2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

A Model-Based Framework for Assessing Operator Performance in Navigational Bronchoscopy

Zhaoxing Deng[1,2], David Hanley[1], Francis Xiatian Zhang[2], Kevin Dhaliwal[2], Mohsen Khadem[1,2]

***Abstract*— Bronchoscopy is a critical procedure for diagnosing and treating pulmonary diseases, but its safe and effective execution demands substantial operator training. Insufficient experience is associated with higher complication rates, including bleeding, pneumothorax, and bronchospasm. Existing assessment tools provide structured evaluations, yet they remain heavily reliant on subjective expert judgment and limited sensory feedback. To address this limitation, we propose a model-based framework for objective performance evaluation in navigational bronchoscopy. Our approach leverages pose data from electromagnetic (EM) trackers, routinely used in clinical navigation, and embeds nonholonomic kinematic constraints that characterize expert-like trajectories. Using the model and a Model Predictive Path Integral (MPPI) control, we generate optimal reference trajectories and define error metrics that quantify deviations between operator-executed and model-predicted motions. We hypothesize that these deviations provide robust discriminative features for distinguishing between expert and novice performance. Experiments on a phantom lung dataset comprising 11 operators and 98 procedures demonstrate that the proposed metrics significantly separate skill levels, enabling the construction of an effective classifier for operator proficiency. This framework offers an interpretable, data-driven alternative to supervisor-dependent assessments and represents a step toward scalable, objective skill evaluation and transfer in bronchoscopy training and robotic platforms.**

I. Introduction

Bronchoscopy is a standard procedure for diagnosing and treating pulmonary diseases, carried out with a flexible instrument that integrates a camera, illumination, and a working channel for therapeutic tools [1]. While generally safe, the procedure carries risks such as bronchospasm, bleeding, pneumothorax, and hypoxemia, with complication rates rising when performed by less experienced operators [2], [3]. To minimize these risks, physicians undergo extensive training; for example, the American College of Graduate Medical Education (ACGME) recommends completion of at least 100 supervised procedures before a trainee is deemed competent [4]. Clinical success, particularly in diagnostic tasks such as sampling, depends strongly on operator proficiency [5], motivating the need for reliable methods to monitor and accelerate training.

Simulation-based training has become an important strategy for developing bronchoscopy skills and reducing adverse outcomes [6]. However, most existing assessment frameworks remain subjective, relying heavily on expert supervision [7]. Widely used tools such as the Bronchoscopy Skills and Tasks Assessment Tool (BSTAT) and the Ontario Bronchoscopy Assessment Tool (OBAT) provide standardized scoring schemes [8], [9], yet many criteria—such as posture, instrument handling, and trauma avoidance—are still based on personal judgment rather than objective measurement. With sensing technologies increasingly integrated into clinical practice, there is a growing opportunity to transition toward quantitative, data-driven approaches.

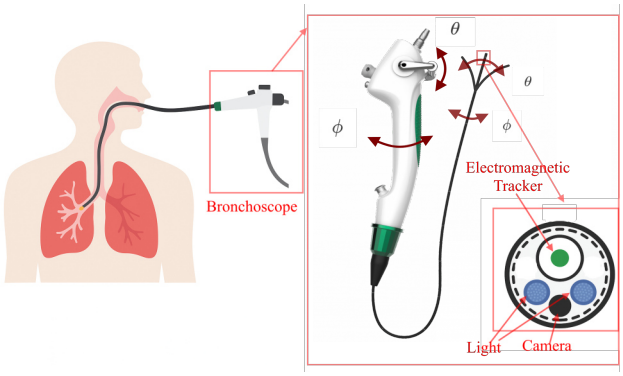

Fig. 1. A bronchoscope is a medical instrument used to examine the inside of the airways and lungs. The figure shows a commercial bronchoscope (Ambu A/S, Denmark) equipped with a 6-degree-of-freedom electromagnetic sensor (NDI, Canada) used for data collection and evaluation experiments. A bronchoscope has a lever that allows the operator to articulate and bend the tip (θ), enabling precise navigation through the airways. By twisting the bronchoscope (ϕ), the operator can change the bending plane, allowing the device to navigate the complex 3D structure of the bronchial tree.

