HHS Public Access

Author manuscript

Neurogastroenterol Motil. Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

Neurogastroenterol Motil. 2018 December ; 30(12): e13462. doi:10.1111/nmo.13462.

Endoflip vs High-Definition Manometry in the Assessment of Fecal Incontinence: A Data-Driven Unsupervised Comparison

Ali Zifan, PhD[1], **Catherine Sun, BSc**[1], **Guillaume Gourcerol, MD**[2], **PROFESSOR Anne M Leroi, MD**[2], and **DR Ravinder K Mittal, MD**[1]

[1]Department of Medicine, Division of Gastroenterology, University of California, LA Jolla, CA, USA

[2]INSERM U1073, Service de Physiologie Digestive, CHU Rouen, INSERM CIC 1404 Rouen, F-76000

Abstract

Background: How much anal sphincter dysfunction contributes to FI is not clear. High definition anorectal manometry (HDAM), and functional luminal imaging probe (Endoflip) are two new techniques to study anal sphincter function.

Aims: The goal was to compare the diagnostic utility of HDAM and Endoflip using optimal feature(s) in each modality for FI diagnosis. Methods: Blinded classification was carried out on 70 female subjects (32 FI & 38 controls), using 3 prominent machine learning clustering techniques, with 3 distance metrics. For HDAM, descriptive statistics, shape, and textural features characterizing the spatial relationship of pixels in the HDAM high pressure zone, and for Endoflip, permutations of pressure and CSA combinations (i.e., multiplication, division, or individually) at rest and squeeze were tested.

Results:

Intra-modality: 1)-Endoflip: Best clustering was obtained using the combination of the ratio of CSA over pressure at 40ml and 50ml at rest, which had significantly better specificity ($p<0.001$) than using only pressure at 50ml, no difference in sensitivity ($p=0.68$). 2)-HDAM: clustering using textural information at rest had significantly higher specificity compared to using only the maximal pressure at rest ($p<0.001$).

Inter-modality: Clustering results using optimal features were not significantly different with respect to sensitivity or specificity ($p>0.05$). Optimal Endoflip feature set differed significantly in specificity compared to HDAM maximal pressure at both rest ($p<0.001$) and squeeze ($p<0.001$).

Conclusion: Defective anal closure function is fairly sensitive and highly specific in diagnosing FI. Using optimal feature-sets, HDAM and Endoflip perform in a similar fashion in diagnosing FI, but are not complementary.

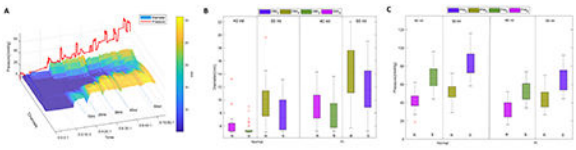
Address for Correspondence: Ravinder K Mittal MD, ACTRI, 9500 Gillman Drive, MC 0061 La Jolla, CA. 92093-0990, rmittal@ucsd.edu, Phone: 858-543-3328.

Author Contributions: AZ: Data analysis, figure preparation, and wrote the manuscript; CS: Data analysis, figure preparation; AML: Data Acquisition, data analysis; GG: Data Acquisition, data analysis; RKM: Data analysis, co-wrote the manuscript.

COI: None of the authors have any conflict of interest

Abbreviated abstract:

High definition anal manometry and functional luminal imaging probe (endoflip) are two novel techniques to assess the anal closure function, however their sensitivity and specificity in diagnosing fecal incontinence (FI) is not clear. • We used machine learning tools to determine the ideal parameters provided by HDAM and endoflip; and performed blinded analysis on a group of FI patients and normal subjects to determine the sensitivity and specificity of two techniques in diagnosing FI. Our data show that both of these techniques are moderately sensitive but highly specific in the FI diagnosing; but they are not complimentary.



Keywords

Functional luminal imaging probe; High definition anorectal monetary; unsupervised learning; cluster analysis

Introduction

The etiology of fecal incontinence (FI) is multifactorial[1–6]; stool volume and liquidity, rectal compliance, rectal inflammation, sensory and motor functions of the rectum and anal closure mechanism, all may play important roles. Several studies show that in majority of the patients, weakness of the anal closure mechanism is a major factor in the pathogenesis of FI[7, 8]. The internal anal sphincter, external sphincter and the puborectalis muscle, contribute to the anal closure mechanism. Studies show that 70% or more of the resting anal canal pressure is related to the internal anal sphincter (IAS), and the remainder due to the external anal sphincter (EAS). Recent studies also show that the increase in anal pressure with voluntary squeeze is due to the puborectalis (PRM) and EAS in the proximal and distal halves of the anal canal respectively[9]. The pressure increase related to EAS contraction is much larger than related to the PRM. The 3 anal closure mechanisms are anatomically overlapping structures[10], and it may be difficult to know the precise contribution of each of them by recording anal pressures using intraluminal pressure recording probes, i.e., manometry technique.

A number of studies show that in majority of FI patients, the anal canal pressure during voluntary squeeze is lower as compared to normal subjects pointing towards the weakness of the EAS as the major cause of FI[7, 8, 11]. However, the manometry techniques used to record anal canal pressure in those studies were not “state of the art”. High resolution anal manometry (HRAM) is the most widely used technique to measure anal canal pressure[12]. More recently, 3D-high definition anal manometry (HDAM) has also been used by several investigators[13–16]. The 3D-HDAM probe is 10 mm in diameter, and it has 16 rings of pressure sensors spaced 6mm apart with 16 sensors in each ring. The advantages of 3D-HDAM over conventional HRAM is that it provides detailed information on the anal canal

pressure profile, such as the area of the HPZ, circumferential and axial pressure asymmetry, the parameters that may be useful to assess the anal closure function [13].

The functional luminal imaging probe (FLIP) is another novel technique to assess anal closure function[17]. It records changes in luminal diameter/cross sectional area during distension of the anal canal[18]. Investigators have used distensibility of the anal canal as an important parameter to distinguish FI patients from normal controls. The distensibility is defined as a ratio of the luminal cross-sectional area (CSA) to pressure. Studies have found that anal canal distensibility is higher in FI patients compared to normal subjects[14, 17, 19]. Gourcerol et al performed both, 3D-HDAM and FLIP studies in a cohort of patients with FI and controls to determine if one technique outperformed the other in distinguishing the two groups[14]. We recently reported that in addition to using the peak pressure from the HDAM recordings (as used by Gourcerol), circumferential and axial asymmetry of the anal pressure profile, and area of the anal HPZ may be useful parameters to distinguish patients from the normal subjects[13]. Leroi et al performed both, 3D-HDAM and FLIP studies in a cohort of patients with FI and controls to determine if one technique outperformed the other in distinguishing the two groups[20, 21]. The limitation of their studies are that, 1) the data analysis was not blinded and, 2) they didn't use "state of the art" machine learning tools and permutation of features to determine the best parameters that may distinguish two groups.

