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

Anjulie S. Agrusa, *Student Member, IEEE*, Armen A. Gharibans, *Member, IEEE*, Alexis A. Allegra, *Student Member, IEEE*, David C. Kunkel, and Todd P. Coleman, *Senior Member, IEEE*

***Abstract*—*Objective*: Gastric slow wave abnormalities have been associated with gastric motility disorders. Invasive studies in humans have described normal and abnormal propagation of the slow wave. This study aims to disambiguate the abnormally functioning wave from one of normalcy using multi-electrode abdominal waveforms of the electrogastrogram (EGG). *Methods*: Human stomach and abdominal models are extracted from computed tomography scans. Normal and abnormal slow waves are simulated along stomach surfaces. Current dipoles at the stomachs surface are propagated to virtual electrodes on the abdomen with a forward model. We establish a deep convolutional neural network (CNN) framework to classify normal and abnormal slow waves from the multi-electrode waveforms. We investigate the effects of non-idealized measurements on performance, including shifted electrode array positioning, smaller array sizes, high body mass index (BMI), and low signal-to-noise ratio (SNR). We compare the performance of our deep CNN to a linear discriminant classifier using wave propagation spatial features. *Results*: A deep CNN framework demonstrated robust classification, with accuracy above 90% for all SNR above 0 dB, horizontal shifts within 3 cm, vertical shifts within 6 cm, and abdominal tissue depth within 6 cm. The linear discriminant classifier was much more vulnerable to SNR, electrode placement, and BMI. *Conclusion*: This is the first study to attempt and, moreover, succeed in using a deep CNN to disambiguate normal and abnormal gastric slow wave patterns from high-resolution EGG data. *Significance*: These findings suggest that multi-electrode cutaneous abdominal recordings have the potential to serve as widely deployable clinical screening tools for gastrointestinal foregut disorders.**

Manuscript received March 15, 2019; revised May 28, 2019; accepted May 29, 2019. Date of publication June 12, 2019; date of current version February 19, 2020. This work was supported in part by the National Science Foundation Center for Science of Information under Grant CCF-0939370. *(Corresponding author: Todd P. Coleman.)*

A. S. Agrusa and A. A. Gharibans are with the Department of Bioengineering, University of California.

A. A. Allegra is with the Department of Electrical and Computer Engineering, University of California.

D. C. Kunkel is with the Division of Gastroenterology, Department of Medicine, University of California.

T. P. Coleman is with the Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: tpcoleman@ucsd.edu).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TBME.2019.2922235

***Index Terms*—Artificial intelligence, convolutional neural network, electrogastrogram, forward model, gastroenterology, gastric slow wave, machine learning, neural network, video classification.**

I. 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 [1], [2]. 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 [3]. 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 Parkinsons and diabetes patients [4], [5]. However, published findings have had limited success in demonstrating correlation between gastric emptying and symptoms [6] or symptom improvement [7]. As some drugs improve symptoms but not gastric emptying and vice versa [8]–[10], the NIH Gastroparesis Consortium has recently recommended that improvement in gastric emptying not be considered a requirement for clinical drug trials in gastroparesis [6].

It has been proposed that gastric motility disorders (including gastroparesis) should be sub-typed to improve treatment efficacy [11], [12]. 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 [13]. Recent findings from invasive electrode recordings [14], [15] 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

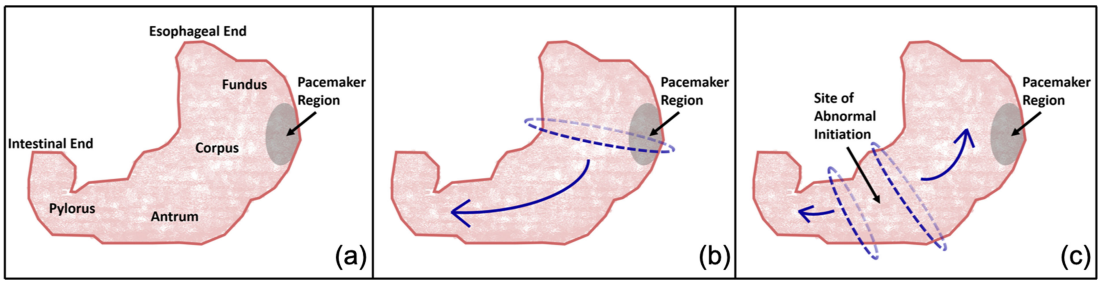

Fig. 1. Anterior view of stomach model extracted from human research subject's abdominal CT scan. (a) Anatomical regions of the stomach with pacemaker region highlighted in gray. (b) Normal gastric slow wave initiation in the pacemaker region and propagation of equipotential rings (dashed) in anterograde 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 [14], [15].

abnormal initiation, with slow waves generated outside the pacemaker region, propagating simultaneously in both anterograde and retrograde directions (Fig. 1).

This archetypal dysrhythmia, as well as others, has been associated with gastric functional and motility disorders, including gastroparesis and functional dyspepsia [16]–[22]. 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 [23]. Recent advances in the development of gastric pacemaking devices [24]–[27] 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 [13]. Whereas some findings based upon these spectral analyses have shown separation between patients with GI symptoms and healthy controls [28]–[30], others have found no such differences [31], [32].

