INTERNATIONAL JOURNAL OF ECONOMIC SOCIAL SCIENCE AND MANAGEMENT LAW
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3070-0558
VOL.7, NO.2 (2026)

A ROBUST DEEP LEARNING APPROACH FOR AUTOMATED DETECTION AND CLASSIFICATION OF GASTROINTESTINAL TRACT DISEASES IN ENDOSCOPY IMAGES

Rajeshree M. Patil
Department of Electronics and Telecommunication Engineering,
Jhulelal Institute of Technology, Nagpur, India
Dr. Shubhangi Giripunje
Department of Electronics and Telecommunication Engineering,
Jhulelal Institute of Technology, Nagpur, India

Abstract— Gastrointestinal (GI) disorders represent a major global health burden, and accurate early diagnosis is critical for effective treatment and improved patient outcomes. Endoscopic imaging is a primary diagnostic tool; however, manual interpretation is subjective and prone to errors. This study presents a deep learning-based approach for automated classification of GI diseases using EfficientNet-B1 and ResNet-101 models on the Kvasir-V2 dataset. Image preprocessing and supervised learning techniques were applied to enhance feature extraction and model generalization. Experimental results indicate that EfficientNet-B1 outperformed ResNet-101, achieving an accuracy of 94.71% with high precision, recall, and F1-score values. The proposed framework demonstrates the effectiveness of deep learning in supporting gastroenterologists by providing reliable and accurate diagnostic predictions. This approach has the potential to improve clinical workflows and reduce diagnostic variability in endoscopic examinations.

Keywords— Gastrointestinal diseases, Endoscopy, Deep learning, EfficientNet-B1, ResNet-101, Kvasir-V2, Medical image classification.

I. INTRODUCTION

Gastrointestinal (GI) tract disorders, including colorectal cancer, inflammatory bowel disease, and gastroesophageal reflux disease, represent a significant global health challenge and contribute substantially to morbidity and mortality worldwide. Early and accurate diagnosis of these disorders is essential for reducing disease progression, improving patient survival, and minimizing long-term complications [11]. Effective disease management depends heavily on timely detection and precise clinical interpretation by medical professionals, which remains a complex and demanding task [11].

Endoscopy is considered one of the most reliable diagnostic tools for GI disorders, as it enables direct visualization of the gastrointestinal tract and allows biopsy-based confirmation of pathological conditions. However, manual interpretation of endoscopic images is subjective, time-consuming, and prone to inter-observer variability, fatigue-related errors, and inconsistent diagnostic outcomes. Therefore, automated and intelligent diagnostic systems are increasingly required to support gastroenterologists in clinical decision-making.

Recent advancements in artificial intelligence, particularly deep learning (DL), have revolutionized medical image analysis by enabling automated feature extraction and classification. Convolutional neural networks (CNNs) and transformer-based models have demonstrated outstanding performance in medical image classification tasks [4,8]. Nevertheless, endoscopic image analysis remains challenging due to variations in illumination, anatomical structures, and imaging artifacts such as bubbles, blur, and noise, which significantly affect model performance and generalization [6].

To address these challenges, this study evaluates the effectiveness of EfficientNet-B1 and Transformer-based networks (TRSNet) for diagnosing and classifying gastrointestinal diseases from endoscopic images. EfficientNet-B1 employs compound scaling to optimize model depth, width, and resolution, enabling high classification accuracy with fewer parameters and computational efficiency.

Received: 21-04-2026 | Accepted: 29-05-2026 | Published: 07-06-2026 | www.ijeml.com

Transformer-based architectures, such as TRSNet, utilize self-attention mechanisms to capture long-range dependencies and global contextual features, providing improved representation learning for complex medical images.

Several studies have demonstrated the potential of deep learning-based frameworks for GI disease diagnosis. Hybrid deep learning and conventional machine learning models have achieved high classification accuracies on benchmark datasets such as Kvasir and HyperKvasir [12]. Optimization techniques, handcrafted feature fusion, and preprocessing strategies have further enhanced diagnostic accuracy and robustness [9,1]. Ensemble learning approaches integrating multiple deep learning models have also shown superior performance in detecting GI abnormalities [12]. Moreover, deep CNN frameworks optimized with evolutionary algorithms and attention mechanisms have achieved accuracy exceeding 95% on multiple datasets [16]. Transformer-based and fusion-based models have been reported to improve classification performance by capturing contextual and spatial features simultaneously [3].

