Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2024)
IEEE Xplore Part Number: CFP24OSV-ART; ISBN: 979-8-3503-7642-5

2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) | 979-8-3503-7642-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/I-SMAC61858.2024.10714870

LungCarcinoGrade-EffNetSVM: A Novel Approach to Lung Carcinoma Grading using EfficientNetB0 and Support Vector Machine

Pragati Patharia
Department of Electronics and Communication Engineering
Guru Ghasidas Vishwaviodyalaya, Bilaspur, C.G., India 495009
Email-pathariapragati@gmail.com

Prabira Kumar Sethy
Department of Electronics and Communication Engineering
Guru Ghasidas Vishwaviodyalaya, Bilaspur, C.G., India 495009
Email-prabirsethy.05@gmail.com

Abstract- Lung carcinoma grading is a critical task in the accurate diagnosis and treatment planning of lung cancer. In this study, we present "LungCarcinoGrade-EffNetSVM", a novel approach that combines the powerful feature extraction capabilities of EfficientNetB0 with the classification ability of Support Vector Machine (SVM) for lung carcinoma grading. The dataset utilized for this study was sourced from the Kaggle repository and includes images representing three types of lung carcinoma—Adenocarcinoma (ACA), Large Cell Carcinoma (LCC), Squamous Cell Carcinoma (SCC)—along with normal cell samples. Our proposed method achieved an Acc. of 86.88%, Sens. of 86.88%, and Spec. of 95.63%. The Prec. and F1 score were 87.06% and 86.64%, respectively, with a false positive rate (FPR) of 4.37%. The model also demonstrated robust performance with a Matthews correlation coefficient (MCC) of 0.8257 and a Kappa statistic of 0.65. The computational time for grading was recorded at 9.3082 seconds. These results indicate that the integration of EfficientNetB0 and SVM provides a reliable and efficient method for lung carcinoma grading, potentially aiding in more accurate and timely diagnosis of lung cancer.

Keywords- Lung carcinoma grading, EfficientNetB0, Support Vector Machine (SVM), Diagnosis, Lung cancer

1. Introduction

One of the most frequent and fatal types of cancer in the world is lung carcinoma, which is simply called lung cancer. Cancer of the lung develops when cells proliferate uncontrollably in the lung tissues [1]. Small cell lung carcinoma (SCLC) and non-small cell lung cancer (NSCLC) are the two primary subtypes of lung carcinoma [2]. "NSCLC accounts for approximately 85% of all lung cancer cases and includes subtypes such as adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Grading of lung carcinoma is a critical component in the clinical management of patients [3]".

