bioengineering

MDPI

Review

Application of Explainable Artificial Intelligence Based on Visual Explanation in Digestive Endoscopy

Xiaohan Cai, Zexin Zhang, Siqi Zhao, Wentian Liu * and Xiaofei Fan *

Department of Gastroenterology, Tianjin Medical University General Hospital, No. 154, Anshan Road, Heping District, Tianjin 300052, China; caixiaohan333@163.com (X.C.); 15504768589@163.com (Z.Z.); 17746161449@163.com (S.Z.)

* Correspondence: lwentian64@163.com (W.L.); ffan@tmu.edu.cn (X.F.)

Abstract

At present, artificial intelligence (AI) has shown significant potential in digestive endoscopy image analysis, serving as a powerful auxiliary tool for the accurate diagnosis and treatment of gastrointestinal diseases. However, mainstream models represented by deep learning are often characterized as complex "black boxes," with decision-making processes that are difficult for humans to interpret. The lack of interpretability undermines physicians' trust in model results and hinders the broader use of models in clinical practice. To address this core challenge, Explainable AI (XAI) has emerged to enhance the transparency of decision-making, thereby establishing a foundation of trust for human–machine collaboration. The review systematically reviews 34 articles (7 articles in esophagogastroduodenoscopy, 13 articles in colonoscopy, 9 articles in endoscopic ultrasonography, and 5 articles in wireless capsule endoscopy), focusing on the research progress and applications of XAI in the field of digestive endoscopic image analysis, with particular emphasis on the visual explanation-based methods. We first clarify the definition and mainstream classification of XAI, then introduce the principles and characteristics of key XAI methods based on visual explanation. Subsequently, we review the applications of these methods in digestive endoscopy image analysis. Lastly, we explore the obstacles presently faced in this domain and the future directions. This study provides a theoretical basis for constructing a trustworthy and transparent AI-assisted digestive endoscopy diagnosis and treatment system and promotes the implementation and application of XAI in clinical practice.

Keywords: explainable artificial intelligence; machine learning; deep learning; visual explanation; digestive endoscopy; medical imaging

check for updates

Academic Editor: Andrea Cataldo

Received: 30 August 2025
Revised: 22 September 2025
Accepted: 25 September 2025
Published: 30 September 2025

Citation: Cai, X.; Zhang, Z.; Zhao, S.; Liu, W.; Fan, X. Application of Explainable Artificial Intelligence Based on Visual Explanation in Digestive Endoscopy. *Bioengineering* **2025**, *12*, 1058. https://doi.org/10.3390/bioengineering12101058

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The continuous development of artificial intelligence (AI) is currently driving significant changes in medicine. It can enhance medical clinical decision-making, reduce medical errors, and improve patient prognosis [1]. Machine Learning (ML) is a subset of AI. Its core lies in constructing a mathematical model through algorithms and optimizing the model's parameters using data, thereby enabling it to make accurate predictions or decisions on unknown data [2,3]. However, traditional ML models often require manual feature extraction, which limits its application in more complex tasks. Deep learning (DL) is a specialized area within ML. Its fundamental structure relies on multi-layer artificial neural networks that mimic neuron activity transmission in the human brain to process information. It can independently learn features from input data and shows significant

advantages in processing complex data, enabling it to handle more challenging tasks, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and so on [4]. Currently, AI is increasingly being used in medical imaging. By training on large amounts of medical image data, it enables the automatic detection and diagnosis of lesions, which helps clinicians make faster and more informed decisions and improves outcomes for patients [5]. In the field of digestive endoscopy, AI-related computer-aided detection (CADe) and computer-aided diagnosis (CADx) have been widely studied. By processing and analyzing vast amounts of digestive endoscopic image data, it is expected to overcome the limitations in endoscopic examinations, thereby improving the quality of clinical diagnosis [6,7]. However, despite the popularity of these AI models, there are still concerns or doubts about their clinical applications. One of the main challenges we face is the inherent "black box" nature of some AI models, especially in complex DL algorithms. DL models contain multiple hidden layers, making it very challenging for humans to comprehend how AI models reach their final conclusions [8].

However, in the field of healthcare, AI-driven medical decisions are closely related to human life and death. Inaccurate or incorrect results can have serious consequences for patients. Therefore, it is essential to provide explanations for AI results to ensure that diagnostic outcomes align with medical expertise, thereby earning the trust of doctors, patients, and regulators. Meanwhile, newer regulations like the European Union's General Data Protection Regulation (GDPR) have made the use of black-box models more challenging in all sectors, including healthcare, as they require traceability of decisions [9,10]. Therefore, while achieving high performance in AI, we believe that research into AI explainability methods is crucial, which has resulted in the rise in explainable AI (XAI) [10]. XAI can enhance the interpretation of the outputs of black-box models and ensure the credibility of the results.

Van der Velden et al. [11] included 223 relevant studies to comprehensively analyze the application of XAI methods in medical image analysis. They discussed and compared the advantages and disadvantages of various XAI methods and put forward their unique insights into the future development of XAI. The results showed that visual explanation is the most commonly used interpretability method in medical image analysis. Salih et al. [12] included 32 articles to conduct a retrospective analysis of the application of XAI in various cardiac imaging methods. Among them, the visual explanation method achieves model interpretability by visualizing relevant features, accounting for approximately half of the article. Similarly, Qian et al. [13] included 56 articles to study the application of XAI in the analysis of MRI images of different human body parts, and identified the evaluation metrics for XAI methods. The results showed that XAI based on visual explanation accounted for more than 60%. The above studies highlight the widespread clinical application of XAI based on visual explanation. However, there are still relatively few reviews on XAI, especially XAI based on visual explanation, in the field of digestive endoscopic image analysis.

Therefore, based on the above background, this review aims to systematically summarize the current research status and application progress of XAI in the field of digestive endoscopy, with a focus on XAI centered on visual explanation. At the same time, this paper will also analyze the clinical application value, existing challenges, and future development directions in this field, so as to provide theoretical basis and practical guidance for promoting the standardized application of XAI in digestive endoscopy diagnosis and treatment. In Section 2, we define XAI and outline its general classification methods. In Section 3, we elaborate on the principles, advantages, and disadvantages of XAI methods based on visual explanation. In Section 4, we review the specific applications of XAI methods based on visual explanation in the analysis of digestive endoscopy images. In

the end, we summarize the benefits, challenges, and future development trends of XAI in clinical practice to promote its broader implementation.

2. Definition and Classification of XAI

2.1. Definition of XAI

Technically speaking, the academic community has not yet formed a unified and clear definition of XAI [10]. The terms "interpretability" and "explainability" are frequently used interchangeably in relevant research, and there are still divergences in the academic community regarding the definition of their concepts [14]. Among them, Rudin believes that AI interpretability refers to the characteristics that the model itself can be directly understood by humans; while AI explainability emphasizes clarifying the working mechanism of the original black box model by constructing an auxiliary explanation model [15]. In this paper, we will use the term "explainable artificial intelligence", which refers to the ability to generate details or reasons to explain how black-box models make decisions, thereby making the results of AI more understandable to humans [16,17].

2.2. Classification of XAI

XAI can be categorized into the following three types based on different classification standards: model-based explanation versus post hoc explanation, model-specific explanation versus model-agnostic explanation, and global explanation versus local explanation (see Figure 1).

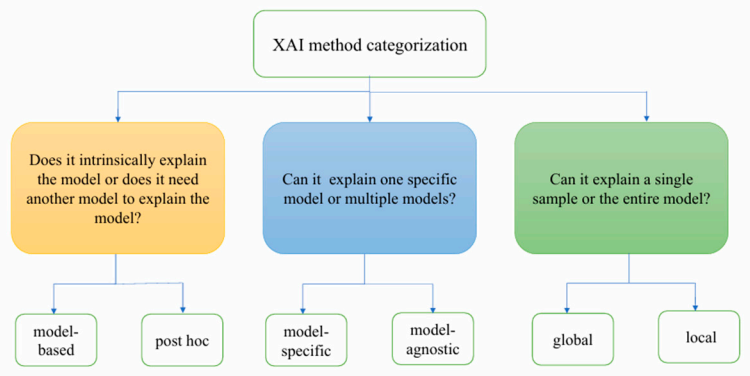


Figure 1. Explainable artificial intelligence method categorization.

2.2.1. Model-Based Explanation vs. Post Hoc Explanation

Model-based explanation involves constructing relatively simple models that allow humans to use their domain knowledge to understand how the model converts inputs into outputs, like traditional linear regression and decision trees. These models not only provide model prediction results but also feature importance scores. Post hoc explanation refers to analyzing a constructed model (such as a CNN model) to gain insights into the model's learning relationships [18]. The difference between it and model-based explanation lies in that this method attempts to train a new model to explain the black-box network; whereas model-based explanation mandates that the model itself is interpretable [11]. For instance, Gradient-weighted Class Activation Mapping (Grad-CAM) is a common post hoc interpretability method applied in DL models.

2.2.2. Model-Specific Explanation vs. Model-Agnostic Explanation

Model-specific explanation is closely related to the internal structure of a particular model and can only be applied to specific models or algorithms, which to a certain extent restricts the freedom of model selection [9]. For example, the Graph Neural Network Explainer (GNN Explainer), as a specialized interpretability tool, provides interpretive

analysis only for the predictions of graph neural network models [19]. In contrast, model-agnostic explanations remain unaffected by which model is selected and do not require probing the internal working mechanisms or parameters of the model. They directly operate on the model's inputs and outputs, observing how modifications to inputs influence outputs to discover certain regions related to the outputs [11].

