



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

⊕ www.ijirset.com 🖂 ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025 |DOI: 10.15680/IJIRSET.2025.1405383

Revolutionizing Lung and Colon Cancer Detection: Deep Learning Applications using Efficient Net on Histopathological Images

P. Balashanmuga¹, VadivuV², Iswarya, S³, Ligeetha, S⁴, Gowsalya⁵

Associate Professor, Department of Electronics and Communication Engineering, Mahendra Engineering College,

Namakkal, Tamil Nadu, India¹

UG Student, Department of Electronics and Communication Engineering, Mahendra Engineering College, Namakkal,

Tamil Nadu, India²⁻⁴

ABSTRACT: Detecting lung and colon cancers in their initial phases is essential to enhance patient outcomes, but conventional histopathological examination is time-consuming and susceptible to human error. To overcome this, we introduce an automated, deep learning-based cancer detection system using histopathological images. The system employs an EfficientNet model, which is trained on an augmented dataset of 25,000 images derived from HIPAA-compliant and validated samples. By using preprocessing techniques like resizing, noise removal, and feature extraction, In terms of classification, the model yields impressive results of benign and cancer tissue samples into five separate classes. We evaluate the model's performance using a range of metrics, including accuracy, specificity, precision, recall, sensitivity, and F1-score. This strategy not only increases diagnostic speed and accuracy but also opens the door to incorporating AI-based support tools into clinical work flows, thus assisting pathologists and minimizing diagnostic errors.

KEYWORDS: Deep Learning, EfficientNet, Histopathological Images, Lung Cancer Detection, Colon Cancer Detection, Image Classification, Data Augmentation, Medical Image Analysis

I. INTRODUCTION

Cancer continues to rank among the foremost causes of death globally, with lung and colorectal cancers contributing a large percentage of mortality [1]. Early and precise detection is essential for enhancing patient prognosis, allowing for early treatment intervention, and mitigating healthcare loads [2]. Conventionally, histopathological review of tissue biopsies under a microscope ranks as the gold standard in cancer diagnosis. But this is a very subjective, timeconsuming, and prone to human error process, particularly when pathologists have to analyze a high number of intricate images [3].Recent technological progress in artificial intelligence, particularly deep learning, has been highly promising for analyzing medical images automatically [4]. Convolutional Neural Networks (CNNs) perform well since they can learn intricate, hierarchical patterns from images, yielding leading results in applications such as detecting cancer [5]. Of the various deep learning models, EfficientNet is notable for its excellent performance, obtained through a brilliant balance of depth, width, and image resolution within the network [6]. Herein, we present a deep learning architecture intended to automatically classify and detect lung and colon cancer based on histopathological images. With the use of an augmented dataset containing 25,000 images drawn from HIPAA-compliant sources, an EfficientNet-based framework is trained to distinguish between malignant and benign tissue samples in five categories. Additional preprocessing methods involving resizing, removal of noise, and feature extraction also improve the performance of the model. To evaluate the performance, strict measures measures like accuracy, sensitivity, specificity, precision, recall, and F1-score. The adoption of AI-based diagnostic systems not only seeks to assist pathologists with the avoidance of workload and decreased diagnostic mistakes but also forms the basis of future digital pathology and precision medicine [7]. In this article, methodology, experimental findings, and possible clinical impact of using EfficientNet models for cancer detection of the lung and colon are discussed.

IJIRSET©2025





[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|

II. RELATED WORKS

Recent years have seen the tremendous rise of research on using deep learning for medical image analysis, especially for cancer detection applications. Histopathological images, which are high in morphological information, are important diagnostic tools, and their interpretation using AI has been a significant research thrust in recent years [3]. Litjens et al. [4] performed an exhaustive review of the use of deep learning in medical image analysis and the success of CNN-based models in histopathology. Tan and Le [6] introduced the EfficientNet architecture, which achieved exceptional performance on many vision tasks with low computational complexity and is a suitable choice for medical imaging where accuracy is paramount.

