J Gastroenterol
https://doi.org/10.1007/s00535-020-01743-2

The Japanese Society of Gastroenterology

ORIGINAL ARTICLE—ALIMENTARY TRACT

Artificial intelligence automates and augments baseline impedance measurements from pH-impedance studies in gastroesophageal reflux disease

Benjamin Rogers[1] · Sabyasachi Samanta[2] · Kevan Ghobadi[2] · Amit Patel[3] · Edoardo Savarino[4] · Sabine Roman[5,6,7] · Daniel Sifrim[8] · C. Prakash Gyawali[1]

Received: 31 August 2020 / Accepted: 19 October 2020

Abstract
Background Artificial intelligence (AI) has potential to streamline interpretation of pH-impedance studies. In this exploratory observational cohort study, we determined feasibility of automated AI extraction of baseline impedance (AIBI) and evaluated clinical value of novel AI metrics.
Methods pH-impedance data from a convenience sample of symptomatic patients studied off ($n = 117$, 53.1 ± 1.2 years, 66% F) and on ($n = 93$, 53.8 ± 1.3 years, 74% F) anti-secretory therapy and from asymptomatic volunteers ($n = 115$, 29.3 ± 0.8 years, 47% F) were uploaded into dedicated prototypical AI software designed to automatically extract AIBI. Acid exposure time (AET) and manually extracted mean nocturnal baseline impedance (MNBI) were compared to corresponding total, upright, and recumbent AIBI and upright:recumbent AIBI ratio. AI metrics were compared to AET and MNBI in predicting $\geq 50\%$ symptom improvement in GERD patients.
Results Recumbent, but not upright AIBI, correlated with MNBI. Upright:recumbent AIBI ratio was higher when AET $> 6\%$ (median 1.18, IQR 1.0–1.5), compared to $< 4\%$ (0.95, IQR 0.84–1.1), 4–6% (0.89, IQR 0.72–0.98), and controls (0.93, IQR 0.80–1.09, $p \leq 0.04$). While MNBI, total AIBI, and the AIBI ratio off PPI were significantly different between those with and without symptom improvement ($p < 0.05$ for each comparison), only AIBI ratio segregated management responders from other cohorts. On ROC analysis, off therapy AIBI ratio outperformed AET in predicting GERD symptom improvement when AET was $> 6\%$ (AUC 0.766 vs. 0.606) and 4–6% (AUC 0.563 vs. 0.516) and outperformed MNBI overall (AUC 0.661 vs. 0.313).
Conclusions BI calculation can be automated using AI. Novel AI metrics show potential in predicting GERD treatment outcome.

Keywords Gastroesophageal reflux disease · pH-impedance monitoring · Mean nocturnal baseline impedance · Artificial intelligence

✉ C. Prakash Gyawali
cprakash@wustl.edu

1 Division of Gastroenterology, Washington University School of Medicine, 660 South Euclid Ave., Campus Box 8124, Saint Louis, MO 63110, USA

2 Crosswave Solutions, LLC., Lexington, KY, USA

3 Division of Gastroenterology, Duke University School of Medicine, The Durham Veterans Affairs Medical Center, Durham, NC, USA

4 Division of Gastroenterology, Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy

5 Digestive Physiology, Hospices Civils de Lyon, Hopital E Herriot, Université de Lyon, 69437 Lyon, France

6 Digestive Physiology, Université de Lyon, Lyon I University, 69008 Lyon, France

7 Université de Lyon, Inserm U1032, LabTAU, 69008 Lyon, France

8 Barts and The London School of Medicine and Dentistry Queen Mary, University of London, London, UK

Published online: 05 November 2020

Introduction

Ambulatory pH-impedance monitoring is frequently used to quantify esophageal reflux burden in suspected gastroesophageal reflux disease (GERD) [1]. While acid exposure time (AET) is a primary metric of interest, baseline impedance (BI) measurements add value to pH-impedance monitoring. When calculated from nocturnal, artifact-free periods, mean nocturnal baseline impedance (MNBI) has been demonstrated to increase GERD diagnostic yield and predict response to therapy, particularly in patients with indeterminate AET [2–4]. However, BI can be influenced by the presence of refluxate or liquid in the esophagus, particularly when measured while supine; measurement while upright, where gravity promotes clearance of esophageal content, could overcome this limitation. However, manual measurement of BI in the upright position is cumbersome due to frequency of esophageal events (swallows, belches, reflux episodes, and movement artifacts) during awake hours.

The field of medicine has recently turned to artificial intelligence (AI) to automate and, where possible, optimize workflows. For instance, convolutional neural networks (CNN) have been utilized for detection of mucosal lesions on endoscopic images [5, 6]. To our knowledge, AI techniques have not been employed for the analysis of ambulatory reflux monitoring studies or for the extraction of BI measurements.

We hypothesized that an AI algorithm could be generated that would allow identification, censoring, and removal of esophageal events so that both upright and recumbent BI could be calculated. We further hypothesized that upright BI would add diagnostic value in assessing BI, because it would eliminate the potential influence of retained esophageal content in the supine position. We tested these hypotheses in this exploratory study using complex decision tree analysis of data from previously performed pH-impedance studies involving healthy normal volunteers and symptomatic patients tested off and on medical reflux therapy. The purpose of this investigation was to evaluate the feasibility of AI rather than to assess clinical outcomes; hence existing convenience patient samples were utilized.

