Acta Scientific Gastrointestinal Disorders (ISSN: 2582-1091)

Volume 8 Issue 1 January 2025

Research Article

Artificial Intelligence in Hill Classification of Hiatal Hernia

Andreas Edenberg*

Neo Clinic, Oslo, Norway

***Corresponding Author:** Andreas Edenberg, Neo Clinic, Oslo, Norway.

Received: November 14, 2024

Published: Dcember 03, 2024

Abstract

Artificial intelligence has made many headlines in the past 12 months, spearheaded by the capabilities demonstrated with generative AI by ChatGPT. Within almost every industry and every branch, there is an ever-rising optimism that AI can solve everything ranging from climate problems to food shortages. Namely in medicine, the most promising uses of AI so far lie within diagnostics utilizing image analysis. To utilize the machine learning abilities and capabilities of AWS' Sage Maker algorithm to improve classification and diagnosis of hiatal hernia. Machine learning prediction models from AWS were used to classify and diagnose hiatal hernia using the Hill classification. The dataset model was previously trained with Sage Maker to achieve high accuracy and sensitivity for analyzing the researched dataset. It was acquired by performing gastroscopic examinations in the Neo Clinic Oslo in 2022. All patients received primary gastroesophageal reflux disease (GERD) evaluation. During the study period, 982 patients underwent gastroscopic examinations. In total, 112 Hill I, 70 Hill II, 35 Hill III and 12 Hill IV hernias were identified and classified. The Hill I hernia was the most commonly found type of hill hernia. The present small, single-center proof-of-concept study shows once more that AI has great potential in aiding modern medicine. Machine learning derived analytic models were able to accurately detect and classify present hill hernias out of many endoscopic findings. Harnessing this ability and expanding it to most other diagnostics using image analysis can further improve adequate diagnosis and classification of certain conditions.

Keywords: Hiatal Hernia; Hill Hernia; Gastroscopy; Ai Machine Learning

Introduction

The protrusion of abdominal contents through the hiatus of the diaphragm into the mediastinum is a common disorder known as hiatal hernia. This condition is usually characterized by involvement of the gastroesophageal junction (GEJ) and a possible inclusion of any other abdominal structure. The extent of which either structure is involved as well as the possibility of other abdominal organs' herniation through the hiatus has resulted in the anatomical classification from types I through IV. To make a statement about the competence of the gastroesophageal closing mechanism, Hill et al. (1996) introduced the classification of hiatal hernias into Hill I through Hill IV [1]. The Hill classification describes the constitution of the gastroesophageal valve (while an endoscope is inserted), ranging from a wall-like and tightly closed valve (Hill 1) up to a basically non-existent valve, leaving the GEJ wide open (Hill 4). Alongside the aforementioned "type" classification, both "Hill" and "type" classifications tend to be vaguely (if at all) classified in diagnoses [2].

Most recently the potential of AI tools was demonstrated by generative AI, i.e. ChatGPT [3-5]. With AI tools on the rise, every industry is optimistic that AI has a solution for most problems they encounter. In life sciences and modern medicine, the most promising uses lie with image analysis [6-8]. Other than image analysis, AI has further enormous potential to support clinical routine workflow and computer-aided diagnosis (CAD) in gastroenterology in general. Outside of the current main application of AI in endoscopy, more recent research showed encouraging results in other upper gastrointestinal settings, such as eosinophilic esophagitis, gastroesophageal reflux disease and motility disorders [9]. Accordingly, AI and CAD systems will most likely be increasingly utilised in daily clinical practice and therefore a basic understanding will be required among physicians.

While there are solutions integrated into modern hardware for automated polyp detection readily available through Olympus and Fujifilm, which already increase the quality of diagnostic endoscopies, even more potential for enhancement comes with machine learning.

