International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-13

Deep Learning Driven Gastrointestinal Disease Diagnosis from WCE Images: A Hybrid Approach

Kaveri Rajaram Bhosle[1], Prashant M Yawalkar[2]
[1]Student Department of Computer Engineering MET Institute of Engineering Nashik
[2]Dept. of AI & Professor and Head Department of Computer Engineering MET Institute of Engineering Nashik

Abstract: - Wireless Capsule Endoscopy (WCE) has revolution-ized gastrointestinal (GI) diagnostics by enabling non-invasive visualization of the entire digestive tract. Despite its clinical advantages, a single WCE examination produces tens of thou-sands of image frames, making manual analysis time-consuming and prone to inter-observer variability. In recent years, arti-ficial intelligence—particularly deep learning—has emerged as a powerful tool for automated GI disease classification. This review presents a comprehensive analysis of existing approaches for WCE-based GI image analysis, including traditional ma-chine learning methods, convolutional neural networks (CNNs), transformer-based architectures, and hybrid models. The paper critically examines commonly used benchmark datasets such as Kvasir, Kvasir-Capsule, HyperKvasir, and WCECCD, along with evaluation metrics and optimization strategies adopted in recent studies. Furthermore, the role of explainable artificial intelligence techniques, including attention mechanisms and Grad-CAM, is discussed in enhancing model interpretability and clinical trust. Key challenges such as class imbalance, limited annotated data, cross-dataset generalization, and real-time deployment constraints are identified. Finally, emerging research directions including multimodal learning, domain adaptation, temporal video modeling, and foundation models for medical imaging are outlined. This review aims to provide researchers and clinicians with a structured understanding of current advancements and future opportunities in automated GI disease diagnosis.

Keywords: — Wireless Capsule Endoscopy, Gastrointestinal Disease Classification, Deep Learning, EfficientNet, Convolu-tional Neural Networks, Medical Image Analysis, Explainable Artificial Intelligence, Grad-CAM, Computer-Aided Diagnosis.

I. Introduction

Gastrointestinal (GI) disorders such as colorectal cancer, inflammatory bowel disease, ulcers, and polyps pose a sig-nificant global health challenge. Early and accurate di-ag-nosis is essential for effective treatment and improved pa-tient outcomes. Conventional diagnostic techniques, including colonoscopy and upper gastrointestinal endoscopy, provide direct visualization but are invasive, costly, and require spe-cialized expertise [1].

Wireless Capsule Endoscopy (WCE) has emerged as an important advancement in GI diagnostics. The capsule device captures tens of thousands of images while traversing

the digestive system, enabling detailed examination of regions that are difficult to access using traditional endoscopy [2]. Despite its clinical advantages, WCE generates a massive volume of image data, making manual review labor-intensive and susceptible to human fatigue [3]. This challenge has motivated the development of automated computer-aided diagnostic systems.

Early efforts in automated GI image analysis relied on traditional machine learning techniques using handcrafted features combined with classical classifiers. Although these approaches demonstrated feasibility, their performance was limited by feature design constraints and poor generalization across datasets [4]. The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), enabled end-to-end feature learning directly from raw images, resulting in significant performance improvements in multiclass GI disease classification tasks [5].

While CNN-based models have achieved high accuracy, they often struggle to capture long-range spatial dependencies and global contextual information present in complex endoscopic imagery. To overcome these limitations, transformer-based architectures have been introduced for medical image analysis [6]. Vision Transformer variants and hybrid CNN–Transformer frameworks have demonstrated improved robustness and representational capability in lesion detection and classification [7].

In addition, explainable artificial intelligence techniques such as Grad-CAM have been incorporated to enhance interpretability and clinical trust in automated diagnostic systems [8]. Given the rapid advancements in deep learning architectures and optimization strategies, a structured synthesis of existing research is essential.

This paper presents a comprehensive review of state-of-the-art deep learning approaches for GI disease diagnosis using WCE images. We analyze traditional machine learning methods, CNN-based models, transformer-driven architectures, hybrid frameworks, and explainability techniques, followed by identification of research gaps and future research directions.

