IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Confidence-based Intent Prediction for Teleoperation in Bimanual Robotic Suturing

Zhaoyang Jacopo Hu, Haozheng Xu, Sion Kim, Yanan Li, Ferdinando Rodriguez y Baena, and Etienne Burdet

***Abstract*—Robotic-assisted procedures offer enhanced precision, but while fully autonomous systems are limited in task knowledge, difficulties in modeling unstructured environments, and generalization abilities, fully manual teleoperated systems also face challenges such as delay, stability, and reduced sensory information. To address these limitations, we propose an interactive control strategy that assists the human operator by predicting their motion plan at both high and low levels. At the high level, a surgeme recognition system is employed through a Transformer-based real-time gesture classification model to dynamically adapt to the operator's actions. At the low level, a Confidence-based Intention Assimilation Controller adjusts robot actions based on inferred user intent and shared control paradigms. The system is built around a robotic suturing task, supported by sensors that capture robot kinematics and task dynamics. Experimental results across users with varying skill levels demonstrate the effectiveness of the proposed approach, yielding statistically significant improvements in task completion time and user satisfaction compared with traditional teleoperation.**

***Index Terms*—Teleoperation, Human-Robot Interaction, Intent Prediction, Gesture Recognition, Robotic Surgery.**

I. Introduction

In traditional teleoperation the human operator fully controls the robot's movements. While offering clear roles and responsibilities, this control under-utilizes both the human and robot capabilities. Robots like the da Vinci Surgical System are equipped with sensors and models offering valuable local information that can complement the human operator's view, such as during visual occlusions or operations with different sensory modalities. By spanning across the spectrum between traditional fully manual teleoperation and full autonomy, shared control combines the benefits of both to enhance teleoperation with the robot's sensory data and control [1]. While demonstrated for suturing assistance [2], [3], these methods overlook the impact on positional uncertainty, environmental unknowns, or instrument errors [4], [5]. For example, robotic surgery cameras are frequently occluded by body tissues or components of the robot [6]. To address these shortcomings, this paper develops an *interaction control* framework for the da Vinci Surgical System, framing shared control as an information exchange between human user and robot agent based on the user's intent [7] with task- and low-level predictions.

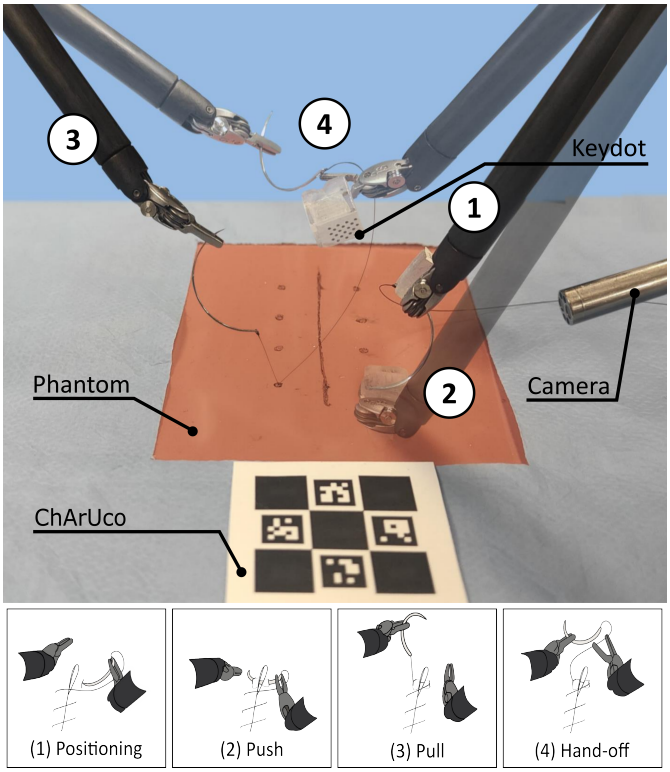

Fig. 1: Bimanual robotic suturing with four key surgemes: (1) Needle tip positioned at the entry point; (2) Needle pushed/inserted reaching the exit point; (3) Needle pulled out; (4) Needle passed to the starting manipulator.

A. Related Work

Research in human-human collaboration has shown that partners share not just their current position but also their motion intention to optimize joint actions [8], [9]. Considering this finding, our approach incorporates shared plans and predictive models to improve coordination in human-robot interaction. At the *task level*, motion intention can be inferred from high-level hand gestures or surgemes [10] that correspond to specific surgical actions. These surgemes can be used to infer the surgeon's task throughout the procedure

Manuscript received: November 9, 2025; Accepted: December 16, 2025. This paper was recommended for publication by Editor Jessica Burgner-Kahrs upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the EPSRC and Intuitive Surgical in the form of an industrial CASE studentship. *(Corresponding authors: Zhaoyang Jacopo Hu, Ferdinando Rodriguez y Baena, Etienne Burdet).*

Zhaoyang Jacopo Hu, Sion Kim and Ferdinando Rodriguez y Baena are with the Department of Mechanical Engineering, Imperial College of Science, Technology and Medicine, London, SW7 2AZ, UK (e-mail: jacopo.hu20@imperial.ac.uk, f.rodriguez@imperial.ac.uk). Haozheng Xu is with the Department of Surgery & Cancer, Imperial College of Science, Technology and Medicine, London, SW7 2AZ, UK. Yanan Li is with the Department of Engineering and Design, University of Sussex, Brighton, BN1 9RH, UK. Etienne Burdet is with the Department of Bioengineering, Imperial College of Science, Technology and Medicine, London, W12 0BZ, UK. (e-mail: e.burdet@imperial.ac.uk).

