

# Role of Artificial Intelligence in the Diagnosis of Gastroesophageal Reflux Disease

Review began 05/28/2024  
Review ended 06/02/2024  
Published 06/11/2024

© Copyright 2024  
Kommuru et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Sravani Kommuru <sup>1</sup>, Faith Adekunle <sup>2</sup>, Santiago Niño <sup>3</sup>, Shamsul Arefin <sup>4</sup>, Sai Prudhvi Thalvayapati <sup>5</sup>, Dona Kuriakose <sup>6</sup>, Yasmin Ahmadi <sup>7</sup>, Suprada Vinyak <sup>8</sup>, Zahra Nazir <sup>9</sup>

1. Medical School, Dr. Pinnamaneni Siddhartha Institute of Medical Sciences & Research Foundation, Vijayawada, IND 2. Medical School, American University of the Caribbean, Cupecoy, SXM 3. Surgery, Colegio Mayor de Nuestra Señora del Rosario, Bogota, COL 4. Internal Medicine, Nottingham University Hospitals NHS Trust, Nottingham, GBR 5. Medical School, Government Kilpauk Medical College, Chennai, IND 6. Internal Medicine, Petre Shotadze Tbilisi Medical Academy, Tbilisi, GEO 7. Medical School, Royal College of Surgeons in Ireland - Medical University of Bahrain, Busaiteen, BHR 8. Internal Medicine, Wellmont Health System/Norton Community Hospital, Norton, USA 9. Internal Medicine, Combined Military Hospital, Quetta, PAK

**Corresponding author:** Zahra Nazir, naz\_zahra95@outlook.com

---

## Abstract

Gastroesophageal reflux disease (GERD) is a disorder that usually presents with heartburn. GERD is diagnosed clinically, but most patients are misdiagnosed due to atypical presentations. The increased use of artificial intelligence (AI) in healthcare has provided multiple ways of diagnosing and treating patients accurately. In this review, multiple studies in which AI models were used to diagnose GERD are discussed. According to the studies, using AI models helped to diagnose GERD in patients accurately. AI, although considered one of the most potent emerging aspects of medicine with its accuracy in patient diagnosis, presents limitations of its own, which explains why healthcare providers may hesitate to use AI in patient care. The challenges and limitations should be addressed before AI is fully incorporated into the healthcare system.

---

**Categories:** Gastroenterology, Internal Medicine, Healthcare Technology

**Keywords:** git endoscopy, ai and machine learning, infectious esophagitis, gastroesophageal reflux disease (gerd), artificial intelligence in medicine

## Introduction And Background

Gastroesophageal reflux disease (GERD) stands out as one of the most widespread gastrointestinal conditions worldwide and one of the most diagnosed digestive pathologies in the United States, with a prevalence of more than 20% [1]. According to the 2006 Montreal consensus, GERD is a series of symptoms and complications secondary to regurgitating gastric contents into the esophagus [2]. The clinical presentation of GERD is manifested by typical symptoms such as heartburn, regurgitation, and esophageal chest pain. However, different atypical presentations have been described, including chest pain, dental erosions, chronic cough, laryngitis, or asthma [3].

GERD can be classified into three different phenotypes based on endoscopic and histopathological appearance, including non-erosive reflux disease (NERD), erosive esophagitis (EE), and Barrett's esophagus (BE) [4]. While it is both treatable and preventable, most patients with GERD often do not seek advice from their primary care physician and usually rely on over-the-counter medications. Hence, patients tend to present at later stages of the disease, when complications such as Barrett's esophagus have already set in, warranting superior investigations and treatments, increasing the healthcare burden. GERD could be misdiagnosed as myocardial infarction, which again increases unwanted hospital visits and diagnostic procedures. The scope of artificial intelligence (AI) within GERD lies in developing effective strategies for screening and early diagnosis and its potential complications.

AI uses computer algorithms that imitate essential human functions like learning and problem-solving. Although it has limitations, AI is more efficient in performing healthcare tasks as compared to the traditional system [5]. There have been studies where AI has been used to depict the anatomical locations in endoscopic images to diagnose GERD accurately [6]. AI has also been used in clinical questionnaires regarding GERD, which are used to diagnose GERD promptly and adequately without invasive procedures. With increasing research being conducted on the use of AI in diagnosing GERD, it also presents its limitations; AI requires a vast amount of data from previous patients to be fed into the system in order to produce more reliable results, it lacks real abstract human-like thinking, and most of the systems available are research-based and may suffer from selection bias [6].

With the recent increase in research on the use of AI in diagnosing various GI diseases, healthcare workers would need to further educate themselves on the method of diagnosis, criteria of diagnosis, and advantages and disadvantages of AI use. We reviewed the available studies regarding the use of AI in diagnosing GERD.