Despite these advancements, several challenges remain, including dataset imbalance, noisy imaging conditions, limited generalization across datasets, and high computational complexity [6]. Therefore, there is a need for efficient and generalized deep learning frameworks that can provide reliable diagnostic performance under real-world clinical conditions.

The primary objective of this research is to develop an intelligent automated diagnostic framework using EfficientNet-B1 and Transformer-based architectures for gastrointestinal disease classification from endoscopic images. The study systematically evaluates model performance, identifies strengths and limitations, and provides insights into their clinical applicability. The anticipated outcomes of this research will contribute to improving automated medical diagnostic systems, enhancing clinical workflows, and supporting gastroenterologists in achieving accurate and consistent diagnostic results. Ultimately, integrating deep learning-based diagnostic tools into healthcare systems has the potential to significantly improve early detection, treatment planning, and patient outcomes.

II. LITERATURE REVIEW

Deep learning-based computer-aided diagnostic (CAD) systems have gained significant attention for automated gastrointestinal (GI) disease detection due to their ability to extract complex features from medical images. Several studies have proposed integrated deep learning frameworks to improve diagnostic accuracy and reduce human dependency in endoscopic examinations. An integrated CNN-based framework incorporating ResNet architectures was developed to classify GI tract pathologies and achieved an accuracy of 98.37%, demonstrating the feasibility of deep learning-based automated diagnosis systems in clinical environments [5,6].

Residual learning frameworks have been widely explored to enhance model stability and performance in medical image analysis. A ResNet-based model was implemented to classify gastrointestinal diseases from endoscopic images, achieving an overall accuracy of 94%. The study reported that deeper residual architectures provided improved stability and classification performance compared to shallower networks, highlighting the effectiveness of residual learning in GI disease detection tasks [3]. Furthermore, research has shown that ResNet-based feature extraction methods are highly efficient for image processing and classification tasks in medical imaging applications [4].

Hybrid frameworks combining deep learning and traditional machine learning techniques have also been investigated. An ensemble-based framework integrating deep learning and conventional machine learning methods demonstrated high classification performance, achieving an accuracy of 98.42% on the Kvasir dataset and 98.53% on the HyperKvasir dataset. These results indicate that hybrid and ensemble approaches can significantly improve classification robustness and generalization performance [12]. Additionally, combining deep learning models with optimization techniques has been reported to enhance detection accuracy and

Received: 21-04-2026 | Accepted: 29-05-2026 | Published: 07-06-2026 | www.ijeml.com

clinical workflow efficiency, enabling reliable diagnostic support for clinicians [8].

Feature engineering and preprocessing techniques have been widely explored to improve model robustness. A wavelet transform-based deep learning framework achieved classification accuracies of 97.25% and 93.75% for multi-level GI disease classification, demonstrating the importance of multiresolution feature extraction in medical image analysis [9]. Another study proposed a deep saliency-based model with Bayesian optimization for feature selection, achieving up to 99.61% accuracy on multiple datasets, highlighting the role of feature selection in improving classification outcomes [16].

Fusion-based deep learning frameworks have also been proposed to enhance disease classification performance. Two-stream fusion models combining features from multiple streams improved classification accuracy and addressed challenges related to imaging artifacts and endoscopic video analysis [3]. These studies emphasize that integrating multiple feature representations can significantly enhance disease discrimination capability in complex medical images.

Several studies have highlighted the importance of developing robust and generalized diagnostic frameworks for clinical applications. Deep learning-based automated diagnostic systems have demonstrated the potential to improve medical diagnosis infrastructure and support clinical decision-making by providing consistent and accurate predictions [2]. Moreover, the integration of advanced computational models into healthcare systems is expected to provide transformative diagnostic capabilities beyond traditional commercial approaches, benefiting both clinicians and patients [6].

Overall, the reviewed literature demonstrates that deep learning architectures, hybrid frameworks, ensemble models, feature engineering techniques, and fusion strategies significantly enhance GI disease classification performance. However, challenges such as computational complexity, generalization across datasets, and robustness to noisy imaging conditions remain open research problems. Therefore, further research is required to develop efficient, generalized, and clinically applicable deep learning frameworks for automated gastrointestinal disease diagnosis.