Computed Tomography (CT) scans are instrumental in detecting and diagnosing lung tumors, including lung cancer, due to their ability to produce high-resolution, detailed cross-sectional images of the lungs [4]. These scans can identify small nodules or tumors that might be missed by standard X-rays, facilitating early detection when the disease is most treatable [5]. CT scans also provide critical information on the shape, size, and location of tumors, aiding in the differentiation between benign and malignant lesions [6]. They play a vital role in staging lung cancer by showing the extent of tumor spread and involvement of nearby lymph nodes and organs [7]. Additionally, CT scans guide needle biopsies for accurate tissue sampling, monitor treatment effectiveness by comparing images over time, and detect recurrence, all while being non-invasive and quick, thus enhancing the overall management and prognosis of lung cancer patients [8]. Many authors are reported of lung carcinoma grading using CT Scan images. In the article "Lung-EffNet: A Novel Approach for Lung Cancer Classification Using EfficientNet," Raza et al. (2023) presents Lung-EffNet, a model for classifying lung cancer that utilises transfer learning. The model is constructed by including the EfficientNet architecture and augmenting it with supplementary classification layers. The model is evaluated using the "IQ-OTH/NCCD" benchmark dataset, which classifies lung cancer patients as benign, malignant, or normal. The evaluation is performed using five different variations of EfficientNet (B0–B4). The study addresses the issue of class imbalance by employing data augmentation approaches. It achieves an impressive Acc. rate of 99.10% and a ROC score ranging from 0.97 to 0.99 on the test set. The findings indicate that Lung-EffNet, specifically the EfficientNetB1 version, surpasses other pre-trained CNN designs in terms of both Acc. and efficiency. Moreover, Lung-EffNet demonstrates superior speed and efficiency by using less parameters. This makes it highly suitable for widespread implementation in clinical settings and a very promising method for automating the detection of lung cancer using CT images [9]. "Tehnan et al. (2023) propose a hybrid approach combining a CNN with the Ebola optimization search algorithm (EOSA). The designed CNN architecture's solution vector is optimized using EOSA to find the best combination of weights and biases for classification. The hybrid EOSA-CNN model, tested on the IQOTH/NCCD lung cancer dataset, achieved a classification Acc. of 0.9321. Compared to other methods such as GA-CNN, LCBO-CNN, MVO-CNN, SBO-CNN, WOA-CNN, and a classical CNN, the EOSA-CNN demonstrated superior performance. Specifically, it achieved a Spec. of 0.7941 for normal cases, 0.97951 for benign cases, and 0.9328 for malignant cases, along with a Sens. of 0.9038 for normal, 0.13333 for benign, and 0.9071 for malignant cases. These results indicate that the hybrid EOSA-CNN algorithm provides an effective solution for the classification of lung cancer from CT images [10]". Lanjewar et al. (2024) proposed a novel DL (DL) method by modifying the DenseNet201 model, enhancing it with additional layers to improve lung cancer identification. "Two feature selection methods were employed to extract the best features from DenseNet201, which were then fed into various ML classifiers. The method executed on Kaggle CT Scan dataset and average Acc. of 95%, and a p-value of less than 0.001 [11]". Quasar et al. (2023) proposed an ensemble model that integrates various DL techniques for enhanced detection and classification from CT scan images. "The ensemble model combines BEiT, DenseNet, and Sequential CNN using methods like AND, OR, Weighted Box Fusion, and Boosting, tested on the Chest CT-Scan Images Dataset. The study reviews the effectiveness of ensemble methods, noting their ability to improve Acc. by leveraging the strengths of multiple classifiers to counteract individual weaknesses. Results show that this ensemble approach outperforms both single-model techniques and other ensemble methods, achieving a high prediction Acc. of 98% [12]". Mamun et al. (2023) proposed a DL-based CNN framework for early detection of lung cancer using CT scan images. "The study also compares the proposed CNN model with other DL models, including Inception V3, Xception, and ResNet-50. Performance metrics such as Acc., Area Under Curve (AUC), recall, and loss were used for comparison. The results demonstrate that the CNN model outperforms the other models, achieving an Acc. of 92%, an AUC of 98.21%, a recall of 91.72%, and a loss of 0.328, indicating its potential superiority over traditional methods for lung cancer detection [13]". Dunn et al. (2023) explore the application of artificial intelligence and data science techniques to interpret medical image scans for lung cancer diagnosis, highlighting the limitations of traditional radiologist interpretations which can be time-consuming and subjective. "It focuses on radiomics, a field combining medical imaging with personalized medicine, to develop automated diagnostic tools. The incremental multiple resolution residual network (iMRRN), a DL segmentation model, was used to automatically segment CT images from 355 lung cancer patients in the "LungPET-CT-Dx" dataset from The Cancer Imaging Archive (TCIA). The iMRRN had a failure rate of 4.35% in segmenting tumor lesions. Seven classification algorithms were then trained on radiomic features extracted from these images to classify different lung cancer subtypes. Over-sampling addressed data imbalance, and chi-square tests identified higher order texture features as the most predictive. The SVM achieved the highest Acc. of 92.7% (0.97 AUC) in classifying adenocarcinoma, small cell carcinoma, and squamous cell carcinoma. The study demonstrates the potential of AI-based diagnostic tools to automatically classify lung cancer subtypes by combining DL segmentation with supervised classification, and discusses practical issues in applying AI in biomedicine [14]". Riquelme et al. (2020) discusses the challenges in detecting malignant lung nodules from CT scans and the role of computer-aided diagnosis (CAD) systems in alleviating this burden for radiologists. "It emphasizes the shift towards DL approaches, which have demonstrated superior performance compared to classical methods in various applications. The study categorizes recent state-of-the-art DL algorithms and architectures proposed for CAD systems into two main types: (1) Nodule detection systems, which identify candidate nodules from CT scans, and (2) False positive reduction systems, which classify these candidates into benign or malignant tumors. Key characteristics of each technique are outlined, including their performance metrics, and the availability of CT lung datasets for research is highlighted. The review concludes with a comparative analysis of the different DL approaches, discussing their strengths and limitations in improving lung cancer screening and diagnosis using CT scans [15]". Shafi et al. (2022) proposed a DL-enabled SVM model for the diagnosis of early-stage lung cancer using CT scans. "The model aims to overcome the challenges of detecting asymptomatic lung cancer nodules, which are complex and prone to errors even for specialists. Trained on CT images from the LIDC/IDRI database, the model identifies physiological and pathological changes in lung cancer lesions by comparing profile values. It achieves a high Acc. of 94% in detecting pulmonary nodules indicative of early-stage lung cancer. Comparative analysis shows its superiority over other methods, including complex DL, simple ML, and hybrid techniques [16]". AlYasriy et al. (2020) introduced a computer-aided system for detecting lung cancer using a CNN with AlexNet architecture. The study utilizes a dataset from Iraqi hospitals to classify patient cases into normal, benign, or malignant categories. The proposed model achieves a high Acc. of up to 93.548%, with additional strong performance metrics including 95.714% Sens. and 95% Spec. [17]. Kareem et al. (2021) proposed for lung cancer detection using image-processing and computer-vision techniques. "The process involves three main stages: image enhancement, image segmentation, and feature extraction. Subsequently, SVM classifiers are employed with various kernels to classify the images into three categories: normal, benign, or malignant. Here, IQ-OTH/NCCD lung cancer dataset was used, comprising 1100 CT-scan images which include both healthy and tumorous chest scans Through rigorous evaluation, the system achieves a maximum Acc. of 89.8876% on the dataset, highlighting its effectiveness in automated lung cancer diagnosis using CT scans [18]". Besides CT scan, Histopathological images are also used for Lung cancer diagnosis [19,20].