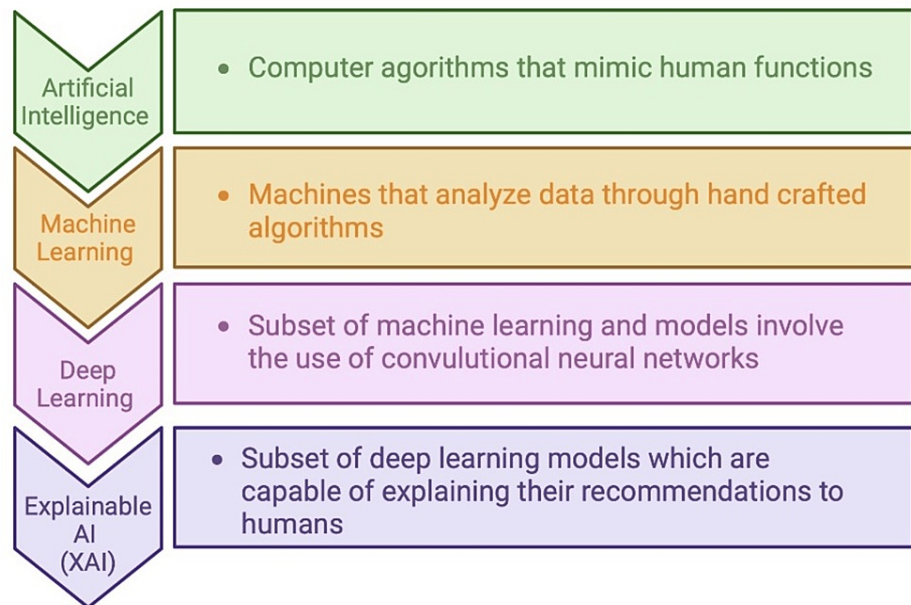
### How to cite this article

Kommuru S, Adekunle F, Niño S, et al. (June 11, 2024) Role of Artificial Intelligence in the Diagnosis of Gastroesophageal Reflux Disease. *Cureus* 16(6): e62206. DOI 10.7759/cureus.62206

## Review

### Technical aspects of AI

AI is a computer algorithm designed to imitate human cognitive processes. Machine learning (ML) and Deep Learning (DL) are the latest advancements in AI. In ML, researchers use their knowledge to identify relevant features in an existing database, which they then use to train a model [5]. The newer and most potential derivative of ML is DL, which autonomously extracts distinctive attributes of input data using artificial neural networks (ANN) [5]. Figure 1 explains AI and its different subsets [7-9].



**FIGURE 1: Artificial intelligence (AI) and its subsets**

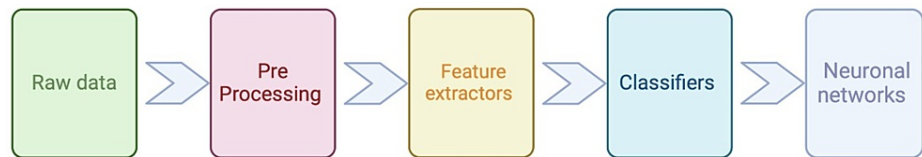
Image Credit: Authors; created using biorender.com

Reference: [10,11]

The components of a DL model include pre-processors, feature extractors, classifiers, and explainable AI (XAI). Few studies may include attention modules to fine-tune the feature extractors. Feature extractors are used to convert raw data into a form that the classifier can process. This feature extraction process not only preserves the original data but also enhances the outcome of the AI model [10,11]. There are three types of feature extractors: (a) supervised models that use labeled data to obtain all the general features of an image, e.g., the edges, angular points, texture, and marginal points of the images, histograms of oriented gradients (HOG), local binary pattern (LBP), scale-invariant feature transform (SIFT), (b) supervised models to obtain the general features of the image using the prior knowledge, and (c) unsupervised feature extraction algorithm that use unlabeled data includes convolutional layers of the convolutional neural network (CNN).

As the information flows through each layer, DL models extract the maximum number of image features from the given input, refining the inferences of the previous layer [12,13]. Classifiers separate the extracted features obtained from the feature extractor into multiple groups including support vector machine (SVM), neural network, gradient boosting classification tree (GB), random forest, K-nearest neighbors (KNN), histogram-based gradient boosting classification tree (HistGB), extreme gradient boosting classification tree (XGB), linear SVM, SVM with radial basis function kernel (RBF), decision tree, Gaussian Naive Bayes, logistic regression, and multilayer perceptron (MLP) [14]. One of the significant concerns of DL models is that the several neuronal layers involved in these models could have explained how the data was interpreted before decision-making. This was considered a black box for DL models [8,9].

Through continuous innovations, a new subset of DL called interpretable DL or XAI came into existence. These models could explain the rationale behind data interpretation and decision-making [7,10]. Biran and Cotton, in their survey on XAI, mentioned three different types of models: (a) interpretable models, (b) prediction interpretation and justification models, and (c) visualization models [10]. All these models are designed to overcome the bias in the CNNs of the DL models. Figure 2 shows the outline of a basic DL model [9].



**FIGURE 2: Outline of a basic DL model**

Image Credit: Authors; Created using biorender.com

Reference: [9]

While AI was initially viewed as a threat to many healthcare jobs, including physicians, it is certainly not [15,16]. Clinical medicine remains an art that depends on the decisions made by licensed practitioners to ensure transparency and maximum patient safety [15]. In today's existing scenario, AI is only being used as a middle ground to find out things that are not easily identifiable by humans. However, the significance of these findings must still be determined by licensed practitioners [17-20].

