A Framework for the Evaluation of Human Machine Interfaces of Robot-Assisted Colonoscopy

Martina Finocchiaro, Tommaso Banfi, Sonia Donaire, Alberto Arezzo, Carlos Guarner-Argente, Arianna Menciassi, *Fellow, IEEE*, Alicìa Casals, *Senior Member, IEEE*, Gastone Ciuti, *Senior Member, IEEE*, and Albert Hernansanz

***Abstract*—The Human Machine Interface (HMI) of intraluminal robots has a crucial impact on the clinician's performance. It increases or decreases the difficulty of the tasks, and is connected to the users' physical and mental stress. *Objective:* This article presents a framework to compare and evaluate different HMIs for robotic colonoscopy, with the objective of identifying the optimal HMI that minimises the clinician's effort and maximises the clinical outcomes. *Methods:* The framework comprises a 1) a virtual simulator (clinically validated), 2) wearable sensors measuring the cognitive load, 3) a data collection unit of metrics correlated to the clinical performance, and 4) questionnaires exploring the users' impressions and perceived stress. The framework was tested with 42 clinicians investigating the optimal device for tele-operated control of robotic colonoscopes. Two control devices were selected and compared: a haptic serial-kinematic device and a standard videogame joypad. *Results:* The haptic device was preferred by the endoscopists, but the joypad enabled better clinical performance and reduced cognitive and physical load. *Conclusion:* The framework can be used to evaluate different aspects of a HMI, both hardware and software, and determine the optimal HMI that can reduce the burden on clinicians while improving the clinical outcome. *Significance:* The findings of this study, and of future studies performed with this framework, can inform the design and development of HMIs for intraluminal robots, leading to improved clinical performance, reduced physical and mental stress for clinicians, and ultimately better patient outcomes.**

***Index Terms*—Human machine interface, robotic colonoscopy, medical simulation, intraluminal robots.**

I. Introduction

IN THE last decades, the increasing number of intraluminal procedures has demonstrated strong benefits for the patients [1]. However, from the clinicians' perspective, this type of procedures is challenging to master due to the limited and complex workspace, unstable control of long flexible scopes, and the loss of direct view over the surgical scene [2], [3]. In addition, the poor ergonomics and intuitiveness of the instruments currently used in the clinical practice (e.g., colonoscopes) contribute to the rise of surgeons' mental and physical burden, having negative effects both on their health and outcome of the procedures [4], [5]. In this scenario, the use of robotic assistants, e.g., multi-steerable snake-like robots and endoscopic capsules, can help addressing the drawbacks, increasing the stability and precision of the tools, and developing a more assistive and user-friendly Human Machine Interface (HMI) [6], [7]. From the design prospective, the introduction of robotic technologies increases the number of degrees-of-freedom (DOF) to control and the sensing information to process, posing the basis for a new framework of human-robot interaction [8]. Besides the mechanical design of the device, the HMI (i.e. the interface used to manoeuvre an endoscope, together with the adopted control strategy and the feedback received during the intervention), has an important impact over the user's experience and procedure outcome [9], [10], [11]. Accordingly, a good HMI can decrease the difficulty of the tasks and reduce the users' physical and mental stress, influencing the clinicians' final performances [12]. In the last decades, several physical interfaces, control strategies, visual and haptic cues, have been developed for robot-assisted intraluminal procedures, but little space has been left for the

Manuscript received 8 April 2023; revised 14 June 2023; accepted 18 July 2023. Date of publication 3 August 2023; date of current version 22 January 2024. This work was supported by the European Union's Horizon 2020 Research and Innovation Program through the ATLAS Project, under the Marie Sklodowska-Curie Grant 813782. The work of Gastone Ciuti and Albert Hernansanz share last (senior) authorship. *(Corresponding author: Gastone Ciuti.)*

Martina Finocchiaro is with the Center of Research on Biomedical Engineering, Universitat Politècnica de Catalunya, Spain, also with the BioRobotics Institute, Italy, and also with the Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Italy.

Tommaso Banfi and Arianna Menciassi are with the BioRobotics Institute, Italy, and also with the Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Italy.

Sonia Donaire, Alicìa Casals, and Albert Hernansanz are with the Center of Research on Biomedical Engineering, Universitat Politècnica de Catalunya, Spain.

Alberto Arezzo is with the Department of Surgical Sciences, University of Torino, Italy.

Carlos Guarner-Argente is with the Unidad de Endoscopia, Servicio de Gastroenterología y Hepatología, Hospital de la Santa Creu i Sant Pau, Spain.

Gastone Ciuti is with the BioRobotics Institute, 56127 Pisa, Italy, and also with the Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna 56127 Pisa, Italy (e-mail: gastone.ciuti@santannapisa.it).

This article has supplementary downloadable material available at https://doi.org/10.1109/TBME.2023.3301741, provided by the authors.

Digital Object Identifier 10.1109/TBME.2023.3301741

TABLE I
DESPCRIPTION OF HUMAN MACHINE INTERFACE (HMI) CHARACTERISTICS AND THEIR EVALUATION WHEN USED WITH STANDARD COLONOSCOPE

Human Machine Interface (HMI)		Space of interaction between the user of a product and the product itself: equipment used to manoeuvre the endoscope, together with the adopted control strategy and the feedback received during the intervention	
		Definition	**Standard colonoscope**
HMI characteristics	Intuitive	provides a familiar interaction means to the user: mapping the movements of the endoscope down to the controller, in a simple way that the users have experienced before	intrinsically non-intuitive mapping of the degrees of freedom (DOF) between the handler and tool tip
	User-friendly	easy-to-use and easy-to-learn: correlated with the complexity the HMI brings on the device, which in the worst cases, moves the focus of the user from the object to control to the interface itself. The more intuitive and user-friendly is the interface, the easier is learning how to use it	learning curve steep and lack of any guidance - requiring more than 100 procedures to acquire competency
	Ergonomic	hardware that mediates the interaction minimizes the user physical stress by attenuating the discomfort and the risk of injury	poor ergonomics - leading to wrist tendons inflammations, back and neck pain

comparison between each solution [13]. Indeed, there is a lack of common knowledge about which are the fundamental features of an optimal HMI for a specific clinical task, i.e. the HMI that minimises the clinicians' cognitive and physical load and maximises the clinicians' acceptance and the outcome of the procedure. Therefore, this article presents a framework for the investigation, analysis, and comparison of different HMIs for robotic colonoscopy. The *HMI evaluation framework* includes: 1) an open and modular virtual simulator of robotic colonoscopy, 2) wearable sensors to measure the clinicians' cognitive load, 3) a data collection and synchronization unit, and 4) surveys to evaluate the users' experience. It provides a platform and a method to compare in details different HMIs and identify the optimal one based on the users' performances, cognitive load and physical stress. The final goal is to extract insights, guidelines, and metrics over the design of the next generation intraluminal robotic devices. The simulator embedded in the framework has been validated by 28 gastrointestinal (GI) endoscopists, and a user case study has been conducted with 42 clinicians from different EU hospitals to compare two of the most used commercial interfaces for teleoperated control in robotic colonoscopy (i.e. an haptic serial-kinematic device and a videogame joypad).

II. THE HMI EVALUATION FRAMEWORK

The *HMI evaluation framework* is a comprehensive platform suited to evaluate and compare different HMIs or parts of them in a controlled simulation environment. It is composed of:

- an open, modular, interactive **virtual simulator of robotic colonoscopy**, allowing to freely connect different input devices, implement different control strategies and provide various feedback to the users; the simulator enables the acquisition of all relevant data related to the execution (e.g., time, distance traveled, force exerted on organs *etc.)*;
- two **wearable sensors**, i.e. heart rate band and eye trackers, to track the cognitive load of the users during the experiments;
- a **data collection and synchronization unit** able to gather all the data coming from the simulator (i.e. metrics related to the users' performance during the medical procedure and quality of control) and from the sensors;
- a set of **questionnaires** to collect users' personal information and impressions in terms of preferences, cognitive and physical load, easiness to use, intuitiveness, and satisfaction regarding each device/system tested;
- **HMIs** to be tested; they could be either the controller devices as the experiments reported in this article or specific modules/features of the interface (e.g., type of haptic feedback, control strategies, augmented reality *etc.*)

The tests conducted with this framework give objective and subjective measures about 1) the performances in the clinical scenario, 2) the quality of control 3) the intuitiveness, 3) the *user-friendliness*, and 5) the *ergonomicity* of the HMI tested. Table I provides a definition of the desired characteristics of a HMI (i.e. intuitiveness [14], user-friendliness [15], ergonomicity [5], [16]). Based on this information, the quality of the HMI tested with the framework is assessed. Table I also reports the evaluation of the desired characteristics on the HMI of a standard colonoscope according to [3], [4], [5].

III. THE SIMULATOR

A. Specification

The simulator is the core of the framework and the fundamental infrastructure used to compare the different HMIs. Its design followed the specific requirements listed below:

- **openly and easily interfaceable** - able to connect, receive input and provide output to different devices and systems, and able to collect different data;
- **modular and scalable** - enabling the activation/deactivation of different modules without altering the basic simulation kernel;
- **realistic** in terms of visual and mechanical rendering - allowing a smooth on-line interactive simulation;
- **robust, controllable and repeatable** - to perform multiple user's tests.

