Spatio-Temporal Characterization of Gastric Distensibility in Upper Endoscopy Identifies the Presence of *Helicobacter pylori*

Josué Ruano-Balseca, Diego Bravo, *Graduate Student Member, IEEE*, Diana L. Giraldo, Martín Gómez, Fabio A. González, and Eduardo Romero

***Abstract*— Assessment of gastric distensibility is crucial for understanding various gastrointestinal disorders, particularly those affecting gastric motility and reflux conditions. Gastric distensibility can be altered by pathological conditions such as linitis plastica or gastric mucosal inflammation, as seen in *Helicobacter pylori* infections. Although distensibility is evaluated during routine endoscopy, the assessment remains purely visual. Moreover, some specialized endoscopic tools assess only specific gastric regions and are often unavailable. Artificial intelligence tools may used to estimate the gastric three-dimensional surface from the endoscopy video and capture temporal distensibility changes of the gastric surface. This paper introduces a novel spatio-temporal characterization of the gastric distensibility during a routine endoscopy procedure. The method reconstructs the gastric surface for each frame of a short endoscopy video using a shape-from-shading network trained with a synthetic endoscopy collection. Local and regional changes in surface bending are then quantified, which herein depend on the distensibility. This spatio-temporal characterization was applied to a group of 160 patients, among them 60 diagnosed with *Helicobacter pylori* confirmed by pathology examination. The obtained characteristics in three different regions: body, antrum, and pylorus, demonstrated significant differences between patients with and without *Helicobacter pylori* under a t-test ($p < 0.05$). The proposed method effectively captures subtle gastric wall bending from routine endoscopy videos, providing a strategy for analyzing distensibility in pathological conditions across various gastric regions without requiring specialized tools.**

***Index Terms*— Endoscopy, gastric distensibility, *Helicobacter pylori*, spatio-temporal characterization, three-dimensional reconstruction.**

I. Introduction

A ROUTINE upper-endoscopy or esophagogastroduodenoscopy procedure (EN) typically informs about the stomach distensibility, i.e. the stomach's ability to stretch or shrink in response to inner air insufflation, since this can be altered by diffuse infiltrative processes such as the linitis plastica, some lymphomas, or gastric mucosal inflammation produced by *Helicobacter pylori* (*H. pylori*) infection [1], [2], [3]. In a routine EN procedure, the gastroenterologist assesses distensibility by observing how the gastric rugae are flattened in response to air insufflation. Gastric distention is reported as normal if the endoscopist observes how the gastric wall passively follows insufflation and no resistance is noticed [3]. Evaluation of this particular motion, yet passive, highly drives the patient management when studying symptoms like bloating, swelling, postprandial fullness, early satiety, abdominal pain, chronic nausea or vomiting. Gastric distention assessment is usually a first step in the study of any upper gastrointestinal motility disorder, typically accompanied by biopsies searching for histological changes. Nevertheless, this evaluation is completely subjective and this is a starting point of this study, a first endeavor to objectively quantify the reported distensibility.

Alternative techniques for assessing gastric distension, such as barostat [4] and EndoFLIP [5], both accessory tools of the endoscope, provide localized pressure measurements using ballooned sensors with piezoelectric or impedance transducers. The resultant measures are then the consequence of the reaction force between two surfaces, the balloon and the gastric wall. Likewise, they require precise anatomical targeting, typically at sites like the gastric sphincters, esophagogastric junction and pylorus, and in the gastric fundus. Despite their utility, these methods yield only local

Received 19 October 2025; revised 28 December 2025; accepted 5 March 2026. Date of publication 9 March 2026; date of current version 3 June 2026. This work was supported by the Program for the Early Detection of Premalignant Lesions and Gastric Cancer in Urban, Rural and Dispersed Areas in the Department of Nariño of call 920 of 2022 of MinCiencias under Project 110192092354. The work of Diana L. Giraldo was supported by the Flemish Government under the "Flanders Artificial Intelligence Research" Program. Recommended by Associate Editor Q. Dou. *(Corresponding author: Eduardo Romero.)*

JosuéRuano-Balseca, Diego Bravo, and Eduardo Romero are with the Computer Imaging and Medical Applications Laboratory (CIM@LAB), Universidad Nacional de Colombia, Bogotá 111321, Colombia (e-mail: jaruanob@unal.edu.co; dbravoh@unal.edu.co; edromero@unal.edu.co).

Diana L. Giraldo is with the imec-Vision Laboratory, Department of Physics, University of Antwerp, 2000 Antwerp, Belgium (e-mail: diana.giraldofranco@uantwerpen.be).

Martín Gómez is with the Gastroenterology Unit, Hospital Universitario Nacional de Colombia, Bogotá 111321, Colombia, and also with the Internal Medicine Department, Universidad Nacional de Colombia, Bogotá 111321, Colombia (e-mail: magomezz@unal.edu.co).

Fabio A. González is with the Machine Learning, Perception and Discovery Laboratory (MindLab), Universidad Nacional de Colombia, Bogotá 111321, Colombia (e-mail: fagonzalezo@unal.edu.co).

Digital Object Identifier 10.1109/TMI.2026.3672198

pressure measurements that fail to represent the stomach's overall distensibility. Moreover, they are often unavailable for routine clinical practice and miss the global-local gastric relations which are essential for understanding impaired distension.

Although distensibility may be understood as the ability of an organ to adapt its size and shape to several conditions, in the case of the stomach this is particularly associated with the elasticity of the gastric tissue. As mentioned, this is subjectively assessed during the endoscopic procedure and any alteration usually impacts the gastric distensibility. Over the past three decades, numerous studies have explored how *H. pylori* affects gastric motility and sensory functions, potentially leading to delayed gastric emptying (GE) or impaired gastric accommodation. Although gastric motility has been reported to be affected by *H. pylori* [6], [7], [8], neither study has demonstrated this effect is significant [9] nor directly related with the infection, at least using ultrasound examination [10], [11]. Nevertheless, recent research conducted in animal models has demonstrated the *H. pylori* significantly accelerated GE [6], hypomotility [7] or delayed emptying [8]. These contradicting results reflect the fact that most of these methods are contaminated by many different types of noise, from patient heterogeneity to environmental influences [5], [8], [12], ending up by assessing uncontrolled scenarios which easily lead to inconclusive interpretations. Finally, very little is known about the relationship between *H. pylori* infection and distensibility distortion, basically because it is difficult to assess during an invasive procedure with minimal available procedure times.

Overall, gastric distensibility is a routine part of the stomach assessment, but this is only a visual opinion which in most cases has no major repercussions. Yet some mechanical devices might approximate how a stomach passively contracts after the organ is insufflated, such implementation is not a minor issue. Instead, the present investigation introduces an AI-based method to reconstruct a $3d$ gastric surface from routine endoscopy videos using a shape-from-shading network trained with synthetic data. Afterwards, a novel dynamic descriptor captures the temporal spatial bending change of a small surface element from the gastric reconstructed surface. This descriptor was computed for a group of 160 subjects, 60 among them with pathology-confirmed *Helicobacter pylori*, and significantly local and regional differences were demonstrated. This investigation introduces a novel quantification of the gastric wall distensibility from the routine endoscopy procedure.

This work presents two main contributions:

- A model which approximates distensibility by the temporal bending change of the gastric wall using only a preselected small video sequence as illustrated in Fig. 1.
- A thorough evaluation in a real population with and without *H. pylori* diagnosed by gastric biopsy and pathological examination. The test demonstrated significant differences between the two groups, applying the presented endoscopic assessment of distensibility.

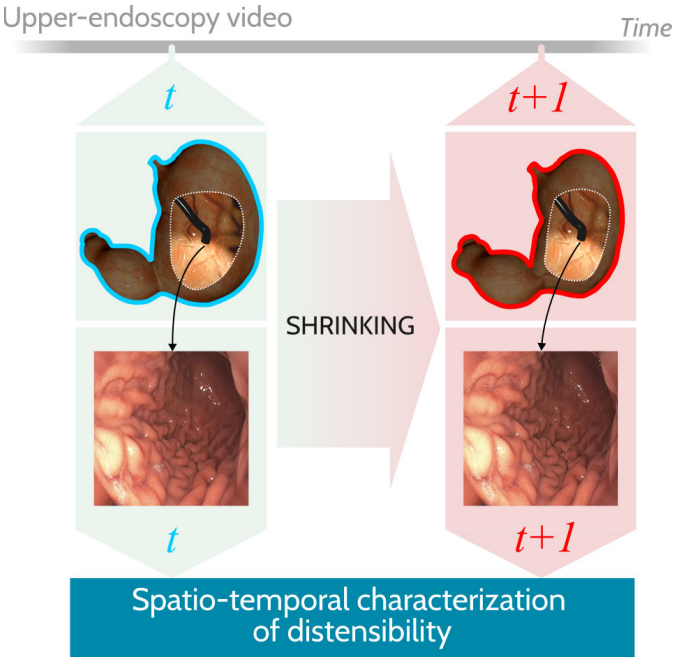

Fig. 1. Overall approach: When exploring the gastrointestinal tract, endoscopists navigate by abrupt transitions between noisy images of the gastric wall, and in certain regions they stabilize the device and capture several frames following a screening protocol. During these steady moments, camera motion is negligible, and changes reflect gastric wall movements in response to insufflation. The two consecutive frames in the figure, taken from the middle-lower gastric body, show minor structural differences. Notice how the same structure presents little differences but some deformation is observed at the left lateral rugae while the center perspective of the gastric wall has changed. As demonstrated later, conditions like *Helicobacter pylori* infection change this response.

