Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

AI4CRC: A Deep Learning Approach Towards Preventing Colorectal Cancer

Samuel Fanijo *

Department of Computer Science, Iowa State University, USA; e-mail: sfanijo@iastate.edu
* Corresponding Author : Samuel Fanijo

Abstract: Each year, over 1.9 million cases of colorectal cancer (CRC) are diagnosed globally, with projections reaching 3.2 million cases and 1.6 million deaths annually by 2040. CRC ranks as the third most common cancer, contributing to over 10% of new cases with a 5-year survival rate of only 65%. Effective early detection could raise this rate to 90%, but current gold-standard methods are manual and prone to errors. To address this, our study introduces AI4CRC, a deep learning-based framework for automating the detection of polyps, the tumors responsible for colorectal cancer, in the human colon. We trained state-of-the-art models—including VGG, ResNet, DenseNet, and EfficientNet—and evaluated them on a publicly available dataset. Our approach leverages fine-tuning, various activation functions (ReLU, ELU, PReLU, Mish, Swish), and optimizers (Adam, RMSProp) to optimize performance. DenseNet and EfficientNet outperformed others, achieving 99% accuracy and a 99.4% F1 score. These results validate the potential of deep learning to enhance colorectal cancer detection, improving diagnosis accuracy and patient outcomes.

Keywords: Colonoscopy; Colorectal Cancer; Deep Learning; Digital Health; Transfer Learning.

1. Introduction

Colorectal cancer (CRC) is the third most prevalent type of cancer, representing approximately 10% of all cancer cases, with more than 1.9 million cases diagnosed each year, worldwide[1]. Colorectal cancer, caused by uncontrolled cell proliferation (or polyp) resulting from genetic mutations, manifests in the colon within the large intestine of the human body (Fig 1). Its primary contributing factors include old age, excessive consumption of processed food, alcohol abuse, smoking, and hereditary[2].

As illustrated in Fig 2, a polyp is an abnormal tissue growth that protrudes from a mucous membrane. Early detection is critical, as polyps can progress to colorectal cancer, significantly impacting the human life expectancy. The cumulative risk of cancer developing in an unremoved polyp is 2.5% at 5 years, 8% at 10 years, and 24% at 20 years post-diagnosis. The 5-year survival rate for polyp-induced cancer is 65%[1], but with early detection mechanisms, this rate can improve to as high as 90%[3].

Early, precise endoscopic excision of precancerous lesions is considered the most effective method for preventing colorectal cancer[4]. While colonoscopy is widely regarded as the gold standard for conducting this excision of precancerous lesions, its manual implementation process by gastroenterologists, who are often overworked, is linked with an increased rate of missed cases[5]–[7]. Moreover, the utter reliance on human expertise could also introduce inconsistencies in polyp detection, especially in the early stages when the polyps are small and less detectable. This variability highlights a significant gap in the current diagnostic approach, which could be bridged through the integration of artificial intelligence (AI) and digital health technologies to improve the efficiency of detection and reduce the incidence of missed cases of polyps during colonoscopy[8], [9].

Received: August, 25th 2024
Revised: September, 4th 2024
Accepted: September, 15th 2024
Published: September, 22nd 2024
Curr. Ver.: September, 22nd 2024

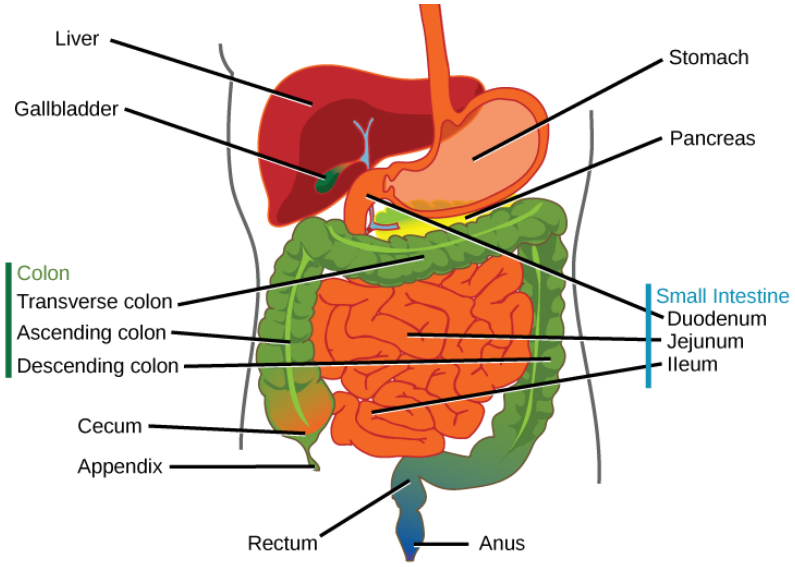

Figure 1. The Human digestive system and location of the colon [10].

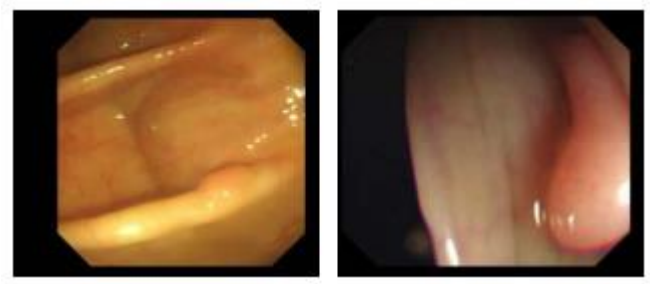

Figure 2. Sample images of polys[11].

Machine learning-based methods have become essential tools in automating disease diagnosis within the digital health domain. Popular approaches include Decision Trees, Naive Bayes, Support Vector Machines, and Random Forests[12], [13]. However, many of these systems rely on preprocessed features for training, which can result in missed opportunities to learn the most critical features necessary to effectively identify polyps[14], [15]. This limitation hinders the adoption of automatic systems in real-world clinical practice. Furthermore, while some existing works have explored deep learning-based approaches[9], [14], [15], they often lack the high accuracy required for robust colorectal cancer prediction. Therefore, developing more accurate AI approaches that can directly learn deep insights and features from diagnostic images is crucial for providing more reliable and efficient diagnostic recommendations for clinical experts.

To address these challenges, we introduce AI4CRC (Artificial Intelligence for Colorectal Cancer), a method designed to leverage deep learning and transfer learning to improve the detection of polyps, which are precursors to colorectal cancer, in colonoscopy images. Deep learning, particularly through Convolutional Neural Networks (CNNs), has shown significant success in clinical decision support tools via computer vision [16], [17]. Unlike traditional neural networks, CNNs can learn complex features from images using a series of convolutional layers with small-sized kernels that apply weights to input data and pass them through an activation function. This approach enhances the model's learning efficiency, leading to more accurate results.

Furthermore, our approach also leverages transfer learning, a method that enables the use of pre-trained models on new, but related tasks, to improve the efficiency and accuracy of our system. This is because while deep learning approaches have shown promise in improving detection rates, they frequently rely on large, annotated datasets, which are not always available. This addition of transfer learning differentiates our method from several existing works in literature.