None of the simulators available in the literature could satisfy the specific requirements listed above [17]. Therefore, a new virtual simulator was designed and developed to be embedded in the *HMI evaluation framework*. This decision was motivated by the need of having a unique platform that could be customised for the different needs, i.e. for testing different components of the interface both hardware and software. To this end, a virtual platform is more convenient than a physical one, since it allows to 1) easily simulate multiple scenarios, 2) turn on/off and add new features, 3) run repeatable and controllable experiments and 4) accurately

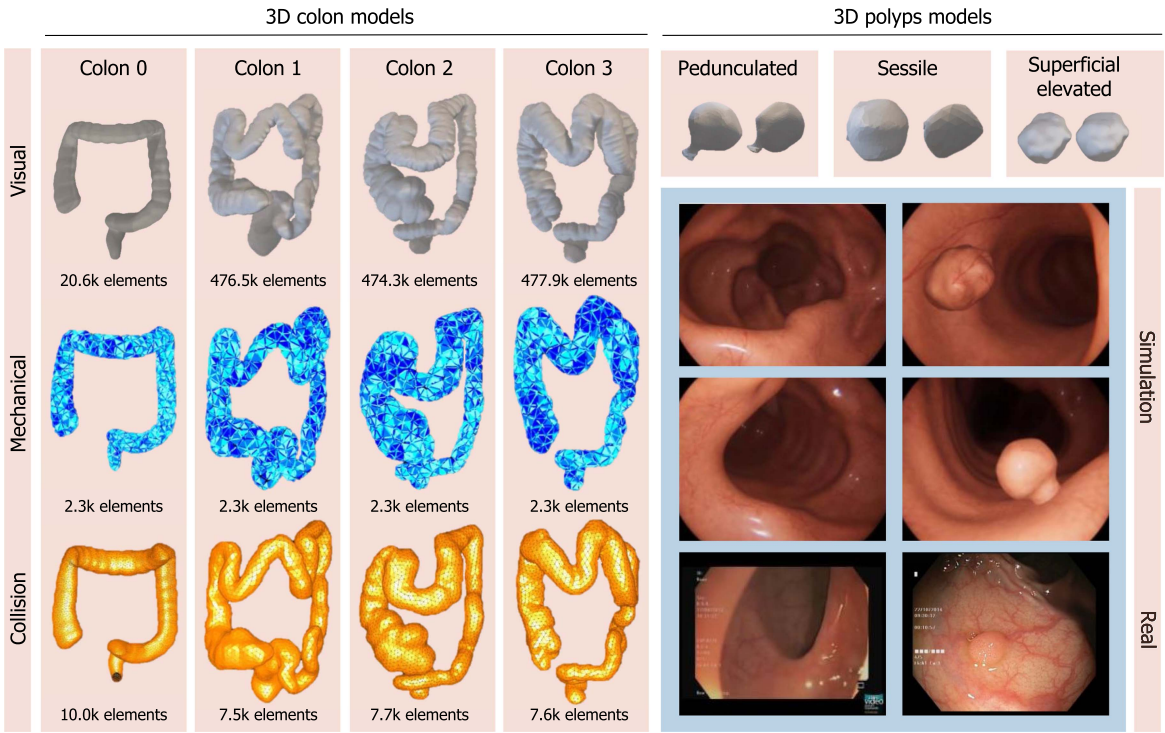


Fig. 1. Four 3-D colon models and three types of polyps models (pedunculated, sessile, and elevated) are reconstructed from CT colonographies images (Colon 0 = C0, Colon 1 = C1, Colon 2 = C2, Colon 3 = C3). For each 3-D colon, three meshes are generated respectively for the visual (high density triangular mesh), mechanical (low density tetrahedral mesh) and collision model (moderate density mesh). Realistic textures are applied on the visual model of both the colon and the polyps to simulate the visual appearance of the colonoscopy videos (blue square: simulated images on the two top rows versus real images on the last bottom row.

track different metrics during the experiments. The simulator was purposely designed to maximise its modularity and ability to be customised for the different testing needs. Hence, it allows to easily load multiple anatomical models, robotic device models and control strategies (i.e. robotic colonoscope), connect different master devices for the guidance of the robot, provide various feedback (e.g., haptic, visual, auditory *etc.*) and record multi-source data. In addition, particular attention was given to realistically reproduce the intraluminal procedure visually and mechanically. In this way, the endoscopists testing the HMIs would feel more prone to perform the simulated procedure as they would do in the real clinical setting [18].

B. Simulator Architecture

The simulation platform was developed under *SOFA (Simulation Open Framework Architecture)* [19], an open and modular development framework oriented to physics simulation. *SOFA* is the central module of the simulator architecture and contains the virtual workspace with the anatomical models and the robotic colonoscope.

The anatomical models are represented with three different meshes: 1) a volumetric tetrahedral mesh representing the mechanical model for deformation and computation of the interaction forces; 2) a low-resolution triangular mesh for collision estimation and 3) a high-resolution triangular mesh for visual rendering. The meshes are obtained using *Autodesk Meshmixer* (Autodesk, San Rafael, CA, U.S.A.) and the open source tool *Gmsh* [20], as shown in Fig. 1. The resolution of the meshes is chosen as a trade-off between simulation accuracy and computational cost, obtaining a realistic visual and force feedback while preserving a real-time simulation.

SOFA physics engine computes collision, deformation and interaction forces between the colonoscope (herein simplified as a capsule with a camera on one side) and the simulated anatomy. The collision endoscope-colon is computed with the *SOFA default pipeline* (contacts solved with the *Lagrange Multiplier method*), using the low-resolution triangular surface.

The platform uses the *SOFA* plugin *SOFAAPAPI-UNITY3D* (InfinyTech3D, Nice, France) to replace the *SOFA* visual rendering module with the *Unity* game engine (Unity Technologies, San Francisco, CA, U.S.A.), to increase the visual realism. A realistic visual feedback is achieved setting the endoscopic camera field of view to 120°, adding the reflection of the light on the surface of the organs, darkening the peripheral of the endoscopic image, and implementing lens distortion and chromatic aberration [21]. *Unity* is also used to interface with the proposed master devices to guide the virtual robotic colonoscope. The endoscopic lightening is rendered with a cone of white light of 140° (intensity = 150 lx and temperature = 7000 K, values obtained by comparing the simulation visual aspect with real endoscopic images). All the software and datasets used for designing the simulator are open source, except for the *SOFAAPAPI-UNITY3D* plugin (connection between *SOFA* and *Unity*). The simulation frequency is 25 Hz in *SOFA* and 20 Hz

in *Unity* under a laptop with Intel(R) Core(TM) i7-10750H processor, CPU of 2.60 GHz, 32 GB of RAM and NVIDIA GeForce RTX 2060 graphic card.

C. Anatomical Model Simulation

The platform has been tested using 3D anatomical models of real patients reconstructed from CT colonographies from the public dataset *Cancer Imaging Archive* [22]. The reconstruction pipeline starts from a pseudo-automatic segmentation of the CT images using the open-source software *3D Slicer* [23], [24]. These models are refined using *Blender* [25], by applying surfaces smoothing, creating the anal sphincter and generating a wall thickness of 2.5 mm [26]). The physical properties of the colon are computed using Finite Element Method solvers provided by *SOFA*, generating realistic deformation of the tissue resulting from the contact with the virtual endoscope tip. The tissue properties, derived from [27] and applied to the tetrahedral mesh, are set as follows: stiffness = 1.5 MPa, mass = 500 g, Poisson Coefficient = 0.3. Physical constraints, modeled as springs with one end in a fixed position and the other end attached to a node of the tetrahedral mesh, are included to constrain maximum colon deformations, generating a more realistic behaviour of the anatomy. The stiffness is fixed to 50 kN/mm following the recommendations of [28]. The deformations are computed using the *SOFA linear Conjugate Gradient solver*. The number of iterations has been empirically set to 25 and has been validated by experienced colonoscopists. The simulated colons have three types of polyps: peduncolated, elevated and sessile, described into the endoscopic classification of superficial neoplastic lesions [29]. The 3D models are obtained following the same procedure used to obtain the colons models, starting from the CT colonography dataset. The polyps are placed in different spots of the lumen, inheriting the same mechanical properties of the colon. For a realistic visual render, the texture of the colon walls and polyps are derived from real endoscopy images of the KVASIR dataset [21], and applied to the high-resolution surface triangular mesh using the *Unity* rendering pipeline.

IV. Data Collection

The simulated environment is designed to perform robotic colonoscopy procedures. In order to compare the users' performances and experiences with the different HMIs, three types of data and metrics are recorded during and after the procedures:

- data correlated with the **clinical outcome** of the procedure (e.g., percentage of total mucosa visualised during the withdrawal, force exerted on the mucosa during the intubation *etc.*);
- data informing about the **cognitive and physical stress** experienced by the users (e.g., gaze entropy, perceived mental demand *etc.*);
- data related to the **quality of control** over the endoscope with the HMI used (e.g., smoothness of the trajectory, control intuitiveness *etc.*).

The execution data is collected from the simulation platform (20 Hz) and the wearable sensors: eye tracking glasses (30 Hz) and heart rate band (1 Hz). The required multi-source data synchronisation is achieved with the *Lab Streaming Layer* [30] and all data are stored into a single data base. In addition, all users are required to fill a pre and post questionnaire. A summary of the most relevant data and metrics is provided in Table II, whereas the surveys are available in Supplementary Materials.