II. Materials and Methods

The proposed methodology consists of five steps and Fig. 2 summarizes the process:

1) A short EN sequence is chosen from anatomical sites after a systematic screening protocol (SSS) [13]. Two consecutive frames from the whole sequence, f_t and f_{t+1} are selected if they show the smallest apparent motion estimated by a dense optical flow.
2) A custom neural network is trained with synthetic video frames to learn a shape-from-shading model that converts brightness variations in actual EN frames into a $3d$ surface representation. The resultant surface is divided in a fixed number of elements. Each element is tracked from f_t to f_{t+1} using the displacement computed from an optical flow-based strategy.
3) Each element is characterized both regionally and locally at times t and $t+1$, regionally by describing the impact of the neighborhood in the element bending, and locally, by establishing how bent the element is. Distensibility is captured by a spatio-temporal characterization of local and regional bending tensors.
4) Statistical analysis is conducted to assess whether the proposed spatio-temporal characterization reveals significant differences in distensibility between the *H. pylori*-positive and -negative groups.

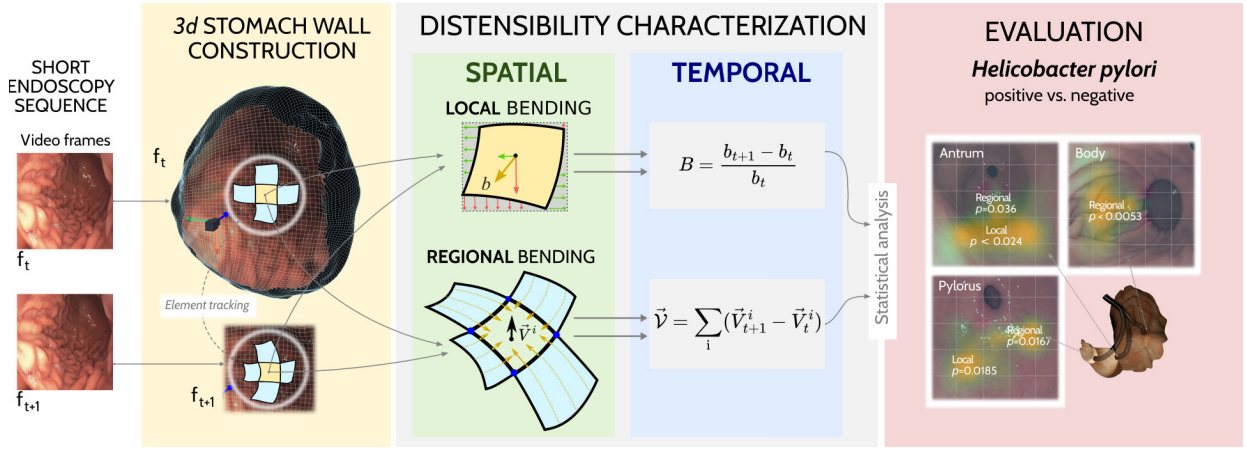

Fig. 2. Overview of the proposed approach: two consecutive frames, f_t and f_{t+1}, are transformed to depth maps which are then used to reconstruct 3d surfaces and each is split into elements (yellow block). Distensibility is quantified at each element, first spatially by approximating local and regional bending, and then temporally by computing the bending variation (gray block). Finally, a statistical analysis shows this characterization is significantly different in pathological conditions: the demonstrated presence of *Helicobacter pylori* (red block).

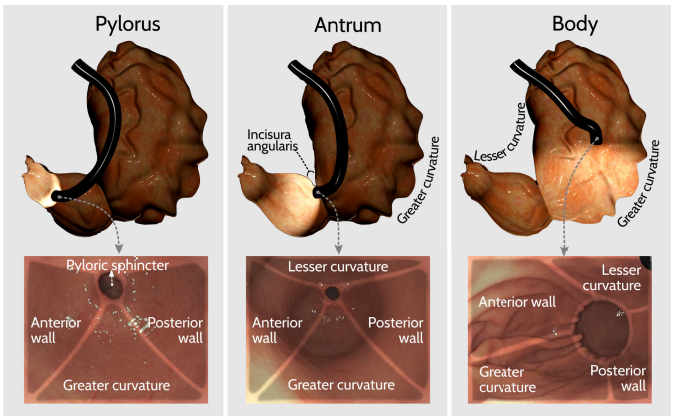

Fig. 3. The three gastric locations selected for spatio-temporal characterization. The first row shows the approximate position of the endoscope tip in the stomach model. The second row displays the endoscopic frames extracted from each anatomical site.

A. Anatomical Sites and Sequence Extraction

As classically described, *H. pylori* is initially concentrated in the pylorus and antrum, before progressing to the stomach body [14], [15]. Three anatomical locations were then selected: the pylorus, antrum, and lower stomach body. The video sequences were provided by a gastroenterologist, who needs to stabilize the camera view from 1 to 3 seconds to ensure the picture is not blurred at each of these sites following Yao's protocol, a SSS for the stomach [13]. The anatomical sites are identified as follows:

- *Pylorus:* Yao's protocol requires the endoscope is positioned to view the pylorus and about 5 *cm* away from it, as shown in left panel of Fig. 3.
- *Antrum:* The endoscope tip is typically looking at the pylorus, at a distance of about 10 *cm* (near the *incisura angularis*), to visualize the entire antrum and collect biopsy samples [16]. At this point, the camera typically visualizes the posterior, anterior, and greater curvatures of the antrum, as shown in central panel of Fig. 3.
- *Stomach body:* The gastroenterologist evaluates the stomach from approximately 20 *cm* away from the pylorus, positioning the endoscope tip at the lesser curvature, after Yao's protocol. From this camera view, the middle/lower section of the stomach body is captured, as shown in the right panel of Fig. 3. In addition, the anterior and greater curvatures cover most of the image.

Once an approximated temporal localization of each anatomical site is identified in the video, two consecutive frames are selected. For doing so, a dense optical flow [17] is computed between consecutive frames of the selected sequence. This dense optical flow estimates the apparent motion or displacement of the scene, calculated by tracking changes of the pixel intensity between consecutive frames. The result is a motion vector field from which the average magnitude serves to choose the pair of frames (f_t, f_{t+1}) with the lowest average magnitude. Finally, each case is documented with three short video sequences, one for each anatomical site.

B. Three-Dimensional Reconstruction of the Stomach Wall

Once f_t and f_{t+1} frames are extracted, three dimensional reconstructions are performed for both frames. Given the stable camera position and orientation, variations between these frames are mainly attributed to bending changes. Three dimensional reconstruction of the stomach wall is achieved for a particular frame of the endoscopy video as follows:

- estimate the frame depth map using the shape-from-shading model (see section II-B.1).
- project each pixel from 2d image coordinates into 3d real-world coordinates to generate a point cloud, using the estimated depth map and camera parameters (see section II-B.2)
- build up a stomach surface with quadrilateral cells by connecting the cloud points, a process applied to both f_t and f_{t+1} (see section II-B.3)

Finally, the surface is divided into elements and each is tracked from f_t to f_{t+1} for later spatio-temporal analysis (see section II-B.4). Fig. 4 shows the overview of the three steps required to reconstruct the stomach wall. These steps hereafter:

1) Depth Estimation of Gastric Sequences: The depth map of an EN frame approximates the 3d wall of the stomach by

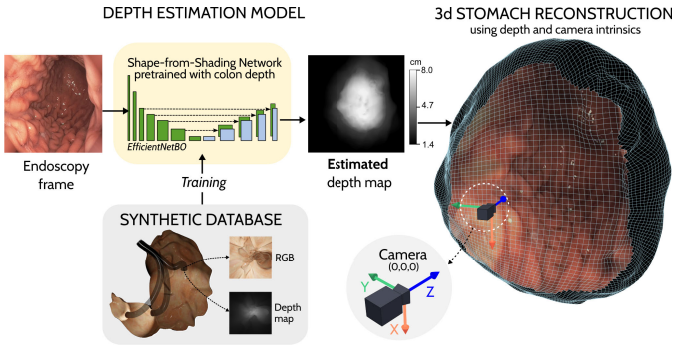

Fig. 4. The 3*d* reconstruction of the stomach wall involves: a) estimating stomach wall depth, b) transform each depth RGB value to the 3*d* camera coordinates and build the gastric surface.

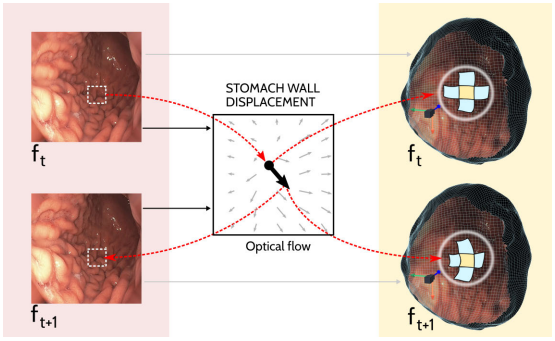

Fig. 5. The optical flow tracks an element from f_t to f_{t+1} by matching their pixel intensities and pairing the most similar regions.

estimating the distance of each visible point of the stomach to the camera's focal plane. Depth is learned from shading variations, a technique known as shape-from-shading which solves the classic ill-posed problem of projecting an infinite set of 3*d* points onto a single 2*d* point, by a set of constraints to regularize the solution:

- the camera follows a pinhole model with a fixed focal length
- the light source is positioned near the optical center of the camera
- most of the mucosa's reflections are Lambertian

These constraints simplify the problem by making the light intensity a distance function, following the inverse square law as attenuation model. Under these conditions, a deep neural network learns a depth map from a single RGB image using a pixel-level regression task. This regression is carried out by supervisedly learning the Shape-from-Shading network (*SfSNet*) with a custom network, the *SfSNet*, an encoder-decoder architecture with the EfficientNetB0 as encoder and four up-sampling blocks in the the decoder component [18]. A customized loss function $\mathbf{L}$ balances the low frequencies of the depth map by the L_z and preserves high-frequency details with the L_e and L_c terms, the edges (gradient ∇) and curves (Hessian matrix H) of the depth maps as:

$$\mathbf{L}(d,\hat{d}) = w_1 L_z(d,\hat{d}) + w_2 L_e(\nabla(d),\nabla(\hat{d})) + w_3 L_c(H(d),H(\hat{d})) \quad (1)$$

being d the ground truth depth and $\hat{d}$ the estimated depth. $w_1 = 0.1$, $w_2 = 0.3$ and $w_3 = 0.6$ were set by exhaustive search of the optimal parameters.