2. Material and Methodology

2.1. About Dataset

The dataset is collected from Kaggle repository. The dataset contains 3 chest cancer types which are ACA, LCC, SCC, and normal. The details of dataset and samples of images are illustrated in Table 1 and Figure 1.

Table 1. Details of Dataset

Types	No. of Images
ACC	338
LCC	200
SCC	260
Normal	202

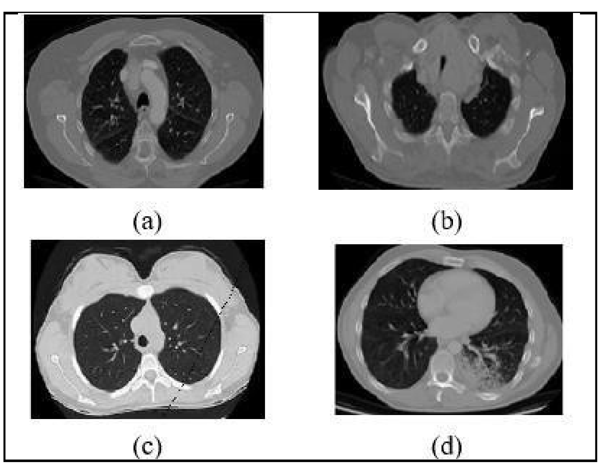

Figure 1. Lung Cancer CT Scan images (a) ACA (b) LCC (c) SCC (d) Normal

2.2. Methodology

In this study, we propose a methodology termed "LungCarcinoGrade-EffNetSVM" for the comprehensive grading of lung carcinoma using CT scan images. The approach begins with EfficientNetB0, a leading CNN renowned for its efficient yet powerful feature extraction capabilities. Leveraging pre-trained weights, EfficientNetB0 extracts discriminative features from the input CT scan images, setting the stage for subsequent classification.

