Virtual disease landscape using mechanics-informed machine learning: application to esophageal disorders

Sourav Halder[1], Jun Yamasaki[2], Shashank Acharya[2], Wenjun Kou[3], Guy Elisha[2], Dustin A. Carlson[3], Peter J. Kahrilas[3], John E. Pandolfino[3], Neelesh A. Patankar[1,2]

[1] Theoretical and Applied Mechanics Program, McCormick School of Engineering, Northwestern University, Evanston IL, USA.

[2] Department of Mechanical Engineering, McCormick School of Engineering, Northwestern University, Evanston IL, USA.

[3] Division of Gastroenterology and Hepatology, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago IL, USA.

Running title: Esophageal virtual disease landscape

Corresponding author:

Neelesh A. Patankar
Professor, Department of Mechanical Engineering
Northwestern University
2145 Sheridan Road
Evanston IL, 60208-3111
Room: Tech L490
E-mail: n-patankar@northwestern.edu
Phone: 847-491-3021 (O)
Fax: 847-491-3915
Web: http://patankar.mech.northwestern.edu

Grant support: This work was supported by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Grants R01-DK079902, P01-DK117824 (to J.E. Pandolfino) and by the National Science Foundation (NSF) grants OAC 1450374 and OAC 1931372 (to N.A. Patankar).

Declaration of interest:

Peter J. Kahrilas, and John E. Pandolfino hold shared intellectual property rights and ownership surrounding FLIP panometry systems, methods, and apparatus with Medtronic Inc.

Dustin A. Carlson: Medtronic (Speaking, Consulting)

Wenjun Kou: Crospon, Inc. (Consulting)

Peter J. Kahrilas: Ironwood (Consulting), Reckitt (Consulting), Johnson & Johnson (Consulting)

John E. Pandolfino: Crospon, Inc (stock options), Given Imaging (Consultant, Grant, Speaking), Sandhill Scientific (Consulting, Speaking), Takeda (Speaking), Astra Zeneca (Speaking), Medtronic (Speaking, Consulting), Torax (Speaking, Consulting), Ironwood (Consulting), Impleo (Grant)

S. Halder, J. Yamasaki, S. Acharya, N.A. Patankar: None

Author contributions: S.H., N.A.P., P.J.K., J.E.P., contributed to study conception and design; S.H. conceptualized and designed the simulation framework; S.H., J.Y. performed simulations and analyzed data; S.H., S.A., W.K., D.A.C., N.A.P., P.J.K., J.E.P., interpreted results of calculations; S.H. prepared figures and drafted the manuscript; D.A.C., W.K. contributed to data acquisition; J.E.P., P.J.K. and N.A.P. contributed to obtaining funding, critical revision of the manuscript and final approval.

ABSTRACT

Esophageal disorders are related to the mechanical properties and function of the esophageal wall. Therefore, to understand the underlying fundamental mechanisms behind various esophageal disorders, it is crucial to map mechanical behavior of the esophageal wall in terms of mechanics-based parameters corresponding to altered bolus transit and increased intrabolus pressure. We present a hybrid framework that combines fluid mechanics and machine learning to identify the underlying physics of various esophageal disorders (motility disorders, eosinophilic esophagitis, reflux disease, scleroderma esophagus) and maps them onto a parameter space which we call the virtual disease landscape (VDL). A one-dimensional inverse model processes the output from an esophageal diagnostic device called the functional lumen imaging probe (FLIP) to estimate the mechanical “health” of the esophagus by predicting a set of mechanics-based parameters such as esophageal wall stiffness, muscle contraction pattern and active relaxation of esophageal walls. The mechanics-based parameters were then used to train a neural network that consists of a variational autoencoder that generated a latent space and a side network that predicted mechanical work metrics for estimating esophagogastric junction motility. The latent vectors along with a set of discrete mechanics-based parameters define the VDL and formed clusters corresponding to specific esophageal disorders. The VDL not only distinguishes among disorders but also displayed disease progression over time. Finally, we demonstrated the clinical applicability of this framework for estimating the effectiveness of a treatment and tracking patients’ condition after a treatment.

KEYWORDS:

Achalasia, FLIP, convolutional neural network, computational fluid dynamics, dysphagia, variational autoencoder

1. INTRODUCTION

Medical diagnoses are often made using devices to measure physical quantities such as pressure, fluid velocity, or tissue geometry and deformation. Such measurements serve as surrogates for fundamental physiomarkers such as tissue integrity and neuromuscular function that define the health of an organ. In fact, since the fundamental physiomarkers are often unknown or unmeasurable, clinical decisions often need to be based on these physical quantities. However, such measurements never precisely track the relevant physiomarkers and this can lead to discrepancies. Experimental and computational frameworks that combine the diagnostic measurements with the physical laws that govern them could potentially minimize these discrepancies leading to more accurate clinical decisions. In this work, we present a novel hybrid approach that uses machine learning and principles of fluid mechanics to process raw data generated from an esophageal diagnostic device to develop a set of mechanics-based parameters as fundamental physiomarkers of disease in a patient-specific manner.

A widely used test for evaluating esophageal dysphagia is high-resolution manometry (HRM) [1-5]. HRM measures swallow-induced pressures at multiple sites within the esophageal lumen including the esophagogastric junction (EGJ) to make clinical diagnoses according to the Chicago Classification v4.0 (CCv4.0) [6], the current worldwide standard classification of esophageal motility disorders. A newer technology for investigating esophageal motility is the functional luminal imaging probe (FLIP) [7, 8] which assesses the response of the esophagus to distention. Figure 1 is a schematic diagram of the FLIP probe incorporating 16 impedance planimetry sensors to measure esophageal luminal cross-sectional area and a pressure sensor at its distal end. The sensors are housed within a compliant bag that is incrementally filled with saline. During measurements, the FLIP probe is passed trans-orally and positioned across the EGJ. Distending the esophagus with the FLIP bag normally induces a contractile response including periods of repetitive antegrade contractions (RACs), but alternative patterns (or no

contractile response) can be seen in patients with esophageal motility disorders or other disease states [9]. Compared to HRM, which evaluates primary peristalsis, FLIP evaluates secondary peristalsis which is physiologically different. Carlson et al. [10] have shown that subjects with normal esophageal motility on HRM exhibit abnormal EGJ opening with FLIP. Other studies [11, 12] have also shown that, unlike other diagnostic tests, FLIP can be used to estimate the in-vivo mechanical properties of the esophageal wall. Savarino et al. [13] and Carlson [14] have published reviews of the clinical applicability of FLIP. Additionally, FLIP has been shown to be used intraoperatively [15, 16] during endoscopic or laparoscopic myotomy to calibrate the procedure. Figure 2 shows the typical FLIP measurements of cross-sectional areas in a normal subject as compared to a type I and III achalasia subject.

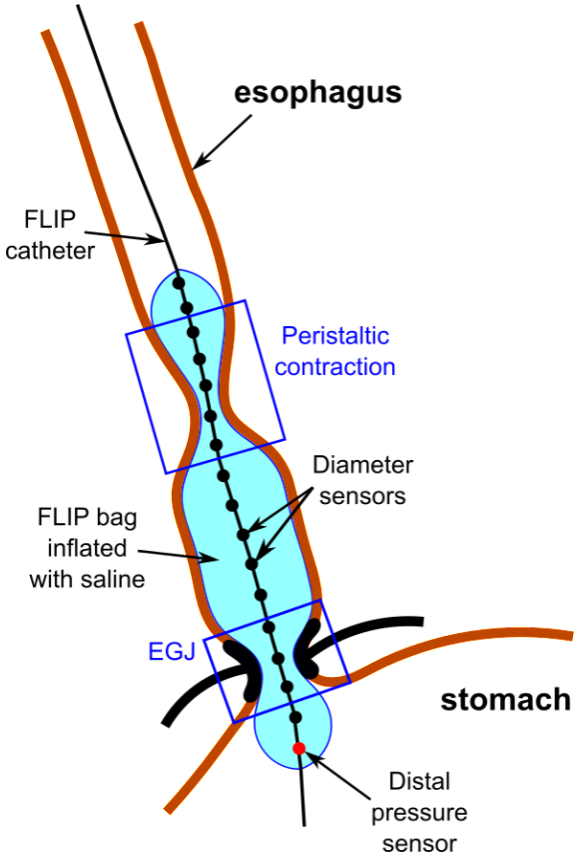


Figure 1: Schematic diagram of the FLIP probe. The secondary peristaltic contraction and EGJ tone are evident inside the blue boxes. EGJ tone results from a complex interaction between the lower esophageal sphincter and the crural diaphragm.

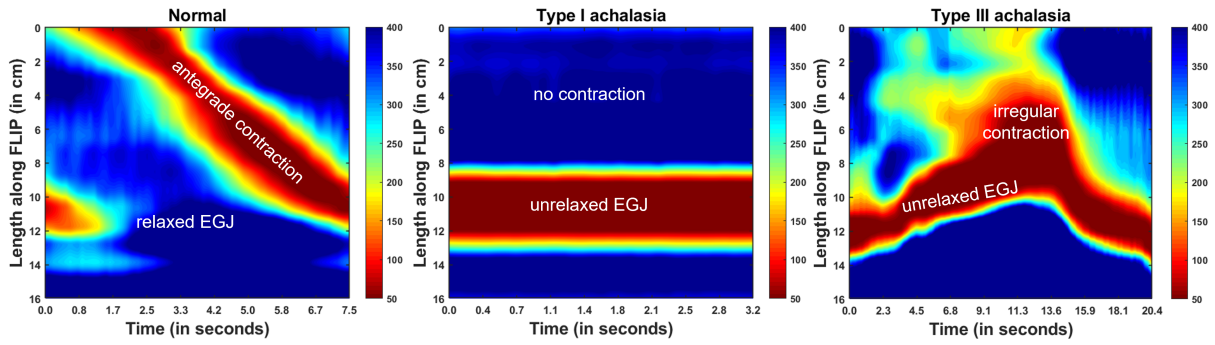


Figure 2: Variation of cross-sectional area (in mm^2) along the FLIP length in a typical (a) normal, (b) type I, and (c) type II achalasia subject. The low cross-sectional areas (in red) show the contraction pattern and EGJ tone, while the high cross-sectional values (in blue) show the relaxed regions. The normal subject shows an antegrade contraction with a relaxed EGJ, type I achalasia shows no contractility and an unrelaxed EGJ, and type III achalasia shows unrelaxed EGJ with irregular contractions in the esophageal body.

In this work, we present a framework that works with FLIP measurements to add precision to clinical interventions. We used the pressure and diameter data obtained from FLIP to calculate mechanics-based parameters such as esophageal wall properties, muscle contraction strength, EGJ tone, and active relaxation of the esophageal musculature. Esophageal biomechanics have been extensively studied using both experimental [17-22] and computational [23-31] approaches. For our analysis, we used the mathematical framework as described in Halder et. al. [26] to calculate the mechanics-based parameters since it worked with clinical fluoroscopy data obtained from the esophagus and makes rapid predictions with limited computational resources.

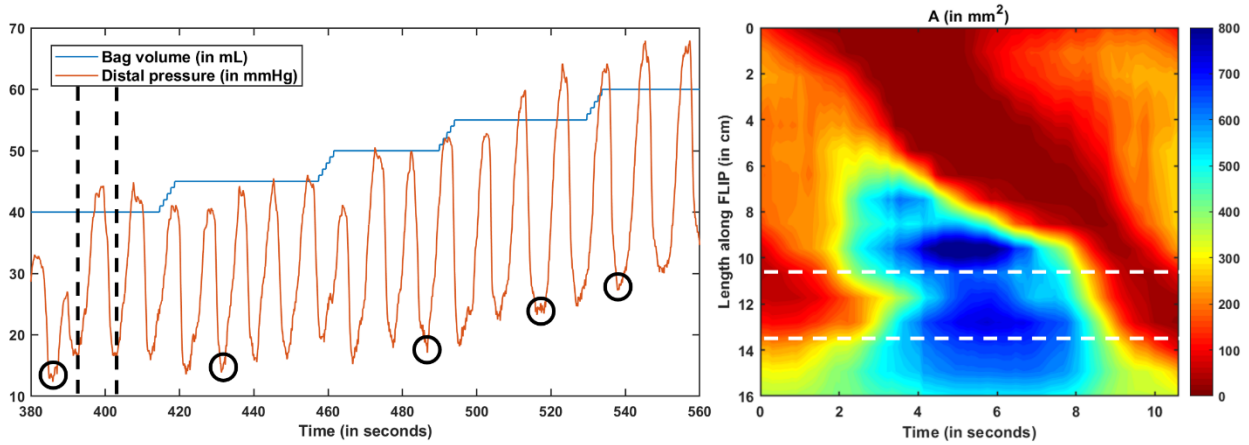
The mechanics-based parameters estimate the mechanical “health” of the esophagus in a patient-specific manner. However, identifying unique patterns in these parameters for specific esophageal disorders is challenging, especially with parameters that are functions of both time and location along the esophagus. It is also important that the methodology adjust to errors in the FLIP device operation such as probe positioning and potential discrepancies introduced by human input such as manually specifying values for mechanics-based parameters. This is analogous to manually identifying landmarks on pressure topography plots during the

interpretation of HRM studies. This challenging task is tackled using machine learning which has been widely applied in medical diagnosis [32-41]. Machine learning techniques have been used both for medical image analysis and raw patient data analysis. In gastroenterology, machine learning has been used mainly for image segmentation and classification tasks [42-45]. The exception is a recent study [46] demonstrating the use of a variational autoencoder (VAE) [47] to identify contractility patterns from raw HRM data. The clusters generated in the latent space of the VAE categorized the raw HRM data into patient groups corresponding to specific motility disorders. However, although the data clusters were beneficial for diagnosis, they do not have a discrete physical meaning. In this work, we present a novel framework, called mechanics-informed variational autoencoder (MI-VAE), which forms clusters in a parameter space corresponding to specific esophageal disorders and these clusters do have physical meaning because they were generated from mechanics-based parameters. We call this parameter space the Virtual Disease Landscape (VDL).

2. METHODS

Distention of the FLIP probe usually elicits esophageal contractions that its sensors characterize by variations of cross-sectional area along the distal 10-13 cm of the esophagus (including the EGJ) and pressure within the probe measured at its distal end. Using these outputs, we estimated the mechanical "health" of the esophagus and identified patterns of output observed with esophageal motility disorders and other esophageal diseases. This was done in two steps: 1) using an inverse model to estimate the mechanical "health" of the esophagus by calculating parameters such as esophageal wall properties, contraction strength, active relaxation, work done while opening the EGJ (EGJW) and work required to open the EGJ (EGJROW) [48], and 2) using the calculated mechanics-based parameters as input to a VAE which generates the VDL in the form of its latent space. The next two subsections discuss these steps in details.

2.1. Mechanics model



(a) (b)

Figure 3: Example of graphical output from a normal FLIP study. (a) Bag volume and pressure in the FLIP probe. The pressure variations are caused by antegrade contractions. The black circles indicate minimum distal pressure at each fill volume. These low-pressure points are assumed to correspond to $\theta \sim 1$, (b) Iso-area contour plot of cross-sectional area data for the time interval delineated by the dashed lines in (a). The EGJ is located between the dashed white lines. It is seen to distend as the contraction moves from the proximal to the distal end of the FLIP probe.

Figure 3 illustrates output from a normal FLIP study during a period of RACs as the distention volume is incrementally increased at about 30 second intervals. With a RAC pattern, repetitive antegrade contractions are observed at a frequency of 6±3 per minute. Each antegrade contraction leads to a rise in bag pressure (red tracing in Figure 3a). Output from the 16 impedance planimetry sensors are displayed as an iso-area topography plot in Figure 3b by interpolating data between sensors. The deep red band in Figure 3b is an antegrade contraction which is associated with an increase in cross-sectional area at the EGJ shown by the blue region near the distal end. Note that since FLIP output is of cross-sectional area there is no information regarding the actual three-dimensional geometry of the esophageal lumen. Hence, for simplicity and to conserve computational resources we modeled the FLIP as a one-dimensional flexible tube.