Transfer learning is particularly advantageous in the context of colorectal cancer detection due to several reasons:

1. Limited Annotated Data: Colorectal polyp detection datasets are often limited in size and annotation due to the invasive manner of the colonoscopy procedure, which poses a challenge for training deep learning models from scratch. Transfer learning mitigates this issue by allowing us to fine-tune pre-trained models that have already learned generalized features from large-scale datasets, such as ImageNet [18]. These pre-trained models can then be adapted to the specific task of polyp detection, requiring fewer annotated samples to achieve high accuracy[19].
2. Faster Convergence: Starting with a pre-trained model enables faster convergence during training, as the model has already learned useful features and only needs to adapt these to the specific domain of colonoscopy images. This reduces the overall training time and computational resources required, making it feasible to develop an accurate model even with limited access to high-performance computing resources[20].
3. Improved Generalization: By leveraging knowledge from related tasks, transfer learning enhances the model's ability to generalize to new, unseen data[21]. This is particularly important in medical applications where variability in image quality, patient demographics, and disease manifestation can affect model performance. Transfer learning helps in building a more robust model that can maintain high accuracy across different clinical scenarios[22].

This use of transfer learning not only differentiates our method from several existing works in the literature but also directly addresses the challenges posed by limited data availability and the need for efficient model training. Hence, in this work, we employ deep learning and transfer learning approaches intending to develop a fast and intelligent diagnostic system that can be efficient and effective in aiding the prevention of colorectal cancer. In summary, the key contributions of this work are:

1. Development of an intelligent diagnostic system for colorectal cancer: We present AI4CRC, a deep learning-based system that automates the detection of colorectal cancer, leveraging state-of-the-art deep learning models such as VGG, ResNet, DenseNet, and EfficientNet.
2. Model Optimization through Transfer Learning: Our approach incorporates transfer learning, allowing pre-trained models to be adapted for polyp detection, thereby improving the system's efficiency and accuracy in predicting colorectal cancer, even with limited annotated datasets.
3. Improvement in Diagnostic Accuracy: Our work demonstrates substantial improvements in accuracy, in prediction of colorectal cancer, compared to existing benchmarks in literature compared.

By developing an intelligent diagnostic system, we aim to improve the early detection of colorectal cancer, ultimately contributing to better clinical outcomes and patient care. The rest of our article is arranged as follows: Section 2 discusses the related works. In Section 3, we present the proposed methods and materials. Section 4 discusses our experimental setup and results. We compare our method to existing works in Section 5; and we wrap up the article in Section 6.

2. Related Works

Colonoscopies play a crucial role in identifying precancerous polyps and abnormalities that may develop into cancer. Most early research works on polyp and colorectal cancer detection focused on analyzing image features using image processing techniques. In the context of image processing, features are individual characteristics of an image that are tied to the visual primitives that make up an object (such as edges, colors, lines, etc.). Methods for detecting features in an image are often used to pinpoint regions of interest (ROIs), which are subsequently described using a single or a set of features. Shape, texture, and color are the most prevalent subcategories of feature descriptions[23]. More recently, over the last decade, as machine and deep learning methods have advanced, more efforts have also been focused on leveraging them to improve the accuracy of predictions for colorectal cancer. Overall, the current body of literature on classifying polyp/colorectal disease using image processing can be categorized into three main subfields. They are (a) approaches using shape descriptors, (b)

methods utilizing texture and color features, and (c) techniques based on machine/deep learning.

Shape descriptors: Shape descriptors have long been a fundamental approach in medical image analysis, particularly for detecting and classifying polyps in colorectal cancer. These algorithms focus on extracting the geometrical features of polyps, such as their size, contour, and curvature, which are critical in distinguishing benign from malignant growths. Early works in this domain utilized simple shape descriptors like circularity, aspect ratio, and area to detect anomalies in colonoscopy images. For example, research [24] developed a method that utilized polyp candidates' curvature and shape features to enhance detection rates in colonoscopy images. Their study demonstrated that shape-based features, when effectively extracted, could significantly improve the accuracy of polyp detection, although the reliance on manual feature extraction posed limitations in handling diverse polyp shapes.

More advanced approaches have employed statistical shape models (SSMs), which provide a more robust framework for capturing the variability in polyp shapes. For example, in [25], the authors introduced a shape model-based algorithm that combined principal component analysis (PCA) with a statistical shape model to detect polyps with varying shapes and sizes. Their model showed improved performance over simpler shape descriptors, particularly in detecting polyps that deviate from standard shapes. However, the dependence on accurate segmentation remains a challenge, as errors in segmentation can lead to misclassification.

2.1. Text and Color descriptors

Textural and color-based descriptors represent another critical area in analyzing colorectal cancer images. These descriptors leverage the visual patterns and color variations within polyps to distinguish them from surrounding tissues. Early approaches in this field focused on gray-level co-occurrence matrices (GLCM) and color histograms to capture polyps' textural and color information. In [26], a texture-based approach using GLCM combined with color histograms was proposed to enhance polyp detection accuracy. This method was particularly effective in identifying subtle texture differences that are often overlooked by shape-based methods, making it a valuable complementary tool.

Furthermore, recent advancements have seen the integration of more sophisticated textural descriptors, such as Local Binary Patterns (LBP) and Gabor filters, which provide multi-scale texture analysis. In [27], the authors employed a combination of LBP and color histogram features to develop a hybrid model for colorectal cancer prediction. Their study highlighted the importance of multi-modal feature extraction, demonstrating that combining textural and color-based descriptors can significantly improve detection accuracy. Despite these advances, text and color-based descriptors often struggle with variability in image quality and lighting conditions, which can impact their robustness in clinical settings.

2.2 Machine and Deep Learning

The application of machine learning and deep learning methods in colorectal cancer prediction has gained significant attention in recent years, offering a data-driven approach to feature extraction and classification. Unlike traditional shape and textural descriptors requiring more labor-intensive manual feature extraction, machine learning models can directly learn key features from large datasets. Early machine learning models, such as Support Vector Machines (SVMs) and Random Forests (RFs), have been pivotal in classifying colorectal polyps. For example, in[28], an SVM-based model was developed that combined multiple features, including texture, shape, and color, to classify polyps in colonoscopy images. The model achieved a classification accuracy of 72%, demonstrating the potential of machine learning in enhancing polyp detection. However, the study also highlighted limitations in its reliance on manually extracted features, making it less adaptable to large datasets. In another study [29], a Random Forest classifier was used to analyze colonoscopy images, integrating both textural and color descriptors to improve polyp detection rates. This approach yielded an accuracy of approximately 83%, showing that RFs can effectively handle the variability in polyp appearance. However, the reliance on handcrafted features for its training still limited the model's scalability and adaptability to new data.

Deep learning, particularly Convolutional Neural Networks (CNNs), has improved performance by automating feature extraction and achieving higher accuracy in polyp detection. For example, in [30], a CNN model designed for polyp detection was introduced and trained

on a dataset of colonoscopy images, and achieved an F1 score of 91%, outperforming traditional machine learning methods. Another noteworthy example is the work presented in [31], where a deep learning model was utilized to detect polyps from endoscopic images with an F1 score of 83.2%. Furthermore, in [32], the authors introduced the U-Net architecture, a CNN-based model designed for medical image segmentation, which has since become a popular model in the detection of polyps. Their model demonstrated high performance in identifying polyps with diverse shapes, sizes, and textures, surpassing traditional machine learning methods. However, it is desirable to have a model that can achieve up to 99+% in accuracy without necessarily expending a ton of computational resources.

