Received: 30 January 2024 | Revised: 22 May 2024 | Accepted: 9 July 2024 | Published online: 15 July 2024

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 3(1) 83–93
DOI: 10.47852/bonviewAIA42022549

BON VIEW PUBLISHING

A Novel Ensemble Deep Learning-Based Polyp Detection Using Colonoscopy Dataset

K. Sai Rakshana[1], Antony Dennis Ananth[2,*] and L. Gowri[2]

[1]*SASTRA (Deemed University), India*

[2]*School of Computing, SASTRA (Deemed University), India*

Abstract: This work addresses the critical task of polyp detection and classification using the SUN colonoscopy video database, which consists of still images annotated with bounding boxes. These images categorize frames into polyp and non-polyp and encompass six distinct classes of polyps: hyperplastic polyp, sessile serrated lesion, low-grade adenoma, traditional serrated adenoma, high-grade adenoma, and invasive carcinoma. The approach involves a two-stage classification process. Initially, MobileNetV2 is employed to distinguish between polyp and non-polyp frames. Subsequently, ResNet50 and GoogLeNet are utilized to classify the identified polyps into the six predefined categories. Data augmentation techniques are implemented to address the inherent imbalance in class distribution within the dataset, enhancing model performance and generalizability. The results highlight the effectiveness of GoogLeNet, which achieved an impressive accuracy of 98%, significantly outperforming ResNet50's accuracy of 76.16%. This substantial improvement underscores the potential of GoogLeNet in enhancing the accuracy of polyp classification. The significance of this work lies in its contribution to advancing automated polyp detection and cancer stage classification, crucial for early diagnosis and treatment. These findings provide a foundation for further research and development in this domain, with the potential to improve clinical outcomes through more accurate and timely identification of colorectal polyps.

Keywords: deep learning, polyp detection, CNN, image classification, colorectal disease

1. Introduction

The danger of colorectal disease as a significant global medical problem features the need for effective early discovery strategies. This goal is significantly supported by the examination project, which is centered around polyp detection using the SUN colonoscopy video dataset where information assortment was done [1–3]. The principal objective is to detect polyps utilizing modern profound learning models, which would enormously support the early finding of colorectal disease and may try and have an impact on how screenings are presently led.

Polyps require brief therapy since they might be early marks of colorectal malignant growth. It is fundamental to recognize and describe them right on time to prevent them from forming into threatening growths [4, 5]. A fundamental asset for this is the SUN colonoscopy video database, which offers a wide assortment of clarified polyp approaches that portray the multiple manners by which these distortions happen. The objective of the review is to find and address polyp development at its weakest places to forestall issues [6–8].

1.1. Polyps: Precursors to colorectal cancer

Polyps are significant marks of colorectal cancer since they are a scope of tissue irregularities that might possibly be harmful [9–11]. The polyps were mostly located in the rectum or colon; these developments are regularly harmless however have a secret possibility of turning destructive. Early location and portrayal of polyps are critical for brief protection activity. Improving deep learning-based polyp detection using feature extraction and data augmentation [12, 13]. Since polyps are in many cases asymptomatic in the beginning phases, routine testing becomes fundamental. For individualized patient treatment that guarantees productive distinguishing proof and hazard appraisal, it becomes basic to understand the particular attributes of different polyp sorts [3, 11, 14].

1.2. Dataset overview

This work depends on the SUN dataset, which contains annotated polyp and non-polyp still images. Highlighting more than 49,000 still images of polyps which are isolated into six classifications that relate to various colorectal issues, this assortment gives a definite knowledge of polyp which is colorectal cancer [15]. The significance of the dataset rests in its amount as well as in its magnificent portrayal of the regular

*Corresponding author: Antony Dennis Ananth, School of Computing, SASTRA (Deemed University), India. Email: dennisananth@it.sastra.edu

fluctuation in polyp shape, which is fundamental for preparing productive AI models. Utilizing a validation dataset alongside training and testing datasets is crucial for developing and evaluating machine learning models effectively. In this study, the dataset was meticulously partitioned, allocating 80% of the data for training purposes and reserving 20% for testing. Within the training set, 5% was specifically earmarked for validation purposes. This meticulous allocation ensures that the validation dataset remains independent from both the training and testing phases, allowing for unbiased model evaluation and hyperparameters tuning. The utilization of a dedicated validation subset enhances the robustness of the model development process, enabling researchers to fine-tune model parameters and assess performance accurately. Consequently, this approach facilitates the creation of reliable diagnostic tools for the detection of colorectal cancer, ensuring the integrity and efficacy of the developed models. Several data augmentation techniques were employed to address class distribution imbalance and enhance the diversity of the dataset. These techniques included rotation, shifting, shearing, zooming, and flipping. Rotation involved rotating the image by a certain angle, introducing variations in the orientation of polyps and non-polyps. Shifting randomly translated the image horizontally and vertically, simulating slight movements in the colonoscopy procedure. Shearing distorted the image along the x or y-axis, mimicking the deformation that can occur during image capture. Zooming adjusted the image's scale, enabling the model to learn from different magnifications of polyps and non-polyps. Finally, flipping horizontally mirrored the image, providing additional variations in the appearance of polyps and non-polyps. Together, these augmentation techniques increased the dataset's size and diversity, allowing the models to learn robust features and generalize better to unseen data. Models are better prepared by considering all the variations in still images to perceive and arrange polyps in true conditions precisely [11, 16].

1.3. Classification with MobileNetV2

Combining MobileNetV2, which is renowned for its efficiency and lightweight architecture, makes sense given the need for continuous image processing in polyp and non-polyp grouping. Using parallel grouping to find polyps, MobileNetV2 is able to discriminate between outlines that include and do not contain polyps. This compact yet accurate model serves as a CNN model for describing polyp and non-polyp outlines and establishes the groundwork for further sophisticated orders [17].

1.4. Staging with ResNet50 and GoogLeNet

The task coordinates ResNet50 and GoogLeNet, high-level profound learning models, for polyp characterization. ResNet50's lingering learning catches fine-grained highlights essential for separating polyp stages. GoogLeNet improves highlight extraction, distinguishing complex examples in polyp pictures. Prepared on an explained dataset, the two models give basic data to polyp distinguishing proof and potential harm evaluation which is the stages of polyps [18]. Their flexibility to the subtleties of clinical imaging positions them as incredible assets for exact and complex polyp organizing.

2. Objectives

1) The fundamental goal of this task is to make a high-level framework for diagnosing colorectal malignant growth, specifically focusing on the identification and classification of polyps utilizing the SUN colonoscopy video dataset.
2) The essential objective is to use modern profound learning algorithms to identify polyps, further developing early recognition endeavors precisely.
3) Using state-of-the-art advancements like MobileNetV2, ResNet50, and GoogLeNet, the intention is to improve indicative accuracy. This undertaking likewise addresses difficulties connected with class irregularity inside the dataset by utilizing a powerful information expansion approach.

3. Methodology

For polyp identification with MobileNetV2, explicit improvements were acquainted with advance execution. This included redoing the architecture with extra layers like Global Average Pooling and dense layers with ReLU activation. A binary cross-entropy loss function tailored for binary classification tasks was employed. Moreover, a custom information generator worked with proficient pre-processing and clump age for preparing, empowering compelling treatment of enormous datasets and increasing strategies.