A. Clinical Performance

A list of relevant metrics correlated with the quality of the colonoscopy is derived during the simulated procedure (full list available in Table II). Indeed, during the robotic colonoscopy, the clinician guides the robotic endoscope from the anal sphincter to the cecum (i.e. intubation), minimising the force exerted on the walls to avoid patients' pain and risks of generating lesions. Once the cecum is reached, the endoscope is pulled back while carefully screening the whole mucosa to find any polyp or lesion (i.e. withdrawal) [2]. Therefore, the metrics extracted are divided in two phases, i.e. 1) intubation, and 2) withdrawal. In the clinical practice, the time of withdrawal and Adenoma Detection Rate (percentage of colonoscopies performed by a particular endoscopist in which at least one adenoma was detected) are the main objective metrics used to evaluate the quality of colonoscopy [31]. However, a virtual simulated scenario allows to track more precise indicators of the performances [32], e.g., force exerted on the mucosa wall, percentage of total mucosa visualised and length of trajectory followed by the endoscope.

B. Control Data

Besides the clinical outcome, data recorded from the simulation are analysed to derive insights on the quality of control of the endoscope, e.g., smoothness of trajectory, target location accuracy, *etc.* The experiments reported in this article included a polyp targeting task, which is a standard colonoscopy precision task during withdrawal phase. Therefore, during the tests, the users had to localise and focus each polyp in the middle of a superimposed marker (in the form of an X) over the endoscopic view. The performance is analysed in terms of time for completing the task and target focusing accuracy. In addition, the smoothness of the whole trajectory is computed (Table II).

C. Physiological Data

Biometrical data is measured to objectively estimate the users' cognitive load during the execution of the trials (Table II). Accordingly, the mental stress imposed by each interface cannot be neglected when designing new human-machine interaction paradigms. From the analysis of the literature, the heart rate and gaze entropy represent a good combination for tracking mental overload [33], [34]. Indeed, studies have shown that the heart rate of surgeons increases during stressful tasks [35]. Similarly, the gaze entropy increases when the users perform more complex tasks [33], showing more random exploration patterns. Herein, the heart rate is measured with a Polar H10 chest strap (Polar, Kempele, Finland) at 1KHz, and its running average is transmitted via Bluetooth to the laptop at 1 Hz with a dedicated program. The eyes movements are recorded at 30 Hz with a binocular wearable eye tracking glasses (Pupil Core, Pupil Labs GmbH, Berlin, Germany) [36]. The gaze entropy gives a measure of the average uncertainty over the direction of the gaze

TABLE II
LIST OF ALL THE OBJECTIVE METRICS EXTRACTED FROM THE SIMULATION AND FROM THE WEARABLE SENSORS TO EVALUATE THE HMI

		Metric	Description
Clinical performances	**Intubation**	Time (min)	Time spent to intubate the colon
		Length of trajectory (m)	Length of the path followed by the endoscope during the intubation phase
		Cumulative Deformation (m)	Sum of the maximum deformation of all the elements of the mesh at any time step
	Withdrawal	Time (min)	Time spent for the withdrawal of the endoscope
		Length of trajectory (m)	Length of the path followed by the endoscope during the withdrawal phase
		% mucosa visualized (%)	Percentage of mesh elements visualized during the withdrawal
Control	**Fine movements**	Time (s)	Time spent for targeting the polyp from the first time it appears on the screen
		Error (mm)	Distance between the center of the target and the center of the polyp
	Overall	Fixations	Number of fixations of the device based on a dispersion-duration detection method (max dispersion = 3.0°, min time = 300ms)
		Rotations (rad)	Sum of 3D angles of rotation in absolute value
		Smoothness	Smoothness of the trajectory computed as the cumulative angular variation between consecutive segments of the spatial trajectories: $index_{smoothness} = \frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{L}\Delta\alpha_{ij}$ L = total number of segments of the trajectory $\Delta\alpha_j$ = angular difference between two consecutive trajectory segments * Lower values of the index → smoother trajectories
Physiological data	**Gaze entropy**	Intubation (bit)	Measure of the uncertainty over the gaze position at any point in time $H_g(X) = -\sum p(x,y)\cdot\log_2 p(x,y)$ $p(x,y)$ is the probability that the gaze falls on a certain point of the screen
		Withdrawal (bit)	
		Whole procedure (bit)	
	Heart rate	Intubation (bpm)	Heart rate mean
		Withdrawal (bpm)	
		Whole procedure (bpm)	

at an instant in time during the simulated tasks/procedures [37]. To compute the gaze entropy, the total visual field allowed by the eye tracker is divided into 80 × 54 degrees of visual angle (DVA), generating 4320 bins of 1×1 DVA. Gaze data with a confidence lower than 0.8 are discarded, following the *Pupil Labs* recommendations [38]. Therefore, the probabilities of the gaze falling on each bin is computed, and the gaze entropy is derived as in [33]. The gaze analysis is also used to derive the number of fixations of specific parts of the HMI (i.e. the controller device) [39], [40]. A fixation is counted when the gaze falls on a point of the controller for a minimum time of 300 ms [38]. The position of the controllers is constantly tracked by the *Pupil Labs Surface Tracking* plugin using printed markers.

D. Surveys

Three sets of questionnaires are used for the users' subjective evaluation of the HMI and the analysis of the usefulness/easiness. The first one, administered at the beginning of the experiments, examines the past experience of the users: 1) experience in colonoscopy, 2) experience with video games, simulators or musical instruments, and 3) level of tiredness at the moment of the experiments, i.e. number of hours slept the night before the tests and the number of hours worked on the same day (full survey available in Supplementary Materials). This information is relevant for the cognitive load analysis. The second set of questions investigates the subjective experience of the user with the HMI, and is administered right after having completed the experiments with a HMI. Inspired by the NASA-Task Load Index [41] and the Borg rating of perceived exertion [42], the questions explore five different areas: 1) mental demand, 2) physical demand, 3) subjective impression on the own performance, 4) effort, 5) frustration. These questions are based on Likert scale (1-5) and are administered right after having tested each HMI (Supplementary materials, survey II). A final questionnaire is conceived to be delivered at the end of all the trials (having tested all the HMI in the study), and review the subjective mental and physical stress associated with each tested platform (Supplementary Materials, survey III). In this way, the users are pushed to compare the HMI and provide a more informed opinion about their favorite one. As further explained in Section V, the order each HMI is tested is random

V. CASE STUDY

The *HMI evaluation framework* was tested for the first time by comparing two different physical devices for teleoperated control of the robotic endoscope. The experiments were designed to answer the following question: *which is the optimal controller for intelligent teleoperation of robotic colonoscopes, i.e. the one minimising the users' cognitive load and maximising the outcome of the procedure?* Considering the different levels of autonomy of medical robots described in [43], this study assumes the "robot assistance" modality. This first level of robotic assistance is called "intelligent teleoperation" for the specific application of robotic colonoscopy [44]. In this case, the operator guides the tip of the robotic endoscope, while the control system generates the required low level control strategy to enable the execution of the desired motion [44], [45]. Taking as an example the magnetic colonoscopes, the teleoperator guides the colonoscope tool tip and the robotic control system computes the movements of the external permanent magnet

to have the robotic capsule reaching the point desired by the operator [44]. Thus, only the tip of the endoscope is simulated in the experiments, since the robotic endoscope control is out of the scope of this study.

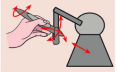
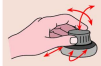
A. HMI Screening Survey

Considering the high number of HMI designed for the teleoperation of robotic endoscope, a survey was conducted to perform a first screening among them. Its goal was to select a set of interfaces with the most interesting features for the GI endoscopists to be tested with the *HMI evaluation framework* (Full survey available in [46]). Although there are few examples in the literature of robotic devices driven by standard endoscopes controllers [47], most of the innovative intraluminal systems introduce new control interfaces [44], [48], [49], [50], [51], [52]. Therefore, most of these controllers were analysed and their main features extracted. The questionnaire administered to the endoscopists required to rate with a Likert scale (1-5) the level of agreement regarding the inclusion of these features on the next generation HMI for robot-assisted colonoscopy. Explanatory graphics were provided to help understanding the questions, which were conceived jointly by GI endoscopists and engineers. The questions inquired about specific parts of each interface (e.g., type of control, shape of the handler, presence of force feedback *etc.*). Each query was not directly linked to the controller itself to avoid biases and to allow participants not familiar with all the interfaces to give their valuable opinion. Consensus measure [53] was used to assess the dispersion of the clinicians' answers. Four different controllers were evaluated, chosen as the ones most used for robot-assisted colonoscopy, and having configurations similar to most of the HMI used in the literature: 1) thumb-driven videogame joypads using the two finger levers for insertion/retraction and deflection, and controlling the roll and the extra functionalities with buttons integrated into the controller; 2) haptic device with a spring-mass mechanism for the insertion/retraction and deflection/roll (proportional control), and extra functionalities controlled with the buttons on the controller; 3) one-hand joystick for the control of all the movements of the endoscope and pedals for extra functionalities; and 4) 3D mouse enabling insertion/retraction by pressing the device inward/outward and extra functionalities by buttons. See Table III for better visualisation of the mapping of DOFs in each controller and the related features. For each HMI, a sum of all the scores obtained by a feature of the interface was computed. Scores were proportionally distributed between *absolutely disagree* $= -2$ and *absolutely agree* $= 2$, and summed for all the participants for each interface. The final interfaces chosen for the testing with the *HMI evaluation framework* better explained in Section VII-A are the videogame joypad and the haptic device.