III. PROPOSED SYSTEM

Localizing and distinguishing gastrointestinal (GI) tract diseases from endoscopic images is a fundamental task in modern medical diagnostics, as it assists clinicians in making timely interventions and improves patient outcomes [11]. This study proposes a deep learning–based framework utilizing EfficientNet-B1 and ResNet-101 architectures for automated GI disease classification using the Kvasir-V2 dataset. Endoscopic images often exhibit subtle inter-class variations, such as dyed-lifted polyps and dyed resection margins, as well as normal anatomical structures including the cecum, pylorus, and z-line. These visual similarities make clinical differentiation challenging and necessitate robust feature learning models. The proposed framework aims to address these challenges and highlights the potential of deep learning-based diagnostic systems for accurate and timely GI disease detection.

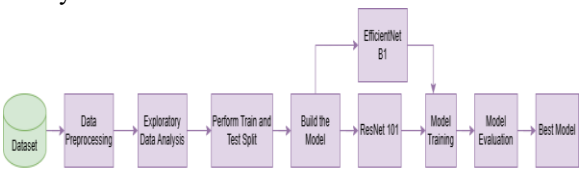

Fig. 1. Workflow diagram for GI disease classification

A. Data Description

The Kvasir-V2 dataset is selected for this study due to its comprehensive collection of endoscopic images representing a wide range of gastrointestinal conditions. The dataset includes multiple disease classes such as polyps, ulcers, and inflammatory conditions, along with normal anatomical variations including the cecum, pylorus, and z-line. In addition to prominent pathological features, the dataset also contains subtle manifestations such as dyed margins and minor lesion characteristics, which are critical for evaluating model robustness. Each image is annotated with a corresponding class label, enabling the use of supervised learning techniques for disease detection and classification [12]. Expanding the dataset with additional normal and pathological cases is identified as a future direction to further enhance model generalization and clinical reliability.

Received: 21-04-2026 | Accepted: 29-05-2026 | Published: 07-06-2026 | www.ijeml.com

B. Image Preprocessing

Image preprocessing is a crucial step in preparing endoscopic images for deep learning model training. In this study, all images are resized to a fixed resolution compatible with EfficientNet-B1 and ResNet-101 input requirements to ensure uniformity and computational efficiency [4]. Resizing also reduces memory consumption and accelerates training without compromising diagnostic features. The dataset is divided into training and validation subsets to assess model generalization on unseen data. Additionally, batch-wise data loading is employed to improve training efficiency and stabilize gradient updates during optimization. These preprocessing steps enhance data quality and improve the overall effectiveness of the learning process [4].

C. Image Visualization

To better understand the dataset characteristics, sample images from the training dataset are visualized along with their corresponding class labels, as shown in Fig. 2.

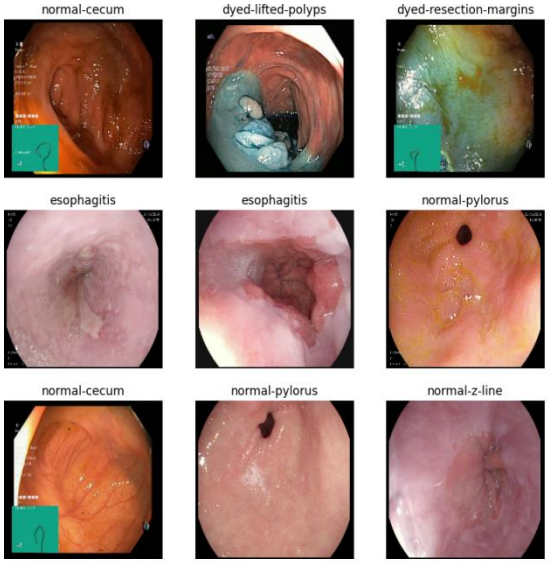

Fig. 2. Classes Visualization of dataset

Each image is displayed within a structured grid layout, allowing visual inspection of class diversity and intra-class variations. Dataset visualization provides valuable insights into feature complexity, class imbalance, and visual similarity among categories, which supports informed model selection and evaluation.

