ISSN: 2577 - 8005

Research Article

Medical & Clinical Research

A Novel Approach to Diagnosis: Using Convolutional Neural Networks to Classify Non-Small Cell Lung Cancers on CT Scans

Siya N Setty[1*], Niyathi Srinivasan[1], Juliette Clausen[1] and Russell J Thompson[2]

[1]*Lexington High School, Lexington 02421, Massachusetts, USA*

[2]*Department of Computer Science, Worcester Polytechnic Institute, Worcester, Massachusetts, USA*

[*]Corresponding Author
Siya N Setty, Lexington High School, Lexington 02421, Massachusetts, USA.

Submitted: 17 Nov 2024; **Accepted**: 25 Nov 2024; **Published**: 04 Dec 2024

Citation: Setty, S.N., Srinivasan, N., Clausen, J., Thompson, R.J. (2024). A Novel Approach to Diagnosis: Using Convolutional Neural Networks to Classify Non-Small Cell Lung Cancers on CT Scans. *Med Clin Res, 9*(12), 01-06.

Abstract

Background: *Lung cancer is the leading cause of cancer-related deaths around the world, making early detection vital to treatment and survival. However, lung cancer has variable clinical presentations, so diagnosis, even by trained medical professionals, is challenging and time- consuming. Deep Learning (DL) has proven effective in training big data to generate accurate diagnoses through classification in a timely manner.*

Aim and Objective: *This study uses data augmentation and hyperparameter optimization methods on convolutional neural networks (CNN) to understand the benefits of various architectures in addition to creating an accurate transfer-based tool for lung cancer diagnosis on CT scans.*

Materials and Methods: *We found a dataset composed of chest CT images of patients with non-small cell lung cancers (NSCLC) categorized into adenocarcinoma, large cell carcinoma, and squamous cell carcinoma and a control group of normal chest CT scans. We tested the CNN architectures VGG16, InceptionV3, ResNet50, and EfficientNetB0 for feature extraction and observed each model on a variety of different metrics such as validation accuracy, statistical errors, and learning rate. Each model was trained on the pre-chosen data set 2 times at 6 and 12 epochs and several evaluation metrics were recorded.*

Observation and Results: *EfficientNetB0, ResNet50, VGG16, InceptionV3 achieved test accuracies of 98.92%, 91.40%, 77.42%, 70.97% respectively when implemented with hyperparameter tuning and dropout layers. EfficientNetB0 also exceeded the other architectures in metrics such as precision, recall, and F1-score.*

Conclusion: *Convolutional neural networks using EfficientNetB0 yielded a higher accuracy in diagnosis of non-small cell cancer on CT scan which encourages further research in developing robust diagnostic tools using DL to expedite diagnosis, especially in locations with inadequate healthcare resources.*

Keywords: Lung cancer, CT scan, Deep learning, Convolutional neural networks, Transfer learning

1. Introduction

Lung cancer is the leading cause of cancer deaths worldwide [1]. Furthermore, cancer diagnosis and mortality rates are increasing due to the high levels of introduced carcinogens in smoking devices, impure water sources, as well as unfiltered exposure to chemicals such as asbestos in work environments [2]. Non-small cell lung cancer accounts for 80% of lung cancer cases [3]. The most common subtypes of NSCLC are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [3]. Each subtype is defined by the difference in tumor characteristics on CT and on histopathology. Adenocarcinoma presents as a solid mass or nodule in the lung tissue which can vary in size and shape and may have irregular or spiculated borders. Further, adenocarcinoma can show signs of lymph node involvement, such as enlarged lymph nodes near the lungs, and areas of consolidation or ground-glass opacities in the lungs. Large cell carcinoma includes the presence of a large, solid mass or nodule in the lung tissue, which appears heterogeneous with irregular borders. Finally, squamous cell carcinoma lung cancer typically appears as a centrally located mass or nodule in the lung tissue which has a round or oval shape with well-defined borders. Squamous cell carcinoma can also manifest as areas of cavitation within the tumor, which can be seen as hollow spaces or air-filled pockets within the mass [4].