2.2.3. Global Explanation vs. Local Explanation

Global explanations provide the general learned relationships of a model, identifying features that contribute to classifying a certain target across all instances. They reveal the general decision-making patterns of the model. In contrast, local explanations present the input-output relationships for individual instances, i.e., the features or attributes that are predicted to influence the specific target output [11,20]. For example, SHapley Additive exPlanations (SHAP) analysis can not only provide global explanations to reveal the ranking of important features related to overall prediction results, but also provide important features associated with the result of individual sample to provide local explanations [21].

3. XAI Methods Based on Visual Explanation

Currently, XAI methods applied in medical image analysis are generally classified into three categories: visual explanation, textual explanation, and case-based explanation [11]. Among them, visual explanation is the most prevalent in medical image analysis [4]. It mainly analyzes which parts of the original image have affected the final output result through backpropagation or perturbation methods [9]. Most of them are displayed in the form of heatmaps, where different colors in the heatmap indicate the degree of influence of that part on the prediction result [22], and generally provide post hoc explanation (Figure 2). Textual explanation refers to providing descriptive text to explain predictions, such as visual question answering tasks, image captions, image captions with visual explanation [23]. It can establish a connection between medical images and semantic information. Case-based explanation makes predictions by analyzing examples or data related to the current task, such as Testing with Concept Activation Vectors (TCAV) [24]. Current research indicates that XAI methods based on visual explanation occupy a dominant position within the domain of analyzing medical images. They display features related to predictions on images in an intuitive and comprehensible manner. Meanwhile, the plug-and-play, model-agnostic characteristics and the availability of open-source implementations further facilitate their widespread adoption [4]. Therefore, this paper mainly reviews the application of XAI methods based on visual explanation in medical images, especially in endoscopic images.

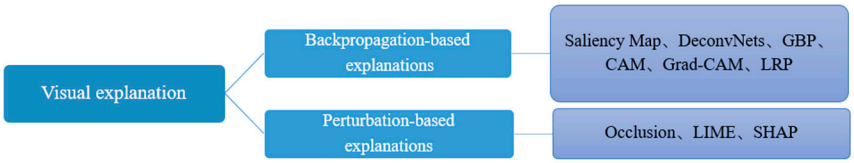


Figure 2. Visual explanation methods categorization.

3.1. Backpropagation-Based Methods

Backpropagation-based methods estimate the impact of gradients, weights, and activations by performing one or more forward propagations in the network and calculating partial derivatives during the backpropagation phase to generate attribution maps [25]. They operate relatively quickly, but the relationship between the results and output changes is weak [9]. The following are several key methods based on backpropagation.

3.1.1. Saliency Map Visualization

Simonyan et al. [26] first proposed a method using backpropagation called saliency map visualization. It calculates gradients through backpropagation and uses the gradients to highlight the correlation between input pixels and prediction results, thereby achieving model visualization [20].

3.1.2. Deconvolution Networks (DeconvNets) and Guided BackPropagation (GBP)

Deconvolution networks (DeconvNets) consist of a series of deconvolution and unpooling layers and generate attribution maps by setting negative gradients to zero during the backward pass to visualize the neural activations of each layer [20,27]. Guided BackPropagation (GBP) is an improvement over the deconvolution method. It visualizes gradients with respect to the image while performing backpropagation through the ReLU activation function. By guiding backpropagation, it suppresses negative gradients from coming back and underscores the pixels that most significantly influence the output, generating a saliency map [13,20]. While both DeconvNets and GBP reveal the fine-grained details in the image, they have limitations: their visualizations cannot distinguish different categories, as the saliency maps for various classes often look very similar [28].

3.1.3. Class Activation Mapping (CAM)

Proposed by Zhou et al. [29], Class Activation Mapping (CAM) is a method designed to interpret CNN models. First, it implements a global average pooling operation in the terminal convolutional layer of the network to replace the fully connected structure, and then projects the output layer's weights back to the feature mapping space, thereby generating a heat map to realize the identification and visualization of important regions in the image.

3.1.4. Gradient-Weighted Class Activation Mapping (Grad-CAM)

Gradient-Weighted Class Activation Mapping (Grad-CAM) is an extension and generalization of CAM. The difference between it and CAM is that it is applicable to various types of CNNs, while CAM can only be used in CNNs with global average pooling [13]. Meanwhile, Selvaraju et al. [28] also proposed Guided Grad-CAM, which combines Grad-CAM with guided backpropagation. This hybrid approach generates high-resolution, fine-grained visualization maps that not only locate the relevant regions but also provide detailed insight into the features influencing the model's decision, uniting the precise localization capability of Grad-CAM with the high-resolution attributes of guided backpropagation.

3.1.5. Layer-Wise Relevance Propagation (LRP)

Bach et al. [30] first proposed layer-wise relevance propagation (LRP) to realize the pixel-wise explanation of the decisions made by nonlinear classifiers. LRP attributes relevance scores to individual input features to explain the predictions of neural networks. This method starts from the output layer of the network and traces back layer by layer to the input layer through backpropagation. During the iteration process of each layer, the algorithm assigns corresponding contribution scores to each neuron in the previous layer and strictly follows the principle of relevance conservation to ensure that the overall correlation strength remains consistent during the transmission process [30,31]. This process is used to show the impact of features on the prediction results.

3.2. Perturbation-Based Methods

Perturbation-based methods can be executed by altering, eliminating, or concealing specific input features and measuring the differences from the original output. Among

them, the features that have the greatest impact on the output results are considered the most important [9]. Since this process does not require access to the internal structure of the model, this method is classified as a model-agnostic explanation. However, since this method needs to make predictions on multiple inputs and outputs, it generally takes longer than backpropagation-based methods [32].

3.2.1. Occlusion

The occlusion predicts the impact of relevant regions on output results by altering input images (features) [4]. Among them, the occluded parts that have a greater impact on the output results are considered to be of high importance, while those have a smaller impact on the output results are assigned low importance [11,33].

3.2.2. Local Interpretable Model-Agnostic Explanations (LIME)

Ribeiro et al. [34] introduced local interpretable model-agnostic explanations (LIME), a method that gives explanations through training a relatively simple model to simulate a complex model, thereby learning the relationship between perturbed input data and output changes. For example, it uses a linear model to approximate and replace a CNN. It utilizes the resemblance between the modified input and the initial input as weights to make sure that the explanations given by the simple model with heavily modified inputs are more trustworthy [11,17]. In image analysis, LIME can highlight the image regions that make the most significant contributions to specific class decisions, thereby achieving local interpretation of the model.

3.2.3. SHapley Additive exPlanations (SHAP)

Lundberg and Lee [35] introduced the concept of SHapley Additive exPlanations (SHAP), which utilizes the principle of Shapley value in cooperative game theory to provide a basis for the prediction results of ML models. This method estimates the marginal contribution of each feature to the model's prediction results by sampling different feature combinations multiple times and evaluating the changes in their outputs [9,11]. Meanwhile, after pixelating the Shapley values, different colors can be used to indicate the active or passive impacts of different characteristics on the model's decision-making. This method can intuitively display the importance ranking of various features, helping to interpret the model's decision-making process comprehensively [11,36].

4. Applications of XAI Methods Based on Visual Explanation in Digestive Endoscopy

We searched articles published in the PubMed database from 2015 to 2025 to identify the current applications of XAI based on visual explanation in the analysis of digestive endoscopic images. The search terms are as follows: (explainable artificial intelligence" OR "interpretable artificial intelligence" OR "artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "convolutional neural network") AND (endoscopy). The inclusion criteria are as follows: (1) The XAI models are applied to digestive endoscopic image analysis; (2) The article adopts specific visual explanation methods to explore the interpretability of AI models; (3) Complete article information is accessible. We also included key articles identified through manual searching or reference lists. Then, the above articles were manually screened and evaluated again, and a total of 34 eligible articles were finally included.

4.1. Applications in Esophagogastroduodenoscopy

Esophagogastroduodenoscopy (EGD) is an important examination method for diagnosing upper gastrointestinal diseases, especially for esophageal and gastric diseases. It

allows direct visualization of the mucosal surface, facilitating the precise diagnosis and assessment of illnesses. Among them, white light endoscopy (WLE) is the most widely used examination method, mainly used to observe and evaluate the gross morphological appearance of lesions. Compared with WLE, chromoendoscopy (CE) and narrow-band imaging (NBI) can make the scope and outline of lesions clearer, thereby improving the ability to identify lesions. Magnifying endoscopy (ME) combines the functions of endoscopy and microscopic imaging, enabling magnified observation of the microstructures of the mucosal surface such as glandular openings and microvascular, thus demonstrating unique diagnostic advantages in the identification of early-stage cancers. It has unique diagnostic value in detecting early-stage cancers. Clinically, endoscopists often combine ME with CE or NBI; upon suspicious findings in WLE, they first use staining or NBI mode to highlight the outline of the lesion tissue and then switch to the magnifying mode to observe the fine structure of the local mucosa. This strategy can improve the detection rate of lesions and the accuracy of biopsy [37]. In addition, with technological advancements, more and more AI and XAI are integrated into EGD to support clinical decision-making. These systems can highlight areas of concern, assist in classifying lesions, and provide visual explanations to endoscopists, enhancing diagnostic reliability and precision [38]. Table 1 summarizes XAI applications and visual explanation methods used in EGD.

Table 1. A Summary of Representative Studies on XAI in Esophagogastroduodenoscopy examination.

Types of Digestive Endoscopy	Author	Year	Aim	Main Visual XAI Method
Esophagogastroduodenoscopy	Ge et al. [39]	2023	Diagnosis of Los Angeles classification of GERD	Grad-CAM
	De Souza et al. [40]	2021	Identification of BE and esophageal adenocarcinoma	Saliency map, GBP, etc.
	García-Peraza-Herrera et al. [41]	2020	Classification of IPCLs	CAM
	Everson et al. [42]	2021	Classification of IPCLs	CAM
	Ueyama et al. [43]	2021	Diagnosis of EGC	Grad-CAM
	Hu et al. [44]	2021	Diagnosis of EGC	Grad-CAM
	Cho et al. [45]	2020	Prediction of submucosal invasion for gastric neoplasms	CAM

BE, Barrett's esophagus; CAM, Class Activation Mapping; EGC, early gastric cancer; GBP, Guided Back Propagation; GERD, gastroesophageal reflux disease; Grad-CAM, Gradient-Weighted Class Activation Mapping; IPCLs, intraepithelial capillary loops; XAI, explainable artificial intelligence.

