

TRUSTX-LUNG: EXPLAINABLE DEEP ENSEMBLE INTELLIGENCE FOR LUNG CANCER DETECTION

Pragnya Das¹, Satya Narayan Tripathy² and Kali Prasad Rath³

¹Department of Computer Science and Engineering, Centurion University of Technology and Management, India

²Department of Computer Science and Engineering, Berhampur University, India

³Department of Mathematics, NIST University, India

Abstract

This study is a proposal of an integrative framework using Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-agnostic Explanations (LIME) to provide trustworthy and explainable explanations in the context of a deep ensemble model designed to classify pulmonary carcinoma. Three different pre-trained models, ResNet50, DenseNet121, and EfficientNetB0, and a deep ensemble model were used in this research approach. The paper outlines the methodological pathway that starts with the preprocessing of the high-resolution computed tomography (CT) images, followed by image enhancement and feature extraction through convolution neural network architectures. As there is no clear suggestion of which pre-trained model gives the best performance as compared to others, an ensemble approach has been introduced. The ensemble approach is a combination of multiple deep learning models that combine predictions to provide better accuracy and less variance between the different model predictions. Two leading post-hoc explainability techniques, Grad-CAM and LIME, are used to make the model interpretable when applied to clinical situations. The experimental assessment performed on image datasets of lung cancer from the computed tomography (CT) modality demonstrates that ensemble learning with these XAI techniques increases the diagnostic outcomes and the transparency. The deep ensemble model achieved 0.972 of accuracy, while ResNet50, DenseNet121, and EfficientNetB0 achieved 0.94, 0.9358, and 0.94 of accuracy, respectively. The results clearly indicate the regions of interest and the diagnostic features that help in the detection of lung cancer and boost the clinical confidence in the system. Possible problems, including image quality differences and the computational load of ensemble methods, are also addressed, and methods of optimization and implementation are suggested. In conclusion, the results illustrate that the use of Grad-CAM and LIME in a deep ensemble framework is beneficial for achieving excellent prediction accuracy and delivering interpretable insights that are essential for clinical decision-making and follow-up treatment of patients.

Keywords:

Lung Cancer CT Imaging, ResNet50, DenseNet121, EfficientNetB0, Ensemble Learning, Grad-CAM, LIME, Model Interpretability

1. INTRODUCTION

In a heavily populated nation similar to India, abundant deaths occur due to inappropriate treatment and late identification of sickness. Cancer, a persistent disease, plays the lead role in this section. It is a challenging task for the physicians to identify it at before time stage. Lung cancer remains the next largely diagnosed cancer worldwide and the foremost source of cancer-related deaths, with over 2.2 million cases and 1.8 million deaths in 2020 alone [1]. The death tempo of pulmonary cancer surpasses breast cancer, and near 2030 it is approximate to attain 2.45 million worldwide. As per an assortment of study, 80% of pulmonary cancer deaths are due to smoking tobacco [2]. According to the

World Health Organization, lung cancer acts as a most important source of cancer-related death, through an estimated 2.21 million new cases and 1.8 million deaths each year [2]. Currently, AI plays a crucial role in the healthcare zone, mainly for diagnosing the disease as well as medical image analysis. Both machine learning and deep learning algorithms have been used for early diagnosis of sickness. Machine learning is used to discover insightful patterns and representations using meaningful input data [3]. Taking a decision based upon a single machine learning or deep learning algorithm for the forecast of cancer may not be considered as wise a decision, as the health check errors may charge the patient's life, and the accuracy is also lost too. Hence, ensemble learning approaches such as bagging, maxvoting, boosting, averaging, etc., are used to make the prediction accurate. Deep learning (DL) methods, basically convolution neural networks (CNNs), have achieved high-tech performance in lung cancer detection [4]. However, "black-box" DL models hinder clinical adoption due to lack of transparency [5]. Ensemble learning reduces predictive variance and enhances robustness by combining multiple DL architectures [6]. To bridge accuracy and interpretability, we integrate two leading post-hoc explainability methods: Grad-CAM (Gradient-weighted Class Activation Mapping) and LIME (Local Interpretable Model – agnostic Explanations) to elucidate quality assistance at both global and local levels.

This study has the following contributions that must be acknowledged.

- The proposed tri level hybrid system of heterogeneous CNN models which have pre-trained models of ResNet50, DenseNet121 and EfficientNetB0 are aimed at extracting important features and subsequently classifying lung cancer CT scan images. ResNet50 and DenseNet121 are added to the good residual learning and stability on deep representations and presented on the improvement of gradient flow, feature reuse of subtle lesion patterns, and compound scaling efficiency and lightweight computing with EfficientNetB0. This hybrid three-level ensemble improves discrimination of features, reduction of convergence bias, and far superior at the discrimination of at early-stage malignant nodules.
- Pipeline (median filter + Gaussian filter): Adaptive hybrid filtering was released with the goal of eliminating spackle and salt-and-pepper noise. This two-filter synergy's noise-free preprocessing output is important for optimizing CNN-enabled feature extraction and is anatomically intact.
- Images are improved using the Contrast Limited Adaptive Histogram Equalization (CLAHE) approach. In addition to improving photographs' aesthetic appeal, it also made it simpler to find and better analyze items. This localized

flexibility is particularly novel for enhancing micro-nodule visibility and ground-glass opacity (GGO) detection in lung CT images.

- The novelty is magnified by a dual explainable AI (XAI) module that can be used to explain two things. Grad-CAM focuses on the edges of the lesion and discriminative areas at the convolutional level, but at the decision space, LIME provides instance-level attribution of features. This combined reasoning mechanism enhances clinical acceptability and trust by giving multi-perspective interpretability by enabling radiologists to evaluate the local decision thinking and global model behavior. The enduring part of the manuscript is structured as shown in the Table.1.