### Role of AI in diagnosing GERD

One of the initial studies conducted on using AI in the detection of GERD was conducted by Pace et al., who developed the Questionario Italiano Diagnostico (QUID) questionnaire to distinguish erosive esophagitis from non-erosive GERD [21]. The data was explored further using ANN, a neuronal network implemented to diagnose the type of GERD. The results obtained were statically significant ( $p < 0.001$ ) for GERD patients as compared to the controls, but it failed to distinguish erosive from non-erosive esophagitis [21]. In a study conducted by the Mayo Clinic, all patients enrolled in the study had to fill out a questionnaire suggested by the Mayo Clinic, the Gastro-Esophageal Reflux Questionnaire (GERQ) [22]. These patients underwent an EGD and a 24-hour esophageal pH meter. GERD was objectively confirmed in 103, and it was customary in 56 patients. The AI system automatically selected the 45 most relevant variables from the GERQ (collectively referred to as QUID), allowing the CAD tool to predict GERD in all 103 patients with an accuracy of 100% [22].

#### Endoscopy

GERD is typically diagnosed based on a clinician's assessment of symptoms and treated empirically. According to the current guidelines by the American College of Gastroenterology, endoscopy is the first line of investigation in patients with symptoms of GERD after the initial eight-week trial of proton pump inhibitors (PPIs) [23]. The scope of AI was extended into endoscopy in the form of two popular techniques: computer-assisted detection (CADE) and computer-assisted diagnosis (CADx). With CADE, the AI model acts as a "second pair of eyes" for the investigator, and the CADx system is used to classify a lesion based on several morphological parameters and generate probability scores for malignancy or non-malignancy [23,24]. Although these techniques have been widely used in the past by colonoscopists to increase adenoma detection, similar techniques can be applied to upper GI endoscopy to detect GERD and its other potential complications like Barrett's esophagus and esophageal adenocarcinoma. A pioneering AI model for diagnosing GERD was the hierarchical heterogeneous descriptor fusion (HHDF)-SVM developed by Huang et al. [25]. This method utilized conventional endoscopic images for diagnosis. Their method could automatically diagnose GERD and achieve better accuracy than state-of-the-art methods [24,25]. The initial AI models used in diagnosing GERD, i.e., the QUID questionnaire and HHDF-SVM, were limited to differentiating patients with GERD from healthy patients but failed to classify GERD.

#### AI Models Used to Classify GERD

One of the first AI models to classify GERD according to the Los Angeles classification was proposed by Wang et al., a DL model with data augmentation GERD-VGGNet that used convolutional neural networks to classify GERD [26]. They selected 2000 cases, of which only 496 images passed after quality checks. Images from 464 patients were used for model training and validation, and the remaining 32 images were used to evaluate and compare the accuracy of both trainees and the proposed model. Among the feature extractors that were compared in this study to extract the image features, the VGG 16 model showed 100% training accuracy and 89.3 validation accuracy. So, upon combining VGG 16 with classifiers (GERD-VGGNet model, RBF-SVM, Decision Tree, Random Forest, and Adaboost), the GERD-VGGNet and VGG 16 combination model outperformed others. The development set's model demonstrated an overall accuracy of 99.2% (grade A-B), 100% (grade C-D), and 100% (standard group) when employing narrow-band image (NBI) endoscopy. Meanwhile, on the test set, the suggested model achieved an accuracy of 87.9% [26].

Another similar model that aimed to classify GERD using narrow-band imaging (NBI) endoscopic images was proposed by Yen et al. at Chung Shan Medical University Hospital in 2022 [14]. The model was

developed using 671 NBI endoscopy images in the development set and 32 images in the test set. The development set consisted of 244 grade A-B images, 229 grade C-D images, and 198 standard endoscopic images, which were classified based on the Los Angeles classification of GERD. The model combined both DL and ML aspects of AI. GerdNet (feature extractor)-RF(classifier) was developed to process the NBI images. The model performed much more efficiently than the GERD-VGGNet model. The overall accuracy increased from 98.9±0.5% to 99.0±0.1%, and the test accuracy also improved from 78.8±8.5% to 92.5±2.1%.

A study was conducted at the Hanoi Medical University Hospital, Vietnam, to develop a method using data augmentation techniques to classify GERD from endoscopic images. ResNet-50 was used for feature extraction and classification. Two data augmentation techniques, Affine Transformation and Generative Adversarial Network (i.e., StyleGAN2-ADA) on different color models (RGB, HSV), were used to improve the classification of GERD due to which there was a significant improvement from 83.2% to 91.7% in accuracy when classifying GERD using StyleGAN-ADA on RGB channels compared to the original data on the dataset [27].

Another study in Xiangya Hospital of Central South University, Hunan, China, added to the use of AI to diagnose GERD. The main aim of the study was to provide a framework for the diagnosis of Barrett's esophagus and reflux esophagitis in endoscopic images using DL techniques. Experimental results showed that this model achieved a classification recognition accuracy of 93.0%, with macro-precision, macro-recall, and macro-F scores of 93.5%, 92.9%, and 93.2%, respectively [28].

Proceeding with more models and comparison studies, we reviewed a study by Liao et al. aimed to compare the classification performance of the regions of interest (ROI)-based CNN models with the LBP-based SVM classifier, the HOG-based SVM classifier, and the SIFT-based SVM classifier [12]. All these were AI-based models for endoscopy. They used a dataset containing 1394 clinical NF-NBI images from 50 subjects, of which 554 images were from 21 patients (NERD+) and 840 were from 29 healthy controls. They observed that in both the subject-dependent and subject-independent experiments, the ROI-based CNN models achieved a higher mean of 10-fold test accuracy compared to the SVM classifiers, with a 20.5% improvement in the former and a 14.0% improvement in the latter. This underscores the superior classification performance of the ROI-based CNN models over the SVM classifier [12].

