UNIVERSITY OF CALIFORNIA SAN DIEGO

Probabilistic Methods to Quantify Spatiotemporal Patterns and Categorize Dysfunction in Cutaneous Multi-Electrode Gastric Neuromuscular Recordings

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Anjulie Agrusa

Committee in charge:

Professor Todd P. Coleman, Chair
Professor Gert Cauwenberghs
Professor Kevin King
Professor Ramesh Rao
Professor Shankar Subramaniam

2021

The dissertation of Anjulie Agrusa is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2021

DEDICATION

To my grandfather, Dadu.

EPIGRAPH

If I have seen further than others,

it is by standing upon the shoulders of giants.

—Sir Isaac Newton

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Todd Coleman. Beyond expanding my technical capabilities, you have taught me how to *really* think: to formulate and solve complex problems. Throughout my PhD, you have always pushed me to accomplish more than I thought I could. I've deeply enjoyed our frequent "nerding out" conversations, and I will forever appreciate your empowering mentorship.

I would like to thank my committee members, Professors Gert Cauwenberghs, Kevin King, Shankar Subramaniam, and Ramesh Rao for our insightful conversations and your guidance throughout this process. To Professors Subramaniam and Cauwenberghs, thank you for your mentorship during my undergraduate years, as it was pivotal in my decision to pursue academic research. Thank you to Professor Bruce Wheeler for showing me the difference between a good and great mentor. I have learned so much from you. I would also like to thank my clinical collaborators, Dr. David Kunkel and Dr. Thomas Abell for your guidance and sharing your contagious passion for this work.

Thank you to my incredible mentors in the lab, Armen, Dae, Justin, Marianne and Phil, for showing me the ropes of research and setting such a great example to follow. And thank you to my lab-mates Gabe and Alexis for being wonderful colleagues. A special thank you to my lab-mate and close friend Gladys for making our office feel like home. I will always cherish our grad school memories, and I will make sure my next office is stocked with chocolate.

I am also grateful to my high school teachers, Dr. Johnson and Mrs. Villa, for showing me how to think critically and pursue knowledge. Thank you to Ken Charhut, Mark Pollard, and Nikolai Liachenko for introducing me to the world of data science. Thank you to my parents for teaching me the importance of education and to the rest of my family for keeping life interesting. Thank you to my wonderful friends: Rachel, Kayla, Julia, Christy, Ismael, Alice, Nick, Sierra, and Madison that have truly shown me unwavering love and support. You have gone above and beyond the call of duty of friendship in so many ways. Finally, to my partner Sydnee. I quite

simply couldn't have done this without you.

Chapter 2, in full, is a reprint of the material as it appears in IEEE Transactions on Biomedical Engineering 2019. Agrusa, Anjulie; Gharibans, Armen; Allegra, Alexis; Kunkel, David; Coleman, Todd, IEEE, 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in the proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2020. Agrusa, Anjulie; Allegra, Alexis; Kunkel, David; Coleman, Todd, IEEE, 2020. The dissertation author was the primary investigator and author of these materials.

Chapter 4, in full, is currently being prepared for submission for publication of the material. Agrusa, Anjulie; Coleman, Todd. The dissertation author was the primary investigator and author of these materials.

Chapter 5, in full, is currently under review for publication of the material. Agrusa, Anjulie; Kunkel, David; Coleman, Todd. The dissertation author was the primary investigator and author of this material.

VITA

2017	B. S. in Bioengineering: Biosystems, University of California, San Diego
2018	Teaching Assistant, University of California, San Diego
2021	Ph. D. in Bioengineering, University of California, San Diego

PUBLICATIONS

A.S. Agrusa and T. P. Coleman, "Unsupervised Anomaly Detection with Wasserstein Barycenters," In Preparation.

A.S. Agrusa, D. C. Kunkel, and T. P. Coleman, "Robust Regression and Optimal Transport Methods to Predict Gastrointestinal Disease Etiology from High Resolution EGG and Symptom Severity," Under Review.

A.S. Agrusa, A. Allegra, D. C. Kunkel, and T. P. Coleman, "Robust Methods to Detect Abnormal Initiation in the Gastric Slow Wave from Cutaneous Recordings," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2020.

J. F. Kurniawan, A.S. Agrusa ,V. Wu, I.R. Lerman, D.C. Kunkel, and T.P. Coleman, "Su1356 STRETCHABLE ADHESIVE-INTEGRATED ELECTRONICS FOR MULTIELECTRODE CUTANEOUS GASTRIC SLOW WAVE MONITORING," Gastroenterology, Volume 168, Issue 6, 2020.

A.S. Agrusa, A. A. Gharibans, A. Allegra, D. C. Kunkel, and T. P. Coleman, "A Deep Convolutional Neural Network Approach to Classify Normal and Abnormal Gastric Slow Wave Initiation from the High Resolution Electrogastrogram," IEEE Transactions on Biomedical Engineering, Volume 67, Issue 3, June 2019.

ABSTRACT OF THE DISSERTATION

Probabilistic Methods to Quantify Spatiotemporal Patterns and Categorize Dysfunction in Cutaneous Multi-Electrode Gastric Neuromuscular Recordings

by

Anjulie Agrusa

Doctor of Philosophy in Bioengineering

University of California San Diego, 2021

Professor Todd P. Coleman, Chair

One of the primary components of GI motility is the gastric slow wave, which travels along the surface of the stomach and locally actuates smooth muscle contractions, known as peristalsis. Pathological slow wave patterns, detected through intrasurgical recordings, have been shown in high incidence with GI disease and symptoms. The ability to non-invasively detect pathological slow wave patterns is of high clinical interest, as this could significantly enhance diagnostics and lead to tailored treatment regimens. In four parts, we develop statistical analysis methods to infer spatial slow wave patterns from cutaneous high-resolution electrogastrogram (HR-EGG) waveforms.

We first establish a deep three-dimensional convolutional neural network (3D-CNN) framework to classify normal and abnormally initiated slow waves from *in silico* multi-electrode waveforms. We develop a linear discriminant classifier using spatial features of wave propagation to compare performance. We report the first successful classification study and show robustness across recording and anatomical variability.

Second, we develop a transfer learning framework, *in silico*, to build a 3D-CNN classifier robust to heterogeneity in both the location of the slow wave abnormal initiation as well as the recording start times with respect to slow wave cycles. We find that by building complexity in training sets, transfer learning one model to the next, the final network exhibits high classification accuracy on data with closer resemblance to human physiologic recordings, suggesting downstream potential for use as widely deployable clinical screening tools for GI disorders.

Third, we establish an optimal transport theory-based philosophical framework for the detection of anomalous probability distributions. We demonstrate a rigorous mathematical proof and show an example algorithm within the framework of our methodology to detect an anomaly in a set of probability distributions that model slow wave direction estimates from HR-EGG recordings on human subjects.

Finally, we implement robust regression and clustering methods on optimal transport-based features of probability density estimates of slow wave directions from HR-EGG recordings on human subjects. We report highly accurate robust regression and suggest the ability to predict treatment response and underlying disease etiology. This methodology may provide clinicians with an opportunity to screen patients and optimize treatment regimens.

Chapter 1

Introduction

1.1 Clinical Background

Gastrointestinal (GI) problems are the second most common reason for someone to miss school or work, after the common cold [29]. Two of the most prevalent GI functional and motility disorders are functional dyspepsia (FD) and gastroparesis (GP), which affect 10% and 1.5-3% of the U.S. population, respectively [42, 77]. Multiple disease etiologies underlie the umbrella diagnoses of GP and FD, including but not limited to i) antral hypomotility, ii) pylorospasms, iii) gastric myoelectric dysfunction [56], and iv) autonomic dysfunction. Tailored treatments for such causes include i) prokinetic medications, ii) botulinum toxin injection [85], gastric neuromodulation [2, 20, 91, 98, 89, 13, 122], and iv) vagus nerve stimulation [58, 59], respectively. The common occurrence of treatment regimens tailored to the wrong etiology or to symptom mitigation results in repetitive visits to the clinic and adverse effects on quality of life with vast economic, social, and healthcare consequences. There is a desire and unmet need to sub-type GP and FD in terms of underlying etiologies to improve prediction of treatment response and enable targeted interventions [86, 73].

Current diagnostic tools are typically not etiology-specific. For instance, the PAGI-SYM

questionnaire on foregut symptoms, which includes the gastroparesis cardinal symptom index, is used to manage treatments but shows a high degree of symptom overlap with underlying disease etiologies [56]. Alternatively, the gastric emptying test, a clinical gold standard, typically involves ingestion of a meal containing a radioactive tracer and imaging at fixed time intervals to assess the duration of time required for the tracer to empty the stomach. However, gastric emptying shows inconsistent correlation with symptoms or symptom improvement [123, 74] and multiple medications have been shown to improve symptoms but not gastric emptying and vice versa [39, 97, 21].

1.2 Gastric Myoelectric Physiology

Gastric neuromuscular activity involves the interplay between electrical activity on the stomach's surface and smooth muscle contractions that propel food from the esophagus to the intestines. Gastric pacemaker cells, or interstitial cells of Cajal (ICCs), receive modulatory inputs from enteric neurons and undergo spontaneous cyclical depolarization, similar to cardiac pacemaker cells in the heart. These oscillatory membrane potentials excite neighboring smooth muscle cells, causing contractions at regular intervals, and excite adjacent downstream ICCs in their network [46, 161, 120]. This results in a traveling electrical wave, called the gastric slow wave [158, 80, 153].

Slow waves, in healthy individuals, originate in the pacemaker region of the stomach and propagate in an anterograde direction toward the duodenum. This propagation occurs in equipotential rings along the outer surface of the stomach, at a frequency of 3 cycles per minute (0.05 Hz) [80, 28, 66]. Abnormalities in slow wave initiation and propagation patterns arise from loss or deterioration of ICCs [113, 19]. These pathological slow waves can spontaneously arise from locations other than the pacemaker region (i.e. the antrum) and/or propagate in aberrant directions (i.e. in a retrograde direction towards the esophagus or a bifurcated combination of

anterograde/retrograde propagation). These abnormal initiation and propagation patterns disrupt the spatial coordination of smooth muscle contractions and have been observed in GP/FD patients [110].

1.3 Previous Work

High-resolution spatiotemporal mapping of the gastric slow wave has been performed with electrodes placed directly on the surface of the stomach during open-abdominal surgery. These findings indicated normal slow wave behavior in control subjects (n=12) [109] and abnormalities in slow wave initiation and propagation in $> 80\%$ of GP subjects (n=12) [110]. Whereas these studies demonstrated spatial slow wave abnormalities as possible explanations for foregut disorders, the involvement of surgical procedures limits the technique's scalability.

A non-invasive alternative is the electrogastrogram (EGG), which involves electrodes placed on the surface of the abdomen (as opposed to the stomach) and measures electrical activity that reaches the abdominal surface via volume conduction. Conventional EGG is comprised of 3-4 electrodes and is analyzed via spectral methods (i.e. power-frequency) [80, 118]. However, it has been shown that i) spectral features do not reliably correlate with symptoms or disease [69, 107] and ii) spatial slow wave abnormalities (which *do* have high correlation with disease and symptoms) are undetectable by spectral analyses [110, 53, 19, 23].

The high-resolution EGG (HR-EGG) has recently been established to overcome the pitfalls of traditional EGG. This approach is comprised of a dense cutaneous electrode array and subsequent signal processing methods for artifact rejection [52]. HR-EGG has been shown to capture slow waves with high spatial resolution and extract meaningful spatial *(as opposed to spectral)* features, including the instantaneous slow wave direction at each electrode [50]. With the advent of these new techniques, in addition to novel artifact rejection methods [52], spatial features have been shown to correlate with symptom incidence and severity in gastroparesis and

functional dyspepsia patients [54].

1.4 Contributions of Chapter 2

Taking it a step further than symptom correlation, it has yet to be determined if one can classify a normal from abnormal slow wave with HR-EGG recordings. Training a machine learning algorithm requires that data be labeled with ground-truth metrics. There are no current methods to obtain cutaneous EGG recordings labeled with spatial gastric slow wave abnormalities. Symptom reports, for example, cannot be used to label data, as symptoms can arise from a variety of disease etiologies [56]. In addition, a small subset of patients with diagnosed functional and motility disorders lack spatial slow wave abnormalities [110], so clinical diagnoses cannot be used to robustly label the data either. Currently accepted approaches to label these spatial abnormalities involve placing electrodes directly on the serosal surface of the stomach during an open abdominal surgical procedure [109, 110]. Because of this, simultaneous cutaneous and serosal recording has yet to be performed in humans. Thus, a simulation for which ground truth labels are established *a priori* makes possible the training of a machine learning algorithm with multi-electrode EGG waveforms.

Recent advances in deep learning have enabled breakthrough performance in the analysis of physiological signals [4, 129, 5, 6, 132, 103]. The state-of-the-art class of neural network architecture used in image-classification tasks is a convolutional neural network (CNN) [82, 152, 37, 149]. Similarly, three-dimensional CNNs (3D CNNs) are an accepted best-practice in video-classification tasks [75, 156, 96]. In a video recognition task, the 3D CNN 'sees' the video as an 'N' by 'N' grid of discrete pixels with varying intensity values over time. The data collected by a square multi-electrode array, as seen in this study, is an 'N' by 'N' grid of voltage values over time. Here, we make the novel analogy of a multi-electrode recording to a video so that we can subsequently apply these video classification techniques from deep learning. We hypothesize that

a 3D CNN, can also be used to accurately classify normal versus abnormal slow wave propagation from cutaneous HR-EGG recordings.

In Chapter 2, we establish a deep CNN framework to disambiguate normal from abnormal slow wave initiation with cutaneous multi-electrode data. We also investigate the effects of non-idealized measurements on classification accuracy, which include shifted array positioning with respect to the true center of the stomach, smaller array sizes, high BMI, and low signal-to-noise ratio (SNR). Finally, we compare the performance of our deep CNN to a traditional machine learning approach using spatial HR-EGG features of slow wave propagation. These objectives are carried out through a series of simulations of the normal and abnormal slow wave on a variety of computed tomography (CT)-extracted stomachs from human research subjects, propagated to virtual electrodes on the surface of subject-specific abdominal models. These findings suggest that the use of multi-electrode cutaneous abdominal recordings combined with machine learning algorithms may serve as a promising avenue of further research in developing widely-deployable clinical screening tools for gastrointestinal foregut disorders.

1.5 Contributions of Chapter 3

In Chapter 2, we demonstrated the first study in the field to use HR-EEG waveforms to classify, with a 3D convolutional neural network machine learning approach, normal versus abnormal spatial slow wave patterns [7]. However, there was still an unmet need to build methods and assess classification performance when considering other forms of variability reflective of human recordings, including location of spatial slow wave abnormalities and recording start times relative to slow wave phase.

In Chapter 3, we use transfer learning from the 3D CNN developed in Chapter 2 [7] to build a method that is robust to heterogeneity in both the location of the abnormal initiation on the stomach surface as well as the recording start times with respect to slow wave cycles. Specifically,

using the simulation framework in Chapter 2 as a starting point, we now introduce variability in five different possible locations of the abnormal initiation. We also shuffle the 'start' time of every recording, so that the gastric slow wave was not at the same point in its cycle at the beginning of every recording.

Transfer learning within the context of deep learning is the process by which a neural network is trained for a particular classification task using 'source' data and then re-trained in some capacity with new 'target' data in order to perform a different classification task, the 'target' task [72, 11, 15, 40, 168, 145]. The theory behind this approach is that information learned in the source task is beneficial for model fitting in the target task [72, 11, 15], even when the source and target tasks vary significantly [145], and with low availability of target data to train with [40, 72, 15]. The first layers in a sequential deep neural network select for general high-level features in the data, whereas the final layers select for patterns that are highly-specific to the training data. As such, the degree of similarity between the source and target classification tasks can impact which features of the source model are transferred to the target task [168].

We find that by starting with training lowest-complexity models and building complexity in training sets, transfer learning one model to the next, the final network exhibits high accuracy and low type-I error probabilities across all locations. Altogether, this work showcases the potential of developing non-invasive assessment tools that are more robust to variability in human EEG recordings, and may propel scalable approaches to non-invasively screen for spatial slow wave abnormalities and improve the tailoring of etiology-specific treatment strategies.

1.6 Contributions of Chapter 4

The methods explored in Chapters 2 and 3 operate under a supervised framework. Supervised methodologies require ground truth labels, which, in this dissertation, were derived from invasive studies [109, 110] that powered *in silico* modeling. While we assert that these invasive

metrics provide a good framework for describing normal and pathological slow wave physiological phenotypes, we also note the fundamental limits with regard to sample sizes within the realm of intrasurgical human recordings. As such, we explored unsupervised and semi-supervised methodologies, which did not require recordings to be labeled with slow wave abnormalities, in the remaining two chapters of this dissertation with analyses in the domain of probability distributions.

We propose an anomaly detection methodology in Chapter 4 that can be applied across a diverse range of fields, such as computer security, defense, and medicine [10, 33, 124, 150], in addition to problems in Gastroenterology discussed in this dissertation. Traditional anomaly detection methods operate under the paradigm of first defining 'normal' behavior and subsequently identifying anomalies that deviate from normal patterns, in both supervised and unsupervised manners [162, 45]. We will restrict our discussion to unsupervised clustering algorithms, as these methods do not required labeled datasets. In clustering methods, such as the K-Means algorithm [64], data are aggregated into groups with high inter-class and low intra-class variability. One of the most significant pitfalls of K-Means in the context of anomaly detection is that it performs grouping over the entire dataset, and does not directly attempt to identify anomalies [64, 43].

Optimal transport theory is a state-of-the-art framework to establish a metric between probability distributions [159] and has powered state-of-the-art studies in broad applications of mathematics, data science, and economics [159, 141, 127, 48]. The 'optimal transport' map is that which minimizes the cost of transporting all of the mass in one distribution to another. The cost associated with this distance-minimizing transportation map is referred to as the Wasserstein (squared) distance. Since clustering operates on a metric between samples, within the context of clustering probability distributions it is natural to use the Wasserstein metric. There have been multiple studies in the field that have followed a K-Means clustering framework with the Wasserstein metric between probability distributions [99, 100, 166]. Instead of cluster centriods, as is standard in K-Means clustering, these methods form clusters around Wasserstein barycenters.

However, the main drawback of the K-Means algorithm as it pertains to anomaly detection still applies: anomalies are not directly accounted for.

There is an unmet need to design an anomaly detection technique that i) operates on a set of probability distributions as an input, ii) directly targets the identification of anomalies, and iii) is unsupervised. One of the well-established principles in the field of anomaly detection is that specific algorithmic needs can vary significantly on a study-by-study basis [10, 68]. As such, instead of proposing a specific algorithm, in this work we propose a philosophical framework for detecting anomalies in a set of probability distributions. Namely, we examine a framework in which we partition subsets of the set of distributions, compute a Wasserstein barycenter between distributions in each subset, and finally compute Wasserstein distances between each distribution in a subset and the subset barycenter. We hypothesize that anomaly distributions will be consistently 'far' from corresponding subset barycenters. The main contribution to this work is a rigorous mathematical proof to show that the largest possible Wasserstein distance between a probability distribution and a subset barycenter, given *theoretical* set parameters, is the distance between an anomaly distribution and the barycenter of a subset containing the anomaly. We propose that the mathematical proof will allow future studies to design algorithms in a tailored way, using this philosophical framework. Finally, we demonstrate an example of an algorithm derived from this framework along with an example in Gastroenterology.

1.7 Contributions of Chapter 5

The final contribution of this dissertation builds off of previous work in the field which established the first preliminary linear correlation between incidence of non-anterograde slow wave propagation and symptom severity [53]. This finding has has some key limitations. Since multiple disease etiologies underlie GP and FD, a 'one size fits all' regression paradigm is likely inadequate to explain symptom severity solely from spatial slow wave features. For instance,

in patients with high symptom severity arising from etiologies other than gastric myoelectric dysfunction, spatial slow wave features would likely resemble those of healthy controls although their symptom severity score would exceed that of control subjects [53].

This line of reasoning calls for the implementation of a "robust regression" method which involves a mathematically rigorous approach to determine an inclusion criterion of data to be used in regression, and an exclusion criterion of data to be considered outliers. Robust regression has been widely used in the current literature to fit highly-accurate regressions in the presence of outliers by assuming the probability distribution of residuals has multiple modes [137, 18, 32, 35]. Optimization of model-fitting parameters typically involves maximizing a multi-modal likelihood function over the residual errors, with at least one mode near zero [137, 18, 32]. This method optimizes regression fit over the data assumed to represent 'signal' rather than 'noise.'

Another potential improvement to the preliminary correlation study in [53] involves estimating an entire probability distribution of wave directions for each recording. This provides more information, pertaining to modality and overall distributional structure, that may lead to improved characterizations of recordings. Furthermore, utilizing the full probability distributions enables the use of an optimal transport framework to build metrics between distributions.

We use robust regression of features pertaining to Wasserstein distances between probability distributions and standardized patient symptom metrics in order to identify and exclude an "outlier" class of patients for which we hypothesize that their underlying GP or FD etiology is *not* gastric myoelectric dysfunction. We then perform clustering of linearly transformed features of the outlier class to isolate a subset of patients with i) features that are indistinguishable from healthy controls and ii) symptoms that exceed regression predictions. While the regression is supervised, the clustering and detection of this patient subset is unsupervised, and as such the methodology in this chapter is semi-supervised (and does not require recordings that are labeled with respect to slow wave abnormalities). We report a highly accurate correlation between symptoms and slow wave features in the non-outlier group of subjects and give an indication

of this method's ability to predict treatment response, which has the potential to rapidly guide effective treatments and promote good clinical outcomes.

Chapter 2

A Deep Convolutional Neural Network Approach to Classify Normal and Abnormal Gastric Slow Wave Initiation from the High Resolution Electrogastrogram

2.1 Introduction

Gastric motility and functional disorders are abundantly common, with functional dyspepsia and gastroparesis, the most common of such disorders, affecting 10% and 1.5-3% of the population, respectively [42, 77]. One way in which these disorders are currently classified is based on symptom criteria, an insufficient metric due to patient subjectivity and symptom overlap between varying disease etiologies [56]. The clinical gold standard for diagnosing upper GI disorders is gastric emptying, which typically involves imaging after ingestion of a meal containing

a radioactive tracer. Gastroparesis, or delayed gastric emptying, is diagnosed if an insufficient amount of the tracer has emptied out of the stomach, and occurs in a majority of Parkinson's and diabetes patients [65, 70]. However, published findings have had limited success in demonstrating correlation between gastric emptying and symptoms [123] or symptom improvement [74]. As some drugs improve symptoms but not gastric emptying and vice versa [39, 97, 21], the NIH Gastroparesis Consortium has recently recommended that improvement in gastric emptying not be considered a requirement for clinical drug trials in gastroparesis [123].

It has been proposed that gastric motility disorders (including gastroparesis) should be sub-typed to improve treatment efficacy [86, 73]. One such etiology arises from abnormalities in the neuromuscular patterns of the gut. Much like the electrical rhythms of the brain and heart, the gastroenterological (GI) tract has electrical activity, governed by the enteric nervous system. The electrical wave propagating along the serosal (outer) surface of the stomach, the gastric slow wave, oscillates at approximately 3 cycles per minute (0.05 Hz) and coordinates the smooth muscle contractions of peristalsis during digestion [80]. Recent findings from invasive electrode recordings [109, 110] placed directly on the stomach surface describe a 'normal functioning' slow wave, with initiation in the pacemaker region and anterograde propagation of equipotential rings, as well as 'abnormal initiation', with slow waves generated outside the pacemaker region, propagating simultaneously in both anterograde and retrograde directions (Fig. A.1).

This archetypal dysrhythmia, as well as others, has been associated with gastric functional and motility disorders, including gastroparesis and functional dyspepsia [90, 34, 25, 144, 87, 111, 114]. Other work, also involving electrodes placed directly on the surface of the stomach, has shown the ability of gastric slow wave abnormalities to predict states of nausea and vomiting [105]. Recent advances in the development of gastric pacemaking devices [2, 108, 12, 125] that can initiate, entrain and/or normalize propagation of the gastric slow wave has added impetus to the identification of patients with gastric arrhythmias who may benefit from an electrical pacing intervention. While monitoring the slow wave with electrodes placed on the stomach

during invasive surgery is not clinically scalable as a screening tool, it does provide validation for electrode-based characterization of the slow wave.

There is an unmet need to develop a widely-deployable screening tool that is i) non-invasive, ii) able to make direct claims on gastric myoelectric dysfunction, and iii) able to guide effective treatments. In a similar fashion to that of the brain and heart, waves generated at the stomach's surface propagate to the skin via volume conduction. These voltages can be measured with cutaneous electrodes, namely the electrogastrogram (EGG) in the case of gastric electrophysiology. Traditionally, the EGG is comprised of a small number of electrodes (typically 3-4) and spectral analysis within the 0.05 Hz frequency range is performed [80]. Whereas some findings based upon these spectral analyses have shown separation between patients with GI symptoms and healthy controls [143, 119, 81], others have found no such differences [69, 107].

Although it is non-invasive, the EGG has not encountered widespread clinical adoption because of these contradictory findings and because of a lack of consistency with invasive clinical gold standards [3, 118]. Numerous findings have shown that spatial dysrhythmias, including the abnormal initiation, typically occur within the normal frequency range of the gastric slow wave (0.05 Hz) [110, 19, 23]. As such, current literature has converged on the notion that spectral-based analyses are unable to capture these spatial abnormalities [41, 112]. This may explain why spectral analyses have been inconsistent in their findings.

Since the most significant features of the slow wave in terms of classifying spatial abnormalities are its initiation location and propagation direction, it is reasonable to hypothesize that a multi-electrode recording system with higher spatial resolution may be beneficial in detecting such abnormalities. Indeed, recently the high-resolution EGG (HR-EGG), a multi-electrode array of 25 or more electrodes, has been shown to capture slow waves with high spatial resolution and extract meaningful spatial *(as opposed to spectral)* features, including the instantaneous slow wave direction at any given point in space [50]. With the advent of these new techniques, in addition to novel artifact rejection methods [52], spatial features have been

shown to correlate with symptom incidence and severity in gastroparesis and functional dyspepsia patients [54].

Taking it a step further than symptom correlation, it has yet to be determined if one can classify a normal from abnormal slow wave with HR-EGG recordings. Since the HR-EGG has been a recent development, modern machine learning has been seldom employed to analyze EGG waveforms and attempt 'normal' and 'abnormal' classification. Training a machine learning algorithm requires that data be labeled with ground-truth metrics. Unlike in the field of cardiology, where electrophysiological abnormalities can be detected by visual inspection of an ECG, there are no current methods to obtain cutaneous EGG recordings labeled with spatial gastric slow wave abnormalities. Symptom reports, for example, cannot be used to label data, as symptoms can arise from a variety of disease etiologies [56], one of which being dysrhythmias of the enteric nervous system. In addition, a small subset of patients with diagnosed functional and motility disorders lack spatial slow wave abnormalities [110], so clinical diagnoses cannot be used to robustly label the data either. Currently accepted approaches to label these spatial abnormalities involve placing electrodes directly on the serosal surface of the stomach during an open abdominal surgical procedure [109, 110]. Because of this, simultaneous cutaneous and serosal recording has yet to be performed in humans. Thus, a simulation for which ground truth labels are established *a priori* makes possible the training of a machine learning algorithm with multi-electrode EGG waveforms.

The simulation of electrical conduction patterns on an organ's surface is not unprecedented. Recent studies in cardiology, for example, simulate abnormal cardiac electrophysiology along with normal cardiac function on subject-specific MRI-constructed and CT-constructed heart models to perform subsequent analysis, some of which even simulate volume conduction and propagation of voltages to virtual cutaneous electrodes [169, 160, 67]. With an analogous methodology, in this work we simulate underlying conduction patterns on the stomach and propagate voltages to virtual cutaneous electrodes in a human subject-specific manner. Once labeled waveforms are obtained,

machine learning is possible. Machine learning studies in cardiology [4, 129, 5, 146, 130] follow the premise that within the same class, the conduction pattern on the organ is consistent, and across different classes, the conduction pattern is different. This same paradigm can be applied to gastric electrophysiology.

Recent advances in deep learning have enabled breakthrough performance in healthcare applications [47], including the analysis of physiological signals such as the electrocardiogram [4, 129, 5], electroencephalogram [6, 132, 103], and electromyogram [14, 165]. We hypothesize that such tools can also be used to accurately classify normal versus abnormal slow wave propagation from cutaneous HR-EGG recordings, with robustness to high anatomical variability between subjects, including BMI, stomach shape, and stomach position relative to cutaneous landmarks. Specifically, we hypothesize that a three-dimensional convolutional neural network (3D CNN) will succeed in such a classification task, with reasoning as follows.

The state-of-the-art class of neural network architecture used in image-classification tasks (i.e. classifying animal images) is a convolutional neural network (CNN) [82, 152, 37, 149]. Similarly, three-dimensional CNNs are an accepted best-practice in video-classification tasks [75, 156, 96]. In a video recognition task, the 3D CNN 'sees' the video as an 'N' by 'N' grid of discrete pixels with varying intensity values over time. The data collected by a square multi-electrode array, as seen in this study, is an 'N' by 'N' grid of voltage values over time. Here, we make the novel analogy of a multi-electrode recording to a video so that we can subsequently apply these video classification techniques from deep learning.

In this paper, we establish a deep CNN framework to disambiguate normal from abnormal slow wave initiation with cutaneous multi-electrode data. We also investigate the effects of non-idealized measurements on classification accuracy, which include shifted array positioning with respect to the true center of the stomach, smaller array sizes, high BMI, and low signal-to-noise ratio (SNR). Finally, we compare the performance of our deep CNN to a traditional machine learning approach using spatial HR-EGG features of slow wave propagation. These objectives

are carried out through a series of simulations of the normal and abnormal slow wave on a variety of computed tomography (CT)-extracted stomachs from human research subjects, propagated to virtual electrodes on the surface of subject-specific abdominal models.

With an ECG, identification of certain cardiac abnormalities can be done via visual inspection of the waveform. This is not the case for the EGG. Therefore, the methodology presented in this study involving machine learning of features was necessary in order to separate normal and spatially abnormal EGG waveforms. This is the first study to attempt, and moreover succeed, in using a deep CNN to disambiguate normal and abnormal gastric slow wave patterns with HR-EGG waveforms. These findings suggest that the use of multi-electrode cutaneous abdominal recordings combined with machine learning algorithms may serve as a promising avenue of further research in developing widely-deployable clinical screening tools for gastrointestinal foregut disorders.

2.2 Methods

2.2.1 The Forward Model

We simulated two modalities of the gastric slow wave on stomach geometry extracted from CT scans from 40 human research subjects: (i) normal initiation and propagation, and (ii) abnormal initiation with bifurcated retrograde and anterograde propagation (Fig. A.1b-c). First, we defined a novel 'Medial Curve' that captured the stomach's characteristic shape (Section 2.2.1). We then sliced the stomach geometry into thin planes, with each plane containing a discrete point on the Medial Curve and oriented organoaxially (such that a normal vector to the plane was parallel to the Medial Curve's tangent vector at the aforementioned point). We solved for the serosal stomach voltage at each point on the Medial Curve, then applied it to all points within the corresponding slice (Section 2.2.1), creating equipotential rings across the entire stomach surface. Next, we calculated subject-specific abdominal boundaries (Section 2.2.1) and propagated the

serosal voltages to electrodes placed at these abdominal boundaries (Section 2.2.1). Finally, we perturbed the simulations by varying electrode placement, abdominal tissue depth, signal-to-noise ratio (SNR), and electrode array size, with combinations of these yielding more than 2,400 independent simulations from each stomach model (Section 2.2.1). Ethical approval for this work was obtained from the institutional review board at the University of California, San Diego.

Defining the Medial Curve

The Medial Curve provided the backbone for traveling slow wave propagation. This curve, $C : [-1,1] \rightarrow \mathbb{R}^3$, is a collection of (x,y,z) points on the surface of the stomach – parameterized from the esophageal to pyloric end – that traces the stomach's unique characteristic shape (Fig A.2). To find this curve, we manually extracted and voxelized the three-dimensional stomach model from CT images [106, 101]. The voxelized representation of the stomach was iteratively thinned to its 'geometric skeleton' [116, 102] (Supplemental Materials Section A), which is a set, $\mathcal{B} = \{p_1, p_2, \ldots, p_M\}$, of M points where each $p_m \in \mathbb{R}^3$. The geometric skeleton preserves stomach topography and consists of a central spine and branches, similar to the vasculature of a leaf. The spine of the geometric skeleton best approximates the characteristic shape of the stomach. M was 5374 points, on average. Although via visual inspection, the spine of the geometric skeleton roughly traces the organoaxis of the stomach, it is unknown which points in $\mathcal{B}$ comprise the spine and which comprise the branches. Furthermore $\mathcal{B}$ is not parameterized from the esophageal end to the pyloric end. Considering this, and since gastric slow wave propagation occurs organoaxially, we developed a method using all points in $\mathcal{B}$ to construct a continuous and differentiable function $C(\zeta)$ that roughly traces the organoaxis of the stomach.

The Medial Curve, $C(\zeta)$, was constructed as a linear combination of Legendre polynomials $(\phi_k : k \geq 0)$:

$$\phi_0(\zeta) = 1, \quad \phi_1(\zeta) = \zeta, \quad \phi_2(\zeta) = \frac{1}{2}(3\zeta^2 - 1), \quad \ldots$$

We chose to construct $C(\zeta)$ using Legendre polynomials because they each are continuous and differentiable and form an orthogonal basis of functions on the [-1,1] interval. As such, the weighted combination of the Legendre polynomials used to define the Medial Curve is still continuous and differentiable. Continuity allowed for mapping of slow wave propagation along the organoaxis. Differentiability allowed us to build equipotential rings as thin planes comprising a point on the Medical Curve and its associated tangent vector, for full serosal propagation.

A vector-valued coefficient $\underline{g}_k \in \mathbb{R}^3$ is associated with the kth Legendre polynomial ϕ_k so that by defining the composite matrix G containing vector-valued coefficients as column vectors, $G = \left[\underline{g}_1\ \underline{g}_2\ \dots\ \underline{g}_K\right]$, we can succinctly describe the curve for any matrix G as:

$$C(\zeta;G) = \sum_{k=0}^{K-1} \underline{g}_k \phi_k(\zeta). \tag{2.1}$$

There exists a matrix, G^*, for which $C(\zeta;G^*)$ best fits the spine of the geometric skeleton. To find G^*, we defined the objective function $J_{\mathcal{B}}(G)$ to be minimized in (2.2) below. It was necessary to incorporate both (2.2a) and (2.2b) into $J_{\mathcal{B}}(G)$ in order to ensure that each point on the geometric spine was near some point on the Medial Curve, (2.2a), *and* each point on the Medial Curve was near some point on the geometric spine, (2.2b). Without the inclusion of (2.2b), there are an infinite number of curves, $C(\zeta)$, for which (2.2a) holds that extend past the esophageal and pyloric boundaries.

$$\begin{aligned} J_{\mathcal{B}}(G) &= \left(\sum_{m=1}^{M} \min_{\zeta\in[-1,1]} \|C(\zeta) - p_m\|_2^2\right) && \text{(2.2a)} \\ &+ \left(\int_{-1}^{1} \min_{m\in\{1,\dots,M\}} \|C(\zeta) - p_m\|_2^2 d\zeta\right) && \text{(2.2b)} \end{aligned}$$

We solved for G^* by minimizing (2.2), using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm:

$$G^* = \underset{G\in\mathbb{R}^{3\times K}}{\operatorname{argmin}} J_{\mathcal{B}}(G). \tag{2.3}$$

The optimal Medial Curve was given by $C^*(\zeta) \equiv C(\zeta; G^*)$.

Serosal Voltage Simulation

We simulated voltage potentials on the full serosal surface of the stomach (Fig A.3). We modeled the normal and abnormal wave initiation and propagation patterns to be consistent with recent findings from invasive human recordings [109, 110]. This was implemented by solving the one dimensional wave equation (2.4) at discrete points along the Medial Curve via finite difference analysis with a temporal step size calculated using the Courant-Friedrichs-Lewy condition [50]:

$$\frac{\partial^2 S(\zeta,t)}{\partial t^2} = c^2(\zeta)\frac{\partial^2 S(\zeta,t)}{\partial \zeta^2}. \tag{2.4}$$

In (2.4), $S(\zeta,t)$ is voltage as a function of both time t and position ζ on the Medial Curve. Wave speed, $c(\zeta)$, is a function of the Euclidean position $C^*(\zeta)$ corresponding to position ζ on the Medial Curve, which is highest in the pacemaker region (6.0 mm/s), second-highest in the antrum (5.9 mm/s), and lowest in the corpus (3.0 mm/s). We also imposed trends in wave amplitude consistent with the current literature [109, 110]; amplitudes in the pacemaker, antrum, and corpus regions were 0.57 mV, 0.52 mV, and 0.25 mV, respectively. At the two boundaries, we employed Mur's boundary condition to prevent waves from reflecting back into the stomach. Finally, we applied each discrete voltage, $S(\zeta,t)$, in equipotential rings oriented organoaxially on the stomach associated with points on the Medial Curve $(C^*(\zeta) : \zeta \in [-1,1])$. Altogether this defined the potential at all points on the stomach's serosal surface.

Abdominal Boundaries and Electrode Array Placement

To approximate any research subject's cutaneous abdominal boundaries, we constructed a dense cylindrical point cloud where each cross section is an ellipse. The major and minor axes of

the ellipse were defined from measurements of the projection of the research subject's abdominal CT scan onto the X,Z plane (Figure A.4a). The point cloud was placed concentrically with the stomach.

We assume the smallest possible thickness of the abdominal wall (skin, adipose tissue, muscle, etc.) to be 1 cm. To create a 1 cm shell between the inner and cutaneous boundaries, we created an inner elliptic cylinder (blue ellipse in Fig. A.4b), with elliptical radii 1 cm less than the radii of the cutaneous elliptic cylinder.

In order for the point on the anterior surface of the stomach closest to the abdominal wall to be just within the inner elliptic cylinder, we laterally translated the two elliptic cylinders in the posterior direction by z_0, which was minimized subject to the constraint that each point from the stomach lies within the inner ellipse shifted by z_0:

$$\frac{(x_i)^2}{(a-1)^2} + \frac{(z_i - z_0)^2}{(b-1)^2} \leq 1, \quad i = 1, \ldots, H. \tag{2.5}$$

where (x_i, z_i) is the ith point from the stomach, a is the major radius and b is the minor radius of the cutaneous ellipse. This constrained convex optimization problem was solved in a subject-specific manner due to subject-to-subject stomach and abdominal variability (Supplemental Materials Section G). H was, on average, 7446 points.

At this point, the cutaneous boundary was centered at $(0, z_0)$. To model BMI trends, the cutaneous boundary was subsequently translated in the anterior direction (increasing the separation between the anterior stomach and abdominal boundaries) by u cm, resulting in its center now being $(0, z_0 + u)$. We defined the 'abdominal tissue depth' to be $1 + u$ cm for $u \geq 0$.

We constructed an electrode array on the anterior face of the cutaneous surface. Electrode 'anchor points' were placed in a grid, spaced 2 cm apart both vertically and along the elliptical arc length. Full electrodes were defined as the closest 32 points within the elliptic cylinder point cloud to each anchor point. When computing electrode voltages (Section 2.2.1), the voltage at

each full electrode is an average of voltages at these these 32 points. The entire 100-channel electrode array was centered with respect to the stomach's (x,y) center.

Propagating Voltage from the Serosal Surface of the Stomach to Cutaneous Electrodes

At each electrode n, we calculated the voltage at time t, $V_n(t)$, using the principle of superposition as a linear combination of D 'source' current dipole moments, $(I_d(t) : d = 1,\ldots,D)$. Each current dipole $I_d(t)$ is taken from the serosal simulation in Section 2.2.1 and either corresponds to the potential along the Medial Curve at time t, $S(\zeta,t)$, or one of the associated equipotential rings on the plane normal to the tangent vector of $S(\zeta,t)$ at point ζ. We also model additive white Gaussian measurement noise $N_n(t)$, altogether giving rise to:

$$V_n(t) = \sum_{d=1}^{D} W_{n,d} I_d(t) + N_n(t). \tag{2.6}$$

Source weights are given by:

$$W_{n,d} = \frac{\cos\theta}{4\pi\sigma r_{n,d}^2} \tag{2.7}$$

where θ is the angle between the organoaxial current dipole and the electrode, $r_{n,d}$ is the distance between the source, d, and electrode, n, and σ = 0.125 S/m is the conductivity of the medium between the stomach and electrode interface (i.e., homogenized between abdominal bone, muscle, adipose tissue, and skin) and was chosen to be between that of muscle (0.5 S/m) and fat (0.1 S/m) [128, 135]. Since EGG is at 0.05 Hz, we did not incorporate capacitive effects, as has been done for volume conduction modeling with electrophysiologic signals at higher frequencies [136, 57]. In this work, when we shift the electrode array on the surface of the abdomen and present findings, we term some electrodes as 'out of range of the stomach's electrical activity.' The rationale of this is due to source weights being governed by i) the distance between the electrode and voltage source on the stomach and ii) the angle between these two points (2.7). As such, in this context,

electrode array shifts will cause high attenuation of these weights. Voltage at these electrodes will be governed primarily by noise, therefore justifying the terminology 'out of range of electrical activity.'

Creating HR-EGG Datasets

For each simulation of the slow wave on the serosal surface of the stomach, we generated several independent HR-EGG datasets via manipulation of electrode array placement, abdominal tissue depth, electrode array size, and signal to noise ratio (SNR). We shifted the electrode array horizontally such that the center of the array moved along the abdominal elliptical arc from -12 cm to 12 cm in increments of 3 cm. Likewise, we shifted the center of the array vertically from -12 cm to 12 cm in increments of 3 cm. Each stomach-abdomen model began with a minimum spacing of 1 cm between stomach and abdominal boundaries (Section 2.2.1). From this initial positioning, we moved the electrode array laterally away from the stomach up to 15 cm in increments of 1 cm, to simulate changes in abdominal tissue depth, with the assumption that this suggests trends in BMI. This method of adding tissue depth created a close resemblance to abdominal tissue regions seen from the abdominal CT scans.

We added white Gaussian noise to all simulated HR-EGG datasets. It should be noted that this 'measurement noise' is an addition to the noise already present in the simulated volume conduction solution (2.6). In the training datasets, we calculated the noise variance such that the ratios of median signal power variance to added noise variance yielded a SNR of 20 dB. In the same fashion, we added white Gaussian noise to the test datasets used in the experiments to assess classifier performance over a SNR range of -40 to 40 dB. We added noise in a related but slightly different manner to test datasets used in the experiments to assess classifier performance over horizontal and vertical electrode array shifts, as well as abdominal tissue modulation. In these latter three test datasets, we calculated the noise variance once for all HR-EGG datasets

resulting from each distinct stomach model such that the resulting SNR was 10 dB at the centered configuration (i.e., horizontal and vertical shifts were 0 cm, and abdominal tissue depth was at its 1 cm baseline). We then added white Gaussian noise with these calculated variances to all horizontally, vertically, and laterally shifted permutations of the HR-EGG dataset generated from the particular stomach model.