In this comprehensive and detailed analysis, we address three lingering questions, 1) in terms of clinical diagnosis, are the two modalities (i.e., Endoflip and HDAM), sensitive, specific (or both), 2) complementary, and 3) what are the optimal features in each modality that discriminate healthy subjects from patients suffering from FI. We use unsupervised learning[22],using advanced machine learning techniques, and include new feature sets (parameters) in both modalities to determine whether there exist naturally occurring patterns among healthy subjects and FI patients that clearly separate the two groups, without any prior knowledge of subject classification (i.e. healthy vs FI). We test different permutations of features, similarity metrics, and recommend the optimal features from each modality that achieved the optimal diagnostic result.

Materials & Methods

Subjects

The details of the patient population and controls have been reported previously by Gourcerol et al[14](See Suppl. File 1). We will focus on the details of the comprehensive data analysis; the novel aspect of our study. Briefly, thirty-two female patients with history of FI and thirty-eight healthy female subjects were recruited at the Physiology Unit of Rouen University Hospital (France) for this prospective study. All patients had suffered from FI for more than 3 months. Severity of incontinence was graded by the Cleveland Clinic Incontinence Scoring System, while quality of life was assessed with the French version of the American Society of Colon and Rectal Surgeons' (ASCRS) Fecal Incontinence Quality of Life Scale (FIQL). The exclusion criteria for the study were (i) fecal impaction; (ii) previous rectal resection, imperforate anus, other congenital anorectal malformation; (iii) rectal prolapse; (iv) stoma; (v) pathologies preventing the insertion of the probe (i.e., anal stenosis, complicated hemorrhoids, anal fissures); and (vi) pregnant or breast-feeding

patients. The healthy subjects had no history of FI, did not take any regular medication related to bowel disorders, had no history of gastrointestinal disease or anorectal surgery, and were not were pregnant or breast-feeding. The two groups were matched for age, parity and BMI.

EndoFLIP assessments

The distensibility of the anal sphincter was determined using the EndoFLIP® system (Crospon Ltd, Galway, UK). The probe consists of a two-lumen polyethylene tube with an outer diameter of 3 mm. It contains an array of 17 ring electrodes placed at 4-mm intervals that measure electrical impedance to estimate the CSA at 16 points, 5 mm apart. The CSA measurements were performed over an 8-cm-long zone. A 12-cm-long bag mounted on the probe was designed to be filled to a maximum diameter of 25 mm. The intra-bag pressure was calculated using a solid-state pressure transducer placed inside the bag. The CSA measurements and pressures were sampled at 10 Hz and were stored in the data acquisition system. The EndoFLIP® probe was inserted in the rectum, while the subjects were in the left lateral position. Before using the system, air was removed from the probe and the baseline intra-bag pressure was set to zero. The EndoFLIP® probe was inserted, with two detection electrodes remaining visible outside the anal verge. During the experiments, the probe was held in place manually. The bag was filled with 10, 20, 30, 40, or 50 mL of a conductive saline solution at a rate of 40 mL/min. The resting measurements were recorded for 30s, following which the subjects were asked to squeeze at the end of each distension.

3D High Resolution Manometry

3D-HRM was performed using a solid-state probe (Given Imaging, Yoqneam, Israel) with a 10-mm external diameter. The probe incorporated 256 pressure sensors arranged in 16 rings, each of which had 16 circumferentially oriented sensors. With the subject in the left lateral position, the lubricated probe was inserted such that a panel of pressure sensors lay across the anal canal, with some extending above and some below the anal canal. The probe was oriented with the marker on the probe pointing to the dorsal aspect of the subject and it was held in place manually during each procedure. Maneuvers were performed with a standard sequence that included a 30-s recovery period between each maneuver: Rest – anorectal pressures were measured with the subject relaxed after 15-min rest period; Squeeze – the subject was instructed to squeeze the anal sphincter as strongly and as long as possible (two attempts).

Data Analysis

For HDAM, descriptive statistics (minimum and maximum pressure, 0.25, 0.50 and 0.75 quantiles, interquartile range), shape (area, symmetry), ratio of peak squeeze pressure over rest pressure, difference of peak squeeze pressure and rest pressure, and textural features[23], characterizing the spatial relationship of pixels in the HDAM high pressure zone(HPZ) were extracted. Below, we discuss some of the important features:

Anal Symmetry Index (ASI): In geometric terms, the binary ASI determines how much a given HPZ is symmetric (e.g., circle/square have perfect symmetry [ASI = 0] assuming only horizontal and vertical rotations). A Gray-scale ASI is a measure of the shape differences

along with the difference in pressures within the region of anal canal (see Figure 1)[13]. Carrying out the above process produces four features for a single HDAM-image HPZ (i.e., binary and gray-horizontal and vertical ASIs).

Textural features: Texture is a measure of the variation of the intensity (in this case pressure in the anal canal HPZ zone) of a surface (Figure 2), quantifying properties such as smoothness, coarseness and regularity. Statistical techniques characterize texture by the statistical properties of the grey levels of the points comprising a surface. Typically, these properties are computed from the, 1) grey level histogram, or 2) grey level co-occurrence matrix (GLCM) of the surface[23]. The GLCM characterizes the texture of an HDAM image by calculating how often pairs of pixel with specific values, and in a specified spatial relationship occur in the HDAM image, creating a GLCM, and then extracting statistical measures from this matrix. For our study, nine textural features were extracted from both the gray level histograms, and the gray level co-occurrence matrix (GLCM) of the HDAM data (See Suppl. File 2).

For Endoflip, permutations of pressure and CSA (i.e., multiplication, division(also known as the distensibility index), combination, or individually) at both rest and squeeze were tested for 40 ml and 50ml balloon distensions, that is balloon distensions that produced luminal diameters larger than the FLIP probe diameter during squeeze. Finally, each clustering was scored based on the actual subject labeling, using external evaluation (Rand-index[24, 25]). The Rand-index is similar to accuracy in measuring predictive performance.

Unsupervised learning

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses (in our case heathy and FI subjects). The most common unsupervised learning method is cluster analysis[26], which is used for exploratory data analysis to discover hidden patterns, structure or grouping in data. In other words, we would like the individuals within a group (e.g., healthy) to be close or similar to one another, but dissimilar from individuals in the other group.

In our simulations, we use three prominent clustering methods to analyze the Endoflip and HDAM data: 1) Exclusive clustering (using K-means algorithm[27]), 2) Hierarchical clustering[27], and 3) Probabilistic clustering (using Gaussian Mixture Models (GMM)). In the first case (i.e., Exclusive clustering), data are grouped in an exclusive way, so that if a certain subject belongs to a definite cluster (e.g., healthy group) it could not be included in another cluster (i.e., FI). On the other hand, in hierarchical clustering the goal is to build a multilevel hierarchy of clusters by creating a cluster tree. The clusters are modeled using a measure of similarity, which is defined upon metrics, such as Euclidean or probabilistic distance. Finally, probabilistic clustering methods use a completely probabilistic approach to grouping. We employ the aforementioned methods to divide the input data into two distinct groups (i.e., Normal and FI). For a detailed overview of each method, the interested reader is referred to the appropriate references[23, 26, 28].

K-means clustering—K-means is one of the simplest unsupervised learning algorithms that solves the well-known clustering problem. In K-means clustering, the data is partitioned

into K (e.g., two in our case), distinct clusters based on distance to the centroid of a cluster. The algorithm starts with an initial solution, which is iteratively improved until a local minimum is reached[28].