Overall, while there has been significant progress in detecting and classifying polyps using traditional and machine/deep learning methods, there are still opportunities for improvement. Traditional methods often fall short in terms of accuracy and robustness, particularly when dealing with the inherent variability and noise in medical images. Moreover, while some studies have explored the use of deep learning for polyp detection, there remains a gap in leveraging transfer learning effectively to enhance detection accuracy without the need for extensive dataset-specific training, leading to high computational resource complexity.

This study seeks to address these gaps by applying transfer learning to fine-tune pre-trained deep learning models for the early detection of polyps in colonoscopy images. Our approach focuses on improving detection accuracy and generalizability while also minimizing the need for large, annotated datasets. By doing so, we aim to create a more reliable and accessible tool for early colorectal cancer detection, which could significantly enhance patient outcomes and reduce the global burden of this disease.

3. Materials and Method

This study proposes a deep learning-based approach to enable and improve the detection of polyps in colonoscopy images for the early detection of colorectal cancer – with a focus on also leveraging transfer learning to improve model accuracy and generalizability. Fig 3 provides a comprehensive overview of the proposed methodology, outlining the sequential steps from data acquisition and preprocessing through to model training, including transfer learning, and final model evaluation.

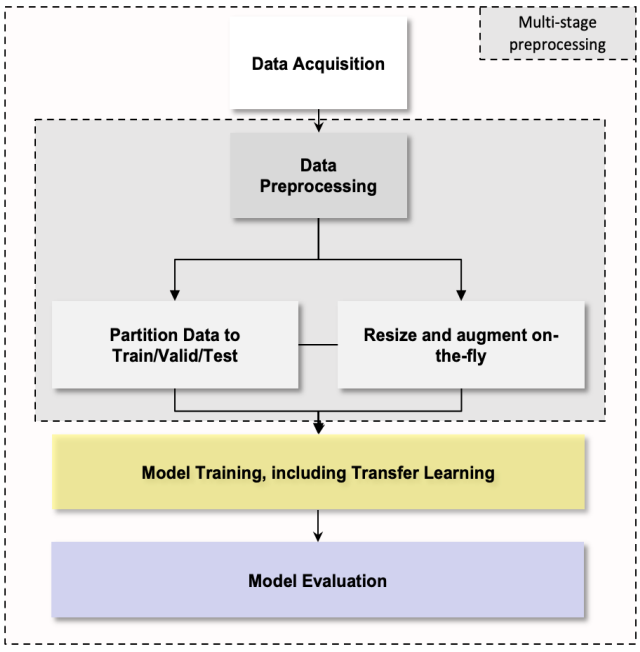

Figure 3. Overview of the proposed method

3.1. Dataset Acquisition

In this study, we use an open-source dataset released by the Computer Vision Center[11] for model training. The dataset comprises real-life patient data with two distinct classes: polyps (positive class) and non-polyps (negative class), both of which are completely labeled.

Due to the high cost of obtaining images for most medical datasets, particularly for colonoscopy, which is a very invasive procedure, the amount of available training data is limited, consisting of approximately 1212 images, as shown in Table 1, each representing a single patient. However, we address this limitation by utilizing transfer learning and implementing on-the-fly data augmentation techniques during training. These approaches enable us to expand the dataset and optimize performance effectively.

Table 1. Dataset split-up.

Class	Training	Validation	Testing	Total
Polyp	424	91	91	606
Non-polyp	424	91	91	606
Total	848	182	182	1212

(a) (b) (c) (d)

Figure 4. Sample images from the dataset. (**a**) and (**b**) show colorectal polyps (positive class); while (**c**) and (**d**) show normal colonoscopy images without polyps (negative class)

Figure 4 illustrates examples from the dataset: Panels A and B represent images from the Polyps class, where colorectal polyps are clearly visible and highlighted. Panels C and D represent images from the non-polyps class, which show normal colonoscopy views without any visible polyps.

3.2. Dataset Preprocessing

Data preprocessing is a crucial step in preparing the raw colonoscopy images for input into the deep learning models. With CNNs' excellent ability to learn directly from image data, we do not need to worry about manually extracting or engineering features, unlike traditional methods. Hence, our data preprocessing process become more seamless. This study employs a multi-stage preprocessing pipeline, as described below:

3.2.1. Partition Data

The dataset is partitioned into three subsets: training, validation, and testing. The training set is used to train the model, the validation set is used to tune hyperparameters and avoid overfitting. The test set is reserved for the final evaluation of the model's performance. We used the "70:15:15" ratio (consistent with common practice in literature) to split our dataset into training, validation, and testing, respectively as shown in Table 1.

3.2.2. Resize and Augment On-the-Fly

To enhance the robustness and generalization of the model, the images are resized and undergo on-the-fly data augmentation during training. Our approach is similar to the augmentation techniques implemented in [33] and [34], where techniques such as random rotations, flips, zooms, and shifts are applied to simulate the variability in real-world clinical settings. This step helps the model become more resilient to variations in polyp appearance.

3.3. Model Selection and Transfer Learning

The core of our proposed method is the training of deep learning models as well as transfer learning. Transfer learning allows us to leverage pre-trained models, specifically those trained on the ImageNet dataset, which contains millions of images across thousands of categories. By importing these pre-trained weights, the models can benefit from the extensive feature extraction capabilities developed during training on diverse image data, which are then

fine-tuned for the specific task of colorectal polyp detection. Below is a description of the models explored in our experiments:

3.3.1. Linear Perceptron

The Linear Perceptron, introduced by Rosenblatt, is a simple neural network model with no hidden layers, used primarily as a baseline for comparison [35]. While it lacks the complexity needed for deep feature extraction, it provides a foundational benchmark for evaluating more sophisticated models.

3.3.2. Multi-layer Perceptron

This model extends the linear perceptron by adding one or more hidden layers, allowing it to capture more complex patterns. However, it is still limited in handling the intricate features present in image data compared to convolutional networks [35].

3.3.3. Logistic Regression

Another baseline model, logistic regression, is a statistical method used for binary classification tasks, which is useful for understanding the model's performance in a linear two-class setting.

3.3.4. Basic CNN

The Basic CNN model (see Figure 5) consists of several convolutional layers followed by pooling layers and fully connected layers. CNNs raised the bar to another level in image recognition tasks, with their core ability to directly identify and learn patterns from images [36]. This makes them more effective for image classification tasks compared to simpler models.

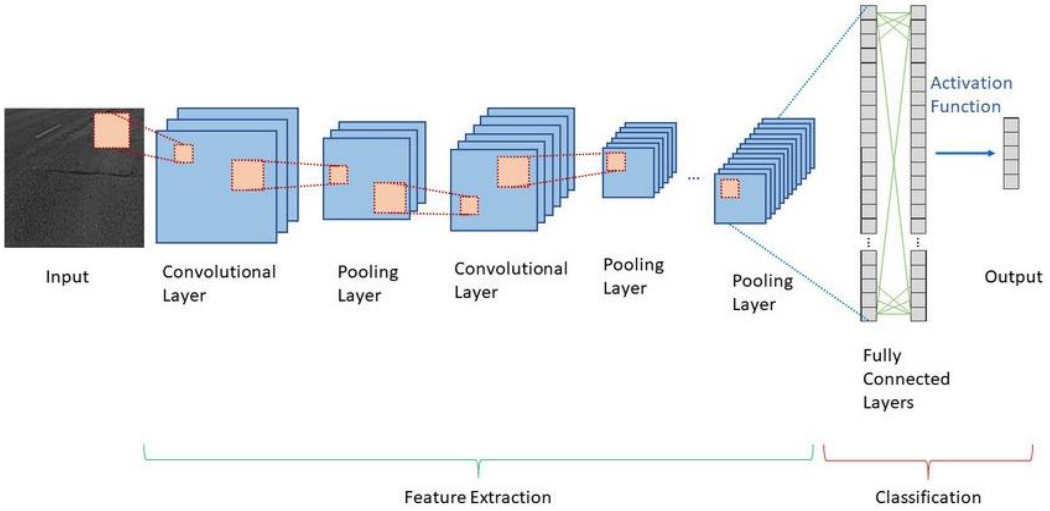

Figure 5. Basic CNN Architecture [Modified from [3]]

3.3.5. VGG19

VGG19 is a deep CNN with 19 layers, known for its uniform architecture that stacks convolutional layers with small 3x3 filters[37]. By importing the ImageNet weights, VGG19 benefits from strong feature extraction capabilities, particularly in identifying fine-grained details within the images, which are crucial for accurate polyp detection.