The surgical training community has made substantial progress in data-driven skill evaluation for minimally invasive surgery. A landmark example is the JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS), which recorded kinematic and video data from robotic-assisted laparoscopic training tasks to enable automated gesture recognition and operator skill classification [10], [11]. This dataset has since driven a wide range of research directions, from early approaches using handcrafted movement features with classical classifiers [12] to more recent convolutional and temporal deep learning architectures that model spatiotemporal patterns in surgical motion [13], [14]. While highly influential, JIGSAWS and related efforts focus on benchtop robotic laparoscopy tasks and do not capture the unique challenges of flexible endoscopy—specifically, navigating nonholonomic, highly compliant instruments through branching and de-

[1]School of Informatics, University of Edinburgh, UK.
[2]Baillie Gifford Pandemic Science Hub, Institute for Regeneration and Repair, University of Edinburgh, UK.
This work was supported by the UKRI Medical Research Council under Grants MR/T023252/1 and MR/Y011694/1, by the Baillie Gifford Pandemic Science Hub, and by the UKRI Engineering and Physical Sciences Research Council under U-Care Programme Grant (EP/T020903/1).

formable anatomical structures [15].

By contrast, work specific to bronchoscopy remains limited [16]. Some prior studies have attempted to assess operator proficiency indirectly by estimating airway trauma or tissue interaction forces [17], [18], but these methods are constrained by practical challenges. Real-time recovery of bronchoscope shape or lung anatomy is difficult [19], [20], and monocular pose estimation in endoluminal scenes remains challenging due to limited texture, specularities, and repetitive anatomy [21]. Additionally, anatomical motion from respiration can displace bronchial structures by up to 25 mm per cycle [22], complicating comparisons with preoperative images. More recently, Deng *et al.* [23] introduced a trajectory-based framework for automated skill evaluation in bronchoscopy, showing the feasibility of objective assessment in the endoluminal setting. However, their analysis was restricted to global trajectory features and lacked the granularity to capture local deviations [21] around bifurcations or complex airway segments. At the same time, the field is experiencing rapid growth in robotic bronchoscopy systems for both diagnosis and therapy. Yet, systematic approaches for skill assessment and, critically, for skill transfer in bronchoscopy have not been developed to the same extent as in robotic laparoscopy and keyhole surgery. This discrepancy underscores the need for methods that leverage sensing technologies and geometric priors to provide objective, explainable evaluations of operator performance in flexible endoscopy.

In this paper, we introduce a model-based framework for operator performance evaluation in bronchoscopy. Our method uses pose data from electromagnetic (EM) trackers—commonly employed in navigational bronchoscopy—to capture endoluminal trajectories and quantify adherence to nonholonomic motion constraints. By embedding kinematic and geometric priors into the analysis, the proposed framework offers an objective and interpretable alternative to traditional supervisor-dependent assessments, while helping to bridge the current gap between conventional training practices in bronchoscopy and the data-driven methodologies that have transformed other areas of surgical education.

A. Problem Statement and Contributions

Bronchoscopy is essential for diagnosing and treating pulmonary diseases, with clinical outcomes closely tied to operator skill. Yet current assessment methods rely heavily on subjective expert evaluations and lack objective, quantitative metrics. This gap hampers consistent training and limits the ability to advance skill transfer in emerging robotic endoscopy platforms. To address these challenges, we present a model-based framework for objective skill assessment in bronchoscopy, leveraging clinically viable sensing technologies to deliver reproducible and interpretable evaluations.

Our contributions are:

1) Nonholonomic Kinematic Model: We formulate an nonholonomic model of bronchoscopy that encodes the geometric constraints expert operators implicitly obey when navigating along airway centerlines. This model provides the foundation for evaluating deviations from expert-like motion.
2) Model-based Semi-Supervised Classification: We couple the kinematic model with a semi-supervised classifier that uses trajectory deviations as discriminative features, enabling systematic separation of expert and novice performance.