Hierarchical clustering—Hierarchical clustering works by grouping data objects into a tree of clusters (also known as dendrogram). This method uses hierarchical decomposition of a given set of data objects. Every cluster node contains child clusters; sibling clusters partition the points covered by their common parent. Such an approach allows exploring data on different levels of granularity[29].

Gaussian Mixture Models—The GMM-based clustering algorithm explores a probability-centered approach to clustering. The most widely used clustering method of this kind is the one based on learning a mixture of Gaussians; we can consider clusters as Gaussian distributions centered on their barycentres. In other words, we model clusters as a mixture of multivariate normal density components[28].

Distance measure

An important component of a clustering algorithm is the distance measure (a notion of dissimilarity) between data points. Based on the values found, subjects are assigned to a certain group (either healthy or FI) with minimum distance. Hence, this distance calculation plays a key role in any clustering algorithm. In this study, we use the following measures: Squared Euclidean (the sum of the squared differences between the values of the data points); Cityblock (or Manhattan), which is the sum of the absolute differences between the values of the data points; and Cosine (similarity) distance (the cosine of the angle between two vectors of values)[23].

Statistical Analysis

Quantitative data are reported as mean (±standard deviation) or median (IQR 25–75) when appropriate. The normality of the distributions was checked by the Shapiro-Wilk test. In the statistical analysis, non-parametric ANOVA (Kruskal–Wallis test) using Dunn's correction was used for multiple comparisons. $P<0.05$ was considered significant. The 95% sensitivity and specificity confidence intervals is also reported, and their values compared using the modified McNemar's test[30] (separately among patients and healthy subjects) to find if the scores from the two modalities are really different from each other. All calculations were performed in Matlab 2017a (Mathworks, Inc, Natick, Massachusetts, United States).

Results

Figure 3 shows a sample reconstruction of a normal subject and a FI patient undergoing endoflip distension study. The variation of pressure with ramp distension of the anal canal is shown in red. Furthermore, the variation of the balloon diameter with the increase of pressure is shown as a surface topograph. The least distensible part of the anal canal is clearly the mid-anal canal resulting in an hourglass shape of the diameter topograph, suggesting that the mid-anal canal is likely the most important part in maintaining continence. Figures 3B & 3C show boxplots of the distribution of the diameter and pressure

at 40ml and 50ml balloon distension, for all subjects in both groups. Overall, the rest and squeeze diameters are higher in the patients compared to normal for both, 40ml and 50ml Endoflip balloon distensions.

Figure 4 shows a sample HDAM and Endoflip responses for a normal and FI subject undergoing both modalities in the dataset. Note, the pressure profile at rest and squeeze in the subject, and distribution of pressure in the anal high-pressure zone, which are markedly different between normal and patient (see Suppl. Video 1 for a sample dynamic Endoflip reconstruction for both a normal and FI patient). Figure 5 shows the box plot of the HDAM rest and squeeze images, including both median and mean(shown in circles) in each of the normal and patient subjects (each HDAM image has been converted into a single column vector of m rows, where m is 16×16).

Clustering

Unsupervised classification (blinded) was carried out using the feature sets, clustering methods, and distance metrics, as described in the methods section. As the feature space was highly dimensional (i.e., the simulations consisted of the use of 3 clustering methods, with 3 different distance metric along with combination of different feature permutation in each modality), only the top ten metrics with the highest accuracy (i.e., Rand Index[31]) were selected for further processing. Sample clustering results are shown in Figure 6 for both HDAM and Endoflip, using the k-means algorithm. In this figure, Endoflip misclassified 16 out of 70 subjects (both normal and patients), using Endoflip features at rest (pressure, CSA and CSA/Pressure at both 40 and50ml), and this number dropped in the HDAM to 12 using the 9 textural features. Figure 7 shows a bar plot of the features used (i.e., textural features for HDAM and multiplication and division of pressure and CSA at 40ml and 5ml for Endoflip), and their statistical significance from each other in each modality.

Once the output of each clustering method is known, the resulting labels from each approach (i.e., clustering method, distance metric and feature set) are summarized in a table (i.e., confusion matrix), from which four conditional probabilities are computed: true positive rate, or sensitivity, which corresponds to the probability of a positive (or abnormal) result in a diseased individual; true negative rate, or specificity, that is the proportion of good items predicted as such; false positives rate, or 1-specificity, being the probability of a negative (or normal) result in a healthy individual; false negatives rate, corresponding to the proportion of bad items predicted as good. Columns 3 and 4, in Table 1 show the best performing clustering methods, alongside the distance measures, and feature sets that produced the highest scores in each modality. Among the clustering methods, probabilistic clustering did not reach the top 10, and the results were dominated by the k-means algorithm, and Hierarchical in a few instances. In terms of distance metric, the Cityblock distance produced the best results, followed by the cosine similarity distance.

Intra-modality

Endoflip: the best clustering result produced a sensitivity of 56% (44% −68%), and specificity of 97% (90% - 100 %). The latter was obtained using the combination of the ratio of CSA over pressure (i.e., distensibility index) at 40ml and 50ml at rest, which had

significantly better specificity ($p<0.001$) than using pressure at 50ml, although there was no difference in sensitivity ($p=0.68$). At squeeze, the combined 40 ml and 50 ml CSA/pressure ratio also produced the highest clustering index, resulting in a sensitivity of 59% (47.0% - 71%), and specificity 89% (79% - 96%). However, the sensitivity and specificity ($p=0.34$, $p=0.11$, respectively) of CSA/pressure ratio was not different compared to pressure at 50ml during squeeze.

HDAM: clustering using textural information at rest yield a sensitivity of 69% (56% - 79%) and specificity of 97% (90% - 100.0%). It also had significantly higher specificity compared to using only the maximal pressure at rest ($p<0.001$), but not sensitivity ($p=0.07$). During squeeze, the combination of all features, was significantly different in specificity than using maximal pressure alone at squeeze ($p<0.001$), but not sensitivity ($p=0.07$).

Inter-modality:

for both modalities, features extracted at rest produced the best clustering result index. Clustering results using the optimal features were not significantly different with respect to sensitivity or specificity ($p>0.05$). Optimal Endoflip feature set differed significantly in specificity compared to HDAM maximal pressure at both, rest ($p<0.001$) and squeeze ($p<0.001$). The optimal Endoflip features (i.e., CSA to pressure ratio at 40ml and 50ml) were different in sensitivity to just using max-rest HDAM pressure ($p<0.01$), but not max-squeeze HDAM pressure ($p=0.35$). Endoflip using pressure at 50ml gave the highest sensitivity for this modality, however, it was not different in sensitivity compared to max HDAM rest pressure ($p=0.15$).

