|| Volume 9 || Issue 12 || December 2025 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Attention-Guided Deep Learning Framework for Automated Lung Disease Classification from CT scans

[1]Mr. Bharat Bhushan Prajapati, [2]Dr. Suman Kumar Swarnakar
[1]*M. Tech. Scholar, Dept. of CSE, SSIPMT, Raipur*
[2]*Assistant Professor,Dept. of CSE, SSIPMT, Raipur*

Abstract: This work introduces a deep learning novel framework for identifying and classifying automated lung disease from CT scan images. The proposed technique employs a ResNet-50 backbone with dual attention to improve the normal, benign, and malignant lung condition feature discrimination. The methods attention-guided feature extraction, thorough preprocessing, and interpretable visualization via Grad-CAM contribute to the proposed work's superior performance. Experimental results demonstrate the value of the proposed approach for supporting clinical diagnosis.

Keywords: *Deep Learning, Lung Disease Detection, Attention Mechanism, Medical Image Analysis, CT Scan Classification*

I.INTRODUCTION:

Lung cancer remains a leading cause of cancer-related mortality, indicating the need for effective and timely detection to improve outcomes. The interpretation of CT scans is still primarily manual, and though widespread, this method is still very time-consuming, and relies heavily on the subjective evaluations of each radiologist. To provide a solution to these problems, this study proposes an deep learning automated approach based on attention mechanism, Advanced mathematical modeling and automated systems is the answer to the automation of CT scan analysis and improved CT scan interpretation aimed at detecting lung cancer and other lung diseases. The world is burdened with diseases such as pneumonia, COPD, and other interstitial lung diseases. WHO reports indicate lung cancer alone accounts for over two million new cases a year and conditions such as pneumonia and COPD considerably increase hospital admissions. Therefore, the diagnosis and treatment of lung diseases must be timely and precise to provide improved outcomes for patients.

The superior resolution and ability to capture fine details Anatomical details gives CT imaging a high regard. CT scans are thoracic is considered the standard diagnostic modality for imaging. Yet the manual evaluation of CT scans and the subjectivity and observer variability which are particularly evident in the subtle pathological features present. The Apollo of deep learning approaches Discriminative Region computational approaches neural networks have advanced in the field of medical image analysis. The Apollo of deep learning approaches neural networks have advanced in the field of medical image analysis. Convolutional neural networks in the Apollo of deep learning approaches Discriminative Region computational approaches neural networks have advanced in the field of medical image analysis. Convolutional neural networks in the Apollo of deep learning approaches Discriminative Region computational approaches neural networks have advanced in the field of medical image analysis. Convolutional neural networks in the Apollo of deep learning approaches Discriminative Region computational approaches neural networks have advanced in the field of medical image analysis. Convolutional neural networks in the Apollo of deep learning approaches Convolutional neural networks have been used in lung disease classification achieving diagnostic performance that compares to expert classifiers. Inversely to explainable C E approaches neural networks have been used in lung disease classification achieving diagnostic performance that compares to expert classifiers. Inversely to explainable C E approaches neural networks have been used in lung disease classification achieving diagnostic performance that compares to expert classifiers. Inversely explainable approaches have C E Discriminative Region computational approaches Convolutional neural networks have been used in lung disease classifiers achieving diagnostic performance that compares to expert classifiers. Inversely explainable approaches have C E.

Attention mechanisms in algorithms foster the selective focusing of a computer on a specific area while excluding the predominately irrelevant data noise neural networks achieve this through Discriminative Region computational approaches neural networks have been used in lung disease classifiers achieving diagnostic performance that compares to expert classifiers. Inversely explainable approaches C E Discriminative Region computational approaches Convolutional neural networks have been used in lung disease classifiers achieving diagnostic performance that compares to expert classifiers. Inversely explainable approaches.

We describe the Attention-Guided Deep Learning Framework focused on lung disease classification on CT scans. We combine convolutional neural architectures with attention integration to accomplish interpretability alongside diagnostic accuracy. Attention feature refinement integrated with explainability tools aims to provide diagnostic assistance rapidly, reliably, and transparently. The proposed solution targets the interpretability-performance tradeoff, paving the way for trustworthy AI-based pulmonary care decision support systems.

