IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 October 2023, accepted 22 November 2023, date of publication 30 November 2023, date of current version 8 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3338545

TOPICAL REVIEW

Generative Adversarial Networks for Augmenting Endoscopy Image Datasets of Stomach Precancerous Lesions: A Review

BRUNO MAGALHÃES, ALEXANDRE NETO, AND ANTÓNIO CUNHA
Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
School of Sciences and Technology, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal

Corresponding author: Bruno Magalhães (bruno.s.magalhaes@inesctec.pt)

This work was supported in part by the National Funds through the Portuguese Funding Agency, FCT—Fundacão para a Ciência e a Tecnologia under Project PTDC/EEI-EEE/5557/2020; in part by the European Union under Grant 101095359; and in part by the U.K. Research and Innovation under Grant 10058099. The work of Alexandre Neto was supported by the Individual FCT under Grant 2022.13770.BD.

ABSTRACT Gastric cancer (GC) is still a significant public health issue, among the most common and deadly cancers globally. The identification and characterization of precancerous lesions of the stomach using endoscopy are crucial for determining the risk of cancer and guiding appropriate surveillance. In this scenario, deep learning (DL)-based computer vision methods have the potential to help us classify and identify particular patterns in endoscopic images, leading to a more accurate classification of these types of lesions. The quantity and quality of the data used heavily influence the classification performance of DL networks. However, one of the major setbacks for developing high-performance DL classification models is the typical need for more available data in the medical field. This review explores the use of Generative Adversarial Networks (GANs) and classical data augmentation techniques for improving the classification of precancerous stomach lesions. GANs are DL models that have shown promising results in generating synthetic data, which can be used to augment limited medical datasets. This review discusses recent studies that have implemented GANs and classical data augmentation methods to improve the accuracy of cancerous lesion classification. The results indicate that GANs can effectively increase the dataset's size, enhance the classification models' performance. In specific applications, such as the augmentation of endoscopic images depicting gastrointestinal polyps and Barrett's esophagus Adenocarcinoma, our review reveals instances where GANs, including models like Deep Convolutional GANs and conditional GANs, outperform classical data augmentation methods. Furthermore, this review highlights the challenges and limitations of the recent works using GANs and classical data augmentation techniques in medical imaging analysis and proposes directions for future research.

INDEX TERMS Dataset augmentation, deep learning, generative adversarial networks, precancerous lesions, upper gastro endoscopy.

I. INTRODUCTION

Gastric cancer (GC) remains a significant global health concern, with over one million new cases reported in 2020 and an estimated 769,000 deaths, ranking fifth in incidence and fourth in mortality among all types of cancer worldwide [1]. Like several malignancies, intestinal-type GC typically progresses through precancerous stages, characterized by conditions that confer an increased risk of cancer development in the long term.

GC is most commonly sporadic in nature, and chronic inflammatory conditions often trigger its development. The progression of GC is a multistep process, starting with chronic gastritis (typically caused by chronic infection with *Helicobacter pylori*) and then progressing to the loss of normal gastric glands, known as atrophic gastritis (AG).

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

In some cases, the atrophic mucosa may develop intestinal-type glands, known as intestinal metaplasia (GIM), which can then progress to low-grade dysplasia, high-grade dysplasia, and eventually, invasive adenocarcinoma. The trigger for GC development, characterized by a multistep process beginning with long-standing inflammation, was first described by Pelayo Correa and is now known as Correa's multi-stage cascade of gastric oncogenesis [2] (Figure 1)(Figure 2).

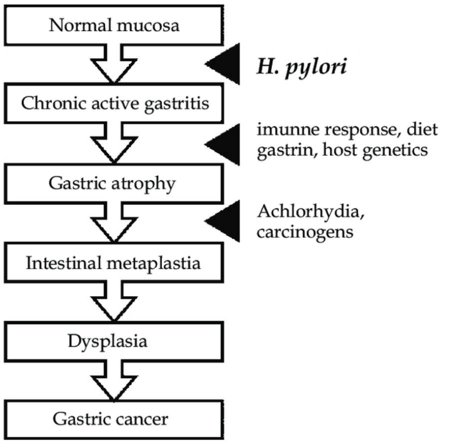


FIGURE 1. Correa's multi-stage cascade of gastric oncogenesis [3].

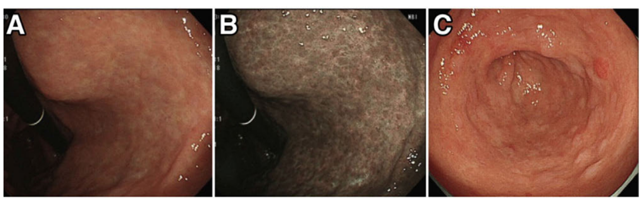


FIGURE 2. Typical endoscopic appearance of chronic AG and GIM. Characteristic endoscopic features of chronic AG include pale appearance of mucosa, loss of gastric rugal folds, and prominence of submucosal blood vessels due to thinning of the atrophied gastric epithelium, as shown in (A) HD-WLE and (B) NBI. Changes representing IM are frequently found in chronic AG. On HD-WLE, the areas with IM typically appear mildly nodular (C) [4].

The prognosis of GC is closely related to the stage of the disease at diagnosis, where cure is only possible when GC is diagnosed in the early stages. However, early diagnosis is difficult since symptoms are rare until the disease becomes locally advanced or disseminated [5]. Early diagnosis is thus only possible if screening programs are adopted, and patients at higher risk with precancerous conditions are identified and submitted to surveillance.

Endoscopy is a diagnostic and therapeutic tool that enables the visualization of the digestive tract. Detecting and characterizing precancerous lesions through endoscopy is essential for determining cancer risk and guiding appropriate surveillance. However, it is crucial to recognize that the prevalence of precancerous gastric lesions may be underestimated, as endoscopic assessment alone is insufficient for their detection in routine practice [5].

Identifying precancerous lesions through endoscopic manual screening is an intensive and time-consuming labor process that heavily relies on clinical expertise since the correlation between endoscopic and pathological findings is suboptimal. However, virtual chromoendoscopy can increase the detection and characterization of gastric precancerous conditions and lesions [6]. Virtual chromoendoscopy still needs expertise and is not widespread among endoscopists. Computer-assisted diagnosis (CAD) using artificial intelligence (AI) can mitigate these challenges. The utilization of CAD systems in medicine has significantly expanded in recent years. The growing interest in these systems is mainly due to their ability to make diagnostic decisions at a level comparable or superior to that of a qualified human expert, thus functioning as an expert system [7].

Deep learning (DL) based computer vision methods can classify and detect specific image patterns or objects. DL-based methods for image classification are a well-established and widely researched area. Supervised image classification with DL requires a dataset and class labels. The network is then trained according to the supervision provided by the labels [8].

Convolutional neural networks (CNNs), a type of DL architecture, have been applied to improve the diagnosis of gastrointestinal lesions such as colorectal polyps, esophageal cancer, *Helicobacter pylori* infection, and GC [9]. Recently, a CNN system based on endoscopic white light images has shown high sensitivity, specificity, and accuracy for recognizing AG and GIM [10], [11], [12]. DL methods have been demonstrated to effectively detect gastrointestinal diseases through endoscopy, with studies showing successful results in identifying gastrointestinal disorders [13].

The quantity and quality of the data used can significantly impact the classification performance of DL networks. However, in the medical domain, the data available is often limited. Other issues, such as data imbalance, where the number of samples in each class category varies significantly, also affect the dataset's quality. The quality of a dataset affects its ability to represent the real world accurately; the samples should reflect a distribution similar to that scenario [14]. It is also important to note that annotating extensive endoscopic data is a costly, time-consuming process requiring specialized expertise [15].

A common strategy to address the limited and imbalanced data availability in the medical domain and improve the performance of DL-based image classification methods is data augmentation. Studies have demonstrated that data augmentation techniques can enhance a dataset's quantity and quality, which involve manipulating or synthesizing existing images [16]. Recently, sub-branches of data augmentation have emerged, such as generative data augmentation, which utilizes generative models to create new data [17].

Generative models can be broadly defined as a network that describes how a dataset is generated using a probabilistic model. By sampling from this model, new data can be generated. In endoscopy image analysis, a common challenge is the lack of large, annotated datasets, which restricts the application of supervised DL. Generating labeled datasets in the medical domain requires significant medical expertise, time, and effort [18].

For example, in the case of a dataset containing a sub-optimal number of upper endoscopic images of stomach precancerous lesions, it may be desirable to build a model that can generate new images of this type of lesion that have never existed but still appear realistic because the model has learned the general rules governing the appearance of this specific type of lesion. This is the type of problem that can be solved using generative models [19].

Generative Adversarial Networks (GANs) have brought a new perspective to the DL community. GANs were first introduced in 2014 [20] and later described and discussed compared to other generative models in detail [21]. The architecture of GANs consists of two neural networks: the generator and the discriminator. The task of each network is to compete against each other so that the generator network's output images are indistinguishable from the discriminator. This adversarial training enables the synthesis of images similar to real images. It is important to remark that challenges in the application of GANs for gastric lesion classification, usually stem from inherent variations in lesion characteristics, intricacies involved in capturing subtle patterns, and the imperative for fine-tuning GAN architectures to align with the nuanced features present in gastric images. However, with the advancements in GAN structure and computing power, generative data augmentation in the image domain has become a reality. The advances in GANs and their various applications in the medical domain, including image classification, are discussed within the context of this paper.

A. MOTIVATION

The need for more medical data available challenges the development of DL classification models. The dataset used in DL models often requires some form of data augmentation to achieve high performance. Classical data augmentation has been the traditional solution to this problem. However, generative models, such as GANs, have emerged as a promising alternative for data generation in DL models.

