2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)
02-04 May 2024, Military Institute of Science and Technology (MIST), Dhaka-1216, Bangladesh

2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) | 979-8-3503-8577-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICEEICT62016.2024.10534491

# Interpreting Multiclass Lung Cancer from CT Scans using Grad-CAM on Lightweight CNN Layers

Fahim Muntasir
*Department of Computer Science and Engineering*
*BRAC University*
Dhaka, Bangladesh
muntasirfahim.niloy@gmail.com

Ayon Datta
*Department of Electrical and Electronic Engineering*
*Ahsanullah University of Science and Technology*
Dhaka, Bangladesh
ayandutt17@ieee.org

Shakib Mahmud
*Department of Biomedical Engineering*
*Bangladesh University of Engineering and Technology*
Dhaka, Bangladesh
1718023@bme.buet.ac.bd

***Abstract*—Lung cancer is one of the most prevalent and deadly cancer kinds. CT scan imaging has made it easier to observe in-depth at different lung cancer conditions. Convolutional Neural Networks have already been utilized for detecting objects in images. This research implements a lightweight CNN architecture in classifying two types of cancerous images and distinguishing them from healthy ones. AlexNet, ResNet101 V2, DenseNet201, MobileNet V2, and other sizable deep learning and transfer learning models were tested against this unique architecture with global max pooling. The proposed CNN model overlaps in performance with much bigger models whilst using about eight times fewer parameters, thus being efficient. Apart from making the architecture smaller, Grad-CAM was used to explain the convolutional flow over different convolution layers. This integration of the Grad-CAM technique into the analysis adds a crucial layer of transparency to the deep-learning models. By visualizing the regions within lung cancer images that significantly influence predictions, we aim to enhance the interpretability of our model, fostering trust and better understanding among healthcare professionals.**

***Index Terms*—Grad-CAM, Lightweight CNN, CT Scan Imaging, Lung Cancer, Transfer Learning**

## I. Introduction

Lung cancer is still a major worldwide health concern that necessitates cutting-edge methods for precise and early detection. In 2020, lung cancer accounted for more cancer-related fatalities than breast, colorectal, and prostate cancer combined. Lung cancer can be seen in both males and females. It also stands second in cancer-related deaths [1]. This burden falls disproportionately on those from lower socioeconomic backgrounds. Although smoking cigarettes can be directly associated with lung cancer, a huge portion of non-smokers have also been found to grow this cancer. [2]. The primary aetiological factor in lung cancer is tobacco use. Together with tobacco use, other variables including poor diet, genetic predisposition, occupational exposures, and air pollution may influence the descriptive epidemiology of lung cancer [3]. Medical professionals have identified different genres of lung cancers. The two primary categories are non-small cell and small cell lung cancers. [4]. Adenocarcinoma falls under NSCLC and is predominant among non-smokers and females [5]. It was also found to be the most prevalent lung cancer in the USA [6]. Lung tumours with large cell carcinoma represent for 5% to 10% of all malignancies. Large, peripheral, necrotic masses that infect surrounding tissues are the most prevalent manifestations of aggressive tumours [7].

A Computed Tomography (CT) scan can identify the cancer level or tumours in human body [8]. With the recent advancements of deep learning and neural network methods, the underlying pattern can be identified from this CT scan image. However, medical professionals still lack the confidence in trusting the results from these approaches. The main reason is the technical confusion between computer methodologies and medical profession. Only numerical results aren't sufficient when dealing with life-threatening disease such as lung cancer. This study harnesses the power of deep learning to classify lung cancer from CT scan images. The focus extends beyond achieving high accuracy, delving into the interpretability of the models through the integration of Grad-CAM, a technique shedding light on the decision-making processes of convolutional neural networks (CNNs). The significance of this project lies not only in its pursuit of heightened accuracy in lung cancer classification, but also in its commitment to making the decision-making process of AI models more accessible and transparent. As we navigate the intersection of deep learning and medical imaging, our objective is to contribute to the ongoing evolution of artificial intelligence in healthcare, with a particular focus on improving the diagnostic landscape for lung cancer. Through this innovative approach, we endeavour to pave the way for more informed and timely interventions, ultimately impacting patient outcomes and advancing the frontier of medical diagnostics.