Table.1. .Enduring part of the manuscript

Section	Description
2	Related Works
3	Proposed Methodology
4	Experimental Results
5	Comparison with benchmark dataset
6	Discussion
7	Potential challenges and Future work
8	Conclusion

2. RELATED WORKS

2.1 DEEP LEARNING FOR LUNG CANCER CLASSIFICATION

Prior work has applied CNNs such as ResNet and DenseNet to classify lung nodules, achieving accuracies above 90% on LIDC datasets [7]. By importing additional layers to the novel DenseNet framework and modifying the DenseNet201 model, Lanjewar et al. [8] projected a deep learning model for detecting lung cancer illness. The top features generated via the DenseNet201 predictive model using two feature selection methods have been evaluated by a number of machine learning classifiers. They have achieved 95% average accuracy using the 5-fold cross-validation technique. Sohaib et al. [9] offered a deep learning framework to classify the lung cancer CT scan imagery collected from Kaggle chest CT scan images. In addition, different other pre-trained models named VGG16, InceptionV3, and ResNet50 have been used to classify the images. The accuracy obtained by the proposed CNN model is 94.2% as compared to accuracies obtained from other pre-trained models. Alshamrani et al. [10] engaged a collaborative deep learning approach to classify the CT scan lung image together from Kaggle chest CT scan images as cancerous or non-cancerous type. This model included six feature patches to improve the network formerly trained with ResNet50. 93.24% accuracy is achieved.

2.2 ENSEMBLE LEARNING +XAI IN MEDICAL IMAGING

Both ensemble learning and explainable AI (XAI) played important roles in the advancement of lung cancer prediction in the modern era. Shah et al. [11] achieved 95.0% accuracy for

classifying the LUNA16 lung images using three 2D CNN models with Grad-CAM visualization. Li et al. [12] achieved 94.3% accuracy by applying a combination of three different models named ResNet50, SVM, and XGBoost along with a LIME-based feature-level explanation on the LIDC-IDRI dataset. Ahmed et al. [13] applied CNN with gradient boosting and saliency maps based on explainability on a private CT dataset of 600 samples. They got 95.28% accuracy. Banerjee et al. [14] got 93.7% accuracy using multi scale CNN and random forest for classifying the LIDC-IDRI dataset. For explanation of the result Grad-CAM techniques used. Shahin et al. [15] used LIME based explanation and achieved 92.5% accuracy on local CT dataset for classifying the images using CNN-SVM based model. Rahman et al. [16] used VGG16-Inceptionv3 for classification and Grad-CAM++ for explanation of the outcome. For their research approach, they used SPIE-AAPM lung dataset and achieved 96.2% accuracy.

3. PROPOSED METHODOLOGY

The common structural design of the planned model comprises data acquisition; image preprocessing, including resizing, filtering, and enhancement; dataset splitting; data augmentation of the training set; training and testing the pre-trained models ResNet50, DenseNet121, and EfficientNetB0; image classification uses the ensemble model; and explanation of results using explainable AI (XAI) techniques. The proposed framework is shown in Fig.1 below.

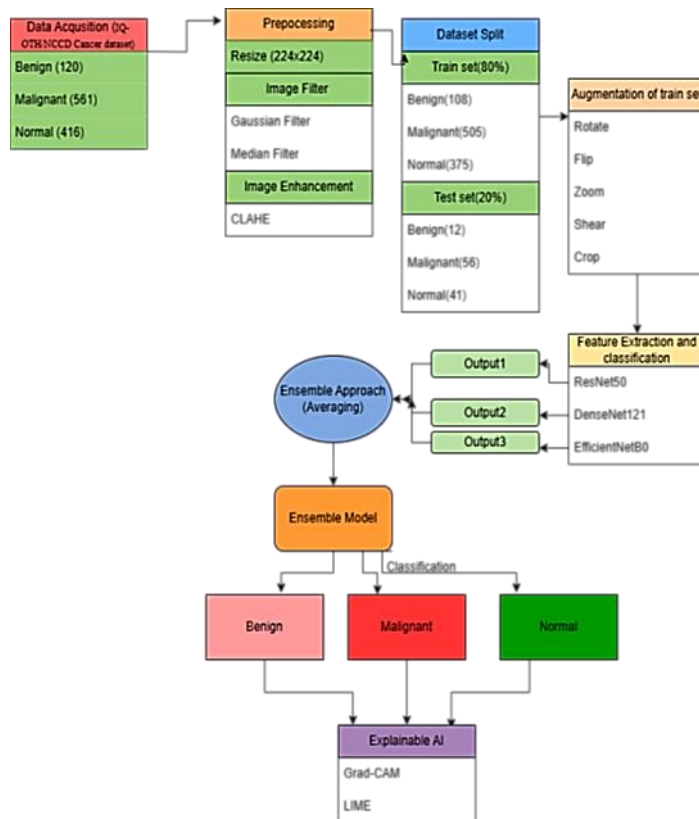


Fig.1. Proposed Framework

The Table.2 shows the algorithm for the proposed methodology.