Previous studies have used configurations with fewer than 100 electrodes. For example, the original HR-EGG recordings utilized 25 electrode arrays [112, 50] and ambulatory systems capable of recording from 9 electrodes have recently been established [52]. As such, we trained and tested smaller square electrode arrays with 25 and 9 channels and added noise for all training and test datasets of the smaller arrays as described above.

2.2.2 Machine Learning Classification of HR-EGG Waveforms

We constructed and trained a convolutional neural network (CNN) to classify normal and abnormal HR-EGG electrode data (Section 2.2.2). For comparison, we computed wave propagation spatial features to train a linear discriminant analysis (LDA) classifier (Section 2.2.2).

Neural Network Architecture

The CNN consisted of four sequentially ordered 3D-convolutional layers, each with 32 filters, followed by two dense (fully connected) layers, with 64 and 2 filters, respectively. A schematic diagram is shown in Fig. A.5. Layers 1 through 5 were activated with a rectified linear unit (ReLU) and the ultimate dense layer was activated with a softmax threshold. We optimized weights via back propagation with an Adam optimizer, using the following hyperparameters: learning rate = 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Loss was computed via categorical cross-entropy. The outputs of layers two and four underwent max pooling (dimension 2,2,2) in order to reduce the run time of the algorithm. We added a dropout function after the second, fourth, and fifth

layers to prevent over-fitting [151]. We carried out all computations through Tensorflow [1] and programmed nodes using the Keras API [36].

The general form of the neural network architecture was adapted from state-of-the-art 2D and 3D CNN published architectures [82, 152, 75, 156, 96]. Namely, this 'general form' is a sequence of several convolutional layers followed by a few dense layers, in which the ultimate dense layer has 'n' filters corresponding to the 'n' classes in the data (in this case 2). The number of filters in each layer is typically task and data-specific, with some studies using 2 filters in a layer and others using 2,000 [82, 152, 37, 149, 75, 156, 96]. As such, there is not a 'standard' number of filters to use in CNN layers. We implemented 32 filters for each convolutional layer, which was sufficient for the CNN to accurately learn features and classify while keeping the computational cost low. State-of-the-art CNN's [82, 152] train on datasets on the order of 1.2 million samples (i.e. ImageNet), which allows a large set of weights arising from a very deep network to be learned. In this study, however, the training set contained only approximately 6,000 samples, which led us to choose the depth of the network to i) be sufficient for accurate feature identification, ii) avoid over-fitting, and iii) be computationally efficient.

We chose a convolution stride of 1 to retain resolution at the convolution stage, since the aforementioned architecture design reduced the computational load enough to make this stride feasible. We used the same kernel size for filters in each layer in accordance with recent findings suggesting that homogeneous kernel sizes throughout the network are optimal when classifying videos through a 3D CNN [156]. The kernel size was chosen by the input size of the data. State-of-the-art CNN's [82] trained on data with image sizes of 256x256 pixels (i.e. ImageNet) and had kernel sizes ranging from 11x11 in beginning layers to 3x3 in latter layers. Each simulated waveform in our study had an input size of 10x10x300, where the third axis was the temporal component. As such, we chose the first two dimensions of our kernel size, 2x2, to be sufficiently smaller. The final dimension of our kernel size, 2, was also chosen to be small in order to preserve temporal resolution, overall yielding a convolutional kernel of 2x2x2.

Before undergoing any training, the cross validation 'accuracy' is 0.5, since CNN weights are randomly initialized. As seen in Fig. A.7, the cross validation accuracy increases to above 0.7 after just one epoch. The accuracy continues to rise throughout the course of training, but this first jump in accuracy ($>$0.25) is the most prominent. Because of this property, we can say the neural network is 'converging' if this jump in cross validation accuracy is seen after 1 epoch. When constructing the CNN architecture, we performed 'convergence tests' in which we tested perturbations of the architecture (i.e. more or less layers, number of filters, etc.) for convergence. While there may be a slightly different 3D CNN architecture (i.e. more filters in one of the layers, or larger kernel size) that slightly improves the performance of the network, we found that convergence is robust across minor perturbations in architecture. A source of future work could involve the tuning of hyperparameters and iteratively testing architectural parameters to slightly improve CNN classification performance.

Linear Discriminant Analysis

We represented the position of each electrode $((\tilde{x}_n, \tilde{y}_n) : n = 1, \ldots, N)$ with respect to the coordinate system of the curved surface on the abdomen (Fig. A.6a). An example of the voltage $V_n(t)$, associated with two electrodes at different vertical locations, where a phase shift is present, is given in Fig. A.6b.

We extracted HR-EGG spatial features of the slow wave from the cutaneous waveforms by first performing the Hilbert transform on each individual waveform in the array to extract instantaneous amplitude and phase information:

$$V_n(t) + iHb[V_n(t)] = a_n(t)e^{i\phi_n(t)}, \qquad n = 1, \ldots, N \tag{2.8}$$

to construct instantaneous phase information as a function of time and space:

$$\phi(\tilde{x}_n, \tilde{y}_n, t) \equiv \phi_n(t), \quad n = 1, \ldots, N.$$

The spatial gradient of instantaneous phase, $\nabla\phi(\tilde{x}, \tilde{y}, t)$, was constructed at each cutaneous electrode $((\tilde{x}_n, \tilde{y}_n) : \ n = 1, \ldots, N)$.

Since the wave velocity vector v is normal to contours of constant phase, it satisfies:

$$v(\tilde{x}, \tilde{y}, t) \propto -\nabla\phi(\tilde{x}, \tilde{y}, t). \tag{2.9}$$

We define features pertaining to the time-averaged direction of the wave velocity at any electrode in terms by exploiting (2.9) as follows:

$$\Lambda_n = \frac{1}{T} \sum_{t=1}^{T} \tan^{-1} \left(\nabla\phi(\tilde{x}_n, \tilde{y}_n, t) \right), \quad n = 1, \ldots, N. \tag{2.10}$$

Speed calculated from measurements directly on the stomach [109] has been shown to follow different trends with regard to disease than speed calculated from the aggregated slow wave activity propagated to the cutaneous abdominal surface via volume conduction [50, 54, 27]. As such, the wave speed was not used as a feature to train our classifier. In order to include a feature that varies between anterograde propagation (normal function) and the combination of retrograde and anterograde propagation (abnormal initiation), we define the phase gradient directionality (PGD) as the ratio of the norm of the spatially averaged electrode velocities with the spatial average of the norm of electrode velocities:

$$PGD(t) = \frac{\| \frac{1}{N} \sum_{n=1}^{N} \nabla\phi(\tilde{x}_n, \tilde{y}_n, t) \|_2}{\frac{1}{N} \sum_{n=1}^{N} \| \nabla\phi(\tilde{x}_n, \tilde{y}_n, t) \|_2}, \quad t = 1, \ldots, T \tag{2.11}$$

where in (2.11), velocities are replaced with $-\nabla\phi$ by virtue of (2.9). The PGD is a measure of

how aligned the wave velocities at different electrodes are at any point in time, and equals 1 for planar waves, since in such cases the wave velocities at each location are equal. From Jensen's inequality, the numerator is upper bounded by the denominator and so $0 \leq PGD(t) \leq 1$. Thus the PGD is a measure of how 'close' the activity is to being a plane wave, which is akin to what occurs in normal slow-wave HR-EGG waveforms, exhibiting predominantly anterograde propagation. As waves stray from this planar phenotype, as in the abnormal slow wave seen in HR-EGG waveforms, the PGD decreases (i.e., the PGD of white Gaussian noise is with low probability above 0.5 [50]).

A conceptual schematic of PGD is shown in Fig. A.6c. The denominator of (2.11) represents the average magnitude of the spatial phase gradient. In Fig. A.6c, the magnitude of the spatial phase gradient is the same at all electrodes (all arrows are of equal magnitude). In the planar wave case, the partial derivative of phase with respect to $\tilde{x}$ is the same at all spatial locations, whereas in the non-planar wave, the partial derivative of phase with respect to $\tilde{x}$ at the two leftmost electrodes is equal to the negative value of the partial derivative of phase with respect to $\tilde{x}$ at the rightmost electrodes. The same pattern is seen in the partial derivative of phase with respect to $\tilde{y}$. Thus (2.11) is zero for the non-planar wave shown in this figure, as wave velocities are of the same magnitude but have opposing directions.

At certain horizontal shifts of the electrode array, the array may be in a range where it records mostly synchronized retrograde activity in the abnormal simulations. This will cause both the normal and abnormal waves to resemble planar waves (though propagating in different directions). The PGD will be indistinguishable between classes in this case but the temporal average of wave directions at each spatial location will be different. As such, we used the time averaged directions at each electrode $(\Lambda_n : n = 1, \ldots, N)$ as well as the median of $(PGD(t) : t = 1, \ldots, T)$ to create $N + 1$ features for training a linear discriminant analysis (LDA) classifier.

Training and Testing Both Models

Prior to training the model, we initialized all neural network weights with a Glorot Uniform distribution. We trained the CNN for 15 epochs in batches of 40 training simulations at a time and cross validated on a random 25% of the training set. The cross validation accuracy and loss at each epoch during training are plotted in Fig. A.7. All values of each of the horizontal and vertical electrode array shifts, as well as abdominal tissue depths, were represented in the training set. When sampling one of these shifts, the other two shifts type would remain centrally located. For example, when we varied the abdominal tissue depth from 1 to 15 cm, horizontal and vertical electrode arrays shifts were kept within -3 to 3 cm. Likewise, in the test datasets, we held the types of perturbations not explicitly under investigation at a range close to zero. This ensured classification performance dependence on one independent variable at a time. A full description of all datasets used in training and test experiments can be found in the Supplemental Materials Section B. Test data for all experiments was completely independent from training data. HR-EGG waveforms, including all array perturbations, from 28 randomly selected stomach models was used solely in training the CNN, while waveforms from the remaining 12 stomach models was used for testing. All training and testing regimes were repeated six times (i.e., six unique random splits of 28:12 stomach models).

2.3 Results

In both the LDA and neural network approaches, the classification accuracy saturates at high and low SNR extrema (Fig A.8a). Within the SNR range of -12 to 16 dB, the CNN outperforms the LDA classifier and both the LDA and neural network accuracy curves increase mechanistically similarly, with approximately equal slopes. When the LDA accuracy curve is shifted by -17 dB, accuracy as a function of shifted SNR more closely resembles that of the neural network. (Supplemental Materials Section C). Thus, the 'gain' from use of the neural network

methodology over the LDA classifier is 17 dB. However, the LDA accuracy still saturates at 90.97%, which is lower than the CNN's accuracy from 0 dB onward. Our group has observed in recordings that a high-quality EGG signal is in the vicinity of 10dB [52]. At 10 dB, neural network accuracy is 95.46% and LDA accuracy is 76.79%, comprising an accuracy improvement of 18.67%.

A full analysis of the vertical array translation and abdominal tissue depth perturbations (Fig. A.8c-d) can be found in the Supplemental Materials Section D. Below we will analyze the LDA and CNN accuracy curves as a function of horizontal shifts of the 100 channel electrode array (Fig. A.8b).

Since CNN filters are stepwise convolved around the entire input space, invariant properties of the input can be extracted, irrespective of their location in space. This robustness for CNN's is well exemplified within the context of animal image classification, where the face of a cat can be localized and corresponding features necessary for classification can be extracted from any spatial location in both testing and training datasets. As such, classification of these images is robust, even if the relative location of the cat's face in the testing dataset differs from its location in the training dataset [82]. Similar to how the cat's face can be anywhere within the frame of the image, the 'anterograde activity' and bifurcated 'anterograde and retrograde' activity need not be in the exact same part of the electrode array for every subject, as evidenced by high classification accuracy across a broad range of electrode array placements relative to anatomy (Fig. A.8b). In the case of the LDA, the features are specific to the spatial location of each electrode. As seen in Figure A.8b, LDA classification accuracy is sensitive to electrode placement. An explanation (admittedly over-simplified for conceptual transparency) of this vulnerability is as follows.

In the training set (identical in both the CNN and LDA training paradigms), there is an over-representation of horizontal shifts within -3 to 3 cm, as compared to other horizontal shifts outside of this range. It follows that the LDA model fit in training will associate, at a high level, with direction feature constructs as seen in the 'Training' abstractions in Fig. A.9c-d.

In this range of shifts, the array is horizontally centered near the location of the abnormal initiation, and so a "learned" representation can be simply described as follows. For the normal wave propagation, waves are propagating in an anterograde direction near the left side of the array and an anterograde direction near the right side. Since both halves of the array have waves propagating in the same direction, the PGD is high. For the abnormal initiation and propagation, there is anterograde activity near the left side of the array and retrograde activity near the right side. Since one side of the array has waves propagating in an opposite direction from waves in the other side of the array, the PGD is distinctly lower than that of the normal case.

This general logic explains the variability in performance during testing at the two annotations in Figure A.9a. At annotation 2 in Fig. A.9, the array is most centered over the abnormal initiation during testing (Fig. A.9e and Supplemental Materials Section E). As such, the learned representation of wave directions is very consistent with training (Fig. A.9d) and indeed the classification performance is highest. Furthermore, the largest class separation between the normal and abnormal PGD values is seen at annotation 2, and the magnitude of the weight associated with PGD in the LDA classifier is 40 times higher than that of other direction features.

As the electrode array is moved slightly in the negative direction (i.e. horizontal shifts of 0 and -3), the direction vectors are still overall consistent with those learned in training. However, poor classification performance is observed because the array is moving to a region over more synchronized retrograde activity. Thus, the abnormal wave resembles a planar wave, similar to the normal wave but propagating in a different direction. Because the PGD is maximized for planar wave propagation (regardless of direction), the PGD for the normal and abnormal waves will begin to be almost indistinguishable. Since the PGD is the highest-weighted feature in the LDA classifier, the sharp decline in performance is explained.

At annotation 1 in Fig. A.9, the center of the electrode array is shifted to the right of the abnormal initiation point (Fig. A.9b). When testing the abnormal waveforms, the left side of the array now records retrograde activity and the right side records noise (Fig. A.9c). Thus none

of the direction features are consistent between testing and training. When testing the normal waveforms, the left side of the array records anterograde activity and the right side records noise. Thus only the left-side direction features are consistent between testing and training. Though some of these direction features are consistent, direction features on the right side of the array are weighted higher in classification than those on the left (Supplemental Materials Section F). Moreover, wave directions are not coherent and the PGD is reduced in comparison to its learned value during training. Altogether, an erosion of classification performance ensues. Though the CNN approach is robust across horizontal shifts, we see a unimodal trend in classification accuracy with respect to electrode array positioning. We hypothesize that the most significant 'feature' used in classifying the abnormal initiation is one extracted by a filter in the shape of retrograde activity. The CNN's peak performance at a horizontal shift value of -3 cm supports this hypothesis. At this shift configuration, the electrode array is centered at a region in which there is mostly retrograde activity in the abnormal simulation (the array is out of range of the anterograde activity seen to the left of the abnormal initiation- refer to Figure A.9b). Thus, during testing, the CNN will only see retrograde activity for the abnormal simulation and only anterograde activity for the normal simulation– yielding high classification accuracy.

Under our hypothesis that the CNN is heavily selecting for retrograde activity to classify the waveforms as 'abnormal', a configuration that is more localized to the abnormal initiation location will 'confuse' the CNN with incidence of features resulting from anterograde waves in addition to retrograde waves. This is consistent with classification performance decline as a function of positive shifts (Fig. A.8b). The gradual decline in CNN classification accuracy as the array is shifted toward -12 cm is due to the array moving out of range of most of the slow wave activity in general and measurements tending toward noise.

As the electrode array is shifted 6 and 9 cm horizontally, the one standard deviation error bars for the CNN and LDA classification performance overlap. This is mainly because the shifts within 6 to 9 cm move the center of the array near the abnormal initiation site, a trend

that generally maximizes LDA performance and minimizes neural network performance. In this case, the CNN‘s worst classification accuracy is still within one standard deviation of the LDA's best classification accuracy, exhibiting the CNN's high performance across horizontal electrode array shifts. In comparison, the LDA performs significantly poorly when the array is not centered around the site of the abnormal initiation (i.e. when it is shifted by -12 cm). Since the abnormal initiation is not anatomy-specific, the corresponding horizontal shift that puts the electrode array in its range will not be known *a priori* for any individual human subject – even if an abdominal image has been obtained. Altogether, this suggests that the CNN methodology we propose would be significantly less error-prone and more robust, in comparison to the the LDA approach operating on spatial features.

Shifting the electrode array by 6, 9, and 12 cm moves the center of the electrode array toward the pyloric end of the stomach. As the electrode array is shifted in this direction, it moves away from regions of the stomach that exhibit retrograde wave activity in the abnormal case. Under our hypothesis that the CNN is heavily selecting for retrograde activity to provide an abnormal classification, moving away from regions with retrograde activity will cause a decline in performance. There is a high amount of stomach size variability across this axis (6.7 cm, Supplemental Materials section G). Thus, these shift configurations will move the electrode array completely out of range of retrograde activity (or even out of range of the entire stomach) for smaller stomachs but not for larger ones, and a large increase in accuracy variance is observed (Fig.A.8b). This split is seen to the fullest extent at shifts of 9 cm, with shifts of 6 cm having more electrode arrays still recording retrograde activity and shifts of 12 cm placing a higher number of electrode arrays out of range of retrograde activity.

Furthermore, this region of the stomach exhibits the highest amount of stomach shape variability (Supplemental Materials Section G), which, in turn, results in less consistent patterns between simulations at the electrode level. We hypothesize that this is why the variance increase is not seen symmetrically (i.e. in shifts of -6, -9, and -12 cm), as the esophageal side of the

stomach is not as anatomically variable. In the LDA classifier with electrode shifts of 6, 9, and 12 cm, the increase in accuracy variance does not mirror that of the CNN. Instead, variability in classification accuracy is seen at shifts of -3 and 0 cm. This is because LDA performance depends on the degree to which the electrode array is centered around the bifurcated anterograde and retrograde activity in the abnormal initiation. Again, due to anatomical size and shape variability, electrode arrays remain centered around the antrum at shifts of -3 and 0 cm for larger stomachs but move away from the wave bifurcation at these shifts for smaller stomachs.

As smaller array sizes are tested, slight differences in performance robustness with respect to the number of channels are observed (Fig. A.10). Performance from waveforms of the 100 channel arrays exceeds that of the 25 and 9 channel arrays by more than 1 standard deviation in the SNR range of -8 to 0 dB (Fig. A.10a). When the electrode array is shifted horizontally, the most noticeable discrepancy in performance lies in the range of -12 to -6 cm (Fig. A.10b). This is likely due to the fact that the 25 and 9 channel electrode arrays are out of range of electrical activity, whereas the leftmost electrodes of the 100 channel electrode array do remain within the range of electrical activity. In a similar fashion, the largest discrepancy in performance when the array is shifted vertically lies in the extrema (Fig. A.10c), as smaller electrode arrays move more so out of the range of electrical activity than the 100 channel electrode array. All three array sizes yield similar model classification performance trends when abdominal tissue depth is increased, as these permutations are tested only at vertically and horizontally centered electrode array configurations.

2.4 Discussion

In this study, machine learning techniques involving a CNN were applied to the slow wave of the stomach to determine if the clinically important marker of abnormal gastric slow wave initiation could be accurately identified. Machine learning was carried out under thousands

of possible test conditions and overall it performed very favorably in accurately identifying cases of abnormal slow wave initiation in these cases.

This works paves the way for development of a second generation model which incorporates additional abnormalities beyond pacemaker initiation, specifically, a conduction block and/or abnormal conduction velocity [110]. The CNN architecture developed here could be slightly modified to accommodate the classification of more than two hypotheses. Specifically, the ultimate fully connected layer would have four, as opposed to two, filters.

Although we hypothesized that modern machine learning techniques (i.e. CNNs) could be used to disambiguate normal and abnormal gastric myoeletctric function, such techniques typically require availability of massive amounts of training data. For instance, in our CNN approach, we had to train close to 175,000 parameters and test the model on several perturbations, which required over 10,000 labeled datasets. If we wanted to first directly test this in humans, this would have been time and cost prohibitive. One value of the work presented here is that we were able to rapidly determine that such approaches are worth further exploration, using human anatomy and physiologically plausible models of normal and abnormal gastric myoelectric function. Furthermore, a recent study in cardiology that used a 2D CNN to classify single-lead ECG [130] suggests that certain cardiac arrhythmias are only detectable via a multi-lead ECG. Since the 3D CNN methodology developed here allows for the analysis of a dynamic process recorded by a multi-lead system, our work paves the way for this paradigm to be applied in fields beyond gastroenterology, such as cardiology.

Another value of this work is that it is, to the best of our knowledge, the first study to explore the use of a deep CNN to disambiguate a normal and abnormally initiated gastric slow wave from multi-electrode cutaneous abdominal waveforms. Although simultaneous serosal and cutaneous electrical recordings to test our methodology in humans require abdominal surgeries and have not yet been accomplished, recent findings demonstrate the feasibility of multi-electrode gastric mucosal recordings [121] and low-resolution simultaneous mucosal and cutaneous record-

ings [148]. As these techniques advance and are used in larger patient cohort sizes, opportunities will arise to validate our methods in humans using multi-electrode mucosal recordings to define ground truth labels.

Furthermore, one recent advancement in machine learning methodologies includes the use of transfer learning, which allows a new classification task to be performed on a pre-trained model. Transfer learning can make it possible to train a neural network and perform a classification task on a dataset containing an otherwise insufficient amount of training data. With this technique, secondary training or fine tuning of previously-learned neural network weights with the new dataset is performed. Secondary training involves freezing low-specificity weights from the first few layers of the neural network while the remaining final layers remain trainable. Data from the new classification task, termed 'target data', can be used to fine-tune the highly-specificity weights corresponding to specific features. Another method involves fine-tuning the entire neural network with target data. Recent studies have demonstrated efficacy in transfer learning even when the trained and targeted datasets vary significantly [145, 40, 168].

Though many state-of-the-art studies involve transfer learning of 2D models, recent work has shown efficacy of transfer learning with 3D-CNNs [72, 104, 11]. It is reasonable to hypothesize that 100-channel EGG data collected from humans could fit within the realm of similarity to the simulation training data, for which transfer learning is effective. We hypothesize that down or up-samping the human EGG data to match the sampling rate of the *in silico* data will preserve the temporal relationship between trained and target data, which is the third axis of the 3D structure. As such, the weights learned from training the CNN in this *in silico* study could be applied with transfer learning to deploy this framework on multi-electrode human data. This transfer-learned study is likely not trivial, but is a future direction within the realm of possibility.

A common problem in medical data is a class imbalance between patients and healthy controls. However, due to the large prevalence of GI disorders, there is no dearth of patients. In addition, since the EGG is noninvasive, the barrier to enroll healthy controls is low. Thus, there

need not be high class imbalances in EGG studies with real data. If there still is a class imbalance, though, one could use the synthetic minority over-sampling technique (SMOTE) and or generate synthetic data in the minority class. Data generation in the minority class has been used in the field of cardiology [5], as well as radiology [139], via a a generative adversarial network (GAN) in the latter.

The approximations and simplifications of our *in silico* study do not, however, capture the following: (i) Slow wave abnormalities may be intermittent [52]. To account for this, one could employ state space models (i.e. hidden Markov models) or their analogues in deep neural networks (i.e. LSTM) to develop probabilistic descriptions of the state across time. (ii) This model does not simulate artifacts arising from movement and/or abdominal muscle contractions that are typically present in human electrical recordings. To address this, our group has recently developed novel artifact rejection methodology that leverages the unique oscillatory nature of the slow wave and preserves the signal even in time windows for which artifact is present [52]. One may employ these methods on the real HR-EGG data recordings to increase their similarity to the simulation datasets. (iii) The stomach geometry may dynamically change, as compared to our assumption of it being static throughout the recording. In practice, the stomach geometry expands and contracts primarily during and after a meal. Notably, this modulation is found mostly in the fundus, a region with lowest electrical activity, and thus a non-region of interest in myoelectric efforts. (iv) This model places the elliptical abdomen concentrically with the stomach, which is not true to the patient's anatomy in most cases. However, our findings pertaining to classification with respect to horizontal and vertical displacement suggest that any trends in classification accuracy are a function of abnormal initiation location, not anatomy-specific locations. (v) The forward model used in our simulation does not model all the ionic conductances underlying the generation of the slow wave. A more realistic forward model that captures normal and abnormal function has recently been established [31] and could be used as further *in silico* validation of our presented approach.

Though this study and others have shown high classification accuracy and robustness, there remains the challenge of low clinical acceptance of deep learning due to its lack of feature interpretability. To address this, rather than inputting the original spatiotemporal voltage time series, 'animated flashlight sequences' [30] may be used as a pre-processing step for which, subsequently, the aforementioned CNN methodology may be applied. This may be particularly attractive for a human-computer interface setting, since the input to the algorithm is analogous to the input that a human would use to identify abnormalities. Even if this slightly sacrifices performance, it could lead to improved interpretability and an increased chance of adoption as a screening tool by clinicians.

In cardiology, identification of patients with sick sinus syndrome, where the cardiac pacemaker cells are not initiating normally, prompts placement of a potentially life-saving pacemaker device. Indeed, a whole field called cardiac electrophysiology has developed around management of cardiac pacemaker cell arrhythmias. While these concepts are still in a nascent phase in the gastrointestinal tract, with initial efforts showing promise [2, 108, 12, 92], the work here with machine learning shows that gastric pacemaker initiation problems can be readily identified with a high accuracy. Advancement towards improved detection of the underlying stomach pathophysiology from which foregut symptoms arise will lead to improved and more etiology-specific treatment.

2.5 Conclusion

To summarize, this *in silico* study demonstrates the efficacy of using machine learning to classify normal and abnormal slow wave activity from EGG data. This technique is particularly relevant because many foregut GI disorders can masquerade as one another when relying on symptoms alone. A recent finding indicates that with imaging-guided placement of multi-electrode arrays, slow wave spatial electrical patterns become associated with disease and symptom severity

[54]. The robustness of the CNN approach shown here provides preliminary evidence that even in the absence of image guidance, and across a wide range of BMIs and SNRs, such associations may still be established. Furthermore, the approach we have developed has an ability to discern, at an individual level and with high accuracy, an abnormally functioning slow wave from one of normalcy. Apart from simply being a proof of concept, the network architecture and learned weights have the potential, through transfer learning, to reduce the amount of training data required for high performance in human studies. Altogether, these findings suggest that multi-electrode cutaneous abdominal recordings, combined with modern machine learning techniques, have the potential to address unmet needs and possibly serve as widely deployable screening tools in gastroenterology.

2.6 Acknowledgments

We'd like to acknowledge Kevin King, MD, PhD, for his advisory contributions to the paper as an expert in cardiology.

Chapter 2, in full, is a reprint of the material as it appears in IEEE Transactions on Biomedical Engineering 2019. Agrusa, Anjulie; Gharibans, Armen; Allegra, Alexis; Kunkel, David; Coleman, Todd, IEEE, 2019. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Robust Methods to Detect Spatial Abnormalities of the Gastric Slow Wave from Cutaneous Recordings

3.1 Introduction

Gastrointestinal (GI) problems are the second leading cause for missing work or school behind the common cold, with the two most common upper GI sensorimotor disorders, gastroparesis (GP) and functional dyspepsia (FD), affecting 3% and 10% of the US population, respectively [154, 42, 77]. The clinical gold standard for diagnosing GP, the gastric emptying test, typically involves ingestion of a meal containing a radioactive tracer and imaging at fixed time intervals to assess the duration of time required for the tracer to empty the stomach. However, gastric emptying shows inconsistent correlation with symptoms or symptom improvement [123, 74] and multiple medications have been shown to improve symptoms but not gastric emptying and vice versa [39, 97, 21]. Further, many patients without delayed gastric emptying but with foregut symptoms that include nausea, vomiting, abdominal pain/discomfort, early satiety and bloating,

termed gastroparesis-like syndromes [16], are indistinguishable symptom-wise from patients with GP. It has been proposed that FD and GP be sub-typed by disease etiology for targeted interventions [86, 73]. Such disease etiologies include, for example, include dysfunction in either the autonomic nervous system, gastric myoelectric activity, or pylorus [56] with tailored treatments including vagus nerve stimulation [58, 59], gastric electrical stimulation [20, 13], and botulinum toxin injection [85], respectively.

Gastric myoelectric activity involves interplay between smooth muscle cells and a traveling electrical wave, the gastric slow wave, that arises from oscillations in the membrane potential of interconnected gastric pacemaker cells, interstitial cells of Cajal (ICCs), that sequentially excite smooth muscle cells to propel food through the GI tract [46, 158, 161].

Normally functioning gastric slow waves in humans are initiated in a pacemaker region within the corpus of the stomach at a frequency of 0.05 Hz (see Fig. B.1a for anatomical landmarks), propagate rapidly in a circumferential direction to form equipotential rings [28, 66], and more slowly (e.g. 3-6 mm/s) along the organoaxial (corpus to pylorus) axis. In healthy individuals, organoaxial propagation occurs in primarily an anterograde fashion (e.g. towards the pylorus) [158, 80, 153].

Slow wave abnormalities underlying gastric myoelectric dysfunction, including abnormal initiation outside of the pacemaker region and retrograde propagation toward the fundus, have been found in patients with GP and FD undergoing surgery with electrode arrays placed directly on the stomach [110]. Since there are several etiologies that underlie upper GI disorders, it would be desirable to non-invasively detect if the underlying cause of symptoms is gastric myoelectric dysfunction, as this could propel scalable approaches to tailor treatment strategies.

Conventional electrogastrography (EGG) is a noninvasive technique for recording gastric myoelectric activity using electrodes placed cutaneously on the abdominal surface overlying the stomach [80]. Although it is easy to administer and interpret with spectral analyses, EGG has limitations that have prevented widespread uptake. While some studies have observed EGG

abnormalities in patients with upper gastrointestinal symptoms, [143, 119, 81] other studies have found no differences between dyspeptic patients and healthy volunteers [69, 107]. Further, it has been shown that spatial abnormalities can occur at normal frequencies and be can undetected by conventional EGG [110, 19, 23]. Its clinical usefulness has been questioned due to insufficient correlation with assessments made with more widely accepted (albeit invasive) clinical gold standards [3, 118]. The inability to capture spatial abnormalities, eliminate abdominal artifacts, and account for variability in stomach location further contribute to its limitations [3].

The recent development of the high-resolution electrogastrogram (HR-EGG), consisting of an array of cutaneous electrodes has enabled the development of modern signal processing tools that can detect spatial features of the slow wave. For example, we have shown that appropriate signal processing of HR-EGG waveforms detects the instantaneous slow wave speed and direction in healthy subjects [51]. In tandem with novel artifact rejection algorithms [52], our group has also shown that spatial features from HR-EGG correlate with symptom severity in a population of healthy controls and patients with GP and FD, spanning a wide range of body mass indices (BMI‘s) during 90-minute postprandial recordings [54].

We recently demonstrated the first study in the field to use HR-EEG waveforms to classify, with a 3D convolutional neural network machine learning approach, normal versus abnormal spatial slow wave patterns [7]. Since obtaining large labeled datasets with simultaneous cutaneous and intrasurgical stomach recordings to train a machine learning method is at the moment a big technological burden, we developed an in silico framework to generate training and testing datasets. Specifically, human computed tomography (CT) scans were used to develop stomach models reflecting normal and abnormal slow wave propagation, with anterograde propagation initiated in the corpus, and bifurcation of anterograde/retrograde propagation initiated in the antrum, respectively, as described from intrasurgical human stomach recordings [110]. Propagation to the cutaneous surface for simulation of EGG was simulated with a standard forward model that takes into consideration distance to the abdominal surface and tissue conductivity. We developed and

assessed performance with a three-dimensional convolutional neural network (3D CNN) approach [7], showcasing the first machine learning classification of normal vs. abnormal spatial patterns of gastric slow wave activity from non-invasive recordings, with highly accurate results that are robust to non-idealized electrode placement and signal-to-noise ratio (SNR). However, there was still an unmet need to build methods and assess classification performance when considering other forms of variability reflective of human recordings, including location of spatial slow wave abnormalities and recording start times relative to slow wave phase.

In this work, we use transfer learning from the 3D CNN in [7] to build a method that is robust to heterogeneity in both the location of the abnormal initiation on the stomach surface as well as the recording start times with respect to slow wave cycles. Specifically, using the simulation framework in [7] as a starting point, we now introduce variability in five different possible locations of the abnormal initiation. We also shuffle the 'start' time of every recording, so that the gastric slow wave was not at the same point in its cycle at the beginning of every recording.

Transfer learning within the context of deep learning is the process by which a neural network is trained for a particular classification task using 'source' data and then re-trained in some capacity with new 'target' data in order to perform a different classification task, the 'target' task [72, 11, 15, 40, 168, 145]. The theory behind this approach is that information learned in the source task is beneficial for model fitting in the target task [72, 11, 15], even when the source and target tasks vary significantly [145]. In addition, transfer learning has proved to enable stellar performance even with low availability of target data to train with [40, 72, 15]. The first layers in a sequential deep neural network, like the CNN developed in [7], select for general high-level features in the data, whereas the final layers select for patterns that are highly-specific to the training data. As such, the degree of similarity between the source and target classification tasks can impact which features of the source model are transferred to the target task [168].

In the work presented here, section 3.2.1 summarizes the simulation framework developed in [7]. Section 3.2.2 and 3.2.3 we describe how the simulations are adapted to allow for spatial and temporal heterogeneity, respectively. In Section 3.2.4, we discuss details of how we generate training and test datasets. Section 3.2.5 discusses how we harness transfer learning of the deep CNN weights learned in [7] in order to refine the CNN such that it has predictive capabilities on the new data that more closely resembles data acquired from human recordings. We find that by starting with training lowest-complexity models and building complexity in training sets, transfer learning one model to the next, the final network exhibits, on average, 80% classification accuracy in all but the most challenging spatial abnormality location, and below 5% type-I error probabilities across all locations. Altogether, this work showcases the potential of developing non-invasive assessment tools that are more robust to variability in human EEG recordings, and may propel scalable approaches to non-invasively screen for spatial slow wave abnormalities and improve the tailoring of etiology-specific treatment strategies.

3.2 Methods

3.2.1 The Simulation Workflow: Summary

In summary, the simulation designed in [7] generates EGG waveforms by i) simulating normal and abnormal gastric slow waves on stomach models constructed from human CT scans, and then ii) propagating electrical activity at the stomach's surface (solved for in (i)) to virtual electrodes on a CT-constructed abdominal surface via a forward model. The stomach potentials in (i) are solved for by (a) computationally determining a unique parameterized curve that describes the organoaxis of the stomach, (b) solving the one dimensional wave equation along the organoaxis found in (a), and (c) propagating voltage potentials from (b) to the gastric serosal surface in equipotential rings. Our forward model in (ii) is built to compute the voltage at each discrete point on the abdominal surface. Each abdominal voltage is calculated from a superposition of

'source' current dipoles, arising from the voltages at points on the stomach and weighted by the distance and angle between the source and abdominal coordinates, as well as the conductivity of abdominal tissue.

Virtual electrodes are placed on the abdominal surface and 'measure' the electrical potentials propagated to the abdominal surface. We added white Gaussian noise to these simulated recordings, to account for noise from other electrophysiological systems in the body as well as measurement noise. A full description of the paradigm by which noise was added, including noise variances, can be found in [7]. Electrode arrays (containing 100 channels) were placed in various locations on the abdomen, each array with a different degree of centeredness with respect to the stomach. We modulated the anterior spacing between the stomach and the abdomen as to mimic trends in body mass index (BMI). A fully detailed explanation of the simulation can be found in [7].

3.2.2 Modulating the Location of the Abnormal Initiation

The simulation summarized in the previous section models normal gastric slow wave physiology, as well as a spatial slow wave abnormality in which the slow wave is initiated near the antrum of the stomach and bifurcates into an anterograde propagating wave below the initiation towards the pylorus, and a retrograde propagation above the initiation towards the fundus.

However, the location of spatial abnormalities need not be specific to the antrum; mucosal biopsies have found depletion of ICCs in other parts of the stomach [113]. Since ICCs are responsible for coordinating the cyclical pacing and propagation, their depletion causes these mechanisms to break down, thus giving rise to spatial slow wave abnormalities [140].

In this work, we simulate the abnormal initiation in five separate locations (denoted 'Position 1-5') between the antrum and the normal pacemaker region in the corpus (Fig. B.1b). Position 1 most closely resembles the location of the initiation modeled in [7]. In each of the five cases, gastric slow wave propagation bifurcates into an anterograde (below) and retrograde

(above) fashion. The remainder of the simulation, propagation of serosal voltages to virtual electrodes, remains identical to the methods summarized in 3.2.1 and described in detail in [7].

3.2.3 Time Start Shuffle

In our previous work, the simulated EGG waveforms acquired begin during the same point of the slow wave in all training and test datasets. However, during recording, one may not necessarily begin recording the slow wave at any particular phase. As such, in order to introduce variability in the phase of the beginning of the recordings we use as training and test data, we shuffled the start time of the recordings as follows. For each simulated recording, consisting of 300 time samples, a new start time, T^*, is chosen at random, equally likely between 0 (the original start of the simulation) and 300 (the end of the simulation). Then, the new dataset is constructed from [T^*:300] and [0: T^*], looping back to $time = 0$ after the 300^{th} time sample. A schematic is shown in Figure B.2.

Since our recordings have two dimensions of space and one dimension of time, the filters for the CNN used in this work are 3-dimensional. If the temporal component of the simulation waveforms holds too much of a pattern that is only seen in simulation, then CNN filters may overfit in the temporal dimension - thus limiting its efficacy when deployed on human physiologic recordings. Therefore, creating new datasets in this way and re-training the CNN will allow the weights to be updated such that the CNN is more robust in detecting useful patterns in slow wave propagation with less overfitting to 'man-made' temporal features of the simulation.

3.2.4 Training and Testing Datasets

The methods outlined in sections 3.2.1 through 3.2.3 explain how we generated individual simulated EGG recordings. Groups of simulated EGG recordings are binned together into larger training and testing datasets. As stated in section 3.2.1 and explained in detail in [7], several

simulated recordings are generated from one CT-derived stomach model. For instance, the gastric slow wave is simulated on 'Stomach 1', then multiple electrode recordings are generated: one recording in which electrodes are centered on the abdomen, one in which the electrode array is shifted to the far right of the abdomen, one with high simulated BMI, etc. Of the 40 total stomach models used to generate simulation data, 28 stomach models were used for the generation of training data and 12 stomach models were used to generate simulation data for testing (see schematic in Fig. B.3). As such, training data always remained independent from test data.

A comprehensive list of the training/testing datasets and their respective properties can be found in Table 3.1. The original CNN trained in [7] was trained and tested on Dataset 0, the least complex dataset. In this work, we sequentially introduce more heterogeneity into the training and test sets. Datasets 1 and 2 introduce either the five abnormal initiation locations or the time start shuffle, respectively. Dataset 3 combines the five abnormal initiation locations with the time start shuffle, comprising the most heterogeneity and complexity.

We created Datasets 1 and 2 so we could a) train and test the two modifications to the data (five abnormal initiation locations and varied recording start time) independently, 'turning one knob' at a time, and b) iterate the process of transfer learning, adding complexity in a stepwise fashion. Since another component to studies with human physiologic data is that there may be a dearth of labeled training data, and since one significant benefit to transfer learning is robustness to small training datasets [40, 72, 15], Datasets 1 and 3 are smaller by design [1], as to both encompass another aspect of human data and also showcase the robustness of our transfer learning workflow with regard to limits in available training data.

All training datasets are class-balanced. For each recording in the training datasets that have five initiation locations (Datasets 1 and 3), the initiation location (if in the abnormal class) was chosen at random from the five locations, and thus each abnormal initiation location was

[1] In training Datasets 1 and 3, the same 28 stomach models were used to generate data as in the training Datasets 0 and 2. The reduction in data size was due to less sampling of recording parameters, such as electrode array positioning or BMI.

represented 20% of the time in these training datasets. In all of the training datasets, the simulated waveforms were generated from the same 28 'training' stomach models. The 12 stomach models used to generate test data were never seen in training at any stage of the entire work (including the original model trained in [7]). The ultimate goal of the work is to achieve high classification performance on Dataset 3, which most closely resembles what would occur in recordings on human subjects.

3.2.5 Transfer Learning Paradigm

The overall paradigm of transfer learning involves: i) training a neural network with 'source' data, ii) using the weights from the model developed in (i) as a starting point, and retraining the model with 'target' data. Depending on the complexity of the target task, the amount of available target training data, and the similarity between source and target data, retraining the neural network with target data could involve, for example, updating weights from all of the layers, or only updating weights from the final layers, which is referred to as 'fine-tuning' [168, 40, 15]. In this work, we use both approaches, depending on the target task.

Model fitting within the context of neural networks involves the optimization of non-convex objective functions [79]. These functions have several local minima that are distinct

Table 3.1: Dataset specifications, in order of increasing complexity. *Dataset used for training/ testing in [7]. **Testing datasets have 108 recordings per abnormal initiation position

Dataset Name	Recording Start Time Heterogeneity (y/n)	Abnormal Initiation Heterogeneity (y/n)	Num. Recordings in Training (per class)	Num. Recordings in Testing (per class)
Dataset 0*	No	No	3220	108
Dataset 1	No	Yes	504	108 X 5**
Dataset 2	Yes	No	3220	108
Dataset 3	Yes	Yes	504	108 X 5**

from the global minimum, which means that the initial starting point can impact whether or not the global minimum is reached. This is counter to optimization of convex objective functions, where local minima are global minima and so careful attention is not needed for initialization of model-fitting. In our transfer learning scheme, even in the case where weights from all of the layers in the neural network are retrained, initializing model-fitting with weights trained from a previous model can be beneficial.

The deep CNN architecture developed in [7] consists of, in sequence: four convolutional layers, one fully connected layer and lastly, a fully connected layer with two filters, which is the classifying layer. A full summary of the architecture, including max pooling, number of filters per layer, kernel size, etc. can be found at [7]. This model, trained with Dataset 0, is referred to as Model A in our work here. We used the transfer learning paradigm as follows. A schematic of the models created via transfer learning is shown in Fig. B.4.

We first developed two models (Model B, Model C) which each introduced data heterogeneity separately. We then introduced heterogeneity sequentially, training lowest-complexity models and then building complexity, transfer learning one model to the next (Model D, Model E) . To create Model B, we transfer learned weights from Model A and fine-tuned the CNN by updating weights from only the final non-classifying layer onward, with training data from Dataset 1. To create Models C, D and E, we transfer learned weights from Models A, C, and D, respectively. We then updated weights from all of the layers, using transfer learned weights as initial conditions, with training data from Datasets 2, 1, and 3, respectively.