More recent additions to the use of AI came to light when a DL model was developed that classified endoscopic images based on the Los Angeles grade [29]. This study compared the performance of XAI models. A total of 2081 endoscopic images were obtained. The compared models are Grad-CAM, information bottlenecks for attribution (IBA), and an intrinsic (Scouter) method. Among the feature extractors and classifiers, dense 121 outperformed the others. So, dense 121 was experimented for comparing the XAI models. In this visualization experiment, the Scouter method and the Grad-CAM had the best w-Precision (weighted precision) and best incremental area under the curve (iAUC), respectively. In terms of the efficiency of the DL model, the classification accuracy in Grade A and Grade C was significantly higher than that of endoscopists. This model showed an accuracy of 86.7% when compared with endoscopists [29].

One of the most recent studies was by Ge et al., which compared the performance of classification by the DL models with that of endoscopists using Hill's grading system [30]. Their study included 3256 samples. The test data set included 333 samples, and the rest of the samples were imparted into training and validation data sets. ResNet-50 and Xception outperformed the other models among the feature extractors and classifiers. Post-feature extraction, the data was subjected to early dynamic attention (ED-A), which primarily worked on the color intensity of the RGB image, and SE attention, which worked on the spatial orientation of the image. Grad-CAM algorithm was used as the XAI technique for visual interpretation. The experiment revealed that attention modules help the model to localize ROI, enhancing classifying abilities. ResNet-50 combined with attention modules performed the best, as is evident from the area under the curve (AUC) reaching 0.989 and an accuracy of 93.39% [30].

A significant long-term complication of chronic untreated GERD is a premalignant condition called Barrett's esophagus. Early detection of neoplasia in these patients can result in improved treatment outcomes and a reduction in healthcare costs. In 2019, de Groof et al. developed the CADe model to differentiate neoplastic and non-dysplastic lesions in patients with Barrett's esophagus [31]. Using this model, they evaluated five independent endoscopy data sets. They found that the CADe system achieved superior results with higher accuracy than 53 nonexpert endoscopists. The model classified images as containing neoplasms or nondysplastic Barrett's esophagus with 89% accuracy, 90% sensitivity, and 88% specificity [31]. In 2020, another DL model was proposed by Hashimoto et al. [32] for early visual detection of esophageal neoplasia in patients. The evaluation encompassed 455 test images, with the detection of early neoplasia yielding a sensitivity of 96.4%, specificity of 94.2%, and accuracy of 95.4%.

One of the main challenges of upper digestive endoscopy is unseen areas which are usually missed by trainee endoscopists referred to as blind spots. Wu et al. proposed a DL model named ENDOANGEL with an aim to decrease the number of blind spots. This AI model is currently approved by the FDA [33]. Although initially used to detect early gastric cancer as a secondary output, this model could be extended to detecting esophageal cancer. In their study, 1050 patients were randomized into the ENDOANGEL group (n=498) and a

control group (n=504), and it was found that the ENDOANGEL group had fewer blind spots (mean 5.38, SD 4.32 vs. 9.82, SD 4.98;  $P < 0.001$ ).

On a side note, de novo GERD can occur after sleeve gastrectomy, a bariatric procedure performed in morbidly obese individuals. A meta-analysis published in 2021 that included 46 studies totaling 10,718 patients stated that the increase of postoperative GERD after sleeve gastrectomy was 19% and de novo reflux was 23% [34]. An ML model was proposed in 2022 to detect the probability of postoperative GERD in sleeve gastrectomy patients by utilizing the SPSS software (IBM Corp., Armonk, New York, United States [35]. The study revealed that the model had excellent accuracy with an AUC of 0.93, yet a moderate sensitivity of 79.2% and specificity of 86.1%.

A lesser-known extraesophageal manifestation of gastroesophageal reflux disease, laryngopharyngeal reflux (LPR) occurs when stomach contents flow backward into the laryngopharynx. This is associated with hoarseness, vocal fatigue, excessive throat clearing, globus pharyngeus, chronic cough, postnasal drip, and dysphagia [36]. Laryngoscopic findings may include erythema, edema, ventricular obliteration, postcricoid hyperplasia, and pseudosulcus change [37]. A unique study, which deviated from the routine endoscopic investigations and impedance studies to diagnose GERD, was conducted at Mayo Clinic, Rochester, which aimed to use voice recordings to distinguish Barrett's esophagus from GERD using ML algorithms [38]. They used a sample of 245 patients (98 vocally normal, 34 GERD patients, and 35 with Barrett's esophagus). Patients categorized as GERD+ exhibited erosive esophagitis (Los Angeles grade B-D), peptic stricture, or acid exposure time exceeding 6%. Barrett's esophagus was delineated as columnar mucosa invasion surpassing 1 cm, accompanied by confirmed specialized intestinal metaplasia. Patients lacking these criteria were classified as GERD-. Voice recording consisted of a six-sentence standard script read over 25-45 seconds from patients undergoing esophagogastroduodenoscopy (EGD) and ambulatory pH monitoring studies due to clinical indications. Random forest models were used as feature extractors and classifiers. They used an F1 (0-100) score to interpret the results of their study and found that the proposed AI had an excellent ability to distinguish Barrett's esophagus and GERD from normal voice, with scores ranging from 60 to 70 [38].