To improve detection and classification of any number of endoscopic findings, the present, small proof-of-concept study utilized the machine learning services hosted by Amazon Web Services (AWS). Machine learning is a new field of computer science, offering a multitude of algorithms capable of "learning" a relation between input and output variables. This learning process takes place using training samples and optimizing a prediction model [10,11]. Using fast and reliable AI tools offers the potential of obtaining detailed and correct diagnoses in a moment's notice – especially in the field of image diagnostics. Feeding those algorithms with big amounts of data should make these type of diagnoses robust and valid alternatives as compared to man-made diagnoses, especially in cancer-specific predictive modellings as well as diagnosis prediction in Alzheimer disease [6-8].

Methods

Patients underwent endoscopic examination assessing the axial length of hiatus hernia as well as the gastroesophageal flap valve using the Hill classification. No primary gastroesophageal reflux disease (GERD) evaluation was performed, although around 50% of patients were reported to actively take proton pump inhibitors. No esophagitis assessment or columnar metaplasia assessment took place. All the 982 gastroscopies were performed by the author of this research paper. The gastroscopic examinations took place in 2022.

All patients examined were older than 18 years and had not undergone any previous hiatus surgeries. Median age of patients was 44 years. Some subjects had to be excluded for technical reasons, i.e. the performing specialist was unable to take an adequate picture during the endoscopic examination. The demographic distribution of the patient population was relatively evenly spread (Table 1).

All of the gastroscopic images were taken with a modern, AI-supported Olympus EVIS X1 CV-1500 endoscopy system. A GIF-1100 EVIS X1 video gastroscope was connected to the endoscopy system and used for image acquisition. None of the included AI-tools were used, as they are pretrained only for polyp detection. The captured images were then uploaded to AWS and previously defined, custom labels, were fitted to all images to identify Hill I through Hill IV hiatal hernias. These custom labels were acquired by initial training of the AWS machine learning model.

Results

The rapidly increasing opportunities offered by AI tools and machine learning algorithms were effectively used in identifying the different types of Hill hernias. In 2022, 982 patients underwent gastroscopic inspection. Neither GERD evaluation, esophagitis assessment nor columnar metaplasia assessments were performed, as previously described. Out of the 982 examined patients, 660 (67.2%) were male and 322 (32.7%) were female. 520 patients were older than 50 years (52.9%) while the remaining 462 patients were younger than 50 years (47%).

Characteristics	n	%
Male	660	67.2
Female	322	32.7
Age < 50	520	52.9
Age > 50	462	47.0

Table 1: Demographics of the patient population. 67.2% were male, 32.7% female. 52.9% were younger than 50 years, 47.0% were older than 50 years.

The score that is used defines an average model performance of the entire test dataset is called the "F1 Score" and is defined by measuring different metric as follows

It is used for binary classification traditionally referred to as positive and negative. Precision measures the quality of the prediction when it predicts the positive class. Recall (or sensitivity in some cases) measures how completely a model predicts the actual class members in a dataset. Combined, these measurements make up the F1 score, which is the harmonic mean of the precision and recall. Explained in another way: The F1 score determines how accurately the model can recognize the different classes. An F1 score of 1 would indicate a 100% success rate, but this is however unlikely. In image recognition a F1 score above 0.8 would be considered fairly accurate.

In total, 112 Hill I, 70 Hill II, 35 Hill III and 12 Hill IV hernias were identified and classified, as can be seen from figure 1A. In unison with the Hill I label being the most extensively trained dataset (129 test images were used), the Hill I hernia was the most commonly found type of hill hernia. Hill I hiatal hernias are by far the most common type – the major significance of this type of hernias is in their association with reflux disease. Hill I hiatal hernias are also the most difficult to objectively define and ever since a major focus of controversy in diagnosis and diagnostics [12]. The proper classification of hiatal hernia is crucial for treatment options.

The dataset model was trained on AWS using Sage Maker for 1.167 hours and achieved a F1 score of 0.825, resulting from an average precision of 0.818 and an overall recall of 0.839. F1 scores for the individual labels/Hill hernias were 0.833 for Hill I, 0.800 for Hill II, 0.667 for Hill III, 0.833 for Hill IV. According to each individual F1 score, the label performance for Hill III hernias should be slightly worse than for the anatomically rather similar Hill II hernia. For training the dataset on AWS Sage Maker, 129 test images for Hill I, 65 for Hill II, 24 for Hill III and 8 for Hill IV were taken and utilised.

