increase sample diversity and improve the model's ability to generalize to unseen CT scans.

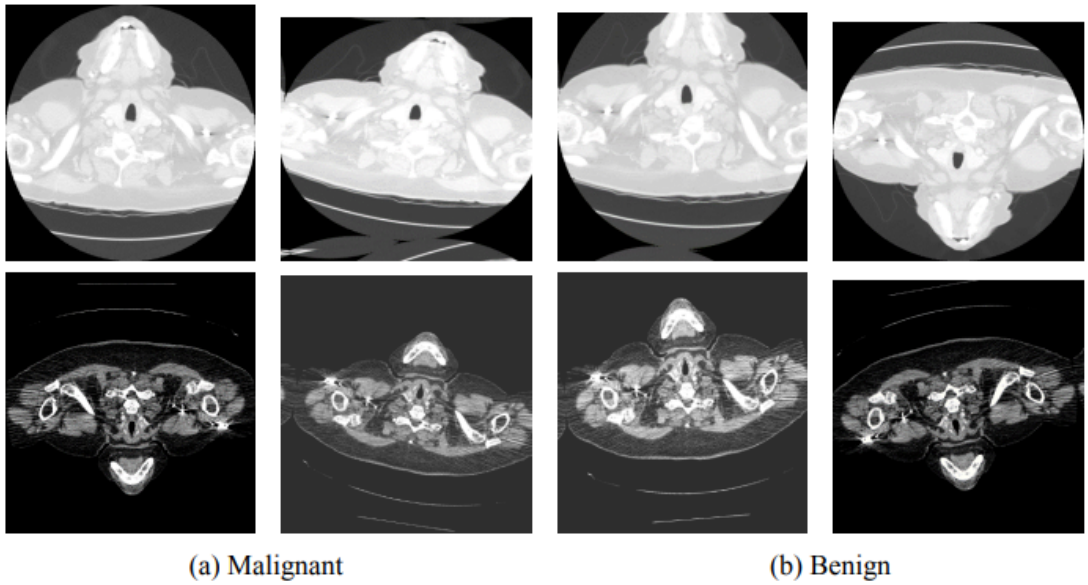


Figure 2. Augmented training samples generated from the LIDC-IDRI dataset: (a) malignant nodule variants (left) and (b) benign nodule variants (right), demonstrating spatial, photometric, and geometric transformations applied to enhance model generalization.

Proposed Architecture

The proposed model architecture is designed to classify lung CT images into benign and malignant categories by combining lightweight and deep convolutional feature extractors within an ensemble learning framework. The overall pipeline begins with the acquisition of lung CT images, followed by preprocessing operations such as resizing, normalization, and augmentation. The preprocessed images are then passed through three pre-trained convolutional neural networks: MobileNetV1, MobileNetV2, and VGG19.

Let $\bar{x}_i$ denote the preprocessed CT image. The feature extraction process of MobileNetV1, MobileNetV2, and VGG19 can be represented as:

$$z_i^{(1)} = \phi_1(\bar{x}_i), \qquad z_i^{(2)} = \phi_2(\bar{x}_i), \qquad z_i^{(3)} = \phi_3(\bar{x}_i), \tag{7}$$

where $\phi_1(\cdot)$, $\phi_2(\cdot)$, and $\phi_3(\cdot)$ denote the feature extraction functions of MobileNetV1, MobileNetV2, and VGG19, respectively. The corresponding feature representations are denoted by $z_i^{(1)}$, $z_i^{(2)}$, and $z_i^{(3)}$.

MobileNetV1 and MobileNetV2 are first combined to form the Hybrid MobileNet (V1+V2) branch. MobileNetV1 provides a simple and computationally efficient structure based on depthwise separable convolutions, while MobileNetV2 improves feature representation through inverted residual blocks and linear bottlenecks. Their fusion allows the model to retain computational efficiency while improving feature diversity. The hybrid representation is defined as:

$$z_i^{(H)} = \alpha z_i^{(1)} + (1-\alpha) z_i^{(2)}, \tag{8}$$

where $z_i^{(H)}$ represents the fused Hybrid MobileNet feature representation and $\alpha \in [0,1]$ controls the contribution of MobileNetV1 and MobileNetV2. In this study, equal contribution is considered, with $\alpha = 0.5$.

In parallel, VGG19 is used as a deeper feature extractor because of its ability to capture hierarchical and high-level spatial patterns from medical images. While the Hybrid MobileNet branch focuses on efficient feature learning, the VGG19 branch provides complementary deep visual representations that are useful for distinguishing subtle differences between benign and malignant nodules.

The final ensemble model combines the outputs of the Hybrid MobileNet branch and the VGG19 branch. Let $p_i^{(H)}$ and $p_i^{(V)}$ denote the predicted malignancy probabilities produced by Hybrid MobileNet and VGG19, respectively. The final ensemble prediction is calculated as:

$$\hat{y}_i = \beta p_i^{(H)} + (1-\beta) p_i^{(V)}, \tag{9}$$

where $\hat{y}_i$ denotes the final predicted probability of malignancy and $\beta \in [0,1]$ controls the contribution of both branches. For equal-weight ensemble learning, $\beta = 0.5$.

The predicted class label is obtained using a binary decision rule:

$$\tilde{y}_i = \begin{cases} 1, & \hat{y}_i \geq 0.5, \\ 0, & \hat{y}_i < 0.5, \end{cases} \tag{10}$$

where $\tilde{y}_i = 1$ represents a malignant nodule and $\tilde{y}_i = 0$ represents a benign nodule. This ensemble strategy allows the proposed model to benefit from the computational efficiency of MobileNet-based architectures and the strong hierarchical feature extraction capability of VGG19. As a result, the final architecture provides an efficient and reliable solution for lung nodule classification in CT images.