II. Background: Wireless Capsule Endoscopy And Gi Image Analysis

Wireless Capsule Endoscopy (WCE) is a minimally invasive diagnostic technology designed to visualize the gastrointestinal tract, particularly the small intestine, which is difficult to access using conventional endoscopy. The capsule is equipped with a miniature camera, light source, battery, and wireless transmitter that capture and transmit images to an external recording device as it travels naturally through the digestive system [9]. A single WCE examination may generate more than 50,000 frames, providing extensive visual information for detecting abnormalities such as bleeding, ulcers, tumors, and inflammatory lesions.

Despite its diagnostic value, WCE presents significant analytical challenges. The enormous volume of image data requires prolonged manual inspection by gastroenterologists, increasing the likelihood of fatigue-related errors and inter-observer variability [10]. Additionally, many gastrointestinal abnormalities exhibit subtle visual

differences, making dif-ferentiation between normal and pathological tissue difficult, especially in early disease stages.

Another technical challenge arises from variations in illumi-nation, motion blur, intes-tinal content, and viewpoint changes during capsule movement. These factors intro-duce noise and variability into captured images, complicating automated fea-ture extraction and classification [11]. Furthermore, publicly available WCE datasets often suffer from class imbalance and limited annotation quality, restricting the generaliza-tion capability of trained models [12].

To address these limitations, researchers have increasingly explored deep learning-based approaches capable of auto-matic feature representation learning. In particular, convo-lutional neural networks and transformer-based architectures have shown promise in extracting discriminative features from complex medical images [13]. However, achieving reliable performance across diverse datasets and real-world clini-cal settings remains an ongoing research challenge.

Understanding these technical and clinical constraints is es-sential before examining the evolution of artificial intelligence techniques applied to GI disease classification. The following sections provide a structured analysis of traditional machine learning methods and modern deep learning architectures used in WCE-based diagnosis.

III. Datasets Used In Gi Disease Research

The development of automated gastrointestinal (GI) disease classification systems relies heavily on the availability of well-annotated benchmark datasets. Publicly ac-cessible datasets have significantly accelerated research in Wireless Capsule Endos-copy (WCE) image analysis by enabling standardized evaluation and fair comparison of different machine learning and deep learning models.

One of the most widely used datasets is the Kvasir dataset, which contains annotated endoscopic images covering mul-tiple GI disease categories such as polyps, esopha-gitis, and ulcerative colitis [1]. Its extended variant, Kvasir-Capsule,

TABLE I: Publicly Available WCE and GI Datasets

Dataset	Imag-es/Frames	Classes	Application
Kvasir	8,000+	8	GI image classification
Kvasir-Capsule	47,000+	14	WCE abnor-mality detec-tion
HyperKvasir	110,000+	10+	Image & video analysis
WCECCD	4,000+	5	Capsule dis-ease detection

focuses specifically on capsule endoscopy images and includes a large number of labeled frames across diverse pathological findings [2]. These datasets are frequently adopted for multi-class classification tasks due to their public availability and structured annotations.

Another large-scale dataset is HyperKvasir, which provides both images and videos from upper and lower GI exami-nations, supporting research in classification, segmentation, and video-based analysis [3]. Similarly, the WCECCD dataset contains capsule endoscopy images categorized into clinically relevant disease classes and has been used for benchmarking automated WCE diagnostic systems [4].

Despite these contributions, challenges remain. Most datasets exhibit class imbalance, with significantly fewer sam-ples for rare abnormalities compared to normal findings. In addition, variations in imaging devices, acquisition conditions, and annotation protocols introduce domain shifts that can affect cross-dataset generalization [5]. Addressing these lim-itations through domain adaptation, data augmentation, and multi-center validation remains an active area of research.

IV. Literature Survey

The rapid advancement of deep learning has significantly transformed gastrointestinal (GI) disease classification using Wireless Capsule Endoscopy (WCE) images. Early deep learn-ing studies primarily relied on convolutional neural networks (CNNs) to automatically extract hierarchical image features. Khan et al. [1] proposed a sparse convolutional DenseNet architecture integrated with attention mechanisms and GRU units to enhance contextual representation. Their approach demonstrated strong generalization performance on Kvasir datasets.