## II. Literature Review

Early detection of lung cancer from CT scan images is an actively explored topic in biomedical image analysis. Various methods have been explored in recent years, and from the literature review, it is suggested that a Convolutional Neural Network (CNN) can effectively be employed for lung cancer classification. Shafi et al. used LIDC/IDRI lung cancer database images. The images are annotated and publicly available. Their approach incorporated Support Vector Machines (SVM) to assist the deep learning model, resulting in an impressive 94% accuracy [9]. In 2019, Causey et al. attempted early detection of lung cancer using a CNN on 1449 CT scan images. Their algorithm employed a series of slices and multitasking features to predict the likelihood of a tumor being cancerous. Their approach also contained a spatial hierarchy to detect tumours at various scales. This work achieved an area under curve of 0.858 and an accuracy of 78.2% [10]. Emphasizing image quality and feature extraction, Shakeel et al. obtained 98% accuracy on CT images collected from the cancer image archive dataset [11]. Sori et al. took into account the noise in the lung cancer images. They used a CNN architecture after denoising the images, resulting in an accuracy of 89% [12]. In recent years, gradient-weighted class activation mapping (Grad-CAM) has been used to provide visual explanations for decisions from CNN-based models. Imawanto et al. utilized Xception and ResNet50V2 transfer models for predicting lung cancer [13]. Grad-CAM was then applied to display the specific areas of focus within the CT scan. Ise et al. achieved an accuracy of 70%, subsequently using Grad-CAM to visualize critical areas [14]. Joshua et al. proved in their study that lightweight models can have higher acceptance by incorporating visualization techniques such as Grad-CAM [15].

From the literature review, it can be concluded that using Grad-CAM can enhance the interpretability of the model's decision-making process, no matter how lightweight or complex the model's processing power is. This, in turn, increases trust among radiologists and promotes adoption in the field.

## III. Methodology

### A. Dataset

The dataset includes CT scan images of two types of lung cancer [16]. There are a total of 505 training images and 225 testing images. The three types of cells were: Adenocarcinoma, Large cell carcinoma and Normal.

### B. Data Preprocessing

The images were of different shapes. For this specific task, the width of the images was set at 200px by 200px. The raw images were converted into training and testing datasets in 16 images per batch. All the sets were rescaled with a factor of 1/255 to normalize the images.

### C. Deep Learning Approach

Deep learning is proven to be better at image classification than traditional machine learning [17]. CNN or Convolutional Neural Networks are unmatched in performance and reliability when it comes to data in array-like format. An image is an array, consisting of three 2D arrays of pixels [18]. That's why

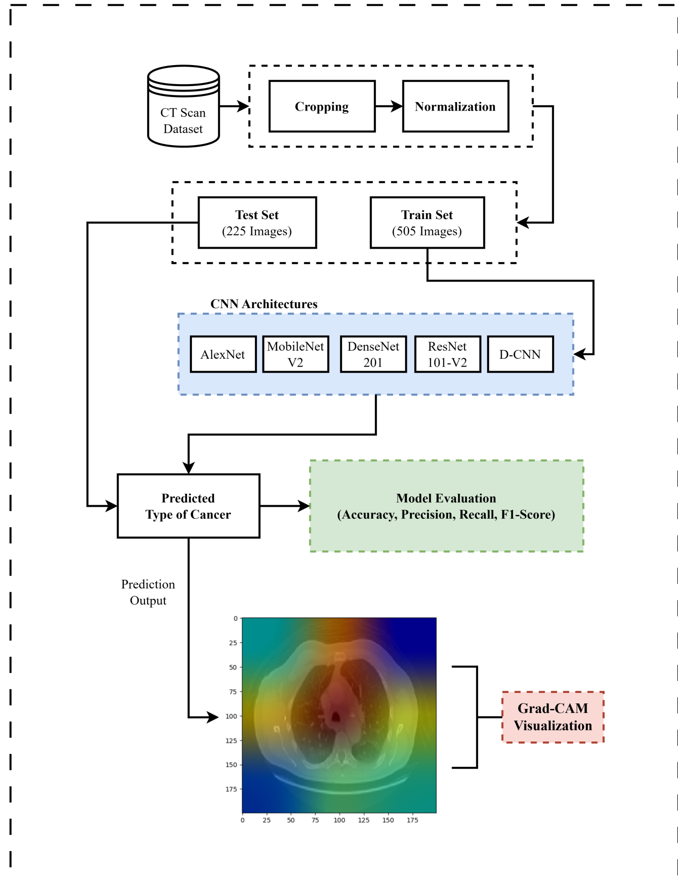


Fig. 1. Proposed methodology

a custom Deep CNN architecture was designed specifically for this classification of lung cancer cells. Moreover, transfer learning has made it easy to use prebuilt larger models for any new classification problems [19]. This work has opted to use the following transfer learning architectures.