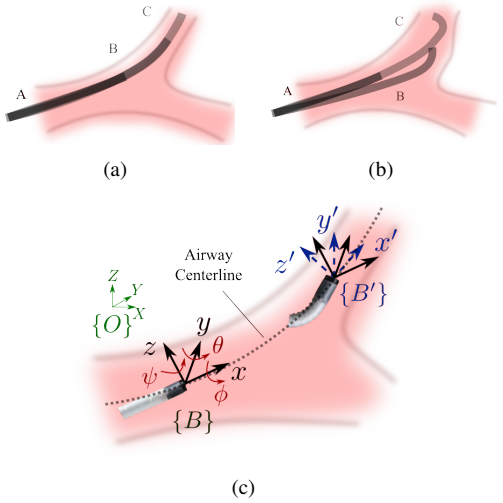

Fig. 2. A schematic of bronchoscopy illustrating our hypothesis. (a) An expert bronchoscopist can smoothly follow the airway centerline by simultaneously bending the bronchoscope's articulated tip while advancing it through the lumen. (b) In contrast, a novice bronchoscopist tends to exhibit less smooth movements, leading to more frequent contact with and pressure on the surrounding tissue, especially around bifurcations. (c) A schematic of optimal bronchoscope steering in the airways. An inertial coordinate frame O is fixed at the entry point of the bronchoscope into the airways at the beginning of the trachea, with the bronchoscope tip position represented by $\mathbf{p} = [x,\ y,\ z,]^T$. A local body-fixed frame B is attached to the bronchoscope tip, initially coinciding with the inertial frame.

The remainder of this paper is organized as follows. Section II introduces the proposed methodology, including the nonholonomic kinematic model of bronchoscopy, the use of Model Predictive Path Integral (MPPI) control to generate expert-like trajectories, and the definition of deviation-based performance features. Section III describes the experimental setup, dataset collection, and evaluation protocol. The results of feature analysis and classification experiments are then presented and discussed. Finally, Section IV provides concluding remarks and outlines future directions for objective skill assessment in bronchoscopy and robotic endoscopy.

II. Methodology

In bronchoscopy, the operator aims to keep the bronchoscope's camera centered along the airway lumen, avoiding contact with the airway wall, as illustrated in Fig. 2(a). However, novice operators may encounter two common issues: misalignment of the camera with the airway centerline (Fig. 2(b)) and abrupt, uncontrolled movements that can result in trauma to the airway wall. Based on these observations, we propose a geometric model of optimal bronchoscopy that experts adhere to. We hypothesize that deviations from this optimal trajectory can effectively indicate

the operator's skill level, with greater deviations suggesting lower proficiency.

Our approach assumes that the bronchoscope's Cartesian coordinates and pose are available during the procedure, achievable using electromagnetic sensors commonly used in navigational bronchoscopy. Additionally, due to the practical limitations of real-time lung movements and mismatches with preoperative images, we assume that the exact centerline of the lumen from medical imaging is not known during the procedure.

A. Kinematic model of optimal bronchoscopy

We present a kinematic model for bronchoscope tip steering that we hypothesize expert clinicians would follow. This model is subsequently used to quantify the performance of operators. Figure 2(c) illustrates the concept of *optimal bronchoscopy*, where the instrument advances smoothly along the airway centerline with minimal tissue interaction. The bronchoscope tip is articulated by a control lever to bend in the desired direction, while axial twisting of the shaft changes the bending plane, enabling 3D navigation through branching structures. To formalize this behavior, we attach a local body-fixed frame $\{B\}$ to the tip. Within this frame, we denote the insertion velocity, twist rate, and bending rate as v, ω, and $\gamma \in \mathbb{R}$, respectively. The tip orientation in space is parameterized using Euler angles—roll (ϕ), pitch (θ), and yaw (ψ)—about the x, y, and z axes.

Since the primary objective is to follow the airway centerline, we assume negligible rotation about the z-axis ($\psi \approx 0$). The bronchoscope tip position is expressed in the inertial frame as $\mathbf{p} = [x, y, z]^T \in \mathbb{R}^3$, and the generalized configuration is defined as $\mathbf{q} = [x, y, z, \phi, \theta]^T \in \mathcal{C} \subset \mathbb{R}^5$.