With respect to APIs, while MobileNetV2 filled in as the base design, huge changes were made to upgrade usefulness. Extra layers were added to the base model, and custom misfortune capabilities were used during preparing. These changes worked on the model's capacity to catch polyp highlights in colonoscopy pictures, displaying a mix of laid out engineering and customized preparing techniques.

For polyp stage characterization, ResNet50 and GoogleNet models were carried out. While both pointed toward characterizing polyp stages, GoogleNet displayed predominant precision, ascribed to building refinements and coordinated information expansion strategies. These improvements, including turn, moving, shearing, zooming, and flipping, reinforced the model's capacity to successfully sum up and group concealed information. Early halting and learning rate booking additionally advanced preparing, adding to its viability.

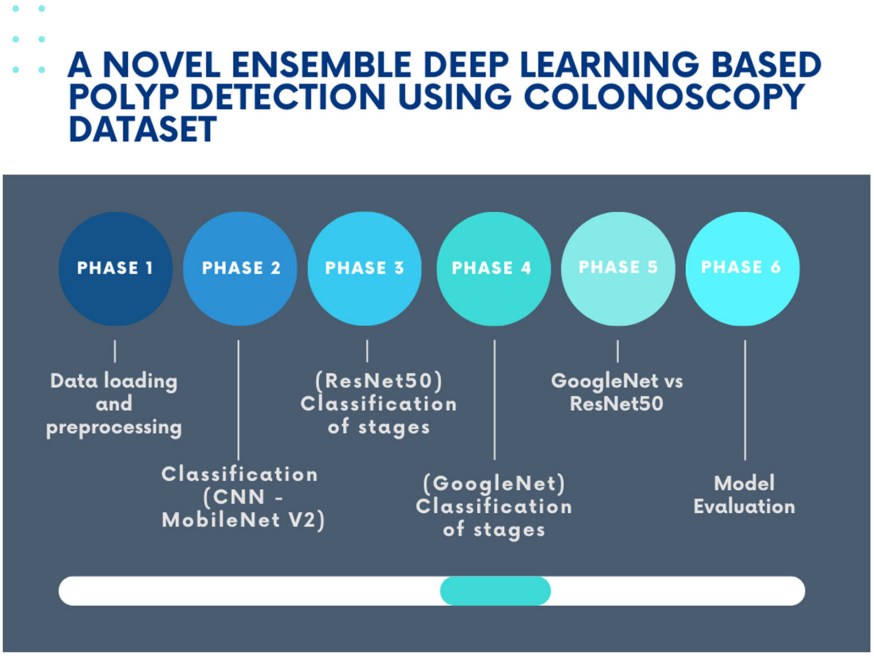
Prominently, the GoogleNet model utilized existing APIs without broad alterations, underlining the significance of custom-made preparing systems. This approach highlights the model's proficiency in accomplishing predominant execution for polyp stages grouping, making it a solid answer for this undertaking.

3.1. Data acquisition and pre-processing

The main era of the philosophy is information gathering and pre-processing. The dataset, which was gotten from the SUN colonoscopy video dataset, comprises an excess of 49,000 commented-on polyp outlines. Each polyp is addressed in one of six groups of polyps. To prepare profound learning models that can dependably recognize and characterize polyps—an essential for solid indicative outcomes—this rich dataset is utilized. Broad pre-processing is finished to ensure information consistency and quality [19–21]. This incorporates taking care of absent or undermined information, adjusting annotations to compare images, and normalizing image configurations to diminish the chance of aspect disparities [7, 10, 22]. To really prepare and test the resulting profound learning models, it is planned to create a reasonable and organized dataset [6].

Normalizing still images of the dataset during the pre-processing stage is fundamental for consistency, staying away from differences that can thwart the models' preparation. Tending to absent or undermined information likewise guarantees the uprightness of the dataset, giving model learning a strong base [4, 6].

Figure 1
Workflow



3.2. Binary classification with MobileNetV2

The venture's most memorable stage centers around binary classification utilizing the MobileNetV2 model, which is an essential stage that lays out the preparation for additional complicated undertakings to come. The objective of binary classification is to recognize the colonoscopy still images of polyp and non-polyp outlines. As a result of its lightweight design, which adjusts computational proficiency and model exactness, MobileNetV2 was painstakingly chosen for this errand. Since the SUN colonoscopy video dataset contains a lot of information, handling outlines rapidly during the preparation and classification stages requires high handling effectiveness [16].

The cautiously pre-processed dataset, in which each still images were normalized and aligned up with matching comments, fills in as the preparation set for MobileNetV2. The model acquires the capacity to identify and arrange polyps during preparing precisely. A significant stage in the symptomatic pipeline is the emphasis on polyp and non-polyp binary classification, which empowers the framework to recognize outlines that might contain irregularities and those that don't [23]. The certainty score that MobileNetV2 gets fills in as a reference point for later periods of the venture, showing how well the model can reliably recognize polyp and non-polyp outlines [24].

3.3. Polyp staging with ResNet50 and GoogLeNet

During the polyp identification stage, the task utilizes two strong profound learning models: ResNet50 and GoogLeNet. These models were chosen in view of their perplexing structures, which are expected to catch unobtrusive highlights that are urgent for separating between polyp stages. The task's purpose to give doctors a more intensive comprehension of polyp qualities—especially with respect to the likely danger of recognized developments—is exhibited by the choice to go past binary characterization. ResNet50 and GoogLeNet are both broadly perceived for their capacity to distinguish complex examples in the still images, which makes them reasonable for the difficult assignment of characterizing polyps into six distinct gatherings [4].

The dataset is utilized as a preparation set for ResNet50 and GoogLeNet subsequent to going through careful pre-processing and effective binary characterization. These models recognize and arrange polyps as indicated by their stages: hyperplastic polyp, sessile serrated lesion, low-grade adenoma, traditional serrated adenoma, high-grade adenoma, and invasive carcinoma. They were autonomously prepared on this advanced dataset. This technique not just enables the models to distinguish polyps with high precision; however, it additionally empowers them to offer canny discourse on the conceivable seriousness of every development that is identified [25]. A significant stage in the choice of a model is the similar examination that follows, which features unobtrusive contrasts in the qualities and shortcomings of ResNet50 and GoogLeNet. By taking on the most exact and powerful design for polyp arranging, the task is directed by this near understanding, ensuring that the symptomatic framework produces precise and reliable outcomes for clinical translation and navigation. Figure 1 highlights the different phases of polyp detection using various deep learning architectures.

3.4. Addressing class imbalance through data augmentation

Guaranteeing the vigor and speculation capacity of the prepared models requires tending to the natural class irregularity in the dataset. The acknowledgment of class irregularity recognizes the likelihood that some polyp classes are underrepresented, which could cause predispositions and diminished precision in the models, particularly for minority classes. Pre-processing includes the execution of an essential information expansion process to

moderate this. To adjust the portrayal of different classes in the preparation information, this involves delivering additional images for each polyp class. The model is presented to a more changed scope of situations by adding varieties of the current images to the dataset utilizing techniques like revolution, flipping, and zooming. By doing this, the probability of the model overfitting to predominant classes is diminished, and its capacity to sum up across the whole scope of polyp classes is moved along [25].