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 [33], [34]. 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) [15], [35], [36]. As such, current literature has converged on the notion that spectral-based analyses are unable to capture these spatial abnormalities [37], [38]. 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 [39]. With the advent of these new techniques, in addition to novel artifact rejection methods [40], spatial features have been shown to correlate with symptom incidence and severity in gastroparesis and functional dyspepsia patients [41].

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 [3], 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 [15], 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 [14], [15]. 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 [42]–[44]. 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 [45]–[49] 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 [50], including the analysis of physiological signals such as the electrocardiogram [45]–[47], electroencephalogram [51]–[53], and electromyogram [54], [55]. 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) [56]–[59]. Similarly, three-dimensional CNNs are an accepted best-practice in video-classification tasks [60]–[62]. 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.

II. Methods

A. 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. 1b-c). First, we defined a novel 'Medial Curve' that captured the stomach's characteristic shape (Section II-A1). 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 II-A2), creating equipotential rings across the entire stomach surface. Next, we calculated subject-specific abdominal boundaries (Section II-A3) and propagated the serosal voltages to electrodes placed at these abdominal boundaries (Section II-A4). 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 II-A5). Ethical approval for this work was obtained from the institutional review board at the University of California, San Diego.

1) 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. 2). To find this curve, we manually extracted and voxelized the three-dimensional stomach model from CT images [63], [64]. The voxelized representation of the stomach was iteratively thinned to its 'geometric skeleton' [65], [66] (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

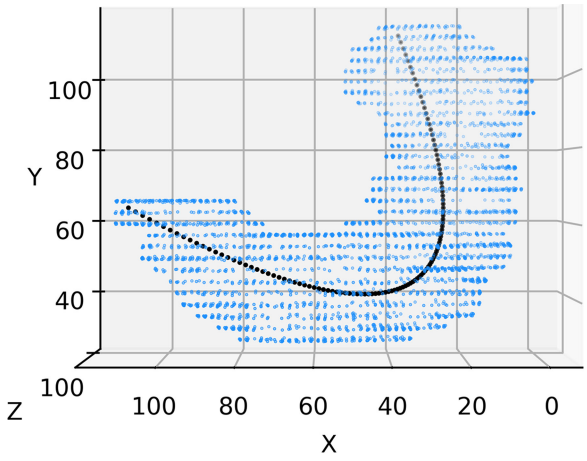

Fig. 2. Medial Curve (black) that captures the characteristic shape of the stomach, solved for with the optimization paradigm defined in (3). Units in mm.

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 $(\phi_k : k \geq 0)$:

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

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, $G = [\underline{g}_1 \ \underline{g}_2 \ \dots \ \underline{g}_K]$, we can succinctly describe the curve for any matrix G as:

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

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) below. It was necessary to incorporate both (2a) and (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, (2a), *and* each point on the Medial Curve was near some point on the geometric spine, (2b). Without the inclusion of (2b), there are an infinite number of curves, $C(\zeta)$, for which (2a) holds that extend past the esophageal and pyloric boundaries.

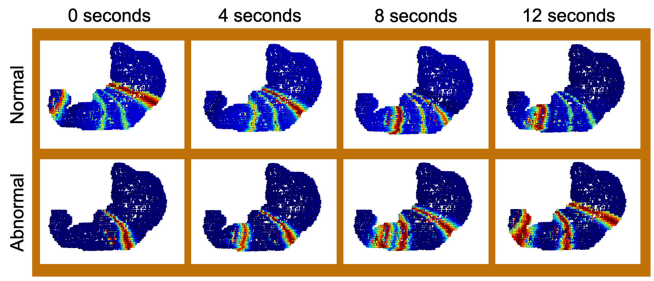

Fig. 3. Wave propagation at the serosal surface of the stomach with voltage values displayed as heat map. Normal slow wave initiation (top) and abnormal initiation (bottom), with each successive snapshot (left to right) at time t = 0, 4, 8, 12 seconds. The voltage at each point on the serosal surface of the stomach is solved for in Section II-A2 in accordance with the description in the most recent literature [14], [15], [39].

$$J_{\mathcal{B}}(G) = \left(\sum_{m=1}^{M} \min_{\zeta \in [-1,1]} \|C(\zeta) - p_m\|_2^2 \right) \tag{2a}$$

$$+ \left(\int_{-1}^{1} \min_{m \in \{1,\dots,M\}} \|C(\zeta) - p_m\|_2^2 d\zeta \right) \tag{2b}$$

We solved for G^* by minimizing (2), using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm:

$$G^* = \underset{G \in \mathbb{R}^{3\times K}}{\arg\min}\, J_{\mathcal{B}}(G). \tag{3}$$

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

2) Serosal Voltage Simulation: We simulated voltage potentials on the full serosal surface of the stomach (Fig. 3). We modeled the normal and abnormal wave initiation and propagation patterns to be consistent with recent findings from invasive human recordings [14], [15]. This was implemented by solving the one dimensional wave equation (4) at discrete points along the Medial Curve via finite difference analysis with a temporal step size calculated using the Courant-Friedrichs-Lewy condition [39]:

$$\frac{\partial^2 S(\zeta,t)}{\partial t^2} = c^2(\zeta) \frac{\partial^2 S(\zeta,t)}{\partial \zeta^2}. \tag{4}$$

In (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 [14], [15]; 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.

