Efficient Lung Cancer Detection Leveraging Hybrid Deep Learning Models

Lavanya G[1[0000-0001-8958-3123]], Krishna Prakaash D M[2], Ishwarya V[3] and Paarkavy B[4]

[1,2,4] Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India
[3] Department of Artificial Intelligence and Machine Learning, Saveetha Engineering College, Chennai, India
lavanyajoyce2021@gmail.com, suruthikrish11805@gmail.com, ishuvasanth05@gmail.com, paarkavy.bv@gmail.com.
Corresponding Author: lavanyajoyce2021@gmail.com

Abstract. Due to late diagnosis and the inability to find precancerous stages, lung cancer remains the leading cause of death from cancer. The research work presented in the article describes the application of Vision Transformers (ViTs) in generating an accessible and efficient method of early lung cancer screening. It accurately defines lung disorders as normal, benign, or malignant by the global attention ability of ViTs. Techniques such as data augmentation for better generalization, transfer learning to get maximum performance on medical datasets, and weighted loss functions in class imbalance have been used. We identified that the performance of the hybrid model of vision transformer with Convolutional Neural Networks (CNN) in the early detection of lung cancer outperforms other deep learning networks like standalone EfficientNet and Convolutional Neural Networks, achieving a diagnostic accuracy of 94.1%. AI-assisted diagnosis realizes enhanced transparency and self-assurance through visual enlightenments with Grad-CAM as interpretable.

Keywords: Lung cancer, Vision Transformers, Deep learning, Grad-CAM, AI-assisted diagnosis

1 Introduction

The primary cause of cancer-related deaths globally is still lung cancer, primarily because of late-stage diagnosis and the difficulty in identifying anomalies at an early stage [1]. A high death rate results from traditional diagnostic techniques like radiology and biopsies frequently failing to detect early symptoms of the disease [17], even with breakthroughs in medical imaging [11]. However early and precise lung cancer identification is essential since prompt care can greatly increase patient survival rates. The use of cutting-edge artificial intelligence (AI) methods to improve diagnosis and help medical practitioners make quicker and more accurate judgments is becoming increasingly popular.

In medical image analysis, deep learning models have demonstrated encouraging outcomes in recent years[5], especially for tasks like radiology image segmentation, anomaly detection, and classification. Convolutional Neural Networks (CNNs) have formed the basis for several state-of-the-art systems to extract spatial information and patterns from images. Nevertheless, CNNs might not be able to capture global relationships in intricate images fully. This restriction is addressed by a more recent class of deep learn- ing models called Vision Transformers (ViTs) [11], which use attention processes to capture contextual information and long-range dependencies—two crucial components for tasks like lung cancer classification.

It presents a novel hybrid framework for pulmonary condition classification in the context of medical image techniques. It integrates Vision Transformers[11] with Convolutional Neural Networks to improve the speed and accuracy of lung disease diagnoses by fusing a global feature understanding provided by Vision Transformers with spatial feature extraction strengths inherent to Convolutional Neural Networks on distinctions to normal, benign, or malignant categories[12]. The challenge of class imbalance and the type of subtle anomalies in the medial datasets is addressed because advanced techniques like weighted loss functions, data augmentation, and transfer learning are applied to this problem. Grad-CAM (Gradient-weighted Class Activation Mapping) [16] also increases the model's interpretability and helps engender trust in AI-assisted diagnosis through visual explanations of decision-making processes for the model.

The present research work is beneficial in enhancing AI-based diagnosis by utilizing a Vision Transformer (ViT)-powered model for lung disease classification toward achieving high diagnostic accuracy. The scalability and ease of accessibility of the model make it even better to deploy on different healthcare systems, especially where resources are scarce and access to specialist doctors is limited. With minimized reliance on manual processes, the proposed approach has the advantage of

a dependable and effective diagnostic tool that allows for earlier detection and timely intervention, thus enhancing patient outcomes.

The review of the latest literature on lung cancer detection is presented in Section 2. The methodology is addressed in Section 3, and the results of applying the vision transformer in lung cancer diagnosis are revealed in Section 4. Section 5 presents the conclusion.5 provides the conclusion.