Digital Object Identifier 10.1109/LRA.2026.3653386

(Fig. 1). For instance, online gesture classification has been used to automate camera motion [11] or switch assistance modes [12]. After recognizing a surgeme, the robot can estimate the operator's motion intention at a lower *control level* to adapt seamlessly during collaborative tasks, akin to the interaction observed in connected humans carrying out a tracking task [7], [13]. The results of [14]–[17] suggest that adaptation and incorporation of motion intention can minimize the operator's effort. In particular, the intention assimilation controller (IAC) of [14] combines human and robot motion plans based on the task demands. However, IAC does not define how to modify the interaction strategy in order to best combine the partners' motion plans, which can be governed by the robot's confidence in its sensor data [7], [18]. Confidence-based approaches have been applied before to improve safety [19] and trajectory tracking [20] by adjusting behavior based on information reliability. In this work, we consider that camera data are more accurate than forward kinematics for position estimation, but prone to occlusions and fluctuations, and combine these sensory modalities correspondingly.

B. Interaction Control Strategy

In this paper we propose to integrate task-level prediction with *confidence-based IAC* (C-IAC), and implement and validate this interaction control strategy on the da Vinci Research Kit (dVRK). The dVRK offers high-definition cameras but limited sensory feedback, making tasks like suturing tedious, time consuming, and attention-intensive [21]. We hypothesize that sensing and motion planning of the human and robot can be combined to enhance suturing performance and comfort. To test this, we developed a kinematic suturing dataset and trained a Transformer model to infer suturing activity in real-time on the dVRK by decomposing it into surgemes. We also designed a mechanical device to automatically orient the needle relative to the manipulator. We then developed an IAC-based controller with confidence level computed from Bayesian inference analyzing kinematic and camera data. Two user studies evaluated the entire system's performance.

The main contributions of this paper are as follows:

- Specialized implementation of shared control with a gesture recognition model for bimanual robotic suturing using the dVRK.
- Integration of a dynamic parameter to compute the confidence component based on Bayesian inference.
- Development of a confidence-based IAC for inferring operators motion plans in robotic surgery teleoperation.
- User studies on the da Vinci robot validating the performance improvement with C-IAC.

II. Surgical Setting

A. Compact Holder for Enhanced Needle Alignment

During suturing, operators often require multiple manipulations of the needle to achieve the optimal orientation [22], [23]. Previous attempts in solving this problem include [24], which created a component whose geometry enabled needle alignment and that could be easily attached to the existing surgical tool without requiring a complete replacement. However, the device increases the size of the end-effector, making needle handling awkward. We developed a *Compact Holder for Enhanced Needle Alignment* (CHENA), improving prior designs [23], [24] by facilitating needle grabbing and holding (Fig. 2). The aligner in [24] was intended for autonomous suturing and required a large rear wall and catching area to guide the needle. The necessity of these parts force the components to extrude over the opposite side of the surgical tool, limiting its maneuverability and maximum jaw open angle. In contrast, CHENA self-aligns and positions the needle without the need for large extrusion in the catching area by using a minimalistic geometry and a magnet to create a driving force for the needle to align even without closing the jaws.

B. dVRK Setup

For the experiments, we use the dVRK equipped with two *Patient Side Manipulators* (PSMs): right PSM1 with DeBakey Forceps and CHENA, and left PSM2 with Large Needle Drivers. The needle used is a $1/2$ circle, 24 mm chord length with triangular cutting point. Robot control is performed in Cartesian space via the dVRK software. Two sigma.7 hand interfaces (Force Dimension; maximal force $20.0N$, maximal torque $400mNm$, maximal grasping force $8.0N$) control the robot using the TCP/IP protocol. As in [25], we renounce to binocular vision and provide a 2D high resolution display to the human user using a RealSense camera, with a 50 ms delay introduced by the teleoperation to observe the effects on performance between controllers [26]. Latency contributes to the complexity and unpredictability of the task and surgeon's performance and is considered one of the biggest hurdles to enable telepresence [27]. [26], [28] showed that a latency below 200 ms is ideal for telesurgery, but even 50 ms can impact performance. This setup aims to emphasize the disparity in sensory information between human and robot, impacting performance outcomes. We maintain pedals for clutching and control of the end-effector orientation.

A surgical phantom is created to perform the suturing task using a 1 mm thick Ecoflex 00-50 layer, replicating the setup in [29]. Similarly, the wound and entry points are manually marked, measured and recorded relative to a ChArUco marker on the phantom. A keydot marker on the CHENA tracks the PSM1 and needle pose. Marker tracking relies on the da Vinci Si endoscope, which provides high-resolution images.

III. Kinematic Gesture Recognition

The Transformer [30] improves upon existing gesture recognition models with its ability to process and analyze kinematic data in parallel. Encoder-only configurations of the Transformer have proven effective in classification tasks [31], suggesting their potential applicability in surgical gesture classification. Therefore, we implement a real-time gesture recognition system that leverages the encoder-only Transformer model. This system classifies gestures by processing a continuous window of kinematic data from the two PSMs and two sigma.7 hand interfaces.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

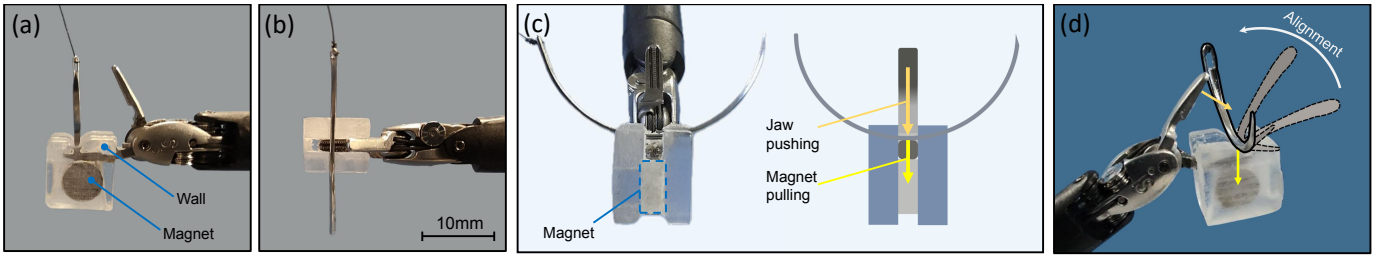

Fig. 2: Design of the Compact Holder for Enhanced Needle Alignment (CHENA). (a-b) The structure comprises of walls that create a concavity where the needle is attracted and maintained fixed by a magnet. (c-d) The needle self-aligns using to two independent acting forces: the jaw pushing and the magnet pulling the needle towards the cavity.