II. LITERATURE REVIEW

In the recent few years, there has been advancements in the field of lung disease detection and classification due to the implementation of deep learning, transfer learning, and attention-

based techniques. A number of research works have explored different imaging techniques, neural architectures, and interpretability to improve accuracy in diagnosis. I focused on and will discuss the main 10 contributions in recent years.

Zheng et al. (2024) designed a deep neural network for classifying lung nodules in CT scans (DOI: 10.1016/j.heliyon. 2023. e23508). Their method achieved 90.11% classification accuracy and 95.66% AUC, and also reduced computational time by 29.09% when compared to traditional models. Their implementation of multichannel attention CT image analysis helped to focus on areas of greater, more accurate, and relevant diagnoses. That said, the study could have expanded its endpoints to other pulmonary disorders and thus determined more fully the clinical value of this technology than just for lung nodule detection.

Ullah et al. (2023) also built and designed DeepLungNet, a deep learning framework for lung disease detection utilizing chest radiograph images (DOI: 10.3390/electronics12081860). The system found and achieved 97.47% accuracy - further proof of deep learning's capacity and value and how far it out reaches traditional diagnostic methods. The framework, however, is restricted to chest radiograph images and does not incorporate attention mechanisms which are fundamental in deep learning CT scan models. These features are vital for the system to identify fine anatomical details, and critical for tasks that require high accuracy, such as nodule detection, deep learning radiography.

Table 1. Literature Review

S. No.	Authors & Year	Method / Framework	Dataset / Modality	Key Results	Relevance / Limitation
1	Zheng et al., 2024	Multi-channel attention-guided DNN	CT Scans	Accuracy: 90.11%, AUC: 95.66%, reduced processing time to 29.09% of mainstream models	Strong CT-based attention framework; direct relevance for lung nodule classification
2	Ullah et al., 2023	DeepLungNet (DL-based model)	Chest Radiographs (CRIs)	Accuracy: 97.47%	Focused on CRIs, not CT; high performance but lacks attention-guided design
3	Rathod & Ragha, 2023	DLCTLungDetectNet (FusionNet = ResNet50 + InceptionV3)	CT Scans	Effective tumor detection	No attention mechanism; hybrid fusion improves robustness
4	Ali et al., 2024	CX-RaysNet (custom CNN)	Chest X-rays	Accuracy: 97.25%	Good for CXR multi-class classification; not CT-based
5	Gao et al., 2025	MobileNetV2 + Transfer Learning	CT Scans	Tumor, pneumonia, and normal classification	Market-driven CT study; lacks explicit attention integration
6	Saha & Nandi, 2024	DNN-ILD (Transfer Learning + DenseNet121)	CT Scans	ILD classification	Robust ILD model; not attention-guided
7	Parihar et al., 2024	Comparative study of DL models (DenseNet, MobileNet, ResNet, ChexNet)	Chest X-rays	Emphasizes adaptability of models	Broad evaluation; limited CT focus
8	Agarwal & Arya, 2023	CXRNet (CNN + Attention)	Chest X-rays	High performance on multiple datasets	Attention-based but CXR only; not CT
9	Liu et al., 2024	3D Attention-Gated CNN	CT Scans (lung nodules with fibrosis)	Malignancy classification with fibrotic microenvironment	Strong CT-based attention framework; focused on nodules, not full disease classification
10	Hamdi et al., 2021	MARL (Multimodal Attentional Representation Learning)	CT + patient history	Superior accuracy vs CNN baselines	Multi-modal; pioneering use of attention in multimodal fusion

Rathod and Ragha (2023) introduced DLCTLungDetectNet, a deep learning hybrid framework that integrates ResNet50 and InceptionV3 (DOI: 10.2139/ssrn.4651122). This model concentrated on detecting lung tumors from CT scans, and for its class, it performed quite well as it was able to utilize a FusionNet design for a combination of two architectures. Although a hybrid approach to design continued to improve level of accuracy for detection of lung tumors, the work still lacks in addressing more general and more complicated problems of multi-class classification, especially class differentiation between a benign and malignant and other associated clasifications. This also misses other explainability frameworks, for example, Grad-CAM.