The orientation of the body frame relative to the inertial frame is described by the rotation matrix $\mathbf{R}_B^O$:

$$\mathbf{R}_B^O = \begin{bmatrix} C_\theta & -S_\theta & 0 \\ C_\phi S_\theta & C_\phi C_\theta & -S_\phi \\ S_\phi S_\theta & S_\phi C_\theta & C_\phi \end{bmatrix}, \tag{1}$$

where $C(\cdot)$ and $S(\cdot)$ denote cosine and sine, respectively. Under the assumption that forward motion occurs exclusively along the local x-axis, the inertial-frame velocity must satisfy

$$\mathbf{R}_B^O \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} v \\ 0 \\ 0 \end{bmatrix}.$$

Substituting (1) yields two independent kinematic constraints:

$$-\dot{x}S_\theta + \dot{y}C_\theta C_\phi + \dot{z}C_\theta S_\phi = 0 \tag{2a}$$

$$-\dot{y}S_\phi + \dot{z}C_\phi = 0. \tag{2b}$$

These constraints are *nonholonomic*, meaning they restrict instantaneous velocities but cannot be directly integrated into pose coordinates. Collectively, they reduce the admissible motions of the tip to a three-dimensional subspace of the five-dimensional configuration space. This can be compactly written in Pfaffian form as $\mathbf{A}(\mathbf{q})\dot{\mathbf{q}} = 0$, with

$$\mathbf{A} = \begin{bmatrix} -S_\theta & C_\theta C_\phi & C_\theta S_\phi & 0 & 0 \\ 0 & -S_\phi & C_\phi & 0 & 0 \end{bmatrix}. \tag{3}$$

At each configuration $\mathbf{q}$, the admissible velocities lie in the null space $\mathcal{N}(\mathbf{A}(\mathbf{q}))$. A convenient basis $\{g_1, g_2, g_3\}$ of this null space yields the reduced-order kinematic model:

$$\dot{\mathbf{q}} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ 0 \\ 0 \end{bmatrix} v + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \omega + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \gamma. \tag{4}$$

Thus, the bronchoscope is governed by three independent inputs: insertion speed v, axial twist ω, and bending rate γ. Together, they span the full configuration space. Applying the accessibility rank condition confirms that the Lie algebra of the control vector fields spans $\dim(\mathcal{C}) = 5$, establishing *full controllability* of the system:

$$\operatorname{rank} \Delta(\mathbf{q}) = \dim(\mathcal{C}). \tag{5}$$

This controllability property underpins our subsequent controller design, which generates reference trajectories that represent *optimal bronchoscopy* and allows comparison against operator-executed paths.

B. Model-Predictive Path Integral Control

In this study, the approximated optimal trajectories are simulated using the Model Predictive Path-Integral (MPPI) controller. MPPI is a sampling-based optimal control method that reformulates trajectory optimization as an information-theoretic inference problem, enabling it to handle nonlinear dynamics and non-quadratic cost functions [24]. Unlike classical trajectory optimization approaches, MPPI is well-suited to problems with complex constraints and discontinuous costs, while its sampling-based structure supports parallel computation for real-time implementation [25]. These features make MPPI particularly suitable for modeling *optimal bronchoscopy*, where motion must respect anatomical constraints while maintaining smooth navigation.

Let the bronchoscope configuration at time t be denoted by $\mathbf{q}_t \in \mathcal{C}$, evolving according to the nonholonomic equations in (4). The control input is defined as $\mathbf{u}_t = [v_t, \omega_t, \gamma_t]^T \in \mathbb{R}^3$, corresponding to insertion speed, twist rate, and bending rate. The state evolution is then expressed as

$$\mathbf{q}_{t+1} = \mathbf{F}(\mathbf{q}_t, \mathbf{u}_t), \tag{6}$$

where $\mathbf{F}(\cdot)$ encodes the reduced-order kinematic model.

At each iteration, MPPI samples K perturbed input sequences $V = (\mathbf{v}_t)_{t=0}^{T-1}$ from a Gaussian distribution centered at a nominal input sequence $U = (\mathbf{u}_t)_{t=0}^{T-1}$ with covariance Σ_t, i.e. $\mathbf{v}_t \sim \mathcal{N}(\mathbf{u}_t, \Sigma_t)$. Each candidate sequence generates a trajectory $\{\mathbf{q}_0, \ldots, \mathbf{q}_T\}$, and its performance is evaluated through a cost function

$$S(V; \mathbf{q}_0) = \phi(\mathbf{q}_T) + \sum_{t=0}^{T-1} c(\mathbf{q}_t, \mathbf{v}_t), \tag{7}$$

where $c(\cdot)$ is the stage cost and $\phi(\cdot)$ is the terminal cost.