Because training Datasets 1 and 3 were smaller, we accounted for variability in weight optimization (arising from the non-convex landscape) by repeating the transfer learning and retraining/ fine-tuning framework five independent times for those training datasets, with each of the five iterations having no memory of a previous iteration.

3.3 Results

Fig. B.5a displays results of testing Model B with Dataset 1 (heterogeneity in the abnormal initiation location, but not in recording start time). We achieved high performance at all five abnormal initiation locations (Fig. B.5a, black) with very little computational load during training (as only fine-tuning of weights was performed). This is shown in comparison to Model A (red), using the same test data in order to show the accuracy 'gain' of our transfer learning and fine-tuning framework. The decline in performance of Model A at abnormal initiation positions outside of Position 1 is to be expected, as Model A was only trained on data in which the abnormal class had an abnormal initiation at Position 1.

Fig. B.5b compares performance of Model C with Model A. Since Dataset 2 only had heterogeneity in recording start times, its test accuracy is shown as a single mark at Position 1. We achieved an accuracy improvement of 32% between Model A (magenta diamond) and Model C (black circle) when testing Dataset 2. Furthermore, we found that Model C showed robustness to heterogeneity in abnormal initiation position (Fig. B.5b, blue) even though Model C had *never* seen any examples of an abnormal initiation outside of Position 1 in training. By comparison, Model A, although able to achieve stellar performance at Position 1 (red), lacks the ability to classify above the probability of chance at all other positions.

Fig. B.5c displays results of our most sophisticated model, Model E, tested on data most akin to human EGG recordings–with heterogeneity in both abnormal initiation position and recording start time (Dataset 3). In all but Position 5, Model E was able to achieve a classification accuracy above 75%. The average classification accuracy at Positions 1-4 is 80%. The variability in classification accuracy among independent iterations of model fitting is highest in Model E, which we expand upon in section 3.4.

Fig. B.5d shows the Type-I (false positive) vs. Type-II (false negative) error probability from testing Model E with Dataset 3 (heterogeneity both spatially and temporally). A false

negative classification is one in which the model incorrectly classifies an abnormal waveform as 'normal.' The disparity between Type-I and Type-II errors increases between Positions 1 and 5 and the incidence of Type-I errors never exceeds 5% across all five abnormal initiation positions.

3.4 Discussion

The primary goal of this work was to develop a machine learning algorithm to accurately classify normal from abnormal simulated EGG waveforms with heterogeneity in the recording start time relative to the slow wave cycle and abnormal initiation position, as is seen in data from human EGG recordings (referred to as 'Dataset 3' throughout this work). Overall, we were successful in this aim, with a mean classification accuracy above 75% for 4 of the 5 abnormal initiation locations.

The novelty of the work presented here lies both within the success of showing the ability to classify simulated waveforms more akin to human recordings and the transfer learning framework we developed to do so. As explained in Section 3.2.5, model fitting in this study involves optimization of non-convex objective functions. We have shown here that transfer learning weights from a model trained on a less complex classification scheme allows for significant improvement in classification performance on challenging tasks. To verify this, we trained the CNN architecture from scratch, in which all weights were initialized with random values, with the most complex dataset, Dataset 3. For the sake of brevity, we did not plot these results, but in summary, the cross validation accuracy never exceeded the probability of chance (0.5) for each of 50 training epochs. Model weights could not successfully converge, even when the learning rate was modulated. We hypothesize that this was a result of both the complexity of the classification task (arising from dataset heterogeneity) and the small size of the training dataset. This demonstrates the importance of the methods designed here in this work and also gives credence to the hypothesis that our most sophisticated model, Model E, can itself be fine-tuned

with data from human HR-EGG recordings, even if only small amounts of labeled training data are available.

The lowest classification accuracy, seen at Position 5, characterizes abnormal slow waves initiated within the range of the pacemaker region. Slow waves did propagate slightly toward the fundus, but most of the 'abnormal' slow wave propagation in this case was indistinguishable from normal propagation. Since we created all of the waveforms via simulation, this was intentional as to test the bounds of our model. This represents the 'worst case' scenario for the classifier.

Most of the variability seen in Fig. B.5c, when testing Model E, actually arises from one outlier model. The mean standard deviation across all five initiation positions is 6.5%, and excluding the one outlier, the mean standard deviation across all five positions drops to 3.7%. We hypothesize that when training models with smaller training datasets (Datasets 1 and 3), updating fewer weights in training led to higher model stability. For example, as seen in Fig. B.5a, Model B–which was fine-tuned as opposed to fully retrained– underwent independent training and testing five times with almost no variability (mean standard deviation $<1\%$). We further confirmed this hypothesis by repeating the training and testing of Model B, except with a full retraining of all of the weights (as opposed to fine-tuning). Under these parameters, the mean classification accuracy was approximately the same as reported in Fig. B.5a, but the mean standard deviation more than doubled. Therefore, if the target task is simple enough such that fine-tuning only some neural network weights is sufficient, then the method of fine-tuning may be superior to updating all of the weights, as it appears to improve model stability.

Though the overall trend of Fig. B.5c shows the highest classification accuracy in classifying 'Position 1' and the lowest accuracy in classifying 'Position 5,' the trend was not monotonic from positions 1-5. A subject of future work would be to determine if these deviations from monotonicity are indicative of a reproducible pattern or merely an artifact of stomach size and shape variability.

Although this work showed the ability to successfully harness information from simpler

classification results and achieve classification in a significantly more challenging task, future studies will entail optimizing neural network architecture and hyperparameters as to increase the overall classification accuracy. We have shown that increasing the complexity of the classification task results in a decline in peak performance. For example, we were not able to classify our most heterogeneous dataset (Dataset 3) with as high of accuracy as that of Model A tested at Position 1 (with no heterogeneity in training/testing). Instead, we observed a 'cap' in performance around 90% when testing Dataset 3. Future work to boost the accuracy could involve modifying the size of the CNN filters, particularly in the time axis. This would entail the re-training of the entire framework, including Model A. Nonetheless, the work presented here gives credibility to the claim that slow wave patterns can be disambiguated via a machine learning framework.

Finally, the low incidence of Type-I errors seen in classifying Dataset 3 with Model E is significant. Type-I errors, or 'false positive' results in this context refer to cases in which the model classified a waveform as abnormal but in reality it was normal. For the ultimate purpose of this work, which is to effectively screen for abnormalities in gastric slow wave initiation location and propagation, it is beneficial to have a low probability of of Type-I errors even at the expense of Type-II errors. Within the context of differential diagnosis, treatments for gastric slow wave abnormalities, such as gastric electrical stimulation, currently require invasive procedures. In contrast, treatment options for GP or FD arising from a different etiology, such as cutaneous vagus nerve stimulation [58, 59] or botulinum toxin injection [85] for treating autonomic or pylorus dysfucntion, respectively, are significantly less invasive. As such, the low Type-I error probabilities of our approach indicate that classification outputs of our method will not likely lead to unnecessary invasive procedures.

3.5 Conclusion

To summarize, this study builds off of a novel result [7] which shows *in silico* that normal vs. abnormal gastric slow wave propagation patterns can be disambiguated via a 3D-CNN framework. The work presented here introduces both spatial and temporal heterogeneity into the *in silico* EGG waveforms, making them more akin to human EGG recordings, and harnesses transfer learning of CNN weights from [7] to achieve robust classification in the presence of this heterogeneity. The results of this study suggest that transfer learning of weights from *in silico* model fitting may show promise in deploying this framework in human multi-electrode recordings.

Chapter 3, in full, is a reprint of the material as it appears in the proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2020. Agrusa, Anjulie; Allegra, Alexis; Kunkel, David; Coleman, Todd, IEEE, 2020. The dissertation author was the primary investigator and author of these materials.

Chapter 4

Unsupervised Anomaly Detection with Wasserstein Barycenters

4.1 Introduction

Detection of anomalies has powered studies across a diverse range of fields, such as computer security, defense, and medicine [10, 33, 124, 150]. Traditional anomaly detection methods operate under the paradigm of first defining 'normal' behavior and subsequently identifying anomalies that deviate from normal patterns, in both supervised and unsupervised manners [162, 45]. Classical studies in this field have defined 'normal' behavior by assuming that data can be generated by some probability distribution and performing numerical techniques, such as kernel density estimation or maximum likelihood estimation, to estimate this underlying distribution, using either i) only data labeled as 'normal' in supervised methods [24] or ii) large datasets (with a low incidence of anomalies compared to normal samples) in unsupervised methods [45, 44]. Once the distribution is estimated, there are well established methods to quantify anomalies based on their proximity to the distributional mean, subject to distributional variance [10, 45]. Estimation of these underlying probability distributions, however, requires some *a priori*

knowledge concerning the class and modality of the generative 'non-anomalous' distributions.

Robust regression is a technique that has been successful in fitting highly-accurate regressions to datasets containing anomalies [32, 35, 8]. These methods rely on the assumption that features correlate with labels (typically linearly), as well as the assumption that residual errors between 'predicted' and true labels follow a multi-modal distribution, with at least one mode typically near zero. Whereas some robust regressions assume an anomaly-containing mode of the residual distribution is also near zero but has higher variance [32], others assume an anomaly-containing mode is centered at a specific location [8]. This yields an inclusion criterion for data to be used in the regression, and all other data are labeled as anomalous [32, 137, 18, 35, 8]. As is the case with distribution-based approximations, these robust regressions rely on assumptions about the underlying structure of the problem.

Another well-studied type of anomaly detection mechanism involves support vector machines (SVMs) [10, 88]. In these studies, an entire anomaly class is identified, in which there is a support vector, or hyperplane, that separates features in the normal and anomaly classes. These methods require labeled training datasets to establish a separation between normal and anomalous features and make predictions on new data. 'Robust' SVMs have been additionally developed to mitigate challenges that arise from poorly labeled training data [71]. Even so, these methods are not employable in cases in which training labels are not feasible to obtain in any capacity.

Since autonomous processes are often desired, we will restrict the remainder of our discussion to unsupervised clustering algorithms, as these methods do not required labeled datasets. In clustering methods, such as the K-Means algorithm [64], data are aggregated into groups such that inter-class variability is high and intra-class variability is low. One of the vulnerabilities of traditional K-Means type clustering is that the number of clusters, K, must be specified. Modern autonomous algorithms have made progress in optimizing the parameter, K, in addition to cluster centriod locations [43, 126, 63]. However, additional challenges arise, such

as proper initialization of K, and formulating upper and lower bounds for K. Possibly one of the most significant pitfalls of K-Means in the context of anomaly detection is that it performs grouping over the entire dataset, and thus does not directly attempt to identify anomalies [64, 43]. For instance, it would be more 'profitable' in a loss context to place another cluster centroid at the location of an anomalous data point, with the downstream effect being that this anomaly would not be far from a cluster center.

An improvement upon this is the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, which, by design, is robust to anomalies [61]. The defining methodology within DBSCAN involves the algorithm identifying 'core' points and forming aggregates of data within a certain proximity to these core points, rather than a single centriod which is the case in the K-Means algorithm. Anomalies are defined as data that lie outside of a specified range of all data within a proximity to core points. This algorithm is highly sensitive to input parameters that determine the minimum number of points to be considered a cluster and the maximum allowable distance between neighboring points in a cluster. As in the case of autonomous K-Means clustering, strategies have been proposed to estimate these parameters using only the input dataset [78].

These aforementioned methods have been widely studied with datasets and features in $\mathcal{R}^n$. However, fewer studies have been published which classify and detect anomalies within a set of probability distributions. The standard clustering and anomaly detection frameworks rely on a metric between elements of the dataset and clustering variables (i.e. cluster centroids). Optimal transport theory is a state-of-the-art framework to establish a metric between probability distributions [159] and has powered state-of-the-art studies in broad applications of mathematics, data science, and economics [159, 141, 127, 48]. The 'optimal transport' map is that which minimizes the cost of transporting all of the mass in one distribution to another. The cost associated with this distance-minimizing transportation map is referred to as the Wasserstein (squared) distance. Since clustering operates on a metric between samples, within the context of

clustering probability distributions it is natural to use the Wasserstein metric.

There have been multiple studies in the field that have followed a K-Means clustering framework with the Wasserstein metric between probability distributions [99, 100, 166]. Instead of cluster centriods, as is standard in K-Means clustering, these methods form clusters around Wasserstein barycenters. These methods still require proper initialization of the number of barycenters, K, but aforementioned approximation methods could be used to estimate this. However, the main drawback of the K-Means algorithm as it pertains to anomaly detection still applies: anomalies are not directly accounted for.

There is an unmet need to design an anomaly detection technique that i) operates on a set of probability distributions as an input, ii) directly targets the identification of anomalies, and iii) is unsupervised. One of the well-established principles in the field of anomaly detection is that specific algorithmic needs can vary significantly on a study-by-study basis [10, 68]. As such, instead of proposing a specific algorithm, in this work we propose a philosophical framework for detecting anomalies in a set of probability distributions. Namely, we examine a framework in which we partition subsets of a set of distributions, compute a Wasserstein barycenter between distributions in each subset, and finally compute Wasserstein distances between each distribution in a subset and the subset barycenter. We hypothesize that anomaly distributions will be consistently 'far' from corresponding subset barycenters. The main contribution to this work is a rigorous mathematical proof to show that the largest possible Wasserstein distance between a probability distribution and a subset barycenter, given *theoretical* set parameters, is the distance between an anomaly distribution and the barycenter of a subset containing the anomaly. We propose that the mathematical proof will allow future studies to design algorithms in a tailored way, using this philosophical framework. And finally we demonstrate an example of an algorithm derived from this framework along with an example in medicine, specifically the field of Gastroenterology.

4.2 Definitions

Define the Wasserstein ball, B, where:

$$B(\Upsilon,\alpha) \triangleq \left\{P \in \mathcal{P}(\mathsf{X}) \;\; s.t. \;\; W_2(\Upsilon,P)^2 \leq \alpha^2\right\} \tag{4.1}$$

where $\alpha \in \mathcal{R}$, $\mathcal{P}(\mathsf{X})$ is the set of all probability distributions on support X (i.e. the set of real numbers) with finite second moment. This ball has a center, Υ, which we assume all non-outlier samples are close to, within W_2 distance of α. The squared Wasserstein distance, or Kantorovich cost, between probability distributions R and Q is given by:

$$W_2(R,Q)^2 = \min_{\pi \in \Pi} \int_{u \in \beta_R} \int_{v \in \beta_Q} d(u,v)^2 d\pi(u,v) \tag{4.2}$$

where Π is the set of all joint probability distributions with marginals R and Q, β_R is the support of R and β_Q is the support of Q. The distance $d(u,v)^2$ is given by $= ||u-v||^2$ if X is the set of all real numbers, but can be generalized to distances along other manifolds (i.e. geodeisc distance).

Let the set Ψ consist of $N+1$ probability distributions, in which N distributions, $\{P_1, P_2, ...P_N\}$, lie within $B(\Upsilon,\alpha)$. Let the $N+1^{th}$ probability distribution of Ψ be an anomaly, denoted by P_O, where P_O deviates from Υ by:

$$W_2(\Upsilon,P_O) = \alpha + \delta \tag{4.3}$$

for some $\delta > 0$. Define $\mathcal{J}$ as the set $\{1,\ldots,J\}$ of indices of subsets, and define $\mathcal{K}$ as the set $\{1,\ldots,K\}$ of indices of distributions in subset j. Let $\mathcal{S}^{(j)}$ be the jth subset and $\mathcal{S}_k^{(j)}$ be the kth index of the jth subset. Define the j^{th} subset barycenter as follows:

$$\chi_j = \arg\min_{P \in \mathcal{P}(\mathsf{X})} \frac{1}{K} \sum_{k=1}^{K} W_2(P, \mathcal{S}_k^{(j)})^2 \tag{4.4}$$

where $2 < K < N+1$. Below we state a theorem that gives justification to an anomaly detection algorithm we later describe. It mathematically justifies the idea of taking subsets of all the distributions and comparing each element to the subset's barycenter. The idea is that if subsets contain the same element whose distance to the barycenter is larger than others, then this is suggestive that such an element is an anomaly.

Theorem 1. *The supremum of distances between elements of a subset and a subset barycenter, over all subsets, is the maximal distance between the anomay and any subset barycenter containing the anomaly.*

$$\sup_{j\in\mathcal{J},k\in\mathcal{K}} W_2(\chi_j, \mathcal{S}_k^{(j)})^2 = W_2(\chi_{j^*}, P_O)^2 \tag{4.5}$$

where j^* is the index of the subset, containing P_O, for which the subset barycenter is maximally far from P_O. We will build up to proving Theorem 1 after demonstrating prerequisite proofs.

4.3 Maximizing Distances Along a Geodesic

A geodesic is defined as a length minimizing path, in this case, between two probability distributions. We define the unit speed geodesic between two distributions, for which

$$W_2(P_s, P_t) = |s-t| W_2(R, Q) \tag{4.6}$$

for $0 \leq s,t \leq 1$ and $R, Q \in \mathcal{P}(\mathsf{X})$. For some λ, where $0 < \lambda < 1$ define:

$$W_2(R, P_\lambda) = \lambda W_2(R, Q) \tag{4.7}$$

and similarly

$$W_2(P_\lambda, Q) = (1-\lambda) W_2(R, Q) \tag{4.8}$$

Therefore:

$$W_2(P,Q) = W_2(R,P_\lambda) + W_2(P_\lambda,Q) \tag{4.9}$$

For any locally compact geodesic space X, the subsequent Wasserstein space of probability measures $\mathcal{P}(\mathsf{X})$ inherits the geodesic space property [93, Lemma 2.4, Prop 2.6]. Thus for any pair of probability distributions $R, Q \in \mathcal{P}(\mathsf{X})$, a Wasserstein geodesic exists between them.

For any distribution lying on a geodesic between R and Q, the distance properties between the three correspond to equality in the triangle inequality. We define a pair of distributions, $(P_{-\alpha}, P_\alpha)$ that lie on the boundary of B, denoted $\partial B(\Upsilon,\alpha)$, for which Υ lies on the geodesic connecting $P_{-\alpha}$ and P_α.

Lemma 1. *If Υ lies on the geodesic that connects $P_{-\alpha}$ and P_α then:*

$$\begin{cases} W_2(P_{-\alpha},P_\alpha)^2 > W_2(P_{-\alpha},\tilde{P})^2 \\ W_2(P_{-\alpha},P_\alpha)^2 > W_2(P_\alpha,\tilde{P})^2 \end{cases} \tag{4.10}$$

where $\tilde{P} \in B(\Upsilon,\alpha)$ and Υ neither lies on the geodesic connecting $P_{-\alpha}$ and $\tilde{P}$ nor the geodesic connecting P_α and $\tilde{P}$.

Proof. The Wasserstein metric obeys the triangle inequality, and as such

$$W_2(P_{-\alpha},\tilde{P}) \leq W_2(P_{-\alpha},\Upsilon) + W_2(\Upsilon,\tilde{P}) \tag{4.11}$$

Since Υ does not lie on the geodesic connecting $P_{-\alpha}$ and $\tilde{P}$, the inequality in (4.11) is strict. Further, by properties of ball $B(\Upsilon,\alpha)$, it follows that $W_2(\Upsilon,\tilde{P}) \leq \alpha$, and so

$$W_2(P_{-\alpha},\tilde{P}) < 2\alpha. \tag{4.12}$$

Since Υ lies on the geodesic connecting $P_{-\alpha}$ and P_{α}, and since $P_{-\alpha}, P_{\alpha} \in \partial B(\Upsilon, \alpha)$,

$$\begin{aligned} W_2(P_{-\alpha}, P_{\alpha}) &= W_2(P_{-\alpha}, \Upsilon) + W_2(\Upsilon, P_{\alpha}) \\ &= 2\alpha \end{aligned} \tag{4.13}$$

Combining (4.12) and (4.13) yields:

$$W_2(P_{-\alpha}, \tilde{P}) < W_2(P_{-\alpha}, P_{\alpha}). \tag{4.14}$$

The second inequality in (4.10) follows similarly. □

Since $\partial B(\Upsilon, \alpha)$ is the boundary of a convex set with nonzero diameter, there are infinite possible '$P_{-\alpha}$'s. Further, from the symmetry of the ball, there are infinite possible $(P_{-\alpha}, P_{\alpha})$ pairs. Define P_{α^*} as follows:

$$P_{\alpha^*} \triangleq \arg \min_{P_{\alpha} \in \partial B(\Upsilon, \alpha)} W_2(P_{\alpha}, P_O)^2. \tag{4.15}$$

Similarly, $P_{-\alpha^*}$ is defined for the boundary distribution that satisfies (4.10) for $P_{\alpha} = P_{\alpha^*}$. The outlier distribution lies on a path that is an extension to the geodesic between $P_{-\alpha^*}$ and P_{α^*}, and as such $P_{-\alpha^*}, \Upsilon, P_{\alpha^*}$ and P_O all lie on a geodesic.

Lemma 2. *The maximum possible Wasserstein squared distance between any two distributions in* Ψ *is* $W_2(P_{-\alpha^*}, P_O)^2$.

Proof. This is an extension of Lemma 1. Let distributions R and Q lie in $B(\Upsilon, \alpha)$, including the boundary. The triangle inequality gives:

$$W_2(R, Q) \leq W_2(R, \Upsilon) + W_2(\Upsilon, Q) \tag{4.16}$$

By virtue of (4.1), $W_2(R, Q) \leq 2\alpha$. (This inequality holds with equality if $(R, Q) = (P_{-\alpha^*}, P_{\alpha^*})$). Therefore, the maximum distance between any two distributions in B is equal to 2α. Since

$P_{-\alpha^*}, P_{\alpha^*}$ and P_O all lie on a geodesic:

$$\begin{aligned} W_2(P_{-\alpha^*}, P_O) &= W_2(P_{-\alpha^*}, P_{\alpha^*}) + W_2(P_{\alpha^*}, P_O) \\ &= 2\alpha + W_2(P_{\alpha^*}, P_O) \end{aligned} \tag{4.17}$$

In addition, since the Wasserstein distance is a metric space, $W_2(P_{\alpha^*}, P_O) = 0$ if and only if $P_{\alpha^*} = P_O$, which is false by definition of P_O. Therefore:

$$W_2(P_{-\alpha^*}, P_O) > W_2(R, Q) \quad \forall R, Q \in B(\Upsilon, \alpha). \tag{4.18}$$

□

4.4 Geodesic Convex Hull Bounds the Barycenter

Define the geodesic convex hull of the j^{th} subset, $H^{(j)}$, to be the smallest set that contains: i) all $M, N \in \mathcal{S}^{(j)}$ and ii) all unit speed geodesics with endpoints $M, N \in \mathcal{S}^{(j)}$.

Lemma 3. *The barycenter χ_j lies within the geodesic convex hull of the j^{th} subset.*

Proof. We will prove this by contradiction. Let the barycenter χ_j, lie outside of $H^{(j)}$. Denote distribution $\hat{G}$ to be the distribution in $H^{(j)}$ (including the boundary) that is closest to χ_j:

$$\hat{G} = \arg\min_{G \in H^{(j)}} W_2(\chi_j, G)^2. \tag{4.19}$$

Current literature has established precedent for describing an angle between two geodesics operating over the Wasserstein metric [55, 155]. These methods use a law of cosines-based approach to compute these angles. The angle between i) the unit speed geodesic connecting $G \in H^{(j)}$ and $\hat{G}$ and ii) the unit speed geodesic connecting $\hat{G}$ and χ_j is denoted as $\angle G\hat{G}\chi_j$ and

given by:

$$\begin{aligned}\angle G\hat{G}\chi_j &= \arccos \frac{W_2(G,\hat{G})^2 + W_2(\hat{G},\chi_j)^2 - W_2(\chi_j,G)^2}{2W_2(G,\hat{G})W_2(\hat{G},\chi_j)} \\ &\triangleq \theta.\end{aligned} \tag{4.20}$$

We will give necessary and sufficient conditions to prove that if (4.19) holds, then $\cos(\theta) \leq 0$, which follows the same logic from a proof of the convex projection theorem in [115]. We will show the necessary condition by contradiction. Let $\cos(\theta) > 0$. Define distribution $G_\lambda \in H^{(j)}$ which lies on the geodesic between G and $\hat{G}$ and thus:

$$\begin{aligned}W_2(G,G_\lambda) &= \lambda W_2(G,\hat{G}) \\ W_2(G_\lambda,\hat{G}) &= (1-\lambda)W_2(G,\hat{G})\end{aligned} \tag{4.21}$$

for $0 \leq \lambda \leq 1$. We define an expression for $W_2(G_\lambda,\chi_j)^2$ by virtue of the law of cosines between the two geodesics: i) between G_λ and $\hat{G}$ and ii) between $\hat{G}$ and χ_j. The angle between these two geodesics is equivalent to θ since G_λ lies on the geodesic between G and $\hat{G}$.

$$W_2(G_\lambda,\chi_j)^2 = l^2 + [(1-\lambda)W_2(G,\hat{G})]^2 - 2l[(1-\lambda)W_2(G,\hat{G})]\cos\theta \tag{4.22}$$

where $l = W_2(\hat{G},\chi_j)$. Since $W_2(G_\lambda,\chi_j)^2$ is a function of λ, we take the partial derivative with respect to λ, evaluated at $\lambda = 1$:

$$\frac{\partial W_2(G_\lambda,\chi_j)^2}{\partial\lambda}|_{\lambda=1} = 2lW_2(G,\hat{G})\cos\theta \tag{4.23}$$

This expression is greater than zero because all Wasserstein distances are non-negative and we

have assumed $\cos(\theta) > 0$. There exists a limit:

$$\lim_{\lambda \to 1^-} \frac{W_2(G_\lambda, \chi_j)^2|_{\lambda=1} - W_2(G_\lambda, \chi_j)^2|_\lambda}{1-\lambda} > 0 \tag{4.24}$$

Combining (4.24) with (4.21) yields:

$$W_2(\hat{G}, \chi_j)^2 > W_2(G_\lambda, \chi_j)^2 \tag{4.25}$$

Since $G_\lambda \in H^{(j)}$, this is a violation of (4.19) and thus a contradiction arises. To prove the sufficient condition, let $\cos\theta \leq 0$. Using this assumption and rearranging (4.20) gives:

$$W_2(\chi_j, G)^2 > W_2(G, \hat{G})^2 + W_2(\hat{G}, \chi_j)^2 \tag{4.26}$$

and subsequently,

$$W_2(\chi_j, G)^2 > W_2(\hat{G}, \chi_j)^2 \tag{4.27}$$

Therefore we have shown necessary and sufficient conditions to prove that $\cos\theta \leq 0$. Another consequence of this, by equations (4.20) and (4.26) is that:

$$W_2(\chi_j, G)^2 > W_2(G, \hat{G})^2 \tag{4.28}$$

This is a contradiction of (4.4), as χ_j would not be the minimizer of the objective function, and therefore the barycenter must lie within the geodesic convex hull, $H^{(j)}$. □

4.5 Extremal Distances Between Barycenter and Boundaries of the Geodesic Convex Hull

It has been shown in Lemma 3 that the barycenter cannot exceed the unit speed geodesic convex hull of points in set $\mathcal{S}$. Therefore, no barycenter can exceed the bounds of distributions in Ψ. We have shown in Lemma 2 that the maximum distance between any two distributions in Ψ is the distance between P_O and $P_{-\alpha*}$. As such, since Theorem 1 maximizes distances to a barycenter, we aim to construct a set for which the barycenter of the set has minimal deviation from $P_{-\alpha*}$.

In order to prove our main result, we first consider an auxiliary lemma which we use to show the above result in this section:

Lemma 4. *In a scalarized optimization problem, if one coefficient increases at the ith index, the vector-valued objective at the ith index decreases.*

Proof. Define the vector-valued objective:

$$f_o(X) = [F_0(X), F_1(X), F_2(X)]^T \tag{4.29}$$

where $f_o(X) \in \mathcal{R}^3$. We analogously define the scalarized optimization problem:

$$\min_X \lambda^T f_o(X) \tag{4.30}$$

where $\lambda \in \mathcal{R}^3$, $\lambda \succ 0$, and λ_0 respresents the first entry of λ. The minimizer is given by

$$X^* = \arg\min_X \lambda^T f_o(X) \tag{4.31}$$

By principles of Pareto optimality, specifically trade-off analysis [26], if we define λ^* such that

$\lambda_0^* > \lambda_0$, and we define

$$\hat{X} = \arg\min_X \lambda^{*T} f_o(X), \tag{4.32}$$

then $F_0(\hat{X}) \leq F_0(X^*)$. □

With this, we can now state the a key result:

Lemma 5. *For any one subset $\mathcal{S}^{(j)}$ that contains P_O, the Wasserstein squared distance between a subset barycenter and $P_{-\alpha^*}$, given by*

$$W_2(P_{-\alpha^*}, \chi_j)^2,$$

is minimal when the remaining $K-1$ entries in $\mathcal{S}^{(j)}$ are $P_{-\alpha^*}$.

Proof. Without loss of generality, let the number of samples (K) in $\mathcal{S}^{(j)}$ be equal to 3. We write the barycenter optimization problem as a scalarized optimization problem with a vector-valued objective, namely:

$$\chi_j = \arg\min_{P \in \mathcal{P}(\mathsf{X})} \lambda^T f_o(P) \tag{4.33}$$

where

$$f_o(P) = [W_2(P, P_{-\alpha^*})^2, W_2(P, P_x)^2, W_2(P, P_O)^2]^T \tag{4.34}$$

and $P_x \in \mathcal{B}, P_x \neq \{P_{-\alpha^*}, P_O\}$. There is an inherent constraint on λ, such that $1^T\lambda = 1$. The third index of λ is fixed at $1/K$ because one entry of the subset must be equal to P_O. Let

$$\lambda = [\frac{K-1}{K} - \varepsilon, \varepsilon, \frac{1}{K}]^T \tag{4.35}$$

Define:

$$\lambda^* = [\frac{K-1}{K} - \varepsilon + \xi, \varepsilon - \xi, \frac{1}{K}]^T \tag{4.36}$$

where $0 < \xi < \varepsilon$. By lemma 4, since $\lambda_0^* > \lambda_0$, then

$$W_2(\chi_j(\lambda^*), P_{-\alpha^*})^2 \leq W_2(\chi_j(\lambda), P_{-\alpha^*})^2. \tag{4.37}$$

This process of adding ξ to the first entry of λ^* can be repeated as λ_1^* approaches 0. We can define the following limit:

$$\min_{\lambda} W_2(\chi_j(\lambda), P_{-\alpha^*})^2 = \lim_{\lambda_0^* \to (\frac{K-1}{K})^-, \lambda_1^* \to 0^+} W_2(\chi_j(\lambda^*), P_{-\alpha^*})^2 \tag{4.38}$$

And as such, the optimal λ^* is given by:

$$\lambda^* = [\frac{K-1}{K}, 0, \frac{1}{K}]^T \tag{4.39}$$

which corresponds to the set in which there are $K-1$ entries of $P_{-\alpha^*}$ and one entry of P_O. □

Through an analogous argument, it follows that the largest possible distance between a distribution in $B(\Upsilon, \alpha)$ and a subset barycenter for which the subset does not contain the outlier distribution occurs between P_{α^*} and the barycenter of a subset that contains K-1 distributions of $P_{-\alpha^*}$ and one distribution of P_{α^*}. With these derivations, we define the following sets:

$$\mathcal{S}^{(j*)} = \{P_{-\alpha^*}, P_{-\alpha^*}, ..., P_{-\alpha^*}, P_O\} \tag{4.40}$$

$$\mathcal{S}^{(\hat{j})} = \{P_{-\alpha^*}, P_{-\alpha^*}, ..., P_{-\alpha^*}, P_{\alpha^*}\} \tag{4.41}$$

We assume that there are at least $K-1$ samples of $P_{-\alpha^*}$ in Ψ.

Remark 1. *Because we assert that $K > 2$ in the definition of the problem, there does not exist a set with K-1 distributions of the anomaly (i.e. the symmetric analogue) and thus the result proved later in Theorem 1 is unique.*

4.6 Localization of the Barycenter

Lemma 6. *Suppose set $\mathcal{S}$ has two distinct entries, which we denote as R and Q. Then $W_2(\chi^*_{\mathcal{S}}, R) \leq W_2(R,Q)$ and $W_2(\chi^*_{\mathcal{S}}, Q) \leq W_2(R,Q)$.*

Proof. For the sake of brevity, throughout the proof we utilize $\chi^* \equiv \chi^*_{\mathcal{S}}$. We define the objective function

$$J_{\mathcal{S}}(P) \triangleq \frac{1}{K}\sum_{k=1}^{K} W_2(P,P_k)^2. \tag{4.42}$$

which is minimized in (4.4) in order to solve for the barycenter. We will prove Lemma 6 by contradiction. Let $W_2(\chi^*,R) > W_2(R,Q)$. The resulting objective function is:

$$J_{\mathcal{S}}(\chi^*) = \frac{1}{K}[(K-1)W_2(\chi^*,R)^2 + W_2(\chi^*,Q)^2] \tag{4.43}$$

Using the assumption $W_2(\chi^*,R) > W_2(R,Q)$, we derive the following inequality:

$$J_{\mathcal{S}}(\chi^*) > \frac{K-1}{K}W_2(R,Q)^2 + \frac{1}{K}W_2(\chi^*,Q)^2 \tag{4.44}$$

By definition of a barycenter, χ^* must minimize the objective function. If there exists χ' for which $J_{\mathcal{S}}(\chi') < J_{\mathcal{S}}(\chi^*)$ then χ^* is not the minimizer. Let $W_2(\chi',R) = W_2(R,Q)$ and as such we will evaluate the objective function for $\chi' = Q$.

$$J_{\mathcal{S}}(Q) = \frac{K-1}{K}W_2(R,Q)^2 \tag{4.45}$$

Since the Wasserstein metric is non-negative:

$$J_{\mathcal{S}}(Q) < J_{\mathcal{S}}(\chi^*) \tag{4.46}$$

and therefore the inequality $W_2(\chi^*,R) \leq W_2(R,Q)$ must hold since χ^* is a barycenter. We

can analogously show $W_2(\chi^*, Q) \leq W_2(R, Q)$. Let Let $W_2(\chi^*, Q) > W_2(R, Q)$, which yields the objective inequality:

$$J_{\mathcal{S}}(\chi^*) > \frac{K-1}{K} W_2(\chi^*, R)^2 + \frac{1}{K} W_2(R, Q)^2 \tag{4.47}$$

Now, let $W_2(\chi', Q) = W_2(P, Q)$; evaluate the objective function for $\chi' = R$.

$$J_{\mathcal{S}}(R) = \frac{1}{K} W_2(R, Q)^2 \tag{4.48}$$

Since $J_{\mathcal{S}}(\chi^*) > J_{\mathcal{S}}(R)$, therefore $W_2(\chi^*, Q) \leq W_2(R, Q)$. □

Lemma 7. *Suppose set $\mathcal{S}$ has two distinct entries, which we denote as R and Q. Then $\chi^*_{\mathcal{S}}$ lies on a geodesic between R and Q.*

Proof. For the sake of brevity, throughout the proof we utilize $\chi^* \equiv \chi^*_{\mathcal{S}}$. By Lemma 6, $W_2(\chi', Q) \leq W_2(R, Q)$ therefore, if χ^* lies on the geodesic between R and Q then the following statement holds:

$$W_2(R, Q) = W_2(\chi^*, R) + W_2(\chi^*, Q) \tag{4.49}$$

as derived in (4.9). Let χ' be a distribution that does not lie on the geodesic. By the triangle inequality:

$$W_2(R, Q) < W_2(\chi', R) + W_2(\chi', Q) \tag{4.50}$$

We can evaluate the objective function at χ^* and χ', respectively, with substitutions given by (4.49) and (4.50).

$$J_{\mathcal{S}}(\chi^*) = \frac{K-1}{K} [W_2(R, Q) - W_2(\chi^*, Q)]^2 + \frac{1}{K} W_2(\chi^*, Q)^2 \tag{4.51}$$

$$\begin{aligned} J_{\mathcal{S}}(\chi') &= \frac{K-1}{K} W_2(\chi', R)^2 + \frac{1}{K} W_2(\chi', Q)^2 \\ &> \frac{K-1}{K} [W_2(R, Q) - W_2(\chi', Q)]^2 + \frac{1}{K} W_2(\chi', Q)^2 \end{aligned} \tag{4.52}$$

Since $W_2(\chi', Q)$ and $W_2(\chi^*, Q)$ have the same domain, let $W_2(\chi', Q) = W_2(\chi^*, Q)$. Therefore, for every possible distribution, χ', with $W_2(\chi', Q)$ there exists a distribution χ^* which lies on the geodesic between R and Q with $W_2(\chi', Q) = W_2(\chi^*, Q)$, for which $J_{\mathcal{S}}(\chi^*) < J_{\mathcal{S}}(\chi')$. Therefore the barycenter must lie on the geodesic that connects the two elements of set $\mathcal{S}$. □

4.7 Wasserstein Distance to the Barycenter

In this section, we will prove that $W_2(\chi_{j^*}, \mathcal{S}_{P_O}^{(j^*)})^2 > W_2(\chi_{\hat{j}}, \mathcal{S}_{P_{\alpha^*}}^{(\hat{j})})^2$, which proves Theorem 1. By virtue of (4.4), we can define the specific barycenter for sets $\mathcal{S}^{(j^*)}$ and $\mathcal{S}^{(\hat{j})}$. For the sake of brevity we will denote the barycenter as $\chi_{\mathcal{S}}$ and we will denote the K^{th} distribution to be P_A, where:

$$\begin{cases} P_A = P_O & \text{for set } \mathcal{S}^{(j^*)} \\ P_A = P_{\alpha^*} & \text{for set } \mathcal{S}^{(\hat{j})} \end{cases} \tag{4.53}$$

By Lemma 7 and by summation properties of geodesics (4.9):

$$W_2(\chi_{\mathcal{S}}, P_A)^2 = [W_2(\Upsilon, P_A) + W_2(\Upsilon, \chi_{\mathcal{S}})]^2 \tag{4.54}$$

We assume for now that the barycenters, χ_{j^*} and $\chi_{\hat{j}}$, lie closer to $P_{-\alpha^*}$ than P_{α^*} with respect to Υ, which means $W_2(P_{\alpha^*}, \chi_{\hat{j}}) \geq W_2(P_{\alpha^*}, \Upsilon)$ and analogously $W_2(P_O, \chi_{j^*}) \geq W_2(P_O, \Upsilon)$, but we show an extension of these methods if this is not the case.

We will first derive an expression for $W_2(\Upsilon, \chi_{\mathcal{S}})$. By use of Lemma 7, summation properties of geodesics (4.9), and shorthand notation given by (4.53):

$$\chi_{\mathcal{S}} = \arg\min_{P \in \mathcal{P}(\mathsf{X})} \frac{K-1}{K}[W_2(\Upsilon, P_{-\alpha^*}) - W_2(P, \Upsilon)]^2 + \frac{1}{K}[W_2(P, \Upsilon) + W_2(\Upsilon, P_A)]^2 \tag{4.55}$$

Rearranging yields:

$$\chi_{\mathcal{S}} = \arg\min_{P \in \mathcal{P}(\mathsf{X})} AW_2(P,\Upsilon)^2 + BW_2(P,\Upsilon) + CW_2(P,\Upsilon) + D \tag{4.56}$$

where:

$$\begin{cases} A = 1 \\ B = -\frac{2(K-1)\alpha}{K} \\ C = \frac{2(\alpha+\delta)}{K} \quad \text{for } \mathcal{S}^{(j^*)} \\ C = \frac{2\alpha}{K} \quad \text{for } \mathcal{S}^{(\hat{j})} \\ D = \alpha^2 + \frac{\delta^2+2\alpha\delta}{K} \quad \text{for } \mathcal{S}^{(j^*)} \\ D = \alpha^2 \quad \text{for } \mathcal{S}^{(\hat{j})} \end{cases} \tag{4.57}$$

We denote the objective function to be minimized in (4.56) as:

$$f(g(P)) = Ag(P)^2 + Bg(P) + Cg(P) + D \tag{4.58}$$

where $g(P) = W_2(P,\Upsilon)$. The only constraint on P is that P is a probability distribution, and thus $P \succcurlyeq 0$. An implicit family of inequalities follow:

$$v_i^T P \leq 0 \quad i = 1,\ldots I, \tag{4.59}$$

where P is vectorized over its support and v_i is a vector, identical in length to P, in which the ith entry of v_i is equal to -1 and the remaining entries are equal to zero. This family of inequalities harnesses the property that the sum of non-positive real numbers is non-positive. Further, since P is a probability distribution, the total mass within the support cannot be zero. Therefore, there exists at least one $i \in 1,\ldots I$, for which $v_i^T P < 0$. Since this inequality is strict and the remaining inequalities are affine, by Slater's condition [26], strong duality holds. The Lagrangian dual is

given by:

$$\mathcal{L}(P,\underline{\kappa}) = f(g(P)) + \sum_{i=1}^{I} \kappa_i v_i^T P \tag{4.60}$$

We assume feasibility of both the primal and dual problems ($\underline{\kappa}^*, P \succcurlyeq 0$), and since strong duality holds, we employ the Karush-Kuhn-Tucker (KKT) conditions [26] by taking the gradient of (4.60) with respect to P and equating it to zero:

$$0 = \frac{\partial g(P)}{\partial P}[2Ag(P) + B + C] + \frac{\partial}{\partial P}[\sum_{i=1}^{I} \kappa_i^* v_i^T P] \tag{4.61}$$

We divide both sides of (4.61) by $\frac{\partial g(P)}{\partial P}$:

$$2Ag(P) + B + C = -\frac{\partial}{\partial g(P)}[\sum_{i=1}^{I} \kappa_i^* v_i^T P] \tag{4.62}$$

Since $v_i^T P$ does not depend explicitly on $g(P)$, its derivative with respect to $g(P)$ is zero, and therefore:

$$W_2(P,\Upsilon) = -\frac{B+C}{2A} \tag{4.63}$$

Substituting this result into (4.54), using i) appropriate values of C defined in (4.57) and ii) proper substitution of $W_2(\Upsilon, P_A)$ yields:

$$\begin{cases} W_2(\chi_{j^*}, P_O)^2 = (\frac{K-1}{K}(\alpha+\delta) - \frac{B}{2})^2 \\ \\ W_2(\chi_{\hat{j}}, P_{\alpha^*})^2 = (\frac{K-1}{K}(\alpha) - \frac{B}{2})^2 \end{cases} \tag{4.64}$$

Since, $\delta > 0$, therefore $W_2(\chi_{j^*}, P_O)^2 > W_2(\chi_{\hat{j}}, P_{\alpha^*})^2$.