*1) AlexNet:* Though AlexNet is not a transfer learning model, it was the first of its kind to use convolution with relu activation function [20]. It outperformed all the other models in the Image net competition by a huge margin. The architecture have five convolutional layers, max-pooling layers, and three fully-connected layers. In this task, it was implemented with the original architecture with a slight change at the last output layer, 3 classes instead of 1000 of the ImageNet dataset.

*2) ResNet101 V2:* ResNet101V2 stands for residual network. It has high-performance multiple residual layers, using batch normalization before each weight layer, unlike the original (V1) network [21]. ResNet offers a unique solution to the vanishing gradient problem by stacking multiple identity mappings, skipping those layers, and reusing the activation function from the previous layer.

*3) MobileNet V2:* Sandler et al. proposed MobileNetV2, a model that integrates inverted residuals and linear bottlenecks into the MobileNetV1 model [22]. This method addresses the limitation of depth-wise convolution on the number of channels, allowing for feature extraction from input channels. Inverted residual blocks, consisting of 1×1 increasing, 3×3 depth-wise, and 1×1 decreasing filters, facilitate learning, suppress gradient distribution, and improve model stability.

*4) DenseNet201:* Dense Convolutional Network (DenseNet) is a deep, accurate, and efficient training method that connects layers in a feed-forward style, reducing the number of parameters and improving feature propagation [23]. The network's deeper, accurate, and efficient training can be achieved with shorter connections between input and output layers. The 201 indicates this is 201 layers deep.

### *D. Grad-CAM*

Gradient-weighted Class Activation Mapping, also known as Grad-CAM, is used to generate visual explanations for decisions from CNN-based models, enhancing their transparent nature [24]. Using the gradients of each target concept streaming into the final convolutional layer, this technique creates a coarse map of localization that emphasizes the key regions of a picture for idea prediction. Grad-CAM can demonstrate how a picture can be complex or grasped by a filter. This visualization is typically shown as a heatmap, which helps the algorithm to determine which portion of the image is most important.

## IV. Proposed CNN Architecture

The Deep CNN architecture consists of several parts. The architecture is described below

1) Input Layer: The input shape is defined as (w, h, 3), where w and h are the width and height of the input pictures, and 3 is the number of colour channels (RGB).
2) First convolutional layer: It consists of 64 filters with kernel sizes of (5, 5) and strides of 2, followed by a ReLU activation function. Following that, there is the first max pooling layer, with a pool size of (2, 2) and a stride of 2.
3) Second convolutional layer: It consists of 256 filters with kernel sizes (3, 3) and strides of 2, followed by a ReLU activation function. The second max pooling layer has a pool size of (3,3) and a stride of 2.
4) Third convolutional layer: This has 128 filters, a kernel size of (3,3), and the ReLU activation function. The batch normalization feature has been included to provide smooth and stable convergence throughout training.
5) Batch Normalization Layer: Between the third and fourth convolution layers, a batch normalization layer was used. The purpose of such a layer was to fasten the training process with higher learning rate. Moreover, this makes the model more stable by recentring and rescaling inputs.
6) Fourth convolutional layer: It has 512 filters, a kernel size of (3, 3) and the ReLU activation function. The last max pooling layer is introduced, with a pool size of (3, 3).