Methods

For this exploratory observational cohort study to assess feasibility of AI in pH-impedance analysis, we utilized a convenience sample of existing patients with pH-impedance data and GERD symptom outcome collected prospectively from two tertiary care centers (St. Louis, USA and Padua, Italy). We analyzed ambulatory pH-impedance studies from symptomatic patients with persisting esophageal symptoms despite medical management studied both off and on proton pump inhibitor (PPI) therapy from previously established datasets that were prospectively collected [4, 7]. Subjects studied off PPI were required to discontinue medication at least 7 days prior to their study. Major motor disorders on high-resolution manometry (HRM) and large hiatus hernias ($\geq$ 3 cm) were excluded on enrollment into these studies. A control cohort was comprised of asymptomatic subjects also undergoing ambulatory pH-impedance studies off PPI, which were collected as part of an international normative study [8]. The study protocol for extraction of data from existing pH-impedance studies, and analysis using both standard and AI software, was approved by the human research protection office (institutional review board) at Washington University in St. Louis. Only shelved de-identified data were used for this study with no links to original patient information. Data sharing agreements were completed by each collaborating institution for pooled analysis of de-identified clinical and esophageal testing data.

Artificial intelligence analysis of pH-impedance studies

Raw data, in the form of comma-separated value (csv) files, were exported from commercial pH-impedance analysis software (Diversatek, Boulder, CO) and uploaded into a dedicated prototypical AI program created specifically for this study. The program consisted of a python-based decision tree analysis (DTA) algorithm that analyzed uploaded csv files. Multiple AI and machine learning (ML) platforms were explored. The more commonly applied machine learning techniques used in medicine (i.e., CNN) rely on existing image recognition algorithms, which are not applicable for impedance tracings. Therefore, DTA was selected, as algorithmic pathways could be designed based on expert panel consensus on parameters defining reflux, belch, air swallows, and liquid swallows. The final DTA algorithm consisted of 24 nodes over nine layers. During the development process, two study investigators (BDR, CPG) manually compared 2049 individual events with those identified by AI from two pH-impedance studies. AI software precisely identified 88.5% of these events, including 119 reflux episodes, 168 air events including belches, and 1528 variations of the impedance baseline including swallows and artifacts, indicating good sensitivity of AI in identification of pH-impedance events. From each study, these identified events and meal times were censored and excluded. The resulting baseline data points were averaged to yield a baseline impedance value, during both upright and recumbent periods, devoid of confounding data, which we termed AI baseline impedance (AIBI). The upright period BI and recumbent period BI were compared by converting these values into a ratio using upright AI divided by recumbent AI (U:R AIBI ratio), as well as by simple subtraction of recumbent AI from upright AI.

Conventional analysis of pH-impedance studies

Conventional analysis of pH-impedance studies consisted of extraction of AET, numbers of reflux episodes, and MNBI. AET was extracted using automated software currently in use in commercially available products (Diversatek, Boulder, CO). Categories of AET included pathologic (> 6%), borderline (4–6%), and physiologic (< 4%), according to the Lyon Consensus [1]. Numbers of reflux episodes were calculated from expert manual analysis of impedance studies after exclusion of artifacts and meal times. MNBI consisted of the average of the baseline impedance from three 10-min periods during quiet rest while recumbent (typically from 1 to 3 a.m.), and 2292 ohms was used as the normative threshold as per published data [2, 9]. Manually obtained MNBI values and a raw total BI (average baseline impedance value across the entire pH-impedance study) were extracted from 3 to 5 cm above the lower esophageal sphincter (LES), both using the same built-in software tool.

Data analysis

Data are reported as mean ± standard error of the mean (SEM) or median and interquartile range (IQR) unless otherwise indicated. Categorical data were compared using the χ^2 and Fisher's exact test, and continuous data were analyzed using the two-tailed Student's *t* test, ANOVA, or Kruskal–Wallis test, with Bonferroni correction for multiple comparisons when appropriate. Correlation between continuous variables was evaluated using Spearman's rank correlation and linear regression.

Symptom data were extracted from existing patient questionnaires that included visual analog scales (VAS) before and after GERD therapy, where patients reported their esophageal symptomatic state over the previous 2 weeks on a 10 cm line anchored by "no symptoms" (0) and "extremely severe symptoms" (100). Our groups have utilized this scale (global symptom severity, GSS), in several reports comparing the esophageal symptomatic state following GERD therapy, in assessing symptom outcome [3, 7, 10]. Treatment response consisted of ≥ 50% symptom improvement on VAS following medical therapy [4, 7]. Receiver operating characteristic (ROC) analyses were performed to assess performance characteristics of acid and impedance metrics in predicting symptom outcomes, with subset analysis between typical and atypical symptoms. A p value of < 0.05 was considered significant. SPSS Statistics v26.0 (Armonk, NY, USA) was utilized for all statistical analyses.