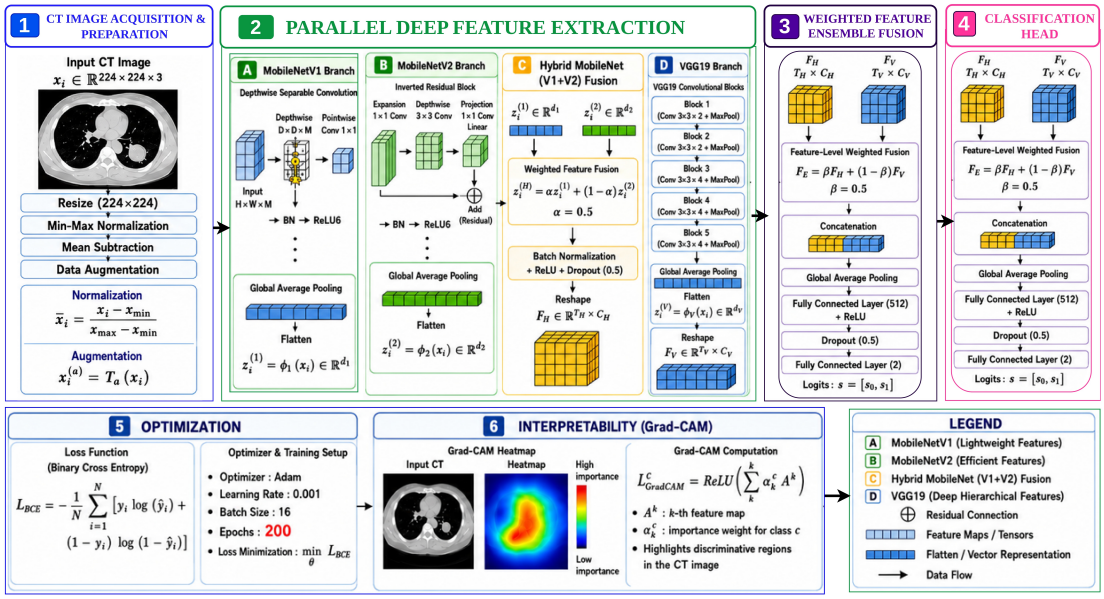


Figure 3. End-to-end proposed ensemble architecture for benign and malignant lung nodule classification.

HYBRID MOBILENET (V1+V2) MODEL ARCHITECTURE

The Hybrid MobileNet (V1+V2) architecture combines MobileNetV1 and MobileNetV2 to obtain an efficient and discriminative convolutional neural network for benign and malignant lung nodule classification. The input CT images are resized to $224 \times 224 \times 3$ and are passed separately through the MobileNetV1 and MobileNetV2 branches. In both branches, the convolutional base is initialized with pre-trained weights and kept frozen during feature extraction. This preserves the learned visual representations of the base models while reducing the number of trainable parameters and improving computational efficiency.

Let $\bar{x}_i$ denote the preprocessed input CT image. The feature representations extracted by MobileNetV1 and MobileNetV2 are defined as:

$$z_i^{(1)} = \phi_1(\bar{x}_i), \qquad z_i^{(2)} = \phi_2(\bar{x}_i), \tag{11}$$

where $\phi_1(\cdot)$ and $\phi_2(\cdot)$ represent the feature extraction functions of MobileNetV1 and MobileNetV2, respectively. The extracted convolutional features are flattened and then passed through a fully connected dense layer with 128 neurons and ReLU activation. Each branch uses a final sigmoid output layer to produce a probability score for binary classification.

The outputs of MobileNetV1 and MobileNetV2 are combined to construct the Hybrid MobileNet representation. The fusion operation is expressed as:

$$z_i^{(H)} = \alpha z_i^{(1)} + (1-\alpha) z_i^{(2)}, \tag{12}$$

where $z_i^{(H)}$ denotes the fused Hybrid MobileNet feature representation and $\alpha \in [0,1]$ controls the contribution of both MobileNet branches. In this study, equal weighting is used with $\alpha = 0.5$, allowing MobileNetV1 and MobileNetV2 to contribute equally to the hybrid representation. This fusion strategy improves feature diversity while preserving the lightweight nature of the MobileNet family. Figure 4 illustrates the workflow of the proposed Hybrid MobileNet (V1+V2) architecture.

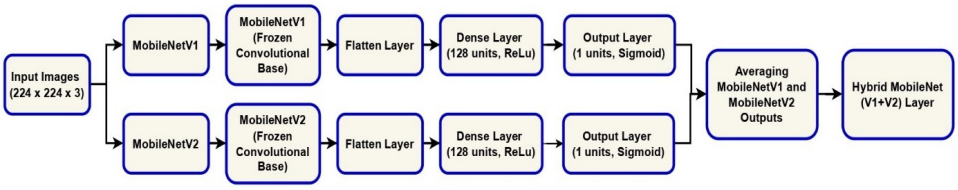


Figure 4. Workflow of the Hybrid MobileNet (V1+V2) architecture.

MobileNetV1 uses depthwise separable convolutions, where a standard convolution is decomposed into a depthwise convolution followed by a pointwise convolution. This design significantly reduces computational complexity while preserving useful spatial features. MobileNetV2 improves on this structure with inverted residual blocks and linear bottlenecks. It first expands the input representation using a 1×1 convolution, then applies depthwise convolution and finally projects the representation back to a lower-dimensional space with another 1×1 convolution. Residual connections are employed when input and output dimensions are compatible, helping the network preserve useful information and maintain stable feature propagation. Therefore, the combination of MobileNetV1 and MobileNetV2 provides a balanced architecture that supports efficient inference and strong feature extraction for lung nodule classification. Figure 5 presents the individual architectural components of MobileNetV1 and MobileNetV2.