The utilization of information expansion decreases class imbalance in two ways. It, most importantly, guarantees a more intensive opportunity for growth by advancing a more equivalent portrayal of every polyp class and keeping the model from being one-sided toward the greater part of class. Second, the models' capacity to adjust to different polyp situations is worked on by the increase of changeability. In genuine applications, where polyp appearances can shift enormously, this flexibility is fundamental [16].

3.5. Training and testing data split

To empower an intensive evaluation of the profound learning models, the dataset is painstakingly partitioned during the essential preparation and testing information split stage. To re-enact genuine circumstances and assurance that the models can sum up their insight beyond the preparation set, this polarity is significant. 48,000 images altogether are painstakingly picked for preparing, giving areas of strength for which the models can develop and adjust. At the same time, 12,000 additional pictures are held for the testing set, which is a significant subset that the models have never found in preparing. This subset is painstakingly chosen to cover an extensive variety of polyp varieties, ensuring that the testing set keeps on being illustrative of the hardships experienced in certifiable circumstances. Testing transforms into an exhaustive assessment of the models' exhibition since they need to accurately distinguish and classify polyps in circumstances they have never seen.

3.6. Model evaluation and comparative analysis

After a thorough preparation stage, a careful model assessment is done to decide how well the created framework performs. The utilization of measurements like precision, accuracy, recall, and *F*1 score takes into consideration a careful assessment of each model's ability to distinguish and organize polyps accurately. As a fundamental measurement, precision shows how exact the models are generally speaking in their expectations [16]. To assess a model's ability to lessen bogus up-sides, accuracy works out the proportion of genuine positive forecasts to all anticipated up-sides. Alternately, review gives understanding of the models' capacity to catch all relevant cases by ascertaining the proportion of genuine positive expectations to every genuine positive. *F*1 score, a consonant mean of accuracy and review, gives a fair assessment, especially valuable in situations with imbalanced datasets. The examination of ResNet50 and GoogLeNet for polyp arranging is the focal part of this evaluation. This investigation investigates the unobtrusive benefits and weaknesses of each model, uncovering how well everyone can arrange polyps into particular stages. In light of the bits of knowledge acquired from the assessment results, the models are calibrated to ensure that the chosen design is upgraded to give best execution in genuine applications, eventually improving the viability of colorectal disease diagnostics.

3.7. Implementation and deployment

During the strategy's last stage, the accentuation moves to the genuine application and sending of the carefully prepared models, transforming hypothetical improvements into a valuable and pragmatic indicative instrument. The center of this coordinated framework is the engineering that was chosen: an insightful mix of GoogLeNet or ResNet50 for polyp organizing and MobileNetV2 for paired grouping. The coordination cycle involves carrying these models into congruity to make a bound-together system that precisely and productively processes new colonoscopy pictures. MobileNetV2, which is notable for its meager engineering, performs well in rapidly recognizing pictures as one or the other polyp or non-polyp, filling in as an underlying channel for additional means. ResNet50 or GoogLeNet then assume the more complicated assignment of polyp arranging, utilizing their complex models to recognize the various phases of polyps.

3.8. Ensemble model

To work on the exactness of polyp identification and classification, an ensemble model was developed by consolidating the forecasts of both the ResNet50 and GoogLeNet (InceptionV3) models. Regardless of individual correctness of 76.16% for ResNet50 and 98% for GoogLeNet, the troupe model accomplished just a 3% expansion in exactness. The troupe model utilized the expectations of both ResNet50 and GoogLeNet to create a last forecast. This was accomplished by averaging the anticipated probabilities from the two models for each information picture. Albeit this troupe approach expected to profit by the qualities of the two models, the peripheral expansion in precision proposes restricted reciprocal execution between the two models.

The procedure included pre-processing the information, preparing the two models freely, and afterward consolidating their expectations. The dataset, got from the SUN colonoscopy video dataset, contained commented-on pictures of polyps and non-polyps across six classes. Broad pre-processing guaranteed information consistency and quality, including taking care of absent or undermined information and normalizing picture designs. Information expansion procedures like turn, moving, shearing, zooming, and flipping were utilized to address class circulation awkwardness and improve the variety of the dataset. These procedures misleadingly expanded the dataset's size and presented fluctuation in the preparation models, permitting the models to learn vigorous elements and sum up better to concealed information.

The ResNet50 and GoogLeNet models were pertained on the ImageNet dataset and afterward tweaked on the polyp dataset utilizing move learning. Custom grouping heads were added to the two models, trailed by assemblage utilizing the Adam analyzer and absolute cross-entropy misfortune capability. Blended accuracy preparing was empowered to advance execution and memory use. During training, early stopping and learning rate scheduling callbacks were employed to prevent overfitting and improve convergence. The models were trained for 10 epochs using a custom data generator, which generated batches of augmented images and their corresponding labels. After preparing fulfillment, the models were assessed on a different test dataset, and expectations were created utilizing the group approach. Disarray lattices and characterization reports were produced to evaluate the models' presentation across various classes. Furthermore, the general Beneficiary Working Trademark (ROC) bend and Region Under the Bend were determined to assess the models' presentation in twofold grouping errands. Notwithstanding the group approach's

minimal improvement in exactness, further examination concerning model engineering, preparing procedures, and outfit strategies might be justified to accomplish more significant execution acquires in polyp recognition and order errands.

4. Results and Discussion

With a score of 98 percent, the principal paired grouping stage utilizing MobileNetV2 showed exceptional precision. This outcome features how well the model recognizes approaches that contain polyps and those that do not, giving areas of strength to additional means in the symptomatic cycle. The ideal precision proposes that MobileNetV2 capabilities as a reliable first channel, ensuring that main certifiable polyp outlines advance to the more complicated organizing stage. This exceptional accuracy plays a fundamental role in reducing false positives in a demonstration tool where accuracy is essential for the accurate diagnosis of colorectal illness.

4.1. Performance evaluation

In Table 1, there is a critical exactness distinction among ResNet50 and GoogLeNet. GoogLeNet performs better compared to ResNet50, beating it with an astonishing exactness of 98.78%, while ResNet50 performs more regrettably, at 76.16%. This distinction shows that the GoogLeNet design is more skilled at distinguishing inconspicuous qualities that are fundamental for exact polyp arranging. Since GoogLeNet is more exact, it can recognize the six distinct sorts of polyps all the more actually, giving specialists more dependable data about whether the developments they have found are harmful. The unpredictable varieties in accuracy among these refined models highlight the meaning of picking reasonable structures dependent upon the complexity of the grouping task. In Table 1, the accuracy for the existing deep learning models is highlighted.

