Enhancing Early Diagnosis of Lung and Colon Cancer Through Deep Learning Analysis of Histopathological Images

1st Amir Keshtkar
Faculty of Engineering
University of Western Ontario
London, Canada
akeshtk@uwo.ca

2rd Michael Ruiz
Faculty of Engineering
University of Western Ontario
London, Canada
mruiz6@uwo.ca

3th William Sharkey
Faculty of Engineering
University of Western Ontario
London, Canada
wsharke@uwo.ca

4th Ramtin Sirjani
Faculty of Science
University of Western Ontario
London, Canada
rsirjani@uwo.ca

5th Morgan Walker
Faculty of Engineering
University of Western Ontario
London, Canada
mwalk56@uwo.ca

***Abstract*—Colon and Lung cancer are among the most critical health challenges globally, requiring early and accurate diagnosis to reduce mortality rates. However, the complexity of diagnosing these cancers often relies on the expertise of histopathologists, where a lack of experience can significantly threaten patient outcomes. With the rise of deep learning technologies, there has been a shift towards leveraging these tools for enhancing medical imaging analysis. This study focuses on employing deep learning to improve the detection of lung and colon cancer through histopathological images. Specifically, we use two pre-trained CNN models, NASNetMobile and InceptionV3, refined for this purpose. These models are trained and evaluated on the LC25000 dataset, focusing on metrics such as precision, recall, F1 score, accuracy, and AUROC to gauge performance. Additionally, we apply GradCAM and SmoothGrad techniques to visualize the decision-making processes of these CNN models, offering insights into their ability to differentiate between malignant and benign cases. This study not only highlights the efficacy of NASNetMobile and InceptionV3 in cancer detection but also showcases the potential of deep learning in medical imaging analysis, paving the way for more reliable and early diagnosis of critical diseases.The results demonstrate exceptional performance of both models across various metrics, with NASNetMobile achieving precision, recall, and F1 score of 1 for detecting benign lung tissue, and InceptionV3 achieving similar scores for detecting benign colon tissue, indicating promising avenues for future research in enhancing diagnostic accuracy and patient outcomes.**

***Index Terms*—Deep Learning, Medical Imaging, CNN, NASNetMobile, InceptionV3, Lung Cancer, Colon Cancer, Histopathology, GradCAM, SmoothGrad.**

I. Introduction

Cancer remains a paramount health concern globally, causing substantial morbidity and mortality. Lung cancer, deemed the most common cancer globally with 2.5 million new cases in 2022, accounts for 12.4% of the total new cases [1]. Both lung and colon cancers represent a significant part of this global cancer burden. Specifically, colorectal cancer, which includes colon cancers, totals 1.9 million new cases, contributing 9.6% of the new cases [1]. This highlights the significant prevalence of these two types of cancer, underscoring the urgent need for advanced diagnostic and treatment strategies. Presently, lung and colon cancer diagnosis heavily relies on biopsy analysis. This involves histopathological examination of the samples, which necessitates a trained professional for accurate interpretation. Owing to the limited healthcare resources and professionals, especially in developing regions, accomplishing this can be challenging. Artificial intelligence presents a promising alternative to these obstacles [2].

In this context, our study employs transfer learning, a potent technique in the realm of deep learning, leveraging pre-existing, robustly trained CNN models - specifically, NASNetMobile and InceptionV3. These models, known for their exceptional performance in previous research, are adapted and refined to suit the specific demands of lung and colon cancer diagnosis. By transferring knowledge from diverse imaging tasks to our focused domain, these models are poised to offer significant improvements in accuracy and reliability over traditional diagnostic methods.

II. Related Work

Several previous studies have explored the application of deep learning techniques in histopathological image analysis for long and colon cancer detection. For instance, Masud et al. used a multi-channel CNN for the detection of lung and colon cancer classification using a dataset of histopathological images [3]. Similarly, Suresh et al. investigated the use of ROI-based feature learning using CNN for the diagnosis of lung cancer, with a classification accuracy of 93.9% [4].