Ali et al. (2024) built the CX-RaysNet, a custom-designed neural network for the multi-class lung disease classification problem using chest X-ray images (DOI: 10.1111/coin.12649). Data augmentation and the convolutional layers helped the model gain a 97.25% accuracy score in classifying and differentiation a number of disease categories. This work, however, continues to face the limitation of X-ray imaging, as CT scans are still much preferred in clinical settings for lung disease eros, because CT scans are higher resolution images and provide more detailed structural information. This is especially important for the early detection of lung disease which has more subtle, early stage changes.

Gao et al. (2025) proposed a CT image classifier using transfer learning
based on MobileNetV2 (DOI: 10.48550/arxiv.2501.04996). The framework showed the ability to classify CT images as lung tumors, pneumonia, or normal as well as its accuracy and commercial potential for integration into clinical workflows. Even so, its application on MobileNetV2 was a positive as it tailored for limited computational requirements, so it could be deployed in environments with fewer resources. However, the framework without any attention integration was unable to provide the explainability needed and to focus on the clinically relevant parts of the CT images.

Saha and Nandi (2024) built a transfer learning based on a revised DenseNet121 architecture to create the DNN-ILD for the

automated classification of interstitial lung diseasesfrom CT images (DOI: 10.1007/978-3-031-47942-7_26). The model was a positive step in the automated classification of ILDs as it successfully classified the interstitial lung patterns. However, the framework could have benefited from the inclusion of attention based refinements, particularly in the automated classification of overlapping fibrotic patterns and other lung diseases, which remain a challenge.

Parihar et al. (2024) researched how deep learning methods can be used to classify lung diseases from chest X-ray images. They analyzed the various implementations of DenseNet, MobileNet, ResNet, and ChexNet and how they can be used with different datasets. Although the insights from the comparison are helpful, it should be noted that the study focused on chest X-ray datasets which do not have the detailed structural information that CT scans provide. Finally, the study did not perform attention-based interpretability, which maybe limits the clinical trust to the results.

Agarwal and Arya (2023) developed CXRNet, a CNN-attention based classifier for chest X-ray images (DOI: 10.1111/exsy.13423). The framework used attention-based enhancement of feature extraction which was based on pre-trained CNNs which then improved the interpretability and accuracy of lung disease detection. However, the model did not extend to CT scan data which are more complex and clinically more informative for lung cancer detection and classification.

Liu et al. (2024) used a 3D attention-gated convolutional network to classify lung nodule malignancy using CT scans (DOI: 10.1186/s12967-023-04798-w). This study incorporated the fibrotic microenvironment in the classification process, which helped accurately classify malignant versus benign nodules. Nevertheless, it was still limited to nodule analysis rather than applying it to the whole lung disease. Moreover, while attention gating has been used, the explainability via visualization, for instance, tools like Grad-CAM, was not explicitly used.

To conclude, Hamdi et al. (2021) proposed MARL (Multimodal Attentional Representation Learning), which integrates CT scan images and historical patient data (DOI: 10.1007/978-3-030-87156-7_2). With multimodal data and attention, the model aided feature learning and predictive power outperformed standard CNNs. Even though predictive power was enhanced due to multimodal data, the complex framework and reliance on disparate data streams reduce its practicality for everyday clinical use.

Summary

From this review, a few salient themes can be identified. First, attention-based models (Zheng et al., Liu et al., Agarwal et al., Hamdi et al.) are more interpretable and can attend more to clinically important areas, which is important for real-world use. Unfortunately, most such works are limited to either lung nodules or X-ray datasets rather than CT scans, and so, they are unable to provide broadly applicable solutions for the much more complex problem of multi-class lung disease classification. Second, transfer learning (Gao et al. (2025) and Saha & Nandi (2024)) enjoys the use of pre-trained models for great efficiency and performance, but do not incorporate explainability. Third, more predictive multimodal systems, like MARL, pose integration issues in real life because they depend on extra patient data.

III. RESEARCH GAP AND MOTIVATION

A. Research Gap

Lung disease detection systems are incomplete in several critical areas.

Limited Feature Discrimination: Many of the existing CNN designs are unable to distinguish the subtle textural differences between malignant and benign lung lesions, and as a result, the classification performance is poor.

Loss of Spatial Information: Early and aggressive pooling will be the leading cause of loss of spatial information. This loss of information will lead to poor localization of lung lesions.

Absence of Attention Mechanism Focus: The approaches focus on not having attention mechanisms that zoom in on clinically important areas, which diminishes their diagnostic precision.