4.2. Polyp detection and evaluation metrics

Figure 2 depicts the number of predictions for each class. Every polyp class is addressed on the x-axis of the diagram, which shows the quantity of expectations on the y-axis. This visual portrayal makes it simple to assess the model's exhibition in a few classes rapidly and naturally, which makes it more straightforward to recognize potential regions for advancement or improvement. In Figure 3, a definite breakdown of true-positive, true-negative, false-positive, and false-negative expectations is likewise given by the confusion matrix. Understanding the specific benefits and hindrances of the models and coordinating future upgrades for expanded indicative accuracy are made conceivable by examining the confusion matrix mentioned in Figure 4.

4.3. PR and ROC curve

From Figures 5–7, the precision-recall curve and ROC curve give the appraisal more profundity. The prejudicial force of the models at various limits is made sense of by the ROC curve, which outlines the compromise between true-positive rate and false-positive rate. Simultaneously, the precision-recall curve offers essential data about finding some kind of harmony among responsiveness and accuracy; this is particularly significant in circumstances where positive class ID is of the most extreme significance. Cautious assessment of these curves makes it simpler to change edges and model boundaries, working on the framework's demonstrative accuracy.

Table 1
Accuracy table

Model	Accuracy
MobileNetV2	98.00%
ResNet50	76.16%
GoogLeNet	98.87%

4.4. Classification report

Figure 8 shows the true and predicted class distribution giving a careful comprehension of how well the models sum up to different polyp classes. Varieties in the dissemination give understanding of potential biases or challenges in classifying specific polyp types. At long last, the grouping report sums up key execution markers for each class, including support, recall, *F*1-score, and accuracy. These exhaustive report capabilities as a careful reference, giving a nuanced comprehension of the models' presentation in each class and coordinating future enhancements for worked-on symptomatic characteristics.

Subsequently, Figures 9 and 10 discover the viability of MobileNetV2 in the primary binary classification and the prevalence of GoogLeNet in the subsequent polyp arranging. The measurements and representations offer a complete image of the models' strengths and areas for improvement, giving wise data to the continuous enhancement of the indicative device for the recognition of colorectal malignant growth. To improve the accuracy of polyp detection and classification, an ensemble model was constructed by combining the predictions of both the ResNet50 and GoogLeNet (InceptionV3) models. Despite individual accuracies of 76.16% for ResNet50 and 98% for GoogLeNet, the ensemble model achieved a 3% increase in accuracy.

4.5. Model comparison

In Table 2, the proposed model (GoogleNet V2) displays remarkable execution across numerous measurements, settling on it a champion decision for object location undertakings. With an amazing accuracy and review of 99.00%, the model shows its capacity to precisely distinguish positive occurrences while limiting misleading up-sides and bogus negatives. This high accuracy and review are supplemented by an ideal *F*1 score of 99.00%, demonstrating an ideal harmony among accuracy and review. The use of the ReLU actuation capability further upgrades the model's productivity and adequacy, empowering quicker union during preparing and better treatment of disappearing angles. Moreover, with an expedient derivation season of 9.8 milliseconds, the model grandstands its capacity to handle input information and produce expectations quickly, making it reasonable for continuous applications where quick reaction times are essential. Generally speaking, the proposed model succeeds in both execution and effectiveness, situating it as an ideal answer for object identification undertakings in different functional settings.

5. Conclusions and Future Work

5.1. Conclusions

In conclusion, this examination denotes a critical step in progressing colorectal disease diagnostics through the formation of a high-level polyp identification and organizing framework.