Combining HDAM and Endoflip

After the unsupervised classification and scoring stage, with access to the actual labeling we determined outliers in each group. Among the best clustering results, from the 70 subjects (32 normal and 38 patients), there were 15 outliers (1 normal misclassified) in the Endoflip group and 11 outliers (1 normal misclassified) in the HDAM. Ten of the 15 outliers were shared by both modalities (see table in Suppl file 3). Whether the combination of best performers in each modality improved upon the result was also determined. The best scoring achieved was combining the HDAM textural features at ‘rest’ alongside the CSA/pressure ‘rest’ ratio of Endoflip at 40 and 50ml, using k-means method, which yielded a sensitivity of 0.69 and specificity of 0.97. There was no significant difference with the optimal performers in each modality. The above was also carried out for the same features during squeeze, which gave sensitivity of 0.63 and specificity of 0.826. The combination of, rest and squeeze textural HDAM features, and CSA to pressure ratio at 40ml and 50ml was fed to the clustering algorithms. It achieved a sensitivity of 0.66 and specificity of 0.79, which underperformed compared to the individual modalities.

Discussion

We used the “state of the art” unsupervised machine learning techniques, on a set of data captured by Gourcerol and colleagues to determine the discriminating ability of HDAM and Endoflip to diagnose FI patients[14]. We used novel parameters provided by HDAM and FLIP

in our analysis. Our results show that both Endoflip and HDAM tests are highly specific tests. Moreover, the addition of features in the HDAM analysis adds to the specificity of diagnosing FI. Even with the optimal clustering approach, there were still misclassifications in the patient group. The latter implies that there are factors other than the anal pressure and distensibility that contributes to the etiology of FI.

Using optimal feature-sets, both modalities performed in a similar fashion in diagnosing FI. The regional distribution of pressure (as shown by textural features) and distension patterns of the anal canal were abnormal in FI patient. The combination of the best features from each modality did not improve upon the result. Somewhat unexpected was the finding that the HDAM and Endoflip indices of the anal canal at rest performed better than the squeeze. The strength of our study as compared to the earlier studies is in the data analysis. We used: 1) regional pressure distribution information provided by the HDAM recordings (i.e., pressure is not uniform over the longitudinal extent of the anal canal), this allowed the inclusion of regional connectivity of the HDAM pixels lying in the HPZ zone, into the feature set., 2) unsupervised learning using machine learning methods allowing a completely blinded data analysis (in the earlier studies, HDAM and Endoflip tracings were read by the expert trained in analyzing the tests, who was not blinded to the type of subject (i.e., FI patient or healthy subject)), and finally, 3) different permutations of features provided by Endoflip (i.e., individually, combination, division, and multiplication) were tested.

What are the similarities and differences in the parameters that the HDAM and FLIP measure? Both of these techniques record the anal closure function but in somewhat different manner. The HDAM is a 10mm diameter probe with 256 closely spaced sensors arranged in an axial and circumferential fashion over the 6.4 cm length of the probe. One can record axial and circumferential asymmetry of the anal sphincter HPZ with the HDAM. The HDAM probe is non-compliant and it records isometric contraction of the anal sphincter muscle. On the other hand, using FLIP, one measure the ability of anal canal to close against a distended balloon with a certain pressure in it. Since the balloon is collapsible, a concentric muscular contraction of the anal sphincters against resistance is recorded by the FLIP, which is different from the isometric contraction recorded by the HDAM probe. In the FLIP recordings, the current methods only use the minimal dimension of the anal canal at rest and squeeze and not axial asymmetry of the anal canal even though the latter is recorded in the hour glass shape of the anal HPZ. The circumferential asymmetry of the HPZ is lost in the FLIP measurements because the system is designed to provide only the mean diameter at each axial location, which may be thought of as a limitation of the FLIP. The FLIP system allows one to record the anal sphincter function with different degrees of anal distension, i.e., different lengths of the sphincter muscles. The latter can allow one to study the length-tension function of the anal sphincter muscle[32, 33]. The 40ml and 50ml distensions produce anal diameters which are greater than that of the HDAM probe. The pressure sensors in the HDAM probe are spaced 6mm apart and therefore one can have small error in measuring pressure at the cranial and caudal edges of the anal HPZ, which is not different from the 5mm spacing between the electrodes in the case of FLIP. Even though, in principle these two techniques are quite different, but it is interesting that they provide very similar specificity in FI diagnosis. Traditionally, manometric probes used for the anal pressure measurements are 4–5mm in diameter and generally the squeeze values have provided better discriminatory

ability than the rest, unlike in the present study. The reason for the above may be related to the greater diameter of the HDAM and FLIP probes that result in stronger force generated by the anal sphincter muscles at longer lengths.

Even with the optimal clustering approach, there were still misclassifications in the patient group. The latter implies that there are factors other than the anal pressure and distensibility that contributes to the etiology of FI. Our observations that the outliers in the two groups were same subjects and mostly patients and not normal subjects may have several interpretations; **1)** that some of the normal subjects studied actually had low normal anal sphincter function, which overlapped with the pressure in FI patients. **2)** The etiology of FI in patient outliers was not related to the anal sphincter dysfunction which is in accordance with the multifactorial etiology of FI. Factors such as rectal inflammation, rectal compliance, excessive amount of liquidity of stool, urge incontinence played bigger role in their anal incontinence symptoms. **3)** Even though, overall there were no significant differences in the age, parity and BMI of two groups, it was not a case controlled study and hence a limitation of the database that we used. Our study was not designed to determine the etiology of the FI or anal sphincter dysfunction, which may be considered as a limitation of our study. However, our study underscores the importance of defective anal closure function in the pathogenesis of idiopathic variety of fecal incontinence. Whether the defect in the anal closure function is neurogenic or myogenic or both, we can’t tell.

From the clinician point of view, we wish for all our diagnostic tests to be highly sensitive as well as highly specific. Anal manometry, as a part of the diagnostic work for FI has been in use for a long time, it is reassuring to know that manometry is a fairly sensitive and highly specific test. The new “kid on the block”, i.e., Endoflip is equally sensitive and specific as well. Both are relatively simple tests, and if the costs were comparable, one can use them interchangeably. It is clear though that these tests are not complimentary. The requirement for an ideal screening test for a malignant condition is that it should be highly sensitive. On the other hand, for a benign condition like FI it may be acceptable for the tests to be less sensitive as long as they are highly specific. It is likely that the reason for the lower sensitivity (compared their specificity) of the anal sphincter function testing by Endoflip and HDAM to diagnose FI is related to its multifactorial etiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support: This work was supported by a NIH Grant DK060733, and a grant from the Haute-Normandie Region in the framework of the Young Emerging Clinical Researchers call for projects, Rouen University Hospital.