These extracted features are then input into a SVM configured with a one-vs-all mapping strategy. This SVM implementation is enhanced with the ErrorCorrecting Output Codes (ECOC) to optimize classification Acc. across the four carcinoma categories: ACA, LCC, SCC, and normal samples.

The model undergoes fine-tuning with specific hyperparameter settings: a mini-batch size of 32, an initial learning rate set at 0.0001, and optimization via Stochastic Gradient Descent with Momentum (SGDM). This ensures the model is optimized for performance in handling the complex task of multi-class carcinoma classification.

Evaluation of "LungCarcinoGrade-EffNetSVM" employs a range of metrics including Acc., Sens., Spec., Prec., F1 score, FPR, MCC, and Kappa statistic. These metrics collectively assess the model's efficacy in accurately grading lung carcinoma, crucial for precise diagnosis and treatment planning.

Furthermore, computational efficiency is measured, with each image graded within 9.3082 seconds, highlighting the methodology's practicality for real-time application in clinical settings. By integrating advanced DL techniques for feature extraction and SVM for classification, "LungCarcinoGrade-EffNetSVM" offers a robust framework for enhancing the Acc. and efficiency of lung carcinoma grading, potentially improving patient outcomes in clinical practice. The architecture of "LungCarcinoGrade-EffNetSVM" is illustrated in Figure 2.

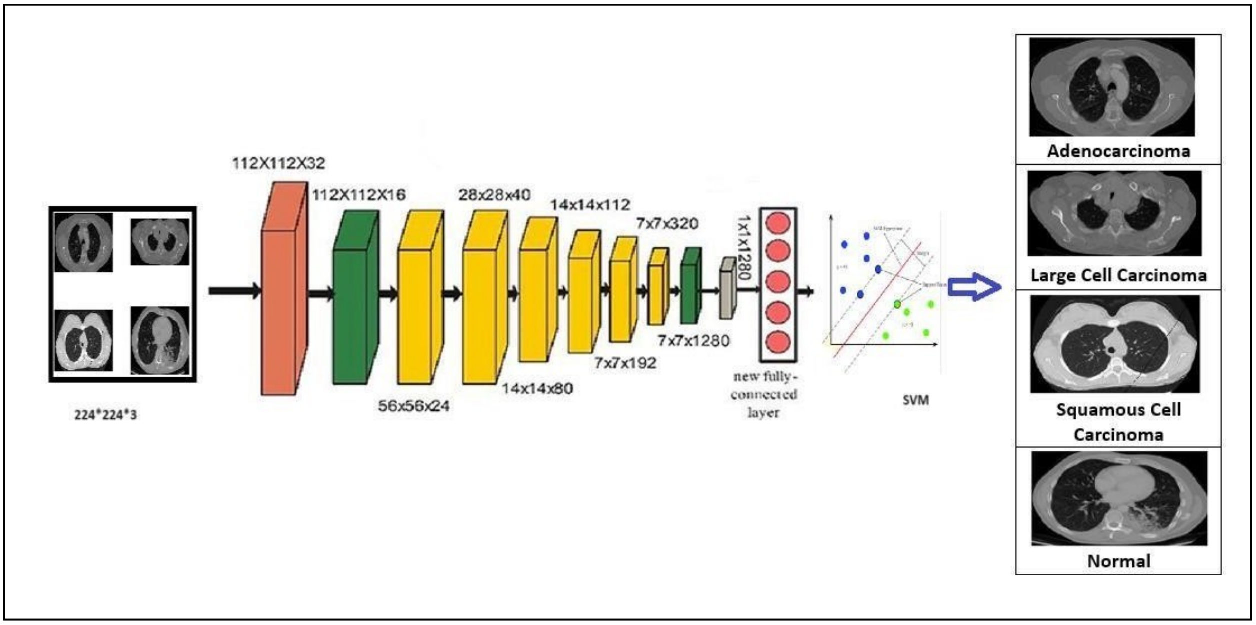

Figure 2. LungCarcinoGrade-EffNetSVM model for lung carcinoma grading.

