ColonScan Pro: An Automated Colon Cancer Classification System Using Deep Learning and Explainable AI

Muhammad Ummar Hisham bin Ahmad Madzlan1, a) and Sabaa Ahmed Yahya Al-Galal2, b)

1, 2Faculty of Computing and Informatics, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia

*b) Corresponding author: algalalsabaa@mmu.edu.my*

*a) ummarhisham157@gmail.com*

**Abstract.** Early diagnosis is the most important means of improving survival rates for this major cause of cancer related morbidity and mortality globally. Colon cancer can be identified and diagnosed by images of colonoscopy and histopathology. Most of these images can be manually interpreted but it is costly in terms of time, labor, and prone to human error in case of subtle differences between benign and malignant tissues. In this project, we build an automated colon cancer detection and classification system using CNN and XAI. It is a system designed to assist medical professionals in the automated analysis of histopathological images for early cancer detection and tackling problems of manual interpretation that can be very time consuming and prone to human error. For the proposed system, two state-of-the-art CNN architectures, namely EfficientNetV2, MobileNetV2 and ResNet-50 are implemented for image classification. In order to increase trust in the model when it is used in a clinical setting, the system is provided with the Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual explanations of how the model comes to its conclusions. There is a user-friendly interface for medical professionals where technicians have one workflow, and doctors have another. Techs can upload histopathological images; they are processed by technicians and doctors can review diagnostic reports with AI generated insights — or they can add professional annotations. Reliable diagnosis is tracked by means of performance metrics such as accuracy, precision, sensitivity and AU-ROC score. This work demonstrates that with the combination of CNN and XAI techniques, we can have an accurate, interpretable, and efficient tool to detect colon cancer, which will help medical professionals make evidence-based diagnostic.

# Introduction

Colon cancer is a large global health problem, constituting about 10% of all cancer cases and is the second most deadly cancer in the world. In 2020, 1.9 million people developed colorectal cancer, and 935000 died from this cancer, according to World Health Organization, and it is anticipated that by 2050 over 2 million people will have colorectal cancer [1]. These alarming statistics highlight the urgent need for early detection which improves a patient’s chances of survival and a better outcome for treatment.

Currently, colon cancer diagnosis relies on expert human interpretation of histopathological examination and colonoscopy. These methods may be labor-intensive, time consuming, and recognizable to human error [2]. CNNs are powerful tools for automating histopathological image diagnosis, accuracy can reach up to 99% in classifying colon malignancies [3]. In addition to speeding things up, CNNs also reduce the workload of the physician. Yet, most clinicians are unable to interpret the decision processes of CNNs. It undermines trust and prevents the integration of AI-driven tools into clinical workflows. To account for this, XAI tools like Grad-CAM have been introduced so that through visual explanations, visualize regions of the image that drive the model decisions to help develop more trust amongst medical professionals [4].

Manual diagnostic approaches are not scalable to meet the growing demand of early detection in large populations and difference in tissue morphology can lead to diagnostic variability [5]. However, CNNs can automate and standardize the interpretation of images, but lack interpretability [2]. Therefore, it’s understandable that clinicians are hesitant to use AI systems for high stakes decision-making without understanding how they make decisions. Therefore, highly accurate and interpretable diagnostic tools are needed that are trustworthy and can be used in clinical practice.

The goal of this project is to design a robust CNN based model for colon cancer image classification, whether the image is benign or malignant, and include XAI methods to make the model transparent and reliable. The key research questions are to identify the most effective CNN models for medical image analysis, to optimize model accuracy, and to leverage XAI for model reasoning explanation, and to validate the model’s reliability.

In order to reach these goals, the project will develop an application that automatically classifies colon cancer images with a hybrid approach that combines CNNs and XAI. The methodology involves collecting and preparing a high-quality dataset of histopathological images, developing and optimizing various CNN architectures and integrating XAI methods to generate visual explanations for each classification. A user-friendly interface that presents both diagnostic results and visual explanation will be obtained from the resulting application, which will be easy to integrate into clinical workflows. Pathologists, clinicians, hospitals, diagnostic centers and medical researchers are all the target audience that can benefit from more accurate, efficient and interpretable diagnostic tools for colon cancer.

