Taylor & Francis
Taylor & Francis Group

Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: www.tandfonline.com/journals/uaai20

AI-Driven DenseResConcatenation Network for Early and Accurate Detection of Gastrointestinal Abnormalities

Maheswari Subburaj , Chinthamani Mohan Krishna , Arun Kumar Sivaraman , Sasi Kumar Periyasamy , Janakiraman Nithiyanantham & Kamalavelu Velayutham

To cite this article: Maheswari Subburaj , Chinthamani Mohan Krishna , Arun Kumar Sivaraman , Sasi Kumar Periyasamy , Janakiraman Nithiyanantham & Kamalavelu Velayutham (2026) AI-Driven DenseResConcatenation Network for Early and Accurate Detection of Gastrointestinal Abnormalities, Applied Artificial Intelligence, 40:1, 2712691, DOI: 10.1080/08839514.2026.2712691

To link to this article: https://doi.org/10.1080/08839514.2026.2712691

APPLIED ARTIFICIAL INTELLIGENCE
2026, VOL. 40, NO. 1, 2712691
https://doi.org/10.1080/08839514.2026.2712691

AI-Driven DenseResConcatenation Network for Early and Accurate Detection of Gastrointestinal Abnormalities

Maheswari Subburaj[a], Chinthamani Mohan Krishna[a], Arun Kumar Sivaraman[b,c], Sasi Kumar Periyasamy[d], Janakiraman Nithiyanantham[e], and Kamalavelu Velayutham[f]

[a]Vellore Institute of Technology, Chennai, India; [b]Algoma University, Brampton, Canada; [c]Gulf College, Al Mabaila, Sultanate of Oman; [d]Vellore Institute of Technology, Vellore, India; [e]K.L.N. College of Engineering, Madurai, India; [f]University of Central Lancashire, Preston, UK

ABSTRACT

The gastrointestinal (GI) tract, encompassing the stomach, liver, pancreas, and other digestive organs, is essential for digestion. Due to their prevalence and impact on quality of life, gastrointestinal diseases pose significant global public health challenges. The Global Cancer Observatory (GLOBOCAN), a project of the International Agency for Research on Cancer (IARC), part of the World Health Organization (WHO), reports over 4.2 million new GI cancer cases and 2.2 million deaths in 2020. Early detection and accurate diagnosis are vital for improving outcomes, supervised learning and pre-trained models like Xception, Alexnet, and VGG16, have demonstrated effectiveness; however, a hybrid model that integrates the capabilities of multiple approaches shows superior performance. This research introduces DenseResConcatenation, a hybrid deep-learning model that combines ResNet50 and DenseNet121 to enhance diagnostic precision in GI disease prediction using endoscopic images. By integrating traditional ML techniques with modern deep learning, the model supports personalized treatment planning and robust classification. Our proposed model achieved a top-1 accuracy of 94.15% and an F1 score of 94.12% on an unseen dataset. The proposed framework significantly improves detection accuracy by leveraging visual patterns in GI-tract imagery, opening the door to more effective and accurate solutions in endoscopic diagnostics.

1. Introduction

The detection and classification of diseases is an important but often overlooked area of research, with the identification and characterization of gastrointestinal (GI) tract issues being critical in modern medical care because these conditions affect millions of individuals worldwide and significantly burden healthcare systems (Sung et al. 2021). The GI tract, which starts in the mouth and finishes at the rectum, supports the transformation of food into nutrients that the body can absorb for energy and other vital processes (Steinway et al.2020. Utilizing advancements in artificial intelligence (AI) and machine learning (ML), especially in medical imaging, holds great promise for improving the efficiency and accuracy of GI disease detection (LeCun et al. 2015; Litjens et al. 2017). As new AI and ML methodologies develop, medical image diagnosis becomes faster and helps physicians analyze the type of disease accurately, facilitating timely decision making and reducing delays in surgery (Esteva et al. 2021). However, the availability of large, labeled datasets for medical image diagnosis remains a significant challenge, impeding the development of robust algorithms (Pogorelov et al. 2017). The global burden of gastrointestinal disorders impacts not only individual health but also imposes heavy financial strains on healthcare systems (Sung et al. 2021). Efficient identification and precise diagnosis are essential for effective treatment and better patient outcomes (Zhou et al. 2021). Therefore, a comprehensive understanding of the structure and interconnected functions of the GI system is critical for both appreciating physiological complexity and managing GI health effectively. Deep learning models have demonstrated excellent diagnostic ability for both classification and segmentation of medical images (Litjens et al. 2017). To drive improvements in diagnostic efficiency and accuracy,

CONTACT Maheswari Subburaj maheswari.s@vit.ac.in School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, India

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08839514.2026.2712691

especially in the early detection of GI tract diseases, ongoing research is crucial (Talo et al. 2019). Endoscopic examination remains the gold standard for GI tract evaluation, providing real-time visualization of internal structures and pathological conditions (Rex et al. 2017). A gastroscopy examines the upper GI tract – including the stomach, esophagus, and the first section of the small intestine – while a colonoscopy evaluates the colon and rectum. Both approaches employ digital high-definition endoscopes for real-time evaluation but require expensive instruments and skilled clinicians (Rex et al. 2017). The GI tract consists of multiple interconnected organs, all vital for digestion, absorption, and processing of food. To improve diagnostic accuracy, several preprocessing techniques must be investigated and implemented to optimize image analysis. Image preprocessing (such as enhancement, normalization, and denoising) contributes greatly to diagnostic model performance (Gonzalez and Woods 2018). Techniques like Gaussian blur can enhance image quality by reducing high-frequency noise, which is crucial in GI imaging for reliable diagnosis (Buades et al. 2005). Additionally, data augmentation plays a critical role in improving model robustness and reducing overfitting, particularly when the available dataset is limited or imbalanced (Ramamurthy et al. 2022). Nonetheless, challenges remain in ensuring that automated diagnosis yields reliable, high-probability predictions to support clinical decision-making. Numerous related studies have developed and tested different models on various datasets to improve disease detection accuracy and class prediction (Sharma, Kumar, and Garg 2023). For example, Zhao et al. (2022) provided a systematic review summarizing the application of CNNs in gastric cancer identification and detection, utilizing multi-center, multi-institutional datasets to enhance model diversity and generalizability (Zhao et al. 2022). The main technical contributions in this line of research include:

1. Gaussian blur preprocessing followed by noise filtering: Smoothing images with Gaussian and median filters improves model training and prediction (Ramamurthy et al. 2022). 2. Data augmentation: Augmentation improves generalization, model robustness, and helps mitigate overfitting, especially with limited or imbalanced data (Talo et al. 2019). 3. Model building: Training various pre-trained models and developing hybrid architectures through concatenation of top-performing networks has yielded improved diagnostic performance (Sivari et al. 2023). 4. Disease detection and evaluation: Employing comprehensive metrics and validation techniques, including cross-dataset analysis, to assess real-world model performance (Gunasekaran et al. 2023; Pogorelov et al. 2017).

- **Novel Hybrid Architecture**: Applying Gaussian blur, followed by noise filtering using median blur, helps to smooth the image. Development of DenseResConcatenation, a hybrid CNN architecture that effectively combines ResNet50 and DenseNet121 through feature concatenation, leverages the residual learning capabilities of ResNet50 and the dense connectivity benefits of DenseNet121.
- **Superior Performance Achievement**: Demonstration of 94.15% accuracy on the Kvasir V1 dataset and 93.62% on the Kvasir V2 dataset, representing significant improvements over individual backbone models (ResNet50: 90.56%, DenseNet121: 90.13%).
- **Comprehensive Experimental Validation**: Implementation of a rigorous experimental methodology, including advanced data augmentation techniques (rotation, flipping, scaling, brightness adjustment) and systematic evaluation across multiple performance metrics.
- **Detailed Comparative Analysis**: Extensive comparison with five state-of-the-art pre-trained models (ResNet50, DenseNet121, EfficientNetB0, InceptionV3, Xception) demonstrates the superiority of the hybrid approach.
- **Clinical Applicability Assessment**: Practical demonstration of disease-prediction capabilities on real endoscopic images with confidence scores, providing a pathway toward clinical implementation.

The remainder of this paper is organized as follows: Related work reviews deep learning applications for GI disease diagnosis, covering CNN architectures, ensemble methods, and current challenges. Proposed methodology presents the DenseRes Concatenation methodology, including the overall workflow and dataset description. Data preprocessing details data preprocessing techniques, including Gaussian blur with noise filtering and comprehensive data augmentation strategies. Model building describes the individual model building process for ResNet50, DenseNet121, and other architectures. Building the DenseResConcatenation hybrid model from the top two evaluated models explains the hybrid model

construction through feature concatenation and training procedures. Results and discussions present experimental results, performance metrics, confusion matrices, and disease prediction capabilities. Finally, Conclusion and future works conclude with a summary of findings, limitations, and comprehensive future work directions.

2. Related work

2.1. Convolutional neural network (CNN) architectures

To address the problems in this study, researchers have examined a number of related works. One such work is a CNN-based system that combines six different types of extracted features and uses feature fusion with various classifiers. This approach likely enhances model accuracy and performance in medical image analysis, as explored by Pogorelov et al. (2017). Researchers have also investigated how well different CNN architectures work for analyzing gastrointestinal (GI) images. They have utilized pre-trained deep-learning (DL) models such as VGG16, ResNet, InceptionV3, and MobileNet. These models have achieved high accuracy in tasks such as polyp detection, esophagitis classification, and landmark detection (Agrawal et al. 2017; Escobar et al. 2021; Sharma, Kumar, and Garg 2023; Yoshiok et al. 2023). Furthermore, ensemble models, which combine multiple pre-trained models, have demonstrated promising results. Gunasekaran et al. (2023), Sivari et al. (2023), Khan et al. (2022) show that ensemble models can outperform individual models, achieving accuracy exceeding 95% for GI disease classification. References (such as Jha et al. 2021; Quindós et al. 2023; Zhao et al. 2022; Zhuang, Zhang, and Liao 2023) provided an inclusive overview of the current state of the art in deep-learning applications for medical image analysis of the digestive system. The study reviews and conducts a meta-analysis on the application of convolutional neural networks for determining the depth of invasion in gastrointestinal cancer (Wu et al. 2024). CNNs improve disease categorization and polyp localization, according to Nogueira-Rodríguez et al. (2022). Specialized approaches, such as attention-guided CNNs and self-supervised learning performed by Quindós et al. (2023), further improve diagnostic accuracy. Current deep-learning approaches for GI disease classification face several critical limitations: (1) Reliance on single CNN architectures that may not capture the diverse feature representations needed for complex medical images (Sharma, Kumar, and Garg 2023); (2) Insufficient handling of class imbalance in medical datasets, leading to poor performance on rare diseases (Gunasekaran et al. 2023); (3) Limited generalization across different clinical settings and imaging equipment (Sivari et al. 2023); (4) Lack of comprehensive comparative studies on standardized datasets (Khan et al. 2022); and (5) Inadequate attention to computational efficiency for clinical deployment (Owais et al. 2020). Our DenseResConcatenation approach addresses these limitations by: (1) Combining the complementary feature-extraction capabilities of ResNet50 and DenseNet121 to capture diverse image characteristics; (2) Implementing robust data-augmentation strategies with plans for advanced synthetic data generation; (3) Providing extensive comparative analysis on benchmark Kvasir datasets; (4) Achieving superior performance (94.15% accuracy) compared to individual backbone architectures; and (5) Establishing a foundation for future optimization toward clinical deployment.

2.2. Applications of DL in GI image analysis

Medical image processing is the foundation for deep learning. CNNs are perfect for tasks like lesion detection and image classification because they are great at extracting characteristics from pictures, including shapes, textures, and patterns that are explored in the papers by Dheir and Abu-Naser (2022) and Sivari et al. (2023). By utilizing various CNN models, including transfer learning with VGG16, InceptionV3, and ResNet50, this study focuses on disease detection within the GI tract. The suggested ResNet50 model does a better job of finding esophagitis, ulcerative colitis, and gastrointestinal polyps (Rezvy et al. 2020). DL excels at identifying abnormalities in GI images. Papers (such as Dheir and Abu-Naser 2022; Sharma, Kumar, and Garg 2023) demonstrate its effectiveness in detecting polyps, ulcers, and cancerous lesions in endoscopy and capsule endoscopy images. This can significantly support doctors or endoscopists during procedures, leading to earlier intervention and improved patient outcomes. Papers (such as Aruna, Puviarasan, and Palaniappan 2007; Attallah and Sharkas 2021; Haile et al. 2022; Vania et al.

2023; Walsh et al. 2014) describe the application of deep learning and machine learning techniques for the detection, classification, and diagnosis of gastrointestinal diseases using endoscopic images, exploring various models and hybrid approaches to enhance accuracy and efficiency in medical image analysis. Furthermore, Bolourani et al. (2021) and Ozawa et al. (2019) demonstrate machine learning applications for diagnosing and predicting outcomes in digestive disorders, highlighting the usefulness of deep learning in medical image processing.

2.3. Recurrent neural networks

RNNs are especially helpful in comprehending sequential data, such as videos captured during endoscopic procedures. RNNs can detect minute alterations or anomalies over time because they record temporal information and analyze the links between frames, as explored in the paper Owais et al. (2020). The paper Lonseko et al. (2021) describes the use of attention-guided convolutional neural networks for gastrointestinal disease classification in endoscopic images. Although the paper primarily focuses on convolutional networks, it frequently employs attention mechanisms in conjunction with RNNs for sequence data.

2.4. Disease classification

DL models may utilize images to categorize various GI diseases. Diseases such as esophagitis, ulcerative colitis, and Crohn's disease can be successfully categorized when paired with other medical data, allowing clinicians to identify patients, as investigated in Ramamurthy et al. (2022). A comprehensive computer-aided diagnosis (CAD) tool is developed for classifying various gastrointestinal (GI) diseases using endoscopy videos explored in the paper Owais et al. (2020). Deep learning-based classification, spatiotemporal feature extraction, and a PCA-based KNN classifier are used in this study to achieve great results with high accuracy and an F1 score. Papers (such as Navale et al. 2024; Naz et al. 2023; Sang et al. 2021; Sekuboyina, Devarakonda, and Seelamantula 2017; Su et al. 2022) focus on applying deep learning and convolutional neural networks (CNNs) for the detection and classification of gastrointestinal diseases using endoscopic images.