2 Literature Survey

The use of advanced imaging methods supplemented with deep learning and big medical data is investigated to detect lung infections such as tuberculosis and pneumo-nia, which are significant global health issues. Thamizhselvi et al examine transfer learning in deep learning models [1] to enhance diagnostic accuracy, helping doctors and radiologists make more precise decisions based on chest X-rays.CT scans [20] of the Lung Image Database Consortium (LIDC) are utilized todetect pulmonary nodules, which occur in about 34% of lung cancercases [2]. The nodule sample database is enriched with data preparation methods like pixel interception and normalization, and data augmentation procedures like rotation and scaling. Abbas et al propose a novel CNN design named Decompose, Transfer, and Compose (DeTraC)[3]. The methodology makes use of class decomposition and transfer learning to improve medical image classification. In spite the substantial improvement achieved by deep CNNs in image classification, the issue of limited annotated data in medical imaging remains. By breaking down classes into more unique subclasses, DeTraC facilitates learning and speeds up convergence. Detection of lung cancer is treated using image processing, segmentation, feature extraction, and classification methods, and stage detection to detect lung cancer from CT images [4]. Cancerous nodule staging and classification are performed using an Artificial Neural Network (ANN). The system is developed using Android smartphone technology to facilitate easier cancer diagnosis. Early detection is crucial for improving survival rates and expanding access to treatment options. Ruprah et al proposed an optimized approach for early-stage lung cancer detection[5] is introduced using the VGG-16 convolutional neural network (CNN). The proposed research work enhances the VGG-16[10] model to differentiate between malignant, benign, and normal conditions in CT scan images[17], integrating several optimization techniques such as Gaussian blur, SMOTE (Synthetic Minority Over-sampling Technique), transfer learning, and early callback[5]. Early detection is crucial for prognosis[18], and AI methods[19], particularly deep learning models[20], are increasingly being used to enhance diagnostic accuracy. Convolutional Neural Networks (CNNs)[8,10] are shown to be particularly efficient for medical image classification, as demonstrated in various lung cancer cases. Metagar and Sayyed present and validate a method of using CNNs to classify lung tumors based on their malignancy, compared to traditional diagnostic methods using CT scan images [7]. After testing several ML models [9,14], the one with the highest accuracy was implemented in real time to detect lung cancer. System performance was evaluated to assess the effectiveness of lung cancer detection. Additionally, a user-friendly interface was developed by Murthy et al for medical experts was incorporated to assist in disease diagnosis.

The literature review highlights how important advanced deep learning methods are for the early identification and categorization of lung conditions such as pneumonia, TB, and cancers. To overcome issues like class imbalance and a lack of annotated data, techniques like CNNs, Vision Transformers, and transfer learning are frequently used. These methods achieve exceptional accuracy by utilizing preprocessing, augmentation, and optimization techniques. By combining transfer learning, class decomposition, and other cutting-edge techniques, well-known models like VGG-16, RESNET, and De-TraC demonstrate increased diagnostic precision. To improve patient outcomes, this proposed research work emphasizes the significance of early-stage detection, precise classification, and easily available diagnostic tools.

The methods of screening for lung cancer have conventionally been achieved through the use of imagery, especially CT scans, and biopsies. Such procedures are very source-intensive and often need professional interpretation. Early-stage lung cancers also become challenging to identify because the symptoms are mild and easily overlooked. These methods have also inter-observer variability, which may likely cause delayed interventions and uneven accuracy in diagnosis. Recent evidence of the presence of Artificial Intelligence and Machine Learning [13,15] has led to the discovery of automated systems that have been designed to help in the diagnostic processes. Among them, impressive potential is found in applications concerning CNNs, which are capable of detecting spatial features in medical imaging. However, CNN-based systems often struggle with key challenges such as Class Imbalance, Subtle Feature Detection, Global Context Limitation, and lack of Interpretability.
Some systems, although applying data augmentation, weighted loss functions, and transfer learning to further their performance [1], lack robustness for cross-dataset use. These limitations thus call for a structure that is going to be strong and superior, as strong as the previous models but without their disabilities. It will focus on accuracy, scalability, and interpretability in realizing early and reliable diagnosis of lung cancer [14].To tackle data imbalance, resampling techniques are employed, followed by advanced deep learning algorithms of ViTs, to achieve high accuracy in classifying lung cancer images. The use of deep learning models to identify lung cancer through Vision Transformers (ViT) and Convolutional Neural Networks (CNNs) is explored, with similarities drawn to air quality prediction [6]. Similar to the approach in air quality prediction, this work emphasizes relevant dataset analysis for feature identification that significantly impacts classification. Preprocessing techniques, including data normalization and augmentation, are applied to create balanced and high-quality datasets.