B. Subjects, Experimental Design, and Procedure

A total of 42 GI endoscopists were enrolled for the experiments: 20 novices ($<$ 1 year of experience with colonoscopy) and 22 experts ($>$ 1 year of experience with colonoscopy and $>$ 150 colonoscopies performed [17]). Considering the speciality, 21 were colorectal surgeons and 21 gastroenterologists. The participants were asked to perform six simulated robotic colonoscopies: three with one device (Videogame Joypad, VJ, i.e. DualShock 4 controller of PlayStation, Sony Interactive Entertainment Inc., Tokyo, Japan) and three with the other one (Haptic Device with serial architecture, HD, i.e. Touch, 3D Systems Corp., Rock Hill, SC, U.S.A.). As mentioned before, the simulation platform assumes level 1 of robotic assistance. Therefore, the inputs on the controller devices were directly mapped over the tip of the endoscope (control in the image frame, see Fig. 2 block "Controllers") with the following control modalities: 1) HD with a spring-mass mechanism for the insertion/retraction and deflection/roll (proportional control), and extra functionalities controlled with the buttons on the controller; 2) VJ using the two finger levers for insertion/retraction and deflection, and controlling the roll and the extra functionalities with buttons integrated into the controller. The study followed a 2 $\times$ 2 mixed factorial design, considering 1) the two levels of experience in colonoscopy (novices vs. experts), and 2) the two devices (HD vs. VJ). For each device, the first two procedures

TABLE III
CLASSIFICATION OF THE COMMERCIAL CONTROLLERS ACCORDING TO THE PRELIMINARY SURVEY

		Videogame Joypad	Haptic Device	Hand Joystick	3D Mouse
					
Control	Control movements of the endoscope	Two hands +44	One hand +14	One hand +14	One hand +14
	Control insertion/retraction of the endoscope	Move a lever placed on top of the controller with a thumb (up/down) +53	Push/pull the manipulator along a specific direction (in/out) -34	Press buttons integrated in the controller +22	Push/pull the manipulator along a specific direction (in/out) -34
	Control deflection/rotation of the tip of the endoscope	Move two levers placed on top of the controller with the thumbs (up/down - right/left) +35	Move a manipulator along its three main axes (up/down - right/left - in/out) and twist it -23	Move a lever along its two main axes (up/down - right/left) and rotate it around itself +27	Move a lever along its two main axes (up/down - right/left) and rotate it around itself +27
	Control of extra functionality	Buttons on the controller +67	Buttons on the controller +67	Pedals +10	External set of buttons -40
Ergonomy	Shape of the controller	Playstation style joypad +57	Cylindric device +34	Cylindric device +34	Knob -39
	Body parts involved in the control	Only wrist +40	Wrist and forearm -40	Wrist and forearm -40	Only wrist +40
Feedback	Modular force feedback	no -93	yes +93	no -93	no -93
	Type of force feedback	Vibration +41	Movement restriction +43	Vibration +41	Visual information +51
	Total	+244	+154	+15	-74

For each of the four configurations, and for each of the questions, the sum of the scores obtained is reported in blue (Likert scale where *absolutely disagree* is -2 and *absolutely agree is* +2). The joypad and the haptic device are the preferred controllers.

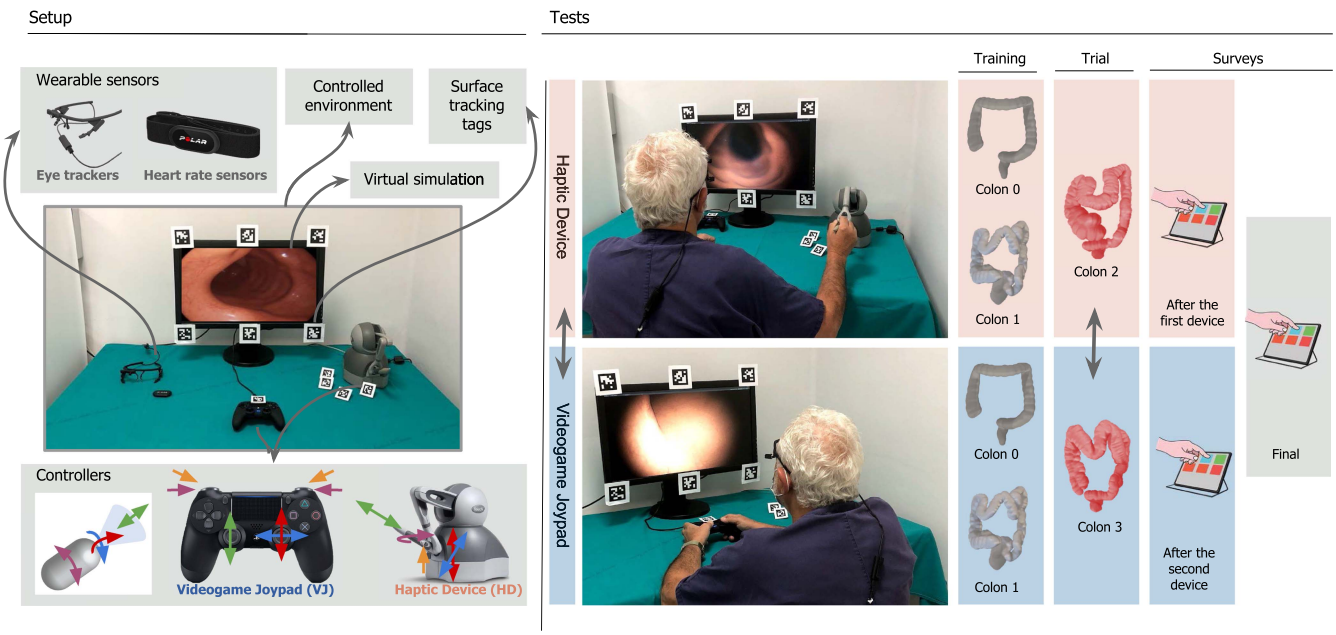


Fig. 2. Experimental setup (left) and experimental design (right). The experiments are performed in a controlled environment (no external disturbances), using the colonoscopy virtual simulator, the eye trackers and chest band to track the cognitive load, and the two controllers: haptic device with serial architecture and video game joypad. The mapping of the degrees of freedom between the controllers and endoscope is shown at the bottom-left, where the yellow arrows represent the extra function of activating/deactivating the polyp target for the targeting task. The users are required to perform three colonoscopies for each device: the first two procedures are for training, whereas the final one is the test. Surveys are administered after each trial and at the end of the whole experiment.

were used as a training phase, while the last one was considered as a valid trial (see Fig. 2). The training was performed always with the same two colons (Fig. 1: C0, C1, where C stands for Colon), while the two testing trials were conducted with two different colons (C2 and C3). Potential practice/learning effects on the medical procedure were controlled by a Latin square design across both the device (half of the participants started with the VJ and the other half with the HD), and across the colon used for the trial (half of the participants performed the trial with the VJ in C2 and the trial with the HD in C3, while the second half followed the opposite sequence). This balance was ensured also among each group with the same level of experience, i.e. novices and experts. Thus, the possible effects of confounding factors, including learning of series effects, and task-switching costs (i.e. the costs associated with going from a complex task to an easy one) were minimised. In addition, the two training procedures before the trial ensure that all the subjects have the same level of experience with the simulator. Each experimental session took place in a dedicated training room inside Hospital de la Santa Creu i Sant Pau (Barcelona, Spain) and A.O.U. Citta della Salute e della Scienza di Torino (Torino, Italy). Before starting the experiments, all the subjects were given the same clear instructions about the tasks to do: 1) perform a complete colonoscopy starting from the rectum and reaching the cecum; 2) once reached the cecum, withdraw the endoscope looking for polyps; and 3) for each polyp found, take a picture by centring the lesion on a specific target (two brackets square on the side and a cross at the centre). A single experimental session lasted around 95 minutes. Subjects were allowed to take a short period of rest (less than 5 minutes) between subsequent experimental trials. Special care was dedicated to avoid any distraction that could interfere with the users' performance and mental stress (i.e. silence, removal of mobile phones/smart watches or any source of notifications, stable light, forbidden entrance to any external person in the room). Surveys, described in Section IV and reported in Supplementary Materials (survey I-III), were administered at the beginning of the tests, after each trial and at the end of the whole experiment (Experimental setup available in Fig. 2). The experimental study was carried out following the principles of the Declaration of Helsinki.

C. Data Collection and Analysis

All the data and metrics described in Section IV were collected from the surveys, the simulation platform, and the wearable sensors. Setup of the eye tracking systems, including calibrations, and of the heart rate band preceded the start of the experiment. To analyse the effect of the device on the clinicians' performances, a series of separate unpaired Mann-Whitney U tests were conducted comparing the distribution of the medians for each metric between the two devices (HD vs. VJ) for 1) all the subjects, 2) only novices and 3) only experts. In addition, considering that one of the two colons used for the trials (C2) resulted to be slightly more difficult to navigate (longer and with more curves than C3), the tests were conducted also for each colon used in the trials: 4) all subjects, 5) novices and 6) experts in colon C2, and 7) all subjects, 8) novices and 9) experts in colon C3. Concerning the surveys and the physiological data, a series of paired Mann-Whitney U tests were run, comparing the distributions of the differences of each subject's metric/answer

for the two devices: VJ vs. HD for 1) all participants, 2) only novices and 3) only experts. Consensus measure as computed in [53] was used to assess the dispersion of the clinicians' answers to the surveys. Each test was considered significant for p-values < 0.05.

VI. Validation of the Simulator

The clinical validation assesses whether the simulator is a realistic representation of the real procedure and therefore can be used to test the HMIs. The realism of the simulator (i.e. face validity) was assessed by a group of GI endoscopists that took part into the experiments comparing the two HMIs. All the participants had prior experience in colonoscopy: 28 subjects with an average of 10 years of experience. At the conclusion of the experimental session, the clinicians were requested to assess on a Likert scale (1-5) the realism of different aspects of the simulation (i.e. overall procedure, visual rendering including illumination and camera field of view, mechanical deformation of the tissue, organ anatomy and appearance and location of polyps). The complete survey can be found in the Supplementary Materials - survey IV. The results were analysed as follows: the consensus measure was used to assess the dispersion of the answers, whereas the mean value of each answer was used to assess the realism.