## Impedance studies

The gold standard for determining whether GERD is pathological is pH impedance testing; however, manual analysis takes a long time. AI can streamline the interpretation of pH-impedance studies [38]. Normal individuals, patients with previous GERD, and patients with esophageal functional disorders have higher baseline impedance values than those with erosive esophagitis and NERD, and low mean nocturnal baseline impedance can independently predict the efficacy of anti-reflux medical or surgical treatment [39]. Acid exposure time (AET) is the most critical parameter obtained from impedance monitoring in diagnosing GERD. According to the Lyon Consensus, AET < 4% is physiological for GERD, and AET > 6% is pathological. However, AET between 4% and 6% is determined to be inconclusive and warrants further evaluation. The other metrics that should be evaluated in such cases include several reflux episodes, mean nocturnal baseline impedance (MNBI), and reflux-induced peristaltic wave (PSPW). AI advances have been made to study the number of reflux episodes, MNBI, and PSPW index, but no significant AI application was developed to measure the AET [40]. This might be one of the most important missing links in diagnosing GERD and a potential future research advancement. We reviewed the existing literature and the models proposed to enhance the detection and interpretation of baseline impedance values.

A study was undertaken to develop an AI model across three tertiary medical centers in Taiwan [36]. The aim was to assess the performance of a cascade multivariate long short-term memory fully convolutional network (MLSTM-FCN) model, which outperformed other models and showed an accuracy of 0.917, sensitivity of 0.818, and specificity of 1.00 when compared with experts in the field [36].

A new research project, carried out jointly by the University of Wisconsin-Madison and Peking University Third Hospital, employed end-to-end DL techniques for diagnosing LPR [41]. According to the pH monitoring results, 37 subjects were diagnosed with LPR, while 38 subjects showed symptoms suggestive of LPR but tested negative. The control cohort comprised 77 individuals who were deemed healthy. Proximal acid episodes, proximal acid exposure time, impedance detected proximal acid exposure, and the reflux score index (RSI) were used to categorize data. ResNet-18, ResNet50, and ResNet101 were feature extractors and classifiers. The study discovered that ResNet50 integrated with the CBAM module achieved the highest classification accuracy, reaching 73.4%. Meanwhile, ResNet18 equipped with the CBAM module demonstrated the highest AUC value at 73.9%, outperforming other models [41].

A study conducted in Taiwan utilized 7939 impedance events from 106 patients who underwent pH-impedance studies to establish an AI model [42]. The data for this study was obtained from patients aged 20-65 years with typical GERD symptoms (heartburn and acid regurgitation) for at least three months and negative endoscopy, who underwent 24-hour impedance-pH monitoring without PPI therapy for at least seven days. Time intervals of 30 seconds before the event start time and 90 seconds after the event start time were considered. The raw waveform signals were then transformed into RGB-colored images for a total time of 120 seconds. The raw data was converted into grayscale images for input into the deep residual learning model for image recognition through attention. The feature extractor cum classifier (ResNet18) was

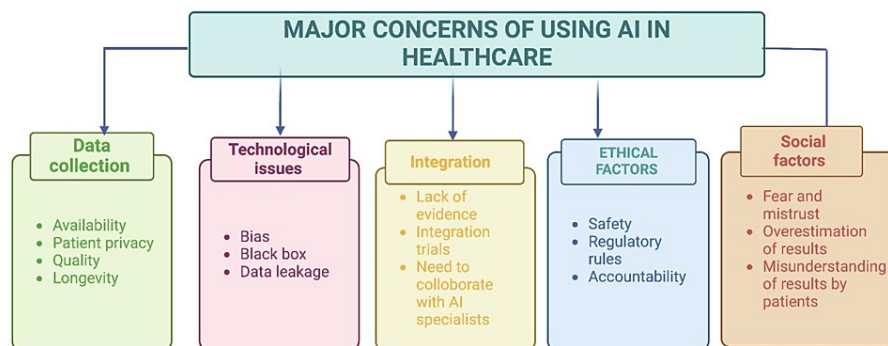
used to identify reflux episodes. It demonstrated that the accuracy of the AI model in recognizing reflux episodes was up to 87% [42].

In conclusion, the ML system is 68.7% sensitive to identifying reflux episodes and 80.8% specific [43].

### Limitations of AI

AI, while a powerful tool in medicine, is not without its limitations. One of the most significant challenges is the economic and regulatory burden it can place on organizations. The approval of safety, efficacy, and transparency regulations is a prerequisite for the utilization of AI technology in clinical settings [44]. However, these systems require constant updating and upkeep, including the integration of large amounts of data, software updates, and hardware repairs. Such tasks necessitate human resources and funding support, which can pose significant economic and regulatory challenges [45–48]. The high cost of developing, validating, and integrating AI into healthcare systems can be a deterrent, and stringent regulatory standards, while necessary for patient safety, can also slow down the process.