3.3.6. MobileNetV2

MobileNetV2 is an efficient CNN model designed for mobile and resource-constrained environments[38]. It uses depth-wise separable convolutions to reduce the number of parameters without sacrificing performance. The use of ImageNet weights enhances its ability to detect polyps in real-time applications, providing a good balance between accuracy and computational efficiency.

3.3.7. ResNetV50

ResNetV50 (50-layer Residual Network) is part of the ResNet family, which introduced the concept of residual learning to address the vanishing gradient problem in deep networks

[39]. By using ImageNet pre-trained weights, ResNetV50 can effectively recognize complex patterns and improve detection accuracy, especially in cases where polyps have irregular shapes or textures.

3.3.8. DenseNet121

DenseNet121 is a CNN that connects each layer to every other layer in a feed-forward approach, which helps in mitigating the vanishing gradient problem and encourages feature reuse[40]. The pre-trained ImageNet weights allow DenseNet121 to excel in detecting subtle features in colonoscopy images, making it a strong candidate for polyp detection tasks.

3.3.9. EfficientNetV2M

EfficientNetV2M is a scalable CNN model that optimizes both accuracy and efficiency [41]. It employs a compound scaling method to balance network depth, width, and resolution. By importing the ImageNet weights, EfficientNetV2M can leverage its efficient architecture to detect polyps with high accuracy while maintaining computational efficiency.

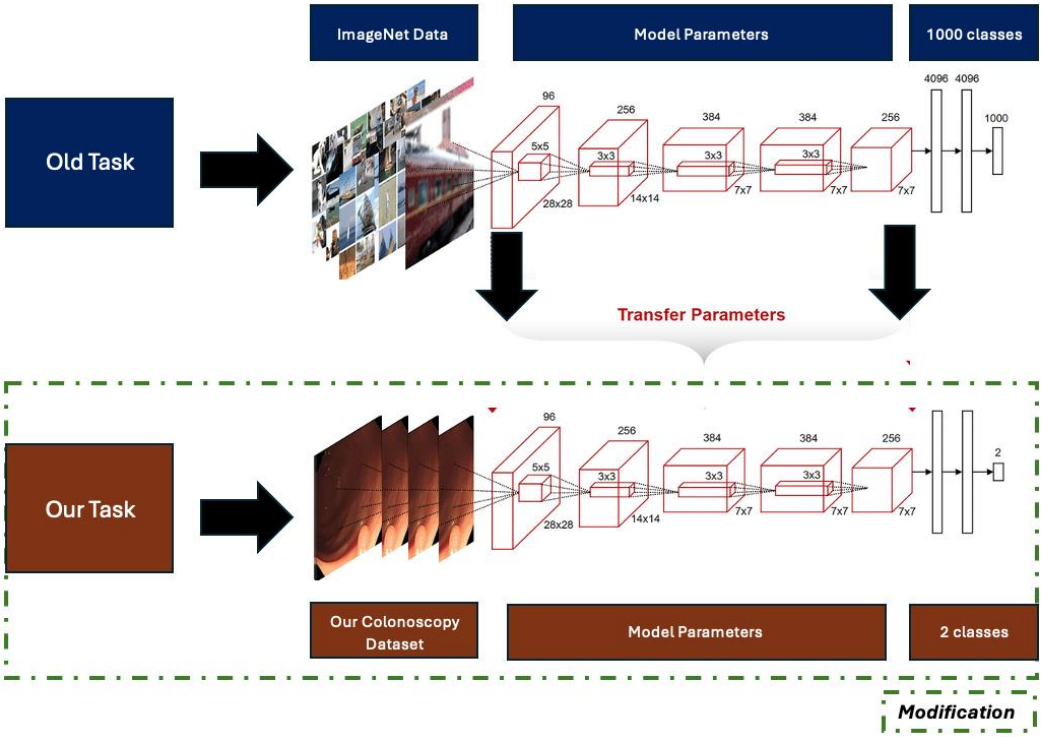

Figure 6. Transfer Learning Architecture (Modified from [42])

Each of these models—VGG19, MobileNetV2, ResNet50, DenseNet121, and EfficientNetV2M—was fine-tuned on the colorectal polyp dataset by retraining the final layers to adapt the pre-trained features to the specific task. It is important to clarify that the models, including the EfficientNetV2M, were not structurally modified but were fine-tuned. This means that the pre-trained base architecture was frozen, and only the final layers were retrained to suit the specific characteristics of the colorectal polyp detection task. Fine-tuning these models offers several advantages. First, it enhances the detection system's accuracy by leveraging the rich feature representations learned from large, general-purpose datasets like ImageNet. Second, fine-tuning significantly reduces the training time and computational resources required compared to training a model from scratch, which is particularly beneficial when working with limited annotated data, as is often the case in medical image analysis. A visual description of the transfer learning architecture is shown in Fig 6.

The top section of the figure shows the original task where the model was pre-trained on the ImageNet dataset, while the bottom section depicts our task, where the final layers were fine-tuned on the colorectal polyp dataset. The modifications made are marked in the figure, highlighting the transfer of parameters and the adaptation of the output layers from 1000 ImageNet classes to 2 classes specific to our task.

3.4. Evaluation Metrics

Following model development, we evaluated their performance on the test dataset using appropriate metrics. We focused on key metrics such as precision, recall, accuracy, and F1 score, shown in Equations 1 – 4, which are commonly used for classification tasks in literature. Other metrics such as AUC[43], [44] can also be used, but was not considered here as the F1 score, which harmonizes recall and precision, already provides a robust and comprehensive measure of the model's performance in this medical context.

The F1 score is particularly important in medical applications where false positives and negatives carry significant consequences. By balancing precision and recall, the F1 score offers a more nuanced evaluation of the model's ability to identify true cases while minimizing errors correctly, which is critical for ensuring patient safety and effective diagnosis[45]. This makes the F1 score a preferred metric in scenarios with high misclassification costs, such as in medical diagnostics.