The optimal controls are then obtained as a weighted average of the sampled inputs:

$$\mathbf{u}_t^* = \frac{\sum_{k=1}^{K} w(V^k)\, \mathbf{v}_t^k}{\sum_{k=1}^{K} w(V^k)}, \tag{8}$$

with importance weights

$$w(V^k) = \exp\left(-\tfrac{1}{\lambda} S(V^k)\right), \tag{9}$$

where λ is a temperature parameter that adjusts exploration versus exploitation.

To adapt MPPI for bronchoscopy trajectories, we segment each recorded operator path into K segments around airway bifurcations . Each segment is further divided into m clips indexed by time steps. For each clip, candidate waypoints are extracted from the operator's trajectory. Only waypoints that lie in the forward direction of motion and are sufficiently distant from the previous waypoint are retained.

The retained waypoints guide MPPI to generate trajectories $\hat{\mathbf{q}}(t)$ that remain close to the operator's original motion while satisfying the nonholonomic constraints in (2). To ensure smoothness, the waypoints are connected by cubic splines, and their tangent directions are used as guidance. The stage cost function used in MPPI implementation is:

$$\begin{aligned} c(\mathbf{q}_t, \mathbf{v}_t) = (\mathbf{q}_t - \mathbf{q}_f)^T Q (\mathbf{q}_t - \mathbf{q}_f) + \\ (\mathbf{v}_t - \mathbf{v}_f)^T R (\mathbf{v}_t - \mathbf{v}_f) + \alpha * \arccos(\mathbf{z}_t \cdot \mathbf{z}_f), \end{aligned} \tag{10}$$

where $\mathbf{q}_t$ is the configuration at time t, $\mathbf{v}_t$ is the input, $\mathbf{z}_t$ is the direction vector at time t which can be calculated from $\mathbf{q}_t$. The quantities R, Q, α are tunable parameters.

The MPPI controller then refines the trajectory by sampling variations of $\mathbf{u}_t = [v_t, \omega_t, \gamma_t]^T$ and selecting those that minimize deviation from the waypoint-guided spline while enforcing kinematic feasibility.

C. Performance Features

To quantify operator proficiency, we evaluate the deviation of each operator-executed trajectory from an optimal reference generated by MPPI under the nonholonomic kinematic model. Since a full trajectory may contain both smooth and irregular segments, we adopt a clip-based approach for finer granularity. Each trajectory segment is divided into four clips, and the features described below are computed independently for each clip. The resulting feature set captures local deviations from expert-like motion patterns.

Feature 1: Positional Deviation. The first feature measures the mean squared positional error between the operator trajectory and the MPPI-generated reference trajectory. Each trajectory is uniformly sampled into $4N$ points in time, and the MPPI trajectory is resampled by arc length to ensure alignment with the operator path. The feature for each clip is defined as

$$F_1 = \frac{1}{N} \sum_{i=1}^{N} \|\Delta \mathbf{p}_i\|^2, \tag{11}$$

where $\Delta \mathbf{p}_i = \mathbf{p}_i - \hat{\mathbf{p}}_i$ denotes the positional deviation at sample i, with $\mathbf{p}_i$ representing the operator trajectory points and $\hat{\mathbf{p}}_i$ the corresponding MPPI-generated points (see Fig. 3). Lower values of F_1 indicate closer adherence to the optimal path.