Results

Of 210 patients in the convenience cohort, 117 had been studied off PPI and 93 on PPI; 115 controls were included. Controls were significantly younger than patient groups and were comprised of fewer females than study subjects ($p < 0.001$ for each, Table 1). Proportions with AET > 6% were highest off PPI, while physiological AET (< 4%) was seen more often in controls and on PPI ($p < 0.001$ for each, Table 1). Similarly, proportions of patients with MNBI < 2292 ohms was highest off PPI and lowest in controls ($p < 0.001$ for all comparisons, Table 1).

Similar to MNBI, AIBI was lowest in the off PPI group and highest in controls ($p < 0.001$ across all comparisons, Table 1). Recumbent AIBI correlated significantly with MNBI equally well at both 3 cm ($r^2 = 0.79$, $p < 0.001$) and 5 cm locations ($r^2 = 0.86$, $p < 0.001$), whereas upright AIBI did not (Fig. 1). Total AIBI also correlated with MNBI at both locations, although the relationship was not as robust (3 cm: $r^2 = 0.45$, $p < 0.001$; 5 cm: $r^2 = 0.37$, $p < 0.001$). In contrast, the raw total BI (3 cm: 2177 ± 47 ohms; 5 cm: 1960 ± 36 ohms) was significantly lower than both MNBI (2361 ± 62 and 2102 ± 50 ohms, respectively) and total AIBI (2337 ± 63 ohms and 2120 ± 56 ohms, respectively, $p < 0.001$ across groups) indicating that esophageal events (including swallows and reflux events) reduced raw total BI, but not MNBI or AIBI.

Relationships between upright and supine AIBI were further analyzed at both 3 and 5 cm locations (Table 2). Controls had very similar total, upright, and recumbent AIBI (Table 2). However, U:R AIBI ratios were higher with total AET > 6% (median 1.18, interquartile range, IQR 1.0–1.5) compared to controls (0.93, IQR 0.80–1.09), or when total AET < 4% (0.95, IQR 0.84–1.1) and total AET was 4–6% (0.89, IQR 0.72–0.98) ($p \leq 0.04$ for each comparison). Similar ranges were noted at 3 cm.

Following medical therapy, 96 (47.8%) patients reported symptom improvement, which consisted of 50% of those tested off PPI and 44.8% of those tested on PPI (Table 1). When tested off PPI, MNBI, total AIBI, and the U:R AIBI ratio were significantly different between responders and non-responders to GERD management ($p < 0.05$ for each comparison); however, there was no difference between responders and non-responders when tested on PPI. Collectively, only U:R AIBI ratio segregated management

Table 1 Clinical characteristics of study groups

	Off PPI $N = 117$	On PPI $N = 93$	Controls $N = 115$	p value
Age (years)	53.1 ± 1.2	53.8 ± 1.3	29.3 ± 0.8	< 0.001
Gender (% F)	66	74	47	< 0.001
AET > 6%	32 (27.4%)	18 (19.4%)	1 (0.9%)	< 0.001
AET 4–6%	17 (14.7%)	5 (5.4%)	4 (3.5%)	< 0.001
AET < 4%	68 (58.1%)	70 (75.3%)	110 (95.7%)	< 0.001
MNBI 3 cm < 2292 ohms	70 (59.8%)	51 (55.4%)	33 (28.7%)	< 0.001
MNBI 5 cm < 2292 ohms	82 (70.1%)	56 (60.9%)	40 (34.8%)	< 0.001
MNBI 3 cm (ohms)	2014 ± 107	2189 ± 113	2839 ± 86	< 0.001
MNBI 5 cm (ohms)	1823 ± 87	1877 ± 94	2547 ± 64	< 0.001
AIBI 3 cm (ohms)	1933 ± 81	2111 ± 95	2931 ± 117	< 0.001
AIBI 5 cm (ohms)	1772 ± 64	1858 ± 83	2683 ± 112	< 0.001
Response to treatment	57 of 114 (50%)	39 of 87 (44.8%)	N/A	

PPI proton pump inhibitor, *AET* acid exposure time, *MNBI* mean nocturnal baseline impedance, *AIBI* artificial intelligence baseline impedance