2.5. Computer-Aided Diagnosis (CAD)

Endoscopists may get real-time feedback during surgeries via DL-powered CAD systems. Frameworks that incorporate deep learning for both feature extraction and classification are discussed in papers (Nagao et al. 2022), potentially leading to better diagnostic accuracy. Some papers, such as Bharadwaj et al. (2018), Chen et al. (2018), Guimarães et al. (2020), Ueyama et al. (2021), examine how AI and deep learning can be used in gastrointestinal endoscopy to find, classify, and diagnose diseases like ulcerative colitis, inflammatory bowel disease, colorectal polyps, stomach problems, and early gastric cancer.

2.6. Data augmentation

Medical image analysis encounters the challenge of limited training data. Data augmentation methods, such as random flipping, rotation, image augmentation, and instance cropping, are employed to artificially generate new data from preexisting datasets. The model is completed in two stages, Phase I and Phase II. In Phase I, the study's average precision was 51.14% for "BE," 50% for "HGD," and 0.62% for "polyp." Lower scores were obtained for'suspicious' and "cancer," with phase II semantic-segmentation scoring increased to 0.62 and detection scoring to 0.29 following the data augmentation covered in Rezvy et al. (2020).

An automated, deep-learning-based method is utilized to classify gastrointestinal disorders from endoscopic images. It makes use of the HyperKvasir dataset, which covers 23 classes of gastrointestinal disorders, and employs data augmentation to address class imbalance. Some preprocessing methods include adding more data, which can involve random geometric changes such as flipping the data horizontally, shifting its width and height, and rotating it. This is done to increase the number of samples and address the class imbalance in the dataset. The study uses cutting-edge deep-learning algorithms, such as EfficientNet-B0 and

a custom Effimix network. The Effimix network has squeeze and excitation layers, background erasing, and custom activation functions that were examined in Ramamurthy et al. (2022).

2.7. Explainable AI (XAI)

Integration of XAI approaches into DL models may improve clinical decision-making transparency and reliability. The "black box" nature of some DL models can hinder trust in clinical settings. Explainable AI (XAI) techniques provide insights into model decision-making, fostering transparency and trust. In Mukhtorov et al. (2023), it is discussed how Grad-CAM can be used to show the areas of interest that a ResNet-152 model uses for wireless capsule image categorization. The model's performance was enhanced, achieving 98.28% training accuracy and 93.46% validation accuracy. The research aimed to improve endoscopic image classification accuracy by utilizing AI techniques and efficient medical data augmentation methods. The study used the deep CNN model ResNet-152 to classify endoscopic images. Furthermore, interpretable insights into the CNN model's decision-making process were obtained using explainable AI techniques such as Grad-CAM, Hires-CAM, Layer-CAM, Grad-CAM++, Xgrad-CAM, and Score-CAM.

2.8. Image segmentation

Deep learning (DL) can segment specific regions of interest (ROIs) in GI images. This review describes the use of transfer learning and deep learning models, such as VGGNet and Inception-V3 networks, to locate landmarks in medical pictures, particularly in the gastrointestinal system. Leveraging pre-trained models with transfer learning overcomes the barrier of insufficient training data, resulting in significant performance metrics. Segmentation is used for applications such as polyp delineation and anatomical landmark detection, as explored in Agrawal et al. (2017). This enables more precise analysis and quantification of lesions, which aids in therapy planning.

2.9. Model performance

High end-to-end computer performance and many epochs of training are required to evaluate the model performance and diagnose the endoscopic image. For example, as explored in the study Yogapriya et al. (2021), they used 32 GB of GPU memory and trained the model for more than 400 epochs, achieving an accuracy of 96%.

2.10. Model generalizability

Model generalizability is a machine learning algorithm's ability to perform well on previously unseen data from the same distribution as its training data. It determines how well a model can produce accurate predictions or classifications on fresh, previously unknown data. Transfer learning is an important approach for adapting models previously trained on common images, such as the ImageNet dataset, to classify clinical images in the gastrointestinal domain accurately. Using DenseNet201, ResNet50, Xception, VGG19, and VGG16, the authors say that their proposed method achieves this level of accuracy while using a tiny portion of the trainable parameters compared to more advanced techniques. This shows how effective and generalizable their approach is, as discussed in Escobar et al. (2021). Also, works like Cheung (2017), Gnanambal and Rayappan (2011), Konstantin et al. (2017) show advanced techniques in deep learning and image processing. They focus on making medical image analysis more general, accurate, and of high quality by using new methods such as DenResNet ensembles, deep learning comparisons, self-supervised anomaly detection, and hybrid filtering methods.

2.11. Augmented reality (AR) in gastrointestinal endoscopy

AR/VR technologies can potentially improve clinical guidance, diagnostic accuracy, and advanced treatment modalities in gastrointestinal endoscopy and inflammatory bowel disease management. The potential

use of augmented reality (AR) in gastrointestinal endoscopy centers on the integration of pathologic correlates and high-definition video data. A wide variety of endoscopic procedures could be computer-guided with AR technology. This is especially true for finding polyps, as examined in Mahmud et al. (2015).

2.12. Role of endoscopy in treating and diagnosing GI diseases

Endoscopy plays a crucial role in the detection and treatment of gastrointestinal (GI) illnesses, particularly ulcerative colitis (UC). The study compares invasive (endoscopy-dependent) and noninvasive disease-activity measures for UC in 66 consecutive individuals. It assesses the relationships and contributions of endoscopic and noninvasive indicators in assessing disease activity, as explored in Higgins et al. (2005). Reference Bharadwaj et al. (2018), which explores Crohn's disease (CD) and ulcerative colitis (UC), looks at the growing use of endoscopy in detecting and treating inflammatory bowel disease (IBD). The research underlines the need for precise diagnosis, especially distinguishing between CD and UC, because their treatment methods differ greatly, requiring both medication and surgery. It also emphasizes developments in endoscopic methods and their positive influence on controlling IBD problems, ultimately leading to improved patient outcomes.

3. Proposed methodology

This section contains the overall research workflow: dataset preparation, preprocessing methods used, and the overall work's model architecture.

3.1. Overall workflow

The overall workflow, illustrated in Figure 1, provides a comprehensive visual representation of the entire process, starting from the initial data-collection phase to the final model evaluation. Each crucial phase of the research is meticulously detailed, highlighting the collection, processing, and use of medical image accumulation. The workflow begins with the collection of a broad group of medical images representing distinct ailments that require diagnosis. Preprocessing techniques and picture-enhancement technologies follow this to ensure the dataset is ready for the next steps. The

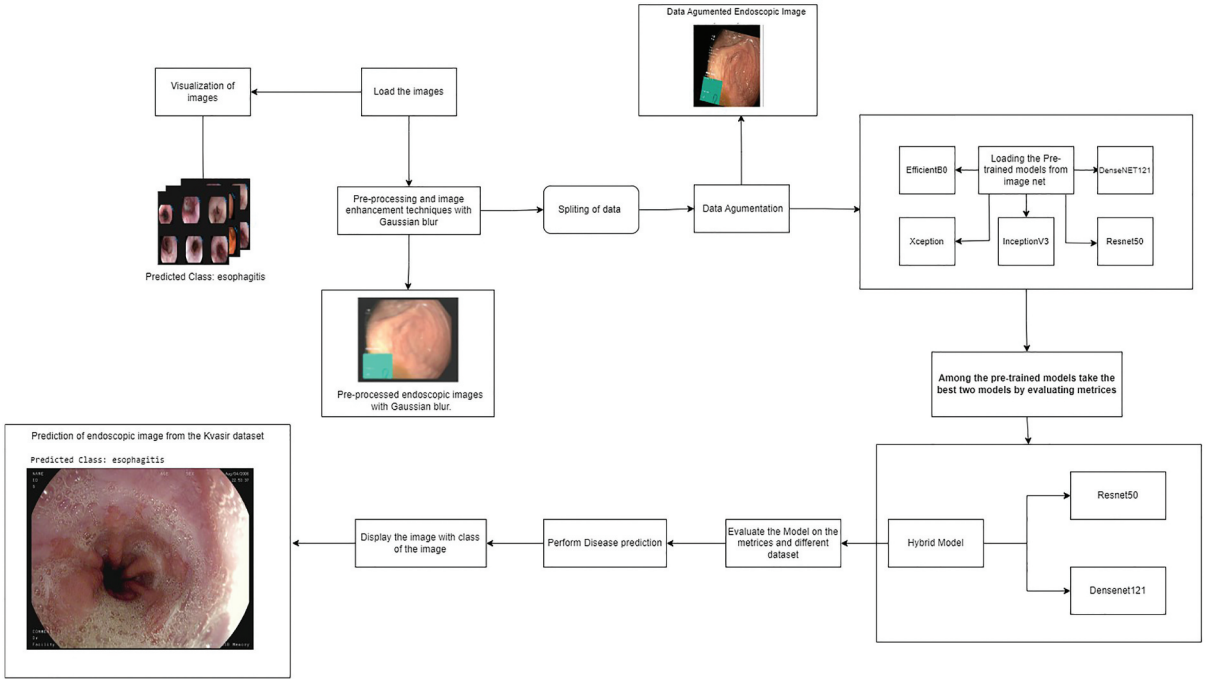

Figure 1. Proposed architecture.

graphic then illustrates the integration of various pre-trained models, including ResNet50, DenseNet121, and Xception, to efficiently train the dataset. After training, these models are thoroughly tested using important performance indicators such as accuracy and F1 score. The top two performing models are then chosen and blended to further improve diagnostic accuracy. The next phase of the approach, represented in Figure 1, is to test this concatenated model on a fresh dataset to ensure its accuracy in identifying medical images.

3.2. Dataset description

This study utilizes the Kvasir datasets V1 and V2 (Pogorelov et al. 2017), which contain labeled endoscopic images representative of various GI diseases. The distribution of images across the different gastrointestinal disease categories in the Kvasir V1 and V2 datasets is illustrated in Figure 2. The Kvasir dataset V1 includes 4,000 images across 8 classes, while Kvasir V2 expands to 8,000 images across multiple disease categories. These images were captured using endoscopic equipment at Vestre Viken Health Trust (VV) in Norway and annotated by medical professionals from VV and the Cancer Registry of Norway (CRN) (Pogorelov et al. 2017). The datasets have been widely adopted as benchmark datasets for GI disease classification research due to their clinical relevance and expert annotation quality.

The CRN, an independent association affiliated with the Oslo University Hospital Trust and the South-Eastern Norway Regional Health Authority, conducts cancer research that provides useful insights. The pictures are split into three groups: anatomical landmarks (Cecum, Pylorus, Z-line), medically significant abnormalities (ulcerative colitis, polyps, esophagitis), and two endoscopic polyp removals. These datasets, annotated by clinical professionals and cancer research groups, help to advance cancer detection, prevention, and care by studying endoscopic pictures. The cecum, pylorus, and Z-line are significant anatomical markers in the GI tract. The Z-line marks the stomach-esophageal junction and assists in diagnosing gastrointestinal disease. The pylorus, positioned at the stomach-duodenum junction, helps to detect problems such as loss, stenosis, and ulceration. The large intestine, located in the right iliac fossa of the abdomen, can be palpated if it becomes enlarged due to inflammation or malignancy. Pathological findings include ulcerative colitis, polyps, and esophagitis. Ulcerative colitis causes chronic inflammation and ulcers in the digestive system. Some polyps on the colon or rectal lining can become cancerous. Acid reflux, infections, drugs, or allergies commonly cause esophagitis, an inflammation of the esophagus. The proposed model architecture is trained using the Kvasir dataset V2, which contains 8000 images ranging in resolution from 720 × 576 to 1920 × 1072 pixels. A bespoke model is built using a variety of pre-trained models to improve detection and classification accuracy, as well as metrics. The completed simulation is tested on the Kvasir dataset V1 to ensure its effectiveness on an unknown dataset.

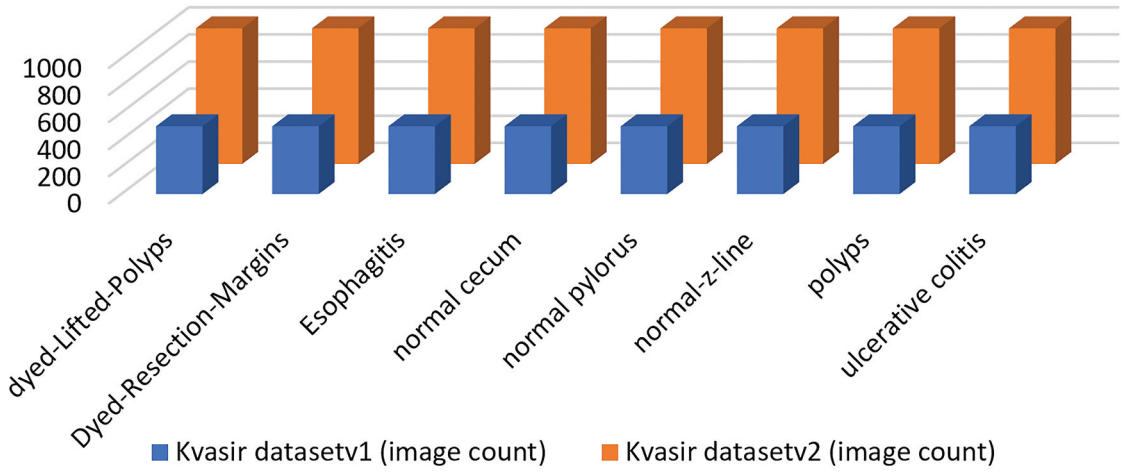

Figure 2. Dataset statistics.

4. Data preprocessing

During data preprocessing, several stages and image-enhancing approaches, including Gaussian blur with noise filtering, CLAHE, and a combination of Gaussian blur with adaptive histogram equalization, were tested. However, Gaussian blur and noise filtering yielded the best results, optimizing feature extraction and enhancing the interpretability of the GI tract images.