The content validity (assessment of the suitability of a simulator as a training tool) [54] is not addressed in this study: the platform is not being evaluated as a training device, but as a framework to develop and analyse HMIs. Concerning the construct validity (ability of a simulator to distinguish novices from experts), this study compared the performances of the novices vs. those of the experts by means of a separate unpaired Mann-Whitney U test for each metric. However, it is worth to note that this platform simulates a robotic colonoscopy at level 1 of autonomy. In consequence, the evaluated procedure differs from the conventional colonoscopy in two basic aspects: 1) input controller, 2) the absence of the long passive endoscope shaft. Therefore, the experts in the conventional procedure might not be considered experts in handling the robotic platform.

In addition, subjects' performances in C2 were compared with those in C3 to check whether a more difficult colon (i.e. longer and with more curves) implied significantly worse performances. This aspect was evaluated with a separate unpaired Mann-Whitney U test on each metric: C2 vs. C3 for 1) all participants, 2) only novices, and 3) only experts (significance for p-values <0.05).

VII. Results of the Tests

A. HMI Screening Survey

Of the 71 participants, 80% were gastroenterologists, while the other 20% were colorectal surgeons. The clinicians had different levels of experience in colonoscopy: 15% had < 2 years, 58% had > 10 years, and 27% were in the middle. The result of the analysis shows that the characteristics of the videogame joypad and the haptic devices are the most preferred by the users, collecting respectively +244 points and +154 points. Therefore, these two controllers were chosen to be tested with the HMI evaluation framework. Both the 3D mouse and the hand joystick were discarded because they reached low scores (-74, +15), and they were considered not worthy for test (Table III). Indeed, the 3D mouse got a negative score while the hand joystick got +15 points (which is less than the 6% of the points collected by the joypad).

B. Validation of the Simulator

Face validity was assessed by 28 GI endoscopists, both gastroenterologists and colorectal surgeons. As shown in Fig. 3, all the questions got an average level of satisfaction ≥ 3 over 5. With a consensus always > 0.8, the survey shows a high level of agreement between the clinicians for each question. The construct validity could not be assessed as no significant difference, in terms of performance metrics, was detected between the novices and experts. This result suggests that the use of two user-friendly interfaces (VJ and HD) and the easing of the procedure provided by the robot autonomy (level 1) decreases the performance gap between experts and novices. However, further validation of the realism of the simulator was given by the fact that the performances between C2 and C3 differ statistically (see Supplementary Materials for the p-values of the statistical tests and for the mean values of the metrics recorded; Fig. 4 for the box plots). Indeed C3 is longer than C2 (C2: 125 cm C3: 135 cm) and has more curvatures (sum of 3D angles C2 $= 48°$ and C3 $= 63°$), making the procedure harder to perform. This was reflected in most of the clinical metrics, which got worse in C3).

C. Best Human Machine Interface

Of the 42 subjects that performed the experiments, 5 were discarded. Among them, 4 participants could not successfully complete one of the two trials, whereas for one subject there was a system failure during the collection of the data. The analysis shows that both the interfaces selected (HD and VJ) represent a valuable solution for teleoperated control of a robotic colonoscope, since none of them provided poor results. However, a few differences were detected between the two options, making the VJ the best option for teleoperated control of a robotic colonoscope. Indeed, the VJ 1) enabled better clinical performances (higher percentage of mucosa visualised in C3, the most complex colon to examine), 2) facilitated the control (lower error in the targeting task in C3), 3) was objectively more user-friendly (less fixations, lower gaze entropy and mean heart rate) and 4) less physically demanding (expressed through the survey). Nevertheless, the HD 1) provided smoother trajectories and 2) was perceived as more user-friendly and intuitive by the users (rated in the survey as less difficult to use, less mentally demanding for the withdrawal phase and enabling better performances). In addition, 3) the majority of the users preferred the HD with respect to the VJ, especially among the experts, as expressed in the final questionnaire. A summary of the overall results is shown in Table IV, whereas the results of the survey are presented in Fig. 6. All the p-values of the statistical tests and the mean and standard deviations of each data recorded are available in Supplementary Materials; the boxplots of the data are shown in Figs. 4 and 5.

1) Clinical Performances: During the intubation phase, no statistical significance was detected between the two devices. However, in the withdrawal phase, the VJ succeeded the HD

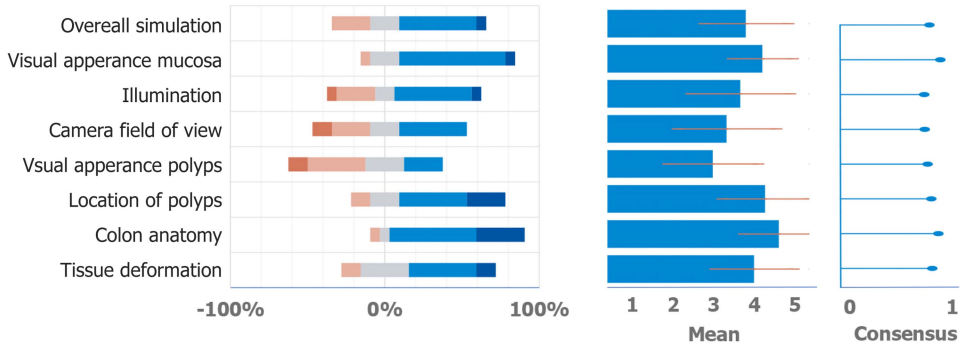


Fig. 3. Content and face validity of the virtual simulator. Distribution of the answers provided by 28 clinicians using the Likert scale (left), average score and standard deviation (centre), and consensus (right). The consensus measures the dispersion of the clinicians' answers [53].

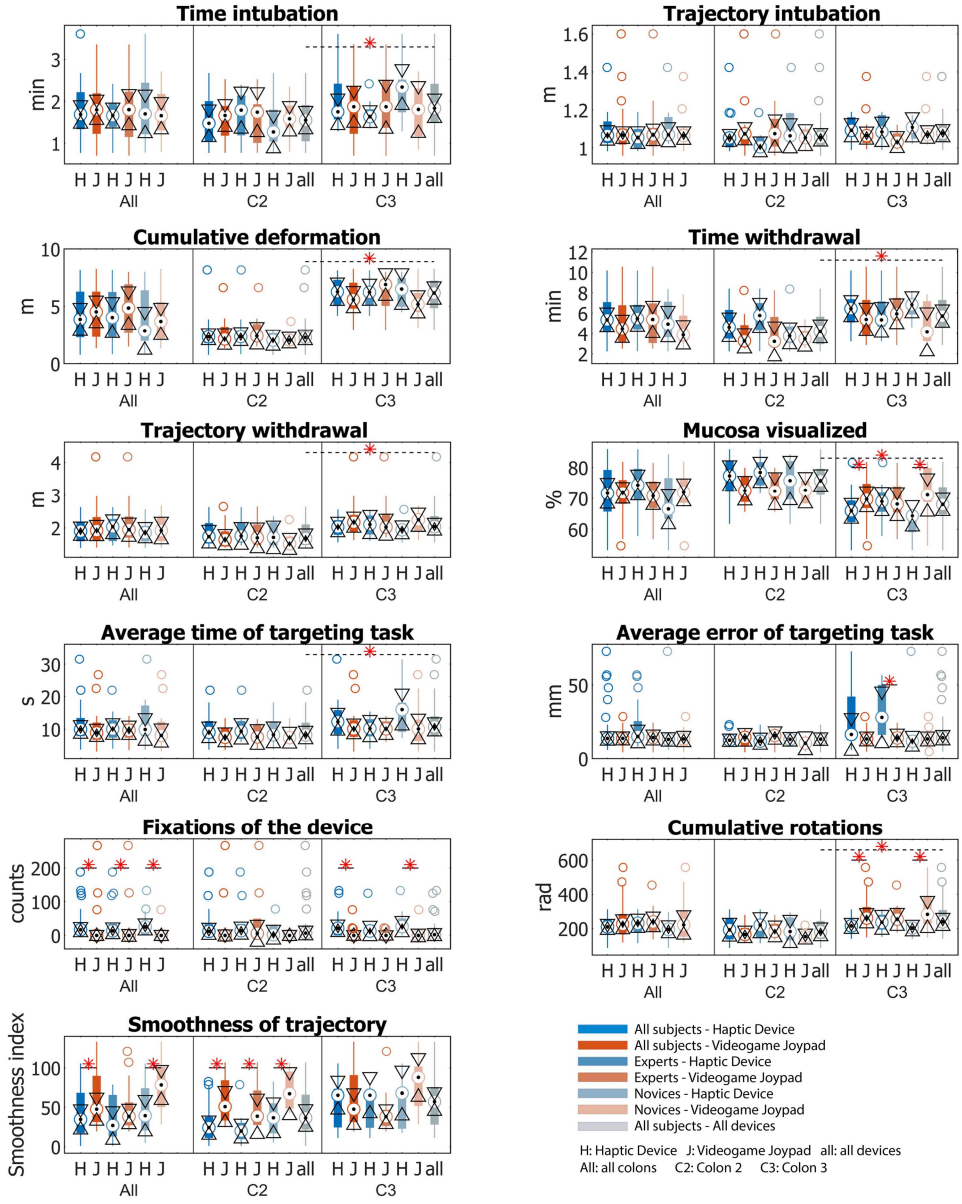


Fig. 4. Metrics recorded during the experiments divided for controller device (H: haptic device, V: videogame joypad, all: all devices), colon (All: all colons, C2: colon 2, C3: colon 3) and level of expertise (All subjects, experts, novices). Statistical significance on the Mann–Whitney test is highlighted with the star (p-value < 0.05). The median is represented by the circles with a black dot inside and its 95% confidence interval is delimited by the triangles.