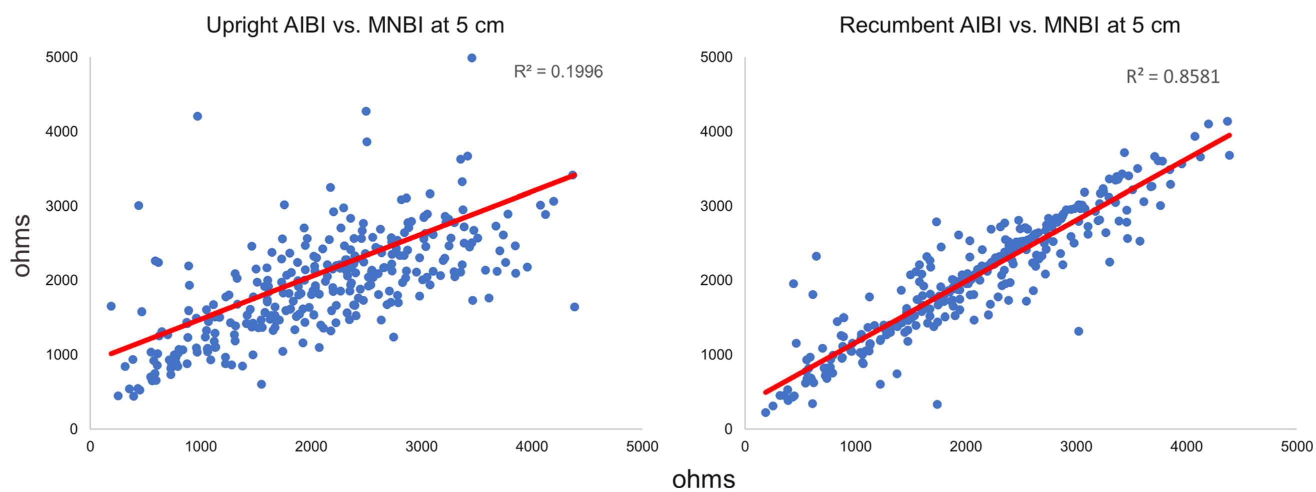


Fig. 1 Correlation between mean nocturnal baseline impedance (MNBI) and artificial intelligence baseline impedance (AIBI) at 5 cm above the lower esophageal sphincter. Upright AIBI (left panel) did not correlate strongly with MNBI ($r^2 = 0.20$); in contrast, recumbent AIBI (right panel) had a very strong correlation ($r^r = 0.86$). Similar results were seen at the 3 cm location (data not shown)

responders from controls and non-responders regardless of treatment status (Table 2).

On ROC analysis of different metrics in predicting response to medical therapy, off therapy U:R AIBI ratio at 5 cm had area under the curve (AUC) of 0.66, comparable with AET (AUC 0.72) and better than MNBI (AUC 0.31) (Fig. 2). When analyzing according to AET-based cohorts, U:R AIBI ratio at 5 cm outperformed total AET in predicting response to medical therapy in those with AET > 6% (AUC 0.766 vs. 0.606 respectively) and AET 4–6% (AUC 0.563 vs. 0.516 respectively) and performed similar to AET in those with AET < 4% (AUC 0.615 vs. 0.634 respectively, Fig. 3). In the cohort with U:R AIBI ratio of ≥ 1.0 and AET > 6%, AUC was 0.938 ($p = 0.019$) for U:R AIBI ratio at 5 cm in predicting symptom response. Performance of U:R AIBI ratio at 5 cm was optimal in individuals with typical symptoms (AUC 0.69), particularly regurgitation (AUC 0.71) where this metric outperformed AET (AUC 0.54). Metrics at 3 cm trended similarly, but were less robust.

Although patients were more likely to have MNBI < 2292 ohms than U:R AIBI ratio ≥ 1 (66.0 vs. 46.1%, $p < 0.001$), both U:R AIBI ratio ≥ 1 and MNBI < 2292 ohms correlated strongly with abnormal acid exposure in patients tested off ($p \leq 0.002$) and on PPI ($p \leq 0.01$). Further ROC analysis of MNBI suggested optimal performance characteristics at MNBI 3 cm of 1700 ohms (AUC 0.66) and 5 cm of 1500 ohms (AUC 0.63) in predicting response to medical therapy in those tested off PPI. Correlation between MNBI and U:R AIBI ratio was not robust ($r^2 = 0.14$). No pattern of upright or recumbent acid burden

Table 2 Comparison of esophageal physiologic parameters between controls and patients tested Off and On PPI

	Controls	Off PPI		On PPI	
		Response	No response	Response	No response
	n = 115	*n = 57*	*n = 57*	*n = 39*	*n = 48*
Total AET (%)	0.6 (0.2–1.2)	4.5 (2.0–9.4) †*	1.9 (0.7–4.2) †	0.3 (0.0–6.9)	0.4 (0.1–1.8)
3 cm					
MNBI (ohms)	2799 (2162–3521)	1537 (819–2337)†*	2390 (1626–3087)	2128 (983–3036)†	2208 (1556–3060)†
Total AIBI (ohms)	2623 (2185–3368)	1752 (1029–2272)†**	2209 (1665–2596)†	2034 (1415–2573)†	2173 (1750–2736)†
Upright AIBI (ohms)	2577 (2012–3375)	1523 (933–2266)†*	2150 (1526–2457)†	1985 (1376–2561)†	2204 (1634–2662)†
Recumbent AIBI (ohms)	2772 (2174–3481)	1760 (899–2462)†*	2320 (1858–2805)†	2102 (1174–2827)†	2226 (1685–2874)†
U:R AIBI ratio	0.89 (0.77–1.13)	1.10 (0.80–1.21)†*	0.90 (0.77–1.06)	0.96 (0.77–1.20)	1.00 (0.84–1.18)
5 cm					
MNBI (ohms)	2497 (2170–3029)	1378 (922–2027)†*	2128 (1493–2885)	1645 (1015–2571)†	1968 (1165–2593)†
Total AIBI (ohms)	2398 (1985–2908)	1513 (1058–2248)†**	1968 (1535–2374)†	1785 (1274–2241)†	1865 (1543–2231)
Upright AIBI (ohms)	2329 (1938–2834)	1450 (1025–2206)†*	1828 (1423–2272)†	1656 (1239–2245)†	1826 (1528–2127)†
Recumbent AIBI (ohms)	2468 (2111–2953)	1425 (1051–2085)†*	2070 (1623–2627)†	1621 (1052–2292)†	1835 (1428–2525)†
U:R AIBI ratio	0.93 (0.80–1.09)	1.05 (0.83–1.22)†*	0.89 (0.79–1.06)	0.96 (0.87–1.24)	0.97 (0.85–1.15)