Early detection of cancer is key for appropriate management and better patient survival. Most lung cancers are discovered at an advanced stage, and the fact that only 15% of lung cancers are discovered in early stages is largely responsible for poor prognosis [4]. However, due to its variable clinical presentations, diagnosis of lung cancer can be difficult, even by trained medical professionals. Generating accurate diagnoses is important for selection of various treatment options.

Over the years, the implementation of computer-aided diagnosis models (CADs) has become an invaluable cost effective help to expedite diagnoses [5]. In our study, we experimented with image classification of lung cancer using convolutional neural networks (CNNs) and transfer learning. The primary function of classification models is to predict the label of a given input of data after being trained with similar images from a labeled dataset (training dataset). Classifying models will self-evaluate their performance on unsupervised machine learning tasks with hyperparameters such as loss functions, activation functions, logarithmic loss functions, learning rate, and validation accuracy. The model's goal is to learn from its sources of error on the training data and use metrics to backpropagate and adjust the importance (i.e. weight) of each pixel on its final guess [6].

Transfer learning models are universally trained, pre-existing architectures that can be imported onto a dataset and trained onto specific data. Each of these models have unique benefits and drawbacks depending on the data. EfficientNet is a high-quality CNN architecture that targets saturation in models that are made with too high resolution and too many parameters to promote efficiency. This design uses compound scaling to highlight important regions without compromising input image quality. ResNet50 is a traditional neural network that notably uses residual connections to skip over redundant layers and max pooling to cut down computational costs. InceptionV3 uses an efficient grid size reduction and an additional factorization in the third iteration to reduce parameters and minimize costs. Lastly, VGG16 limits hyperparameters and increases convolutional layers. CNNs place convolutional filters on an image to detect important regions and adjust weights accordingly to classify images accurately. These models have proven to be efficient in processing big data but have yet to be assessed for their potential clinical applications. Nearly 49.9% of lung cancer is seen in patient's from underdeveloped countries [3]. Given the many disparities in the healthcare field in under- developed areas, stemming from a lack of diagnosticians and medical resources, machine learning and deep learning models can be used to increase efficiency and recommend treatment options [3].

In this study, we explore the power of previously validated architectures, hyperparameter optimization techniques, and dropout layers on our model in a holistic format to combat these skyrocketing fatalities. We hypothesize that if we use hyperparameter optimization paired with the EfficientNetB0 architecture to classify types of lung cancer on computed tomography (CT) scans, then we can create a model with improved accuracy compared to other tried and true CNN architectures.

2. Materials and Methods

2.1 Data Collection

We used a pre-existing dataset to train our model, which was made publicly available [7]. This data was composed of chest CT scans of the following lung cancer subtypes: adenocarcinoma, large cell carcinoma, normal chest CT scans (without tumor), and squamous cell carcinoma, indexed as 0, 1, 2, 3 respectively. The 1000 images were separated by class into training, validation, and testing datasets as follows: 70% training (700), 20% validation (200), and 10% testing (100). The dataset was compiled from national databases at random to remove bias and was cited in multiple peer-reviewed journal articles.

2.2 Data Augmentation

Data augmentation, a useful tool in classification to create generalizable models, was utilized to introduce various orientations and motions into the images. Due to the high risk of bias and overfitting within our small dataset, we implemented the Keras Image Data Generator library and used it to augment the number of unique images the model was trained on. The data was preprocessed through the Image Data Generator function, which arbitrarily modified certain images with one or more transformations (rotation, width shift, height shift, zoom, shear (distortion), flip). Parameters included: rotation range = 20, width shift range = 0.2, height shift range = 0.2, shear range = 0.2, zoom range = 0.2. The model was then trained on these modified images, along with images called directly from the database, which diversified the images we could pass through to get an accurate diagnosis.