A. Suturing Dataset

A major obstacle in robotic surgery gesture recognition is the scarcity of annotated datasets [10], with one of the few being JIGSAWS [29], which is widely used as a benchmark dataset. To accommodate our specific settings, which are slightly different from JIGSAWS, we constructed our own kinematic dataset made of bimanual suturing motions, referred to as *Suturing Task in Imperial's Tracking Collection for Hamlyn's Evaluation Set* (STITCHES). It includes kinematic and video data recorded at 20 Hz from a user with over 900 hours of dVRK experience. The dataset captures movements from the two PSMs and two sigma.7s, totaling 19 kinematic features per device: end effector position (3), rotation matrix (9), linear velocity (3), angular velocity (3), and gripper angle (1). Gesture labels, manually annotated using video references, add a final feature, resulting in 77 features. We collected 10 recordings, each having four sutures, or throws, defined as the combination of surgemes completing one loop.

B. Gesture Recognition Algorithm

The kinematic window captures linear velocity (3), rotational velocity (3), and gripper angle (1) from each of the four devices at 20 Hz, resulting in 28 features per time step. This window spans 60 time steps (3 seconds) and is fed into a classification network composed of two Transformer encoder blocks, each with attention heads, followed by two fully connected layers of 64 units each and a softmax activation. To improve stability in real-time, we use an exponential moving average (EMA) window of size 10 over the output probabilities and a 0.8 probability threshold on averaged probabilities. The EMA filters noise and transient errors, ensuring stable gesture predictions while adapting quickly to new gestures by weighting recent data more heavily. The threshold ensures that only high-probability gestures are recognized. We validated the model using 5-fold cross-validation, dividing the 10-recordings dataset into five folds, each containing two recordings.

C. Suturing Gesture Classification

Suturing is a fundamental laparoscopic surgery task [32]. [29] identified 15 different gestures G in robotic surgery, 10 of which apply to suturing. However, [10], [24] suggested that suturing could be represented with just 4 more general fundamental gestures. We demonstrate the viability of this approach by using the JIGSAWS dataset [29], which has been widely recognized as a benchmark for the dVRK, and proposing two strategies to group the original gesture labels G from JIGSAWS into 5 classes (Table I). Both include the fundamental gestures (1) *Positioning*, (2) *Push*, (3) *Pull*, and (4) *Hand-off*, as illustrated in Fig. 1, and an additional *Other* class to group the rest of the gestures. In strategy 1, only the original four gestures G are labeled with the corresponding surgeme while the rest are labeled *Other*. Strategy 2 groups gestures similar in motion to the four original gestures and the rest is labeled *Other*. If the Transformer-based model can classify both strategies with similar accuracy, then this enables us to verify whether four surgemes can represent suturing adequately. We can then implement strategy 2, which more broadly represents each gesture, in the shared control system's Transformer model.

TABLE I. Proposed Surgeme Labels (a-d) using JIGSAWS

Surgeme	Strategy 1	Strategy 2
(1)	G2	G2, G5
(2)	G3	G3
(3)	G6	G6, G10
(4)	G4	G4, G8
Other	G1, G5, G8, G9, G10, G11	G1, G9, G11

IV. Confidence-based IAC

A. Intention Assimilation Controller

IAC enables the robot to extract the human's motion plan from the interaction force and then combine it with its own plan [14]. Given the operator's applied force u_h, system position x, and velocity $\dot{x}$, IAC predicts the human's motion plan and ensures stable human-robot interaction [14]:

$$u_h = -L_1(x - \tau) - L_2\dot{x} \tag{1}$$

where L_1, L_2 are stiffness and viscosity matrices [14]. The robot target τ is extracted from Equation 1 during the movement using a Kalman filter to compute the predicted state estimate of the human target $\hat{\tau}_h$. The robot is then controlled via:

$$\tau = \lambda\tau_r + (1 - \lambda)\hat{\tau}_h \tag{2}$$

where the robot target is the convex combination of the estimated human target $\hat{\tau}_h$ and its own target τ_r, selected based to its task and sensing information. The confidence factor λ balances the robot's reliance on information from its target

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

τ_r versus the human target τ_h, influencing assistance behavior during interaction.

When $\lambda = 0$, the robot considers its own information unreliable and fully follows the predicted human target position $\hat{\tau}_h$. $0 < \lambda < 1$ sets a cooperation between human and robot control inputs, considering both of their targets. This allows to benefit from the speed of reaching the predicted human target $\hat{\tau}_h$ instead of the current position x_h while dynamically adjusting confidence based on the sensory information detected by the robot. At $\lambda = 1$, the robot disregards the human interaction, adopting a coactive behavior to achieve its own target, assuming sufficient information to complete tasks without teleoperation delays. To support the user, $0 \leq \lambda \leq 1$ must be ensured. The implementation of C-IAC compared to traditional teleoperation is shown in Fig. 3. Dynamic and continuous adjustment of λ are key for smooth and situation-specific human-robot interaction. This method allows to optimize the control authority in real-time based to the system's confidence on either human or robot.

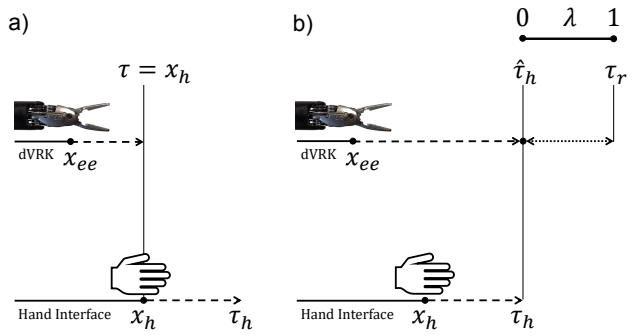

Fig. 3: Comparison between a) traditional and b) C-IAC teleoperation. While traditional teleoperation has inherent delay to reach the current position of the user's hand, C-IAC directly moves towards the predicted position. λ is used in the shared control scheme to adjust the assistance. With high confidence ($\lambda \approx 1$), delay is minimized as the robot to follow its own target.