Label	F1 Score	Test Images	Precision	Recall	Threshold
Hill I	0.833	129	0.750	0.983	0.359
Hill II	0.800	65	0.086	0.750	0.719
Hill III	0.667	24	0.667	0.667	0.175
Hill IV	0.833	8	0.833	0.667	0.112

Table 2: Label performance for each individual Hill type hernia during dataset model training. Corresponding to Hill I hernias being the most common hernias, 129 test images were fed to the AWS Sage Maker algorithm.

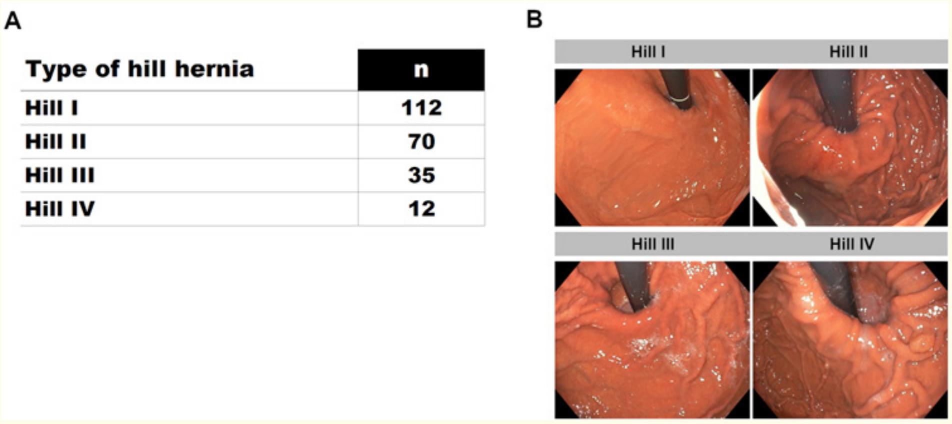

Figure 1: Hill hernias identified in this project. 112 Hill I hernias, 70 Hill II hernias, 35 Hill III hernias and 12 Hill IV hernias were identified (A). Representative images of Hill I-IV hernias were acquired during gastroscopic examinations (B).

Discussion

Most hiatal hernias are found incidentally and are usually discovered during routine chest radiographs, computed tomographies or routine gastroscopies. The aim of the study was not to debate the optimal hernia classification but rather a proof of concept demonstrating how a relatively small dataset can be trained to recognize a condition.

This presence of hiatus hernia is rarely categorized in size when they are found, and this is an area where automated AI aided classficiation can provide useful. Not knowing the grade of hernia can potentially have implications later in those cases where a patient develops further symptoms and is considered for surgery.

Several classification systems have been proposed for the gastroesophageal junction but we have emphasized the Hill classification here. There is evidence to suggest that this is the most practical classification to assess the reflux barrier [13], in part due to easily differentiated classes visually. The classification consist of grade I to grade IV where the degree of fold closure around the scope provide a fast and easy way to separate smaller hernia from the larger. Using our algorithm instant recognition and classification of the type of hiatal hernia, if present, can ensure a much more adequate therapy down the road. Determining the appropriate course of treatment for this condition can be complex, even with proper classification. While smaller hernias or asymptomatic ones can be managed through conservative therapy with medications and without a significant impact on patients' lifestyles, symptomatic and/or larger hernias may require surgical intervention. Assessing the necessity of a surgical intervention involves careful evaluation of individual symptoms, the size of the hernia, possible complications and the overall health status of the patient. Postoperative complications such as infection, recurring hernias, difficulties while swallowing or injury to surrounding structures can be avoided if no surgical intervention is needed. An early diagnosis may negate the need for a surgical intervention, which is especially desirable for patients who are not suitable for surgery due to existing health conditions, progressed age or comorbidities.