One study compared different architectures on a standardized lung and colon cancer dataset, measuring the performance

of 8 different CNN-based pre-trained models [5]. In addition to assessing and comparing pre-trained models, we aim to extract feature activations from the convolutional layers.

III. Data

The dataset employed in this study encompasses 25,000 histopathological images, each measuring 768 x 768 pixels and stored in JPEG format. Originating from HIPAA compliant and validated sources, the dataset initially consisted of 750 images, comprising varying proportions of benign and malignant lung and colon tissue samples [6]. Through augmentation techniques utilizing the Augmentor package, the dataset was expanded to 25,000 images, maintaining a balanced distribution across five distinct classes: Lung Benign Tissue, Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Colon Adenocarcinoma, and Colon Benign Tissue. This comprehensive dataset offers researchers an invaluable resource for exploring histopathological patterns across different tissue types and pathological conditions, fostering advancements in medical imaging and pathology research.

IV. Methods

Research Objectives

This work aims to address the critical challenge in medical diagnosis of lung and colon cancer by assessing the capability of a deep convolutional neural network to perform classification between benign and cancerous histopathological images. The performances of two DCNNs, InceptionV3 and NASNet-Mobile, were bolstered through training on augmented images from the LC25000 dataset. The models were subsequently used to classify test and validation sets and their performances were documented and compared with one another. The performance of the models was evaluated through metrics such as accuracy, precision, recall, f1-score, and support. From these metrics we can assess if a DCNN can reliably diagnose instances of lung and colon cancer, which is particularly useful for areas of the world where availability to healthcare is scarce.

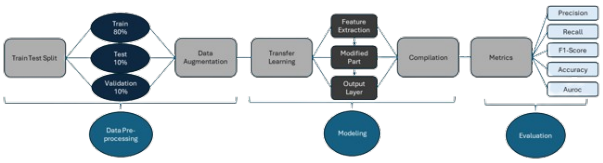

Fig. 1: Schematic diagram outlining the three-stage research framework employed for classification. The stages include Data Preprocessing, Modeling, and Evaluation.

Dataset Configuration for Multiclass Classification

The LC25000 dataset, a well-known lung and colon cancer histopathological image dataset, forms the basis of this study. This dataset, made publicly available to address the scarcity of accessible clinical imaging datasets, comprises 15,000 lung and 10,000 colon images. These are evenly distributed across five distinct classes: lung adenocarcinoma, lung squamous cell carcinoma, lung benign, colon carcinoma, and colon benign, with each class containing 5,000 images.

This research adopts a multi-class classification framework. This involves classifying images into one of the five aforementioned categories. The dataset is split in the following manner for this purpose: 80% of the data is allocated for training, 10% for testing, and the remaining 10% for validation.

Data Augmentation

In this study, we're working with the LC25000 dataset [17], where each image starts at 768x768 pixels. We've resized these images to 224x224 pixels, a necessary step for consistency and computational efficiency when feeding them into our convolutional neural network (CNN) models. The dataset comes with some pre-applied augmentations: rotations up to 25 degrees, and flips horizontally and vertically, each with specific probabilities [17]. These initial augmentations are essential as they introduce a level of variability that helps in preventing the model from overfitting on overly uniform data.

To take our augmentation process further, we utilize the imgaug library [5]. This library allows us to introduce a more diverse range of image transformations, significantly enhancing our data's variability (figure). By doing so, we aim to train our CNN models to recognize and classify images more robustly, improving their ability to generalize to new, unseen images. This is particularly crucial in medical imaging, where subtle variations can be key to accurate diagnoses.

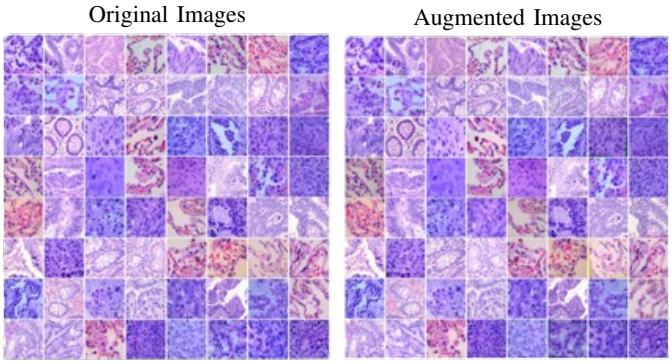

Fig. 2: This figure illustrates the Original Images (left) and potential Augmentations to the Images (right), demonstrating the visual differences of possible augmentations applied in our work