2.1.1. Governing equations

The one-dimensional mass and momentum equations that describe fluid flow through a flexible tube [49-52] are as follows:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = 0, \tag{1}$$

$$\frac{\partial Q}{\partial t} + \frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + \frac{A}{\rho}\frac{\partial P}{\partial x} + \frac{8\pi\mu Q}{\rho A} = 0, \tag{2}$$

where A is the cross-sectional area of the FLIP, Q and P are the flowrate and pressure in the fluid inside the FLIP, respectively. x and t represent the position on the probe and time, respectively. ρ and μ are the density and dynamic viscosity of the fluid, respectively. In equations (1) and (2), A was known for all x and t.

The fluid pressure inside the esophagus has been found to be linearly proportional to its cross-sectional area [11, 12]. Hence, it is possible to relate the fluid pressure inside the esophagus with its stiffness as shown below:

$$P = P_{ex} + K\left(\frac{A}{\theta A_{ref}} - 1\right), \tag{3}$$

where K is the stiffness of the esophageal walls, P_{ex} is the pressure outside the esophagus (mostly thoracic pressure), A_{ref} is the relaxed cross-sectional area of the esophageal lumen, and θ is the activation parameter. When the esophagus is in its inactive state, $\theta = 1$. At this state, $A = A_{ref}$, and $P = P_{ex}$ i.e., the pressure inside the esophagus is equal to the pressure outside it. When $\theta < 1$, a contraction is induced in the esophagus and the pressure rises. $\theta > 1$ indicates active relaxation characterized by reduced pressure and increased cross-sectional area.

The non-dimensional form of equations (1)-(3) can be written as follows:

$$\frac{\partial \alpha}{\partial \tau} + \frac{\partial q}{\partial \chi} = 0, \tag{4}$$

$$\frac{\partial q}{\partial \tau} + \frac{\partial}{\partial \chi}\left(\frac{q^2}{2}\right) + \alpha \frac{\partial}{\partial \chi}\left(\frac{\alpha}{\theta}\right) + \varphi \frac{q}{\alpha} = 0, \tag{5}$$

wherein $\chi = \frac{x}{L}$, $\tau = \frac{ct}{L}$, $q = \frac{Q}{cA_0}$, $\alpha = \frac{A}{A_{ref}}$, $p = \frac{P}{\rho c^2}$, $A_0 = \frac{\pi D_0^2}{4}$, and $\varphi = \frac{12\mu L}{\rho c D_0^2}\left(\frac{1}{\theta}\right)$. Equation (3) was used to replace P in equation (2) and non-dimensionalized to obtain equation (5). The velocity scale c was taken to be 3 cm /s which is typically the speed of a peristaltic contraction. The ratio K/A_{ref} can be considered as a measure of the stiffness of the esophageal walls.

2.1.2 Numerical implementation

The FLIP is closed at its two ends, so it is necessary to enforce zero flow rate boundary conditions at $\chi = 0$ and $\chi = 1$. Since equation (4) requires only one boundary condition for q, we differentiate it with respect to χ to obtain a second order form as follows:

$$\frac{\partial^! \alpha}{\partial \tau \partial \chi} + \frac{\partial^! q}{\partial \chi^!} = 0. \tag{6}$$

We do the same for equation (5) to specify a Dirichlet boundary condition for pressure at the distal end (as measured by the pressure sensor) and zero pressure gradient at the proximal end as typically observed in practice. Equation (5) takes the following form after differentiating with respect to χ:

$$\frac{\partial^! q}{\partial \tau \partial \chi} + \frac{\partial^!}{\partial \chi^!} / \frac{q^!}{\alpha} 0 + \frac{\partial}{\partial \chi} \mathrm{M} \alpha \frac{\partial}{\partial \chi} \mathrm{C} \frac{\alpha}{\theta} \mathrm{DN} + \varphi \frac{\partial}{\partial \chi} \mathrm{C} \frac{q}{\alpha} \mathrm{D} = 0. \tag{7}$$

With α known for all χ and τ, equation (6) was solved to calculate q. Following this, equation (7) was solved to calculate $\frac{1}{2}$ with the known values of α and the calculated values of q. Equations (6) and (7) are solved using the finite volume method as described in Halder et. al. [26] on a grid as shown in Figure 4.

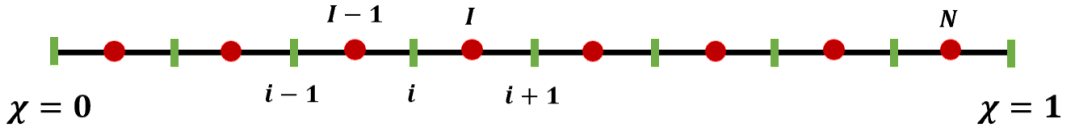


Figure 4: Schematic of the discretized domain. The green markers show the cell boundaries, and the red markers show the cell centers.

The discretized form of equations (6) and (7) are shown below:

$$q_3 - \frac{1}{2}(q_{345} + q_{365}) = \frac{\Delta \chi}{2 \Delta \tau}(\alpha_7 - \alpha_{765} - \alpha_7^{"} + \alpha_{765}^{"}), \tag{8}$$

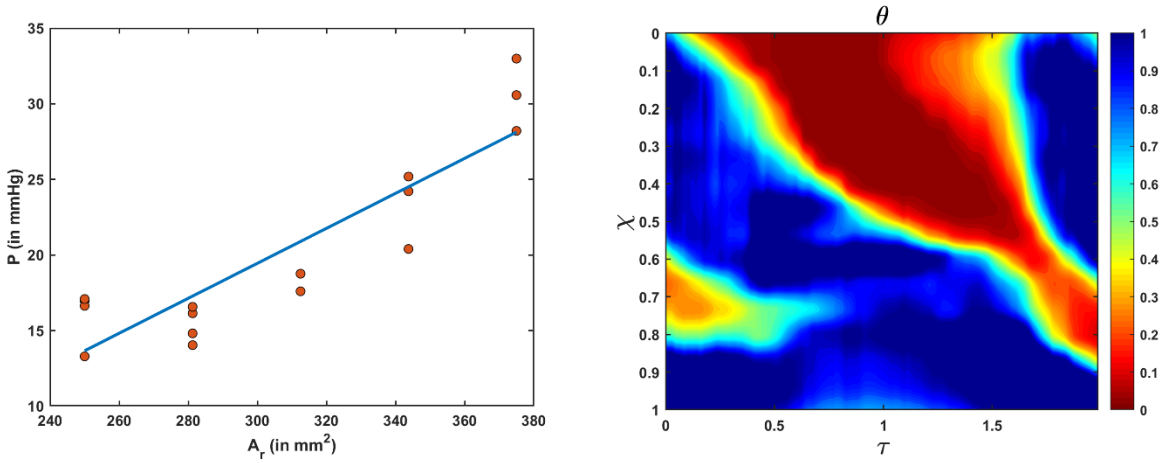
$$(\alpha_7 + \alpha_{765}) \mathrm{C} \frac{\alpha}{\theta} \mathrm{D}_3 - \alpha_7 \mathrm{C} \frac{\alpha}{\theta} \mathrm{D}_{345} - \alpha_{765} \mathrm{C} \frac{\alpha}{\theta} \mathrm{D}_{365} = \frac{\Delta \chi}{2 \Delta \tau}(q_{345} - q_{365} - q_{345}^{"} + q_{365}^{"}) + / \frac{q_{345}^!}{\alpha_{345}} + \frac{q_{365}^!}{\alpha_{365}} - \frac{2 q^!}{\alpha_3^3} 0 + \frac{\varphi}{2} 8 \frac{q_{345}}{\alpha_{345}} - \frac{q_{365}}{\alpha_{365}} :, \tag{9}$$

where, $i, I = 1,2, \ldots, N$. The non-dimensional cross-sectional areas α were known from the impedance sensors.

2.1.3 Initial and boundary conditions

Equations (6) and (7) were used to solve for q and $\frac{1}{2}$ respectively. Note that α and θ were grouped here for simplicity. Once $\frac{1}{2}$ is known, calculating θ is straightforward with the known values of α. During operation, the FLIP probe is a flexible tube closed at both ends. Since equations (6) and (7) are second-order in χ, we need two boundary conditions for q and $\frac{1}{2}$ at $\chi = 0$ and $\chi = 1$. Additionally, we need one initial condition for q since equation (7) is first-order in τ. At $\tau = 0$, we specified $q = 0$ for all χ. Since the two ends of the FLIP bag are closed, $q = 0$ at $\chi = 0$ and $\chi = 1$. Additionally, with the distal pressure (P_8), $P = P_8(t)$ at $\chi = 1$. Therefore at $\chi = 1$, the corresponding value of $\frac{1}{2}$ is $\frac{*_{\&}6(*\$6+)}{-\%\emptyset^{\#}}$. At $\chi = 0$, we specified $\frac{;}{;<}\mathrm{C}\frac{1}{2}\mathrm{D} = 0$. This follows directly from equation (5) since $q = 0$ at $\chi = 0$ for all values of τ.

2.1.4. Calculation of the primary mechanics-based parameters



(a) (b)

Figure 5: Mechanics-based parameters. (a) The variation FLIP pressure with the reference cross-sectional area. The slope and y-intercept of the fitted line yield $K/A_{"}$ and $P_{"} - K$, respectively. (b) Variation of the activation parameter corresponding to the cross-sectional area variation shown in Figure 3(b).

Solving equations (6) and (7) requires the knowledge of the parameters $K/A_{"}$ and $P_{"} - K$ in a patient-specific manner. They were calculated using the approach as described in Acharya et al.

[48]. As described in equation (3), P and A are linearly proportional to each other. Thus, in a scenario where the esophagus is inactive (which corresponds to $\theta = 1$) and distended, a plot of P vs. A should be a straight line with its slope and intercept as K/A_0 and $P_0 - K$, respectively. This inactive but distended esophagus would take the shape of a cylinder with a reference cross-sectional area $A_=$ as follows:

$$A_= = \frac{V}{L}, \tag{10}$$

where V is the FLIP bag volume and L is the length of the FLIP. But identifying the corresponding pressure inside the esophagus is not so straightforward. Usually, the time instants at which the distal pressure (P_8) readings are the lowest values for every bag volume, correspond to the time instants when it is reasonable to assume $\theta \sim 1$ as shown in Figure 3(a). We selected these pressures readings and the corresponding reference cross-sectional areas $A_=$ and fitted a straight line between them, and the slope and intercept of this line estimated K/A_0 and $P_0 - K$, respectively as shown by Figure 5(a). Note that there is only one value for $A_=$ for each bag volume but multiple pressure readings based on the time instant at which the pressure was considered. This is why there is a vertical spread of data in Figure 5(a) for each value of $A_=$.

With K/A_0 known, we calculated the scaling cross-sectional area $A_,$ and the viscosity parameter φ in equation (7). With $(P_0 - K)$ known, and P_8 known at the distal end, we calculated the boundary condition for $\frac{1}{2}$ as described in Section 2.1.3.

In general, the mechanics model described above works on FLIP data for selected segments of recordings, for instance, the time taken for an antegrade contraction to traverse the probe and merge with the EGJ as shown in Figure 3b.

2.1.5 Physical significance of the primary mechanics-based parameters

In summary, there are seven primary mechanics-based parameters: a measure of stiffness (K/A_m), an estimate of the external pressure ($P_m - K$), the maximum distal pressure recorded during the selected time interval of interest ($P_{>?@}$), the time taken by an antegrade contraction to traverse the FLIP (T), FLIP bag volume (V), and the contraction and relaxation pattern described through the activation parameter ($\theta(x,t)$). The parameters K/A_m and $P_m - K$ were calculated in a patient-specific manner and had a single value for each patient, while the other parameters had different values for each patient based on the time interval of data considered. The parameter K/A_m estimates the mechanical properties of the esophagus and helps in identifying the relation between the wall properties with the esophageal function. The parameter $P_m - K$ quantifies the state outside the esophagus. The maximum distal pressure $P_{>?@}$ is the net effect of the contraction strength and EGJ tone, and hence is an important parameter in estimating esophageal motility. The parameter T indirectly estimates the speed of an antegrade contraction i.e., slower contractions have higher T and vice versa, thus estimating an important feature of the esophageal function. The bag volume V controls the extent to which the esophagus is distended which impacts on the passive behavior of the esophageal walls. The activation parameter estimates the contraction strength and pattern, the EGJ tone, and the active relaxation of the esophageal walls. The activation parameter essentially drives the mechanics of esophageal transport and in the context of the FLIP, it helps estimate esophageal contractility. Active relaxation can be estimated through the maximum value of θ, $\theta_{>?@}$, which was considered as another primary mechanics-based parameter because active relaxation aids in bolus transport and EGJ relaxation. We call these parameters primary since they completely define the mechanical state of the esophagus, and other mechanics-based parameters can be calculated using combinations of them.

2.1.6. Calculation of the secondary mechanics-based parameters: EGJ work metrics

The EGJ work metrics as described in Acharya et. al. [48] were considered for secondary parameters. The first EGJ work metric i.e., the work done in opening the EGJ is defined as follows:

$$EGJW = \int_{t_0}^{t_1} \int_{x_0}^{x_1} P \frac{\partial A}{\partial t} dx dt, \quad (11)$$

where t_0 and t_1 are the time instants between which the EGJ opens from its least to greatest diameter, and x_0 and x_1 are the proximal and distal boundaries of the EGJ. The second EGJ work metric i.e. the work required to open the EGJ is defined as follows:

$$EGJROW = \int_{A_0}^{A_1} \left(\frac{K}{A^2 \theta} \right) A + (P^2 - K) (x_1 - x_0) dA, \quad (12)$$

where A_0 and A_1 are the least and greatest reference cross-sectional areas at the EGJ as measured on FLIP corresponding to diameters of 3 mm and 22 mm, respectively. Since θ is a function of both x and t, $EGJROW$ is also a function of x and t. For simplicity, we choose three values of $EGJROW$ corresponding to: 1) θ calculated at t_0 ($EGJROW_0$), 2) median value of θ calculated between t_0 and t_1 ($EGJROW_1$), and 3) minimum value of θ calculate between t_0 and t_1 ($EGJROW_A$).

2.2. Mechanics-informed variational autoencoder

The mechanics-based parameters calculated by the 1D inverse model yield a quantitative estimate of the mechanical “health” of the esophagus through the wall mechanical properties and esophageal contractility. Identifying similarities and dissimilarities of these parameters across patient groups is a crucial step in the development of the VDL. This was done in an unsupervised manner with the help of a VAE. Since this neural network works

entirely on the mechanics-based parameters, we call it the mechanics informed variational autoencoder (MI-VAE).

2.2.1. Network Architecture

The mechanics-based parameters are as follows: $\theta(x,t)$, $K/A_{"}$, $P_{"}-K$, $P_{>?@}$, T, V, $\theta_{>?@}$, $EGJW$, and the 3 measures of $EGJROW$. These are all scalar values except θ, which varies with x and t so simple statistics can be used to identify patterns of these quantities. However, identifying patterns quantitatively with the activation parameter θ, which describes the esophageal contractility, requires a different approach. Since the variation of θ takes the form of a matrix, as shown in Figure 5b, we used a convolutional neural network-based VAE, which we called network 1, to identify the unique patterns of disordered contractility through the generated latent space.