VGG19 Model Architecture

VGG19 is a deep convolutional neural network developed by the Visual Geometry Group at the University of Oxford. It follows a simple, uniform architecture based on repeated convolutional blocks, small 3×3 convolutional kernels, ReLU activation functions, and max-pooling layers for spatial downsampling. The architecture contains 16 convolutional layers followed by three fully connected layers, making a total of 19 weight layers.

In this study, VGG19 is used as a complementary deep feature extractor for lung nodule classification. While the MobileNet-based branch provides computational efficiency, VGG19 captures deeper hierarchical image representations that are useful for identifying subtle visual differences between benign and malignant nodules. For a preprocessed input CT image $\bar{x}_i$, the VGG19 feature extraction process can be expressed as:

$$z_i^{(V)} = \phi_V(\bar{x}_i), \tag{13}$$

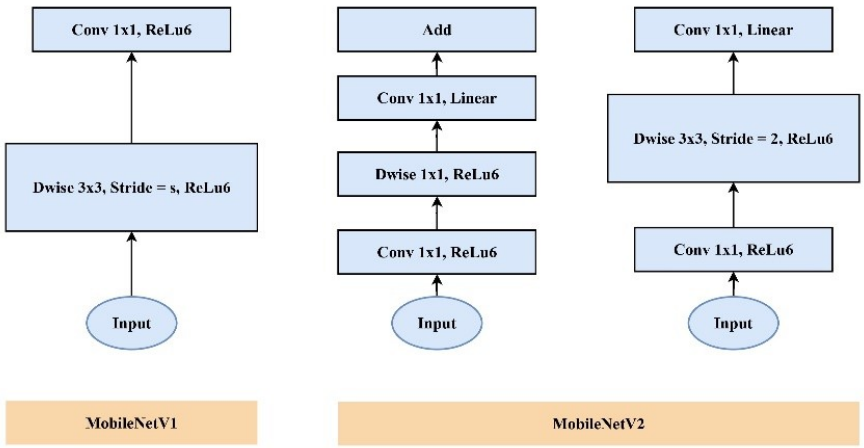


Figure 5. Individual architectural components of MobileNetV1 (left) and MobileNetV2 (right)

where $\phi_V(\cdot)$ denotes the VGG19 feature extraction function and $z_i^{(V)}$ represents the extracted deep feature representation.

After feature extraction, the VGG19 branch produces a malignancy probability using a sigmoid activation function:

$$p_i^{(V)} = \sigma(W_V z_i^{(V)} + b_V), \tag{14}$$

where W_V and b_V denote the trainable weight matrix and bias term of the classification layer, respectively, and $\sigma(\cdot)$ represents the sigmoid activation function. The output $p_i^{(V)} \in [0,1]$ indicates the probability that the input CT image belongs to the malignant class.

Using VGG19 strengthens the ensemble model by providing deeper spatial and semantic features. These features complement the lightweight representations learned by MobileNetV1 and MobileNetV2, thereby improving the final benign and malignant classification performance.

Final Ensemble Model Architecture

The final ensemble model combines the predictive outputs of Hybrid MobileNet (V1+V2) and VGG19 to improve benign- malignant lung nodule classification. The input image, of size $224 \times 224 \times 3$, is processed independently by the Hybrid MobileNet and VGG19 branches. The Hybrid MobileNet branch provides lightweight, efficient feature representations, while the VGG19 branch captures deeper hierarchical image features. The predictions from both branches are then combined using an equal-weight averaging strategy.

Let $p_i^{(H)}$ denote the predicted malignancy probability generated by the Hybrid MobileNet branch, and let $p_i^{(V)}$ denote the predicted malignancy probability generated by the VGG19 branch. The final ensemble prediction is defined as:

$$\hat{y}_i = \beta p_i^{(H)} + (1-\beta) p_i^{(V)}, \tag{15}$$

where $\hat{y}_i$ represents the final predicted probability of malignancy, and $\beta \in [0,1]$ controls the contribution of both branches. In this study, equal contribution is used with $\beta = 0.5$.

The final class label is obtained using a binary decision rule:

$$\tilde{y}_i = \begin{cases} 1, & \hat{y}_i \geq 0.5, \\ 0, & \hat{y}_i < 0.5, \end{cases} \tag{16}$$

where $\tilde{y}_i = 1$ represents a malignant nodule and $\tilde{y}_i = 0$ represents a benign nodule.

During training, the model is optimized using binary cross-entropy loss, which is suitable for two-class classification tasks. The objective function is expressed as:

$$\mathscr{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right], \tag{17}$$

where N denotes the total number of training samples, y_i is the actual class label, and $\hat{y}_i$ is the predicted malignancy probability. By minimizing this loss, the ensemble model learns to reduce the difference between the predicted and true benign and malignant labels.

The proposed ensemble model achieved accuracy, precision, recall, and F1-score values of 96.13%, 97.10%, 98.57%, and 96.16%, respectively. These results indicate that combining the computational efficiency of Hybrid MobileNet with the deeper feature extraction capability of VGG19 improves the reliability of lung nodule classification. Figure 6 presents the workflow architecture of the proposed ensemble learning framework.

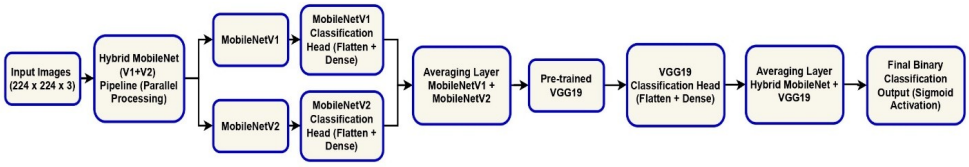


Figure 6. Workflow architecture of the proposed Hybrid MobileNet (V1+V2) and VGG19 ensemble model.