References

1. Hayden DM, Weiss EG. Fecal Incontinence: Etiology, Evaluation, and Treatment. Clinics in Colon and Rectal Surgery 2011;24:64–70. [PubMed: 22379407]
2. Weledji EP. Electrophysiological Basis of Fecal Incontinence and Its Implications for Treatment. Annals of Coloproctology 2017;33:161–168. [PubMed: 29159162]

3. Saldana Ruiz N, Kaiser AM. Fecal incontinence - Challenges and solutions. World Journal of Gastroenterology 2017;23:11–24. [PubMed: 28104977]
4. Rao SSC. Endpoints for Therapeutic Interventions in Fecal Incontinence: Small Step or Game Changer. Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society 2016;28:1123–1133. [PubMed: 27440495]
5. Rao SSC, Bharucha AE, Chiarioni G, et al. Anorectal Disorders. Gastroenterology 2016:S0016-5085(16)00175-X 10.1053/j.gastro.2016.02.009.
6. Mittal RK, Sheean G, Padda BS, et al. Length Tension Function of Puborectalis Muscle: Implications for the Treatment of Fecal Incontinence and Pelvic Floor Disorders. Journal of Neurogastroenterology and Motility 2014;20:539–546. [PubMed: 25273124]
7. Bharucha AE, Fletcher JG, Harper CM, et al. Relationship between symptoms and disordered continence mechanisms in women with idiopathic faecal incontinence. Gut 2005;54:546–55. [PubMed: 15753542]
8. Read NW, Haynes WG, Bartolo DC, et al. Use of anorectal manometry during rectal infusion of saline to investigate sphincter function in incontinent patients. Gastroenterology 1983;85:105–13. [PubMed: 6852445]
9. Liu J, Guaderrama N, Nager CW, et al. Functional correlates of anal canal anatomy: puborectalis muscle and anal canal pressure. Am J Gastroenterol 2006;101:1092–7. [PubMed: 16606349]
10. Raizada V, Bhargava V, Karsten A, et al. Functional morphology of anal sphincter complex unveiled by high definition anal manometery and three dimensional ultrasound imaging. Neurogastroenterol Motil 2011;23:1013–9, e460. [PubMed: 21951657]
11. Kim YS, Weinstein M, Raizada V, et al. Anatomical Disruption & Length-Tension Dysfunction of Anal Sphincter Complex Muscles in Women with Fecal Incontinence. Diseases of the colon and rectum 2013;56:1282–1289. [PubMed: 24105004]
12. Bharucha AE, Rao SS. An update on anorectal disorders for gastroenterologists. Gastroenterology 2014;146:37–45 e2. [PubMed: 24211860]
13. Zifan A, Ledgerwood-Lee M, Mittal RK. A Predictive Model to Identify Patients With Fecal Incontinence Based on High-definition Anorectal Manometry. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2016;14:1788–1796.e2. [PubMed: 27464594]
14. Gourcerol G, Granier S, Bridoux V, et al. Do endoflip assessments of anal sphincter distensibility provide more information on patients with fecal incontinence than high-resolution anal manometry?
15. Cheeney G, Remes-Troche JM, Attaluri A, et al. Investigation of anal motor characteristics of the sensorimotor response (SMR) using 3-D anorectal pressure topography. Am J Physiol Gastrointest Liver Physiol 2011;300:G236–40. [PubMed: 21109594]
16. Coss-Adame E, Rao SS, Valestin J, et al. Accuracy and Reproducibility of High-definition Anorectal Manometry and Pressure Topography Analyses in Healthy Subjects. Clin Gastroenterol Hepatol 2015;13:1143–50 e1. [PubMed: 25616028]
17. Sorensen G, Liao D Fau - Lundby L, Lundby L Fau - Fynne L, et al. Distensibility of the anal canal in patients with idiopathic fecal incontinence: a study with the Functional Lumen Imaging Probe.
18. McMahon BP, Frokjaer Jb Fau - Kunwald P, Kunwald P Fau - Liao D, et al. The functional lumen imaging probe (FLIP) for evaluation of the esophagogastric junction.
19. Luft F, Fynne L, Gregersen H, et al. Functional luminal imaging probe: a new technique for dynamic evaluation of mechanical properties of the anal canal. Tech Coloproctol 2012;16:451–7. [PubMed: 22936582]
20. Gourcerol G, Granier S, Bridoux V, et al. Do endoflip assessments of anal sphincter distensibility provide more information on patients with fecal incontinence than high-resolution anal manometry? Neurogastroenterol Motil 2016;28:399–409. [PubMed: 26670599]
21. Leroi AM, Melchior C, Charpentier C, et al. The diagnostic value of the functional lumen imaging probe versus high-resolution anorectal manometry in patients with fecal incontinence. Neurogastroenterol Motil 2018.
22. Duda RO, Hart PE, Stork DG. Pattern Classification: Wiley, 2012.

23. Gonzalez RC, Woods RE, Eddins S. Digital Image Processing Using MATLAB: Pearson Prentice Hall Upper Saddle River, New Jersey 2004.
24. Hubert L, Arabie P. Comparing partitions. Journal of Classification 1985;2:193–218.
25. Clifford H, Wessely F, Pendurthi S, et al. Comparison of Clustering Methods for Investigation of Genome-Wide Methylation Array Data. Frontiers in Genetics 2011;2:88. [PubMed: 22303382]
26. Everitt B, Landau S, Leese M, et al. Cluster analysis: Wiley, 2011.
27. Schonlau M Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. Computational Statistics 2004;19:95–111.
28. Bishop CM. Pattern recognition and machine learning: New York : Springer, [2006] ©2006, 2006.
29. Nielsen F Hierarchical Clustering In: Nielsen F, ed. Introduction to HPC with MPI for Data Science. Cham: Springer International Publishing, 2016:195–211.
30. Trajman A, Luiz RR. McNemar $\chi 2$ test revisited: comparing sensitivity and specificity of diagnostic examinations. Scandinavian Journal of Clinical and Laboratory Investigation 2008;68:77–80. [PubMed: 18224558]
31. Rand WM. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 1971;66:846–850.
32. Mittal RK, Sheean G, Padda BS, et al. The external anal sphincter operates at short sarcomere length in humans. Neurogastroenterol Motil 2011;23:643–e258. [PubMed: 21418426]
33. Rajasekaran MR, Jiang Y, Bhargava V, et al. Length-tension relationship of the external anal sphincter muscle: implications for the anal canal function. Am J Physiol Gastrointest Liver Physiol 2008;295:G367–73. [PubMed: 18599590]

Key points:

- High definition anal manometry and functional luminal imaging probe (endoflip) are two novel techniques to assess the anal closure function, however their sensitivity and specificity in diagnosing fecal incontinence (FI) is not clear.
- We used machine learning tools to determine the ideal parameters provided by HDAM and endoflip; and performed blinded analysis on a group of FI patients and normal subjects to determine the sensitivity and specificity of two techniques in diagnosing FI. Our data show that both of these techniques are moderately sensitive but highly specific in the FI diagnosing; but they are not complimentary.
- We provide ideal parameters that are useful in distinguishing FI patients from controls when using HDAM and endoflip. Our findings highlight the significance of defective anal closure function as a major contributor to the pathophysiology of FI.