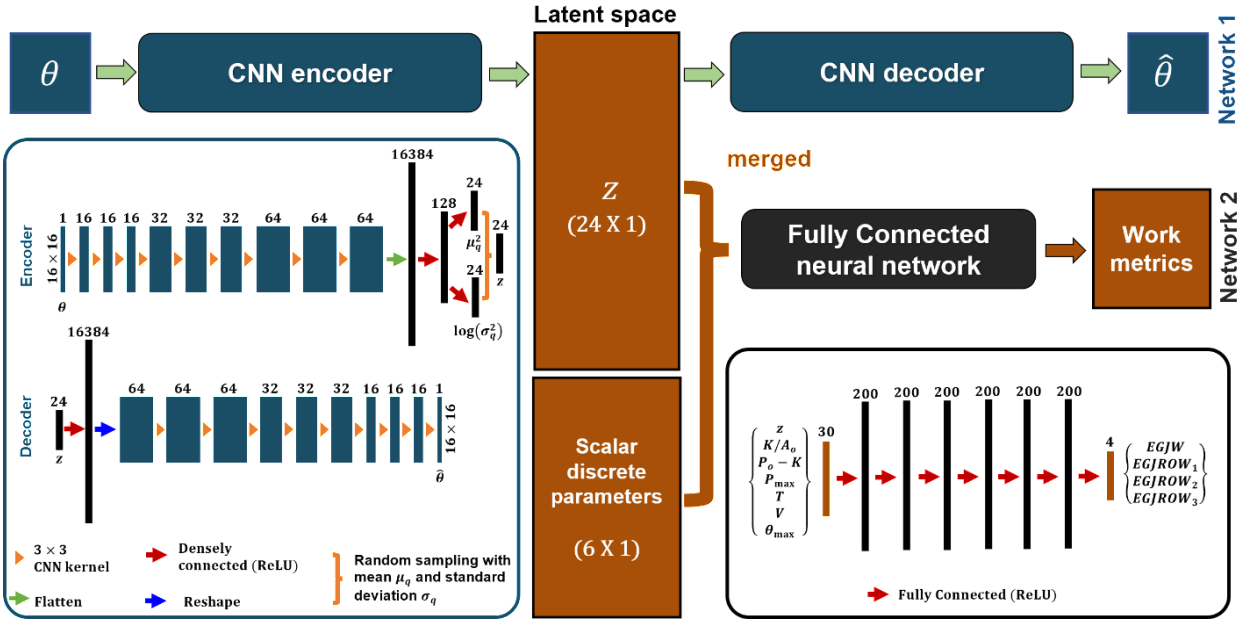


Figure 6: Network architecture of the mechanics-informed variational autoencoder. In the architecture for network 1, the numbers on the top of the boxes represent the number of channels; the numbers and the output size are represented on the side. For network 2, the numbers on the top represent the number of hidden units.

We used a latent space of 24 dimensions. We used ReLU as the activation function for all layers of network 1. Additionally, we merged a 6-dimensional vector consisting of a set of discrete parameters to the 24-dimensional vectors generated in the latent space of network 1. Combined, these 30-dimensional vectors populate a parameter space that forms the VDL. The details of the network architecture are shown in Figure 6.

The combined 30-dimensional vector generated with the latent space of network 1 and the vector of discrete parameters became input to a second neural network, which we called network 2, to predict $EGJW$ and the 3 measures of $EGJROW$. Network 2 consisted of a densely connected neural network of 3 hidden layers with 75 hidden units each. We used ReLU as the activation function for all layers of network 2. Together, networks 1 and 2 form the MI-VAE that develops the VDL and gives physical significance to the vectors in the VDL.

The work metrics, $EGJW$ and $EGJROW$, were not included in the 30-dimensional vectors of VDL but kept as separate entities to be predicted by network 2. This was done so that the VDL was based entirely on primary mechanics-based parameters which define the mechanical state and functioning of the esophagus. The work metrics are secondary mechanical parameters that were kept separate to avoid unnecessary biasing of VDL. Network 2 not only predicts the work metrics, but also forms a framework to test different derivable parameters that can potentially be used as physiomarkers for disease states.

2.2.2. Data

FLIP data used to train the MI-VAE were collected from a cohort of 804 volunteer subjects and patients, the details of which are provided in Table 1. Note that we have included the ‘HRM inconclusive’ to investigate how they appear on FLIP, which is potentially different. The MI-VAE was trained without the disease labels. Hence, the inclusion of the ‘HRM inconclusive’ has no impact on the training but increases the generalizability of the predictions.

Table 1: Cohort details that contributed to training MI-VAE

Group	**# Subjects**
Normal	237
Type I achalasia (Ach 1)	76
Type II achalasia (Ach 2)	148
Type III achalasia (Ach 3)	47
EGJ outflow obstruction (EGJOO)	27
Hypercontractile esophagus (HC)	13
Distal esophageal spasm (DES)	11
Ineffective esophageal motility (IEM)	44
Absent contractility (AC)	17
Eosinophilic esophagitis (EoE)	45
Gastroesophageal reflux disease (GERD)	11
Scleroderma (SSc)	5
Inconclusive (Inc)	123

Pressure vs distention plots (e.g. Figure 3a) from each FLIP study were visually inspected to isolate 6-12 areas of interest for analysis (e.g. Figure 3b), characterized by antegrade contractions and/or maximal pressure variations. HRM studies were also done on all subjects and patient groups were based on their CCv4.0 HRM diagnoses or medical history. Within this cohort, there were 24 achalasia patients with data available before and after treatment with pneumatic dilation, laparoscopic Heller myotomy (LHM) or peroral endoscopic myotomy (POEM) and 20 eosinophilic esophagitis (EoE) patients with data before and after treatment with swallowed topical steroid. Additionally, there was one achalasia patient for whom post-POEM data were available for tracking (years 1, 4, and 7). The distributions of the primary and secondary mechanical parameters are shown in Figures 7 and 8, respectively. The work metrics show trends similar to those reported in Acharya et. al. [48].

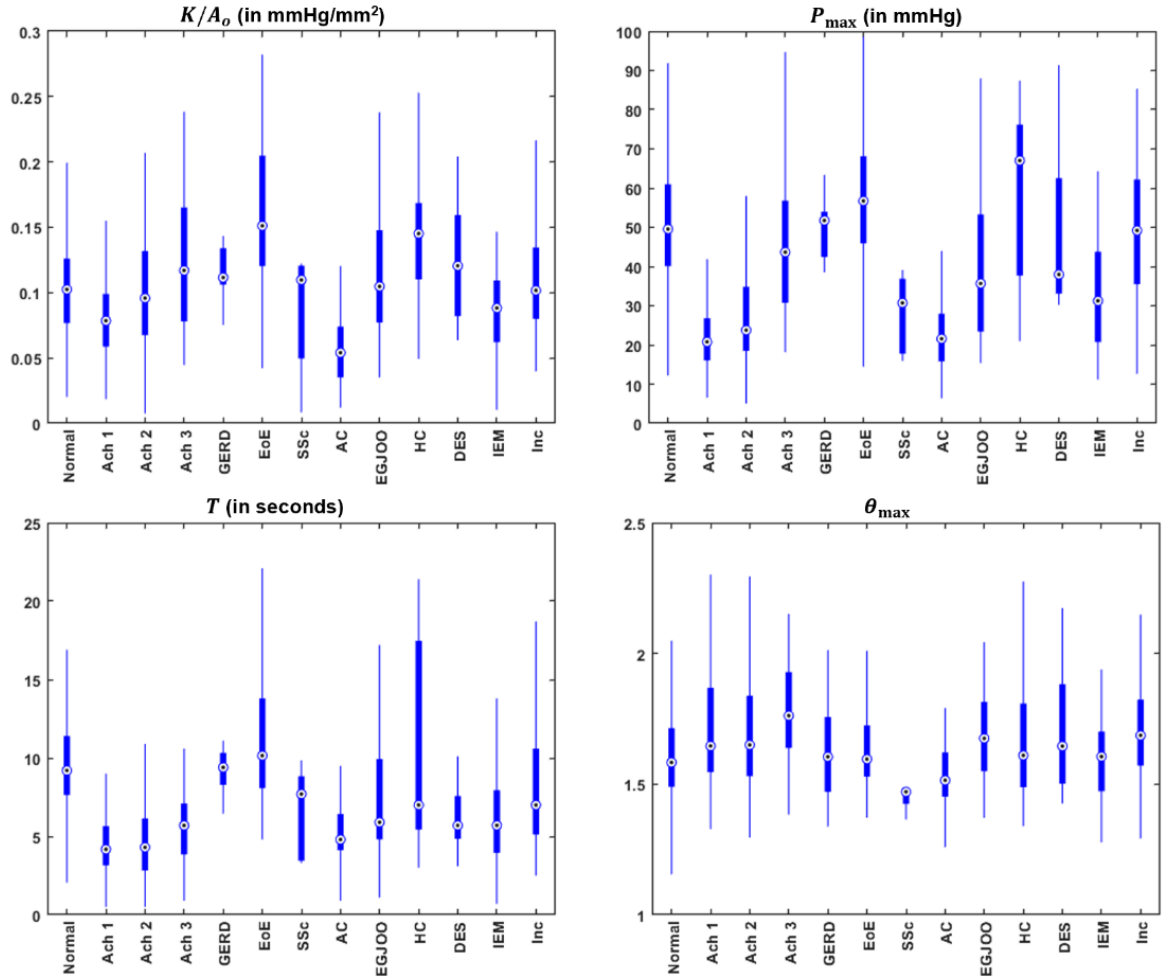


Figure 7: Box plots showing the distribution of the 4 mechanics-based parameters: $K/A_{"}$, $P_{\mathrm{B\ C\#}}$, T, and $\theta_{\mathrm{BC\#}}$. The groups are: Normal esophageal motility, Type I achalasia (Ach 1), Type II achalasia (Ach 2), Type III achalasia (Ach 3), gastroesophageal reflux disease (GERD), eosinophilic esophagitis (EoE), scleroderma (SSc), absent contractility (AC), EGJ outflow obstruction (EGJOO), hypercontractile esophagus (HC), distal esophageal spasm (DES), ineffective esophageal motility (IEM), and inconclusive (Inc).

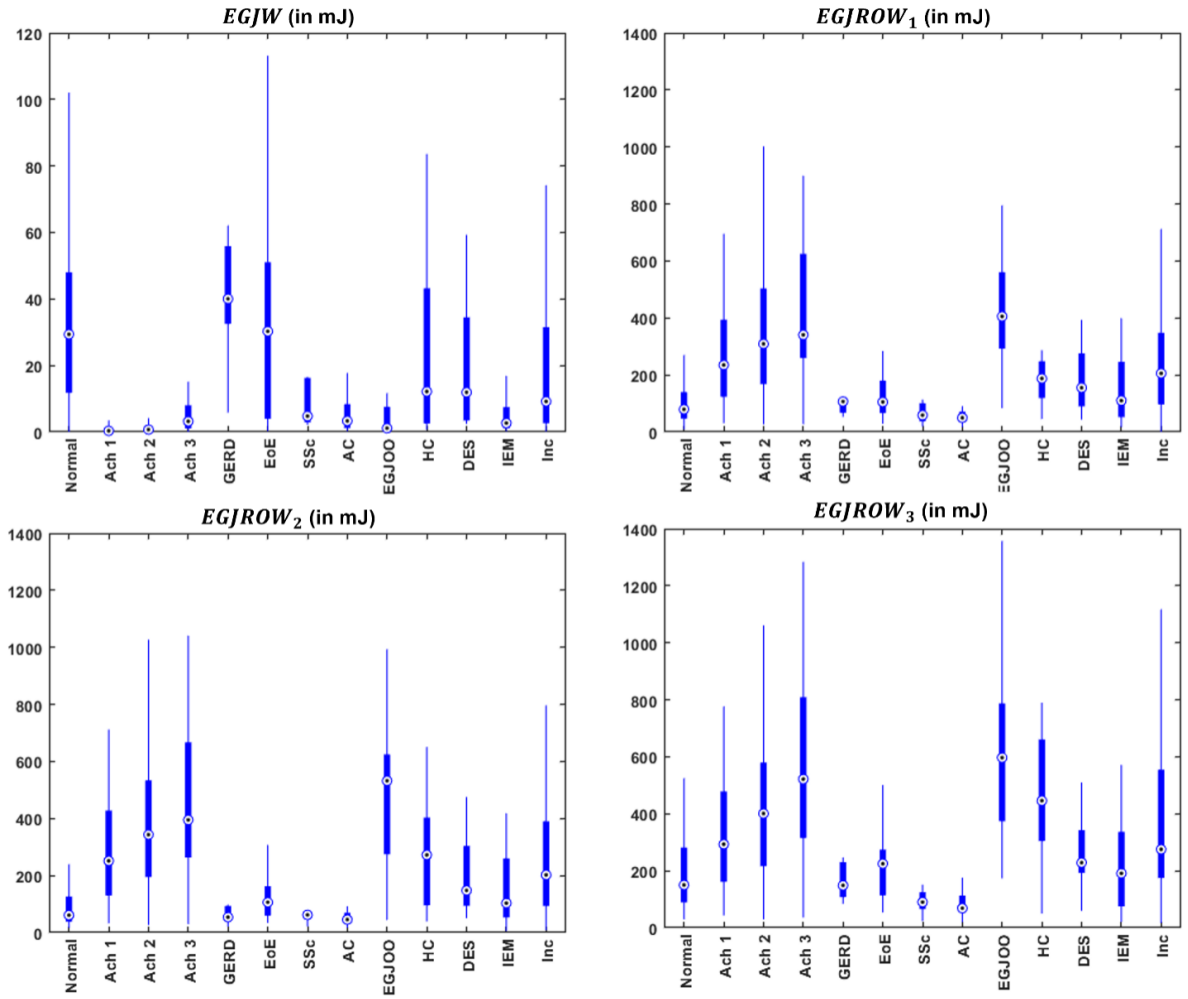


Figure 8: Box plots showing distribution of the EGJ work metrics. The groups are: Normal esophageal motility, Type I achalasia (Ach 1), Type II achalasia (Ach 2), Type III achalasia (Ach 3), gastroesophageal reflux disease (GERD), eosinophilic esophagitis (EoE), scleroderma (SSc), absent contractility (AC), EGJ outflow obstruction (EGJOO), hypercontractile esophagus (HC), distal esophageal spasm (DES), ineffective esophageal motility (IEM), and inconclusive (Inc).

The patient cohort generated a total dataset of size 7,187 parameter sets (mean of 8.9 per subject). This was augmented to generate a larger dataset of 222,797 parameter sets. The parameters $K/A_{"}$, $P_{"} - K$, P_8, and $T_{>?@}$ were augmented by multiplying with a factor f calculated as follows:

$$f = 1 + 0.05\mathbb{N}, \tag{13}$$

where $\mathbb{N}$ is a random sampling from a normal distribution with its magnitude less than 2. The raw cross-sectional area variation $A(x,t)$ was augmented using a combination of grid distortional, elastic transformation, and motion blur available in the opensource Python library Albumentations [53], the details of which are provided in Table 2.

Table 2: Augmentation details for cross-sectional area

Augmentation type	**Albumentations parameters**
Grid distortion	p=0.9, num_steps=4, distort_limit = (-0.2,0.2)
Elastic transformation	p=0.8, alpha=5.0, sigma=100, alpha_affine = 2.0
Motion blur	p=0.7, blur_limit = (3,6)

Using the augmented cross-sectional areas and the augmented values for $K/A_{"}$, $P_{"}-K$, P_8 and T, we solved equations (6) and (7) to obtain $\theta(x,t)$. Finally, the corresponding $\theta_{>?@}$ and $P_{>?@}$ were calculated by taking the maximum of θ and P_8, respectively.