Figure 2
Number of predictions for each class

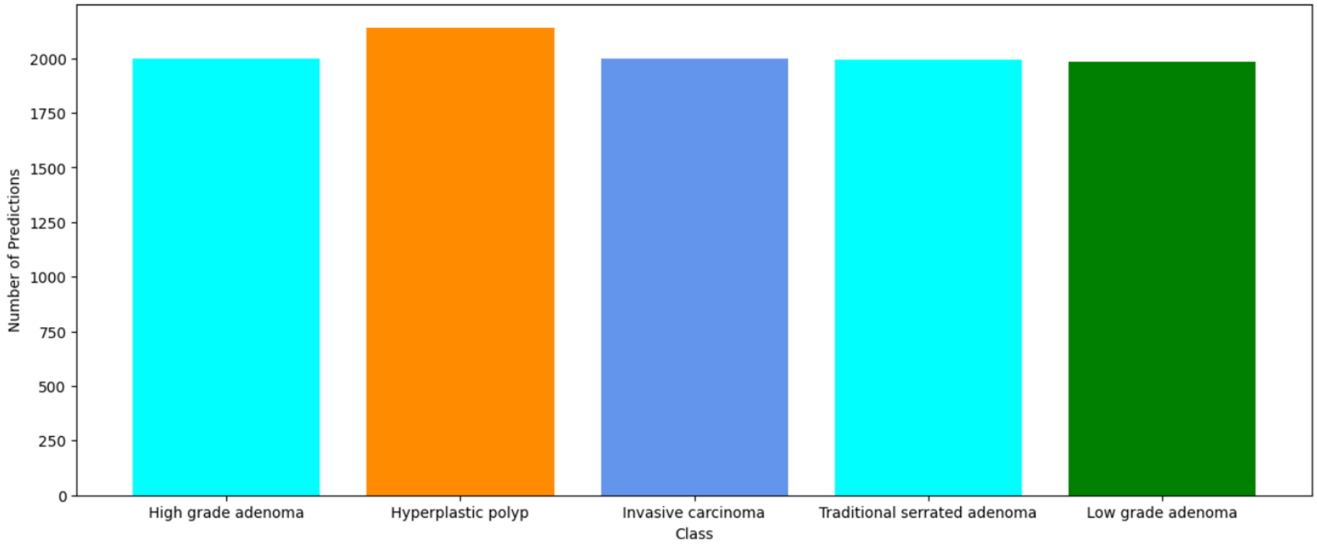


Figure 3
Confusion matrix of GoogLeNet

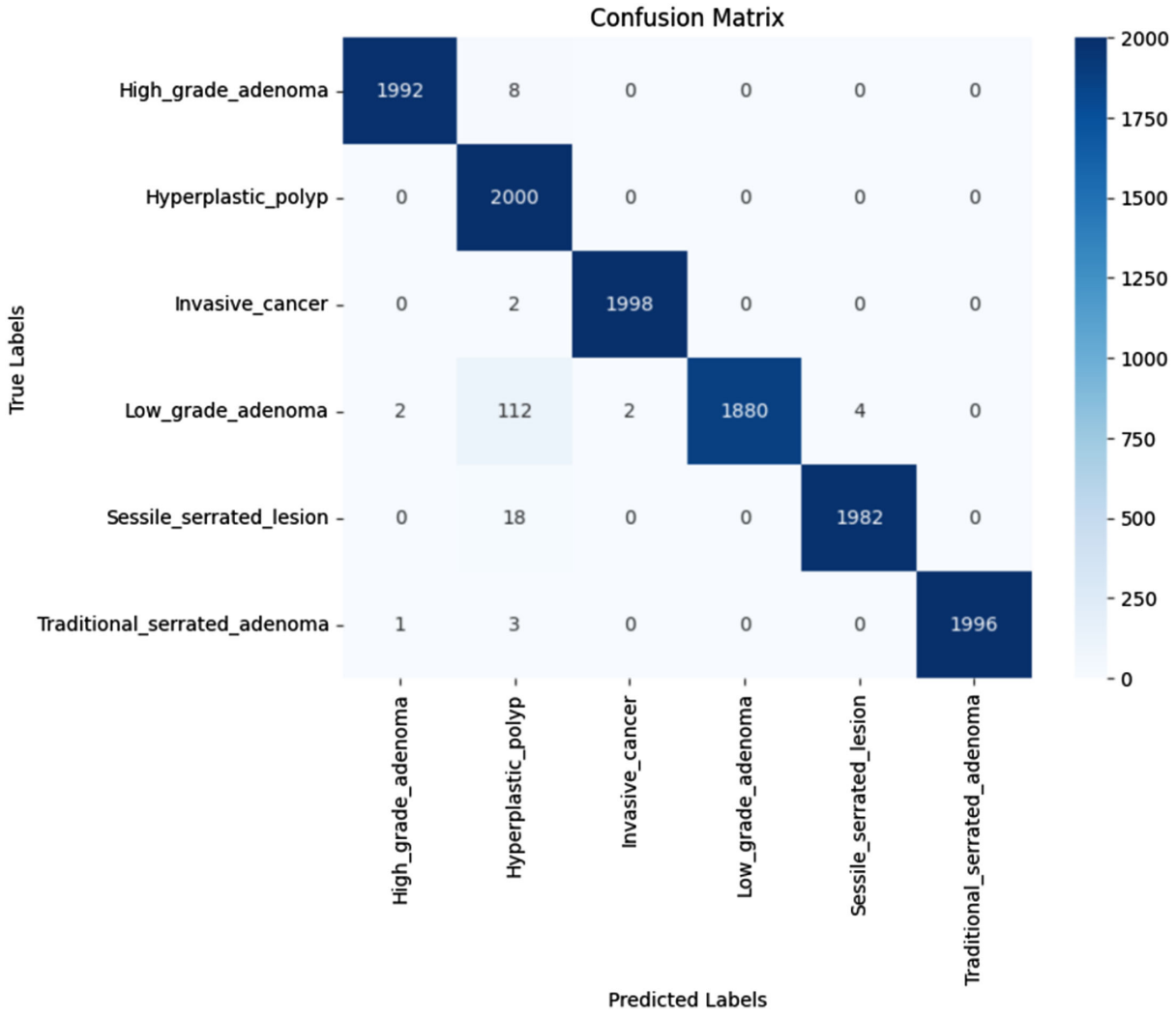


Figure 4
Ensemble model's confusion matrix

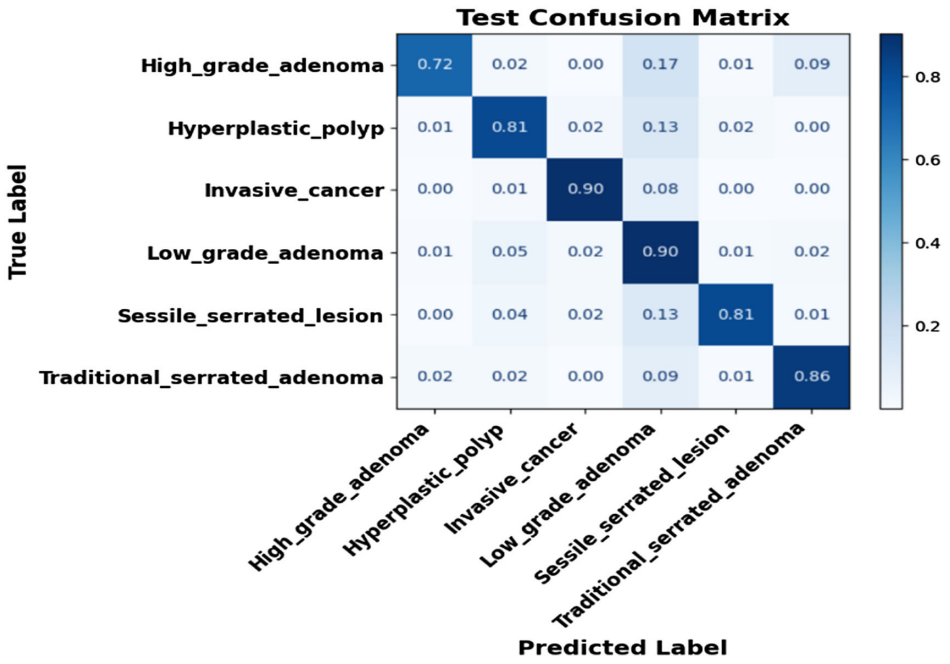


Figure 5
ROC curve of GoogLeNet

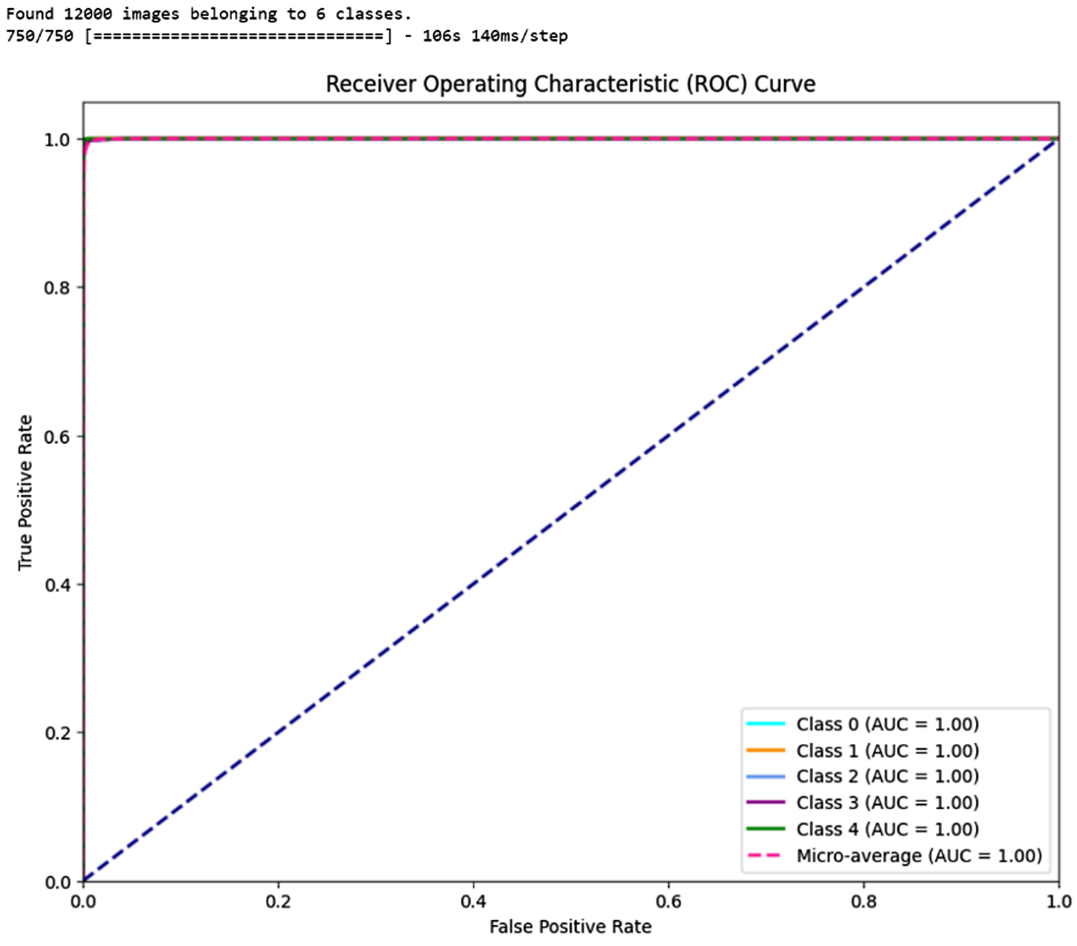


Figure 6
Precision-recall curve of GoogLeNet

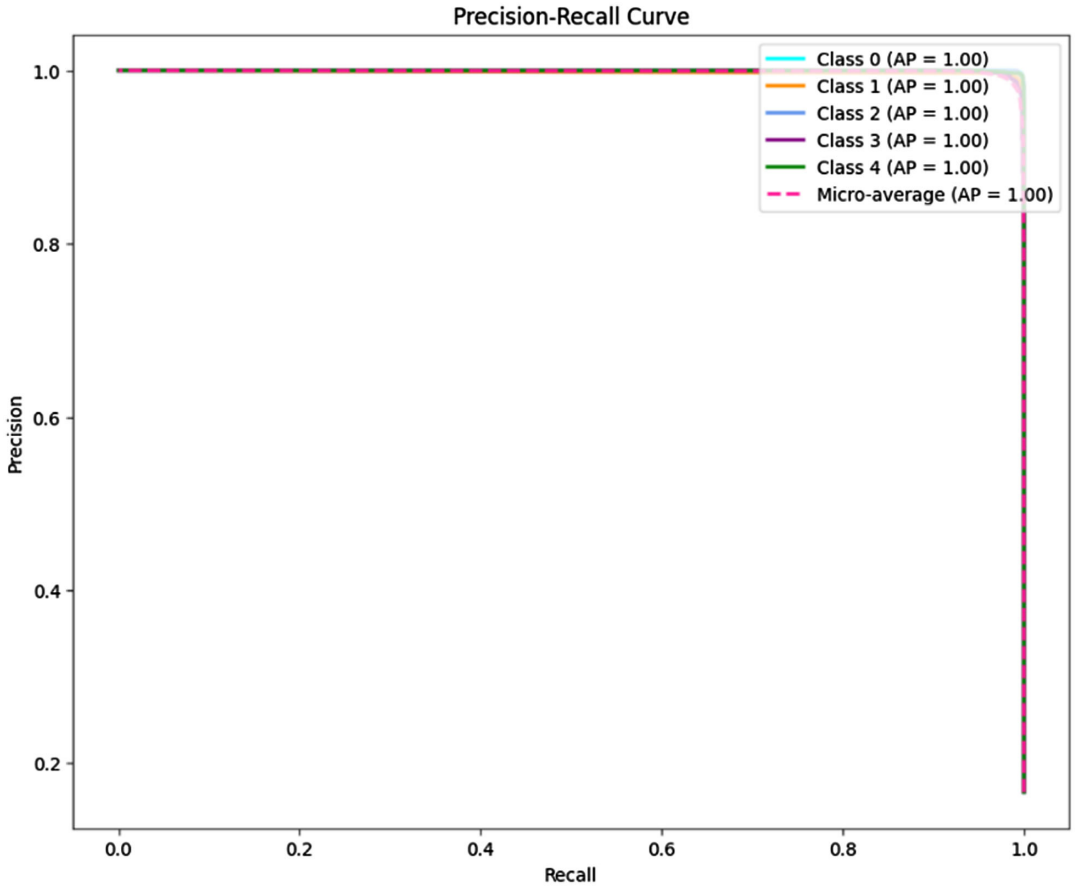
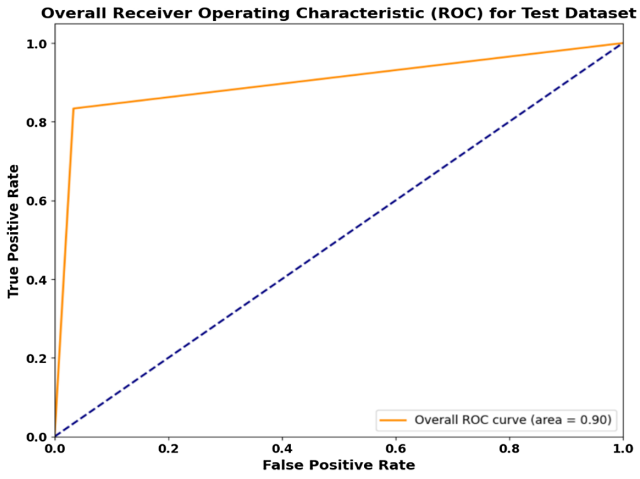


Figure 7
ROC curve of the ensemble model



By utilizing state-of-the-art profound learning models—in particular GoogLeNet, ResNet50, and MobileNetV2—the review has yielded promising results in exactly distinguishing and sorting polyps. Eminently, MobileNetV2's immaculate presentation, accomplishing an ideal exactness score in the underlying paired grouping, lays out a vigorous starting point for ensuing polyp classification. Stunningly, GoogLeNet arose as the predominant model, bragging a precision 98.78% and ResNet50's 76.16%. The coordination of these models into the symptomatic cycle holds extraordinary potential to upgrade patient results and reinforce early identification endeavors in colorectal malignant growth essentially.

5.2. Future work

Looking forward, there are invigorating possibilities for refining and growing this work. One promising road includes making the created models available to medical care experts through coordination into an instinctive web application. The shift from an examination system to a viable, client-focused instrument requires tending to constant handling capacities, computational proficiency, and in general ease of use. Sending the framework on a web stage improves openness as well as works with far-reaching reception, adding to the democratization of cutting-edge symptomatic innovations. In conclusion, this work is ready for a promising future as it changes from research discoveries to a down-to-earth web application. Zeroing in on convenience, nonstop model upgrade, and viable application, the framework can possibly assume a critical part in the early distinguishing model and organizing of colorectal disease, eventually adding to further developed medical services results.