Accuracy is the ratio of correctly identified samples to the total number of samples in a specific class. Recall measures the proportion of correctly identified samples for a particular class out of the total number of samples that actually belong to that class. Precision is the number of correctly classified samples for a specific class divided by the total number of samples classified as that class[23]. F1 Score combines precision and recall into a single metric by taking their harmonic mean. It is calculated as the weighted average of precision and recall, providing a balanced assessment of a classifier's performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1\,Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

4. Implementation and Results

4.1. Experimental Setup and Parameter Settings

In order to optimize the performance of the proposed model, we conducted two separate series of experiments – Stage-I and Stage-II. In Stage-I, the models used were entirely neural network-based, with the exception of the Logistic Regression model. For the Logistic Regression model, we reshaped the 3D image data (100x100 pixels with 3 color channels) into 2D arrays by flattening each image into a single vector of 30000 features (3*100*100). This transformation enabled the Logistic Regression model to process the image data in a format suitable for linear classification.

In Stage-I, we perform (global) training across all models, and Table 2 provides an overview of the parameter configurations used for training the models, which are critical for optimizing their learning performance. The logistic regression model requires no parameter selection as its implementation function is straightforward using the TensorFlow library. For the rest of the models in Stage-I experiments, we chose Binary Cross-Entropy as the loss function which is perfect for the two-class classification task, distinguishing between Polyps and Non-Polyps. We set the maximum number of training epochs to 50, with early stopping implemented to prevent overfitting by halting training when no significant improvement was observed in validation accuracy. We selected the Adam optimizer for its adaptive learning rate, efficiently adjusting model weights during training. By setting the learning rate to 0.0001, we aimed to strike a balance between optimization speed and stability, ensuring smooth convergence. We applied the ReLU activation function in the convolutional layers to introduce non-linearity, which enhances the model's capacity to learn intricate patterns from the images. To further refine the models, we employed a learning rate scheduler, dynamically adjusting the learning rate during training to optimize performance.

Table 2. Stage-I and Stage-II Experiments Parameter Settings.

Parameter	Stage-I Values	Stage-II Values
Loss	Binary Cross entropy	Binary Cross entropy
Optimizer	Adam	Adam, RMSProp
Activation	ReLU	ReLU, ELU, PReLU, Swish, Mish
Max Epochs	50	50
Learning Rate Scheduler	Yes. Minimum LR = 0.00001, Maximum LR = 0.001	Yes. Minimum LR = 0.00001 Maximum LR = 0.001

In Stage-II, we selected the top two performing models for further refinement through hyper-parameter tuning. This phase focused on adjusting several key parameters: (a) the learning rate, managed via a learning rate scheduler, (b) the activation functions, including ReLU, ELU, PReLU, Mish, and Swish, and (c) the optimizer, comparing the performance of Adam and RMSProp, as shown in Table 2 above. In all experimental setups (apart from the Logistic Regression model), we resized images to 100×100×3 pixels and applied image augmentation directly using TensorFlow's functions. We used Binary cross-entropy as the loss function across all models. Initially, we set the training to run for 50 epochs, but we actively managed this using early stopping and the reduce-learning-rate-on-plateau (ReduceLRR) function. We set the patience threshold to 5 epochs, allowing the training process to halt or adjust if it failed to improve validation accuracy within this period significantly. Through empirical observation, we found that the model peaked around epoch 25, after which the network began to overfit. To avoid overfitting and maintain model performance, our final epoch was set to 30, as further training showed no significant benefits. All models were developed using the Tensorflow-Keras library (in Python environment), run on a Tesla GPU.

4.2. Experimental Results and Discussion

In Table 3, we show the summary of Stage-I experimental results. We analyzed the performance of the models on the test set while also monitoring the training accuracy in case of any signs of overfitting. As expected, the transfer learning models outperformed the baseline models. The DenseNet and EfficientNet models achieved the highest accuracy and F1 score, both at 99%. A graphical summary of the F1 scores across all models is shown in Fig 7.

Training time varied significantly across the models. The Linear Perceptron, being the simplest, required only 3.6 seconds for training but yielded a training accuracy of just 46%, underscoring its limited capability in handling the complex task of polyp detection. In contrast, more advanced models like DenseNet121 and EfficientNetV2M demonstrated much higher training accuracy, though at the cost of significantly longer training times of 120 seconds and 306 seconds, respectively. These results highlight the trade-off between training time and model performance, with more complex models requiring more time to train but offering superior accuracy.

Table 3. Summary of Stage-I Experiments and Results.

Model	Training time (secs)	Training accuracy	Precision	Recall	F1 Score	Accuracy
Linear Perceptron	3.6	46	44	55	49	47
Multilayer Perceptron	32.8	47	46	100	63	47
Logistic Regression	1.6	50	53	64	39	49
Basic CNN	30.9	75	73	73	73	73
VGG19	161	89	87	88	87	88
MobileNetV2	94	98	96	96	96	96
ResNet50	130	99	98	98	98	98
DenseNet121	120	98	99	99	99	99
EfficientNetV2M	360	98	99	99	99	99

When examining precision, recall, and F1 score, which are critical metrics for evaluating the efficacy of models in medical diagnosis, the differences between the models become even

more pronounced. The Linear Perceptron, with a precision of 44%, recall of 55%, and an F1 score of 49%, performed poorly across the board, making it clear that this model is unsuitable for the task. The Multi-layer Perceptron showed a perfect recall of 100%, indicating that it could identify all true positives. However, its low precision of 46% resulted in an F1 score of 63%, reflecting a high false positives rate.

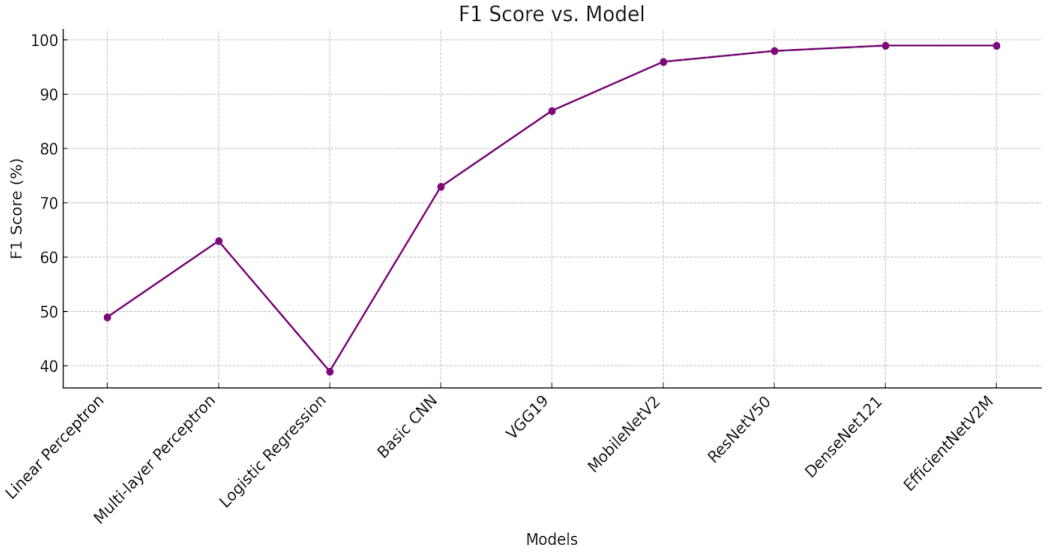

Figure 7. F1 Score performance of each model in Stage-I experiments.

The Logistic Regression model performed slightly better, achieving a precision of 53% and a recall of 64%, but it still struggled with an overall F1 score of 39%, indicating limited effectiveness. The Basic CNN model showed a marked improvement, with both precision and recall at 73%, resulting in a balanced F1 score of 73%. This suggests that while the basic CNNs are effective, deeper and more sophisticated architectures are necessary for achieving higher accuracy and reliability in polyp detection.