2.2.3. Training and prediction

The final form of the loss function used for network 1 is as follows:

$$L_{\mathrm{D5}} \coloneqq \frac{1}{2M}\mathrm{g}_{3}^{\mathrm{G}}\ \mathrm{h1}+\log\sigma_{\mathrm{E},3}^{!}-\sigma_{\mathrm{E},3}^{!}-\mu_{\mathrm{E},3}^{!}\mathrm{m}+\frac{\beta}{N}\mathrm{g}_{\mathrm{H}}^{\mathrm{D}}\ \mathrm{o}\theta_{\mathrm{H}}-\theta_{\mathrm{H}}\mathrm{q}^{!}, \tag{14}$$

wherein i represents each component of the latent space and j represents each component of the matrices θ and θ. M is the dimension of the latent space and N is the product of the two spatial dimensions of the input and generated output. In this case, $M=24$ and $N=256$. The first term is the Kullback-Leibler divergence (KLD), and the second term is the reconstruction loss. Here $q(z|\theta)$ is the approximate posterior distribution which is a Gaussian distribution with a mean and standard deviation of μ_{E} and σ_{E}, respectively. The details of the derivation of the KLD term are provided in the Appendix. β is a scaling parameter used to balance the magnitudes of the reconstruction loss and the KLD for proper training of network 1. We found that $\beta=1000$

resulted in a good balance between the two losses. The loss function shown in equation (14) is described for each input. While training, we defined the total loss as the mean of L_{D5} calculated over the mini-batch dataset.

We trained network 1 first followed by network 2 with the mean latent space variables and the discrete mechanics-based parameters merged as its input. The input of network 1 i.e., θ, was scaled to lie between 0 and 1, and then subtracted from 1 so that the lesser values at the contraction zones would have greater values instead. Thus, network 1 focusses on minimizing the reconstruction error at the contraction zones since they have the most impact in the variation of θ. Of the 222,797 mechanics-based parameters, 204,000 were used for training and the remainder for testing. Network 1 was trained for 250 epochs. We used a learning rate of $1 \times 10^{6\mathrm{I}}$ for the first 100 epochs, $3.3 \times 10^{6\mathrm{J}}$ for the next 100 epochs, and finally $5 \times 10^{6\mathrm{K}}$ for the final 50 epochs. Adam [54] was used as the optimizer. Network 1 achieved an average mean-squared error between input and generated output as $1.36 \times 10^{6\mathrm{A}}$ and a KLD loss of 1.54.

The input and output of network 2 are shown in Figure 6. The input and output were scaled to lie between 0 and 1. We used a mean-squared error loss function for network 2. Of the total dataset, 2048 were used for validation and the remainder for training. To train this network, we used a learning rate of 10^{-3} and used RMSprop as optimizer. Network 2 was trained for 1000 epochs, and it achieved a mean-squared error accuracy of $1.88 \times 10^{6\mathrm{J}}$ on the validation dataset. We used Keras [55] running on top of TensorFlow [56] to train network 1 and 2. The learning curves for network 1 and 2 are provided in the Appendix.

2.3. Post-processing

The 30-dimensional VDL was reduced to 3 dimensions for visualization. This dimensional reduction was done using two methods: linear discriminant analysis (LDA) and principal component analysis (PCA). LDA is a supervised approach whereby HRM CCv4.0

diagnoses were used. LDA minimizes the distance between VDL points with the same CCv4.0 diagnosis and maximizes the distance between points with different diagnoses. This was done by finding directions in the high-dimensional space that most effectively separated the data into groups. Alternatively, PCA is an unsupervised approach that reduces dimensions by projecting the 30-dimensional VDL vectors onto 3 vectors with the greatest variances in data. We used an open source Python library Scikit-learn [57] to perform dimension reduction using LDA and PCA. The proportion of variance attributable to the greatest three principal components in PCA were 0.151, 0.100, and 0.074, respectively. The proportion of variance attributable to each of the selected components in LDA were 0.639, 0.154, and 0.076. Note that the proportion of variance for LDA is different from that of PCA. In PCA, the proportional of variance attributable to the 3 principal components are the 3 greatest eigenvalues of the covariance matrix for the whole dataset. In LDA, the proportion of variance attributable to its 3 components are the 3 greatest eigenvalues of the product of the inverse of the within-class covariance matrix and the between-class covariance matrix. Note that when we used LDA for the dimension reduction, we did not use the ‘inconclusive’ label for training, but only the conclusive diagnoses on HRM. The LDA trained on conclusive diagnoses was used to predict on the ‘inconclusive’ data. Thus, the inconclusive data labels did not add noise to our analysis.

LDA and PCA are both based on linear models to find optimal projection from high-dimensional space to low-dimensional space, in which the former aims to maximize the cross-class variation constrained by in-class variation, and the latter aims to maximize the variation without class information (i.e., unsupervised). Singular value decomposition (SVD) on covariance matrix will be identical to PCA. Compared with other techniques such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP), which involve nonlinear projection (or dimension reduction), LDA and PCA are simpler and more easily explained. Moreover, the current projection starts from latent space after data

transformation by encoder model. The entire transformation from raw data to projected space is actual a composite mapping, consisting of an encoder mapping followed by projection mapping. With a well-trained sophisticated encoder model, nonlinear transformation, by t-SNE or UMAP is unnecessary, although they may be more useful when directly projecting the raw data. Also, t-SNE and UMAP creates a low-dimensional visualization of the high-dimensional data through a process of iteration. This leads to not only a general increase in computational time, but also necessitates the search and selection of the maximum number of iterations. These two methods have additional parameters including perplexity (in the case of t-SNE) and number of neighbors (UMAP), which require interpretation by the user and tuning. The resulting visualizations may differ based on the parameters. It is also worth noting that the initialization of t-SNE is random, causing potential issues with robustness/repeatability. Last but not the least, we used LDA and PCA just to have a simple visualization tool after the latent space is generated using the MI-VAE. The actual VDL and its application is in the higher dimension of 30 and does not change based on how the data are visualized.

We also generated probability-based boundaries of the VDL points corresponding to different disease groups to quantify the distribution of patient groups as captured in the VDL. This is essentially a supervised classification task where a probability is assigned to every VDL point for each of the 12 patient groups shown in Table 1. This was done using a Random Forest classifier labeling each 30-dimensional VDL vector with the patient diagnosis. For this training, only the original dataset was used and not the augmented ones since the diagnoses of the augmented ones were unknown. 25% of the original dataset of 6244 were used for testing and the remainder for training. We used Scikit-learn for the Random Forest analysis. We tried all combinations of the numbers of estimators (100, 250, 500, 1000, 2000, 3000) and maximum depth (64, 128, 256, and nodes are expanded until all the leaves are pure or all leaves contain less than 2 samples) of each tree using the Scikit learn python library. We selected 1000

estimators and the maximum depth of 128 because they gave the greatest accuracy (Subset accuracy or Exact match of 0.74) on the test dataset. We also trained a Random Forest classifier to predict the probability of a point in the VDL to be an antegrade contraction or not. The parameters of the classifier as well as the train-test split were the same as described above. Each data point was manually labelled as 1 for antegrade contraction or 0 for not an antegrade contraction. The Jaccard score on this test set was 0.94. Further details on accuracy are provided in the Appendix. The trained Random Forest classifiers add another capability to the MI-VAE framework to automate the diagnostic process such that it can predict a diagnosis as well its peristaltic behavior without manual intervention.

3. RESULTS AND DISCUSSION

3.1. Virtual disease landscape in reduced dimensions

The VDL was populated with points corresponding to the original dataset rather than the augmented data since only their diagnoses were known. Figures 9 and 10 show the dimension reduced VDL generated by LDA and PCA, respectively.

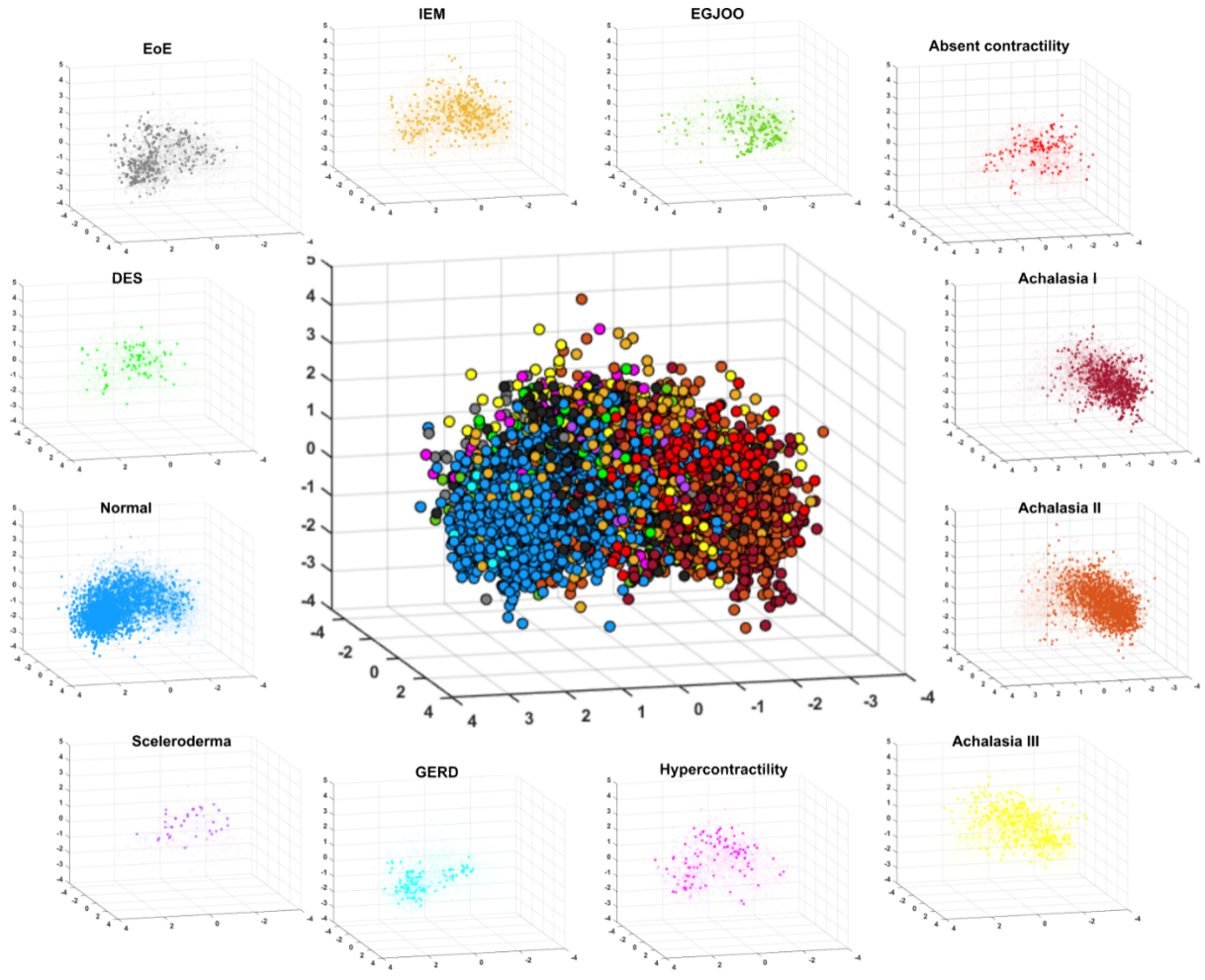


Figure 9: Dimension-reduced VDL using LDA. Patient groups are shown with different colors and each group is shown in isolation in the smaller plots. The black points are the 'inconclusive' subjects.

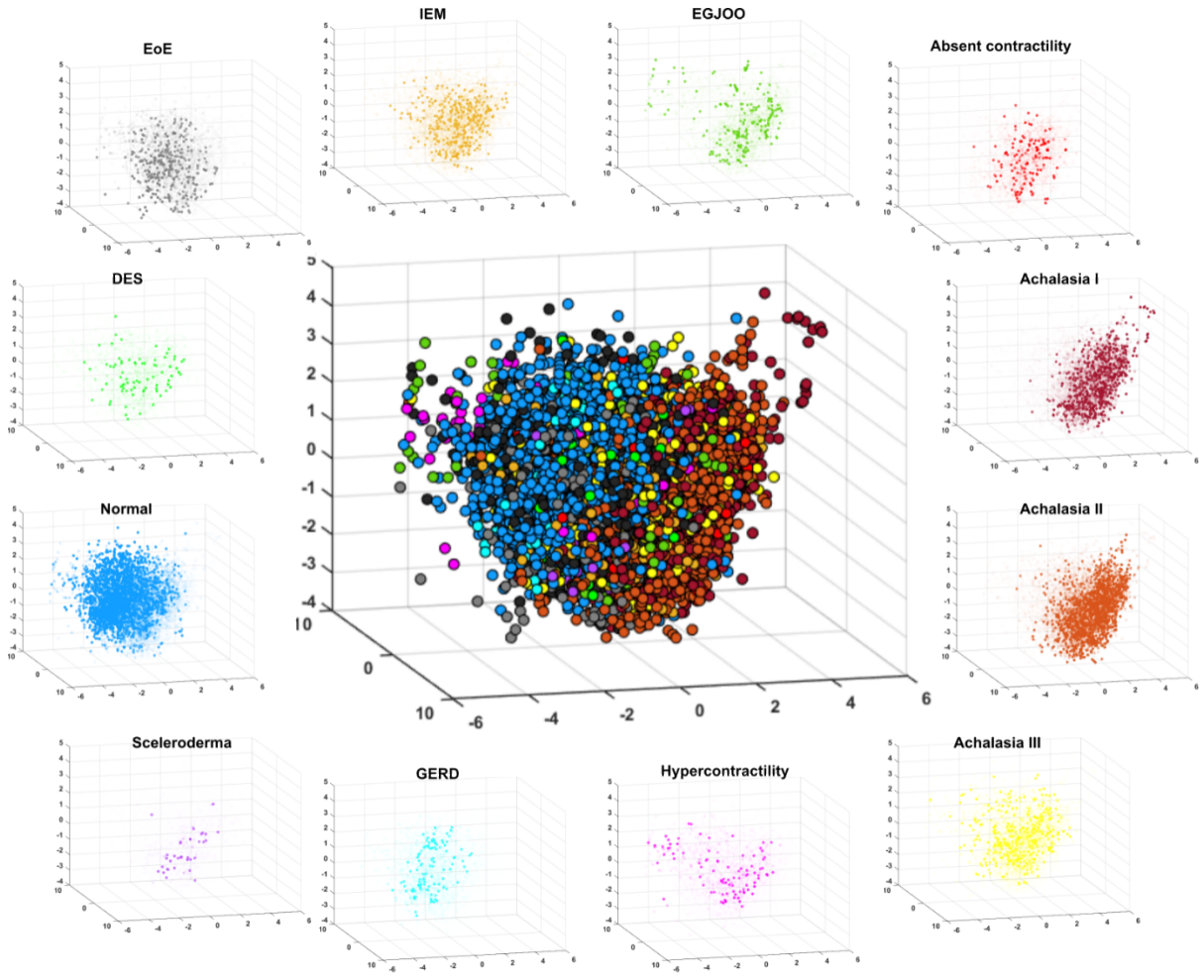
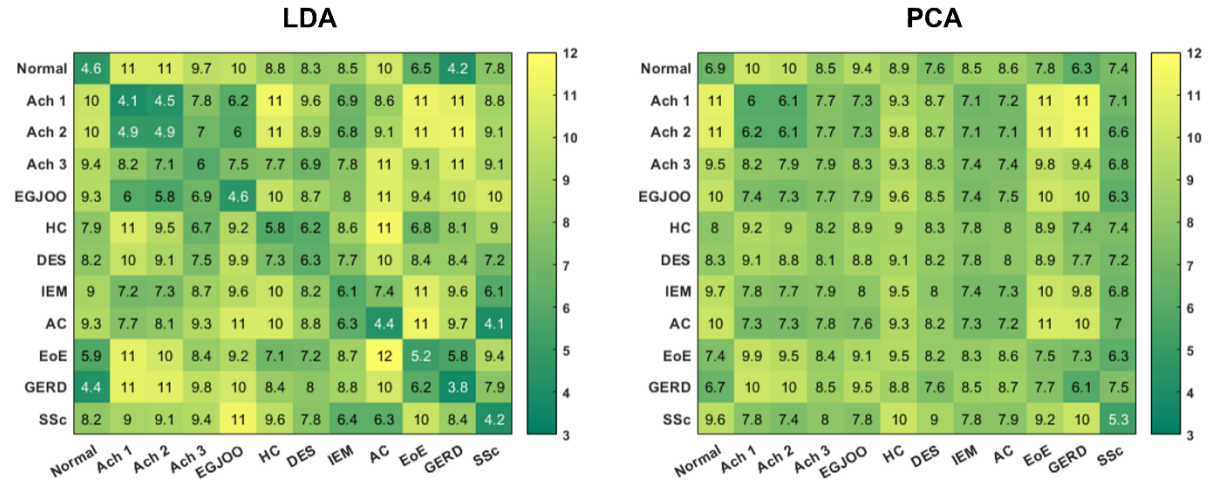


Figure 10: Dimension-reduced VDL using PCA. Patient groups are shown with different colors and each group is shown in isolation in the smaller plots. The black points are the 'inconclusive' subjects.