B. Confidence with Bayesian Inference

We use vision and proprioception as sensory inputs for the robot. Visual information comes from tracking markers placed on the robot's manipulator and phantom (keydot kd and ChArUco ch, respectively). Proprioception is estimated using the robot joint angles and is consistently available, unlike vision that can be obstructed, leading to a tracking failure. These two sensory modalities are integrated to ensure continuous state estimation during the task, especially in the presence of visual occlusions or sensor noise. Visual tracking (via kd and ch markers) provides the primary estimate of the manipulator's position. When visual data is available, it is used both for real-time tracking and to recalibrate the kinematic estimates obtained from the robot's joint encoders, mitigating the cumulative errors introduced by the cable-driven architecture of the dVRK. When vision is unavailable, position estimation integrates the velocity from the robot proprioception. We define the robot's *confidence* in its sensing accuracy or reliability by using a Beta distribution $B(\alpha, \beta)$ and analyzing the robot's performance history, as was proposed in [33]. As shown in [33], the Beta distribution effectively infers the human's trust in the robot's performance by analyzing the robot's performance history. Additionally, the distribution is bounded within the interval $[0, 1]$ which is consistent with the confidence boundaries to achieve the interaction behavior needed. Furthermore, the distribution dynamics enables the property of influencing the confidence at the current moment i by previous time steps $i - 1$ [34]. In our case we define:

$$\lambda = E[B(\alpha_i, \beta_i)] = \frac{\alpha_i}{\alpha_i + \beta_i}, \tag{3}$$

where the shape parameters α_i, β_i are defined iteratively through the performance p_i, which is 1 if the robot has confidence in its positional information as kd or ch are visible, and 0 otherwise:

$$\alpha_i = \begin{cases} \alpha_{i-1} + w_1 & p_i = 1 \\ \alpha_{i-1} & p_i = 0 \end{cases}$$
$$\beta_i = \begin{cases} \beta_{i-1} + w_0 & p_i = 0 \\ \beta_{i-1} & p_i = 1 \end{cases} \tag{4}$$

where the weights w_0, w_1 were tuned empirically. As demonstrated in [33], Equations 3 and 4 estimate the confidence based on the performance history, as a successful marker detection causes an increase in α_i by w_1, while occlusions or missed detections cause an increase in β_i by w_0.

C. Shared Control Paradigms

We propose the following shared control paradigms, generalized into the four fundamental surgemes and a general one. The constraints and targets in these paradigms have been developed in consultation with a neurosurgeon from Imperial College London. The following terminology represents quantities measured relative to the task frame in Fig. 4, defined with respect to the manipulator tracked by the keydot:

- $\tau = [x, y, z] \in \mathbb{R}^3$: target position computed by C-IAC.
- $\lambda = [x, y, z] \in \mathbb{R}^3$: confidence applied on end-effector.

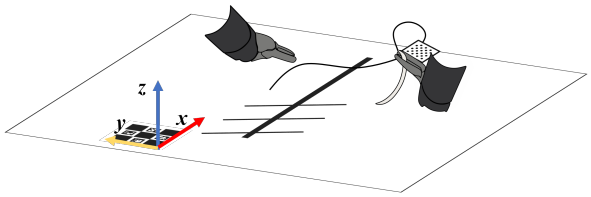

Fig. 4: Task frame position on the tissue surface using a ChArUco with x-axis parallel to the wound, y-axis perpendicular to it, and z-axis perpendicular to the surface. During suturing, the robot refers to the task frame Cartesian coordinates to determine wound orientation and insertion point.

1) Positioning: The user aligns the needle's plane perpendicular to the wound, ensuring the tip touches the insertion point at a fixed height. Since the entry point can vary along the tissue near the wound, the manipulator moves freely along a plane parallel to the tissue and at fixed target height.

$$\tau = [\tau_h^x, \tau_h^y, \tau^z], \quad \lambda = [0, 0, \lambda^z] \tag{5}$$

As the needle plane needs to be perpendicular to the wound for an effective needle insertion, we allow the human to press a pedal to trigger autonomous orientation of the end-effector.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

2) Push: The needle plane must remain perpendicular to the tissue surface and the wound direction. To prevent laceration, the manipulator movements along the wound axis are restricted, while having free motion in other directions. This enables to perform a correct circular motion for the needle to emerge from the opposite side of the wound.

$$\tau = [\tau^{x_j}, \tau_h^y, \tau_h^z], \quad \lambda = [\lambda^x, 0, 0] \tag{6}$$

where x_j is the jth entry point position in x during suturing. As in *Positioning*, the end-effector orientation can be controlled using a pedal to autonomously achieve perpendicularity.

3) Pull: When one hand is stationary and the other moves with higher velocity, stability in the stationary hand improves task performance [12]. In suturing, this corresponds to *Pull*, where the right hand is kept steady above the tissue, allowing users to focus on pulling with the left. To facilitate the next throw, the robot target is set to the next entry point position.

$$\tau = [\tau^{x_{j+1}}, \tau^y, \tau^z], \quad \lambda = [\lambda^x, \lambda^y, \lambda^z] \tag{7}$$

4) Hand-off: The hand-off of the suturing needle usually requires coordination between the two manipulators to allow correct grasping. As the CHENA facilitates this, we simply ensure that the *Hand-off* occurs in proximity of the next entry point, facilitating the next *Positioning* while allowing the PSM1 manipulator to freely approach the PSM2.

$$\tau = [\tau^{x_{j+1}}, \tau_h^y, \tau^z], \quad \lambda = [\lambda^x, 0, \lambda^z] \tag{8}$$

5) Other: When the user's kinematic inputs to the PSMs do not match any of the first four surgemes, no constraint is applied to the manipulator, which corresponds to $\lambda = 0$ for all Cartesian coordinates.

$$\tau = [\tau_h^x, \tau_h^y, \tau_h^z], \quad \lambda = [0, 0, 0] \tag{9}$$

Notice that when a new gesture is detected, the confidence is reset to zero, leaving the human in full control and preventing short and transient misclassifications from causing disruptions. Furthermore, it would be possible to set hard constraints to impose forbidden movements to the user for each control paradigm, but interaction forces with the human may disrupt the control. Instead, the C-IAC allows for the creation of softer and more dynamic constraints by predicting the intent and considering the system's confidence throughout the task. The complete system framework is shown in Fig. 5.