Thus, this literature review aims to assess the current state of the utilization of GANs for enhancing DL models in detecting and classifying stomach precancerous lesions through endoscopy image analysis. The review will provide a comprehensive understanding of the significance of GANs in detecting cancerous lesions and their impact on improving the performance of DL models in this area. The information obtained from the state-of-the-art literature review will bring us closer to determining the best GAN approach for stomach precancerous lesion detection. Our objective is to evaluate and compare the effectiveness of classical dataset augmentation methods and GAN-based dataset augmentation techniques on the performance of DL models.

In the end, this literature review will provide valuable insights into the current state of GANs in enhancing DL models for detecting stomach precancerous lesions and contribute to developing more effective DL models in this area.

B. CONTRIBUTIONS

Use This study undertakes a literature review to evaluate the current state of the usage of GANs in enhancing DL models for the detection and classification of stomach precancerous lesions through endoscopic image analysis. Research questions are formulated to address the identified issues and try to provide a comprehensive understanding of the problem and pave the way for future research by answering the following questions:

Q1: How often and effectively are GANs used to improve the classification performance of stomach precancerous lesions through dataset augmentation in endoscopic image analysis?

Q2: How do we best apply GANs to improve pattern detection in endoscopic images?

Q3: Do the results obtained with dataset augmentation through GANs surpass the ones obtained with classical data augmentation?

II. RELATED WORK

In recent years, there has been a remarkable surge of interest and activity in the domain of generative dataset augmentation for medical images. Researchers and practitioners have begun to recognize the potential of generative DL models in augmenting medical image datasets, leading to a burgeoning body of research and exploration. This resurgence in interest signifies a pivotal turning point in the field, as it is increasingly recognized as a promising avenue for advancing medical image analysis and enhancing the performance of related models. Therefore, this section aims to bridge this gap by presenting a collection of pertinent studies that delve into the application of generative DL models within this domain. By exploring these studies, we aim to acquire a thorough and nuanced understanding of the advancements, challenges, and potential benefits of leveraging generative DL techniques to augment medical image datasets. This analysis will provide valuable insights that inform future research endeavors and facilitate the development of practical approaches in generative dataset augmentation for medical images.

A. REVIEWS

Li et al. [22] investigated the research status of GANs in the context of medical images. The study analyzed several GAN methods commonly applied in the medical image domain. It examined the applications of GANs for medical image synthesis and adversarial learning in various medical image tasks. The study also discussed the existing open challenges and identified future research directions in this field, aiming to provide insights into the potential advancements and opportunities for GANs in medical image analysis and synthesis. The study highlights several challenges of using GANs in medical imaging, including datasets, training methods, reliability, and legality. Additionally, it discusses future directions for unsupervised learning, advancements in addressing clinical needs, and the necessity for GANs tailored explicitly for medical imaging applications.

In applying generative dataset augmentation to medical images, Chen et al. [23] conducted a comprehensive and systematic review encompassing various topics, including the benefits of different augmentation models, loss functions, and evaluation metrics. Additionally, the study examined 105 research papers specifically focused on medical image augmentation. These papers were categorized based on the specific anatomical regions represented in the images. The review also documented the medical image datasets utilized in the studies, the loss functions employed during model training, and the quantitative evaluation metrics applied for image augmentation. Overall, GAN-based medical image augmentation was recognized as a promising approach to effectively address the challenge of limited training samples in medical image diagnosis and treatment models.

Garcea et al. [24] conducted a systematic literature review to examine the data augmentation strategies employed in the medical domain and their impact on the performance of clinical tasks, including classification, segmentation, and lesion detection. The study encompassed a rigorous analysis of over 300 articles published between 2018 and 2022. The review's findings underscore the effectiveness of data augmentation techniques across different organs, imaging modalities, tasks, and dataset sizes. Furthermore, the study identifies potential directions for future research in data augmentation in the medical domain.

B. PRACTICAL WORKS

In a more practical approach, Bissoto et al. [25] critically reviewed using GANs for data augmentation and anonymization in skin lesion analysis. The study shed light on the persisting challenge of data scarcity in this domain despite high-quality public datasets. The findings suggest caution in adopting GANs for medical applications, as favorable results primarily surfaced in out-of-distribution test sets. Notably, preliminary experiments uncovered noise and experimental protocol flaws in GAN-based data augmentation, emphasizing the complexity of transforming synthetic images into reliable performance enhancements. Furthermore, GANs, with their substantial computational requirements, bear the potential to introduce biases and spurious correlations. While GAN-based data anonymization showed promise, particularly for out-of-distribution data, further research is required to evaluate their effectiveness in preserving patient privacy. These insights encourage researchers to exercise prudence, carefully assess the justification for GAN usage, and explore avenues for enhancing the reliability of GANs in data augmentation and anonymization. Despite the challenges, exploring scenarios where GANs enhance results can provide valuable insights into fundamental aspects of DL within the context of skin lesion analysis.

The study by Motamed et al. [26] addresses the critical challenge of training CNNs effectively with limited data. Data augmentation techniques are commonly used to enhance the generalizability of neural networks; however, traditional methods have limitations in generating plausible alternative data. The paper introduces GANs to generate new data effectively. The proposed GAN-based augmentation method is designed explicitly for chest X-ray images, focusing on detecting pneumonia and COVID-19 cases. Comparative experiments are conducted, pitting the GAN-based augmentation against Deep Convolutional GAN (DCGAN) and traditional augmentation methods such as rotation and zoom. The results showcase the superiority of the proposed GAN-based augmentation method, particularly in improving sensitivity and specificity. Statistical analysis, including the calculation of the area under the ROC curve (AUC), demonstrates the significance of this augmentation approach in enhancing the performance of GANs for anomaly detection in chest X-rays. In summary, the study presents IAGAN as a promising semi-supervised GAN-based augmentation method, showing statistical significance in improving GAN's performance for detecting pneumonia and COVID-19 anomalies in chest X-ray images.

In medical image analysis, the importance of substantial annotated data for the success of DL methods, particularly CNNs, is undeniable. However, obtaining sufficient data remains a significant challenge, especially for specific medical learning tasks. Addressing this limitation, Xu et al. [27] presents a novel data augmentation solution called semi-supervised attention-guided CycleGAN (SSA-CycleGAN). This approach leverages cycle-consistency GANs to generate synthetic tumor and normal medical images from their counterparts. A semi-supervised attention module is also introduced to enhance the model's ability to capture critical details, resulting in more realistic synthetic images. Experimental studies across three medical image datasets with limited MRI images demonstrate the effectiveness of SSA-CycleGAN in augmenting data. It excels in adding or removing tumor lesions and generating realistic tumor and normal images.

Furthermore, SSA-CycleGAN surpasses classic data augmentation methods in ResNet18-based MRI image classification tasks, emphasizing its potential to enhance classification performance. Notably, this work distinguishes itself by focusing on generating full-sized medical images and addressing the challenge of generating abnormal images robustly, making it a promising approach for data augmentation in the medical domain. Future extensions may explore tumor/lesion segmentation applications and expand to other medical domains needing improved training performance through generated abnormal images.

Overall, the existing medical image synthesis technology is highly reliable, and combining GANs with other medical image models demonstrates promising outcomes. Consequently, GANs possess significant potential and promising prospects for further development in medical imaging.

The current section gives us a compelling image that the field of generative dataset augmentation for upper gastrointestinal endoscopic images has been inadequately explored in recent years regarding the positive evaluation of its influence on model performance. Surprisingly, only a few reviews have been conducted in this domain, and more notably, no prior studies have been found investigating the application of generative dataset augmentation for endoscopic images of precancerous lesions in the stomach.

III. LITERATURE REVIEW METHODOLOGY

A. ELIGIBILITY CRITERIA

The eligibility criteria (outlined in Table 1) were established to identify relevant studies that address the research questions and exclude studies that are not applicable. These criteria were explicitly designed to capture the relevant literature while limiting the number of documents to those specific to addressing research questions. Implementing these criteria before conducting the literature search helps minimize bias in the selection process [28]

TABLE 1. Inclusion (IC) and exclusion (EC) criteria.

Criteria	Description
IC0	Published since 2017
IC1	The title, abstract, or keywords match the search query
IC2	Work published in refereed journals or conference
IC3	Relevant and transposable methods using GANs for dataset augmentation of endoscopic images
EC0	Work not published in refereed journal or conference
EC1	Literature/Systematic Review
EC2	Full text is not available
EC3	The paper is not written in English
EC4	Does not consider the use of GANs or Data Augmentation
EC5	The dataset is not obtained from gastrointestinal endoscopy
EC6	Dataset obtained from capsule endoscopy
EC7	Out of scope

1) IDENTIFICATION PHASE

Three well-established databases were included in the search phase: IEEE Xplore, Web of Science, and MEDLINE (through PubMed). The search was conducted through Titles/Abstracts/Keywords. After several iterations, the final version of the query was the following:

> *("Generative Adversarial Network" OR "Convolutional Neural Network" OR "Generative Models" OR "Data Augmentation" OR "Deep Learning" OR "Machine Learning") AND (Intestinal OR Gastr* OR Stomach) AND (cancer* OR lesions OR "Atrophic Gastritis" OR "metaplasia" OR Dysplasia) AND (Endosc*)*

The search was conducted on Jun. 14, 2023, and ranged from documents available from 2017 to 2023. A total of 1448 documents were added to the screening phase. To maintain the contemporaneity and relevance of our findings, a restriction was applied to the search process, confining the scope exclusively to papers published since 2017. This approach guarantees the inclusion of the most recent DL algorithms in our results.