Data expressed as median (interquartile range)

GERD gastroesophageal reflux disease, *PPI* proton pump inhibitor, *AET* acid exposure time, *MNBI* mean nocturnal baseline impedance, *AIBI* artificial intelligence baseline impedance, *U:R AIBI* upright: recumbent AIBI

*$p < 0.05$ compared to no response, **$p = 0.06$ compared to no response; †$p < 0.05$ compared to controls; no difference between subgroups on PPI across all measures

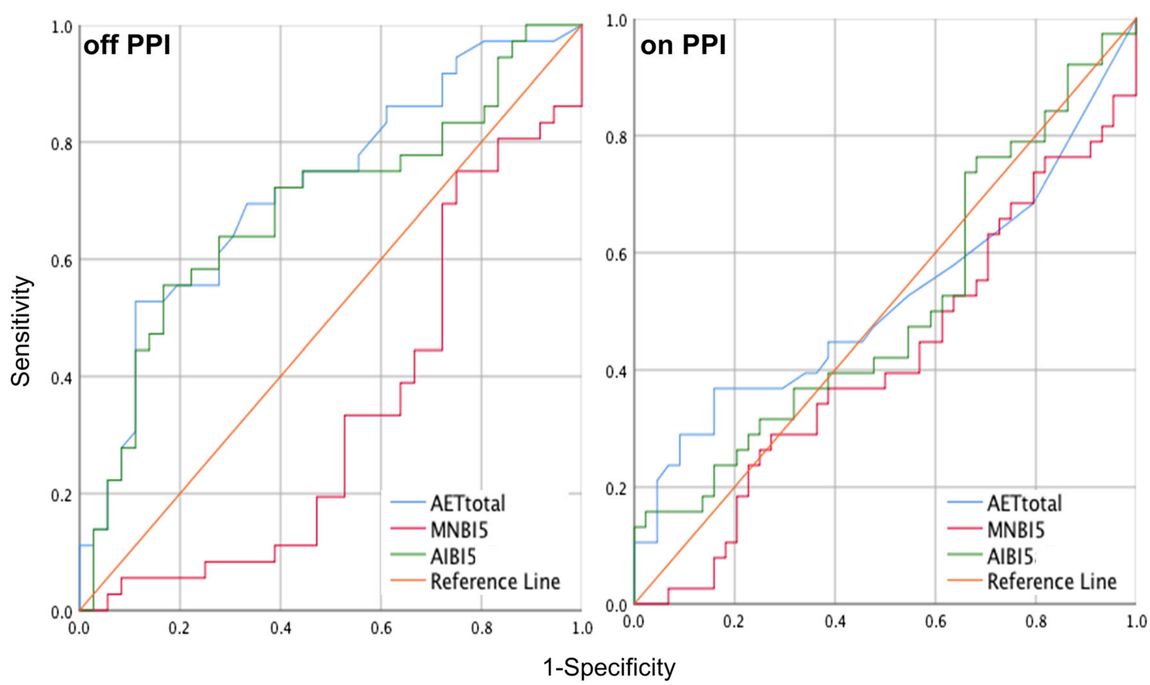


Fig. 2 On receiver operating characteristic (ROC) analysis, among those tested while off therapy, upright to recumbent AIBI ratio (U:R AIBI ratio) of ≥ 1.0 had area under the curve (AUC) of 0.661, comparable with AET (AUC 0.715) and better than MNBI (AUC 0.313) in predicting response to medical therapy. In those tested while on therapy, none of the traditional or novel metrics associated reliably with outcome