TABLE IV
SUMMARY OF THE METRICS IN WHICH A STATISTICAL DIFFERENCE WAS FOUND BETWEEN THE TWO DEVICES DURING THE CASE STUDY

		Metric	VJ	HD
Clinical performances		% mucosa visualized	C3 N/All	
Control	**Fine**	Error	C3 E/All	
	Overall	Smoothness		All C N/All C3 N/E/All
		Fixations	All C N/E/All C3 N/All	
Physiological data	**Gaze entropy**	Withdrawal	N/All	
		Whole procedure	N/All	
	Heart rate	Intubation	E/All	
		Whole procedure	E/All	
Survey	**Mental demand**	5.Intubation		E/All
		9. Success		E
		10. Difficulty		All
		8. Physical demand	N/E/All	
Final survey		13. Physical demand	N/E/All	

Videogame Joypad (VJ) vs Haptic Device (HD). For each metric, the coloured cell under one of the two devices (blue for VJ and green for HD) shows that it performed statistically better than the other. The specific condition in which the statistical difference was found is reported inside each coloured cell (C: colon, N: novices, E: experts, All: both experts and novices, *e.g.,* novices performed statistically better in terms of percentage of mucosa visualised in colon 3 with the VJ with respect to the HD).

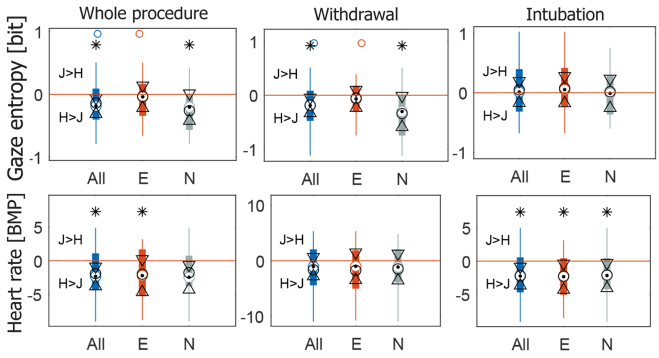


Fig. 5. Differences on the metrics recorded during the experiments between the videogame joypad (V) and the haptic device (H) divided for level of expertise (All: all subjects, E: experts, N: novices). Statistical significance on the paired Mann-Whitney test is highlighted with the star (p-value < 0.05). The data above the orange line shows higher values for the videogame joypad, whereas those under the line for the haptic device. The median is represented by the a circles with a black dot inside and its 95% confidence interval is delimited by the triangles.

in the most complex colon (i.e. C3) with a higher percentage of mucosa visualised (mean values for novices: 72% vs. 62%, and for all groups: 71% vs. 66%, p-value < 0.05, Fig. 4). The percentage of mucosa visualised during the withdrawal, as the cumulative deformation of the colon walls during the intubation phase, is considered the most important metrics to evaluate the quality of the procedure. Indeed, to maximise the diagnostic outcome of the colonoscopy, the mucosa visualised should be 100%. Whereas the force exerted on the walls should be minimised to avoid patient's discomfort and risk of lesions.

2) Control Precision: The polyp targeting task showed that the VJ is slightly more precise than the HD, achieving lower mean errors in C3 (for experts: 14 mm vs. 33 mm, p-values < 0.05). However, clinicians did not feel a difference in the difficulty of performing the required fine movements with the two devices (Fig. 6). Whereas, the results of the smoothness metric suggest that the HD enables smoother trajectories (All colon mean values for novices: 47 vs. 77, and all participants: 40 vs. 61; in C2 mean values for experts: 24 vs. 50, for novices: 39 vs. 71 and for all participants: 31 vs. 59; p-values < 0.05; lower values of the index means higher smoothness levels).

3) Intuitiveness: No statistical differences were detected in the questions regarding the intuitiveness of the devices (Fig. 6). However, 11 of the 21 clinicians that preferred the HD device said it was due to its intuitiveness "feeling as they had the tip of the endoscope in their hand". In contrast, only 4 clinicians claimed they preferred the VJ for its intuitiveness. All of them have had previous experience with the VJ playing at video games, therefore feeling more familiar with it.

4) User-Friendliness: The number of fixations of the HD was higher than the VJ suggesting that the HD was less easy-to-use, and required more visual supervision (mean values in all C for experts: 23 vs. 34, novices 7 vs. 29, and all participants 16 vs. 31; in C3 for novices: 10 vs. 40 and all participants: 7 vs. 31; p-values < 0.05). All the clinicians felt the HD was easier to use (Q10 Survey II, p-value < 0.05 for all participants) and less mentally loading in the intubation phase (Q5 Survey II, p-values < 0.05 for experts and all participants). However, these results were not confirmed by the final questionnaire in which both the devices were overall rated as they implied the same difficulty level. Additionally, the experts felt to be more successful with the HD (Q9 Survey II, p-value < 0.05 for experts) despite the clinical performances do not reflect this impression (Fig. 6). Regarding the objective measure of the cognitive load, both the gaze entropy and the heart rate suggest that the VJ is less cognitively stressful (see Fig. 5 and the Supplementary Material). For the gaze entropy the VJ implied an average reduction of 0.3 b for novices and 0.2 b for all participants during the withdrawal, and 0.3 b for novices and 0.1 b for all participants in the whole procedure (p-value < 0.05). Whereas, the mean heart rate was reduced of about 2 BPM for the experts and all the participants in both the intubation and the overall procedure (p-value < 0.05).

5) Ergonomics: The surveys clearly reveal that the HD is less comfortable and ergonomic than the VJ (Q8 Survey II and Q13 Survey III for all groups p-value < 0.05; see Fig. 6). However, during the experiments, the HD was fixed in a place for standardising the experience, whereas the possibility to better adjust its position for each user could reduce the discomfort.

VIII. DISCUSSION

The optimal HMI for robot-assisted intraluminal procedures should minimise the users' cognitive and physical load while maximise the outcome of the procedure. Considering the wide range of HMIs developed in the last decades for robotic colonoscopes, it is hard to determine which is the optimal one. Indeed, the terms "ergonomic", "intuitive", "user-friendly" are used across several scientific articles with different meaning, and referring to different characteristics of the final interface. Therefore, this article proposes a rigorous protocol to design HMI based on objective and quantifiable measures of the "ergonomy", "intuitiviness", "userfriendliness", and more in general of the quality of the final interface. A fundamental piece of this workflow is the inclusion of clinicians in the design loop of the medical device (i.e. robot-assisted system), from the initial phases by conducting surveys and performing user tests studies.