2) Projection From 2d to 3d: The estimated depth map of the RGB image is used to project from the 2*d* image coordinate system to the 3*d* world coordinate system. Therefore, a pixel in 2*d* coordinates (u, v) with deprmm-cu11 == 24.8.*th d is mapped to the $(x, y, z)3d$ world coordinates in millimeters using the intrinsic parameters of the camera, as follows $\begin{pmatrix} x & y & z \end{pmatrix}^T = dK^{-1} \begin{pmatrix} u & v & 1 \end{pmatrix}^T$ with:

$$K^{-1} = \begin{pmatrix} 1/f & 0 & -c_x/f \\ 0 & 1/f & -c_y/f \\ 0 & 0 & s \end{pmatrix} \quad (2)$$

where f, the focal distance, is set to $0.631\,cm$ (231.631 pixels) corresponding to the endoscope wide angle of 140°, a scale factor $s = 1$, and the focal center $(c_x, c_y) = (0, 0)$. The result of this transformation is a cloud of points.

3) Mesh Construction and the Set of Elements: Once the cloud of points is obtained, these points are vertices of quadrilaterals which form the stomach mesh. Since each vertex corresponds to a pixel in the EN frame, the 4-neighborhood of each pixel provides connections or edges between vertices, producing quadrilateral mesh cells. The resulting mesh represents the 3*d* reconstruction of the stomach wall.

This 3*d* surface is split (see Fig. 5) in a set of elements used to estimate how distensibility changes. Notice each element is oriented: the major axis is usually parallel to any of the two stomach curvatures while the minor axis is orthogonal. The number of elements is beforehand fixed and depends on the type of evaluation, as explained in section III-C.3.

4) Element Tracking: The spatio-temporal descriptor is obtained by tracking an element from f_t to f_{t+1}. Correspondence between elements is set by an optical flow strategy [19] applied to a 40×40 window centered at the pixel under analysis. The pixel intensities within this window are compared to those in a window of the same size in f_{t+1}. Paired pixels are determined by finding the smallest intensity difference (see Fig. 5).

C. Spatio-Temporal Characterization of Distensibility

The stomach is a hollow organ which constantly accommodates to break down food. This motion depends on the organ elasticity, the number of folds and their location, and the muscle arrangement of the gastric wall (*gastric rugae*). Overall, the tissue expands in response to the volume of partially digested food and the contraction pattern of the *muscularis propria*, which regulates stomach motility, or peristalsis. Typically, distensibility is hidden by the complex interaction between gastric container/contents [20], so this particular characteristic is subjectively evaluated during EN, when the stomach is exposed to radial forces by insufflation and the gastric wall motion is basically passive. The presented spatio-temporal characterization at both local and regional levels aims to quantify this passive motion as hereafter explained.

1) Local Bending: Local distensibility is herein approximated by tracking the temporal bending change of the gastric surface for a particular element. For so doing, an element of the stomach surface is projected onto a 2*d* plane defined by the two largest eigenvectors of the covariance matrix computed

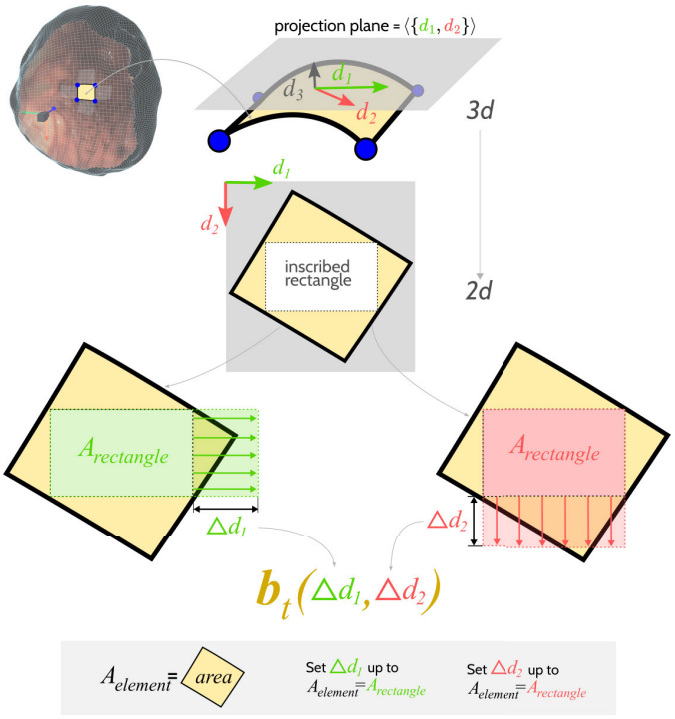

Fig. 6. Local bending is approximated by tracking the temporal bending of a gastric surface element. Spatial bending is estimated by first projecting the element into the 2d plane defined by the two largest vectors of the covariance matrix of the element coordinates. The bending, denoted as b, is calculated as the change in direction d that makes the inscribed rectangle's area equals the element's area.

from the cloud of points of the element. In this reference frame, d_1 and d_2 largest variance directions typically point out to the element's diagonals. Once the element is projected, a rectangle inscribed within this projection follows the reference directions d_1 and d_2, and the spatial bending, denoted as b, is estimated as the rectangle proportion in a given direction d between the original length and the length that makes the rectangle area equals the element area (see Fig. 6). This estimation is performed for f_t and f_{t+1} obtaining $b_t(d)$ and $b_{t+1}(d)$. Finally, the temporal local bending $B(d)$ is computed as:

$$B(d) = \frac{b_{t+1}(d) - b_t(d)}{b_t(d)} \tag{3}$$

This bending B in the d_1 and d_2 directions, normalized by its magnitude, is hereafter referred to as $Local_{longitudinal}$ and $Local_{transversal}$ respectively.

2) Regional Bending: Once an element of the gastric surface is chosen, the element bending associated with the four connected neighbors approximates spatial regional bending, the blue surface in Fig. 7-a. Basically, each of the connected neighbors is split into a set of narrow strips perpendicular to the inter-element boundary (yellow line), each described by the corresponding spatial gradients, the blue vectors in Fig. 7-b. Each of these strips is then represented by a 3d vector $\vec{v}$, which is the sum of the spatial gradients, the red vectors in Fig. 7-b. The complete element boundary of f_t, connected to the four direct neighbors, is described by the vector $\vec{V}_t$, which is the summation of the obtained red vectors. This spatial regional bending is also computed for f_{t+1}, resulting in $\vec{V}_{t+1}$.

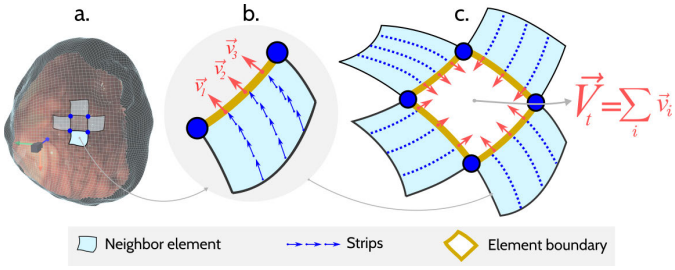

Fig. 7. The spatial regional bending of the surface element in f_t is approximated by the bending associated with its four neighbors, each being divided into narrow strips which are described by the series of blue small vectors, whose summation is represented by the red vectors in the yellow inter-element boundary. The sum of red vectors corresponds to the final element representant $\vec{V}_t$. This spatial regional bending is also computed for f_{t+1}, resulting in $\vec{V}_{t+1}$.

Finally, the temporal regional bending is simply the difference between these two 3d vectors $\vec{V}_{t+1}$ and $\vec{V}_t$, and the resultant vector is normalized by its magnitude. This vector summarizes the influence of the direct neighbors in three component: $\mathcal{V}_x$ for $x - axis$, $\mathcal{V}_y$ for $y - axis$, and $\mathcal{V}_z$ for $z - axis$, from now on referred to as $Regional_x$, $Regional_y$, and $Regional_z$ respectively.

D. Statistical Analysis

The proposed spatio-temporal characterization was compared between the two groups, *H. pylori*-positive and *H. pylori*-negative, using t-tests. For each feature, the null hypothesis assumes the two groups have equal mean feature value, implying observed differences are due to random variation. The significance level, i.e., the probability of incorrectly rejecting the null hypothesis, is thresholded at $p < 0.05$.

Since each element is analyzed with five features, two local and three regional, the risk of false positives is controlled by applying False Discovery Rate (FDR) correction across the five p-values per element. This correction consists of adjusting the p-values following the Benjamini-Hochberg procedure. First, the original p-values are sorted in ascending order $p_1 \leq p_2 \leq \ldots \leq p_m$, where m is the number of tests. Then, the adjusted p-values are computed as $p'_i = min\left(1, \frac{p_i * m}{i}\right)$, which represents the p-value scaled by the ratio of its rank to the total number of tests. These element-wise adjusted p-values are compared against the type I error threshold of 0.05 ($p < 0.05$) to determine for which features the group difference remains statistically significant.

E. Datasets

1) Endoscopy Procedures for Gastric Distensibility Analysis: The proposed spatio-temporal characterization and the subsequent statistical analysis were applied to 160 EN videos from a public database [21], [22]. Inclusion criteria were: (i) a signed informed consent form, and (ii) histopathology examination to confirm the presence or absence of *H. pylori* following Sydney protocol [16], i.e., five biopsy samples at specific locations: two from pylorus, one from *incisura angularis*, and two from stomach body. Likewise, exclusion criteria included cases without clinical history or conclusive pathological diagnosis. The cohort comprises 160 cases from the Hospital

Universitario Nacional de Colombia collected from June 2019 to August 2023. Among them, 100 subjects were reported as negative for *H. pylori* and 60 as positive. The positive group comprises 11 males and 24 females, with a mean age of 64.2±14.3 years, and the negative group includes 16 males and 44 females, with a mean age of 66.7 ± 11.0 years. Each of the EN procedures was conducted following a SSS that included 22 anatomical sites [13]. The endoscopist ensured that the gastric mucosa was clean before examining specific anatomical sites. During intubation, any mucus was cleared with the water jet, and any injected water that temporarily accumulates in the fundus was fully aspirated. Additionally, patients received a preparation solution containing N-acetylcysteine and simethicone, which reduced mucus viscosity and improved mucosal visibility. Consequently, no fluid accumulation was observed at the anatomical sites analyzed in this study.

