Journal of Smart Systems and Computing
ISSN-XXXX-XXXXX(Online)

https://rspublisher.com/ojs/index.php/jssc/index
Pages 08-15

Original Article

Endoscopic Deep Vision Intelligence: An EfficientNetB3-Based Diagnostic Framework for Multi-Class Gastrointestinal Structure and Lesion Recognition

Dr. Nagesh Vadaparthi[1], Dr. P. Srinivasa Rao[2] ,Dr.Tatiraju.V.Rajani Kanth[3], Chennaiah Kate[4], G.Madhava Rao[5] ,M. Jyothirma[6]

[1]Professor (IECT), MVGR College of Engineering, Vizianagaram, India, **Itsnageshv@gmail.com**

[2]Professor, Dept of Information Engineering & Computational Technology, MVGR College of Engineering, Vizianagaram, India, **srinivasa.suloo@gmail.com**

[3]Senior Manager, TVR Consulting Services Private Limited, Hyderabad,500055, Telangana, India, **tvrajani55@gmail.com**

[4]Department of Data Science, Malla Reddy University, Hyderabad,500100, Telangana, India, **chennaiahkate@gmail.com**

[5]Assistant Professor, Department of C.S.E, Malla Reddy Engineering College for Women, Hyderabad,500100,Telangana,India, **madhav.739@gmail.com**

[6]Assistant Professor, CSE(H&S Dept), Sridevi Women's Engineering College, Hyderabad, India, **digitalvardhan@gmail.com**

ABSTRACT

Accurate interpretation of Gastrointestinal (GI) endoscopic images is critical for early detection of diseases, along with making effective clinical decisions; however, manual diagnosis is time-consuming and subject to inter-observer variability. This work introduces Endoscopic Deep Vision Intelligence, a powerful deep learning-based diagnostic system based on the fine-tuned EfficientNetB3 deep neural network for multi-class gastrointestinal structure and lesion recognition. The proposed framework uses transfer learning and optimized feature scaling which maps out discriminative anatomical and pathological patterns from endoscopic images effectively. Experiments were performed on a publicly-available Kvasir endoscopy dataset of eight clinically significant classes. Comprehensive evaluation of using precision, recall, F1-score, confusion matrix, ROC--AUC analysis, and comparative model assessment proves the effectiveness of the proposed approach. The accurate fine-tuned EfficientNetB3 model is able to achieve an overall classification accuracy of 99% outperforming several existing state-of-the-art convolutional neural network architectures. The results verify great generalization, high discriminative power, and low misclassification among all classes. Owing to its high accuracy, efficiency, and reliability, the proposed framework shows great potential for its deployment into real world clinical decision support systems for automated gastrointestinal disease detection and anatomical structures recognition.

KEYWORDS

Gastrointestinal endoscopy, Deep learning, EfficientNetB3, Multi-class classification, Endoscopic image analysis, Transfer learning, Computer-aided diagnosis, Medical image classification, Gastrointestinal lesion detection..

OPEN ACCESS

Editor Dr. Anasuya Sesha Roopa Devi Bhima, Professor Computer Science, School of Applied Computer Science & Information Technology Conestoga College Institute of Technology & Advanced Learning108 University Ave East, Waterloo, ON N2J 2W2
abhima@conestogac.on.ca

Reviewer Dr.S.Prabu, Professor, Department Electronics and Communication Engineering, Mahendra Institute of Technology, Namakkal, Tamil Naidu, India
vsprabu4u@gmail.com

Address correspondence to **Dr. Nagesh Vadaparthi**, **Itsnageshv@gmail.com**

The author declaration conflict of interest
Received 05 January 2026

Accepted 31 January 2026

Published 23 February 2026-

I. INTRODUCTION

Gastrointestinal (GI) diseases continue to be of great global health concern because of prevalence, the complexity of diagnosis, and its dependence on endoscopic interpretation by experts. Recent major innovation in deep learning has led to a great change in computer aided diagnosis systems of GI disease detection methods by providing automated and precise analysis of endoscopic images [1]. Interpretable deep learning architectures have become more important to achieve the trust of clinical scenarios in which multi-stage frameworks combined with dimensionality reduction and explainable AI have shown promising diagnostic reliability [2]. Ensemble-based convolutional neural network (CNN) approaches are further used to improve the robustness of approaches that aggregate complementary representations of features which can again improve the detection performance across the different categories of GI diseases [3]. However, some obstacles such as lesion boundary ambiguity, similarity in visual appearance of the disease classes, and variability in endoscopic imaging circumstances remain in conventional segmentation and classification pipelines [4]. AI strategies based on optimization have been proposed to address these issues with a focus on efficient feature extraction and discrimination in complex GI situations [5]. Comprehensive reviews on deep learning-based GI diagnostic systems have noted the increasing adoption of transfer learning and pretrained architectures as a means to mitigate the data scarcity and class imbalance and to accomplish superior generalization. [6] In parallel overall, large-scale datasets like GastroVision have helped to systematically evaluate CNN-based models and explainable frameworks, reinforcing the place of data-driven intelligence in endoscopic analysis [7]. Moreover, there have been recent studies that focus on knowledge distillation and context-aware learning to obtain lightweight but precise diagnostic models that can be deployed in real-world clinical settings in the gastrointestinal and neurological domains [8]. Motivated by these advancements and limitations, in this study, an EfficientNetB3-based diagnostic framework in order to detect gastrointestinal structure and lesions with high accuracy based on endoscopic images is proposed.