Table.2. Algorithm for proposed methodology

Algorithm:
<pre> # Data preprocessing Load image_dataset() Resize all images to 224x224 Filter images [Gaussian filter, Median Filter] Image enhancement [CLAHE] Normalize image pixel values to range [0,1] # dataset split Split the dataset [80% train set, 20% test set] # Data augmentation Augment train set images [RandomFlip('horizontal', RandomRotation(0.2), RandomZoom(0.2))] # Load pre-trained models (DenseNet121, ResNet50, EfficientNetB0) Models = [DenseNet121, ResNet50, EfficientNetB0] For each model in Models: Freeze all convolutional layers Add GlobalAveragePooling2D () Add Dense (256, activation='relu') Add Dropout(0.3) Add Dense (num_classes, activation='softmax') # compile models For each model in Models: Use optimizer with learning rate le-3 Use Categorical cross entropy as loss parameter # train models For each model in Models: Train model with 50 epochs and 32 batch size # fine tune of the parameters For each model in Models: For layer in model.layers: Layer.trainable=True Optimizer=Adam(learning_rate=le-5) Train for 50 more epochs #save train weights Model1=model.save('densenet121.h5') Model2=model.save('resnet50.h5') Model3=model.save('efficientnetb0.h5') #generate prediction using each model For image in testset: Predict1=Model1.predict(image) Predict2=Model2.predict(image) Predict3=Model3.predict(image) # ensemble model creation Predict_ensemble= Average(Predict1, Predict2, Predict3) Ensemble_model designed using Predict_ensemble and implemented on same dataset # compile ensemble_model Use optimizer with learning rate le-3 Use Categorical cross entropy as loss parameter #train ensemble_model Train model with 50 epochs and 32 batch size # Save train weights Ensemble=ensemble_model.save('ensemble_model.h5') #generate prediction For image in testset: Predict_result=ensemble_model(image) # performance evaluation </pre>

```

For each model in [Model1, Model2, Model3,
ensemble_model]:
    Develop confusion matrix
    Accuracy=correct_prediction/total_samples
    Precision =TP/(TP+FP)
    Recall = TP/(FN+TP)
#Grad-CAM visualization
For each model in [Model1, Model2, Model3,
ensemble_model]:
    Extract conv layer
    Compute gradients
    Generate heatmap
#LIME visualization
For each model in [Model1, Model2, Model3,
ensemble_model]:
    Perturb images
    Fit the model
    Identify and visualize the important super pixels

```

3.1 DATA ACQUISITION

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) CT scan lung cancer image collection is accessible to the public for free through the Kaggle data repository [17]. Of the 1097 CT scan lung images in the image dataset, 120 are of the benign type, 561 are of the malignant type, and 416 are of the normal kind. Variability in imaging, including camera variations, noise like artifacts, coverage problems, etc., presents obstacles for photographs. Designing preprocessing, segmentation, and augmentation procedures specifically for IQ-OTH/NCCD can constitute a novel methodological contribution in and of itself because the dataset has distinct picture features (contrast, noise level, and resolution).

3.2 DATA PREPROCESSING

In order to well train the model, preprocessing of the images is highly necessary.

3.3 RESIZE

As the dataset contains images of various dimensions, resizing the images is highly necessary to improve computational efficiency as well as effective learning of the model without compromising key visual information. All the images are resized to 224x224.

3.4 IMAGE FILTER

To improve smoothness and reduce noise of the image, a hybrid filtering pipeline (Gaussian filter + median filter) is implemented. Images first passed through a Gaussian filter for removing Gaussian-type noise and then followed with a median filter for removing salt-and-pepper-type noise from them. In addition, the blurriness of the image increased by the Gaussian filter while preserving the borders using the median filter. This approach assist to have more accurate convergence, reduced false positives and more successful micro-nodules detection which were generally suppressed by noisy data.

3.5 IMAGE ENHANCEMENT

Unlike the conventional histogram equalization approach, the Contrast Limited Adaptive Histogram Equalization (CLAHE) procedure is implemented in addition to the hybrid filtering pipeline. This tri-level approach not only makes the noise-free smooth images but also enhances the contrast of each part of the images. As a result, high-quality images are generated for analysis and classification purposes. Fig.2 shows Gaussian filtered, median filtered and enhancement using CLAHE technique of benign, malignant and normal type CT scan lung cancer images respectively.

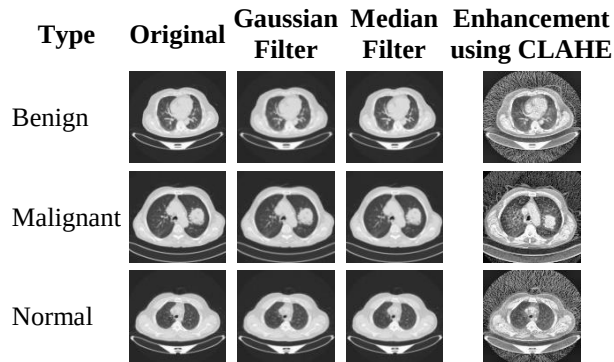


Fig.2. Gaussian filtered, median filtered and enhancement using CLAHE technique of benign, malignant and normal type CT scan lung cancer images

3.6 DATASET SPLIT

The image dataset has splitted in an 80:20 ratio such that 80% of the image data used as a training set along with the left over 20% used as a test set. Out of 1097 images, 988 images were used for training, and the remaining 109 imagery were used for testing purposes. The splitting of the dataset is shown in Table.3.

Table.3. Dataset split

Split set	Benign	Malignant	Normal
Train Set (=998)	108	505	375
Test Set (=109)	12	56	41
Total (=1097)	120	561	416

3.7 DATA AUGMENTATION OF TRAIN SET

As deep learning models need huge amounts of data for training purposes, data augmentation techniques such as rotate, flip, zoom, and shearing; along with crop have been applied to the training set. It helps to increase the size of the training set as well, and the models are able to learn variability in the data.

3.8 FEATURE EXTRACTION AND ENSEMBLE MODEL DESIGN

We employ three backbone CNNs: ResNet-50, DenseNet-121, and EfficientNet-B0 to extract deep features from segmented nodules. Each model is pretrained on ImageNet and fine-tuned on our CT dataset. Extracted feature vectors are integrated and proceed to a fully connected meta-learner, forming a stacking ensemble that outputs the final malignancy probability.