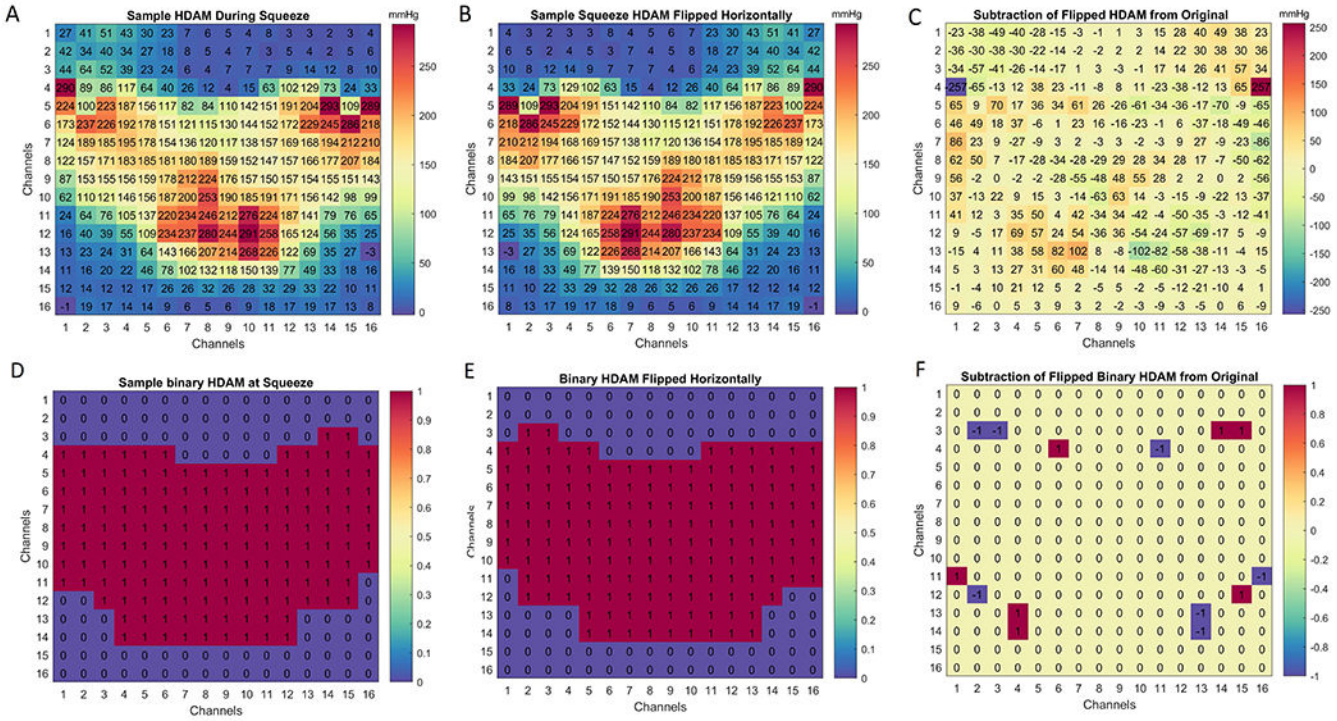


Figure 1.

ASI illustration, **(A-C)** the original raw unfolded image is flipped horizontally, and the results subtracted from it, producing the top right subtracted image panel. Note, the binary ASI **(D-F)** is also extracted in a similar way, however, this time only the binary mask prior to the calculation.

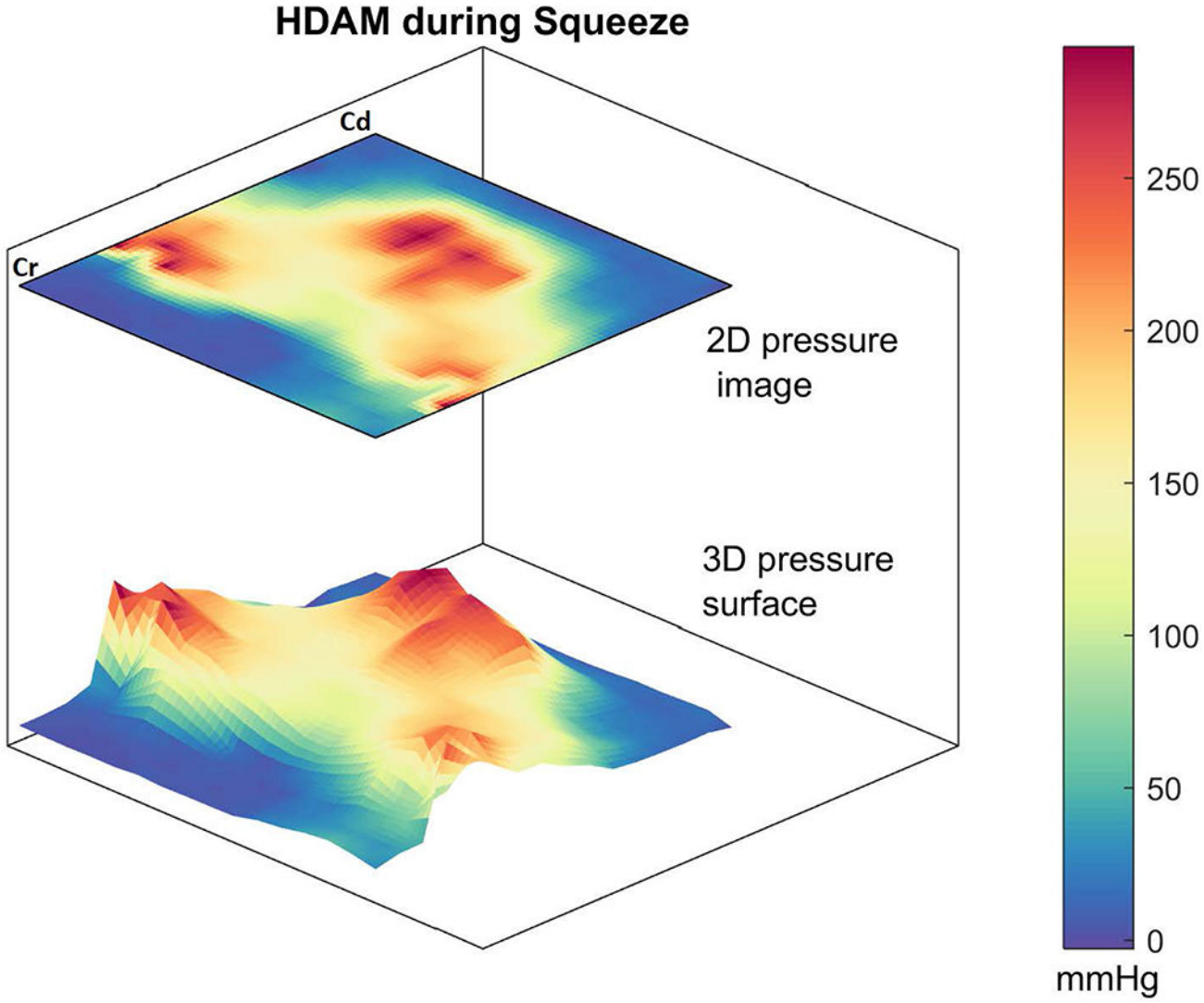


Figure 2.
A sample normal HDAM image during squeeze, and its corresponding 3D pressure surface.