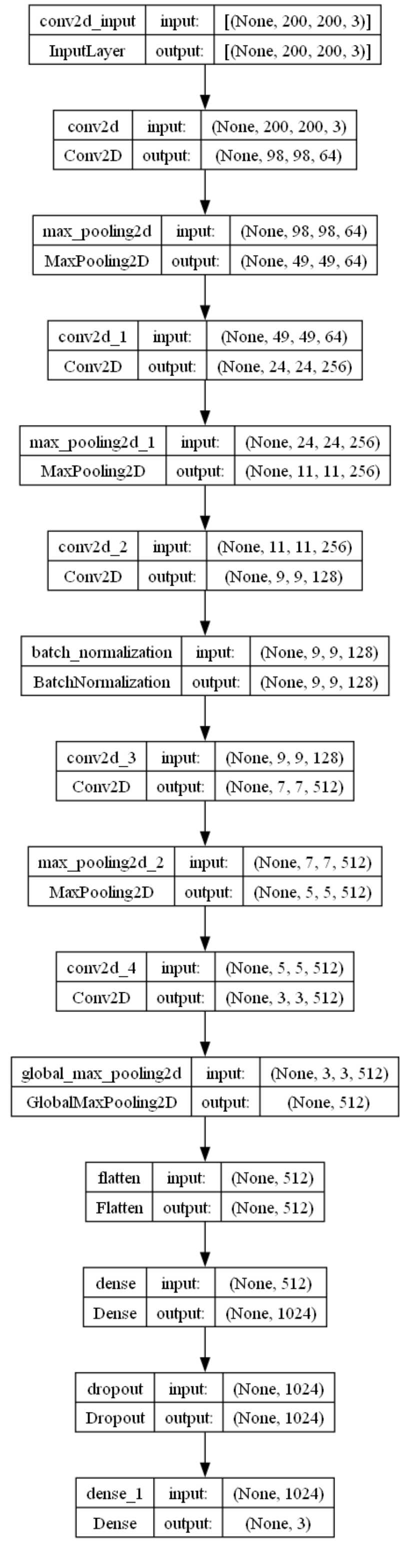


Fig. 2. Proposed CNN Architecture

7) Final convolutional layer: The fifth and final convolution layer uses 512 filters, a kernel dimension of (3, 3), and the same ReLU activation function.
8) Global Max Pooling layer: This enables the networks to focus on the overall presence or absence of features and sharpen those features throughout the images rather than specific locations.
9) Flatten Layer: This layer flattens the output from the convolutional layers into a 1D array.
10) Dense Layers: The dense (fully connected) layer is added with 1024 neurons, ReLU activation functions, and dropout layers with a dropout rate of 0.5 to prevent overfitting.
11) Output Layer: The second and final dense layer, specifically the output layer, has three neurons (as it's a multi-class classification problem) with a soft max activation function which is commonly suitable for such problems as the weight of three neurons will be obtained, and the maximum weighted neuron will be predicted according to its class later.
12) Compilation: The model uses the Adam optimizer with a learning rate of 3e-4, categorical cross entropy as the loss function, and accuracy as the evaluation metric.

This resulted in a much lower parameter compared to all the larger transfer learning models.

## V. Performance Analysis

The performance of different architectures was measured using accuracy, precision, recall, and F1 score. A comparison between the parameter count was also viewed to see if similar performance could be achieved with a smaller, lightweight model. All the models were tested with an exact setup of, batch size: 16, image size: 200x200x3 and 30 epochs. The training and validation accuracy-loss curves show that the D-CNN model had a stable learning curve. Both training and testing accuracies increased in each epoch, with training reaching 100% accuracy. The losses decreased with epochs.

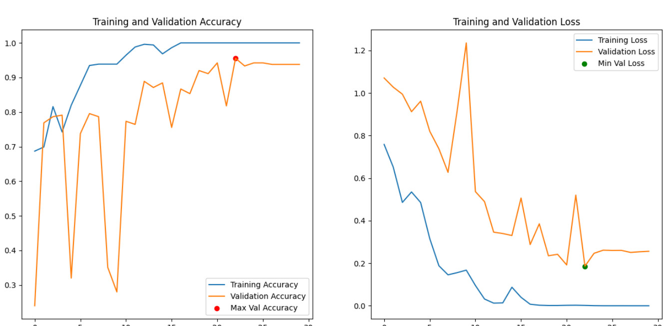


Fig. 3. Training and validation curves

The following are some key observations from the performance comparison table:

1) All the architecture performed closely, with the proposed customized Deep CNN (D-CNN) outperforming others.

TABLE I
Performance comparison of All Models

| Architecture | Acc. % | Pre. % | Rec. % | F1 % | Parameter (Millions) |
|---|---|---|---|---|---|
| MobileNet V2 | 88 | 88 | 89 | 88 | 3.17 |
| ResNet101 V2 | 86 | 86 | 85 | 86 | 44.73 |
| DenseNet201 | 88 | 87 | 90 | 88 | 19.57 |
| AlexNet | 95 | 94 | 95 | 95 | 24.74 |
| **D-CNN** | **96** | **96** | **96** | **96** | **3.93** |

2) Apart from the proposed D-CNN architecture, the AlexNet architecture was superior to the other larger architectures in terms of accuracy.
3) Taking almost eight times fewer parameters than AlexNet (24.74 million for AlexNet, 3.93 million for D-CNN), D-CNN showed the same accuracy as the AlexNet model showed.
4) The quality and amount of the images greatly impact the models' performance.