2.3 Hyperparameter Optimization

We resize each image to a standard size of 224 pixels by 224 pixels and performed hyperparameter tuning using the hyper band search algorithm, powered by Keras [6]. The dropout rate, the number of times the model can hide nodes and layers, was optimized to prevent over fitting and ensure thorough feature extraction.

2.4 CNN Architectures

We experimented with many transfer learning models, including ResNet50, InceptionNetV3, EfficentNetB0 and VGG16 optimizing neural networks around those base models. Each model was selected based on various researched benefits and drawbacks. After testing the various models to identify the optimal architecture, a final convolutional neural network was structured and implemented, with an EfficientNetB0 base model, global average pooling, batch normalization, dropout layers, and dense layers added in. A learning rate of 0.001 and the stochastic gradient descent (SGD) optimizer were also used as presets.

2.5 Gradient-weighted Class Activation Mapping

Gradient-weighted Class Activation Mapping, or Grad-CAM, is a revolutionary technique that helps the viewer understand which parts of a given image a neural network is focusing on

when making classification and weight-adjusting decisions. Specifically, when classifying lung cancer CT scans, Grad-CAM can show diagnosticians the specific areas in the scans that the model is looking at to identify cancerous tissue. Grad-CAM was implemented in our final neural network to analyze diagnosis and explain reasoning to diagnosticians.

2.6 Large Language Models

We verified and elaborated on our results by implementing large language models (LLMs) to explain our results in conjunction with Grad-CAM. Large language models refer to large AI platforms that are pre-trained on vast amounts of data (i.e. ChatGPT, Google Gemini, Microsoft Copilot) to generate text and speech-based responses to a series of prompts. We used Google Gemini, calling gen_model.generate_content() on the prompt and the Grad-CAM image, then printing it to the console using response.resolve () and response.text. We fed our model's predicted cancer classification and our grad-CAM heatmap into a large language model to interpret and output an explanation for the errors in our model and directions for diagnosticians viewing the image. We used clinical definitions and symptoms of the 3 cancers, as well as normal CT scans as inputs to our LLM to make sure our program did not incorporate unnecessary or unverifiable data.

3. Results

We tested the CNN architectures VGG16, InceptionV3, ResNet50, and EfficientNetB0 to compare their efficacy on our augmented dataset. The models were chosen based on their diverse set of strengths and weaknesses. Distinguishing the most efficient and accurate model is important to identify optimal architectures for future diagnostic tools. Each model was trained on the pre-chosen data set 2 times at 6 and 12 epochs and several evaluation metrics were recorded.