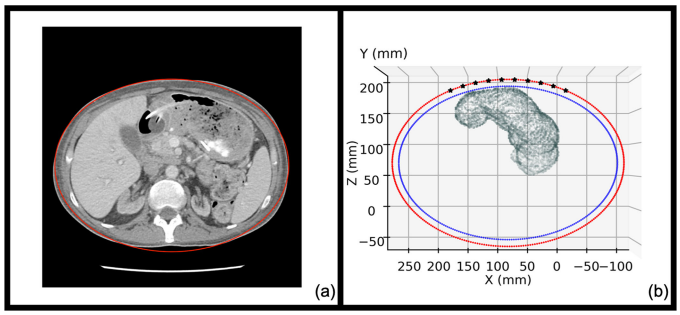

Fig. 4. (a) Top-down projection of human research subject's abdominal CT scan onto the X,Z plane. Approximated elliptical cross section (red) from which elliptical radii were measured. (b) Top-down projection of stomach and elliptic cylindrical point clouds onto the X, Z plane. Abdominal ellipse (red) 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 II-A3. 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.

3) 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 (Fig. 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. 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:

$$\frac{(x_i)^2}{(a-1)^2} + \frac{(z_i - z_0)^2}{(b-1)^2} \leq 1, \quad i = 1, \ldots, H. \tag{5}$$

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 II-A4), 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.

4) 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 II-A2 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:

$$V_n(t) = \sum_{d=1}^{D} W_{n,d} I_d(t) + N_n(t). \tag{6}$$

Source weights are given by:

$$W_{n,d} = \frac{\cos\theta}{4\pi\sigma r_{n,d}^2} \tag{7}$$

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 $\sigma = 0.125$ 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) [67], [68]. 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 [69], [70]. 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 (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.'

5) 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 II-A3). From this initial positioning, we moved the electrode array laterally away

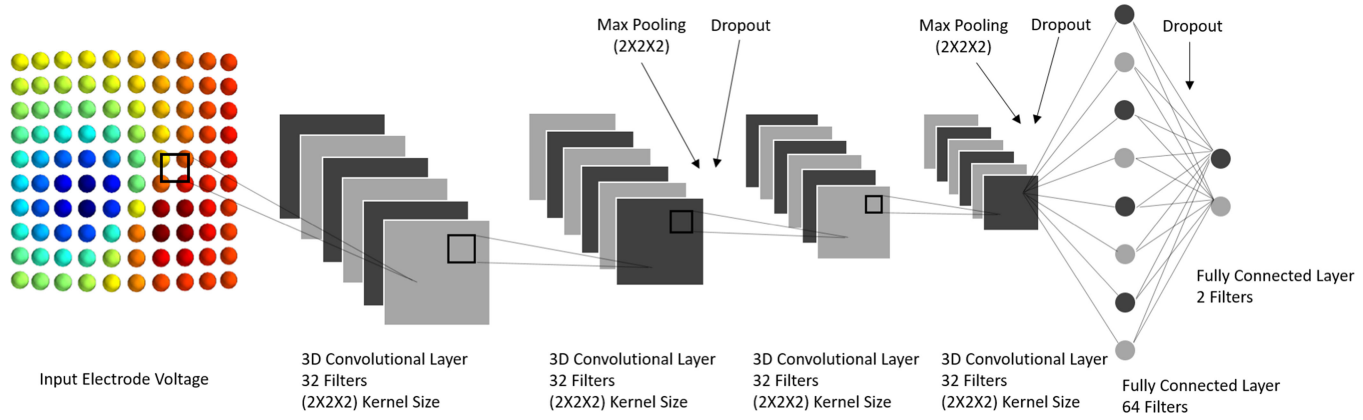

Fig. 5. Convolutional neural network architecture schematic, described in detail in Section II-B1.

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 (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 [38], [39] and ambulatory systems capable of recording from 9 electrodes have recently been established [40]. 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.

B. 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 II-B1). For comparison, we computed wave propagation spatial features to train a linear discriminant analysis (LDA) classifier (Section II-B2).

1) 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. 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 [71]. We carried out all computations through Tensorflow [72] and programmed nodes using the Keras API [73].

The general form of the neural network architecture was adapted from state-of-the-art 2D and 3D CNN published architectures [56], [57], [60]–[62]. 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 [56]–[62]. 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 CNNs [56], [57] 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

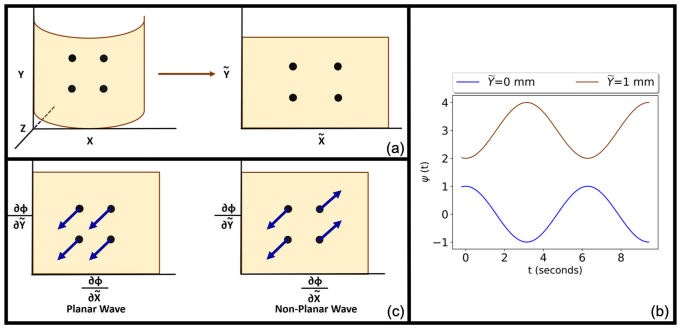

Fig. 6. (a) Cartesian coordinates (x, y, z) translated to $(\tilde{x}, \tilde{y})$ coordinates along the curved abdominal surface. Spatial analyses in Section II-B2 are performed in $(\tilde{x}, \tilde{y})$ coordinates. (b) Conceptual schematic of two out-of-phase oscillatory signals at different vertical 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.

