Early Detection of Lung Cancer Using Pyramid Vision Transformer (PVT v2): A Comparative Analysis of Deep Learning Models

1st Md. Khurshid Jahan
Dept.of ECE
North South University
Dhaka, Bangladesh
khurshid.jahan@northsouth.edu

2nd Ahmed Faizul Haque Dhrubo
Dept.of ECE
North South University
Dhaka, Bangladesh
ahmed.dhrubo@northosuth.edu

3rd Maher Ali Rusho*
Dept. of Lockheed Martin Engineering Management
University of Colorado, Boulder
Boulder, Colorado
maher.rusho@colorado.edu

4th Ashfaqur Rahman Chowdhury
Dept.of ECE
North South University
Dhaka, Bangladesh
ashfaqur.chowdhury@northsouth.edu

5th Farhana Sharmin
Dept. of Public Health
North South University
Dhaka, Bangladesh
farhnana.sharmin.242@northsouth.edu

6th Mohammad Abdul Qayum
Dept.of ECE
North South University
Dhaka, Bangladesh
mohammad.qayum@northsouth.edu

***Abstract*—Lung cancer remains one of the most prevalent and life-threatening diseases worldwide, with early detection being critical for improved survival rates. This paper explores the application of advanced vision transformers, specifically the Pyramid Vision Transformer (PVT v2), to enhance lung cancer detection from Computed Tomography (CT) scan images. We comprehensively compare state-of-the-art deep learning models, including MaxViT, XCiT, and NextViT, and we investigate their performance in classifying lung cancer into malignant, benign, and standard categories. Utilizing the IQ-OTH/NCCD dataset, we fine-tune PVT v2 alongside other models, implementing rigorous preprocessing, augmentation, and regularization techniques to achieve robust results. Our experimental findings demonstrate that PVT v2 achieves superior accuracy with 99.09% and generalization, outperforming other models on multiple evaluation metrics such as precision, recall, Receiver-operating Characteristic curve (ROC) and Area under the curve (AUC) compared to published research. Despite challenges like dataset imbalance and computational costs, our work highlights the potential of vision transformers to revolutionize lung cancer detection and its application in public health.**

***Index Terms*—Lung Cancer Detection, Pyramid Vision Transformer (PVT v2), Vision Transformers, Deep Learning, CT Scans, MaxViT, XCiT, NextViT, Image Classification, Medical Imaging, Early Detection, Receiver-operating Characteristic curve (ROC) and Area under the curve (AUC).**

I. Introduction

In the world, lung cancer is one of the most critical diseases nowadays [1]. World Health Organization (WHO) estimated that around 7.6 million deaths worldwide per year are caused by lung cancer [2]. In 2030, humanity because of cancer is supposed to continue rising, to become around 17 million worldwide [3]. The only method of its cure is to find out lung cancer in the early stage [4]. MRI, isotope, and CT methods are available for the diagnosis of lung cancer, but X-ray chest radiography and Computer Tomography (CT) are the two familiar anatomic imaging modalities that are regularly used in the recognition of different lung diseases [5], [6]. Physicians and radiologists use CT images to identify and recognize the presence of lung cancer [7].

Lung cancer disease cannot be identified easily in CT images. Because of that, low-dose helical computed tomography (LDCT) is applied as a modality [8], [9]. The computer-aided automatic detection (CAD) process must be applied to the clinical center to develop an effective cancer prediction system using an optimized and intelligent technique [10]. Using deep learning in CAD systems has many advantages, like it can perform end-to-end detection by learning the most salient features during training time [11]. This makes the network robust to variations as it captures nodules' features in various CT scans with varying parameters.

Recently, vision transformers have also become very popular in medical sector research [12]. The Swin transformer was a popular choice among researchers; many other architectures were also found where the vision transformer was combined with convolution neural networks or the UNet model [13]. The advent of transformers apprised researchers of CNNs' major drawback, the inability to capture long-range dependencies such as the extraction of contextual information and the non-local correlation of objects [13].