Various XAI methods are used in EGD image analysis to assist in clinical disease diagnosis. Regarding the identification of esophageal-related diseases, gastroesophageal reflux disease (GERD) is a common digestive system disease disorder. It can be diagnosed through endoscopic examination, which reveals characteristic mucosal damage, or through reflux testing that detects abnormal esophageal acid exposure [46]. Ge et al. [39] utilized 2081 endoscopic images to establish a deep learning model based on DenseNet-121 for identifying the Los Angeles classification (LA-grade) of GERD. The model achieved an area under the curve (AUC) of 0.968, and its classification accuracy (86.7%) was significantly higher than that of junior endoscopists (71.5%) and senior endoscopists (77.4%). Meanwhile, heatmaps were generated based on Grad-CAM to solve the black-box problem of the model. Barrett's esophagus (BE) is a disease in which the mucosal cells of the lower esophagus undergo significant changes, which can progress to esophageal adenocarcinoma in severe cases. De Souza et al. [40] constructed various CNN models for the identification of BE and esophageal adenocarcinoma, and adopted multiple XAI tools such as saliency map and GBP to explain the model decisions. The results showed that the regions highlighted by the saliency map had a high consistency with human segmentation, achieving

the best explanation results. This study not only utilized multiple XAI methods but also evaluated different XAI methods by comparing XAI outputs with expert annotations. The histological staging of early squamous cell neoplasia (ESCN) can be predicted by observing the morphological characteristics of its specific microstructure, namely the intraepithelial capillary loops (IPCLs), which are regarded as endoscopic markers of ESCN [47]. Currently, several studies have constructed XAI models to improve the identification of IPCLs. For example, García-Peraza-Herrera et al. [41] used 67,742 video frames from 114 patients to build a CNN for binary classification of IPCLs, achieving an average accuracy of 91.7%. CAM maps were employed to highlight the parts contributing to the results. However, the above-mentioned model can only identify static video frames and cannot run in real-time, which limits their application in clinical practice. Subsequently, Everson et al. [42] used 67,742 high-quality magnifying endoscopy with narrow band imaging (ME-NBI) images from 115 patients to train a CNN to classify IPCLs and predict ESCN histological staging. With an average diagnostic accuracy of 91.7%, the model's performance closely approximated the comprehensive diagnostic level (94.7%) of endoscopic experts. Crucially, this CNN operated at video rate, possessing the ability of real-time prediction, thus overcoming the temporal limitations of previous static-frame models. Additionally, CAM was used to highlight the features affecting classification prediction outcomes.

In terms of assisting in the identification of stomach-related diseases, multiple studies have constructed XAI models based on CAM to achieve diagnosis of early gastric cancer (EGC) and prediction of gastric cancer invasion depth. Ueyama et al. [43] developed a CAD system using 5574 ME-NBI images to identify EGC and achieved high diagnostic accuracy. The comprehensive performance evaluation metrics of this CNN model yielded excellent results: the area under the curve (AUC) reached 99%, overall accuracy was 98.7%, sensitivity achieved 98%, specificity attained 100%. Finally, Grad-CAM was used to visualize the image regions affecting the classification results, and these regions were consistent with those identified by endoscopists, providing an interpretable analysis of the model. However, although the model achieved excellent results, the study was a single-center study and lacks external validation, limiting the generalizability of the findings. Hu et al. [44] constructed an AI model (EGCM) for identifying EGC using 1777 ME-NBI images from 3 centers and conducted a human–machine comparison. The results showed that the AUC of the model in both internal and external validation sets was as high as approximately 80%. Meanwhile, the diagnostic performance of the model (with an accuracy of 0.77) was similar to that of senior experts (with an accuracy of 0.755) and better than that of junior experts (with an accuracy of 0.728). The diagnostic performance of all doctors was improved with the assistance of EGCM, and Grad-CAM highlighted the abnormal areas of the lesions. Accurate assessment of the invasion depth of gastric cancer is critical for selecting the optimal treatment strategy. Cho et al. [45] trained two CNN models using WLE images of gastric cancer patients to classify gastric tumors as either confined to the mucosal layer or demonstrating submucosal invasion, and validated them in external datasets. Among them, the DenseNet-161 model performed better, with an AUC as high as 0.887 in both internal and external validation sets. CAM effectively displayed the characteristic regions of the tumor to explain the model.

4.2. Applications in Colonoscopy

Colorectal cancer (CRC) is the fourth leading cause of death among malignant tumors worldwide [48]. The 5-year survival rate can reach 91% in the early stage, but drops to approximately 14% in the advanced stage [49]. Most of them follow the adenoma-carcinoma progression sequence [50]. Therefore, timely detection and treatment of polyps and/or adenomas through colonoscopy can significantly decrease the incidence rate of colorectal

malignant tumors. However, the accuracy of polyps and/or adenomas detection is affected by factors such as manual fatigue and physician experience. Studies have shown that the missed diagnosis rate of adenomas in serial colonoscopy is 26% [51,52]. Meanwhile, the removal of polyps and/or adenomas must take into account the increased medical costs, including those related to pathological examinations [53]. Therefore, comprehensive identification of lesions and accurate optical diagnosis can significantly reduce costs. In this context, XAI can be used to assist physicians in performing colonoscopies, improving the quality of detection while providing physicians with diagnostic insights. Table 2 summarizes XAI applications and visual explanation methods used in colonoscopy.

Table 2. A Summary of Representative Studies on XAI in Colonoscopy examination.

Types of Digestive Endoscopy	Author	Year	Aim	Main Visual XAI Method
Colonoscopy	Chen et al. [54]	2024	Evaluation of key quality indicators for colonoscopy	Grad-CAM, Guided Grad-CAM, SHAP
	Wickstrom et al. [55]	2020	Polyp recognition	GBP
	Jin et al. [56]	2020	Polyp classification	Grad-CAM
	Zhang et al. [57]	2024	Polyp classification	Grad-CAM
	Choi et al. [58]	2020	Polyp classification	CAM
	Chierici et al. [59]	2022	Identification of CD and UC	GBP
	Sutton et al. [60]	2022	Diagnosis and grading of UC	Grad-CAM
	Weng et al. [61]	2022	Differentiation of CD and ITB	SHAP
	Gabralla et al. [62]	2023	Identification of colon cancer	Grad-CAM
	Binzagr F et al. [63]	2024	Classification of gastrointestinal cancer	SHAP
	Auzine MM et al. [64]	2024	Classification of gastrointestinal cancer	SHAP, LIME
	Dahan et al. [65]	2025	Detection of gastrointestinal disease	Grad-CAM
	Gideon et al. [66]	2025	Detection of gastrointestinal disease	Grad-CAM, SHAP

CAM, Class Activation Mapping; CD, Crohn's disease; GBP, Guided Back Propagation; Grad-CAM, Gradient-Weighted Class Activation Mapping; ITB, intestinal tuberculosis; LIME, Local Interpretable Model-agnostic Explanations; SHAP, SHapley Additive exPlanations; UC, ulcerative colitis; XAI, explainable artificial intelligence.

In terms of assisting in the identification of colorectal polyps, various different XAI models have been developed so far and have achieved excellent results. Chen et al. [54] collected 4189 colonoscopic images containing polyps, cecum, and different levels of bowel cleanliness to train models based on CNN and Transformer. These models are used to intelligently evaluate key quality indicators of colonoscopy. The results show that the EfficientNetB2 model exhibited excellent performance on both the validation set and the test set. Grad-CAM, Guided Grad-CAM, and SHAP methods revealed the regions that have an impact on the prediction results. Wickstrom et al. [55] used three network architectures to realize the semantic segmentation of colorectal polyps and achieve polyp identification at the pixel level. Meanwhile, GBP was used to highlight the important features for polyp prediction, indicating that the model leverages the contours and morphological characteristics of polyps for its predictive analysis. However, the aforementioned models only identify polyps and do not distinguish between different types of polyps. Since hyperplastic polyps rarely undergo malignant transformation, unnecessary endoscopic resection can increase medical costs without added benefit. Therefore, accurate diagnosis of polyp types is important to avoid inappropriate resection. In this context, Jin et al. [56] collected images of 1100 small adenomatous polyps and 1050 small hyperplastic polyps under NBI to train a CNN model for further distinguishing whether colorectal polyps are adenomatous or hyperplastic. The results of the study showed that the accuracy of the model in differentiation was 86.7% (95%CI: 82.3–90.3). Moreover, with the help of AI, the overall diagnostic accuracy of endoscopists was significantly improved (from 82.5% to 88.5%, $p < 0.05$), and the overall diagnostic time was significantly shortened (from 3.92 s to 3.37 s, $p < 0.05$). Meanwhile this study generated heatmaps using Grad-