3.8.1 ResNet50:

In order to prevent vanishing of gradients during training, the deep CNN ResNet50 has been used. It uses residual connection, which helps the information bypass some layers. More reliable learning of image information is done. This model contains 50 layers comprising 49 convolution layers, one maxpooling layer, one average pooling layer, along with one fully connected layer. Convolutional layers are used to extract feature maps, and pooling layers are used to reduce the parameters and speed up the computation. The fully connected layer classifies the lung image as benign, malignant, or normal type based upon the extracted features. At first the model is trained up using an augmented training set of images for 50 epochs. In order to help the model to better adapt it to the dataset, fine-tuning of the parameters has been done and the model trained up for a further 50 epochs. The model is optimized in order to expedite the training process, and efficient updating of the weights is done using the Adam optimizer with a low learning rate.

3.8.2 DenseNet121:

Another pre-trained CNN model named DenseNet121 has been included to improve the reuse of features and efficiency of parameters. Each layer accepts inputs from all subsequent layers. This model contains a total of 120 convolution layers, one max pooling layer, four average pooling layers, along with one fully connected layer. The total architecture comprises 4 dense blocks with multiple convolution layers and 3 transition layers (1 convolution layer followed by 1 average pooling layer). Feature reuse facility and propagation are achieved using dense blocks. Transition layers are used to connect dense blocks and help to minimize the number of feature maps and down sample the spatial dimensions of the feature map. Flattening of all the features is done using a flatten layer by accepting all feature maps. A fully connected layer with a softmax activation function is used for classifying a lung cancer image as benign, malignant, or normal type. Here the model is trained up using the augmented training set for 50 epochs, and later, with fine parameter tuning, the model is trained for a further 50 epochs. It helped to reduce overfit and improve accuracy.

3.8.3 EfficientNetB0:

The foundational model of the EfficientNet family (EfficientNetB0) is used to improve efficiency and accuracy. In order to have systematic balances, compound scaling is applied, through which scaling of the system in deepness, breadth, and resolution is done. GlobalAveragePooling2D and Dropout layers are used to reduce overfitting and improve the healthiness of the model. The Adam optimizer along with the categorical cross entropy loss function has been used to speed up the training process and classify the image into a particular category. Initially, the model is trained up using an augmented training set for 50 epochs. With fine parameter tuning, the model is trained for a further 50 epochs. This approach achieved high accuracy for image classification.

3.8.4 Ensemble Approach:

The validation accuracy and training accuracy achieved by all the pretrained models are nearer to each other. In order to achieve the best accuracy, the ensemble learning approach has been applied. The deep learning neural networks commonly bring into being high variance with generalization error. The high variance

is due to different weights yielded by the model when training the model at different times. In order to avoid this high variance and to achieve better accuracy, the ensemble approach has been used. Training efficiency of the model has improved using the Adam optimizer with a 0.003 learning rate. In this research framework, all the pre-trained models, ResNet50, DenseNet121, and EfficientNetB0, have been trained individually using an augmented training set of the image dataset for 50 epochs and evaluated using the remaining 20% of the image dataset. Ensemble technique averaging is applied on the outcomes generated from all the pretrained models. This ensemble model performs better as compared to the performance of a single model.

3.9 EXPLAINABILITY MODULE

3.9.1 Grad-CAM:

The explainable AI technique Grad-CAM mainly gives importance to the feature maps generated by the final convolution layer, as the preceding layer captures high-level semantic information. Gradient value is calculated for each filter part of the feature map. The particular regions having higher gradient values are considered as highly influenced regions for the model’s prediction. In order to get a single scalar value for each feature map, global average pooling of gradients is performed. The scalar value represents neuron importance weight, which indeed tells how important the corresponding feature map is for predicting the target class. The weighted sums of all such relevant feature maps show the most important regions having a significant role in the target class prediction. The brighter region in the heat map represents the significant regions used for forecast.

3.9.2 LIME:

Another explainable AI technique LIME is used to make transparent the black box nature of the model for prediction. A set of contiguous semantically meaningful regions called superpixels have been generated by applying a segmentation technique on the image. This approach creates perturbed versions of the original image by randomly turning on or off a subset of superpixels. Proximity weight assigned to each performed sample. Super pixels having higher positive weight influence more the model’s prediction; that is, the super pixels strongly contribute to the prediction of the target class. In reverse, superpixels having higher negative weights strongly contribute against the target class forecast.

4. EXPERIMENTAL RESULT

The dataset has splitted into 80 % train, and 20 % test, ensuring patient-level separation. Data augmentation includes random rotations, flips, and intensity shifts. Models are trained with Adam optimizer (learning rate = 1e-3). The training accuracy, validation accuracy, training loss and validation loss achieved by ResNet50, DenseNet121, EfficientNetB0 and ensemble model are represented using graphs. The line plots developed between training accuracy and validation accuracy as well as line plots developed between training loss and validation loss of ResNet50, DenseNet121, EfficientNetB0 and Ensemble model is shown Fig.3(a), Fig.3(b), Fig.3(c), Fig.3(d), Fig.3(e), Fig.3(f), Fig.3(g) and Fig.3(h).