The public video collection used in this work received approval from the Institutional Review Board of the Research Ethics Committee at the Hospital Universitario Nacional de Colombia (No. CEI-2019-06-10), and adhered to the Declaration of Helsinki. Written informed consent was obtained from all participants or their authorized representatives. Each image was anonymized to safeguard participant identity [23]. Each video was captured at 15 frames per second with a spatial resolution of 1920 × 1080 pixels. After removing the typical black borders of endoscopic images, the effective resolution was 1376 × 1076 pixels.

2) Collection of Synthetic Endoscopy Videos for Three-Dimensional Reconstruction: The depth estimation model used for three dimensional reconstruction of the stomach wall was trained, validated, and tested using a publicly available collection of 50 synthetic EN videos [24]. Further details of these partitions are in section III-A. Videos were virtually rendered to construct 3*d* meshes using average stomach measurements from 50 adult cadavers [25]. Real textures of the stomach taken from 50 videos were mapped to the 3*d* meshes for mimicking real appearances. Gastric rugae were approximated by synthetic patterns parameterized according to anatomical configurations from each stomach region. The 7-minute average duration of an actual EN withdrawal was virtually simulated, moving a virtual camera and light source following a path after the SSS with 22 locations [13]. During this virtual exploration, 31,500 RGB synthetic frames with their respective depth maps were generated by respecting the camera's perspective at a spatial resolution of 1280 × 1280.

III. Evaluation and Results

First, performance of depth estimation and three-dimensional reconstruction is presented in section III-A. Second, results of statistical analysis between *H. pylori*-positive and -negative groups from two-fold configurations of the spatio-temporal characterization in sections III-D and III-E.

A. Training and Evaluation of the Three-Dimensional Reconstruction

The depth estimation model (*SfSNet*), described in section II-B.1, was previously validated in colonoscopy images [18], [26]. *SfSNet* was trained using a synthetic colonoscopy dataset and validated at three levels: a) qualitative and quantitative evaluation with twelve videos from a synthetic dataset [18], b) comparison with five state-of-the-art methods in a challenging real dataset composed of 196 frames extracted from 34 videos [27], [28], [29], [30], [31], and c) quantitative validation using two CT-based synthetic collections which provide real distances given by the real size of voxels [18], [28]. Additionally, the estimated depth maps were used to estimate the size of several polyps in millimeters using 100 real cases [26]. *SfSNet* obtained reliable performance, accurately capturing the shape of folds and small abnormalities, while providing a 3*d* representation of gastrointestinal tract in real world coordinates. The colonoscopy pre-trained weights were used to initialize *SfSNet* for this work [18]. These weights were then fine-tuned using a synthetic set of upper EN videos, as described in section II-E.2. The synthetic dataset provides 31,500 video frames with depth maps generated by placing a virtual camera in the scene and from which distances were estimated. The synthetic dataset was split into 70% (22,050 frames) for training, 10% (3,500 frames) for validation, and 20% (6,300 frames) for testing. During training, the network was configured with the Adam optimizer and a learning rate decay that reduced the learning rate by a factor γ after p consecutive epochs without improvement in the validation loss. Additionally, five common network hyperparameters were optimized over 40 trials to prevent overfitting or underfitting, issues frequently reported with synthetic datasets. In addition, a Bayesian optimizer was used to find the optimal combination of the following hyperparameters:

- *Batch size:* Tested from 4 to 10.
- *Initial learning rate:* Tested from 10^{-6} to 10^{-3} in steps of 10^{-6}
- *Weight decay regularization:* This consists in adding the term $\omega\|W\|^2$ to the loss function, to bind the range of the weight values W. Likewise ω was tested from 10^{-6} to 10^{-4} in steps of 10^{-6}
- *Reduction factor λ of learning rate decay:* Tested from 0.4 to 0.9.
- *Patience p of learning rate decay:* Tested from 4 to 10 epochs.

SfSNet performance was assessed in the testing set of the synthetic collection of upper EN videos using two pixel-wise metrics, namely Root mean squared error (*RMSE*) and Threshold accuracy (*ThAcc*), by comparing the estimated depth image y and the ground truth $\hat{y}$. Both metrics are described below:

- *Root mean squared error (RMSE), in centimeters*:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

- *Threshold accuracy (ThAcc), in %:*

$$\frac{100 \times \left|\left\{i; \max\left(\frac{y_i}{\hat{y}_i}, \frac{\hat{y}_i}{y_i}\right) < \delta\right\}\right|}{n} \tag{5}$$

where δ is a decision threshold of the y and $\hat{y}$ ratio, herein set to 1.25 as the strictest used in the literature. For both

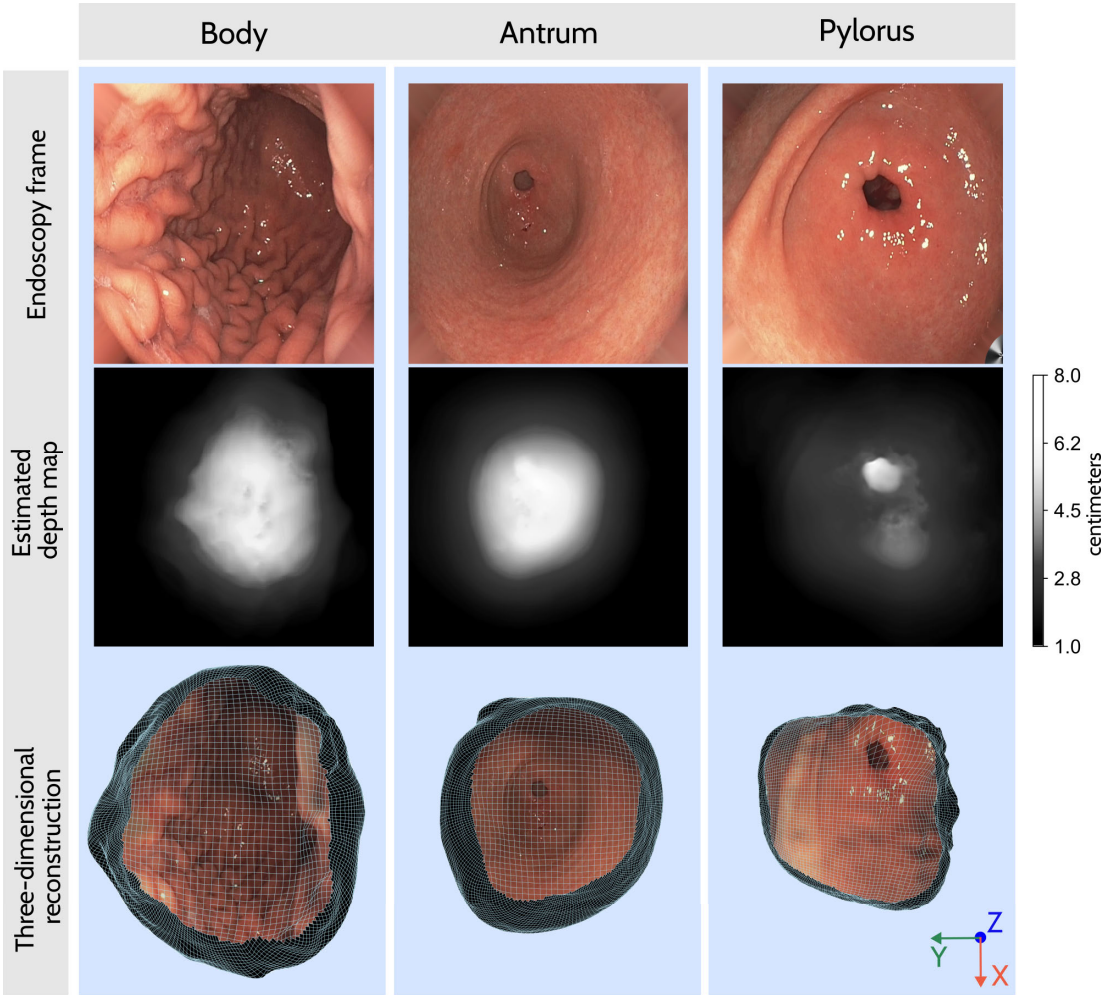

Fig. 8. Depth map and three-dimensional reconstruction of frames from body, antrum, and pylorus.

TABLE I
COMPARISON BETWEEN THE DEPTH ESTIMATION MODEL USED IN THIS WORK, SFSNET, WITH TWO STATE OF THE ART STRATEGIES UNDER THE SAME VALIDATION SCHEME

Method	*RMSE* (mm)	*ThAcc* (%)
SfSNet (proposed)	3.66	98.6
EndoDAC [32]	7.72	87.8
Budd et al. [33]	10.99	68.3

metrics, i is the pixel index, n is the total number of pixels.

Table I compares results obtained by *SfSNet* and two state-of-the-art methods [32], [33], all trained and evaluated with the same training and testing sets, to provide an additional validation of the depth estimation method. *SfSNet* reliably estimated depth in upper EN, achieving an *RMSE* of 3.66 mm and a *ThAcc* of 98.65%, outperforming the other two methods. It is important to note that quantitative evaluation in real images is not possible, since ground truth depth maps can not be obtained from actual upper endoscopic procedures.