Tanwar and Sharma [2] introduced EfficientViT, a hybrid framework combining EfficientNet-B0 and Vision Transformer (ViT), effectively capturing both local texture and global contextual relationships. Similarly, Dahan et al. [3] incor-porated explainability into transformer-based frameworks us-ing Swin Transformer and Grad-CAM, improving diagnostic transparency and reducing false negatives.

Malik et al. [4] explored ensemble CNN models such as VGG-19 combined with recurrent layers for multiclass GI disease classification, achieving competitive accuracy. El-Ghany et al. [5] proposed an intelligent learning rate control mechanism to stabilize CNN training, improving convergence speed and classification accuracy.

Transformer-based and hybrid architectures have recently gained attention due to their ability to model long-range spatial dependencies. Rajput et al. [10] developed a CNN–ViT fu-sion architecture that improved classification robustness across multiple GI classes. Sharma et al. [13] integrated DenseNet with Swin Transformer and saliency-based interpretability modules to enhance both performance and transparency.
To address data imbalance challenges, Rehman et al. [11] proposed a CapsuleNet-based classifier optimized for im-balanced datasets. Furthermore, Diamantis et al. [15] intro-duced a multiscale residual variational autoencoder (TIDE) for synthetic WCE image generation, mitigating limited dataset availability.

Overall, the literature demonstrates a clear evolution from traditional CNN architectures toward hybrid transformer-based, attention-enhanced, and explainable AI frameworks, aiming to improve robustness, interpretability, and clinical applicability.

V. Traditional Machine Learning Approaches

Before the widespread adoption of deep learning, gastroin-testinal (GI) image analysis primarily relied on handcrafted feature extraction combined with classical machine learning classifiers. These approaches focused on designing discrimina-tive texture, color, and shape descriptors to represent patholog-ical patterns in endoscopic and Wireless Capsule Endoscopy (WCE) images.

Commonly used feature extraction techniques included Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and wavelet-based texture descriptors [6]. These features were typically fed into classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forests, and Decision Trees for disease classification [7].

Several early studies demonstrated that handcrafted texture and color features could distinguish between bleeding, ulcers, and normal mucosa with moderate accuracy. Wavelet trans-forms, in particular, were widely explored due to their ability to capture multi-scale texture information and frequency-domain characteristics in GI images [8]. However, these approaches required careful feature engineering and domain expertise.

Despite their initial success, traditional machine learning methods exhibited significant limitations. Feature design was often dataset-specific and lacked robustness across different imaging conditions. Variations in illumination, motion blur, and intestinal content in WCE recordings negatively impacted performance. Furthermore, handcrafted descriptors struggled to capture complex hierarchical patterns inherent in medical images, limiting scalability to large and diverse datasets [9].

These limitations motivated the transition toward deep learning-based approaches capable of automatic feature repre-sentation learning. The following section discusses the emer-gence of convolutional neural networks and their impact on automated GI disease classification.

TABLE II: Comparison of Representative GI Disease Classification Models

Author	Model Type	Acc. (%)	Key Contribution
Khan [1]	Sparse CNN + GRU	96.7	Attention-enhanced hybrid model
Tanwar [2]	EfficientViT	95.2	CNN + Transformer fusion
Dahan [3]	Swin-T + ResNet	94.5	Explainable AI integration

Malik [4]	VGG-19 Ensem-ble	97.2	Multi-model ensemble learn-ing
Rajput [10]	CNN–ViT Hybrid	94.1	Global-local feature fusion
Rehman [11]	CapsuleNet	84.5	Class imbal-ance handling
Diamantis [15]	VAE (TIDE)	–	Synthetic data gen-eration