Results

The proposed Hybrid MobileNet (V1+V2) + VGG19 ensemble model was evaluated for benign and malignant lung nodule classification using accuracy, precision, recall, F1-score, confusion matrix analysis, Receiver Operating Characteristic (ROC) curve analysis, Area Under the Curve (AUC), Grad-CAM visualization, and execution time comparison. This model combines the computational efficiency of the Hybrid MobileNet branch and the deeper hierarchical feature-extraction capability of VGG19. This combination allows the ensemble model to learn both lightweight discriminative features and high-level spatial representations of lung CT images.

The classification performance was evaluated in terms of standard evaluation metrics calculated from confusion matrix. Accuracy is the fraction of correctly classified samples over all samples. It is given by:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{18}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives and false negatives respectively.

Precision and recall are defined as:

$$Precision = \frac{TP}{TP+FP}, \qquad Recall = \frac{TP}{TP+FN}. \tag{19}$$

Precision is the ratio of correctly identified malignant cases to the total number of cases identified as malignant, whereas recall is the number of actual malignant cases correctly identified by the model. The F1-score is a measure that considers both precision and recall and is defined as:

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{20}$$

The proposed ensemble model obtained 96.13% accuracy, 97.10% precision, 98.57% recall, 98.6% specificity, 0.493 Matthews Correlation Coefficient (MCC) and 96.16% F1-score. The high accuracy indicates that the model can correctly classify most benign and malignant lung CT images. The high precision means the model is better at reducing false-positive predictions, which is important to avoid benign nodules being falsely predicted as malignant. The recall value indicates that the model can detect a

large proportion of malignant cases, which is clinically significant, as missing malignant nodules can lead to delayed diagnosis and treatment. Therefore, these results generally indicate that the proposed ensemble model provides a reliable trade-off between sensitivity and specificity for lung nodule classification.

The confusion matrix provides a detailed view of the classification behavior of the proposed model. It correctly classified 197 true-positive and 200 true-negative cases. There were 6 false-positive predictions, where benign cases were misclassified as malignant, and 10 false-negative predictions, where malignant cases were misclassified as benign. These values suggest that the model performs well at distinguishing benign from malignant nodules, but there might be room for further improvement to reduce false-negative cases. The obtained confusion matrix by the predictions on the test set is shown in the Figure 7.

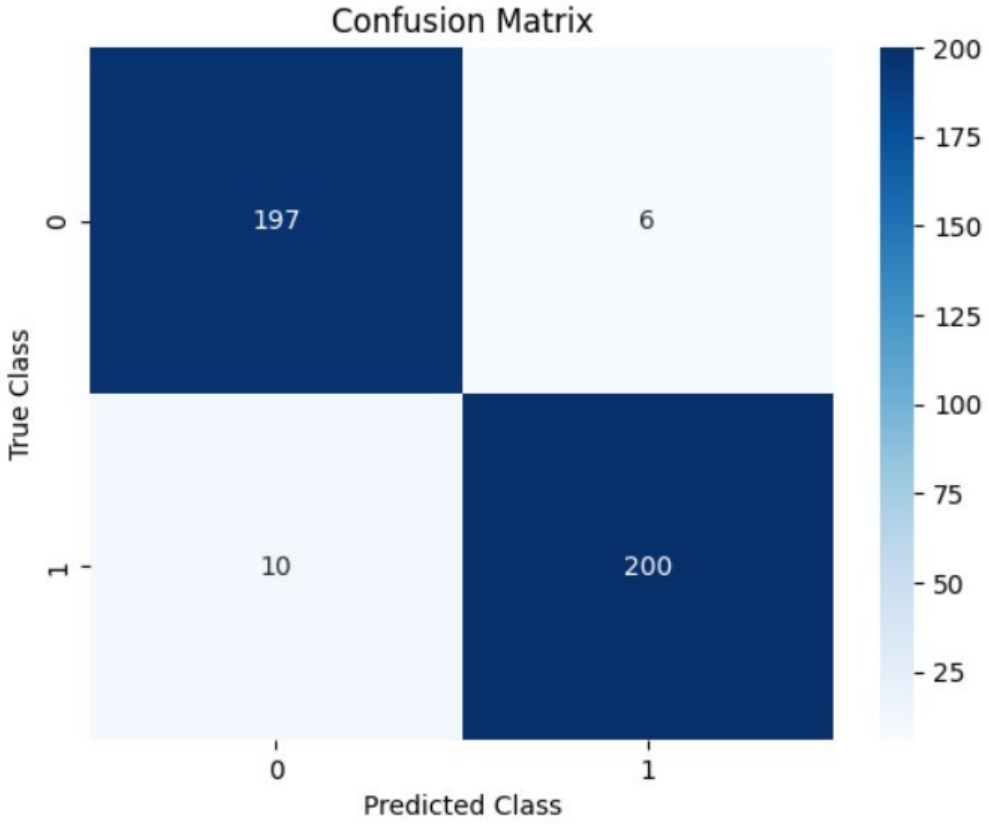


Figure 7. Confusion matrix of the proposed ensemble model on the LIDC-IDRI test set

The ROC curve was used to evaluate the discriminative ability of the proposed model with different classification thresholds. The ROC curve plots the true positive rate versus the false positive rate, demonstrating the model's ability to distinguish between benign and malignant classes. A curve closer to the upper-left corner indicates stronger classification performance. The proposed model achieved an AUC of 0.989, indicating excellent discrimination between benign and malignant lung nodules. The ROC curve of the proposed model is shown in Figure 8.