3. Result and Discussion

Implemented on a standard HP laptop with an Intel Core i7 (11th generation) processor, 8GB of RAM, and integrated NVIDIA graphics using MATLAB 2022a, the "LungCarcinoGrade-EffNetSVM" methodology achieved good results in lung carcinoma grading from CT scan images. The methodology exhibited an Acc. of 86.88%, Sens. of 86.88%, Spec. of 95.63%, Prec. of 87.06%, F1 score of 86.64%, and demonstrated a low FPR of 4.37%. With a MCC of 0.8257 and a Kappa statistic of 0.6500, the model's robust performance highlights its capability to effectively classify ACA, LCC, SCC, and normal samples. Computational efficiency was notable, with each image processed in an average time of 9.3082 seconds, highlighting its potential for real-time clinical application. This approach combines EfficientNetB0 for feature extraction and SVM for classification, offering a reliable framework to enhance lung carcinoma diagnosis and treatment planning, thereby contributing to improved patient care and outcomes in oncology practice. The confusion matrix is illustrated in Figure 3.

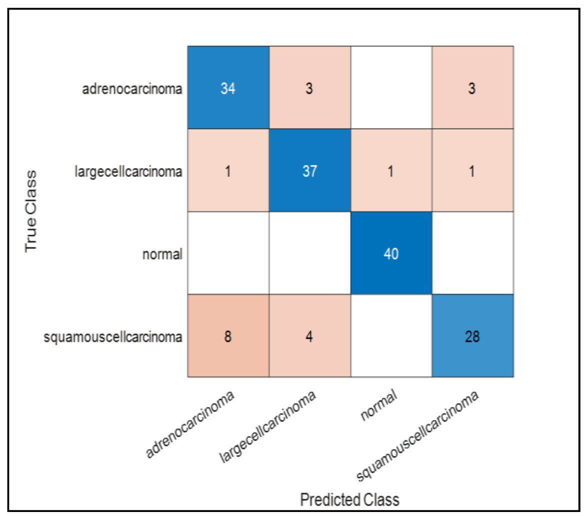

Figure 3. Confusion matrix of LungCarcinoGradeEffNetSVM model.

Further, comparative analysis is carried out with the existing work and tabulated in Table 2.