Lu et al. [5] created an AI model that can predict the sites of cancers with unknown primaries, showing the vast potential of deep learning in clinical diagnosis. Their research highlights the significance of fine-grained image analysis, a principle that can also be applied in distinguishing between lung and colon cancer tissues. Furthermore, Echle et al. [7] demonstrated that deep learning models could achieve clinical-grade performance in detecting microsatellite instability directly from colorectal cancer tissue slides, paving the way for non-invasive diagnostic alternatives. Along with these initiatives, more recent studies have aimed at improving model performance via sophisticated augmentation and attention mechanisms. For example, Talo [8] used deep learning models with large data augmentation to improve classification accuracy in histopathological cancer detection, highlighting the significance of pre-processing techniques. Furthermore, AbdulJabbar et al. [9] presented new attention-based approaches for enhancing the interpretability and diagnostic performance of deep learning models used in lung and colon cancer datasets. In general, the literature points towards a clear trend of utilizing deep learning models, specifically EfficientNet and attention-augmented CNNs, to enhance the speed, accuracy, and reliability of cancer detection systems.

III. PROPOSED WORK

A. Overview

We're putting forward a system that uses an EfficientNet-based deep learning model to automatically classify lung and colon cancer from histopathological images. The goal is to minimize human error, speed up diagnosis, and boost classification accuracy. The approach is split into two main stages: training and testing, with clear steps for preprocessing, feature extraction, and classification.

Β.	Dataset	Descri	ption
<u> </u>	Databet	Deberr	pulon

Disease name	File count	Sample images
Lung benign tissue	5000	
Lung adenocarcinoma	5000	TE C
Lung squamous cell carcinoma	5000	
Colon adenocarcinoma	5000	
Colon benign tissue	5000	

Figure 1. Block Diagram of Training and Testing Phases





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|

The lung and colon cancer data sets used here were obtained from the publicly available Kaggle platform. The dataset is made up of five unique classes, each with 5,000 images following data augmentation.

- Lung benign tissue
- Lung adenocarcinoma
- Lung squamous cell carcinoma
- Colon adenocarcinoma
- Colon benign tissue

This data set employed consists of 750 lung tissue images (250 benign, 250 adenocarcinoma, 250 squamous cell carcinoma), 500 colon tissue images (250 benign & 250 adenocarcinoma). To enhance the dataset's diversity and size, the images were augmented to a total of 25,000 samples using the Augmentor package [8].

C. Preprocessing

Prior to feeding the images into the deep learning model, we apply a set of pre-processing steps:

- 1. Image Resizing: All the images are resized to a consistent size appropriate for the Efficient Net model's input layer
- 2. Noise Filtering: A median filter is used in each image to filter out noise and retain important details.
- 3. Normalization: Pixel values are normalized to enhance model convergence during training.

These pre-processing operations provide consistency throughout the dataset and enhance the quality of feature extraction.

D. Model Architecture and Training



Figure 2. Model Architecture and Training

The Efficient Net architecture is selected due to its high accuracy and efficient scaling [6]. A pre-trained EfficientNet model (e.g., EfficientNet-B0) is fine-tuned with the histopathological dataset:

- Transfer Learning: Pre-trained ImageNet weights are used to take advantage of learned low-level features [6].
- Fine-tuning: The network is again trained on the histopathological dataset to fit the new classification task.
- Data Augmentation: Rotation, flipping, and zooming are used during training to improve model generalization [8]. The training process is tracked by validation accuracy and loss to prevent overfitting. Hyperparameters like batch size, learning rate, and epoch count is fine-tuned to attain optimal model accuracy.

E. Feature Extraction and Classification

Following training, the model learns to extract discriminative features from test images. These features are classified into one of the five classes discussed above.

The output includes:





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|

• Predicted class label

• Probability score for each class

This not only helps in cancer detection but also in giving confidence levels for medical interpretation.

F. Performance Evaluation Metrics

We assess the performance of our proposed model using several widely accepted metrics[14-15]:

Accuracy: This shows how frequently the model is correct overall. It's the proportion of accurate predictions (both positive and negative) out of all predictions made.

Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives) (1)

Precision: This measures how reliable the model's positive predictions are. It shows what fraction of the predicted positives are actually correct.

Precision = TP/(TP + FP)(2)

Recall (Sensitivity): This checks how good the model is at catching all the actual positives. It's the proportion of true positives the model correctly identifies out of all real positives.

Recall = True Positives / (True Positives + False Negatives) (3)

Specificity: This looks at how well the model identifies negatives. It's the fraction of actual negatives that the model correctly predicts as negative.