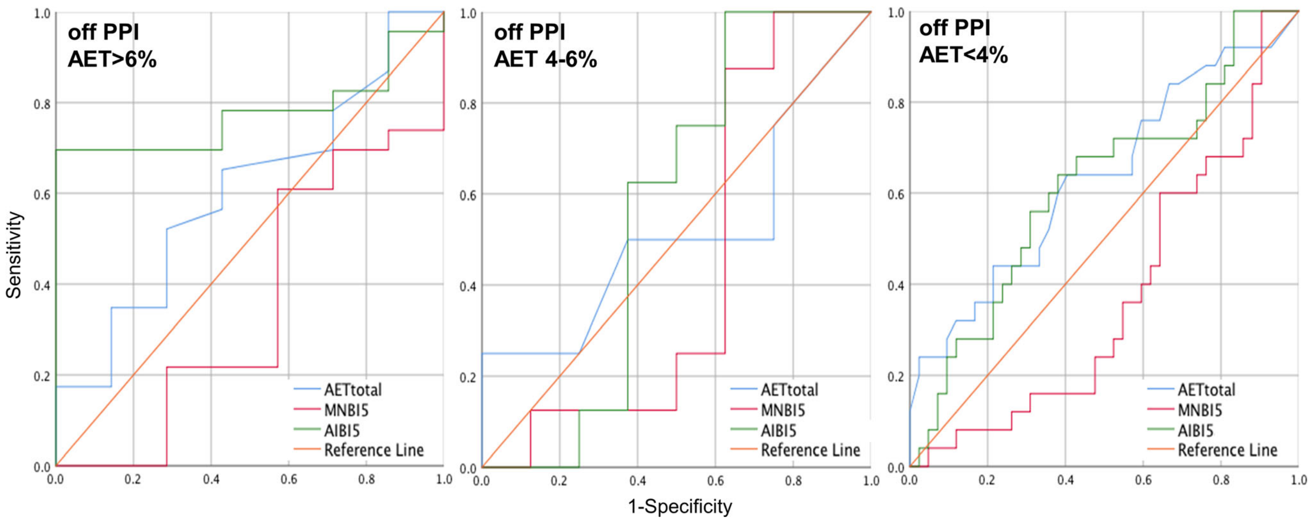


Fig. 3 Receiver operating characteristic (ROC) analysis limited to patients within acid exposure time (AET) categories demonstrates comparable performance of upright to recumbent AIBI ratio (U:R AIBI ratio) $\geq$ 1.0 and total AET in predicting response to medical therapy. U:R AIBI ratio was better than total AET in the AET $> 6\%$ category (AUC 0.766 vs. 0.606, respectively) and AET 4–6% category (AUC 0.563 vs. 0.516, respectively) and compared well with AET in the $< 4\%$ category (AUC 0.615 vs. 0.634, respectively)

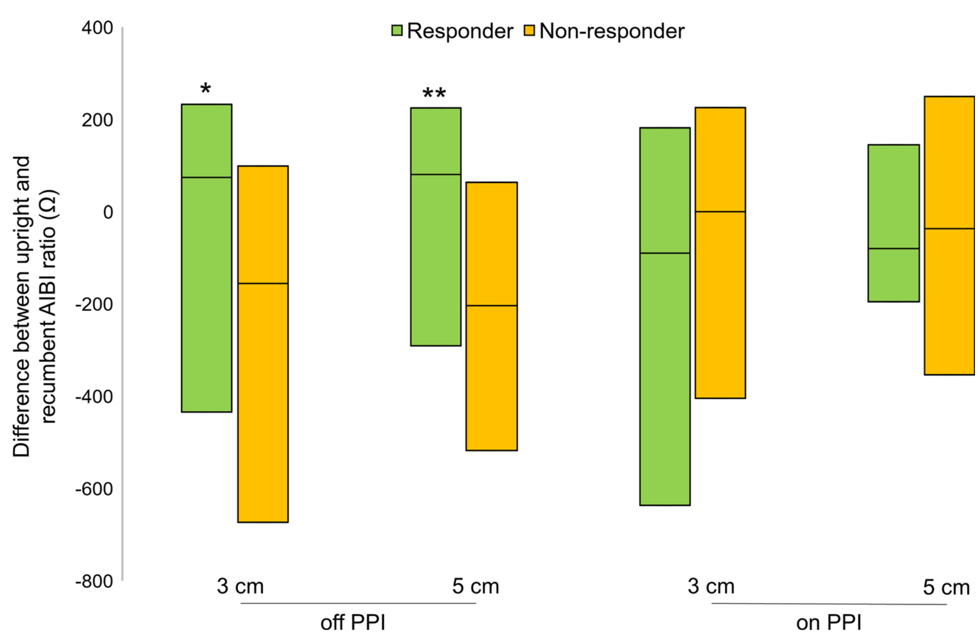
could be found to correlate with U:R AIBI ratio or outcome. On therapy metrics were not predictive of outcome.

To further assess upright versus recumbent BI, the differences between AIBI in the upright and recumbent positions were calculated at the 3 and 5 cm locations using simple subtraction (uprigt AIBI minus recumbent AIBI) (Fig. 4). Individuals tested on PPI showed low variability between upright and recumbent values; a similarly low variability was seen in individuals tested off PPI who responded to medical therapy. In contrast, non-responders showed increased variability between upright and recumbent AIBI values (Fig. 4). Additionally, only responders tested off PPI were found to have upright values higher than recumbent, keeping with the elevated U:R AIBI ratio.