4.1. Workflow of the preprocessing techniques

It was found that a combination of Gaussian blur and noise filtering is the optimal preprocessing approach. This blend efficiently smooths the pictures, lowers noise, and boosts overall data quality for model training. Combining Gaussian blur and noise filtering reduces high-frequency noise and improves image quality, resulting in a cleaner dataset. This preprocessing enables the model to learn more robust features from the images, improving performance while minimizing the danger of overfitting.

The approach begins by iterating over the various categories of images found within a directory. Initially, two empty lists named labels and images are created. These lines specify the lists in which the pre-processed images and their labels will be stored. Initially, the images are reduced to a consistent size of 224 × 224 pixels via bilinear interpolation, ensuring uniform dimensions across all photographs. The original image is loaded and transformed into a numerical array format. Equation (1) provides the parameters for modifying the brightness and contrast of the photos.

$$I_{adjusted} = \alpha \cdot I + \beta \tag{1}$$

Where α represents the contrast control parameter, β represents the brightness control parameter, and "adjusted" represents the adjusted image. The formula sets the parameters for the contrast and brightness adjustments. This change intends to normalize the enlightenment and improve the visual highlights of the images. Next, a middle channel, known as a median filter, is applied to reduce noise and enhance image quality, using cv2.medianBlur() to smooth out inconsistencies. Once more, the images are scaled to the optimal dimensions to maintain consistency. This is generated using Equation (2), where Ifiltered is the filtered image and k denotes the kernel size for the median filter. And then the filtered image is resized to a consistent size of (224 × 224) pixels using bilinear interpolation. These procedures are carried out to guarantee uniform dimensions in all images, promote consistent model training, and Equation (3) shows how that formula is generated.

$$I_{filtered} = MedianBlur\left(I_{adjusted}, k\right) \tag{2}$$

$$I_{resized} = Resize\left(I_{filtered}, (224, 224)\right) \tag{3}$$

To further improve image quality, cv2. GaussianBlur is implemented to reduce high-recurrence noise and emphasize important image features, as illustrated in Figure 3. This step supports creating a more refined representation of the image. This is done by Equation (4) as shown, where k represents the kernel size and σ represents the standard deviation of the Gaussian blur. Equation (5) provides the Gaussian kernel.

$$I_{blurred} = GaussianBlur(I_{resized}, (k, k), \sigma) \tag{4}$$

$$G(u, v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \quad //Gaussian\ kernel \tag{5}$$

Furthermore, one more round of Median Blur is utilized to ensure thorough noise reduction. The formula for this process is illustrated in Equation (6). The blurred image at position (i, j) is calculated using Equation (7).

$$I_{noise\_filtered} = MedianBlur(I_{blurred}, k) \tag{6}$$

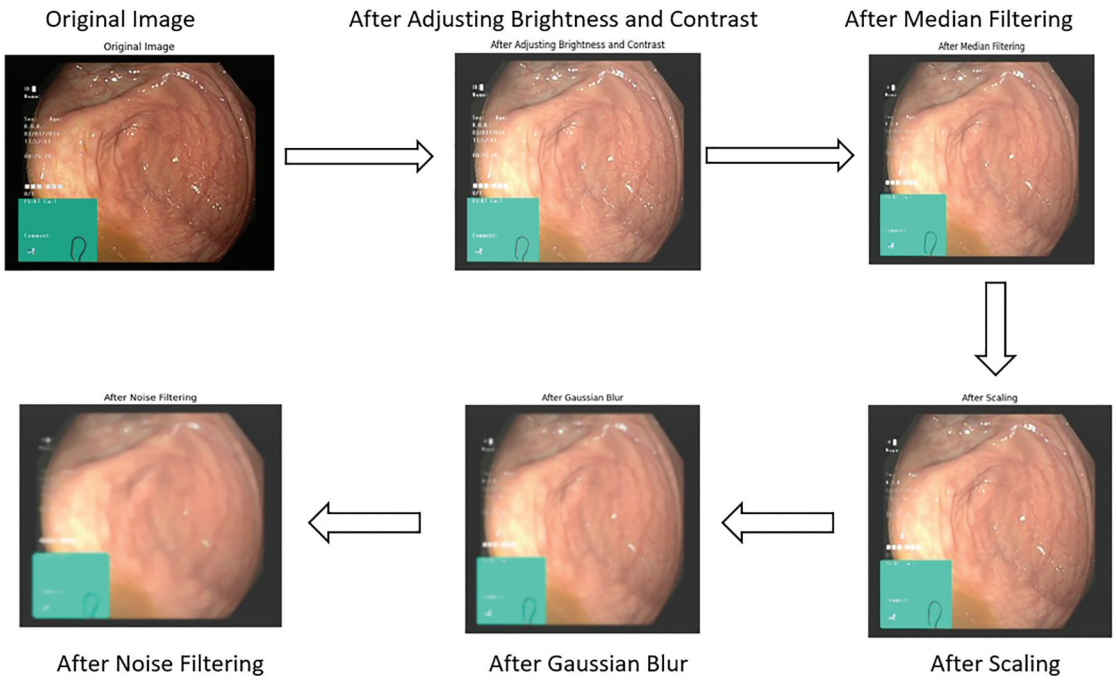

Figure 3. Workflow of the preprocessing techniques.

$$blur\_img(i,j) = \sum_{u=-k}^{k} \times \sum_{v=-k}^{k} \times adj\_img(i+u, j+v) \cdot G(u,v) \tag{7}$$

These preprocessing techniques increase image quality, reduce noise, and successfully prepare images for deep-learning model training. The pre-processed images are then added to the images list, preparing them for training the model architecture. The primary goal of these procedures is to improve image quality, reduce noise, and prepare images for training and testing purposes. In addition to these preparation strategies, one can use the Keras library's preprocessing function $images = preprocess\_input(pictures)$. This line, added after preprocessing, conducts the necessary preprocessing procedures before training the images with the model. Figure 4 illustrates how the pre-trained architecture preprocesses the images. This step is required when utilizing pre-trained models such as ResNet50, DenseNet121, etc. This line prepares images to match the deep-learning models' expected format, modifying pixel values to ensure that the model interprets the data correctly during training and prediction. This line includes steps like normalization, mean subtraction, and channel ordering.

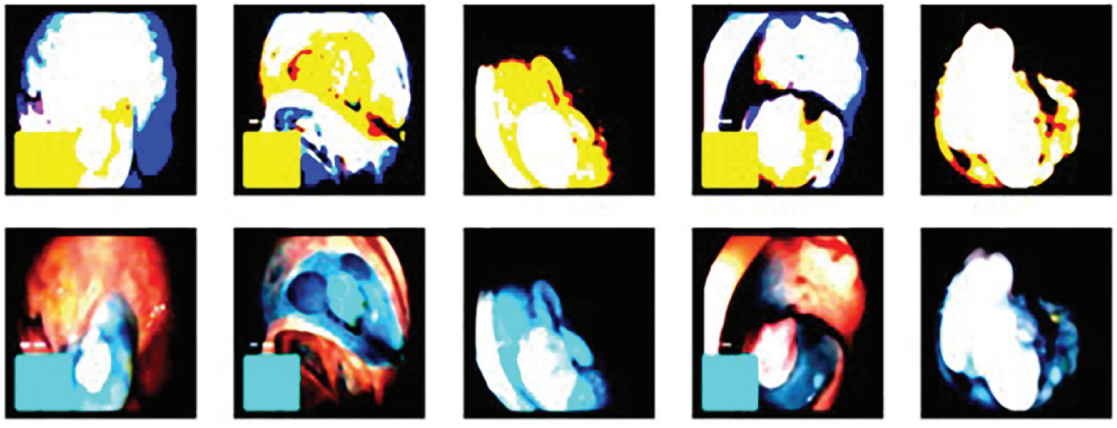

Figure 4. Preprocess input from the Keras library of ResNet50 and DenseNet121.

Normalization is the process of adjusting the values of the pixels in images to a specific range, which is typically between 0 and 1 or −1 and 1. Equation (8) specifies the normalizing formula as follows:

$$normalized\_pixel = \frac{original_{pixel} - mean}{standard_{deviation}} \tag{8}$$

Mean subtraction involves subtracting the mean pixel values of the images from each pixel in the image. The formula goes as shown in Equation (9).

$$adju\_pixel = original\_pixel - mean\_value \tag{9}$$

Channel ordering guarantees the structure of image data in the expected format of the model, such as RGB (Red, Green, Blue) or BGR (Blue, Green, Red). This reordering guarantees that there is constant coordination with the model's design.

With regard to medical image analysis, where annotated datasets are often rare and obtaining new information can be costly and time-consuming, data augmentation offers a powerful means to expand the accessible information and improve the presentation of AI models. It produces different varieties of existing medical images through transformations such as rotation, scaling, flipping, cropping, and adding noise. Data augmentation enables models to learn from a more comprehensive and representative set of examples. Table 1 provides an overview of data augmentation techniques, along with their descriptions for training the model. It makes it easier to know the exact image changes used and how each approach contributes to the augmentation process, as shown in Figure 5.

Table 1. Data augmentation applied to the images.

Technique	Description	Value	Action on Images
Rotation	Rotates the image by a specified angle range in degrees	Rotation range: 20^0	Rotates the images
Width Shift	Shifts the image horizontally by a fraction of its width	Width shift range: 0.2	Shifts the images horizontally
Height Shift	Shifts the image vertically by a fraction of its height	Height shift range: 0.2	Shifts the images vertically
Horizontal Flip	Flips the image horizontally	Enabled	Flips the images horizontally

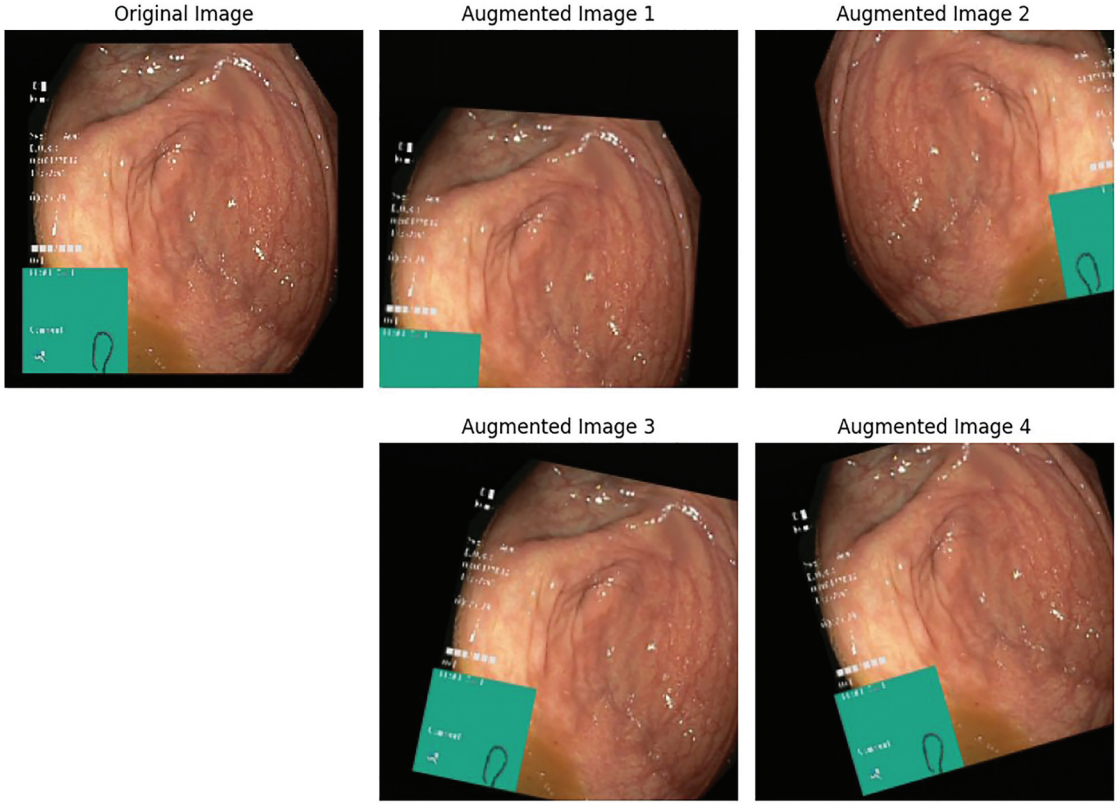

Figure 5. Image after data augmentation.

4.2. Splitting of data

Data splitting is a frequent technique in machine learning models that divides a dataset into training and testing sets. This split allows the model to undergo training on one set of data and testing on another, thereby revealing its performance on previously unknown data. Here, the photos and labels are separated into training and testing sets, with $test_{size} = 0.2$, indicating that 20% of the data utilized for training. Table 2 illustrates the arrays' forms and descriptions as a consequence of picture splitting. Figure 6 also depicts the splitting of images for testing and training using the DenseNet121 model.

Where,

$$X_{train} \in R(1-t) \times n \quad (Training\ set\ features)$$

$$X_{test} \in Rt \times n \quad (Testing\ set\ features)$$

$$y_{train} \in R(1-t) \times 1 \quad (Training\ set\ labels)$$

$$y_{test} \in Rt \times 1 \quad (Testing\ set\ labels)$$

Table 2. Splitting of data.