we apply a series of affine transformations. These include slight scaling of images between 95% and 105% of their

original size, simulating the zoom variations commonly seen in image capture. Additionally, minor translations of up to 5% in any direction are used to mimic small shifts in patient positioning. Rotations between -5 and +5 degrees are also included to represent the variability in how patients are positioned during imaging. These transformations are executed with different interpolation methods and warping modes, adding realism to the simulation.

To further enhance the dataset's realism, we adjust image brightness and contrast. Brightness is varied by multiplying pixel values by factors between 0.9 and 1.1, simulating different exposure levels. Similarly, contrast adjustments are made within the same range, reflecting the potential differences in imaging conditions.

Lastly, we incorporate mild elastic transformations to the images. By applying random, subtle elastic deformations, we simulate the natural flexibility of biological tissues. These are controlled by the parameters alpha, dictating the intensity of the deformation, and sigma, which determines the smoothness of the applied transformation.

These augmentations are selectively applied, introducing a level of randomness that enhances the dataset's diversity. This approach is crucial in training CNN models, as it helps them learn to generalize more effectively across a variety of imaging conditions and patient-specific variations.

Selected Models

we selected the NasNetMobile and InceptionV3 models as representatives based on their high performance in previous research. These models have historically emerged as frontrunners in the prediction of lung and colon cancer from histopathological images and thus are ideal targets for further analysis.

The rationale for choosing pre-trained CNN models lies in the concept of transfer learning. Transfer learning involves utilizing the knowledge (weights) gained while solving one problem and applying it to a different but related problem. In the context of our study, this means leveraging the intricate patterns and features learned by these models in various tasks and applying them to the specific challenge of diagnosing lung and colon cancer.

Compilation

In compiling the InceptionV3 model, the Adam optimizer was selected for its proven efficacy in diverse deep learning applications. We set a learning rate of 0.0001 to ensure gradual, precise weight adjustments. The training extends over 30 epochs, allowing comprehensive learning while avoiding overfitting. A batch size of 64 is chosen, optimizing computational efficiency and gradient estimation effectiveness. These parameters are crucial for the model's optimal training and performance.

Metrics

The evaluation of the models is conducted using a set of key metrics: precision, recall, F1-score, accuracy, and AUROC (Area Under the Receiver Operating Characteristic). Precision is the ratio of correctly predicted positive observations to the total predicted positives, calculated as $\frac{TP}{TP+FP}$, where TP and FP are the numbers of true positives and false positives, respectively. Recall, or sensitivity, measures the proportion of actual positives correctly identified, computed as $\frac{TP}{TP+FN}$. Accuracy assesses the overall correctness of the model and is given by $\frac{TP+TN}{TP+TN+FP+FN}$, incorporating both true negatives (TN) and false negatives (FN). The F1-score provides a balance between precision and recall, calculated as $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. Lastly, AUROC represents the relationship between the true positive rate and the false positive rate across various thresholds, serving as a crucial metric for evaluating the model's discriminative ability. These metrics collectively enable a comprehensive assessment of the models' performance.

V. Experimental Result

The procedure for comparing the InceptionV3 and NASNet-Mobile model performance involves the measurement of the precision, recall and F1 score of the on the test data. The test data is made up of 10% of the total dataset, which is made up of 2 000 images, and has the same preprocessing as the training data. The remaining dataset is split between training and validation, with the training data making up 80% of the dataset and the validation data making up 10% of the dataset. In addition, the plots for training loss, validation loss, training accuracy and validation accuracy are given for each model.