The fine-tuned *SfSNet* is used to absolute depth maps, normalized from 0 to 150 mm, in real EN frames. Fig. 8 shows EN frames from the pylorus, antrum, and stomach in the first row, estimated depth maps in the second row, and 3d reconstructions in the third row. The model successfully estimated accurate depth maps, remarkably capturing the shape and size variations of cavities.

SfSNet was trained on a server equipped with an *NVIDIA Tesla T4* GPU (16 GB VRAM). *SfSNet* reached stable training and validation loss values after 51 epochs, completed in approximately 76 hours. During testing, *SfSNet* estimates a depth map from a frame in 200 ms.

B. Experimental Setup With Phantoms

To prove the presented method captures distensibility, 50 insufflated stomach phantoms were generated from averages of nine real measures of different stomach regions from 50 adult cadavers [24], [25], all of them distended with water. Likewise, to emulate different distensibility degrees after insufflation, the phantoms were modified by radially displacing the mesh cells towards the phantom's center of mass by a parameter varying from 0.1 to 1.0 at steps of 0.1. This factor modelled distensibility in different conditions, from stomachs with rugae or folds nearly intact (poor distension) to stomachs with no folds (maximum distension). Blue outlined frames in Fig. 9 show the appearance of the gastric wall with different distensibility conditions: from the lowest (0.1) to the highest (1.0) distension. Afterwards, the stomach motion was simulated

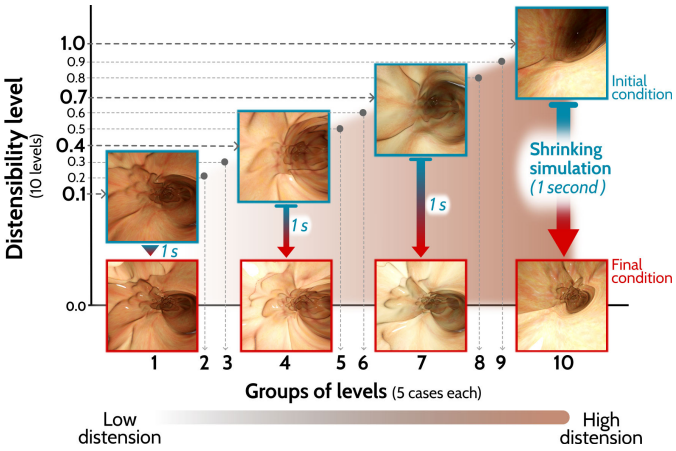

Fig. 9. Synthetic collection of stomach phantoms divided into 10 groups, each representing a different distensibility degree. The shrinking simulation lasted 1 second for all phantoms, starting from complete insufflation to the moment folds start to overlap.

TABLE II
MEAN ABSOLUTE ERROR (MAE) FOR DIFFERENT FACTORS

Factor	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
MAE	0.131	0.039	0.038	0.019	0.047	0.019	0.023	0.144	0.014	0.053

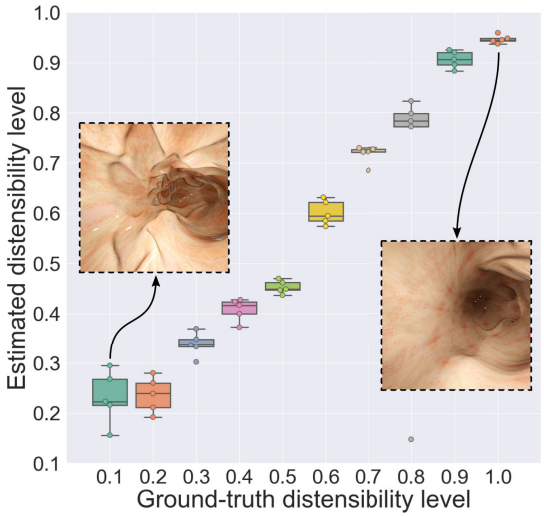

Fig. 10. Boxplots of the estimated distensibility degrees vs. ground-truth distensibility levels (0.1–1.0). Synthetic endoscopy frames illustrating the appearance of stomach phantoms under minimal (left) and maximal (right) distension. These results indicate spatio-temporal features provide robust approximation to several degrees of gastric distension.

from complete insufflation until the moment folds begin to overlap, i.e., 1 second for all phantoms, the final condition observed in the red outlined frames of Fig. 9. Finally, the 50 phantoms were grouped into 10 groups, five phantoms per group, each with a different distensibility degree (Group 1 = 0.1, Group 2 = 0.2, …, Group 10 = 1.0), as displayed in Fig. 9. These phantoms represent a subset of the synthetic collection previously used to train and validate the depth estimation method.

For each phantom, spatio-temporal features were computed at times t_0 and t_1 from 10 elements randomly selected from the stomach body and antrum meshes. These features fed a Random Forest regressor to estimate the distensibility degree. The regression task was evaluated using a 5-fold stratified cross-validation scheme, with 80% of the phantoms used to train and 20% to test at each fold. The performance metric used to compare estimated and ground-truth distensibility degrees was the mean absolute error (MAE). The proposed model, trained with the spatio-temporal descriptors, achieved reliable performance with an average-MAE of 0.053. Table II shows MAE for each distensibility degree along the 5-folds.

Fig. 10 presents boxplots of the estimated distensibility levels (y-axis) against the ground-truth levels (x-axis), showing the spatio-temporal features robustly characterize different levels of gastric distension. These results indicate that the method is sensitive to subtle distensibility changes.

C. Estimating Distensibility in a Cohort With Helicobacter Pylori

1) Extraction of Video Sequences: Sequences from three anatomical sites per patient were extracted from 160 EN videos, collection described in section II-E.1. These videos were recorded during routine EN procedures following the SSS protocol, that is to say they were recorded in real conditions. Video sequences for the three anatomical sites were extracted when the endoscopist stabilized the camera at the following views: (a) the lower or middle upper body in an antegrade view looking at the anterior wall, (b) the pylorus in an antegrade view looking at the greater curvature, and (c) near the incisura angularis looking at the pylorus to capture the entire antrum before collecting biopsy samples (see section II-A). A frame from each of these views was marked by the gastroenterologist and, using this frame as reference, videos of three seconds were selected, i.e., 1.5 s forwards and 1.5 s backwards. Subsequently, applying a dense optical flow [19], inter-frame pixel displacement was computed and pairs of frames were sorted out by the average displacement magnitude $\overline{d}$ in pixels, from lowest to highest. Then, some pairs were discarded based on the following criteria:

- Pairs with no displacement or $\overline{d} < 1.4$ pixels: These typically correspond to moments the gastroenterologist froze the endoscope image to capture one out of the 22 reference pictures in the SSS protocol.
- Pairs with $\overline{d} > 15$ pixels: Despite the gastroenterologist's efforts to stabilize the camera, these involuntary movements result in high displacements.

Finally, the first pair of frames with the smallest but non-null displacement within the range $2 < \overline{d} < 10$ pixels was picked for later analysis.

The time elapsed between the two selected consecutive frames, f_t and f_{t+1}, is 66.6 milliseconds, since these videos were captured at 15 frames per second. During this interval, both the stomach volume (or size) and intragastric pressure decrease rapidly after the endoscopist stops insufflating air [4], [34], coinciding with the period in which the SSS is performed. These volume and pressure changes lead to stomach shrinkage, observable during EN by the reappearance of gastric folds; this is the variation captured by the proposed approach.

2) Construction of Meshes and Elements: For the two selected consecutive frames, f_t and f_{t+1}, a mesh was generated for each frame as explained in section II-B.3. Each mesh was

constructed from an estimated depth map and initialized with $1,280\times1,280$ vertices, matching the frame's spatial resolution. This mesh, consisting of $1,635,841$ cells or an equivalent area of $0.3\pm0.1\,mm^2$, was prone to high-frequency artifacts due to the limited 8-bit precision of the depth map. To mitigate this, the mesh was subsampled by removing vertices, increasing the mesh cell size while preserving sufficient resolution to capture stomach wall rugae and reducing high-frequency artifacts. Subsampling was applied dyadically: by a factor of 2 (320×320 vertices, $101,761$ cells, $4.8\pm1.9\,mm^2$ cell area), and by a factor of 3 (160×160 vertices, $25,281$ cells, $19.4\pm6.4\,mm^2$ cell area). These subsampling factors were selected and applied in the next section to define the two configurations of the spatio-temporal descriptor.

3) Evaluation per Anatomical Site: Provided the surface is reconstructed with respect to the camera, most reconstructions capture information of the gastric wall in the x coordinate while depth information of the stomach is acquired in the z direction.

Two different configurations of the spatio-temporal descriptor, centered evaluation and grid evaluation, were assessed for each anatomical site to evaluate regional and local distensibility. These configurations are described below and illustrated in Fig. 11:

- *Centered evaluation:* in this case the reconstructed surface was analyzed from a single element standpoint, centered in the frame, and its neighborhood. An element in this case was composed of 29×29 mesh cells. The mesh was subsampled by a factor of 3 to capture global distensibility variations.
- *Grid evaluation:* the frame was divided into a 3×3 grid tiles and the spatio-temporal characterization was applied to each of the grid tiles. Therefore the resultant element size was set to 7×7 mesh cells. The subsampled factor of the mesh was 2 since this ensures fine mesh details are captured.

Depending on the region, the projection of the gastrointestinal tract (the black spot in the frame) is displaced by the position of the endoscope with respect to the gastric wall. After segmenting this black spot, this position changes for each of the region views as follows:

- in the body this black spot is typically observed at the right of the scene, i.e., tiles 3, 6, and 9 of grid evaluation in Fig. 12.
- in the antrum and pylorus the black spot is mostly located in the upper row of the scene, i.e., tiles 1, 2, and 3 of grid evaluation in Fig. 12.

These tiles were excluded from the analysis, no neighbor from these tiles was taken into account.