3 Methodology

Using such cutting-edge deep learning techniques for lung disease classification, the proposed approach combines Vision Transformers (ViTs) and Convolutional Neural Networks (CNN). The process flow of the proposed hybrid model is shown in Fig. 1.

3.1. Data gathering and preliminary processing

Three different kinds of CT scan images are included in the dataset: positive, negative, and normal images. Suitable pre-processing, augmentation, and normalization methods are applied to prepare the data for the proposed method of lung cancer detection. In addition to applying pixel normalization, images are scaled to 224x224. After that, the various RGB channels' durations were set to 0.485, 0.456, and 0.406, with corresponding cycle lengths of 0.229, 0.224, and 0.225 to prevent inconsistent inputs to the model.

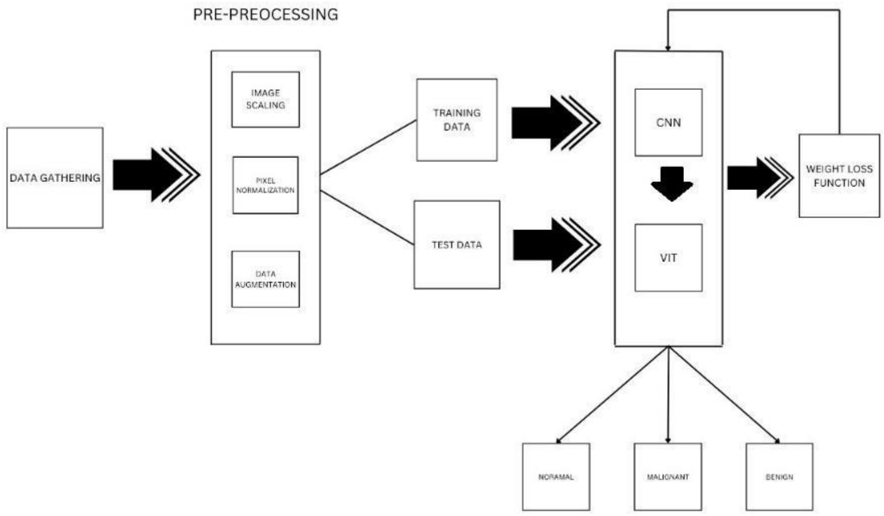

Fig. 1. Process flow of the proposed system

Affine transformations, random rotations, and random flips in horizontal and vertical directions are the augmentation techniques used to increase the dataset's diversity. This reduces overfitting and improves the likelihood that the trained model will generalize to fresh data. Twenty percent of the original data is used for testing, while the remaining eighty percent is used to create the final training set. Fig.2 shows the sample CT scan images extracted from the given dataset for each class in the classification task

(a) Normal (b) Malignant (c) Benign

Fig. 2. Different Classes of Lung Cancer

3.2 Model Architecture

The proposed architecture for the model efficiently extracts global and local characteristics that would allow for accurate lung condition categorization, utilizing the combination of both Vision Transformers (ViT) and Convolutional Neural Networks

(CNNs). The CNN component focuses on extracting spatial features, while the ViT model employs its self-attention mechanism to collect far-off dependencies within the image. The ViT classification head is pre-trained on ImageNet and fine-tuned on the lung condition dataset to predict three classes: normal, malignant, and benign. The hybrid model of the architecture captures the global context and fine-grained image data for improvement in classification performance.