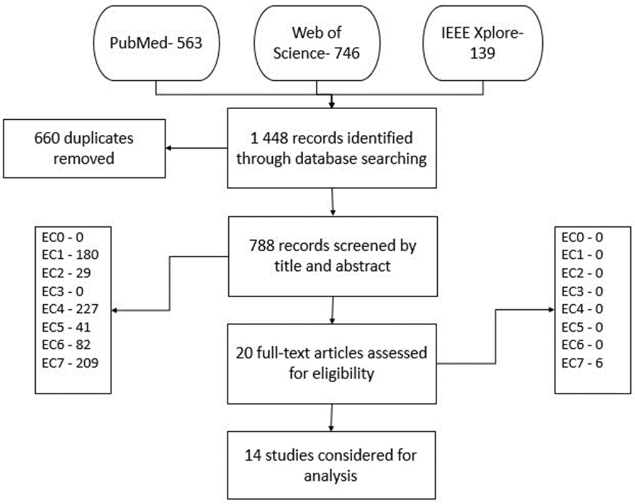


FIGURE 3. Study selection flow diagram.

IV. RESULTS

This section will present our results based on the eligibility criteria outlined in Section III. We will begin by providing a brief motivation for each problem, followed by a critical examination of the algorithmic solutions reviewed, including a thorough analysis of the datasets used, the number of samples, modalities, classes, and the authors' reported results. The outcomes of this review will be summarized in Table 2 and Table 3.

A. SCREENING PHASE

The initial search in the three databases retrieved 1448 documents, of which 660 were duplicates, resulting in 788 unique works for the screening phase. The duplicates were identified through individual titles and abstract reading. The screening phase excluded 768 documents (through exclusion criteria), resulting in 20 documents for full-text reading.

B. ELIGIBILITY PHASE

The resulting papers underwent a full-text assessment guided by the exclusion criteria to assess their eligibility. The exclusion criteria excluded six of the 20 documents left from the screening phase. The papers that fulfilled the eligibility criteria were processed to extract all the data necessary to answer the proposed research questions and conduct a quality assessment. Given the limited number of studies investigating the application of GANs for dataset augmentation in the classification of precancerous stomach lesions, studies were included in this review despite needing to focus on these lesions explicitly. These studies were chosen due to the similarity in methods, datasets, and problem areas addressed, which were deemed relevant to our research (IC3). Thus, of the initial 1284 documents, 14 were deemed for data extraction for the scope of this literature review.

C. CLASSICAL DATA AUGMENTATION

Data augmentation is a technique employed to artificially increase the size of a dataset by applying various modifications to the original images. This strategy is particularly useful in medical imaging as it can improve the performance of machine-learning models by increasing data available for training and reducing overfitting. In the literature, a series of approaches for detecting and classifying gastrointestinal lesions using DL techniques have been proposed, utilizing data augmentation techniques to counteract the issue of insufficient sample availability. These techniques include image rotation, image inversion, color transformation, and noise addition. A summary of the works that have proposed these processes of data augmentation can be found in Table 2.

1) ALGORITHMS

Standard, well-known data augmentation methods were used by five works considered in this review, including image rotation [29], [30], [31], [32], [33], image flipping [33], image cropping [29], [32], image inversion [31], color transformation and noise addition [31].

A common approach observed by Itoh et al. [30], Yan et al. [32], and Qiu et al. [33] was the use of transfer learning along with fine-tuning techniques to increase the model's performance and dispute the lack of available data. A pre-trained model was obtained using a large dataset of ImageNet natural images and freezing the shallow network for fine-tuning. Transfer learning has emerged as a popular approach for enhancing the performance of deep CNNs in computer vision tasks. By leveraging the knowledge learned from pre-trained CNN models on large-scale datasets, transfer learning enables fine-tuning the model to a new task using a smaller dataset. This approach can significantly boost the model's performance and alleviate the scarcity of training data. Fine-tuning allows the model to adapt to the specific characteristics of the new task while still retaining the generic features learned from the pre-trained model. This results in improved accuracy compared to training a model from scratch with a limited dataset. Furthermore, transfer learning reduces the risk of overfitting and can reduce the training time required to perform well. These advantages make transfer learning using fine-tuning a valuable approach for CNNs in practical applications with limited training data [34], [35], [36].

On a different approach, Ham et al. [37] conducted a study utilizing AutoAugment to identify the most suitable data augmentation approach for a specific dataset, followed by an additional augmentation technique using Grad-CAM. The authors introduced an image augmentation method based on Grad-CAM to preserve lesion characteristics while enhancing anomaly classification performance based on the work previously made by [38]. This approach improved the classification performance of all the models evaluated. The AutoAugment increases data by applying geometric transformations, such as rotation, to images. These changes can cause severe transformations and slow down training, leading to poor weight convergence and limited performance improvement.

In contrast, the Grad-CAM augmentation method creates an image by synthesizing the Region of Interest (ROI) in the original image, thus allowing for augmentation while preserving the original image's color and lesion morphological characteristics. This allows for increasing the data significantly without interfering with the training convergence. The Grad-CAM augmentation method also allows for a more significant data increase than the general augmentation method. The study extracted ROI from 551 selected images, increasing 655 abnormal images to 360,905.

In their work, Chae and Cho [39] employed Multi-Filter AutoAugment (MFAA) in conjunction with AutoAugment to enhance the Vision Transformer's classification performance for identifying healthy tissue, gastric lesions, and early GC. They recognized that although AutoAugment aids in data augmentation, it can introduce erroneous or misplaced features in the augmented data, potentially impeding performance or failing to yield substantial improvements. To address this challenge, the authors proposed MFAA as a solution. MFAA incorporates a data filtering procedure that utilizes the classification model trained with original data to process the augmented data generated by AutoAugment. The weights of the classification model retain the crucial feature information necessary for classifying the target object. Leveraging this insight, the authors filtered the augmented data, allowing only the relevant data generated by AutoAugment to remain. This methodology ensures that only the augmented data containing the appropriate features are retained, overcoming the limitations posed by the indiscriminate use of all augmented data. Furthermore, classification models rely on distinct architectural designs and objectives to extract and employ diverse features for object classification. Taking advantage of this aspect, the authors employed various classification models for filtering. This approach selectively retains augmented data with essential features for object classification, presenting the primary advantage of MFAA. Two classification models, ViT and BiT, were trained and used for MFAA, and notable performance improvements were achieved in the classification of healthy tissues, gastric lesions, and early GC using the training data for the classification model. The results demonstrated an impressive F1-score of 0.92 and an AUC of 0.97. These findings highlight the approach's efficacy in enhancing the accuracy and precision of the classification task across all target categories;

TABLE 2. Classical Data Augmentation reviewed articles. Acronyms used: Image Modality (IM); Institutional Review Board (IRB); White light endoscopy (WLE); Narrow band imaging (NBI); Magnifying endoscopy with NBI (ME-NBI); Region of interest (ROI); Area under the ROC Curve (AUC); Intersection over Union (IoU); Data Augmentation (DA); Accuracy (Acc); Sensitivity (S); Specificity (Sp); Precision (P); Recall (Rc); F1-score (F1).

Authors	Data Annotation Protocol	IM	Dataset	Classes	Algorithm	Augmentation	Results
Itoh et al., (2018) [30]	N.D.	WLE	(Private Dataset) **Pre-DA:** 179 images **Post-DA:** 596 images	Positive for H. pylori infection Negative for H. pylori infection	GoogleNet DCNN+transfer learning	Rotation	S=86.7 % Sp=86.7 % AUC=0.956
Hatami et al., (2020) [29]	Images annotated by two expert clinicians	N.D.	(Private Dataset) **Pre-DA:** 1331 **Post-DA:** 3673	Erosion Polyps Ulcer	AlexNet+ 2 fire modules	ROI extraction, Cropping, Rotation	Acc=89%
Qiu et al., (2022) [31]	Experienced senior endoscopist with significant clinical expertise	N.D.	(Private Dataset) **Pre-DA:** 3591 images **Post-DA:** 5261 images	Advanced GC Early cancer Precancerous lesions Normal Benign lesions	DLU-Net	Rotation, Inversion, Color transformation, Noise addition	P=91.1%/ Rc=92.4%/ IoU=91.1%/ Average Acc=94.1%
H. -S. Ham et al., (2022) [37]	Images approved by the IRB and biopsy-verified	WLE	(Private Dataset) **Pre-DA:** 1638 images **Post-DA:** 360,905 images	Normal Abnormal	Inception-v3; EfficientNet-B3; ResNet-152; VGG-16; ViT-B; Xception.	AutoAugment and GRAD-CAM	Acc=84.2% P=86.8% S=80.5% Sp=81.8% F1=83.5% AUC=90.3%
Yan et al., (2020) [32]	A pathologist conducted a reevaluation of two experienced endoscopists	NBI ME-NBI	(Private Dataset) **Pre-DA:** 1880 images **Post-DA:** 11280 images	GIM Non-GIM	Xception; NASNet; EfficientNetB4 + Transfer learning.	Cropping; rotation; Flipping; shifting, zooming	**Xception:** S=95%/Sp=95%/ Acc=95% **NASNet:** S=95%/Sp=95%/ Acc=95% **EfficientNetB4:** S=95%/Sp=95%/ Acc=95%
K. Qiu et al., (2022) [33]	The labeling results provided by endoscopy doctors serve as a reference	NBI ME-NBI	(Private Dataset) **Pre-DA:** 8592 images **Post-DA:** 34000 images	Inflammation Intestinal metaplasia Low-grade neoplasia Early cancer	ATP-UNet + transfer learning	Flipping, Rotation	mIoU=91% mAP=95% F-score=96% V-score=95%
Chae e Cho., (2023) [39]	The IRB has granted approval for all gastroscopic images, ensuring that they are biopsy-verified.	N.D.	(Private Dataset) **Pre-DA:** 900 images **Post-DA:** 27045 images	**Dataset A:** Healthy condition; Abnormalities **Dataset B:** Early Gastric cancer; Gastric Lesions	ViT-H/14	Multi-Filter AutoAugment	P=91.19% S=84.22% F1=86.76% AUC=92.27%
Lee e Cho., (2023) [40]	The IRB has granted approval for all gastroscopic images, ensuring that they are biopsy-verified.	WLE	(Private Dataset) **Pre-DA:** 1200 images **Post-DA:** 115016 images	**Dataset A:** Normal; Abnormal **Dataset B:** Early Gastric Cancer; Abnormal; Normal	EfficientNetV2	Augmented Method Using Poisson Blending	**Dataset A:** P=98.2%/S=91.7%/F1=94.8%/Acc=95% **Dataset B:** P=82.4%/S=82.3%/F1=82.3%/Acc=82.3%

moreover, the effectiveness of the proposed method extended beyond the original dataset used for training. Significant performance improvements were observed when applied to additional gastroscopy data obtained from an independent institution. This outcome underscores the robustness and generalizability of the method, as it demonstrated consistent success in classifying external data sources.