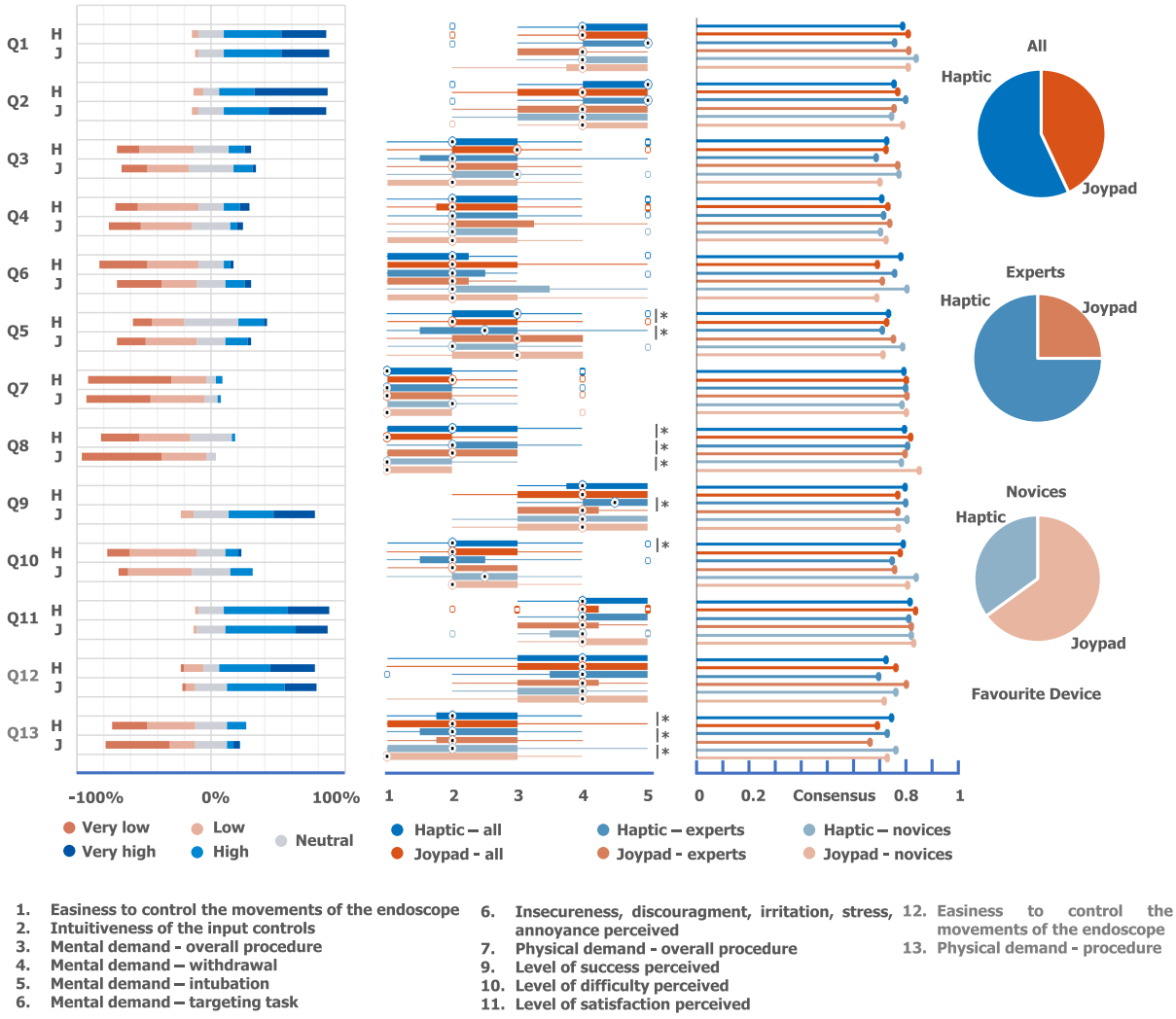


Fig. 6. Summary of the questions and answers to the survey administered to the clinicians after each trial (1-11) and at the end of all the experiments (12-13): distribution of answers (right), boxplots divided by level of experience (centre-right), consensus (centre-left) and pie plot of the favorite device (right). Statistical significance on the paired Mann–Whitney test is highlighted with the star on the boxplot (p-value < 0.05).

The article proposes a list of required metrics to evaluate HMIs in the context of robotised intraluminal procedures with certain degree of autonomy (teleoperator guiding the endoscopic tip). These metrics can differ from those used to evaluate the performance of colonoscopists in the clinical practise and those measured in training simulators. Indeed, both the simulated procedure (conventional vs. robotised) and the final goal (training vs. testing HMI) are different. A virtual platform used jointly with physiological sensors enables the analysis of multiple factors defining the performance of each HMI in different aspects (e.g., clinical outcome, intuitiveness, etc.).

The HMI requirements are directly related with the level of autonomy provided by each robotic platform. In this context, the case-study presented assumes the level 1 of autonomy. Indeed, intelligent telemanipulation is the level of autonomy most easily translated to the clinical practise. Considering the wide amount of controllers for telemanipulation of intraluminal robots, a preliminary survey was run to choose a set of devices with interesting features to test. The survey was conceived by a board of technicians and clinicians, which evaluated different HMI at both technical and medical level. A limitation of the survey is that it does not take into account all the possible controllers available, and focus more on traditional commercial platforms. This choice was motivated by the need of limiting the possibilities by giving priority to the most used interfaces. However, any other interface not considered in this study could be easily tested in the future with the *HMI evaluation framework*.

Indeed, a variety of controllers and emerging technologies, e.g., eye trackers [55], gesture recognition [56], voice/speech recognition [57], mixed reality headsets [58], and handheld motion-sensing controllers [59], have recently been investigated for endoscope control in research domains. These technologies provide means for endoscopists to navigate, manipulate instruments, and interact with the endoscope with eye movements, hand gestures, spoken commands, augmented visualization, and natural hand movements. Such HMIs have the potential to enhance procedural outcomes and improve the endoscopist's experience. Therefore, further exploration and testing of them with the proposed framework could be warranted.

The preliminary survey confirms the results of the study conducted in [13]: haptic device outperforming the 3D mouse and hand-held control. However, our study introduces a controller not considered in [13], i.e. the videogame joypad, which results to be the preferred one from the survey and the optimal one from the experiments.

Regarding the comparison between the videogame joypad and the haptic device, no massive differences were detected. However, both the HMIs were selected after a preliminary screening survey involving the clinicians, therefore they had many of the characteristics requested by the users. Nevertheless, interesting results were observed from the experiments. The HD was preferred and generally felt as more intuitive, less hard to control, and more empowering, especially by the expert clinicians. However, the VJ allowed better clinical performances, finer control and lower objective cognitive load (measured through the sensors) and perceived physical load. Although many young clinicians might have been biased by their previous experience with the VJ, also the other participants (without prior experience with video games) had similar results. In the VJ the directional commands are decoupled between the two hands: one hand controls the rotation while the other one is in charge of the insertion/retraction. Although this paradigm might be seen as less intuitive, it could have eased the control of the movements of the endoscope. Additionally, the HD requires more physical effort to be controlled being less ergonomic than the VJ. Therefore, the increasing fatigue during its use could have had a harmful impact on the users' performances.

A commercial intraluminal robot using the joypad is the MONARCH Platform (Johnson and Johnson, NJ, U.S.A.), designed for bronchoscopy. A similar configuration is also adopted by the Ion robot (Intuitive Surgical, Inc., CA, U.S.A.) and by the Corindus Vascular (Siemens Healthineers, Erlanger, Germany). Indeed, the Ion replaces the joypad with two spheres moved with two fingers (two indices). Whereas, the Corindus Vascular uses two hand joysticks. In both the configurations there is a decoupling of the controls (i.e. insertion/retraction is controlled with one hand/finger, while the deflection with the other) and same type of inputs to control the endoscope (i.e. movements of two "levers").

The drawback of the standard joypad is the impossibility to provide haptic constraints, which is the main advantage of the haptic device, and, as reported in the HMI evaluation survey, is highly requested by the endoscopist. Indeed, an excessive pressure on the colonic wall is the main cause of perforation, the most feared adverse event during a diagnostic colonoscopy. Force feedback is therefore important for preventing the surgeon from causing perforations. Thus, future studies involving the *HMI evaluation framework* could focus on the optimal way to provide the force feedback (i.e. haptic feedback, augmented reality, visual warnings, auditory alerts, *etc*.), and how to embed this feature on the "joypad-style configuration control".

IX. Conclusion

This article presents the first complete framework designed to rigorously compare different HMIs for robot-assisted colonoscopy. The *HMI evaluation framework* allows to design and evaluate different HMI in a controlled environment to derive insights about their performance. It comprises 1) an endoluminal virtual simulator (configured to simulate robotic colonoscopy procedure), 2) physiological sensors, 3) surveys exploring the subjective evaluation of the users, 4) objective metrics to evaluate the HMI, and 5) a protocol for the testing of the interfaces. The *HMI evaluation framework* provides a method to study the performance of the interfaces both at the execution and physiological level, and combines all these metrics in a multi-parametric analysis. A complete study investigating the optimal device for intelligent teleoperated control of robotic colonoscopes was run with 42 GI endoscopists. Nevertheless, the proposed experimental protocol is meant to be applied for testing different components of the HMI: the usability of a new assistive tool (e.g., autonomous polyp detection), the optimal way to convey a piece of information (e.g., haptic feedback vs. augmented reality), the usefulness of autonomous navigation, etc. Indeed, this framework, and more in general the method proposed, enables to analyse in detail the different aspects of the HMI to determine the optimal interface in terms of clinical outcome, intuitiveness, user-friendliness, and ergonomics.

Another important contribution is the development of a simulation platform for colonoscopy, allowing to test the HMI in a controlled and safe pre-clinical environment. The simulator was designed with a modular and flexible architecture, in order to be adaptable to different simulation and evaluation requirements. It has been successfully validated by 28 GI endoscopists and resulted to provide realistic visual and mechanical rendering.

References

[1] V. Vitiello et al., "Emerging robotic platforms for minimally invasive surgery," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 111–126, 2013.

[2] S.-H. Lee et al., "Colonoscopy procedural skills and training for new beginners," *World J. Gastroenterol.*, vol. 20, no. 45, 2014, Art. no. 16984.

[3] B. Spier et al., "Colonoscopy training in gastroenterology fellowships: Determining competence," *Gastrointestinal Endoscopy*, vol. 71, no. 2, pp. 319–324, 2010.

[4] D. E. Yung et al., "Musculoskeletal injuries in gastrointestinal endoscopists: A systematic review," *Expert Rev. Gastroenterol. Hepatol.*, vol. 11, no. 10, pp. 939–947, 2017.

[5] A. K. Shergill and K. R. McQuaid, "Ergonomic endoscopy: An oxymoron or realistic goal?," *Gastrointestinal Endoscopy*, vol. 90, no. 6, pp. 966–970, 2019.

[6] G. Ciuti et al., "Frontiers of robotic colonoscopy: A comprehensive review of robotic colonoscopes and technologies," *J. Clin. Med.*, vol. 9, no. 6, 2020, Art. no. 1648.

[7] N. Mahmud et al., "Computer vision and augmented reality in gastrointestinal endoscopy," *Gastroenterol. Rep.*, vol. 3, no. 3, pp. 179–184, 2015.

[8] J. P. Desai et al. *Encyclopedia of Medical Robotics*. Singapore: World Science, 2017.

[9] O. F. Ahmad et al., "Artificial intelligence and computer-aided diagnosis in colonoscopy: Current evidence and future directions," *Lancet Gastroenterol. Hepatol.*, vol. 4, no. 1, pp. 71–80, 2019.

[10] E. Rozeboom et al., "Intuitive user interfaces increase efficiency in endoscope tip control," *Surg. Endoscopy*, vol. 28, no. 9, pp. 2600–2605, 2014.

[11] C. Fan et al., "Control devices and steering strategies in pathway surgery," *J. Surg. Res.*, vol. 193, no. 2, pp. 543–553, 2015.

