2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES) | 979-8-3315-3669-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCIES63851.2025.11032814

Enhancing Diagnostic Precision in Oncology: A Machine Learning Approach for Tissue Sample Classification

D Archana
Assistant Professor (Senior Grade)
Department of ECE
PSG Institute of Technology and Applied Research
archana@psgitech.ac.in

Dr S Kanthalakshmi
Professor
Department of EEE
PSG College of Technology
skl.eee@psgtech.ac.in

A Vennila
Department of ECE
PSG Institute of Technology and Applied Research
22l151@psgitech.ac.in

R Deeksha
Department of ECE
PSG Institute of Technology and Applied Research
22l11@psgitech.ac.in

P Harsha
Department of ECE
PSG Institute of Technology and Applied Research
22l121@psgitech.ac.in

S Neya
Department of ECE
PSG Institute of Technology and Applied Research
22l131@psgitech.ac.in

***Abstract*—Lung cancer significantly affects the respiratory system and is characterized by prodromes such as a prolonged cough, dyspnea, and angina. Smoking remains the leading risk factor for lung cancer. In contrast, colon cancer affects the digestive system. A Computed Tomography (CT) scan can provide valuable insights for diagnosing lung diseases. This study aims to utilize advanced ML and deep learning techniques to classify the histopathological images of tissue samples into various classes, determining whether the cells are cancerous based on distinct morphological patterns. These patterns may include irregular shapes, increased density, and inconsistent sizes compared to normal cells. The goal is to reduce the reliance on pathologists to make manual diagnoses, which can be both time-consuming and subjective. In the proposed approach, a comparative analysis is conducted on image classes that have been preprocessed and trained from scratch using various models namely Support Vector Machine (SVM), Convolutional Neural Networks (CNN), and XGBoost. A CNN based on the EfficientNetB0 architecture is employed, which is the most minor and most efficient version in the EfficientNet family. This model is pre-trained on the ImageNet dataset for feature extraction, omitting the top fully connected layer to enable transfer learning. As a result, this model achieves an overall accuracy of 99.54%, surpassing all other models tested.**

***Index Terms*—Histopathological images, SVM, CNN, XGBoost, EfficientNetB0**

I. Introduction

Cancer is a significant global health issue, with lung and colorectal cancers ranking among the most prevalent and fatal types. [1] Lung cancer accounts for 18.4% of worldwide cancer-related deaths, while colon cancer accounts for 9.2%. [2] The concurrent occurrence of these two cancers is rare; however, their high metastatic potential, particularly between the lungs and colon, emphasizes the vital need for early and precise diagnosis. Timely detection not only improves survival rates but also facilitates more effective management of these malignancies.

Traditional diagnostic approaches, such as imaging techniques and biopsy analysis, rely on expert pathologists to evaluate histopathological slides. Although these methods are effective, they are labor intensive, susceptible to human error and time consuming particularly when processing large volumes of data. The growing complexity and scale of medical data necessitate innovative solutions to complement and enhance traditional practices.

Machine learning models have demonstrated a remarkable ability to analyze complex patterns within histopathological images, delivering faster and more consistent results. Within this domain, deep learning (DL) [3], a subset of ML, has excelled in large-scale image analysis through architectures namely convolutional neural networks (CNNs) [4]. However, despite their high performance, deep learning models offers limited transparency in feature interpretation — an important factor in clinical decision-making. [5].

This study aims to address these challenges by developing a robust automated diagnostic support system for multi-class classification of histopathological images of lung and colon cancers. Our approach integrates traditional feature engineering methods, namely Support Vector Machines (SVM) [6] and XGBoost, with advanced deep learning models, including CNN and EfficientNetB0. [7] By combining these methodologies, the proposed system seeks to balance interpretability and accuracy while leveraging their strengths for cancer subtype classification. The ultimate goal is to enhance diagnostic precision and streamline the evaluation of histopathological

TABLE II: Comparison of Model Performance

Metric	CNN	XGBoost	SVM	EfficientNetB0
Overall Accuracy	95.60%	91.48%	73.56%	99.54%
Precision (Macro Avg)	0.96	0.92	0.74	1.00
Recall (Macro Avg)	0.96	0.91	0.74	1.00
F1-Score (Macro Avg)	0.96	0.91	0.74	1.00
Execution time	~9 sec/epoch	Faster than CNN	Fastest	Slower
Best Performing Class	High in all (especially "Lung N")	"Lung N"	"Lung N"	High in all
Worst Performing Class	Slight drop for "Lung ACA" (F1:0.91)	"Colon ACA" (F1:0.87)	"Lung ACA" (F1:0.64)	"Lung ACA" (F1:0.99)