Grad-CAM was used to improve the interpretability of the proposed model by highlighting the regions of the CT image that contributed most strongly to the classification decision. This technique uses the gradients of the target class with respect to the final convolutional feature maps to generate a class-discriminative heatmap. The Grad-CAM heatmap is computed as:

$$L^{c}_{\text{Grad-CAM}} = ReLU\left(\sum_{k} \omega^{c}_{k} A^{k}\right), \tag{21}$$

where A^k denotes the k-th feature map and ω^c_k represents the importance weight of that feature map for class c. The resulting heatmap highlights the regions that have the strongest influence on the model's prediction. In this study, Grad-CAM was applied to lung CT images to visualize the regions associated

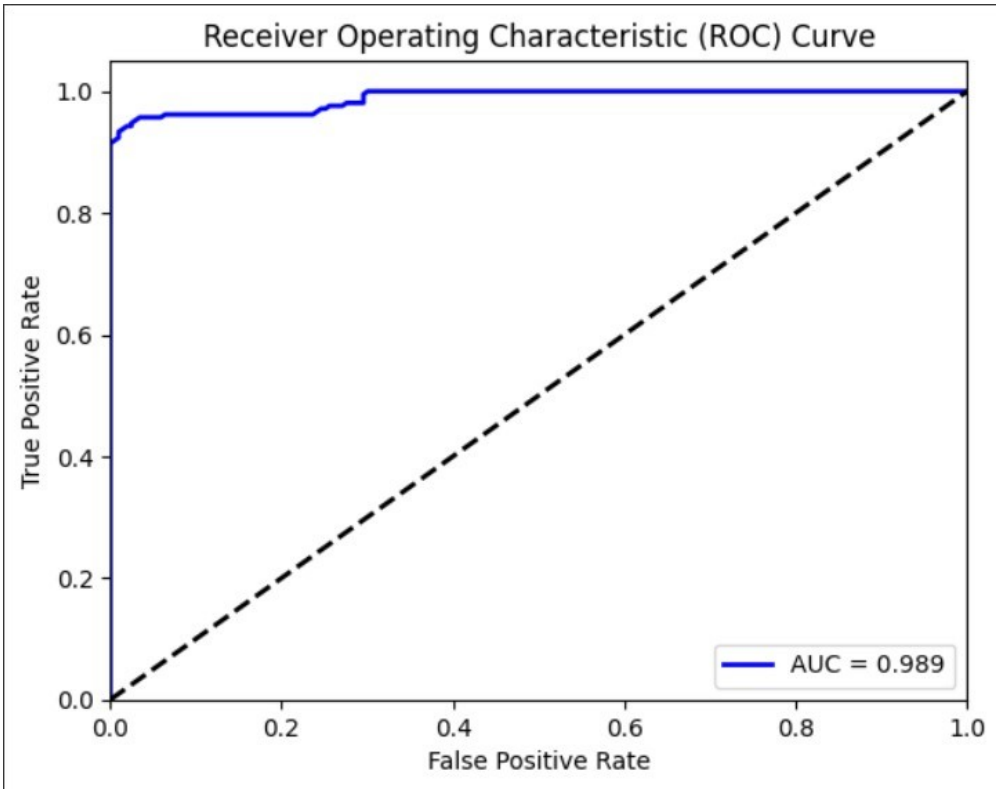
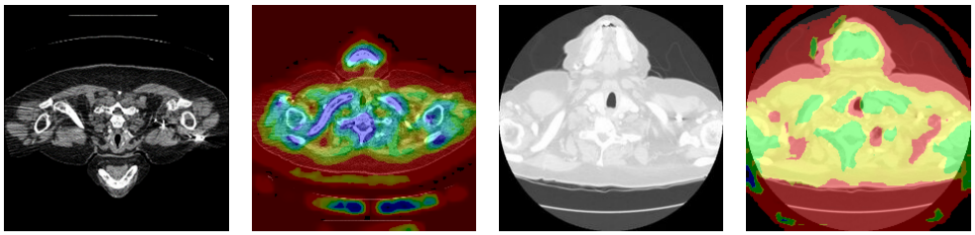


Figure 8. Receiver Operating Characteristic (ROC) curve of the proposed ensemble model.

with benign and malignant classification decisions. The Grad-CAM visualization results are shown in Figure 9.



(a) Benign (b) Malignant

Figure 9. Grad-CAM heatmap visualizations overlaid on benign (a) and malignant (b) lung CT images, highlighting the discriminative regions most influential in the model's classification decisions and confirming its clinically interpretable focus on nodule-relevant areas.

Table 3 presents a comparative analysis of top-performing backbone architectures in terms of training time (in seconds) and memory footprint (in MB). Among the evaluated models — including UNet, LinkNet, FPN, and PSPNet, each paired with various backbones such as EfficientNet-b3, Xception, InceptionV4, and DenseNet121 — the proposed model demonstrates clear superiority across both efficiency metrics. In particular, the proposed architecture achieves the shortest training time of 1347.89 seconds, which is approximately 2.1× faster than the next-best model (FPN-EfficientNet-b3 at 6315.67s) and more than 7× faster than the slowest configuration (LinkNet-InceptionV4 and PSPNet-InceptionV4 at 9527.25s). The proposed model is also memory-efficient with only 16.6 MB of memory, the smallest footprint of all the compared architectures, beating even PSPNet-DenseNet121 (31.7 MB) by almost half. The results indicate that the proposed model is computationally efficient and well-suited for deployment in resource-constrained environments without compromising the architecture's performance.