In line with the significance of training data quality and quantity for CAD systems, Lee and Cho [40] proposed a novel data augmentation technique. Their approach involved synthesizing abnormal and normal endoscopic images using Poisson blending, aiming to generate visually natural images that capture the pronounced intensity variation along lesion boundaries. By adopting this method, the authors sought to improve the training data by introducing more realistic and diverse samples. The proposed technique yielded remarkable results, as the EfficientNetV2 model trained with the augmented dataset achieved the highest classification performance across all evaluation metrics. Notably, this model exhibited an average improvement of approximately 29% compared to the model trained solely on the original dataset. Such a substantial enhancement in classification performance underscores the effectiveness of the proposed data augmentation technique.

The other works reviewed in this section [29], [31] relied on more classic data augmentation methods, such as image rotation, cropping, inversion, color transformation, and noise addition. The improvement of the classification performance was primarily obtained through improvements in the algorithm of the classification models. It is shown in Table 2 that different combinations of classical data augmentation were utilized for dataset expansion. All the works resorted to different CNN frameworks for their classification challenge, applying different changes for a more specific network.

Different methods of dataset preprocessing were observed in the papers reviewed. Some of the dataset preparations for the training phase included the cropping of redundant image parts [33], for example, black frames of the selected original images for the development of the ID system [32]. In a different work, realized by Hatami et al. [29], the images available on the dataset presented a size of 460×475, and resources to the ROI and processes such as crop and rotation by image preprocessing experts ended up reaching 3673 images with a size of 32×32. Finally, in two of the works reviewed [31], [32], the endoscopic images were randomly divided into K groups as a form of cross-validation to help prevent overfitting.

It is worth mentioning that only the Yan et al. [32] work resorted to some explainable AI to help us understand and interpret predictions made by the CNN model. Attention maps were generated using the Grad-CAM method, allowing us to observe where on the image was a higher contribution to the classification decision given.

2) DATASETS

Medical imaging datasets are often small due to the high cost and complexity of collecting and annotating medical data. Data augmentation can be used to artificially increase the size of the dataset, which can help to improve the performance of DL models. All the considered works used private data to train and validate the models. The datasets varied considerably in size, the types of imaging modalities, and the number of classes associated with different anatomical landmarks (see Table 2). Dataset size ranged from a few hundred (e.g., 596 in [30]) to hundreds of thousands (e.g., 369,905 in [37]) of images after the implementation of data expansion.

Various reviewed works used different criteria for selecting images for their datasets. They all included upper endoscopic images focused on the stomach and obtained written consent from patients or protected their identities for ethical reasons. While some works prioritized obtaining high-quality images only [33], others focused on the specific endoscopic technique used for acquiring the images [37]. In Yan et al. [32] work, both criteria were followed, as only images obtained with NBI or ME-NBI without any blurring, lack of focus, halation, or mucus were accepted from their dataset.

Each dataset's total number of classes was small (2-5). Itoh et al. [30] focused on detecting Helicobacter pylori infection in the stomach by constructing a CNN optimized for the purpose in question by learning endoscopic images. Only two classes were present, ''Positive'' and ''Negative'' for HP infection. In the case of [31], the model can output five classification results: advanced ''Gastric Cancer'', ''Early Gastric Cancer'', ''precancerous lesions'', ''normal'', and ''benign lesions''.

In the works reviewed, the datasets were composed of different image modalities, where some considered only white light endoscopy (WLE) frames [30], [37], [39] while others considered narrow-band imaging (NBI) and magnifying endoscopy with NBI (ME-NBI) data [32], [33]. Furthermore, three need to define from what endoscopic technique the dataset was obtained [29], [31], [40].

D. DATA AUGMENTATION USING GANs

The use of GANs to generate high-quality endoscopic images could be used as a method to reduce dataset bias by generating new images to generate large amounts of data for the performance improvement of DL models.

A summary of the works that have proposed GAN approaches as a method of expanding datasets can be found in Table 3. It is worth mentioning that all of the works presented in Table 3 were the result of the applied methodology in Section III. Three of the six works presented do not directly discuss stomach precancerous lesions [42], [43], [44]. However, they seemed worth analyzing due to their relevant methodology using GANs on augmenting datasets containing endoscopic images and the lack of available bibliography concerning precancerous stomach lesions.

1) ALGORITHMS

The selected studies apply GANs for medical dataset expansion, followed by a DL algorithm to classify or detect cancerous lesions. To improve the model's generalization

TABLE 3. Data Augmentation through GANs reviewed articles. Acronyms used: Image Modality (IM); High-definition white-light endoscopic (HD-WLE), Blue light imaging (BLI), and Linked color imaging (LCI), Data Augmentation (DA); Accuracy (Acc); Sensitivity (S); Specificity (Sp); Precision (P); Recall (Rc); F1-score (F1).

Authors	Dataset Annotation Protocol	IM	Dataset	Classes	Algorithm	Augmentation	Results
Cui et al, (2022) [45]	Two senior endoscopists, with the more experienced endoscopist participating in case of ambiguity.	WLE BLI LCI	(Private Dataset) **PA dataset:** 2078 images to 11704 images after DA **NPA dataset:** 909 images to 8699 images after DA	**PA dataset:** Type I (early cancer areas) Type 0 (non-cancer areas) **NPA dataset:** Class1 (suspicious areas) Class2 (highly suspicious areas) Class3 (confirmed areas)	Mask R-CNN + BiFPN	**PA dataset:** DCGAN + classic DA **NPA dataset:** noise addition, brightness change, randomly change pixel value, flipping and translation	**PA dataset:** Acc=89.42%/S=91.67% /Sp=88.95% **NPA dataset:** Average Acc=86.93%
Kim et al, (2021) [46]	Senior endoscopist specialist	WLE	(Private Dataset) **Pre-DA:** 1638 images **Post-Da:** 25854 images	Normal Abnormal	Xception	DCGAN and AutoAugment	**DCGAN+CIFAR-10:** Acc=0.851 (0.820)/P= 0.896 (0.800)/Rc=0.793 (0.854)/F1=0.841 (0.826)
Kanayama et al, (2019) [48]	The patient's attending doctor annotated each image, with bounding boxes on the lesion parts in images displaying lesions.	WLE	(Private Dataset) **Pre-DA:** 131007 images **Post-DA:** Plus 128387 images generated	Normal Lesion	CNN	DCGAN	Average P: 0.632 ± 0.013
P. Sasmal et al, (2020) [44]	Expert's annotations	NBI; WLE	(Private and Public Dataset) **Pre-DA:** Dataset 1: 61 images/ Dataset 2: 581 images **Post-DA:** 4486 IMAGES	Adenoma polyps Hyperplastic polyps Serrated polyps	CNN	DCGAN	**Dataset 1:** Acc=88.33%/S=90.0%/ Sp=86.60% **Dataset 2:** Acc=88.33%
de Souza Jr et al, (2020) [43]	Various experts independently outlined suspicious regions identified in the cancerous images.	HD-WLE	(Public and Private Dataset) **Pre-DA:** Dataset 1(100 images); Dataset 2(76 images) **Post-DA:** N.D.	Barrett's esophagus Adenocarcinoma	LeNet-5 AlexNet	Rotation, Zooming, and horizontal mirroring vs DCGAN	**MICCAI 2015:** Acc= 0.90 ± 0.03 **Augsburg:** Acc=0.88 ± 0.05
A. Samse H. H. Shomee, (2022) [42]	The Kvasir and CVC-ClinicDB datasets feature images that undergo meticulous annotations by one or more medical experts.	N.D.	(Public Dataset) **Pre-DA:** 1612 images **Post-DA:** N.D.	Gastrointestinal Polyps	YOLOv4	StyleGAN2-ada + Conditional GAN	P=84% Rc=74% F1=79% IoU=64.83%

ability and reduce the risk of over-fitting, Cui et al. [45] collected a data set of gastroscopy images and enhanced it. An improved DCGAN method was combined with traditional methods to expand the finely annotated early GC data set. The Mask R-CNN+BiFPN model was proposed to enhance feature fusion and improve the detection effect of early GC lesions. This paper improves the feature pyramid network (FPN) in Mask R-CNN into a BiFPN network. BiFPN network was first proposed as a weighted bidirectional feature pyramid network, an efficient bidirectional cross-scale connection and weighted feature fusion network.