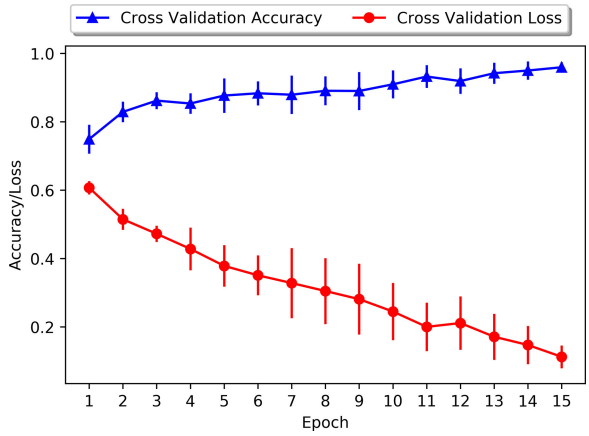

Fig. 7. Cross validation accuracy (blue) and loss (red) during training. Accuracy is shown on a [0, 1] scale.

kernel sizes throughout the network are optimal when classifying videos through a 3D CNN [61]. The kernel size was chosen by the input size of the data. State-of-the-art CNNs [56] trained on data with image sizes of 256 × 256 pixels (i.e. ImageNet) and had kernel sizes ranging from 11 × 11 in beginning layers to 3 × 3 in latter layers. Each simulated waveform in our study had an input size of 10 × 10 × 300, where the third axis was the temporal component. As such, we chose the first two dimensions of our kernel size, 2 × 2, 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 2 × 2 × 2.

Before undergoing any training, the cross validation 'accuracy' is 0.5, since CNN weights are randomly initialized. As seen in Fig. 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.

2) 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. 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. 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:

$$V_n(t) + iHb[V_n(t)] = a_n(t)e^{i\phi_n(t)}, \qquad n = 1, \ldots, N \quad (8)$$

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:

$$v(\tilde{x}, \tilde{y}, t) \propto -\nabla\phi(\tilde{x}, \tilde{y}, t). \quad (9)$$

We define features pertaining to the time-averaged direction of the wave velocity at any electrode in terms by exploiting (9) as follows:

$$\Lambda_n = \frac{1}{T}\sum_{t=1}^{T} \tan^{-1}\left(\nabla\phi(\tilde{x}_n, \tilde{y}_n, t)\right), \quad n = 1, \ldots, N. \quad (10)$$

Speed calculated from measurements directly on the stomach [14] 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 [39], [41], [74]. 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:

$$PGD(t) = \frac{\|\frac{1}{N}\sum_{n=1}^{N} \nabla\phi(\tilde{x}_n, \tilde{y}_n, t)\|_2}{\frac{1}{N}\sum_{n=1}^{N} \|\nabla\phi(\tilde{x}_n, \tilde{y}_n, t)\|_2}, \quad t = 1, \ldots, T \quad (11)$$

where in (11), velocities are replaced with $-\nabla\phi$ by virtue of (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 [39]).

A conceptual schematic of PGD is shown in Fig. 6c. The denominator of (11) represents the average magnitude of the spatial phase gradient. In Fig. 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 (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.

3) 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. 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).

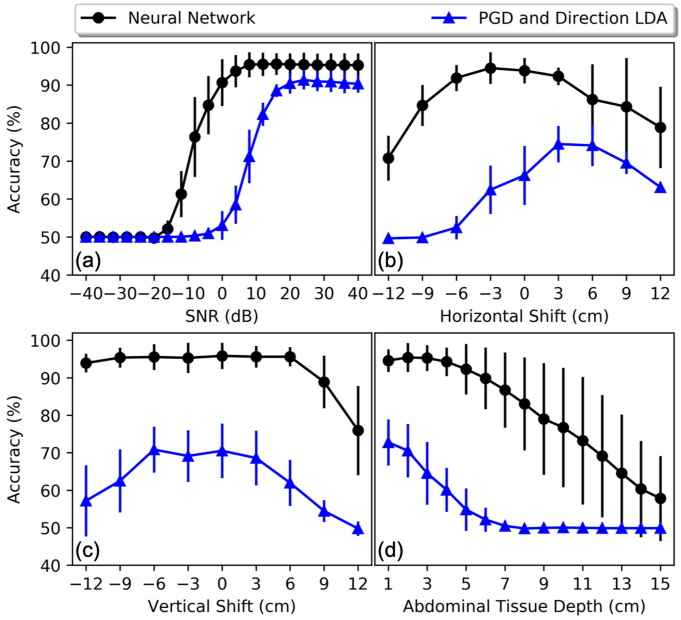

Fig. 8. Comparison of neural network performance with LDA performance. (a) Neural network and LDA performance as a function of SNR. (b) Neural network and LDA performance as a function of horizontal electrode array shifts along the curved surface of the abdomen. Shift distances correspond to arc lengths 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 II-A5).

III. Results

In both the LDA and neural network approaches, the classification accuracy saturates at high and low SNR extrema (Fig 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 10 dB [40]. 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. 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. 8b).