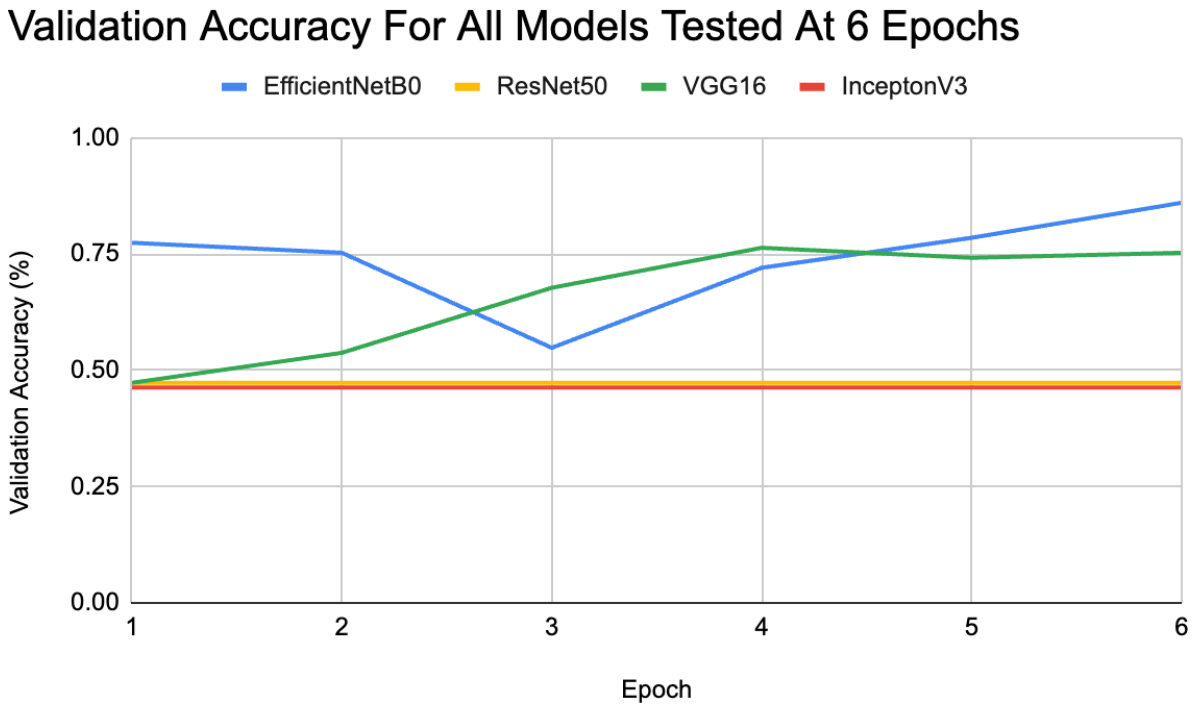

Figure 1: Validation accuracy for all models tested at 6 epochs. Line graph shows percentage of correct classifications for each model. Accuracy of 1 indicates no incorrect classifications. Accuracy of 0 indicates no correct classifications. Measured at intervals of 2 training cycles for a total of 6 training cycles. No patience was set.

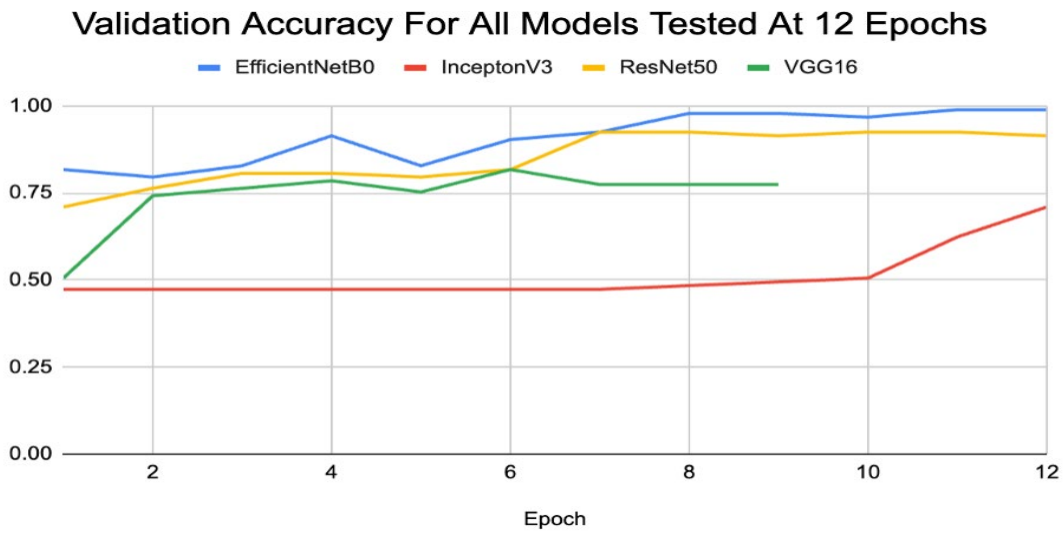

Figure 2: Validation accuracy for all models tested at 12 epochs. Line graph showing percentage of correct classifications for each model. Accuracy of 1 indicates no incorrect classifications. Accuracy of 0 indicates no correct classifications. Measured at intervals of 2 training cycles for a total of 12 training cycles.