The DCGAN architecture replaces pooling layers with fully convolutional ones. It produces high-quality results while stabilizing training more effectively than the original GAN approach. This model or improved versions of it were used by several of the works considered in this review ([43], [44], [46], [48]). Kim et al. [46] combined the DCGAN model with 25 optimized augmentation policies for the CIFAR-10 dataset through AutoAugment to augment the data, and gastroscopy images were classified as normal or abnormal through the Xception network. Sasmal et al. [44] used the DCGAN architecture to train each polyp class separately, using the same 3-fold cross-validation process and data partition. In all the steps of the learning procedure, a complete separation between train and test subsets was taken. After the generator had learned each polyp class data distribution separately, it was able to synthesize new examples by using an input vector of Gaussian distributed samples, i.e., noise. The classification was done using a fully trained CNN to classify endoscopic polyps. A variation of this model was also used by de Souza Jr et al. [43] to cope with the small number of samples and to evaluate the robustness of data-augmented databases concerning Barrett's esophagus and adenocarcinoma using different CNN architectures. To fulfill that purpose, they considered the DCGAN architecture for the data augmentation due to its simple implementation and high generalization, as well as two CNN architectures, i.e., LeNet-5 and AlexNet.

While DCGAN presents the advantage of being suitable for a wide range of applications, other GAN architectures offer different features in their concepts. Sams and Shomee [42] set the objective of proposing GAN-based methods to fight the lack of substantial amount of data required to use in any DL-based detection algorithms for polyp detection. Their work initially used a StyleGAN2-ada to generate random polyp masks, followed by a conditional GAN that was used to translate these composite images into synthetic polyp images. The StyleGAN2-ada is then used to create composite images with healthy GI images. At the same time, the conditional GAN (cGAN) in this work adds a constraint model variable to guide the data generation process, making the convergence to a specific target faster. To evaluate the effect of synthetic images on detecting polyps from gastrointestinal tract images, the YOLOv4 object detection model was used.

In this section, the works reviewed involved, in some cases, the strict filtering, processing, and annotation of the dataset [45]. In the work done by de Souza Jr et al. [43], two different dataset preprocessing methods were applied to each respective approach: the image-based and the patch-based approach. For the first approach, the preprocessing step considered resizing the images to feed the data augmentation and the classification networks. For the patch-based approach, the images were split into patches, with the idea of covering the entire image with a sliding window of 200 × 200 pixels and overlapping 50 pixels in horizontal and vertical directions. Finally, in two of the works reviewed [43], [44], the endoscopic images were randomly divided into K groups as a form of cross-validation to help prevent overfitting.

2) DATASETS

The considered works reviewed in this section use a mix of private and public datasets. All the works in this section considered only WLE frames except the work done by de Souza Jr et al. [43], where there is a combination of WLE and NBI images. Furthermore, in Sams and Shomee [42], there needs to be a mention of the endoscopic technique in which the dataset was obtained. Even though the initial number of samples presented in each dataset from the reviewed articles differs considerably, all of them suffer an increase of more than double the final number of samples through synthetic data generation. The same variability was observed in size, the types of imaging modalities, and the number of classes associated with different anatomical landmarks from each dataset studied.

V. DISCUSSION

Based on the findings reported in Section IV, in this section, we describe the main trends revealed by the analysis of the works reviewed for the impact of classical data augmentation and GANs in gastric lesion detection and lesion characterization. Moreover, we set the pillars for a roadmap towards applying GANs in generating synthetic data for detecting and classifying gastric precancerous lesions.

A. CLASSICAL DATA AUGMENTATION

For medical images, the small volume of data and the need for specialist physician assistance in labeling often make the collection of data sets difficult, and it is often impossible to collect a large number of images. To improve the model's generalization ability and reduce the risk of over-fitting, data enhancement should be used to expand the data. Classical augmentation techniques on endoscopic images mainly include affine transformations such as translation, rotation, scaling, flipping, shearing, color transformation, and noise addition.

The authors used the augmentation methods applied to the works reviewed in section IV to contradict the lack of data available. The considered works report an average classification performance with sensitivity and specificity values often more than 89%, which indicates the relevance of AI algorithms for gastric lesion detection and the need for some data augmentation methods on the respective datasets. Ham et al. [37] achieved a performance improvement of 0.835 and 0.903 in terms of F1 score and AUC on their Xception model through a data augmentation approach using Grad-CAM. In the case of Yan et al. [32], the initial number of images acquired needed to be increased to train the CNN models. The gastric GIM dataset was augmented to increase the training samples' size and improve the ID system's robustness. Overall, in medical image classification and detection, data augmentation has shown that it is an effective method for improving the performance of CNN algorithms. Also, by synthesizing images with realistic intensity variations and blending abnormal and normal samples, Lee and Cho. [40] method effectively enriched the training dataset. The resulting model showcased superior classification capabilities, demonstrating the significant impact of the augmented data on the CAD system's performance. Nevertheless, some limitations can be registered when applying data augmentation methods on a dataset, e.g., improvements in image synthesis methods are required for more realistic augmented images and the support of transfer learning through fine-tuning.

It was impossible to draw statistically significant conclusions regarding the performance obtained by the different methods considered for the augmentation of the dataset. This was due to the significant differences in the datasets used for training and validations. These differences involve the number of images used, the number of classes considered restrictions in the imaging modalities adopted, and the pre-selection of informative frames. Thus, the performance of the different classification tasks cannot be directly compared.

However, as highlighted in Ham et al. [37], Itoh et al. [30] and Yan et al. [32], there is a pressing need for more efficient data augmentation methods in the medical field. One critical challenge classical data augmentation methods face is the limited generalization of the augmented dataset's outcomes. Many ID systems utilized in the reviewed papers rely on high-quality images for training and testing CNN classifiers. While this approach ensures accuracy, it may lead to limited generalization when faced with the variability of real-world data, which often includes lower-quality images. To address this limitation, [32] proposed incorporating poor-quality images in the data augmentation process. By introducing such images, they aim to enhance the generalization capabilities of ID systems. In particular, they suggest synthesizing new endoscopic images with a high Signal-to-noise Ratio (SNR), achieved through the utilization of additional GAN architectures such as CycleGAN or Multi-Scale Gradients GANs. These advanced GAN-based models can generate high-quality yet diverse synthetic images that bridge the gap between the need for high SNR images and the benefits of introducing variation through data augmentation. By thoughtfully integrating such synthetic images into training datasets, researchers can improve the robustness and generalization of CNN classifiers in the challenging domain of stomach precancerous lesion identification.

B. DATA AUGMENTATION USING GANs

Generative DL techniques, specifically GANs, have been widely investigated in generating synthetic data in recent years. The objective of GANs is to learn the underlying distribution patterns of input data through a set of images and subsequently generate new samples that closely resemble the learned distribution. GANs have gained significant attention in medical applications due to their potential to generate synthetic data. The study of the applicability of GANs for synthetic data augmentation and its effect on classification performance was reviewed in section IV, Table 3, of this work.

Two of the works reviewed adopt the use of some classic data augmentation methods to complement the synthetization of data via Generative models ([45], [46]). At the same time, the other four relied solely on synthetic data generated by the GANs to augment the dataset used. Kim et al. [46] achieved the best classification performance in the Xception proposed model when the DCGAN and the 25 augmentation policies were implemented to augment the training data. The same was observed in the work presented by Cui et al. [45], where the improved DCGAN was first used to expand the unsupervised data of the first type of lesions in the fine labeling data. After the expansion, all images were expanded in a ratio of 1:5 using classical data augmentation. The final accuracy achieved by the classification models of this work was 85.10% and 89.42%.

The performances of the models reviewed verified the advantage of using synthetic data over classical data augmentation. Sasmal et al. [44], when training the CNN used for polyp classification with only data generated with GAN, i.e., using synthetic data augmentation, the classification accuracy for the public dataset named DB1 and the private dataset, DB2 (DB1 75%, DB2 88.33%) was found to be higher compared to the dataset where only classical data augmentation was used (DB1 61.29%, DB2 69.7%). de Souza Jr et al. [43] also achieved better classification performance from the synthetic dataset, where 25%, 50%, and 75% of the synthetic samples were randomly added to the original training set. The more synthetic samples were added for learning purposes, the higher the model's accuracy. Such a finding reinforces the impact of high-quality data on the model's generalization when dealing with endoscopic imaging. The same results were observed in Sams and Shomee. [42], where the peak results on the Precision, Recall, F1-score, and IoU level of detection of polyps from the GI tract of the YOLOv4 were obtained in the dataset containing the most significant number of synthetic polyp images (84%; 74%; 79% and 64.83%).

However, having access to significant amounts of synthetic data is only sometimes correlated with better performance. In the work realized by Kanayama et al. [48], when the number of synthesized images input to the training dataset was changed, the model achieved the highest Average Precision score when 20,000 synthesized images were added (0.632 ± 0.013), and the performance of the model was lowered when added a more significant number of images. This indicates that the synthesized images have biases, and this causes poor effects when an excessive number of synthesized images is added.

It is also worth noting that de Souza Jr et al. [43] evaluated whether generating synthetic patches or the whole image is more effective when augmenting the dataset with synthetic data augmentation. The proposed approach in this work was validated over two datasets of endoscopic images, with the experiments conducted over the full and patch-split images. The best results were obtained using the patch-based approach. The same statement holds for both classes of interest, i.e., Barrett's esophagus and adenocarcinoma. Working with patches allows us to access considerably larger datasets, strongly influencing the GANs training step. It is also critical to define that the generation of full images presents a high computational cost, while patch generation requires less computational power due to the lower output resolution. Patches consider local information only while using the full-image content, which may make the learning process prone to errors during the approximation of the distribution of the pixels that belong to healthy and cancerous regions. The same method of generating synthetic patches was implemented by Kanayama et al. [48], where the dataset bias was lessened because the method allows lesion patches to be attached to various parts in normal images.