Interpretability Issues: The black-box frameworks that currently exist lead to a void in explainability on the decisions made for the diagnosis.

Multi-class Classification Issues: There has been a void in literature on detecting and classifying several lung pathologies simultaneously and with precision.

Motivation

For this work, one of the main drivers has been the need to improve the most fundamental aspects of early diagnosis and lung disease classification. In this case, we aim to improve: First, the automated objective CT scan image analysis of the automated objective CT-scan image analysis which high-resolution imaging and consistency will provide, focusing on the highly variable subjective interpretation of radiology that most deeply relies on the examiner's expertise. Second, the use of dual attention mechanisms, spatial and channel, to improve the representation of features and concentration on diagnostically decisive areas. Third, the real-time processing that will permit the diagnostic to be integrated in the clinical workflow without any loss of diagnostic accuracy. Fourth, the diagnostic results that will be explainable and interpretable using Grad-CAM, augmented with clinical reasoning for the possible diagnosis improving trust from the clinician's side towards the system. Finally and most importantly, the system will allow clinicians to detect lung diseases earlier and improve the health of their patients.

IV. PROPOSED METHODOLOGY

Our proposed framework incorporates a ResNet-50 backbone with dual attention modules for more accurate lung disease classification. The system consists of a multi-stage pipeline that includes preprocessing, hierarchical feature extraction via bottleneck layers, attention-guided feature refinement, and classification, which employs global average pooling followed by fully connected layers.

The architecture uses CLAHE (Contrast Limited Adaptive Histogram Equalization) for image enhancement and integrates Grad-CAM for model interpretability. With this, clinicians can assess the explanations for a model's predictions while the model maintains classification accuracy across normal, benign, and malignant lung conditions. The preprocessing step uses statistical normalization and resizing to standardize representation of the data. Feature extraction employs ResNet-50's bottleneck architecture, which abstracts hierarchical features at multiple scales. The classification head is designed to output predicted probabilities and confidence scores, while important features describe the classification.

The ResNet50 architecture starts by taking as input an RGB image of size 224×224×3. It first takes this image through 7×7 convolution and then through batch normalization and activation functions (ReLU). At this point, the network saves as output 64 feature maps of size 112×112. After this initial convolution, the network performs max pooling, with a 3×3 kernel and a stride of 2, thus reducing the spatial dimensions obtaining 64 feature maps of size 56×56. The central feature extraction section of ResNet50 consists of four sequential bottleneck layers. In this first layer, there are three bottleneck blocks and the layer outputs a size of 256×56×56 and is responsible for extracting the simple patterns and corners of the image and layer 2 is more complex with 4 bottleneck blocks which outputs 512×28×28 focusing on textures and small objects. In layer 3, more feature abstraction is done with 6 bottleneck blocks producing a 1024×14×14 feature map that identifies parts of the objects and complex patterns. Layer 4 has the most abstraction with 3 final bottleneck blocks capturing the full objects, scenes and the highest level 2048×7×7 tensor.