Additionally, the language used in AI tools can be a limitation. Tools developed in one language may not immediately apply in different linguistic or cultural contexts, necessitating further adaptation and resources to ensure broad usability [47]. Another challenge in AI-assisted medicine is creating a communication channel between the model and physicians. This involves translating digital information into medical language, aiding diagnosis and treatment [24]. Figures 3 briefly summarize the disadvantages of integrating AI into clinical practice [45]. At the same time, AI can reduce backlogs and mundane tasks for healthcare workers and cut costs apart from several other pathbreaking innovations [49] and there needs to be large-scale research to make it more accessible to the end users.



**FIGURE 3: Concerns of AI in healthcare**

Image Credit: Authors; Created using biorender.com

Source reference: [45]

AI: artificial intelligence

### Conclusions

Integrating AI into the diagnosis process of GERD offers enormous potential to transform clinical practices and patient outcomes positively. AI technologies have found their place in image analysis, interpretation, and classification of GERD, which can give valuable insights to help in early detection, accurate diagnosis, and tailored treatment. However, it is crucial to understand the limitations and challenges that may arise. Addressing these challenges through continued research, development, and training is essential for harnessing the full capabilities of AI to improve patient outcomes in GERD management.

### Additional Information

#### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** Zahra Nazir, Sravani Kommuru, Suprada Vinyak, Shamsul Arefin, Faith Adekunle, Santiago Niño, Sai Prudhvi Thalvayapati, Dona Kuriakose, Yasmin Ahmadi

**Acquisition, analysis, or interpretation of data:** Zahra Nazir, Sravani Kommuru, Suprada Vinyak,

Shamsul Arefin, Faith Adekunle, Santiago Niño , Sai Prudhvi Thalvayapati, Dona Kuriakose, Yasmin Ahmadi

**Drafting of the manuscript:** Zahra Nazir, Sravani Kommuru, Suprada Vinyak, Shamsul Arefin, Faith Adekunle, Santiago Niño , Sai Prudhvi Thalvayapati, Dona Kuriakose, Yasmin Ahmadi

**Critical review of the manuscript for important intellectual content:** Zahra Nazir, Sravani Kommuru, Suprada Vinyak, Shamsul Arefin, Faith Adekunle, Santiago Niño , Sai Prudhvi Thalvayapati, Dona Kuriakose, Yasmin Ahmadi