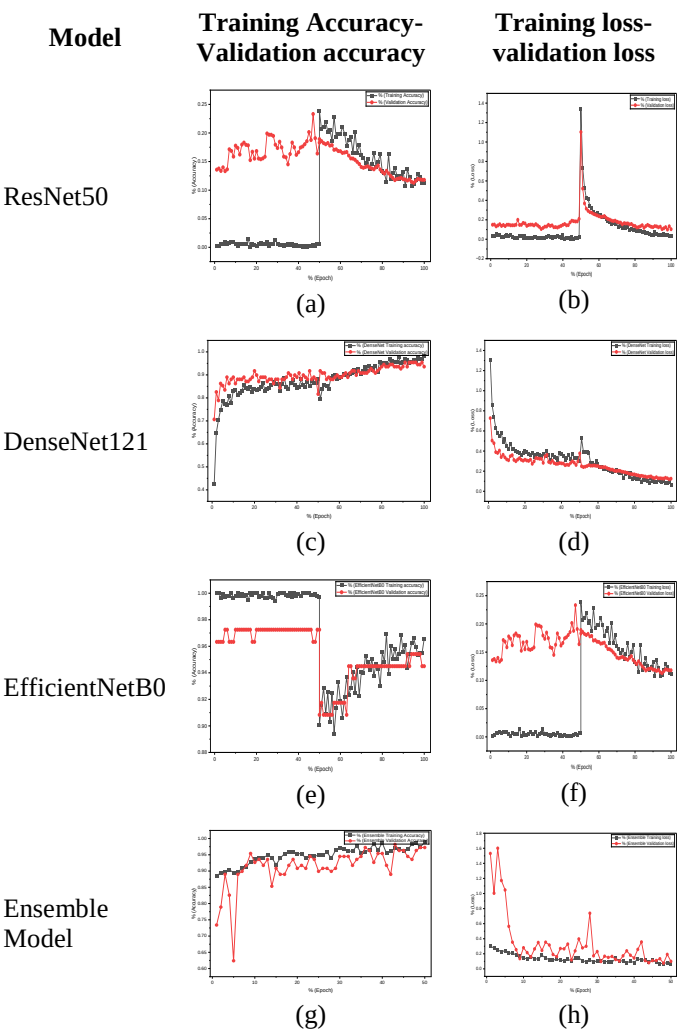


Fig.3. (a) Line plot between Training accuracy and validation accuracy of ResNet50. (b) Line plot between Training loss and validation loss of ResNet50. (c) Line plot between Training accuracy and validation accuracy of DenseNet121. (d) Line plot between Training loss and validation loss of DenseNet121. (e) Line plot between Training accuracy and validation accuracy of EfficientNetB0. (f) Line plot between Training loss and validation loss of EfficientNetB0. (g) Line plot between Training accuracy and validation accuracy of Ensemble model. (h) Line plot between Training loss and validation loss of Ensemble Model

The Table.4 shows the training accuracy, validation accuracy, training loss and validation loss obtained during the final epoch during training of each model.

Table.4. Training accuracy, validation accuracy, training loss and validation loss obtained during the final epoch during training of each model

Model	Epoch	Accuracy		Loss	
ResNet50	100	0.9875	0.945	0.0385	0.1031
DenseNet121	100	0.9831	0.9358	0.0601	0.125
EfficientNetB0	100	0.9655	0.945	0.1119	0.1179
Ensemble Model	50	0.9893	0.9725	0.0612	0.0994

4.1 EVALUATION METRICS

Each model is evaluated using the remaining 20% of the image dataset. On performance evaluation, a confusion matrix is developed for each model. The number of images correctly classified by the proposed ensemble model is more as compared to other single pre-trained models. The confusion matrix generated using ResNet50, DenseNet121, EfficientNetB0, and the proposed ensemble model are shown in Fig.4(a), Fig.4(b), Fig.4(c) and Fig.4(d) respectively.

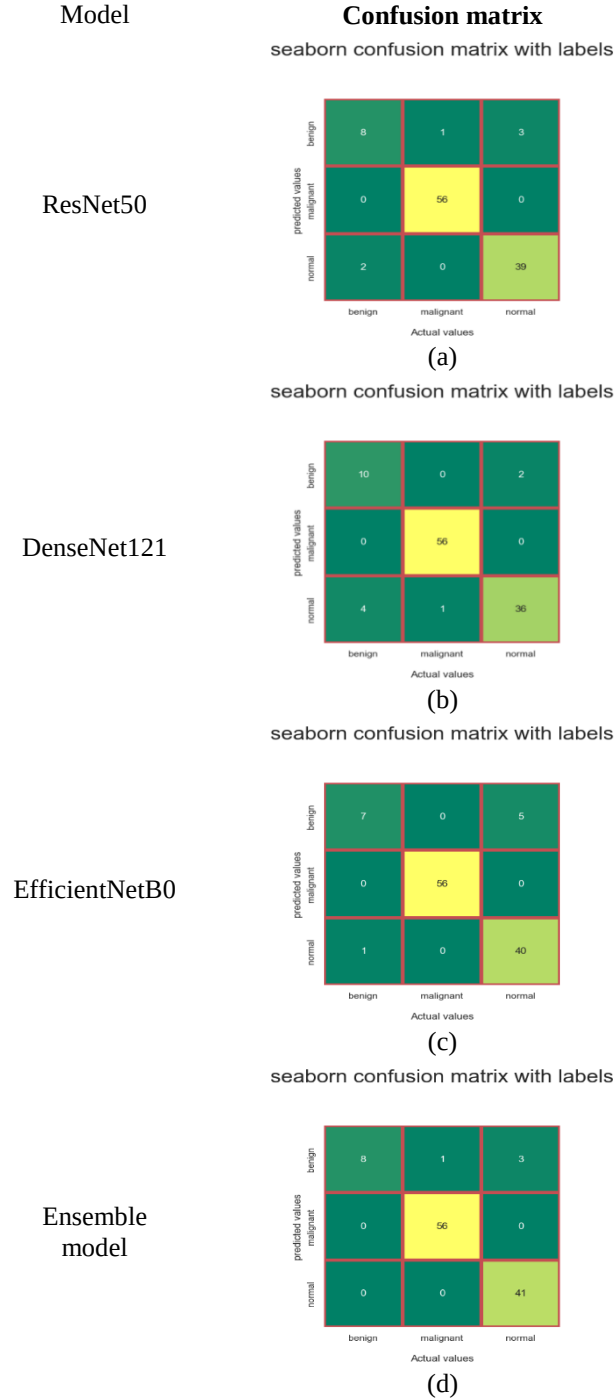


Fig.4. (a) Confusion matrix generated using ResNet50, (b) Confusion matrix generated using DenseNet121, (c) Confusion matrix generated using EfficientNetB0, and (d) Confusion matrix generated using the proposed ensemble model.