V. Experiments

The experiments were approved by the College Research Ethics Committee of Imperial College London (21IC7042). Participants were briefed on the experiments' purpose and protocol, and signed a consent form before starting. The first set of experiments was carried out to validate our Transformer-based model, the STITCHES dataset, and needle pose estimation accuracy. Subsequently, we conducted two user studies to compare traditional teleoperation against C-IAC in two tasks: (i) target reaching, to systematically analyze the controller performance, and (ii) four surgical throws, to represent a key surgical function. For consistency, the PSM1 was equipped with a CHENA in all experiments. This setup implies that while C-IAC controls both PSMs, only the one with the keydot uses confidence-based properties, since it is tracked both kinematically and visually.

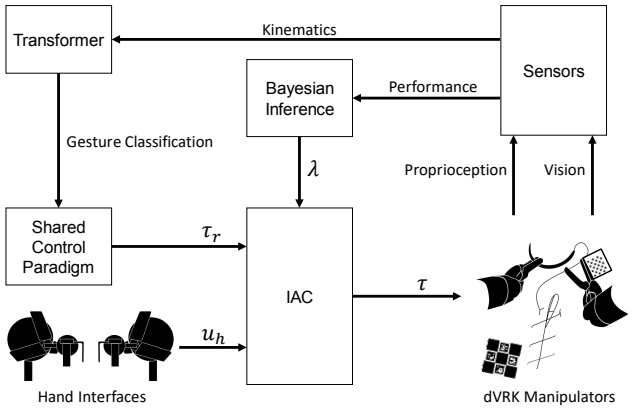

Fig. 5: Bimanual suturing control scheme using C-IAC and Transformer for gesture classification. Sensory information from markers and robot kinematics undergoes Bayesian inference to compute the confidence parameter λ for the IAC. The Transformer utilizes the kinematics to classify the user's gesture and define the target τ_r. In IAC, the user input u_h is used to predict the human's target τ_h, which, together with τ_r, determines the robot manipulator's target.

A. Target Reaching User Study

To perform a detailed comparison between C-IAC and traditional teleoperation without the complexities of a complete suturing task, we design a target reaching experiment consisting of *Positioning* and *Push*. We recruited 8 participants (1 female and 7 males, age 23-30, all right-handed) without dVRK experience. After 10 minutes of familiarization, the subjects controlled the PSM1 to reach four entry points on the phantom, starting from a fixed pose. In each trial, the experimenter triggered a button to indicate successful completion of reaching and needle insertion. The starting pose remained consistent, with entry points positioned on a straight line at $\{15, 30, 45, 60\}$mm from the starting pose, thus challenging the user's depth perception.

We compared traditional teleoperation versus C-IAC, each tested with four subjects. To ensure consistency across participants, λ in C-IAC was gradually increased from 0 to 0.8 using a linear function instead of Bayesian Inference. This ensures uniform assistance across trials. We then use Equation 6 as the control paradigm during all C-IAC trials. This and $\lambda = 0.8$ were selected pragmatically to ensure the human retains a degree of control throughout the experiment.

B. Suturing Task User Study

This study compared the proposed C-IAC against traditional teleoperation in a full suturing task with four entry points that require four suturing throws. The experiment aimed to assess the performance of seven novice users (age 21-29, no medical background), as well as one intermediate (age 23, fifth-year medical student) and one expert (age 41, neurosurgeon with practice in the medical field). None had experience with surgical robots. Before the experiments, each subject trained on the system for 8 sessions of 20 minutes each.

A clinical criterion for successful suturing is maintaining the needle plane perpendicularity to the wound and skin [22], [23]. Since the CHENA keeps the needle perpendicular to the manipulator, we tracked its orientation via the keydot marker to measure the needle perpendicularity during the