CAM methods and visualized the highlights that overlaid the polyp, which can aid the endoscopist to accept AI insights. However, this model was only validated in a single center, so its generalization ability in clinical practice may be unsatisfactory. Although the risk of small polyps progressing to CRC is very low, those with advanced features still pose a high risk [67]. Therefore, to further identify colorectal polyps with advanced features, Zhang et al. [57] constructed a CNN model using NBI images to classify colorectal polyps as either with or without advanced features. The model output included NBI images with Grad-CAM heatmaps to highlight the interpretability of the model. The accuracy of the model in the internal and external validation sets was 0.880 (0.839–0.916) and 0.870 (0.843–0.896), respectively, with AUC values of 0.942 (0.847–0.961) and 0.926 (0.846–0.946), respectively, indicating good robustness. Compared with 19 endoscopists, the AI model demonstrated a greater diagnostic performance ($p < 0.05$). Furthermore, in the prospective test, compared with the non-assisted group, the endoscopists in the AI-assisted group identified more polyps with advanced features and showed higher accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) (all $p < 0.001$). The study achieved excellent results in the internal and external validation sets as well as the prospective test set, demonstrating the model's strong generalization ability and making the model results more credible. Meanwhile, human–machine comparison further illustrated the model's auxiliary role for clinicians. However, both studies focus on binary classification of polyps and do not involve multiple classifications of colorectal polyps. Addressing this, Choi et al. [58] used 3000 endoscopic images of colorectal adenomas to construct three CNN models based on Inception-v3, ResNet-50, and DenseNet-161, respectively, and compared the performance of the CNN models with endoscopists with different years of experience. The models aimed to classify images into multiple pathological categories: normal, tubular adenoma with low-grade dysplasia (TALGD), tubular adenoma with high-grade dysplasia (TAHGD), and carcinoma (CA). The results showed that the three CNN models outperformed the endoscopy expert group, and the model based on DenseNet-161 achieved the best performance. CAM was used to highlight the relevant areas of the images, enhancing model interpretability. Although these models have achieved multi-classification of the pathological types of colorectal adenomas, the models' accuracy in identifying TAHGD and CA lesions is lower than that in identifying normal and TALGD lesions. Future improvements can be made to enhance its performance.

At present, several studies have utilized heatmap-based methods to construct XAI for assisting in the identification of inflammatory bowel disease (IBD). For example, Chierici et al. [59] constructed a DL model to identify ulcerative colitis (UC) and Crohn's disease (CD). GBP showed that typical endoscopic features of IBD, such as mucosal erythema, appear to have higher attribution values, achieving direct visual interpretation. Sutton et al. [60] utilized images from the HyperKvasir dataset to construct an interpretable CNN model for the diagnosis and grading of UC. The Grad-CAM displayed the image regions used for predicting the output such as fibrin covering ulcers, in the form of a heat map, which was consistent with the pathology of UC. However, results in the form of heatmaps can only provide local explanations and do not provide a comprehensive understanding of the entire model's decision process. To address this limitation, Weng et al. [61] employed SHAP method to construct an interpretable XGBoost model integrating endoscopic features to distinguish CD from intestinal tuberculosis (ITB), providing both local and global explanations for the model to identify important features.

In addition to the above models, there are currently various ensemble models built based on datasets for identifying gastrointestinal lesions. By constructing ensemble models, the advantages of multiple models are utilized to improve diagnostic performance.

Gabralla et al. [62] integrated the output of the pretrained CNN models with a meta-learner (Support Vector Machine, SVM) to construct a stacking model (Stacking-SVM) for colon cancer identification. The results showed that compared with the CNN models, the Stacking-SVM model achieved the highest performance in two different datasets. Additionally, Grad-CAM generated heatmaps to visualize the impact of different regions on the prediction results. Binzagr F et al. [63] developed an ensemble model comprising 3 CNN models-InceptionV3, InceptionResNetV2, and VGG16-based on the KvasirV2 dataset to identify polyps, UC, and esophagitis. The hybrid architecture yielded a classification accuracy of 93.17%, with its F1 score reaching 97%. SHAP was used to provide explanations for the model's predictions to doctors. Similarly, Auzine et al. [64] also constructed an ensemble model for identifying polyps, UC, and esophagitis based on the aforementioned CNN models. The accuracy of this ensemble model reaches 96.89%. SHAP and LIME methods was used to enhance the comprehensive understanding of the model's decision-making. Unlike the aforementioned methods for constructing ensemble models, Dahan et al. [65] integrated the deep convolutional neural network model with Swin Transformer to build a hybrid model. This hybrid model is designed to extract more comprehensive endoscopic image features, thereby achieving better identification of gastrointestinal diseases. The accuracy of this hybrid model reaches 93.43%. Grad-CAM highlights the regions that have the most significant impact on the model's results. Similarly, Gideon et al. [66] utilized the Kvasir and GastroNet datasets to develop a deep learning model incorporating CNNs, RNNs, and the transformers to detect gastroenterological diseases while providing explainability through Grad-CAM and SHAP. The Ensemble model outperforms the individual model showing an accuracy of 92.6%. Grad-CAM heatmaps are used to show which parts of medical images are most relevant to the model predictions. SHAP analysis finds the most important features driving the model's predictions are texture and color. These interpretability methods help doctors interpret model decisions and build trust in their predictions. In conclusion, the accuracy rates of the above five ensemble models all exceed 90%, suggesting that ensemble models have great development potential in disease diagnosis.

4.3. Applications in Endoscopic Ultrasonography

Endoscopic ultrasonography (EUS) integrates ultrasonic imaging and endoscopic visualization functions, providing high-quality ultrasonic images for organs such as the gastrointestinal tract and pancreas. This examination enables high-resolution real-time imaging of the digestive tract's mucosal architecture, significantly enhancing the detection of characteristics and ranges of lesions [68]. Therefore, it has important application value in the diagnosis of pancreatic diseases, evaluation of submucosal lesions, and determination of the depth of tumor invasion. Compared to conventional ultrasound (US), computed tomography (CT), and magnetic resonance imaging (MRI), EUS may detect smaller lesions [69]. Currently, EUS serves as a vital tool in the diagnosis of diverse gastrointestinal disorders, improving the disease detection rates [70]. However, the diagnostic accuracy of EUS largely depends on the professional knowledge, practical experience and technical level of the operating physicians. The considerable training costs and prolonged learning curve make it hard to master EUS diagnostic skills [71]. Therefore, combining EUS with XAI may help physicians improve the accuracy of clinical diagnosis while enhancing detection efficiency. Table 3 summarizes XAI applications and visual explanation methods used in EUS.

Table 3. A Summary of Representative Studies on XAI in Endoscopic Ultrasonography examination.

Types of Digestive Endoscopy	Author	Year	Aim	Main Visual XAI Method
Endoscopic Ultrasonography	Gu et al. [72]	2023	Diagnosis of PDAC	Grad-CAM
	Yi et al. [73]	2025	Identification of PDAC and PNETs	Grad-CAM, SHAP
	Marya et al. [74]	2021	Diagnosis of pancreatic diseases	Occlusion
	Mo et al. [75]	2025	Prediction of PNETs pathological grading	SHAP
	Cai et al. [76]	2025	Diagnosis of pancreatic diseases	SHAP
	Cui et al. [77]	2024	Diagnosis of pancreatic diseases	Grad-CAM, SHAP
	Liang et al. [78]	2025	Preoperative risk prediction of GISTs	SHAP
	Liu et al. [79]	2022	Identification of the lesion invasion depth and lesion source of esophageal submucosal tumors	CAM
	Uema et al. [80]	2024	Identification of the invasion depth of EGC	CAM

CAM, Class Activation Mapping; EGC, early gastric cancer; GISTs, gastrointestinal stromal tumors; Grad-CAM, Gradient-Weighted Class Activation Mapping; PDAC, pancreatic ductal adenocarcinoma; PNETs, pancreatic neuro endocrine tumors; SHAP, SHapley Additive exPlanations; XAI, explainable artificial intelligence.

Explainable methods based on Grad-CAM and SHAP are widely used in constructing XAI models to help diagnose diseases related to the pancreas. Gu et al. [72] developed a deep learning radiomics model based on EUS images to effectively identify pancreatic ductal adenocarcinoma (PDAC), and its performance is superior to most clinical experts. Meanwhile, the heatmaps generated by Grad-CAM show that the low to mixed echo areas within the tumor and the tumor boundary areas are of great value to the model's diagnosis, which helps to understand the diagnostic results. Yi et al. [73] constructed explainable ML models based on DL features extracted from EUS images to distinguish pancreatic neuroendocrine tumors (PNETs) from pancreatic cancer, with Grad-CAM and SHAP clarifying and visualizing the model outputs. However, the above models are all based on static image recognition and have not been validated on dynamic video sets, limiting their further clinical application. Marya et al. [74] addressed this limitation by collecting EUS static images and video information to train a CNN model for distinguishing autoimmune pancreatitis (AIP) from PDAC, chronic pancreatitis (CP), and normal pancreas (NP). This model achieved excellent results in both image datasets and video datasets. Meanwhile, by occluding different pixels, heatmaps are generated to assess the features of the model in identifying AIP and PDAC. The heatmap analysis shows that patients with AIP have significantly more high-scoring subregions in the pancreas than those with PDAC, while PDAC patients have more high-score subregions in the retroperitoneum. Nonetheless, the aforementioned studies mainly focus on the lesion area and ignore the surrounding area of the lesion, which may contain valuable diagnostic information. Therefore, Mo et al. [75] constructed a model based on multilayer perception (MLP) integrating radiomics features from intratumoral and peritumoral regions to predict the pathological grade of PNETs. This study provides insights into the value of peritumoral regions, especially the tumor-adjacent parenchyma, in disease diagnosis. Meanwhile, it uses SHAP values to visualize the importance of features, thereby providing interpretability for the model's predictions. However, the aforementioned models only utilize single-modal radiomics information and ignore the potential diagnostic impact of other aspects, such as laboratory test results. Integrating multi-modal information may improve the accuracy and robustness of the model's diagnosis. For instance, Cai et al. [76] constructed a multimodal ML model that combined radiomics features from EUS images of pancreatic lesions and clinical characteristics to identify pancreatic lesions. The findings indicated that the multimodal model performed better. Meanwhile, SHAP method was used to achieve interpretability analysis at both the overall and individual levels (Figure 3). However, the model lacks multi-center external validation and may have overfitting issues. Therefore, Cui and

colleagues [77] constructed a multimodal AI model using EUS images and clinical date of patients with pancreatic lesions from multiple centers to distinguish pancreatic cancer from non-cancerous lesions, which effectively solved this problem and greatly improved the credibility of the model's performance. Meanwhile, this study conducted interpretability analysis using Grad-CAM and SHAP methods and demonstrated that such interpretability analysis improved physicians' acceptance of AI prediction results.