# LITERATURE REVIEW

Medical imaging technologies like X-rays, CT scans, MRI, PET, and ultrasound help to identify tumors in patients at early stages and subsequently improves the outcomes of the patients. Since the integration of AI and deep learning further promotes this field, rapid and highly accurate tumor identification and classification has been made possible. Real time polyp detection rates of the CNN based systems for colonoscopy tests range from 93.6 to 96 percent and less experienced endoscopists can detect as much as 80 percent of the adenomas as they would normally [6]. In particular, CNNs are very effective in detection rates in colonoscopy and histopathological image analysis. For example, lightweight CNNs now surpass 99% accuracy in distinguishing adenocarcinoma from benign tissue, and real time polyp detection systems have greatly reduced missed diagnoses to such an extent that they aid the less experienced endoscopists [3]. Similarly, deep learning systems can view tissue images and reliably assist pathologists in determining if the tissue is healthy or cancerous and make diagnosis that is more accurate and consistent [7].

While these advances have been made, deep learning models are still complex and fail to be adopted clinically because they are still ‘black boxes’ in nature. In attempts to make the models more explainable, researchers such as Grad-CAM and SHAP were developed to show users how their models come to their decisions. The ability of these methods to support clinical practice is clear but their use in real-world medical tools needs improvement to build better relationships between AI solutions and medical professionals [8]. XAI techniques such as Grad-CAM, SHAP and LIME introduce visual and quantitative explanations of model predictions [9]. They then create heatmaps or feature importance scores that allow clinicians to understand and verify the AI’s decision-making process [4]. Such transparency encourages more trust and helps embedding of AI tools in everyday clinical workflows since clinicians can compare AI provided results with their own expertise.

However, lack of large, diverse, and well annotated datasets restricts model training and generalization, and computational demands prevent deployment of advanced models in resource constrained healthcare environment. Even class imbalance in the datasets can distort model performance, in its worst possible predicate way, when models ignore rare but critical events such as early cancer stages [10]. Additionally, EfficientNet and Inception-V3 show excellent performance but need powerful computing equipment which smaller healthcare settings cannot afford [11].

To solve this problem, the latest research explores the development of lightweight and efficient CNN architecture, with the latest version ResNet-50 and EfficientNetV2, that still has a high accuracy but with a lower computation. Model robustness and generalizability are also enhanced by using transfer learning and data augmentation strategies [12]. These frameworks require, however, the integration of XAI methods aiming at interpretability and at reaching clinical acceptance [4]. It is shown that in addition to improving diagnostic performance, building diagnosing CNN models with XAI techniques closes the gap between AI systems and medical staff by making predictions more transparent and trustworthy.

# methodology

In this section we will be discussing the LC-25000 dataset that will be used to train and validate the CNN models, the architecture of the CNN models used in our project and the XAI technique used to provide explainability of the CNN models.

## LC-25000 Dataset

The LC25000 dataset is a collection of histopathological images for lung and colon cancer research. It contains 25,000 images, 768 x 768 pixels, divided into five distinct classes. Of these, two classes are specifically about colon tissue, which are necessary for cancer detection. The first class is Colon Adenocarcinoma (malignant) with 5,000 images of malignant tissue. About 95% of all colon cancer cases are this type of cancer, which is characterized by the presence of irregular glandular structures, nuclear pleomorphism, and stromal invasion in the histopathological images. The second class is Benign Colon Tissue, a class which also has 5,000 images, that represents non-cancerous tissue. These are benign images, and they are used as a control group to help distinguish between healthy tissue and malignant features. These classes, used together, are a beneficial resource for developing and testing diagnostic tools in colon cancer research [13].