4.8 Extension

Let $W_2(P_{\alpha^*}, \chi_{\hat{j}}) < W_2(P_{\alpha^*}, \Upsilon)$ and $W_2(P_O, \chi_{j^*}) < W_2(P_O, \Upsilon)$. Now, the equation described by the geodesic between P_A and $P_{-\alpha^*}$, which is an analogue of (4.54), becomes:

$$W_2(\chi_j, P_A)^2 = [W_2(\Upsilon, P_A) - W_2(\Upsilon, \chi_{\mathcal{S}})]^2 \tag{4.65}$$

Equivalently, (4.55) is rewritten as:

$$\chi_{\mathcal{S}} = \arg\min_{P \in \mathcal{P}(\mathsf{X})} \frac{K-1}{K}[W_2(\Upsilon, P_{-\alpha^*}) + W_2(P, \Upsilon)]^2 + \frac{1}{K}[W_2(\Upsilon, P_A) - W_2(P, \Upsilon)]^2 \tag{4.66}$$

And with identical methodology to Section 4.7, the analogue of (4.58) is:

$$f(g(P)) = Ag(P)^2 + B'g(P) + C'g(P) + D \tag{4.67}$$

where $B' = -B$ and $C' = -C$ in (4.57). Combining (4.63) and (4.65) yields an identical result to (4.64).

4.9 Application in GI

Using the philosophical approach derived in this work, we developed an algorithm to detect an anomalous probability distribution. We will demonstrate our algorithm with an example in the field of Gastroenterology (GI). One of the components of GI functionality is the gastric slow wave. This wave travels along the surface of the stomach and locally actuates the smooth muscle contractions that help propel food through the GI tract [158, 80]. This process is essential in healthy digestion. Recently, it has been shown that pathological slow wave patterns can occur and may give rise to GI disease and symptoms [110, 53].

It is of current interest to be able to non-invasively detect pathological slow wave patterns,

as this could aid in diagnostics and tailored treatment regimens [86, 73]. Contemporary work in the field has made use of the high-resolution electrogastrogram (HR-EGG), which involves a dense array of electrodes placed on the abdomen (directly on the skin) [49, 53]. HR-EGG analysis methods involve filtering, novel artifact rejection methods [52], and the calculation of instantaneous wave directions at each electrode via sinusoid functions of estimated phase gradients [49, 53, 7].

We used methods analogous to those presented in [8] to obtain a set of probability distributions. In summary, we studied 32 human research subjects (n=25 with gastric disease, n=7 asymptomatic controls). We performed approximately 60 minute HR-EGG recordings directly after a meal and calculated instantaneous wave directions at each electrode. We performed maximum likelihood density estimation of von Mises mixtures, since direction data lie on a circular support, and thus had a set of 32 probability distributions.

There are multiple classes of wave propagation patterns across the set of subjects, with distinct phenotypes presented in the probability distributions. For example, a class of distributions were similar to that shown in Fig. C.2A, in which each distribution was generally unimodal with a strong peak near zero. The frame of reference of the direction values (and thus support of the probability distribution) was such that a direction value of zero corresponded to the subject's anatomical angle to the interface between the stomach and intestines. As such, the distribution shown in Fig. C.2A is physiologically indicative of healthy wave propagation toward the intestines. A second class of wave propagation involved bifurcated propagation both toward the intestines and esophagus, seen in Fig. C.2B. A third class of wave propagation involved stochastic and uncoordinated wave propagation, seen in Fig. C.2C.

All subjects fit within these three classes except for one. This final subject, had anomalous wave physiology, as seen in Fig. C.2D, in which there was a strong indication of wave propagation only toward the esophagus. We developed an algorithm based on the philosophical themes presented in this work in order to detect the anomalous distribution. The general framework for

our algorithm involved i) selecting a randomized subset of $K = 7$ probability distributions, ii) computing a Wasserstein barycenter between subset distributions, and iii) computing Wasserstein distances between each subset distribution and the subset barycenter. We formulated the Wasserstein distance as a linear program over a compact set, identical in methodology to [8], which has been shown to greatly reduce computational cost and run time [166, 9, 17]. We repeated these three steps a total of 300 times and plotted each distance between a distribution and the respective subset barycenter. We hypothesized that the anomaly distribution would have distances to respective subset barycenters that could exceed all other distances between distributions and barycenters. We have summarized our algorithm in pseudocode shown in Algorithm 1.

Algorithm 1 Calculating Large Set of Distances Between Subset Distributions and Subset Barycenters

Input: Set of N probability distributions, Ψ, and integer, K
$j \leftarrow 0$
repeat
  Select subset of distributions $L_k \in \{1,...,N\}$ where $k = 1,...K$, $L_u \neq L_v$.
  Denote $\mathcal{S}_k^{(j)} = \Psi_{L_k}$
  Compute barycenter of subset, χ_j, given by (4.4).
  $k \leftarrow 1$
  while $k \leq K$ **do**
    Compute distances to barycenter, $W_2(\chi_j, \mathcal{S}_k^{(j)})^2$, given by (4.2).
    $k \leftarrow k+1$
  end while
  $j \leftarrow j+1$
until $j = 300$

As seen in Figure C.3, there is a 'maximum' distance between a distribution and a subset barycenter, which we will denote $D*$, that is only attainable by one subject's probability distribution. This is marked with a red arrow in Fig. Figure C.3. Further, there are several distances less than $D*$ that are only attainable by the same individual subject's probability distribution. This subject is labeled anomalous, and has the probability distribution seen in Fig. C.2D, which was the one distribution that did not fit into the three main classes of distributions. The anomalous distribution can be identified after only 100 iterations (Fig. C.3A) but emerges

even more distinctly after 300 iterations (Fig. C.3B).

Since our study involved n=32 subjects, we chose a randomized approach to select subset probabilities. In future studies, if the number of distributions is significantly higher and there is concern over whether or not all distributions will be represented, a slightly different approach can be defined in which the first distribution in every subset is pre-specified and the remaining K-1 distributions are selected randomly. This would ensure that each distribution was selected however many times as is desired. In addition, larger datasets may call for a larger subset dimension, K.

4.10 Acknowledgements

We would like to acknowledge Dr. David Kunkel for his expertise in gastric functional and motility disorders as a Gastroenterologist at University of California, San Diego. The HR-EGG recordings collected and development of analysis methodologies for the application section of this work were done so in collaboration with Dr. Kunkel. We would additionally like to acknowledge Dr. Piya Pal as an expert and professor in convex optimization at the University of California, San Diego. Dr. Pal designed lecture notes, cited in this work, that heavily influenced the structure of the proof of Lemma 3.

Chapter 4, in full, is currently being prepared for submission for publication of the material. Agrusa, Anjulie; Coleman, Todd. The dissertation author was the primary investigator and author of these materials.

Chapter 5

Robust Regression and Optimal Transport Methods to Predict Gastrointestinal Disease Etiology from High Resolution EGG and Symptom Severity

5.1 Introduction

5.1.1 Background and Unmet Need

Gastrointestinal (GI) problems are the second most common reason for someone to miss school or work, after the common cold [29]. Two of the most prevalent GI functional and motility disorders are functional dyspepsia (FD) and gastroparesis (GP), which affect 10% and 1.5-3% of the U.S. population, respectively [42, 77]. Multiple disease etiologies underlie the umbrella diagnoses of GP and FD, including but not limited to i) antral hypomotility, ii) pylorospasms, iii) gastric myoelectric dysfunction [56], and iv) autonomic dysfunction. Tailored treatments for

such causes include i) prokinetic medications, ii) botulinum toxin injection of the pylorus [85], gastric neuromodulation [2, 20, 91, 98, 89, 13, 122], and iv) vagus nerve stimulation [58, 59], respectively. The common occurrence of treatment regimens tailored to the wrong etiology or to symptom mitigation results in repetitive visits to the clinic and adverse effects on quality of life with vast economic, social, and healthcare consequences. There is a desire and unmet need to sub-type GP and FD in terms of underlying etiologies to improve prediction of treatment response and enable targeted interventions [86, 73].

5.1.2 Previous Work

Current diagnostic tools are typically not etiology-specific. For instance, the PAGI-SYM questionnaire on foregut symptoms, which includes the gastroparesis cardinal symptom index, is used to manage treatments but shows a high degree of symptom overlap with underlying disease etiologies [56]. Alternatively, the gastric emptying test, a clinical gold standard that entails post-fasting ingestion of a meal containing a radioactive tracer, and measuring the amount of tracer that has left the stomach at fixed intervals, shows inconsistent correlation with symptom severity [123, 74]. Further, some medications improve symptoms but not gastric emptying, and vice versa [39, 97, 21]. As such, there is an unmet need to develop widely-deployable screening tools that correlate with symptom severity and sub-type GP/FD in terms of disease etiology.

Gastric neuromuscular activity involves the interplay between electrical activity on the stomach's surface and smooth muscle contractions that propel food from the esophagus to the intestines. Gastric pacemaker cells, or interstitial cells of Cajal (ICCs), receive modulatory inputs from enteric neurons and undergo spontaneous cyclical depolarization, similar to cardiac pacemaker cells in the heart. These oscillatory membrane potentials excite neighboring smooth muscle cells, causing contractions at regular intervals, and excite adjacent downstream ICCs in their network [46, 161, 120]. This results in a traveling electrical wave, called the gastric slow wave [158, 80, 153].

Slow waves, in healthy individuals, originate in the pacemaker region of the stomach and propagate in an anterograde direction toward the duodenum. This propagation occurs in equipotential rings along the outer surface of the stomach, at a frequency of 3 cycles per minute (0.05 Hz) [80, 28, 66] . Abnormalities in slow wave initiation and propagation patterns arise from loss or deterioration of ICCs [113, 19]. These pathological slow waves can spontaneously arise from locations other than the pacemaker region (i.e. the antrum) and/or propagate in aberrant directions (i.e. in a retrograde direction towards the esophagus or a bifurcated combination of anterograde/retrograde propagation). These abnormal initiation and propagation patterns disrupt the spatial coordination of smooth muscle contractions and have been observed in GP/FD patients [110].

High-resolution spatiotemporal mapping of the gastric slow wave has been performed with electrodes placed directly on the surface of the stomach during open-abdominal surgery. These findings indicated normal slow wave behavior in control subjects (n=12) [109] and abnormalities in slow wave initiation and propagation in $>$ 80% of GP subjects (n=12) [110]. Whereas these studies demonstrated spatial slow wave abnormalities as possible explanations for foregut disorders, the involvement of surgical procedures limits the technique's scalability.

A non-invasive alternative is the electrogastrogram (EGG), which involves electrodes placed on the surface of the abdomen (as opposed to the stomach) and measures electrical activity that reaches the abdominal surface via volume conduction. Conventional EGG is comprised of 3-4 electrodes and is analyzed via spectral methods (i.e. power-frequency) [80, 118]. However, it has been shown that i) spectral features do not reliably correlate with symptoms or disease [69, 107] and ii) spatial slow wave abnormalities (which *do* have high correlation with disease and symptoms) are undetectable by spectral analyses [110, 53, 19, 23].

The high-resolution EGG (HR-EGG) has recently been established to overcome the pitfalls of traditional EGG. This approach is comprised of a dense cutaneous electrode array and subsequent signal processing methods for artifact rejection [52] and estimation of spatial slow

wave directions [49]. It has been shown, both *in silico* and in human subjects, that estimated wave directions orient in an anterograde direction towards the duodenum in healthy controls [49, 7].

HR-EGG was also recently used in human subjects with GP (n=18) and FD (n=7) to establish a correlation between incidence of non-anterograde slow wave propagation and symptom severity [53]. Specifically, a linear correlation (r=0.57) was established between the percentage of the recording for which slow wave directions propagated in a non-anterograde direction and the mean Gastroparesis Cardinal Symptom Index (GCSI) symptom score, a standardized foregut symptom severity score that takes into account nausea, vomiting, early satiety, bloating, and abdominal distention [134]. Although this key finding suggests that gastric slow wave features determined from cutaneous multi-electrode recordings may provide an explanation for GP and FD symptoms, it has has some key limitations. Since multiple disease etiologies underlie GP and FD, a 'one size fits all' regression paradigm is likely inadequate to explain symptom severity solely from spatial slow wave features. For instance, in patients with high symptom severity arising from etiologies other than gastric myoelectric dysfunction, spatial slow wave features would likely resemble those of healthy controls although their symptom severity score would exceed that of control subjects [53]. We hypothesize that if a patient's symptom severity score exceeds what would be 'expected' from slow wave patterns, then the underlying cause of symptoms is other than gastric myoelectric dysfunction. In line with this reasoning, if we can isolate and exclude groups of patients for which symptoms exceed regression predictions, one can reasonably hypothesize that a higher-accuracy correlation can be drawn between abnormal wave propagation and symptoms in the remaining subjects.

5.1.3 Our Proposed Approach

The aforementioned drawbacks call for the implementation of a "robust regression" method which involves a mathematically rigorous approach to determine an inclusion criterion of data to be used in regression, and an exclusion criterion of data to be considered outliers.

Robust regression has been widely used in the current literature to fit highly-accurate regressions in the presence of outliers by assuming the probability distribution of residuals has multiple modes [137, 18, 32, 35]. Optimization of model-fitting parameters typically involves maximizing a multi-modal likelihood function over the residual errors, with at least one mode near zero [137, 18, 32]. This method optimizes regression fit over the data assumed to represent 'signal' rather than 'noise.'

We use robust regression in our work presented here in order to identify and exclude an "outlier" class of patients for which we hypothesize that their underlying GP or FD etiology is *not* gastric myoelectric dysfunction. This separation of patients whose symptoms are adequately explained by slow wave features, and those whose symptoms are not, may have the potential to predict treatment response to, for instance, gastric neuromodulation (a more invasive intervention) or medications (a more conservative approach). Prediction of treatment response has the potential to rapidly guide effective treatments, as well as reduce the risk of performing unnecessary procedures.

Another potential improvement to the preliminary correlation study in [53] involves estimating an entire probability distribution of wave directions for each recording. This provides more information, pertaining to modality and overall distributional structure, that may lead to improved characterizations of recordings. Furthermore, utilizing the full probability distributions enables the use of an optimal transport framework to build metrics between distributions. 'Optimal transport' refers to the process of minimizing the cost of transporting all of the mass in one probability distribution to another, and has powered state-of-the-art studies in broad applications of mathematics, data science, and economics [159, 141, 127, 48]. In this study, we make use of the Wasserstein metric between probability distributions to develop unique spatial slow wave features to be used downstream for robust regression.

We calculate Wasserstein barycenters [9] between probability distributions of control subjects, which serves as a 'model' of normalcy. We then compute the optimal transport cost

between each subject's probability distribution and the controls' barycenter to quantify the deviation of subject's probability distribution from 'normalcy'. In addition, we compute distances between each subject's probability distribution and a uniform distribution over directions as a secondary feature to quantify the deviation from chaotic or uncoordinated propagation. As previously mentioned, abnormal slow wave propagation can take on several forms. One such form is an abnormal initiation near the antrum with a large degree of coordinated retrograde propagation, in addition to some anterograde propagation [110]. Probability distributions of wave directions in these types of cases would differ in shape and modality from recordings of healthy controls, which are typically dominated by anterograde propagation. In addition, this physiology is distinct from chaotic or uncoordinated propagation, which would result in wave directions having probabilistic signatures more akin to a uniform distribution. As such, these features derived from a sophisticated optimal transport framework follow insights from slow wave physiology.

Lastly, previous HR-EGG methods [49, 53] required computed tomography (CT) scans for both accurate placement of electrodes over the stomach, as well as measurements related to anatomical landmarks in order to define 'normal' propagation. The need for a CT presents a current bottleneck in these existing HR-EGG methods because CT is costly for patients and requires otherwise unnecessary exposure to radiation. We have addressed this in our present work by taking into account the features and modalities of full probability distributions, which results in a method that is robust to an estimation of these anatomical measurements.

5.1.4 Key Findings

In summary, our contributions in this manuscript involve building a full probability distribution from measured HR-EGG wave directions, developing an optimal transport framework for features pertaining to distances between probability distributions, and performing robust regression to a) isolate a group of patients in which we hypothesize that gastric myoelectric

dysfunction is not a prominent etiology in their disease and b) form a more accurate correlation between symptoms and features of wave directions in the remaining group of subjects. Our methods are robust to a barycenter-type estimate of the angle that defines 'normal' propagation and, if used in conjunction with recent methods in stomach identification with portable ultrasound [84], would abolish the need for a CT scan. Overall, we propose a method to screen patients for which aggressive intervention, such as gastric neuromodulation, would be most effective in promoting good clinical outcomes.

5.2 Methods

5.2.1 Data Acquisition and Signal Processing

We enrolled healthy control subjects with no prior GI disease (n=7), GP patients (n=18), and FD patients (n=7), for a total of 32 subjects. We performed a PAGI-SYM questionnaire on all subjects, with repeat PAGI-SYM questionnaires on 5 GP/FD patients after pyloric or prokinetic interventions. The PAGI-SYM questionnaire is a standardized, validated patient survey to rate symptoms such as nausea, bloating, early satiety, etc [133]. We performed 25-channel HR-EGG recordings on all subjects for 90 minutes (approximately 30 minutes preprandial and approx. 60 minutes postprandial). Each subject was asked to fast for 12 hours prior to ingestion of the meal at the 30 minute time point.

The ingestion of food activates excitatory neural responses that interface with the stomach to increase the magnitude of gastric smooth muscle contractions [163]. Given this fact, and since gastric contraction strength correlates with EGG power [147, 167], we solely use the postprandial portion of the recording the remainder of this section to operate in a regime of high signal-to-noise ratio (SNR) for our HR-EGG recordings.

HR-EGG recording and prepossessing methods were identical to those published here [53], which included filtering in the range of 0.05 Hz, artifact rejection, etc. For generalizability,

we refer to the number of subjects as S and the number of electrodes as N throughout this manuscript. Ethical approval for this work was obtained from the institutional review board at the University of California, San Diego, USA.

5.2.2 Instantaneous Wave Direction at Every Sensor

We performed the Hilbert transform on mean-centered HR-EGG waveforms to extract the instantaneous wave phase, $\phi_n^{(s)}(t)$, at the n^{th} channel ($n = 1,...N$) and the s^{th} subject ($s = 1,...S$).

$$V_n^{(s)}(t) + iHb[V_n^{(s)}(t)] = a_n^{(s)}(t)e^{i\phi_n^{(s)}(t)} \tag{5.1}$$

In accordance with current HR-EGG analysis studies [53, 49, 7], we calculated the instantaneous wave direction, $\Lambda_n^{(s)}(t)$, at each channel by taking the inverse tangent of phase gradients.

$$\Lambda_n^{(s)}(t) = \tan^{-1}\left(\frac{\nabla\phi^{(s)}(y_n,t)}{\nabla\phi^{(s)}(x_n,t)}\right) \tag{5.2}$$

In order to exclude segments of recordings dominated by noise, we defined a notion of a 'sustained wave' and compiled subsets of the recording in which the 'sustained' criterion was met, explained as follows. We defined the instantaneous phase gradient directionality (PGD) for the s^{th} subject as:

$$\mathrm{PGD}^{(s)}(t) = \frac{\|\frac{1}{N}\sum_{n=1}^{N}\nabla\phi^{(s)}(x_n,y_n,t)\|_2}{\frac{1}{N}\sum_{n=1}^{N}\|\nabla\phi^{(s)}(x_n,y_n,t)\|_2} \tag{5.3}$$

where $t = 1,...T^{(s)}$. The PGD has been used in neuroscience [138] as well as previous HR-EGG analyses [7, 49, 53] as a measure of cohesivity among wave directions. From Jensen's inequality, it is upper bounded by 1 which occurs when a plane wave, which has the highest degree of direction agreement, is present. In order to reduce the false discovery of noise to below 1% with 25 channel arrays, we categorize a wave as satisfying the 'sustained' criterion if the PGD exceeds

0.5 for longer than 2 seconds [49]. As such, direction values calculated in (5.2) that indicate sporadic or otherwise uncoordinated wave propagation must have some inherent pattern and coordination in order to meet the sustained criterion, as is consistent with physiology (i.e. an ectopic event results in local coordination distinct from patterns elsewhere on the stomach [110]).

5.2.3 Direction Values Relative to the Duodenum

The aforementioned wave direction values are measured with respect to the cutaneous recording array. However, the stomach has a characteristic 'J' shape, with subject-to-subject anatomical variability, and further is not always perfectly in plane with the sensor array. Previous studies have addressed this by translating direction values by the subject-specific angle to the duodenum, derived from CT measurements [53, 49]. Thus, a translated direction value of zero radians would indicate wave propagation toward the duodenum (or anterograde propagation).

Instead of translating each subject's direction values by their unique duodenum angle, we translated direction data by an average of all S subjects' duodenum angle. We then slightly adjusted the translation, such that the location of maximum incidence of the histogram of translated directions within a range of $\pm\pi/6$ occurred at zero radians. There is precedent for this approach within the literature that gives a $\pm\pi/3$ degree window of tolerance around the subject-specific angle to the duodenum in order to define 'normal' wave propagation [53]. Our adjustment to the translation operated on a more conservative window while still accounting for anatomical variability. Translating directions in this way eliminates the need for a CT going forward, since we can utilize measurements from CTs that have already been performed in calculation of averages, as well as incorporate data from large imaging databases to further improve the average if desired. Because the downstream analysis methods presented in this work estimate the entire probability distribution and exploit characteristics of the shape and modalities of the distribution, we obviate the need for high precision with the angle by which we translate direction values.

We collapsed sustained wave direction values across both spatial and temporal axes, giving

rise to set $\mathcal{D}^{(s)}$, for the s^{th} subject, as seen in (5.4). The set $\mathcal{D}^{(s)}$ is shown for two human research subjects as a histogram in the circular domain in Fig. D.2A.

$$\mathcal{D}^{(s)} = \{\Lambda_n^{(s)}(t)\} \quad \forall n \in \{1,...N\}, \forall t \in \{1,...T^{(s)}\} \tag{5.4}$$

5.2.4 Probability Density Estimation with von Mises Mixtures

Since direction data in set $\mathcal{D}^{(s)}$ lie on a circle, a natural family of probability density functions on which we can perform density estimation is the family of von Mises distributions, which are commonly referred to as the "Gaussians" of circular statistics [94, 95]. The von Mises probability density function is given by:

$$f(\theta|\mu,\kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)} \tag{5.5}$$

where μ is the mean of the distribution, κ is the inverse variance, and $I_0(\kappa)$ is the 0^{th} order Bessel function. See Fig. D.1A for examples.

We performed maximum likelihood density estimation of von Mises mixtures (vMM), given by (5.6), in order to capture multiple modalities in $\mathcal{D}^{(s)}$, which could indicate multiple modes of wave propagation (i.e. bifurcated anterograde and retrograde propagation). An example of a von Mises mixture is shown in Figure D.1B.

$$f_{vMM}(\theta|\Theta) = \sum_{k=1}^{K} \lambda_k \frac{e^{\kappa_k\cos(\theta-\mu_k)}}{2\pi I_0(\kappa_k)} \tag{5.6}$$

where $\Theta = (\lambda_1,...\lambda_K,\mu_1,...\mu_K,\kappa_1,...\kappa_K)$, and

$$\sum_{k=1}^{K} \lambda_k = 1. \tag{5.7}$$

The log likelihood, given the set of vMM parameters, Θ, and set $\mathcal{D}^{(s)}$ is:

$$L(\Theta) = \sum_{r=1}^{R} \log \sum_{k=1}^{K} \lambda_k \frac{e^{\kappa_k \cos(\mathcal{D}_r^{(s)} - \mu_k)}}{2\pi I_0(\kappa_k)}. \tag{5.8}$$

where R is the length of $\mathcal{D}^{(s)}$, or $NxT^{(s)}$. Optimal parameters, $\hat{\Theta}$, are those which maximize (5.8).

$$\hat{\Theta} = \arg\max_{\Theta} L(\Theta; \mathcal{D}^{(s)}) \tag{5.9}$$

5.2.5 Model Selection Procedure

Solving (5.9) with no restriction on the number of mixture distributions, K, may cause the model to over-fit to variants arising from noise or human subject-to-subject variability, which results in a loss of generalizability . Adding a penalty to the number of distributions prevents over-fitting while still capturing multiple modalities in $\mathcal{D}^{(s)}$. Since all weights are non-negative and sum to 1, by virtue of (5.7), traditional ℓ_1 penalized model fitting procedures are not applicable to this context. We implemented a pruning method to iteratively reduce the number of mixture distributions, K, and subsequently perform dimensionality-based penalization.

We began the iterative pruning method by executing maximum likelihood optimization in (5.9), given the set $\mathcal{D}^{(s)}$ and a large dictionary of "M" pairs of von Mises density parameters, $(\mu, \kappa)_m$, to over-represent the problem space [76]. The parameter dictionary consisted of samples of μ within a range of $[-\pi, \pi]$, sampled in increments of $2\pi/9$. Similarly, κ was sampled within a range of $[0, 15]$, in increments of $15/9$. The complete dictionary consisted of a mesh grid of all possible combinations of μ and κ, with size $M = 100$. Since parameters μ_k and κ_k were specified by this dictionary, optimization in (5.9) solves for the mixture probabilities, $\lambda_k (k = 1, ...K)$, which maximize (5.8).

We then eliminated the set $(\mu, \kappa)_{k'}$ which was lowest-weighted, or

$$k' = \arg\min_{k=1,..K} \lambda_k, \tag{5.10}$$

since $(\mu,\kappa)_{k'}$ contributed the least to the overall log likelihood. We repeated this pruning process until the dimensionality of the parameter dictionary, K, was equal to 1. We computed the negative normalized log likelihood (NNLL) at each pruning iteration, given by:

$$\text{NNLL}(K) = -\frac{1}{R} L(\hat{\Theta}_K) \tag{5.11}$$

where $\hat{\Theta}_K$ is the set of parameters that maximizes the log likelihood in (5.8) when the parameter dictionary is pruned to K specific (μ,κ) pairs. We implemented Bayesian Inference Criterion (BIC) [142] penalization where the penalized normalized negative log likelihood (PNNLL) is given below.

$$\text{PNNLL}(K) = \text{NNLL}(K) + \frac{K \log R}{2R} \tag{5.12}$$

We obtained the optimal number of von Mises densities K^*. which minimizes (5.12), and then obtained the corresponding set of optimal (μ,κ) density parameters, to construct the density estimate.

5.2.6 Maximum Likelihood Estimation with the EM Algorithm

During each step of the pruning method, we used the Expectation Maximization (EM) for mixtures [131] to carry out the maximum likelihood estimation in (5.9).

The EM algorithm makes use of a latent variable, $\underline{Z} \in \{1,2,...K\}^R$, for which $Z_r = k$ encodes which distribution $\mathcal{D}_r^{(s)}$ is drawn from in the von Mises mixture model. The complete log likelihood (conditioned on $\underline{Z}$) is given by:

$$p(\underline{Z}, \mathcal{D}^{(s)}; \Theta_K) = \sum_{r=1}^{R} \sum_{k=1}^{K} 1_{\{Z_r=k\}} \log[\lambda_k f(\mathcal{D}_r^{(s)} | \mu_k, \kappa_k)] \tag{5.13}$$

where $1_{\{A\}}$ equals 1 if event A occurs and 0 otherwise. For the sake of brevity we will use the notation $\mu_k, \sigma_k, \lambda_k$ to index specific parameters in the iteration of the pruning method in which

there are K mixture distributions. It is implied that μ_k refers to the specific $\mu_k \in \Theta_K$, for instance. The posterior probabilities of latent variable, $\underline{Z}$, conditioned on set $\mathcal{D}^{(s)}$ are given by:

$$\Gamma_{r,k}^{(m)} \triangleq P(Z_r = k|\mathcal{D}_r^{(s)};\Theta_K^{(m)}) \tag{5.14}$$

$$= \frac{\lambda_k f(\mathcal{D}_r^{(s)}|\mu_k^{(m)},\kappa_k^{(m)})}{f_{vMM}(\mathcal{D}_r^{(s)}|\Theta_K^{(m)})} \tag{5.15}$$

at the mth iteration of the EM algorithm. We define the expectation of the complete log likelihood given by (5.13) as:

$$Q(\Theta_K^{(m+1)},\Theta_K^{(m)}) = \sum_{r=1}^{R}\sum_{k=1}^{K}\Gamma_{r,k}^{(m)}\log[\lambda_k f(\mathcal{D}_r^{(s)}|\mu_k,\kappa_k)] \tag{5.16}$$

where $\Theta_K^{(m)}$ is the vector of parameters used to compute the posterior probability, $\Gamma_{r,k}$, and vector $\Theta_K^{(m+1)}$ houses each λ_k,μ_k,σ_k in (5.16). Maximizing this expectation yields parameter updates:

$$\lambda_k^{(m+1)} = \frac{\sum_{r=1}^{R}\Gamma_{r,k}}{R} \tag{5.17}$$

for $k = 1,...K$ where R is a normalization constant to guarantee the sum of $\underline{\lambda}^{(m+1)}$ is equal to 1.

We performed 20 iterations of parameter updates.

5.2.7 Wasserstein Distances: Optimal Transport Theory

Assuming that probability distributions P and Q have densities p and q respectively, the Wasserstein squared distance between probability distributions P and Q is defined as:

$$W_2(P,Q)^2 = \min_{\pi\in\Pi[P,Q]}\int_{u\in U}\int_{v\in V}c(u,v)\pi(u,v)dudv \tag{5.18}$$

where $\Pi[P,Q]$ is the set of all possible joint probability densities with marginals P and Q, U is the support of P and V is the support of Q. The function $c(u,v)$ is a cost function pertaining to the squared distance between u and v.

In the work presented here, data are direction values, which lie on a circular manifold. As such, we computed a quadratic cost

$$c(u,v) = d_c(u,v)^2 \tag{5.19}$$

operating on $d_c(u,v)$

$$d_c(u,v) = \min\{|u-v|, 2\pi - |u-v|\} \tag{5.20}$$

which is the geodesic distance on the unit circle. We formulated the optimization in (5.18) as a linear program, discretized over the compact set with boundary points $-\pi$ and π subject to:

$$\begin{aligned}
&\pi(u,v) \geq 0 \\
&\int_{u\in U}\int_{v\in V} \pi(u,v)dudv = 1 \\
&\int_{v\in V} \pi(u,v)dv = p(u) \\
&\int_{u\in U} \pi(u,v)du = q(v)
\end{aligned}$$

where the first two constraints hold because $\pi(u,v)$ is a joint probability density, and the latter two hold because $\pi \in \Pi[P,Q]$ and thus it must preserve marginal densities p and q.

5.2.8 Wasserstein Barycenter of Probability Distributions from Control Subjects

We constructed a Wasserstein barycenter (denoted P_B) between healthy control subjects. This barycenter represents an 'average' distribution for the control group. Other aggregate

methods, such as ℓ_2 averaging of probability distributions, fail to capture modality and shapes of individual distributions (see Figure D.3). This motivated our use of a Wasserstein barycenter approach, as distributional shape and modality are paramount to the downstream feature space in this work.

$$P_B = \arg\min_{P \in \Omega[P,P_c]} \frac{1}{C} \sum_{c=1}^{C} W_2(P,P_c)^2 \tag{5.21}$$

where C is the number of control subjects (n=7) and P_c refers to individual probability distributions of control subjects, solved for in Section 5.2.4. For each subject (including controls) we computed Wasserstein squared distances, via (5.18), between i) each probability distribution, P_s and the control barycenter, P_B and ii) each probability distribution and a uniform distribution, denoted as P_U, over the support $[-\pi,\pi]$. This first feature is a metric of how close each distribution is to an approximation of normalcy, whereas the second feature measures closeness to complete randomness, as a uniform distribution models the general distributional shape of aberrant wave propagation (see Fig. D.4 for such an outlier). These distances were the basis for our features in the following section.

5.2.9 Robust Regression and EM Clustering

We identified an outlier in the control group. This outlier probability distribution deviated in modality from both a) the expected probability distribution of healthy controls with respect to physiology and b) the other control subjects (see Fig. D.4). In fact, this probability distribution showed aberrant, chaotic slow wave propagation similar to distributions estimated from from GP and FD HR-EGG recordings. In this section, we have removed the outlier from calculation in the robust regression algorithm. It is plotted in the results figures but not used for computation of regression coefficients and residuals. As such, the indexing notation of human research subjects going forward is $d \in 1,...D$, where $D = 31$.

To carry out robust regression in this work, we defined a regressor matrix F, of size $D \times 3$,

with the d^{th} row given by:

$$\underline{F}_d^T = [1, W_2(P_d, P_U)^2, W_2(P_d, P_B)^2]. \tag{5.22}$$

We defined the response vector, $\underline{Y} \in \mathcal{R}^D$ as the Mean GCSI (Gastroparesis Cardinal Symptom Index) score for all subjects, which is a subset of the PAGI-SYM questionnaire [133, 134]. We modeled $\underline{Y}$ to linearly relate to F such that:

$$\underline{Y} = \underline{\beta}^T F + \underline{N} \tag{5.23}$$

where $\underline{N}$ is the vector of residual errors. Since spatial gastric slow wave abnormalities are only *one* type of etiology of gastrointestinal foregut disorders, we hypothesize that this linear model will fit a subset of the data for which spatial slow wave abnormalities are a dominant cause of symptoms. For the rest of the subjects, however, we expect their symptom and HR-EGG features to be anomalous to the linear trend. For instance, this anomalous class would encompass patients who have normal slow waves, with probability distributions indistinguishable from healthy controls, but high symptom severity scores. Following this logic, we assume residual errors from the linear regression, $\underline{N}$, will follow a bi-modal distribution: with one mode near zero and another mode near some $\alpha > 0$. This bi-modal distribution can be modeled as a Gaussian mixture model (GMM)

$$f_{GMM}(x; \underline{\eta}, \underline{\sigma}, \underline{\phi}) = \sum_{i=1}^{2} \phi_i \mathcal{N}(x; \eta_i, \sigma_i) \tag{5.24}$$

where ϕ_1 and ϕ_2 are non-negative weights that add to 1 and $\mathcal{N}(x|\eta, \sigma)$ is the density of a Gaussian random variable with expectation η and variance σ^2 evaluated at x:

$$\mathcal{N}(x; \eta, \sigma) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\eta)^2}{2\sigma^2}}. \tag{5.25}$$

Since we hypothesize that the residuals $\underline{N}$ follow a bi-modal GMM, $\underline{Y}$ follows a bi-modal GMM,

with one mode at $\underline{\beta}^T F$ and the other mode at $\underline{\beta}^T F + \alpha$. Because we hypothesize that the outlier class will be comprised of patients for which symptoms exceed what would be 'expected' from features of the slow wave, we aimed to identify outliers that lie *above* the regression line. As such, we defined $\alpha > 0$. To find the optimal offset, α^*, we created a dictionary of offset values, with entries α_h (for $h = 1, ...H$), on the interval $(0, 1]$ in increments of $\frac{1}{H}$. The EM algorithm was used to solve for the remainder of the density estimation parameters in (5.26). EM-model fitting was performed H times, for each α_h in the dictionary. The optimal offset, α^*, was chosen to be the value of α which maximized the log likelihood given in (5.26). We implemented this method with $H = 10$

We used methods similar to those presented in [32]. However, deviating from the approach in [32], where regression coefficients, $\underline{\beta}$, for each mode were distinct from one another, we solved for the regression coefficients only for the non-outlier class. This decision rests upon our hypothesis that the non-outlier class corresponds to patients where slow wave patterns are the dominant etiology for their symptoms, which implies that the outlier class pertains to patients where the dominant cause of symptoms is an etiology other than slow wave abnormalities (i.e. dysfunction of the pylorus, or autonomic dysfunction). Because there are multiple such etiologies, and given that our dataset contained $n = 31$ subjects, we determined that this approach minimizes opportunities for overfitting or spurious calculations.

Similar to the von Mises mixture density estimation procedure discussed in Section 5.2.4, optimization of the GMM problem can be solved with the EM algorithm. We initialized the weights as $\phi_1 = 0.9, \phi_2 = 0.1$ since we can reasonably hypothesize that the majority of the data fits within Class 1. We initialized $\underline{\beta}$ as the coefficients of the linear least squares regression of the entire dataset with symptoms. We initialized σ_1^2 and σ_2^2 as the variance of the residuals given by (5.23) with $\underline{\beta}$ given by its initialized value.

We define a latent binary random vector $\underline{Z} \in \{1, 2\}^D$ for which $Z_d = 1$ if N_d is drawn from a $\mathcal{N}(0, \sigma_1^2)$ distribution and $Z_d = 2$ if N_d is drawn from a $\mathcal{N}(\alpha, \sigma_2^2)$ distribution. As such,

the joint distribution on $(\underline{Z},\underline{Y})$ for parameters $\theta = (\underline{\eta},\underline{\sigma},\underline{\phi})$ is given by:

$$p(\underline{Z},\underline{Y};\theta) = \sum_{d=1}^{D}\sum_{i=1}^{2} 1_{\{Z_d=i\}} \log\left[\phi_i \mathcal{N}(Y_d|\eta_i,\sigma_i)\right] \tag{5.26}$$

where $1_{\{A\}}$ equals 1 if event A occurs and 0 otherwise.

For the E step, we define $\gamma_{d,i}^{(m)}$ to be the posterior probability at the mth iteration that the dth sample is generated from the ith Gaussian component, given by [60]:

$$\begin{aligned}
\gamma_{d,i}^{(m)} &\triangleq P(Z_d = i|Y_d;\theta^{(m)}) && (5.27)\\
&= \frac{\phi_i \mathcal{N}(Y_d|\eta_i,\sigma_i)}{\phi_1 \mathcal{N}(Y_d|\eta_1,\sigma_1) + \phi_2 \mathcal{N}(Y_d|\eta_2,\sigma_2)} && (5.28)
\end{aligned}$$

where

$$\eta_1 = \underline{\beta}^T F \tag{5.29}$$

$$\eta_2 = \underline{\beta}^T F + \alpha_h \tag{5.30}$$

where $h \in 1...H$. For the M step, the standard EM parameter updates for the GMM problem at iteration $m+1$, $\theta^{(m+1)} = (\eta^{(m+1)},\sigma^{(m+1)},\phi^{(m+1)})$, are given as follows [60]. We first define $n_D^{(m)}$ as $n_D^{(m)} = \sum_{d=1}^{D} \gamma_{d,i}^{(m)}$ and then

$$\begin{aligned}
\phi_i^{(m+1)} &= \frac{n_D^{(m)}}{D} && (5.31)\\
\eta_i^{(m+1)} &= \frac{1}{n_D^{(m)}} \sum_{d=1}^{D} \gamma_{d,i}^{(m)} (i=1) Y_d && (5.32)\\
\sigma_i^{2\,(m+1)} &= \frac{1}{n_D^{(m)}} \sum_{d=1}^{D} \gamma_{d,i}^{(m)} \left(Y_d - \eta_i^{(m+1)}\right)^2 && (5.33)
\end{aligned}$$

We used (5.32) to find the regression coefficients by virtue of (5.29). We performed 100 iterations of the EM algorithm (for each $h \in 1,...H$) and found the optimal offset, $\alpha*$ to be 0.8. We

performed maximum likelihood classification [164] to assign each subject to either 'Class 1' or 'Class 2,' where Class 2 included data that lies far above the regression line. Subject p was assigned to Class 2 if:

$$\frac{\mathcal{N}(Y_d|\eta_1,\sigma_1)}{\mathcal{N}(Y_d|\eta_2,\sigma_2)} \leq 1 \tag{5.34}$$

We performed clustering to isolate a subset of Class 2 with features that most closely resembled healthy controls. We refer to this subset of Class 2 as the 'anomalous class.' In total, this anomalous class involved subjects which a) lie far above the regression line (i.e. high symptom severity with respect to regression predictions from slow wave features) and b) had slow wave features that were indistinguishable from controls. We used two different clustering algorithms: i) DBSCAN [61, 62] with parameters: epsilon=0.2, minimum samples=2 and ii) K-Means [64] with K=3, to cluster the transformed features ($\underline{\beta}^T F$) within Class 2. As shown in Fig D.5, both algorithms isolated the same group of seven data points which met criteria i) and ii). In addition, these seven data are the only data whose transformed features fit within a range of the controls. Thus, we labeled these seven subjects as "anomalous" and hypothesize that their prominent disease etiology is not gastric myoelectric dysfunction. Lastly, we performed a final linear least squares regression on the complete dataset labeled as 'non-anomalous' (post anomaly-exclusion) to adjust the regression coefficients and compute r values.

5.3 Results

Figure D.6 shows the result of the robust regression and identification of the anomaly class. We show, in orange, the patients for which we hypothesize that spatial gastric slow wave abnormalities are not a prominent etiology of their disease. These patients exhibit slow wave patterns that closely resemble healthy controls, as is seen with feature overlap, but exhibit high symptom scores, which is not a characteristic of controls. We measured an r value of 0.8 from

regression on only non-anomalous data.

Figure D.7 shows the symptom prognosis of five patients after treatment. Two patients, shown in triangle markers, had no improvement, and three patients, shown in hexagon markers, did see improvements. Interventions included treatments of the pylorus and prokinetic medications. The three patients that experienced symptom improvement lie farther away from the regression line than the patients that did not improve. A natural hypothesis is that these treatments helped mitigate other etiologies responsible for the patients' GP and FD. However, for patients that lie directly on the regression line, it is hypothesized that their symptoms are a direct cause of spatial gastric slow wave abnormalities and as such treating other etiologies may not result in improvement of symptoms.

Figure D.8 compares methods to previous work that pioneered linear regression of spatial gastric slow wave features with patient symptom scores [53]. These previous findings utilize raw direction values, where wave directions are calculated analogously to those described in (2.10), have an inclusion criteria given by (2.11) and are translated by the patient-specific angle to the duodenum, derived from the CT. 'Normal' wave directions are defined to be those which lie with in $\pm\pi/3$ of the angle to the duodenum., and direction values are labeled as 'abnormal' otherwise. To create a comparison plot, we integrated the probability distributions found in Section 5.2.4 on the closed interval $\pm\pi/3$, and defined this as the percentage of 'normal' slow wave activity. Since the integral of a probability density function over its entire support is equal to one, this integration technique captures a percentage of 'normal' and 'abnormal,' and yields very similar results to [53, Fig. 4]. Fig. D.8A shows the linear least squares (LLSQ) regression (blue line) with the shaded 95% confidence interval. Data closely followed trends from [53]. We report a r value of 0.56 and previous findings report r=0.57 [53, Fig. 4]. This firstly validates our approach of probability density estimation because it shows that probability distributions are able to capture signatures of the histogram well. Secondly, this provides a good basis to make comparisons between methods in [53] and the design of our new methods presented here.

Figure D.8 B shows, in pink, the GP and FD subjects with percentages of abnormal slow waves that coincide within the range of healthy controls, excluding the control outlier. It is unlikely that a cluster of patients which have normally functioning slow waves but high symptom scores could be identified with these types of methods because the control group spans a large portion of the total data. For example, data marked as pink (within range of controls) are directly adjacent to data marked in purple (out of range of controls) and as such, there is no clear separation between a healthy and pathological percentage of abnormal slow waves.