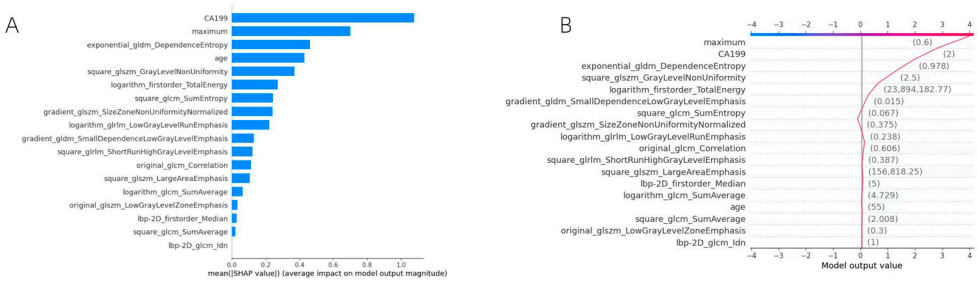


Figure 3. Global explanation and local explanation provided by SHAP. (**A**) The SHAP bar chart is sorted by the mean absolute SHAP values of features to provide global explanation. (**B**) The SHAP decision plot calculates the magnitude and direction of each feature's contribution to the prediction result in an individual sample to achieve local explanation.

Since EUS can identify the morphological characteristics of gastrointestinal tumors and the layered architecture of the gastrointestinal wall, it is also applied in evaluating subepithelial lesions and the depth of tumor invasion. Liang et al. [78] collected various clinical and pathological data from patients with gastrointestinal stromal tumors (GISTs), including high-risk features on EUS, to construct a model for predicting the malignant potential of gastric GISTs. SHAP indicated that high-risk features on EUS, tumor size, tumor boundaries, and monocyte-to-lymphocyte ratio are key variables affecting the model results. In terms of evaluating the depth of tumor invasion, Liu et al. [79] constructed a DL model to identify the lesion invasion depth and lesion source of esophageal submucosal tumors. The overall accuracy of the model can reach 82.49%, and CAM identifies the lesion area in the image to improve the interpretability of the model. However, it remains a single-center study with a single source of training data. Subsequently, Uema et al. [80] developed an AI-based EUS system aimed at diagnosing the invasion depth of EGC and collected data from 10 institutions for adequate external validation. The results showed that the AUC of the model in the internal and external validation sets was 87% and 81.5%, respectively, with accuracies of 82.2% and 74.1%, respectively. Notably, in the external validation set, the diagnostic performance of the model was comparable to experts. CAM visualized the regions of interest of the AI model.

4.4. Applications in Wireless Capsule Endoscopy

Wireless capsule endoscopy (WCE) is a non-invasive diagnostic technique used for gastrointestinal examinations. After the examinee swallows the capsule, the device can move along with gastrointestinal peristalsis and continuously collect images of the digestive tract mucosa during this process, so as to realize the exploration of the interior of the gastrointestinal tract. It has been proven to be of great value in evaluating focal lesions of the digestive tract, such as gastrointestinal bleeding and ulcer identification [81]. However, a single WCE examination generates thousands of video images, which require physicians to spend a considerable amount of time interpreting them, significantly increasing the diagnostic burden. Therefore, establishing XAI models for automated image analysis can greatly save time and energy and improve the diagnostic speed. Table 4 summarizes XAI applications and visual explanation methods used in WCE.

Table 4. A Summary of Representative Studies on XAI in Wireless Capsule Endoscopy examination.

Types of Digestive Endoscopy	Author	Year	Aim	Main Visual XAI Method
Wireless Capsule Endoscopy	Malhi et al. [82]	2019	Detection of gastrointestinal bleeding	LIME
	Wang et al. [83]	2019	Recognition of peptic ulcer	CAM
	Mukhtorov et al. [84]	2023	Detection of gastrointestinal disease	Grad-CAM
	Mascarenhas et al. [85]	2022	Diagnosis of colonic protruding lesions	Grad-CAM
	Nadimi et al. [86]	2020	Detection of colorectal polyps	Saliency map

CAM, Class Activation Mapping; Grad-CAM, Gradient-Weighted Class Activation Mapping; LIME, Local Interpretable Model-agnostic Explanations; XAI, explainable artificial intelligence.

In the domain of gastrointestinal bleeding detection, Malhi et al. [82] constructed a CNN model using 3895 capsule endoscopy images to identify gastrointestinal bleeding. In the validation set, the model's accuracy reached 97.92%. Meanwhile, to explain the model's prediction results, the article selected the LIME method, which has highly reliable results, relatively low computational cost, and model-agnostic nature, to mark the boundaries of gastric bleeding areas [87]. In terms of identifying gastrointestinal ulcers, Wang et al. [83] constructed an ulcer recognition network HAnet-34 (480) with a hyperconnection architecture (HAnet) by using 1157 ulcer videos and 259 normal videos from WCE. The model overall test accuracy is 92.05%. Additionally, they used CAM to display relatively important parts in the images and provide target location information to visualize the model results. For other digestive system diseases, XAI in the form of heatmaps has also been extensively studied. For example, Mukhtorov et al. [84] trained a ResNet152-based CNN model by using an open-source database containing 8000 capsule endoscopy images to identify gastrointestinal diseases. They used the Grad-CAM method to display the image regions that contributed the most to a given classification decision, thereby highlighting the model's interpretability. On the validation set, the model's accuracy was as high as 93.46%. Mascarenhas et al. [85] constructed a CNN-based model utilizing colon capsule endoscopy images to automatically detect protruding lesions in the colonic lumen. The model exhibits a high level of performance and excellent image processing capabilities, with both outstanding recognition accuracy and processing speed. The Grad-CAM method generates heatmaps to emphasize regions of diagnostic significance relevant to the prediction of lesions. However, this model was limited to identifying protruding lesions without differentiating their specific types. Furthermore, Nadimi et al. [86] constructed a CNN model trained on 11,300 data-augmented wireless colon capsule endoscopy images to automatically identify colorectal polyps. The model achieved an accuracy of 98%, a sensitivity of 98.1%, and a specificity of 96.3%, which outperformed all the latest results. Additionally, activation maps were used to display the parts that contributed significantly to the results.

5. Discussion

Currently, XAI is developing rapidly in the field of digestive endoscopy, with more and more XAI methods emerging. Among them, visual explanation methods employed in medical image analysis are widely applied owing to their plug-and-play characteristics and easily accessible open-source implementations. Therefore, this review outlines the latest progress of XAI based on visual explanation in the field of digestive endoscopy image analysis. Notably, approximately 39% of studies used Grad-CAM, about 22.5% used CAM, and CAM-based cases constituted the majority of the investigations analyzed. A potential reason for this prevalence could be the excellent capabilities of Grad-CAM and CAM in interpreting image analysis models, such as high reliability, high efficiency, and intuitive

interpretability. This section will systematically explore the core value of current progress, existing limitations, and future development directions.

Black-box models used to be the primary obstacle hindering the clinical implementation of AI. For clinicians, XAI methods based on visual explanations such as Grad-CAM and LIME visualize lesion-focused regions (such as the morphology of gastrointestinal glandular duct openings, abnormal microvessels, and the abnormal echogenic area of pancreatic lesions, etc.), enabling doctors to intuitively verify the decision-making logic of AI. This visualization has significantly enhanced physicians' trust and acceptance of AI-assisted diagnostic results, which is crucial in high-risk medical fields. Furthermore, XAI methodologies not only bolster clinical confidence but also serve as educational tools. Physicians, particularly junior physicians, may learn from the model's decision-making process, which can improve diagnostic accuracy to a certain extent and help discover new imaging markers. Finally, XAI may help physicians identify areas where the model may have focused incorrectly or failed to focus sufficiently, providing insights to refine and improve model performance [88].

However, in the investigation of XAI applications, this study also found some limitations and challenges. First of all, many studies utilize datasets that are relatively small, lack external validation cohorts, or exhibit unbalanced sample distributions among different disease types, so that the quality and quantity of the data are not guaranteed. In ML, a model's performance is strongly influenced by both the adequacy and representativeness of its training dataset. Poor data quality and insufficient data can affect both the predictive accuracy and the interpretability of the models. Secondly, the types of data applied in most articles are relatively limited in diversity, which may restrict the diagnostic performance of the models. Thirdly, in this survey, most of the XAI articles based on visual explanations provide interpretability by emphasizing regions that critically influence the prediction outcomes, which offers a certain degree of acceptance and credibility for the model results. However, most articles only use XAI methods and lack evaluations of these XAI methods. We cannot determine whether the explanations generated by the models are correct, which makes it necessary to be cautious when interpreting high-risk decisions. As Rudin et al. [15] proposed, the explanations provided by some interpretive models may not accurately represent the original models, and some XAI methods may offer meaningless or insufficient details in interpreting the models. For example, the saliency map is regarded as an interpretability tool, which can highlight the pixels that have the greatest impact on the output and weaken the unnecessary parts. It can tell us where the model is focusing, but we do not understand why the highlighted parts are related to the final result, and the diagnostic basis of the model cannot be fully comprehended. Meanwhile, because saliency maps tend to highlight edges, they may provide similar explanations for each analogy. The reliability of explainable methods remains debatable. Finally, the involvements of medical experts in the design and evaluation of XAI frameworks are limited. Many XAI articles focus on the development of ML models and the deployment of XAI methods, lacking the participation of medical experts. This disconnect may result in XAI applications failing to meet the actual needs of clinicians.