Splitting of Data	Shape	Description
X_{train}	(1 - t) × n	Input features for training set from dataset
X_{test}	t × n	Input features for testing set from dataset
y_{train}	(1 - t) × 1	Labels for training set from dataset
y_{test}	t × 1	Labels for testing set from dataset

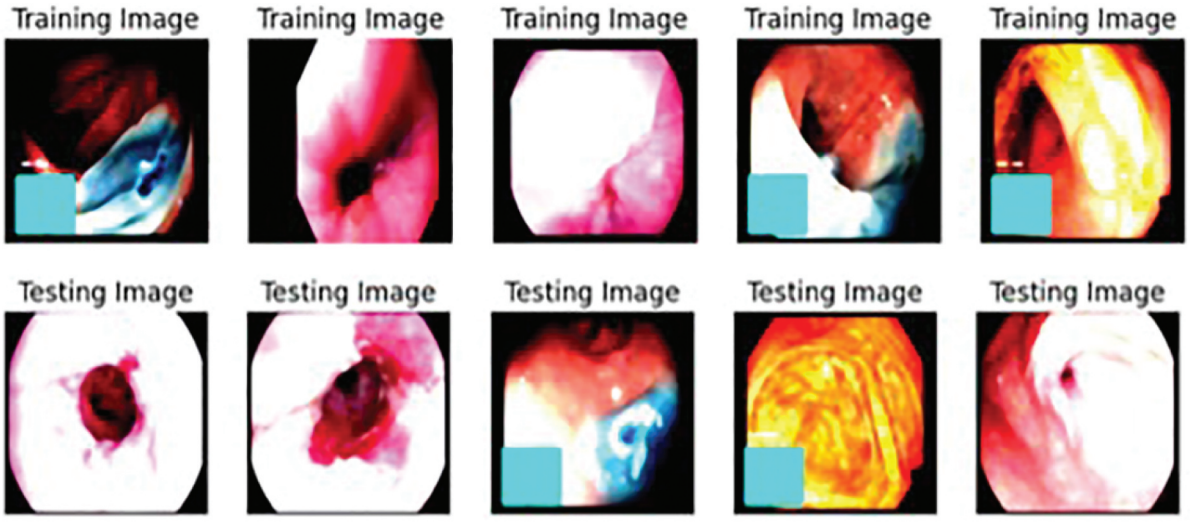

Figure 6. Splitting for DenseNet121.

Algorithm 1. Data preprocessing and image enhancement with Gaussian blur

Input:
Inputs considered: Dataset directory (data_dir), Number of classes (k), Test size (test_size), Random state (random_state), Model paths for ResNet (Rn_path) and DenseNet (Dn_path)

Output:
CM: Trained combined model

1: Initialize:
C ← categories (data_dir) ; [Extract categories from dataset directory]
I, L ← [ ], [ ] ; [Initialize empty lists for images and labels

2: Preprocessing of endoscopic images:
for $c \in C$ **do**
  $i \leftarrow index(c)$; [Get index of category]
  for $p \in paths(data\_dir, c)$ **do**
    img_array ← to_array(img) ; [Convert image to array]
    $\alpha, \beta \leftarrow$ scale_params( ) ; [Get scaling parameters]
    adj_img ← $\alpha \times img\_array + \beta$; [Adjust image array]
    $G(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$; [Gaussian kernel]
    blur_img ← convolve (adj_img, G) ; [Apply Gaussian blur]
    $blur\_img(i,j) = \sum_{u=-k}^{k} \times \sum_{v=-k}^{k} \times adj\_img(i+u, j+v) \cdot G(u,v)$
    I[i] ← I · append(blur_img) ; [Append processed image to list]
    L[i] ← L · append(i) ; [Append label index to list]
  end
  End for
end
End for

$X, Y \leftarrow to\_array(I), to\_categorical(L,k)$; [Convert lists to arrays and make labels categorical]
$X_{train}, X_{test}, Y_{train}, Y_{test} \leftarrow split(X, Y, test\_size, random\_state)$; [Split data into training and testing sets]

5. Model building

The TensorFlow application imports various libraries such as Dense, GlobalAveragePooling2D, and Dropout from the Keras library for model building. This analysis utilizes and evaluates several pre-trained models, such as ResNet50, DenseNet121, EfficientNetB0, InceptionV3, and Xception, for their effectiveness. These models were trained on the dataset to construct the final model that achieves the top two accuracy levels for disease detection. The weights of these pre-trained models were imported from ImageNet for feature extraction. Applications such as image classification, object detection, segmentation, and other computer vision models utilize these models.

5.1. ResNet50

He et al. (2015) proposed a 50-layer deep convolutional network model called ResNet50. It is known for its ability to train very deep networks effectively. ResNet50 engages residual connections, allowing the model to learn and hop over certain layers, which helps to ease the challenges associated with training extremely deep neural networks. In this analysis, ImageNet weights are loaded into the model's architecture, excluding the top layers. This is done to grasp the lower-level feature maps that were acquired by the model during its training on the ImageNet dataset, which encompasses a diverse range of images. The top layers are excluded, which enables the incorporation of custom layers appropriate for the particular task.

5.2. DenseNet121

DenseNet121 is a CNN model that contains 121 layers of a deep convolutional neural network. It uses dense blocks, where each layer receives input from all previous layers, to aid feature reuse and efficient model learning. DenseNet121 connects each layer directly and in a forward manner, unlike other models. ImageNet weights also load this architecture, excluding the top layers of the model. This allows the model to leverage the lower-level layers for feature extraction while training on the datasets. It provides a better representation of images and has enhanced parameter efficiency, making the model network simple to train.

5.3. EfficientNetB0

It is a variant of the EfficientNet architecture. It employs a compound scaling method that consistently scales all dimensions of the network's depth, width, and resolution. ImageNet weights are also loaded into this architecture, excluding the top layers. Language processing also utilizes this model. EfficientNetB0 achieves extraordinary levels of efficiency without compromising accuracy.

5.4. InceptionV3

According to Equation (3), the InceptionV3 model is a CNN deep learning model that uses a blend of 1×1, 3×3, and 5×5 convolutions, as well as 3×3 max pooling. ImageNet weights are also loaded into this architecture, with the top layers excluded. InceptionV3 uses a mix of various mixed convolutions within a similar module, which is unique compared to other models that use fixed-size convolutions.

5.5. Xception

Xception, also known as the "extreme" version of the Inception module, has a deep CNN architecture that includes layers with depth-wise separable convolutions. This architecture allows for significantly faster training of each layer, as it requires fewer parameters and matrix multiplications. ImageNet weights are also loaded into this architecture, with the top layers excluded.

The top layers are excluded in all of these models, allowing custom layers suitable for classification and prediction to be added. The same type of pseudocode is used in all the models. First, it loads the pre-trained models with ImageNet weights, excluding the top layers defined as the base model. A Global Average Pooling (GAP) layer processes the output of these models. The GAP layer reduces each feature map's spatial dimensions to a unique value by taking the average of all the values. Equation (10) provides the mathematical representation.

$$F \in \mathbb{R}^{H \times W} \tag{10}$$

The resultant of the GAP layer is calculated as Equation (11),

$$out = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{ij} \tag{11}$$

A drop-out layer is added for regularization next. During each training iteration, the dropout layer randomly deactivates a selected number of the network's neurons, depending on the specific neurons. The model learns more robust properties, and a reduction in the rate parameter, now set at 0.5, indicates the percentage of input units to drop. A fully connected dense layer, consisting of 1024 units, and a ReLU (Rectified Linear Unit) activation function are added to the dropout layer next. This layer learns features from the output of the previous layers, introducing non-linearity into the model. Equation (12) provides the mathematical representation of the model.

$$f(x) = maximum(0, x) \tag{12}$$

Here, x is the output tensor of the base model is denoted as X. Finally, a logistic layer with a number of units equal to the number of classes in the dataset is added. This layer uses the SoftMax activation function to generate the class probabilities for each category in the classification and detection task. The mathematical representation of the SoftMax function is defined in Equation (13).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{13}$$

Figure 7 depicts the accuracy and loss of each model, demonstrating its performance over more than 50 epochs. This graph indicates that the models are performing well, with no instances of overfitting or underfitting in the hybrid model. Table 3 summarizes the performance measures for the models trained on

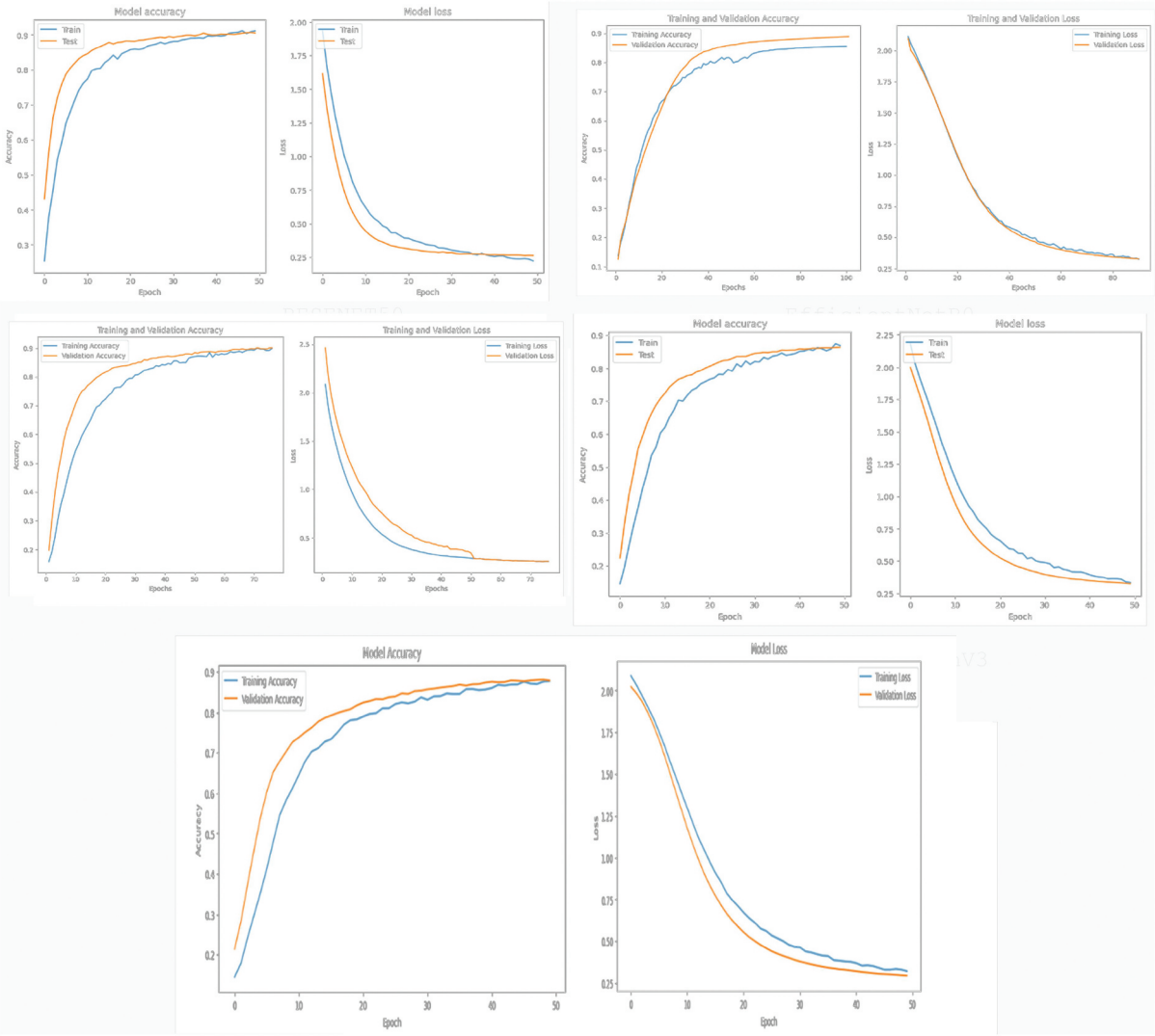

Figure 7. Model accuracy and loss of each model.

Table 3. Model metrics of the Kvasir dataset v2.

Model	Accuracy	F1-Score	Precision	Recall
Resnet50	90.56%	90.55%	90.81%	90.57%
DenseNET121	90.13%	90.00%	90.74%	90.12%
EfficientB0	87.19%	87.00%	87.72%	87.27%
InceptionV3	86.44%	86.37%	86.59%	86.45%
Xception	87.88%	87.78%	88.55%	87.95%

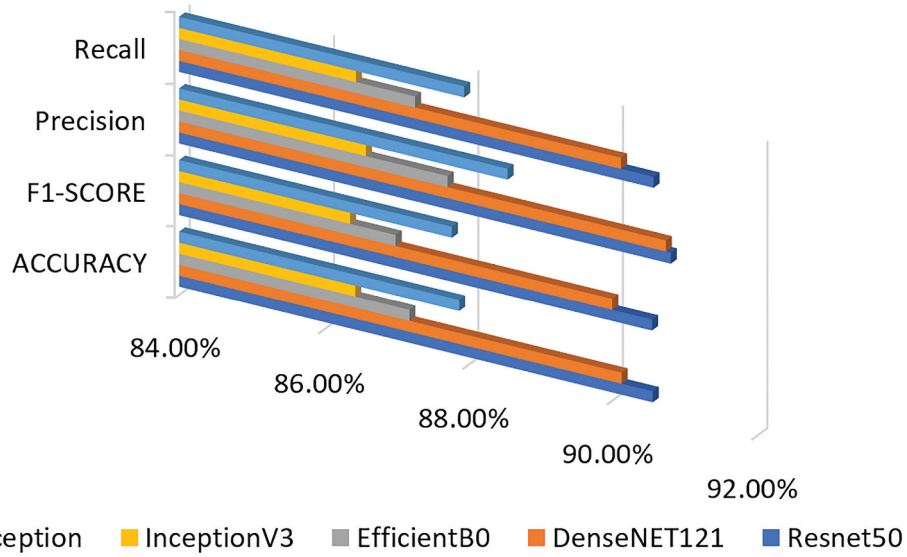

Figure 8. Graphical representation of metrics.

the Kvasir Dataset V2, such as accuracy, F1-score, precision, recall, and others. By examining Figures 7 and 8 (showing the graphical representation of Table 3), one can determine the best two models in terms of accuracy, precision, recall, and F1-score. These models were chosen to form a hybrid model that aims to improve metrics such as F1-score, recall, accuracy, and so on. By leveraging the strengths of the chosen models, the hybrid model aims to enhance performance on previously encountered datasets.