D. Model Selection and Implementation

Deep learning architectures were selected in this study to effectively capture complex spatial and contextual features from gastrointestinal endoscopic images. Two state-of-the-art convolutional neural network models, EfficientNet-B1 and ResNet-101, were employed for GI disease classification due to their proven performance and robustness in medical image analysis tasks.

EfficientNet-B1 belongs to the EfficientNet family, which introduces a compound scaling strategy to uniformly scale network depth, width, and input resolution to achieve optimal accuracy and computational efficiency [16]. Unlike traditional scaling methods that arbitrarily increase model size, EfficientNet uses a principled approach to balance model capacity and efficiency. The compound scaling is defined as:

where is the scaling coefficient, and , , and are constants determined empirically to maintain a balanced trade-off between accuracy and computational cost. EfficientNet-B1 is a scaled version of the baseline model that offers improved representational power and classification accuracy, making it suitable for complex medical imaging datasets such as Kvasir-V2.

ResNet-101 is a deep residual learning architecture designed to address the vanishing gradient problem in very deep networks through identity-based skip connections. These residual connections enable the network to learn deeper hierarchical representations by facilitating effective gradient flow during backpropagation. ResNet-101 has demonstrated strong performance in medical image classification due to its ability to extract discriminative and high-level semantic features [4].

In this study, both EfficientNet-B1 and ResNet-101 were fine-tuned using transfer learning to adapt pretrained weights to GI-specific features. The models were trained on the Kvasir-V2 dataset using

Received: 21-04-2026 | Accepted: 29-05-2026 | Published: 07-06-2026 | www.ijeml.com

supervised learning, and their performance was evaluated using standard classification metrics. The comparative analysis of these architectures allows the assessment of accuracy, computational efficiency, and generalization capability for automated gastrointestinal disease diagnosis.

IV. RESULTS AND DISCUSSION

The Results and Discussion section involves the analysis of the outcomes and the implications of the method that was proposed for gastrointestinal tract disease detection and classification from endoscopy images using EfficientNet B0 and B1 as architectures.

A. Analysis of EfficientNet B1

The EfficientNet-B1 model demonstrated strong performance in gastrointestinal disease classification on the Kvasir-V2 dataset. The model achieved a loss value of 0.2298, indicating stable training convergence and effective feature learning. The classification accuracy reached 94.71%, confirming the model's ability to correctly classify multiple GI disease and normal anatomical categories.

In terms of precision, EfficientNet-B1 obtained 94.88%, indicating a low false-positive rate and reliable prediction of disease classes. The recall value of 94.79% shows that the model effectively identified true pathological cases without missing significant abnormalities. The F1-score of 94.81% further demonstrates a balanced performance between precision and recall, which is essential for clinical diagnostic applications.

The superior performance of EfficientNet-B1 can be attributed to its compound scaling strategy, which optimally balances network depth, width, and resolution. This enables the model to capture fine-grained visual patterns in endoscopic images, including subtle lesion boundaries and texture variations. Overall, EfficientNet-B1 proved to be a robust and efficient model for automated GI disease detection.

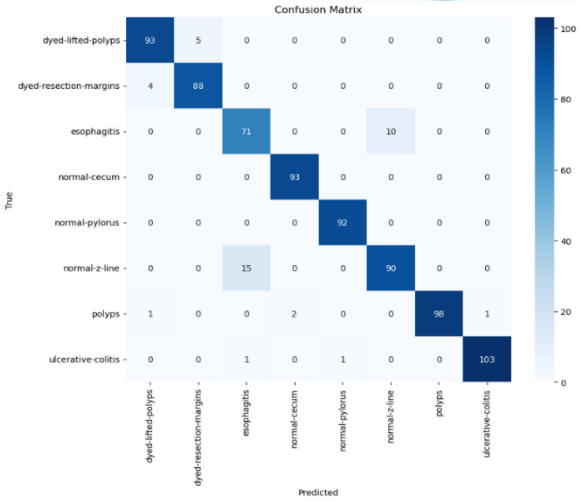

Fig. 3. Confusion matrix for EfficientNet B1

EfficientNet-B1 showed high diagonal dominance in the confusion matrix, indicating strong class-wise prediction accuracy. Minor misclassifications were observed between visually similar classes such as esophagitis and normal z-line, due to subtle texture and color similarities in endoscopic images.