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

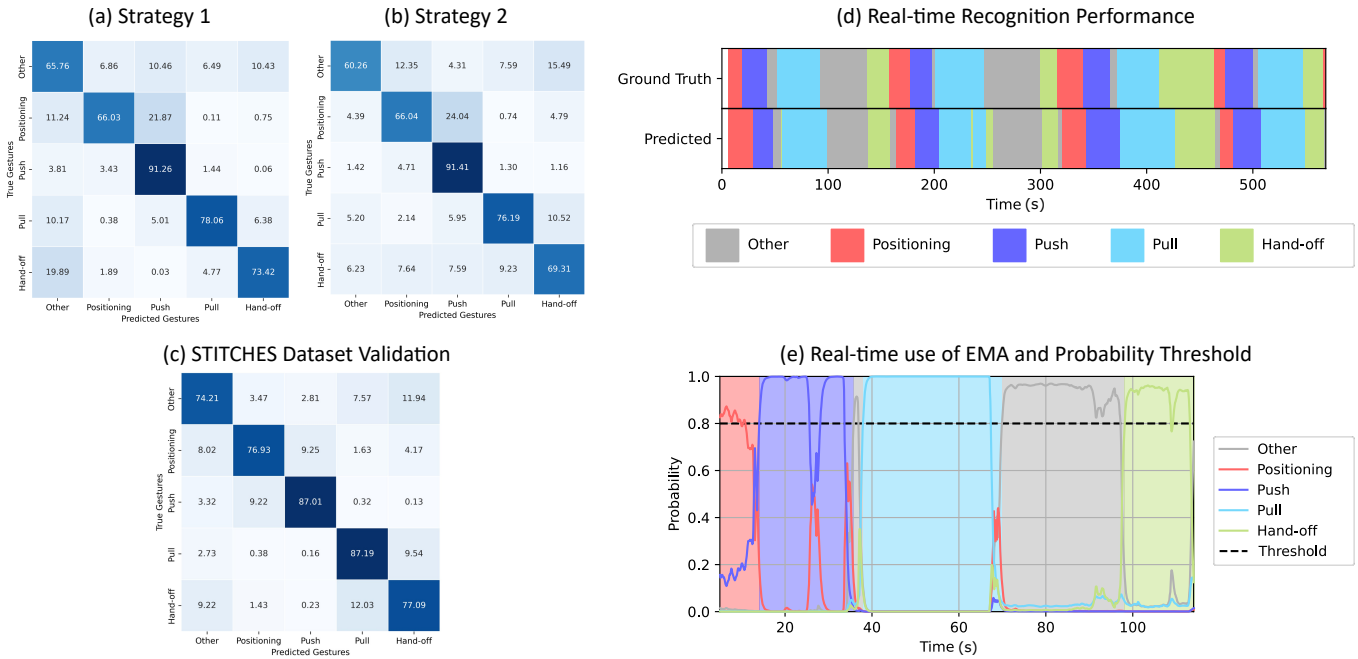

Fig. 6: Transformer model classification on JIGSAWS and STITCHES datasets. (a,b) Transformer gesture classification using labeling strategies 1 and 2. (c) 5-fold cross validation average accuracy of STITCHES. The overall accuracy is 81.19% using the proposed labeling. (d) Ground truth and predicted gesture comparison with STITCHES. (e) Real-time suturing throw with EMA and probability threshold.

task. Additionally, we measured the Transformer's average accuracy, task completion time, and participants' perceived task load through a NASA-TLX questionnaire.

VI. Results

A. Transformer for Two Classification Strategies

The Transformer model is applied to two strategies in JIGSAWS (Section III-B) to demonstrate the feasibility of reducing the suturing task into four fundamental motions and a general one. Fig. 6 (a-b) compares their confusion matrices. The model achieves an overall accuracy of 77.67% and 76.72% for strategy 1 and 2, respectively. Both show consistent trends, with some misclassifications in *Positioning* and *Push*, which are both predominantly right-hand gestures. *Other* has the lowest accuracy in both strategies, while *Push* was the highest. The comparable results confirm that strategy 2 effectively generalizes the task, making it the basis for subsequent experiments.

B. STITCHES Dataset Validation

Training and validation of our Transformer model were conducted on STITCHES, which was then used in the user studies. Using post-processing techniques, i.e. EMA and probability threshold, the model achieved an accuracy of 81.19%. Fig. 6 (c-d) shows the confusion matrix obtained from validation and a comparison between the actual gestures and the predicted in a sample. Additionally, Fig. 6 (e) provides a graphical representation of the post-processing applied in real-time during a suturing throw.

C. Needle Pose Estimation Error

With the CHENA, the needle plane can be approximately considered in a fixed orientation perpendicular to the gripper, similarly to [23], [24]. To compute the needle pose error, we compared the CHENA's pose estimation from visual tracking of the keydot with the robot's kinematic data. Table II presents the mean and standard deviation of position and orientation errors across 80 random robot positions, showing errors in the order of 1 mm and 2°.

TABLE II. Needle Pose Estimation Error Results

	Position (mm)			Orientation (○)		
DoF	**x**	**y**	**z**	**Roll**	**Pitch**	**Yaw**
Mean	0.652	0.933	0.843	2.510	2.299	1.238
Std.	0.997	0.873	0.676	2.727	1.828	1.375

D. Target Reaching User Study

During this study, we measured the time required for the reaching movement to perform accurate insertions. As the data was not normally distributed, a Wilcoxon rank-sum test was conducted, revealing a difference between the controllers ($p = 0.004$). Table III shows that for all entry points, C-IAC reduced the average time required for target reaching and precise insertion relative to traditional teleoperation, with a total average and standard deviation of $(\sigma, \mu) = (36.1, 15.4)$ s, and $(61.9, 39.6)$ s for traditional control.

TABLE III. Duration of Target Reaching (in seconds)

Controller	Entry point 1	Entry point 2	Entry point 3	Entry point 4	Total
Traditional	61 ± 6.8	61 ± 33.3	41 ± 10.6	84.8 ± 72.8	61.9 ± 39.6
C-IAC	25 ± 5.4	45 ± 24.4	39.5 ± 13.5	35 ± 9.8	36.1 ± 15.4

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE IV. User Study Suturing Task Performance based on Participant Suturing Experience

	Novices		Intermediate		Expert	
	Traditional	**C-IAC**	**Traditional**	**C-IAC**	**Traditional**	**C-IAC**
Avg. ⊥ Error in Push (○)	28.83±10.11	7.27±5.26	21.98	4.25	34.47	2.19
Real-time Classification Accuracy (%)	—	65.50±15.25	—	67.37	—	67.17
Avg. Positioning Time (s)	26.31±18.99	19.94±9.23	24.82	22.87	36.95	32.93
Avg. Push Time (s)	43.07±18.97	38.05±21.95	42.00	47.83	47.33	40.07
Avg. Pull Time (s)	76.91±40.68	48.26±17.35	65.65	50.97	87.10	66.12
Avg. Hand-off Time (s)	54.09±19.21	35.29±14.17	36.86	27.68	74.84	49.66
Avg. Throw Time (s)	222.70±73.80	155.94±40.41	214.66	181.15	290.27	218.46
Total Suturing Time (s)	892.18±204.53	652.55±158.71	859.01	724.95	1161.51	873.87