3.3 Weighted Loss Function

The weighted loss function helps to address the class imbalance problem. The weighted cross-entropy loss function is defined as shown in equation (1)

$$L(w) = -\sum_{i=1}^{3} w_i y_i \log p_i \qquad (1)$$

w_i refers weight assigned to each class i,
y_i refers to the true label, and
p_i refers predicted probability of class i

3.4 Model Training

The ViT model uses a transfer learning process during which the initial layers are frozen at the beginning of training for ten epochs, ensuring that previously learned features are preserved. The network gradually unfreezes over epochs, adapting to patterns unique to the disease. Flips, rotation, and changes in brightness are a few types of data augmentation operations that enhance the diversity of the dataset. Encouraging convergence is given by the Adam optimizer and the cosine annealing scheduler, which are balanced out by weighted cross-entropy loss for class imbalance. Early pausing reduces the overfitting risk and saves the best model for evaluation.

4 Results and Discussions

The nuanced and intricate patterns found in medical imaging, predominantly in the early stages of the disease, make lung cancer diagnosis a difficult undertaking. The proposed research work is a sophisticated hybrid framework that combines the recompences of convolutional neural networks (CNNs) and vision transformers (ViTs) to categorize lung disorders into benign, malignant, and normal categories with exceptional accuracy. By fusing CNNs' established spatial feature extraction skills with ViTs' capacity to garner worldwide attention, the model provides a reliable and efficient way to address the complexities of lung cancer diagnosis. developed Vision transformer model. For proper validation of the model's performance, 80% of the data is utilized for training, and 20 % is reserved for testing. Weighted loss functions are used to handle class imbalance problems that are prevalent in medical datasets. The strategy improves the generalization of the model and its classifications by reducing the bias against the majority classes. A complete framework for data augmentation is used, including the techniques of flipping, affine transformation, and random rotation, as a means to simulate a variety of imaging conditions. These are stronger means of enhancing the model's robustness over variations in real-world data.

Pre-trained ImageNet weights are cast-off to initialize the Vision Transformer model, which uses transfer learning to adjust the network to medical imaging data. Stepwise training is used, where frozen layers are first used and then progressively unfrozen to enable deep feature fine-tuning. When paired with sophisticated optimization methods like the Adam optimizer and Cosine Annealing Learning Rate Scheduling, this gradual unfreezing guarantees effective training and guards against overfitting. In addition to optimization, dropout regularization is employed to enhance generalization, especially in the classification head. Fig. 3 represents the loss and accuracy variation during train- ing. The curve indicates the training is complete with a rate of convergence of around 32 epochs.

Fig. 3. Loss and accuracy variation during training

The model provides the classification accuracy for normal, malignant, and benign classifications at 94.14%. Metrics that prove the fair performance of this model in terms of recall, F1-score and precision. Fig. 4 depicts the confusion matrix, showcasing minimal misclassifications, while Grad-CAM visualizations enhance interpretability, validating the model's high accuracy and trustworthiness in AI-assisted lung dis- ease diagnosis.

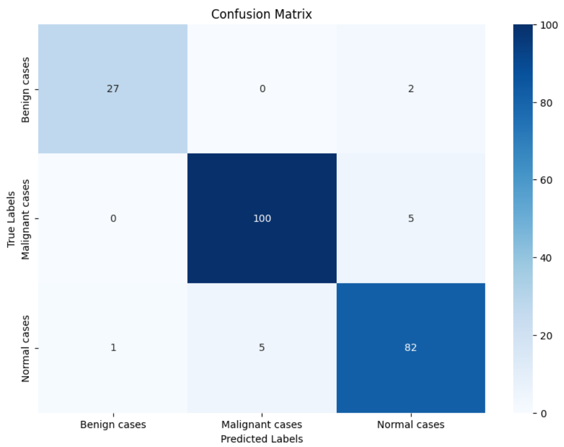

Fig. 4. Confusion Matrix of the classification process using Vision transformer

The combination of Vision Transformers (ViT) with Convolutional Neural Networks (CNN) demonstrates superior performance in lung cancer detection as proved by the relative analysis of accuracy, recall, F1-score, and precision. The ViT-CNN hybrid model achieved notable scores (Refer Fig. 5) across all metrics significantly outperforming the separate CNN and EfficientNet deep learning models. The hybrid approach in this research work influences the spatial feature mining capabilities of CNNs and the global framework understanding of ViTs. This comprehensive analysis of CT images is performed by robust architecture discourses class imbalance effectively and progresses generalization, constructing as reliable solution for early and precise lung cancer classification hypothetically transforming diagnostic practices.