Regarding the future development of XAI, we believe that it will play a crucial role in the auxiliary identification and diagnosis of diseases, as it serves to elucidate the decision-making processes of AI, thereby fostering greater confidence among medical practitioners. To tackle current limitations, we consider that, first of all, in terms of data collection and processing, we should collect multi-center datasets as much as possible, strengthen the utilization of data from public databases, and reasonably employ technologies such as data augmentation and transfer learning to cope with the challenges posed by small datasets. Meanwhile, methods such as denoising and reasonable annotation should be

adopted to improve the quality of image data. Secondly, we can integrate multimodal data, such as different types of image data or combining image and non-image data, to build models. This will not only further improve the diagnostic performance of the models but also enhance the clinical acceptance of AI technologies. Thirdly, in the process of data collection and processing, the data quality can be publicly reported to help identify potential inaccuracies or biases, and relevant standards should be formulated to protect user information so as to ensure data security [89,90]. Fourthly, attention should be paid to program resource constraints when developing XAI algorithms. For instance, real-time reporting systems require strong computing and processing capabilities to respond quickly. Therefore, challenges such as high computing demand should be considered during the model development process, and resource-intensive explanation methods should be used cautiously [89]. Fifth, regarding the evaluation methods of XAI, there is currently no universally accepted evaluation standard. Doshi-Velez and Kim [91] once proposed a method for evaluating interpretability, which conducts the evaluation from three levels: application-based evaluation, human-based evaluation, and function-based evaluation. However, the evaluation of XAI methods remains a relatively young research field at present. It is still important to involve experts in the evaluation of XAI results. Developing standards for assessing the interpretability of different models will facilitate the application and development of clinical XAI. Meanwhile, provided that diagnostic efficacy remains comparable, more inherently interpretable models can be constructed in the future, as they seem to be more reasonable when dealing with high-risk decisions. For example, Dong et al. [92] developed a high-performance explainable AI system for diagnosing early gastric tumors based on feature extraction and multi-feature fitting. The result interface intuitively displays 6 characteristics of lesions and final diagnosis results, which significantly improves the transparency of the model. Sixth, incorporating multiple interpretability methods, such as textual explanations and case-based explanations, may yield more comprehensive and accurate explanations in future research. Finally, healthcare professionals ought to play an integral role in the conception and engineering of XAI systems to realize human–machine interaction. While trained XAI models can provide explainable results to support physicians in making correct decisions, this is not sufficient to fully rely on algorithms for medical diagnosis. Clinicians should also provide medical knowledge to guide the design and modification of AI algorithms. By promoting human–machine interaction, XAI models can achieve more successful application in the medical field.

We made great efforts to include as many relevant articles as possible in our study. However, specific visual explanation methods are sometimes not mentioned in the titles or keywords of papers. Therefore, there may be omissions of relevant literature during the review process. Meanwhile, this review only searched the PubMed database. Given the differences in literature inclusion criteria among various databases, some potentially relevant studies may not have been included in this article. Finally, in this review, XAI refers to the relevant research and applications that use XAI methods to explain the output of original models, and does not include models that are inherently interpretable. This serves as a crucial premise for our subsequent elaboration and analysis in the article.

6. Conclusions

This article summarizes the clinical applications of XAI employing visual explanations in digestive endoscopic image analysis. Structurally, we first introduce the definition and classification of XAI; secondly, we discuss several commonly used XAI methods based on visual explanations and elaborate on their applications in different digestive endoscopies, respectively; ultimately, we summarize the value, limitations and future development prospects of XAI. We put forward our own insights into the current situation of XAI

based on visual explanations in the field of digestive endoscopic image analysis. In the future, in the field of medical image analysis an increasing number of XAIs will emerge to effectively achieve clinical translation. We believe that clinicians should participate in the development, design, and use of models to guide the design of XAIs that conform to clinical workflows and meets clinical needs, so as to give full play to the auxiliary role of XAIs in clinical diagnosis and treatment.

Author Contributions: Conceptualization, methodology, and writing—original draft preparation: X.C.; writing—original draft preparation: S.Z. and Z.Z.; conceptualization, methodology, writing—review & editing, project administration and funding acquisition: X.F. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Tianjin Key Medical Discipline (Specialty) Construction Project, grant number TJYXZDXK-002A (2023074, 2023075).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We sincerely thank the reviewers for their valuable feedback which helped improve the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rosenbacke, R.; Melhus, Å.; McKee, M.; Stuckler, D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. *JMIR AI* **2024**, *3*, e53207. [CrossRef]
2. Kröner, P.T.; Engels, M.M.; Glicksberg, B.S.; Johnson, K.W.; Mzaik, O.; van Hooft, J.E.; Wallace, M.B.; El-Serag, H.B.; Krittanawong, C. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J. Gastroenterol.* **2021**, *27*, 6794–6824. [CrossRef] [PubMed]
3. Yang, Y.J.; Bang, C.S. Application of artificial intelligence in gastroenterology. *World J. Gastroenterol.* **2019**, *25*, 1666–1683. [CrossRef] [PubMed]
4. Borys, K.; Schmitt, Y.A.; Nauta, M.; Seifert, C.; Krämer, N.; Friedrich, C.M.; Nensa, F. Explainable AI in medical imaging: An overview for clinical practitioners-Saliency-based XAI approaches. *Eur. J. Radiol.* **2023**, *162*, 110787. [CrossRef] [PubMed]
5. Chow, J.C.L. Quantum Computing and Machine Learning in Medical Decision-Making: A Comprehensive Review. *Algorithms* **2025**, *18*, 156. [CrossRef]
6. Sinonquel, P.; Eelbode, T.; Bossuyt, P.; Maes, F.; Bisschops, R. Artificial intelligence and its impact on quality improvement in upper and lower gastrointestinal endoscopy. *Dig. Endosc.* **2021**, *33*, 242–253. [CrossRef]
7. Kudo, S.E.; Mori, Y.; Misawa, M.; Takeda, K.; Kudo, T.; Itoh, H.; Oda, M.; Mori, K. Artificial intelligence and colonoscopy: Current status and future perspectives. *Dig. Endosc.* **2019**, *31*, 363–371. [CrossRef]
8. Lekadir, K.; Osuala, R.; Gallin, C.; Lazrak, N.; Kushibar, K.; Tsakou, G.; Aussó, S.; Alberich, L.C.; Marias, K.; Tsiknakis, M.; et al. FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging. *arXiv* **2019**, arXiv:2109.09658.
9. Singh, A.; Sengupta, S.; Lakshminarayanan, V. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging* **2020**, *6*, 52. [CrossRef]
10. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [CrossRef]
11. van der Velden, B.H.M.; Kuijf, H.J.; Gilhuijs, K.G.A.; Viergever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. [CrossRef]
12. Salih, A.; Galazzo, I.B.; Gkontra, P.; Lee, A.M.; Lekadir, K.; Raisi-Estabragh, Z.; Petersen, S.E. Explainable Artificial Intelligence and Cardiac Imaging: Toward More Interpretable Models. *Circ. Cardiovasc. Imaging* **2023**, *16*, e014519. [CrossRef]
13. Qian, J.; Li, H.; Wang, J.; He, L. Recent Advances in Explainable Artificial Intelligence for Magnetic Resonance Imaging. *Diagnostics* **2023**, *13*, 1571. [CrossRef]
14. Luo, Y.; Tseng, H.-H.; Cui, S.; Wei, L.; Ten Haken, R.K.; El Naqa, I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR Open* **2019**, *1*, 20190021. [CrossRef]

15. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef]
16. Hoque, R.A.; Yadav, M.; Yadava, U.; Rai, N.; Negi, S.; Yadav, H.S. Active site determination of novel plant versatile peroxidase extracted from Citrus sinensis and bioconversion of β-naphthol. *3 Biotech* **2023**, *13*, 345. [CrossRef]
17. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion.* **2020**, *58*, 82–115. [CrossRef]
18. Murdoch, W.J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 22071–22080. [CrossRef] [PubMed]
19. Ying, R.; Bourgeois, D.; You, J.; Zitnik, M.; Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. *Adv. Neural Inf. Process Syst.* **2019**, *32*, 9240–9251. [PubMed]
20. Salahuddin, Z.; Woodruff, H.C.; Chatterjee, A.; Lambin, P. Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Comput. Biol. Med.* **2022**, *140*, 105111. [CrossRef] [PubMed]
21. Liu, Z.; Luo, C.; Chen, X.; Feng, Y.; Feng, J.; Zhang, R.; Ouyang, F.; Li, X.; Tan, Z.; Deng, L.; et al. Noninvasive prediction of perineural invasion in intrahepatic cholangiocarcinoma by clinicoradiological features and computed tomography radiomics based on interpretable machine learning: A multicenter cohort study. *Int. J. Surg.* **2024**, *110*, 1039–1051. [CrossRef]
22. Ancona, M.; Ceolini, E.; Öztireli, C.; Gross, M. Towards Better Understanding of Gradient-Based Attribution Methods for Deep Neural Networks. 2018. Available online: https://openreview.net/forum?id=Sy21R9JAW (accessed on 9 July 2025).
23. Borys, K.; Schmitt, Y.A.; Nauta, M.; Seifert, C.; Krämer, N.; Friedrich, C.M.; Nensa, F. Explainable AI in medical imaging: An overview for clinical practitioners–Beyond saliency-based XAI approaches. *Eur. J. Radiol.* **2023**, *162*, 110786. [CrossRef]
24. Kim, B.; Wattenberg, M.; Gilmer, J.; Cai, C.; Wexler, J.; Viegas, F.; Sayres, R. Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018. PMLR 80.
25. Das, A.; Rad, P. Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. *arXiv* **2020**, arXiv:2006.11371. [CrossRef]
26. Simonyan, K.; Vedaldi, A.; Zisserman, A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv* **2014**, arXiv:1312.6034. [CrossRef]
27. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision–ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, The Netherlands, 2014; pp. 818–833.
28. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: New York, NY, USA, 2017; pp. 618–626.
29. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 2921–2929.
30. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE* **2015**, *10*, e0130140. [CrossRef]
31. Nadimi, E.S.; Braun, J.-M.; Schelde-Olesen, B.; Khare, S.; Gogineni, V.C.; Blanes-Vidal, V.; Baatrup, G. Towards full integration of explainable artificial intelligence in colon capsule endoscopy's pathway. *Sci. Rep.* **2025**, *15*, 5960. [CrossRef]
32. Jin, W.; Li, X.; Fatehi, M.; Hamarneh, G. Generating post-hoc explanation from deep neural networks for multi-modal medical image analysis tasks. *MethodsX* **2023**, *10*, 102009. [CrossRef] [PubMed]
33. Huff, D.T.; Weisman, A.J.; Jeraj, R. Interpretation and visualization techniques for deep learning models in medical imaging. *Phys. Med. Biol.* **2021**, *66*, 04TR01. [CrossRef] [PubMed]
34. Ribeiro, M.T.; Singh, S.; Guestrin, C. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 1135–1144.
35. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2017. Available online: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (accessed on 13 July 2025).
36. Ali, S.; Akhlaq, F.; Imran, A.S.; Kastrati, Z.; Daudpota, S.M.; Moosa, M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Comput. Biol. Med.* **2023**, *166*, 107555. [CrossRef]
37. Wang, T.T.; Zhu, S.L. Overview of the Types and Applications of Digestive Endoscopy. *World Latest Med. Inf.* **2019**, *19*, 114–117. [CrossRef]
38. Cao, J.S.; Lu, Z.Y.; Chen, M.Y.; Zhang, B.; Juengpanich, S.; Hu, J.H.; Li, S.J.; Topatana, W.; Zhou, X.Y.; Feng, X.; et al. Artificial intelligence in gastroenterology and hepatology: Status and challenges. *World J. Gastroenterol.* **2021**, *27*, 1664–1690. [CrossRef]

39. Ge, Z.; Wang, B.; Chang, J.; Yu, Z.; Zhou, Z.; Zhang, J.; Duan, Z. Using deep learning and explainable artificial intelligence to assess the severity of gastroesophageal reflux disease according to the Los Angeles Classification System. *Scand. J. Gastroenterol.* **2023**, *58*, 596–604. [CrossRef]
40. de Souza, L.A.; Mendel, R.; Strasser, S.; Ebigbo, A.; Probst, A.; Messmann, H.; Papa, J.P.; Palm, C. Convolutional Neural Networks for the evaluation of cancer in Barrett's esophagus: Explainable AI to lighten up the black-box. *Comput. Biol. Med.* **2021**, *135*, 104578. [CrossRef]
41. García-Peraza-Herrera, L.C.; Everson, M.; Lovat, L.; Wang, H.-P.; Wang, W.L.; Haidry, R.; Stoyanov, D.; Ourselin, S.; Vercauteren, T. Intrapapillary capillary loop classification in magnification endoscopy: Open dataset and baseline methodology. *Int. J. CARS* **2020**, *15*, 651–659. [CrossRef] [PubMed]
42. Everson, M.A.; Garcia-Peraza-Herrera, L.; Wang, H.-P.; Lee, C.-T.; Chung, C.-S.; Hsieh, P.-H.; Chen, C.-C.; Tseng, C.-H.; Hsu, M.-H.; Vercauteren, T.; et al. A clinically interpretable convolutional neural network for the real-time prediction of early squamous cell cancer of the esophagus: Comparing diagnostic performance with a panel of expert European and Asian endoscopists. *Gastrointest. Endosc.* **2021**, *94*, 273–281. [CrossRef] [PubMed]
43. Ueyama, H.; Kato, Y.; Akazawa, Y.; Yatagai, N.; Komori, H.; Takeda, T.; Matsumoto, K.; Ueda, K.; Matsumoto, K.; Hojo, M.; et al. Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging. *J. Gastroenterol. Hepatol.* **2021**, *36*, 482–489. [CrossRef] [PubMed]
44. Hu, H.; Gong, L.; Dong, D.; Zhu, L.; Wang, M.; He, J.; Shu, L.; Cai, Y.; Cai, S.; Su, W.; et al. Identifying early gastric cancer under magnifying narrow-band images with deep learning: A multicenter study. *Gastrointest. Endosc.* **2021**, *93*, 1333–1341.e3. [CrossRef]
45. Cho, B.-J.; Bang, C.S.; Lee, J.J.; Seo, C.W.; Kim, J.H. Prediction of Submucosal Invasion for Gastric Neoplasms in Endoscopic Images Using Deep-Learning. *J. Clin. Med.* **2020**, *9*, 1858. [CrossRef]
46. Katz, P.O.; Dunbar, K.; Schnoll-Sussman, F.H.; Greer, K.B.; Yadlapati, R.; Spechler, S.J. ACG Clinical Guideline: Guidelines for the Diagnosis and Management of Gastroesophageal Reflux Disease. *Am. J. Gastroenterol.* **2022**, *117*, 27–56. [CrossRef]
47. Hatta, W.; Koike, T.; Ogata, Y.; Kondo, Y.; Ara, N.; Uno, K.; Asano, N.; Imatani, A.; Masamune, A. Comparison of Magnifying Endoscopy with Blue Light Imaging and Narrow Band Imaging for Determining the Invasion Depth of Superficial Esophageal Squamous Cell Carcinoma by the Japanese Esophageal Society's Intrapapillary Capillary Loop Classification. *Diagnostics* **2021**, *11*, 1941. [CrossRef]
48. Dekker, E.; Tanis, P.J.; Vleugels, J.L.A.; Kasi, P.M.; Wallace, M.B. Colorectal cancer. *Lancet* **2019**, *394*, 1467–1480. [CrossRef]
49. Siegel, R.L.; Wagle, N.S.; Cercek, A.; Smith, R.A.; Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 233–254. [CrossRef] [PubMed]
50. Leslie, A.; Carey, F.A.; Pratt, N.R.; Steele, R.J.C. The colorectal adenoma–carcinoma sequence. *Br. J. Surg.* **2002**, *89*, 845–860. [CrossRef]
51. Zhao, S.; Wang, S.; Pan, P.; Xia, T.; Chang, X.; Yang, X.; Guo, L.; Meng, Q.; Yang, F.; Qian, W.; et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. *Gastroenterology* **2019**, *156*, 1661–1674.e11. [CrossRef]
52. Maas, M.H.J.; Neumann, H.; Shirin, H.; Katz, L.H.; Benson, A.A.; Kahloon, A.; Soons, E.; Hazzan, R.; Landsman, M.J.; Lebwohl, B.; et al. A computer-aided polyp detection system in screening and surveillance colonoscopy: An international, multicentre, randomised, tandem trial. *Lancet Digit. Health* **2024**, *6*, e157–e165. [CrossRef]
53. Corley, D.A.; Jensen, C.D.; Marks, A.R.; Zhao, W.K.; Lee, J.K.; Doubeni, C.A.; Zauber, A.G.; de Boer, J.; Fireman, B.H.; Schottinger, J.E.; et al. Adenoma detection rate and risk of colorectal cancer and death. *N. Engl. J. Med.* **2014**, *370*, 1298–1306. [CrossRef]
54. Chen, J.; Wang, G.; Zhou, J.; Zhang, Z.; Ding, Y.; Xia, K.; Xu, X. AI support for colonoscopy quality control using CNN and transformer architectures. *BMC Gastroenterol.* **2024**, *24*, 257. [CrossRef]
55. Wickstrøm, K.; Kampffmeyer, M.; Jenssen, R. Uncertainty and interpretability in convolutional neural networks for semantic segmentation of colorectal polyps. *Med. Image Anal.* **2020**, *60*, 101619. [CrossRef] [PubMed]
56. Jin, E.H.; Lee, D.; Bae, J.H.; Kang, H.Y.; Kwak, M.-S.; Seo, J.Y.; Yang, J.I.; Yang, S.Y.; Lim, S.H.; Yim, J.Y.; et al. Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations. *Gastroenterology* **2020**, *158*, 2169–2179.e8. [CrossRef]
57. Zhang, Q.-W.; Zhang, Z.; Xu, J.; Dai, Z.-H.; Zhao, R.; Huang, J.; Qiu, H.; Tang, Z.-R.; Niu, B.; Zhang, X.-B.; et al. Multi-step validation of a deep learning-based system with visual explanations for optical diagnosis of polyps with advanced features. *iScience* **2024**, *27*, 109461. [CrossRef] [PubMed]
58. Choi, K.; Choi, S.J.; Kim, E.S. Computer-Aided Diagonosis for Colorectal Cancer using Deep Learning with Visual Explanations. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2020**, *2020*, 1156–1159. [CrossRef]
59. Chierici, M.; Puica, N.; Pozzi, M.; Capistrano, A.; Donzella, M.D.; Colangelo, A.; Osmani, V.; Jurman, G. Automatically detecting Crohn's disease and Ulcerative Colitis from endoscopic imaging. *BMC Med. Inf. Decis. Mak.* **2022**, *22*, 300. [CrossRef]
60. Sutton, R.T.; Zai Ane, O.R.; Goebel, R.; Baumgart, D.C. Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images. *Sci. Rep.* **2022**, *12*, 2748. [CrossRef]