## CNN Models

EfficientNetV2 uses Fused MBConv blocks in early layers, fusing depthwise and pointwise convolutions into a single 3×3 convolution to reduce memory overheads and accelerate processing. Traditional MBConv blocks leverage smaller 3×3 kernels and lower expansion ratios to balance accuracy and computational cost by handling deeper features through expansion, depthwise convolution, and projection steps. To reduce memory usage, the architecture omits the last stride-1 stage of EfficientNetV1, and to support lightweight or high-accuracy applications, progressive scaling of width, depth, and resolution is applied across variants (S/M/L). Global average pooling and a softmax activated fully connected layer is used for classification. Through training-aware neural architecture search, and in combination with these innovations, we achieve up to 11× faster training, and 6.8× better parameter efficiency.

MobileNetV2 introduces two key innovations: We use linear bottlenecks and inverted residual blocks which decrease computational costs without hurting performance. The architecture starts with downsampling input, then a sequence of inverted residual blocks. Input channels are expanded with 1×1 convolutions, spatial filtering is done with depthwise convolutions, and linear bottlenecks (without nonlinear activation to preserve important information) project back to a lower dimensional space. We add shortcut connections when input/output dimensions match to help gradient flow. The last layers of feature aggregation, global average pooling, and a softmax activated fully connected layer for classification with 1×1 convolutions. The model uses depthwise separable convolutions and ReLU6 (for low precision stability) to minimize parameters, which is suitable for resource constrained environments such as mobile devices. It is used in image classification, object detection, and semantic segmentation.

ResNet-50 Residual learning with skip connections and bottleneck blocks allows training of very deep networks and addresses the vanishing gradient problem. In the 50-layer model, bottleneck blocks are employed, which have three consecutive convolutions: 1×1 to reduce the dimensions, 3×3 to filter the spatial information, and 1×1 to expand the channels. The novelty is shortcut connections which append the input directly to the block output, enabling the network to learn residual functions F(x) = H(x) - x instead of full mappings H(x). This makes optimization easier and allows much deeper networks. The last layers consist of the global average pooling and the fully connected layer with the softmax activation. In contrast to mobile architectures, ResNet-50 employs normal convolutions and is, therefore, computationally expensive yet precise. It is strong at learning hierarchical complex features, and is popular in image classification, object detection, semantic segmentation, and transfer learning.

## Grad-CAM

The goal of XAI is to increase the transparency and interpretability of the AI systems to its users. One such popular XAI technique for CNNs is Grad-CAM. With this method we can visualize where input images are the most important areas for the model decision making process. To generate the coarse localization map, the gradient information flowing into the last layer of a CNN is used to highlight the important regions in the image for the CNNs’ process prediction.

There are several steps to implement Grad-CAM in CNN. First, we must choose the target layer, which is usually the last convolutional layer of our CNN. Next, we use this Grad-CAM model to take the original inputs and then the chosen convolutional layer and final prediction. Next, we define hook functions to store the activations and gradients of the target layer. After we get the gradients of the predicted class score with respect to the feature maps of the selected convolutional layer, we get a heatmap by computing a weighted sum of the feature maps with the gradients.

# Results and discussion

Three CNN architectures were evaluated for the classification of colon cancer histopathological images: EfficientNetV2, MobileNetV2 and ResNet-50. The models are evaluated based on the accuracy, precision, sensitivity and AU-ROC score.

Based on Figure 1, EfficientNetV2 shows fast initial convergence, as the training and validation accuracy approaches 0.95-0.98 after the first 10-15 epochs, and the difference between training and validation accuracy is small indicating that the model has not overfitted much and hence good generalization. The loss curves show a sharp decrease during the initial epoch and the training loss as well as the validation loss converge to the range of 0.1-0.2 after epoch 20. In contrast, from Figure 2, MobileNetV2 exhibits great convergence behavior, as the training and validation accuracy curves are close to each other during the training process, and it reaches high accuracy (>0.95) comparatively early and demonstrates stable performance without significant overfitting. The loss curves show a steady decrease in both training and validation set and the validation loss is nearly equal to the training loss which shows strong generalization ability.