II. Literature Review

Roy, Sirohi, and Patle [14]developed a system to detect lung cancer nodules using a fuzzy inference system for classification. This method uses gray transformation for image contrast enhancement. The resulting image is segmented using an active contour model. Features like area, mean, entropy, correlation, primary axis length, and minor axis length are extracted to train the classifier. The limitation of this method

is that it does not classify the cancer as benign or malignant, which is the future scope of this proposed model.

Hiram et al. [15] have classified lung nodules using the Frequency domain and SVM with RBF. Neural Ensemble-based Detection (NED) was proposed by Zhi-Hua et al. [16], which utilized an artificial neural network ensemble to identify lung cancer cells. This method provides high accuracy in the identification of cancer cells. Hui Chen et al. [17] have provided a computerized scheme for formatting a lung nodule's classification on a thin-section CT scan using a Neural Network Ensemble (NNE).

Parmar et al. [18] have been used to derive features to describe images quantitatively. The integration of AI-enabled diagnostic tools improves the precision of tumor characterization and paves the way for personalized interventions in lung cancer. A pre-trained GoogleNet architecture was proposed by Sajja et al [19]. They tested the method on the LIDC dataset of CT scans and achieved the highest success compared to several pre-trained CNN models like ResNet 50, AlexNet, and GoogleNet. Lyu [20] proposed an ensemble approach to the IQ-QTH/NCCD dataset using four different CNN models: AlexNet, VGG, DCNN, and DenseNet. It can be seen that the ensemble approach improves the performance of the single model, and the DenseNet architecture provides the highest accuracy by using GoogleNet Model Al-Hussein et al. [21], achieving 94.38% accuracy. Raza et al. [22] proposed EfficientNetB1 architecture using a novel transfer learning-based approach for lung cancer diagnosis. With the proposed Lung-EffNet, they achieved 99.10% accuracy on the test set. They also used various data augmentation techniques to solve the imbalance problem in the dataset.

III. Background Study

A. Existing Technologies

Many deep learning models have been developed to aid in the detection of lung cancer, primarily using CT and MRI scans. EFFI-CNN uses lung CT scan images from LIDC-IDRI and Mendeley data sets [23]. EFFI-CNN has a unique combination of CNN layers with parameters. AlexNet CNN is also used to detect lung cancer because the proposed CNN achieves a high degree of accuracy, which is more effective [24]. CNN architecture, a DL algorithm, helps detect lung cancer. As the methods deal with binary classification, which confirms yes/no of the lung cancer presence in the human body, both SVM and CNN methods are more straightforward than any other ML/DL algorithms for this lung cancer data considered [25]. ResNet-50/101 and EfficientNet-B3 are also commonly used for lung cancer diagnosis because these models utilize transfer learning, where they are pre-trained on large datasets like ImageNet and then fine-tuned for medical applications, such as identifying cancerous nodules from CT images [26]. YOLOv8 has been shown to achieve high accuracy in lung cancer detection, especially in CT images [27]. 3D Residual CNN is employed to reduce false positives in lung nodule detection [27]. DenseNets have been used to classify lung nodules, crucial for early cancer detection [27]. ARNN has an attention layer that performs the encoding and decoding process within single sequences using variable-length vectors. It also has the advantage of performing automatic feature extraction [28]. A hybrid bidirectional Long-Short-Term-Memory (BiDLSTM)-Mask Region-Based Convolutional Neural Network (Mask-RCNN) model proposes a lung disease prediction framework [29].