As expected, the transfer learning models all outperformed the baseline models. For example, the VGG19, MobileNetV2, ResNetV50, DenseNet121, and EfficientNetV2M all demonstrated superior performance. VGG19 achieved a precision of 87% and a recall of 88%, leading to an F1 score of 87%. MobileNetV2, which is designed for efficiency, showed an impressive balance between precision and recall, both at 96%, resulting in a high F1 score of 96%. This makes MobileNetV2 a strong candidate for real-time applications where computational efficiency is crucial.

ResNetV50 and DenseNet121 both exhibited outstanding performance with precision and recall rates of 98% and 99%, respectively, and corresponding F1 scores of 98% and 99%. These results underscore the models' ability to detect polyps with minimal errors accurately. EfficientNetV2M matched DenseNet121 with 99% precision, recall, and F1 score, though it required the longest training time. This suggests that while EfficientNetV2M offers the highest accuracy, its longer training time could be a limitation in time-sensitive environments.

In summary, the results indicate that advanced deep learning models, particularly ResNetV50, DenseNet121, and EfficientNetV2M, are highly effective for colorectal polyp detection, offering high accuracy and balanced precision-recall trade-offs. The final choice of model, however, may depend on the specific requirements of the deployment environment. While EfficientNetV2M provides the best overall performance, its longer training time might make models like MobileNetV2, which is both efficient and accurate, more suitable for resource-constrained settings. These findings highlight the potential of deep learning models in improving the early detection of colorectal cancer, with DenseNet121 and EfficientNetV2M emerging as strong candidates for clinical deployment.

In the Stage-II experiments, we focused on the top two models from Stage-I and tuned their hyperparameters to strengthen performance further. This phase was focused on optimizing the choice of optimizer, activation functions, and learning rate via a learning rate scheduler. We did not adjust the epochs and batch size in Stage-II, as the epochs were already

optimized during Stage-I. Tables 4 and 5 summarize the stage-II experiments and the results of model performance.

Table 4. Stage-II experiment: DenseNet121 Model.

Optimizer	Activation Function	Training Accuracy	Test Accuracy	Test F1 Score
Adam	ReLU	99	99.4	99.4
	Elu	100	99	99
	PReLU	100	99.4	99.4
	Mish	100	99.3	99.3
	Swish	100	99	99
RMSProp	ReLU	99	99.3	99.3
	Elu	100	99	99
	PReLU	100	99	99
	Mish	100	99.3	99.3
	Swish	100	99.3	99.3

Table 5. Stage-II experiment: EfficientNetV2M Model.

Optimizer	Activation Function	Training Accuracy	Test Accuracy	Test F1 Score
Adam	ReLU	100	99	99
	Elu	100	99	99
	PReLU	100	99.4	99.4
	Mish	99	99	99
	Swish	100	99	99
RMSProp	ReLU	100	99.4	99.4
	Elu	100	99.4	99.4
	PReLU	99.8	99.4	99.4
	Mish	99.7	99.3	99.3
	Swish	100	99.3	99.3

For DenseNet121, we experimented with two optimizers—Adam and RMSProp—combined with various activation functions, including ReLU, Elu, PReLU, Mish, and Swish. We chose to explore Adam and RMSProp as the optimizers for hyperparameter tuning due to their proven effectiveness in training deep learning models as reported in the literature[1], [3], [12] especially in scenarios with complex datasets like medical images. Under the Adam optimizer, all activation functions achieved high training accuracy, with ReLU and PReLU delivering the highest test F1 score of 99.4%, corresponding to a test accuracy of 99.4%. The RMSProp optimizer showed a slightly varied performance, with test F1 scores of 99% for ReLU and Elu, and 99.3% for PReLU, Mish, and Swish. However, the differences between these scores are minimal, reflecting the overall robustness of DenseNet121 across different configurations. Although RMSProp produced slightly lower test F1 scores than Adam, the difference is negligible (only 0.1%), suggesting that both optimizers are highly effective, with Adam offering a marginal advantage.

The hyperparameter tuning of EfficientNetV2M followed a similar approach. The Adam optimizer again paired effectively with PReLU, achieving the highest test F1 score of 99.4% and a perfect training accuracy. ReLU, Elu, and Swish also performed strongly, each attaining a test F1 score of 99%, showcasing EfficientNetV2M's reliability across various activation functions. With the RMSProp optimizer, Elu matched the 99.4% F1 score, while PReLU and Swish achieved 99.3%. As with DenseNet121, the slight variations in F1 scores between different configurations are minimal, indicating that EfficientNetV2M maintains strong generalization across a range of hyperparameters.

Overall, the results of the hyperparameter tuning in Stage-II experiments indicate that both DenseNet121 and EfficientNetV2M can achieve near-perfect performance with careful

selection of hyperparameters. Notably, the Adam optimizer combined with the ReLU and PReLU activation function for DenseNet121 produced the highest test F1 scores, meanwhile, the RMSProp optimizer combined with ReLU, Elu, and PReLU had the highest test F1 scores for EfficientNetV2M. This suggests that both models can perform exceptionally well under various configurations, but the minor differences observed, particularly in test F1 scores, highlight the importance of hyperparameter tuning in achieving optimal model performance. However, overall, the EfficientNetV2M model, with its consistently highest three activations under RMSProp optimizer (compared to the rest), reaffirmed its robustness.

Given these results after the two-stage experiments, EfficientNetV2M, with RMSProp optimizer and ReLU activation function emerged as our final model. But again, overall, the result shows that these two models are competitive, and the eventual choice will depend on the computational resources available at the point of deployment. This is due to the fact that the Stage-I result indicates that DenseNet's architecture is optimized for speed, resulting in faster training times. On the other hand, EfficientNet's architecture is geared towards accuracy optimization.

4.3. Impact of Stage-II Tuning

Stage-II tuning focused on optimizing specific hyperparameters such as the optimizer and activation functions for DenseNet121 and EfficientNetV2M. This stage provided a critical opportunity to refine the already strong models and push their performance slightly higher, specifically focusing on F1 score optimization, which is crucial in medical diagnostics where precision and recall are paramount.

The impact of Stage-II is particularly evident when examining the test F1 scores and training accuracies. While the Stage-I results already had DenseNet121 and EfficientNetV2M performing at a near-perfect level, Stage-II tuning allowed for a more nuanced exploration of hyperparameter settings. For instance, under Stage-II results in Table 4 and Table 5, we observed that the Adam and RMSprop optimizers combined with the PReLU activation function consistently yielded a test F1 score of 99.4% for both models, slightly edging out their Stage-I counterparts. These marginal gains might appear small, but they are significant in a clinical context where the highest possible accuracy is critical to patient outcomes.

Moreover, the Stage-II experiments confirmed the robustness of these models under various configurations, providing evidence that certain combinations of hyperparameters (like RMSProp with Elu or ReLU) can slightly improve model generalization without sacrificing training accuracy. The ability to fine-tune and achieve even minimal improvements highlights the critical role of hyperparameter optimization in leveraging the full potential of advanced models like DenseNet121 and EfficientNetV2M.