6. Building DenseResConcatenation hybrid model from the top two evaluated models

Based on the image in Fig. 7 and the metrics in Figure 8, ResNet50 and DenseNet121 are the best-performing models. To build a hybrid model, one can use both models' structures. ResNet50 uses 48 convolutional layers, 1 MaxPool layer, and 1 Average Pool layer to do 3.8×10^9 floating-point calculations. Skip connections help mitigate the vanishing-gradient problem. Meanwhile, DenseNet121 has dense connections between its 121 layers, which promote feature reuse and simplify gradient flow. ResNet50 is a residual network. A residual network is a type of DAG (Direct Acyclic Graph) network that has residual (or shortcut) connections that bypass the main network layers. Initially, ResNet50 worked for image-recognition tasks, but as the study Wahab et al. (2024) notes, it can also be used to improve accuracy in non-computer-vision applications. First, as shown in Figure 9, the model receives images in the format (None,

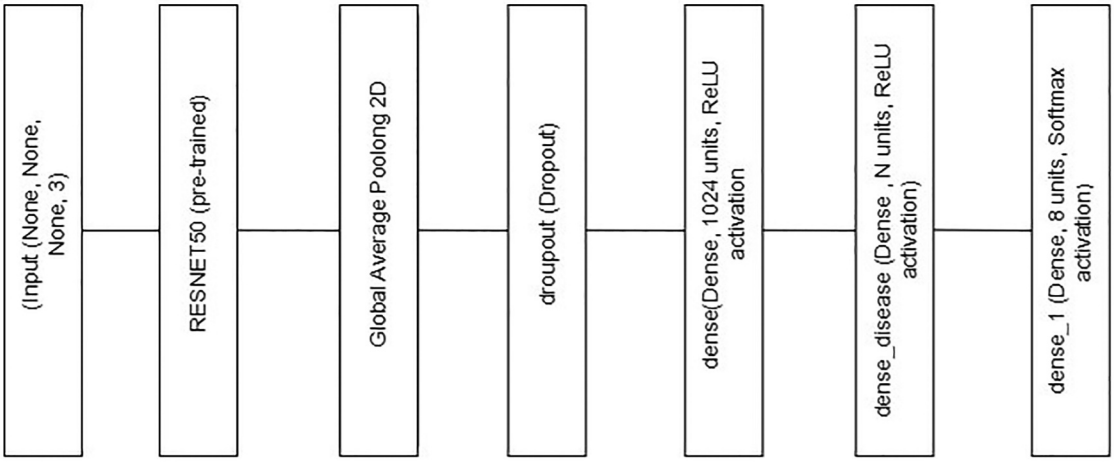

Figure 9. Workflow of ResNet50.

None, None, 3) as input. This input shape allows the model to handle input images of varying sizes, as the images from the dataset vary from 720 × 576 up to 1920 × 1072 pixels. Also, during preprocessing, the images are resized to a fixed size (224 × 224) to ensure consistency across all input images. The preprocessing techniques ensure that all input images fed into the model have the same dimensions. The images are then sent to the defined ResNet50 architecture. As shown in Figure 10, ResNet50's use of "skip connections" or "shortcuts" is a key component that enables the gradient to backpropagate to prior layers directly.

$$y = F(h, W_i) + h \tag{14}$$

The layer's initial and output vectors are represented by h and y in Equation (12). The function is defined as the residual mapping that needs to be learned in Equation (14). After learning the residual mapping, it makes pushing the weights easier, which helps it contest the vanishing gradient problem. Then the images are passed through GAP, Dropout, and SoftMax, as explained in Equations (10–13). DenseNet121 is a dense CNN network that has 121 layers that highlight the dense connections between layers. Figure 11 illustrates that DenseNet121 connects layers directly with each other, a departure from the natural CNN architecture. For every 'L' layer, there are $L(L+1)/2$ direct connections. This connectivity promotes feature reuse, simplifies gradient flow, and helps lessen the vanishing gradient problem. Figure 12 illustrates that preprocessing techniques initially take the image as input, assigning the input shape as None and resizing it to the desired size. Formally, the lth layer in a DenseNet receives the feature maps of all previous layers, $h_0, h_1, \ldots, h_{l-1}$, as input shown in Equation (15).

$$h_l = H_l([h_0, h_1, \cdots, h_{l-1}]) \tag{15}$$

Where $[h_0, h_1, \cdots, h_{l-1}]$ is the concatenation of the feature maps produced in layers from $0, \cdots, l-1$ and $Hl(\cdot)$ is a combined function of the operations.

In Figure 11, the architecture of DenseNet121 heavily relies on transition layers, which act as intermediary layers between the blocks to regulate the expansion of feature map dimensions and minimize computational complexity. First, there is a batch normalization layer. Next is a 1 × 1 convolutional layer to reduce the number of channels. Finally, there is a pooling layer (such as average pooling) to reduce the spatial dimensions.

Figure 13 shows that the process starts with loading the pre-trained ResNet50 and DenseNet121 models that were trained on the dataset for picture detection and classification. The sample input shape is given to

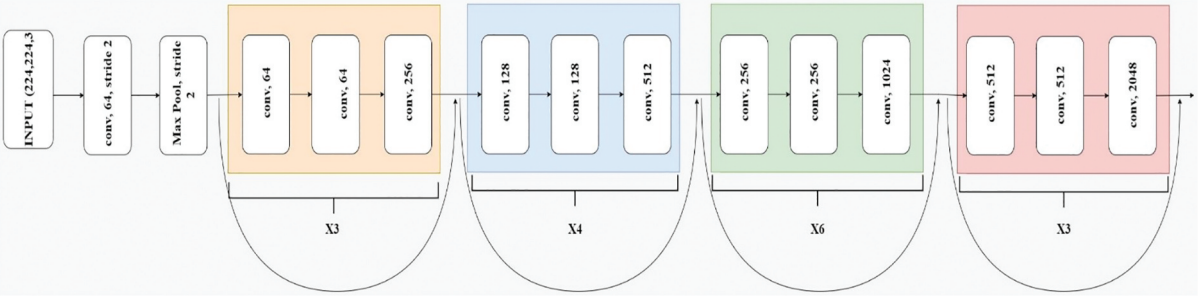

Figure 10. ResNet50 architecture.

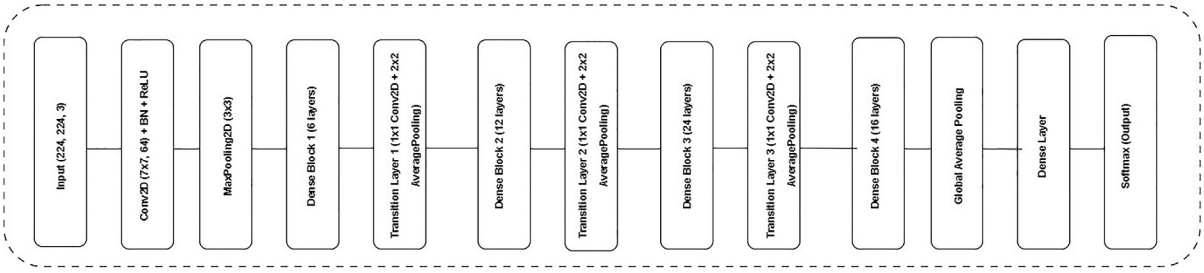

Figure 11. Desnet121 architecture.

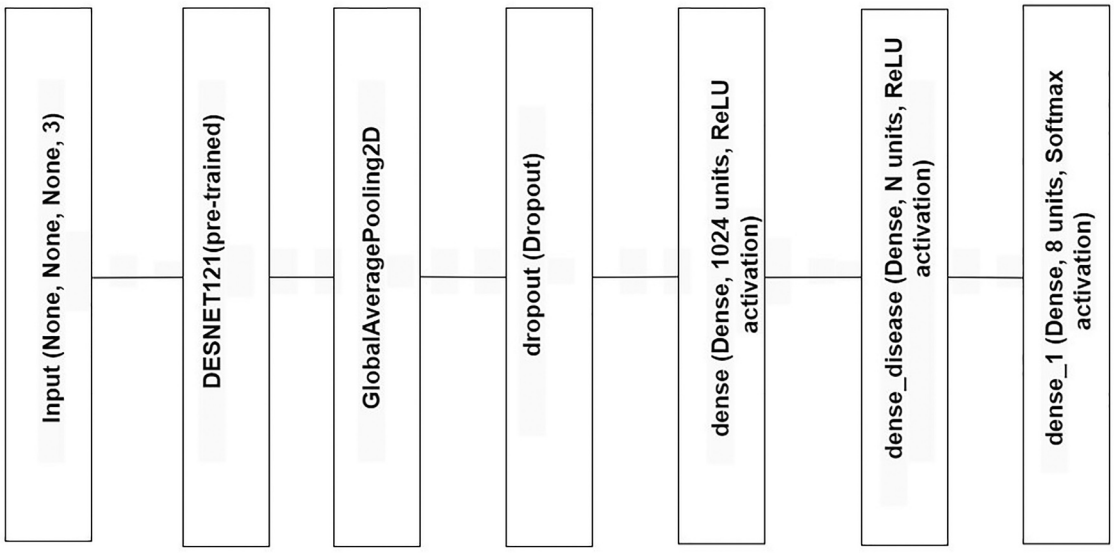

Figure 12. Workflow of DenseNet121.

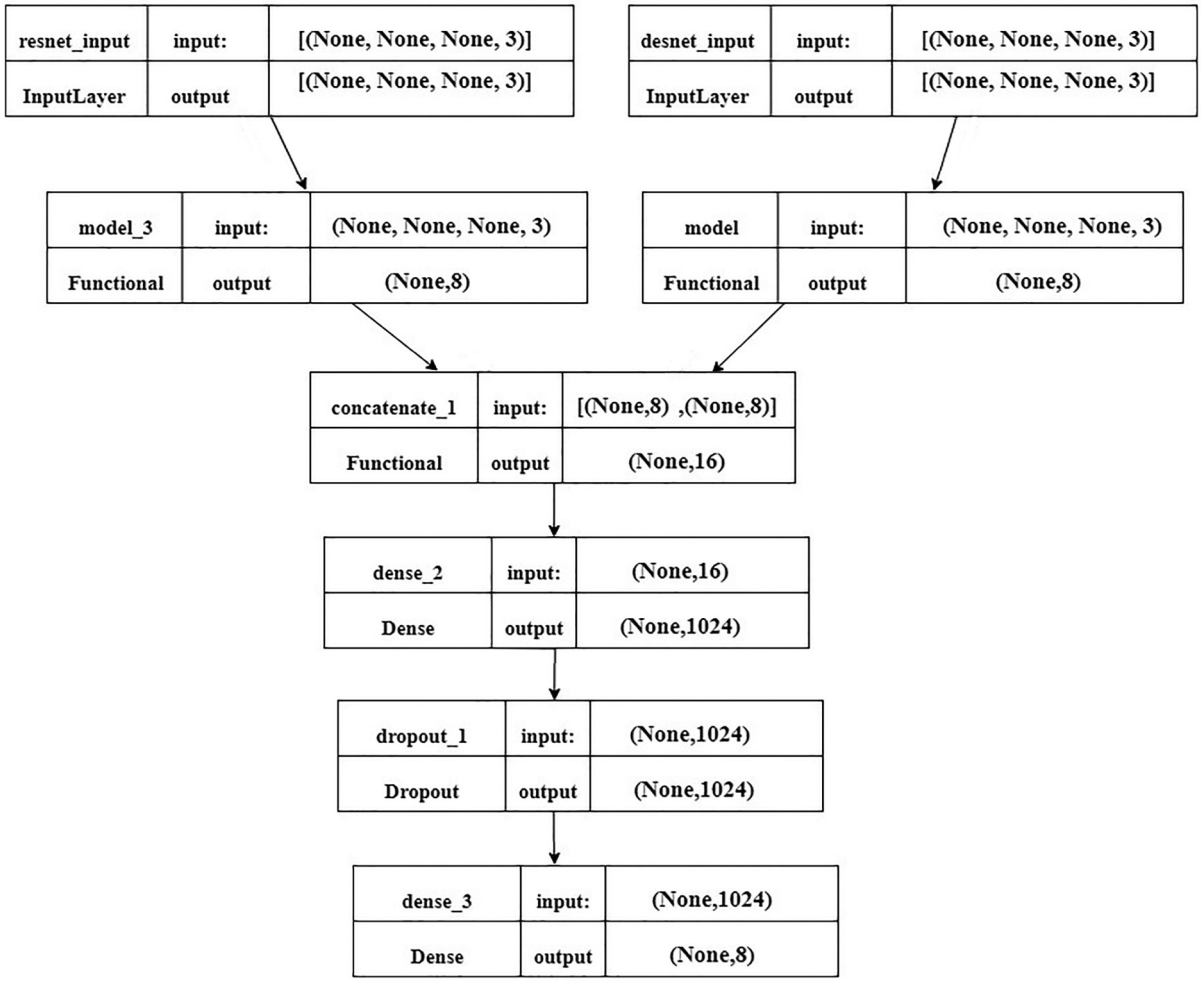

Figure 13. Workflow of the DenseResConcatenation hybrid model.

both models as a pre-processed image (224, 224). The layers of both models are then frozen to prevent further training, and their pre-trained weights are utilized as permanent features. The mathematical representation of these qualities is as follows:

$$C(x) = F_{ResNet} \oplus F_{DenseNet} \tag{16}$$

For example, in Equation (16), FResNet stands for the output feature maps of ResNet50, and FDenseNet for the output feature maps of DenseNet121. $C(x)$ is the output of the hybrid model, and $F_{ResNet} \oplus F_{DenseNet}$ is the sum of the outputs of ResNet50 and DenseNet121. The $C(x)$ is then passed through a dense layer with 50% dropout to prevent overfitting of the model. Another dense layer, consisting of several units, corresponds to the number of categories, and a SoftMax activation function generates probabilities for each category. In the hybrid model, "(None, None, None, 3)" represents a batch of images, where each image is in the form of RGB, i.e., 3 color image channels. The hybrid model concatenates the outputs from both the ResNet50 and DenseNet121 models to form a single layer with the shape "(None, 16)." In this layer, the hybrid model learns the combined complex features from both models. Finally, a concatenated layer is connected to a dense (fully connected) layer with 1024 units. This layer learns to make predictions based on the combined high-level features. To avoid overfitting, this layer randomly sets a portion of the input units to 0 at each update during training. The model requires an Adam Optimizer with a lower learning rate to update the parameters, such as weights and biases. This optimizer combines the benefits of both momentum and RMS crop techniques for efficient neural network optimization. Equation (17) provides the mathematical representation.

$$\Theta_{ti+1} = \theta_{ti} - \alpha \times m_{ti} / \sqrt{(v_{ti} + \varepsilon)} \tag{17}$$

Here, θ represents the parameters of the neural network model, and α represents the learning rate; in this case, it is 1×10^{-6}, and θ_{ti+1} is the most recent value of the parameters at time step t_{i+1}. The parameters' current value at time step ti is denoted by θ_{ti}. m_t is the exponentially moldy average of previous gradients (first moment estimate). v_t is the exponentially moldy average of past squared gradients (second moment estimate). ε is a small constant (typically 10^{-7}) added to the denominator for numerical stability.