B. Our Approach

As we know, CNN, RNN, and other deep learning models are very common for detecting lung cancer and other medical diseases [30]. So, we want to introduce everyone to some new models of vision transformers. The first is PVT v2, which is known as the Pyramid Vision Transformer. PVT v2 can obtain more local image and feature map continuity than PVT v1 [31]. It can process variable-resolution input flexibly and enjoy the same linear complexity as CNN. Secondly, ViTamin, with only 436M parameters and trained on the public DataComp-1B dataset, achieves an impressive 82.9% zero-shot ImageNet accuracy [32]. Thirdly, MaxViT uses a hierarchical backbone similar to standard ConvNet practices where the input is first downsampled using Conv3x3 layers in the stem stage. The body of the network contains four stages, with each stage having half the resolution of the previous one with a doubled number of channels [33]. Fourthly, NextViT is a next-generation vision Transformer for efficient deployment in realistic industrial scenarios, which dominates both CNNs and ViTs from the perspective of latency/accuracy trade-off [34]. Fifthly, XCiT is known as the Cross-Covariance Image Transformer. It is built upon XCA and combines conventional transformers' accuracy with convolutional architectures' scalability. XCiT by reporting excellent results on multiple vision benchmarks, including self-supervised learning for image classification on ImageNet-1k, object detection, instance segmentation on COCO, and semantic segmentation on ADE20k [35].

IV. Methodology

A. Data Analysis

For our research, we collected the dataset from Kaggle [36]–[38]. This dataset is based on cancer diseases (IQ-OTH/NCCD) lung cancer, and it was collected in the specialist hospitals mentioned above over three months in the fall of 2019. CT scans of different patients with lung cancer in various stages are included here. The dataset has three classes of norm: abnormal, benign, and malignant. The dataset contains 1190 images representing CT scan slices of 110 cases. Among them, 40 cases are diagnosed as malignant, 15 cases are diagnosed as benign, and 55 cases are classified as typical cases. Here, Figures 1, 2, and 3 represent three states of lung cancer from the dataset.

B. System Design

We utilized the Pvt v2 for the image classification model, especially the PVT-v2-b0 architecture from OpenGVLab, which

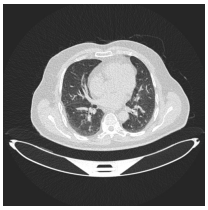

Fig. 1. Bengin Case

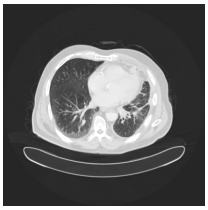

Fig. 2. Malignant Case

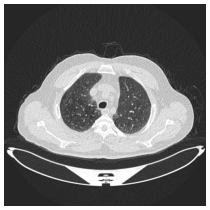

Fig. 3. Normal Case

we fine-tuned to classify lung cancer images into three categories: benign, malignant, and normal. Data was preprocessed using image transformations, including resizing to 224x224, grayscale conversion, Gaussian blur, random horizontal flips, rotations, and color jitter to enhance generalization. The dataset was split into training into 80%, validation into 10%, and test into 10% sets, with appropriate normalization applied. Figure 4 shows all the procedures for designing our models and training the dataset.

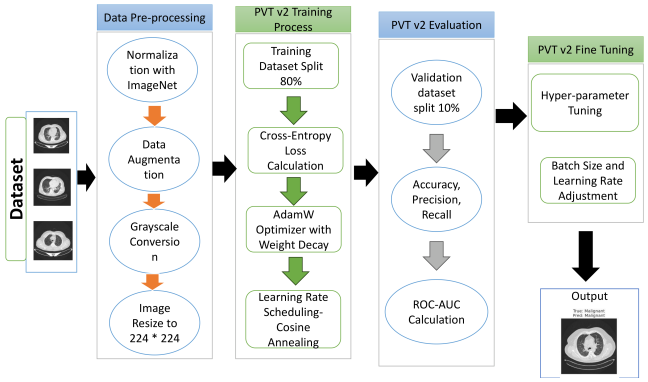

Fig. 4. Proposed System Architecture of Pyramid Vision Transformers.