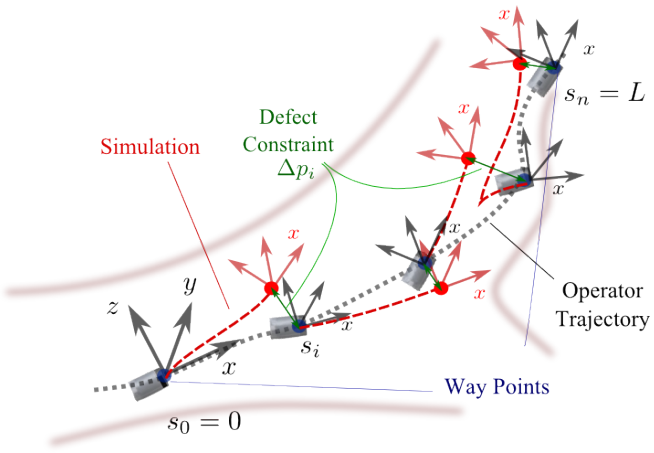

Fig. 3. Schematic of the proposed classification framework. Starting from the operator's initial pose $\mathbf{q}(t_0)$, an MPPI-generated trajectory is produced that satisfies the nonholonomic constraints of the kinematic model. Deviations between the operator trajectory and the reference trajectory are quantified through positional and orientation-based features (F_1, F_2, F_3). These features serve as discriminative metrics for evaluating operator performance.

Feature 2: Orientation Deviation. The second feature quantifies the mismatch in heading direction between the operator trajectory and the MPPI reference. Using the same temporal sampling as above, let $\mathbf{v}_j$ and $\hat{\mathbf{v}}_j$ denote the heading vectors of the operator and reference trajectories at point j, respectively. The orientation deviation is defined as

$$F_2 = \frac{1}{N} \sum_{j=1}^{N} \arccos\left(\frac{\mathbf{v}_j \cdot \hat{\mathbf{v}}_j}{\|\mathbf{v}_j\| \, \|\hat{\mathbf{v}}_j\|} \right), \tag{12}$$

which measures the average angular misalignment of the operator's trajectory relative to the optimal reference.

Feature 3: Orientation Fluctuation. The third feature captures the smoothness of the operator trajectory by measuring frame-to-frame fluctuations in orientation. Let $\mathbf{v}_i$ denote the heading direction at point i. The feature is defined as

$$F_3 = \frac{1}{N-1} \sum_{i=2}^{N} \arccos\left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{v}_i}{\|\mathbf{v}_{i-1}\| \, \|\mathbf{v}_i\|} \right). \tag{13}$$

Large fluctuations reflect abrupt or unstable motions, which are more common among novice operators.

III. Experiments

A. Experiment setup

The bronchoscopy dataset was collected using a Bronchoscopy Training Model (KOKEN Co., LTD.) and a commercial Ambu bronchoscope (Ambu A/S, Denmark). A 6-degree-of-freedom (6-DOF) electromagnetic (EM) sensor (NDI, Canada) was attached near the bronchoscope tip to record real-time pose information. In parallel, a monocular camera located at the distal end of the bronchoscope provided video recordings at a resolution of 480×480 pixels and a

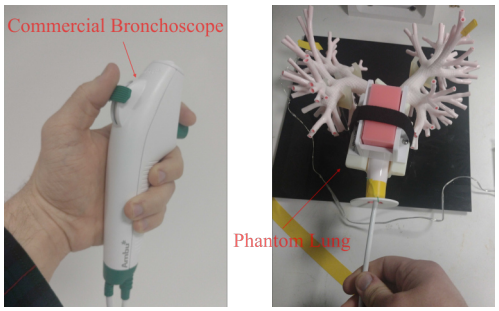

Fig. 4. Experimental setup showing the bronchoscope, the phantom lung model, and the electromagnetic (EM) tracker embedded near the bronchoscope tip.

frame rate of 16 Hz. All video frames were synchronized with time-stamped EM sensor data. The experimental setup is shown in Fig. 4.

A total of eleven participants took part in the study: five skilled pulmonologists with extensive clinical bronchoscopy experience and six novices with limited prior exposure. Each operator controlled the bronchoscope using the three primary degrees of freedom (DoFs): insertion/retraction, articulation of the distal tip via the control lever, and axial rotation of the handle to change the bending plane. These DoFs correspond to the input variables (v, ω, γ) of the nonholonomic kinematic model described in Section II-A.

The lung phantom was securely positioned within the working volume of the magnetic field generator to ensure accurate EM tracking of the bronchoscope tip. During each trial, operators were instructed to navigate through predefined airway paths. The experimental protocol comprised six distinct trajectories designed to test maneuverability in different anatomical scenarios, including straight passages, smooth bends, and bifurcations. For each trajectory, synchronized video and pose data were recorded to build the dataset.