The patient groups clustered into different regions of the VDL, but these clusters overlapped. As shown in Figure 9 with the dimension reduced VDL using LDA (ldaVDL), most of the normal subjects, GERD, and EoE patients lay on the left side of the VDL, whereas achalasia, EGJOO, and absent contractility patients lay on the right side. The other diseases were distributed between the extremes. The overlap and separation among groups mirrored the similarity of their contractile responses on FLIP testing. For instance, normal subjects and GERD usually have similar FLIP patterns with both exhibiting antegrade contractions and normal EGJ opening. On the other hand, type 1 and 2 achalasia patients both show no antegrade contractions and

reduced EGJ opening. The other patient groups exhibit varied contractile behavior ranging from weak antegrade contractions to an irregular contractile response along with normal or borderline EGJ opening. Similar behavior was observed in the dimension reduced VDL using PCA (pcaVDL) shown in Figure 10. However, the overlap among patient groups in pcaVDL was greater compared to that of ldaVDL. This is visualized more quantitatively in the distance matrix heatmaps of Figure 11.



(a) (b)

Figure 11: Distance matrix showing the median distance between points of each patient group specified by columns with the centroid of disease cluster specified by the rows. Each row has been normalized to represent distances as percentage and so each row adds up to 100.

An element in the i-th row and j-th column of the distance matrix is the median distance between the points of the j-th cluster and the centroid of the i-th cluster. Thus, for a good separation between clusters, the diagonal elements should have lesser values compared to the off-diagonal elements. As shown in Figure 11, the diagonal elements of distance matrix for ldaVDL were less than those of pcaVDL. Also, the off-diagonal elements of the distance matrix for ldaVDL were greater compared to those of pcaVDL. Hence, the LDA better segregated the patient groups consistent with it being based on a supervised approach of labeling points. However, it should be noted that the 30-dimensional VDL was generated in an unsupervised manner. This high-dimensional VDL captures the similar and dissimilar features of the input

corresponding to the patient groups. Thus, dimension reduction based on the variance of the data in an unsupervised manner as done in PCA provides insight into the structure of the 30-dimensional VDL. The clusters still observable after PCA demonstrated that the VDL successfully identified features distinguishing among these esophageal disorders. The distance matrix shown in Figure 11a also quantifies the similarities that some patient groups share as can be seen qualitatively in Figure 9. For instance, in row 1 (normal subjects), normal and GERD had lower values and in row 11 (GERD), the median values for normal and GERD subjects had lower values. Similarly, in rows 2-3 which correspond to achalasia Type I and Type II, the median distance values were very similar and much lower than the other patient groups. Similar, though less prominent, characteristics can be observed in Figure 11b. Hence, the VDL captured both similarities and dissimilarities among these esophageal disorders. The choice between ldaVDL and pcaVDL depends on the application; pcaVDL should be chosen if there is low confidence in diagnostic distinctions within the data.

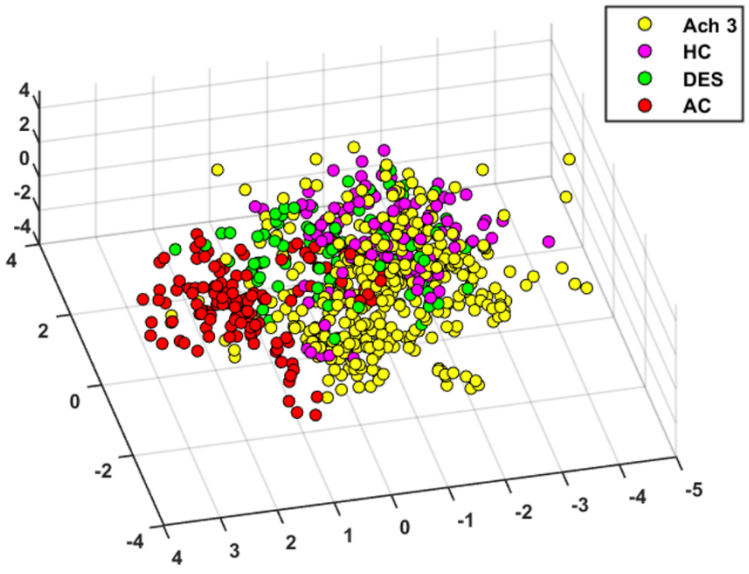
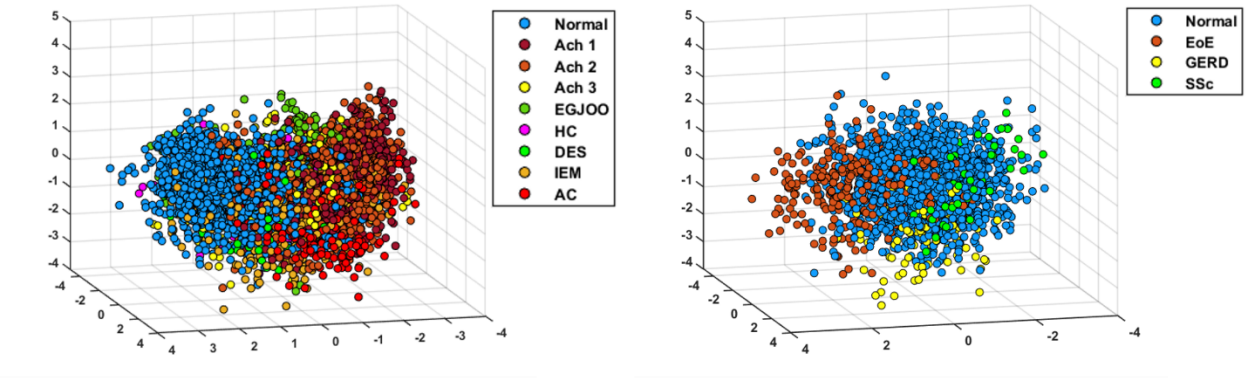


Figure 12: Specific motility groups represented in the ldaVDL. Type III achalasia, hypercontractile esophagus, distal esophageal spasm, and absent contractility are specially compared here. Type III achalasia, hypercontractile esophagus, and distal esophageal spasm have significant overlap while absent contractility cluster distinctly from the rest.

The availability of the data for the various disease groups determined the simple sizes. The main purpose for the MI-VAE is to generate the VDL and not predictive analysis. Thus, the diseases with smaller sample sizes appear as smaller clusters on the VDL and their location with respect to the other groups is determined by their contraction pattern. Additionally, the most distinct responses to FLIP distention occurs for normal subjects and achalasia groups. Both groups are of comparable sizes (normal: 237, achalasia (including type I and II): 224). Rest of the diseases lie somewhere between these two groups and often have very similar FLIP responses in terms of their contraction pattern as shown in Figure 10. Since no labels were used in training the MI-VAE, this proves the accurate representation of the diseases through the VDL by the MI-VAE. Augmenting the dataset also makes the MI-VAE more generalizable. Thus, a high training and validation accuracy (or low values of losses) ensures that all the diseases are accurately represented on the VDL (through the MI-VAE's latent space).

The differences between the esophageal motility disorders cannot be visualized clearly in the ldaVDL shown in Figure 10 because of the large number of groups included. Figure 12 isolates four motility disorders: absent contractility, type III achalasia, hypercontractile esophagus, and distal esophageal spasm (DES). Absent contractility exhibits no contractile activity and low EGJ tone whereas type III achalasia, hypercontractile esophagus, and DES show excessive or aberrant contractile activity with variable, often increase EGJ tone. These disorders are displayed in Figure 13 wherein absent contractility points cluster separately from the other three groups which extensively overlap each other. Of note, type III achalasia, which shows great variability in its HRM contraction pattern, exhibits the greatest scatter of points in Figure 12. Figure 13 extends this concept by subdividing all of the patient groups in Figures 9 and 10 into motility disorders (type I, II, III achalasia, EGJOO, hypercontractile esophagus, DES, IEM, and absent contractility) and not primarily defined by abnormal esophageal motility (EoE, GERD, and scleroderma).



(a) (b)

Figure 13: LDA performed separately for: (a) esophageal motility disorders vs normal esophageal motility and (b) other esophageal diseases vs normal esophageal motility. Esophageal motility disorders display similar segregation from the normal motility as in Figure 9 whereas the other diseases extensively overlap with the normal controls, which now display better segregation.

In Figure 13, each of these subgroups is compared to normal controls. Comparing Figure 13a (motility disorders) to Figure 13b (other disorders), the LDA segregates the motility disorders from the normal controls much more effectively than it does the other disorders which extensively intermix with the normal controls.

3.2. Antegrade contractions represented in the VDL

Most of the patient groups in Figures 9 and 10 correspond to diagnoses based on CCv4.0. Motility diagnoses were based on HRM rather than the FLIP study because: 1) a formal classification standard for FLIP has not been validated as with the CCv4.0 for HRM and 2) some diagnoses distinguished using HRM are not differentiated with FLIP e.g., achalasia type I and II. However, when diagnoses were used to visualize the VDL in ldaVDL, the similarities and dissimilarities between the disease groups could be quantified thereby providing an estimate of the capability and limitations of FLIP in distinguishing these disorders.

Even though the colors representing patient groups provide valuable insight into the distribution of each condition on the VDL, the diagnoses do not directly represent any specific characteristic of esophageal contractility or esophageal wall properties. For instance, a normal subject might

have a normal antegrade contraction only 70% of the time. In such a case, all points corresponding to this subject would be labelled the same irrespective of the abnormal contractions in 30% of them. Hence, the designation ‘normal subject’ is not a fundamental physiomarker, but a normal antegrade contraction is. Additionally, some subjects might respond differently on FLIP compared to HRM [9]. We, therefore, investigated the contraction characteristics of all the points that populate the VDL.

Figure 14 shows the point cloud of ldaVDL and pcaVDL with labels indicating contraction characteristics. A Random Forest classifier predicted the probability of a point in the VDL being a normal antegrade contraction or not. The probabilities of the points in the VDL are shown in reduced dimensions in Figures 14c and 14f for ldaVDL and pcaVDL, respectively. The VDL points for normal antegrade contractions and abnormal contractions cluster nicely without significant overlap, which is also evident from the high accuracy of the Random Forest classifier. This was because the distinction between normal antegrade contractions and abnormal contractions is a more fundamental criterion than the CCv4.0 label imposed through manometry. The contractile characteristics as identified by network 1 was an important hidden feature of θ variations, and therefore, the latent space variable captured this and led to the clustering as shown in Figure 14b,c and 14e,f. The effectiveness of clustering based on contractile characteristics is illustrated more quantitatively in the distance matrices in Figure 14h. Clearly there was a significant difference between the diagonal and off-diagonal values of the distance matrices, which signifies that the points in each grouping (normal antegrade contraction or abnormal contraction) remain close to each other and segregate from the other group. The bar plot in Figure 14(g) quantitatively estimates the percentage of each disease group that exhibit normal antegrade contractions.

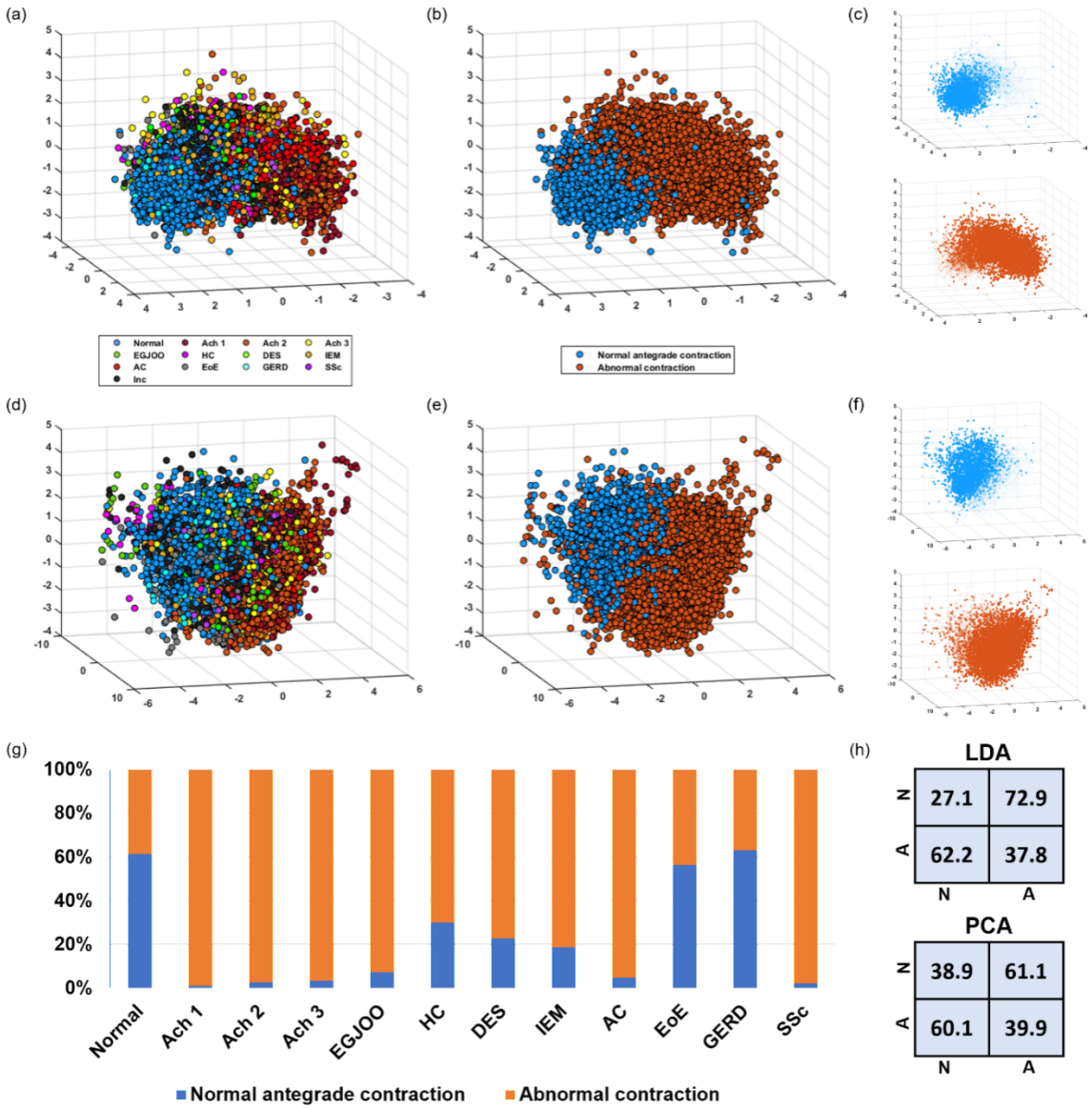


Figure 14: Antegrade contractions within disease groups. (a) LDA representation of the VDL for all disease groups (ldaVDL, i.e Figure 9); (b) data in panel a redisplayed as normal antegrade contractions (blue) or not a normal antegrade contraction (red) irrespective of disease group; (c) normal antegrade contractions (upper) vs abnormal contractions (lower) from panel b classified using Random Forest; (d) PCA representation of the VDL for all disease groups (pcaVDL, ie Figure 10); (e) data in panel d redisplayed as normal antegrade contractions (blue) or not a normal antegrade contraction (red) irrespective of disease group; (f) normal antegrade contractions (upper) vs abnormal contractions (lower) from panel e classified using Random Forest; (g) Bar plots showing the percentage of θ variations of each disease group being either normal antegrade contraction or abnormal contraction; (h) Distance matrices measuring the effectiveness of clustering for contractile behavior in ldaVDL and pcaVDL. The intensity of each point in panels c and f corresponds to the probability of that point being in each of the two groups as predicted by the Random Forest classifier.