Discussion

In this exploratory proof-of-concept study to determine feasibility of AI in pH-impedance analysis, we show that AI can accurately and rapidly extract clinically meaningful

Fig. 4 Box plots comparing the mathematical difference between upright and recumbent AIBI values (ohms, median and interquartile range) between responders and non-responders to medical GERD therapy, at the 3 and 5 cm locations. Median values were significantly different at both 3 and 5 cm locations in responders compared to non-responders among patients tested off proton pump inhibitor (PPI) therapy, but not in patients tested on therapy. (*p = 0.01, **p = 0.004)



metrics from standard ambulatory pH-impedance tracings, and can censor events to generate an artifact-free BI across the entire study. Recumbent BI extracted using AI strongly correlates with MNBI [11] indicating that the extracted BI is a valid measure. We demonstrate that the relationship of upright to supine AIBI predicts symptom response to medical therapy. These findings show that the performance of AI-based interpretation of pH-impedance studies is not inferior to the current reflux monitoring standards utilized to diagnose GERD, and novel AI-based metrics not available from standard analysis, such as the AIBI ratio, have potential to predict treatment outcome. Finally, while this report focused on BI after censoring esophageal events, our findings open the door to investigation of AI in identification of esophageal events on pH-impedance studies. Of further value, AI metrics are extracted instantaneously, reproducibly, and offer an advantage to current MNBI extraction methods.

Esophageal events, both liquid and gas, follow stereotypical patterns. Reflux episodes are defined by sharp retrograde decline in impedance $\geq 50\%$ with subsequent return to baseline [12] as outlined by the recently published Wingate Consensus [13]. Liquid swallows, alternatively, are defined by sharp drops in impedance that propagate in the anterograde direction. Sharp rises in impedance define air events, and the directionality determines whether the event is an air swallow or belch (anterograde or retrograde, respectively) [14]. These distinct features allow for the characterization tasks at which computers excel. In the development phase of our AI software, we were able to achieve nearly 90% sensitivity in identification of these esophageal events using a complex DTA. This approach was purposefully selected as DTA increases reproducibility while reducing the possibility of over-fitting and magnifying inherent biases, which are the major limitations in the clinical implementation of machine learning techniques.

For this proof-of-concept study, we automated the recognition, censoring, and removal of esophageal events in order to focus on BI values. We noted a strong correlation between our recumbent AIBI and MNBI calculated according to current standards. However, the association between AIBI in the upright position and MNBI was much less robust, which supported our hypothesis that upright BI would differ from recumbent BI. This prompted us to investigate the relationship of upright to supine AIBI. Based on ROC analysis of patients studied off PPI, an upright to supine AIBI ratio ≥ 1 was associated with better outcome with medical therapy. The U:R AIBI ratio trended closely with AET in predicting outcomes, with improved performance in individuals with regurgitation and borderline AET, highlighting the fact that the field is ripe for further investigations and expanded AI utilization.

The fact that the raw BI value prior to AI censoring and removal of esophageal events is lower than the final AIBI highlights the fact that esophageal retention of content from swallows and reflux episodes contributes to measured BI; this is a situation where AI ensures accuracy of BI measurement. We demonstrate that a differential between upright and supine BI is relevant to outcome. Interestingly, this may suggest that individuals with stability between upright and recumbent BI values (U:R AIBI ratio ≥ 1, raw subtraction values closer to 0) have fewer confounding mechanisms. Stated simply, non-responders, noted to have higher volatility in their BI values, may have increased likelihood of alternative explanations for symptoms, particularly since the pH-impedance catheter is presumably more physically stable in the recumbent position.

This proof-of-concept study demonstrates the potential of AI in interpretation of pH-impedance studies. Our intent in presenting our findings is not to supplant AET or MNBI, which have been demonstrated to correlate with cross-sectional and longitudinal reflux burden, respectively [15]. Rather, we wish to establish that AI has potential to predict GERD outcome and also demonstrates potential for the identification of novel metrics that could be reported instantaneously and seamlessly. By showing that we could identify and censor all types of esophageal episodes, we have also introduced an advanced method for analyzing the basic components of impedance testing. This mechanism provides an avenue for directing a computer's attention so that more advanced machine learning algorithms can be applied.

Our study is not without limitations. We used a convenience cohort of normative data and existing data sets of patients studied off and on PPI. Because of the exploratory nature of this study, we included patients with diverse presentations. While appropriate for this analysis, which did include typical symptom subgroups, larger prospective and consecutively collected patients will provide a more robust confirmation of our hypotheses, particularly in patients with proven GERD and typical reflux symptoms. We also did not have patients with large hiatus hernias or surgically managed patients, which would have provided an additional level of confidence in post-therapy analysis; we did not segregate subjects by EGJ morphology. We used VAS scales to determine GERD treatment outcome, which are validated tools for assessment of change in perceptive symptoms and function within several domains [16, 17]; however, other GERD questionnaire scores such as the GERDQ and reflux disease questionnaire (RDQ) were not available for this report. Esophageal events of all types could induce prolonged effects on baseline impedance beyond the acute change induced by bolus presence; this has not been applied using our current algorithm, but will need to be evaluated in future iterations. Our AI algorithm has laid the groundwork for projects, some already underway, which will build upon the current framework by incorporating advanced machine learning techniques. These techniques may allow for clinically relevant improvements on current analysis paradigms, including automated meal and artifact identification and extraction, instantaneous analysis, and improved accuracy with confidence predictions based on aggregated outcomes data.