Model-Backbone	Training Time (s)	Memory Footprint (MB)
UNet-efficientnet-b3	7427.580034	53.3
LinkNet-xception	7649.485346	109.4
UNet-xception	8914.528674	115.5
FPN-efficientnet-b3	6315.674771	50.5
LinkNet-InceptionV4	9527.253391	185.4
FPN-xception	2816.464562	93.3
PSPNet-InceptionV4	9527.253391	167.4
PSPNet-Densenet121	4471.666690	31.7
Proposed	**1347.888067**	**16.6**

Table 3. Comparison of top-performing backbone architectures based on training time and memory footprint

We summarize a detailed comparison of the performance of the proposed model against existing research on six standard evaluation metrics: Recall, Specificity, Accuracy, Precision, F1-Score, and MCC in Table 4.

Overall Performance of the Proposed Model

The proposed model achieves competitive and, in several cases, superior results across most of the metrics, recording a Recall of **98.57%**, Specificity of **98.6%**, Accuracy of **96.13%**, Precision of **97.10%**, F1-Score of **96.16%** and an MCC of **0.493**. Here is the Metric-wise Analysis of our proposed model:

- The recall score (98.57%) of our proposed model outperforms most of the other methods, e.g., (21) (53.25%), (13) (85.1%), and (32) (75.2%) and is comparable with (20) (92.5%) and (24) (93.00%) indicating that it can correctly identify the positive cases and reduce the false negative cases.
- The proposed model achieves the highest specificity of **98.6%** among all compared works that reported this metric, outperforming (20) (95.8%), (13) (97.4%), (28) (97.30%), and (32) (96.2%). This, in turn, shows the model's extraordinary potential to correctly identify negative cases, which is critical in medical or safety-sensitive applications.
- The best performer is (21) with **99.81%** accuracy. Nevertheless, it is worth mentioning that (21) has a very low Recall (53.25%) and F1-Score (58.55%), indicating a class imbalance problem, where high accuracy is achieved by mostly predicting the majority class. The balance across all metrics, with an accuracy of **96.13%**, makes the proposed model a more reliable and trustworthy result.
- The proposed model has the highest precision of **97.10%** as compared to all the compared references including (33) (92.0%), (20) (82.3%), and (13) (87.3%). This indicates that if the model predicts a positive case, it is highly trustworthy, with a very low false-positive rate.
- The proposed model shows the highest F1-Score of **96.16%** among all the existing works (20) (87.1%), (13) (86.2%), (33) (86.5%) and (32) (77.66%). As the F1-Score is the harmonic mean of Precision and Recall, this result confirms that the proposed model maintains an excellent trade-off between sensitivity and precision.
- The proposed model reports an MCC of **0.493** which is lower than (33) (0.840), (20) (0.845), and (13) (0.833). MCC is sensitive to the class distribution and balance of the dataset. The relatively

low MCC may indicate that the dataset used for the proposed model has some degree of class imbalance or that the MCC penalizes trade-offs differently across thresholds. Nevertheless, the high F1-Score and Precision values validate the model's real-world efficacy.

Key Takeaways

The proposed model generally has a good, robust performance profile. It achieves the best Specificity, Precision and F1-Score among all compared methods and remains very competitive in Recall and Accuracy. While the MCC is relatively modest, the holistic metric performance strongly supports the reliability and generalizability of the proposed model, especially in situations where both false positives and false negatives are consequential.

Ref.	Recall	Specificity	Accuracy	Precision	F1-Score	MCC
(35)	93.97	89.83	88.79	–	–	–
(21)	53.25	–	**99.81**	65.02	58.55	0.59
(33)	81.60	98.5	95.56	92.0	86.5	0.840
(24)	93.00	92.10	92.90	–	–	–
(20)	92.5	95.8	95.25	82.3	87.1	0.845
(13)	85.1	97.4	95.25	87.3	86.2	0.833
(28)	**96.00**	97.30	97.20	–	–	–
(32)	75.2	96.2	92.47	80.3	77.66	0.732
Proposed	98.57	**98.6**	96.13	**97.10**	**96.16**	**0.493**

Table 4. Performance comparison of the proposed model with existing research works

Discussion

The proposed Hybrid MobileNet (V1+V2) + VGG19 ensemble model exhibits robust and consistent performance in classifying benign and malignant lung nodules on the LIDC-IDRI dataset. The ensemble approach of combining lightweight MobileNet-based architectures with the deeper hierarchical feature extractor VGG19 is a pragmatically efficient design decision, yielding an accuracy of 96.13%, a precision of 97.10%, a recall of 98.57%, an F1-score of 96.16%, a specificity of 98.6%, and an AUC of 0.989. The main advantage of the proposed model is its ability to optimize computational efficiency and classification reliability simultaneously. As shown in Table 3, the proposed model achieves the lowest training time (1347.89 s) and the lowest memory footprint (16.6 MB) among all backbone architectures, making it significantly more practical for deployment in resource-constrained clinical environments. This efficiency advantage does not come at the expense of predictive performance, unlike heavier architectures that achieve marginal accuracy gains at substantially higher computational cost.

The high precision (97.10%) is particularly relevant for clinical practice, as it indicates that the model makes very few false-positive predictions, reducing the likelihood of unnecessary follow-up procedures or patient anxiety from misclassified benign nodules. Correspondingly, the 98.57% recall can prove the ability of the model to detect the majority of true malignant cases, which is critical for early-stage lung cancer detection, considering that missed diagnosis can have a considerable impact on the patient outcome (32). The specificity of 98.6%, the highest among all methods compared in Table 4, further supports the model's ability to rule out malignancy in benign cases correctly. This trade-off between sensitivity and specificity is reflected in the AUC of 0.989, indicating near-perfect discrimination ability across classification thresholds. This is supported by the confusion matrix analysis, which shows only 6 false positives and 10 false negatives out of 413 test samples, indicating a well-calibrated and reliable model.