Table 2. Comparative analysis with state-of-art

Author and Ref.	Methodology and Model Used	Dataset Used	Number of Classes/Types of Images	Number of Images	Metrics Reported	Key Findings and Remarks
Raza et al. (2023) [9]	Lung-EffNet model, EfficientNet B0–B4 versions with data augmentation	IQ-OTH/NCCD	Benign, malignant, normal	Not mentioned	Acc. 99.10%, ROC 0.97–0.99	EfficientNetB1 outperformed other models, achieving superior Acc. and efficiency with fewer parameters, making it suitable for clinical implementation.
Tehnan et al. (2023) [10]	Hybrid EOSA-CNN model, EOSA optimized CNN weights	IQ-OTH/NCCD	Benign, malignant, normal	Not mentioned	Acc. 93.21%, Spec. (normal: 0.7941, benign: 0.9795, malignant: 0.9328), Sens. (normal: 0.9038, benign: 0.1333, malignant: 0.9071)	EOSA-CNN model outperformed other CNN variants (GA-CNN, LCBO-CNN, etc.), though benign sensitivity was relatively low, indicating possible imbalance.
Lanjewar et al. (2024) [11]	Modified DenseNet201 model with feature selection methods and various ML classifiers	Kaggle CT Scan dataset	Not mentioned	Not mentioned	Acc. 95%, p-value < 0.001	DenseNet201 was enhanced with additional layers, improving feature extraction and classification results, leading to high statistical significance in performance.
Quasar et al. (2023) [12]	Ensemble model integrating BEiT, DenseNet, and Sequential CNN using multiple fusion methods	Chest CT-Scan Images Dataset	Not mentioned	Not mentioned	Acc. 98%	The ensemble approach outperformed single models by leveraging multiple classifiers, improving accuracy through methods like Weighted Box Fusion and Boosting.
Mamun et al. (2023) [13]	DL-based CNN framework, compared against Inception V3, Xception, and ResNet-50	Not mentioned	Not mentioned	Not mentioned	Acc. 92%, AUC 98.21%, recall 91.72%, loss 0.328	The CNN model outperformed other DL models, showing potential superiority for early lung cancer detection using CT scan images.
Dunn et al. (2023) [14]	iMRRN for segmentation, SVM classification on radiomic features	LungPET-CT-Dx (TCIA)	Adenocarcinoma, small cell carcinoma, squamous cell carcinoma	355	Acc. 92.7%, AUC 0.97	iMRRN effectively segmented CT images, and SVM achieved the highest Acc. in classifying lung cancer subtypes using radiomic features and chi-square feature selection.
Riquelme et al. (2020) [15]	Review of DL methods in CAD systems for lung nodule detection and false positive reduction	Various CT lung datasets	Benign, malignant	Not mentioned	Acc. and performance metrics varied across reviewed models	DL approaches in CAD systems demonstrate superior performance for nodule detection and classification, significantly aiding radiologists in improving lung cancer screening and diagnosis.
Shafi et al. (2022) [16]	DL-enabled SVM model for early-stage lung cancer detection	LIDC/IDRI database	Pulmonary nodules	Not mentioned	Acc. 94%	The DL-SVM model proved effective for detecting early-stage lung cancer, surpassing traditional methods in accuracy and reliability.
AlYasriy et al. (2020) [17]	CNN using AlexNet architecture for lung cancer detection	Dataset from Iraqi hospitals	Normal, benign, malignant	Not mentioned	Acc. 93.55%, Sens. 95.71%, Spec. 95%	The AlexNet-based CNN showed strong performance, with high Acc. and sensitivity in distinguishing between normal, benign, and malignant cases.
Kareem et al. (2021) [18]	SVM classifiers with various kernels for lung cancer detection	IQ-OTH/NCCD	Normal, benign, malignant	1100	Acc. 89.89%	The image-processing and SVM approach was effective for lung cancer detection using CT scans, though Acc. was lower than that of more recent models.

From Table 2, it is evident that only two studies, namely Lanjewar et al. (2024) and Quasar et al. (2023), focused on four-way classification of carcinoma detection using CT scan images with a dataset size of 1000 images, both employing augmentation techniques to enrich the dataset. These studies utilized DenseNet201 for their classification tasks. In contrast, our proposed method utilizes EfficientNetB0, which offers distinct advantages in CNN architectures due to its efficient parameter utilization and balanced scaling of depth, width, and resolution dimensions. This efficiency not only optimizes computational resources but also enhances adaptability to diverse hardware constraints, making it suitable for deployment on edge devices and mobile platforms. Augmentation techniques, while beneficial in some contexts, can alter the appearance of medical images, potentially introducing artifacts that might mislead the model during training and compromise diagnostic Acc.. Moreover, improper application of augmentation may skew the data distribution within the training set, artificially inflating variability and biasing the model towards recognizing augmented patterns rather than genuine medical conditions. Therefore, our approach avoids augmentation, aiming to develop a more robust method for lung carcinoma grading in future iterations.

4. Conclusion

Our study presents “LungCarcinoGrade-EffNetSVM”, a novel approach for lung carcinoma grading using EfficientNetB0 and SVM, achieving an Acc. of 86.88% without employing augmentation techniques. EfficientNetB0's superior parameter efficiency and scalability optimize computational resources and enhance adaptability across diverse hardware platforms.By avoiding augmentation, we mitigate risks associated with altering medical image integrity and biases in data distribution, thereby preserving diagnostic Acc.. Comparison with existing studies highlighted in Table 2 underscores the potential advantages of our approach in achieving reliable and efficient lung carcinoma grading. Moving forward, future iterations will focus on refining our methodology to further enhance diagnostic Prec. and expand applicability in clinical settings, aiming to contribute significantly to advancements in lung cancerdiagnosis and treatment planning.

References

[1] Cruz, C. S. D., Tanoue, L. T., & Matthay, R. A. (2011). Lung cancer: epidemiology, etiology, and prevention. Clinics in chest medicine, 32(4), 605-644.