Heterogeneity in the kinds of cancerous lesions classes considered in the different works makes it difficult to compare their performance fairly. Other differences involve the number of images used and the number of anatomical landmarks considered. However, one common factor can be observed in all the works reviewed, which is the improvement of the classification performance of the DL model with the addition of GAN-generated Synthetic Data. The literature study revealed a consistent trend of improved results when utilizing synthetic data augmentation generated by GANs compared to traditional data augmentation techniques. Despite some studies incorporating both methods, it was observed that the utilization of GANs alone for dataset expansion yielded satisfactory results and superior performance compared to traditional methods. The main GAN framework utilized in the reviewed papers was the DCGAN framework, even though some level of transformation on the original framework of the DCGAN was applied to make it more specific for the dataset in hands. To generate data through DCGAN, there is a limitation in that the more significant the amount of data, the more diverse and high-quality data can be generated.

A DCGAN constitutes a foundational architecture within the GAN framework. They distinguish themselves using deep convolutional layers in the generator and discriminator networks. Renowned for their simplicity and effectiveness, DCGANs are a robust baseline for numerous image-generation tasks. To ensure the successful training of DCGANs, several prerequisites must be considered. These include access to a moderately sized dataset characterized by sufficient diversity.

Furthermore, tuning critical hyperparameters such as learning rates and batch sizes is imperative to attain stable training outcomes. Cui et al. [45] recognized the importance of optimizing the architecture of the DCGAN generator for proficient image generation. They initially configured the convolutional layers with dimensions of 1024, 512, 256, 128, and 3. However, they refined this architecture, adjusting it to 512, 256, 128, 64, and 3, which played a pivotal role in achieving notable outcomes. In the DCGAN architecture by Sasmal et al. [44], they employed a unique design that included a fully connected layer reshaped to dimensions of $8 \times 8 \times 1024$, along with the incorporation of five fractionally-strided convolutional layers. This unique architectural configuration significantly contributed to generating high-fidelity synthetic polyp images.

Furthermore, the accompanying discriminator network featured four convolutional layers, enhancing its ability to discriminate between genuine and synthetic images. In both de Souza Jr et al. [43] and Kanayama et al. [48] studies, the emphasis was placed on hyperparameter optimization rather than detailed architectural modifications. Kanayama et al. [48] work prioritized the fine-tuning of hyperparameters, specifically adopting a learning rate of 0.0002 and a minibatch size of 64. Although the study did not delve into extensive architectural specifics, this optimization strategy likely played a pivotal role in achieving superior results. Similarly, Souza Jr et al. [43] investigated the influence of batch size and other hyperparameters within the DCGAN-based data augmentation framework. While this research did not explicitly outline architectural alterations, it centered on the empirical exploration of varying batch sizes, particularly examining batch sizes of 16 and 32 samples. This exploration yielded valuable insights into the nuanced effects of batch size on the resulting output. Notably, both studies underscored the critical importance of hyperparameter optimization as a fundamental factor contributing significantly to their favorable research outcomes Sams and Shomee. [42] used a different approach, setting them apart from conventional methods. Their approach entailed a multifaceted utilization of GANs, combining the capabilities of StyleGAN2-ada, CycleGAN, and cGAN frameworks. StyleGAN2-ada, known for its style-based generator, enabled precise control over image attributes, which is crucial for creating diverse and high-quality synthetic images. CycleGAN, on the other hand, excels in image-to-image translation tasks, making it valuable for synthesizing realistic images. Including a conditional GAN further extended their capabilities, allowing for generating images with specific desired characteristics. Therefore, the StyleGAN2-ada introduced a novel architectural paradigm, incorporating a data augmentation layer with 18 transformations before the discriminator layer. While not elaborated in exhaustive detail, this innovative modification markedly improved the generation of diverse and lifelike polyp masks, mimicking the intricacies of real polyps. This advancement substantially elevated the quality and diversity of synthetic data, forming the foundation for subsequent phases of their research. Moreover, their study seamlessly integrated the CycleGAN framework, featuring an extended generator and discriminator architecture, to synthesize realistic polyp images from composite inputs. Notably, the generator network was meticulously designed, commencing with a 7×7 convolutional layers, followed by 15 ResNet blocks and fractional-strided convolutional layers. Although precise architectural specifics were not provided, these modifications to the CycleGAN framework likely contributed significantly to generating high-fidelity synthetic polyp images. Additionally, Sams and Shomee. [42] harnessed the power of a cGAN with tailored hyperparameter settings, including a cycle consistency weight (λ) of 50 and gradual learning rate reduction. They introduced identity mapping loss, optimally weighted at 0.3 times the cycle loss, to ensure the faithful reproduction of color profiles from source images within their synthetic polyp images. This identity mapping loss, represented as Liden (G, F), was pivotal in preserving consistent color profiles. In the end, this amalgamation of architectural enhancements, meticulous hyperparameter optimization, and the judicious integration of StyleGAN2-ada, CycleGAN, and cGAN frameworks, alongside the utilization of identity mapping loss, played a pivotal role in advancing the quality, diversity, and realism of synthetic polyp images. These innovations ultimately culminated in heightened research outcomes, underlining the transformative potential of GANs in medical image synthesis.

There were also different approaches utilized by the authors regarding whether it is more effective to generate synthetic patches or the whole image. One advantage of working with patches instead of whole image approaches in GAN models is that it allows for more fine-grained control over the generated images. By working with smaller patches of images, the GAN model can focus on specific regions or details of the image rather than trying to generate the entire image at once. This can lead to more realistic and higher-quality generated images, as the model can pay more attention to the specific features and characteristics of the patches it works with. Additionally, working with patches can reduce the computational requirements of the GAN model, as it processes smaller amounts of data at a time. This can be useful in cases where computational resources are limited, or the dataset is particularly large.

One of the possible limitations when utilizing GANs for dataset augmentation was found in the work presented here by Kanayama et al. [48], where they encountered a phenomenon known as 'mode collapse.' Mode collapse occurs when a GAN generates a limited subset of the potential output space, failing to capture the full diversity of the underlying data distribution. This limitation can result in the production of synthetic data that needs more variability, potentially hindering the ability of DL models to generalize effectively. In his case, this issue stemmed from the limited size of the lesion image dataset employed.

In this sense, there is a clear opportunity for applying more modern and sophisticated generative DL architectures that could provide increased results in detecting and classifying stomach precancerous lesions through endoscopy image analysis. In this field, novel contributions in the development of machine learning algorithms for dataset augmentation can be envisioned in the following directions:

1. Beyond classical data augmentation: Classical data augmentation is a great tool to approach the common problem of lack of quality data in the medical field. However, generative DL architectures are powerful tools for data set augmentation that can provide more realistic, diverse, and high-quality synthetic data compared to classical data augmentation techniques. Ultimately, this can improve machine learning models' performance and robustness.
2. Generative models for data augmentation: Generative models can be utilized to address the dearth of publicly available data and the lack of representation of multiple lesion classes by synthesizing images of lesions. GANs can generate synthetic data that is highly similar to real-world data, which can lead to better performance in classification and detection tasks.
3. Test of new GANs: Most works reviewed in this paper resorted to the DCGAN framework for dataset expansion. However, it is essential to note that different types of GANs are available, such as Wasserstein GAN, cGAN, and others, each with unique characteristics and capabilities. There is a need to emphasize the importance of evaluating the performance of different types of GANs in the classification improvement of stomach precancerous lesions. The use of conditional cGANs is suggested, as the samples generated with this type of GAN are annotated and of high quality for further training DL models. This could improve the performance, robustness, and generalization of the results.

VI. CONCLUSION

This literature review evaluated the current state of the usage of GANs in enhancing DL models for the detection and classification of stomach precancerous lesions through endoscopic image analysis. Twelve articles were considered relevant per the inclusion criteria, where a few were included even though they did not focus specifically on stomach precancerous lesions. This gave us the perception that this area is attractive for future research. From the studies identified with this review, we were able to answer the proposed questions as follows:

- **Q1:** How often and effectively are GANs used to improve the classification performance of stomach precancerous lesions through dataset augmentation in endoscopic image analysis?
- **R:** A limited number of available studies have explored the utilization of generative models for dataset augmentation, specifically in the context of stomach precancerous lesions. Among the reviewed works, various GAN frameworks, including DCGAN, StyleGAN, and CycleGAN, have been employed to synthesize new data for augmenting the dataset of stomach precancerous lesions. These GAN-based approaches have demonstrated promising results, contributing to improvements in the performance of DL classification models. Notably, using GANs for dataset expansion has increased diversity in the augmented data, ultimately enhancing the robustness and generalization capabilities of DL models. Some classification models achieved accuracy rates of up to 90% when trained on datasets that included synthetic images. The classification results improved in most cases as more synthetic images were integrated into the dataset.
- **Q2:** How do we best apply GANs to improve pattern detection in endoscopic images?
- **R:** The optimal application of GANs for enhancing pattern detection in endoscopic images depends on the specific characteristics of the images and patterns in question. When considering GAN-based improvements, assessing whether generating synthetic patches or entire images is more suitable is essential. Additionally, the practical application of GANs for enhancing pattern detection in endoscopic images is contingent on various factors, with a crucial consideration being the customization of GAN architecture and hyperparameters. While established baseline architectures can serve as a valuable starting point, it is paramount to recognize that each dataset and pattern detection task may necessitate personalized adjustments. The unique characteristics of endoscopic images, such as variations in lighting, tissue types, and lesion appearances, demand a tailored