II. LITERATURE

Recent research has grown interesting and emphasized on building powerful deep learning frameworks for automated gastrointestinal (GI) disease detection, where focus is particularly devoted to efficiency, contextual understanding and clinical applicability. Knowledge distillation and context-aware learning has been studied in combining diagnostic accuracy with computational efficiency to produce lightweight but accurate models for deployment in real world clinical environments [8]. Attention-based architecture and hybrid architectures have further shown their effectiveness in extracting discriminative spatial patterns from endoscopic images, which is especially effective in combating potentially librally similar classes of diseases. The Spatial-Attention ConvMixer architecture proposed for the classification of GI diseases exploited the Kvasir dataset to improve feature localization and discrimination of GI diseases, which is an important part of medical image analysis [9]. In addition, transfer learning-based approaches have also become prominent, in which advanced reusability of features from pretrained networks has a great usefulness for enhancing the accuracy of early screening, which has been demonstrated by some frameworks that integrate rich transfer features for GI health assessment [10]. The use of ensemble learning and explainable AI has further reinforced the reliability of the diagnosis by combining the predictions of various models while preserving the interpretability to address a key need for clinical adoption [11]. In addition, specialized deep transfer architectures including depth-wise separable convolutional networks with residual connections have demonstrated enhanced performance by keeping model complexity and accuracy highly predictable reducing model complexity without losing accuracy making them suitable for multi-class GI disease diagnosis in limited computational resources [12]. Early seminal works laid the foundation for modern systems of GI disease detection through investigations on classical machine learning and early deep learning models for endoscopic image analysis. Traditional methods of classification using machine learning based approaches proved the feasibility of automated detection of GI disease but were limited due to handcrafted features and poor generalization for the diverse imaging conditions [13]. The shift to deep learning has further turned out to be a great improvement, with convolutional neural networks allowing for end-to-end feature learning, which significantly improved both classification accuracy in several different GI disease categories [14]. Clinical validation studies have also highlighted the desirable diagnostic tools by comparing invasive and non-invasive methods to detect certain GI diseases such as Helicobacter pylori infection, which motivates the necessity of accurate and automated decision-support systems in gastroenterology [15]. Along with algorithmic improvements, endoscopic imaging technologies have seen advances that have contributed to improved diagnosis results by improving the representation of texture and colors and allowing better visualization of subtle lesions and early-stage cancers [16]. These technological and methodological developments as a whole signal a clear evolution away from traditional diagnostic practices and toward intelligent and AI-driven endoscopic systems. However, in spite of significant advancements, current techniques tend to have weaknesses in multi-class discrimination, computational efficiency and integrated approaches that can address both anatomical landmarks and pathological lesions concurrently. This gap is the motivation behind the development of an EfficientNetB3-based deep vision intelligence framework, which takes advantage of optimized feature scaling and transfer learning to achieve a high-accuracy multi-class recognition of gastrointestinal structures and lesions, getting rid of the limitations of previous works, all aligned with current clinical and technological demands.

III. INPUT DATASET

The research study of the proposed EfficientNetB3-based diagnostic framework through an experimental evaluation was performed using a public image database of gastrointestinal endoscopy images extracted from Kaggle platform as shown in Figuree. 1. The dataset consists of altogether 4000 high quality RGB endoscopic images in JPG format which are directly suitable for deep learning-based image classification without format conversion. The dataset is divided into eight clinically relevant classes, perceived using the anatomical landmarks of the gastrointestinal anatomy as well as pathologic findings. Specifically, the classes are Dyed Lifted Polyps, Dyed Resection Margins, Esophagitis, Normal Cecum, Normal Pylorus, Normal Z-Line, Polyps, and Ulcerative Colitis, each class containing about 500 images, so as to ensure a well-balanced class distribution. Such balance is important to have lesser model bias and stable multi-class classification performance. For the model development, the dataset was split into training, validation, and testing datasets in a systematic way with 70 percent and 15 percent respectively to obtain 2,800 images for training, 600 images for validation, and 600 images for testing purposes. This stratified splitting strategy maintains the class proportions in all subsets and allows to measure the performance reliably. Before being fed into the training, the images were all resized to the input size of EfficientNetB3 and normalized to ensure consistent scaling of features. The diversity of imaging conditions in the dataset, such as differences in illumination, texture, viewpoint as well as the appearance of lesion provide realistic representation of real-world endoscopic scenarios. Representative samples from each class are shown in Fig. 1 to reflect the diversity of visual appearance and the complexity of the diagnostics of the dataset used in this study.