The values of m_{ti} and v_{ti} are calculated using Equations (18) and (19), respectively.

$$m_{ti} = \beta_1 \cdot m_{ti-1} + \left(1 - \beta_1\right) \cdot \nabla Jl(\theta_{ti}) \tag{18}$$

$$v_{ti} = \beta_2 \cdot v_{ti-1} + \left(1 - \beta_2\right) \cdot (\nabla Jl(\theta_{ti}))^2 \tag{19}$$

$Jl(\theta t)$ is the gradient of the loss function, considering the parameters θ_{ti} and the exponential decay rates for the first and second moments estimated at time step $t - 1$, and ($J(\theta t))2$, typically set to 0.90 and 0.999, respectively.

This DenseResConcatenation hybrid model has a dense layer activated by "ReLU." It can be shown in Equation (20), where W_1 and b_1 are the dense layer's weights and bias, and Z_1 is the input to the ReLU function. In Equation (21), the output after applying the ReLU function is denoted by X_1. The dropout layer can be represented as shown in Equation (22), where $X_1^{'}$ is the output after applying dropout. The dense layer with the SoftMax function can be represented as shown in Equations (23) and (24), where W_2 and b_2 are the weights and bias of the final dense layer, and where Z_2 is the input to the SoftMax function and X_2 is the output after applying the SoftMax function, respectively. Therefore, the final output of the hybrid model can be written as $Y(x)$ in Equation (25).

$$Z_1 = (W_1 \times C(x) + b_1) \tag{20}$$

$$X_1 = max(0, Z_1) \tag{21}$$

$$X_1^{'} = \frac{X_1 \odot Bernoulli(1 - p)}{(1 - p)} \tag{22}$$

Where, $\odot$ denotes the element-wise multiplication.

$$Z_2 = W_2 \times X_1^{'} + b_2 \tag{23}$$

$$X_2 = \frac{eZ_2}{\sum eZ_2} \tag{24}$$

$$Y(x) = X_2 \tag{25}$$

Algorithm 2. Implementation of Hybrid model

Input:
d ← dataset

Output:
CM: Trained combined model

for *each epoch* ∈ *[1, E]:* **do**
  R ← load_model ('resnet_path')
  D ← load_model ('densenet_path') [Load pre-trained models]
  for *each layer* ∈ *R:* **do**
    layer.trainable ← False [Freeze layers in resnet models]
  end
  end for
  for *each layer* ∈ *D:* **do**
    layer.trainable ← False [Freeze layers in Denset models]
  end
  end for
  for *each image* ∈ *d:* **do**
    if *model is* R_n*:* **then**
      R_output ← R.predict(image) [Generate predictions using ResNet model]
    end
    end if
    if *model is* D_n*:* **then**
      D_output ← D.predict(image) [Generate predictions using DenseNet model]
    end
    end if
    combined_output ← concatenate (R_output, D_output) [Store the combined output for further processing]
    optimizer = Adam (η)
    where $\eta = 1 \times 10^6$
  end
  end for
end
end for

7. Results and discussions

The hybrid model was assessed using metrics such as F1-score, precision, recall, and confusion matrix, among others. The model's performance was evaluated on two different datasets: Kvasir Dataset v2 and Kvasir Dataset v1. These metrics demonstrated a significant rise in values compared to ResNet50 and DenseNet121 on an unknown dataset, as well as a high improvement in accuracy and other metrics.

Accuracy is defined as the percentage of properly categorized instances among the total number of instances examined. Our model obtained an accuracy of 93.63% on the Kvasir Dataset v2, suggesting that it accurately identified the vast majority of occurrences. The precision and recall scores averaged 94.04% and 93.63%, respectively, demonstrating the model's dependability in predicting true positive occurrences while accurately detecting actual positive cases. Table 4 shows that the F1 score, which considers both precision and recall, was 93.57%, indicating high overall performance. A detailed review of the classification report indicated excellent accuracy and recall scores across multiple classes, suggesting the model's ability to properly identify positive examples while avoiding false positives. Notably, the model showed excellent sensitivity and specificity, demonstrating its ability to effectively distinguish between positive and negative

Table 4. Evaluation metrics of the DenseResConcatenation hybrid model over 2 datasets.

Metrices	Kvasir Dataset v2	Kvasir Dataset v1
Accuracy	93.62%	94.15%
F1 - score	93.56%	94.12%
Precision	94.03%	94.28%
Recall	93.62%	94.15%
Average Sensitivity	93.62%	94.15%
Average Specificity	99.08%	99.16%

events. Although overall performance remained excellent, the model faced issues in specific categories, including "esophagitis," where the recall rate was significantly lower at 72%. Despite this, the model achieved strong accuracy and recall scores in other classes, with an average sensitivity of 93.625% and a specificity of 99.089%, as shown in Figure 14.

In comparison, the model trained on the Kvasir Dataset v1, a dataset containing 4000 images, achieved a slightly higher accuracy of 94.15% and an F1 score of 94.12%. The confusion matrix for this model revealed fewer misclassifications and a better-balanced performance across all categories. For example, the "dyed-lifted-polyps" and "dyed-resection-margins" classes had precision and recall scores between 93% and 94%, demonstrating the model's ability to reliably anticipate these situations. As shown in Figure 15, the "normal-pylorus" and "normal-cecum" classes had nearly flawless precision and recall ratings, demonstrating that the model is very good at distinguishing normal anatomical components.

Overall, the model showed excellent sensitivity and specificity, demonstrating its ability to effectively distinguish between positive and negative situations. Figure 16 shows the evaluation metrics by class for the Kvasir dataset v2, which is also similar to dataset v1.

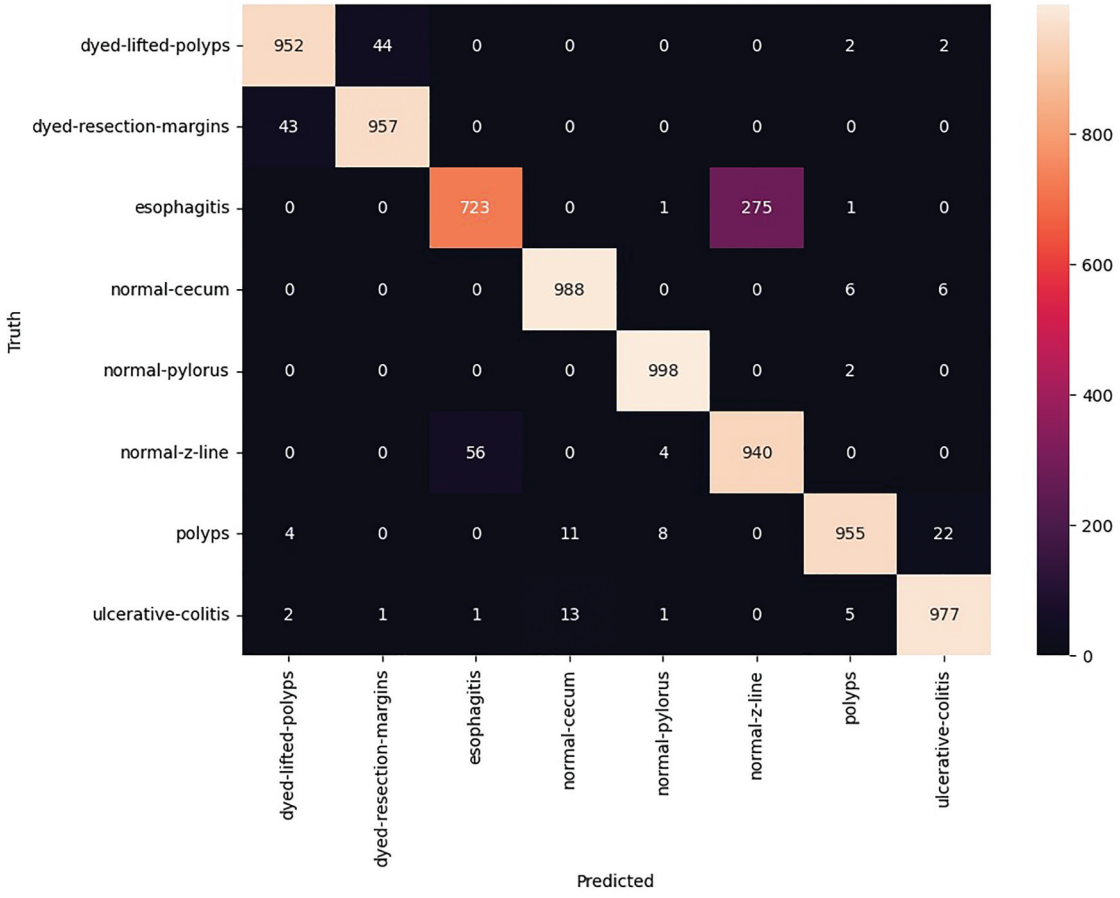

Figure 14. Visualization of the confusion matrix on the Kvasir dataset v2.

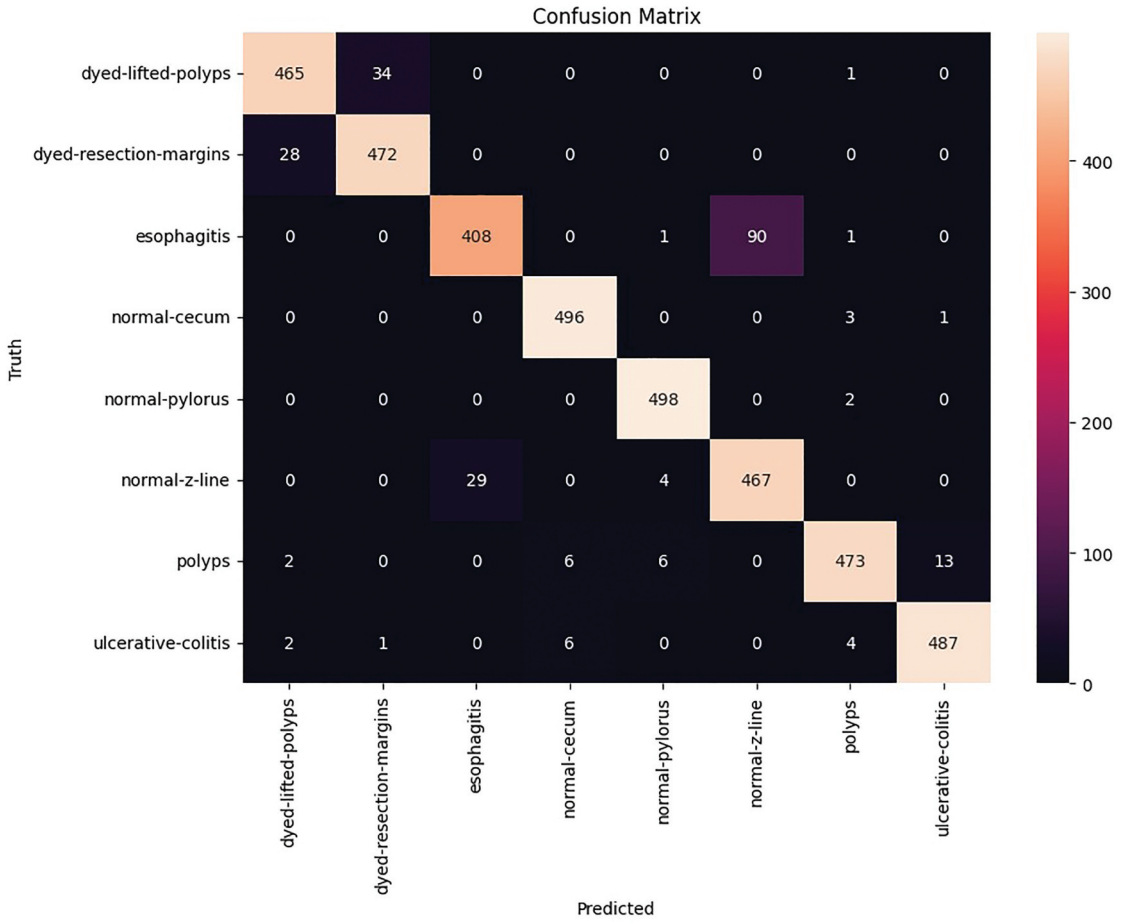

Figure 15. Visualization of the confusion matrix on the Kvasir dataset v1.

The DenseResConcatenation architecture, while achieving superior accuracy (94.15% on Kvasir V1), requires substantial computational resources due to the concatenation of ResNet50 and DenseNet121 feature maps. This computational complexity may limit real-time deployment during endoscopic procedures. The hybrid model demands approximately 3.8 × 109 floating-point operations from ResNet50 alone, plus additional computations from DenseNet121. Future work will explore model-compression techniques, including pruning, quantization, and knowledge distillation, to reduce computational overhead. We will also investigate lightweight architectural modifications and ensemble methods that maintain diagnostic accuracy while enabling real-time clinical deployment. This study is limited to static endoscopic image classification and does not address real-time video-sequence analysis. Clinical endoscopy involves continuous video streams with temporal information that could enhance diagnostic consistency and reduce false positives through temporal context. Future iterations will incorporate temporal-modeling approaches such as 3D convolutional neural networks (3D-CNNs), long short-term memory (LSTM) networks, or transformer-based architectures to process video sequences. This extension will be crucial for clinical implementation, where endoscopists work with continuous video feeds rather than individual frames.