The confusion matrix is a more clear representation of the predicted outcomes. As the numbers show, our model correctly identifies 115 out of 120 adenocarcinoma cases and 47 out of 51 large cell carcinoma. Interestingly, no cancerous image was detected as a normal or healthy image. This shows the quality of the predictions, and thus the D-CNN less involved normal tissues.

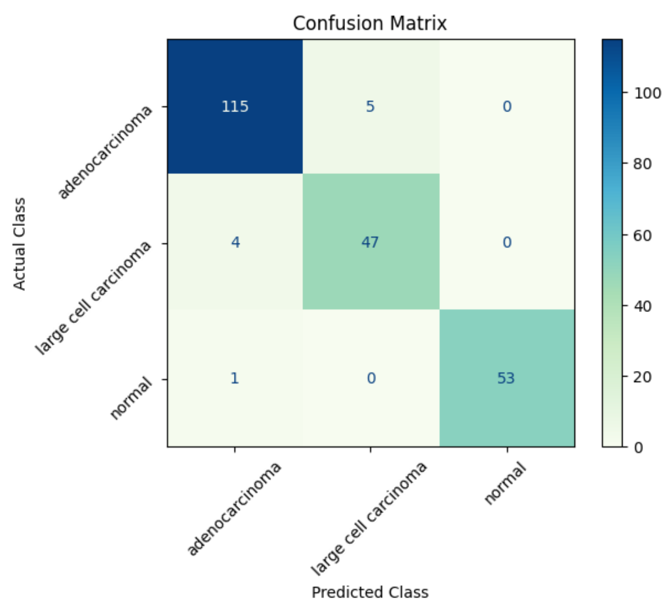


Fig. 4. Confusion matrix of the test set

In comparison with research works from the literature reviews and existing approaches, it is evident that our lightweight CNN approach is very efficient. It even surpassed the performance of a much larger vision transformer (ViT) model. The available works mostly focused on the binary classifications. Multiclass lung cancer detection has just been approached in recent years. It is also visible that, a model accurate enough to be reliable alongside explained results was missing so far. This work successfully established the connection.

TABLE II
COMPARISON WITH PREVIOUS WORKS

| Reference | Approach | Accuracy | Remarks |
|---|---|---|---|
| [10] 2019 | Spatial pyramid in neural network | 78.2% | Binary classification of cancer nodule |
| [12] 2021 | Denoising images and classification with CNN | 89% | Image preprocessing failed to attain higher accuracy. |
| [14] 2021 | Grad CAM with 3D CNN | 70% | Mediocre accuracy tends to mispredict, thus explanations are error-prone. |
| [9] 2022 | SVM assisted neural network | 94% | Binary classification with no explanation |
| [25] 2024 | Vision Transformer with Grad CAM | 84.8% | ViT model used here has more than 100Million parameters making it extremely computationally complex |
| This Work [2024] | Grad CAM with Lightweight D-CNN | 96% | Classification of two different cancer types along with differentiating normal images. Grad-CAM explained the predicted outcome giving a clear understanding. |

### A. Grad-CAM Analysis

Gaining trust and spotting any problem often depend on knowing where the model is gazing in a picture. Here, Grad-CAM allows us to visually confirm the locations of our network's vision, ensuring that it is focused on the right patterns in the image and activating around it. Moreover, it enhances the interpretability of overall CNNs, making them more transparent and aiding in model debugging. This method takes the gradients from the last convolution layer and doesn't need any particular training. When superimposed over the main image as a heatmap, it can provide informative details about the decision-making process of CNN models.

According to the heatmap of Fig. 5 and Fig. 6, the red-marked areas are the most important zones for classifying the image into certain categories. On the other hand, the blue areas are the least significant. To be specific, the change is remarkable in the color intensity of analyzed images of both Large Cell Carcinoma and Adenocarcinoma. Because, the convolutional layers have been changed from layer 1 to 5, where layer 1 is just the beginning phase of Grad-CAM analysis, and from layer 2 to layer 5, it gradually turned out to be the particular predictions of the labeled classes using gradient information. Among these, for both cases, layer 4 shows us the most convincing result of our required visualization.