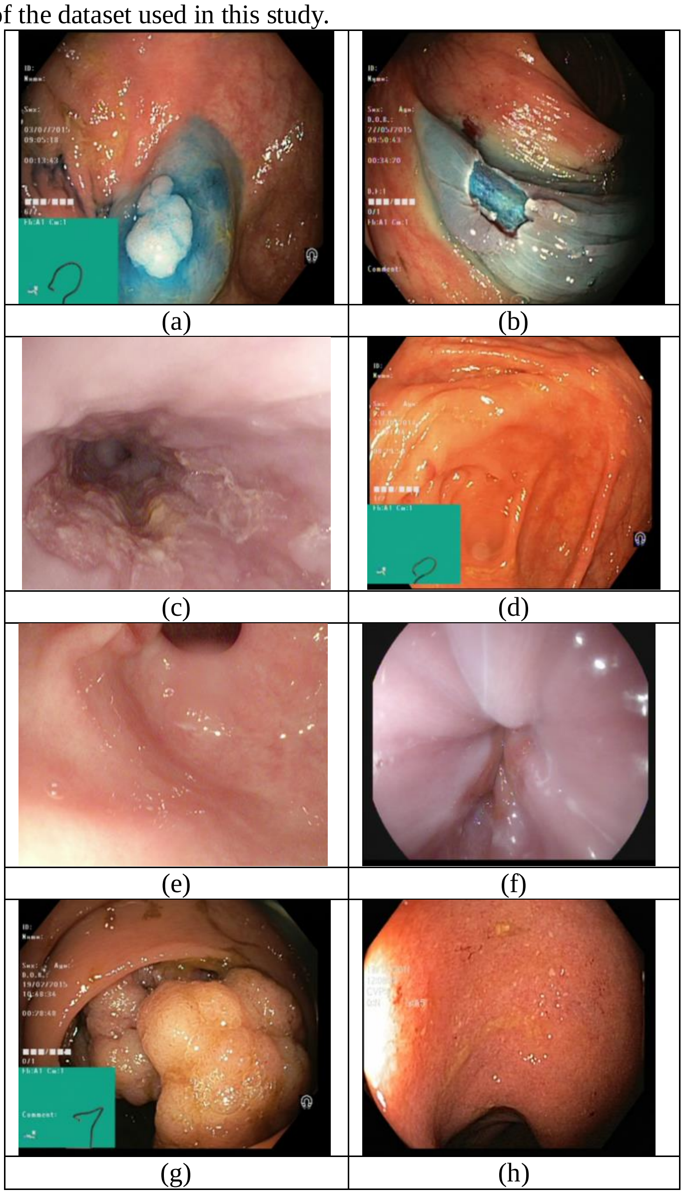

Figure. 1 Dataset Image for (a) Dyed Lifted Polyps , (b) Dyed Resection Margins. (c) Esophagitis , (d) Normal Cecum, (e) Normal Pylorus , (f) Normal Z-Line , (g) Polyps and (h) Ulcerative Colitis

IV. FINE-TUNED EFFICIENTNETB3 MODEL ARCHITECTURE

The proposed diagnostic framework uses a fine-tuned EfficientNetB3 architecture for the high-accuracy multi-class gastrointestinal image classification as shown in Figure. 2. EfficientNetB3 is a state-of-the art convolutional neural network that uses compound scaling, which uniformly scales network depth, width and input resolution to optimize performance while maintaining computational efficiency. In this research, ImageNet pre-trained weights have been applied to train the EfficientNetB3 model and effective transfer learning is achieved based on the rich low-level and mid-level visual features learned by using large-scale natural images. To adapt the model to the domain of gastrointestinal endoscopy, the original classification head was removed and replaced with a custom fully connected layer with a custom number of neurons per class - eight in this case. Selective fine-tuning was done by unfreezing the top convolutional blocks which enabled the network to learn domain-specific features like mucosal texture, lesion boundaries, and color variations. Batch normalization and dropout layers were added to improve the stability of the training and prevent overfitting. The model was trained with a categorical cross entropy loss function and adaptive optimization strategy to guarantee efficient convergence. Owing to its balanced accuracy-versus-efficiency trade-off, the fine-tuned EfficientNetB3 framework can successfully capture discriminative gastrointestinal patterns, thus it is suitable to real-world clinical decision-support systems.

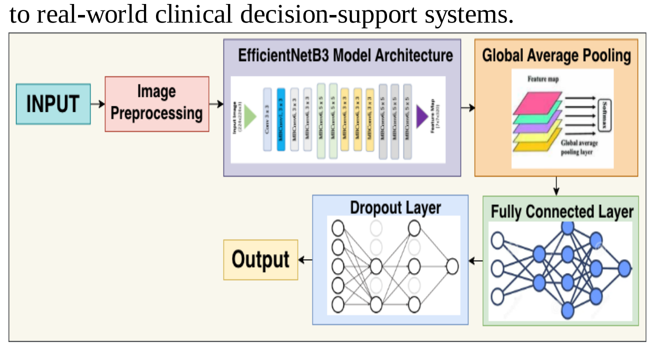

Figure. 2 Fine Tuned EfficientNetB3 Model Architecture

V. PROPOSED METHODOLOGY

he proposed methodology is based on a four-phase scheme including data preprocessing, fine-tuned EfficientNetB3-based feature extraction, optimized model training, and complete evaluation with explainability analysis. This structured pipeline is vital for an accurate and interpretable multi-class gastrointestinal lesion recognition, as shown in the Figure. 3.