Specificity = True Negatives / (True Negatives + False Positives)

(4)

F1-Score: This is a balance between Precision and Recall, especially handy when your data has uneven class sizes

F1-Score = 2 × (Precision × Recall) / (Precision + Recall) (5)

IV. RESULTS AND DISCUSSION

The model used in this research, which is based on EfficientNet, performs an excellent task of classifying lung and colon cancer tissue images into five categories: benign lung tissue, lung squamous cell carcinoma, colon adenocarcinoma, lung adenocarcinoma, and benign colon tissue. Its performance is quite impressive for this type of classification task. The model is measured by a wide range of metrics drawn from the A3.15 classification report, including precision, accuracy, F1-score, recall, and support[16]. Also, the confusion matrix, training/validation accuracy, and training/validation loss curves give more insight into the performance of the model[10],[11].

The report of classification (A3.15) shows a total accuracy of 96%, based on 3,000 test samples. Class-specific performance metrics are excellent: Class 0 (lung benign tissue) has a precision is 0.93, recall is 0.96, and F1-score is 0.94 with 964 samples; Class 1 (lung adenocarcinoma) has a precision and recall of 0.99 and 1.00, respectively, and F1-score of 0.99 and 1,026 samples; Class 2 (lung squamous cell carcinoma) has a precision of 0.97, recall of 0.94, and F1-score is 0.95 with 1,010 samples. These findings establish great reliability in differentiating malignant from benign tissues, with Class 1 having almost flawless classification [12-13].





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025





The confusion matrix also confirms the model performance. In the case of the original dataset, it indicates 985 true negatives and 3 false positives for Class 0, 12 false negatives and 1,000 true positives for Class 1, 31 false positives for Class 2, and total 945 true negatives. In a follow-up assessment, the matrix shows 923 true negatives and 10 false positives for Class 0, 5 false negatives and 1,021 true positives for Class 1, and 65 false positives for Class 2, with 945 true negatives. These findings reflect the model's capacity to reduce misclassifications, especially for adenocarcinoma cases, although minor improvements may be aimed at Class 0 and Class 2 to decrease false Positives.



Figure 4. Training and Validation Accuracy

Accuracy curves for training and validation show steady model convergence. Training accuracy increases sharply from 0.96 to 0.995 within 4 epochs, whereas validation accuracy plateaus between 0.94 and 0.96, reflecting good generalization. A second test indicates training accuracy improving from 0.89 to 0.96 and validation accuracy plateauing at 0.94, supporting the model's strength on varying dataset splits.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|





Loss curves for training and validation provide additional insight into the accuracy results. The initial loss falls from 0.300 to 0.125 for training and from 0.275 to 0.150 for validation across 4 epochs, indicating effective learning. A repeat run indicates training loss falling from 0.300 to 0.175 and validation loss from 0.250 to 0.125, further validating the stability of the model and minimal overfitting.



Figure 6. Training and Validation Loss

The trained EfficientNet model achieved outstanding classification performance on the test set.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|

Class Label	Precision	Recall	F1- Score	Support
0 (Benign Tissue)	0.93	0.96	0.94	964
1 (Cancer Type 1 - e.g., Adenocarcinoma)	0.99	1.00	0.99	1026
2 (Cancer Type 2 - e.g., Squamous Cell Carcinoma)	0.97	0.94	0.95	1010
Overall Accuracy	0.96	3000		

Table 1 Classification Performance Report

The total accuracy achieved was 96%, This demonstrates the model's strong capability to accurately distinguish between different tissue types[17].Table 1 provides the classification accuracy of the suggested EfficientNet-based model over three main classes.The model's accuracy was 0.93, 0.99, and 0.97 for Class 0, Class 1, and Class 2 respectively, reflecting a high rate of correct positive predictions. The recall scores, especially the perfect 1.00 score for Class 1 (cancer type), reflect the model's high sensitivity to detect cancerous tissues. The F1-scores, equally weighing precision and recall, were remarkably high across all classes, and the overall model accuracy was 96% on a test set of 3000 images.

These results prove the suggested system to be extremely reliable for automated lung and colon cancer detection from histopathological images[18].

V. CONCLUSION

The creation and testing of an EfficientNet-powered deep learning algorithm for the automated identification and lung classification, colon cancer from histopathological images represent a major break-through in digital pathology. Through the application of a data set of 25,000 augmented images and the use of preprocessing methods such as resizing, noise removal, and normalization, the proposed algorithm displays a high overall accuracy of 96%, with excellent performance among five different classes. The classification performance, with accompanying confusion matrices and training/validation loss and accuracy plots, illustrate the model's competency in identifying benign and malignant tissue, with superior performance in identifying lung adenocarcinoma with virtually perfect precision and recall. This research identifies the use of EfficientNet architectures as an ability to boost diagnostic accuracy, minimize human bias, and hasten the process of histopathological analysis, thus aiding pathologists in clinical pipelines. The slight variations in how well the model performs for benign lung tissue and squamous cell carcinoma suggest there's room for further improvement. by using more sophisticated data augmentation, hyperparameter tuning, or attention mechanisms. Future studies might also investigate larger databases and real-time clinical testing to further establish the model's relevance in personalized medicine Finally, this work provides a foundation for AI-based tools that might revolutionize cancer diagnosis, enhance patient care, and push the boundaries of medical image analysis.