7.1. Disease prediction

To demonstrate their practical use, trained algorithms were employed to predict diseases based on endoscopic images. The images were also preprocessed before being fed into the applicable models for prediction. The anticipated disease categories were then analyzed, and appropriate interventions were recommended based on the findings, as shown in Figure 17. The model confidently predicted that the sample image belonged to the "ulcerative colitis" class. Given the severity of ulcerative colitis, it is recommended that a healthcare expert be consulted for further evaluation and treatment. Although overall

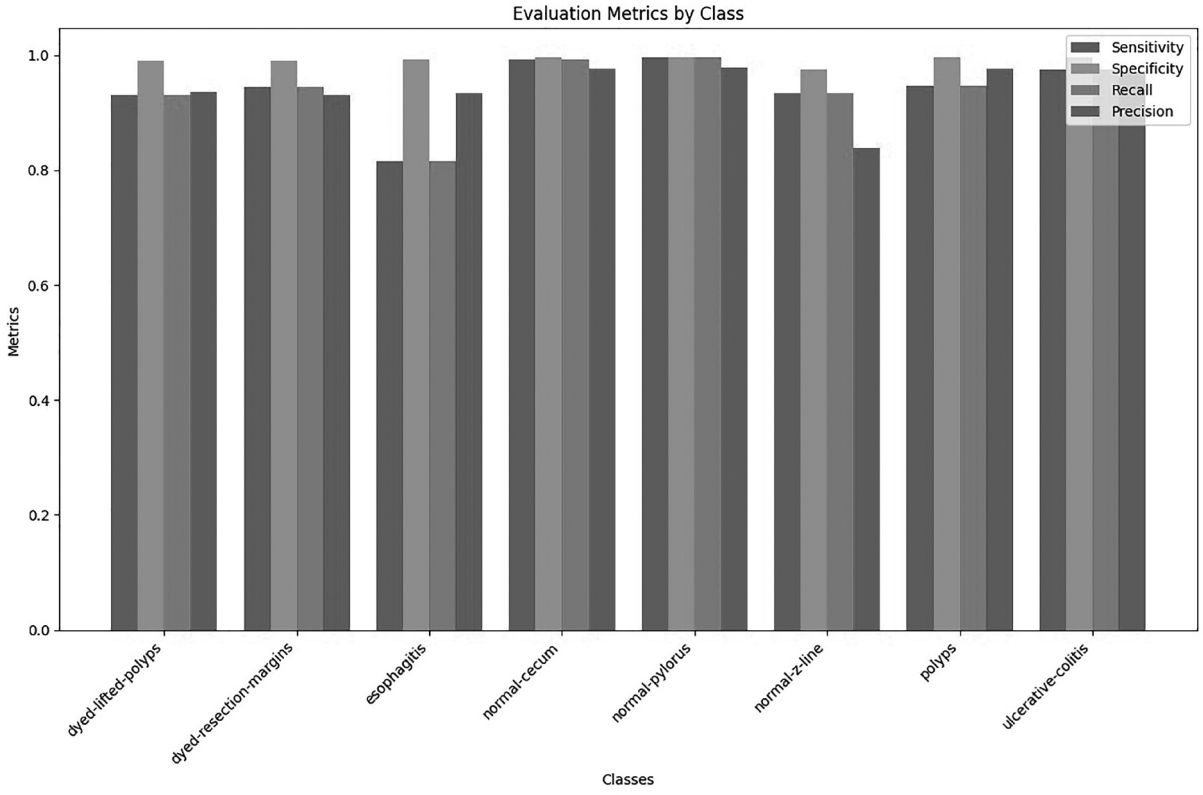

Figure 16. Evaluation metrics by class on the Kvasir dataset v2.

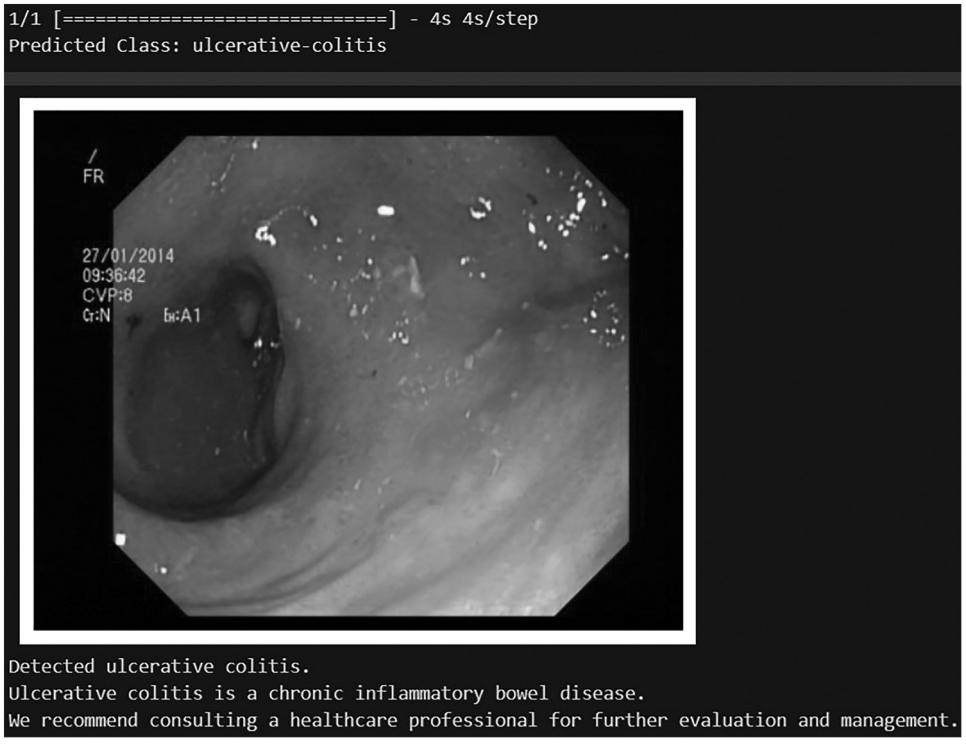

Figure 17. Image of disease prediction.

Table 5. Comparative performance analysis on Kvasir datasets.

Method	Dataset	Accuracy(%)	Precision(%)	Recall (%)	F1-Score (%)	Reference
ResNet50 (Baseline)	Kvasir V2	85.2	84.1	83.6	83.8	Rahim et al. (2020)
DenseNet121 (Baseline)	Kvasir V2	86.7	85.3	84.9	85.1	Talo et al. (2019)
EfficientNetB0	Kvasir V2	87.1	86.5	86.2	86.3	Gunasekaran et al. (2023)
Ensemble CNN	Kvasir V2	89.5	88.9	88.7	88.8	Sivari et al. (2023)
Hybrid Stacking Model	Kvasir V2	91.2	90.8	90.5	90.6	Khan et al. (2022)
DenseRes Concatenation	**Kvasir V2**	**93.62**	**94.03**	**93.62**	**93.56**	**Proposed work**
DenseRes Concatenation	**Kvasir V1**	**94.15**	**94.28**	**94.15**	**94.12**	**Proposed work**

performance remained excellent, the model faced challenges in specific categories, including "esophagitis," where the recall rate was significantly lower at 72%. This suboptimal performance on underrepresented classes highlights the need for more advanced data-augmentation strategies. Future work will incorporate GAN-based synthetic image generation and SMOTE (Synthetic Minority Over-sampling Technique) to better address class imbalance and improve performance on minority classes. Additionally, we plan to implement targeted augmentation techniques specifically designed for underrepresented disease categories.

DenseResConcatenation model in Table 5 demonstrates superior performance compared to existing methods on the same datasets. The hybrid architecture achieves a 2.42% improvement in accuracy over the best-performing ensemble method (Sivari et al. 2023) and a 6.92% improvement over single-backbone approaches (Rahim et al. 2020). This superior performance is attributed to the complementary feature-extraction capabilities of ResNet50's residual connections and DenseNet121's dense connectivity, enabling more comprehensive representation learning for complex endoscopic images.

8. Conclusion and future works

The DenseResConcatenation hybrid model, which was trained on both Kvasir Dataset v2 and Kvasir Dataset v1, can accurately predict gastrointestinal disorders from endoscopic images, as shown in a comparison study. The model trained on Kvasir Dataset v1 had somewhat greater accuracy (94.15%) than that trained on Kvasir Dataset v2 (93.63%), however, the model performed well overall on both datasets. An important feature of this research was the use of preprocessing and data augmentation approaches to improve the models' performance. Preprocessing stages included image resizing, color modification, and noise reduction, while data augmentation comprised techniques like rotation, flipping, and scaling to improve the variety of the training data. These approaches played a critical role in increasing the robustness and generalizability of models such as ResNet50 and DenseNet121. This hybrid model used a design that incorporated characteristics from ResNet50 and DenseNet121, harnessing the capabilities of both networks.

This hybrid model outperformed prior efforts by the authors Trinh and Nguyen (2021), indicating improved generalization across varied datasets. In the future, the main focus will be on improving the models even more to fix problems that were found during evaluation, such as the model trained on the Kvasir Dataset version 2 incorrectly labeling a condition as "esophagitis." Efforts will involve improving feature differentiation and balancing the dataset to enhance model performance across all domains. A key limitation of this study is the evaluation on a single-source dataset (Kvasir V1/V2) from one medical institution. Future work will focus on validating the DenseResConcatenation model across multi-center, diverse clinical datasets to assess generalizability across different patient populations, imaging equipment, and clinical protocols. This validation across diverse clinical environments will help establish the model's robustness and real-world applicability in various healthcare settings.

While the current model was evaluated using a standard train-test split, future research will incorporate 5-fold or 10-fold cross-validation to provide a more robust estimate of the model's generalization performance and ensure greater statistical stability.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Maheswari Subburaj http://orcid.org/0000-0001-6848-2032
Arun Kumar Sivaraman http://orcid.org/0000-0003-0514-484X
Sasi Kumar Periyasamy http://orcid.org/0000-0003-3510-3694
Janakiraman Nithiyanantham http://orcid.org/0000-0001-6616-5340
Kamalavelu Velayutham http://orcid.org/0009-0005-3111-4574

Acknowledgements

The authors wish to express their thanks to Dr. J. Jamuna, Department of Pathology, Velammal Medical College Hospital and Research Institute, Madurai, 625009, India, who provided valuable advice and oversight in guiding our work.

Author contributions

CRediT: **Maheswari Subburaj:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing; **Chinthamani Mohan Krishna:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing; **Arun Kumar Sivaraman:** Investigation, Project administration, Resources, Supervision; **Sasi Kumar Periyasamy:** Investigation, Methodology, Supervision, Writing – review & editing; **Janakiraman Nithiyanantham:** Data curation, Formal analysis, Software, Validation, Writing – review & editing; **Kamalavelu Velayutham:** Investigation, Supervision, Visualization, Writing – review & editing.

Ethical statements

The authors acknowledge that this research work falls outside the purview of requiring ethical approval, as it does not involve human participants or animal subjects.

Data availability statement

The data supporting the findings of this study are available in the public domain , with details provided in the reference section.

The Kvasir Dataset v2, used for gastrointestinal disease diagnosis through endoscopic images, is publicly available and curated under real-world clinical conditions. Dataset access link: https://datasets.simula.no/kvasir/

Code availability

A partial implementation of the DenseResConcatenation deep learning architecture proposed in this study is publicly available on GitHub. Repository link: https://github.com/mohankrishna6534/DenseResConcatenation/tree/main

References

Agrawal, T., R. Gupta, S. Sahu, and C. Y. Espy-Wilson. 2017. "SCL-UMD at the Medico Task-MediaEval 2017: Transfer Learning Based Classification of Medical Images." *MediaEval Benchmarking Initiative for Multimedia Evaluation* 17:13–15. http://slim-sig.irisa.fr/me17/Mediaeval_2017_paper_21.pdf.

Aruna, P., N. Puviarasan, and B. Palaniappan. 2007. "Diagnosis of Gastrointestinal Disorders Using DIAGNET." *Expert Systems with Applications* 32 (2): 329–335. https://doi.org/10.1016/j.eswa.2005.11.039.

Attallah, O., and M. Sharkas. 2021. "GASTRO-CADx: A Three Stages Framework for Diagnosing Gastrointestinal Diseases." *PeerJ Computer Science* 7:e423. https://doi.org/10.7717/peerj-cs.423.

Bharadwaj, S., N. Narula, P. Tandon, and M. Yaghoobi. 2018. "Role of Endoscopy in Inflammatory Bowel Disease." *The Gastroenterology Report* 6 (2): 75–82. https://doi.org/10.1093/gastro/goy006.

Bolourani, S., M. A. Tayebi, L. Diao, P. Wang, V. Patel, F. Manetta, and P. C. Lee. 2021. "Using Machine Learning to Predict Early Readmission Following Esophagectomy." *Journal of Thoracic and Cardiovascular Surgery* 161 (6): 1926–1939,e8. https://doi.org/10.1016/j.jtcvs.2020.04.172.

Buades, A., B. Coll, and J.-M. Morel. 2005. "A Non-Local Algorithm for Image Denoising." In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 60–65. San Diego, CA: IEEE. https://doi.org/10.1109/CVPR.2005.38.

Chen, P. J., M. C. Lin, M. J. Lai, J. C. Lin, H. H. Lu, and V. S. Tseng. 2018. "Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis." *Gastroenterology* 154 (3): 568–575. https://doi.org/10.1053/j.gastro.2017.10.010.

Cheung, V. 2017. DenResNet: Ensembling Dense Networks and Residual Networks. https://cs231n.stanford.edu/reports/2017/pdfs/933.pdf.

Dheir, I. M., and S. S. Abu-Naser. 2022. "Classification of Anomalies in Gastrointestinal Tract Using Deep Learning." *International Journal of Academic Engineering Research* 6 (3): 15–28.

Escobar, J., K. Sanchez, C. Hinojosa, H. Arguello, and S. Castillo. 2021. "Accurate Deep Learning-Based Gastrointestinal Disease Classification via Transfer Learning Strategy." In *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, 1–5. Popayán, Colombia: IEEE. https://doi.org/10.1109/stsiva53688.2021.959199.