B. Results

Figure 5 illustrates representative examples of operator trajectories compared with MPPI-generated references. Expert trajectories tend to closely follow the optimal paths with smooth curvature and minimal oscillation. By contrast, novice trajectories often exhibit erratic deviations, especially around bifurcations where navigation is more challenging. Nevertheless, both groups follow broadly similar global paths, reflecting that the primary differences between experts and novices are local and episodic rather than global.

This observation is corroborated in the feature analysis (Fig. 6). The mean feature values for experts and novices are often comparable, indicating that the majority of the trajectory is executed similarly. However, novices exhibit substantially higher variance, with outlier clips corresponding to abrupt deviations in tip orientation or unstable steering. These deviations typically arise in anatomically complex regions such as bifurcations. Thus, the discriminative information lies not in the global average performance but in the distributional spread and the presence of high-error events.

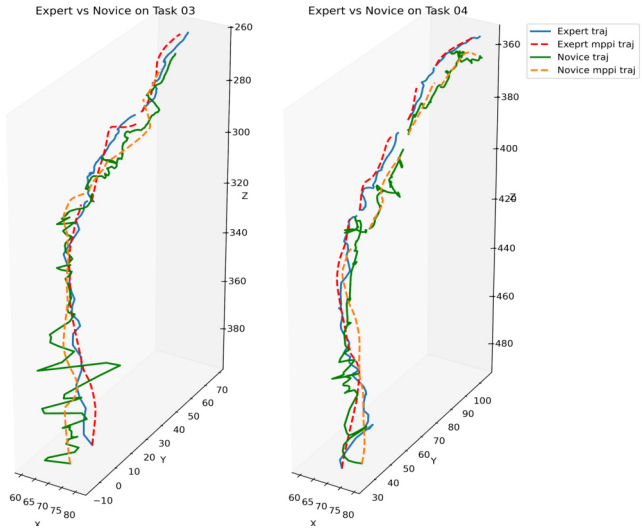

Fig. 5. Examples of original trajectories and MPPI-generated trajectories for experts and novices. Expert trajectories exhibit smooth, stable motion, whereas novice trajectories show greater oscillatory deviations, particularly near bifurcations.

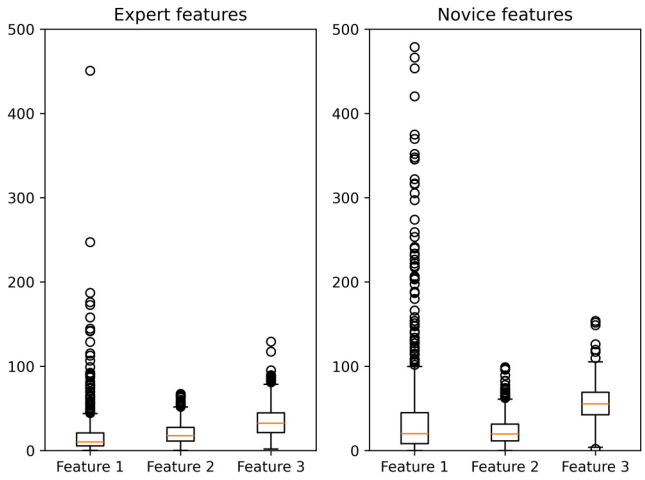

Fig. 6. Boxplots of feature distributions for expert and novice operators. Experts display lower variance across features, while novices exhibit occasional large deviations.

To quantify these differences, we performed a Mann–Whitney U test (also known as the Wilcoxon rank-sum test), a non-parametric method that compares the rank distributions of two groups without assuming normality. This test is well-suited to our setting because novice feature distributions are skewed and contain outliers. The results confirm that novices exhibit significantly higher positional deviation (F_1: median 20 vs. 10, $U = 118$, $p = 0.008$, $r = 0.46$) and steering fluctuation (F_3: median 50 vs. 30, $U = 104$, $p = 0.002$, $r = 0.55$). In contrast, orientation deviation (F_2) showed comparable medians between groups (20 vs. 20), but novices displayed greater variance, with occasional high-error outliers near bifurcations ($U = 162$, $p = 0.11$, $r = 0.21$). These findings support our hypothesis that experts and novices differ primarily in the spread and episodic

TABLE I

PERFORMANCE METRICS OF CLASSIFIERS ACROSS ITERATIONS. ITERATIVE REFINEMENT IMPROVES DISCRIMINATION UNTIL CONVERGENCE (ITERATION 3), BEYOND WHICH TEST ACCURACY BEGINS TO DECLINE DUE TO LABEL IMBALANCE.