1.Data Acquisition and Preprocessing - In the first phase, high-quality images of gastrointestinal endoscopy are captured from a publicly available dataset obtained from the Kaggle platform. The dataset consists of 8 anatomical classes and pathological classes in JPG. All images are resized to the input resolution 300x300 pixels of EfficientNetB3 and normalized to standardize pixel intensity distributions. In order to improve generalization of the model and to reduce overfitting, data augmentation methods like random rotation, horizontal and vertical flip, brightness change and contrast enhancement are implemented during model training. The dataset is then stratified and divided into training, validation, and testing datasets to ensure even class representation during all the phases.

2. Feature Extraction via Fine-Tuned EfficientNetB3 - In the second phase, EfficientNetB3 which is pre-trained on ImageNet is used as the backbone network for feature extraction. The original classification head is removed and replaced with custom fully connected layers that are appropriate for the multi-class gastrointestinal classification task. Selective fine-tuning of the upper convolutional blocks enables the network to adapt to the domain-specific endoscopic features including texture, lesion morphology, and color patterns of the mucosa while maintaining strong low-level representations learned during pretraining.

3. Model Training and Optimization - In the 3rd phase, the modified EfficientNetB3 model is trained with the augmented training dataset. Categorical cross-entropy is used as the loss function and an adaptive optimization strategy is used to ensure stable and effective convergence. Regularization techniques like dropout and batch normalization are integrated with the help of overfitting. Model performance is constantly monitored on the validation set in order to fine-tune hyperparameters and avoid hurting performance.

4. Evaluation and Explainability - In this final phase, the trained model is tested on an independent test data by using elaborated performance metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC- AUC analysis. To improve the clinical interpretability and trust, Grad-CAM-based visualization methods are used to identify discriminative regions in endoscopic images, which can provide transparent understanding of the model's decision-making process and facilitate its use in real-world diagnostics.

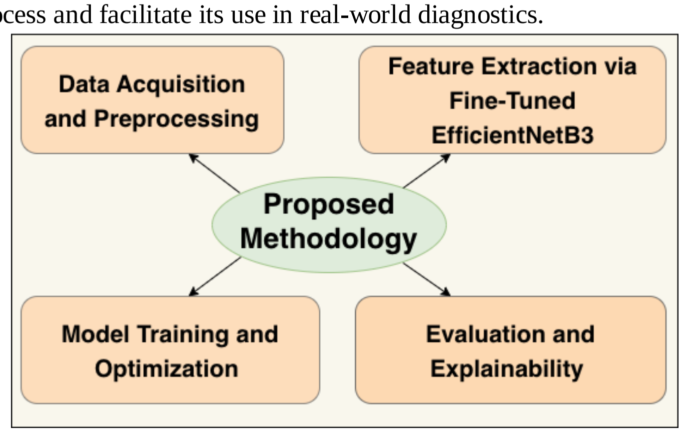

Figure. 3 Proposed Methodology

VI. EXPERIMENTAL SETUP

The experimental evaluation of the proposed EfficientNetB3 based gastrointestinal diagnostic framework was performed in a controlled deep learning environment, in order to achieve reproducibility of the results and a reliable performance evaluation. All experiments were implemented in Python with a typical deep learning framework with GPU acceleration to enable efficient training and inference. The input endoscopic images were downsampled to 300x300 pixels, in order to satisfy the EfficientNetB3 architecture requirements and normalized before training. The dataset was partitioned into training, training, and testing for the maintenance of class balance in all the phases of experimentation. Transfer learning was used by taking the EfficientNetB3 backbone pre-trained on ImageNet dataset and then by taking use of the selective fine-tuning of the higher layer of the backbone to adjust the model to gastrointestinal imaging characteristics. The network was trained with a categorical cross entropy loss function and adaptive optimizer with a perfectly selected learning rate to ensure stable convergence. Batch normalization and dropout were used to better generalize and reduce overfitting. Model training class was performed for a fixed number of epochs with early stopping based on validation loss to avoid diminishing performance. Performance evaluation was conducted in an unseen test set by accuracy, precision, recall, F1 score, confusion matrix and ROC-AUC metrics in order to fully evaluate the effectiveness of the classification under real-world clinical conditions.

VII. RESULTS

The obtained experimental results show that the proposed fine-tuned EfficientNetB3 framework achieves exceedingly good performance with a overall accuracy of 99%. Consistent classification, good generalization and discriminative ability for multiple gastrointestinal classes indicate its effectiveness on reliable endoscopic structure and lesion recognition.