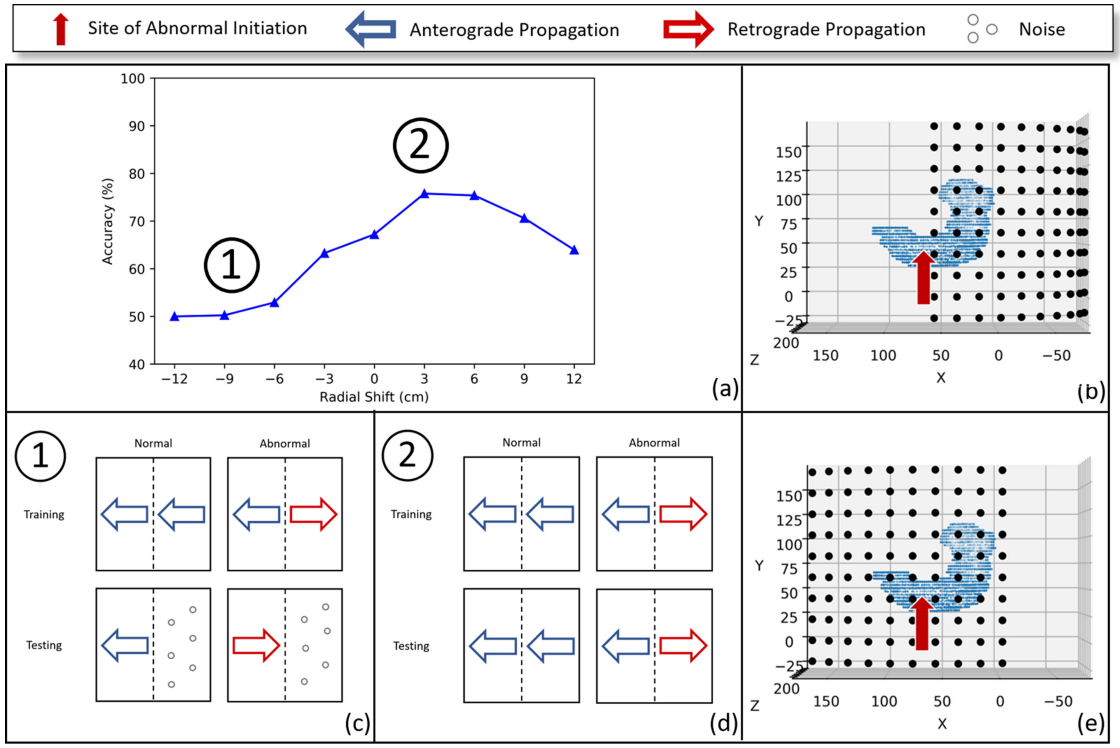

Fig. 9. Vulnerability of the LDA classifier to horizontal electrode array position. (a) LDA accuracy as a function of horizontal electrode array shifts. (b) Front-facing view of electrode array horizontally shifted by −9 cm. The electrode array is to the right of the abnormal initiation, with rightmost 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.

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 [56]. 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. 8b). In the case of the LDA, the features are specific to the spatial location of each electrode. As seen in Fig. 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. 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 Fig. 9a. At annotation 2 in Fig. 9, the array is most centered over the abnormal initiation during testing (Fig. 9e and Supplemental Materials Section E). As such, the learned representation of wave directions is very consistent with training (Fig. 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. 9, the center of the electrode array is shifted to the right of the abnormal initiation point (Fig. 9b). When testing the abnormal waveforms, the left side of the array now records retrograde activity and the right side records noise (Fig. 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 Fig. 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. 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. 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. 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. 10a). When the electrode array is shifted horizontally, the most noticeable discrepancy in performance lies in the range of -12 to -6 cm (Fig. 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. 10c),

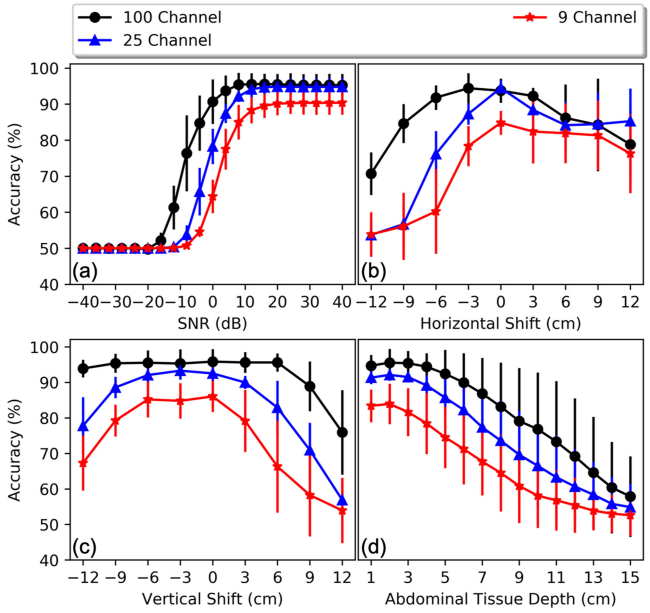

Fig. 10. Comparison of neural network accuracy between 100 channel (black), 25 channel (blue), and 9 channel (red) electrode arrays with respect to: (a) SNR, (b) horizontal electrode array shifts, (c) vertical electrode array shifts, and (d) abdominal tissue depth.