We observed that the average accuracies of all models increased as the number of epochs increased (Figure 1 & 2). However, our EfficientNetB0 model achieved the highest validation accuracy (98.92%) when implemented with our hyperparameter tuner and trained at 12 epochs. The InceptionV3 had an initial accuracy of below 50%, which could mean that the model was just making lucky guesses. VGG16 accuracy was not recorded at all epochs because the patience expired after 2 stagnant epochs (Figure 2). Learning rate was optimized relative to each model, which explained the low patience on VGG16 but higher patience on InceptionV3. Traditional metrics such as precision (number of true positive images by total positive images for each class), recall (number of true positives divided by true positives plus false negatives for each class), and F1-score (the harmonic mean of precision and recall) were calculated using the computer-generated confusion matrices for each model (Table 1). We also generated a confusion matrix for our EfficientNetB0 (Figure 3) which shows the number of images in our test dataset that the model classified as false positive, false negative, true positive, and true negative for each class.

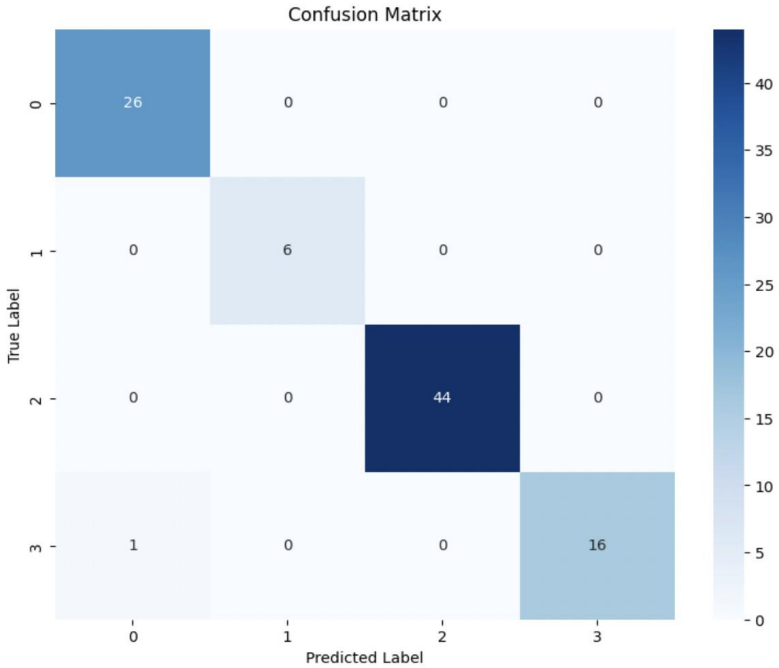

Figure 3: Confusion Matrix for the EfficientNetB0 model. Indexes of 0, 1, 2, 3 correspond with labels Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma, respectively. Sum of quantities in each column indicates how many scans the model categorized into each label (predicted label). Sum of quantities per row indicates how many scans belong to each label (true label).

Model	Class	Precision	Recall	F1-Score
VGG16	Adenocarcinoma	0.636	0.807	0.711
	Large Cell	undefined (no predictions)	0	undefined
	Normal	0.977	1	0.988
	Squamous Cell	0.733	0.647	0.687
InceptionV3	Adenocarcinoma	0.5	0.08	0.138
	Large Cell	0.235	0.666	0.347
	Normal	0.635	0.909	0.748
	Squamous Cell	0.888	0.471	0.616
ResNet50	Adenocarcinoma	0.857	0.923	0.889
	Large Cell	1	0.333	0.499
	Normal	1	1	1
	Squamous Cell	0.789	0.882	0.833
EfficientNetB0	Adenocarcinoma	0.963	1	0.982
	Large Cell	1	1	1
	Normal	1	1	1
	Squamous Cell	1	0.941	0.970

Table 1: Calculated Precision, Recall, and F1-Scores based on generated confusion matrices for each of the tested models (VGG16, InceptionV3, ResNet50, and EfficientNetB0) in each of the classes (Adenocarcinoma, Large Cell Carcinoma, Normal, and Squamous Cell Carcinoma). Score of 0 indicates all classifications were missed and score of 1 indicates no errors.