**Supervision:** Zahra Nazir

## Disclosures

**Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## References

1. Antunes C, Curtis SA, Aleem A: Gastroesophageal reflux disease . Statpearls [Internet]. StatPearls Publishing, Treasure Island (FL); 2023.
2. Vakil N, van Zanten SV, Kahrilas P, Dent J, Jones R: The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol.* 2006, 101:1900-20; quiz 1945. [10.1111/j.1572-0241.2006.00630.x](https://doi.org/10.1111/j.1572-0241.2006.00630.x)
3. Zuhair V, Babar A, Ali R, et al.: Exploring the impact of artificial intelligence on global health and enhancing healthcare in developing nations. *J Prim Care Community Health.* 2024, 15:21501319241245847. [10.1177/21501319241245847](https://doi.org/10.1177/21501319241245847)
4. Fass R, Ofman JJ: Gastroesophageal reflux disease--should we adopt a new conceptual framework? . *Am J Gastroenterol.* 2002, 97:1901-9.
5. Visaggi P, de Bortoli N, Barberio B, et al.: Artificial intelligence in the diagnosis of upper gastrointestinal diseases. *J Clin Gastroenterol.* 2022, 56:23-35. [10.1097/MCG.0000000000001629](https://doi.org/10.1097/MCG.0000000000001629)
6. Parasher G, Wong M, Rawat M: Evolving role of artificial intelligence in gastrointestinal endoscopy . *World J Gastroenterol.* 2020, 26:7287-98. [10.3748/wjg.v26.i46.7287](https://doi.org/10.3748/wjg.v26.i46.7287)
7. Clarrett DM, Hachem C: Gastroesophageal reflux disease (GERD). *Mo Med.* 2018, 115:214-8.
8. Lamy JB, Sekar B, Guezennec G, Bouaud J, Séroussi B: Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach. *Artif Intell Med.* 2019, 94:42-53. [10.1016/j.artmed.2019.01.001](https://doi.org/10.1016/j.artmed.2019.01.001)
9. van der Velden BH, Kuijf HJ, Gilhuijs KG, Viergever MA: Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal.* 2022, 79:102470. [10.1016/j.media.2022.102470](https://doi.org/10.1016/j.media.2022.102470)
10. Biran O, Cotton C: Explanation and Justification in Machine Learning: A Survey . Columbia University, New York, NY; 2017.
11. MathWorks: Feature extraction. (2024). Accessed: 2024: <https://www.mathworks.com/discovery/feature-extraction.html>.
12. Liao J, Lam HK, Jia G, et al.: A case study on computer-aided diagnosis of nonerosive reflux disease using deep learning techniques. *Neurocomputing.* 2021, 28:149-66. [10.1016/j.neucom.2021.02.049](https://doi.org/10.1016/j.neucom.2021.02.049)
13. Zhou J, Hu N, Huang ZY, Song B, Wu CC, Zeng FX, Wu M: Application of artificial intelligence in gastrointestinal disease: a narrative review. *Ann Transl Med.* 2021, 9:1188. [10.21037/atm-21-3001](https://doi.org/10.21037/atm-21-3001)
14. Yen HH, Tsai HY, Wang CC, Tsai MC, Tseng MH: An improved endoscopic automatic classification model for gastroesophageal reflux disease using deep learning integrated machine learning. *Diagnostics (Basel).* 2022, 12:2827. [10.3390/diagnostics12112827](https://doi.org/10.3390/diagnostics12112827)
15. Gniadek T, Kang J, Thepatee T, Krive J: Framework for classifying explainable artificial intelligence (XAI) algorithms in clinical medicine. *Online J Public Health Inform.* 2023, 15:e50934. [10.2196/50934](https://doi.org/10.2196/50934)
16. Holzinger A: From machine learning to explainable AI. *IEEE Xplore.* 2018, 55-66. [10.1109/DISA.2018.8490530](https://doi.org/10.1109/DISA.2018.8490530)
17. Miller DD, Brown EW: Artificial intelligence in medical practice: the question to the answer? . *Am J Med.* 2018, 131:129-33. [10.1016/j.amjmed.2017.10.035](https://doi.org/10.1016/j.amjmed.2017.10.035)
18. Li D, Kulasegaram K, Hodges BD: Why we needn't fear the machines: opportunities for medicine in a machine learning world. *Acad Med.* 2019, 94:623-5. [10.1097/ACM.0000000000002661](https://doi.org/10.1097/ACM.0000000000002661)
19. Tacchella A, Romano S, Ferraldeschi M, Salvetti M, Zaccaria A, Crisanti A, Grassi F: Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle study. *F1000Res.* 2017, 6:2172. [10.12688/f1000research.13114.2](https://doi.org/10.12688/f1000research.13114.2)
20. Mayo RC, Leung J: Artificial intelligence and deep learning - radiology's next frontier? . *Clin Imaging.* 2018, 49:87-8. [10.1016/j.clinimag.2017.11.007](https://doi.org/10.1016/j.clinimag.2017.11.007)
21. Pace F, Riegler G, de Leone A, Pace M, Cestari R, Dominici P, Grossi E: Is it possible to clinically differentiate erosive from nonerosive reflux disease patients? A study using an artificial neural networks-assisted algorithm. *Eur J Gastroenterol Hepatol.* 2010, 22:1163-8. [10.1097/MEG.0b013e32833a88b8](https://doi.org/10.1097/MEG.0b013e32833a88b8)
22. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR: A new questionnaire for gastroesophageal reflux disease . *Mayo Clin Proc.* 1994, 69:539-47. [10.1016/s0025-6196\(12\)62245-9](https://doi.org/10.1016/s0025-6196(12)62245-9)
23. Katz PO, Dunbar KB, Schnoll-Sussman FH, Greer KB, Yadlapati R, Spechler SJ: ACG clinical guideline for the diagnosis and management of gastroesophageal reflux disease. *Am J Gastroenterol.* 2022, 117:27-56.