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.

IV. 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 [15]. 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 [49] 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 [75] and low-resolution simultaneous mucosal and cutaneous recordings [76]. 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 [77]–[79].

Though many state-of-the-art studies involve transfer learning of 2D models, recent work has shown efficacy of transfer learning with 3D-CNNs [80]–[82]. 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 oversampling 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 [47], as well as radiology [83], 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 [40]. 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 [40]. 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 [84] and could be used as further *in silico* validation of our presented approach.

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 [24]–[26], [85], 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.

V. 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 [41]. 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.

Acknowledgment

The authors would like to thank K. King, M.D., Ph.D., for his advisory contributions to the paper as an expert in cardiology.

References

[1] H. El-Serag and N. Talley, "The prevalence and clinical course of functional dyspepsia," *Alimentary Pharmacol. Therapeutics*, vol. 19, no. 6, pp. 643–654, 2004.

[2] H.-K. Jung *et al.*, "The incidence, prevalence, and outcomes of patients with gastroparesis in olmsted county, minnesota, from 1996 to 2006," *Gastroenterology*, vol. 136, no. 4, pp. 1225–1233, 2009.

[3] A. Gikas and J. K. Triantafillidis, "The role of primary care physicians in early diagnosis and treatment of chronic gastrointestinal diseases," *Int. J. General Med.*, vol. 7, pp. 159–173, 2014.

[4] Z. S. Heetun and E. M. Quigley, "Gastroparesis and Parkinsons disease: A systematic review," *Parkinsonism Related Disorders*, vol. 18, no. 5, pp. 433–440, 2012.

[5] M. Horowitz *et al.*, "Gastroparesis: Prevalence, clinical significance and treatment," *Can. J. Gastroenterol. Hepatol.*, vol. 15, no. 12, pp. 805–813, 2001.

[6] P. J. Pasricha *et al.*, "Characteristics of patients with chronic unexplained nausea and vomiting and normal gastric emptying," *Clin. Gastroenterol. Hepatol.*, vol. 9, no. 7, pp. 567–576, 2011.

[7] P. Janssen *et al.*, "The relation between symptom improvement and gastric emptying in the treatment of diabetic and idiopathic gastroparesis," *Amer. J. Gastroenterol.*, vol. 108, no. 9, pp. 1382–1391, 2013.

[8] R. Corinaldesi *et al.*, "Effect of chronic administration of cisapride on gastric emptying of a solid meal and on dyspeptic symptoms in patients with idiopathic gastroparesis," *Gut*, vol. 28, no. 3, pp. 300–305, 1987.

[9] 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 Pharmacol. Therapeutics*, vol. 26, no. 8, pp. 1121–1130, 2007.

[10] M. E. Barton *et al.*, "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*, vol. 146, no. 5, p. S–20, 2014.

[11] J. Langworthy, H. P. Parkman, and R. Schey, "Emerging strategies for the treatment of gastroparesis," *Expert Rev. Gastroenterol. Hepatol.*, vol. 10, no. 7, pp. 817–825, 2016.

[12] E. J. Irvine *et al.*, "Design of treatment trials for functional gastrointestinal disorders," *Gastroenterology*, vol. 150, no. 6, pp. 1469–1480, 2016.