The EfficientNetB0 model was able to generate near-perfect results, with limited false positive and false negative rates for each class (Table 1). It only misclassified one case of squamous cell cancer as adenocarcinoma on CT (Figure 3). The specifics of this oversight are observable in our t-distributed stochastic neighbor embedding (t-SNE) feature plot (Figure 4), which is an advanced visualization technique designed to map high-dimensional data into a lower-dimensional space while preserving the structure and relationships within the data. The algorithm minimizes the divergence between these probability distributions using a cost function, ensuring that points that are close in the original high-dimensional space remain close in the low-dimensional projection. The squamous cell cancer was misinterpreted to be an adenocarcinoma likely because the tumor characteristics were near the edge of the image, causing the model to extract features that led to misdiagnosis.

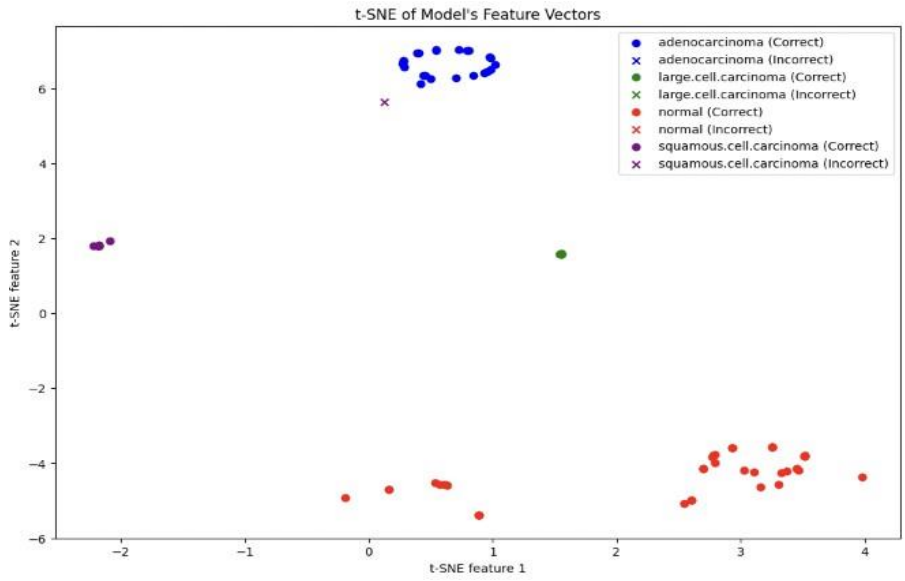

Figure 4: t-distributed stochastic neighbor embedding (t-SNE) plot for the EfficientNetB0 model. Model of pairwise similarities between data points in the high-dimensional mirrored in a lower- dimensional representation. Circles indicate correct classifications of scan; Xs indicate scans classified as wrong label.

4. Discussion

After experimentation, we determined that our most effective model architecture was EfficientNetB0, which was able to accurately classify 98.92% of images and was consistent with our hypothesis,. Other tested models fell short in multiple ways. Architectures such as ResNet50, which had a relatively high validation accuracy at all epochs, fell short in being able to accurately distinguish between various cancer subtypes. The max pooling in the ResNet50 model increased overfitting to the dataset, similar to InceptionV3. VGG16 was irrelevant to our data as it limited the hyperparameters we set.

Overall, our experimentation with these commonly used model architectures to design our model was successful in evaluating the constraints of each design in a practical application. Evaluating each model with various metrics clarifies each architecture's facets and flaws on our dataset. By incorporating hyperparameter optimization, we were able to effectively increase our model's accuracy and case-by-case efficacy by automizing feature extraction through the learning rate, loss function, activation function, optimization algorithm, and dropout rate [6].