We employed a custom dataset class to load images and labels, leveraging a Trainer from the Hugging Face library to streamline the training process. The training was conducted over 50 epochs, using a cosine learning rate scheduler with restarts and weight decay for regularization. The model was optimized and evaluated based on key metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, computed on both validation and test datasets, providing robust insights into the model's performance. We also apply this cross-entropy loss formula in our models, which calculates the difference between the predicted class probabilities and the ground truth:

$$L = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

To find out the accuracy of every model, we use this formula:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} 1(\hat{y}_i = y_i)$$

For other models such as NextViT, MaxViT, XCiT, and ViTamin, we set the batch size to 32, the learning rate at 0.001. To calculate the learning rate, we applied this Learning Rate Scheduling (Cosine Annealing) formula:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{T_{\text{cur}}}{T_{\max}}\pi\right)\right)$$

We use 50 epochs, and the weight decay of 1e-4(L2 regularization) and applied this formula:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda\sum_{w} w^2$$

We use the AdamW optimizer, image size into 224 * 224 as input resolution for models. The augmentation we use in the dataset is a random resized crop, random horizontal flip, random rotation, random grayscale conversion, and color jitter-like brightness, contrast, and hue adjustment. We also resized all images to 224x224 resolution and normalized them using ImageNet mean and standard deviation values: [0.485, 0.456, 0.406] for mean and [0.229, 0.224, 0.225] for standard deviation.

V. Result and Analysis

Our research focused on the PVT v2 model, but we also used many pre-trained models in the same dataset, such as MaxViT, NextViT, Vitamin, and XCiT. All models were assessed based on training accuracy, validation accuracy, training loss, and validation loss. This way, we can understand each model's performance on the training dataset and its capacity to generalize to unseen data. Table I represents the performance metrics for all models across these tasks.

TABLE I
Overall Model Performance Results

Model	Training Loss	Training Accuracy	Validation Loss	Validation accuracy
MaxViT	0.0330	98.29%	0.0749	98.17%
NextViT	0.0620	97.49%	0.2774	91.74%
Vitamin	0.1961	92.12%	0.2774	91.74%
XCiT	0.0535	98.06%	0.1164	97.25%
PVT v2	0.1278	99.09%	0.0507	98.18%

Here, in training accuracy, PVT v2 achieved the highest accuracy with 99.09%, and MaxViT was in second place with 98.29%. XCiT and NextViT also reached good accuracy with 98.06% and 97.49%. Vitamin lagged with a training accuracy of 92.12%. Regarding validation accuracy, PVT v2 performed best at 98.18%, and with a bit of low accuracy, MaxViT performed at 98.17%. XCiT maintained a vital validation accuracy of 97.25%, while NextViT and Vitamin performed less than other models at 91.74%.

Figure 5 shows the output of our model, which detects lung cancer. MaxViT exhibited a low value of 0.0749 regarding validation loss, indicating good generalization capabilities. PVT v2 showed an even lower validation loss of 0.0507, which suggests that it adapts well to unseen data. XCiT recorded a validation loss of 0.1164, indicating a solid performance. On the other hand, NextViT and Vitamin faced challenges with

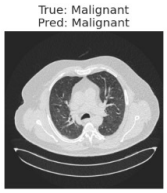

Fig. 5. Performance Outcomes of the Lung Cancer Detection System.

higher validation losses of 0.2774 and 0.2774, respectively, which represent these two models could imply overfitting issues or difficulty capturing the dataset's complexities. Figure 6 represents the training and validation loss and accuracy curves, which shows how our model performs well.

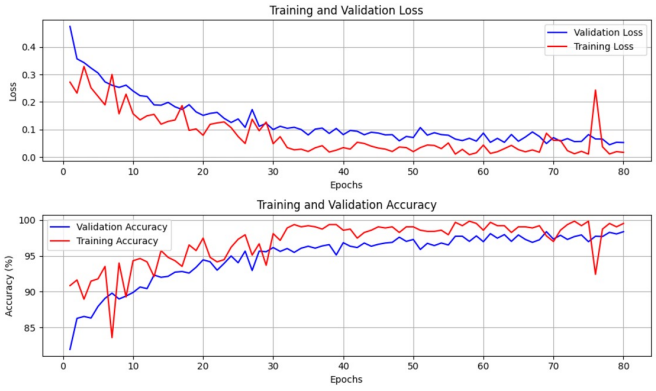

Fig. 6. Training and Validation Loss & Accuracy Curves for PVT v2.