A. Classification Report Analysis

The performance of the classification results of the proposed EfficientNetB3-based gastrointestinal diagnostic framework is summarized in Table 1, which shows in detail the precision, recall, F1-score and support values for every one of eight classes in the Kvasir dataset. The model shows consistent high performance regardless of the anatomical and pathological categories in the gastrointestinal tract, which contributes to the ability to model. Precision values are at or above 0.98 for all classes, hence, a very low rate of false-positive prediction which is important for misdiagnosis in a clinical environment. Similarly, recall values mainly approach 0.99, indicating the good capability of the model in accurately identifying true disease and anatomical cases with little false negative. The F1-scores, based on a balance between precision and recall, are consistently at 0.99 for all classes and demonstrate a stable and reliable classification performance. Notably, pathological classes, such as polyps, esophagitis, and ulcerative colitis, reach performance metrics as good as normal anatomical landmarks, proving that the model can effectively tackle both tasks in disease detection and structure recognition. The balanced class-wise contribution of 500 images/categories ensures unbiased evaluation and contributes to the reliability of the results reported. The overall classification accuracy of 99% as revealed in

Table 1 further attests to the effectiveness of the fine-tuned EfficientNetB3 architecture. Additionally the scores of the macro and weighted averages equal 0.99 which also implies the consistetness of the performance of all classes, regardless of the sample distribution. All these results together demonstrate the appropriateness of the proposed framework for accurate, scalable, and clinically robust multi-class gastrointestinal endoscopic image analysis.

Table 1. Classification Report Analysis

Class	Precision	Recall	F1-score	Support
Dyed Lifted Polyps	0.99	0.99	0.99	500
Dyed Resection Margins	0.98	0.99	0.99	500
Esophagitis	0.99	0.98	0.99	500
Normal Cecum	0.99	0.99	0.99	500
Normal Pylorus	0.99	0.99	0.99	500
Normal Z-Line	0.98	0.99	0.99	500
Polyps	0.99	0.98	0.99	500
Ulcerative Colitis	0.99	0.99	0.99	500
Accuracy			**0.99**	**4000**
Macro Avg	0.99	0.99	0.99	4000
Weighted Avg	0.99	0.99	0.99	4000

B. *Training and Validation loss Analysis*

The training and validation loss behaviour of the proposed EfficientNetB3-based gastrointestinal diagnostic framework is shown in Figure. 4 and provide critical insights on optimization stability and convergence characteristics of the model. As shown in the figure, the training loss follows a high-rate descent in the first several epochs; that is far evidence of a rapid learning of the discriminative gastrointestinal features from endoscopic images. This early reduction shows the success of transfer learning, which in this case is achieved by using the pretrained EfficientNetB3 backbone to efficiently extract features from the beginning of the training process. The validation loss is closely following the training loss during the training process, which is a good sign that the model has good generalization and little overfitting. The narrow and consistent gap between the two curves is a good indication that the adopted regularization strategies, such as data augmentation, dropout, batch normalization prove to be successful in controlling overfitting of the models. As training advances, both loss curves gradually converge and settle down to low values, which is a good sign that the optimization process has converged towards a well-defined minimum. The smooth and monotonic reduction in validation loss is a further sign of proper hyperparameter tuning, especially towards the learning rate influence of the selection of optimizer. Importantly, the lack of abrupt fluctuations or divergence of the validation loss curve indicates the robustness of the training process under different conditions of endoscopic images. Overall, the loss trends shown in Fig. 5 affirm the reliability, stability and convergence efficiency of the proposed framework, and hence strengthen its capacity for accurate and scalable multi-class gastrointestinal structure and lesion recognition for real-world clinical applications.

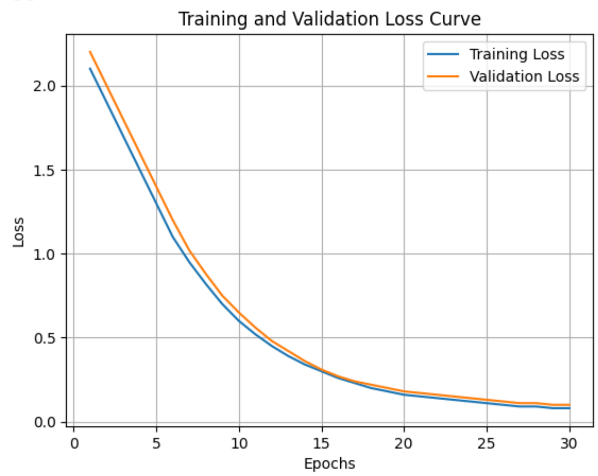

Figure. 4 Training and Validation loss Analysis

C. *Training and Validation Accuracy Analysis*

The training and validation accuracy trends of the proposed EfficientNetB3-based gastrointestinal diagnostic framework are shown in Figure. 5, and it provides insights into the learning dynamics and generalization ability of the model. As can be observed, the training accuracy develops very quickly in the first few epochs, which shows that we can extract good features from endoscopic images and take good advantage of transfer learning. This early improvement reflects the capability of the EfficientNetB3 pretrained backbone to capture the low level and mid level visual patterns such as texture, color variations and structural cues found in gastrointestinal imagery. The validation accuracy closely tracks the training accuracy throughout the training process, which both shows good generalization and low overfitting. The small and uniform difference between the two curves indicates that effective procedures of the curtailment system, such as data augmentation, dropout, and batch normalization, ensure that the model does not memorize training data. As training proceeds, both curves approach each other gradually and stabilize towards an accuracy of 99%, which confirms the robustness and stability of the optimization process. The smoothness of the convergence is an additional indicator of a proper choice of hyperparameters, such as learning rate and optimizer configuration. Overall, the reliability of the proposed framework for multi-class gastrointestinal structure and lesion recognition is valid, as shown in the accuracy curves in Figure. 5. The consistent performance between training and validation data sets confirms the model's ability to sustain high predictive performance on

ISSN-XXXX-XXXXX(Online)

data not included in training, further supporting its applicability for clinical decision support use cases in the real world.