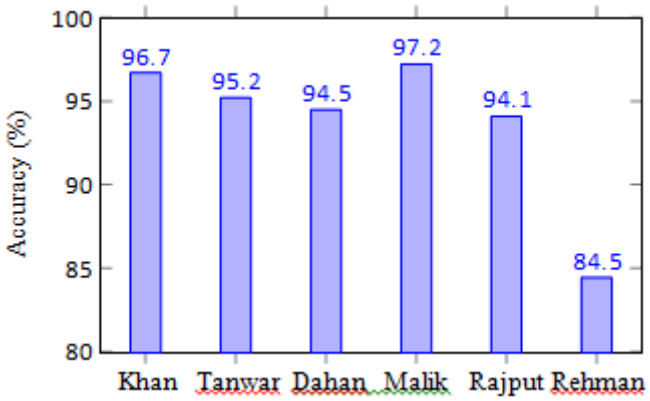

Fig. 1: Accuracy comparison of representative GI disease classification models.

The comparative analysis in Table II and Figure 1 high-lights that hybrid CNN–Transformer models consistently out-perform traditional standalone CNN architectures. Ensemble approaches achieve high accuracy but increase computational complexity. Transformer-based frameworks provide improved contextual modeling but require larger datasets and compu-tational resources. Recent research trends emphasize explain-ability, data efficiency, and domain generalization as critical factors for real-world clinical deployment.

VI. Deep Learning-Based Gi Disease Classification

The emergence of deep learning has significantly trans-formed automated gastrointestinal (GI) image analysis. Unlike traditional machine learning approaches that rely on hand-crafted features, deep learning models—particularly Convo-lutional Neural Networks (CNNs)—learn hierarchical feature representations directly from raw image data. This capability has led to substantial improvements in classification accuracy, robustness, and scalability for Wireless Capsule Endoscopy (WCE) applications.

Early CNN-based approaches focused on binary classifica-tion tasks such as bleeding detection. With the availability of larger annotated datasets, researchers extended these methods to multi-class GI disease classification. Pretrained architectures

TABLE IV: Limitations Of Existing GI Classification Approaches

Model Type	Strength	Limitation
CNN	High accuracy	Weak global context modeling
Transformer	Global feature learning	Data hungry, computationally expensive
Hybrid	Balanced feature extraction	Increased complexity
CapsuleNet	Spatial preservation	Limited scalability
VAE-Based	Data augmentation	Synthetic realism concerns

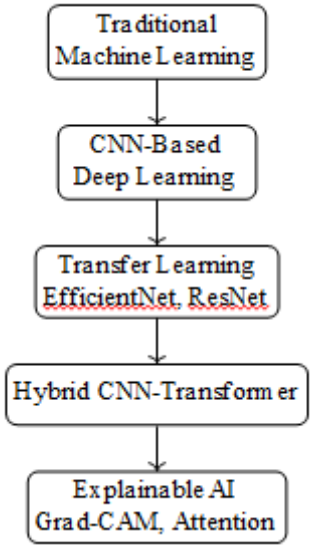

Fig. 2: Taxonomy of AI approaches in GI disease classification.

TABLE III: Research Gap Analysis in GI Disease Classification

Type	Strength	Limitation	Gap
Traditional ML	Low computation	Poor generalization	Not scalable
CNN Models	High accuracy	Limited global context	Subtle lesion confusion
Transformer	Global learning	High computation	Needs large data
Hybrid Models	Balanced features	Increased complexi-	Limited real-time

		ty	validation
Explaina-ble AI	Interpretability	No stand-ard metrics	Needs clinical validation

such as VGG, ResNet, DenseNet, and Inception have been widely adopted using transfer learning strategies to address limited dataset sizes [10]. Transfer learning enables models trained on large-scale natural image datasets to be fine-tuned for medical image analysis, accelerating convergence and improving performance.

Residual networks and densely connected networks have demonstrated strong capability in distinguishing visually sim-ilar abnormalities, including polyps, erosions, and inflamma-tory lesions [11]. EfficientNet architectures further improved performance by balancing network depth, width, and resolu-tion scaling, offering competitive accuracy with fewer param-eters [12]. These models have shown promise for real-time clinical deployment due to improved computational efficiency. Despite their success, CNN-based models exhibit certain limitations. Their receptive fields primarily focus on local features, which may restrict the ability to capture global contextual relationships within complex endoscopic imagery. Additionally, performance can degrade when applied to cross-domain datasets due to variations in acquisition conditions and annotation protocols [13].