3.3. Continuous behavior of the VDL

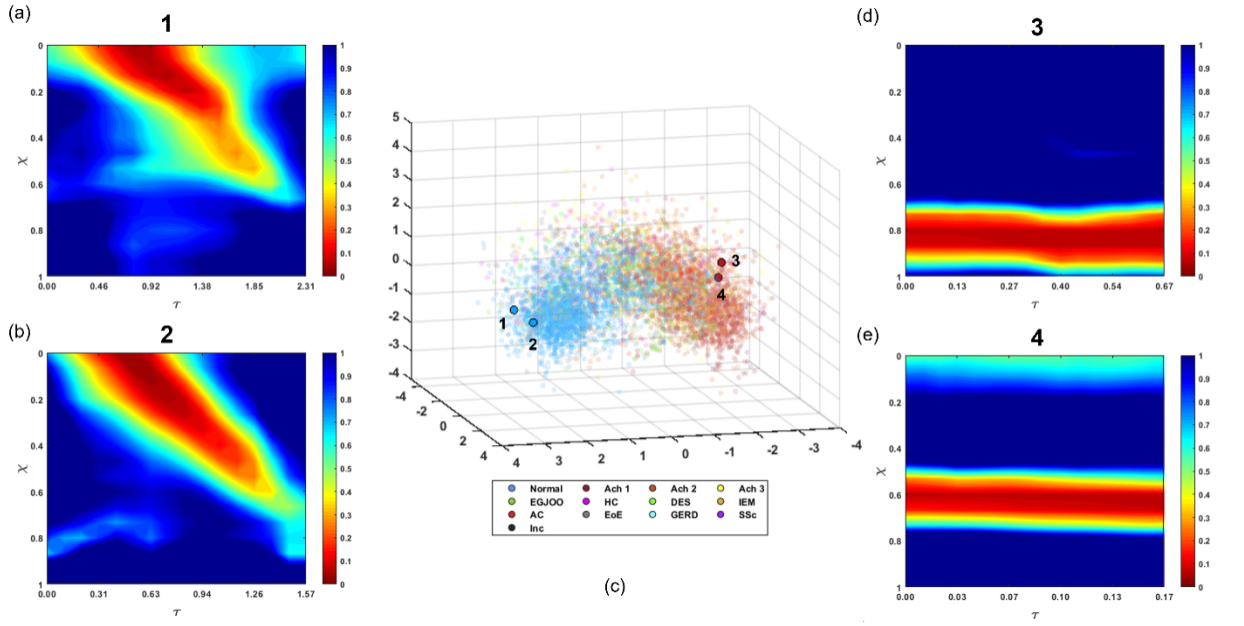


Figure 15: The continuous behavior of the VDL is shown through points selected from its two extremes: Normal subjects and achalasia Type I and II patients. (a) and (b) show a typical contraction pattern in normal subjects with strong antegrade contraction and relaxed EGJ. (d) and (e) show the characteristics of typical achalasia patients with no contraction at the esophageal body and a strong tone at the EGJ marked by the horizontal red zone. (c) VDL represents of the contraction patterns 1-4.

The ability of the VDL to identify the similarities and dissimilarities among patient groups is further illustrated in Figure 15. Two points in close proximity were selected from the extreme left end of the normal subjects and two from the extreme right end of the achalasia type II. The normal subjects both showed normal antegrade contractions (oblique red band) and a relaxed EGJ (blue region ahead of the contraction). Similarly, the two points of achalasia type II showed similar variation of the activation parameter; both exhibited no contraction and the EGJ remained closed as evident by the horizontal red band at the distal esophagus. The large separation between points 1 and 2 from points 3 and 4 indicates that they displayed completely different behavior. On the other hand, comparing points 3 and 4, we see that the location of the EGJ band was positioned differently in the two cases, they were still in close proximity on the VDL. Hence, we conclude that the VDL visualized in reduced dimensions using LDA is not dependent on consistent placement of the FLIP eliminating that as a potential source of error.

3.4 Generative property of the MI-VAE

An important feature of the MI-VAE is its generative capability. Due to the continuous nature of the VDL, new vectors from the VDL i.e., those which were not present in the training set can generate meaningful representations of the mechanics-based parameters. Figure 16 illustrates the transition from a point at the extreme of the normal cohort to another point at the extreme of achalasia type II.

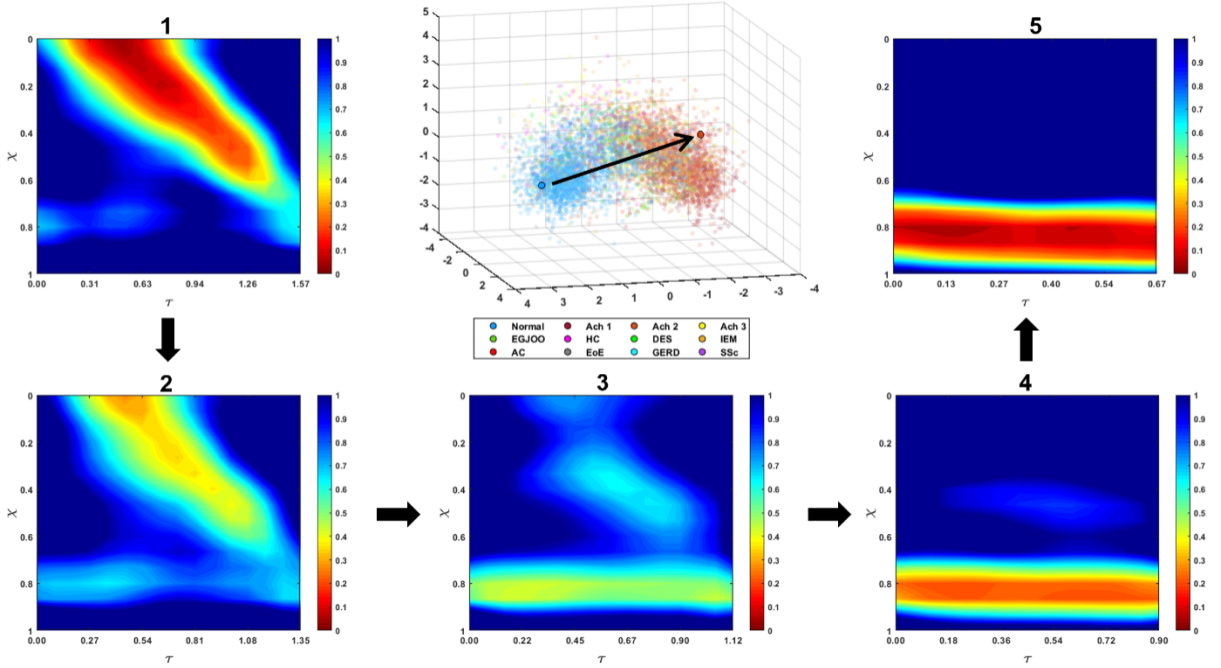


Figure 16: Example to describe the continuous nature of the VDL. Points 1 and 5 are chosen at the extremes of normal subjects and achalasia Type II patients, respectively. Points 2-4 are equi-spaced along the vector joining 1 and 5. The θ variations generated for 2-4 show the transition from normal to achalasia characteristics.

A 30-dimensional vector was calculated between these points and 3 equispaced intermediate points along this vector. Thus, the 2 endpoints were known but the 3 intermediate points were new and generated mechanics-based parameters not known before. As the transition occurs between the extremes, we see that the antegrade contraction progressively weakens and EGJ tone strengthens. Note that the vector (or line) shown in Figure 16 does not represent an actual case, but a hypothetical one to demonstrate the capability of the VDL. In an

actual scenario, the disease progression won't necessarily be linear as suggested in Figure 16. In fact, it won't necessarily progress at all. However, if the disease remains stable or progresses in an irregular manner, we can still use the VDL to extrapolate to the likely future 'mechanical' state in an average sense based on the pattern of past data.

Additionally, we can retrieve the other mechanics-based parameters apart from the θ variation ($K/A^{"}$, $P_{>?@}$, T, $\theta_{>?@}$) and the four EGJ work metrics ($EGJW$, $EGJROW_5$, $EGJROW_!$, and $EGJROW_A$) through network 2. This was the main reason for choosing mechanics-based parameters as the input of the MI-VAE rather than the raw distal pressure and cross-sectional area variations. The generated parameters reflect on the mechanical health of the organ unlike the raw data generated from FLIP. Additionally, with the application of mechanics, the two distributed measurements (cross-sectional area, $A(x,t)$, and distal pressure $P_8(t)$) were combined into one activation parameter $\theta(x,t)$, which not only simplified application of the MI-VAE, but also has physical meaning since it estimates esophageal muscle contraction.

With this framework, it is possible to track the progression of an esophageal disorder over time and, using this time-series data, it is possible to extrapolate in the VDL the likely future state of the esophagus. For instance, if the contraction pattern of a subject is found to move from 1 to 2 in Figure 16 in year 1, and from 2 to 3 in year 2, then it can be extrapolated that the state of the esophagus most likely progress to contraction pattern 4 and eventually 5.

3.5. Estimating the effectiveness of a treatment

Treatment strategy varies depending on the mechanical 'health' of the esophagus. For example, achalasia shows a θ variation with a strong tone (low values of θ) at the EGJ that does not relax, and the severity of the scenario is greater with progressively lower values of θ. Less 'potent' treatment with pneumatic dilation [58-60] could suffice for less severe cases while using laparoscopic Heller myotomy (LHM) myotomy [61, 62], or peroral endoscopic myotomy

(POEM) [63, 64] for severe cases. EoE usually show higher values of stiffness (K/A^n) as shown in Figure 7 and is treated with topical steroids. We hypothesize that our framework could be used to guide the need for treatment by quantifying the severity as well as aid in early discovery of the disorders. FLIP studies can be obtained before and after these procedures to evaluate the effectiveness of a treatment as well as for tracking the esophageal condition for years after treatment. In the next two subsections, we discuss two such scenarios where the MI-VAE framework can be applied to aid FLIP diagnosis.

3.5.1. Pre- and post-treatment state of the esophagus in achalasia patients

With a myotomy, the circular muscle fibers of the lower esophageal sphincter are cut to weaken the inherent tone at the EGJ making it easier for swallowed food and fluid to empty from the esophagus. POEM is an endoscopic myotomy wherein the circular muscle fibers of the lower esophageal sphincter ± the distal esophagus are cut. Using the MI-VAE framework, we present a quantitative approach for assessing the effectiveness of a POEM procedure. We tested this on three achalasia patients, one for each achalasia subtype. Figure 17 illustrates the contraction patterns before and after POEM along with their placement on the VDL. We selected the "typical" θ variation for each case by identifying the most effective contraction and the least EGJ tone observed at the same 50 mL FLIP bag volume. In Figure 17a, it is evident that after POEM, the EGJ tone had decreased significantly with improvement in contraction strength up to the proximal limit of the myotomy. On the VDL, this corresponded to the movement from achalasia zone toward the normal subjects. For the type II patient shown in Figure 17b, lower esophageal sphincter tone was clearly diminished by treatment but there was minimal recovery of the esophageal contraction and movement on the VDL is toward the absent contractility sector.

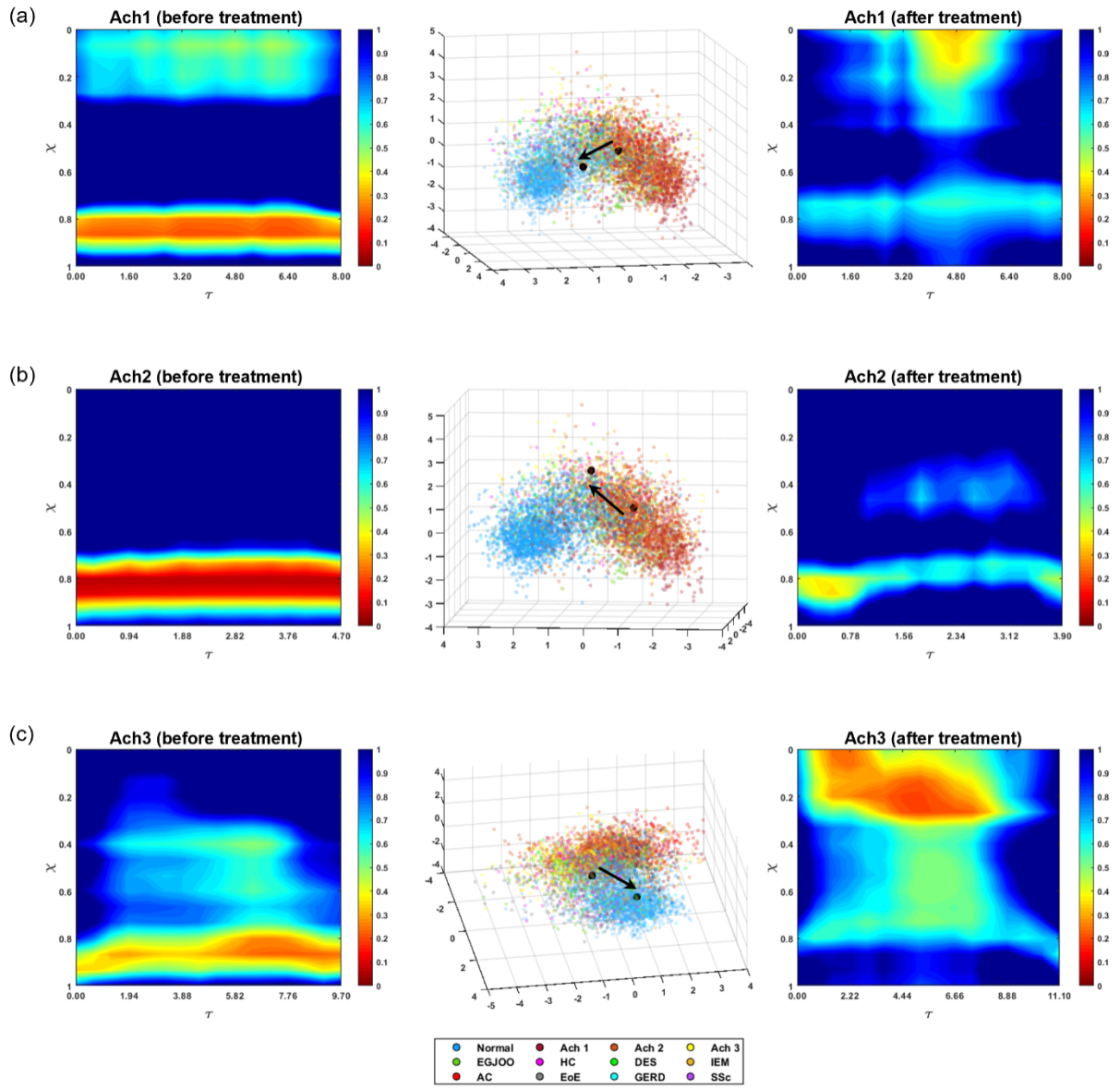


Figure 17: Estimating the effectiveness of a POEM procedure on achalasia patients using MI-VAE. (a)-(c) show the θ variation before treatment in achalasia types I, II, and III, respectively and the figures on the right show the θ variations after treatment. The figures in the center show the VDL representations of the contour plots with the arrows going from before to after treatment.