Using the confusion matrix, accuracy, precision, recall and F1-score have been calculated for ResNet50, DenseNet121, EfficientNetB0 and the ensemble model respectively. Table.5 represents the accuracy, precision, recall and F1-score obtained by the pre-trained models along with the ensemble model. The precision, recall and F1-score calculated class-wise for all models.

Table.5. The accuracy, precision, recall and F1-score obtained by ResNet50, DenseNet121, EfficientNetB0 and the ensemble model

Model		Precision	Recall	F1-score	Accuracy
ResNet50	Benign	0.8	0.67	0.73	0.94
	Malignant	0.98	1	0.99	
	Normal	0.93	0.95	0.94	
DenseNet121	Benign	0.71	0.83	0.77	0.9358
	Malignant	0.98	1	0.99	
	Normal	0.95	0.88	0.91	
EfficientNetB0	Benign	0.88	0.58	0.7	0.94
	Malignant	1	1	1	
	Normal	0.89	0.98	0.93	
Ensemble Model	Benign	1	0.75	0.86	0.972
	Malignant	1	1	1	
	Normal	0.93	1	0.96	

The Fig.5 shows the line graph developed using accuracy, precision, recall and F1-score obtained by ResNet50, DenseNet121, EfficientNetB0 and the proposed ensemble model.

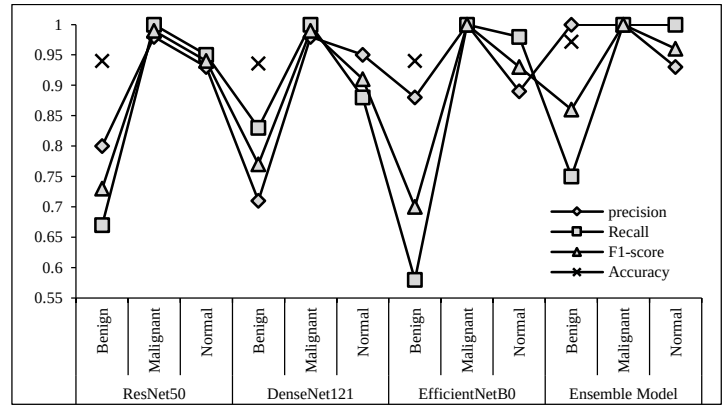


Fig.5. The line graph developed using accuracy, precision, recall and F1-score obtained by ResNet50, DenseNet121, EfficientNetB0 and the proposed ensemble model

4.2 INTERPRETABILITY ANALYSIS

4.2.1 Grad-CAM:

High-level semantic information present in the feature maps generated from the final convolution layer is specified significant importance by the Grad-CAM explainable technique. For each filtered region of the feature map, a gradient value is calculated. By applying Global Average Pooling on gradient values, a single scalar value is determined. The explainable AI technique Grad-CAM mainly gives importance to the feature maps generated by the last convolution layer, as the last layer captures high-level semantic information. Gradient value is calculated for each filter

part of the feature map. The particular regions having higher gradient values are considered as highly influenced regions for the model’s prediction. In order to get a single scalar value for each feature map, global average pooling of gradients is performed. The scalar value represents neuron importance weight, which indeed tells how important the corresponding feature map is for predicting the target class. The weighted sums of all such relevant feature maps show the most important regions having a significant role in the target class prediction. The brighter region in the heat map represents the significant regions used for prediction. A color-coded heatmap is overlaid on the original image. It shows which regions were most important for the model’s prediction. The colors follow a gradient scale, where warmer colors (red and yellow) signify high importance and cooler colors (blue and green) signify low importance. In this proposed experiment, the ensemble model shows most of the significant regions that have major contributions towards the prediction of the lung cancer image as the malignant type better as compared to ResNet50, DenseNet121, and EfficientNetB0, respectively. Original image along with Grad-CAM visual explanations for a lung cancer image classified as malignant type using ResNet50, DenseNet121, EfficientNetB0 and ensemble model are shown in Fig.6(a), 6(b), 6(c), 6(d) and 6(e) respectively.

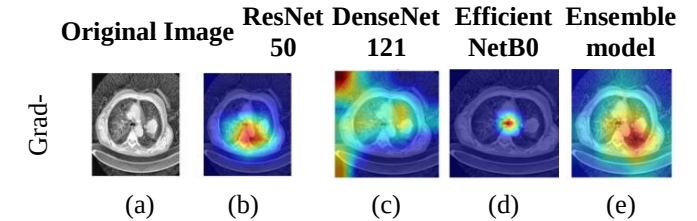


Fig.6. (a) Original image, (b) Grad-CAM visual explanation for a lung cancer image classified as malignant type using ResNet50, (c) Grad-CAM visual explanation for a lung cancer image classified as malignant type using DenseNet121, (d) Grad-CAM visual explanation for a lung cancer image classified as malignant type using EfficientNetB0, (e) Grad-CAM visual explanation for a lung cancer image classified as malignant type using ensemble model

4.2.2 LIME

Another explainable AI technique LIME is also used to make transparent the black box nature of the model for prediction. A set of contiguous semantically meaningful regions called super pixels have been generated by applying a segmentation technique on the image. This approach creates perturbed versions of the original image by randomly turning on or off a subset of super pixels. Proximity weight assigned to each perturbed sample. Super pixels having higher positive weight influence more the model’s prediction, i.e., the super pixels strongly contribute to the prediction of the target class. In reverse, super pixels having higher negative weights strongly contribute against the target class prediction. LIME visual explanation for a lung cancer image classified as malignant type using ResNet50, DenseNet121, EfficientNetB0 and ensemble model are shown in Fig.7(a), Fig.7(b), Fig.7(c) and Fig.7(d) respectively.