The MCC of 0.493 is relatively lower than those reported in (33) (0.840), (20) (0.845), and (13) (0.833), and should be interpreted with caution. MCC is a metric that is highly sensitive to the distribution of all four entries in the confusion matrix. MCC can be disproportionately affected by class imbalance or differences in dataset size and composition. Since we observe high F1-score and high precision

simultaneously for the proposed model, the MCC discrepancy could be due to differences in the test set class distribution rather than a lack of discriminative ability in the proposed model. This observation is in line with the fact that (21) reports 99.81% accuracy with only 53.25% recall and 58.55% F1-score. It reveals that using aggregate metrics alone can be misleading without considering class balance. The Grad-CAM visualization further supports the model's clinical interpretability by confirming that the classification decisions are based on diagnostically relevant regions of the CT image rather than background artefacts or noise. Such transparency is essential for gaining physicians' trust and regulators' acceptance of AI-supporting diagnostic tools in clinical workflows (28).

The practical scalability of the proposed approach is also achieved through multithreaded training optimization during model development. This is a major advantage for adapting the system to diverse clinical environments with different scanner protocols and patient demographics, as retraining or fine-tuning the model on new institutional datasets incurs less overhead due to the parallelization of computationally intensive operations. Compared to ensemble methods in previous studies, such as the VGG, Xception and ResNet ensemble (13) and the VGG16, ResNet50V2 and DenseNet201 ensemble that resulted in 91% accuracy, the proposed model provides a more specific and compact fusion method by merging MobileNet variants, which are efficiency-centric, with the deep representational power of VGG19. This design is more consistent with the practical requirements of clinical deployment, where inference speed, memory efficiency and predictive reliability must be balanced concurrently.

CONCLUSIONS AND FUTURE WORKS

In this work, a novel hybrid ensemble model was proposed for benign-malignant lung nodule classification on the LIDC-IDRI dataset. The model is improved by integrating MobileNetV1 and MobileNetV2 into a Hybrid MobileNet (V1+V2) branch, and then using an equal-weight ensemble with VGG19. The Hybrid MobileNet branch provides lightweight, computationally efficient feature representations, while VGG19 provides deeper hierarchical spatial features as a supplementary, robust classification system. The proposed model achieves an accuracy of 96.13%, precision of 97.10%, recall of 98.57%, F1-score of 96.16%, specificity of 98.6% and AUC of 0.989 on the LIDC-IDRI test set. It also achieves the shortest training time of 1347.89 s and the smallest memory footprint of 16.6 MB among all compared backbone architectures, showing that high classification performance can be achieved without sacrificing computational efficiency. The Grad-CAM visualization further enhances the model's clinical utility by providing interpretable heatmaps that pinpoint the most diagnostically relevant regions in lung CT images. The comparison with current methods shows that the proposed model achieves the best specificity, precision and F1-score among all works, while maintaining competitive recall and accuracy. Collectively, these results validate the proposed approach as a highly efficient, effective and interpretable tool for early lung cancer detection and clinical decision support.

The proposed approach can be extended and strengthened along several directions. First, although the LIDC-IDRI dataset provides a well-established benchmark, future work can evaluate the model on other publicly available and multi-institutional CT datasets to assess cross-domain generalizability and robustness to scanner variability. Second, the current binary classification scheme (benign vs. malignant) can be generalized to multi-class nodule characterization, comprising subtypes such as ground-glass opacity, part-solid, and solid nodules, each with different clinical implications. Third, the equal-weight averaging strategy for ensemble fusion ($\beta = 0.5$, $\alpha = 0.5$) is a fixed combination scheme, but adaptive or learned fusion weights, optimized via Bayesian optimization or meta-learning, can further improve classification performance. Fourth, using 3D volumetric analysis with CT rather than the current 2D slice-based approach can provide richer spatial context and nodule morphology across axial planes, potentially improving sensitivity for smaller or irregularly shaped nodules. Fifth, an important step towards clinical translation can be prospective clinical validation in a real-world radiology workflow, including radiologist agreement and decision time with and without AI assistance. Furthermore, we can also investigate model compression techniques, such as knowledge distillation and quantization, to further reduce the memory footprint and inference latency of the ensemble model for deployment on edge devices and low-resource healthcare infrastructure.

ACKNOWLEDGMENTS

So long and thanks for all the fish.

REFERENCES

[unk] Checking your browser - reCAPTCHA.

[2] (2021). Lung cancer classification system for ct images using deep convolutional neural network. In *Lecture Notes in Computer Science*, volume 10356, pages 412–423.

[3] (2022). Lung cancer classification using fuzzy hyperline segment neural network (fhlsnn). *International Journal of Advanced Trends in Computer Science and Engineering*, 8(3).

[4] (2026). Key statistics for lung cancer.

[5] Abdollahi, J. (2023). Evaluating lenet algorithms in classification lung cancer.

[6] Agarwal, A., Patni, K., and Rajeswari, D. (2021). Lung cancer detection and classification based on alexnet cnn. In *IEEE Conference on Computer Communication Technology*.

[7] AL-Huseiny, M. S. and Sajit, A. S. (2021). Transfer learning with googlenet for detection of lung cancer. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2):1078–1086.

[8] Al-Tawalbeh, J., Alshargawi, B., Alquran, H., et al. (2022). Classification of lung cancer using machine learning algorithms. In *Proceedings of IICETA*.

[9] Ardimento, P., Aversano, L., Bernardi, M. L., and Cimitile, M. (2021). Deep neural networks ensemble for lung nodule detection on chest ct scans. In *IEEE International Joint Conference on Neural Networks*.