A. InceptionV3

The precision, recall and F1 score of the InceptionV3 model for the five classes are shown in figure 5. The model performed well on every metric, with scores between 0.98 and 1 for each class. The model performed particularly well on detecting benign colon tissue, with a precision, recall and f1-score of 1.

	precision	recall	f1-score	support
Lung_benign_tissue	0.998004	1.0000	0.999001	500.0000
Lung_adenocarcinoma	0.993964	0.9880	0.990973	500.0000
Lung_squamous_cell_carcinoma	0.992016	0.9940	0.993007	500.0000
Colon_adenocarcinoma	0.994012	0.9960	0.995005	500.0000
Colon_benign_tissue	1.000000	1.0000	1.000000	500.0000
accuracy	0.995600	0.9956	0.995600	0.9956
macro avg	0.995599	0.9956	0.995597	2500.0000
weighted avg	0.995599	0.9956	0.995597	2500.0000

Fig. 3: Precision, recall, F1 score of InceptionV3 model on the test set.

The confusion matrix is shown in figure 6, providing a breakdown of the model's classification results,

Figure 7 shows the plots for training loss, validation loss, training accuracy and validation accuracy.

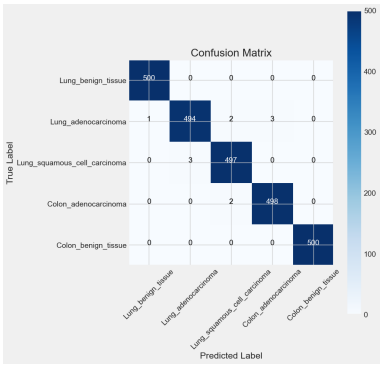

Fig. 4: Confusion Matrix of InceptionV3 model on the test set.

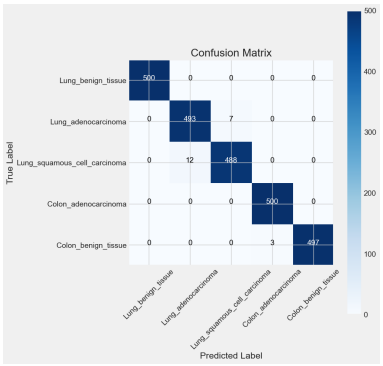

Fig. 7: Confusion Matrix of NASNetMobile model on the test set.

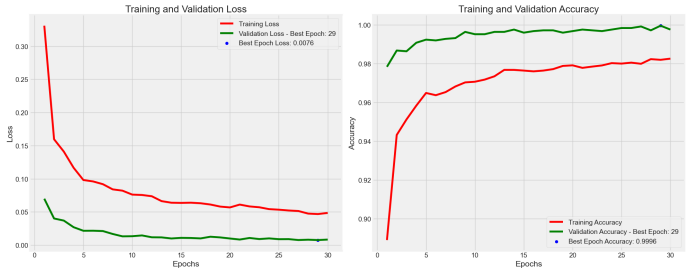

Fig. 5: InceptionV3 Training and Validation Loss and Training and Validation Accuracy.

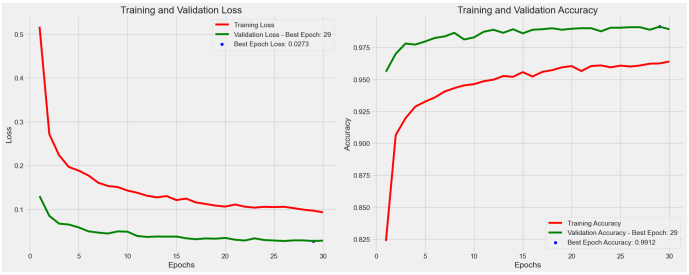

Fig. 8: NASNetMobile Training and Validation Loss and Training and Validation Accuracy.

B. NASNetMobile

The precision, recall and F1 score of the NASNetMobile model for the five classes are shown in figure 8. The model performed well on every metric, with scores between 0.98 and 1 for each class. The model performed particularly well on detecting benign lung tissue, with a precision, recall and f1-score of 1.

	precision	recall	f1-score	support
Lung_benign_tissue	1.000000	1.0000	1.000000	500.0000
Lung_adenocarcinoma	0.976238	0.9860	0.981095	500.0000
Lung_squamous_cell_carcinoma	0.985859	0.9760	0.980905	500.0000
Colon_adenocarcinoma	0.994036	1.0000	0.997009	500.0000
Colon_benign_tissue	1.000000	0.9940	0.996991	500.0000
accuracy	0.991200	0.9912	0.991200	0.9912
macro avg	0.991226	0.9912	0.991200	2500.0000
weighted avg	0.991226	0.9912	0.991200	2500.0000

Fig. 6: Precision, recall, F1 score of NASNetMobile model on the test set.