In Figure 17c, we see that for the Type III patient, the contraction strength improved significantly after POEM (albeit not antegrade) and EGJ was reduced such that the corresponding point in the VDL lay more in the normal cohort zone than that seen in Figures 17a and 17b. The improvement after treatment can be estimated quantitatively by the magnitude of the vector drawn from the initial to the final point in the VDL. The direction of the vector quantitatively

estimates the direction of improvement. This vector corresponds not only to the θ variation, but also to the discrete mechanics-based parameters adding physical meaning to the quantitative assessment to treatment effectiveness.

3.5.2. Post-treatment tracking

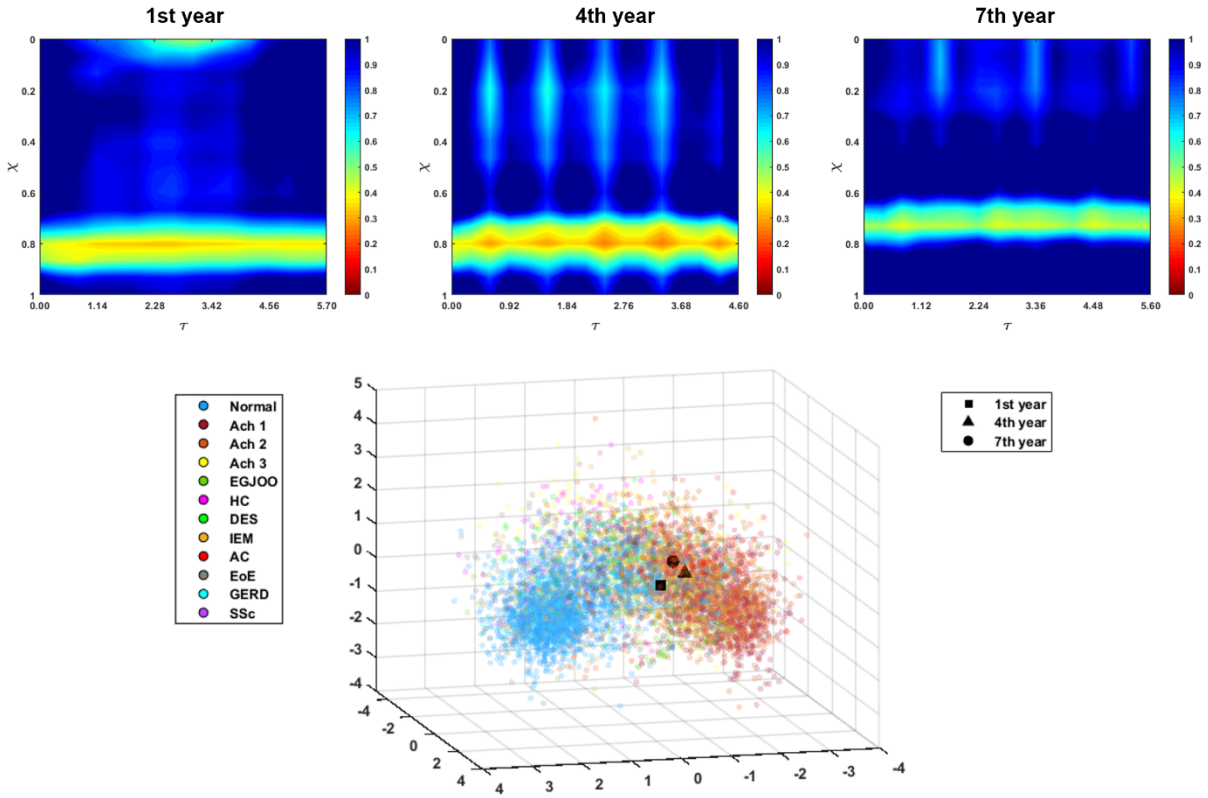


Figure 18: Tracking the state and motility of the esophagus after POEM for 7 years. Three θ variations are shown as observed in the 1st, 4th, and 7th year. These are plotted in their VDL representation. The three points lie close to each other indicting that the patient has stable esophageal motility charcateristics for the 7 years they have been tracked.

After treatment, it is often necessary to periodically re-evaluate patients over time. We tested our MI-VAE framework by tracking the condition of one patient who had undergone POEM and had FLIP data was available for years 1, 4, and 7 after POEM. The θ variations for each of these three years are shown in Figure 18 and are in close proximity to each other suggesting that the patient's condition was stable over the years. This example shows how the

MI-VAE framework can be used for post-POEM tracking of patient condition but can be applied very easily to other treatment procedures as well.

FLIP has been shown to uniquely identify various esophageal disorders in some cases that appeared normal on HRM and other diagnostic tests. Hirano et al. [65] reports the application of FLIP in the clinical management of esophageal disorders. Although, the application of FLIP has been shown in various clinical situations, it is still relatively new and lacks a standard classification as available for HRM, i.e., CCv4.0. With the help of VDL developed by the MI-VAE, we demonstrated a formal approach to use FLIP data that can be translated to clinical applications through quantitative assessment of esophageal disorders through FLIP. However, it must be noted that at its current state, AI by itself does not have an advantage over human judgement based on FLIP panometry. The framework we have presented is not meant to replace human judgement in interpreting FLIP data, but rather to provide a computational tool that can formalize the diagnostic process based on mechanics-informed machine learning. This framework is intended as an aid to medical professionals in their assessment of various esophageal disorders by adding quantitative elements to their diagnostic approach as well as speed up the process.

3.6. Limitations

The MI-VAE framework provides a technique to map esophageal disorders onto a parameter space called the VDL based on their mechanical characteristics estimating the mechanical ‘health’ of the esophagus thereby aiding in diagnosis and directing treatment. However, the VDL also has limitations. First, the raw data output from the FLIP study cannot be used directly by this framework. The MI-VAE requires the manual identification of the time instants between which readings are to be considered. This manual intervention might introduce differences in prediction based on which time instants are chosen. Second, the FLIP device has some technical limitations. It has an upper measurement limit for diameter which corresponds to

the maximum distension of the FLIP bag. Strong contractions might sometimes lead to the bag distending greater that this upper limit and cause incorrect readings. It also has a lower limit for its measurements owing to the catheter diameter onto which the bag is mounted. Again, strong contractions might cause full collapse of the FLIP bag on the catheter leading to incorrect readings. Additionally, it is sometimes observed that the bag volume calculated using the diameter readings might not be equal to the actual recorded bag volume. These factors might cause errors in the prediction of MI-VAE. Third, all esophageal disorders are not well represented in the dataset and there is a wide range of sample sizes. The characteristics of the disorders represented by a smaller dataset (like scleroderma) might not be learned properly by the MI-VAE. Therefore, the predictions of the Random Forest classifier and the relative placement of the points on the VDL through LDA might not be as accurate as the disorders represented by a larger dataset. Fourth, reduction of the VDL dimensions using LDA and PCA for visualization might lead to loss to important features that define the state and functioning of the esophagus. Fifth, as described earlier, the labels used for dimension reduction using LDA as well as training the random forest classifier is patient-specific, and not specific to the mechanics-based parameters. For instance, some θ variations of normal subjects might not exhibit peristaltic behavior. This might introduce some errors in the prediction of MI-VAE. Finally, the applicability of MI-VAE for predicting future disease progression is not possible because, as described in section 3.4, there is a lack of relevant chronological data which remains a limitation of this work.

4. CONCLUSION

In this work, we presented a framework called mechanics-informed variational autoencoder (MI-VAE) that quantitatively identified and distinguished among esophageal disorders based on their physical characteristics through a parameter space called the virtual

disease landscape (VDL). The physical characteristics were estimated through a set of physical parameters such as esophageal wall stiffness, contraction pattern, active relaxation of the esophageal wall muscles, and work metrics estimating EGJ behavior. These parameters were solved in a patient-specific manner from FLIP data using a one-dimensional mechanics-based inverse model. The VDL identified similarities and dissimilarities among the esophageal disorders as well as classified them based on their contractile characteristics. Additionally, Random Forest classifiers trained on the data represented in the VDL add a predictive capability to this framework to identify esophageal disorders and their contractile characteristics. We also described how the generative property of the MI-VAE gives it the capability to predict disease progression in time. Furthermore, we demonstrated through clinical applications that the MI-VAE can estimate the effectiveness and stability of a treatment over time. Finally, since the MI-VAE framework uses mechanics-based predictions of physiomarkers to develop a VDL, it can be extended to be used with other diagnostic technologies (and organs) as long as mechanics-based physiomarkers can be derived with them. For instance, a similar MI-VAE can be developed using high resolution impedance manometry (HRIM) data using the same governing equations with pressure and cross-sectional area measured from the HRIM catheter. Additionally, mechanics-based analysis has been shown to be applied to fluoroscopy [26] to predict physiomarkers which can also be used as described in this paper. There has also been active cardiovascular [66] and respiratory [67] research to develop physiomarkers for quantifying the course of diseases such as aortic aneurysms through wall shear stress, or the effectiveness of drug delivery in the lungs [68]. These mechanics-based physiomarkers could be used to develop a VDL for these organs following a similar approach as described in this paper.

REFERENCES

1. Fox, M., et al., *High-resolution manometry predicts the success of oesophageal bolus transport and identifies clinically important abnormalities not detected by conventional manometry.* Neurogastroenterology & Motility, 2004. **16**(5): p. 533-542.
2. Pandolfino, J.E., et al., *High-Resolution Manometry of the EGJ: An Analysis of Crural Diaphragm Function in GERD.* Official journal of the American College of Gastroenterology | ACG, 2007. **102**(5).
3. Fox, M.R. and A.J. Bredenoord, *Oesophageal high-resolution manometry: moving from research into clinical practice.* Gut, 2008. **57**(3): p. 405.
4. Pandolfino, J.E., et al., *Achalasia: A New Clinically Relevant Classification by High-Resolution Manometry.* Gastroenterology, 2008. **135**(5): p. 1526-1533.
5. Pandolfino, J.E., et al., *High-resolution manometry in clinical practice: utilizing pressure topography to classify oesophageal motility abnormalities.* Neurogastroenterology & Motility, 2009. **21**(8): p. 796-806.
6. Yadlapati, R., et al., *Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0©.* Neurogastroenterology & Motility, 2021. **33**(1): p. e14058.
7. Gyawali, C.P., et al., *Evaluation of esophageal motor function in clinical practice.* Neurogastroenterology & Motility, 2013. **25**(2): p. 99-133.
8. Carlson, D.A., et al., *Evaluation of Esophageal Motility Utilizing the Functional Lumen Imaging Probe.* The American journal of gastroenterology, 2016. **111**(12): p. 1726-1735.
9. Carlson, D.A., et al., *The Functional Lumen Imaging Probe Detects Esophageal Contractility Not Observed With Manometry in Patients With Achalasia.* Gastroenterology, 2015. **149**(7): p. 1742-1751.
10. Carlson, D.A., et al., *Evaluating esophageal motility beyond primary peristalsis: Assessing esophagogastric junction opening mechanics and secondary peristalsis in patients with normal manometry.* Neurogastroenterology & Motility, 2021. **33**(10): p. e14116.
11. Orvar, K.B., Gregersen, H., & Christensen, J., *Biomechanical characteristics of the human esophagus.* Digestive diseases and sciences, 1993. **38**(2): p. 197-205.
12. Kwiatek, M.A., et al., *Mechanical Properties of the Esophagus in Eosinophilic Esophagitis.* Gastroenterology, 2011. **140**(1): p. 82-90.
13. Savarino, E., et al., *Use of the Functional Lumen Imaging Probe in Clinical Esophagology.* Official journal of the American College of Gastroenterology | ACG, 2020. **115**(11).
14. Carlson, D.A., *Functional lumen imaging probe: The FLIP side of esophageal disease.* Current Opinion in Gastroenterology, 2016. **32**(4).
15. Teitelbaum, E.N., et al., *Comparison of esophagogastric junction distensibility changes during POEM and Heller myotomy using intraoperative FLIP.* Surgical Endoscopy, 2013. **27**(12): p. 4547-4555.
16. Campagna, R.A.J., et al., *Intraoperative assessment of esophageal motility using FLIP during myotomy for achalasia.* Surgical Endoscopy, 2020. **34**(6): p. 2593-2600.
17. Fan, Y., H. Gregersen, and G.S. Kassab, *A two-layered mechanical model of the rat esophagus. Experiment and theory.* BioMedical Engineering OnLine, 2004. **3**(1): p. 40.
18. Natali, A.N., E.L. Carniel, and H. Gregersen, *Biomechanical behaviour of oesophageal tissues: Material and structural configuration, experimental data and constitutive analysis.* Medical Engineering & Physics, 2009. **31**(9): p. 1056-1062.

19. Sokolis, D.P., *Structurally-motivated characterization of the passive pseudo-elastic response of esophagus and its layers.* Computers in Biology and Medicine, 2013. **43**(9): p. 1273-1285.
20. Stavropoulou, E.A., Y.F. Dafalias, and D.P. Sokolis, *Biomechanical and histological characteristics of passive esophagus: Experimental investigation and comparative constitutive modeling.* Journal of Biomechanics, 2009. **42**(16): p. 2654-2663.
21. Yang, W., et al., *Directional, Regional, and Layer Variations of Mechanical Properties of Esophageal Tissue and its Interpretation Using a Structure-Based Constitutive Model.* Journal of Biomechanical Engineering, 2005. **128**(3): p. 409-418.
22. Yang, W., et al., *3D Mechanical Properties of the Layered Esophagus: Experiment and Constitutive Model.* Journal of Biomechanical Engineering, 2006. **128**(6): p. 899-908.
23. Acharya, S., et al., *Pumping Patterns and Work Done During Peristalsis in Finite-Length Elastic Tubes.* Journal of Biomechanical Engineering, 2021. **143**(7).
24. Brasseur, J.G., *A fluid mechanical perspective on esophageal bolus transport.* Dysphagia, 1987. **2**(1): p. 32.
25. Ghosh, S.K., et al., *The mechanical basis of impaired esophageal emptying postfundoplication.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2005. **289**(1): p. G21-G35.
26. Halder, S., et al., *Mechanics informed fluoroscopy of esophageal transport.* Biomechanics and Modeling in Mechanobiology, 2021. **20**(3): p. 925-940.
27. Kou, W., et al., *A fully resolved active musculo-mechanical model for esophageal transport.* Journal of Computational Physics, 2015. **298**: p. 446-465.
28. Kou, W., et al., *A continuum mechanics-based musculo-mechanical model for esophageal transport.* Journal of Computational Physics, 2017. **348**: p. 433-459.
29. Li, M. and J.G. Brasseur, *Non-steady peristaltic transport in finite-length tubes.* Journal of Fluid Mechanics, 1993. **248**: p. 129-151.
30. Li, M., J.G. Brasseur, and W.J. Dodds, *Analyses of normal and abnormal esophageal transport using computer simulations.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 1994. **266**(4): p. G525-G543.
31. Yang, W., et al., *Finite element simulation of food transport through the esophageal body.* World journal of gastroenterology, 2007. **13**(9): p. 1352-1359.
32. Ching, T., et al., *Opportunities and obstacles for deep learning in biology and medicine.* Journal of The Royal Society Interface, 2018. **15**(141): p. 20170387.
33. Esteva, A., et al., *A guide to deep learning in healthcare.* Nature Medicine, 2019. **25**(1): p. 24-29.
34. Faust, O., et al., *Deep learning for healthcare applications based on physiological signals: A review.* Computer Methods and Programs in Biomedicine, 2018. **161**: p. 1-13.
35. Hamet, P. and J. Tremblay, *Artificial intelligence in medicine.* Metabolism, 2017. **69**: p. S36-S40.
36. Jiang, F., et al., *Artificial intelligence in healthcare: past, present and future.* Stroke and Vascular Neurology, 2017. **2**(4): p. 230.
37. Miotto, R., et al., *Deep learning for healthcare: review, opportunities and challenges.* Briefings in Bioinformatics, 2018. **19**(6): p. 1236-1246.
38. Piccialli, F., et al., *A survey on deep learning in medicine: Why, how and when?* Information Fusion, 2021. **66**: p. 111-137.
39. Rajkomar, A., J. Dean, and I. Kohane, *Machine Learning in Medicine.* New England Journal of Medicine, 2019. **380**(14): p. 1347-1358.
40. Topol, E.J., *High-performance medicine: the convergence of human and artificial intelligence.* Nature Medicine, 2019. **25**(1): p. 44-56.
41. Yu, K.-H., A.L. Beam, and I.S. Kohane, *Artificial intelligence in healthcare.* Nature Biomedical Engineering, 2018. **2**(10): p. 719-731.