[12] F. O. Marcondes et al., "Adenoma detection rate falls at the end of the day in a large multi-site sample," *Dig. Dis. Sci.*, vol. 63, no. 4, pp. 856–859, 2018.

[13] G. Ciuti et al., "A comparative evaluation of control interfaces for a robotic-aided endoscopic capsule platform," *IEEE Trans. Robot.*, vol. 28, no. 2, pp. 534–538, Apr. 2012.

[14] A. Blackler, V. Popovic, and D. Mahar, "Intuitive interaction applied to interface design," in *Proc. New Des. Paradigms: Proc. Int. Des. Congr.. Int. Des. Congr.*, 2005, pp. 1–10.

[15] J. Lowgren et al., *The Encyclopedia of Human-Computer Interaction*, 2nd ed. Aarhus, Denmark: Interact. Des. Found., 2012.
[16] M. R. Lehto and S. J. Landry, *Introduction to Human Factors and Ergonomics for Engineers*. Boca Raton, FL, USA: CRC Press, 2012.
[17] M. Finocchiaro et al., "Training simulators for gastrointestinal endoscopy: Current and future perspectives," *Cancers*, vol. 13, no. 6, 2021, Art. no. 1427.
[18] E. Wilson et al., "Is that realistic? the development of a realism assessment questionnaire and its application in appraising three simulators for a gynaecology procedure," *Adv. Simul.*, vol. 3, no. 1, pp. 1–7, 2018.
[19] F. Faure et al., "SOFA: A multi-model framework for interactive physical simulation," in *Soft Tissue Biomech. Model. Comput. Assist. Surg.*, vol. 11, Y. Payan, Ed. Berlin, Germany: Springer, Jun. 2012, pp. 283–321. [Online]. Available: https://hal.inria.fr/hal-00681539
[20] C. Geuzaine and J.-F. Remacle, "Gmsh: A three-dimensional finite element mesh generator with built-in pre- and post-processing facilities," *Int. J. Numer. Methods Eng.*, vol. 79, no. 11, pp. 1309–1331, 2009.
[21] K. İncetan et al., "VR-Caps: A virtual environment for capsule endoscopy," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101990.
[22] K. Clark et al., "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.
[23] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, "3D slicer: A platform for subject-specific image analysis, visualization, and clinical support," in *Intraoperative Imaging and Image-Guided Therapy*. Berlin, Germany: Springer, 2014, pp. 277–289.
[24] "3D slicer image computing platform," 2022. [Online]. Available: https://www.slicer.org/
[25] B. O. Community, *Blender - A 3D Modelling and Rendering Package*. Amsterdam, The Netherlands: Blender Foundation, 2018. [Online]. Available: http://www.blender.org
[26] W. Wiesner et al., "Normal colonic wall thickness at CT and its relation to colonic distension," *J. Comput. Assist. Tomogr.*, vol. 26, no. 1, pp. 102–106, 2002.
[27] M. B. Christensen, K. Oberg, and J. C. Wolchok, "Tensile properties of the rectal and sigmoid colon: A comparative analysis of human and porcine tissue," *Springerplus*, vol. 4, no. 1, pp. 1–10, 2015.
[28] A. J. Loeve, P. Fockens, and P. Breedveld, "Mechanical analysis of insertion problems and pain during colonoscopy: Why highly skill-dependent colonoscopy routines are necessary in the first place... and how they may be avoided," *Can. J. Gastroenterol.*, vol. 27, no. 5, pp. 293–302, 2013.
[29] R. Lambert, "The Paris endoscopic classification of superficial neoplastic lesions: Esophagus, stomach, and colon: November 30 to december 1, 2002," *Gastrointest Endosc*, vol. 58, pp. S3–S43, 2003.
[30] "Lab streaming layer," 2022. [Online]. Available: https://github.com/sccn/labstreaminglayer
[31] A. H. Calderwood and B. C. Jacobson, "Colonoscopy quality: Metrics and implementation," *Gastroenterol. Clin.*, vol. 42, no. 3, pp. 599–618, 2013.
[32] C. M. Zupanc et al., "Assessing colonoscopic inspection skill using a virtual withdrawal simulation: A preliminary validation of performance metrics," *BMC Med. Educ.*, vol. 17, pp. 1–10, 2017.
[33] L. L. Di Stasi et al., "Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: Gaze-based indices," *Appl. Ergonom.*, vol. 65, pp. 168–174, 2017.
[34] A. Rieger et al., "Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress," *Int. Arch. Occup. Environ. Health*, vol. 87, no. 2, pp. 165–174, 2014.
[35] K. I. Jones et al., "Assessing surgeon stress when operating using heart rate variability and the state trait anxiety inventory: Will surgery be the death of us?," *Colorectal Dis.*, vol. 17, no. 4, pp. 335–341, 2015.
[36] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.: Adjunct Pub.*, 2014, pp. 1151–1160.
[37] L. L. Di Stasi et al., "Gaze entropy reflects surgical task load," *Surg. Endoscopy*, vol. 30, no. 11, pp. 5034–5043, 2016.
[38] "Pupil labs documentation," 2022. [Online]. Available: https://docs.pupil-labs.com/core/
[39] V. K. Dik et al., "Measuring gaze patterns during colonoscopy: A useful tool to evaluate colon inspection?," *Eur. J. Gastroenterol. Hepatol.*, vol. 28, no. 12, pp. 1400–1406, 2016.
[40] M. J. Edmondson et al., "Looking towards objective quality evaluation in colonoscopy: Analysis of visual gaze patterns," *J. Gastroenterol. Hepatol.*, vol. 31, no. 3, pp. 604–609, 2016.
[41] S. G. Hart and L. E. Staveland, "Development of nasa-TLX (task load index): Results of empirical and theoretical research," in *Advances in Psychology*. Amsterdam, The Netherlands: Elsevier, 1988, vol. 52, pp. 139–183.
[42] G. Borg, *Borg's Perceived Exertion and Pain Scales*. Champaign, IL, USA: Hum. Kinetics, 1998. [Online]. Available: https://books.google.es/books?id=MfHLKHXXlKAC
[43] P. E. Dupont et al., "A decade retrospective of medical robotics research from 2010 to 2020," *Sci. Robot.*, vol. 6, no. 60, 2021, Art. no. eabi8017.
[44] J. W. Martin et al., "Enabling the future of colonoscopy with intelligent and autonomous magnetic manipulation," *Nature Mach. Intell.*, vol. 2, no. 10, pp. 595–606, 2020.
[45] A. Pore et al., "Colonoscopy navigation using end-to-end deep visuomotor control: A user study," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 9582–9588.
[46] M. Finocchiaro et al., "Human machine interfaces for robot-assisted colonoscopy: A clinical survey," *Proc. 14th Hamlyn Symp. Med. Robot.*, vol. 4, no. 1, p. 53, 2022.
[47] A. Eickhoff et al., "Computer-assisted colonoscopy (the NeoGuide Endoscopy System): Results of the first human clinical trial ("PACE study")," *Official J. Amer. College Gastroenterol.*, vol. 102, no. 2, pp. 261–266, 2007.
[48] M. Verra et al., "Robotic-assisted colonoscopy platform with a magnetically-actuated soft-tethered capsule," *Cancers*, vol. 12, no. 9, 2020, Art. no. 2485.
[49] M. Morino, E. Forcignanò, and A. Arezzo, "Initial clinical experience with a novel flexible endoscopic robot for transanal surgery," *Techn. Coloproctol.*, vol. 26, no. 4, pp. 301–308, 2022.
[50] L. Zorn et al., "A novel telemanipulated robotic assistant for surgical endoscopy: Preclinical application to esd," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 797–808, Apr. 2018.
[51] Y. Huang et al., "A three-limb teleoperated robotic system with foot control for flexible endoscopic surgery," *Ann. Biomed. Eng.*, vol. 49, no. 9, pp. 2282–2296, 2021.
[52] T. Iwasa et al., "A new robotic-assisted flexible endoscope with single-hand control: Endoscopic submucosal dissection in the ex vivo porcine stomach," *Surg. Endoscopy*, vol. 32, no. 7, pp. 3386–3392, 2018.
[53] W. J. Tastle and M. J. Wierman, "An information theoretic measure for the evaluation of ordinal scale data," *Behav. Res. Methods*, vol. 38, no. 3, pp. 487–494, 2006.
[54] K. A. Waschke, "Will you be my first colonoscopy patient? planning simulator training for novice endoscopists," *Gastrointestinal Endoscopy*, vol. 86, no. 5, pp. 890–891, 2017.
[55] A. Sivananthan et al., "A novel gaze-controlled flexible robotized endoscope; preliminary trial and report," *Surg. Endoscopy*, vol. 35, no. 8, pp. 4890–4899, 2021.
[56] L. Santos et al., "Dynamic gesture recognition using a smart glove in hand-assisted laparoscopic surgery," *Technologies*, vol. 6, no. 1, 2018, Art. no. 8.
[57] K. Zinchenko, C.-Y. Wu, and K.-T. Song, "A study on speech recognition control for a surgical robot," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 607–615, Apr. 2017.
[58] Y. X. Mak et al., "Experimental evaluation using head motion and augmented reality to intuitively control a flexible endoscope," in *Proc. IEEE 9th RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, 2022, pp. 1–7.
[59] P. Berthet-Rayne et al., "The i2Snake robotic platform for endoscopic surgery," *Ann. Biomed. Eng.*, vol. 46, pp. 1663–1675, 2018.

Open Access provided by 'Scuola Superiore "S.Anna" di Studi Universitari e di Perfezionamento' within the CRUI CARE Agreement