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

C. Recall (Sensitivity)

One important indicator for evaluating the model's capacity to identify good cases is recall. Recall is frequently given priority in the medical field, as it shows the ability to recognize all pertinent examples, such as the identification of all malignant tissues. The ratio of real positive instances to all actual positive instances is known as recall [24].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

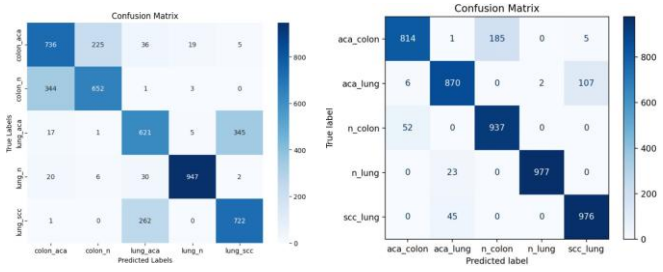
A higher recall indicates fewer false negatives, ensuring that most positive cases are correctly detected.

D. F1-Score

The F1-score provides a comprehensive evaluation of a model's effectiveness by calculating the harmonic mean of recall and accuracy. It is especially beneficial for imbalanced datasets [3], since it includes both false positives and false negatives in calculation. By balancing the trade-off between precision and recall, the F1-score offers a single, unified performance metric [24]:

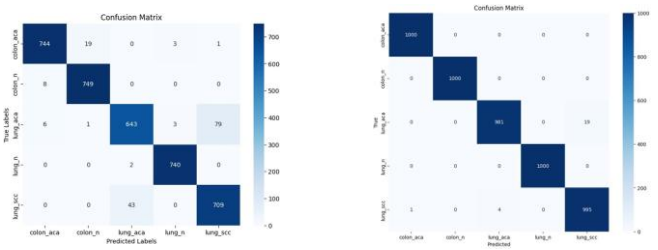
$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

A higher F1 score indicates a better trade-off between precision and recall, which is essential for tasks involving both false positives and false negatives carry crucial weight [24].



(a) (b)

Fig. 3: Confusion matrices of feature selection methods: a) SVM, b) XGBoost



(a) (b)

Fig. 4: Confusion matrices of feature selection methods: a) CNN, b) EfficientNetB0

V. Result

The performance of the four models — Convolutional Neural Network (CNN), XGBoost, Support Vector Machine (SVM), and CNN-based EfficientNetB0—was assessed using key standards such as accuracy, precision, recall, and F1-score. The outcomes are outlined in Table 2.

EfficientNetB0 outperforms CNN, XGBoost, and SVM in overall accuracy, achieving 99.54%, which is significantly higher than CNN's 95.60%, XGBoost's 91.48%, and SVM's 73.56%. It also maintained high F1-scores, with a peak of 1.00 for most classes except "Lung ACA" (0.99). CNN performed well, especially in the "Lung N" class with an F1-score of 1.00, while XGBoost showed moderate performance, excelling in the "N Lung" class (F1-score of 0.99) but struggling with "Colon ACA" (F1-score of 0.87).

SVM recorded the lowest accuracy at 73.56%. While it performed reasonably well in the "N Lung" class (F1-score of 0.96), the results are poor in "Lung ACA" (F1-score of 0.64) and "Colon ACA" (F1-score of 0.69), indicating challenges with complex datasets.

Regarding efficiency, CNN required approximately 9 seconds per epoch, making it the most computationally intensive. XGBoost was faster than CNN but more resource-intensive than SVM. EfficientNetB0 had a training time of around 5 minutes per epoch. SVM, with its simpler structure, was the fastest but traded off accuracy for speed.

VI. Conclusion

EfficientNetB0 is recommended for applications where accuracy is critical, particularly in medical diagnostics.CNN provides an optimal trade-off between accuracy and efficiency,

979-8-3315-3669-5/25/$31.00 ©2025 IEEE

making it well-suited for environments with moderate computational resources. XGBoost is appropriate for scenarios demanding faster execution with acceptable accuracy. However, SVM's poor classification performance makes it unsuitable for complex datasets.

Future research may focus on optimizing EfficientNetB0's training time while maintaining accuracy. Additionally, hybrid approaches combining CNN and XGBoost could be explored to capitalize on their strengths. Efficient hardware implementations and further hyperparameter tuning may also enhance model performance.

REFERENCES

[1] S. Mehmood, T. M. Ghazal, M. A. Khan, M. Zubair, M. T. Naseem, T. Faiz, and M. Ahmad, "Malignancy detection in lung and colon histopathology images using transfer learning with class selective image processing," *IEEE Access*, vol. 10, pp. 25657–25668, 2022.