The super pixels having maximum contribution towards the prediction of cancer image type are represented with green colour, while the super pixels having negative contribution are

represented using yellow color. The weights obtained by the super pixels having a significant role during prediction using ResNet50, DenseNet121, EfficientNetB0, and the ensemble model are shown in Table.6.

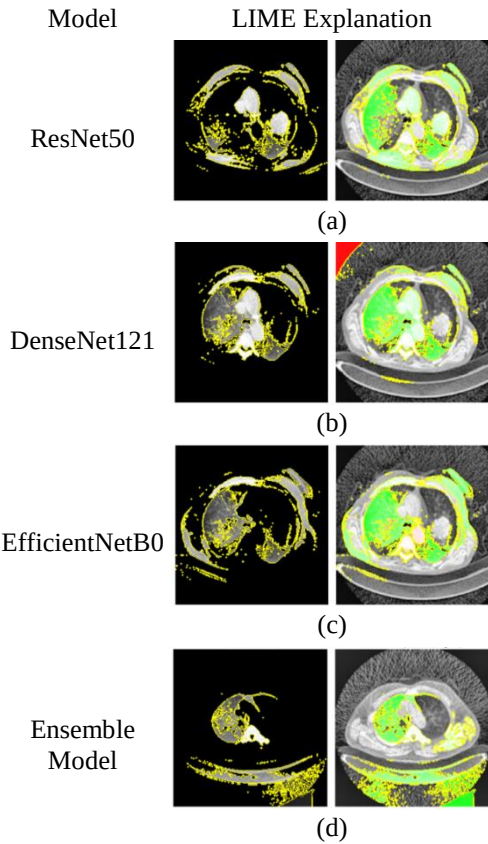


Fig.7. (a) LIME visual explanation for a lung cancer image classified as malignant type using ResNet50, (b) LIME visual explanation for a lung cancer image classified as malignant type using DenseNet121, (c) LIME visual explanation for a lung cancer image classified as malignant type using EfficientNetB0, (d) LIME visual explanation for a lung cancer image classified as malignant type using ensemble model

Table.6. Weights obtained by the super pixels having a significant role during prediction using ResNet50, DenseNet121, EfficientNetB0 and Ensemble Model

ResNet50		DenseNet121		EfficientNetB0		Ensemble	
Super pixel	Weight	Super pixel	Weight	Super pixel	Weight	Super pixel	Weight
51	0.0338	70	0.0036	41	0.0057	67	0.0434
41	0.0330	88	0.0029	82	0.0047	54	0.0426
96	0.0275	77	0.0021	70	0.0047	60	0.0418
25	0.0275	23	0.0021	29	0.0039	102	0.0387
83	0.0217	51	0.0021	50	0.0038	101	0.0313
103	0.0213	41	0.0021	35	0.0036	39	0.0271
65	0.0201	71	-0.0018	88	0.0034	100	0.0270
88	0.0200	117	-0.0018	56	-0.0030	116	0.0262
50	0.0193	11	-0.0017	77	-0.0028	111	0.0249
93	0.0186	50	0.0017	52	0.0027	64	0.0242

5. COMPARISON WITH BENCHMARK DATASET

The proposed ensemble framework is used on a benchmark dataset, LIDC-IDRI to compare the model performance. The model was highly accurate on the experimental data set because of clear classes specification; reduced inter-patient variability and lesser size of the dataset enabled the proposed model to identify and analyze discriminative features with few false positives. Conversely, since the imaging conditions are heterogeneous, the nodule shapes vary, and the scanner is noisy, and the boundaries of the images are difficult to recognize in the benchmark data, the proposed model was slightly less accurate but more generalizable. As a result of more training diversity, the benchmark dataset achieves better sensitivity, whereas the low specificity is reached because of the ambiguity in differentiating between benign and malignant training. In general, the ensemble is performing better in numerical performance regarding IQ-OTH/NCCD than it performs in benchmark dataset. Table.7 shows an experimental comparison of the results of IQ-OTH/NCCD and the benchmark dataset.

Table.7. Experimental comparison of IQ-OTH/NCCD and the benchmark dataset

Performance Metric	IQ-OTH/NCCD dataset	Benchmark Dataset (LIDC-IDRI)
Accuracy	0.972	0.904
Sensitivity	0.75-1	0.7-0.8
Precision	0.93-1	0.8-0.87
F1-score	0.86-1	0.8-0.85
Ambiguity in class	Lower	Higher

6. DISCUSSION

It has been evaluated that proposed model is advanced with 2-4% higher accuracy as compared to conventional ones of single CNN backbone models[7-10] as it shows the advantage of combined diverse feature representations. The ensemble with many diverse fusion features overcomes the struggle of traditional CNN models to detect the whole variety of nodules variations [11-16]. This method is used to create inequity among normal, malignant and benign nodules through their increased accuracy, sensitivity and F1-score. Dual explainability (Grad-CAM +LIME) can be included and results in increased robustness and clinical trust of the model. Single explainable approach is used, in which we have partial understanding of the internal reasoning of the model. Such a dual explainability methodology enhanced transparency as well as diminished diagnostic ambiguity. Using Grad-CAM, Robust spatial features are obtained by highlighting clinically significant lesion areas and fine-grained justification of the model's result are provided using LIME.