However, from Figure 3, the training process of ResNet-50 is the most stable and consistent among the three models, and the accuracy curves are steadily increasing, and training and validation measures are well-aligned, indicating possible perfect generalization without the risk of overfitting. The loss curves exhibit monotonic, smooth decline on both the training and validation sets, which shows stable optimization during training.

A graph of a training and validation accuracy

AI-generated content may be incorrect.A graph of training and validation

AI-generated content may be incorrect.

(a) (b)

**FIGURE 1.** Epoch trends of EfficientNetV2: (a) Epoch-accuracy plot, (b) Epoch-loss plot

A graph of a training and validation accuracy

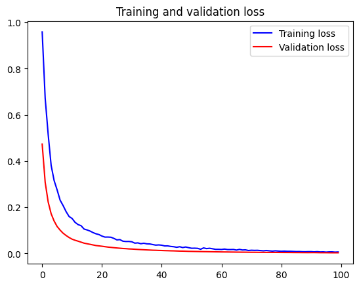
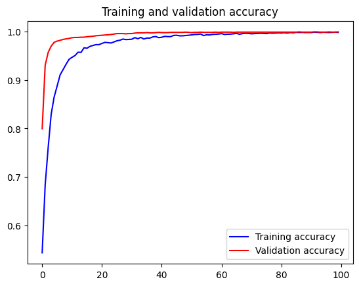
AI-generated content may be incorrect.A graph of training and validation

AI-generated content may be incorrect.

(a) (b)

**FIGURE 2.** Epoch trends of MobileNetV2: (a) Epoch-accuracy plot, (b) Epoch-loss plot

The trends on epochs demonstrate that the three models converge quite early with MobileNetV2 and EfficientNetV2 converging a bit faster in the early epochs than ResNet-50. All three architectures have strong generalization abilities according to the near parallelism of training and validation curves, but ResNet-50 has the most steady and stable training dynamics with the smoothest convergence trend. The small difference between training and validation scores of all models indicate that the transfer learning strategy with pre-trained weights is suitable to avoid overfitting and still achieve good classification performance to detect colon cancer.



(a) (b)

**FIGURE 3.** Epoch trends of ResNet-50: (a) epoch-accuracy plot; (b) epoch-loss plot

Table 1 shows outstanding performance in the three neural network structures, namely EfficientNetV2, MobileNetV2, and ResNet-50. The three models have incredibly high-performance measures with accuracy rates of 99.60%, 99.95%, and 99.95 percent respectively. Precision scores were also quite remarkable, with EfficientNetV2 displaying 99.60 percent, and both MobileNetV2 and ResNet-50 having 99.95 percent. The same tendencies were observed in sensitivity where EfficientNetV2 got 99.60% and both MobileNetV2 and ResNet-50 obtained 99.95%. First and most impressive, the AU-ROC (Area Under the Receiver Operating Characteristic curve) scores were excellent across all models with EfficientNetV2 scoring 99.99%, and MobileNetV2 and ResNet-50 getting a perfect 100% score.

**TABLE 1.** Performance of EfficientNetV2, MobileNetV2 and ResNet-50

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Sensitivity | F1-Score | AU-ROC |
| EfficientNetV2 | 99.60% | 99.60% | 99.60% | 99.60% | 99.99% |
| MobileNetV2 | 99.95% | 99.95% | 99.95% | 99.95% | 100% |
| ResNet-50 | 99.95% | 99.95% | 99.95% | 99.95% | 100% |

Such findings indicate that the three architectures worked remarkably well on classifying colon cancer, with MobileNetV2 and ResNet-50 demonstrating slightly better results than EfficientNetV2 on the majority of metrics. The almost 100 scores in terms of accuracy, precision, sensitivity, and AU-ROC show that the models performed very well in terms of correctly labeling the target classes and ensuring that the false positives and false negatives were minimal. The AU-ROC scores of 100% of MobileNetV2 and ResNet-50 especially indicate their high discriminative power, which implies that these models might have been able to discriminate between various classes in the dataset flawlessly. These high-performance values indicate the model architecture capabilities in classifying colon cancer.