To overcome these challenges, researchers have explored advanced architectures incorporating attention mechanisms and transformer-based components, which are discussed in the following section. The literature analysis reveals a pro-

gressive transition from handcrafted feature-based machine learning approaches toward deep convolutional networks, and more recently toward hybrid transformer-driven architectures enhanced with explainability modules. While classification accuracy has improved significantly over the years, sev-eral critical challenges persist, including dataset imbalance, cross-domain generalization, computational efficiency, and clinically validated interpretability. Most existing studies fo-cus primarily on frame-level accuracy without comprehen-sive cross-institutional evaluation. Therefore, future research must emphasize robust validation, multimodal integration, and lightweight deployable models to ensure reliable real-world clinical integration.

VII. Transformer-Based and Hybrid Models For Gi Diagnosis

Transformer architectures, originally introduced in natural language processing, have recently gained significant atten-tion in computer vision through the development of Vision Transformers (ViTs). Unlike Convolutional Neural Networks (CNNs), which rely on localized receptive fields, transform-ers utilize self-attention mechanisms to model long-range dependencies and global contextual relationships across im-age patches [14]. This capability is particularly beneficial in gastrointestinal (GI) image analysis, where subtle spatial variations and distributed lesion patterns may be clinically relevant.

In Wireless Capsule Endoscopy (WCE) applications, transformer-based models have demonstrated improved ro-bustness in distinguishing visually similar disease categories. By dividing an image into patches and processing them as a sequence, Vision Transformers can capture structural and contextual interactions that may be overlooked by purely convolutional architectures [15]. Recent studies have reported enhanced performance in multi-class GI disease classification using transformer-based frameworks, particularly in complex lesion detection scenarios.

To address limitations associated with limited medical datasets, hybrid CNN–Transformer architectures have been proposed. These models combine CNN layers for efficient local feature extraction with transformer modules for global feature refinement [13]. Such hybrid frameworks aim to balance computational efficiency and representational power, achieving improved classification accuracy while maintaining scalability.

Despite these advancements, transformer-based models typ-ically require substantial computational resources and large-scale training data. In medical imaging contexts where anno-tated data is limited, careful regularization, transfer learning, and lightweight attention mechanisms are necessary to ensure stable performance [12].

The integration of attention mechanisms has also con-tributed to improved interpretability, as attention maps can highlight diagnostically important regions within GI images. However, achieving clinically reliable and explainable decision-making remains an active area of research.

The following section discusses explainable artificial intel-ligence techniques and their importance in enhancing trans-parency and trust in automated GI disease diagnosis systems.

VII. Explainable Artificial Intelligence In Gi Disease Diagnosis

While deep learning models have demonstrated remarkable performance in gastrointestinal (GI) disease classification, their black-box nature raises concerns regarding transparency, reliability, and clinical acceptance. In medical applications, interpretability is essential to ensure that automated decisions align with clinically meaningful features. Consequently, ex-plainable artificial intelligence (XAI) techniques have gained increasing importance in Wireless Capsule Endoscopy (WCE) research.

One of the most widely adopted interpretability techniques is Gradient-weighted Class Activation Mapping (Grad-CAM), which generates heatmaps highlighting image regions that contribute most to model predictions [13]. In GI disease classification, Grad-CAM visualizations help verify whether the model focuses on pathological structures such as inflamed mucosa, ulcer margins, vascular irregularities, or polyp pro-trusions. Such visual explanations enhance clinician trust and support decision validation.

Beyond Grad-CAM, attention-based visualization methods and saliency maps have been integrated into transformer and hybrid architectures to improve interpretability

[14]. These ap-proaches provide insight into feature importance across spatial regions and image patches, offering better understanding of model behavior.