In summary, while Stage-I established a strong baseline, Stage-II fine-tuning was essential in squeezing out the last bit of performance from the models, ensuring that they are not only accurate but also well-tuned for generalization across diverse scenarios. This two-stage approach underlines the importance of iterative refinement in machine learning pipelines, especially in high-stakes applications such as medical diagnostics.

5. Comparison

With the EfficientNetV2M as our final model, which achieved both test F1 score and an accuracy of 99.4%, we demonstrated significant advancements in the task of colorectal cancer detection. As shown in Table 6, the performance of this model not only meets but surpasses the results of many existing works in the domain, highlighting the effectiveness of our approach.

Traditional machine learning models, such as SVMs and RFs, have been widely used in previous studies for polyp detection. For example, [28] achieved an accuracy of 72% using an SVM-based model that combined multiple features such as texture, shape, and color. Although effective, these models required manual feature extraction, making them less adaptable and scalable.

Similarly, [29] employed a Random Forest classifier to analyze colonoscopy images, achieving an accuracy of approximately 83%. While RFs can handle variability in polyp appearance, they still fall short of the generalization capabilities seen in deep learning models like EfficientNetV2M. Another study by [46] utilized a cascade of classifiers, including SVMs and decision trees, achieving an accuracy of 88%. This approach involved complex feature

engineering and model stacking, yet it did not reach the performance levels that our deep learning model could achieve with automated feature extraction. Furthermore, [47] explored a hybrid model combining Bayesian networks with traditional classifiers, achieving an accuracy of 84.5%.

Table 6. Comparison of our work with existing methods.

Existing Works	Method	Accuracy
Sasmal et al. [28]	SVM	72
Grosu et al. [29]	Random Forest	83
Tajbakhsh et al. [46]	Cascade of Classifiers (SVM, Decision Trees)	88
Dehghani et al.[47]	Hybrid Model (Bayesian Networks + Classifiers)	84.5
Ellahyani et al. [30]	Deep CNN	91
Urban et al. [48]	CNN in Real-time Colonoscopy	96.4
Öztürk & Özkaya [49]	CNN + LSTM	97.9
Ours	Deep + Transfer Learning: EfficientNetV2M	99.4

Despite these efforts, traditional models like these are often limited by the quality of manually extracted features and do not match our model's F1 score of 99.4%, which benefits from end-to-end learning capabilities. Furthermore, several CNN-based learning approaches in literature have shown promising results in polyp detection but were limited by their architectural simplicity. For example, [30] reported a deep CNN model achieving an accuracy of 91%. However, this performance still lags behind the results of more advanced models like ours, EfficientNetV2M. The authors of [48] also took a significant step by integrating a CNN-based system into real-time colonoscopy procedures, achieving an accuracy of 96.4%. This model was a considerable advancement, but our EfficientNetV2M model surpasses it with a higher F1 score, indicating better precision and recall balance. Another notable study by [49] employed a CNN with a Long Short-Term Memory (LSTM) network to analyze colonoscopy videos, achieving an accuracy of 97.9%. While this approach combined spatial and temporal features, it still fell short of the 99.4% accuracy achieved by our method, which demonstrates superior performance in static image analysis.

Finally, regarding clinical applicability, the high F1 score of 99.4% of our final model underscores its reliability in clinical settings, where the cost of misclassification is often high. Compared to existing works, which often tradeoff between precision and recall, our model's balanced performance ensures that clinicians can trust the results for making critical decisions during colonoscopy procedures.

6. Conclusion, Limitations, and Future Work

In this study, we have presented a comprehensive approach to early detection of colorectal cancer using advanced deep learning techniques. We evaluated various models through a series of experiments, ranging from traditional machine learning algorithms to state-of-the-art deep learning architectures. Among these, EfficientNetV2M emerged as the top performer, achieving a high test F1 score of 99.4%. This model demonstrated not only superior accuracy but also a balanced performance across precision and recall, making it highly reliable for clinical applications. The results of our work highlight the significant advancements that can be achieved through the application of deep learning in medical image analysis, particularly in improving diagnostic accuracy and patient outcomes in real-world medical settings.

While our study has shown remarkable results, several limitations should be acknowledged. The current study relied on a relatively limited dataset for training and validation, which could affect the model's generalizability across different clinical settings. Additionally, our model was tested primarily on static images, which may not fully capture the complexity and variability encountered in real-time diagnostics, such as video endoscopy procedures.

To further enhance and extend the capabilities of our approach, we propose the following directions for future research: 1) Expanding Dataset Diversity: One of the limitations of the current study is the reliance on limited datasets for training and validation due to the expensive nature of acquiring and labeling medical datasets. Our future work will focus on incorporating a more diverse set of images from different populations, sources, and imaging conditions. This will help further improve the model's generalizability across various clinical

settings and patient demographics. 2) Integration with Video Analysis: Extending the current model to handle video data rather than only static images could significantly enhance its utility in real-time diagnostics. We will explore this to capture temporal information in video streams and further guarantee high detection accuracy in dynamic scenarios. 3) Exploring Hybrid Approaches: Our future research will also explore hybrid models that combine the strengths of EfficientNetV2M with other architectures, such as Transformer models or ensemble methods. This could lead to further improvements in robustness, particularly in challenging cases where polyps are difficult to detect.

In conclusion, our study has demonstrated the potential of deep learning and transfer learning as a robust method to develop effective and accurate diagnostic tools for polyp detection, as part of efforts to prevent colorectal cancer – setting a new benchmark in the task. Furthermore, by addressing our outlined areas for future work, we can continue to push the boundaries of what's possible in medical image analysis, ultimately contributing to better patient care and outcomes through deep learning in artificial intelligence.

Funding: This research has no external funding.

Data Availability Statement: The dataset used is a public dataset which was published by [11]. The direct access to its official repository is available at: http://vi.cvc.uab.es/colon-qa/cvccolondb/.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] National Cancer Institute, "Cancer stat facts: Colorectal cancer," *SEER*. https://seer.cancer.gov/statfacts/html/colorect.html (accessed Aug. 01, 2024).

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA. Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, Jan. 2019, doi: 10.3322/caac.21551.

[3] A. Jacovi, O. S. Shalom, and Y. Goldberg, "Understanding Convolutional Neural Networks for Text Classification," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Sep. 2018, pp. 56–65. [Online]. Available: http://arxiv.org/abs/1809.08037

[4] H. Zhu, Y. Fan, and Z. Liang, "Improved Curvature Estimation for Shape Analysis in Computer-Aided Detection of Colonic Polyps," in *Virtual Colonoscopy and Abdominal Imaging. Computational Challenges and Clinical Opportunities*, 2011, pp. 9–14. doi: 10.1007/978-3-642-25719-3_2.

[5] K. Abdeljawad, K. C. Vemulapalli, C. J. Kahi, O. W. Cummings, D. C. Snover, and D. K. Rex, "Sessile serrated polyp prevalence determined by a colonoscopist with a high lesion detection rate and an experienced pathologist," *Gastrointest. Endosc.*, vol. 81, no. 3, pp. 517–524, Mar. 2015, doi: 10.1016/j.gie.2014.04.064.

[6] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014, doi: 10.1109/RBME.2013.2295804.

[7] M. Veta *et al.*, "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Med. Image Anal.*, vol. 20, no. 1, pp. 237–248, Feb. 2015, doi: 10.1016/j.media.2014.11.010.

[8] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 141–152, Sep. 2003, doi: 10.1109/TITB.2003.813794.