[13] K. L. Koch and R. M. Stern, *Handbook of Electrogastrography*. London, U.K.: Oxford Univ. Press, 2004.
[14] G. O'Grady *et al.*, "Origin and propagation of human gastric slow-wave activity defined by high-resolution mapping," *Amer. J. Physiol.-Gastrointestinal Liver Physiol.*, vol. 299, no. 3, pp. G585–G592, 2010.
[15] G. O'Grady *et al.*, "Abnormal initiation and conduction of slow-wave activity in gastroparesis, defined by high-resolution electrical mapping," *Gastroenterology*, vol. 143, no. 3, pp. 589–598, 2012.
[16] Z. Lin *et al.*, "Gastric myoelectrical activity and gastric emptying in patients with functional dyspepsia," *Amer. J. Gastroenterol.*, vol. 94, no. 9, pp. 2384–2389, 1999.
[17] J. Chen *et al.*, "Abnormal gastric myoelectrical activity and delayed gastric emptying in patients with symptoms suggestive of gastroparesis," *Digestive Diseases Sci.*, vol. 41, no. 8, pp. 1538–1545, 1996.
[18] M. Bortolotti *et al.*, "Gastric myoelectric activity in patients with chronic idiopathic gastroparesis," *Neurogastroenterol. Motility*, vol. 2, no. 2, pp. 104–108, 1990.
[19] W. Sha, P. J. Pasricha, and J. D. Chen, "Rhythmic and spatial abnormalities of gastric slow waves in patients with functional dyspepsia," *J. Clin. Gastroenterol.*, vol. 43, no. 2, pp. 123–129, 2009.
[20] A. Leahy *et al.*, "Abnormalities of the electrogastrogram in functional gastrointestinal disorders," *Amer. J. Gastroenterol.*, vol. 94, no. 4, pp. 1023–1028, 1999.
[21] G. O'Grady *et al.*, "Recent progress in gastric arrhythmia: Pathophysiology, clinical significance and future horizons," *Clin. Exp. Pharmacol. Physiol.*, vol. 41, no. 10, pp. 854–862, 2014.
[22] C. Ouyang, "Clinical significance of gastric dysrhythmias," *World J. Gastroenterol.*, vol. 2, no. Suppl 1, pp. 5–6, 1996.
[23] A. C. Nanivadekar *et al.*, "Machine learning prediction of emesis and gastrointestinal state in ferrets," *PLOS One*, vol. 14, no. 10, 2019, Art. no. e0223279.
[24] T. L. Abell *et al.*, "Gastric electrical stimulation in intractable symptomatic gastroparesis," *Digestion*, vol. 66, no. 4, pp. 204–212, 2002.
[25] G. O'Grady *et al.*, "High-resolution entrainment mapping of gastric pacing: A new analytical tool," *Amer. J. Physiol.-Gastrointestinal Liver Physiol.*, vol. 298, no. 2, pp. G314–G321, 2009.
[26] S. Alighaleh *et al.*, "Real-time evaluation of a novel gastric pacing device with high-resolution mapping," *Gastroenterology*, vol. 154, no. 6, p. S–39, 2018.
[27] S. C. Payne, J. B. Furness, and M. J. Stebbing, "Bioelectric neuromodulation for gastrointestinal disorders: Effectiveness and mechanisms," *Nature Rev. Gastroenterol. Hepatol.*, vol. 16, pp. 89–105, 2019.
[28] W. Sha, P. J. Pasricha, and J. D. Chen, "Correlations among electrogastrogram, gastric dysmotility, and duodenal dysmotility in patients with functional dyspepsia," *J. Clin. Gastroenterol.*, vol. 43, no. 8, pp. 716–722, 2009.
[29] H. P. Parkman *et al.*, "Electrogastrography and gastric emptying scintigraphy are complementary for assessment of dyspepsia," *J. Clin. Gastroenterol.*, vol. 24, no. 4, pp. 214–219, 1997.
[30] K. L. Koch *et al.*, "Gastric emptying and gastric myoelectrical activity in patients with diabetic gastroparesis: Effect of long-term domperidone treatment," *Amer. J. Gastroenterol.*, vol. 84, no. 9, pp. 1069–1075, 1989.
[31] 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 J. Gastroenterol.*, vol. 37, no. 10, pp. 1144–1148, 2002.
[32] A. Oba-Kuniyoshi *et al.*, "Postprandial symptoms in dysmotility-like functional dyspepsia are not related to disturbances of gastric myoelectrical activity," *Brazilian J. Med. Biol. Res.*, vol. 37, no. 1, pp. 47–53, 2004.
[33] S. Abid and G. Lindberg, "Electrogastrography: Poor correlation with antro-duodenal manometry and doubtful clinical usefulness in adults," *World J. Gastroenterol.*, vol. 13, no. 38, pp. 5101–5107, 2007.
[34] H. Parkman *et al.*, "Electrogastrography: A document prepared by the gastric section of the american motility society clinical GI motility testing task force," *Neurogastroenterol. Motility*, vol. 15, no. 2, pp. 89–102, 2003.
[35] T. R. Angeli *et al.*, "Loss of interstitial cells of Cajal and patterns of gastric dysrhythmia in patients with chronic unexplained nausea and vomiting," *Gastroenterology*, vol. 149, no. 1, pp. 56–66, 2015.
[36] R. Berry *et al.*, "Patterns of abnormal gastric pacemaking after sleeve gastrectomy defined by laparoscopic high-resolution electrical mapping," *Obesity Surgery*, vol. 27, no. 8, pp. 1929–1937, 2017.
[37] P. Du *et al.*, "Progress in mathematical modeling of gastrointestinal slow wave abnormalities," *Frontiers Physiol.*, vol. 8, 2018, Art. no. 1136, doi: 10.3389/fphys.2017.01136.
[38] G. O'Grady *et al.*, "Methods for high-resolution electrical mapping in the gastrointestinal tract," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 287–302, 2019, doi: 10.1109/RBME.2018.2867555.
[39] A. A. Gharibans *et al.*, "High-resolution electrogastrogram: a novel, non-invasive method for determining gastric slow-wave direction and speed," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 807–815, Apr. 2017.
[40] A. A. Gharibans *et al.*, "Artifact rejection methodology enables continuous, noninvasive measurement of gastric myoelectric activity in ambulatory subjects," *Scientific Rep.*, vol. 8, no. 1, 2018, Art. no. 5019.
[41] A. A. Gharibans *et al.*, "Spatial patterns from high-resolution electrogastrography correlate with severity of symptoms in patients with functional dyspepsia and gastroparesis," *Clin. Gastroenterol. Hepatol.*, vol. 17, no. 13, pp. 2668–2677, Dec. 2019.
[42] S. Zahid *et al.*, "Patient-derived models link re-entrant driver localization in atrial fibrillation to fibrosis spatial pattern," *Cardiovascular Res.*, vol. 110, no. 3, pp. 443–454, 2016.
[43] C. T. Villongco *et al.*, "Non-invasive, model-based measures of ventricular electrical dyssynchrony for predicting CRT outcomes," *EP Europace*, vol. 18, pp. iv104–iv112, 2016.
[44] G. Ho *et al.*, "Rotors exhibit greater surface ECG variation during ventricular fibrillation than focal sources due to wavebreak, secondary rotors, and meander," *J. Cardiovascular Electrophysiology*, vol. 28, no. 10, pp. 1158–1166, 2017.
[45] U. R. Acharya *et al.*, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, pp. 81–90, 2017.
[46] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.
[47] U. R. Acharya *et al.*, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Medicine*, vol. 89, pp. 389–396, 2017.
[48] S. P. Shashikumar *et al.*, "A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2017, pp. 141–144.
[49] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.
[50] O. Faust *et al.*, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomedicine*, vol. 161, pp. 1–13, 2018.
[51] U. R. Acharya *et al.*, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Medicine*, vol. 100, pp. 270–278, 2018.
[52] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of EEG signal," in *Proc. Int. Joint Conf. Neural Netw.*, 2014, pp. 2850–2853.
[53] P. W. Mirowski *et al.*, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG," in *Proc. IEEE Workshop Mach. Learn. Signal Process.*, 2008, pp. 244–249.
[54] U. C. Allard *et al.*, "A convolutional neural network for robotic arm guidance using SEMG based frequency-features," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 2464–2470.
[55] P. Xia, J. Hu, and Y. Peng, "EMG-Based estimation of limb movement using deep learning with recurrent convolutional neural networks," *Artificial Organs*, vol. 42, no. 5, pp. E67–E77, 2018.
[56] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[57] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2015, pp. 1–9.
[58] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 3642–3649.
[59] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556v6*.
[60] S. Ji *et al.*, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
[61] D. Tran *et al.*, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 4489–4497.
[62] D. Maturana and S. Scherer, "Voxnet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 922–928.