Despite these advancements, several challenges remain. Explanations may vary across models, and heatmaps do not always guarantee causal reasoning. Additionally, standardized evaluation metrics for interpretability are still lacking in med-ical imaging research [15]. Future work must focus on devel-oping robust and clinically validated explanation frameworks that combine quantitative assessment with expert evaluation. Improving interpretability is essential not only for regula-tory approval but also for the safe integration of AI-driven diagnostic systems into real-world clinical workflows.

IX. Open Challenges and Research Gaps

Despite significant progress in automated gastrointestinal (GI) disease classification using deep learning, several technical and clinical challenges remain unresolved. Addressing these limitations is essential for developing robust and clini-cally deployable diagnostic systems.

- **Class Imbalance and Limited Annotations**
  Most publicly available Wireless Capsule Endoscopy (WCE) datasets exhibit severe class imbalance, where normal frames significantly outnumber pathological findings. Rare abnormalities such as early-stage lesions or subtle inflamma-tory patterns often contain limited samples, leading to biased model learning and reduced sensitivity. Furthermore, manual annotation of WCE images requires expert gastroenterolo-gists, making large-scale labeled dataset creation both time-consuming and expensive.

- **Cross-Dataset Generalization**
  Models trained on a single dataset frequently experience performance degradation when evaluated on data collected from different medical centers or imaging devices. Variations in resolution, illumination, acquisition protocols, and patient demographics introduce domain shifts that affect model gen-eralization. Cross-dataset validation and domain adaptation remain underexplored areas in GI image analysis.

- **Temporal Information Utilization**
  Most current studies focus on frame-level image classifica-tion, treating WCE frames independently. However, WCE is inherently a video-based modality, where temporal continuity may provide valuable contextual cues for disease detection. In-corporating sequential modeling techniques to leverage inter-frame relationships remains an open research direction.

- **Computational Efficiency and Real-Time Deployment**
  Deep learning architectures, particularly transformer-based models, require substantial computational resources. Deploy-ing such models in real-time clinical settings or edge-based healthcare systems demands lightweight architectures, model compression, and efficient inference strategies.

- **Explainability and Clinical Validation**
  Although explainable AI techniques such as Grad-CAM provide visual insights, standardized frameworks for eval-uating interpretability are still lacking. Clinical validation through prospective studies and expert-in-the-loop evaluation is necessary to ensure reliability and regulatory acceptance.

  Addressing these challenges will be critical for transitioning AI-based GI disease classification systems from experimental research to routine clinical practice.

X. Future Research Directions

The rapid advancement of artificial intelligence in gas-trointestinal (GI) image analysis presents several promising research directions aimed at improving clinical applicability and diagnostic performance. Future research should move beyond frame-level classification toward more comprehensive and clinically integrated systems.
A key direction involves incorporating temporal modeling into Wireless Capsule Endoscopy (WCE) analysis. Since WCE generates continuous video streams, integrating sequence-based architectures such as recurrent neural networks, tem-poral convolutional networks, or transformer-based temporal attention mechanisms can enhance disease localization and contextual understanding across consecutive frames.

Multimodal learning is another important area, where image data can be combined with clinical metadata such as patient history, laboratory reports, and demographic information. This integration has the potential to improve diagnostic accuracy and enable personalized treatment strategies. In addition, self-supervised and semi-supervised learning approaches can re-duce dependency on large annotated datasets, which are often difficult and expensive to obtain.

Domain adaptation and cross-institutional validation are essential for improving model generalization across differ-ent clinical environments. Developing standardized evaluation frameworks and multi-center datasets will further enhance robustness and reproducibility.

Another promising direction is automated video summa-rization, which focuses on extracting diagnostically relevant frames while reducing redundant information, thereby mini-mizing clinician workload.

Finally, lightweight model architectures and optimization techniques such as pruning, quantization, and knowledge distillation are crucial for enabling real-time deployment in clinical and edge-based systems. Enhancing interpretability through clinically validated explainable AI frameworks will also play a vital role in increasing physician trust and facili-tating regulatory approval of automated GI diagnostic systems.