Table 2 shows that after cross-validation, all the three models, i.e., EfficientNetV2, MobileNetV2, and ResNet-50, recorded a slightly different performance than they would have with only the training dataset, which highlights the necessity to test the robustness of the model on out-of-sample data. EfficientNetV2 showed the least decline and still had relatively high accuracy, precision, and sensitivity of 99.55 and AU-ROC of 99.99%, testifying to its relatively good generalizability. In comparison, MobileNetV2 and ResNet-50 had a more pronounced decrease, as the accuracy, precision, and sensitivity dropped to 98.67 percent and an AU-ROC of 99.79 percent. This performance shows that, although the initial performance was close to statistical perfection, cross-validation showed that EfficientNetV2 is less prone to overfitting than MobileNetV2 and ResNet-50.

**TABLE 2.** Performance of EfficientNetV2, MobileNetV2 and ResNet-50 after cros validation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Sensitivity | F1-Score | AU-ROC |
| EfficientNetV2 | 99.55% | 99.55% | 99.55% | 99.55% | 99.99% |
| MobileNetV2 | 98.67% | 98.67% | 98.67% | 98.67% | 99.79% |
| ResNet-50 | 98.67% | 98.67% | 98.67% | 98.67% | 99.79% |

Based on Figure 4, 5 and 6, Grad-CAM visualization was applied to a histopathology image to further interpret the model's predictions. The model is interested in meaningful areas. Seeing the blue-highlighted areas in the heatmap are the areas with a clear cell morphology and tissue architecture in the original histopathological image. The model seems to focus on the regions that have tightly crowded cells with hyperchromatic nuclei and abnormal cellular arrangement. These are major histological findings that pathologists consider in making diagnostic decisions. The fact that they are selectively highlighting these regions with unusual cellular properties, as opposed to attending background tissue or staining artifacts, indicates that the model has learned to attend to clinically relevant features.

A close-up of a microscope

AI-generated content may be incorrect.

**FIGURE 4.** Visual saliency image generated by Grad-CAM; EfficientNetV2

A close-up of a microscope

AI-generated content may be incorrect.

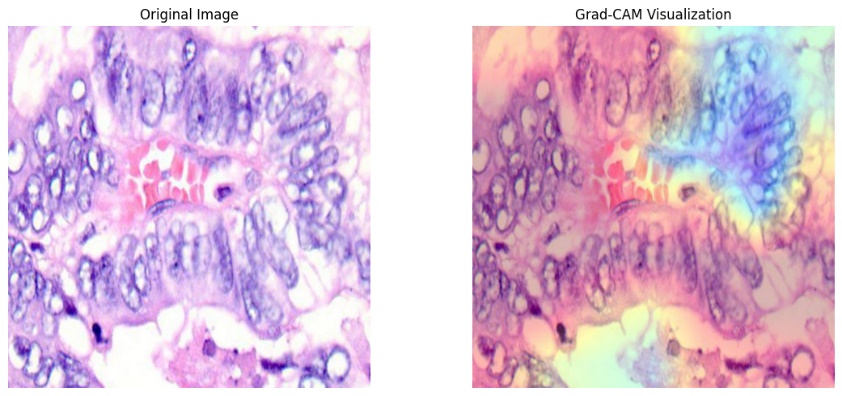
**FIGURE 5.** Visual saliency image generated by Grad-CAM; MobileNetV2

# CONClusion

The objective of the project was to meet the urgent need of early and precise colon cancer detection using an automated diagnostic system based on state-of-the-art CNNs and XAI techniques. Due to the realization of the disadvantages of traditional manual interpretation, such as time consuming, labor intensive, and vulnerable to human error, the most important goal was to build a robust, reliable, and interpretable tool that could assist medical personnel to analyze histopathological images. The project aims to implement and evaluate the two advanced CNN architectures, EfficientNetV2, and ResNet-50 for colon cancer image classification with high accuracy rates and computational efficiency that is suitable for clinical environments.