In conclusion, we demonstrate that AI can successfully and instantaneously extract BI from the entire pH-impedance study by identifying and censoring esophageal events. While recumbent AIBI resembles MNBI, the ratio of upright to recumbent AIBI augments BI evaluation and predicts outcome from medical GERD management. Further research and additional analyses will be required to fully appreciate AI potential in simplifying pH-impedance interpretation and to determine whether this approach will have a meaningful impact in clinical practice.

Author contributions BDR: study concept, data analysis, drafting of manuscript, and critical review of manuscript; SS: creation of AI algorithm, data analysis, and critical review of manuscript; KG: creation of AI algorithm, data analysis, and critical review of manuscript; AP: data collection and final review of manuscript; ES: data collection, key intellectual content, critical review of manuscript; SR: key intellectual content and critical review of manuscript; DS: key intellectual content and critical review of manuscript; CPG: study concept, data analysis, key intellectual content, drafting, and finalization of manuscript.

Compliance with ethical standard

Conflict of interest No conflicts of interest exist. Partial funding through the Thomas and Mary Eckert Fellowship Clinical Research Fund, through the Division of Gastroenterology, Washington University School of Medicine, St. Louis, awarded to BDR. BDR: no disclosures; SS: no disclosures; KG: co-founder of ClearifiRx; ES: Lecture Fee: Medtronic, Takeda, Janssen, MSD, Abbvie, Malesci; Consulting: Medtronic, Takeda, Janssen, MSD, Reckitt Bencikser, Sofar, Unifarco, SILA, Oftagest ; SR: consulting Medtronic, research support Diversatek Healthcare, Medtronic; DS: research grants: Reckitt Benckiser UK, Jinshan Technology China, Alfa Sigma Italy; CPG: Consulting: Medtronic, Diversatek, Isothrive, Ironwood, Quintiles.

References

1. Gyawali CP, Kahrilas PJ, Savarino E, et al. Modern diagnosis of GERD: the Lyon Consensus. Gut. 2018;67:1351–62.
2. Frazzoni M, Savarino E, de Bortoli N, et al. Analyses of the post-reflux swallow-induced peristaltic wave index and nocturnal baseline impedance parameters increase the diagnostic yield of impedance-pH monitoring of patients with reflux disease. Clin Gastroenterol Hepatol. 2016;14:40–6.
3. Patel A, Wang D, Sainani N, et al. Distal mean nocturnal baseline impedance on pH-impedance monitoring predicts reflux burden and symptomatic outcome in gastro-oesophageal reflux disease. Aliment Pharmacol Ther. 2016;44:890–8.
4. Rengarajan A, Savarino E, Della Coletta M, et al. Mean nocturnal baseline impedance correlates with symptom outcome when acid exposure time is inconclusive on esophageal reflux monitoring. Clin Gastroenterol Hepatol. 2020;18:589–95.
5. Jin EH, Lee D, Bae JH, et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. Gastroenterology. 2020;158:2169–79.
6. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. Gastroenterology. 2020;158:2150–7.
7. Patel A, Sayuk GS, Gyawali CP. Parameters on esophageal pH-impedance monitoring that predict outcomes of patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol. 2015;13:884–91.
8. Sifrim D, Roman S, Savarino E, et al. Normal values and regional differences in oesophageal impedance-pH metrics: a consensus analysis of impedance-pH studies from around the world. Gut. 2020; https://doi.org/10.1136/gutjnl-2020-322627.
9. Martinucci I, de Bortoli N, Savarino E, et al. Esophageal baseline impedance levels in patients with pathophysiological characteristics of functional heartburn. Neurogastroenterol Motil. 2014;26:546–55.
10. Rogers BD, Patel A, Wang D, et al. Higher esophageal symptom burden in obese subjects results from increased esophageal acid exposure and not from dysmotility. Clin Gastroenterol Hepatol. 2020;18:1719–26.
11. Hoshikawa Y, Sawada A, Sonmez S, et al. Measurement of esophageal nocturnal baseline impedance: a simplified method. J Neurogastroenterol Motil. 2020;26:241–7.
12. Sifrim D, Castell D, Dent J, et al. Gastro-oesophageal reflux monitoring: review and consensus report on detection and definitions of acid, non-acid, and gas reflux. Gut. 2004;53:1024–31.
13. Gyawali CP, Rogers B, Frazzoni M, et al. Inter-reviewer variability in interpretation of pH-impedance studies: the Wingate Consensus. Clin Gastroenterol Hepatol. 2020;S1542–3565(20):31230–1. https://doi.org/10.1016/j.cgh.2020.09.002.
14. Sifrim D, Silny J, Holloway RH, et al. Patterns of gas and liquid reflux during transient lower oesophageal sphincter relaxation: a study using intraluminal electrical impedance. Gut. 1999;44:47–54.
15. Gyawali CP. Redeeming clinical value of esophageal pH impedance monitoring. Clin Gastroenterol Hepatol. 2016;14:47–9.
16. Bengtsson M, Ohlsson B, Ulander K. Development and psychometric testing of the visual analogue scale for irritable bowel syndrome (VAS-IBS). BMC Gastroenterol. 2007;7:16.
17. Guyatt GH, Townsend M, Berman LB, et al. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis. 1987;40:1129–33.