D. Centered Evaluation Results

Overall, 80% of the entire frame is divided into five elements. Evaluation was performed only on the central element and both regional and local analysis was carried out. Table III presents the statistical analysis of the five spatio-temporal features in the rows and the anatomical sites in the columns. Four features show significant differences between the *H. pylori* positive and negative groups. Specifically, a significant local feature was observed in the stomach body while regional features corresponded to directions in x- and $z-axis$. Likewise, local feature in transversal direction was significantly different in the antrum. After FDR correction, three out of four of these features still show significant differences, namely the three features of the body marked with * in the table.

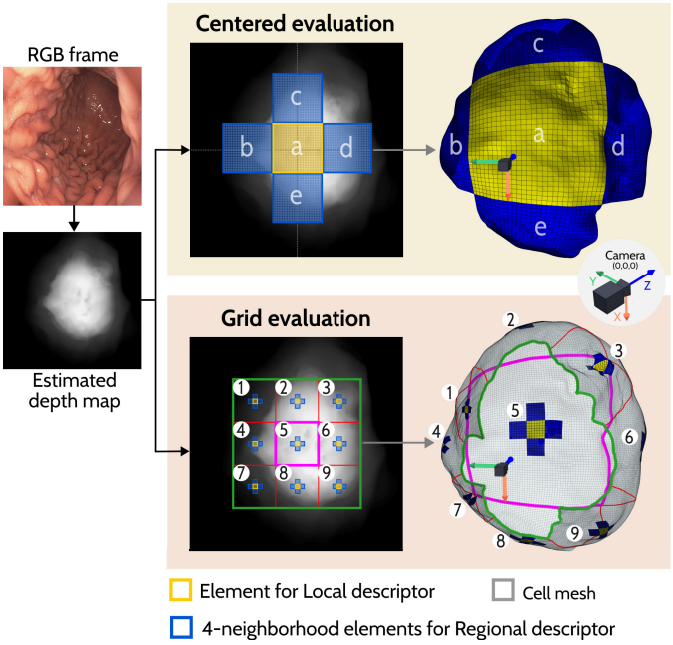

Fig. 11. Illustration of the two evaluation strategies by anatomical site. The first row displays the centered evaluation of a frame and its corresponding mesh. The second row shows a superimposed 3×3 grid, a local neighborhood of elements within each grid tile and the corresponding reconstructed surface. Notice how the green outer grid border and the pink central tiles are deformed in the surface view.

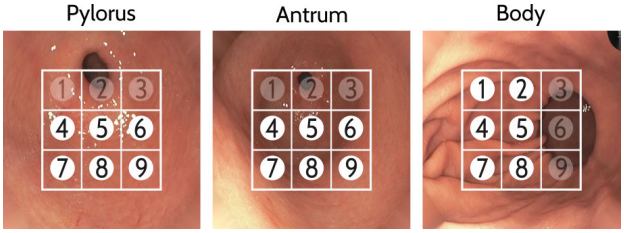

Fig. 12. Tiles shown as transparent were excluded from the grid evaluation because they superimpose the black intestinal tract.

TABLE III
RESULTS OF CENTERED EVALUATION: *p*-VALUES FROM T-TESTS BETWEEN GROUPS WITH H. PYLORI -POSITIVE AND H. PYLORI -NEGATIVE FOR EACH ANATOMICAL SITE

Descriptor	Anatomical site		
	Stomach body	Antrum	Pylorus
$Local_{longitudinal}$	0.266	0.060	0.107
$Local_{transversal}$	**0.010***	**0.011**	0.551
$Regional_{x-axis}$	**0.011***	0.601	0.672
$Regional_{y-axis}$	0.463	0.955	0.928
$Regional_{z-axis}$	**0.013***	0.974	0.922

1) Specific Diagnostic Thresholds: The presence of *H. pylori* is diagnosed by pathology examination after applying the Sydney protocol, that is to say five biopsies were taken and *H. pylori* presence was detected at least in one of them. Clinical usefulness was completed by a new experiment to set diagnostic thresholds. Specifically, a decision tree was trained

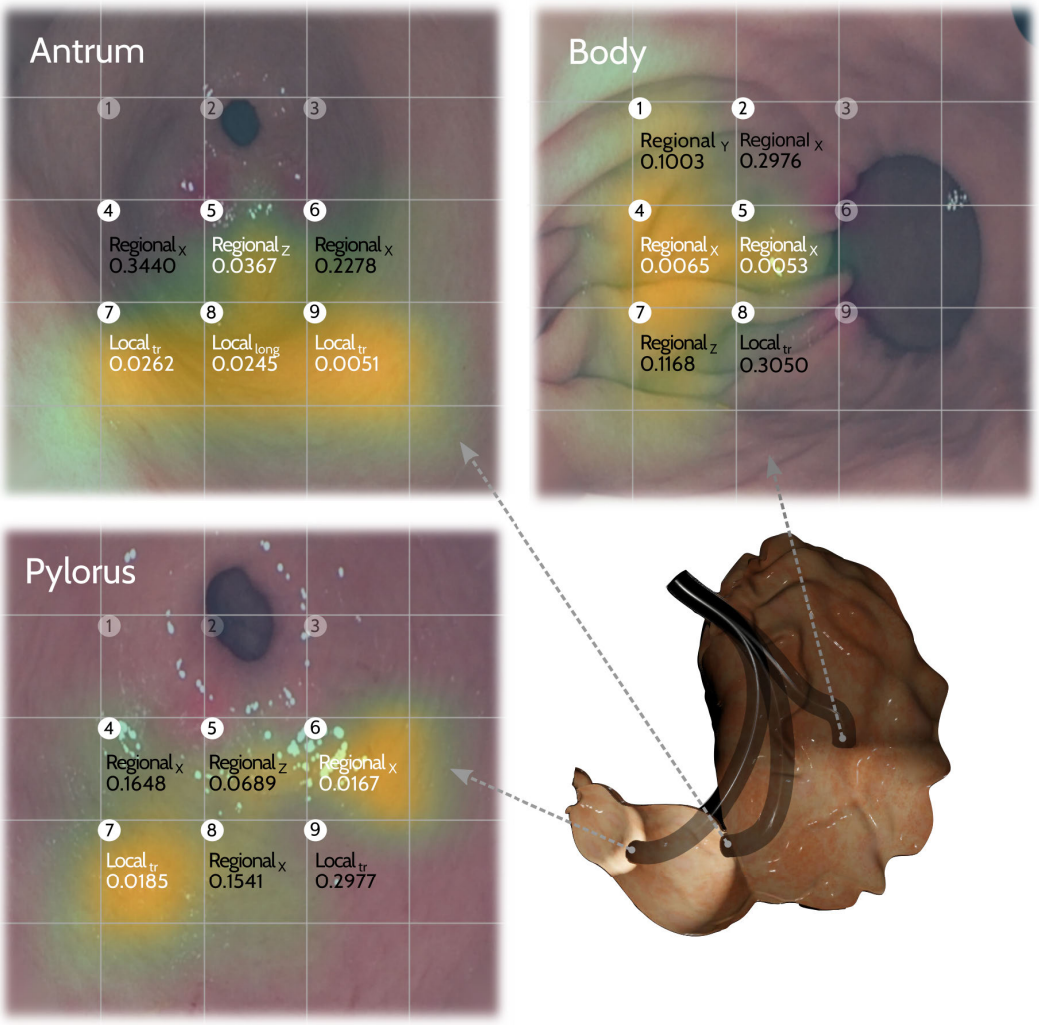

Fig. 13. Summary of results for grid evaluation: heat-maps highlight the most discriminant features for each grid tile in pylorus, antrum, and stomach body. Features with a p-value $\leq$ 0.05 are highlighted in white.

to predict the presence of *H. pylori* when it was fed by features separately extracted from the antrum and body. After 100 bootstrap resamples, the binary classification model reached an accuracy of 82.9% $\pm$ 2.1% for body and 82.6% $\pm$ 3.0% for antrum. The decision branches of the tree were taken as the diagnostic thresholds, they were distributed as shown in Table IV:

TABLE IV
SPECIFIC DIAGNOSTIC THRESHOLDS FOR EACH FEATURE IN THE STOMACH BODY AND ANTRUM

	Anatomic sites	
Feature	**Body**	**Antrum**
$Local_{longitudinal}$	0.14 (0.13, 0.16)	0.07 (0.05, 0.09)
$Local_{transversal}$	0.27 (0.26, 0.28)	0.06 (0.04, 0.08)
$Regional_{x-axis}$	0.26 (0.25, 0.28)	0.23 (0.21, 0.24)
$Regional_{y-axis}$	0.17 (0.16, 0.18)	0.30 (0.28, 0.32)
$Regional_{z-axis}$	-0.14 (-0.14, -0.13)	0.01 (0.01, 0.03)
	Mean (95% Confidence interval)	

E. Grid Evaluation Results

Distensibility was also assessed at a smaller scale, specifically using the same surface partition but this time each of the elements of the centered evaluation is independently tested, meaning each of them is in due turn subdivided and characterized by the herein introduced motion descriptor. Fig. 13 highlights the descriptors with significant differences per grid tile, shown as a heatmap constructed using the highest and lowest p-values obtained from each anatomical site. As explained before, tiles from the upper row of the antrum and right column of the body were removed from the analysis since they typically contain the black spot.