[2] L. T. Omar, J. M. Hussein, L. F. Omer, A. M. Qadir, and M. I. Ghareb, "Lung and colon cancer detection using weighted average ensemble transfer learning," in *2023 11th international symposium on digital forensics and security (ISDFS)*, pp. 1–7, IEEE, 2023.

[3] R. D. Mohalder, J. P. Sarkar, K. A. Hossain, L. Paul, and M. Raihan, "A deep learning based approach to predict lung cancer from histopathological images," in *2021 international conference on electronics, communications and information technology (ICECIT)*, pp. 1–4, IEEE, 2021.

[4] F. Demir, A. M. Ismael, and A. Sengur, "Classification of lung sounds with cnn model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105376–105383, 2020.

[5] B. O. Soufiene and C. Chakraborty, *Machine learning and deep learning techniques for medical image recognition*. CRC Press, 2023.

[6] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (svm) learning in cancer genomics," *Cancer genomics & proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[7] A. Bala, A. H. Muqaibel, N. Iqbal, M. Masood, D. Oliva, and M. Abdullahi, "Machine learning for drone detection from images: A review of techniques and challenges," *Neurocomputing*, p. 129823, 2025.

[8] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic lung cancer prediction from chest x-ray images using the deep learning approach," in *2018 11th biomedical engineering international conference (BMEiCON)*, pp. 1–5, IEEE, 2018.

[9] A. Farhadipour, "Lung and colon cancer detection with convolutional neural networks and adaptive histogram equalization," *Iran Journal of Computer Science*, vol. 7, no. 2, pp. 381–395, 2024.

[10] S. Postavaru, R. Stoean, C. Stoean, and G. J. Caparros, "Adaptation of deep convolutional neural networks for cancer grading from histopathological images," in *Advances in Computational Intelligence: 14th International Work-Conference on Artificial Neural Networks, IWANN 2017, Cadiz, Spain, June 14-16, 2017, Proceedings, Part II 14*, pp. 38–49, Springer, 2017.

[11] R. Mormont, *Addressing Data Scarcity with Deep Transfer Learning and Self-Training in Digital Pathology*. PhD thesis, Universite de Liege (Belgium), 2022.

[12] R. Mormont, P. Geurts, and R. Mare´e, "Comparison of deep transfer learning strategies for digital pathology," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 2262–2271, 2018.

[13] M. Ali and R. Ali, "Multi-input dual-stream capsule network for improved lung and colon cancer classification," *Diagnostics*, vol. 11, no. 8, p. 1485, 2021.

[14] M. M. Hossain, M. R. Islam, M. F. Ahamed, M. Ahsan, and J. Haider, "A collaborative federated learning framework for lung and colon cancer classifications," *Technologies*, vol. 12, no. 9, p. 151, 2024.

[15] J. S. Syal, A. Jain, A. K. Dubey, and V. Jain, "Improving lung and colon cancer detection using ensemble method approach," in *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1767–1773, IEEE, 2024.

[16] D. Albashish, "Ensemble of adapted convolutional neural networks (cnn) methods for classifying colon histopathological images," *PeerJ Computer Science*, vol. 8, p. e1031, 2022.

[17] E. Deniz, A. S¸engür, Z. Kadirog˘lu, Y. Guo, V. Bajaj, and U¨ . Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health information science and systems*, vol. 6, pp. 1–7, 2018.

[18] M. Tog˘ac¸ar, "Disease type detection in lung and colon cancer images using the complement approach of inefficient sets," *Computers in Biology and Medicine*, vol. 137, p. 104827, 2021.

[19] M. Hasan, M. S. Rahman, S. Islam, T. Ahmed, N. Rifat, M. Ahsan, R. Gomes, and M. Chowdhury, "Vision transformer-based classification for lung and colon cancer using histopathology images," in *2023 International Conference on Machine Learning and Applications (ICMLA)*, pp. 1300–1304, IEEE, 2023.

[20] K. Zhao, Y. Si, L. Sun, and X. Meng, "Portnet: Achieving lightweight architecture and high accuracy in lung cancer cell classification," *Heliyon*, vol. 11, no. 3, 2025.

[21] A. Singh, S. Kumar, J. Gangrade, Y. P. Singh, and T. K. Dey, "Integrating convolutional and recurrent neural networks for lung and colon disease detection," *Neural Computing and Applications*, pp. 1–22, 2025.

[22] Q. Li, F. Li, J. Zhang, J. Ye, W. Song, and A. Mantooth, "Data-driven cyberattack detection for photovoltaic (pv) systems through analyzing micro-pmu data," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 431–436, IEEE, 2020.

[23] J. Sanyal, D. Rubin, and I. Banerjee, "A weakly supervised model for the automated detection of adverse events using clinical notes," *Journal of Biomedical Informatics*, vol. 126, p. 103969, 2022.

[24] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.