[10] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P. Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Casteele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., and Clarke, L. P. (2015). Data from lidc-idri. Data set.

[11] Asuntha, A. and Srinivasan, A. (2020). Deep learning for lung cancer detection and classification. *Multimedia Tools and Applications*, 79(11):7731–7762.

[12] Benso, F. et al. (2026). Comprehensive characterization of oncogene-addicted non-small cell lung cancer and its immune microenvironment.

[13] Gogineni, A. K., Kishore, R., Raj, P., Naik, S., and Sahu, K. K. (2019). Unsupervised clustering algorithm as region of interest proposals for cancer detection using cnn. In *International Conference On Computational Vision and Bio Inspired Computing*, pages 1386–1396. Springer.

[14] Hossain, M., Ahmad, S., Hasan, S., Rahim, M., and Shin, J. (2024). Lung cancer detection and classification using lenet-lstm model on computed tomography images. *Multidisciplinary Sciences Journal*, 7(4):2025188.

[15] Hu, H., Li, Q., Zhao, Y., and Zhang, Y. (2021). Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors. *IEEE Transactions on Industrial Informatics*, 17:2880–2889.

[16] Huang, Q., Xu, N., and Xu, K. (2026). Development and validation of an inflammatory-immune-nutritional integrated scoring system: a novel strategy for predicting postoperative survival in non-small cell lung cancer. *PeerJ*, 14:e21122.

[17] Jena, S. R., Thomas, G. D., and Narain, P. (2021). Lung cancer detection and classification with dgmm-rbcnn technique. *Neural Computing and Applications*.

[18] Krishna, M. N. and Puviarasi, R. (2024). Resnet50-based lung cancer prediction using ct images compared with inceptionv3. In *AIP Conference Proceedings*, volume 2816.

[19] Krishnan, S. and Meganathan, N. T. (2026). Disease diagnosis and prediction using deep learning: a review. *PeerJ Computer Science*, 12:e3484.

[20] Lai, K. D., Nguyen, T. T., and Le, T. H. (2021). Detection of lung nodules on ct images based on the convolutional neural network with attention mechanism. *Annals of Emerging Technologies in Computing (AETiC)*, 5(2):78–89.

[21] Luo, D., Yang, I., Bae, J., and Woo, Y. (2024). Research on performance metrics and augmentation methods in lung nodule classification. *Applied Sciences*, 14(13):5726.

[22] Mamun, M., Farjana, A., Miraz, M. A., and Ahammed, S. (2022). Lung cancer prediction model using ensemble learning techniques and systematic review analysis. *IEEE AIOT*, 5(1):135–142.

[23] McMains, V. (2016). Johns hopkins study suggests medical errors are third-leading cause of death in

u.s.

[24] Nair, M., Svedberg, P., Larsson, I., and Nygren, J. M. (2024). A comprehensive overview of barriers and strategies for ai implementation in healthcare: Mixed-method design. *Plos one*, 19(8):e0305949.

[25] Naveen, V. and Manisundaram, M. (2023). Lung cancer detection using ensemble technique of cnn. In *Studies in Autonomous Data-driven Industrial Computing*, pages 57–61.

[26] Nooreldeen, R. and Bach, H. (2021). Current and future development in lung cancer diagnosis. *International Journal of Molecular Sciences*, 22(16):8661.

[27] Ormeño-Arriagada, P., Taramasco, C., Marquez, G., Araya, D., Alcocer, D., Toro, C., Gatica, G., and Hurtado, J. P. V. (2026). Comparison of deep learning and machine learning architectures for early oral cancer diagnosis. *PeerJ Computer Science*, 12:e3468.

[28] Ozdemir, O., Russell, R. L., and Berlin, A. A. (2019). A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans. *IEEE transactions on medical imaging*, 39(5):1419–1429.

[29] Phankokkruad, M. (2021). Ensemble transfer learning for lung cancer detection. In *Proceedings of 4th International Conference on Data Science and Information Technology*, pages 438–442.

[30] Rashed, M., Hossain, M., Mahdi, A., et al. (2025). Hybrid machine learning models for accurate type 2 diabetes mellitus prediction using a stacking classifier and a meta-model approach. *Cureus Journal of Computer Science*.

[31] Shakeel, P. M., Burhanuddin, M. A., and Desa, M. I. (2022). Automatic lung cancer detection from ct image using improved dnn and ensemble classifier. *Neural Computing and Applications*, 33:923–937.

[32] Song, Q., Zhao, L., Luo, X., and Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. *Journal of healthcare engineering*, 2017(1):8314740.

[33] Susan, S., Sethi, D., and Arora, K. (2023). Cross-domain learning for pulmonary nodule detection using gestalt principle of similarity: S. susan et al. *Soft Computing*, pages 1–12.

[34] Tang, G., Du, L., Ling, S., Che, Y., and Chen, X. (2024). Multi-type classification of lung nodules based on ct radiomics and ensemble learning for diversity weighting. *Quantitative Imaging in Medicine and Surgery*, 14(12):8942–8965.

[35] Tsuchiya, N., Kobayashi, S., Nakachi, R., Tomori, Y., Yogi, A., Iida, G., Ito, J., and Nishie, A. (2025). Application of a pulmonary nodule detection program using ai technology to ultra-low-dose ct: differences in detection ability among various image reconstruction methods. *Japanese Journal of Radiology*, 43(8):1303–1312.

[36] Venkatesh, C., Al-Wesabi, F. N., Alshahrani, H. M., Yahya, A. E., Kiran, A., Sivayamini, L., Alam, A., et al. (2026). An optimal diagnosis model for pancreatic cancer prediction based on deep learning method. *PeerJ Computer Science*, 12:e3541.