[2] Araujo, L. H., Horn, L., Merritt, R. E., Shilo, K., Xu-Welliver, M., & Carbone, D. P. (2020). Cancer of the lung: Non–small cell lung cancer and small cell lung cancer. In Abeloff's clinical oncology (pp. 1108-1158).

[3] Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., ... & Rosell, R. (2015). Non-small-cell lung cancer. Nature reviews Disease primers, 1(1), 1-16.

[4] Hoffman, E. A. (2022). Origins of and lessons from quantitative functional X-ray computed tomography of the lung. The British Journal of Radiology, 95(1132), 20211364.

[5] Del Ciello, A., Franchi, P., Contegiacomo, A., Cicchetti, G., Bonomo, L., & Larici, A. R. (2017). Missed lung cancer: when, where, and why?. Diagnostic and interventional radiology, 23(2), 118.

[6] Millet, I., Doyon, F. C., Hoa, D., Thuret, R., Merigeaud, S., Serre, I., & Taourel, P. (2011). Characterization of small solid renal lesions: can benign and malignant tumors be differentiated with CT?. American journal of roentgenology, 197(4), 887-896.

[7] Woodard, G. A., Jones, K. D., & Jablons, D. M. (2016). Lung cancer staging and prognosis. Lung cancer: treatment and research, 47-75.

[8] Heck, S. L., Blom, P., & Berstad, A. (2006). Accuracy and complications in computed tomography fluoroscopy-guided needle biopsies of lung masses. European radiology, 16, 13871392.

[9] Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence, 126, 106902.

[10] Mohamed, T. I., Oyelade, O. N., & Ezugwu, A. E. (2023). Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm. Plos one, 18(8), e0285796.

[11] Lanjewar, M. G., Panchbhai, K. G., & Charanarur, P. (2023). Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers. Expert Systems with Applications, 224, 119961

[12] Quasar, S.R., Sharma, R., Mittal, A. et al. Ensemble methods for computed tomography scan images to improve lung cancer detection and classification. Multimed Tools Appl 83, 52867–52897 (2024). https://doi.org/10.1007/s11042-023-17616-8

[13] Mamun, M., Mahmud, M. I., Meherin, M., & Abdelgawad, A. (2023, March). Lcdctcnn: Lung cancer diagnosis of ct scan images using cnn based model. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 205-212).

[14] Dunn, B., Pierobon, M., & Wei, Q. (2023). Automated classification of lung cancer subtypes using deep learning and CT-scan based radiomic analysis. Bioengineering, 10(6), 690.

[15] Riquelme, D., & Akhloufi, M. A. (2020). Deep learning for lung cancer nodules detection and classification in CT scans. Ai, 1(1), 28-67.

[16] Shafi, I., Din, S., Khan, A., Díez, I. D. L. T., Casanova, R. D. J. P., Pifarre, K. T., & Ashraf, I. (2022). An effective method for lung cancer diagnosis from ct scan using deep learning- based support vector network. Cancers, 14(21), 5457.

[17] Al-Yasriy, H. F., Al-Husieny, M. S., Mohsen, F. Y., Khalil, E. A., & Hassan, Z. S. (2020). Diagnosis of lung cancer based on CT scans using CNN. In IOP conference series: materials science and engineering (Vol. 928, No. 2, p. 022035).

[18] Kareem, H. F., AL-Husieny, M. S., Mohsen, F. Y., Khalil, E. A., & Hassan, Z. S. (2021). Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indonesian Journal of Electrical Engineering and Computer Science, 21(3), 1731.

[19] Zhang, Y., Yang, Z., Chen, R., Zhu, Y., Liu, L., Dong, J., ... & Zhou, M. (2024). Histopathology images-based deep learning prediction of prognosis and therapeutic response in small cell lung cancer. NPJ digital medicine, 7(1), 15.

[20] Sethy, P. K., Geetha Devi, A., Padhan, B., Behera, S. K., Sreedhar, S., & Das, K. (2023). Lung cancer histopathological image classification using wavelets and AlexNet. Journal of XRay Science and Technology, 31(1), 211-221.