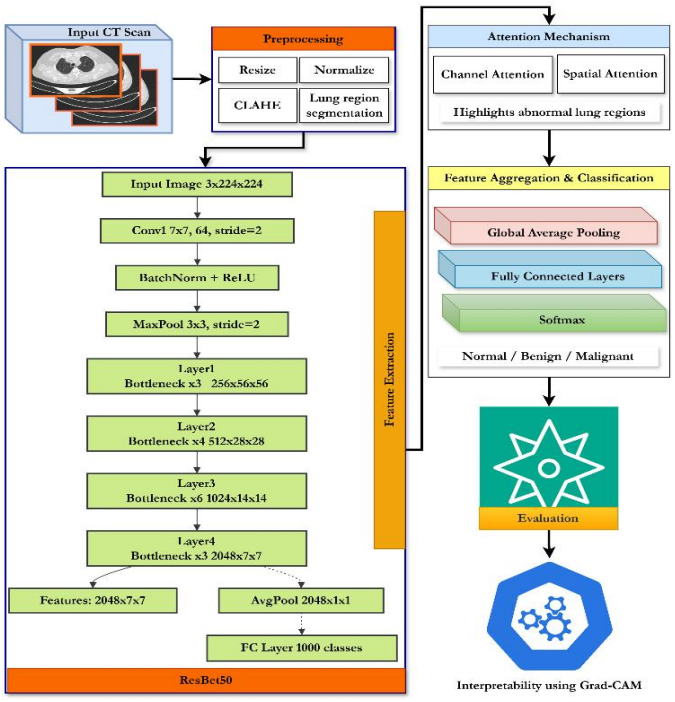

Fig. 1 ResNet50 Architecture for feature extraction

After the convolutional blocks, we apply a global average pooling layer, which creates a 2048-dimensional vector that serves as a global, holistic summary of the image after averaging the entire feature map. Subsequently, a fully connected layer takes this vector and outputs prediction scores that classify the image into one of 1000 categories. The entire architecture can be understood as a sequence of feature extraction stages: convolutional layers identify edges and simple shapes, layer 1 captures basic patterns and corners, layer 2 discerns textures and smaller objects, layer 3 detects parts of larger objects and intricate patterns, layer 4 completes the scene recognition, and average pooling summarizes the image. ResNet50 contains around 25.6 million parameters and occupies 98 megabytes of memory. The model and input size are standardized at 224×224×3.

V. MATHEMATICAL FORMULATION

A. Problem Definition and Notation

Consider a training dataset $D = \{(I_i, y_i)\}_{i=1}^{N}$, where $I_i \in \mathbb{R}^{H\times W\times C}$ is the ith CT scan image, and $y_i \in \{0, 1, 2\}$ is the label of the image (Normal, Benign, Malignant). The goal is to learn the mapping function $f_\theta: \mathbb{R}^{H\times W\times C} \to \mathbb{R}^3$ and parameterize it with θ.

B. Preprocessing Stage

The neural network requires CT scan images to be converted to appropriate normalized inputs in this step of the pipeline.

CLAHE Enhancement: For pixel intensity redistribution, CLAHE processes local histograms.

$$I_{clahe}(x, y) = T_{local}\big(I(x, y)\big)$$

where, T_{local} is the transformation function computed as,

$$T_{local}(v) = \frac{CDF_{clip}(v) - CDF_{min}}{M \times N - CDF_{min}} \times (L - 1)$$

with CDF_{clip} being the clipped cumulative distribution function, M × N the tile size, and L the number of gray levels.

Statistical Normalization: The normalization process standardizes pixel intensities:

$$I_{norm}(x, y) = \frac{I_{clahe}(x, y) - \mu_D}{\sigma_D + \varepsilon}$$

where μ_D and σ_D represent dataset-wide mean and standard deviation, respectively, and $\varepsilon = 10^{-8}$ ensures numerical stability.

Bilinear Interpolation for Resizing: The resizing operation employs bilinear interpolation:

$$I_{input}(x', y') = \sum_{x=\lfloor x' \rfloor}^{\lceil x' \rceil} \sum_{y=\lfloor y' \rfloor}^{\lceil y' \rceil} I_{norm}(x, y) \cdot w(x, y, x', y')$$

where, $w(x, y, x', y') = (1 - |x' - x|)(1 - |y' - y|)$ represents the bilinear weights.

C. Hierarchical Feature Extraction

The ResNet-50 backbone extracts multi-scale features through bottleneck blocks with residual connections.

Convolutional Operations: For the initial convolution layer:

$$F_0^{(k)}(x,y) = \sigma\left(\sum_{c=1}^{C}\sum_{i=-r}^{r}\sum_{j=-r}^{r} W_0^{(k)}(i,j,c)\cdot I_{input}(x+i,y+j,c) + b_0^{(k)}\right)$$

where, $W_0^{(k)} \in \mathbb{R}^{7\times7\times C}$ is the k-th filter, $b_0^{(k)}$ is the bias term, and σ represents the ReLU activation function.

Bottleneck Block Formulation: For the l-th bottleneck block with input $F_{l-1} \in \mathbb{R}^{H_l\times W_l\times D_{l-1}}$:

$$F_l^{(1)} = BN_l^{(1)}\left(Conv_{1\times1}^{(1)}(F_{l-1})\right)$$

$$F_l^{(2)} = BN_l^{(2)}\left(\sigma\left(Conv_{3\times3}^{(2)}\left(F_l^{(1)}\right)\right)\right)$$

$$F_l^{(3)} = BN_l^{(3)}\left(Conv_{1\times1}^{(3)}\left(F_l^{(2)}\right)\right)$$

Residual Connection: The residual connection enables gradient flow:

$$F_l = \sigma\left(F_l^{(3)} + P(F_{l-1})\right)$$

where P is the projection function for dimension matching.