7. POTENTIAL CHALLENGES AND FUTURE WORK

The primary pitfalls are the computational cost of calculating Grad-CAM values of large CT volumes and the possible

instability of LIME when there are strongly correlated super pixels. Future research will consider better GRAD-CAM approximations and hierarchical LIME segmentations to be deployed in real-time .The ensemble model also produces much less disturbed samples when compared to samples that were generated by individual pre-trained models. The ensemble model predictions were less noisy and consistent. Consequently, LIME has less jagged local decision boundaries compared to DenseNet121, ResNet50, and EfficientNetB0. Since the procedure of grad and feature map is done simultaneously, Grad-CAM elucidation of the ensemble model is quicker compared to other analysis. Since ensemble prediction produces more smooth and steady gradients, the gradient with Grad-CAM is more distinct and does not have a lot of noise. The inconsistency in CT acquisition parameters, model and slice thickness can limit the generalizability of the model in different clinical settings. Grad-Cam and LIME can be used to improve trust, but they have serious computational overheads in generating an explanation that can only be used with real-time processing. The small dataset can result in the bias of the model. Diagnosis procedure can be strengthen and uncertainty in the borderline can be minimized using a combined data of medical imaging as well as clinical information (age, smoking strategy, radiologist score) To offer high-quality explanations at a lower computational cost, future studies can use more informative and faster XAI techniques, including SHAP DeepExplainer or real-time saliency mapping.

8. CONCLUSION

In case of classification and transparent explanations on lung cancer prediction, our proposed model reaches the state of art. also It is a powerful interpretable framework that can be used to assist radiologists in their clinical practices and improve patient outcomes through the combination of accuracy and trustworthiness. The proposed ensemble model, which combines ResNet50, DenseNet121, and EfficientNetB0, turns out to be quite powerful and successful in lung cancer prediction by capturing a variety of imaging features and being better than a number of single-model-based and state-of-the-art models. By adding Grad-CAM and LIME, the interpretability factors become better, showing regions of clinical interest and justifying the classification choice to promote trust and assist radiologists. Despite the challenges associated with the approach in terms of computational cost and variability of the data, the findings clearly show that the approach is a promising and explainable early lung cancer detector, and the future research will involve optimization, domain adaptation, and the expansion of the clinical areas of application.

REFERENCES

- [1] M.S. Ahmed, "Interpretable Lung Cancer Detection using Explainable AI Methods", *Proceedings of International Conference for Advancement in Technology*, Vol. 21, pp. 1-11, 2023.
- [2] "World Lung Cancer Day 2020 Fact", Available at: <https://www.chestnet.org/News/CHEST-News/2020/07/WorldLung-Cancer-Day-2020-Fact-Sheet>, Accessed in 2020.

- [3] D.D. Martinelli, "Machine Learning for Metabolomics Research in Drug Discovery", *Intelligence Based Medicine*, Vol. 8, pp. 1-24, 2023.
- [4] I. Shafi, "Ensemble CNN with SHAP and LIME for Histopathological Cancer Classification", *BMC Medical Informatics and Decision Making*, Vol. 24, pp. 1-11, 2024.
- [5] S.M. Lundberg and S.I. Lee, "A Unified Approach to Interpreting Model Predictions", *Proceedings of International Conference on Neural Information Processing Systems*, Vol. 30, pp. 4768-4777, 2017.
- [6] M.A. Thanoon, "Deep Learning Techniques for Lung Cancer Screening and Diagnosis based on CT Images: A Systematic Review", *Diagnostics*, Vol. 13, No. 16, pp. 1-17, 2023.
- [7] J. Ker, L. Wang, J. Rao and T. Lim, "Deep Learning Applications in Medical Image Analysis", *IEEE Access*, Vol. 6, pp. 9375-9389, 2017.
- [8] M.G. Lanjewar, K.G. Panchbhair and P. Charanarur, "Lung Cancer Detection from CT Scans using Modified DenseNet with Feature Selection Methods and ML Classifiers", *Expert Systems with Applications*, Vol. 224, pp. 1-13, 2023.
- [9] M. Sohaib and M. Adewunmi, "Artificial Intelligence based Prediction on Lung Cancer Risk Factors using Deep Learning", *International Journal of Informatics and Communication Technology*, Vol. 12, No. 2, pp. 188-194, 2023.
- [10] K. Alshamrani and H.A. Alshamrani, "Classification of Chest CT Lung Nodules using Collaborative Deep Learning Model", *Journal of Multidisciplinary Healthcare*, Vol. 17, pp. 1-15, 2024.
- [11] V. Shah, "Ensemble Deep Convolutional Neural Networks for Lung Nodule Detection", *Journal of Medical Imaging*, Vol. 23, pp. 1-15, 2023.
- [12] H. Li, "Hybrid Ensemble of ResNet50 and ML Classifiers for Lung Cancer Prediction", *IEEE Access*, Vol. 12, pp. 1-17, 2022.
- [13] M. Ahmed, "Deep CNN-Boosting Framework for Early Lung Cancer Diagnosis", *Computers in Biology and Medicine*, Vol. 24, pp. 1-17, 2021.
- [14] A. Banerjee, "Multi-Scale CNN and Random Forest Ensemble for Pulmonary Nodule Classification", *Pattern Recognition Letters*, Vol. 5, pp. 1-17, 2022.
- [15] M. Shahin, "CNN-SVM Ensemble Approach for Lung Cancer CT Image Classification", *Multimedia Tools and Applications*, Vol. 1, pp. 1-7, 2021.
- [16] S. Rahman, "Hybrid VGG16-Inception Ensemble for Lung Lesion Detection", *Biomedical Signal Processing and Control*, Vol. 2, pp. 1-8, 2023.