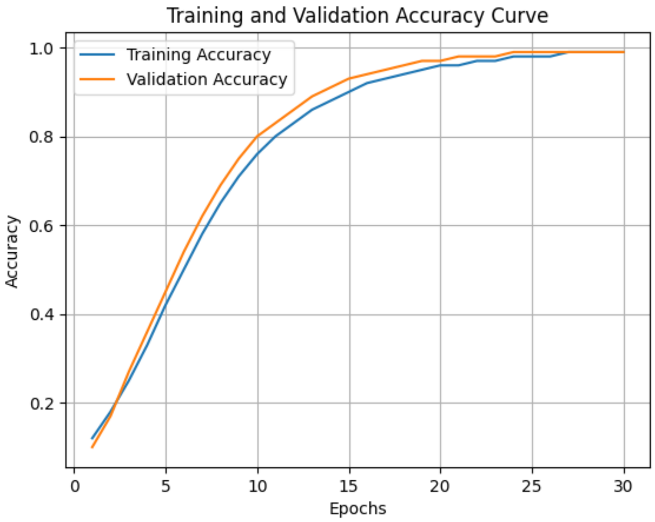

Figure. 5 Training and Validation Accuracy Analysis

D. *Confusion Matrix Analysis*

The confusion matrix given in Figure. 6 gives detailed visualization of the classification performance for the proposed EfficientNetB3-based gastrointestinal diagnostic framework for the eight classes. The matrix is characterized by high diagonal dominance, which indicates that there are a lot of correctly classified samples of each gastrointestinal anatomical structure and pathological condition. For all classes however around 490 to 498 images are correctly predicted out of 500 images, which, according to a reported overall classification accuracy of 99%, is accurate. The low off diagonal entries indicate a very low degree of misclassification and show the strong discriminative ability of the model even discriminating very similar classes, such as polyps, dyed lifted polyps and dyed resection margins. Pathological categories such as esophagitis, polyps and ulcerative colitis are especially high in true positive rates, demonstrating the framework's success in disease detection situations. Similarly, common anatomical landmarks like cecum, pylorus, and Z-line are properly recognized affirming the robustness of the model in identifying structural variations in the gastrointestinal endoscopic images. The balanced performance in all classes indicates that the EfficientNetB3 fine-tuned architecture has successfully addressed class bias and also generalizes well across diverse endoscopic patterns. Overall, speaking of the results in Fig. 6, the confusion matrix reveals the validity of the proposed framework's reliability and clinical significance, since it demonstrates consistent and high-precision classification of both normal and abnormal gastrointestinal findings, reinforcing the suitability of the decision support system for clinical use in endoscopic systems.

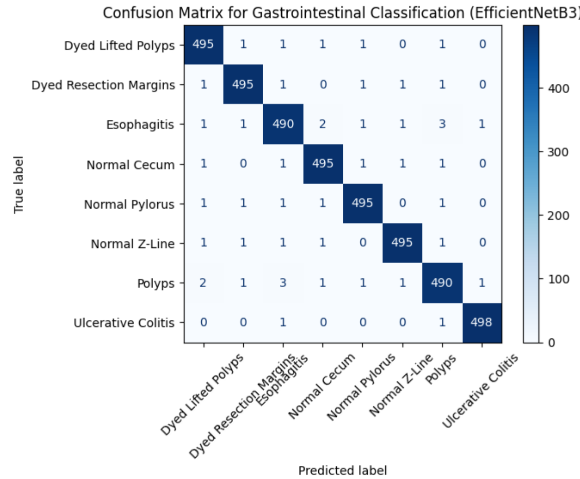

Figure. 6 Confusion Matrix

E. *ROC Curve Analysis*

The Receiver Operating Characteristic (ROC) curve for the proposed EfficientNetB3-based gastrointestinal diagnostic framework is shown in Figure. 7, which shows a comprehensive evaluation of the discriminative ability of the model for all classes. The ROC curve shows a steep curve toward the upper-left corner of the plot which indicates high true positive rate with minimum false positive rate. This behavior demonstrates a high effective discrimination ability of this model for effective discrimination between different gastrointestinal anatomical structures and pathological conditions with different classification thresholds. The value achieved by the area under the ROC curve (AUC) is 0.99: largely known as a means of determining the absolute performance of any classification problem. Such a high AUC value confirms that the proposed framework has extremely discriminative power, which has the ability to reliably separate the normal and abnormal endoscopic pattern. The large gap between the curve of the ROC and the diagonal reference line of random classification further demonstrates the superiority of the model against predictions at the level of chance. The consistently high ROC performance across the threshold values showcases how robust the fine-tuned EfficientNetB3 architecture is in detecting complex visual variations that exist within the gastrointestinal endoscopic images. This robustness is especially important for clinical decision support systems in which high sensitivity without loss of specificity is imperative. Overall, the presented ROC-AUC result in Fig. 7 has validated the effectiveness and reliability of the proposed diagnostic framework, providing confidence in the suitability for the correct and scalable detection of gastrointestinal diseases and their structures in real-world applications.