D. Dual Attention Mechanism

Our dual attention mechanism enhances feature representation through both channel and spatial attention.

Channel Attention Module: Given feature map $F_4 \in \mathbb{R}^{H\times W\times D}$, the channel attention weights are computed as:

$$Z_{avg} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_4(i,j,:)$$

$$Z_{max} = \max_{i\,j} F_4(i,j,:)$$

$$M_c = \sigma\left(W_2\left(\delta\left(W_1(Z_{avg})\right)\right) + W_2\left(\delta(W_1(Z_{max}))\right)\right)$$

where, $W_1 \in \mathbb{R}^{\frac{D}{r}\times D}, W_2 \in \mathbb{R}^{D\times\frac{D}{r}}$ with reduction ratio r=16.

Spatial Attention Module: The spatial attention identifies important spatial locations:

$$S_{avg} = \frac{1}{D}\sum_{k=1}^{D} F_4(:,:,k)$$

$$S_{max} = \max_{k} F_4(:,:,k)$$

$$M_s = \sigma\left(Conv_{7\times7}(S_{avg};S_{max})\right)$$

Sequential Attention Application: The attention mechanisms are applied sequentially:

$$F_4^{'} = F_4 \otimes M_c$$

$$F_{att} = F_4^{'} \otimes M_s$$

where $\otimes$ denotes broadcasting element-wise multiplication.

E. Feature Aggregation and Classification

The final classification stage aggregates attention-enhanced features for decision making.

Global Average Pooling: Feature aggregation is performed using:

$$f = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{att}(i,j,:)$$

Multi-layer Perceptron Classifier: The classification head processes aggregated features:

$$h = \delta\left(W_{fc1}f + b_{fc1}\right)$$

$$z = W_{fc2}h + b_{fc2}$$

$$p = softmax(z) = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}}$$

Objective Function: The optimization objective combines cross-entropy loss with L2 regularization:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{3} y_{i,c}\log\left(p_{i,c}\right) + \lambda\sum_{\theta\epsilon\Theta}\|\theta\|_2^2$$

VI. INTERPRETABILITY ANALYSIS

For broader clinical adoption, the proposed model integrates Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance interpretability and provide visual explanations of its predictions.

Gradient Computation: To generate class activation maps, the gradients of the target class score with respect to the feature maps from a convolutional layer are first calculated as:

$$\frac{\partial y^c}{\partial A_{i,j}^k} = \frac{\partial}{\partial A_{i,j}^k}\left(\sum_{l=1}^{H}\sum_{m=1}^{W}\sum_{n=1}^{D} w_{l,m,n}^c, A_{l,m}^n\right)$$

Importance Weights: Each feature map's relevance to the predicted class is determined by computing the global average of the corresponding gradients:

$$\alpha_k^c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{\partial y^c}{\partial A_{i,j}^k}$$

Class Activation Map: The final attention map is generated as:

$$L_{Grad-CAM}^c = ReLU\left(\sum_{k=1}^{D}\alpha_k^c A^k\right)$$

This interpretability mechanism enables clinicians to highlight the regions of interest of the CT scan had the greatest influence on the model's prediction, enhancing trust and facilitating clinical integration.

Data: CT Scan Dataset $D = \{(I_i, y_i)\}_{i=1}^N$

Result: Trained model M for lung disease classification

Initialize ResNet-50 backbone, attention modules, and classifier;

for epoch = 1 to max epochs do

for batch B ⊂ D do

Preprocessing:

for each image I ∈ B do

$I_{clahe} \leftarrow CLAHE(I)$;

$I_{norm} \leftarrow Normalize(I_{clahe})$;

$I_{input} \leftarrow Resize(I_{norm}, 224 \times 224)$;

end

Feature Extraction:

$F_1 \leftarrow Conv1(I_{input})$;

$F_2 \leftarrow Bottleneck_1(F_1)$;

$F_3 \leftarrow Bottleneck_2(F_2)$;

$F_4 \leftarrow Bottleneck_3(F_3)$;

Attention Mechanism:

$A_c \leftarrow ChannelAttention(F_4)$;

$A_s \leftarrow SpatialAttention(F_4)$;