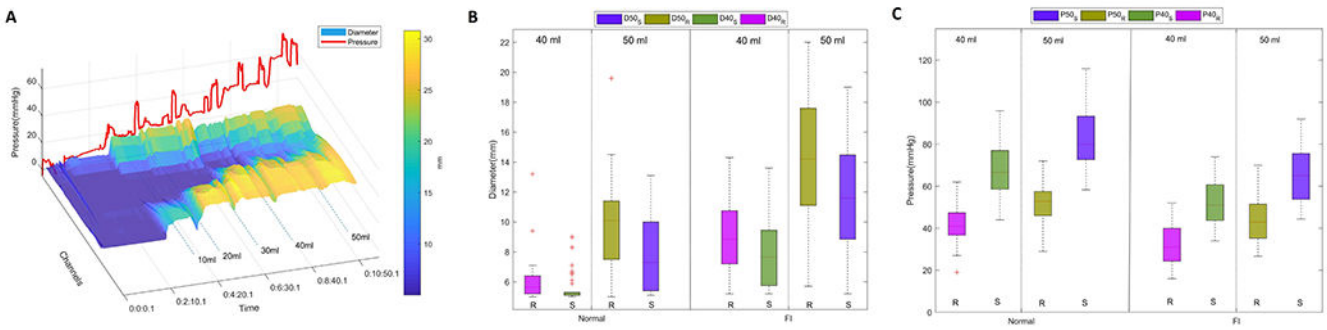


Figure 3.

(A) A sample subject undergoing through the Endoflip stepped volume distension protocol. The subject was asked to perform a voluntary contraction following a 30-s rest period. The pressure change for this recording is shown as the red waveform, and the surface topograph represents the change in the balloon diameter. The changing colors from dark blue to yellow illustrate increasing diameter. The subject squeeze can be identified from the peaks of the red line, and the corresponding reduction in diameter of the anal canal and color change toward dark blue on the diameter topograph. **(B)** Boxplots showing the distribution of diameter values at 40ml and 50ml across the two groups, **(C)** Boxplots showing the distribution of balloon pressure values at 40ml and 50ml across the two groups.

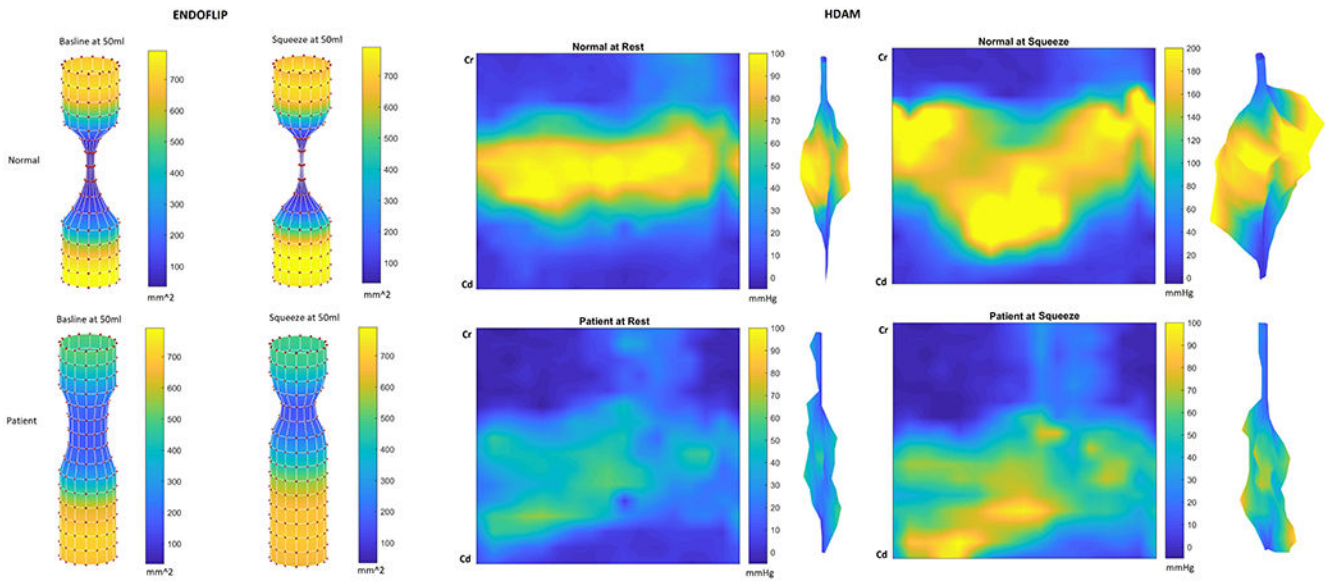


Figure 4.

Sample reconstruction of both HDAM and Endoflip modalities for both a normal and FI patient. For HDAM, aside the unfolded version of the cylindrical reconstruction, the 3D extruded HDA version is also shown. For Endoflip, cylindrical reconstruction of the diameter variation is shown for both rest and squeeze states. Note, the variation in topology between the two groups, especially the higher diameter baseline for the FI patient.

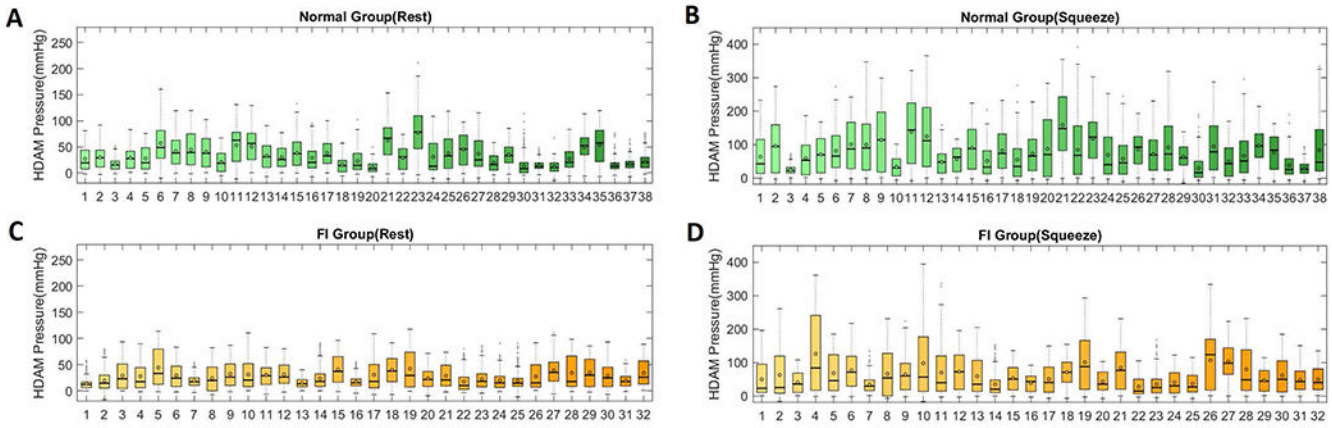


Figure 5.

Boxplot showing the distribution of raw HDAM pressure values for the normal group at rest **(A),** and squeeze **(B).** The same pressure variation for the FI population during rest, **(C)** and **(D)** squeeze.