In terms of training loss, MaxViT achieved the lowest value at 0.0330, indicating excellent learning and minimal error during training. XCiT placed second by a training loss of 0.0535, showcasing its effective training process. Although PVT v2 had a higher training loss of 0.1278, it maintained competitive training accuracy. NextViT and Vitamin experienced even higher training losses of 0.0620 and 0.1961, respectively, which correlate with their lower training accuracies, suggesting challenges in capturing the dataset's complexities.

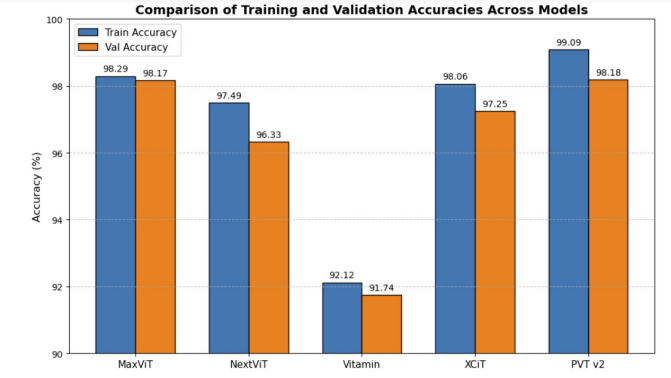

Fig. 7. Analyzing Accuracy Trends in Different Models

Figure 7 represents all the model's training and validation accuracies. Finally, MaxViT and PVT v2 are the best-performing models, showcasing superior training, validation accuracy, and low validation loss. MaxViT efficiently extracts complex features, leading to high accuracy and minimal errors. PVT v2 also demonstrates strong generalization capabilities, indicating its potential for practical applications. On the other hand, XCiT also shows promising results; NextViT and Vitamin highlight the challenges of achieving effective performance and generalization, suggesting that these models need further optimization. Overall, the findings underscore the advantage of utilizing advanced architectures like MaxViT and PVT v2 for tackling complex datasets, particularly in contexts requiring robust feature extraction and generalization.

VI. Comparison

Prity et al. [39] use this same dataset, and they apply CNN-based model XceptionNet where they achieve 99.64% in Validation accuracy and 99.51% in training accuracy. They trained their model for 80 epochs with early stopping, 8 batch sizes, and for learning rate 1e-5. For comparison, they also applied Inception V3, DenseNet201, EfficientNetB7, MobileNetV2, NASNetMobile, and ResNet101v2. Jafar Abdollahi et al. [40] also used the same dataset to evaluate LeNet algorithms for classification. Here, they Achieved 97.88% accuracy, 83.17% validation accuracy, and 93.17% sensitivity.

TABLE II
ANALYZING PERFORMANCE MEASURES IN COMPARISON TO OTHER TECHNIQUES

Study	Year	Methodology	Accuracy
Prity et al.	2023	CNN-Based Model (XceptionNet)	99.64&
Jafar Abdollahi et al.	2023	LeNet Algorithm	83.17%
Abdalbasit Mohammed Qadir et al.	2023	Hybrid-LCSCDM	98.54%
Our proposed method (PVT v2)	2024	Pyramid Vision Transformer	99.09%

Abdalbasit Mohammed Qadir et al. [41] used an innovative lung cancer detection system known as the Hybrid Lung Cancer Stage Classifier and Diagnosis Model known as Hybrid-LCSCDM. Here, they approach in two ways. The first approach used a pre-trained model VGG-16 for detecting key features in lung cancer, and secondly, the classifier used machine learning, which is XGBoost. They addressed the class imbalance challenge in our dataset by applying Stratified 5-fold Cross-Validation. Their strategy achieved an overall accuracy of 98.54%. In that case, we worked with different approaches. We used transformer-based models here to determine which model gave us better performance. After analyzing other models in Table II and all our model's performances, we noticed that the Pyramid vision transformer, known as PVT v2, gave us better training and validation accuracy. We got 99.09% in training accuracy and 98.18% in validation accuracy.