Figure D.8 C replicates the approach shown in Panel A, but we have marked the seven subjects that were flagged as 'anomalous' in the methods presented in this work. Some of these subjects lie either close to or directly on the regression line and thus do not appear anomalous at all in this previous approach. This shows that using the previous methodology, this group of patients for which it is hypothesized that spatial slow wave abnormalities are not a likely cause of disease would be missed.

In Fig. D.8 D, we replicate panel B with only the postprandial period of the recordings. This was done to show that the paradigm shift within our methodology (i.e. looking at signatures of the full probability distribution and performing robust regression to account for different disease etiologies) was the key driver of our novel findings presented in this work. While we assert that performing HR-EGG studies during an immediate postprandial period is a best practice moving forward (with justification in 5.2.1), isolating the postprandial portion of the recording alone was not the mechanism behind our results.

We show in Fig. D.8, in red, the outlier within the control group that we identified after estimating probability distributions. Interestingly, this outlier does not appear to separate from the remainder of the control subjects in Fig. D.8 panels A-D. We show the probability density function for this outlier subject in Fig. D.4. This density lacks a cohesive single mode centered near zero and more closely resembles a probability distribution that would be expected from chaotic propagation. However, summing energy across the multiple modes within the range of

$\pm\pi/3$, gives a false indication of normalcy since this total energy is closer to a range of the other controls. As such, investigating signatures and modalities of the entire probability distribution allowed us to recognize an outlier that would have otherwise gone undetected.

5.4 Discussion

In this work, we were able to show a strong correlation (r=0.8) between features relating to spatial gastric slow wave abnormalities and standardized patient symptom scores, improving upon previous published findings' r value by 0.23 [53]. We were also able to isolate a group of patients for which we hypothesize spatial gastric slow wave abnormalities are not likely a predominant etiology of their disease. We have shown results that allow us to hypothesize that treating a patient's GP or FD with a treatment regimen that does not directly target gastric myoelectric dysfunction may only produce results in the case where other underlying etiologies are present, and patients who lie on the regression line (i.e. we hypothesize that their symptoms can be explained entirely based on spatial gastric slow wave abnormalities) will see limited success in such treatments.

Previous studies have used a patient CT to guide electrode array placement over the antrum of the stomach and determine the angle towards the duodenum [53, 49]. The advent of portable ultrasound devices and their usage to locate the antrum [83, 84] suggest placement of electrodes to cover the stomach need not require a CT. Since the methods presented here do not make use of the exact angle to the duodenum, from CT reconstructed geometry, but rather are robust to a barycenter of duodenum angles across several human subjects, a patient CT scan may no longer be needed for future studies. In addition, even if the electrode placement is less precise with the ultrasound localization method, we hypothesize, based on what has been shown here, that designing analysis methods to find patterns within the overall shape and modalities of probability distributions may provide robustness to slight deviations from optimal electrode placement, as

these perturbations would not likely change distribution modalities. Since the HR-EGG device is non-invasive, eliminating the need for a CT will eliminate a significant barrier in bringing this technology to the clinic as a widely-deployable screening tool.

By virtue of using histograms and performing density estimation, we made the implicit assumption in this work that wave direction values can be treated as independent, identically distributed samples collapsed across temporal and spatial axes (see Section 5.2.2). A subject of future work would be to solve for conditional probabilities in both the temporal and spatial axes using Markov modeling techniques.

It is well documented that the vulnerabilities of the EM algorithm arise from the non-convexity of the log likelihood surface and thus the presence of local minima/ maxima [22, 117]. Two of these vulnerabilities are: the need for appropriate parameter initialization as well as a proper stopping criterion to limit the number of extra iterations performed once convergence has been reached [22, 117, 170]. Failure to implement these regulations can result in parameters converging to one of their most extreme values. As such, it is typically standard to apply targeted parameter initializations [117]. In our work presented here, we initialized parameters with insights specific to the problem in the GMM implementation of the EM algorithm. However, when using the EM algorithm for the vMM problem, we initialized distribution weights randomly and we observed that convergence did not depend on proper initialization of weights, as was expected since this problem is convex. In addition, published studies have established precedent for enforcing a 'stopping criterion' for the EM algorithm [38]. As such, in both implementations of the EM algorithm (vMM and GMM problems), we chose a maximum number of parameter updates that allowed the algorithm to achieve convergence. More robust parameter initializations have been proposed [117], as well as competitive EM (CEM) [170] and other 'split and merge' algorithms [157], which offer an approach to address both the challenges of proper parameter initialization and developing a stopping criterion. A subject of future work could entail implementing one of these algorithms in lieu of the standard EM algorithm presented here.

Through the methods presented in this work, we identified an outlier within the control group. As discussed previously, identification of this outlier was only made possible once we estimated the full probability distribution of wave directions. The outlier, which deviates from from the 'normal' slow wave propagation pattern typically seen in controls, may be a result of some noise-based artifact of the recording. However, this finding could instead suggest discovery of an asymptomatic class of patients that have spatial slow wave abnormalities. This suggestion is not meant to imply certainty of any kind, but rather to motivate a future inquiry in the GP and FD research community. Future work could entail enrolling of a large cohort of healthy controls, performing HR-EGG and analysis methods analogous to those presented here to assess probability distributions of wave directions.

5.5 Implications of This Work

Currently there is not a reliable method to determine GP and FD etiology before performing treatment. It is hypothesized that gastric neuromodulation is able to entrain the slow wave. The methods and results presented here suggest that we may be able to improve identification of candidates for neuromodulation. Furthermore, we may be able to identify patients for which existing conservative treatments would yield symptom improvement. Finally, we may be able to provide an explanation as to why certain treatments either only partially alleviate symptoms or fail to do so completely.

Chapter 5, in full, is currently under review for publication of the material. Agrusa, Anjulie; Kunkel, David; Coleman, Todd. The dissertation author was the primary investigator and author of this material.

Appendix A

Appendix to Chapter 2

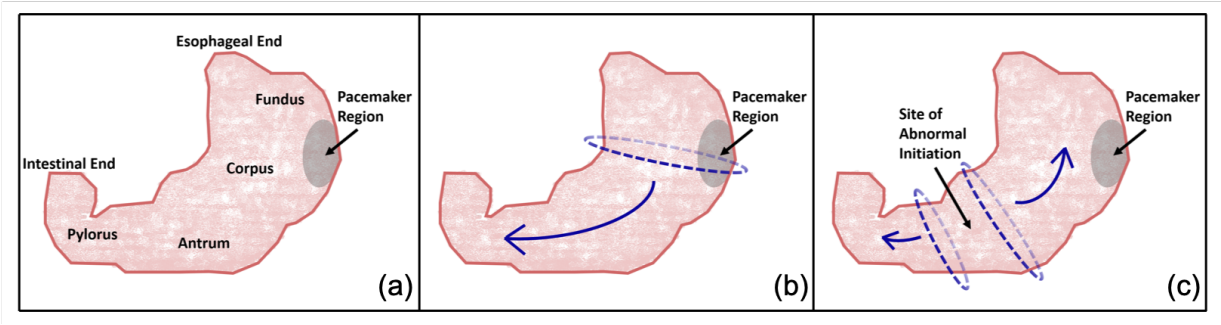

Figure A.1: Anterior view of stomach model extracted from human research subject's abdominal CT scan. a) Anatomical regions of the stomach with pacemaker region highlighted in gray. b) Normal gastric slow wave initiation in the pacemaker region and propagation of equipotential rings (dashed) in anterograde fashion toward the intestinal end. c) Abnormal initiation of the gastric slow wave outside of the pacemaker region. Propagation of distinct equipotential rings in the anterograde direction (left ring) and retrograde direction (right ring). These initiation and propagation patterns were characterized from invasive studies [109, 110].

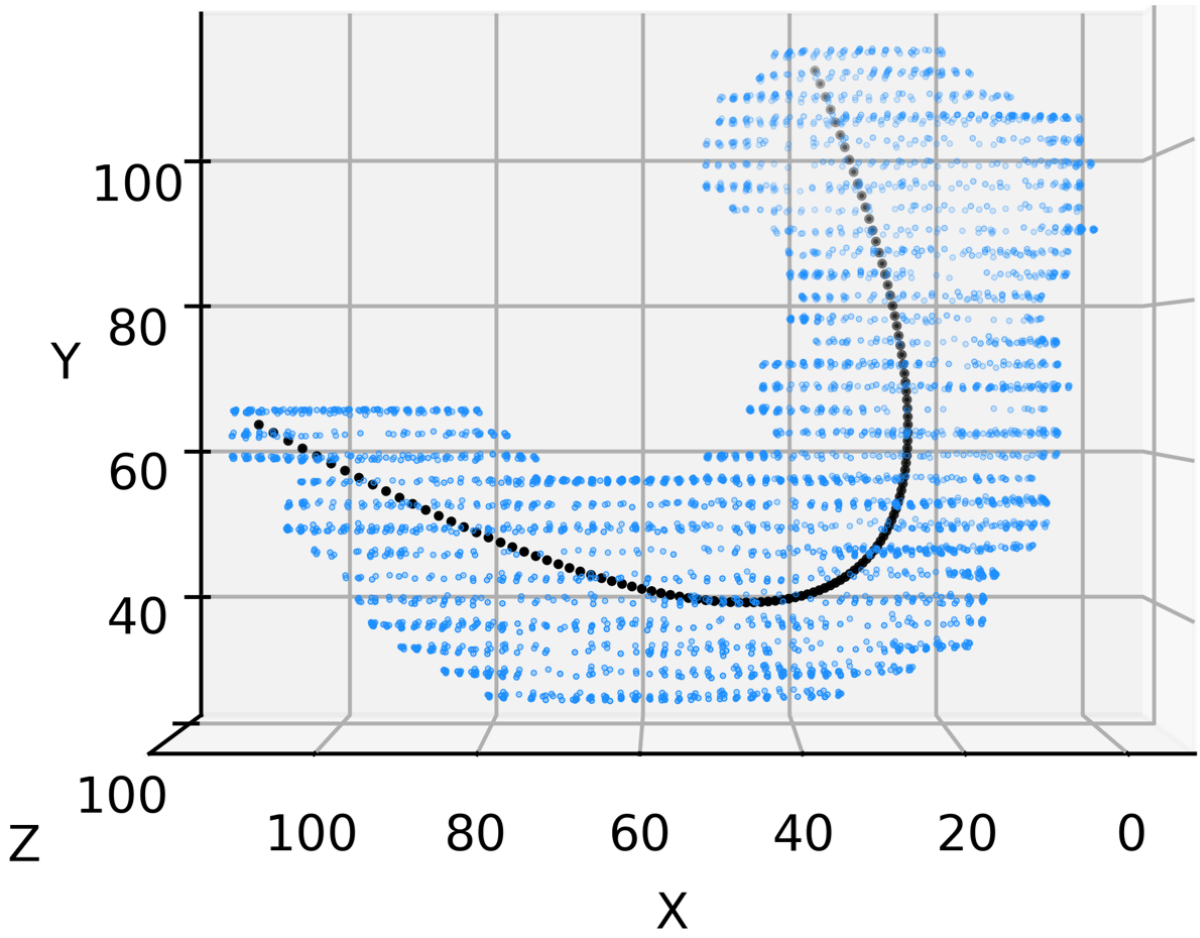

Figure A.2: Medial Curve (black) that captures the characteristic shape of the stomach, solved for with the optimization paradigm defined in (2.3). Units in mm.

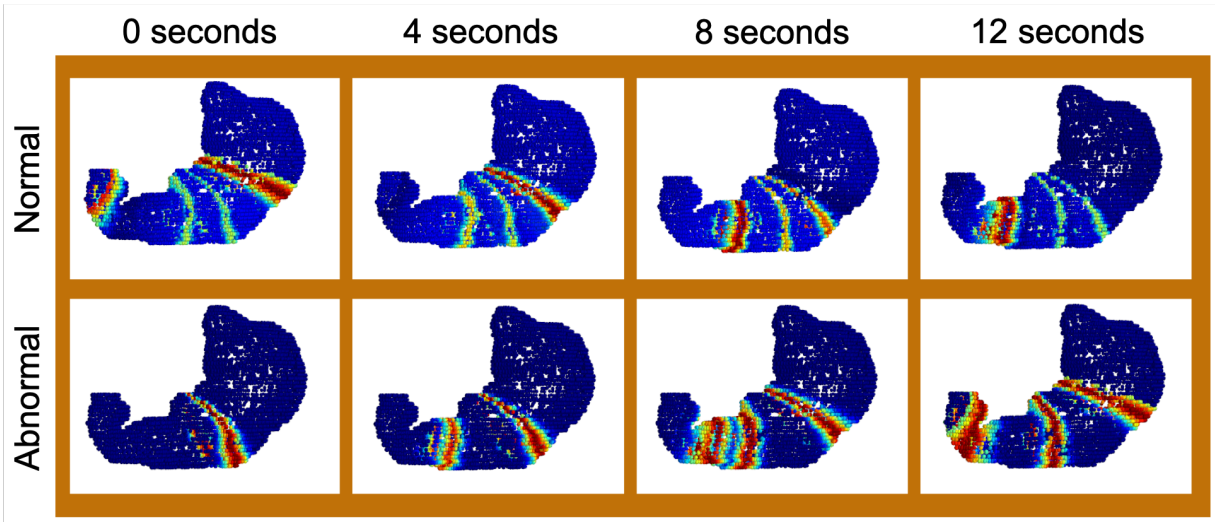

Figure A.3: Wave propagation at the serosal surface of the stomach with voltage values displayed as heat map. Normal slow wave initiation (top) and abnormal initiation (bottom), with each successive snapshot (left to right) at time t = 0,4,8,12 seconds. The voltage at each point on the serosal surface of the stomach is solved for in Section 2.2.1 in accordance with the description in the most recent literature [109, 110, 50].

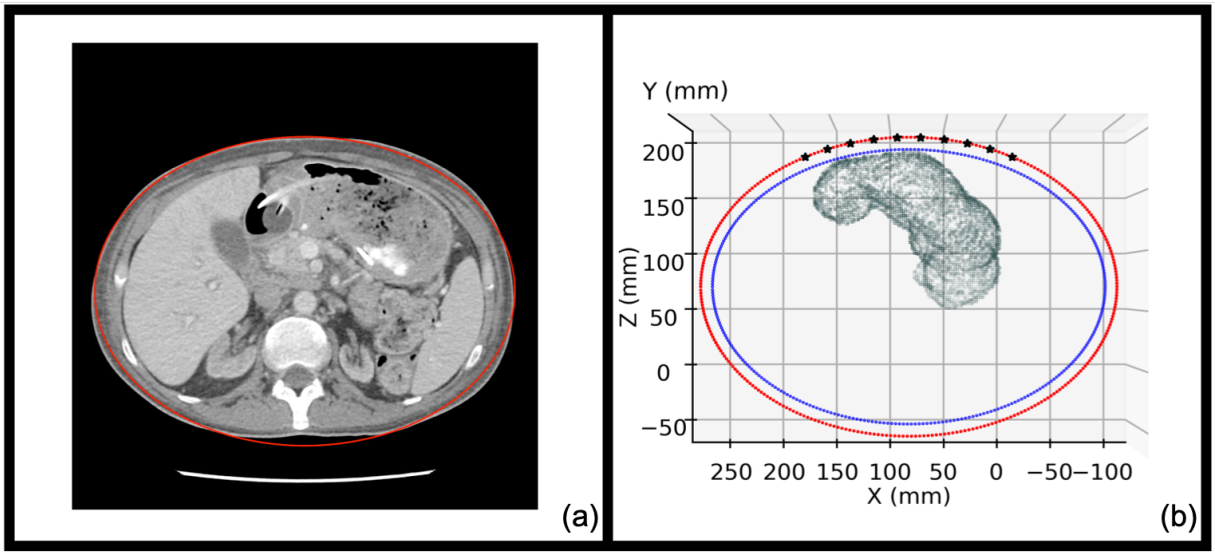

Figure A.4: a) Top-down projection of human research subject's abdominal CT scan onto the X,Z plane. Approximated elliptical cross section (red) from which elliptical radii were measured. b) Top-down projection of stomach and elliptic cylindrical point clouds onto the X,Z plane. Abdominal ellipse (red) and inner abdominal boundary (blue), separated by a 1 cm cavity. Abdominal and inner ellipses are translated laterally by z_0 via optimization described in Section 2.2.1. The anterior stomach and abdominal surfaces are as close as they can be to one another while imposing the constraint that no point on the stomach crosses the inner abdominal boundary. This is the 1 cm 'baseline' abdominal tissue depth.

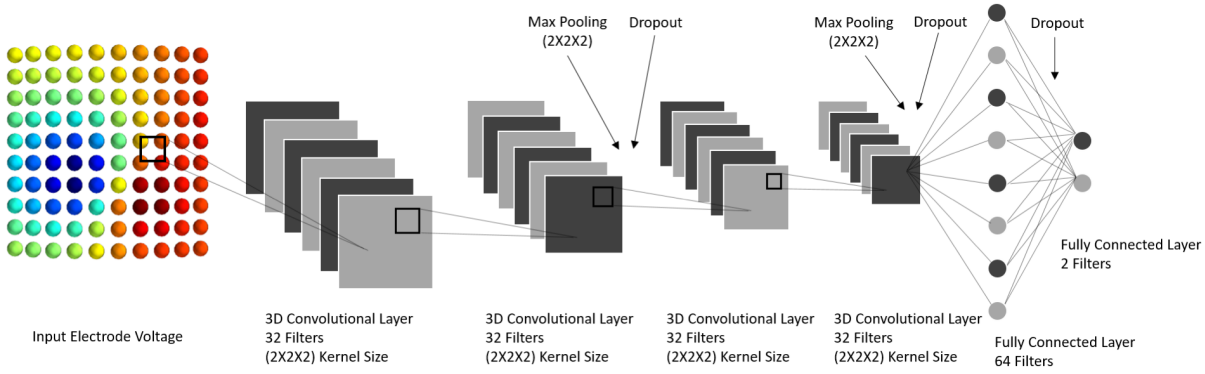

Figure A.5: Convolutional neural network architecture schematic, described in detail in Section 2.2.2.

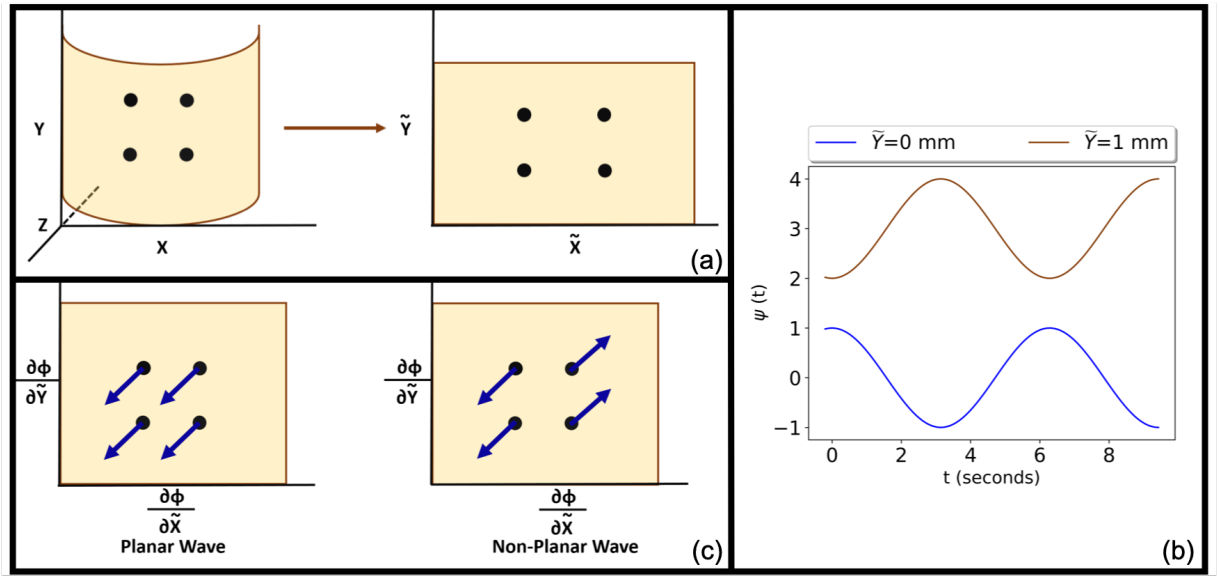

Figure A.6: a) Cartesian coordinates (x,y,z) translated to ($\tilde{x}$,$\tilde{y}$) coordinates along the curved abdominal surface. Spatial analyses in Section 2.2.2 are performed in ($\tilde{x}$,$\tilde{y}$) coordinates. b) Conceptual schematic of two out-of-phase oscillatory signals at different vertical spatial locations, $\tilde{y}$=0 (blue) and $\tilde{y}$=1 (brown). At time $t = 0$, the change in the oscillatory signal with respect to a unit change in $\tilde{y}$ is π. c) Phase gradient abstraction with wave velocity vector shown in blue. Both planar and non-planar waves have equal speeds (velocity vector magnitude) but the non-planar wave has non-homogeneous wave directions.

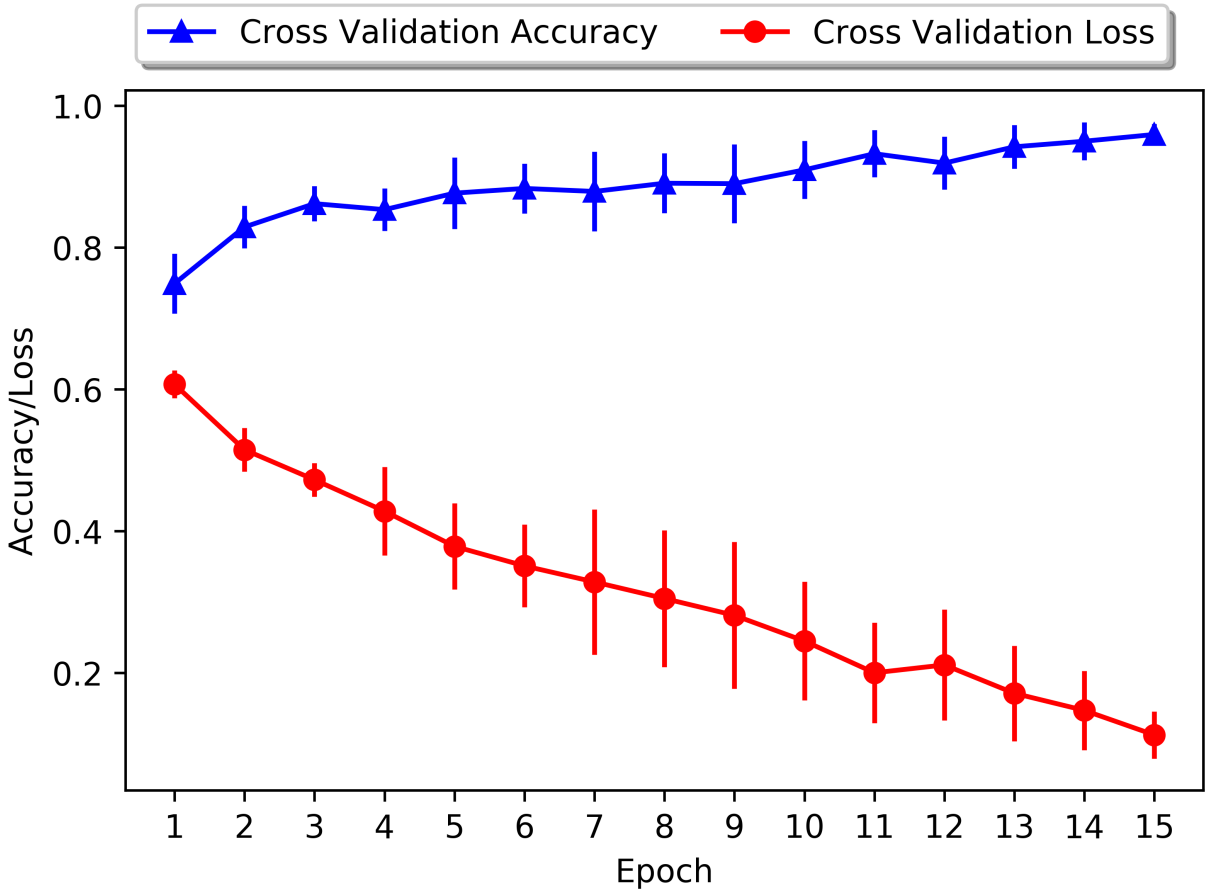

Figure A.7: Cross validation accuracy (blue) and loss (red) during training. Accuracy is shown on a [0,1] scale.

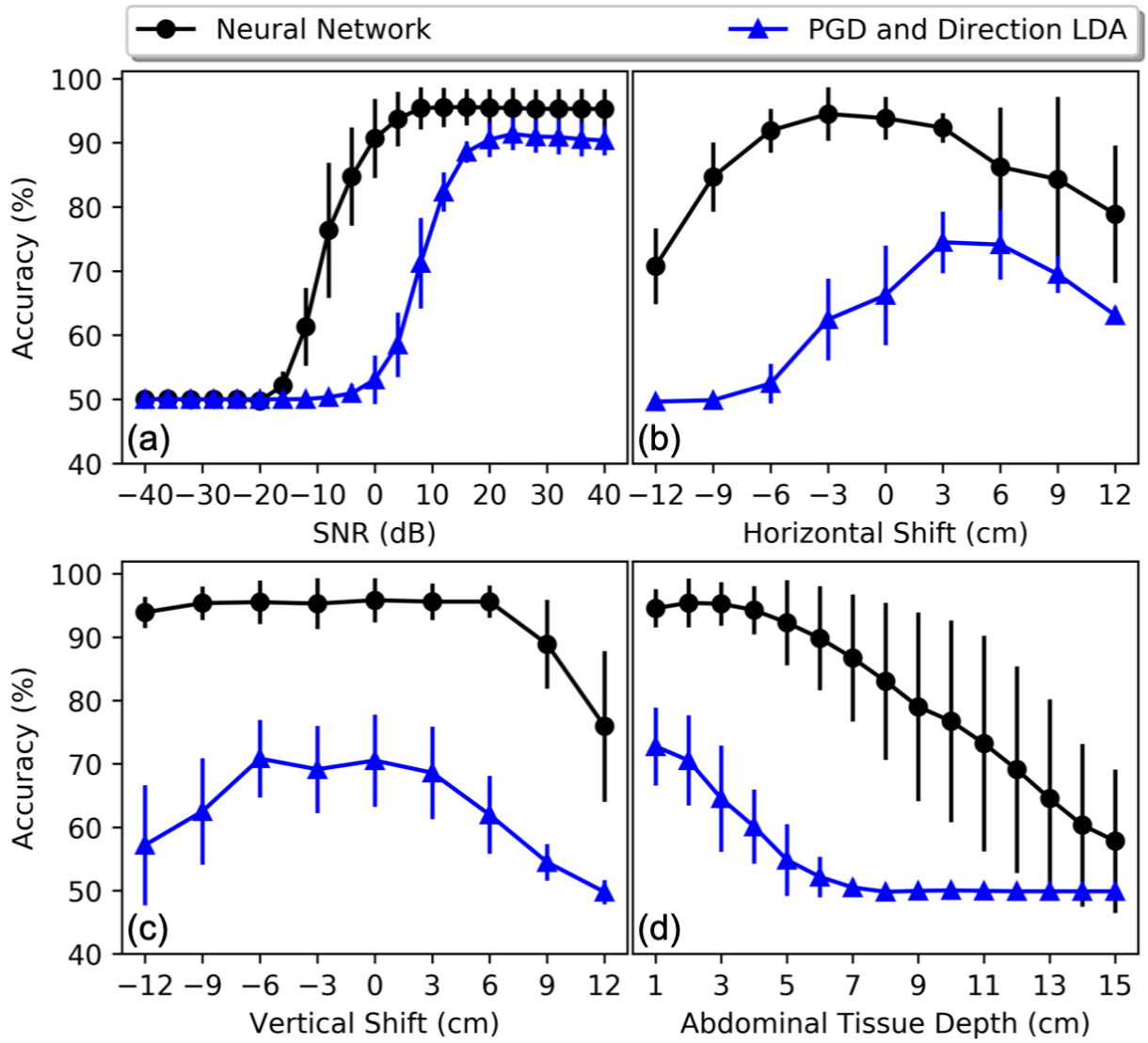

Figure A.8: Comparison of neural network performance with LDA performance. (a) Neural network and LDA performance as a function of SNR. (b) Neural network and LDA performance as a function of horizontal electrode array shifts along the curved surface of the abdomen. Shift distances correspond to arc lengths along the cylindrical ellipse. (c) Neural network and LDA performance as a function of vertical electrode array shifts. (d) Neural network and LDA performance as a function of abdominal tissue depth. All test datasets (b-d) had a baseline SNR of 10 dB (Explained in detail in Section 2.2.1).

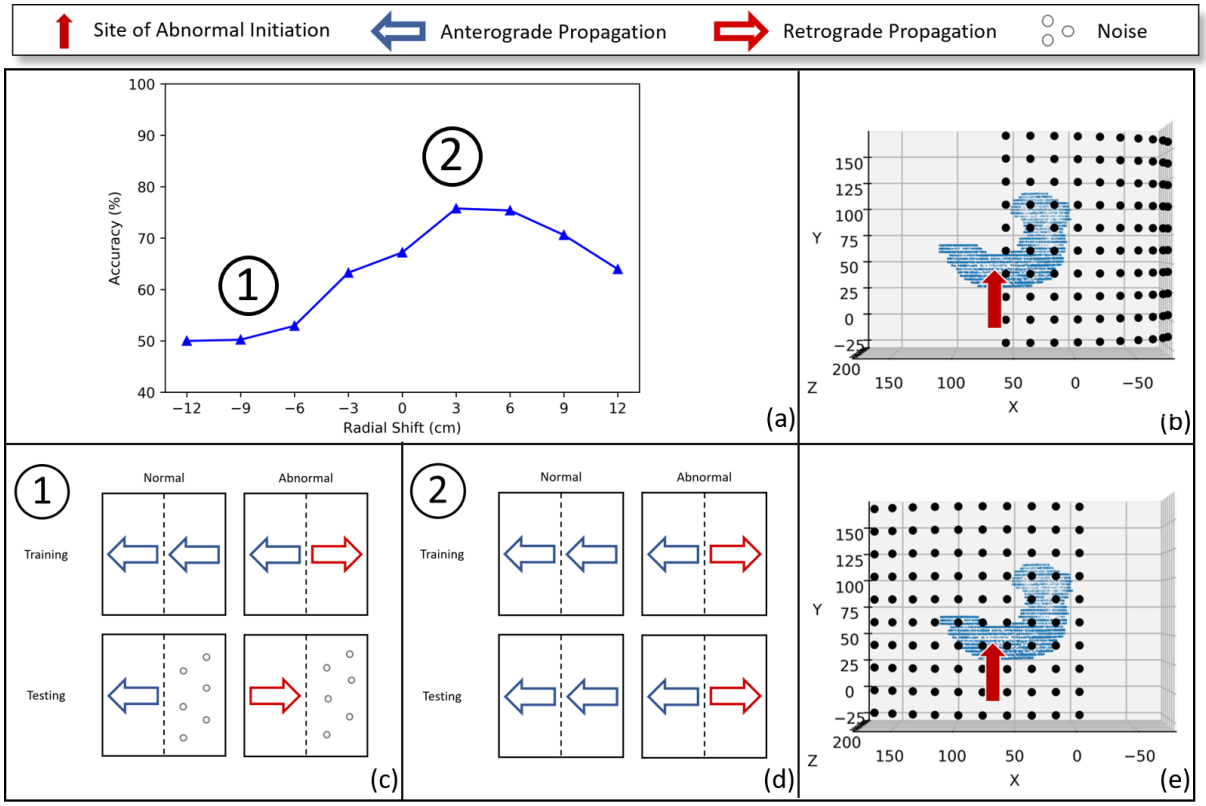

Figure A.9: Vulnerability of the LDA classifier to horizontal electrode array position. a) LDA accuracy as a function of horizontal electrode array shifts. b) Front-facing view of electrode array horizontally shifted by -9 cm. The electrode array is to the right of the abnormal initiation, with rightmost electrodes out of range of electrical activity–resulting in electrodes picking up primarily noise. Panels c-d show training (top) and testing (bottom) abstractions of direction features, with: c) annotation 1 and d) annotation 2. All test sets were trained on the same training set, so abstractions of learned direction features during training are the same for both annotations. Direction features corresponding to anterograde propagation are shown in blue, retrograde propagation in red, and noise-dominated measurements depicted as gray dots. e) Front-facing view of electrode array horizontally shifted by 3 cm. The electrode array is horizontally centered over the abnormal initiation.

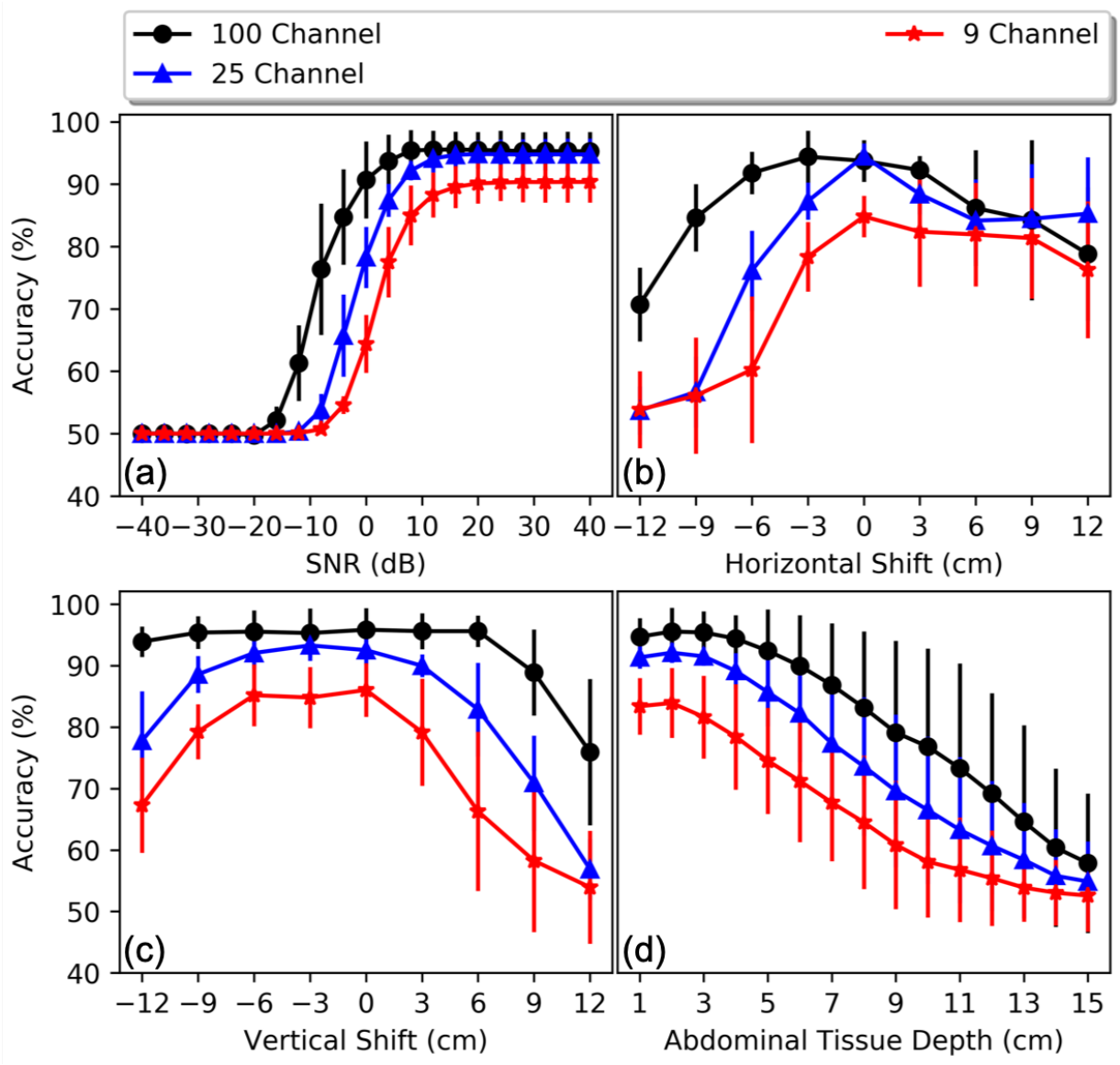

Figure A.10: Comparison of neural network accuracy between 100 channel (black), 25 channel (blue), and 9 channel (red) electrode arrays with respect to: a) SNR, b) horizontal electrode array shifts, c) vertical electrode array shifts, and d) abdominal tissue depth.

Appendix B

Appendix to Chapter 3

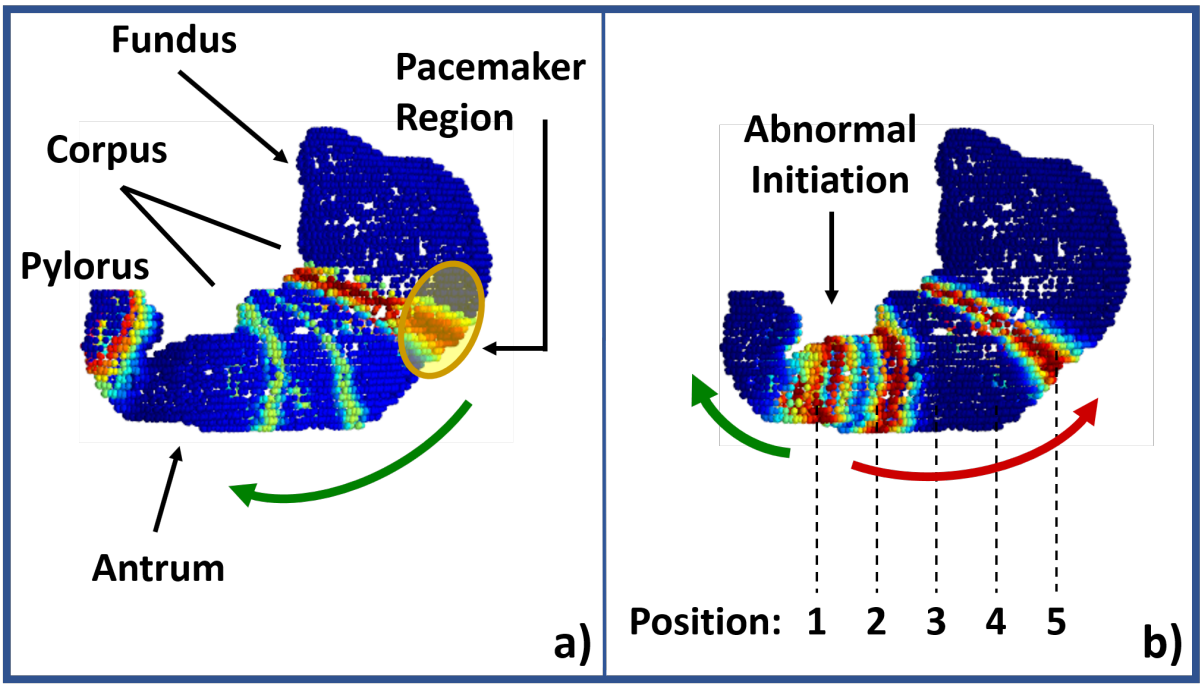

Figure B.1: a) Overview of stomach anatomy and normal gastric slow wave propagation. Slow waves are initiated in the pacemaker region of the corpus (yellow) and propagate in an anterograde direction toward the pylorus (green arrow). b) Overview of abnormal initiation and propagation, as well as experimental design in Sec. 3.2.2. Slow waves are initiated outside of the pacemaker region (i.e. the antrum) and propagate simultaneously in a retrograde (red arrow) and anterograde (green arrow) direction. In section 3.2.2, we discuss simulating the abnormal initiation in five locations: Positions 1-5, shown above. Both a) and b) are taken from the slow wave simulation at one instance in time. Slow waves propagate in equipotential rings (seen in both a) and b)).

Figure B.2: Schematic of methods used to simulate recordings with different start times, as described in Section 3.2.3. Each new start time, T^*, was drawn at random from a uniform distribution between 0 and 300.

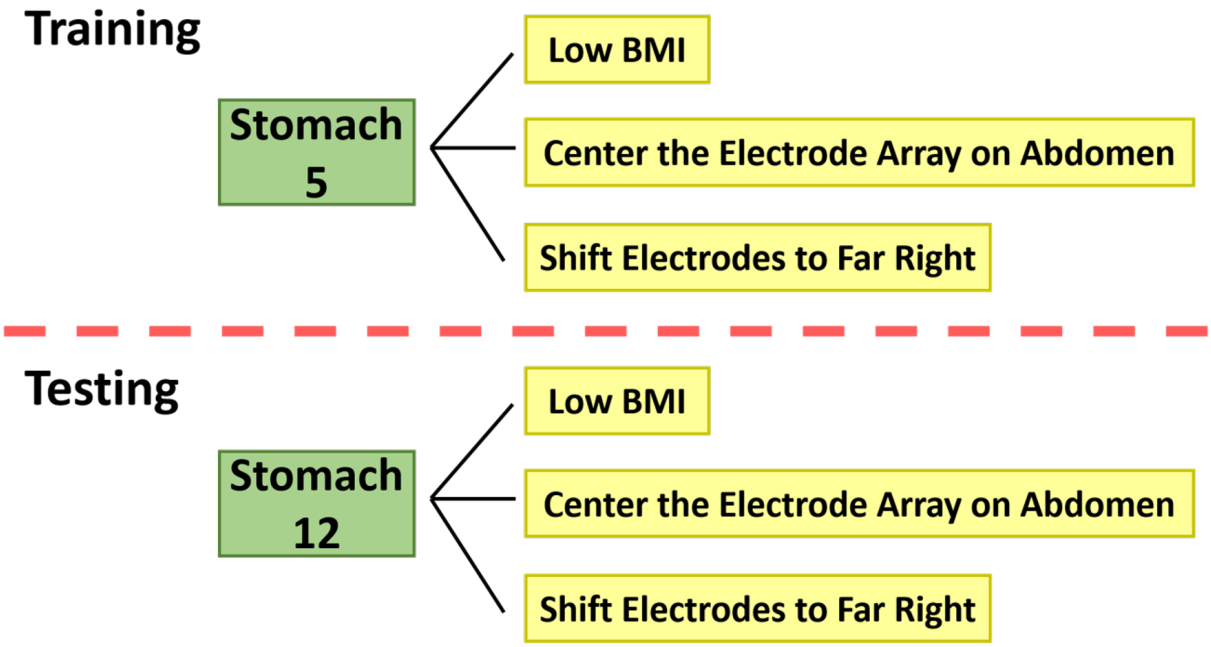

Figure B.3: Schematic of methods used to simulate HR-EGG recordings, described in detail in Section 3.2.4. Simulation of slow waves on one stomach model, i.e. 'Stomach 5,' was used to generate multiple recordings. Stomach models used to generate recording for training were never used to generate recordings for testing, as to ensure independence between training and testing sets.