approach. This leads to the realization that there is no universal 'one-hit formula' when employing GANs in this context. In the works we have reviewed, it becomes evident that researchers often embark on a journey of architectural and hyperparameter customization to align the GAN model with the specific nuances of their dataset. Moreover, hyperparameters like learning rates, batch sizes, and regularization terms are not standardized but meticulously adjusted to balance model stability and convergence. This level of personalized adaptation extends to conditional GANs, where setting cycle consistency weights and identity mapping loss factors is guided by the specific color and pattern requirements of endoscopic images. The successful application of GANs in endoscopy necessitates recognizing that the architecture and hyperparameters should be fine-tuned to meet the distinct challenges posed by each dataset and pattern detection task.
- **Q3:** Do the results obtained with dataset augmentation through GANs surpass the ones obtained with classical data augmentation?
- **R:** An examination of the comparative effectiveness of GANs and traditional data augmentation techniques reveals nuanced outcomes across different scenarios, consistently showcasing the advantages of synthetic data over classical augmentation. Specifically, when models were trained for specific tasks using solely GAN-generated data, superior performance was observed, outperforming datasets augmented only with classical methods. Comparable trends were noted in other works, where increased utilization of synthetic samples consistently led to elevated model accuracy, precision, recall, F1-score, and IoU levels when compared to datasets augmented through classical methods. The notable superiority of GANs in these comparative analyses emphasizes their potential for more effective augmentation, reinforcing the significance of synthetic data in enhancing model performance. However, it is crucial to acknowledge that the correlation between access to significant amounts of synthetic data and improved performance is not universal. Some works found an optimal balance, achieving better results with a specific quantity of synthesized images, while excessive synthesis led to diminished performance. This highlights the nuanced nature of data augmentation, emphasizing the need for a thoughtful approach that recognizes the varying effectiveness of GANs and classical augmentation methods based on dataset characteristics and model requirements.

It is important to note that the use of GANs for data augmentation is still an evolving field, and there are limitations to consider, such as computational complexity and the potential for mode collapse. Nevertheless, the work presented in this literature review shows that GANs are a viable alternative to classical data augmentation techniques and offer significant potential for improving the performance of DL models to classify stomach precancerous lesions in endoscopic images. Future research in the application of GANs for enhancing DL models in stomach precancerous lesion classification could explore novel GAN architectures beyond the DCGAN framework. While the reviewed works predominantly employed DCGAN, the exploration of alternative architectures such as conditional GAN, Wasserstein GAN, and Big GAN holds promise for further improving the classification performance. This avenue of investigation could provide valuable insights into the comparative effectiveness of different GAN architectures in augmenting endoscopic datasets of precancerous stomach lesions.

ACKNOWLEDGMENT

National Funds finance this work through the Portuguese funding agency, FCT—Fundacão para a Ciência e a Tecnologia, within project PTDC/EEI-EEE/5557/2020. Co-funded by the European Union (grant number 101095359) and supported by the UK Research and Innovation (grant number 10058099). However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible. Alexandre Neto's work is also funded by the individual FCT grant reference number 2022.13770.BD.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[2] P. Correa and M. B. Piazuelo, "The gastric precancerous cascade," *J. Digestive Diseases*, vol. 13, no. 1, pp. 2–9, Jan. 2012, doi: 10.1111/j.1751-2980.2011.00550.x.

[3] M. Kargar, M. Baghernejad, A. Doosti, and S. Ghorbani-Dalini, "Clarithromycin Resistance and 23S rRNA Mutations in Helicobacter pylori," in *Gastrointestinal Endoscopy*, O. Pascu, Ed. InTech, 2011, doi: 10.5772/22655.

[4] *AGA Clinical Practice Update on the Diagnosis and Management of Atrophic Gastritis: Expert Review | Elsevier Enhanced Reader*. Accessed: Nov. 5, 2022. [Online]. Available: https://reader.elsevier.com/reader/sd/pii/S0016508521032364?token=4017E4CD66258FD7C47067DBBDEFC6F575CA7ED72F33BAC5AE639674C39511B5B8678B2843FA157FE6FA86FB83E63B76&originRegion=eu-west-1&originCreation=20221105153615

[5] N. Chapelle, M. Péron, J.-F. Mosnier, L. Quénéhervé, E. Coron, A. Bourget, E. Cauchin, Y. Touchefeu, and T. Matysiak-Budnik, "Prevalence, characteristics and endoscopic management of gastric premalignant lesions in France," *Digestive Diseases*, vol. 38, no. 4, pp. 286–292, 2020, doi: 10.1159/000503748.

[6] P. Pimentel-Nunes, D. Libânio, R. Marcos-Pinto, M. Areia, M. Leja, G. Esposito, M. Garrido, I. Kikuste, F. Megraud, T. Matysiak-Budnik, B. Annibale, J.-M. Dumonceau, R. Barros, J.-F. Fléjou, F. Carneiro, J. van Hooft, E. Kuipers, and M. Dinis-Ribeiro, "Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European society of gastrointestinal endoscopy (ESGE), European helicobacter and microbiota study group (EHMSG), European society of pathology (ESP), and sociedade portuguesa de endoscopia digestiva (SPED) guideline update 2019," *Endoscopy*, vol. 51, no. 4, pp. 365–388, Apr. 2019, doi: 10.1055/a-0859-1883.

[7] Y. Mori, S. Kudo, H. E. N. Mohmed, M. Misawa, N. Ogata, H. Itoh, M. Oda, and K. Mori, "Artificial intelligence and upper gastrointestinal endoscopy: Current status and future perspective," *Digestive Endoscopy*, vol. 31, no. 4, pp. 378–388, Jul. 2019, doi: 10.1111/den.13317.

[8] W. Du, N. Rao, D. Liu, H. Jiang, C. Luo, Z. Li, T. Gan, and B. Zeng, "Review on the applications of deep learning in the analysis of gastrointestinal endoscopy images," *IEEE Access*, vol. 7, pp. 142053–142069, 2019, doi: 10.1109/ACCESS.2019.2944676.

[9] C. Le Berre, W. J. Sandborn, S. Aridhi, M.-D. Devignes, L. Fournier, M. Smaïl-Tabbone, S. Danese, and L. Peyrin-Biroulet, "Application of artificial intelligence to gastroenterology and hepatology," *Gastroenterology*, vol. 158, no. 1, pp. 76.e2–94.e2, Jan. 2020, doi: 10.1053/j.gastro.2019.08.058.

[10] N. Lin et al., "Simultaneous recognition of atrophic gastritis and intestinal metaplasia on white light endoscopic images based on convolutional neural networks: A multicenter study," *Clin. Transl. Gastroenterol.*, vol. 12, no. 8, Aug. 2021, Art. no. e00385, doi: 10.14309/ctg.0000000000000385.

[11] L. Wu, J. Wang, X. He, Y. Zhu, X. Jiang, Y. Chen, Y. Wang, L. Huang, R. Shang, Z. Dong, B. Chen, X. Tao, Q. Wu, and H. Yu, "Deep learning system compared with expert endoscopists in predicting early gastric cancer and its invasion depth and differentiation status (with videos)," *Gastrointestinal Endoscopy*, vol. 95, no. 1, pp. 92.e3–104.e3, Jan. 2022, doi: 10.1016/j.gie.2021.06.033.

[12] M. Xu, W. Zhou, L. Wu, J. Zhang, J. Wang, G. Mu, X. Huang, Y. Li, J. Yuan, Z. Zeng, Y. Wang, L. Huang, J. Liu, and H. Yu, "Artificial intelligence in the diagnosis of gastric precancerous conditions by image-enhanced endoscopy: A multicenter, diagnostic study (with video)," *Gastrointestinal Endoscopy*, vol. 94, no. 3, pp. 540.e4–548.e4, Sep. 2021, doi: 10.1016/j.gie.2021.03.013.

[13] S. Ali et al., "Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102002, doi: 10.1016/j.media.2021.102002.

[14] D. Banik and D. Bhattacharjee, "Mitigating data imbalance issues in medical image analysis," in *Advances in Data Mining and Database Management*, D. P. Rana and R. G. Mehta, Eds. Hershey, PA, USA: IGI Global, 2021, pp. 66–89, doi: 10.4018/978-1-7998-7371-6.ch004.

[15] S. K. Thapa, P. Poudel, B. Bhattarai, and D. Stoyanov, "Task-aware active learning for endoscopic image analysis," Apr. 2022, *arXiv:2204.03440*. Accessed: Feb. 15, 2023.

[16] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *J. Med. Imag. Radiat. Oncol.*, vol. 65, no. 5, pp. 545–563, Aug. 2021, doi: 10.1111/1754-9485.13261.

[17] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552, doi: 10.1016/j.media.2019.101552.

[18] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016, doi: 10.1109/TMI.2016.2553401.

[19] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," *Artif. Intell. Med.*, vol. 109, Sep. 2020, Art. no. 101938, doi: 10.1016/j.artmed.2020.101938.

[20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Jun. 2014, *arXiv:1406.2661*. Accessed: Feb. 15, 2023.

[21] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," Apr. 2017, *arXiv:1701.00160*. Accessed: Nov. 2, 2022.

[22] X. Li, Y. Jiang, J. J. Rodriguez-Andina, H. Luo, S. Yin, and O. Kaynak, "When medical images meet generative adversarial network: Recent development and research opportunities," *Discover Artif. Intell.*, vol. 1, no. 1, p. 5, Dec. 2021, doi: 10.1007/s44163-021-00006-0.

[23] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, "Generative adversarial networks in medical image augmentation: A review," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105382, doi: 10.1016/j.compbiomed.2022.105382.