VII. Applications, Limitations, and Future work

The application of advanced deep learning models, such as PVT-V2, holds significant potential in the early detection of lung cancer, a leading cause of mortality worldwide. Early and accurate diagnosis is critical for improving survival rates, and PVT-V2 can enhance the analysis of CT scans by enabling precise, rapid, and automated detection of cancer nodules, as well as their classification into benign or malignant categories. This AI-driven diagnostic approach has substantial public health implications, particularly in resource-limited settings where trained radiologists and advanced diagnostic tools are scarce.

Beyond lung cancer, the methodology can be extended to other critical diseases, laying the groundwork for AI-driven solutions to complex health challenges. Such applications support broader public health goals, addressing healthcare disparities, fostering innovation in disease prevention, and enhancing management strategies. However, the integration of AI models like PVT-V2 into clinical practice poses challenges, including the need for seamless compatibility with existing healthcare infrastructure and fostering collaboration among clinicians, technologists, and policymakers. Addressing these challenges is essential to maximizing the impact of AI in advancing equitable and effective public health interventions.

Our significant limitation is that we only work on one dataset based on Iraq. Still, lung cancer is a worldwide problem, so we have to increase our dataset and combine other countries' datasets. Although the images were resized to 224x224 pixels, this could have caused a loss of critical details, especially when detecting smaller nodules or more intricate features of malignant cases. The reduced resolution may affect the model's ability to effectively differentiate between benign and malignant tumors. NextViT and Vitamin exhibited higher training losses, suggesting that these models might overfit the training data. This could reduce their ability to generalize to unseen data, leading to a drop in performance during real-world deployment. The dataset is imbalanced, with more benign and normal cases than malignant ones. We also should have to include MRI-based datasets in our new dataset for our models. We will develop real-time lung cancer detection models integrated with cloud-based systems to assist radiologists in quicker diagnosis and treatment decisions. These systems should be capable of handling high volumes of patient data, and further work should be done to ensure the model's scalability and integration with hospital infrastructure.

VIII. Conclusion

Early detection is critically important if a significant reduction in lung cancer morbidity and mortality is to be realized. Although some patients at risk of lung cancer have other comorbidities that could preclude surgery, new approaches such as stereotactic ablative body radiotherapy (SABR) have had promising results in treating patients with stage. It also needs to be clear how many CT scans patients should have throughout their lifetime. The proposed technique gives promising results compared with other methods that have been used. Vision transformer models are developed daily, so we should not use old techniques in the future because old models will not handle upcoming advanced machine images. Moving in vision transformers instead of old models will bring a massive revolution. It is more efficient than other models to detect even more minor things that old models cannot do.

References

[1] S. Sun, B. Zheng, and W. Qian, "Computer-aided lung cancer diagnosis with deep learning algorithms," in Medical Imaging: Computer-Aided Diagnosis, vol. 9785, International Society for Optics and Photonics, 2016.

[2] L. Nie, L. Zhang, Y. Yang, M. Wang, R. Hong, and T.-S. Chua, "Beyond doctors: Future health prediction from multimedia and multimodal observations," in Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 591–600.

[3] B. U. Dhaware and A. C. Pise, "Lung cancer detection using Bayesian classifier and FCM segmentation," in 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), pp. 170–174, IEEE, 2016.

[4] G. L. F. da Silva, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, and M. Gattass, "Taxonomic indexes for differentiating malignancy of lung nodules on CT images," Research on Biomedical Engineering, vol. 32, no. 3, pp. 263–272, 2016.

[5] S. C. Park, J. Tan, X. Wang, D. Lederman, J. K. Leader, S. H. Kim, and B. Zheng, "Computer-aided detection of early interstitial lung diseases using low-dose CT images," Physics in Medicine & Biology, vol. 56, pp. 1139–1153, 2011. doi: 10.1088/0031-9155/56/4/016.

[6] Q. Z. Song, L. Zhao, X. K. Luo, and X. C. Dou, "Using deep learning for classification of lung nodules on computed tomography images," Journal of Healthcare Engineering, 2017. doi: 10.1155/2017/8314740.

[7] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification," Multimedia Tools and Applications, vol. 79, no. 11–12, pp. 7731–7762, Jan. 2020, doi: https://doi.org/10.1007/s11042-019-08394-3.