REFERENCES

- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660
- Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer Statistics, 2022. CA: A Cancer Journal for Clinicians, 72(1), 7–33.
- 3. Tizhoosh, H. R., & Pantanowitz, L. (2018). Artificial Intelligence and Digital Pathology: Challenges and Opportunities. Journal of Pathology Informatics, 9, 38.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88. https://doi.org/10.1016/j.media.2017.07.005





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405358|

- Lu, M. Y., Chen, T. Y., Williamson, D. F. K., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). AI-based pathology predicts origins for cancers of unknown primary. Nature, 594(7861), 106–110. https://doi.org/10.1038/s41586-021-03512-4
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105-6114. https://arxiv.org/abs/1905.11946
- Echle, A., Grabsch, H. I., Quirke, P., van den Brandt, P. A., West, N. P., Hutchins, G. G. A., ... & Kather, J. N. (2021). Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology, 159(4), 1406–1416.
- 8. Talo, M. (2022). Automated classification of histopathological images using deep learning and data augmentation. Applied Sciences, 12(3), 1372.
- AbdulJabbar, K., Raza, S. E. A., Rosenthal, R., Jamal-Hanjani, M., Veeriah, S., Akarca, A. U., ... & Turajlic, S. (2023). Pathway-based attention neural networks identify cancer-specific immune archetypes. Nature Cancer, 4(2), 256–270.
- N. N. Saranya, A. Alkhayyat, S. A. V, R. M. L, and T. Jayachandran, "Contrastive Loss Function Siamese Network for Change Detection on Remote Sensing Images," in 2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2024, pp. 1-4.
- R. M. Naga, T. Jayachandran, L. H. Alzubaidi, D. Mazumder, and H. D. Praveena, "Residual Network Based Bidirectional Gated Recurrent Unit for Speech Recognition Using Speech Signals," in Proc. Int. Conf. on Distributed Computing and Optimization Techniques (ICDCOT 2024), 2024.
- T. Jayachandran and S. Kavitha, "A Development and Analysis of Color Image New Blind Watermarking Based n DWT-SVD Swapping and 3-Dimensional Cryptography Technique," Fluctuation and Noise Letters, vol. 22, no. 5, p. 2350026, 2023.
- T. Jayachandran, A. Logeswaran, S. Thillaikkarasi, B. Sakthikumar, and S. Karthick, "Energy-Efficient Clustering and Routing Algorithm using Deer Hunting Optimization in Wireless Sensor Networks," in Proc. 7th Int. Conf. on Electronics, Communication and Aerospace Technology (ICECA 2023), 2023, pp. 1522–1527.
- 14. J. Wang, Y. Zhang, and H. Li, "Deep Learning for Histopathological Image Analysis: A Review of Recent Advances and Challenges," Medical Image Analysis, vol. 92, p. 103058, 2024.
- 15. s. Kumar, R. Sharma, and A. Gupta, "Optimized EfficientNet for Lung Cancer Detection Using Histopathological Images," Journal of Medical Imaging, vol. 10, no. 3, pp. 033501, 2023.
- M. A. Rahman, T. Hossain, and F. Ahmed, "Attention-Enhanced Deep Learning for Colon Cancer Classification in Histopathology Images," IEEE Transactions on Biomedical Engineering, vol. 71, no. 5, pp. 1456–1465, 2024.
- V. Parameshwari, R. Jayanthi, M. Sudha, M. Srinevasan, S. Gladson, and O. C. Mathew, "Enhancing Patient Tracking with Advanced Deep Transfer Learning," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp. 868-872.
- V. Parameshwari, T. Abirami, M. Karthikkumar, V. Logeswari, T. Jayachandran, and G. Rathanasabhapathy, "Investigation of Healthcare Tracking System for Diabetic Patients Utilizing Machine Learning Algorithms to Integrate the Internet of Things," 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 2024, pp. 821-826.









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com