We can implement such methods in designing more robust diagnostic tools to be used in clinical settings. Across all tested models, there was very little misclassification of normal scans as cancerous and cancerous as normal. Further, our best-performing model limited misclassification of different types of cancers. This differentiated our EfficientNetB0 model from many other pre-existing classification tools in the literature.

The addition of data augmentation and transformation was analogous to diverse CT imaging techniques. Technicians have various methods of photographing with the CT machine, putting patients in different orientations, capturing blurry images, etc. By randomizing images, our model was able to account for this variability.

Using data visualization methods such as Grad-CAM and t-SNE plots to validate our model's generalizability further refined our final EfficientNetB0 model and was especially useful for interpreting the model's predictions. Moreover, use of t-SNE plots was useful to reduce the high-dimensional data of our image to 2 dimensions to spot patterns and clusters, helping viewers visualize how data is organized at a high level, which determined the efficacy of our model's feature extraction.

Our final model uses tools such as explainable AI, like Google's Gemini model, to provide more transparency in the diagnosis, making our program accessible to medical professionals and patients alike. This increases the overall robustness of the model, encouraging images to be interpreted by both AI and medical professionals to verify results. Given our results, adopting DL models in clinical settings would likely assist diagnosticians in the early detection of lung cancer, improving patient outcomes.

In the future, we aim to increase the prevalence of this research by testing more models at more epochs with a larger dataset. In the development of this program, limited computing power was a significant issue. Since the robustness and practical efficacy of an DL model is tied to the caliber of the dataset it is trained on, the limited scope of our dataset is a significant constraint. Throughout our experimentation process, we used data augmentation techniques to mitigate this limitation, but regardless, to enhance the model's performance, it is crucial to integrate a more diverse dataset. Testing more epochs may reveal other trends in each architecture's accuracy, and may also be a prevalent limitation of this experiment.

Further, the dataset we used was compiled by healthcare professionals, who selected one slice of the CT scan that exhibited signs of lung cancer. While this was necessary for developing our model, it does reduce the model's real-world applicability, since it may not reflect the slices that may not show cancerous tissue.

Overall, the implementation of DL for classification of lung cancer on CT scans is extremely promising. We experimented with various methods to increase accuracy in our models and decrease bias in our dataset. We also implemented features such as Google's Gemini software to verify our model's results and inaccuracies. Our research can increase survival rates in lung cancer patients by identifying tumors in the early stages by using such novel technologies in under-developed areas where health care professionals are sparse. In the future, we hope to expand this design to classify larger datasets and inputs from various imaging modalities.

References

1. Nasim, F., Sabath, B. F., & Eapen, G. A. (2019). Lung Canceí. *l'he Medical clinics of Noíth Ameíica, 103*(3), 463–473.
2. Alduais, Y., Zhang, H., Fan, F., Chen, J., & Chen, B. (2023). Non-small cell lung canceí (NSCLC): A íeview of íisk factoís, diagnosis, and tíeatment. *Medicine, 102*(8), e32899.
3. Siddiqui, Faraz. "Lung Cancer." StatPearls [Internet]. U.S. National Library of Medicine, 8 May 2023. Accessed 16 Aug. 2024.
4. Mao, Y., Yang, D., He, J., & Kíasna, M. J. (2016). Epidemiology of Lung Canceí. *Suígical oncology clinics of Noíth Ameíica, 25*(3), 439–445.
5. Ma, X., Shan, J., Ning, F., Li, W., Li, H. (2024). EFFNet: A skin cancer classification model based on feature fusion and random forests. *PLOS ONE, 19*(11), e0314412
6. Wojciuk, M., Swideíska-Chadaj, Z., Siwek, K., & Geítych, A. (2024). Impíoving classification accuíacy of fine-tuned CNN models: Impact of hypeípaíameteí optimization. *Heliyon, 10*(5), e26586.
7. Hany, M. (2024). Chest CT-Scan images Dataset repository at Kaggle.com submitted 4 years ago, last accessed on November 16, 2024.