B. Analysis of ResNet101

The ResNet-101 model also achieved competitive performance in classifying gastrointestinal diseases. The model recorded a loss value of 0.3445, which indicates effective learning but slightly less optimal convergence compared to EfficientNet-B1. ResNet-101 achieved an accuracy of 92.19%, demonstrating its capability to extract discriminative features from endoscopic images.

The precision of ResNet-101 was 92.31%, indicating good classification reliability with a moderate false-positive rate. The recall value of 92.19% suggests that the model successfully detected most disease cases, although some pathological instances were misclassified. The F1-score of 92.16% reflects a balanced but slightly lower performance compared to EfficientNet-B1.

ResNet-101 benefits from its deep residual learning architecture, which mitigates the vanishing gradient problem and allows effective training of very deep networks. The residual connections facilitate

hierarchical feature extraction, enabling the model to learn complex spatial representations from medical images. However, its higher computational complexity and lack of compound scaling may contribute to slightly lower performance compared to EfficientNet-B1.

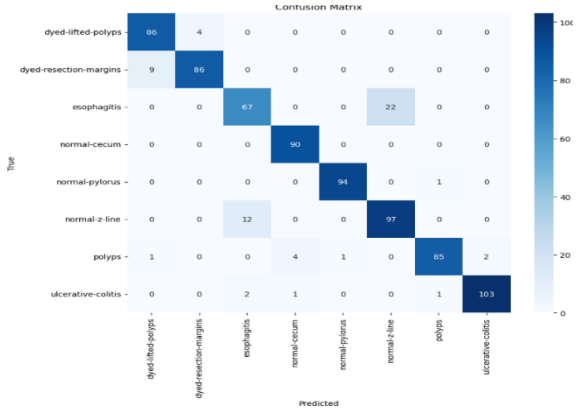

Fig. 4. Confusion matrix of ResNet 101

ResNet-101 also demonstrated reliable classification performance, but with slightly higher confusion among subtle lesion and normal anatomical classes. Misclassification was mainly observed between esophagitis, dyed margins, and normal structures, highlighting the challenge of distinguishing fine-grained mucosal variations.

V. CONCLUSION

EfficientNet B1 achieved a comparable performance to EfficientNet B0 with fewer correct predictions, although its statistical composition matched the precision and F1-score at the same level of accuracy at 94.71%. The decision algorithm indicates that EfficientNet B1 is a dependable solution because it maintains precision while preserving the generalization effectiveness.

ResNet 101 achieved an accuracy rate of 92.19%, yet revealed higher loss values at 0.3445, which indicates a possible susceptibility to overfitting or limited ability compared with EfficientNet models when dealing with complex datasets. Research demonstrates that DenseNet 121 achieves satisfactory results with an accuracy of 93.63%, yet falls behind EfficientNet models in precision, recall, and F1-score examinations, indicating limited robustness during this application procedure.

Future Research

Future research will focus on expanding the dataset with diverse clinical cases, integrating attention mechanisms and ensemble learning, and improving model interpretability. Real-time clinical validation and deployment in endoscopic systems will also be explored to enhance practical diagnostic reliability.

ACKNOWLEDGMENT

I sincerely acknowledge the Department of Electronics and Telecommunication Engineering, Jhulelal Institute of Technology, Nagpur, India, for providing the necessary facilities and support to carry out this research. I am deeply thankful to Dr. Shubhangi Giripunje for her valuable guidance and continuous support throughout this study. I also express my gratitude to the creators of the Kvasir-V2 dataset for making the dataset publicly available, which significantly contributed to this research.