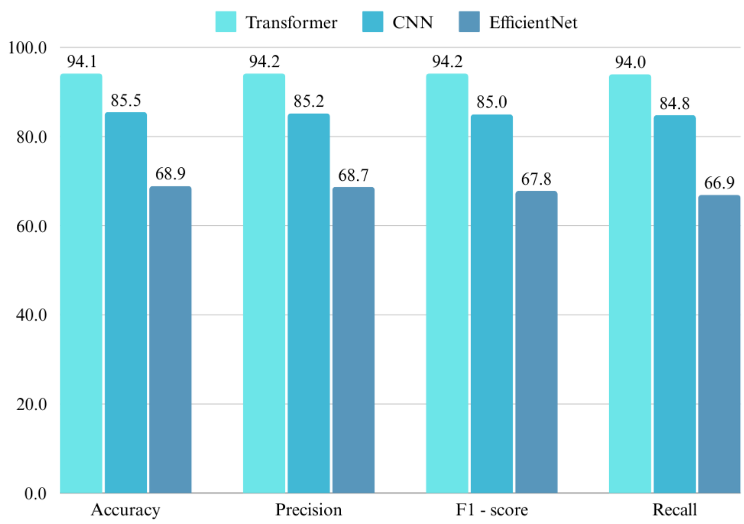

Fig. 5. Performance Analysis of hybrid vision transformer model and CNN compared with other deep learning models

These measures highlight the framework's superiority over current models in classifying lung conditions. One of its finest features is the interpretability, achieved by incorporating Gradient-weighted Class Activation Mapping (Grad-CAM). It generates the heatmaps and underscores the areas of the input image that have a stronger influence on the model's decision-making process. Therefore, aside from the visually disclosing reasoning mechanisms, it also provides assurance in the accuracy of AI-supported diagnostics,which is more palatable to healthcare professionals. This model is placed as a benchmark in AI-driven medical imaging because of the hybrid of ViTs and CNNs, with advanced training methods and interpretability tools.

5 Conclusion

Cancer challenge early diagnosis has been effectively dealt with by the hybrid model for classifying lung cancer syndicating Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs). With the help of CNNs to detect spatial patterns and ViTs to use global information, the model classifies lung diseases into normal, malignant, or benign. Classification performance has been significantly improved with the use of advanced methods like weighted loss functions, data augmentation, and transfer learning, which addresses the issues of detection of subtle anomalies and class imbalance. Further, Grad-CAM, providing visual insight into the decision-making process, enhances model interpretability and provides transparency and confidence in AI-assisted diagnosis. The model produced an exceptional accuracy rate of 94.14%, outpacing other conventional techniques and showcasing its potency and practicability as a scalable lung cancer identifier. In healthcare, the research indicates how the diagnostic accuracy will be enhanced by the AI models, promising so much to enable early interventions and better patient outcomes.

References

1. Thamizhselvi, D., Saranya, A., & Dhanusha, R. (2024). Lung cancer detection using a transfer learning algorithm. In 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT) (pp. 1-5). IEEE, Chennai. https://doi.org/10.1109/IC3IoT60841.2024.10550399
2. Zhang, Y., Dai, B., Dong, M., Chen, H., Zhou, M.: A lung cancer detection and recognition method combining convolutional neural network and morphological features. In: 2022 IEEE 5th International Conference on Computer and Communication Engineering Technology (CCET), pp. 145-149. IEEE, Beijing (2022). https://doi.org/10.1109/CCET55412.2022.9906329.
3. Abbas, A., Abdelsamea, M. M., Gaber, M. M.: DeTrac: Transfer learning of class decomposed medical images in convolutional neural networks. IEEE Access 8, 74901-74913 (2020). https://doi.org/10.1109/ACCESS.2020.2989273.
4. Jawaid, A., Hafeez, J., Khan, S., Ur Rehman, A.: Lung cancer detection using artificial neural network on Android. In: 2023 Global Conference on Wireless and Optical Technologies (GCWOT), pp. 1-8. IEEE, Malaga (2023). https://doi.org/10.1109/GCWOT57803.2023.10064658.
5. Ruprah, T. S., Regmi, B., Jadhav, S. B., Singh, S.: Early stage lung cancer detection using deep learning. In: 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), pp. 1-6. IEEE, Pune (2024). https://doi.org/10.1109/MIT- ADTSoCiCon60330.2024.10575345.
6. Thamizhselvi, D., Kasi, B., Kamalakkannan, K., Bharath, S., Gowtham, S., Kishore, M. A.: Air quality prediction using Adaboost. In: 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), pp. 1-6. IEEE, Chennai (2023). https://doi.org/10.1109/ICCEBS58601.2023.10448934.
7. Metagar, S. M., Sayyed, F. R.: Lung cancer detection and classification using machine learning algorithm (CNN). In: 2024 Second International Conference on Advances in Information Technology (ICAIT), pp. 1-5. IEEE, Chikkamagaluru (2024). https://doi.org/10.1109/ICAIT61638.2024.10690646.
8. Gupta, D., Dawn, S.: Detection and staging of lung cancer from CT scan images by deep learning. In: 2023 International Conference on Disruptive Technologies (ICDT), pp. 274-