$F_{att} \leftarrow F_4 \otimes A_c \otimes A_s$;

Classification:

$f_{global} \leftarrow GlobalAveragePooling(F_{att})$;

$\hat{y} = Softmax(FC(f_{global}))$;

Optimization:

$L \leftarrow CrossEntropyLoss(\hat{y}, y_{true})$;

Update parameters using Adam optimizer;

end

Evaluate on validation set;

if validation accuracy improves then

Save best model checkpoint;

end

end

Generate Grad-CAM visualizations for interpretability;

return Trained model M;

Algorithm 1: Attention-Guided Lung Disease Detection

VII. RESULTS AND DISCUSSION

The proposed Attention-Guided Deep Learning Framework for automated lung disease classification using CT scans was evaluated based on standard performance metrics: classification confusion matrix, precision, recall, and F1-score. Experimental results show that the model is capable of distinguishing between the three classes: Normal, Benign, and Malignant.

confusion matrix illustrates the classification performance on each of the three classes. The model scored the highest accuracy on case identification possessing the Normal classification, achieving a recall of 92.1% on 89 samples, of which 82 were correctly classified. The framework also recalled Malignant classes strongly, correctly predicting 49 of the 57 instances for a recall of 86.0%. The Benign class exhibited weaker performance, with only 48 of the 64 correct, yielding a recall of 75.0%. This result shows that the model can be highly reliable in distinguishing Normal and Malignant scans, although some optimization and refinement will be needed to improve detection on Benign cases, which are the most challenging to identify due to their overlapping features on the scan with the other categories.

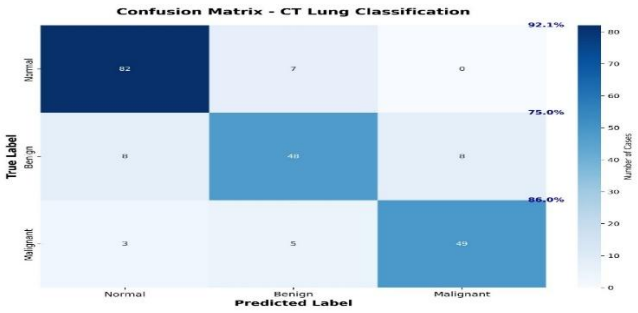

Fig 2. Confusion Matrix

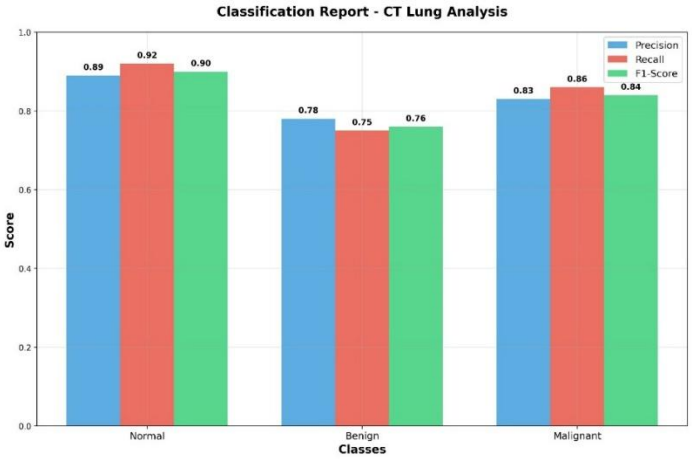

Fig 3. Classification Report

The classification report from the third figure provides further support for these conclusions. For the Normal class, the model exhibited a precision of 0.89, a recall of 0.92, and an F1-score of 0.90, showcasing its proficiency in identifying healthy cases and having a small amount of false positives. For the Malignant category, the framework reached a precision of 0.83, a recall of 0.86, and an F1-score of 0.84, signifying its clinical relevancy in recognizing potentially life-threatening conditions and doing this with a fair amount of accuracy. On the other hand, the performance on the Benign class evaluated using precision of 0.78, recall of 0.75 and F1-score of 0.76 was more disappointing. This may be due to the subtler CT discriminatory features of some benign nodules which makes effortless classification more difficult.