[9] Peng Li, Kap Luk Chan, and S. M. Krishnan, "Learning a Multi-Size Patch-Based Hybrid Kernel Machine Ensemble for Abnormal Region Detection in Colonoscopic Images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 2, pp. 670–675. doi: 10.1109/CVPR.2005.201.

[10] Wikimedia Commons, "Category: Intestines." https://commons.wikimedia.org/wiki/Category:Intestines (accessed Aug. 01, 2024).

[11] J. Bernal, J. Sánchez, and F. Vilariño, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, Sep. 2012, doi: 10.1016/j.patcog.2012.03.002.

[12] G. Chartrand *et al.*, "Deep Learning: A Primer for Radiologists," *RadioGraphics*, vol. 37, no. 7, pp. 2113–2131, Nov. 2017, doi: 10.1148/rg.2017170077.

[13] H. Zhu, Y. Fan, H. Lu, and Z. Liang, "Improved Curvature Estimation for Computer-aided Detection of Colonic Polyps in CT Colonography," *Acad. Radiol.*, vol. 18, no. 8, pp. 1024–1034, Aug. 2011, doi: 10.1016/j.acra.2011.03.012.

[14] M. P. Tjoa and S. M. Krishnan, "Feature extraction for the analysis of colon status from the endoscopic images," *Biomed. Eng. Online*, vol. 2, no. 1, p. 9, Apr. 2003, doi: 10.1186/1475-925X-2-9.

[15] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-Based Polyp Detection in Colonoscopy," in *Bildverarbeitung für die Medizin 2009*, 2009, pp. 346–350. doi: 10.1007/978-3-540-93860-6_70.

[16] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Inform.*, vol. 7, no. 1, p. 29, Jan. 2016, doi: 10.4103/2153-3539.186902.

[17] B. Korbar *et al.*, "Deep Learning for Classification of Colorectal Polyps on Whole-slide Images," *J. Pathol. Inform.*, vol. 8, no. 1, p. 30, Jan. 2017, doi: 10.4103/jpi.jpi_34_17.

[18] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1717–1724. doi: 10.1109/CVPR.2014.222.

[19] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/TMI.2016.2535302.

[20] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, Nov. 2014, pp. 3320–3328. [Online]. Available: http://arxiv.org/abs/1411.1792

[21] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[22] H.-C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[23] J. Bernal, F. Vilariño, and J. Sánchez, "Feature detectors and feature descriptors: Where we are now," Universitat Autonoma de Barcelona, 2010. [Online]. Available: https://refbase.cvc.uab.es/files/BVS2009.pdf

[24] S. M. Krishnan, X. Yang, K. L. Chan, S. Kumar, and P. M. Y. Goh, "Intestinal abnormality detection from endoscopic images," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, 2002, vol. 2, pp. 895–898. doi: 10.1109/IEMBS.1998.745583.

[25] A. Al Mamun, P. P. Em, T. Ghosh, M. M. Hossain, M. G. Hasan, and M. G. Sadeque, "Bleeding recognition technique in wireless capsule endoscopy images using fuzzy logic and principal component analysis," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 3, p. 2688, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2688-2695.

[26] J.-X. Liu, C.-H. Zheng, and Y. Xu, "Extracting plants core genes responding to abiotic stresses by penalized matrix decomposition," *Comput. Biol. Med.*, vol. 42, no. 5, pp. 582–589, May 2012, doi: 10.1016/j.compbiomed.2012.02.002.

[27] Z. Lv, Y. Wang, C. Zhang, X. Gao, and X. Wu, "An ICA-based spatial filtering approach to saccadic EOG signal recognition," *Biomed. Signal Process. Control*, vol. 43, pp. 9–17, May 2018, doi: 10.1016/j.bspc.2018.01.003.

[28] P. Sasmal, M. K. Bhuyan, Y. Iwahori, and K. Kasugai, "Colonoscopic Polyp Classification Using Local Shape and Texture Features," *IEEE Access*, vol. 9, pp. 92629–92639, 2021, doi: 10.1109/ACCESS.2021.3092263.

[29] S. Grosu *et al.*, "Machine Learning–based Differentiation of Benign and Premalignant Colorectal Polyps Detected with CT Colonography in an Asymptomatic Screening Population: A Proof-of-Concept Study," *Radiology*, vol. 299, no. 2, pp. 326–335, May 2021, doi: 10.1148/radiol.2021202363.

[30] A. Ellahyani, I. El Jaafari, S. Charfi, and M. El Ansari, "Fine-tuned deep neural networks for polyp detection in colonoscopy images," *Pers. Ubiquitous Comput.*, vol. 27, no. 2, pp. 235–247, Apr. 2023, doi: 10.1007/s00779-021-01660-y.

[31] J. González-Bueno Puyal *et al.*, "Polyp detection on video colonoscopy using a hybrid 2D/3D CNN," *Med. Image Anal.*, vol. 82, p. 102625, Nov. 2022, doi: 10.1016/j.media.2022.102625.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[33] M. A. Hambali and P. A. Agwu, "Adversarial Convolutional Neural Network for Predicting Blood Clot Ischemic Stroke," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 51–64, Jun. 2024, doi: 10.62411/jcta.10516.

[34] M. T. H. Khan Tusar, M. T. Islam, A. H. Sakil, M. N. H. N. Khandaker, and M. M. Hossain, "An Intelligent Telediagnosis of Acute Lymphoblastic Leukemia using Histopathological Deep Learning," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 1–12, May 2024, doi: 10.62411/jcta.10358.

[35] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.

[36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *arXiv*, Sep. 2014, pp. 1–14. [Online]. Available: https://arxiv.org/abs/1409.1556

[38] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[41] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[42] Y. J. Kim, J. P. Bae, J.-W. Chung, D. K. Park, K. G. Kim, and Y. J. Kim, "New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images," *Sci. Rep.*, vol. 11, no. 1, p. 3605, Feb. 2021, doi: 10.1038/s41598-021-83199-9.

[43] D. R. I. M. Setiadi, H. M. M. Islam, G. A. Trisnapradika, and W. Herowati, "Analyzing Preprocessing Impact on Machine Learning Classifiers for Cryotherapy and Immunotherapy Dataset," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 39–50, Jun. 2024, doi: 10.62411/faith.2024-2.

[44] T. R. Noviandy, G. M. Idroes, and I. Hardi, "An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 84–95, Aug. 2024, doi: 10.62411/faith.2024-16.

[45] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.
[46] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information," *IEEE Trans. Med. Imaging*, vol. 35, no. 2, pp. 630–644, Feb. 2016, doi: 10.1109/TMI.2015.2487997.
[47] N. Dehghani *et al.*, "Robust Colorectal Polyp Characterization Using a Hybrid Bayesian Neural Network," in *Cancer Prevention Through Early Detection*, 2022, pp. 108–117. doi: 10.1007/978-3-031-17979-2_11.
[48] G. Urban *et al.*, "Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy," *Gastroenterology*, vol. 155, no. 4, pp. 1069-1078.e8, Oct. 2018, doi: 10.1053/j.gastro.2018.06.037.
[49] Ş. Öztürk and U. Özkaya, "Gastrointestinal tract classification using improved LSTM based CNN," *Multimed. Tools Appl.*, vol. 79, no. 39–40, pp. 28825–28840, Oct. 2020, doi: 10.1007/s11042-020-09468-3.