278. IEEE, Greater Noida (2023). https://doi.org/10.1109/ICDT57929.2023.10151194.
9. Murthy Nimmagadda, S., Likhitha, K., Srilatha, G., Sree, S. M.: Lung cancer prediction and classification using machine learning algorithms. In: 2024 International Conference on Expert Clouds and Applications (ICOECA), pp. 1012-1015. IEEE, Bengaluru (2024). https://doi.org/10.1109/ICOECA62351.2024.00176.
10. Al-Shouka, T. T., Alheeti, K. M. A.: A transfer learning for intelligent prediction of lung cancer detection. In: 2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT), pp. 54-59. IEEE, Al-Muthana (2023). https://doi.org/10.1109/AICCIT57614.2023.10217967.
11. Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Jiang, Z., Hou, Q., Feng, J.: Deepvit: Towards deeper vision transformer. arXiv preprint arXiv:2103.11886 (2021) https://doi.org/10.48550/arXiv.2103.11886
12. Travis WD. Pathology of lung cancer. Clin Chest Med. 2011 Dec;32(4):669-92. doi: 10.1016/j.ccm.2011.08.005. PMID: 22054879.
13. Kadir T, Gleeson F. Lung cancer prediction using machine learning and advanced imaging techniques. Transl Lung Cancer Res. 2018 Jun;7(3):304-312. doi: 10.21037/tlcr.2018.05.15. PMID: 30050768; PMCID: PMC6037965.
14. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, Shete S, Etzel CJ. A risk model for prediction of lung cancer. J Natl Cancer Inst. 2007 May 2;99(9):715-26. doi: 10.1093/jnci/djk153. PMID: 17470739.
15. S. S. Raoof, M. A. Jabbar and S. A. Fathima, "Lung Cancer Prediction using Machine Learning: A Comprehensive Approach," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 108-115, doi: 10.1109/ICIMIA48430.2020.9074947.
16. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
17. Oudkerk, M., Liu, S., Heuvelmans, M. A., Walter, J. E., Field, J. K.: Lung cancer LDCT screening and mortality reduction-evidence, pitfalls and future perspectives. Nature Reviews Clinical Oncology 18, 135-151 (2021). https://doi.org/10.1038/s41571-020-00432-6.
18. Li, Y., Wu, X., Yang, P., Jiang, G., Luo, Y.: Machine learning for lung cancer diagnosis, treatment, and prognosis. Genomics, Proteomics & Bioinformatics 20(5), 850-866 (2022). https://doi.org/10.1016/j.gpb.2022.11.003.
19. Svoboda E. Artificial intelligence is improving the detection of lung cancer. Nature. 2020 Nov;587(7834):S20-S22. doi: 10.1038/d41586-020-03157-9. PMID: 33208974.
20. Thanoon, M. A., Zulkifley, M. A., Mohd Zainuri, M. A. A., Abdani, S. R.: A review of deep learning techniques for lung cancer screening and diagnosis based on CT images. Diagnostics 13(16), 2617 (2023). https://doi.org/10.3390/diagnostics13162617.