REFERENCES

Afraz, M., Fayyaz, A. M., & Haseeb, A. (2023). A unified paradigm of classifying GI tract diseases in endoscopy images using multiple features fusion. Azerbaijan Journal of High Performance Computing, 6(1), 49–76. https://doi.org/10.32010/26166127.2023.6.1.49.76

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010. https://doi.org/10.1093/database/baaa010

Amin, M. S., Shah, J. H., Yasmin, M., Ansari, G. J., Khan, M. A., Tariq, U., Kim, Y. J., & Chang, B. (2022). A two stream fusion assisted deep learning framework for stomach diseases classification. CMC-Computers, Materials &

Continua, 73, 4423–4439. https://doi.org/10.32604/cmc.2022.030432

Ashraf, R., Habib, M. A., Akram, M., Latif, M. A., Malik, M. S. A., Awais, M., Dar, S. H., Mahmood, T., Yasir, M., & Abbas, Z. (2020). Deep convolution neural network for big data medical image classification. IEEE Access, 8, 105659–105670. https://doi.org/10.1109/ACCESS.2020.2998808

Benali Amjoud, A., & Amrouch, M. (2020). Convolutional neural networks backbones for object detection. In Image and signal processing: 9th International Conference, ICISP 2020 (pp. 282–289). Springer International Publishing. https://doi.org/10.1007/978-3-030-51935-3_30

Browning, C. M., Cloutier, R., Rich, T. C., & Leavesley, S. J. (2022). Endoscopy lifetime systems architecture: Scoping out the past to diagnose the future technology. Systems, 10(5), 189. https://doi.org/10.3390/systems10050189

De Las Casas, L. E., & Hicks, D. G. (2021). Pathologists at the leading edge of optimizing the tumor tissue journey for diagnostic accuracy and molecular testing. American Journal of Clinical Pathology, 155(6), 781–792. https://doi.org/10.1093/ajcp/aqaa212

Du, W., Rao, N., Liu, D., Jiang, H., Luo, C., Li, Z., Gan, T., & Zeng, B. (2019). Review on the applications of deep learning in the analysis of gastrointestinal endoscopy images. IEEE Access, 7, 142053–142069. https://doi.org/10.1109/ACCESS.2019.2944676

Escobar, J., Sanchez, K., Hinojosa, C., Arguello, H., & Castillo, S. (2021, September). Accurate deep learning-based gastrointestinal disease classification via transfer learning strategy [Paper presentation]. XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), IEEE. https://doi.org/10.1109/STSIVA53688.2021.9591995

Galati, F., Ourselin, S., & Zuluaga, M. A. (2022). From accuracy to reliability and robustness in cardiac magnetic resonance image segmentation: A review. Applied Sciences, 12(8), 3936. https://doi.org/10.3390/app12083936

Gikas, A., & Triantafillidis, J. K. (2014). The role of primary care physicians in early diagnosis and treatment of chronic gastrointestinal diseases. International Journal of General Medicine, 159–173. https://doi.org/10.1080/00016489.2021.1982148

Gunasekaran, H., Ramalakshmi, K., Swaminathan, D. K., & Mazzara, M. (2023). GIT-Net: An ensemble deep learning-based GI tract classification of endoscopic images. Bioengineering, 10(7), 809. https://doi.org/10.3390/bioengineering10070809

Jie, Y., Ji, X., Yue, A., Chen, J., Deng, Y., Chen, J., & Zhang, Y. (2020). Combined multi-layer feature fusion and edge detection method for distributed photovoltaic power station identification. Energies, 13(24), 6742. https://doi.org/10.3390/en13246742

Kaur, M., Singh, D., Roy, S., & Amoon, M. (2023). Efficient skip connections-based residual network (ESRNet) for brain tumor classification. Diagnostics, 13(20), 3234. https://doi.org/10.3390/diagnostics13203234

Kalshetty, R., & Parveen, A. (2023). Abnormal event detection model using an improved ResNet101 in context aware surveillance system. Cognitive Computation and Systems, 5(2), 153–167. https://doi.org/10.1049/ccs2.12084

Khan, M. A., Sahar, N., Khan, W. Z., Alhaisoni, M., Tariq, U., Zayyan, M. H., Kim, Y. J., & Chang, B. (2022). GestroNet: A framework of saliency estimation and optimal deep learning features based gastrointestinal diseases detection and classification. Diagnostics, 12(11), 2718. https://doi.org/10.3390/diagnostics12112718

Land, K. J., Boeras, D. I., Chen, X. S., Ramsay, A. R., & Peeling, R. W. (2019). REASSURED

diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. Nature Microbiology, 4(1), 46–54. https://doi.org/10.1038/s41564-018-0295-3