[63] F. S. Nooruddin and G. Turk, "Simplification and repair of polygonal models using volumetric techniques," *IEEE Trans. Visualization Comput. Graph.*, vol. 9, no. 2, pp. 191–205, Jun. 2003.
[64] P. Min, "binvox," 2004–2019. [Online]. Available: http://www.patrickmin.com/binvox, Accessed on: May 1, 2018.
[65] K. Palágyi and A. Kuba, "Directional 3D thinning using 8 subiterations," in *Proc. Int. Conf. Discrete Geometry Comput. Imagery*, Springer, 1999, pp. 325–336.
[66] P. Min, "thinvox," 2017. [Online]. Available: http://www.patrickmin.com/thinvox, Accessed on: May 1, 2018
[67] R. Plonsey, "Bioelectric phenomena," in *Wiley Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ, USA: Wiley 2001.
[68] K. Roeleveld *et al.*, "Volume conduction models for surface EMG; Confrontation with measurements," *J. Electromyogr. Kinesiol.*, vol. 7, no. 4, pp. 221–232, 1997.
[69] B. J. Roth and F. L. Gielen, "A comparison of two models for calculating the electrical potential in skeletal muscle," *Ann. Biomed. Eng.*, vol. 15, no. 6, pp. 591–602, 1987.
[70] T. Gootzen *et al.*, "A comprehensive model description for motor unit action potentials," *Electroencephalogr. Clin. Neurophysiol*, vol. 66, no. 5, p. S40, 1987.
[71] N. Srivastava *et al.*, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[72] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: http://tensorflow.org/
[73] F. Chollet *et al.*, "Keras," 2015. [Online]. Available: https://keras.io
[74] L. Bradshaw *et al.*, "Diabetic gastroparesis alters the biomagnetic signature of the gastric slow wave," *Neurogastroenterol. Motility*, vol. 28, no. 6, pp. 837–848, 2016.
[75] N. Paskaranandavadivel *et al.*, "Multi-day, multi-sensor ambulatory monitoring of gastric electrical activity," *Physiological Meas.*, vol. 40, no. 2, 2019, Art. no. 025011.
[76] A. Shine *et al.*, "Electrogastography at baseline and response to temporary gastric electrical stimulation–a comparison of cutaneous with mucosal recordings," *Digestive Disease Week*, vol. 156, pp. S30–S31, 2019.
[77] A. Sharif Razavian *et al.*, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2014, pp. 806–813.
[78] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
[79] J. Yosinski *et al.*, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
[80] S. Hussein *et al.*, "Risk stratification of lung nodules using 3D CNN-based multi-task learning," in *Information Processing in Medical Imaging*. Cham, Switzerland: Springer, 2017, pp. 249–260.
[81] P. Molchanov *et al.*, "Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 4207–4215.
[82] L. A. Alexandre, "3D object recognition using convolutional neural networks with transfer learning between input channels," in *Intelligent Autonomous Systems 13*, E. Menegatti, N. Michael, K. Berns, and H. Yamaguchi, Eds. Cham, Switzerland: Springer, 2016, pp. 889–898.
[83] H. Salehinejad *et al.*, "Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process*, 2018, pp. 990–994.
[84] S. Calder *et al.*, "Torso-tank validation of high-resolution electrogastrography (EGG): Forward modelling, methodology and results," *Ann. Biomed. Eng.*, vol. 46, pp. 1183–1193, 2018.
[85] Y.-K. Lo *et al.*, "A wireless implant for gastrointestinal motility disorders," *Micromachines*, vol. 9, no. 1, p. 17, 2018, doi: 10.3390/mi9010017.