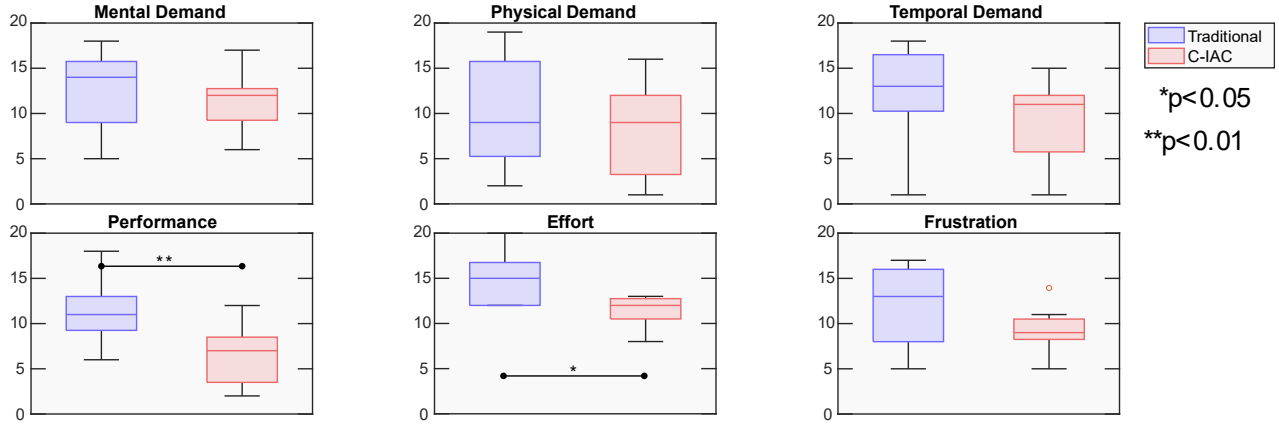

Fig. 7: Subjective assessment of C-IAC vs traditional teleoperation with NASA-TLX with the seven novice participants after four throws suturing user study. Lower values indicate better performance.

E. Suturing Task User Study

During the study with users of different suturing skill levels, participants performed four consecutive throws. The results in Table IV highlight C-IAC's smaller perpendicular error and reduced task duration across all surgemes relative to traditional teleoperation. As the seven novices' data is not normally distributed, Wilcoxon Signed Rank test was used, which revealed a significant reduction in perpendicular error during *Push* with C-IAC ($p < 0.001$). Furthermore, C-IAC yields faster performance than traditional teleoperation ($p < 0.001$) in singular throw time, in particular during *Pull* ($p < 0.001$) and *Hand-off* ($p < 0.001$). Interestingly, the intermediate and expert surgeon exhibit similar trends. The NASA-TLX responses from the novices (Fig. 7) indicate a lower perceived workload with C-IAC. After confirming data normality, we performed t-test to compare the scores, summarized in Fig. 7. The analysis found that users felt they performed better with C-IAC ($p < 0.01$) and required less effort ($p < 0.05$). Similar trends were observed for temporal demand ($p = 0.079$) and frustration ($p = 0.083$).

VII. Discussion

The classification accuracy results confirmed the feasibility of using four fundamental surgemes and a general one. While the *Push* surgeme had the highest overall accuracy, *Other* had the lowest, likely due to the aggregation of diverse gestures. These results are consistent with the literature [35], [36], with the lower performance attributed to the circular dependence of the C-IAC's Transformer model, which was trained on traditional teleoperation data. To train the model, we created STITCHES and implemented strategy 2, which yielded a similar accuracy trend as with JIGSAWS. In creating STITCHES, we observed bimanual hand movements to pull enough thread for subsequent throws for the first 1-2 throws, as seen in Fig. 6 (d-e), where *Other* follows *Pull*. While this movement is not fundamental to a single stitch, it is essential for consecutive throws. Our online gesture recognition more accurately reflects the real task than [11], where this property was missed due to using only two throws or short thread. Additionally, needle pose estimation is consistent with the literature [23], [24].

In the first user study, C-IAC proved to enable faster operation than traditional teleoperation, leading to a smaller standard deviation. We conducted a second user study involving four consecutive throws with participants of different skill levels, without prior experience in robotic suturing. C-IAC enabled faster task completion than traditional teleoperation, both in average throw and total task time. For needle insertion, C-IAC significantly reduced perpendicularity error while maintaining similar *Push* times. Additionally, it improved performance in *Pull*, *Hand-off*, and throw time, demonstrating seamless integration of the autonomous component into the workflow.

In our experiments, the model showed lower accuracy than during dataset validation, as STITCHES was recorded from a different user. Additionally, intermediate and expert, having already developed their own suturing styles, were more affected than novices. The participants also required more suturing time than the data in STITCHES, impacting classification accuracy. Notably, the expert took the longest, perhaps due to bias towards non-robotic laparoscopic suturing [37].

Fig. 7 reports NASA-TLX scores given by the novices, indicating that C-IAC imposes less workload than traditional

control, with significant improvements in performance and effort. The higher suturing time compared to previous works [24], [29] reflects the challenging surgical setting designed to encourage human-robot interaction. Latency, 2D display, different hand interfaces, and inexperience with the surgical setup are the main contributors to this increased time.

C-IAC relies on a trained gesture recognition model, which itself requires a dataset. To avoid circular dependence, C-IAC's Transformer model was trained on traditional teleoperation movements. This solution demonstrates the C-IAC ability to handle lower accuracy models by adjusting the confidence λ parameter. In our tests, setting λ below 1 proved preferable due to human and robot sensor limitations. This suggests that enhanced sensory input, such as additional cameras, force sensors, or advanced path planning, could enable higher confidence (i.e. $\lambda = 1$), potentially improving the performance.

VIII. Conclusion

At the extremes of the control spectrum, traditional teleoperation faces delay and sensory limitations, while full automation is limited by sensory inaccuracies and lack of broad task knowledge. These challenges were addressed here by developing a unified human-robot system with high and low level intent detection to improve suturing performance while relieving cognitive workload. To our knowledge, this represents the first implementation of real-time intent detection for the interaction control in bimanual robotic suturing, using confidence across multiple sensory modalities. The proposed teleoperation system allows the operator to perform more efficient gestures, with reduced cognitive strain, which can be easily extended to other fields requiring remote teleoperation. In the future, a larger user sample size should be acquired to analyze the C-IAC and search for optimal probability thresholds and λ limits.