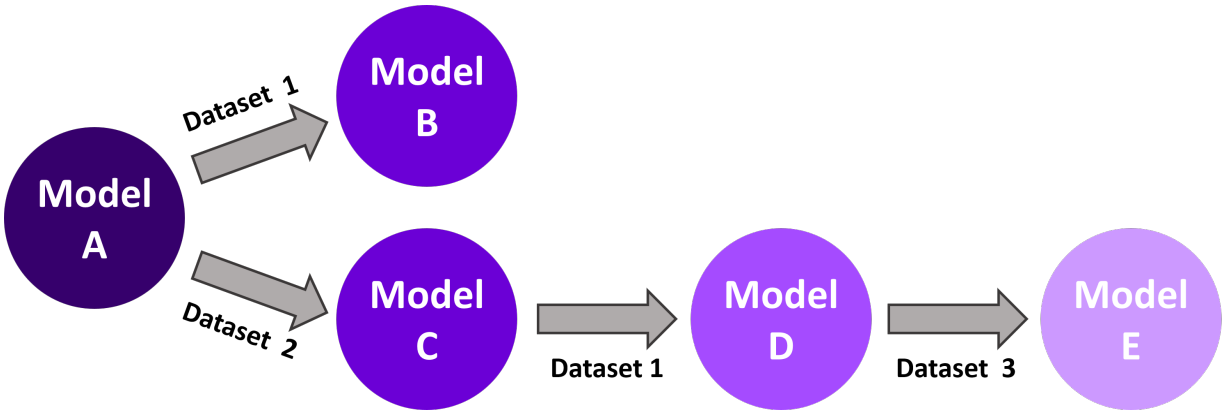

Figure B.4: Transfer learning framework described in Sec. 3.2.5. Model A was developed in [7] (trained and validated with Dataset 0). Each model (B-E) underwent transfer learning of weights from the model that precedes it. Models C-E were developed via sequential transfer learning.

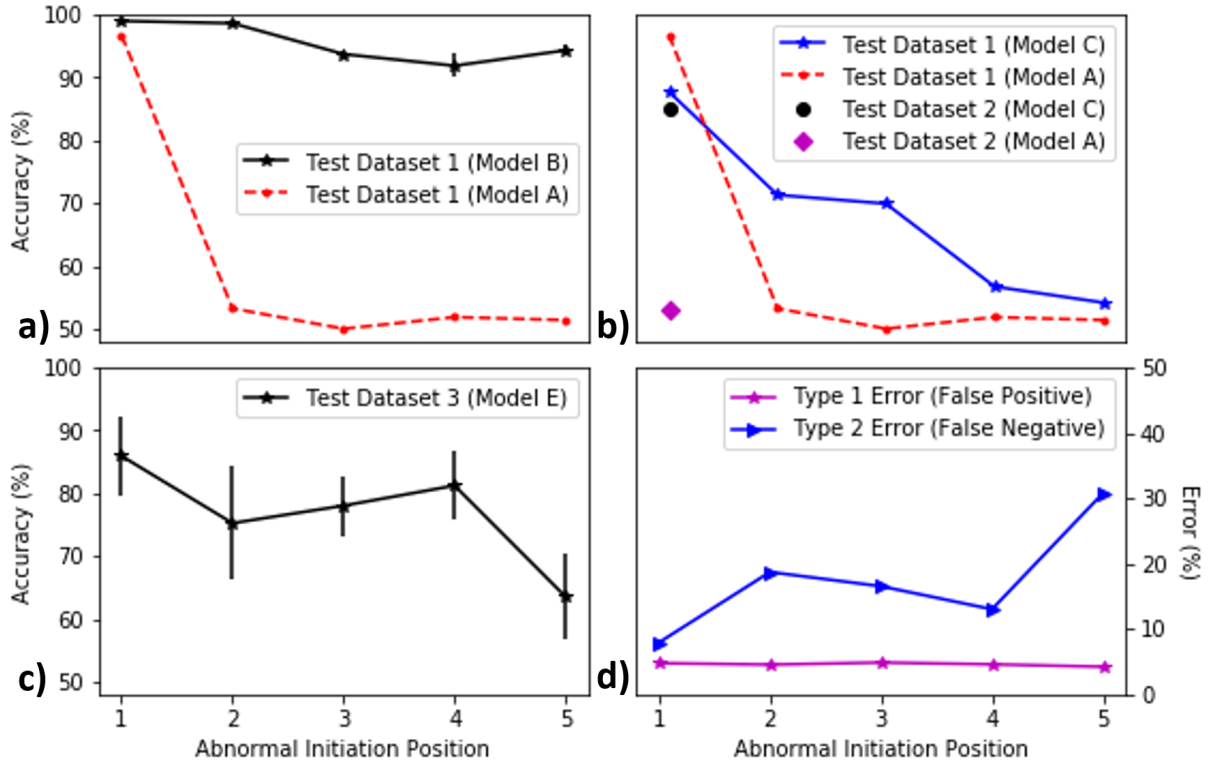

Figure B.5: Results of models generated in this work, shown in comparison to Model A from [7]. Panels a-b show results of models developed with only one type of data heterogeneity. Panels c-d show results of sequential transfer learning and resulting error probability. Results of Models B and E (shown in a) and c)) display the mean performance over the five iterations of training and testing as described in Sec. 3.2.5, with error bars indicating variability.

Appendix C

Appendix to Chapter 4

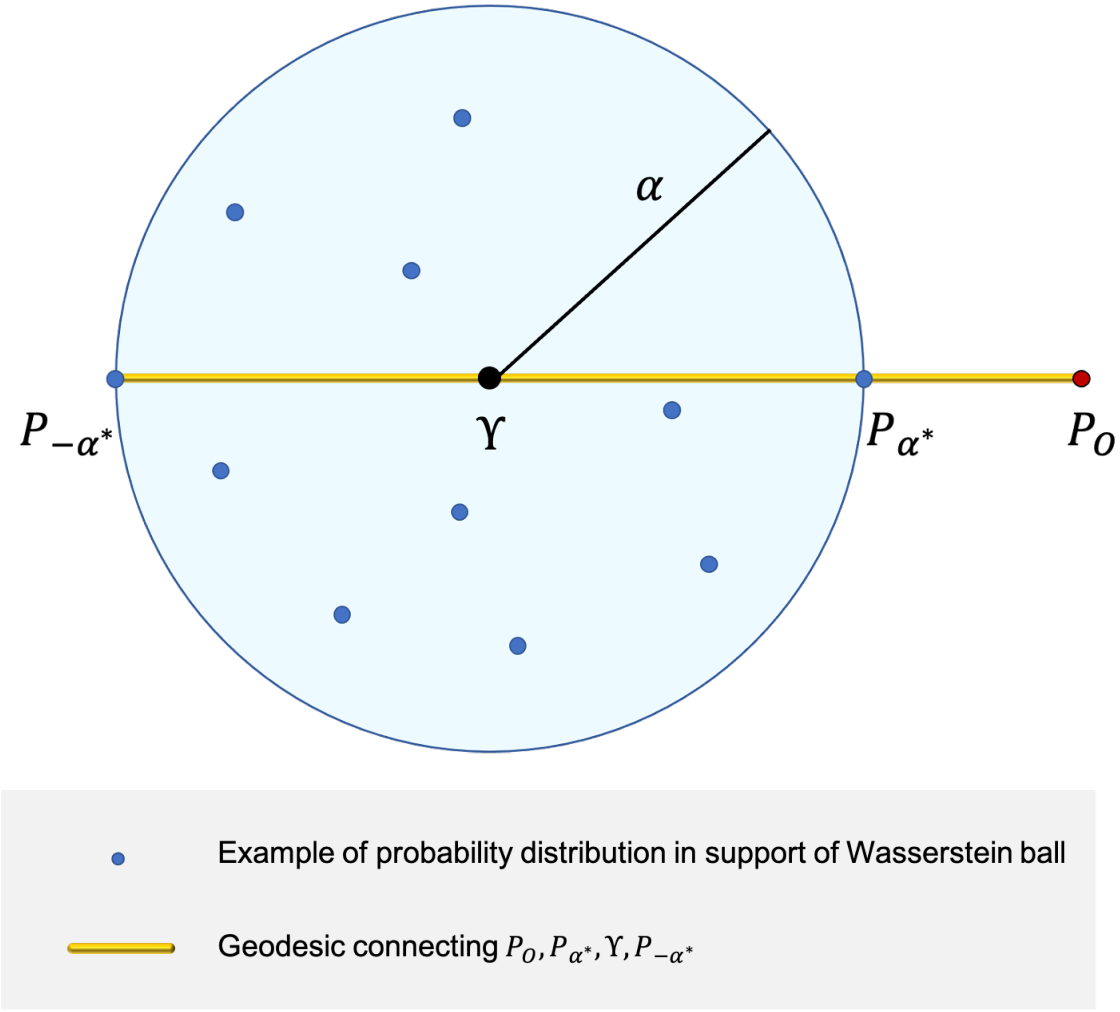

Figure C.1: Wasserstein ball $B(\Upsilon, \alpha)$, as shown in (4.1).

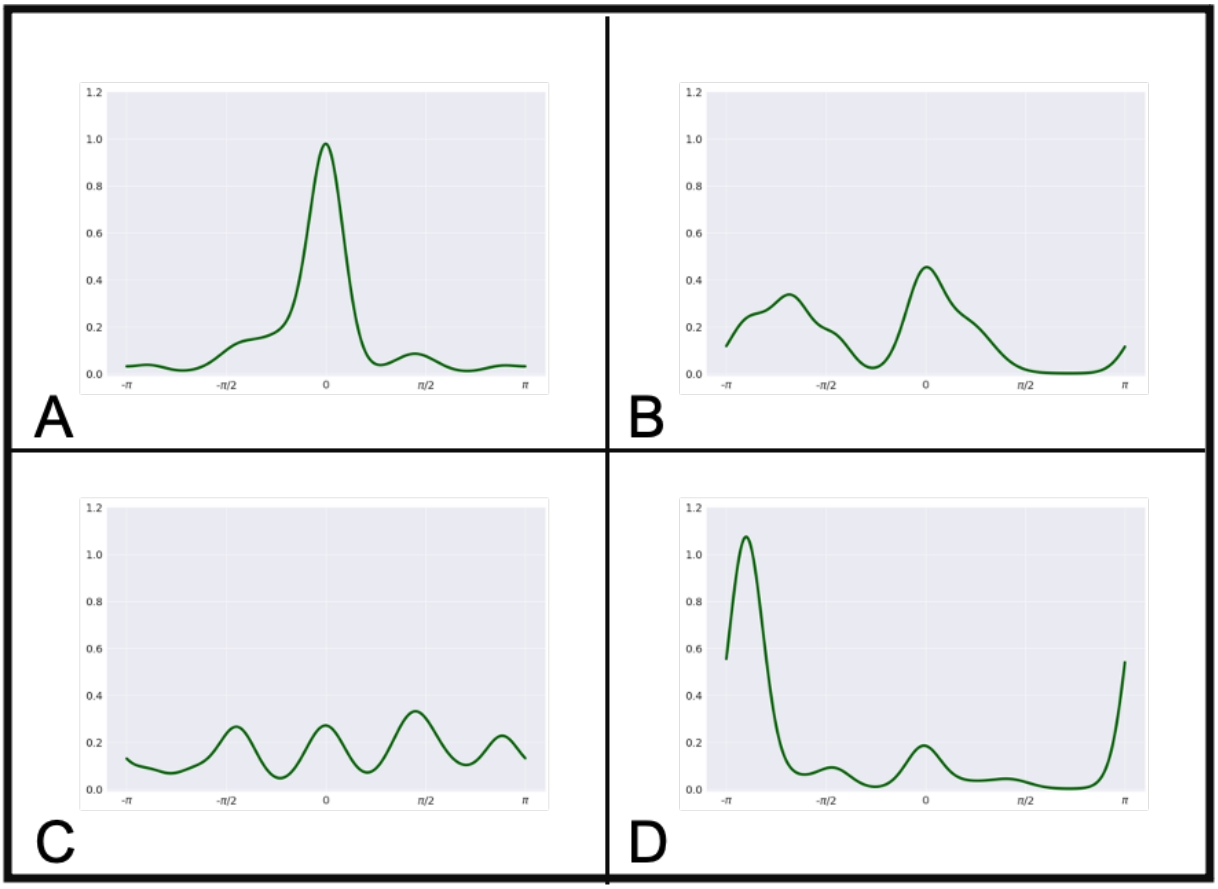

Figure C.2: Classes of probability distributions estimated from measured wave directions. A) Physiologically indicative of healthy wave propagation toward the intestines. B) Physiologically indicative of bifurcated wave propagation toward the intestines and esophagus. C) Physiologically indicative of stochastic wave propagation. D) Anomaly probability distribution. Physiologically indicative of wave propagation primarily toward esophagus.

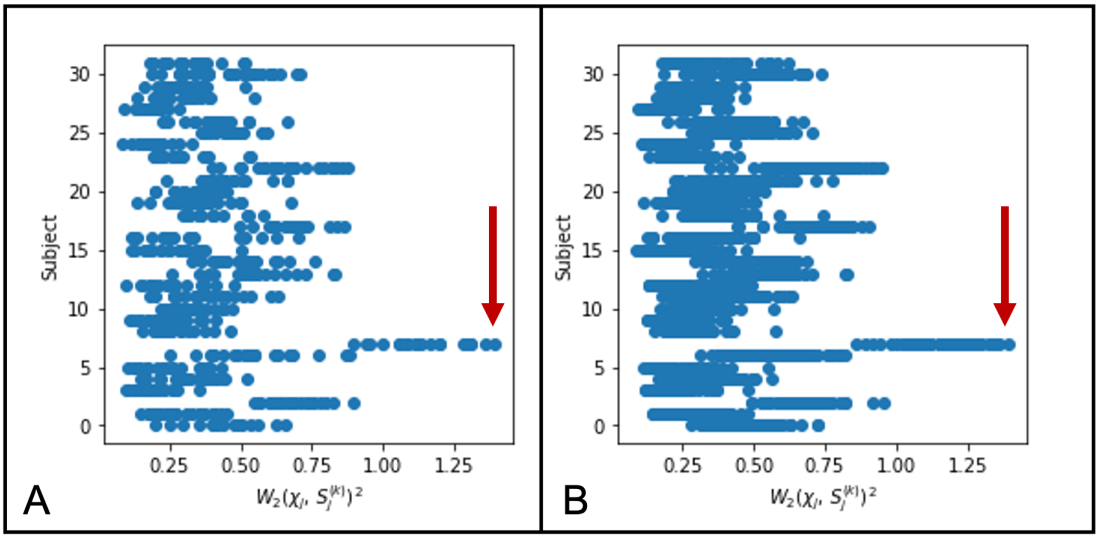

Figure C.3: Wasserstein distances between probability distributions and respective subset barycenters computed in Algorithm 1, run for A)100 iterations and B) 300 iterations. Shown for each human subject.

Appendix D

Appendix to Chapter 5

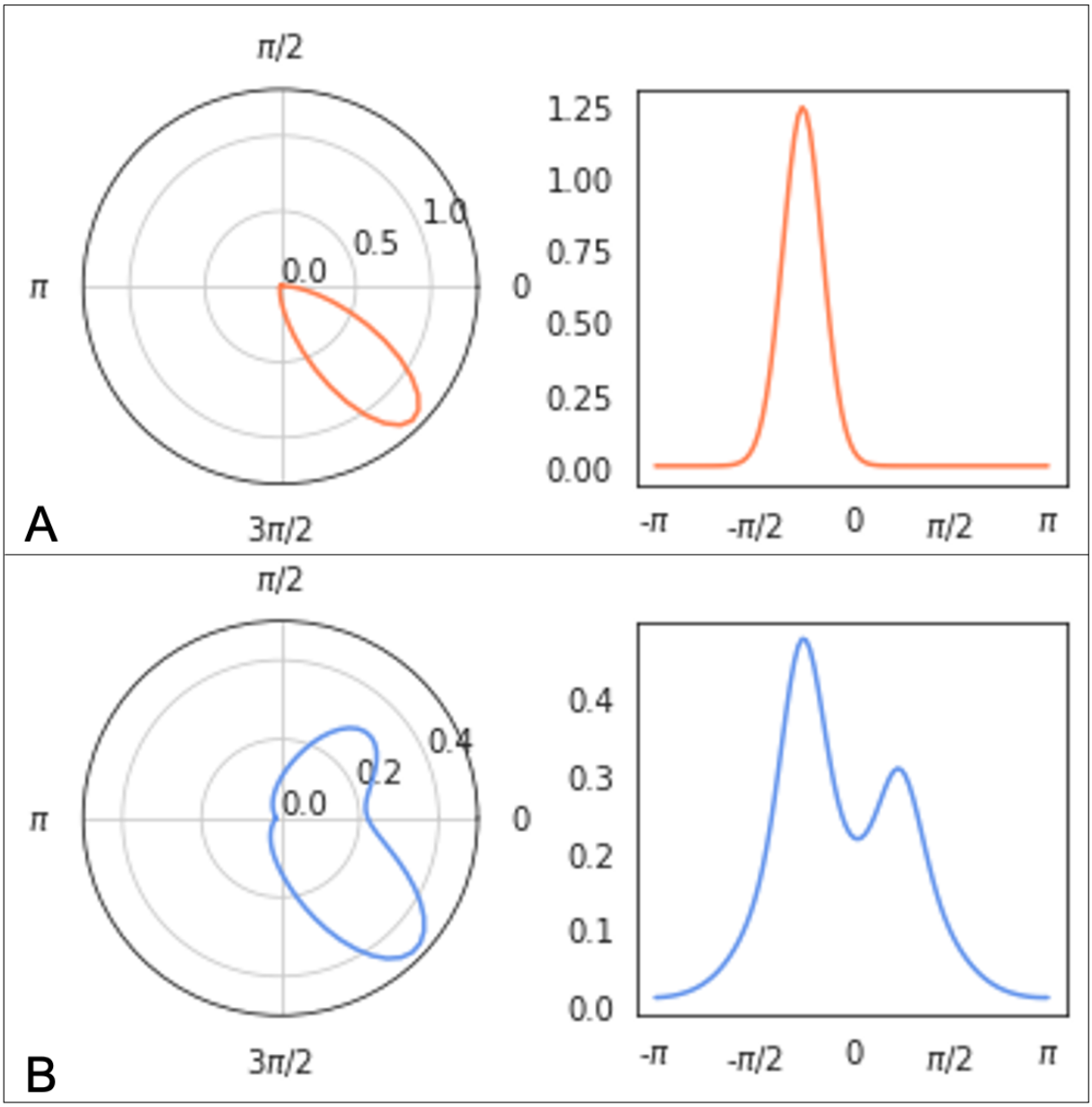

Figure D.1: A) Probability density function of a unimodal von Mises distribution, given by (5.5). b) von Mises mixture of two distributions, given by (5.6). Both the unimodal and multimodal distributions are shown in polar (left) and Euclidean (right) axes.

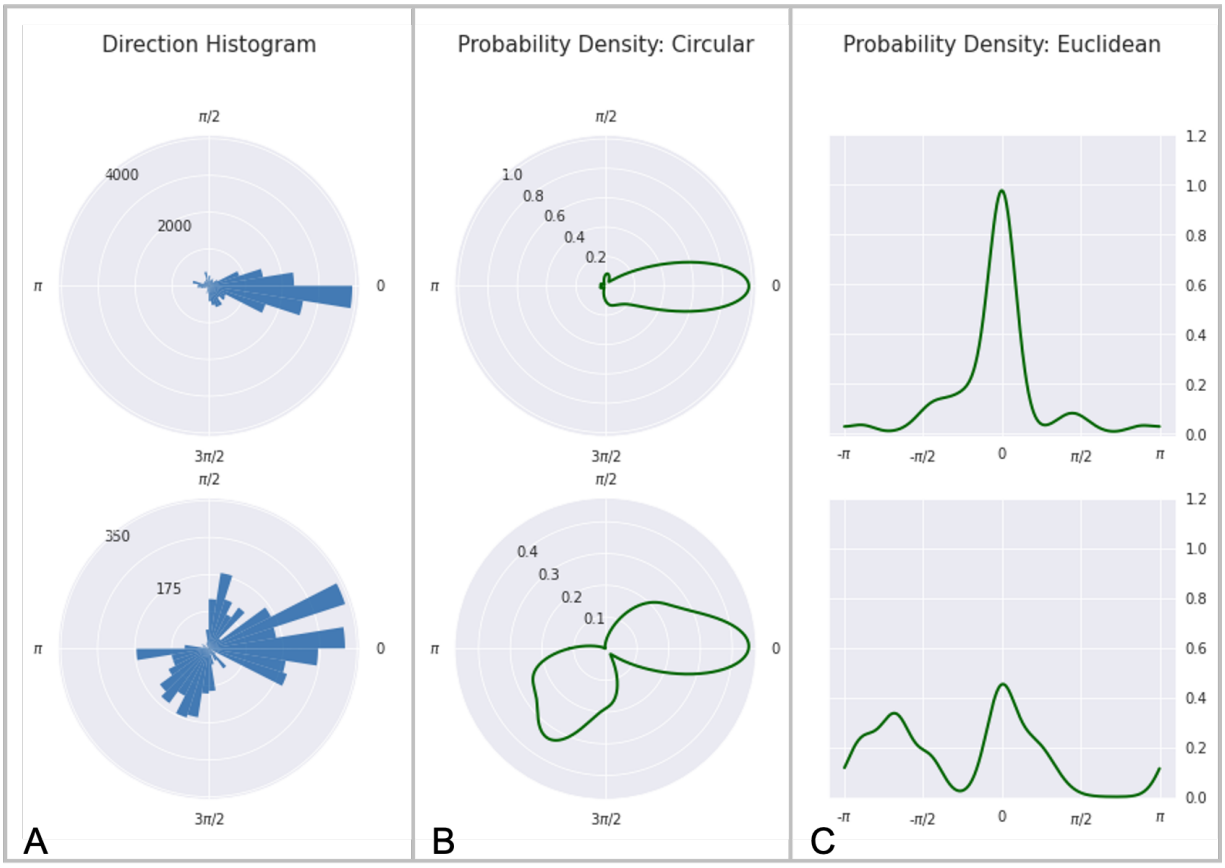

Figure D.2: A) Histograms of wave directions, $\mathcal{D}^{(s)}$. B) Estimated probability density function, from methods described in Section 5.2.4, shown in polar axes. C) Estimated probability density function, from methods described in Section 5.2.4, shown in Euclidean axes. Panels A-C are shown for a healthy control subject with normal slow wave physiology (top) and a symptomatic patient with indications of abnormal gastric slow wave function (bottom).

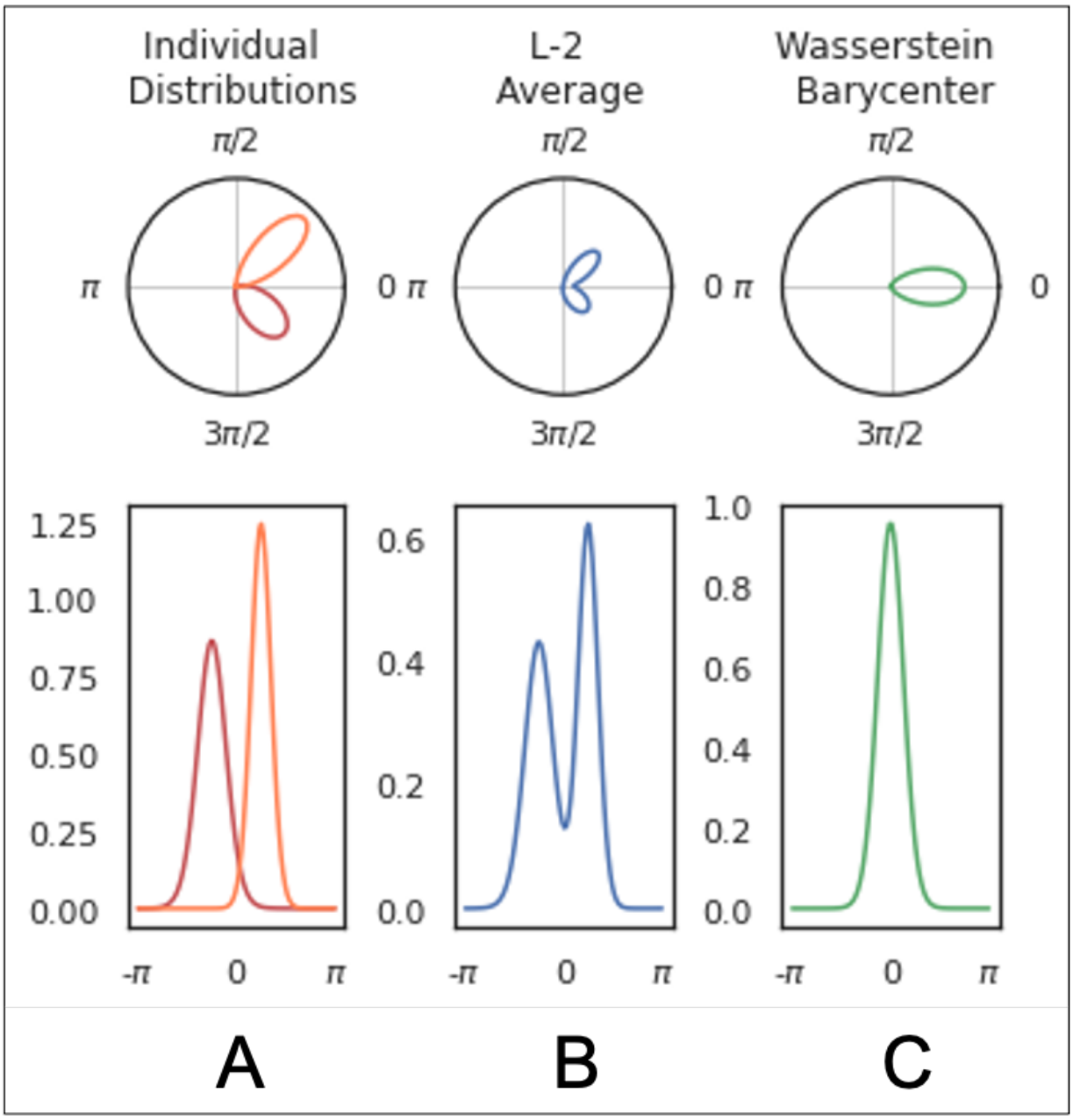

Figure D.3: Example highlighting the importance of using a Wasserstein barycenter to represent a model of healthy controls in Section 5.2.8. A) Two unimodal von Mises probability distributions. B) The L-2 average of the two distributions is bi-modal, with one mode centered at each individual distribution's mean. C) The Wasseerstein barycenter between the two distributions is unimodal and therefore better represents the structure of both of the original distributions. Panels A-C are shown in uniform-radius polar (top) and Euclidean (bottom) axes.

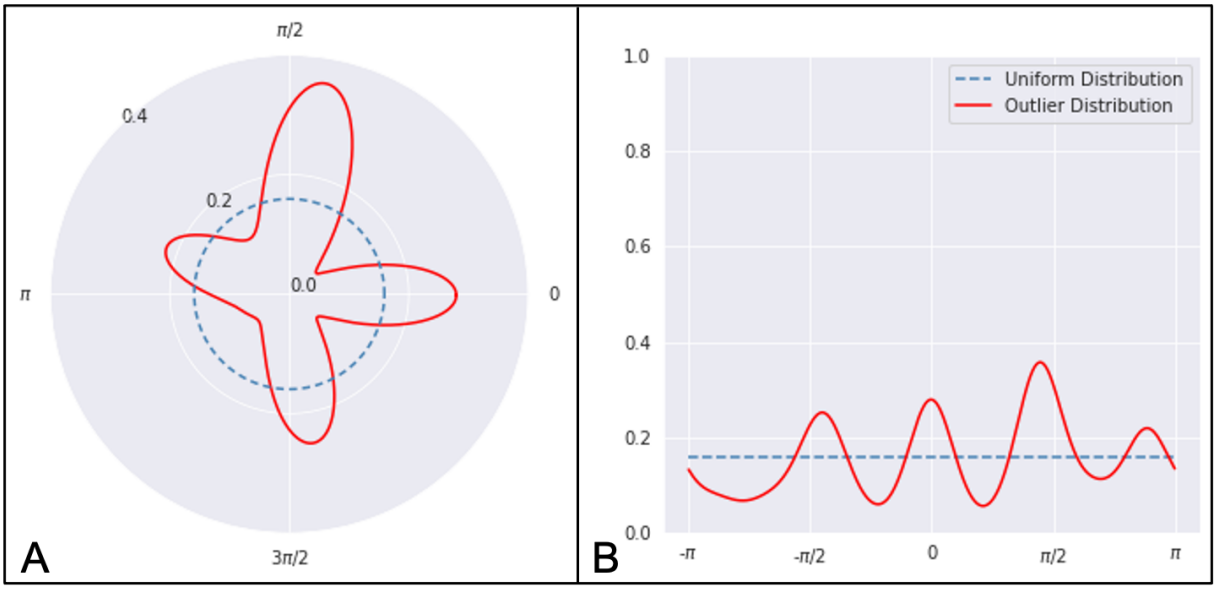

Figure D.4: Probability density function (found in Section 5.2.4) of outlier within the control group, with uniform distribution along the support $[-\pi, \pi]$ for reference. The outlier distribution lacks the high-energy mode near zero that is expected of healthy controls exhibiting primarily anterograde wave propagation (i.e. as seen in Fig. D.2C), and instead exhibits overall structure similar to a uniform distribution. Shown in A) polar axis and B) Euclidean axis.

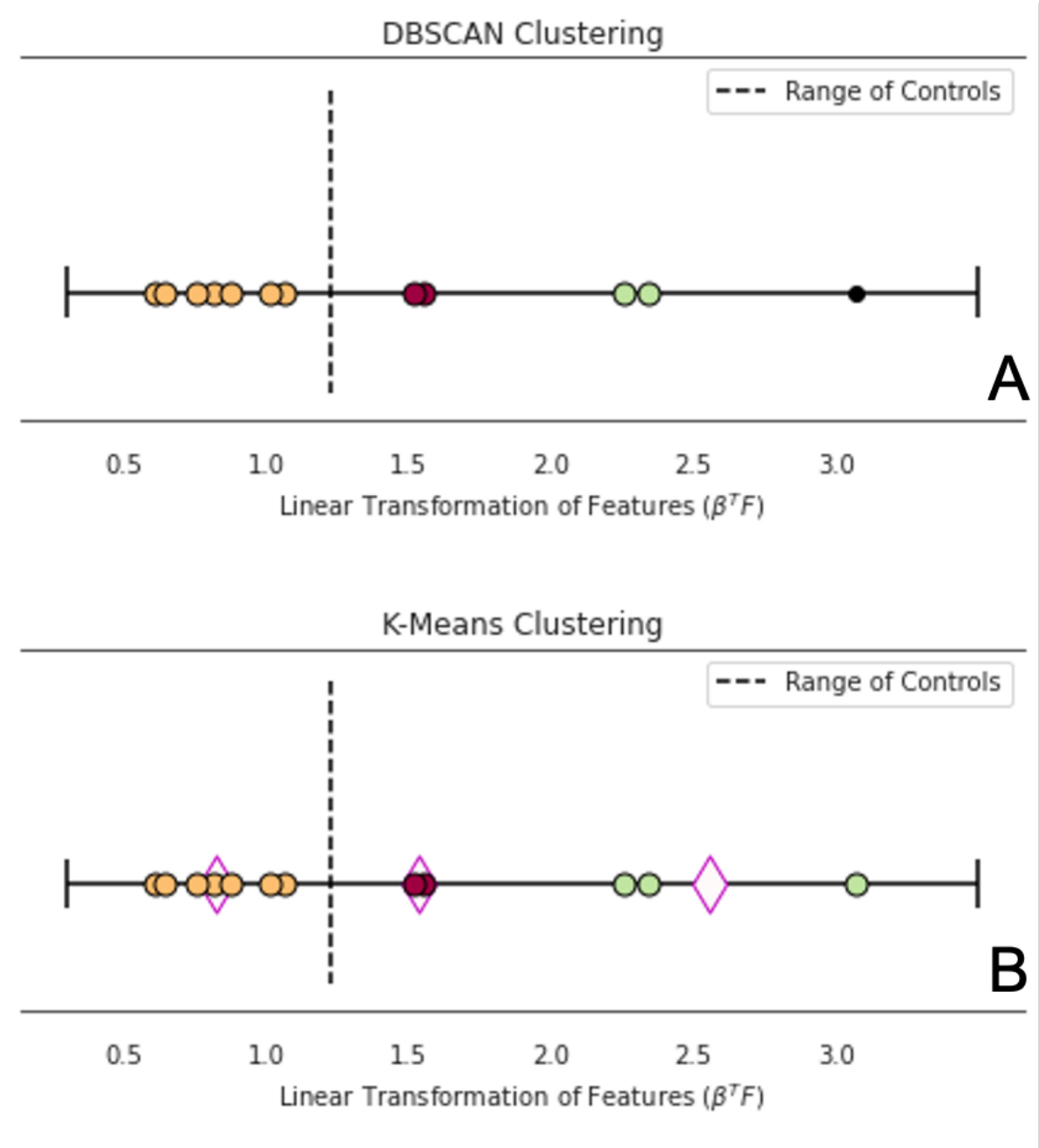

Figure D.5: DBSCAN (A) and K-Means (B) clustering of linear transformation of features. K-Means centers plotted in magenta (diamonds). Both clustering algorithms identify the same anomalous class (orange). The upper limit of transformed features within the control group (outlier excluded) is plotted as a dashed line to show that all anomalous class features fit within a boundary established by controls.

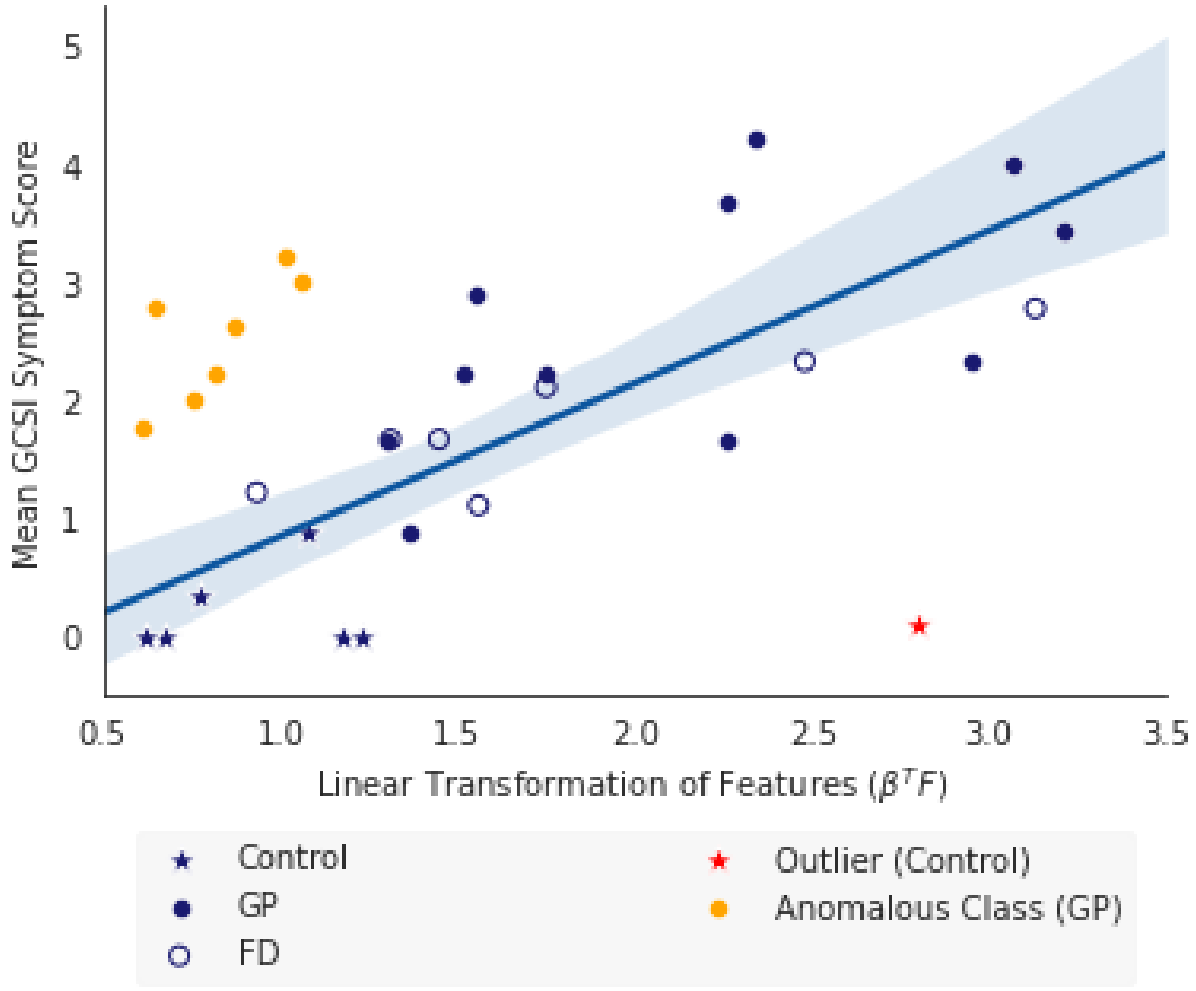

Figure D.6: Identification of anomalous class (orange) for which we hypothesize that the underlying GP/FD etiology was not spatial gastric slow wave abnormalities. Regression with non-anomalous data (blue). We show 95% confidence interval (shaded) and report r=0.8. Outlier in the control group is shown in red but was not used in regression.

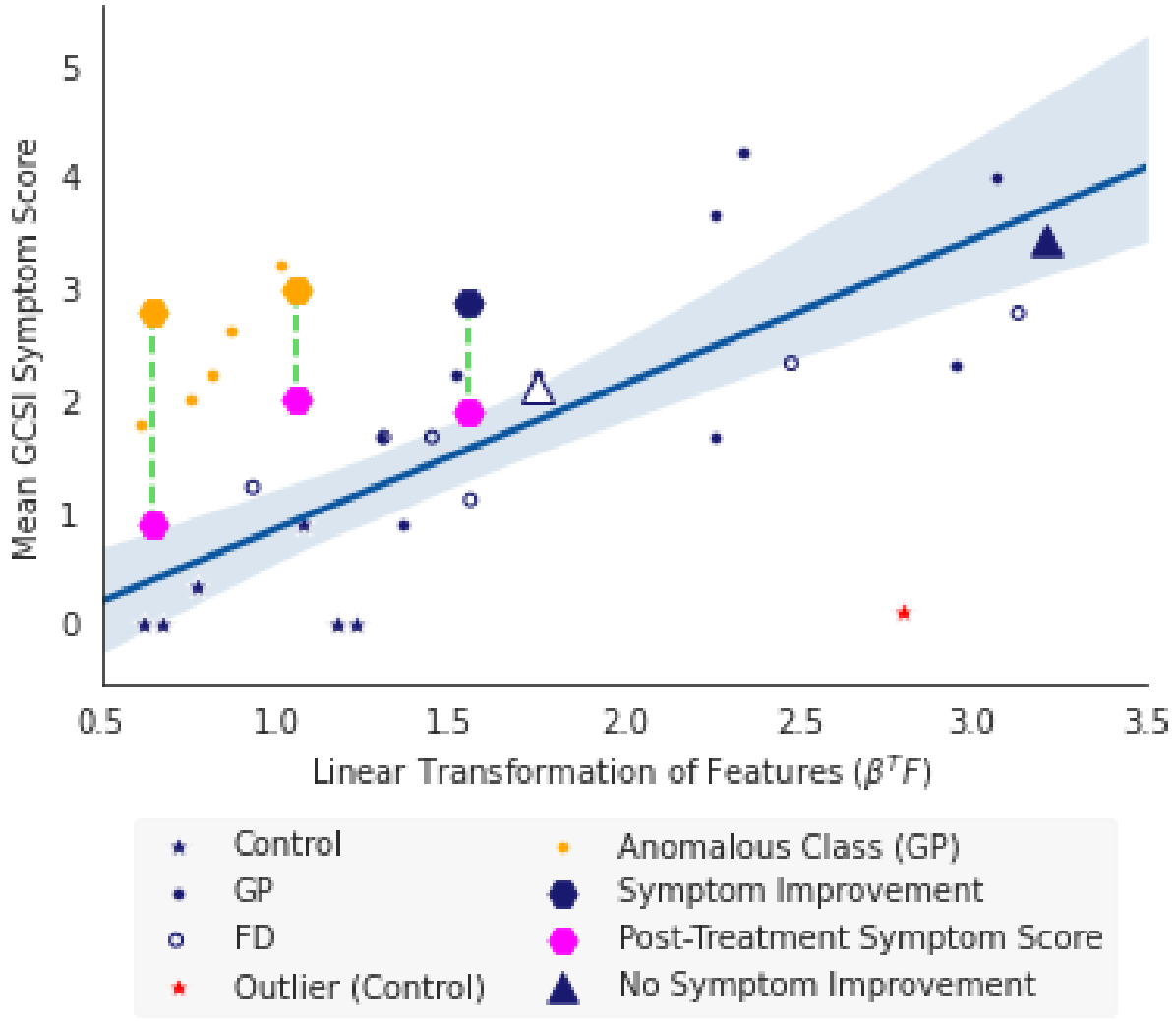

Figure D.7: Symptom progression after pyloric and prokinetic interventions (n=5). These interventions did not directly target spatial gastric slow wave abnormalities. Three patients reported improvement of symptoms (marked with a hexagon), and two patients saw no improvement (marked with a triangle). Patients who did not improve after treatment were in close proximity to the regrssion line, whereas patients who did improve were far above the regression line, and have post-treatment scores that lie closer to the regression line.