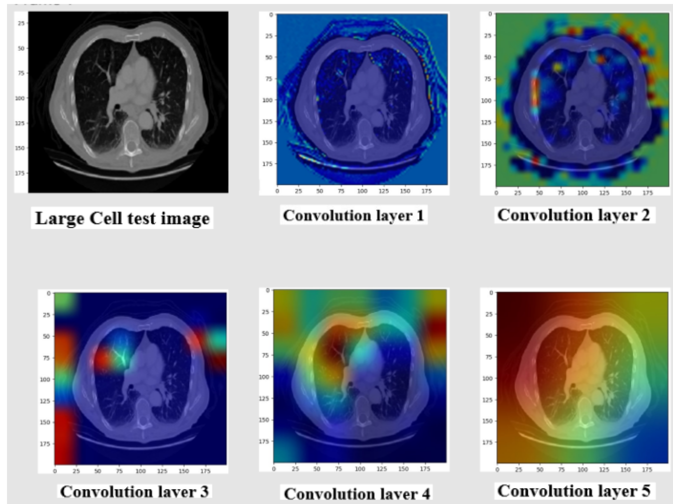


Fig. 5. Grad-CAM Visualization of Large Cell Carcinoma

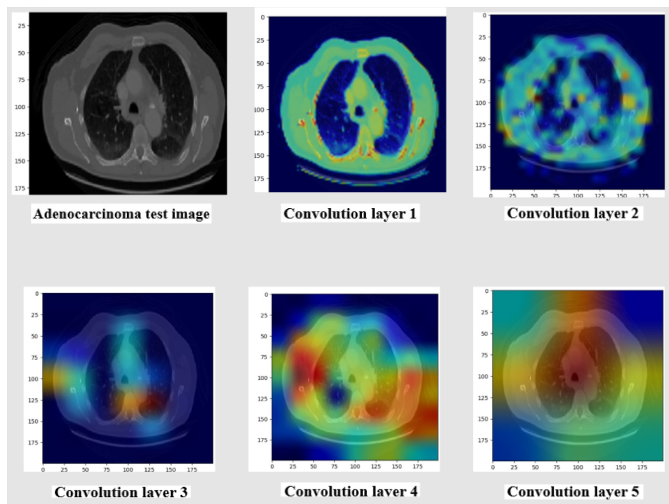


Fig. 6. Grad-CAM Visualization of Adenocarcinoma

## VI. CONCLUSION

This study has leveraged deep learning techniques, including AlexNet, ResNet101 V2, DenseNet201, MobileNet V2, and a lightweight CNN architecture, for the classification of lung cancer images successfully. From the findings presented in Table 1, it is evident that existing convolutional neural network (CNN) architectures have demonstrated commendable accuracies, ranging from 86% (ResNet101 V2) to 95% (AlexNet). Notably, MobileNetV2 and DenseNet201 have exhibited comparable accuracies of 88%. However, it is noteworthy that these models are often characterized by substantial parameter counts. The primary objective of the proposed D-CNN architecture is to enhance optimization efficiency over significantly larger transfer learning models. Initial modifications made to the architecture preceding the dense layer have yielded promising outcomes. As illustrated in Table 1, the parameter count serves as an indicator of model complexity. Among the models evaluated, MobileNetV2 possesses the closest parameter count to that of D-CNN. Nevertheless, D-CNN surpasses MobileNetV2 in terms of accuracy. Conversely, AlexNet, while demonstrating performance proximity to our proposed model, entails more than eight times the parameter count of D-CNN. In addition to accuracy and parameter count considerations, the interpretability of model decisions holds considerable importance. The incorporation of the Grad-CAM technique for interpretability purposes has significantly augmented the project's significance. By offering transparent insights into the decision-making processes of the models, this interpretability framework not only enhances confidence in model predictions but also bolsters the clinical relevance of the project, thus fostering trust in its applicability within medical contexts. The success of accurate lung cancer image classification holds potential benefits for healthcare professionals, enabling early identification of cancer cases and facilitating timely interventions. In conclusion, this project represents a notable stride in the application of deep learning to medical imaging, offering accurate classification coupled with model interpretability. The combination of advanced deep learning models and visualization tools contributes to the ongoing evolution of artificial intelligence in healthcare, promising improved diagnostics reports. In further study, an approach for more sophisticated and accurate mapping of cancer cells can be made. Also, multimodal neural networks can be used for combining images with information from medical personnel for better feature extraction.