Acknowledgments

This work was conducted at Imperial College London's Hamlyn Centre for Robotic Surgery. Special thanks to Kiran Bhattacharyya and Anton Deguet for their support.

References

[1] N. Jarrasse *et al.*, "Slaves no longer: review on role assignment for human–robot joint motor action," *Adaptive Behavior*, pp. 70–82, 2014.

[2] M. Selvaggio *et al.*, "Haptic-guided shared control for needle grasping optimization in minimally invasive robotic surgery," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[3] G. A. Fontanelli, G.-Z. Yang, and B. Siciliano, "A comparison of assistive methods for suturing in mirs," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[4] B. Li, H. Lin, F. Zhong, and Y. Liu, "Real-time geometric joint uncertainty tracking for surgical automation on the dvrk system," in *IEEE International Conference on Robotics and Biomimetics*, 2024.

[5] J. A. Barragan, H. Ishida, A. Munawar, and P. Kazanzides, "Improving the realism of robotic surgery simulation through injection of learning-based estimated errors," in *IEEE International Symposium on Medical Robotics*, 2024, pp. 1–7.

[6] S. Tukra, H. J. Marcus, and S. Giannarou, "See-through vision with unsupervised scene occlusion reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 3779–3790, 2021.

[7] Y. Li *et al.*, "A review on interaction control for contact robots through intent detection," *Progress in Biomedical Engineering*, 2022.

[8] G. Ganesh *et al.*, "Two is better than one: Physical interactions improve motor performance in humans," *Scientific Reports*, pp. 1–7, 2014.

[9] A. Takagi, M. Hirashima, D. Nozaki, and E. Burdet, "Individuals physically interacting in a group rapidly coordinate their movement by estimating the collective goal," *Elife*, vol. 8, 2019.

[10] B. van Amsterdam *et al.*, "Gesture recognition in robotic surgery: a review," *IEEE Transactions on Biomedical Engineering*, 2021.

[11] N. Pasini *et al.*, "Grace: Online gesture recognition for autonomous camera-motion enhancement in robot-assisted surgery," *IEEE Robotics and Automation Letters*, 2023.

[12] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared control–based bimanual robot manipulation," *Science Robotics*, 2019.

[13] A. Takagi *et al.*, "Physically interacting individuals estimate the partner's goal to enhance their movements," *Nature Human Behaviour*, 2017.

[14] ——, "Flexible assimilation of human's target for versatile human-robot physical interaction," *IEEE Transactions on Haptics*, 2020.

[15] L. Chen, Z. J. Hu, Y. Huang, E. Burdet, and F. R. y Baena, "Human robot shared control in surgery: A performance assessment," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[16] H. Zheng, Z. J. Hu *et al.*, "A user-centered shared control scheme with learning from demonstration for robotic surgery," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[17] Z. J. Hu, Z. Wang *et al.*, "Towards human-robot collaborative surgery: Trajectory and strategy learning in bimanual peg transfer," *IEEE Robotics and Automation Letters*, 2023.

[18] M. Kam, H. Saeidi *et al.*, "A confidence-based supervised-autonomous control strategy for robotic vaginal cuff closure," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.

[19] S. Chernova *et al.*, "Interactive policy learning through confidence-based autonomy," *Journal of Artificial Intelligence Research*, 2009.

[20] H. Saeidi *et al.*, "A confidence-based shared control strategy for the smart tissue autonomous robot (star)," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[21] B. T. Ostrander, D. Massillon *et al.*, "The current state of autonomous suturing: a systematic review," *Surgical Endoscopy*, pp. 1–15, 2024.

[22] A. Guni *et al.*, "Development of a technical checklist for the assessment of suturing in robotic surgery," *Surgical Endoscopy*, 2018.

[23] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous suturing framework and quantification using a cable-driven surgical robot," *IEEE Transactions on Robotics*, 2020.

[24] S. Sen *et al.*, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[25] A. Saracino *et al.*, "Haptic feedback in the da vinci research kit (dvrk): A user study based on grasping, palpation, and incision tasks," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2019.

[26] F. Richter, R. K. Orosco, and M. C. Yip, "Motion scaling solutions for improved performance in high delay surgical teleoperation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[27] G. T. Sung and I. S. Gill, "Robotic laparoscopic surgery: a comparison of the da vinci and zeus systems," *Urology*, 2001.

[28] S. Xu *et al.*, "Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dv-trainer® simulator," *Surgical Endoscopy*, 2014.

[29] Y. Gao *et al.*, "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling," in *MICCAI workshop: M2cai*, 2014.

[30] A. Vaswani, N. Shazeer *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[31] T. Lin *et al.*, "A survey of transformers," *AI open*, 2022.

[32] E. M. Ritter and D. J. Scott, "Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery," *Surgical innovation*, vol. 14, no. 2, pp. 107–112, 2007.

[33] Y. Guo and X. J. Yang, "Modeling and predicting trust dynamics in human–robot teaming: A bayesian inference approach," *International Journal of Social Robotics*, vol. 13, no. 8, pp. 1899–1909, 2021.

[34] J. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, 1992.

[35] B. Van Amsterdam *et al.*, "Gesture recognition in robotic surgery with multimodal attention," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1677–1687, 2022.

[36] Y. Zheng and A. Majewicz-Fey, "Transformer-based automated skill assessment and interpretation in robot-assisted surgery," in *International Symposium on Medical Robotics (ISMR)*. IEEE, 2024.

[37] S. Kaul, N. L. Shah, and M. Menon, "Learning curve using robotic surgery," *Current Urology Reports*, vol. 7, no. 2, pp. 125–129, 2006.