Regarding these results, the free wall of the stomach in the antrum shows significant local differences at the greater curvature portion (lower row of the Antrum panel in Fig. 13). The local bending feature likely captures subtle, localized changes in distension, particularly in areas without prominent folds such as the antrum, a smooth stomach wall portion with a flat mucosa [35]. While regional differences are observed at the center of the grid in the $z-axis$, suggesting a more curved greater curvature, aligned with the direction of the endoscope exploration. In contrast, the stomach body shows regional differences only at the mid level of the scene as shown

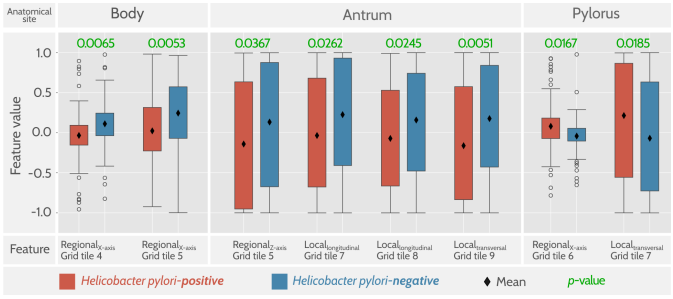

Fig. 14. Boxplots showing the values of the significantly different features from grid evaluation, comparing *H. pylori*-positive (red) and *H. pylori*-negative (blue) groups for the stomach body, antrum, and pylorus.

in Body panel in Fig. 13, corresponding to the stomach wall region between the greater curvature and anterior wall, likely the region which should show more motion [36]. This is, in fact, the region where the stomach wall typically stretches or compresses its folds [36], which is the most affected region when distensibility is impaired [8], [37]. Also, direction of the regional differences, i.e., $x-axis$ which runs nearly perpendicular to the greater curvature, seems to represent the primary direction in which the stomach loses its ability to distend. Finally, in the pylorus, significant differences were observed regionally in the posterior wall and locally in the greater curvature (tile 7 and 6 of the Pylorus panel in Fig. 13 respectively).

Additionally, Fig. 14 shows the spread of *H. pylori*-positive and *H. pylori*-negative groups of the significantly different features per anatomical site. Distribution of both groups show slight overlap in the interquartile ranges of stomach body and pylorus, while the overlap is greater in case of antrum. In the stomach body, the *H. pylori*-positive group shows less regional distensibility compared to the negative group. Similarly, in the antrum, the *H. pylori*-positive group presents smaller local and regional distensibility than the negative group. In contrast, the *H. pylori*-positive group shows greater regional and local distensibility in the pylorus when compared to the negative group. These results may look contradictory but they are just reflecting the physiology of the stomach. In fact, in cases without inflammatory response (*H. pylori*-negative), the antroduodenal junction or pylorus is nothing but a sphincter and therefore it naturally moves less [38] than the antrum or body which together make the antral pump, a mechanism in charge of initiating the circumferential contraction to achieve peristalsis [39]. Since the pylorus moves less, it is then reasonable to find that an infected and swollen pylorus is more distensible, offers less resistance to the food transit and therefore the temporal bending change should be greater [40].

Overall, the statistical analysis for the grid evaluation identified significant differences in ten features, as listed below. Three out of these features continue to show significant differences after FDR correction and are marked with *.

- Body:
  - Two regional features in $x-axis$ direction at grid tile 4 (p=0.0065*) and 5 (p=0.0053*).
  - Regional feature in $y-axis$ direction at grid tile 5 (p=0.038).
- Antrum:
  - Two local features in the longitudinal direction at grid tile 8 (p=0.024) and 9 (p=0.0050*).
  - Two local features in the transversal direction at grid tile 7 (p=0.026) and 9 (p=0.040).
  - Regional feature in $z-axis$ direction at grid tile 5 (p=0.036).
- Pylorus:
  - Local feature in longitudinal direction at grid tile 7 (p=0.018).
  - Regional feature in $x-axis$ direction at grid tile 6 (p=0.016).

IV. Discussion

This work introduces a novel spatio-temporal characterization of the gastric distensibility, a key indicator of stomach condition which is subjectively reported after insufflating the organ in EN procedures. The underlying hypothesis is that changes of the stomach bending during steady periods are influenced by several pathologies which produce stiffer gastric walls. Briefly, after a pair of consecutive frames with little change is selected for analysis, a three-dimensional gastric surface is reconstructed for each frame. Afterwards, a model approximates the temporal bending change of the surface elements and their neighborhoods. This characterization has revealed significant differences in passive motility between a group of patients with *H. pylori* and subjects without infection, diagnosed after the routine pathology microscopy examination. It is worthy to clarify that distensibility is considered as a clinic sign reported by endoscopists rather than a pure measurement of the gastric elasticity or even the gastric dynamics. The relation of this observation with the amount of gastric bending is established by the statistical analysis of the *H. pylori* presence since it has been demonstrated in other studies that this bacteria is associated with a greater rigidity of the gastric wall, or a diminished distensibility.

In clinical practice, motility disorders are associated with symptoms like abdominal pain and nausea. A routine study of these symptoms starts with an EN to exclude structural issues or obstructions along with biopsy samples to test for *H. pylori* [5], [16]. Additionally, distensibility, i.e. the stomach's ability to expand in response to inner volume changes, is assessed during EN by observing how the gastric rugae are flattened in response to air insufflation [3]. This response of gastric rugae has been characterized by inserting pressure transducers through the biopsy channel to measure intra-gastric pressure during air insufflation, identifying three groups: well distended F-I, partially distended F-II, and No distension F-III. However, provided these pressure transducers are not available in routine EN, the distensibility assessment has ended up by being merely visual. On the other hand, different studies have employed balloon distension techniques such as barostat [4] or endoluminal functional lumen imaging probe (EndoFLIP) [41]. These methods, by impedance sensors, capture very local pressure patterns with the consequent necessity of estimating specific anatomical landmarks and therefore missing the complexity of the entire gastric disorder [5], [42]. Specifically,

the barostat system consists of a thin, compliant balloon with a computer-controlled air pump, and pressure and volume sensors to manipulate intraluminal pressure. The barostat produces stomach distension and measures changes in tissue response, usually located in the proximal stomach or gastric fundus [43]. Similarly, EndoFlip estimates the gastric dynamics from very local measures, in particular the esophagus and pyloric sphincters [5] and again it misses the entire gastric dynamics. In summary, while these methods provide insight about distensibility, they are unavailable in many real scenarios [5], [42]. Overall, these methods capture either very local pressure patterns or particular dynamic properties of the gastric distension but they miss the local-global relationships which might quantify the physiopathology of the impaired distension. Furthermore, these methods estimate an active distensibility since the device must be insufflated and the measured force is a result of this interaction. Aside from the fact that there are no public measurements of these methods, comparison with the proposed approach would not be fair since the presented method estimates a passive distensibility, i.e., the wall is passively moving after insufflation is over (close to what clinicians do). Instead, this work builds up a new measurement which approximates the visual judgement of gastric dynamics and demonstrates correlation with pathological findings (presence of *H. pylori*), the most important indicator of abnormality. Gastroenterologists use this distensibility evaluation as a clinical observation to globally capture altered conditions as the linitis plastica or infection processes which alter the expected passive response. EndoFlip or barostat methods hardly capture the different stages of these abnormalities. Unlike these strategies, the present proposal builds up a local approximation of the stomach surface and by combining local-regional information, the method shows significant differences between cases with and without *H. pylori*.

To the best of our knowledge, this is the first method devoted to estimate changes of the local stomach bending during relatively motionless periods of a routine EN. The whole strategy combines customized machine learning models to reconstruct both the stomach surface and distention changes. Moreover, the reconstruction strategy of the stomach wall from depth estimation has been extensively validated [18], [26], and then fine-tuned with a synthetic collection of endoscopy videos constructed from nine averaged measurements derived from 50 adult cadavers [24]. In addition, this study includes a method for identifying two consecutive frames with enough stability to ensure that the presented analysis captures stomach motion at precise anatomical sites. This configuration prevents capture of false gastric movements, namely illumination changes due to occlusions and abrupt camera movements. Artifacts caused by mucus or fluids, including injected water, were prevented by the patient's preparation, as detailed in section II-E.1.

This constraint nevertheless was a bottleneck to apply the method to the entire organ since pairs of frames which simultaneously meet this condition are difficult to find for different regions of the stomach. Overall, the presented spatio-temporal descriptor effectively captured subtle local and regional bending which was proven to be significantly different between groups with and without *H. pylori*. Nevertheless, certain variability might be controlled to improve the proposed method, for instance the specific navigation protocol, the model characterization and the device parameters during navigation to diminish the impact of the intrinsic biological variability.

All methods were implemented in Python and, when executed on a standard laptop (*Intel Core i7*, four cores at 1.90 GHz, 16 GB RAM), required an average of 27.25 seconds to analyze one of the elements of the reconstructed surface. This total processing time is distributed as follows: optical flow computation required 0.10 seconds; 3*d* reconstruction of the frame sequence, 1.93 seconds; local feature analysis, 1.36 seconds; and regional feature analysis, 23.82 seconds. These short execution times indicate this method may be compatible with future clinical application, since it can be implemented as a post-procedure action, immediately after the expert registers the gastric anatomical sites defined by exploration protocol and the corresponding sequences have been extracted. Code for reproducing the reported results is available at https://github.com/Cimalab-unal/3dSpatioTempFeatures

In conclusion, the proposed method effectively captured subtle bending of the gastric wall from routine endoscopy videos, demonstrating aptness to detect differences associated with pathological conditions, specifically the presence of *H. pylori*. Remarkably, the method uses only the spatio-temporal information from endoscopy videos and therefore it may be a great asset in clinical practice. Specifically, clinical guidance requires at least three steps: 1) once the stomach is totally insufflated and the specialist is at the bottom of the pyloros, the method computes the optical flow for any pair of frames, 2) stop in the video when sequences with minimal motion are detected since they may correspond to the steady moments of the examination, 3) apply the proposed strategy to the sequence of frames corresponding to this window, and 4) interpret the obtained feature values in the context of the patient.

Furthermore, obtaining this characterization does not impose any additional tasks or protocols on clinicians, since it can be computed during standard endoscopic navigation. Likewise, different application scenarios might be envisaged, starting with the automatic classification or grading of pathological conditions in any anatomical site when the camera is stabilized. Another potential application is the complete 3*d* reconstruction of the stomach for tracking passive gastric dynamics to characterize pathological conditions. Moreover, the reconstruction of the navigation path would give the actual possibility of recording the exact location of the biopsy as a landmark to follow up any pathology by analyzing the gastric dynamics, comparing with other subjects, and mapping the histologic quantification to the gastric wall.