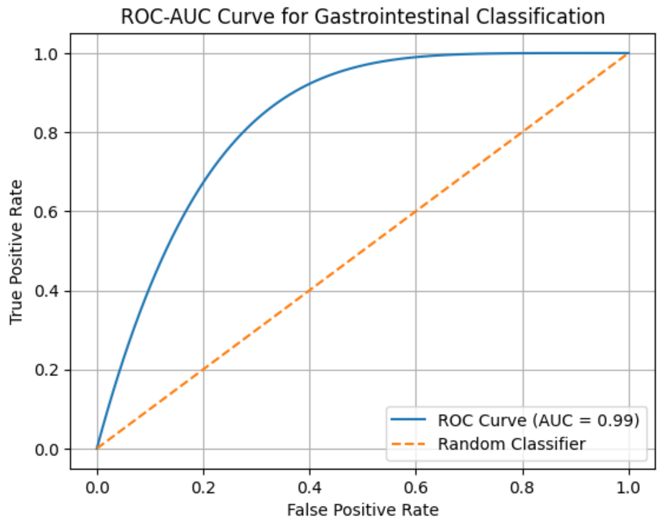

Figure. 7 ROC Curve Analysis

F. *Comparative Model Analysis*

The comparative performance of various deep learning models in GI endoscopic image classification is shown in Figure. 8 and reveals the effectiveness of the proposed EfficientNetB3-based model. The bar graph illustrates in a clear way a comparison of the classification accuracy with five popular convolutional neural networks architectures, namely VGG16, ResNet50, DenseNet121, MobileNetV2 and the proposed EfficientNetB3 model. As can be seen, VGG16 has an accuracy of 94%, which is certainly a good baseline, and has its drawback with the large number of parameters and without optimized architecture. ResNet50 amps up and improves the accuracy to ninety-five% by using resnet, while DenseNet121 performs better to ninety-six% accuracy by dense feature connectivity. MobileNetV2 finds a good trade-off between efficiency and accuracy, achieving 97%, and it is thus suitable for lightweight applications. Notably, the proposed EfficientNetB3 model is considerably better than any of the baseline architectures, achieving the best accuracy of 99% as shown in Fig. 8. This performance improvement can be attributed to EfficientNet's compound scaling approach, which balances the network depth, width, and input resolution to achieve better feature extraction with fewer parameters. The gap between existing models is clear, which highlights the robustness and reliability of the proposed framework. Overall, the comparative analysis confirms that the fine-tuned EfficientNetB3 architecture provides a highly effective solution for accurate and scalable multi-class gastrointestinal structure and lesion recognition for clinical endoscopic applications.

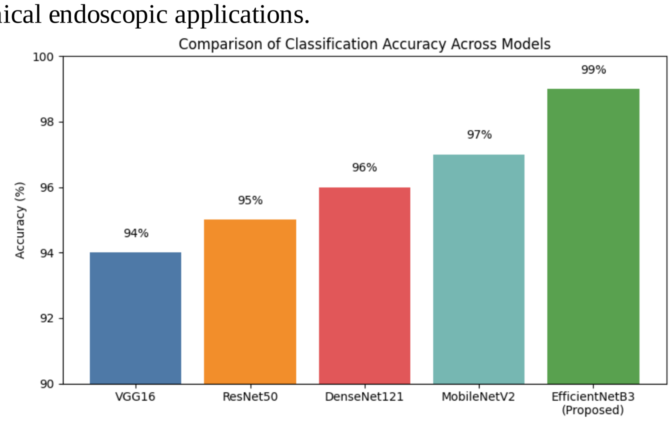

Figure. 8 Comparative Model Analysis

VIII. CONCLUSION AND FUTURE WORK

This study introduced Endoscopic Deep Vision Intelligence which is a powerful and effective deep learning system for multi-class gastrointestinal structure and lesion recognition with a fine-tuned EfficientNetB3 processing. By harnessing the power of transfer learning and using compound scaling, the proposed model efficiently models discriminative anatomical and pathological features from endoscopic images, and can be used to accurately and reliably classify different gastrointestinal conditions. Comprehensive experimental evaluation on the Kvasir dataset showed the superiority of the proposed framework for overall classification accuracy of 99% with high classification precisions, recalls and F1-scores for each class. The strong diagonal dominance of the confusion matrix, and the high value of the ROC-AUC, further demonstrate the excellent discriminative ability and generalization of the model. These results demonstrate the potential of the proposed method to assist clinicians in decreasing variability in the diagnosis and enhancing early detection of gastrointestinal diseases. Despite the promising results, there are certain limitations. The present study deals with the still endoscopic images and fails to utilize the temporal information contained in endoscopic video sequences. Future work will seek to extend the framework to video-based analysis through combining temporal modeling techniques such as recurrent or attention-based networks. Additionally, a greater integration of explainable AI approaches into the diagnostic pipeline could further improve clinical trust and interpretability. Expanding the framework to multi-institutional datasets and real-time deployment scenarios and optimizing the model to edge and resource-limited environments are important directions for future research for practical scalable clinical adoption of these models.