## Acknowledgement

The authors confirm that ChatGPT or any other LLMs weren't used for any intellectual purposes in the preparation of the research work.


## References


[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[2] F. Islami, A. Goding Sauer, K. D. Miller, R. L. Siegel, S. A. Fedewa, E. J. Jacobs, M. L. McCullough, A. V. Patel, J. Ma, I. Soerjomataram *et al.*, "Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the united states," *CA: a cancer journal for clinicians*, vol. 68, no. 1, pp. 31–54, 2018.

[3] J. Malhotra, M. Malvezzi, E. Negri, C. La Vecchia, and P. Boffetta, "Risk factors for lung cancer worldwide," *European Respiratory Journal*, vol. 48, no. 3, pp. 889–902, 2016.

[4] H. A. Wakelee, E. T. Chang, S. L. Gomez, T. H. Keegan, D. Feskanich, C. A. Clarke, L. Holmberg, L. C. Yong, L. N. Kolonel, M. K. Gould *et al.*, "Lung cancer incidence in never-smokers," *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, vol. 25, no. 5, p. 472, 2007.

[5] H. Nakamura and H. Saji, "A worldwide trend of increasing primary adenocarcinoma of the lung," *Surgery today*, vol. 44, pp. 1004–1012, 2014.

[6] D. J. Myers and J. M. Wallen, "Lung adenocarcinoma," in *StatPearls [Internet]*. StatPearls Publishing, 2023.

[7] C. I. S. Silva and N. L. Müller, "Online case 24," in *The Teaching Files: Chest*, ser. Teaching Files in Radiology, C. I. S. Silva and N. L. Müller, Eds. Philadelphia: W.B. Saunders, 2010, pp. e48–e49.

[8] S. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on ct images," *Future Generation Computer Systems*, vol. 92, pp. 374–382, 2019.

[9] I. Shafi, S. Din, A. Khan, I. D. L. T. Díez, R. d. J. P. Casanova, K. T. Pifarre, and I. Ashraf, "An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network," *Cancers*, vol. 14, no. 21, p. 5457, 2022.

[10] J. L. Causey, Y. Guan, W. Dong, K. Walker, J. A. Qualls, F. Prior, and X. Huang, "Lung cancer screening with low-dose ct scans using a deep learning approach," *arXiv preprint arXiv:1906.00240*, 2019.

[11] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Lung cancer detection from ct image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, 2019.

[12] W. J. Sori, J. Feng, A. W. Godana, S. Liu, and D. J. Gelmecha, "Dfd-net: lung cancer detection from denoised ct scan image using deep learning," *Frontiers of Computer Science*, vol. 15, pp. 1–13, 2021.

[13] R. E. Imawanto, G. Rifqialdi, A. Yudith, and A. Sinaga, "Computer-aided detection of lung cancer from ct-scan images with visual insights using deep convolutional neural network," 2021.

[14] E. Ise, S. Kido, and Y. Hirano, "Classification of lung cancer into adenocarcinoma and squamous cell carcinoma, and visualization of the grounds of classification," in *International Forum on Medical Imaging in Asia 2021*, vol. 11792. SPIE, 2021, pp. 52–61.

[15] E. S. Neal Joshua, D. Bhattacharyya, M. Chakkravarthy, and Y.-C. Byun, "3d cnn with visual insights for early detection of lung cancer using gradient-weighted class activation," *Journal of Healthcare Engineering*, vol. 2021, p. 6695518, Mar 2021. [Online]. Available: https://doi.org/10.1155/2021/6695518

[16] M. Hany, "Chest ct-scan images dataset," Aug 2020. [Online]. Available: https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images/data

[17] J. Chai, H. Zeng, A. Li, and E. W. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, p. 100134, 2021.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[19] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and building materials*, vol. 157, pp. 322–330, 2017.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[21] R. Mar-Cupido, V. García, G. Rivera, and J. S. Sánchez, "Deep transfer learning for the recognition of types of face masks as a core measure to prevent the transmission of covid-19," *Applied Soft Computing*, vol. 125, p. 109207, 2022.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[24] 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," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[25] L. Weng, Y. Xu, Y. Chen, C. Chen, Q. Qian, J. Pan, and H. Su, "Using vision transformer for high robustness and generalization in predicting egfr mutation status in lung adenocarcinoma," *Clinical and Translational Oncology*, pp. 1–8, 2024.