- [10.14509/ajg.0000000000001538](https://doi.org/10.14509/ajg.0000000000001538)
24. Vulpoi RA, Luca M, Ciobanu A, Olteanu A, Barboi OB, Drug VL: Artificial intelligence in digestive endoscopy-where are we and where are we going?. *Diagnostics (Basel)*. 2022, 12:927. [10.3390/diagnostics12040927](https://doi.org/10.3390/diagnostics12040927)
  25. Huang CR, Chen YT, Chen WY, Cheng HC, Sheu BS: Gastroesophageal reflux disease diagnosis using hierarchical heterogeneous descriptor fusion support vector machine. *IEEE Trans Biomed Eng*. 2016, 63:588-99. [10.1109/TBME.2015.2466460](https://doi.org/10.1109/TBME.2015.2466460)
  26. Wang CC, Chiu YC, Chen WL, Yang TW, Tsai MC, Tseng MH: A deep learning model for classification of endoscopic gastroesophageal reflux disease. *Int J Environ Res Public Health*. 2021, 18:2428. [10.3390/ijerph18052428](https://doi.org/10.3390/ijerph18052428)
  27. Nguyen PT, Tran TH, Dao VH, et al.: Improving gastroesophageal reflux diseases classification diagnosis from endoscopic images using StyleGAN2-ADA. *Artificial Intelligence in Data and Big Data Processing [Proceedings of ICABDE 2021]*. Dang NH, Zhang YD, Tavares JM, Chen BH (ed): Springer International Publishing, Cham; 2022. 381-93. [10.1007/978-3-030-97610-1\\_30](https://doi.org/10.1007/978-3-030-97610-1_30)
  28. Hou M, Wang J, Liu T, et al.: A graph-optimized deep learning framework for recognition of Barrett's esophagus and reflux esophagitis. *Multimedia tools and applications*. *Multimed Tools Appl*. 2024, [10.1007/s11042-024-18910-9](https://doi.org/10.1007/s11042-024-18910-9)
  29. 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. [10.1080/00365521.2022.2163185](https://doi.org/10.1080/00365521.2022.2163185)
  30. Ge Z, Fang Y, Chang J, et al.: Using deep learning to assess the function of gastroesophageal flap valve according to the Hill classification system. *Ann Med*. 2023, 55:2279239. [10.1080/07853890.2023.2279239](https://doi.org/10.1080/07853890.2023.2279239)
  31. de Groof AJ, Struyvenberg MR, van der Putten J, et al.: Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology*. 2020, 158:915-29.e4. [10.1053/j.gastro.2019.11.030](https://doi.org/10.1053/j.gastro.2019.11.030)
  32. Hashimoto R, Requa J, Dao T, et al.: Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc*. 2020, 91:1264-71.e1. [10.1016/j.gie.2019.12.049](https://doi.org/10.1016/j.gie.2019.12.049)
  33. Wu L, He X, Liu M, et al.: Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. *Endoscopy*. 2021, 53:1199-207. [10.1055/a-1350-5583](https://doi.org/10.1055/a-1350-5583)
  34. Yeung KT, Penney N, Ashrafian L, Darzi A, Ashrafian H: Does sleeve gastrectomy expose the distal esophagus to severe reflux?: a systematic review and meta-analysis. *Ann Surg*. 2020, 271:257-65. [10.1097/SLA.0000000000003275](https://doi.org/10.1097/SLA.0000000000003275)
  35. Emile SH, Ghareeb W, Elfeki H, El Sorogy M, Fouad A, Elrefai M: Development and validation of an artificial intelligence-based model to predict gastroesophageal reflux disease after sleeve gastrectomy. *Obes Surg*. 2022, 32:2537-47. [10.1007/s11695-022-06112-x](https://doi.org/10.1007/s11695-022-06112-x)
  36. Fu J, Lee PH, Wang CC, et al.: A cascade deep learning model for diagnosing pharyngeal acid reflux episodes using hypopharyngeal multichannel intraluminal Impedance-pH signals. *Intell.-Based Med*. 2023, 8:100131. [10.1016/j.ibmed.2023.100131](https://doi.org/10.1016/j.ibmed.2023.100131)
  37. Patel DA, Blanco M, Vaezi MF: Laryngopharyngeal reflux and functional laryngeal disorder: perspective and common practice of the general gastroenterologist. *Gastroenterol Hepatol (N Y)*. 2018, 14:512-20.
  38. Kamboj AK, Yarlagadda M, Pietrowicz M, et al.: S402 voice enabled artificial intelligence for detection of pathologic gastroesophageal reflux disease and Barrett's esophagus. *Am J Gastroenterol*. 2022, 117:281-2. [10.14509/01.ajg.0000858248.86805.d7](https://doi.org/10.14509/01.ajg.0000858248.86805.d7)
  39. Rogers B, Samanta S, Ghobadi K, et al.: Artificial intelligence automates and augments baseline impedance measurements from pH-impedance studies in gastroesophageal reflux disease. *J Gastroenterol*. 2021, 56:34-41. [10.1007/s00535-020-01743-2](https://doi.org/10.1007/s00535-020-01743-2)
  40. Wong MW, Rogers BD, Liu MX, et al.: Application of artificial intelligence in measuring novel pH-impedance metrics for optimal diagnosis of GERD. *Diagnostics (Basel)*. 2023, 13:960. [10.3390/diagnostics13050960](https://doi.org/10.3390/diagnostics13050960)
  41. Ye G, Du C, Lin T, et al.: Deep Learning for Laryngopharyngeal Reflux Diagnosis. *Appl Sci*. 2021, 11:4753. [10.3390/app11114753](https://doi.org/10.3390/app11114753)
  42. Wong MW, Liu MX, Lei WY, et al.: Artificial intelligence facilitates measuring reflux episodes and postreflux swallow-induced peristaltic wave index from impedance-pH studies in patients with reflux disease. *Neurogastroenterol Motil*. 2023, 35:e14506. [10.1111/nmo.14506](https://doi.org/10.1111/nmo.14506)
  43. Zhou MJ, Zikos T, Goel K, et al.: Development and validation of a machine learning system to identify reflux events in esophageal 24-hour pH/impedance studies. *Clin Transl Gastroenterol*. 2023, 14:e00634. [10.14509/ctg.0000000000000634](https://doi.org/10.14509/ctg.0000000000000634)
  44. Charow R, Jeyakumar T, Younus S, et al.: Artificial intelligence education programs for health care professionals: scoping review. *JMIR Med Educ*. 2021, 7:e31043. [10.2196/31043](https://doi.org/10.2196/31043)
  45. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K: The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019, 25:30-6. [10.1038/s41591-018-0307-0](https://doi.org/10.1038/s41591-018-0307-0)
  46. Data Science Central: Challenges to successful AI implementation in healthcare . (2022). Accessed: October 30, 2022: <https://www.datasciencecentral.com/challenges-to-successful-ai-implementation-in-healthcare/>.
  47. Petersson L, Larsson I, Nygren JM, et al.: Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Serv Res*. 2022, 22:850. [10.1186/s12913-022-08215-8](https://doi.org/10.1186/s12913-022-08215-8)
  48. Aung YY, Wong DC, Ting DS: The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *Br Med Bull*. 2021, 139:4-15. [10.1093/bmb/ldab016](https://doi.org/10.1093/bmb/ldab016)
  49. Deloitte Insights: AI-augmented government. Using cognitive technologies to redesign public sector work . (2017). Accessed: 2024: <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/artificial-intelligence-government.html>.