[8] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in Proceedings of the 12th IEEE Conference on Computer and Robot Vision (CRV), 2015, pp. 133–138. doi: 10.1109/CRV.2015.25.

[9] K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," IEEE Transactions on Medical Imaging, vol. 24, no. 9, pp. 1138–1150, Sep. 2005.

[10] R. Manser et al., "Screening for lung cancer," Cochrane Database of Systematic Reviews, Jun. 2013.

[11] D. Riquelme and M. A. Akhloufi, "Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans," AI, vol. 1, no. 1, pp. 28–67, Mar. 2020, doi: https://doi.org/10.3390/ai1010003.

[12] Tolgahan Gulsoy and Elif Baykal Kablan, "Diagnosis of lung cancer based on CT scans using Vision Transformers," Nov. 2023, doi: https://doi.org/10.1109/eleco60389.2023.10416046.

[13] H. Ali, F. Mohsen, and Z. Shah, "Improving diagnosis and prognosis of lung cancer using vision transformers: a scoping review," BMC Medical Imaging, vol. 23, no. 1, Sep. 2023, doi: https://doi.org/10.1186/s12880-023-01098-z.

[14] T. S. Roy, N. Sirohi, and A. Patle, "Classification of lung image and nodule detection using fuzzy inference system," in Proceedings of the 2015 International Conference on Computing, Communication and Automation (ICCCA), 2015, pp. 1204–1207.

[15] H. M. Orozco, O. O. V. Villegas, L. O. Maynez, V. G. C. Sanchez, and H. de Jesus Ochoa Dominguez, "Lung nodule classification in frequency domain using support vector machine," in Proceedings of the International Conference on Information Science, Signal Processing and Their Applications (ISSPA), 2012.

[16] Z.-H. Zhou, Y. Jiang, Y.-B. Yang, and S.-F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artificial Intelligence in Medicine, vol. 24, pp. 25–36, 2002.

[17] H. Chen, Y. Xu, Y. Ma, and B. Ma, "Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on CT images," Academic Radiology, vol. 17, no. 5, 2010.

[18] C. Parmar, J. D. Barry, A. Hosny, J. Quackenbush and H. J. Aerts, "Data analysis strategies in medical imaging", Clinical cancer research, vol. 24, no. 15, pp. 3492-3499, 2018.

[19] T. Sajja, R. Devarapalli and H. Kalluri, "Lung cancer detection based on ct scan images by using deep transfer learning", Traitement du Signal, vol. 36, no. 4, pp. 339-344, 2019.

[20] L. Lyu, "Lung cancer diagnosis based on convolutional neural networks ensemble model", 2021 2nd International Seminar on Artificial Intelligence Networking and Information Technology (AINIT), pp. 360-367, 2021.

[21] M. S. Al-Huseiny and A. S. Sajit, "Transfer learning with googlenet for detection of lung cancer", Indonesian Journal of Electrical Engineering and Computer Science, vol. 22, no. 2, pp. 1078-1086, 2021.

[22] R. Raza, F. Zulfiqar, M. O. Khan, M. Arif, A. Alvi, M. A. Iftikhar, et al., "Lung-effnet: Lung cancer classification using efficientnet from ct-scan images", Engineering Applications of Artificial Intelligence, vol. 126, pp. 106902, 2023.

[23] "Efficient CNN for Lung Cancer Detection," International Journal of Recent Technology and Engineering, vol. 8, no. 2, pp. 3499–3505, Jul. 2019, doi: https://doi.org/10.35940/ijrte.b2921.078219.

[24] A. Agarwal, Kritik Patni, and D Rajeswari, "Lung Cancer Detection and Classification Based on Alexnet CNN," Jul. 2021, doi: https://doi.org/10.1109/icces51350.2021.9489033.

[25] Chintakayala Tejaswini, P. Nagabushanam, Priyadharshini Rajasegaran, Palyam Rohith Johnson, and S. Radha, "CNN Architecture for Lung Cancer Detection," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Apr. 2022, doi: https://doi.org/10.1109/csnt54456.2022.9787650.