61. Weng, F.; Meng, Y.; Lu, F.; Wang, Y.; Wang, W.; Xu, L.; Cheng, D.; Zhu, J. Differentiation of intestinal tuberculosis and Crohn's disease through an explainable machine learning method. *Sci. Rep.* **2022**, *12*, 1714. [CrossRef]
62. Gabralla, L.A.; Hussien, A.M.; AlMohimeed, A.; Saleh, H.; Alsekait, D.M.; El-Sappagh, S.; Ali, A.A.; Refaat Hassan, M. Automated Diagnosis for Colon Cancer Diseases Using Stacking Transformer Models and Explainable Artificial Intelligence. *Diagnostics* **2023**, *13*, 2939. [CrossRef]
63. Binzagr, F. Explainable AI-driven model for gastrointestinal cancer classification. *Front. Med.* **2024**, *11*, 1349373. [CrossRef] [PubMed]
64. Auzine, M.M.; Heenaye-Mamode Khan, M.; Baichoo, S.; Gooda Sahib, N.; Bissoonauth-Daiboo, P.; Gao, X.; Heetun, Z. Development of an ensemble CNN model with explainable AI for the classification of gastrointestinal cancer. *PLoS ONE* **2024**, *19*, e0305628. [CrossRef] [PubMed]
65. Dahan, F.; Shah, J.H.; Saleem, R.; Hasnain, M.; Afzal, M.; Alfakih, T.M. A hybrid XAI-driven deep learning framework for robust GI tract disease diagnosis. *Sci. Rep.* **2025**, *15*, 21139. [CrossRef]
66. Gideon, S.G.; Princess, P.J.B. Explainable AI for Gastrointestinal Disease Detection using Ensemble Deep Learning Techniques. In Proceedings of the 2025 International Conference on Visual Analytics and Data Visualization (ICVADV), Tirunelveli, India, 4–6 March 2025; IEEE: New York, NY, USA, 2025; pp. 1113–1119.
67. Vleugels, J.L.A.; Hazewinkel, Y.; Dijkgraaf, M.G.W.; Koens, L.; Fockens, P.; Dekker, E.; DISCOUNT study group. Optical diagnosis expanded to small polyps: Post-hoc analysis of diagnostic performance in a prospective multicenter study. *Endoscopy* **2019**, *51*, 244–252. [CrossRef] [PubMed]
68. Huang, J.; Fan, X.; Liu, W. Applications and Prospects of Artificial Intelligence-Assisted Endoscopic Ultrasound in Digestive System Diseases. *Diagnostics* **2023**, *13*, 2815. [CrossRef]
69. Yoshida, T.; Yamashita, Y.; Kitano, M. Endoscopic Ultrasound for Early Diagnosis of Pancreatic Cancer. *Diagnostics* **2019**, *9*, 81. [CrossRef]
70. Sooklal, S.; Chahal, P. Endoscopic Ultrasound. *Surg. Clin. North. Am.* **2020**, *100*, 1133–1150. [CrossRef]
71. Fan, X.; Huang, J.; Cai, X.; Maihemuti, A.; Li, S.; Fang, W.; Wang, B.; Liu, W. Clinical value of the nomogram model based on endoscopic ultrasonography radiomics and clinical indicators in identifying benign and malignant lesions of the pancreas. *Front. Oncol.* **2025**, *15*, 1504593. [CrossRef] [PubMed]
72. Gu, J.; Pan, J.; Hu, J.; Dai, L.; Zhang, K.; Wang, B.; He, M.; Zhao, Q.; Jiang, T. Prospective assessment of pancreatic ductal adenocarcinoma diagnosis from endoscopic ultrasonography images with the assistance of deep learning. *Cancer* **2023**, *129*, 2214–2223. [CrossRef] [PubMed]
73. Yi, N.; Mo, S.; Zhang, Y.; Jiang, Q.; Wang, Y.; Huang, C.; Qin, S.; Jiang, H. An endoscopic ultrasound-based interpretable deep learning model and nomogram for distinguishing pancreatic neuroendocrine tumors from pancreatic cancer. *Sci. Rep.* **2025**, *15*, 3383. [CrossRef] [PubMed]
74. Marya, N.B.; Powers, P.D.; Chari, S.T.; Gleeson, F.C.; Leggett, C.L.; Abu Dayyeh, B.K.; Chandrasekhara, V.; Iyer, P.G.; Majumder, S.; Pearson, R.K.; et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* **2021**, *70*, 1335–1344. [CrossRef] [PubMed]
75. Mo, S.; Huang, C.; Wang, Y.; Qin, S. Endoscopic ultrasonography-based intratumoral and peritumoral machine learning ultrasomics model for predicting the pathological grading of pancreatic neuroendocrine tumors. *BMC Med. Imaging* **2025**, *25*, 22. [CrossRef]
76. Cai, X.H.; Fan, X.F.; Li, S.; Fang, W.L.; Wang, B.M.; Wang, Y.F.; Feng, Y.; Mu, J.B.; Liu, W.T. Construction of a multimodal interpretable machine learning model based on radiomics and clinical features for distinguishing benign and malignant pancreatic lesions. *World Chin. J. Dig.* **2025**, *33*, 361–372. [CrossRef]
77. Cui, H.; Zhao, Y.; Xiong, S.; Feng, Y.; Li, P.; Lv, Y.; Chen, Q.; Wang, R.; Xie, P.; Luo, Z.; et al. Diagnosing Solid Lesions in the Pancreas With Multimodal Artificial Intelligence. *JAMA Netw. Open* **2024**, *7*, e2422454. [CrossRef]
78. Liang, S.Q.; Cui, Y.T.; Hu, G.B.; Guo, H.Y.; Chen, X.R.; Zuo, J.; Qi, Z.R.; Wang, X.F. Development and validation of a machine-learning model for preoperative risk of gastric gastrointestinal stromal tumors. *J. Gastrointest. Surg.* **2025**, *29*, 101864. [CrossRef]
79. Liu, G.S.; Huang, P.Y.; Wen, M.L.; Zhuang, S.S.; Hua, J.; He, X.P. Application of endoscopic ultrasonography for detecting esophageal lesions based on convolutional neural network. *World J. Gastroenterol.* **2022**, *28*, 2457–2467. [CrossRef] [PubMed]
80. Uema, R.; Hayashi, Y.; Kizu, T.; Igura, T.; Ogiyama, H.; Yamada, T.; Takeda, R.; Nagai, K.; Inoue, T.; Yamamoto, M.; et al. A novel artificial intelligence-based endoscopic ultrasonography diagnostic system for diagnosing the invasion depth of early gastric cancer. *J. Gastroenterol.* **2024**, *59*, 543–555. [CrossRef]
81. Liao, Z.; Hou, X.; Lin-Hu, E.-Q.; Sheng, J.-Q.; Ge, Z.-Z.; Jiang, B.; Hou, X.-H.; Liu, J.-Y.; Li, Z.; Huang, Q.-Y.; et al. Accuracy of Magnetically Controlled Capsule Endoscopy, Compared With Conventional Gastroscopy, in Detection of Gastric Diseases. *Clin. Gastroenterol. Hepatol.* **2016**, *14*, 1266–1273.e1. [CrossRef]
82. Malhi, A.; Kampik, T.; Pannu, H.; Madhikermi, M.; Framling, K. Explaining Machine Learning-Based Classifications of In-Vivo Gastral Images. In *2019 Digital Image Computing: Techniques and Applications (DICTA)*; IEEE: Perth, Australia, 2019; pp. 1–7.

83. Wang, S.; Xing, Y.; Zhang, L.; Gao, H.; Zhang, H. Deep Convolutional Neural Network for Ulcer Recognition in Wireless Capsule Endoscopy: Experimental Feasibility and Optimization. *Comput. Math. Methods Med.* **2019**, *2019*, 7546215. [CrossRef]
84. Mukhtorov, D.; Rakhmonova, M.; Muksimova, S.; Cho, Y.-I. Endoscopic Image Classification Based on Explainable Deep Learning. *Sensors* **2023**, *23*, 3176. [CrossRef] [PubMed]
85. Mascarenhas, M.; Afonso, J.; Ribeiro, T.; Cardoso, H.; Andrade, P.; Ferreira, J.P.S.; Saraiva, M.M.; Macedo, G. Performance of a Deep Learning System for Automatic Diagnosis of Protruding Lesions in Colon Capsule Endoscopy. *Diagnostics* **2022**, *12*, 1445. [CrossRef] [PubMed]
86. Nadimi, E.S.; Buijs, M.M.; Herp, J.; Kroijer, R.; Kobaek-Larsen, M.; Nielsen, E.; Pedersen, C.D.; Blanes-Vidal, V.; Baatrup, G. Application of deep learning for autonomous detection and localization of colorectal polyps in wireless colon capsule endoscopy. *Comput. Electr. Eng.* **2020**, *81*, 106531. [CrossRef]
87. Kakogeorgiou, I.; Karantzalos, K. Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102520. [CrossRef]
88. Haupt, M.; Maurer, M.H.; Thomas, R.P. Explainable Artificial Intelligence in Radiological Cardiovascular Imaging—A Systematic Review. *Diagnostics* **2025**, *15*, 1399. [CrossRef]
89. Budhkar, A.; Song, Q.; Su, J.; Zhang, X. Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics. *Comput. Struct. Biotechnol. J.* **2025**, *27*, 346–359. [CrossRef]
90. Markus, A.F.; Kors, J.A.; Rijnbeek, P.R. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J. Biomed. Inform.* **2021**, *113*, 103655. [CrossRef] [PubMed]
91. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608. [CrossRef]
92. Dong, Z.; Wang, J.; Li, Y.; Deng, Y.; Zhou, W.; Zeng, X.; Gong, D.; Liu, J.; Pan, J.; Shang, R.; et al. Explainable artificial intelligence incorporated with domain knowledge diagnosing early gastric neoplasms under white light endoscopy. *NPJ Digit. Med.* **2023**, *6*, 64. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.