The goal of this work is to bridge the gap between the diagnostic performance of AI for these models and clinical trust of these models, by using XAI methods, especially Grad-CAM, to provide a visual explanation of each model decision. It allows clinicians to have confidence and evidence based medical practice. It results in an application that provides a user-friendly interface to both technicians and doctors to simplify the workflow of the process from image upload to automated analysis, expert review, and annotation.

Finally, this overall work demonstrates that this is feasible and has the potential value of combining highly efficient deep learning models with classifiable AI for cancer diagnostics. Not only is it good at solving one of the problems we have with manual image interpretation, but it is also an app that we can use in a practical way, and it provides a starting point for using and adopting more trustworthy AI tools in healthcare.



**FIGURE 6.** Visual saliency image generated by Grad-CAM; ResNet-50

# References

1. World Health Organization, "Colorectal cancer," WHO Fact Sheets, https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer.
2. M.I. Hasan, M.S. Ali, M.H. Rahman, and M.K. Islam, “Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks,” J Healthc Eng 2022, (2022).
3. M. Di Giammarco, F. Martinelli, A. Santone, M. Cesarelli, and F. Mercaldo, “Colon cancer diagnosis by means of explainable deep learning,” Sci Rep 14(1), (2024).
4. S. Tummala, S. Kadry, A. Nadeem, H.T. Rauf, and N. Gul, “An Explainable Classification Method Based on Complex Scaling in Histopathology Images for Lung and Colon Cancer,” Diagnostics 13(9), (2023).
5. H. Zhang, and K. Ogasawara, “Grad-CAM-Based Explainable Artificial Intelligence Related to Medical Text Processing,” Bioengineering 10(9), (2023).
6. N. Georgiou, P. Kolias, and I. Chouvarda, “Machine Learning Models for the Classification of Histopathological Images of Colorectal Cancer,” Applied Sciences (Switzerland) 14(22), (2024).
7. A.A.A. Mohamed, A. Hançerlioğullari, J. Rahebi, R. Rezaeizadeh, and J.M. Lopez-Guede, “Colon Cancer Disease Diagnosis Based on Convolutional Neural Network and Fishier Mantis Optimizer,” Diagnostics 14(13), (2024).
8. L.A. Gabralla, A.M. Hussien, A. AlMohimeed, H. Saleh, D.M. Alsekait, S. El-Sappagh, A.A. Ali, and M. Refaat Hassan, “Automated Diagnosis for Colon Cancer Diseases Using Stacking Transformer Models and Explainable Artificial Intelligence,” Diagnostics 13(18), (2023).
9. F. Binzagr, “Explainable AI-driven model for gastrointestinal cancer classification,” Front Med (Lausanne) 11, (2024).
10. E. Hamed, M. Tolba, N. Badr, and M.A.-M. Salem, “Large-scale Histopathological Colon Cancer Annotation Model Using Machine Learning Techniques,” International Journal of Intelligent Computing and Information Sciences 23(3), 73–82 (2023).
11. 10 A. Abdelrahman, and S. Viriri, “EfficientNet family U-Net models for deep learning semantic segmentation of kidney tumors on CT images,” Front Comput Sci 5, (2023).
12. M.S.N. Raju, and M.S.B. Srinivasa Rao, “Colorectal Cancer Disease Classification Using Mobilenetv2 Based on Deep Learning,” International Journal of Software Innovation 10(1), (2022).
13. A.A. Borkowski, M.M. Bui, L. Brannon Thomas, C.P. Wilson, L.A. Deland, and S.M. Mastorides, *Lung and Colon Cancer Histopathological Image Dataset (LC25000)* (n.d.).