[26] V. Kumar, Chander Prabha, P. Sharma, N. Mittal, S. S. Askar, and M. Abouhawwash, "Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images," BMC medical imaging, vol. 24, no. 1, Mar. 2024, doi: https://doi.org/10.1186/s12880-024-01241-4.

[27] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," Diagnostics, vol. 13, no. 16, p. 2617, Jan. 2023, doi: https://doi.org/10.3390/diagnostics13162617.

[28] S. Sasikumar, P.N. Renjith, K. Ramesh, and K. Sakthidasan Sankaran, "Attention Based Recurrent Neural Network for Lung Cancer Detection," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Oct. 2020, doi: https://doi.org/10.1109/i-smac49090.2020.9243556.

[29] V. Indumathi and R. Siva, "An efficient lung disease classification from X-ray images using hybrid Mask-RCNN and BiDLSTM," Biomedical Signal Processing and Control, vol. 81, p. 104340, Mar. 2023, doi: https://doi.org/10.1016/j.bspc.2022.104340.

[30] Seyed Hesamoddin Hosseini, Reza Monsefi, and Shabnam Shadroo, "Deep learning applications for lung cancer diagnosis: A systematic review," Multimedia Tools and Applications, Jul. 2023, doi: https://doi.org/10.1007/s11042-023-16046-w.

[31] W. Wang et al., "PVT v2: Improved baselines with Pyramid Vision Transformer," Computational Visual Media, vol. 8, no. 3, pp. 415–424, Mar. 2022, doi: https://doi.org/10.1007/s41095-022-0274-8.

[32] Beckschen, "GitHub - Beckschen/ViTamin: [CVPR 2024] Official implementation of 'ViTamin: Designing Scalable Vision Models in the Vision-language Era,'" GitHub, 2024. https://github.com/Beckschen/ViTamin (accessed Sep. 20, 2024).

[33] Z. Tu et al., "MaxViT: Multi-Axis Vision Transformer." Accessed: Sep. 20, 2024. [Online]. Available: https://arxiv.org/pdf/2204.01697

[34] J. Li et al., "Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios." Accessed: Sep. 20, 2024. [Online]. Available: https://arxiv.org/pdf/2207.05501

[35] A. El-Nouby et al., "XCiT: Cross-Covariance Image Transformers." Accessed: Sep. 20, 2024. [Online]. Available: https://arxiv.org/pdf/2106.09681

[36] H. F. Al-Yasriy, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of Lung Cancer Based on CT Scans Using CNN," IOP Conference Series: Materials Science and Engineering, vol. 928, 2020.

[37] H. F. Kareem, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of SVM Performance in the Detection of Lung Cancer in Marked CT Scan Dataset," Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 3, pp. 1731-1738, 2021, doi: 10.11591/ijeecs.v21.i3.pp1731-1738.

[38] H. Alyasriy and M. Al-Huseiny, "The IQ-OTH/NCCD Lung Cancer Dataset," Mendeley Data, V4, 2023, doi: 10.17632/bhmdr45bh2.4.

[39] Afrina Jannat Prity, M. Firoz Mridha, and Md. Abdullah-Al-Jubair, "CT Image-Based Lung Cancer Detection Using a CNN-Based Model," pp. 488–494, Oct. 2023, doi: https://doi.org/10.1109/icdabi60145.2023.10629354.

[40] J. Abdollahi, "Evaluating LeNet Algorithms in Classification Lung Cancer from Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases," arXiv.org, 2023. https://arxiv.org/abs/2305.13333 (accessed Nov. 29, 2024).

[41] Abdalbasit Mohammed Qadir, Peshraw Ahmed Abdalla, and Dana Faiq Abd, "A Hybrid Lung Cancer Model for Diagnosis and Stage Classification from Computed Tomography Images," Iraqi Journal for Electrical And Electronic Engineering, vol. 20, no. 2, pp. 266–274, Aug. 2024, doi: https://doi.org/10.37917/ijeee.20.2.23.