[24] F. Garcea, A. Serra, F. Lamberti, and L. Morra, "Data augmentation for medical imaging: A systematic literature review," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106391, doi: 10.1016/j.compbiomed.2022.106391.

[25] A. Bissoto, E. Valle, and S. Avila, "GAN-based data augmentation and anonymization for skin-lesion analysis: A critical review," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Nashville, TN, USA: IEEE, Jun. 2021, pp. 1847–1856, doi: 10.1109/CVPRW53098.2021.00204.

[26] S. Motamed, P. Rogalla, and F. Khalvati, "Data augmentation using generative adversarial networks (GANs) for GAN-based detection of pneumonia and COVID-19 in chest X-ray images," *Informat. Med. Unlocked*, vol. 27, 2021, Art. no. 100779, doi: 10.1016/j.imu.2021.100779.

[27] Z. Xu, C. Qi, and G. Xu, "Semi-supervised attention-guided CycleGAN for data augmentation on medical images," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*. San Diego, CA, USA: IEEE, Nov. 2019, pp. 563–568, doi: 10.1109/BIBM47256.2019.8982932.

[28] B. Kitchenham, "Procedures for performing systematic reviews," Keele Univ., Keele U.K., Tech. Rep., Aug. 2004, vol. 33.

[29] S. Hatami, D. R. Shamsaee, and M. H. Olyaei, "Detection and classification of gastric precancerous diseases using deep learning," in *Proc. 6th Iranian Conf. Signal Process. Intell. Syst. (ICSPIS)*, Dec. 2020, pp. 1–5, doi: 10.1109/ICSPIS51611.2020.9349603.

[30] T. Itoh, H. Kawahira, H. Nakashima, and N. Yata, "Deep learning analyzes helicobacter pylori infection by upper gastrointestinal endoscopy images," *Endoscopy Int. Open*, vol. 6, no. 2, pp. E139–E144, Feb. 2018, doi: 10.1055/s-0043-120830.

[31] W. Qiu, J. Xie, Y. Shen, J. Xu, and J. Liang, "Endoscopic image recognition method of gastric cancer based on deep learning model," *Expert Syst.*, vol. 39, no. 3, p. e12758, Mar. 2022, doi: 10.1111/exsy.12758.

[32] T. Yan, P. K. Wong, I. C. Choi, C. M. Vong, and H. H. Yu, "Intelligent diagnosis of gastric intestinal metaplasia based on convolutional neural network and limited number of endoscopic images," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104026, doi: 10.1016/j.compbiomed.2020.104026.

[33] K. Qiu, X. Bu, H. Zhou, and L. Liu, "Research on ME-NBI gastric lesion recognition system based on improved UNet structure," in *Proc. 3rd Int. Conf. Electron. Commun. Artif. Intell. (IWECAI)*, Jan. 2022, pp. 482–485, doi: 10.1109/IWECAI55315.2022.00099.

[34] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[35] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Three aspects on using convolutional neural networks for computer-aided detection in medical imaging," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing* (Advances in Computer Vision and Pattern Recognition), L. Lu, Y. Zheng, G. Carneiro, and L. Yang, Eds. Cham, Switzerland: Springer, 2017, pp. 113–136, doi: 10.1007/978-3-319-42999-1_8.

[36] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/TMI.2016.2535302.

[37] H.-S. Ham, H.-S. Lee, J.-W. Chae, H. C. Cho, and H.-C. Cho, "Improvement of gastroscopy classification performance through image augmentation using a gradient-weighted class activation map," *IEEE Access*, vol. 10, pp. 99361–99369, 2022, doi: 10.1109/ACCESS.2022.3207839.

[38] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," Apr. 2019, *arXiv:1805.09501*. Accessed: Feb. 2, 2023.

[39] J.-W. Chae and H.-C. Cho, "Enhanced classification of gastric lesions and early gastric cancer diagnosis in gastroscopy using multi-filter AutoAugment," *IEEE Access*, vol. 11, pp. 29391–29399, 2023, doi: 10.1109/ACCESS.2023.3260983.

[40] H.-S. Lee and H.-C. Cho, "Improving classification performance in gastric disease through realistic data augmentation technique based on Poisson blending," *J. Electr. Eng. Technol.*, vol. 18, no. 4, pp. 3127–3134, May 2023, doi: 10.1007/s42835-023-01543-7.

[41] T. Itoh, H. Kawahira, H. Nakashima, and N. Yata, "Deep learning analyzes helicobacter pylori infection by upper gastrointestinal endoscopy images," *Endoscopy Int. Open*, vol. 6, no. 2, pp. E139–E144, Feb. 2018, doi: 10.1055/s-0043-120830.

[42] A. Sams and H. H. Shomee, "GAN-based realistic gastrointestinal polyp image synthesis," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–4, doi: 10.1109/ISBI52829.2022.9761447.

[43] L. A. de Souza, L. A. Passos, R. Mendel, A. Ebigbo, A. Probst, H. Messmann, C. Palm, and J. P. Papa, "Assisting Barrett's esophagus identification using endoscopic data augmentation based on generative adversarial networks," *Comput. Biol. Med.*, vol. 126, Nov. 2020, Art. no. 104029, doi: 10.1016/j.compbiomed.2020.104029.

[44] P. Sasmal, M. K. Bhuyan, S. Sonowal, Y. Iwahori, and K. Kasugai, "Improved endoscopic polyp classification using GAN generated synthetic data augmentation," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 247–251, doi: 10.1109/ASPCON49795.2020.9276732.

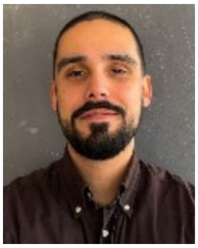
[45] Z.-H. Cui, Q.-Y. Zhang, J.-W. Zhang, X. Sun, Q. Wang, Y. Lei, L. Zang, L. Zhao, and J.-Ji. Yang, "Application of improved mask R-CNN algorithm based on gastroscopic image in detection of early gastric cancer," in *Proc. IEEE 46th Annu. Comput., Softw., Appl. Conf. (COMPSAC)*, H. Leong, S. Sarvestani, Y. Teranishi, A. Cuzzocrea, H. Kashiwazaki, D. Towey, J. Yang, and H. Shahriar, Eds., 2022, pp. 1396–1401, doi: 10.1109/COMPSAC54236.2022.00221.

[46] Y.-J. Kim, H. C. Cho, and H.-C. Cho, "Deep learning-based computer-aided diagnosis system for gastroscopy image classification using synthetic data," *Appl. Sci.*, vol. 11, no. 2, p. 760, Jan. 2021, doi: 10.3390/app11020760.

[47] D. Shen et al., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V* (Lecture Notes in Computer Science), vol. 11768. Cham, Switzerland: Springer, 2019, doi: 10.1007/978-3-030-32254-0.

[48] T. Kanayama, Y. Kurose, K. Tanaka, K. Aida, S. Satoh, M. Kitsuregawa, and T. Harada, "Gastric cancer detection from endoscopic images using synthesis by GAN," in *Medical Image Computing and Computer Assisted Intervention*, in Lecture Notes in Computer Science, vol. 11768, D. Shen, T. Liu, T. Peters, L. Staib, C. Essert, S. Zhou, P. Yap, and A. Khan, Eds. 2019, pp. 530–538, doi: 10.1007/978-3-030-32254-0_59.

BRUNO MAGALHÃES received the bachelor's degree in genetics and biotechnology. He is currently pursuing the master's degree in biomedical engineering with the University of Trás-os-Montes and Alto Douro. He is also a Researcher of the Computer Assisted Gastric Cancer Diagnosis, a project at INESC TEC and the University of Trás-os-Montes and Alto Douro. His research interests include generative models research for dataset augmentation and deep learning methods for detecting gastric lesions.



ALEXANDRE NETO received the bachelor's degree in genetics and biotechnology and the master's degree in bioinformatics and applications in life sciences from the University of Trás-os-Montes and Alto Douro, in March 2021. For the master's thesis dissertation, a project titled "Automatic Glaucoma Screening with Low-Cost Devices" was conducted in which deep learning and computer vision techniques were applied to biomedical images. His Ph.D. degree is focused on deep learning methods for detecting gastric lesions and generating synthetic images for better model generalization. He is currently a Ph.D. Researcher. He has participated in more than ten publications, being the first author in several of them and attending international conferences. He is a part of the Computer Assisted Gastric Cancer Diagnosis Project at INESC TEC and the University of Trás-os-Montes and Alto Douro. His research interests include bioinformatics, computer vision, machine learning, artificial intelligence, and data science.

ANTÓNIO CUNHA is currently a Ph.D. Senior Researcher and an Auxiliary Professor with the Engineering Department, University of Trás-os-Montes and Alto Douro (UTAD). He participated as a member in seven funded research projects. He has been a member with the Centre for Biomedical Engineering Research (C-BER), INESC TEC, since 2015, where one of the final goals is to create a CAD system based on DL approaches to assist clinical doctors in different biomedical image-related diagnoses/screening tasks. Thus, he has plenty of experience in medical imaging analysis and student supervision. In the last five years, he supervised two Ph.D. students and 22 M.Sc. students and published 17 journal articles and 34 Scopus conference papers in this area. His research interests include medical image analysis, bio-image analysis, computer vision, machine learning, and artificial intelligence, particularly in computer-aided diagnosis applied in several imaging modalities, e.g., computed tomography of the lung and endoscopic videos. He is a part of the Organization Committee HCIST-International Conference on Health and Social Care Information Systems and Technologies (2013–2015, 2020–2023), and the Organization Chair (2012) and an Advisory Board (2016–2023). He is a referee for several international journals and conferences and participated yearly in organizing several international scientific events. He is the General Chair of the MobiHealth2022 International Conference.