XI. Conclusion

Wireless Capsule Endoscopy (WCE) has significantly im-proved gastrointestinal (GI) diagnostics by enabling non-invasive and comprehensive imaging. However, the large vol-ume of generated data necessitates efficient automated analysis techniques. This

review presented a structured overview of artificial intelligence-based approaches for GI disease classifi-cation, including traditional machine learning methods, convolutional neural networks, transformer-based architectures, and hybrid models.

Although recent advancements have achieved high clas-sification performance, several challenges remain, including data imbalance, limited annotations, domain variability, com-putational constraints, and lack of standardized interpretability methods. Future research should focus on temporal modeling, domain adaptation, lightweight model design, and clinically validated explainability frameworks to ensure reliable real-world deployment.

.

References

1. Khan, S. Hussain, and A. Rehman, "Sparse Convolutional Network and Attention-Based Deep Hybrid Model for GI Tract Disease Classifi-cation," IEEE Access, vol. 13, pp. 1122–1136, 2025.
2. S. Tanwar and M. Sharma, "EfficientViT: A Hybrid Vision Transformer Model for GI Tract Disease Detection," Biomedical Signal Processing and Control, vol. 87, p. 104406, 2025.
3. N. Dahan, A. Elharrouss, F. Almaadeed, and A. A. Mohammed, "Ex-plainable AI-Based Transformer Framework for GI Lesion Classifica-tion," Expert Systems with Applications, vol. 228, 2025.
4. A. Malik, P. Ramesh, and A. Ghosh, "Deep Learning-Based Multiclass GI Tract Disease Classification with VGG-19-CNN Ensemble," Com-puters in Biology and Medicine, vol. 172, 2024.
5. A. El-Ghany, M. Tarek, and H. Abdelaziz, "Intelligent Learning Rate Control with Efficient CNN Architectures for GI Disease Detection," IEEE Access, vol. 12, 2024.
6. H. Nouman, S. Qureshi, and M. Shahid, "Contrast Enhancement and Deep Learning for WCE Disease Classification," IEEE Transactions on Instrumentation and Measurement, vol. 72, 2023.
7. J. He, Y. Li, and X. Wang, "Dynamic Two-Branch Supervised Networks for GI Lesion Detection in Wireless Capsule Endoscopy," Medical Image Analysis, vol. 90, 2024.
8. F. Kumar, R. Singh, and D. Patel, "Pediatric Wireless Capsule En-doscopy Lesion Classification Using DenseNet and ResNet," Journal of Pediatric Gastroenterology, vol. 78, 2023.
9. Y. Liu, K. Chen, and L. Zhang, "Ulcer Detection from Endoscopy Using Deep Learning in Clinical Workflow," Nature Communications, vol. 14, 2023.
10. S. Rajput, V. Desai, and R. Kulkarni, "CNN–Vision Transformer Fusion Architecture for GI Tract Disease Detection," Scientific Reports, vol. 15, 2025.
11. A. Rehman, S. Iqbal, and M. Ahmad, "CapsuleNet-Based Efficient GI Tract Disease Classifier in Imbalanced Dataset," arXiv:2410.01537, 2024.
12. Y. Zhou and T. Zhang, "CASCRNet: Compact ASPP-Enhanced CNN for WCE Classification," arXiv:2410.01421, 2024.
13. P. Sharma, A. Verma, and N. Joshi, "Hybrid Transformer-CNN with Explainability for GI Disease Detection," IEEE Journal of Biomedical and Health Informatics, vol. 28, 2024.
14. J. R. Lewis, M. Thompson, and E. Clarke, "Artificial Intelligence in Endoscopic Gastrointestinal Diagnosis: A Systematic Review," IEEE Access, vol. 12, 2024.

15. D. Diamantis, K. Papadopoulos, and G. Karagiannis, “TIDE: Multiscale Residual Variational Autoencoder for Wireless Capsule Endoscopy Im-age Generation,” IEEE Access, vol. 11, pp. 10850–10860, 2024