Figure 8
True and predicted class distribution

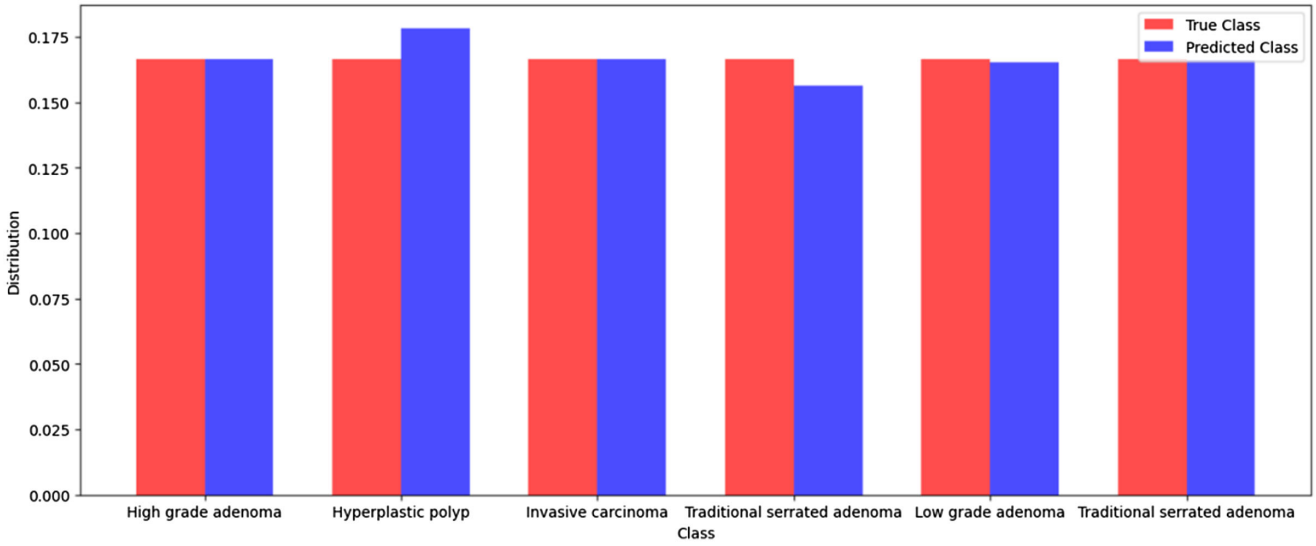


Figure 9
Classification report of GoogLeNet

Classification Report:

	precision	recall	f1-score	support
High_grade_adenoma	1.00	1.00	1.00	2000
Hyperplastic_polyp	0.93	1.00	0.97	2000
Invasive_cancer	1.00	1.00	1.00	2000
Low_grade_adenoma	1.00	0.94	0.97	2000
Sessile_serrated_lesion	1.00	0.99	0.99	2000
Traditional_serrated_adenoma	1.00	1.00	1.00	2000
accuracy			0.99	12000
macro avg	0.99	0.99	0.99	12000
weighted avg	0.99	0.99	0.99	12000

Figure 10
Classification report of the ensemble model

Classification Report for Test Dataset:

	precision	recall	f1-score	support
High_grade_adenoma	0.96	0.72	0.82	2000
Hyperplastic_polyp	0.86	0.81	0.84	2000
Invasive_cancer	0.94	0.90	0.92	2000
Low_grade_adenoma	0.60	0.90	0.72	2000
Sessile_serrated_lesion	0.94	0.81	0.87	2000
Traditional_serrated_adenoma	0.87	0.86	0.87	2000
accuracy			0.83	12000
macro avg	0.86	0.83	0.84	12000
weighted avg	0.86	0.83	0.84	12000

Table 2
Performance comparison of the existing models with the proposed model

Model	Activation function	Precision %	Recall %	*F*1 %	Speed (ms)
YOLOv3	ReLU	79.44	76.10	77.73	10.3
YOLOv3-SPP	ReLU	80.19	78.61	79.39	10.4
YOLOv3-SPP	Mish	83.01	80.66	81.82	10.4
YOLOv3-SPP	Mish	83.43	77.28	80.24	10.5
YOLOv5s	ELU	88.97	64.82	–	–
Proposed Model (GoogLeNet)	ReLU	99.00	99.00	99.00	9.8

6. Challenges and Solutions

1) **Data Acquisition and Pre-processing:** The underlying test lies in social occasion and setting up the dataset. For this situation, the SUN colonoscopy video dataset containing more than 49,000 clarified polyp frames is used. The difficulties tended to here incorporate taking care of absent or ruined information, adjusting comments to relating pictures, and normalizing picture configurations to guarantee consistency. These means are fundamental for guaranteeing the trustworthiness and nature of the dataset. Furthermore, the production of an organized dataset works with the preparation and testing of profound learning models.
2) **Binary Classification with MobileNetV2:** The test in this stage is to precisely arrange colonoscopy actual pictures into polyp and non-polyp classifications. MobileNetV2 is picked for its lightweight plan and harmony between computational proficiency and model precision. By using the cautiously pre-handled dataset, MobileNetV2 is prepared to recognize and characterize polyps precisely. The certainty score got from MobileNetV2 fills in as a kind of perspective for later stages, showing the model's capacity to reliably recognize polyp and non-polyp frames.
3) **Polyp Staging with ResNet50 and GoogLeNet:** This stage includes utilizing ResNet50 and GoogLeNet to group polyps into six unique stages. These models are chosen for their capacity to catch unpretentious elements significant for recognizing polyp stages. The dataset, after careful pre-handling and parallel grouping, is utilized to autonomously train these models. The models distinguish polyps with high exactness as well as give experiences into the expected seriousness of each recognized development.
4) **Addressing Class Imbalance through Data Augmentation:** Class unevenness inside the dataset represents a test as it might prompt inclinations and diminished precision, particularly for minority classes. Information expansion procedures like pivoting, flipping, and zooming are utilized to relieve this awkwardness. By creating extra pictures for every polyp class, the models are presented to a more different scope of situations, diminishing the gamble of overfitting and further developing speculation across all polyp classes.
5) **Execution and Organization:** In the last stage, the center movements to carrying out and conveying the prepared models into down-to-earth analytic apparatuses. The picked design, consolidating MobileNetV2 for twofold characterization and ResNet50 or GoogLeNet for polyp organizing, is coordinated to make a bound-together framework for precisely handling colonoscopy pictures progressively.

7. Potential Benefits and Considerations for Practical Implementation

1) **Improved Diagnostic Accuracy:** Using advanced computer models can help doctors spot colorectal polyps more accurately, potentially leading to better treatment outcomes for patients.
2) **Efficiency in Diagnosis:** These models can speed up the process of analyzing colonoscopy images, making it quicker for doctors to identify and classify polyps. This can save time and improve workflow in hospitals.
3) **Enhanced Clinical Decision Support:** By providing additional support to doctors, these computer models can assist in making more informed decisions about patient care based on the analysis of colonoscopy images.
4) **Potential for Telemedicine and Remote Healthcare:** With these models, patients in remote areas can benefit from expert-level diagnostic capabilities without needing to travel to specialized medical facilities, making healthcare more accessible.
5) **Continuous Improvement and Adaptation:** These computer models can continuously learn and improve over time as they are exposed to more data, ensuring they stay relevant and effective in diagnosing colorectal polyps.

Acknowledgement

We would especially like to thank Prof. Hayato Itoh of Nagoya University, Japan, for granting us access to the SUN Colonoscopy Video Database, which allowed us to complete this research project successfully.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

K. Sai Rakshana: Data curation, Writing – original draft, Writing – review & editing, Visualization. **Antony Dennis Ananth:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources. **L. Gowri:** Supervision, Project administration.

