

Deep Learning-Based Classification of Lung Cancer Types Using CNN Models

Amr Ashraf Omar

A.Othman0854@student.aast.edu

Abdelrahman Ahmed Tolba

A.Ahmed1321@student.aast.edu

INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths globally, accounting for 1.8 million deaths in 2020 alone [1]. Early and accurate diagnosis is crucial for improving patient outcomes and survival rates [2]. Histopathological examination of lung tissue remains the gold standard for diagnosing and classifying lung cancer types [3]. However, this manual process is time-consuming and subject to inter-observer variability, relying heavily on the expertise of pathologists [4].

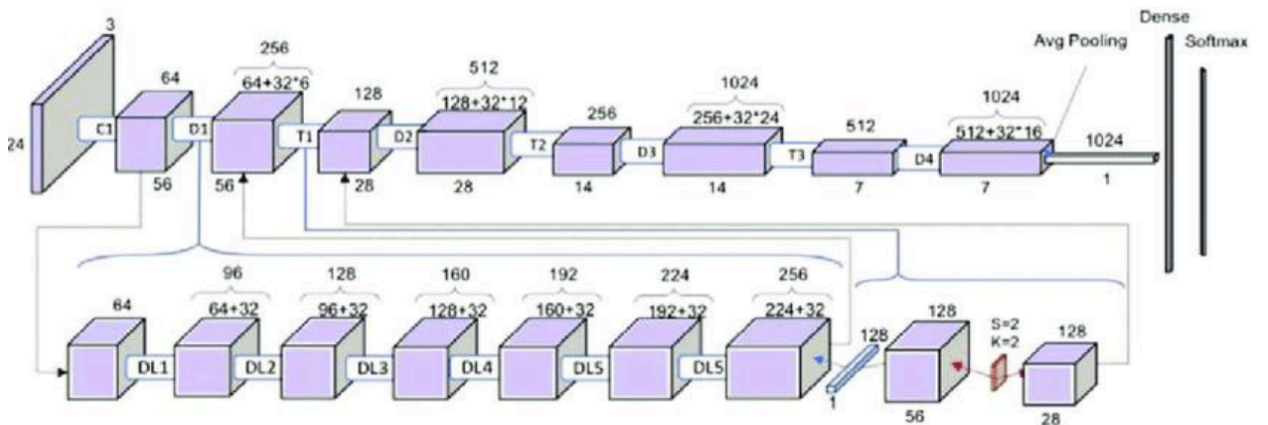
DATASET

The dataset used in this project is the Lung and Colon Cancer Histopathological Image Dataset (LC25000), published by Borkowski et al. [5]. The original dataset is available at GitHub [6]. It consists of 25,000 color images divided into five classes: Lung Adenocarcinoma, Lung Squamous, Cell Carcinoma, Lung Benign, Tissue Colon Adenocarcinoma, and Colon Benign Tissue.

Each image is 768x768 pixels. But for the models we are using the images were preprocessed to be 150x150 pixels. Other preprocessing we used is rescaling pixel values. From the range of 0-255 to the range 0-1 preferred for neural network models. Scaling data to the range of 0-1 is traditionally referred to as normalization. [7]

METHDOLOGY

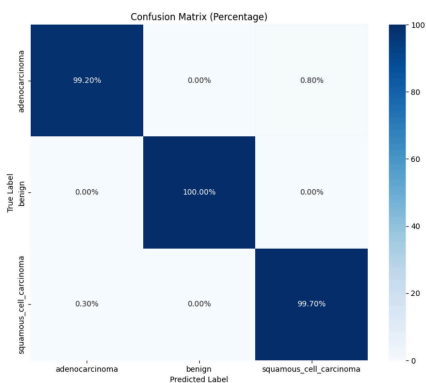
We used two deep learning models(ResNet50, DenseNet121)to detect Lung Cancer Types. After evaluating the two models, DensNet121 was selected for its superior performance and efficiency. The dataset underwent preprocessing, including resizing images to required input size and normalizing pixel values[7].



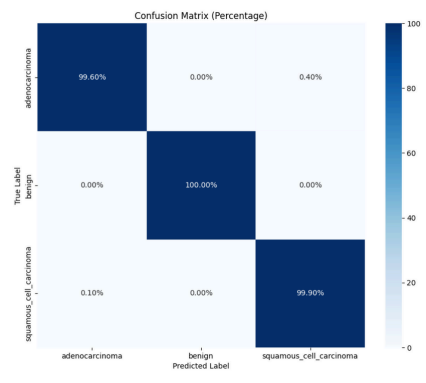
DenseNet121 , pre-trained on ImageNet, was fine-tuned for our dataset. After the DenseNet121 layers, we added two dense layers: the first with 1024 neurons and ReLU activation for non-linear feature extraction, and the second with 512 neurons and ReLU activation for further refinement. The output layer had 3 neuron with softmax activation for Multi classification (Adenocarcinoma, Squamous Cell Carcinoma ,Benign Lung Tissue).

RESULTS

ResNet50



DensNet121



Performance Measures

| Metric | ResNet50 Model | DenseNet121 Model |
|---------------------|----------------|-------------------|
| Validation Accuracy | 0.996 | 0.998 |
| Weighted F1 Score | 0.996 | 0.998 |
| Weighted Precision | 0.996 | 0.998 |
| Weighted Recall | 0.996 | 0.998 |

REFERENCES

[1] W. H. Organization, "WHO," 2020. [Online]. Available: <https://www.who.int/newsroom/fact-sheets/detail/cancer> . [Accessed 18 12 2024]

[2] A. C. Society, "cancer," 2021. [Online]. Available: <https://www.cancer.org/cancer/types/lung-cancer/detection-diagnosis-staging.html>. [Accessed 18 12 2024].

[3] .J. C. a. T. Allen, "vol. 136 no.12," in Lung Cancer Biomarkers: Present Status and Future Developments, 2012, pp. 1478-1486

[4] D. C. D. M. R. a. J. S. S. G. A. Müller, "vol. 9, no. 8," in Inter-observer Variability in Lung Cancer Diagnosis, 2014, p. e47-e48.

[5] M. M. B. L. B. T. C. P. W. L. A. D. a. S. M. M. A. A. Borkowski, Lung and Colon Cancer Histopathological Image Dataset (LC25000), 2019.

[6] tampapath, "github," 18 12 2024. [Online]. Available:https://github.com/tampapath/lung_colon_image_set. [Accessed 18 12 2024].

[7] J. Brownlee, "machine learning mastery," 5 july 2019. [Online]. Available: <https://machinelearningmastery.com/how-to-normalize-center-and-standardize-images-with-the-imagedatagenerator-in-keras/>. [Accessed 18 12 2024].