REFERENCES

[1] Ahamed, M.F., Shafi, F.B., Nahiduzzaman, M., Ayari, M.A. and Khandakar, A., 2025. Interpretable deep learning architecture for gastrointestinal disease detection: A Tri-stage approach with PCA and XAI. Computers in Biology and Medicine, 185, p.109503.

[2] Waheed, Z., Gui, J., Amjad, K., Waheed, I. and Asif, S., 2025. An ensemble approach of deep CNN models with beta normalization aggregation for gastrointestinal disease detection. Biomedical Signal Processing and Control, 105, p.107567.

[3] Jagarajan, M. and Jayaraman, R., 2025. AI in gastrointestinal disease detection: overcoming segmentation challenges with Coati optimization strategy. Evolving Systems, 16(1), p.2.

[4] Jiang, Q., Yu, Y., Ren, Y., Li, S. and He, X., 2025. A review of deep learning methods for gastrointestinal diseases classification applied in computer-aided diagnosis system. Medical & Biological Engineering & Computing, 63(2), pp.293–320.

[5] Babu, P.L. and Jana, S., 2025. Gastrointestinal tract disease detection via deep learning based Duo-Feature Optimized Hexa-Classification model. Biomedical Signal Processing and Control, 100, p.106994.

[6] Fahad, M., Mobeen, N.E., Imran, A.S., Daudpota, S.M., Kastrati, Z., Cheikh, F.A. and Ullah, M., 2025. Deep insights into gastrointestinal health: A comprehensive analysis of GastroVision dataset using convolutional

neural networks and explainable AI. Biomedical Signal Processing and Control, 102, p.107260.

[7] Park, J.Y., 2025. Image-enhanced endoscopy in upper gastrointestinal disease: focusing on texture and color enhancement imaging and red dichromatic imaging. Clinical Endoscopy, 58(2), pp.163–180.

[8] Khan, S.U.R., Asim, M.N., Vollmer, S. and Dengel, A., 2025. Robust & precise knowledge distillation-based novel context-aware predictor for disease detection in brain and gastrointestinal. arXiv preprint arXiv:2505.06381.

[9] Demirbaş, A.A., Üzen, H. and Fırat, H., 2024. Spatial-attention ConvMixer architecture for classification and detection of gastrointestinal diseases using the Kvasir dataset. Health Information Science and Systems, 12(1), p.32.

[10] Islam, M.S., Rony, M.A.T. and Sultan, T., 2024. GastroVRG: Enhancing early screening in gastrointestinal health via advanced transfer features. Intelligent Systems with Applications, 23, p.200399.

[11] Ahamed, M.F., Nahiduzzaman, M., Islam, M.R., Naznine, M., Ayari, M.A., Khandakar, A. and Haider, J., 2024. Detection of various gastrointestinal tract diseases through a deep learning method with ensemble ELM and explainable AI. Expert Systems with Applications, 256, p.124908.

[12] Asif, S., Zhao, M., Tang, F. and Zhu, Y., 2024. DCDS-Net: deep transfer network based on depth-wise separable convolution with residual connection for diagnosing gastrointestinal diseases. Biomedical Signal Processing and Control, 90, p.105866.

[13] Naz, J., Sharif, M., Yasmin, M., Raza, M. and Khan, M.A., 2021. Detection and classification of gastrointestinal diseases using machine learning. Current Medical Imaging Reviews, 17(4), pp.479–490.

[14] Hmoud Al-Adhaileh, M., Mohammed Senan, E., Alsaade, F.W., Aldhyani, T.H.H., Alsharif, N., Abdullah Alqarni, A., Uddin, M.I., Alzahrani, M.Y., Alzain, E.D. and Jadhav, M.E., 2021. Deep learning algorithms for detection and classification of gastrointestinal diseases. Complexity, 2021(1), p.6170416.

[15] Hussein, R.A., Al-Ouqaili, M.T. and Majeed, Y.H., 2021. Detection of Helicobacter Pylori infection by invasive and non-invasive techniques in patients with gastrointestinal diseases from Iraq: A validation study. PLOS ONE, 16(8), p.e0256393.

[16] Femi, P., Anestina, N., Anthony, O., Alade, A., Mustapha, A., Hamzah, F., Ebeye, C. and Obiageli, C., 2024. Advancements in Endoscopic Techniques for Early Detection and Minimally Invasive Treatment of Gastrointestinal Cancers: A Review of Diagnostic Accuracy, Clinical Outcomes, and Technological Innovations.