References

[1] Mamonov, A. V., Figueiredo, I. N., Figueiredo, P. N., & Tsai, Y. H. R. (2014). Automated polyp detection in colon capsule endoscopy. *IEEE Transactions on Medical Imaging*, *33*(7), 1488–1502. https://doi.org/10.1109/TMI.2014.2314959

[2] Tajbakhsh, N., Gurudu, S. R., & Liang, J. (2016). Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging*, *35*(2), 630–644. https://doi.org/10.1109/TMI.2015.2487997

[3] Park, S., Lee, M., & Kwak, N. (2015). *Polyp detection in colonoscopy videos using deeply-learned hierarchical features*. Retrieved from: https://www.semanticscholar.org/paper/Polyp-detection-in-Colonoscopy-Videos-Using-Park-Lee/c808f33e1cce8eecdca85d13a47bfa53bfc2a88f

[4] Pacal, I., & Karaboga, D. (2021). A robust real-time deep learning based automatic polyp detection system. *Computers in Biology and Medicine*, *134*, 104519. https://doi.org/10.1016/j.compbiomed.2021.104519

[5] Katai, H., Ishikawa, T., Akazawa, K., Isobe, Y., Miyashiro, I., Oda, I., . . . , & Registration Committee of the Japanese Gastric Cancer Association. (2018). Five-year survival analysis of surgically resected gastric cancer cases in Japan: A retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001–2007). *Gastric Cancer*, *21*, 144–154. https://doi.org/10.1007/s10120-017-0716-7

[6] Nisha, J. S., Gopi, V. P., & Palanisamy, P. (2022). Automated colorectal polyp detection based on image enhancement and dual-path CNN architecture. *Biomedical Signal Processing and Control*, *73*, 103465. https://doi.org/10.1016/j.bspc.2021.103465

[7] Su, J. R., Li, Z., Shao, X. J., Ji, C. R., Ji, R., Zhou, R. C., . . . , & Li, Y. Q. (2020). Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: A prospective randomized controlled study (with videos). *Gastrointestinal Endoscopy*, *91*(2), 415–424. https://doi.org/10.1016/j.gie.2019.08.026

[8] Vinsard, D. G., Mori, Y., Misawa, M., Kudo, S. E., Rastogi, A., Bagci, U., . . . , & Wallace, M. B. (2019). Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointestinal Endoscopy*, *90*(1), 55–63. https://doi.org/10.1016/j.gie.2019.03.019

[9] Tanabe, S., Ishido, K., Matsumoto, T., Kosaka, T., Oda, I., Suzuki, H., . . . , & Yoshifuku, Y. (2017). Long-term outcomes of endoscopic submucosal dissection for early gastric cancer: A multicenter collaborative study. *Gastric Cancer*, *20*, 45–52. https://doi.org/10.1007/s10120-016-0664-7

[10] Misawa, M., Kudo, S. E., Mori, Y., Cho, T., Kataoka, S., Yamauchi, A., . . . , & Mori, K. (2018). Artificial intelligence-assisted polyp detection for colonoscopy: Initial experience. *Gastroenterology*, *154*(8), 2027–2029. https://doi.org/10.1053/j.gastro.2018.04.003

[11] Hirasawa, T., Aoyama, K., Tanimoto, T., Ishihara, S., Shichijo, S., Ozawa, T., . . . , & Tada, T. (2018). Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer*, *21*, 653–660. https://doi.org/10.1007/s10120-018-0793-2

[12] Chou, Y. C., & Chen, C. C. (2023). Improving deep learning-based polyp detection using feature extraction and data augmentation. *Multimedia Tools and Applications*, *82*(11), 16817–16837. https://doi.org/10.1007/s11042-022-13995-6

[13] Tanabe, S., Hirabayashi, S., Oda, I., Ono, H., Nashimoto, A., Isobe, Y., . . . , & Katai, H. (2017). Gastric cancer treated by endoscopic submucosal dissection or endoscopic mucosal resection in Japan from 2004 through 2006: JGCA nationwide registry conducted in 2013. *Gastric Cancer*, *20*, 834–842. https://doi.org/10.1007/s10120-017-0699-4

[14] Ellahyani, A., Jaafari, I. E., Charfi, S., & Ansari, M. E. (2023). Fine-tuned deep neural networks for polyp detection in colonoscopy images. *Personal and Ubiquitous Computing*, *27*(2), 235–247. https://doi.org/10.1007/s00779-021-01660-y

[15] Gangrade, S., Sharma, P. C., Sharma, A. K., & Gangrade, J. (2023). Computer-aided polyps classification from colonoscopy using stacking-based deep learning model. *Authorea Preprints*.

[16] Liu, X., & Yuan, Y. (2022). A source-free domain adaptive polyp detection framework with style diversification flow. *IEEE Transactions on Medical Imaging*, *41*(7), 1897–1908. https://doi.org/10.1109/TMI.2022.3150435

[17] Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. *BMC Medical Imaging*, *20*, 83. https://doi.org/10.1186/s12880-020-00482-3

[18] Hasan, M. M., Islam, N., & Rahman, M. M. (2022). Gastrointestinal polyp detection through a fusion of contourlet transform and neural features. *Journal of King Saud University-Computer and Information Sciences*, *34*(3), 526–533. https://doi.org/10.1016/j.jksuci.2019.12.013

[19] Bernal, J., Sánchez, J., & Vilariño, F. (2012). Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, *45*(9), 3166–3182. https://doi.org/10.1016/j.patcog.2012.03.002

[20] Bae, S. H., & Yoon, K. J. (2015). Polyp detection via imbalanced learning and discriminative feature learning. *IEEE Transactions on Medical Imaging*, *34*(11), 2379–2393. https://doi.org/10.1109/TMI.2015.2434398

[21] Tajbakhsh, N., Gurudu, S. R., & Liang, J. (2015). Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In *2015 IEEE 12th International Symposium on Biomedical Imaging*, 79–83. https://doi.org/10.1109/ISBI.2015.7163821

[22] Dayyeh, B. K. A., Thosani, N., Konda, V., Wallace, M. B., Rex, D. K., Chauhan, S. S., . . . , & ASGE Technology Committee. (2015). ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointestinal Endoscopy*, *81*(3), 502–e1–502.e16. https://doi.org/10.1016/j.gie.2014.12.022

[23] Branch, M. V., & Carvalho, A. S. (2021). Polyp segmentation in colonoscopy images using U-Net-MobileNetV2. *arXiv Preprint: 2103.15715*.

[24] Jha, D., Ali, S., Tomar, N. K., Johansen, H. D., Johansen, D., Rittscher, J., . . . , & Halvorsen, P. (2021). Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access*, *9*, 40496–40510. https://doi.org/10.1109/ACCESS.2021.3063716

[25] Li, K., Fathan, M. I., Patel, K., Zhang, T., Zhong, C., Bansal, A., . . . , & Wang, G. (2021). Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations. *PLoS ONE*, *16*(8), e0255809. https://doi.org/10.1371/journal.pone.0255809

How to Cite: Sai Rakshana, K., Ananth, A. D., & Gowri, L. (2025). A Novel Ensemble Deep Learning-Based Polyp Detection Using Colonoscopy Dataset. *Artificial Intelligence and Applications*, *3*(1), 83–93. https://doi.org/10.47852/bonviewAIA42022549