The confusion matrix is shown in figure 9, providing a breakdown of the model's classification results.

Figure 10 shows the plots for training loss, validation loss, training accuracy and validation accuracy.

Both the InceptionV3 and NASNetMobile models performed well on the test data, with each models obtaining scores of over 98% for each class in every metric assessed. The InceptionV3 model performed slightly better on average across each metric in comparison to the NASNetMobile model, however, the difference in performance is minimal. The plots for loss and accuracy of training and validation data was comparable for each model, with both plots demonstrating early convergence.

VI. Conclusions

In conclusion, this study underscores the significant potential of deep learning techniques, specifically convolutional neural networks (CNNs), in enhancing the early diagnosis of lung and colon cancer through histopathological image analysis. By leveraging CNN models, InceptionV3 and NASNetMobile, and training them on the LC25000 dataset, we have demonstrated strong performance in accurately classifying tissue samples into benign and malignant categories.

The results obtained from our experiments reveal the robustness and effectiveness of both InceptionV3 and NASNetMobile models in distinguishing between different tissue types, as evidenced by high precision, recall, and F1 scores across all classes. Notably, both models exhibited exceptional performance in detecting specific types of cancerous tissue, such as benign colon tissue for InceptionV3 and benign lung tissue for NASNetMobile, highlighting their versatility and applicability in clinical settings.

Furthermore, our study contributes to the existing body of research by providing insights into the decision-making processes of these CNN models through visualization tech-

niques like GradCAM and SmoothGrad. These techniques offer valuable interpretability, enabling a deeper understanding of how the models differentiate between benign and malignant tissue samples.

The comprehensive evaluation of these deep learning models on a diverse and well-curated dataset underscores their potential as reliable tools for aiding histopathologists in cancer diagnosis. By automating and augmenting the diagnostic process, these models hold promise for improving patient outcomes, particularly in regions with limited access to healthcare resources and expert pathologists.

Moving forward, further research can explore avenues for optimizing model performance, fine-tuning architectures, and integrating additional data sources to enhance the generalizability and scalability of these deep learning solutions. Additionally, efforts towards seamless integration of these AI-driven tools into clinical workflows and regulatory frameworks will be crucial for realizing their full potential in transforming cancer diagnosis and treatment paradigms.

This study not only advances our understanding of deep learning applications in medical imaging but also underscores the transformative impact of AI technologies in addressing critical healthcare challenges, ultimately leading to improved patient care and outcomes.

REFERENCES

[1] World Health Organization, *Global cancer burden growing, amidst mounting need for services*, February 2024. [Online]. Available: https://www.who.int/news/2024/02/01/global-cancer-burden-growing/

[2] M. Al-Jabbar, M. Alshahrani, E.M. Senan, I.A. Ahmed, *Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features*, Bioengineering (Basel), vol. 10, no. 3, pp. 383, March 2023.

[3] M. Masud, N. Sikder, A. A. Nahid, A. K. Bairagi, and M. A. AlZain, "A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework," Sensors, vol. 21, no. 3, p. 748, 2021. [Online]. Available: https://doi.org/10.3390/s21030748

[4] S. Suresh and S. Mohan, "ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis," Neural Comput. Appl., vol. 32, pp. 15989–16009, 2020. [Online]. Available: https://doi.org/10.1007/s00521-020-04787-w

[5] S. Garg and S. Garg, "Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps," in Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference (AICCC '20), New York, NY, USA, 2021, pp. 38–45. [Online]. Available: https://doi.org/10.1145/3442536.3442543

[6] "LC25000 Lung and Colon Cancer Histopathological Image Dataset." [Online]. Available: https://github.com/tampapath/lung_colon_image_set/blob/master/README.md