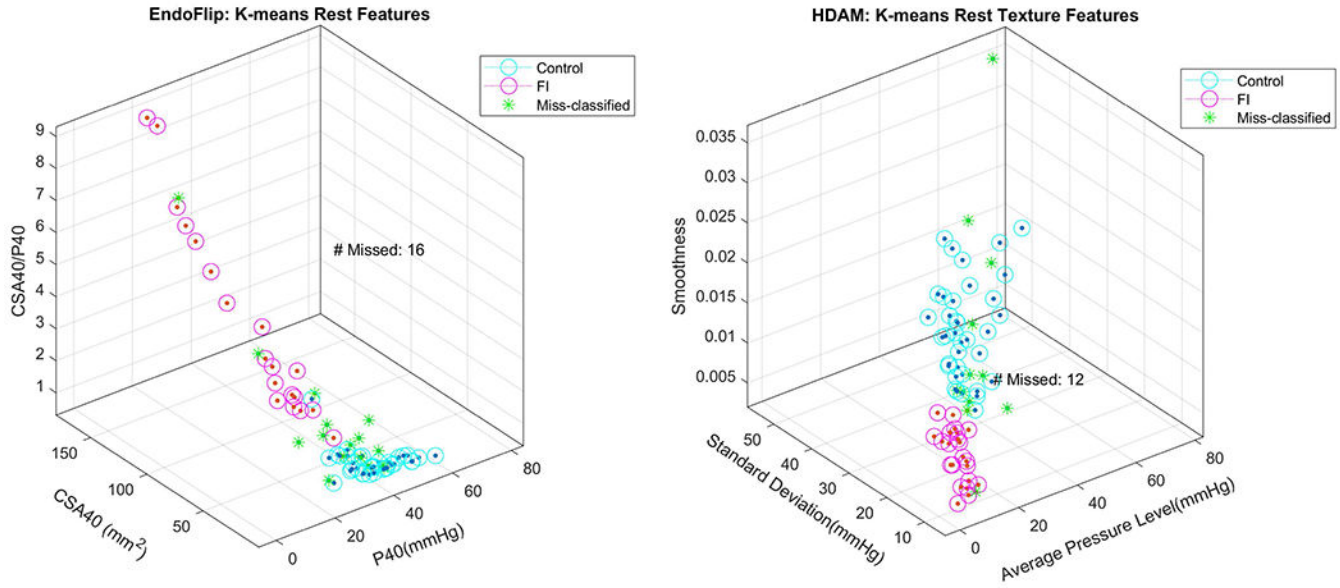


Figure 6.

(A) Sample clustering results using Kmeans, for Endoflip using a combination of rest features, namely p40, CSA40, p40/CSA40 & p50, CSA50, and p50/CSA50. **(B)** Sample clustering results using Kmeans, for the textural features described in Figure 5A. Note, for visualization purposes the actual multi-dimension space has been projected to only three-dimensions (x-y and z-axes).

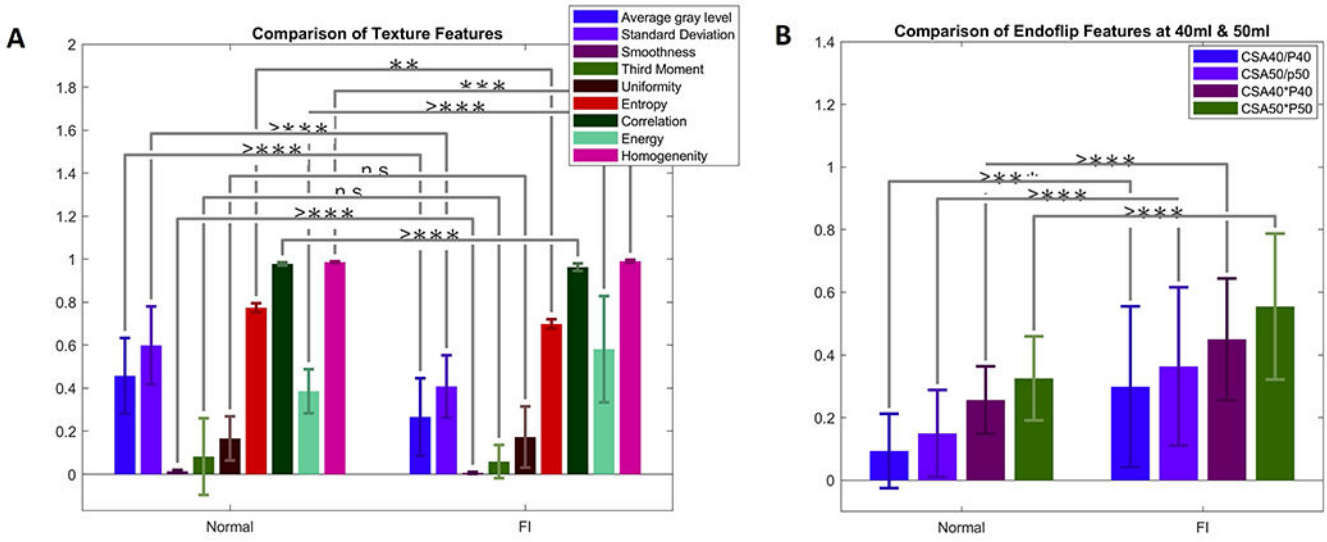


Figure 7.

(A) Comparison of textural features for both groups. In this figure a single star (*), denotes p-values less than 0.05, a double star (**) denotes p-values less than 0.01, a triple star denotes values under 0.0001, and finally indicated with (e.g.) '>***' instead of '****', to show the maximum measured precision has been exceeded. Note for HDAM, aside from Uniformity and Third moment (as described in the method section), all the other texture features were statistically different from eachother. **(B)** Multiplication and division of the Endoflip features, i.e., pressure and CSA at 40 and 50ml.

Table 1.

Summary estimates of sensitivity and specificity of the top performing clustering results, indicating e the sensitivity and specificity, accuracy with corresponding 95% CIs.

		Features	Method	Sensitivity	Specificity
HDAM	Squeeze	All Features	K-means Cosine	53.12% (40.9% - 65.0%)	89.47% (79.3% - 95.6%)
		Texture	k-means Cityblock	62.5% (50.1% - 73.6%)	81.58% (70.1% - 89.7%)
		Max squeeze	k-means Cityblock	68.8% (56.4% - 79.1%)	63.16 (50.7% - 74.2%)
	Rest	Texture	k-means Cityblock	68.76% (56.4% - 79.1%)	97.37% (89.5% - 100.0%)
		All Features	K-means Cityblock	68.75% (56.4% - 79.1%)	89.47% 79.3% - 95.6%)
		Max rest	K-means Cityblock	87.5% (76.9% - 94.2%)	52.6% 40.4% - 64.6%)
Endoflip	Squeeze	CSA/P40 and CSA/P50	K-means Cityblock	59.38% (47.0% - 70.8%)	89.47% 79.3% - 95.6%)
		P50	Hierarchical Cityblock	68.75% (56.4% - 79.1%)	81.58% (70.1% - 89.7%)
		CSA/P50	K-means Sqeuclidean	43.75% (32.1% - 56.1%)	100% (93.5% - 100.0%)
		CSA/P40	K-means Cityblock	53.13%(40.9% - 65.0%)	92.11%82.5% - 97.3%)
	Rest	CSA/P40 and CSA/P50	K-means Cityblock	56.25% (43.9% - 67.9%)	97.37% 89.5% - 100.0%)
		All Features	K-means Cityblock	59.38% (47.0% - 70.8%)	92.11% 82.5% - 97.3%)
		CSA/P50	K-means Cityblock	37.5% (26.5% - 49.8%)	97.37% 89.5% - 100.0%)
		P50	K-means Sqeuclidean	56.25% (43.9% - 67.9%)	78.95% (67.2% - 87.6%)