As expected, the number of smaller hernias and early-stage hernias was much higher compared to more progressed ones (Hill I: 112, Hill II: 70, Hill III: 35, Hill IV: 12; see fig. 1B). In retrospective, this finding proves the utilized machine learning model to be much more accurate and sensitive than previously anticipated. There was a discrepancy in the F1 score of the Hill IV hernia being higher than for Hill 3 where one would expect lower F1 score when n is lower. One possible explanation for this is that Hill IV stand out more, but probably would be lower with a larger dataset. Being able to classify and identify small and early-stage hernias has the potential to guarantee early and cushioning therapy. Especially Hill I hiatal hernias are difficult to objectively define, which has led to major controversies in diagnosis and diagnostics [12]. This type of hiatal hernia makes up more than 95% of all identified hernias – the major problematic of the most common type of is it being asymptomatic, thus mostly being discovered through routine checks or simply after the patient experiences symptoms.

The effectiveness of AI in identifying hernias during endoscopic examinations depends on different factors such as the advancement and sophistication of the utilized algorithms, the quality of training data and training itself as well as the integration of such technologies into daily clinical practice and the specialist's knowledge about them. While AI showed promising results in previous research settings and clinical applications such as cancer-specific predictive modeling and diagnosis prediction of Alzheimer disease utilizing deep learning algorithms [7,8], the widespread incorporation into daily clinical practice requires ongoing validation through large-scale studies and supportive regulatory in order to ensure and prove its effectiveness, reliability and safety.

Conclusion

This small proof of concept study shows how a relatively small dataset can be used to automate classification of a condition using artificial intelligence. This opens up the possibility of future off-the-shelf solutions where clinicians can train dataset on the fly to recognize various conditions in their practice.

Consent

No informed consent was necessary as the study did not use biological material and did not in any way involve identifiable patient data in accordance with the declaration of Helsinki.

Authors' Contributions

Andreas Edenberg performed all of the 982 gastroscopies including exemplary hernia images, dataset training with AWS SageMaker as well as data evaluation himself.

Conflicts of Interest

The author declares that he does not have a conflict of interest.

Financial Support and sponsorship

The study was not funded by any organization.

Bibliography

1. Hill LD and RA Kozarek. "The gastroesophageal flap valve". *Journal of Clinical Gastroenterology* 28.3 (2019): 194-197.
2. Goodwin M., *et al*. "Atypical and typical manifestations of the hiatal hernia". *Annals of Laparoscopic and Endoscopic Surgery* (2020): 6.
3. Chiesa-Estomba CM., *et al*. "Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support". *European Archives of oto-Rhino-Laryngology: Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS): Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery* (2023).
4. ChatGPT: Optimizing Language Models for Dialogue. OpenAI (2023).
5. Parviainen J and Rantala J. "Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care". *Medicine, Health Care and Philosophy* 25 (2022): 61-71.
6. Rahwan Iyad., *et al*. "Machine behaviour". *Nature* 568 (2019): 477-568.
7. Vial A., *et al*. "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review". *Translational Cancer Research* 7.3 (2018): 803-816.
8. Yiming Ding., *et al*. "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain". *Radiology* 290.2 (2019): 456-464.
9. Visaggi P., *et al*. "Artificial Intelligence in the Diagnosis of Upper Gastrointestinal Diseases". *Journal of Clinical Gastroenterology* 56.1 (2022): 23-35.
10. Kononenko I. "Machine learning for medical diagnosis: history, state of the art and perspective". *Artificial Intelligence in Medicine* 23.1 (2001): 89-109.
11. Magoulas GD and Prentza A. "Machine learning in medical applications". *Lecture Notes in Computer Science* (2001): 300-307.

12. Kahrilas PJ., *et al*. "Approaches to the diagnosis and grading of hiatal hernia". *Best Practice and Research. Clinical Gastroenterology* 22.4 (2008): 601-616.

13. Hansdotter I., *et al*. "Hill classification is superior to the axial length of a hiatal hernia for assessment of the mechanical anti-reflux barrier at the gastroesophageal junction". *Endoscopy International Open* 4.3 (2016): E311-317.