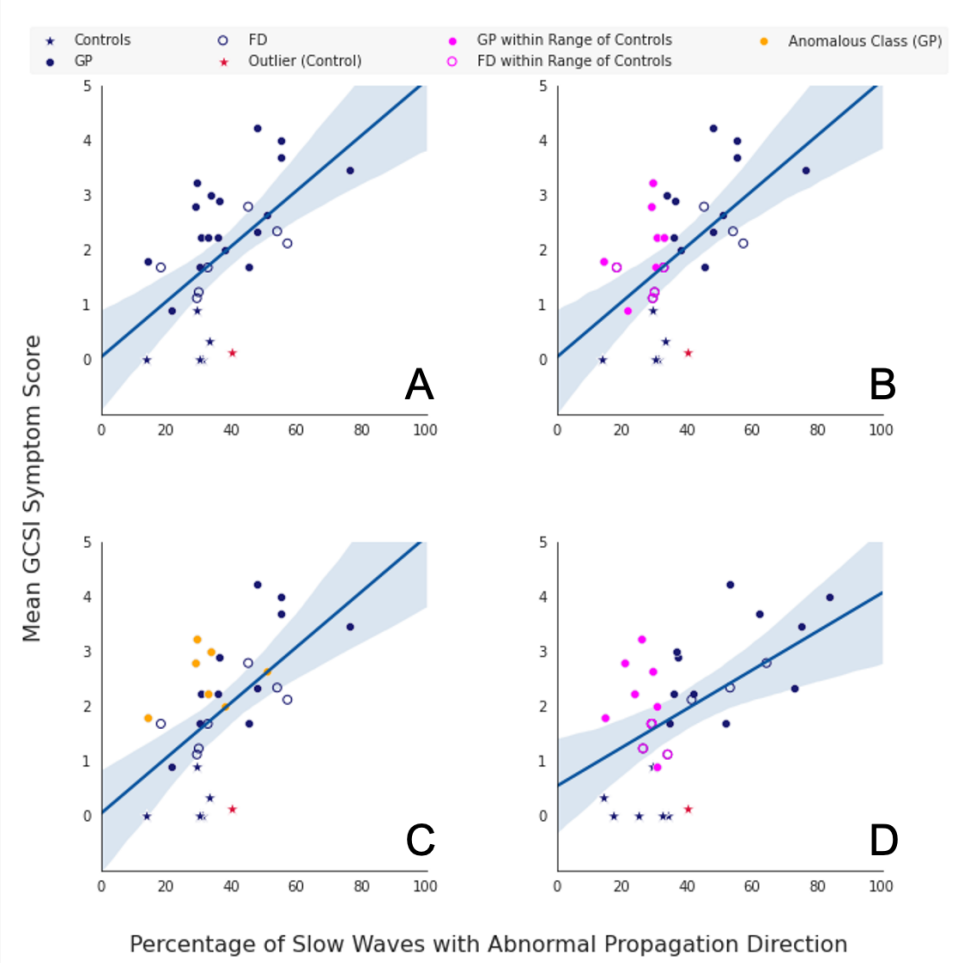

Figure D.8: Linear least squares (LLSQ) regression of percentage of slow waves propagating in an abnormal direction with mean GCSI symptom score (shown as blue line in panels A-D with 95% confidence interval shaded in lighter blue). Percentage of slow waves exhibiting abnormal propagation was found by integrating each probability distribution on the interval $[-\pi/3, \pi/3]$ (same interval as in [53]). Panels A-D serve as a comparison to results published in [53]. Panels A-C are computed from a 90 minute recording, with approx. 30 minutes preprandial and approx. 60 minutes postprandial. (A) We calculated an r value of 0.56 if all 32 subjects were included (consistent with [53]) in the regression and an r value of 0.59 if control outlier was excluded. (B) Pink data show GP and FD subjects whose percentage of slow waves overlap with the range of controls (excluding the outlier). (C) Orange data show the seven subjects labeled as "anomalous" from the robust regression framework presented here. It is hypothesized that these subjects have GP/FD arising from an etiology distinct from gastric myoelectric dysfunction. (D) Percentage of slow waves propagating in an abnormal direction was calculated for only postprandial recording and subsequent regression was performed. Analogous to panel B, pink data show GP and FD subjects whose percentage of slow waves overlap with the range of controls (excluding the outlier).

Appendix E

Supplemental Materials for Chapter 2

E.1 Branched Geometric Skeleton of the Stomach

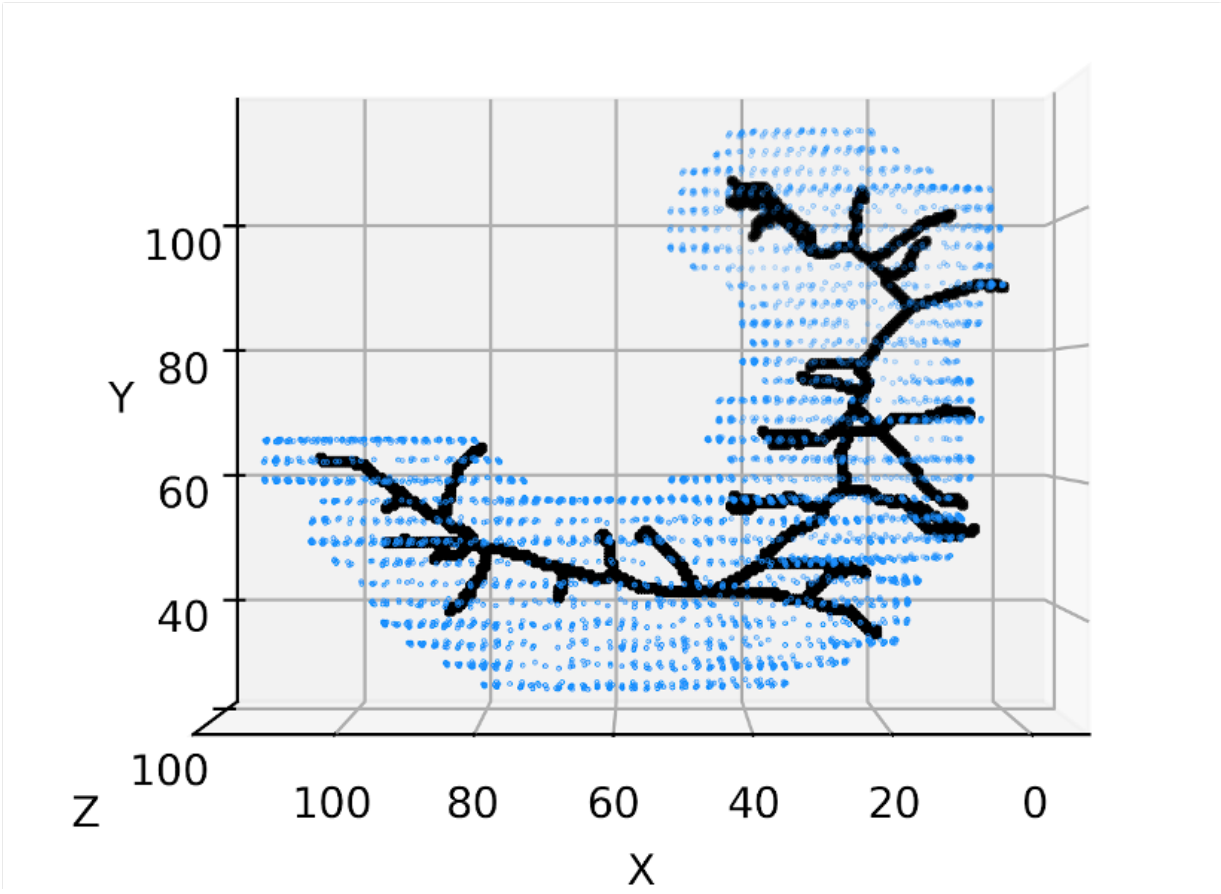

Figure E.1: Geometric Skeleton (black) plotted on the stomach (blue) as a result of iterative thinning described in Section II-A1. The spine of this skeleton best approximates the characteristic shape of the stomach. Units in mm.

E.2 All Datasets Included in Training and Testing Sets

Table E.1: The training set was created with datasets from all combinations in each row.

Horizontal Shift (cm)	Vertical Shift (cm)	Abdominal Tissue Depth (cm)
-12, -9, -6, -3, 0, 3, 6, 9, 12	-3, 0, 3	1
-3, 0, 3	-12, -9, -6, -3, 0, 3, 6, 9, 12	1
-3, 3	-3, 3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
0	0	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Table E.2: Testing datasets used to determine classification performance in response to modulations of each 'Test Variable' were created with datasets from all combinations in respective rows.

Test Variable	Horizontal Shift (cm)	Vertical Shift (cm)	Abdominal Tissue Depth (cm)
SNR	-3, 0, 3	-3, 0, 3	1,2
Horizontal Shift	-12, -9, -6, -3, 0, 3, 6, 9, 12	-3, 0, 3	1,2
Vertical Shift	-3, 0, 3	-12, -9, -6, -3, 0, 3, 6, 9, 12	1,2
Abdominal Tissue Depth	-3, 0, 3	-3, 0, 3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

E.3 'SNR Gain' Seen by Use of the Convolutional Neural Network Methodology as Opposed to the Linear Discriminant Analysis of Spatial Features

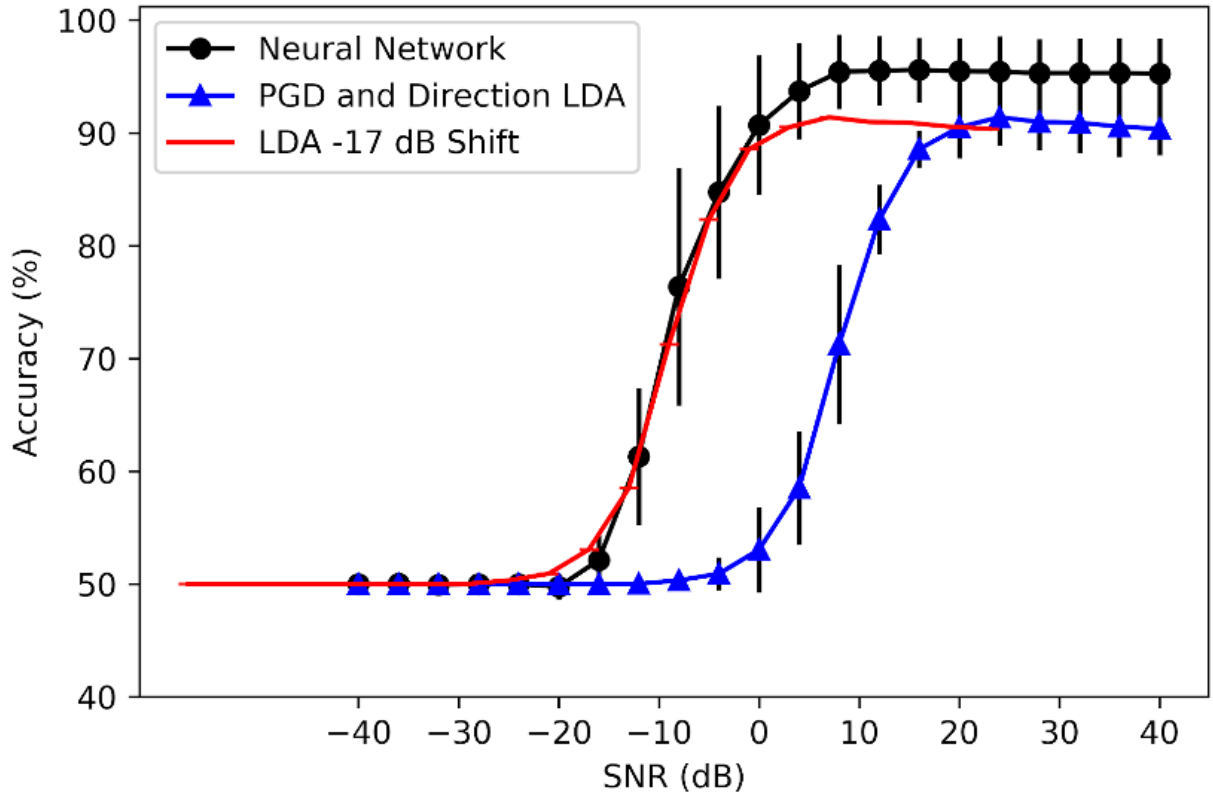

Figure E.2: Classification Accuracy as a function of SNR. Accuracy of linear discriminant analysis on spatial wave propagation features (red) is shifted to the left by -17 dB. This shift of the curve causes the linear discriminant analysis' performance to more closely resemble that of the convolutional neural network (black). Thus, the 'SNR gain' seen by use of convolutional neural network methodology is 17 dB.

E.4 Trends in Classification Accuracy as a Function of Vertical and Abdominal Tissue Depth Perturbations Explained

The convolutional neural network (CNN) was robust to vertical electrode array shifts (Fig. 8c). A decline in classification performance is observed when the electrode array is shifted more than 9 cm vertically. Due to the general 'J' shape of the stomach, 9 and 12 cm vertical shifts upward will move the electrode array such that it is no longer over the stomach– and thus out of range of electrical activity (Fig. S-3). There is also the highest accuracy variability at these two shift configurations, likely induced by stomachs that do not exhibit the general 'J' shape, along with the high amount of stomach size variability. Stomach geometries with a larger pocket in the 'J' shape will produce recordings primarily governed by noise at these electrode configurations. On the contrary, stomach geometries with a fuller top half will have a higher (though still small) percent of electrodes still in range of the electrical activity of the stomach. The linear discriminant analysis (LDA) on spatial features of wave propagation was more vulnerable to vertical electrode array positioning than the CNN (Fig. 8c). Similar to trends in accuracy seen in horizontal shifts (explained in detail in Section III), this vulnerability arose from the electrode array's placement with respect to the site of the abnormal initiation– namely as the electrode array is vertically centered around the abnormal initiation site, the LDA classification accuracy increases. Since the LDA's variability in performance primarily arose from horizontal electrode array positioning, the variance in classification accuracy was fairly consistent across vertical shifts. The variability appears to decrease at shifts of 9 and 12 cm, but this can be attributed to the fact that the majority of electrodes are governed by noise for all stomach geometries at these shifts and the LDA fails entirely to disambiguate normal and abnormal waveforms. Thus, the accuracy is that of 'chance' (0.5) and should not dip below 0.5. Both the CNN and LDA classifiers' performance suffers as the abdominal tissue depth increases (Fig. 8d). As electrodes move farther from serosal slow waves, signal power decreases and as a result, SNR decreases. In the test of classification performance

as a function of SNR (Fig. 8a), it was shown that there is an intermediate range of SNR values for which both the CNN and LDA classifiers were sensitive to perturbations in SNR. Increasing the abdominal tissue depth puts the measurement recordings in a SNR range for which classification accuracy suffers. The threshold in abdominal tissue depth for which the classifier is 'successful' (accuracy $> 50\%$) is lower for the LDA approach than this threshold for the CNN approach. Thus, the CNN methodology is more robust across a range of abdominal tissue depths, suggesting robustness over a wider range of body mass index.

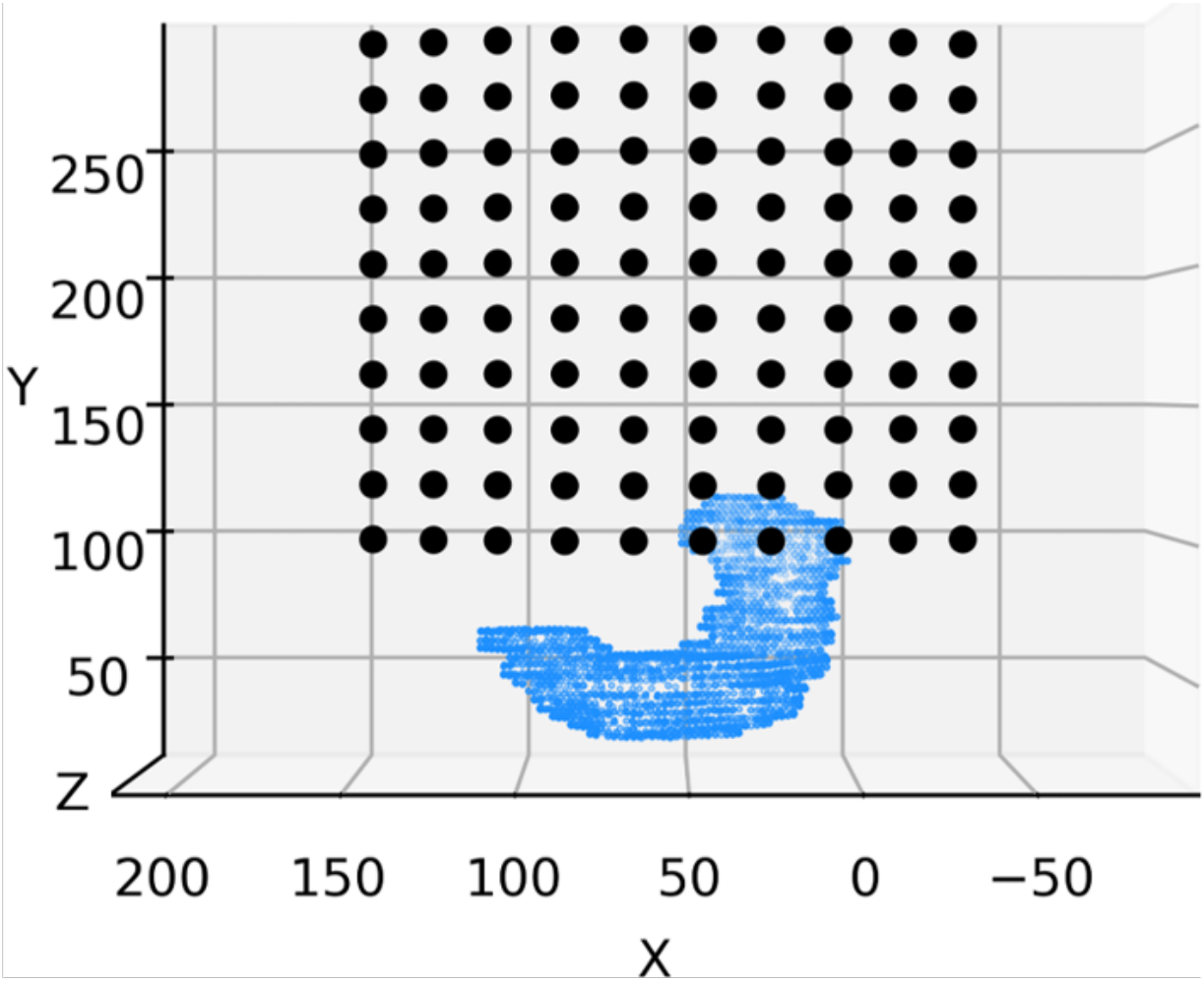

Figure E.3: Electrode array (black) shifted vertically by 12 cm. This generally puts the electrode array out of range of 'J'-shaped stomachs.

E.5 Localization of Electrode Array to Site of Abnormal Initiation

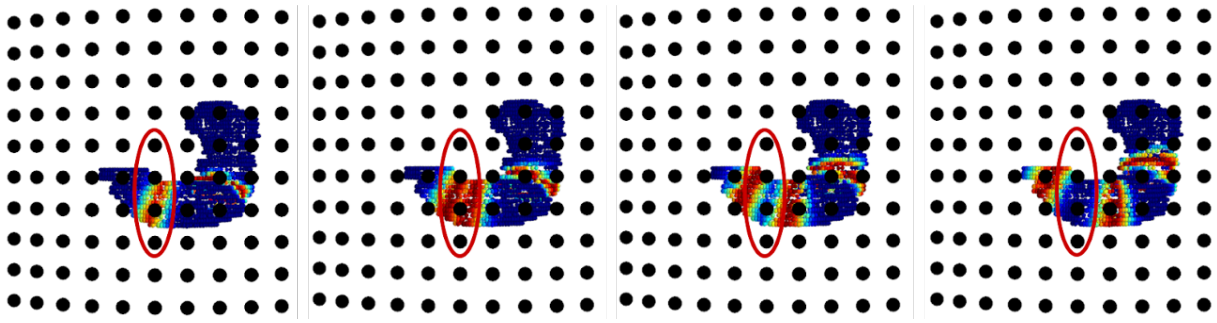

Figure E.4: E Shifting the electrode array horizontally by 3 cm localizes the electrodes around the site of the abnormal initiation (red oval).

E.6 Linear Discriminant Analysis Weights

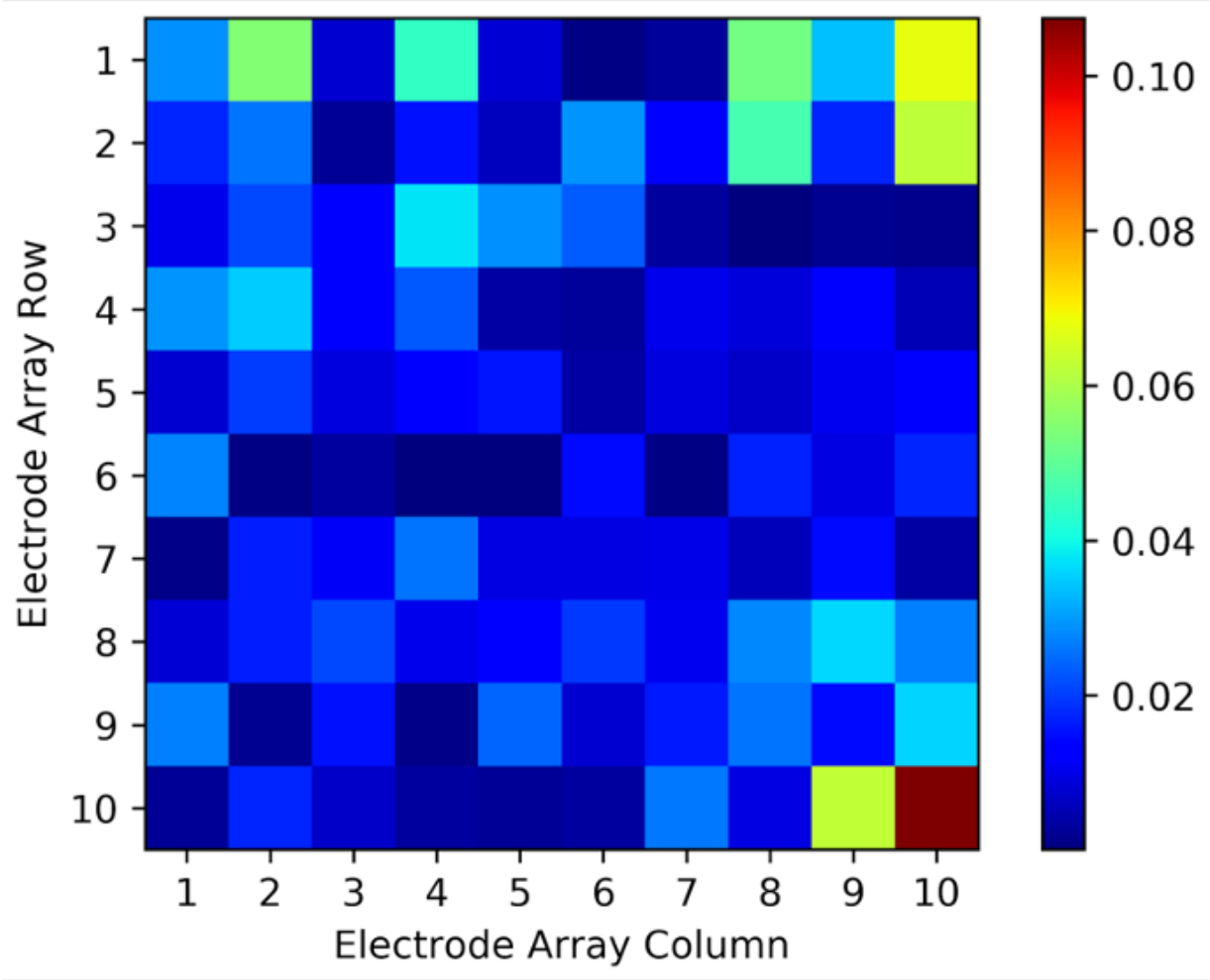

Figure E.5: LDA direction feature weights (magnitude) plotted as a heatmap. Higher intensity (red) indicates that direction features on the right side of the electrode array are weighted higher in classification than direction features on the left side of the array.

E.7 Stomach Size and Shape Variability

Below, each human research subject's stomach model, extracted from their abdominal CT scan, is plotted to show the stomach size and shape variability across the research subjects. Measurements of stomach height, width, and depth are indicated in mm. This supports the claim that simulation datasets that arose from different stomach models are independent. Furthermore, the data used in training the models is independent from data used to test the models.