Future work includes analysis of distensibility under unstable camera positions, when gastroenterologists do not follow a SSS, integrating: (a) a more robust optical flow algorithm capable of dealing with large camera displacements and illumination changes, and (b) a camera pose estimation algorithm to compensate for camera motion or to track points of interest in the stomach wall. Finally, this method may boost a quantitative analysis of other gastric pathological conditions, since it can

be applied to any region of the gastric wall and at different scales.

REFERENCES

[1] Y. Tseng et al., "Unveiling the neuroinflammatory pathogenesis of persistent functional dyspepsia in H. Pylori infection: Insights on MMP-9 as a therapeutic target," *Clin. Transl. Med.*, vol. 13, no. 11, p. e1456, Nov. 2023.

[2] G. Corso et al., "Hereditary gastric and breast cancer syndromes related to CDH1 germline mutation: A multidisciplinary clinical review," *Cancers*, vol. 12, no. 6, p. 1598, Jun. 2020.

[3] S. Suzuki, K. Yamada, K. Yamashita, M. Endo, and T. Takemoto, "Gastroscopic dynamics of gastric mucosa in relation to intragastric pressure," *Endoscopy*, vol. 4, no. 1, pp. 20–28, Feb. 1972.

[4] T. Suzuki, M. Hirano, and Y. Yamamoto, "Examination of visceral perception and gastric tone by gastric stimulation using air inflation during endoscopy," *J. Int. Med. Res.*, vol. 33, no. 2, pp. 160–169, Mar. 2005.

[5] J. H. Noh and H.-Y. Jung, "Role of endoscopy in motility disorders of upper gastrointestinal tract," *J. Neurogastroenterology Motility*, vol. 29, no. 1, pp. 7–19, Jan. 2023.

[6] Y. Saito et al., "Dysfunctional gastric emptying with down-regulation of muscle-specific MicroRNAs in Helicobacter pylori-infected mice," *Gastroenterology*, vol. 140, no. 1, pp. 189–198, Jan. 2011.

[7] A. A. Ashraf et al., "Investigating Helicobacter pylori-related pyloric hypomotility: Functional, histological, and molecular alterations," *Amer. J. Physiology-Gastrointestinal Liver Physiol.*, vol. 321, no. 5, pp. G461–G476, Nov. 2021.

[8] B. Liu et al., "Helicobacter pylori causes delayed gastric emptying by decreasing interstitial cells of cajal," *Exp. Therapeutic Med.*, vol. 22, no. 1, Apr. 2021.

[9] G. Sarnelli, R. Cuomo, J. Janssens, and J. Tack, "Symptom patterns and pathophysiological mechanisms in dyspeptic patients with and without Helicobacter pylori," *Digestive Diseases Sci.*, vol. 48, no. 12, pp. 2229–2236, Dec. 2003.

[10] F. Perri, "Patterns of symptoms in functional dyspepsia: Role of Helicobacter pylori infection and delayed gastric emptying," *Amer. J. Gastroenterol.*, vol. 93, no. 11, pp. 2082–2088, Nov. 1998.

[11] M. Chiloiro et al., "Effect of Helicobacter pylori infection on gastric emptying and gastrointestinal hormones in dyspeptic and healthy subjects," *Digestive Diseases Sci.*, vol. 46, no. 1, pp. 46–53, Jan. 2001.

[12] A. Zullo et al., "Helicobacter pylori and functional dyspepsia: An unsolved issue?," *World J. Gastroenterol.*, vol. 20, p. 8957, Jul. 2014.

[13] K. Yao, "The endoscopic diagnosis of early gastric cancer," *Ann. Gastroenterol., Quart. Publication Hellenic Soc. Gastroenterol.*, vol. 26, no. 1, p. 11, 2013.

[14] D. Keilberg and K. M. Ottemann, "How Helicobacter pylori senses, targets and interacts with the gastric epithelium," *Environ. Microbiol.*, vol. 18, no. 3, pp. 791–806, Mar. 2016.

[15] L. E. Martínez, V. P. O'Brien, C. K. Leverich, S. E. Knoblaugh, and N. R. Salama, "Nonhelical Helicobacter pylori mutants show altered gland colonization and elicit less gastric pathology than helical bacteria during chronic infection," *Infection Immunity*, vol. 87, no. 7, pp. 904–922, Jul. 2019.

[16] N. Faknak et al., "Performance status of targeted biopsy alone versus Sydney protocol by non-NBI expert gastroenterologist in gastric intestinal metaplasia diagnosis," *Endoscopy Int. Open*, vol. 10, no. 4, pp. E273–E279, Apr. 2022.

[17] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds., Berlin, Germany: Springer, 2003, pp. 363–370.

[18] J. Ruano, M. Gómez, E. Romero, and A. Manzanera, "Leveraging a realistic synthetic database to learn shape-from-shading for estimating the colon depth in colonoscopy images," *Computerized Med. Imag. Graph.*, vol. 115, Jul. 2024, Art. no. 102390.

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Arti. Intell. (IJCAI)*, 1981, pp. 674–679.

[20] G. Bassotti, "Biomechanics of the gastrointestinal tract," *Digestive Liver Disease*, vol. 35, no. 11, p. 841, Nov. 2003.

[21] D. Bravo et al., "GastroHUN an endoscopy dataset of complete systematic screening protocol for the stomach," *Sci. Data*, vol. 12, no. 1, pp. 1–14, Jan. 2025.

[22] D. Bravo et al., *GastroHUN an Endoscopy Dataset of Complete Systematic Screening Protocol for the Stomach*. London, U.K.: Springer, 2025, doi: 10.6084/m9.figshare.27308133.

[23] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: Cancer J. Clinicians*, vol. 74, no. 3, pp. 229–263, May 2024.

[24] J. Ruano et al., "Generating synthetic endoscopy videos following a systematic screening protocol," in *Proc. 19th Int. Symp. Med. Inf. Process. Anal. (SIPAIM)*, Nov. 2023, pp. 1–5.

[25] A. M. Karnul and C. K. Murthy, "A study of variations of the stomach in adults and growth of the fetal stomach," *Cureus*, vol. 14, no. 8, Aug. 2022, Art. no. e28517.

[26] J. Ruano et al., "Estimating polyp size from a single colonoscopy image using a shape-from-shading model," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2024, pp. 1–5.

[27] F. Mahmood and N. J. Durr, "Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy," *Med. Image Anal.*, vol. 48, pp. 230–243, Aug. 2018.

[28] A. Rau et al., "Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 7, pp. 1167–1176, Jul. 2019.

[29] D. Freedman et al., "Detecting deficient coverage in colonoscopies," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3451–3462, Nov. 2020.

[30] S. Mathew, S. Nadeem, S. Kumari, and A. Kaufman, "Augmenting colonoscopy using extended and directional cyclegan for lossy image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4696–4705.

[31] H. Itoh et al., "Unsupervised colonoscopic depth estimation by domain translations with a lambertian-reflection keeping auxiliary task," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 6, pp. 989–1001, Jun. 2021.

[32] B. Cui, M. Islam, L. Bai, A. Wang, and H. Ren, "EndoDAC: Efficient adapting foundation model for self-supervised depth estimation from any endoscopic camera," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2024, pp. 208–218.

[33] C. Budd and T. Vercauteren, "Transferring relative monocular depth to surgical vision with temporal consistency," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2024, pp. 692–702.

[34] E. Distrutti, F. Azpiroz, A. Soldevilla, and J.-R. Malagelada, "Gastric wall tension determines perception of gastric distention," *Gastroenterology*, vol. 116, no. 5, pp. 1035–1042, May 1999.

[35] J. B. O'Connor, "Upper gastrointestinal endoscopy," in *Encyclopedia Gastroenterology*. San Diego, CA, USA: Elsevier Academic, 2004, pp. 566–572.

[36] S. Hosseini, R. Avci, N. Paskaranandavadivel, V. Suresh, and L. K. Cheng, "Quantification of the regional properties of gastric motility using dynamic magnetic resonance images," *IEEE Open J. Eng. Med. Biol.*, vol. 4, pp. 38–44, 2023.

[37] H. Suzuki and P. Moayyedi, "Helicobacter pylori infection in functional dyspepsia," *Nature Rev. Gastroenterol. Hepatol.*, vol. 10, no. 3, pp. 168–174, Mar. 2013.

[38] D. Ramkumar and K. S. Schulze, "The pylorus," *Neurogastroenterol. Motility*, vol. 17, pp. 22–30, Jun. 2005.

[39] B. Surjanhata and B. Kuo, "Gastrointestinal motility and enteric neuroscience in health and disease," in *Reference Module in Biomedical Sciences*. Amsterdam, The Netherlands: Elsevier, 2014, doi: 10.1016/B978-0-12-801238-3.00081-9.

[40] T. L. Jue and W. L. Hasler, "The impact of pyloric directed therapy for gastroparesis," *Foregut: J. Amer. Foregut Soc.*, vol. 4, no. 2, pp. 191–199, Jun. 2024.

[41] S. Mussad, R. Maradiaga, L. Navari, L. Kanzaki, Y. Zhang, and S. Chakraborty, "S590 examining the correlation between gastroesophageal sphincter distensibility measured by EndoFLIP and esophageal acid exposure," *Amer. J. Gastroenterol.*, vol. 118, no. 10S, pp. S431–S432, 2023.

[42] Y. Wang, J. D. Z. Chen, and B. Nojkov, "Diagnostic methods for evaluation of gastric motility—A mini review," *Diagnostics*, vol. 13, no. 4, p. 803, Feb. 2023.

[43] G. Holtmann, J. Gschossmann, G. Guerra, H. Goebell, and N. J. Talley, "Perception of gastric distension: Influence of mode of distension on perception thresholds and gastric compliance," *Digestive Diseases Sci.*, vol. 40, pp. 2673–2677, Dec. 1995.