Esteva, A., K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. J. Topol, J. Dean, and R. Socher. 2021. "Deep Learning-Enabled Medical Computer Vision." *NPJ Digital Medicine* 4:5. https://doi.org/10.1038/s41746-020-00376-2.

Gnanambal, I., and M. Rayappan. 2011. "New Hybrid Filtering Techniques for Removal of Gaussian Noise from Medical Images." *ARPN Journal of Engineering and Applied Sciences* 6 (2): 8–12.

Gonzalez, R. C., and R. E. Woods. 2018. *Digital Image Processing*. 4th ed. NY: Pearson.

Guimarães, P., A. Keller, T. Fehlmann, F. Lammert, and M. Casper. 2020. "Deep-Learning Based Detection of Gastric Precancerous Conditions." *Gut* 69 (1): 4–6. https://doi.org/10.1136/gutjnl-2019-319347.

Gunasekaran, H., K. Ramalakshmi, D. K. Swaminathan, M. Mazzara, and M. Mazzara. 2023. "GIT-Net: An Ensemble Deep Learning-Based GI Tract Classification of Endoscopic Images." *Bioengineering* 10 (7): 809. https://doi.org/10.3390/bioengineering10070809.

Haile, M. B., A. O. Salau, B. Enyew, and A. J. Belay. 2022. "Detection and Classification of Gastrointestinal Disease Using Convolutional Neural Network and SVM." *Cogent Engineering* 9 (1): 2084878. https://doi.org/10.1080/23311916.2022.2084878.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Deep Residual Learning for Image Recognition." In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. IEEE. https://doi.org/10.1109/cvpr.2016.90.

Higgins, P. D., M. Schwartz, J. Mapili, and E. M. Zimmermann. 2005. "Is Endoscopy Necessary for the Measurement of Disease Activity in Ulcerative Colitis?" *The American Journal of Gastroenterology* 100 (2): 355–361. https://doi.org/10.1111/j.1572-0241.2005.40641.x.

Jha, D., S. Ali, S. Hicks, V. Thambawita, H. Borgli, P. H. Smedsrud, T. D. Lange, et al.. 2021 "A Comprehensive Analysis of Classification Methods in Gastrointestinal Endoscopy Imaging." *Medical Image Analysis* 70:102007. https://doi.org/10.1016/j.media.2021.102007.

Khan, M. A., N. Sahar, W. Z. Khan, M. Alhaisoni, U. Tariq, M. H. Zayyan, Y. J. Kim, and B. Chang. 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.

Konstantin, P., R. Michael, H. Pål, G. Carsten, T. de Lange, R. Kristin, E. Sigrun, et al.. 2017 "A Comparison of Deep Learning with Global Features for Gastrointestinal Disease Detection." *Multimedia Tools & Applications* 76 (21): 22493–22525. https://doi.org/10.1007/s11042-017-4989-y.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521:436–444. https://doi.org/10.1038/nature14539.

Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. 2017. "A Survey on Deep Learning in Medical Image Analysis." *Medical Image Analysis* 42:60–88. https://doi.org/10.1016/j.media.2017.07.005.

Lonseko, Z. M., P. E. Adjei, W. Du, C. Luo, D. Hu, L. Zhu, T. Gan, and N. Rao. 2021. "Gastrointestinal Disease Classification in Endoscopic Images Using Attention-Guided Convolutional Neural Networks." *Applied Sciences* 11 (23): 11136. https://doi.org/10.3390/app112311136.

Mahmud, N., J. Cohen, K. Tsourides, and T. M. Berzin. 2015. "Computer Vision and Augmented Reality in Gastrointestinal Endoscopy." *The Gastroenterology Report* 3 (3): 179–184. https://doi.org/10.1093/gastro/gov027.

Mukhtorov, D., M. Rakhmonova, S. Muksimova, and Y. I. Cho. 2023. "Endoscopic Image Classification Based on Explainable Deep Learning." *Sensors* 23 (6): 3176. https://doi.org/10.3390/s23063176.

Nagao, S., Y. Tani, J. Shibata, Y. Tsuji, T. Tada, R. Ishihara, and M. Fujishiro. 2022. "Implementation of Artificial Intelligence in Upper Gastrointestinal Endoscopy." *DEN Open* 2 (1): e72. https://doi.org/10.1002/deo2.72.

Navale, S., A. Anjanikar, A. Mohammed, V. A. Sairam, S. Gite, S. Mahajan, and K. Nandhini. 2024. "Automated Classification of Gastrointestinal Abnormalities Using Convolutional Neural Networks." *International Journal of Intelligent Systems and Applications in Engineering* 12 (12s): 348–360. https://ijisae.org/index.php/IJISAE/article/view/4521.

Naz, J., M. I. Sharif, M. I. Sharif, S. Kadry, H. T. Rauf, and A. E. Ragab. 2023. "A Comparative Analysis of Optimization Algorithms for Gastrointestinal Abnormalities Recognition and Classification Based on Ensemble XcepNet23 and ResNet18 Features." *Biomedicines* 11 (6): 1723. https://doi.org/10.3390/biomedicines11061723.

Nogueira-Rodríguez, A., M. Reboiro-Jato, D. Glez-Peña, and H. López-Fernández. 2022. "Performance of Convolutional Neural Networks for Polyp Localization on Public Colonoscopy Image Datasets." *Diagnostics* 12 (4): 898. https://doi.org/10.3390/diagnostics12040898.

Owais, M., M. Arsalan, T. Mahmood, J. K. Kang, and K. R. Park. 2020. "Automated Diagnosis of Various Gastrointestinal Lesions Using a Deep Learning-Based Classification and Retrieval Framework with a Large Endoscopic Database: Model Development and Validation." *Journal of Medical Internet Research* 22 (11): e18563. https://doi.org/10.2196/18563.

Ozawa, T., S. Ishihara, M. Fujishiro, H. Saito, Y. Kumagai, S. Shichijo, K. Aoyama, and T. Tada. 2019. "Novel Computer-Assisted Diagnosis System for Endoscopic Disease Activity in Patients with Ulcerative Colitis." *Gastrointestinal Endoscopy* 89 (2): 416–421.e1. https://doi.org/10.1016/j.gie.2018.10.020.

Pogorelov, K., K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, et al. 2017. "Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection." In *Proceedings of the 8th ACM on Multimedia Systems Conference*, Taipei, Taiwan, June 20–23. 164–169, NY: Association for Computing Machinery. https://doi.org/10.1145/3083187.3083212.

Quindós, A., P. Laiz, J. Vitrià, and S. Seguí. 2023. "Self-Supervised Out-Of-Distribution Detection in Wireless Capsule Endoscopy Images." *Artificial Intelligence in Medicine* 143:102606. https://doi.org/10.1016/j.artmed.2023.102606.

Rahim, T., M. A. Usman, and S. Y. Shin. 2020. "A Survey on Contemporary Computer-Aided Tumor, Polyp, and Ulcer Detection Methods in Wireless Capsule Endoscopy Imaging." *Computerized Medical Imaging and Graphics* 85:101767. https://doi.org/10.1016/j.compmedimag.2020.101767.

Ramamurthy, K., T. T. George, Y. Shah, and P. Sasidhar. 2022. "A Novel Multi-Feature Fusion Method for Classification of Gastrointestinal Diseases Using Endoscopy Images." *Diagnostics* 12 (10): 2316. https://doi.org/10.3390/diagnostics12102316.

Rex, D. K., C. R. Boland, J. A. Dominitz, F. M. Giardiello, D. A. Johnson, T. Kaltenbach, T. R. Levin, D. Lieberman, and D. J. Robertson. 2017. "Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer." *The American Journal of Gastroenterology* 112 (7): 1016–1030. https://doi.org/10.1038/ajg.2017.174.

Rezvy, S., T. Zebin, B. Braden, W. Pang, S. Taylor, and X. W. Gao. 2020. "Transfer Learning for Endoscopy Disease Detection and Segmentation with Mask-RCNN Benchmark Architecture." *CEUR Workshop Proceedings* 2595:68–72. http://ceur-ws.org/Vol-2595/endoCV2020_paper_id_17.pdf.

Sang, D. V., T. Q. Chung, P. N. Lan, D. V. Hang, D. V. Long, and N. T. Thuy. 2021. "AG-CUResnest: A Novel Method for Colon Polyp Segmentation." ArXiv, abs/2105.00402. https://doi.org/10.48550/arXiv.2105.00402.

Sekuboyina, A. K., S. T. Devarakonda, and C. S. Seelamantula. 2017. "A Convolutional Neural Network Approach for Abnormality Detection in Wireless Capsule Endoscopy." In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI)*, Melbourne, Australia, 1057–1060. Piscataway, NJ: IEEE. https://doi.org/10.1109/ISBI.2017.7950698.

Sharma, A., R. Kumar, and P. Garg. 2023. "Deep Learning-Based Prediction Model for Diagnosing Gastrointestinal Diseases Using Endoscopy Images." *International Journal of Medical Informatics* 177:105142. https://doi.org/10.1016/j.ijmedinf.2023.105142.

Sivari, E., E. Bostanci, M. S. Guzel, K. Acici, T. Asuroglu, and T. Ercelebi Ayyildiz. 2023. "A New Approach for Gastrointestinal Tract Findings Detection and Classification: Deep Learning-Based Hybrid Stacking Ensemble Models." *Diagnostics* 13 (4): 720. https://doi.org/10.3390/diagnostics13040720.

Steinway, S. N., J. Saleh, B.-K. Koo, D. Delacour, and D.-H. Kim. 2020. "Human Microphysiological Models of Intestinal Tissue and Gut Microbiome." *Frontiers in Bioengineering and Biotechnology* 8:725. https://doi.org/10.3389/fbioe.2020.00725.

Su, Q., F. Wang, D. Chen, G. Chen, C. Li, and L. Wei. 2022. "Deep Convolutional Neural Networks with Ensemble Learning and Transfer Learning for Automated Detection of Gastrointestinal Diseases." *Computers in Biology and Medicine* 150:106054. https://doi.org/10.1016/j.compbiomed.2022.106054.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. 2021. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries." *CA: A Cancer Journal for Clinicians* 71 (3): 209–249. https://doi.org/10.3322/caac.21660.

Talo, M. 2019. "Automated Classification of Histopathology Images Using Transfer Learning." *Artificial Intelligence in Medicine* 101:101743. https://doi.org/10.1016/j.artmed.2019.101743.

Trinh, Q., and M. L. Nguyen. 2021. "Dense-Res Net for Endoscopic Image Classification." *Computer Science and Information Technology (CS & IT)*: 101–108. https://doi.org/10.5121/csit.2021.111108.

Ueyama, H., Y. Kato, Y. Akazawa, N. Yatagai, H. Komori, T. Takeda, K. Matsumoto, et al. 2021. "Application of Artificial Intelligence Using a Convolutional Neural Network for Diagnosis of Early Gastric Cancer Based on Magnifying Endoscopy with Narrow-Band Imaging." *Journal of Gastroenterology & Hepatology* 36 (2): 482–489. https://doi.org/10.1111/jgh.15190.

Vania, M., B. A. Tama, H. Maulahela, and S. Lim. 2023. "Recent Advances in Applying Machine Learning and Deep Learning to Detect Upper Gastrointestinal Tract Lesions." *IEEE Access* 11:66544–66567. https://doi.org/10.1109/access.2023.3290997.

Wahab, H., I. Mehmood, H. Ugail, J. Del Ser, and K. Muhammad. 2024. "Federated Deep Learning for Wireless Capsule Endoscopy Analysis: Enabling Collaboration Across Multiple Data Centers for Robust Learning of Diverse Pathologies." *Future Generation Computer Systems* 152:361–371. https://doi.org/10.1016/j.future.2023.10.007.

Walsh, A. J., A. Ghosh, A. O. Brain, O. Buchel, D. Burger, S. Thomas, L. White, G. S. Collins, S. Keshav, and S. P. Travis. 2014. "Comparing Disease Activity Indices in Ulcerative Colitis." *Journal of Crohn's and Colitis* 8 (4): 318–325. https://doi.org/10.1016/j.crohns.2013.09.010.

Wu, R., K. Qin, Y. Fang, Y. Xu, H. Zhang, W. Li, X. Luo, Z. Han, S. Liu, and Q. Li. 2024. "Application of the Convolution Neural Network in Determining the Depth of Invasion of Gastrointestinal Cancer: A Systematic Review and Meta-Analysis." *Journal of Gastrointestinal Surgery* 28 (4): 538–547. https://doi.org/10.1016/j.gassur.2023.12.029.

Yogapriya, J., V. Chandran, M. G. Sumithra, P. Anitha, P. Jenopaul, and C. Suresh Gnana Dhas. 2021. "Gastrointestinal Tract Disease Classification from Wireless Endoscopy Images Using Pretrained Deep Learning Model." *Computational & Mathematical Methods in Medicine* 2021:5940433. https://doi.org/10.1155/2021/5940433.

Yoshiok, K., K. Tanioka, S. Hiwa, and T. Hiroyasu. 2023. "Deep-Learning Models in Medical Image Analysis: Detection of Esophagitis from the Kvasir Dataset." ArXiv, abs/2301.02390. https://api.semanticscholar.org/CorpusID:255522609.

Zhao, Y., B. Hu, Y. Wang, X. Yin, Y. Jiang, and X. Zhu. 2022. "Identification of Gastric Cancer with Convolutional Neural Networks: A Systematic Review." *Multimedia Tools & Applications* 81 (8): 11717–11736. https://doi.org/10.1007/s11042-022-12258-8.

Zhou, J., N. Hu, Z.-Y. Huang, B. Song, C.-C. Wu, F.-X. Zeng, and M. Wu. 2021. "Application of Artificial Intelligence in Gastrointestinal Disease: A Narrative Review." *Annals of Translational Medicine* 9 (14): 1188. https://doi.org/10.21037/atm-21-3001.

Zhuang, H., J. Zhang, and F. Liao. 2023. "A Systematic Review on Application of Deep Learning in Digestive System Image Processing." *The Visual Computer* 39 (6): 2207–2222. https://doi.org/10.1007/s00371-021-02322-z.