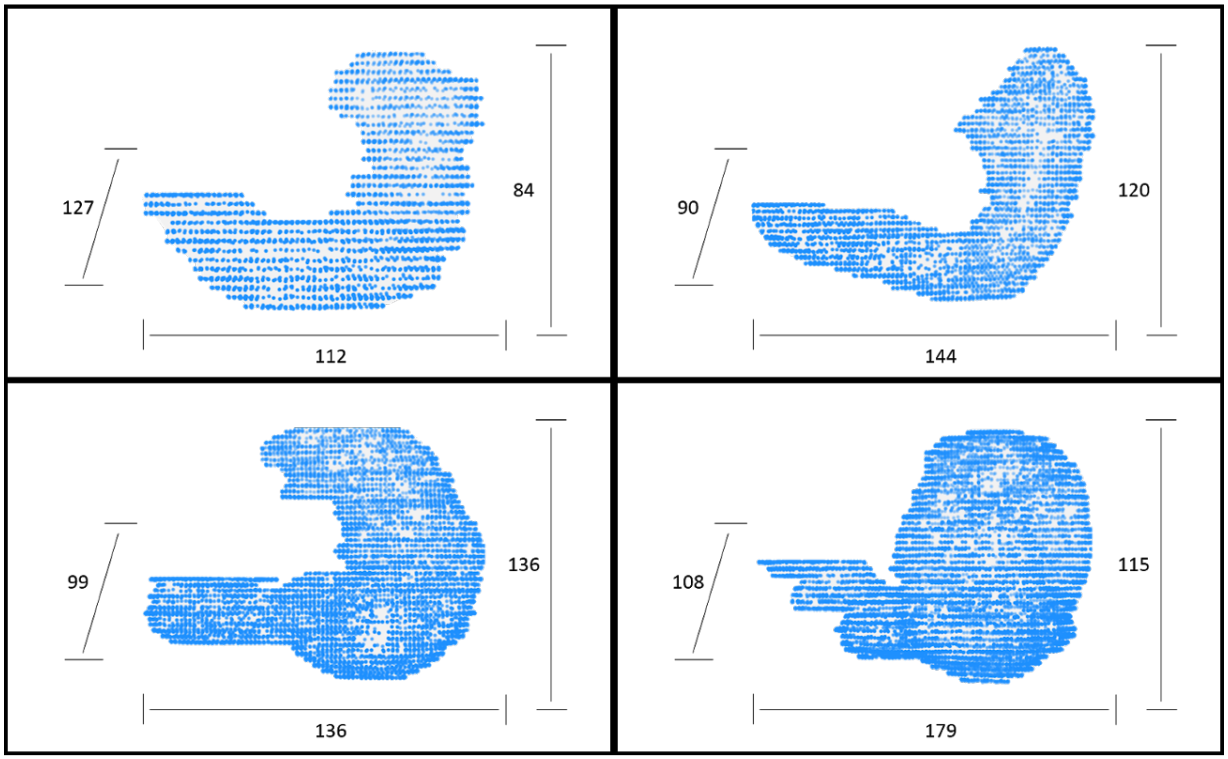

Figure E.6: Subjects 1-4: Stomach reconstruction from CT scan

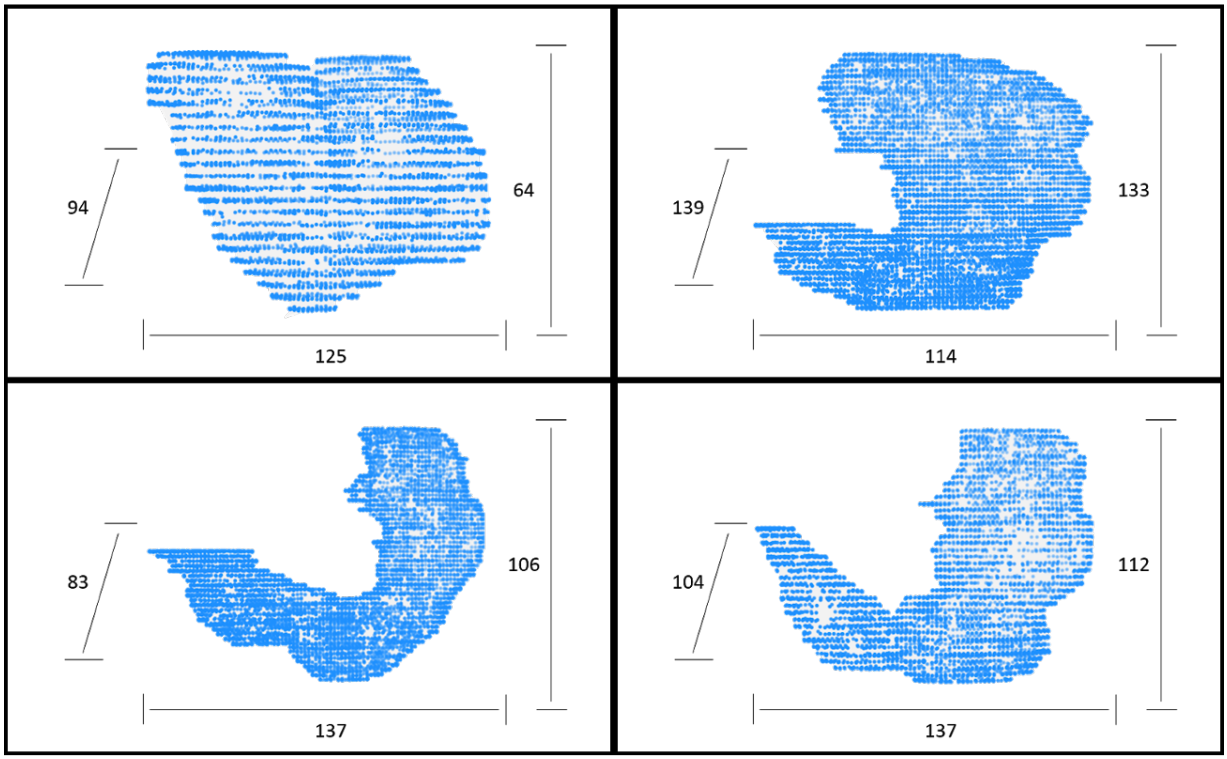

Figure E.7: Subjects 5-8: Stomach reconstruction from CT scan

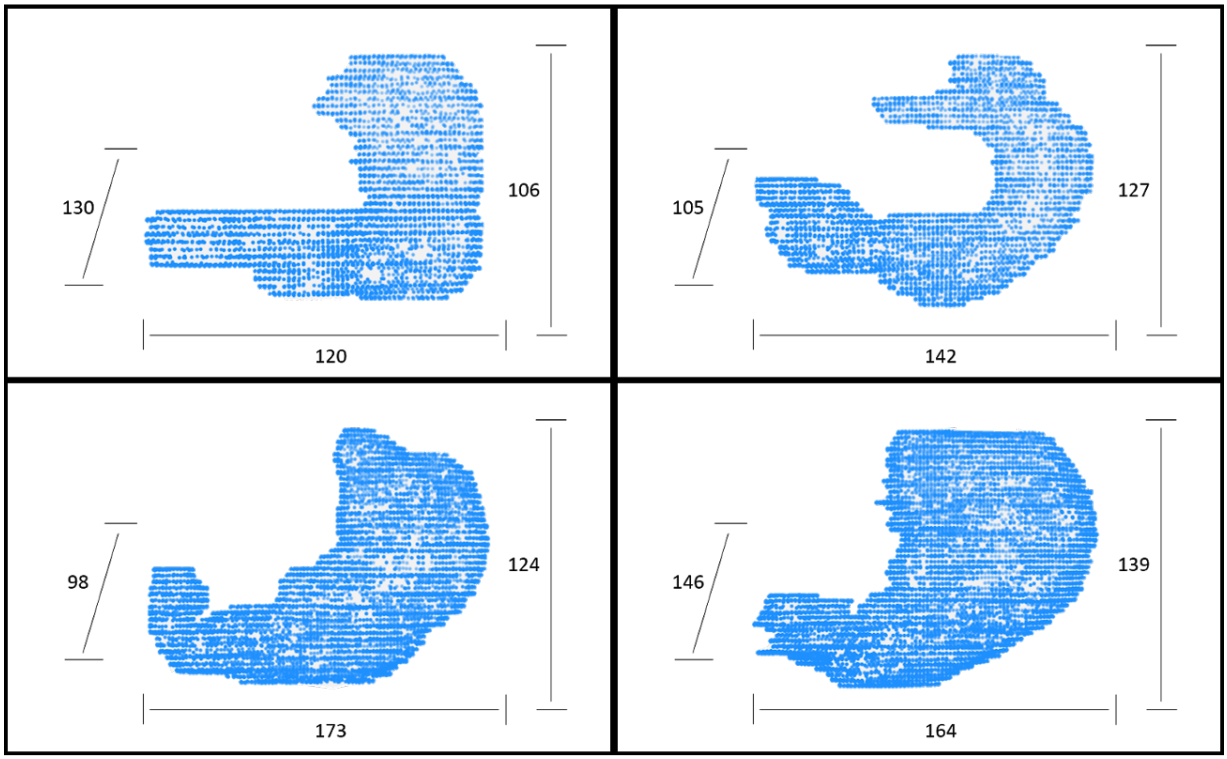

Figure E.8: Subjects 9-12: Stomach reconstruction from CT scan

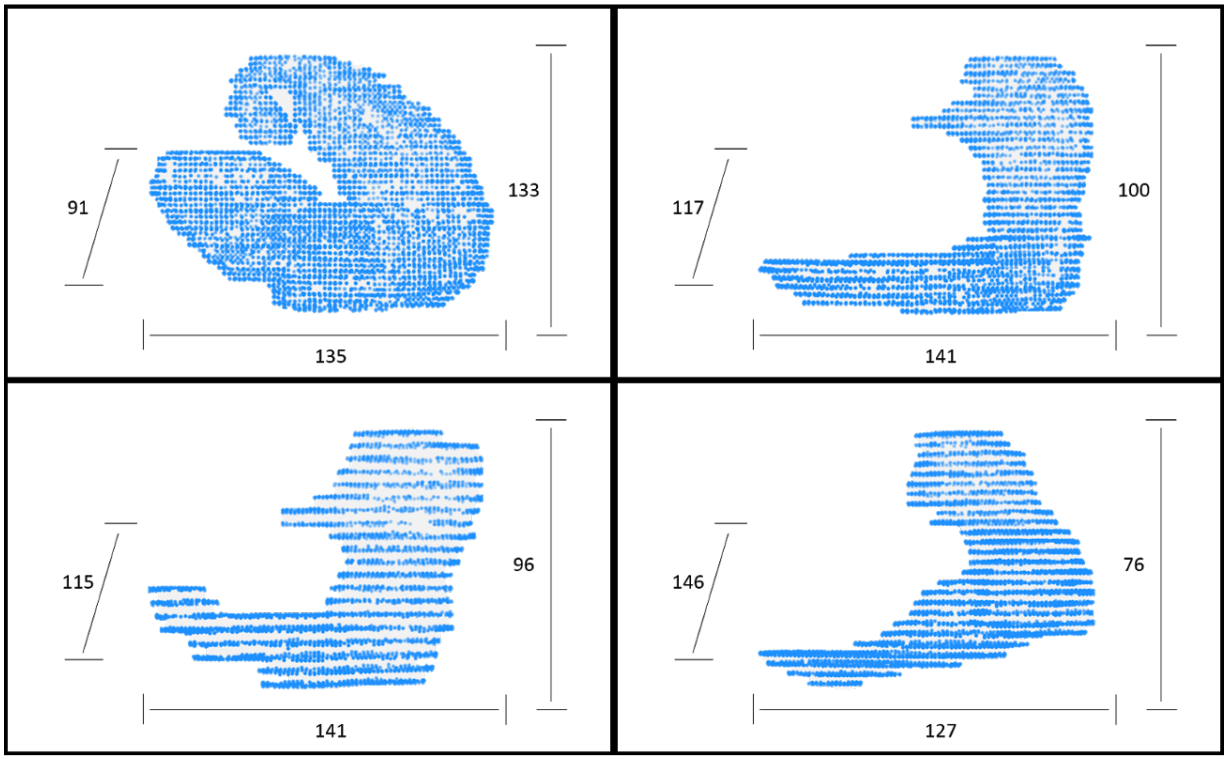

Figure E.9: Subjects 13-16: Stomach reconstruction from CT scan

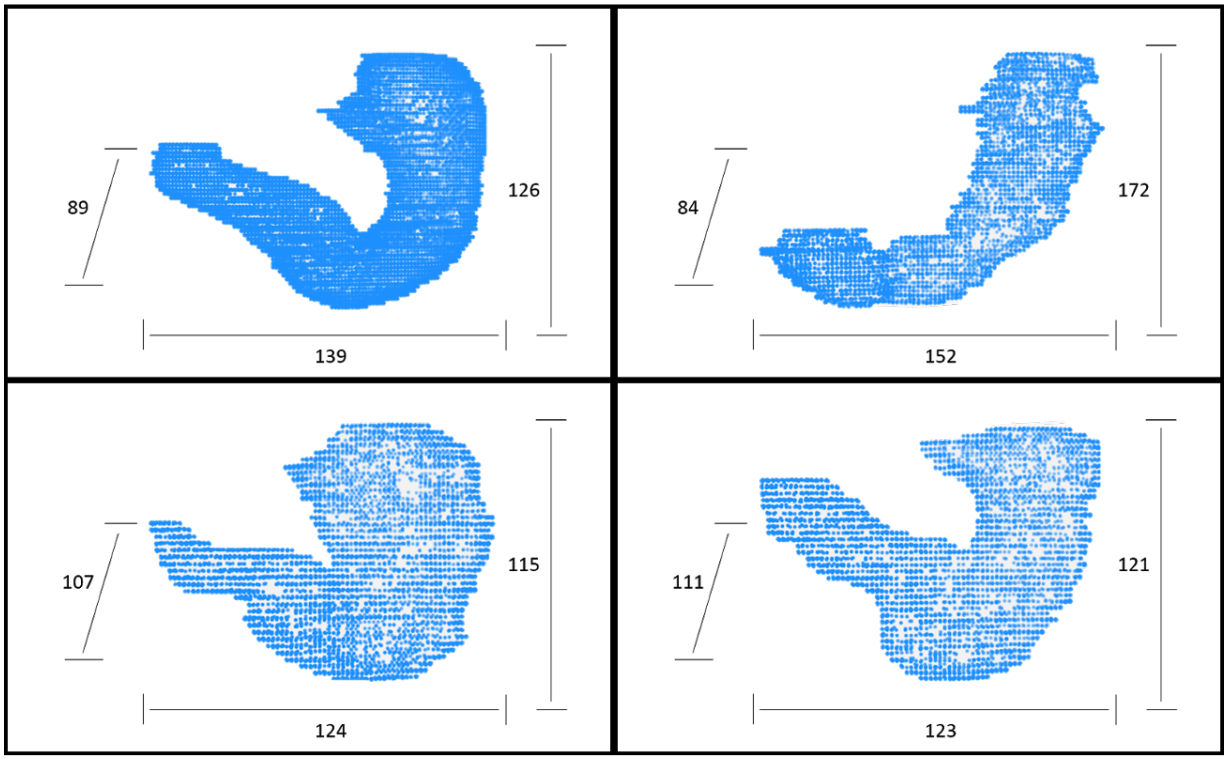

Figure E.10: Subjects 17-20: Stomach reconstruction from CT scan

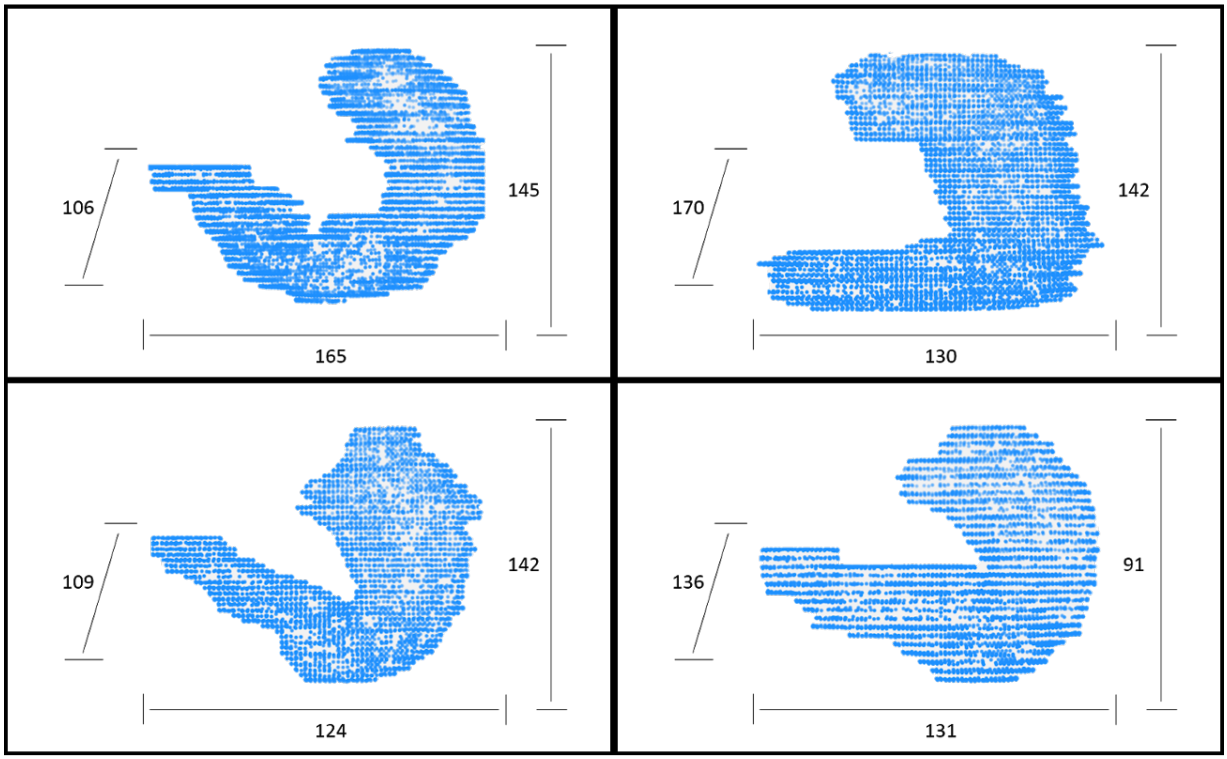

Figure E.11: Subjects 21-24: Stomach reconstruction from CT scan

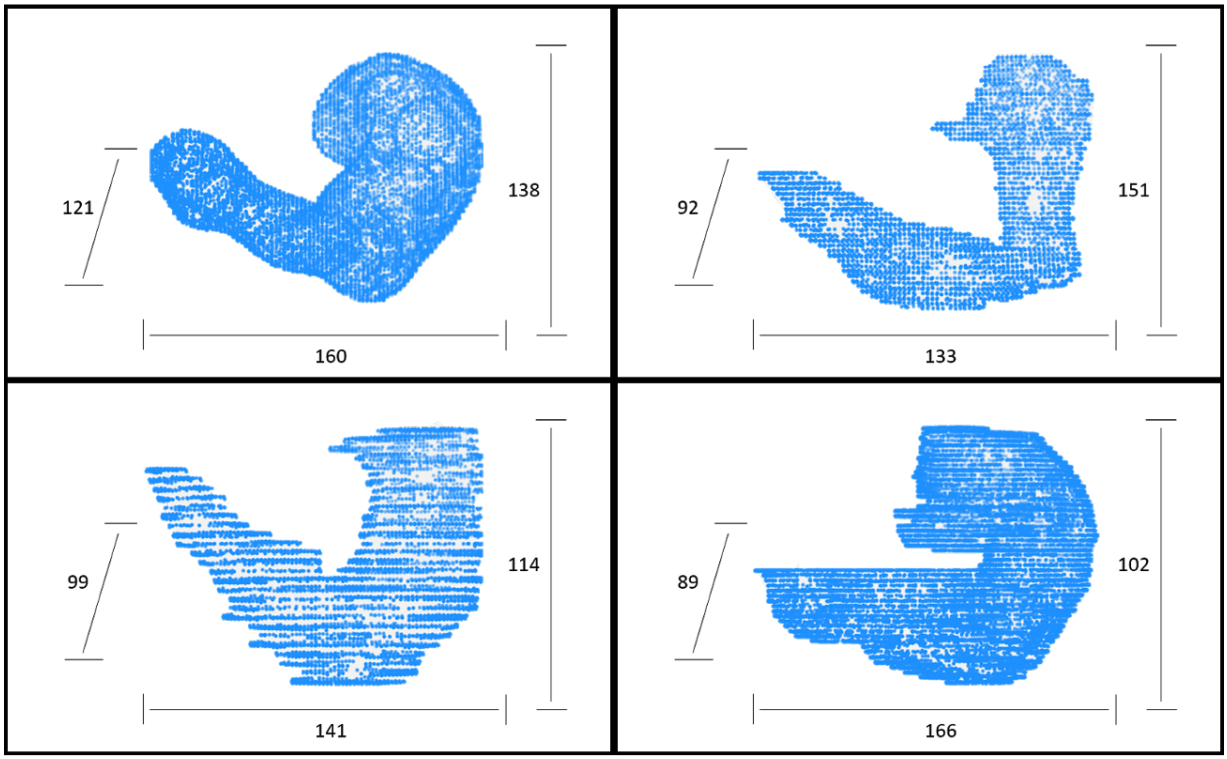

Figure E.12: Subjects 25-28: Stomach reconstruction from CT scan

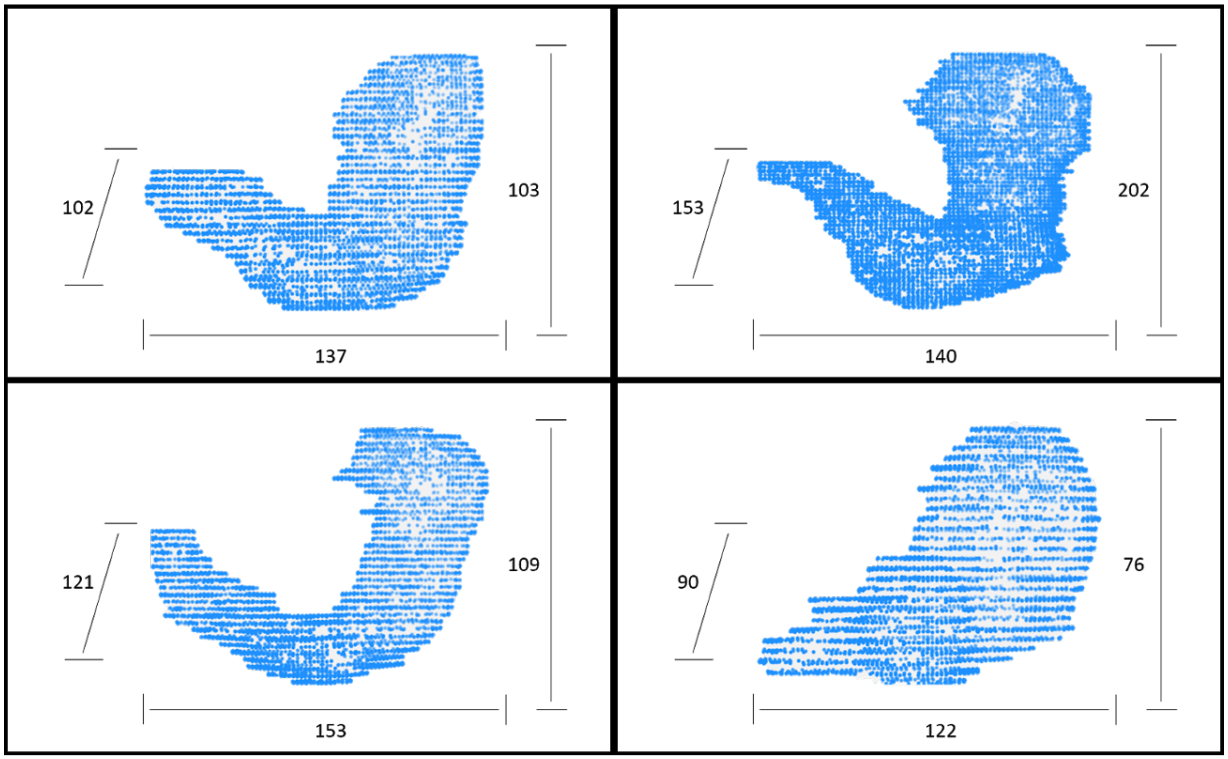

Figure E.13: Subjects 29-32: Stomach reconstruction from CT scan

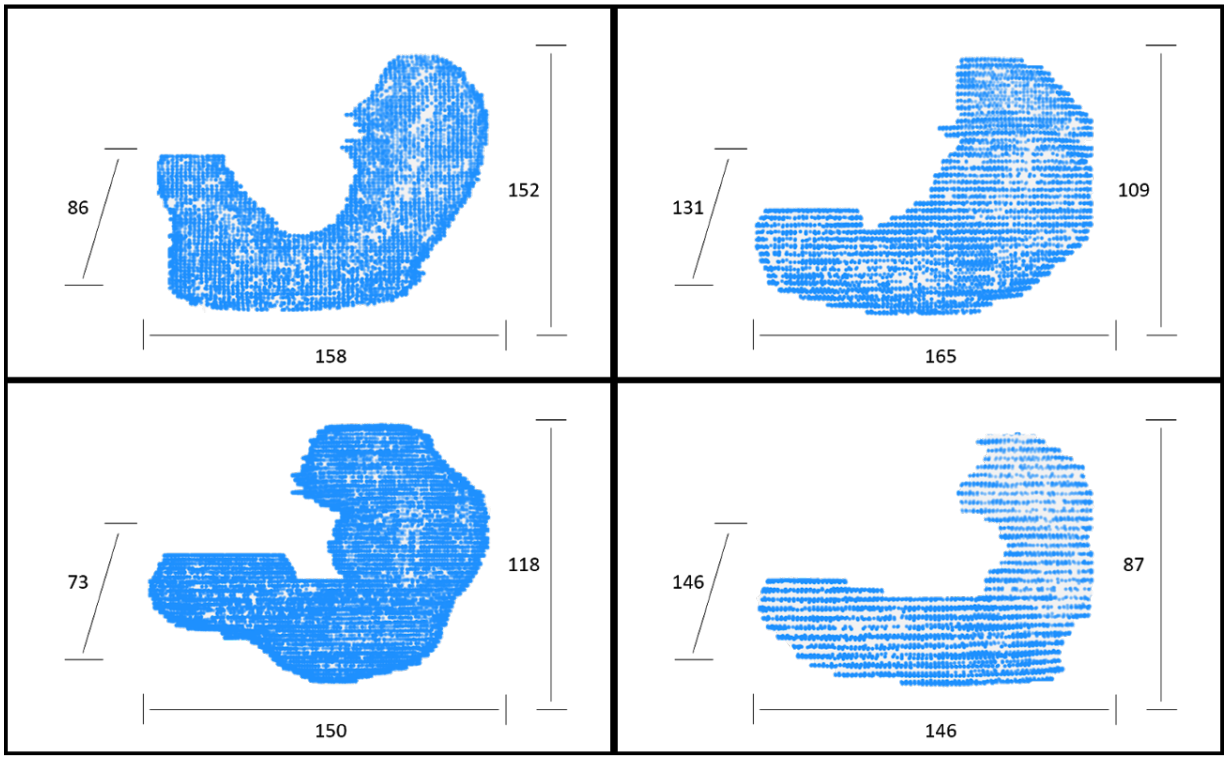

Figure E.14: Subjects 33-36: Stomach reconstruction from CT scan

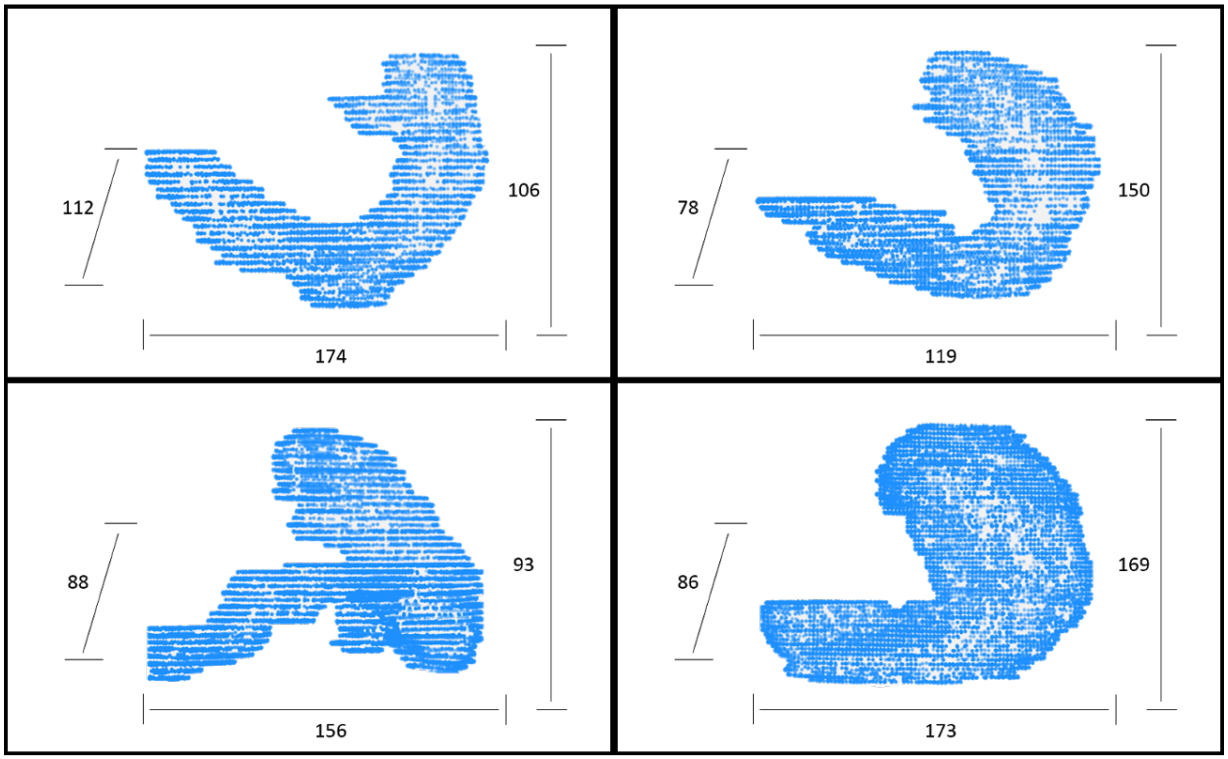

Figure E.15: Subjects 37-40: Stomach reconstruction from CT scan

Bibliography

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, and M. Devin. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `http://tensorflow.org/`. Software available from tensorflow.org.

[2] T. L. Abell, E. Van Cutsem, H. Abrahamsson, J. D. Huizinga, J. Konturek, J. P. Galmiche, G. VoelIer, L. Filez, B. Everts, and W. E. Waterfall. Gastric electrical stimulation in intractable symptomatic gastroparesis. *Digestion*, 66(4):204–212, 2002.

[3] S. Abid and G. Lindberg. Electrogastrography: poor correlation with antro-duodenal manometry and doubtful clinical usefulness in adults. *World Journal of Gastroenterology: WJG*, 13(38):5101, 2007.

[4] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences*, 405:81–90, 2017.

[5] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. San Tan. A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, 89:389–396, 2017.

[6] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, 100:270–278, 2018.

[7] A. S. Agrusa, A. A. Gharibans, A. Allegra, D. C. Kunkel, and T. Coleman. A deep convolutional neural network approach to classify normal and abnormal gastric slow wave initiation from the high resolution electrogastrogram. *IEEE Transactions on Biomedical Engineering*, 2019.

[8] A. S. Agrusa, D. C. Kunkel, and T. P. Coleman. Robust regression and optimal transport methods to predict gastrointestinal disease etiology from high resolution egg and symptom severity. *In Review*, 2021.

[9] M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[10] M. Ahmed, A. N. Mahmood, and J. Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.

[11] L. A. Alexandre. 3D object recognition using convolutional neural networks with transfer learning between input channels. In E. Menegatti, N. Michael, K. Berns, and H. Yamaguchi, editors, *Intelligent Autonomous Systems 13*, pages 889–898, Cham, 2016. Springer International Publishing. ISBN 978-3-319-08338-4.

[12] S. Alighaleh, T. R. Angeli, S. Sathar, G. O'Grady, L. K. Cheng, and N. Paskaranandavadivel. Real-time evaluation of a novel gastric pacing device with high-resolution mapping. *Gastroenterology*, 154(6):S–39, 2018.

[13] S. Alighaleh, L. K. Cheng, T. R. Angeli, M. Amiri, S. Sathar, G. O'Grady, and N. Paskaranandavadivel. A novel gastric pacing device to modulate slow waves and assessment by high-resolution mapping. *IEEE Transactions on Biomedical Engineering*, 66(10):2823–2830, 2019.

[14] U. C. Allard, F. Nougarou, C. L. Fall, P. Giguère, C. Gosselin, F. Laviolette, and B. Gosselin. A convolutional neural network for robotic arm guidance using SEMG based frequency-features. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 2464–2470. IEEE, 2016.

[15] T. Amaral, C. Kandaswamy, L. M. Silva, L. A. Alexandre, J. M. de Sá, and J. M. Santos. Improving performance on problems with few labelled data by reusing stacked autoencoders. In *2014 13th International Conference on Machine Learning and Applications*, pages 367–372. IEEE, 2014.

[16] R. Anaparthy, N. Pehlivanov, J. Grady, H. Yimei, and P. J. Pasricha. Gastroparesis and gastroparesis-like syndrome: response to therapy and its predictors. *Digestive diseases and sciences*, 54(5):1003–1010, 2009.

[17] E. Anderes, S. Borgwardt, and J. Miller. Discrete wasserstein barycenters: Optimal transport for discrete data. *Mathematical Methods of Operations Research*, 84(2):389–409, 2016.

[18] R. Andersen. *Modern methods for robust regression*. Number 152. Sage, 2008.

[19] T. R. Angeli, L. K. Cheng, P. Du, T. H.-H. Wang, C. E. Bernard, M.-G. Vannucchi, M. S. Faussone-Pellegrini, C. Lahr, R. Vather, and J. A. Windsor. Loss of interstitial cells of Cajal and patterns of gastric dysrhythmia in patients with chronic unexplained nausea and vomiting. *Gastroenterology*, 149(1):56–66, 2015.

[20] H. Atassi and T. L. Abell. Gastric electrical stimulator for treatment of gastroparesis. *Gastrointestinal Endoscopy Clinics*, 29(1):71–83, 2019.

[21] M. E. Barton, T. Otiker, L. V. Johnson, D. C. Robertson, R. L. Dobbins, H. P. Parkman,

P. M. Hellström, J. F. Tack, B. Kuo, and A. Hobson. A randomized, double-blind, placebo-controlled phase II study (MOT114479) to evaluate the safety and efficacy and dose response of 28 days of orally administered camicinal, a motilin receptor agonist, in diabetics with gastroparesis. *Gastroenterology*, 146(5):S–20, 2014.

[22] J.-P. Baudry and G. Celeux. Em for mixtures. *Statistics and computing*, 25(4):713–726, 2015.

[23] R. Berry, L. K. Cheng, P. Du, N. Paskaranandavadivel, T. R. Angeli, T. Mayne, G. Beban, and G. O'Grady. Patterns of abnormal gastric pacemaking after sleeve gastrectomy defined by laparoscopic high-resolution electrical mapping. *Obesity surgery*, 27(8):1929–1937, 2017.

[24] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita. Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials*, 16(1):303–336, 2013.

[25] M. Bortolotti, P. Sarti, L. Barbara, and F. Brunelli. Gastric myoelectric activity in patients with chronic idiopathic gastroparesis. *Neurogastroenterology and Motility*, 2(2):104–108, 1990.

[26] S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[27] L. Bradshaw, L. Cheng, E. Chung, C. Obioha, J. Erickson, B. Gorman, S. Somarajan, and W. Richards. Diabetic gastroparesis alters the biomagnetic signature of the gastric slow wave. *Neurogastroenterology and Motility*, 28(6):837–848, 2016.

[28] S. Brandstaeter, S. L. Fuchs, R. C. Aydin, and C. J. Cyron. Mechanics of the stomach: A review of an emerging field of biomechanics. *GAMM-Mitteilungen*, page e201900001, 2019.

[29] R. Brun and B. Kuo. Functional dyspepsia. *Therapeutic advances in gastroenterology*, 3 (3):145–164, 2010.

[30] S. Calder, G. O'Grady, L. K. Cheng, and P. Du. Wave front tracking and velocity profiling of EGG signatures. In *2018 International Gastrointestinal Electrophysiology Society Annual Meeting on*. iGES, 2018.

[31] S. Calder, G. O'Grady, L. K. Cheng, and P. Du. Torso-tank validation of high-resolution electrogastrography (EGG): Forward modelling, methodology and results. *Annals of Biomedical Engineering*, pages 1–11, 2018.

[32] F. Chamroukhi. Unsupervised learning of regression mixture models with unknown number of components. *Journal of Statistical Computation and Simulation*, 86(12):2308–2334, 2016.

[33] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[34] J. Chen, Z. Lin, J. Pan, and R. W. McCallum. Abnormal gastric myoelectrical activity and delayed gastric emptying in patients with symptoms suggestive of gastroparesis. *Digestive Diseases and Sciences*, 41(8):1538–1545, 1996.

[35] R. Chen and I. Paschalidis. Outlier detection using robust optimization with uncertainty sets constructed from risk measures. *ACM SIGMETRICS Performance Evaluation Review*, 45(3):174–179, 2018.

[36] F. Chollet. Keras. `https://keras.io`, 2015.

[37] D. Cireşan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. *ArXiv*, 2012.

[38] K. J. Coakley. A cross-validation procedure for stopping the em algorithm and deconvolution of neutron depth profiling spectra. *IEEE transactions on nuclear science*, 38(1):9–15, 1991.

[39] R. Corinaldesi, V. Stanghellini, C. Raiti, E. Rea, R. Salgemini, and L. Barbara. Effect of chronic administration of cisapride on gastric emptying of a solid meal and on dyspeptic symptoms in patients with idiopathic gastroparesis. *Gut*, 28(3):300–305, 1987.

[40] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *International Conference on Machine Learning*, pages 647–655, 2014.

[41] P. Du, S. Calder, T. R. Angeli, S. Sathar, N. Paskaranandavadivel, G. O'Grady, and L. K. Cheng. Progress in mathematical modeling of gastrointestinal slow wave abnormalities. *Frontiers in physiology*, 8:1136, 2018.

[42] H. El-Serag and N. Talley. The prevalence and clinical course of functional dyspepsia. *Alimentary pharmacology & therapeutics*, 19(6):643–654, 2004.

[43] T. Elomaa and H. Koivistoinen. On autonomous k-means clustering. In *International Symposium on Methodologies for Intelligent Systems*, pages 228–236. Springer, 2005.

[44] E. Eskin. Anomaly detection over noisy data using learned probability distributions. In *Proceedings of the Seventeenth International Conference on Machine Learning (ICML-2000)*, pages 255–262. Stanford University (San Francisco: Kaufmann, 2000), 2000.

[45] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo. A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security*, pages 77–101. Springer, 2002.

[46] G. Farrugia. Interstitial cells of cajal in health and disease. *Neurogastroenterology &*

Motility, 20:54–63, 2008.

[47] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya. Deep learning for healthcare applications based on physiological signals: a review. *Computer Methods and Programs in Biomedicine*, 2018.

[48] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

[49] A. A. Gharibans, S. Kim, D. C. Kunkel, and T. P. Coleman. High-resolution electrogastrogram: a novel, noninvasive method for determining gastric slow-wave direction and speed. *IEEE Transactions on Biomedical Engineering*, 64(4):807–815, 2017.

[50] A. A. Gharibans, S. Kim, D. C. Kunkel, and T. P. Coleman. High-resolution electrogastrogram: a novel, noninvasive method for determining gastric slow-wave direction and speed. *IEEE Transactions on Biomedical Engineering*, 64(4):807–815, 2017.

[51] A. A. Gharibans, S. Kim, D. C. Kunkel, and T. P. Coleman. High-resolution electrogastrogram: a novel, noninvasive method for determining gastric slow-wave direction and speed. *IEEE Transactions on Biomedical Engineering*, 64(4):807–815, 2017.

[52] A. A. Gharibans, B. L. Smarr, D. C. Kunkel, L. J. Kriegsfeld, H. M. Mousa, and T. P. Coleman. Artifact rejection methodology enables continuous, noninvasive measurement of gastric myoelectric activity in ambulatory subjects. *Scientific Reports*, 8(1):5019, 2018.

[53] A. A. Gharibans, T. P. Coleman, H. Mousa, and D. C. Kunkel. Spatial patterns from high-resolution electrogastrography correlate with severity of symptoms in patients with functional dyspepsia and gastroparesis. *Clinical Gastroenterology and Hepatology*, 2019.

[54] A. A. Gharibans, T. P. Coleman, H. Mousa, and D. C. Kunkel. Spatial patterns from high-resolution electrogastrography correlate with severity of symptoms in patients with functional dyspepsia and gastroparesis. *Clinical Gastroenterology and Hepatology*, 2019.

[55] N. Gigli and S.-i. Ohta. First variation formula in wasserstein spaces over compact alexandrov spaces. *Canadian Mathematical Bulletin*, 55(4):723–735, 2012.

[56] A. Gikas and J. K. Triantafillidis. The role of primary care physicians in early diagnosis and treatment of chronic gastrointestinal diseases. *International Journal of General Medicine*, 7:159, 2014.

[57] T. Gootzen, D. Stegeman, A. Van Oosterom, and H. Vingerhoets. A comprehensive model description for motor unit action potentials. *Electroenceph Clin Neurophysiol*, 66:S40, 1987.

[58] A. C. Gottfried, E. P. Adler, N. Fernandez-Becker, J. O. Clarke, A. Habtezion, and L. A. B. Nguyen. Transcutaneous vagal nerve stimulation improves symptoms and gastric emptying in patients with idiopathic gastroparesis. In *Gastroenterology*, volume 156, pages S789–

S790, 2019.

[59] A. Gottfried-Blackmore, E. P. Adler, N. Fernandez-Becker, J. Clarke, A. Habtezion, and L. Nguyen. Open-label pilot study: Non-invasive vagal nerve stimulation improves symptoms and gastric emptying in patients with idiopathic gastroparesis. *Neurogastroenterology & Motility*, page e13769, 2019.

[60] M. R. Gupta and Y. Chen. *Theory and use of the EM algorithm*. Now Publishers Inc, 2011.

[61] M. Hahsler, M. Piekenbrock, and D. Doran. dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, 91(1):1–30, 2019. doi: 10.18637/jss.v091.i01.

[62] M. Hahsler, M. Piekenbrock, and D. Doran. dbscan: Fast density-based clustering with r. *Journal of Statistical Software*, 91(1):1–30, 2019.

[63] G. Hamerly and C. Elkan. Learning the k in k-means. In *Advances in neural information processing systems*, pages 281–288, 2004.

[64] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[65] Z. S. Heetun and E. M. Quigley. Gastroparesis and Parkinson's disease: a systematic review. *Parkinsonism and Related Disorders*, 18(5):433–440, 2012.

[66] G. D. S. Hirst, A. P. Garcia-Londoño, and F. R. Edwards. Propagation of slow waves in the guinea-pig gastric antrum. *The Journal of physiology*, 571(1):165–177, 2006.

[67] G. Ho, C. T. Villongco, O. Yousefian, A. Bradshaw, A. Nguyen, Y. Faiwiszewski, J. Hayase, W.-J. Rappel, A. D. McCulloch, and D. E. Krummen. Rotors exhibit greater surface ECG variation during ventricular fibrillation than focal sources due to wavebreak, secondary rotors, and meander. *Journal of cardiovascular electrophysiology*, 28(10):1158–1166, 2017.

[68] V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial intelligence review*, 22(2):85–126, 2004.

[69] P. Holmvall and G. Lindberg. Electrogastrography before and after a high-caloric, liquid test meal in healthy volunteers and patients with severe functional dyspepsia. *Scandinavian Journal of Gastroenterology*, 37(10):1144–1148, 2002.

[70] M. Horowitz, Y.-C. Su, C. K. Rayner, and K. L. Jones. Gastroparesis: prevalence, clinical significance and treatment. *Canadian Journal of Gastroenterology and Hepatology*, 15 (12):805–813, 2001.

[71] W. Hu, Y. Liao, and V. R. Vemuri. Robust anomaly detection using support vector machines. In *Proceedings of the international conference on machine learning*, pages 282–289. Citeseer, 2003.

[72] S. Hussein, K. Cao, Q. Song, and U. Bagci. Risk stratification of lung nodules using 3D CNN-based multi-task learning. In *Information Processing in Medical Imaging*, pages 249–260, Cham, 2017. Springer International Publishing. ISBN 978-3-319-59050-9.

[73] E. J. Irvine, W. E. Whitehead, W. D. Chey, K. Matsueda, M. Shaw, N. J. Talley, S. J. V. van Zanten, and D. of Treatment Trials Committee. Design of treatment trials for functional gastrointestinal disorders. *Gastroenterology*, 150(6):1469–1480, 2016.

[74] P. Janssen, S. M. Harris, M. Jones, T. Masaoka, R. Farré, H. Törnblom, L. Van Oudenhove, M. Simrén, and J. Tack. The relation between symptom improvement and gastric emptying in the treatment of diabetic and idiopathic gastroparesis. *The American Journal of Gastroenterology*, 108(9):1382, 2013.

[75] S. Ji, W. Xu, M. Yang, and K. Yu. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 221–231, 2013.

[76] T. D. Johnson, T. P. Coleman, and L. M. Rangel. A flexible likelihood approach for predicting neural spiking activity from oscillatory phase. *Journal of neuroscience methods*, 311:307–317, 2019.

[77] H.-K. Jung, G. R. Locke III, C. D. Schleck, A. R. Zinsmeister, L. A. Szarka, B. Mullan, and N. J. Talley. The incidence, prevalence, and outcomes of patients with gastroparesis in olmsted county, minnesota, from 1996 to 2006. *Gastroenterology*, 136(4):1225–1233, 2009.

[78] A. Karami and R. Johansson. Choosing dbscan parameters automatically using differential evolution. *International Journal of Computer Applications*, 91(7):1–11, 2014.

[79] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[80] K. L. Koch and R. M. Stern. *Handbook of Electrogastrography*. Oxford University Press, 2004.

[81] K. L. Koch, R. M. Stern, W. R. Stewart, and M. W. Vasey. Gastric emptying and gastric myoelectrical activity in patients with diabetic gastroparesis: effect of long-term domperidone treatment. *American Journal of Gastroenterology*, 84(9), 1989.

[82] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[83] R. Kruisselbrink, A. Gharapetian, L. E. Chaparro, N. Ami, D. Richler, V. W. Chan, and A. Perlas. Diagnostic accuracy of point-of-care gastric ultrasound. *Anesthesia & Analgesia*, 128(1):89–95, 2019.

[84] J. F. Kurniawan, B. Tjhia, A. Shin, V. Wu, N. Sit, T. Pham, A. Nguyen, C. Li, R. Kumar, M. Aguilar Rivera, I. Lerman, D. C. Kunkel, and T. P. Coleman. An adhesive-integrated stretchable Ag/AgCl electrode array for unobtrusive monitoring of gastric neuromuscular activity. *Under Review*, 2020.

[85] B. E. Lacy, M. D. Crowell, A. Schettler-Duncan, C. Mathis, and P. J. Pasricha. The treatment of diabetic gastroparesis with botulinum toxin injection of the pylorus. *Diabetes Care*, 27(10):2341–2347, 2004.

[86] J. Langworthy, H. P. Parkman, and R. Schey. Emerging strategies for the treatment of gastroparesis. *Expert Review of Gastroenterology and Hepatology*, 10(7):817–825, 2016.

[87] A. Leahy, K. Besherdas, C. Clayman, and O. Epstein. Abnormalities of the electrogastrogram in functional gastrointestinal disorders. *The American Journal of Gastroenterology*, 94(4):1023, 1999.

[88] K.-L. Li, H.-K. Huang, S.-F. Tian, and W. Xu. Improving one-class svm for anomaly detection. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693)*, volume 5, pages 3077–3081. IEEE, 2003.

[89] Z. Lin, R. McCallum, B. Schirmer, and J. Chen. Effects of pacing parameters on entrainment of gastric slow waves in patients with gastroparesis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 274(1):G186–G191, 1998.

[90] Z. Lin, E. Y. Eaker, I. Sarosiek, and R. W. McCallum. Gastric myoelectrical activity and gastric emptying in patients with functional dyspepsia. *The American Journal of Gastroenterology*, 94(9):2384, 1999.

[91] Z. Lin, I. Sarosiek, J. Forster, R. A. Ross, J. D. Chen, and R. W. McCallum. Two-channel gastric pacing in patients with diabetic gastroparesis. *Neurogastroenterology & Motility*, 23(10):912–e396, 2011.

[92] Y.-K. Lo, P.-M. Wang, G. Dubrovsky, M.-D. Wu, M. Chan, J. C. Dunn, and W. Liu. A wireless implant for gastrointestinal motility disorders. *Micromachines*, 9(1):17, 2018.

[93] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[94] K. V. Mardia. Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(3):349–371, 1975.

[95] K. V. Mardia. *Statistics of directional data*. Academic press, 2014.

[96] D. Maturana and S. Scherer. Voxnet: A 3D convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 922–928. IEEE, 2015.

[97] R. McCallum, O. Cynshi, and I. Team. Clinical trial: effect of mitemcinal (a motilin agonist) on gastric emptying in patients with gastroparesis–a randomized, multicentre, placebo-controlled study. *Alimentary Pharmacology and Therapeutics*, 26(8):1121–1130, 2007.

[98] R. W. McCallum, J. D. Z. Chen, Z. Lin, B. D. Schirmer, R. D. Williams, and R. A. Ross. Gastric pacing improves emptying and symptoms in patients with gastroparesis. *Gastroenterology*, 114(3):456–461, 1998.

[99] L. Mi, W. Zhang, X. Gu, and Y. Wang. Variational wasserstein clustering. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 322–337, 2018.

[100] L. Mi, T. Yu, J. Bento, W. Zhang, B. Li, and Y. Wang. Variational wasserstein barycenters for geometric clustering. *arXiv preprint arXiv:2002.10543*, 2020.

[101] P. Min. binvox. `http://www.patrickmin.com/binvox`, .

[102] P. Min. thinvox. `http://www.patrickmin.com/thinvox`, .

[103] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. In *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, pages 244–249. IEEE, 2008.

[104] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz. Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4207–4215, 2016.

[105] A. C. Nanivadekar, D. M. Miller, S. Fulton, L. Wong, J. Ogren, G. Chitnis, B. McLaughlin, S. Zhai, L. E. Fisher, and B. J. Yates. Machine learning prediction of emesis and gastrointestinal state in ferrets. *bioRxiv*, page 607242, 2019.

[106] F. S. Nooruddin and G. Turk. Simplification and repair of polygonal models using volumetric techniques. *IEEE Transactions on Visualization and Computer Graphics*, 9(2): 191–205, 2003.

[107] A. Oba-Kuniyoshi, J. Oliveira Jr, E. Moraes, and L. Troncon. Postprandial symptoms in dysmotility-like functional dyspepsia are not related to disturbances of gastric myoelectrical activity. *Brazilian Journal of Medical and Biological Research*, 37(1):47–53, 2004.

[108] G. O'Grady, P. Du, W. J. Lammers, J. U. Egbuji, P. Mithraratne, J. D. Chen, L. K. Cheng, J. A. Windsor, and A. J. Pullan. High-resolution entrainment mapping of gastric pacing: a new analytical tool. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 298(2):G314–G321, 2009.

[109] G. O'Grady, P. Du, L. K. Cheng, J. U. Egbuji, W. J. Lammers, J. A. Windsor, and A. J. Pullan. Origin and propagation of human gastric slow-wave activity defined by high-resolution mapping. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 299(3):G585–G592, 2010.

[110] G. O'Grady, T. R. Angeli, P. Du, C. Lahr, W. J. Lammers, J. A. Windsor, T. L. Abell, G. Farrugia, A. J. Pullan, and L. K. Cheng. Abnormal initiation and conduction of slow-wave activity in gastroparesis, defined by high-resolution electrical mapping. *Gastroenterology*, 143(3):589–598, 2012.

[111] G. O'Grady, T. H.-H. Wang, P. Du, T. Angeli, W. J. Lammers, and L. K. Cheng. Recent progress in gastric arrhythmia: pathophysiology, clinical significance and future horizons. *Clinical and Experimental Pharmacology and Physiology*, 41(10):854–862, 2014.

[112] G. O'Grady, T. R. Angeli, N. Paskaranandavadivel, J. C. Erickson, C. I. Wells, A. A. Gharibans, L. K. Cheng, and P. Du. Methods for high-resolution electrical mapping in the gastrointestinal tract. *IEEE reviews in biomedical engineering*, 12:287–302, 2019.

[113] T. Ördög. Interstitial cells of cajal in diabetic gastroenteropathy. *Neurogastroenterology & Motility*, 20(1):8–18, 2008.

[114] C. Ouyang. Clinical significance of gastric dysrhythmias. *World Journal of Gastroenterology*, 2(Suppl1):5–6, 1996.

[115] P. Pal. Projection and supporting hyperplanes". Lecture notes for Convex Optimization, Spring 2020.

[116] K. Palágyi and A. Kuba. Directional 3D thinning using 8 subiterations. In *International Conference on Discrete Geometry for Computer Imagery*, pages 325–336. Springer, 1999.

[117] B. Panić, J. Klemenc, and M. Nagode. Improved initialization of the em algorithm for mixture model parameter estimation. *Mathematics*, 8(3):373, 2020.

[118] H. Parkman, W. Hasler, J. Barnett, and E. Eaker. Electrogastrography: a document prepared by the gastric section of the American Motility Society Clinical GI Motility Testing Task Force. *Neurogastroenterology and Motility*, 15(2):89–102, 2003.

[119] H. P. Parkman, M. A. Miller, D. Trate, L. C. Knight, J.-L. Urbain, A. H. Maurer, and R. S. Fisher. Electrogastrography and gastric emptying scintigraphy are complementary for assessment of dyspepsia. *Journal of Clinical Gastroenterology*, 24(4):214–219, 1997.

[120] S. P. Parsons and J. D. Huizinga. Phase waves and trigger waves: emergent properties of oscillating and excitable networks in the gut. *The Journal of physiology*, 596(20): 4819–4829, 2018.

[121] N. Paskaranandavadivel, T. R. Angeli, T. Manson, A. Stocker, L. McElmurray, G. O'Grady,

T. Abell, and L. K. Cheng. Multi-day, multi-sensor ambulatory monitoring of gastric electrical activity. *Physiological measurement*, 40(2):025011, 2019.

[122] N. Paskaranandavadivel, R. Avci, N. Nagahawatte, A. Farajidavar, and L. K. Cheng. Electroceutical approaches for gastroparesis. In *Obesity and Diabetes*, pages 967–982. Springer, 2020.

[123] P. J. Pasricha, R. Colvin, K. Yates, W. L. Hasler, T. L. Abell, A. Ünalp-Arida, L. Nguyen, G. Farrugia, K. L. Koch, and H. P. Parkman. Characteristics of patients with chronic unexplained nausea and vomiting and normal gastric emptying. *Clinical Gastroenterology and Hepatology*, 9(7):567–576, 2011.

[124] A. Patcha and J.-M. Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer networks*, 51(12):3448–3470, 2007.

[125] S. C. Payne, J. B. Furness, and M. J. Stebbing. Bioelectric neuromodulation for gastrointestinal disorders: effectiveness and mechanisms. *Nature Reviews Gastroenterology & Hepatology*, page 1, 2018.

[126] D. Pelleg. Extending k-means with efficient estimation of the number of clusters in icml. 2000.

[127] G. Peyré and M. Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[128] R. Plonsey. Bioelectric phenomena. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2001.

[129] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, (99):1–10, 2017.

[130] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*, 2017.

[131] R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the em algorithm. *SIAM review*, 26(2):195–239, 1984.

[132] Y. Ren and Y. Wu. Convolutional deep belief networks for feature extraction of EEG signal. In *Neural Networks (IJCNN), 2014 International Joint Conference on*, pages 2850–2853. IEEE, 2014.

[133] A. Rentz, P. Kahrilas, V. Stanghellini, J. Tack, N. Talley, E. Trudeau, D. Dubois, and D. Revicki. Development and psychometric evaluation of the patient assessment of upper gastrointestinal symptom severity index (pagi-sym) in patients with upper gastrointestinal

disorders. *Quality of Life Research*, 13(10):1737–1749, 2004.

[134] D. A. Revicki, A. M. Rentz, D. Dubois, P. Kahrilas, V. Stanghellini, N. J. Talley, and J. Tack. Gastroparesis cardinal symptom index (gcsi): development and validation of a patient reported assessment of severity of gastroparesis symptoms. *Quality of Life Research*, 13 (4):833–844, 2004.

[135] K. Roeleveld, J. Blok, D. Stegeman, and A. Van Oosterom. Volume conduction models for surface EMG; confrontation with measurements. *Journal of Electromyography and Kinesiology*, 7(4):221–232, 1997.

[136] B. J. Roth and F. L. Gielen. A comparison of two models for calculating the electrical potential in skeletal muscle. *Annals of biomedical engineering*, 15(6):591–602, 1987.

[137] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.

[138] D. Rubino, K. A. Robbins, and N. G. Hatsopoulos. Propagating waves mediate information transfer in the motor cortex. *Nature neuroscience*, 9(12):1549–1557, 2006.

[139] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett. Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 990–994. IEEE, 2018.

[140] K. M. Sanders, S. M. Ward, and S. D. Koh. Interstitial cells: regulators of smooth muscle function. *Physiological reviews*, 94(3):859–907, 2014.

[141] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63): 94, 2015.

[142] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[143] W. Sha, P. J. Pasricha, and J. D. Chen. Correlations among electrogastrogram, gastric dysmotility, and duodenal dysmotility in patients with functional dyspepsia. *Journal of Clinical Gastroenterology*, 43(8):716–722, 2009.

[144] W. Sha, P. J. Pasricha, and J. D. Chen. Rhythmic and spatial abnormalities of gastric slow waves in patients with functional dyspepsia. *Journal of Clinical Gastroenterology*, 43(2): 123–129, 2009.

[145] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.

[146] S. P. Shashikumar, A. J. Shah, Q. Li, G. D. Clifford, and S. Nemati. A deep learning

approach to monitoring and detecting atrial fibrillation using wearable technology. In *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 141–144. IEEE, 2017.

[147] Y. Shimada, M. Watanabe, N. Shibahara, T. Kita, T. Itoh, and K. Terasawa. Electrogastrographic power ratio in humans is not related to changes in antrum-skin distance but to antral motility. *Journal of gastroenterology*, 33(3):310–317, 1998.

[148] A. Shine, S. Ahmed, L. McElmurray, A. Stocker, C. Pinkston, and T. Abell. Electrogastrography at baseline and response to temporary gastric electrical stimulation – a comparison of cutaneous with mucosal recordings. *Digestive Disease Week*, 2019.

[149] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *ArXiv*, 2014.

[150] H. Sivaraks and C. A. Ratanamahatana. Robust and accurate anomaly detection in ecg artifacts using time series motif discovery. *Computational and mathematical methods in medicine*, 2015, 2015.

[151] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[152] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[153] J. Szurszewski. A 100-year perspective on gastrointestinal motility. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 274(3):G447–G453, 1998.

[154] J. Tack. Gastric motor and sensory function. *Current opinion in gastroenterology*, 25(6): 557–565, 2009.

[155] A. Takatsu. Equivalence of two orthogonalities between probability measures. *arXiv preprint arXiv:1110.3036*, 2011.

[156] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4489–4497, 2015.

[157] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. Smem algorithm for mixture models. *Neural computation*, 12(9):2109–2128, 2000.

[158] D. F. Van Helden, D. R. Laver, J. Holdsworth, and M. S. Imtiaz. Generation and propagation of gastric slow waves. *Clinical and Experimental Pharmacology and Physiology*, 37(4): 516–524, 2010.

[159] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[160] C. T. Villongco, D. E. Krummen, J. H. Omens, and A. D. McCulloch. Non-invasive, model-based measures of ventricular electrical dyssynchrony for predicting CRT outcomes. *EP Europace*, 18(suppl_4):iv104–iv112, 2016.

[161] S. M. Ward and K. M. Sanders. Interstitial cells of cajal: primary targets of enteric motor innervation. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, 262(1):125–135, 2001.

[162] C. Warrender, S. Forrest, and B. Pearlmutter. Detecting intrusions using system calls: Alternative data models. In *Proceedings of the 1999 IEEE symposium on security and privacy (Cat. No. 99CB36344)*, pages 133–145. IEEE, 1999.

[163] K.-J. Won, K. M. Sanders, and S. M. Ward. Interstitial cells of cajal mediate mechanosensitive responses in the stomach. *Proceedings of the National Academy of Sciences*, 102 (41):14913–14918, 2005.

[164] B. Woolf. The log likelihood ratio test (the g-test). *Annals of human genetics*, 21(4): 397–409, 1957.

[165] P. Xia, J. Hu, and Y. Peng. EMG-Based estimation of limb movement using deep learning with recurrent convolutional neural networks. *Artificial Organs*, 42(5):E67–E77, 2018.

[166] J. Ye, P. Wu, J. Z. Wang, and J. Li. Fast discrete distribution clustering using wasserstein barycenter with sparse support. *IEEE Transactions on Signal Processing*, 65(9):2317–2332, 2017.

[167] J. Yin and J. D. Chen. Electrogastrography: methodology, validation and applications. *Journal of neurogastroenterology and motility*, 19(1):5, 2013.

[168] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, pages 3320–3328, 2014.

[169] S. Zahid, H. Cochet, P. M. Boyle, E. L. Schwarz, K. N. Whyte, E. J. Vigmond, R. Dubois, M. Hocini, M. Haïssaguerre, and P. Jaïs. Patient-derived models link re-entrant driver localization in atrial fibrillation to fibrosis spatial pattern. *Cardiovascular research*, 110 (3):443–454, 2016.

[170] B. Zhang, C. Zhang, and X. Yi. Competitive em algorithm for finite mixture models. *Pattern recognition*, 37(1):131–144, 2004.