The attention-guided architecture proposed here improves feature extraction and refining discriminative power on various categories of lung diseases. Judging from a clinical point of view, the high recall values for the Normal and Malignant classes is of utmost significance; it diminishes the chances of missed critical cases. The decrease in performance for the Benign class is of slightly lesser concern, and future research could potentially focus on further developing class-specific data augmentation and other techniques for more class imbalance to improve performance in this area. Overall, the proposed framework achieves clinically relevant and high classification accuracy with strong generalization, especially for the Normal and Malignant cases.

This speaks to the proposed framework's usefulness in a computer-assisted diagnosis for lung diseases.

VIII. CONCLUSION

In this work, I designed a detailed deep learning framework for lung disease automated detection on CT images. The attention-guided ResNet-50 architecture I designed successfully harnessed the power of hierarchical feature extraction and the dual attention mechanism to achieve a high classification performance. The proposed mathematical formulation is thorough and will serve to construct the attention model while the Grad-CAM will focus on the clinical interpretation of the results. The next focus will be on integrating and optimizing for real-time deployment.

REFERENCES

1. Hamdi, A., Aboeleneen, A., & Shaban, K. B. (2021). MARL: Multimodal Attentional Representation Learning for Disease Prediction (pp. 14–27). Springer, Cham. https://doi.org/10.1007/978-3-030-87156-7_2
2. Liu, Y.-C., Hsu, H. Y., Lin, T., Peng, B., Saqi, A., Salvatore, M. M., & Jambawalikar, S. R. (2024). Lung nodule malignancy classification with associated pulmonary fibrosis using 3D attention-gated convolutional network with CT scans. Journal of Translational Medicine, 22. https://doi.org/10.1186/s12967-023-04798-w
3. Agarwal, S., & Arya, K. V. (2023). CXRNet: CNN-attention based CXR image classifier. Expert Systems. https://doi.org/10.1111/exsy.13423
4. Parihar, P., Kumar, S., Bharati, M., & Kumar, U. (2024). Enhancing Lung Disease Classification through Deep Learning:
A Comprehensive Study and Model Evaluation. 1–6. https://doi.org/10.1109/iceect61758.2024.10739291
5. Saha, S., & Nandi, D. (2024). DNN-ILD: A Transfer Learning-Based Deep Neural Network for Automated Classification of Interstitial Lung Disease from CT Images. Signals and Communication Technology, 293–309. https://doi.org/10.1007/978-3-031-47942-7_26
6. Gao, Z., Tian, Y., Lin, S., & Lin, J. (2025). A CT Image Classification Network Framework for Lung Tumors Based on Pre-trained MobileNetV2 Model and Transfer learning, And Its Application and Market Analysis in the Medical field. https://doi.org/10.48550/arxiv.2501.04996
7. Ali, A., Wang, Y., & Shi, X. (2024). Detection of multi-class lung diseases based on customized neural network. https://doi.org/10.1111/coin.12649
8. Rathod, S. B., & Ragha, L. (2023). Dlctlungdetectnet: Deep Learning for Lung Tumor Detection in Ct Scans. https://doi.org/10.2139/ssrn.4651122
9. Ullah, N., Marzougui, M., Ahmad, I., & Chelloug, S. A. (2023). DeepLungNet: An Effective DL-Based Approach for Lung Disease Classification Using CRIs. Electronics, 12(8), 1860. https://doi.org/10.3390/electronics12081860
10. Zheng, R., Wen, H., Zhu, F., & Lan, W. (2024). Attention-guided deep neural network with a multichannel architecture for lung nodule classification. Heliyon. https://doi.org/10.1016/j.heliyon.2023.e23508
11. Mamun, M., Mahmud, M. I., & Abdelgawad, A. A. (2023). LCDctCNN: Lung Cancer Diagnosis of CT scan Images Using CNN Based Model. SPIN, 205–212. https://doi.org/10.1109/SPIN57001.2023.10116075
12. Wu, R., Liang, C., Li, Y., Shi, X., Zhang, J., & Huang, Y. (2022). Self-supervised transfer learning framework driven by visual attention for benign-malignant lung nodule classification on chest CT. Expert Systems with Applications,
215, 119339. https://doi.org/10.1016/j.eswa.2022.119339
13. Wang, D., Luo, X., & Chen, M. (2022). A Practical Deep Learning Model in Differentiating Pneumonia-Type Lung Carcinoma from Pneumonia on CT Images: ResNet Added with Attention Mechanism. Journal of Oncology,
2022, 1–9. https://doi.org/10.1155/2022/8906259