42. Le Berre, C., et al., *Application of Artificial Intelligence to Gastroenterology and Hepatology.* Gastroenterology, 2020. **158**(1): p. 76-94.e2.
43. Min, J.K., M.S. Kwak, and J.M. Cha, *Overview of Deep Learning in Gastrointestinal Endoscopy.* Gut and liver, 2019. **13**(4): p. 388-393.
44. Ruffle, J.K., A.D. Farmer, and Q. Aziz, *Artificial Intelligence-Assisted Gastroenterology—Promises and Pitfalls.* Official journal of the American College of Gastroenterology | ACG, 2019. **114**(3).
45. Yang, Y.J. and C.S. Bang, *Application of artificial intelligence in gastroenterology.* World journal of gastroenterology, 2019. **25**(14): p. 1666-1683.
46. Kou, W., et al., *A deep-learning-based unsupervised model on esophageal manometry using variational autoencoder.* Artificial Intelligence in Medicine, 2021. **112**: p. 102006.
47. Kingma, D.P. and M. Welling, *Auto-Encoding Variational Bayes.* arXiv e-prints, 2013: p. arXiv:1312.6114.
48. Acharya, S., et al., *Estimation of mechanical work done to open the esophagogastric junction using functional lumen imaging probe panometry.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2021. **320**(5): p. G780-G790.
49. Barnard, A.C.L., et al., *A Theory of Fluid Flow in Compliant Tubes.* Biophysical Journal, 1966. **6**(6): p. 717-724.
50. Kamm, R.D. and A.H. Shapiro, *Unsteady flow in a collapsible tube subjected to external pressure or body forces.* Journal of Fluid Mechanics, 1979. **95**(1): p. 1-78.
51. Manopoulos, C.G., D.S. Mathioulakis, and S.G. Tsangaris, *One-dimensional model of valveless pumping in a closed loop and a numerical solution.* Physics of Fluids, 2006. **18**(1): p. 017106.
52. Ottesen, J.T., *Valveless pumping in a fluid-filled closed elastic tube-system: one-dimensional theory with experimental validation.* Journal of Mathematical Biology, 2003. **46**(4): p. 309-332.
53. Buslaev, A., et al., *Albumentations: Fast and Flexible Image Augmentations.* Information, 2020. **11**(2).
54. Kingma, D.P. and J. Ba, *Adam: A Method for Stochastic Optimization.* arXiv e-prints, 2014: p. arXiv:1412.6980.
55. Chollet, F.a.o. *Keras*. 2015; Available from: https://github.com/fchollet/keras.
56. Martín Abadi, A.A., Paul Barham, Eugene Brevdo,, et al. *TensorFlow: Large-scale machine learning on heterogeneous systems*. 2015; Available from: https://www.tensorflow.org/.
57. Pedregosa, F.a.V., G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.
58. Al-Jafar, H., et al., *Severe Recurrent Achalasia Cardia Responding to Treatment of Severe Autoimmune Acquired Haemophilia.* Case Reports in Gastroenterology, 2012. **6**(3): p. 618-623.
59. Savarino, E., et al., *Achalasia With Dense Eosinophilic Infiltrate Responds to Steroid Therapy.* Clinical Gastroenterology and Hepatology, 2011. **9**(12): p. 1104-1106.
60. Spechler, S.J., V. Konda, and R. Souza, *Can Eosinophilic Esophagitis Cause Achalasia and Other Esophageal Motility Disorders?* Official journal of the American College of Gastroenterology | ACG, 2018. **113**(11).
61. Hunter, J.G., et al., *Laparoscopic Heller myotomy and fundoplication for achalasia.* Annals of surgery, 1997. **225**(6): p. 655-665.

62. Richards, W.O., et al., *Heller myotomy versus Heller myotomy with Dor fundoplication for achalasia: a prospective randomized double-blind clinical trial.* Annals of surgery, 2004. **240**(3): p. 405-415.
63. Inoue, H., Minami, H., Kobayashi, Y., Sato, Y., Kaga, M., Suzuki, M., Satodate, H., Odaka, N., Itoh, H., & Kudo, S., *Peroral endoscopic myotomy (POEM) for esophageal achalasia.* Endoscopy, 2010. **42**(4): p. 265-271.
64. Von Renteln, D., et al., *Peroral Endoscopic Myotomy for the Treatment of Achalasia: An International Prospective Multicenter Study.* Gastroenterology, 2013. **145**(2): p. 309-311.e3.
65. Hirano, I., J.E. Pandolfino, and G.E. Boeckxstaens, *Functional Lumen Imaging Probe for the Management of Esophageal Disorders: Expert Review From the Clinical Practice Updates Committee of the AGA Institute.* Clinical Gastroenterology and Hepatology, 2017. **15**(3): p. 325-334.
66. Taylor, C.A. and M.T. Draney, *EXPERIMENTAL AND COMPUTATIONAL METHODS IN CARDIOVASCULAR FLUID MECHANICS.* Annual Review of Fluid Mechanics, 2004. **36**(1): p. 197-231.
67. Grotberg, J.B., *Respiratory Fluid Mechanics and Transport Processes.* Annual Review of Biomedical Engineering, 2001. **3**(1): p. 421-457.
68. Chakravarty, A., et al., *Pulmonary drug delivery and retention: A computational study to identify plausible parameters based on a coupled airway-mucus flow model.* PLOS Computational Biology, 2022. **18**(6): p. e1010143.

APPENDIX

1. Derivation of loss function

The VAE representing network 1 learns to model an input dataset as a distribution, $p(\theta)$ such that the input variable θ is generated from a likelihood distribution $p(\theta|z)$, where z is the latent variable. This distribution $p(\theta)$ is parameterized by the weights of the neural network. The decoder yields the likelihood distribution $p(\theta,z)$, i.e., it takes z as input and outputs θ. The encoder should ideally yield the posterior distribution $p(z|\theta)$. Unfortunately, $p(z|\theta)$ is computationally intractable in general. So, in practice, the encoder yields a conditional distribution $q(z|\theta)$ which approximates $p(z|\theta)$. The Kullback-Leibler divergence (KLD) provides a measure of the difference between the two distributions $q(z|\theta)$ and $p(z|\theta)$, and leads to the following relation:

$$D_{+\$}[q(z|\theta)||p(z|\theta)] = -\,]\,q(z|\theta)[\log p(\theta|z) + \log p(z) - \log q(z|\theta)]dz + \log p(\theta)\,. \qquad (A1)$$

Since KLD is always positive, the right-hand side of the above expression can be written as follows:

$$p(\theta) \geq]\,q(z|\theta)[\log p(\theta|z) + \log p(z) - \log q(z|\theta)]dz. \qquad (A2)$$

The above equation can be re-written in terms of a new KLD form as follows:

$$p(\theta) \geq -D_{+\$}[q(z|\theta)||p(z)] + \mathbb{E}_{\mathrm{L}\sim\mathrm{ENZO}\theta\mathrm{P}}[\log p(\theta|z)], \qquad (A3)$$

where $\mathbb{E}_{\mathrm{L}\sim\mathrm{ENZO}\theta\mathrm{P}}[\log p(\theta|z)]$ is the joint log-likelihood of the input θ and the latent variable z. The right-hand side is called the Evidence Lower bound (ELBO) and is named so since it estimates the lower bound of the likelihood of the data. Thus, maximizing ELBO maximizes the likelihood of the data. The KLD term works like a regularizer and forces the approximate posterior $q(z|\theta)$ to be as close to the prior $p(z)$ as possible. This term causes the posterior

$q(z|\theta)$ to enforce a high probability to z values that can generate the point θ without collapsing to a single point like an autoencoder. This gives a continuous behavior to the latent space so that meaningful generations are possible from points in the latent space which are not related to any input in the training dataset. The second term of equation A3 is the reconstruction error between in the input and the generated output of the entire network. It is possible to derive a closed form solution for the KLD term if we choose the approximate posterior $q(z|\theta)$ to have a Gaussian distribution and choose the prior $p(z)$ to have a standard normal distribution as shown below:

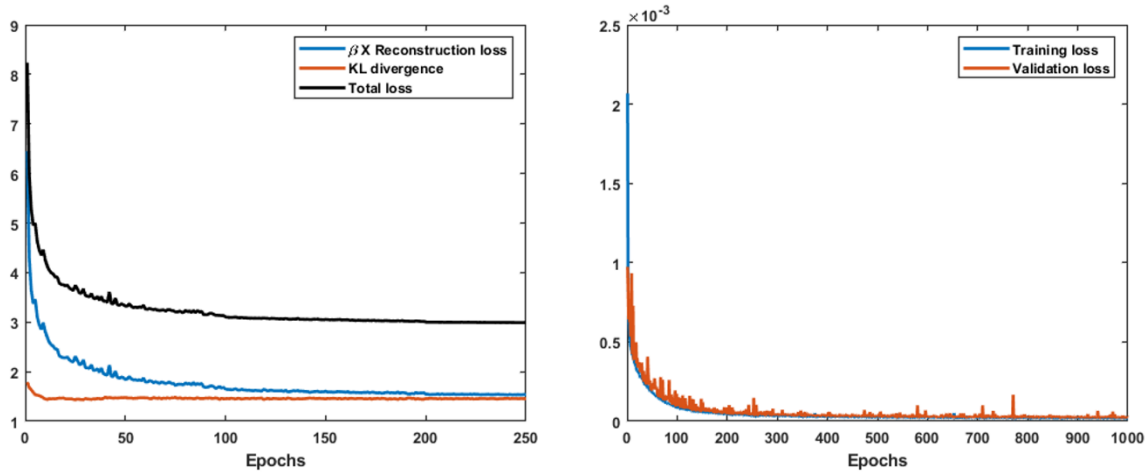
$$-D_{+\$}[q(z|\theta)||p(z)] = \frac{1}{2}\mathrm{h}1 + \log\sigma_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}} - \sigma_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}} - \mu_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}}\mathrm{m}. \tag{A4}$$

The encoder, as shown in Figure 6, outputs $\mu_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}}$ and $\log\sigma_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}}$. Although we have an analytical form for the KLD term of equation A4, the reconstruction error requires to be estimated by sampling. Sampling z from $q(z|\theta)$ directly leads to a problem in implementing backpropagation since the network would have a random node at the input of decoder. This problem can be tackled by a reparameterization trick where z is sampled from the mean and log variance parameters of $q(z|\theta)$ as estimated by the encoder shown as follows:

$$z = \mu_{\mathrm{E}} + \varepsilon \cdot \exp 8\frac{1}{2}\log\sigma_{\dot{\mathrm{E}}}^{\dot{\mathrm{l}}}\mathrm{:}\ , \tag{A5}$$

wherein ε is a random number generated from a standard normal distribution. This step makes it possible to backpropagate in a deterministic manner by considering ε as an extra input. Since ε is sampled from a different distribution which is not a function of any variables with respect to which derivatives might be required, stochasticity is introduced in the network without affecting backpropagation.

2. Training details of MI-VAE



(a) (b)

Figure I: Training details of MI_VAE. a) Learning curve for network 1, b) Learning curve for network 2.

The learning curves of network 1 and 2 are shown in Figure I (a) and (b), respectively. Because of the scaling parameter β as described in equation 14, the KL divergence and the reconstruction loss have similar magnitude. If there was a significant difference between their magnitudes, the total loss would be represented mainly by the larger of the two and proper minimization of both the loss would not happen. The final magnitudes of reconstruction loss, KL divergence, and the total loss after 250 epochs were $1.36 \times 10^{6\mathrm{A}}$, 1.54, and 2.9, respectively. The training and validation losses for network 2 show steps due to the change in learning rates at those points. These learning rates were 10^{-3}, 10^{-4} and 10^{-5}. The final magnitudes of the losses converge to $9 \times 10^{6\mathrm{J}}$.

3. Training details of the Random Forest classifiers

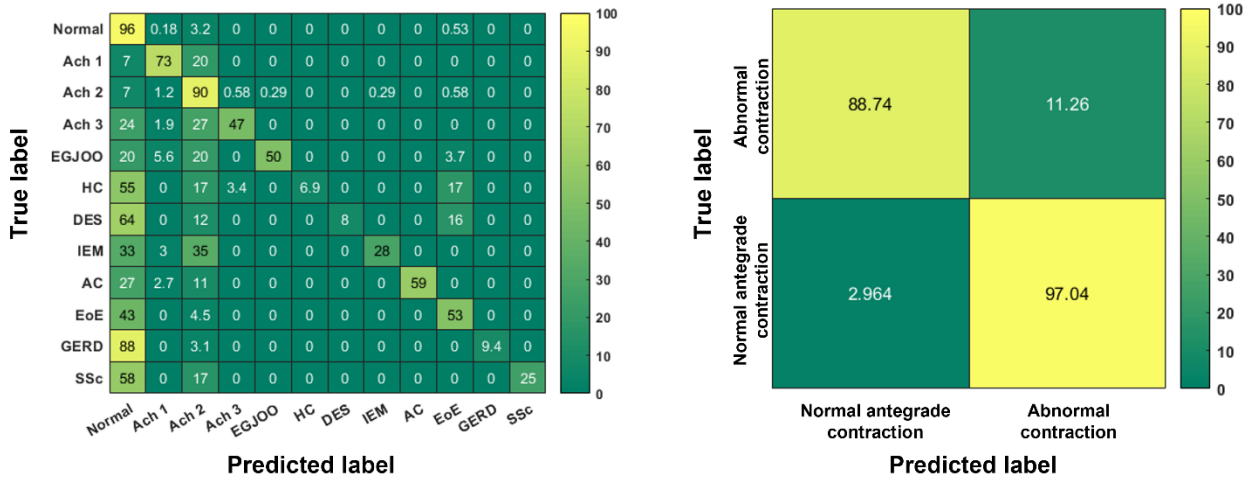


Figure II: Confusion matrices for Random Forest classifiers. a) Confusion matrix for the 12 groups, b) Confusion matrix for contractile behavior.

The confusion matrices corresponding to the two random forest classifiers are shown in Figure II. The numbers in the matrices are represented in terms of percentages. The classifier for the different disease groups (as shown in Figure II(a)) shows reasonable accuracy for most groups. As already discussed, some groups exhibit very similar behavior on FLIP although they were classified as different groups in through HRM. Additionally, a subject might have a variety of contraction patterns. This reduces the overall accuracy of the Random Forest classifier to some extent. The final Subset accuracy observed on the test set was 0.74. The Random Forest classifier to identify contractile behavior performs better as shown by the confusion matrix in Figure II(b). This is because the labels for peristaltic behavior were specified for each data point (corresponding to each θ variation) and labeled just by subject diagnosis. Also, there is less overlap in the VDL between the two groups, i.e., normal antegrade contraction and abnormal contraction. The Jaccard accuracy observed on the test set was 0.94.