Best Thre	Acc	Precision	Recall	F1	Test Acc
0.4892	0.73	0.72	0.69	0.70	0.7962
0.4986	0.77	0.75	0.74	0.75	0.7962
0.5237	**0.80**	**0.81**	**0.80**	**0.81**	**0.7870**
0.7129	0.84	0.86	0.86	0.86	0.6481
0.8581	0.86	0.90	0.89	0.89	0.6388

excursions of their trajectories rather than in global averages.

To capture these subtleties, we use the clip-level dataset by dividing each segment into four clips and extracting features for each clip. Because trajectory-level labels (expert or novice) are only weakly informative at the clip level—i.e., most clips may appear expert-like even for novices—we adopt a multi-stage refinement strategy. Specifically, we train an initial logistic regression classifier, then iteratively identify and down-weight outlier clips whose features are inconsistent with their assigned labels. This process effectively implements a semi-supervised learning approach: clips that are clearly separable reinforce the decision boundary, while ambiguous clips are gradually reassigned, yielding a more robust classifier (Fig. 7).

Table I summarizes classifier performance over iterations. During the first three iterations, accuracy and F1-score steadily improve, while the optimal decision threshold stabilizes near 0.5. Beyond this point, performance on the held-out test set begins to degrade due to over-filtering of clips and an imbalance toward novice labels. For this reason, the third iteration is selected as the final model.

In summary, our results demonstrate that expert and novice operators can be distinguished not by global trajectory averages, but by the presence and frequency of local deviations. The proposed semi-supervised refinement of the clip-level classifier exploits this property to provide reliable discrimination, achieving balanced accuracy, precision, recall, and F1-scores of approximately 0.80 at convergence. Importantly, this approach yields a granular and objective performance score at the clip level, enabling fine-grained assessment of operator skill rather than coarse binary labeling. Such quantitative scoring provides a pathway toward scalable and unbiased evaluation of bronchoscopy training, complementing existing supervisor-dependent methods with data-driven feedback.

IV. CONCLUSION AND FUTURE WORK

In this paper, we presented a model-based framework for assessing operator performance in navigational bronchoscopy. By embedding nonholonomic kinematic constraints into a Model Predictive Path Integral (MPPI) control framework, we generated expert-like reference trajectories and defined objective features that quantify local deviations in operator motion. Using these features, we demonstrated that expert and novice performance can be robustly separated not by global averages, but by episodic deviations, particularly in anatomically complex regions such as bifurcations. A semi-supervised refinement of a logistic regression classifier further enabled clip-level classification, yielding balanced accuracy, precision, recall, and F1-scores of approximately 0.80.

This work contributes toward a shift from subjective, supervisor-dependent evaluations to objective, interpretable, and granular performance metrics in bronchoscopy training. Such quantitative assessments can complement simulation-based curricula, provide automated feedback to trainees, and support skill transfer in emerging robotic bronchoscopy systems.

Future research will extend this framework in several directions. First, validation on larger and more diverse operator cohorts, including multi-center clinical data, will be necessary to ensure generalization. Second, longitudinal studies are needed to investigate how the proposed metrics evolve with training and whether they correlate with clinical outcomes. Finally, we aim to integrate the framework into robotic bronchoscopy platforms, where model-based scoring could provide autonomous assistance, adaptive training feedback, and formal skill transfer between manual and robotic systems. Furthermore, by releasing our dataset and methods publicly, we hope to catalyze the development of standardized benchmarks and foster a research community around objective skill assessment in endoluminal endoscopy.

REFERENCES

[1] P. L. Shah, F. J. Herth, Y. G. Lee, and G. J. Criner, Eds., *Essentials of Clinical Pulmonology*. Boca Raton, FL: CRC Press, 2018.

[2] E. O. Leiten, E. M. H. Martinsen, P. S. Bakke, T. M. L. Eagan, and R. Grønseth, "Complications and discomfort of bronchoscopy: a systematic review," *European Clinical Respiratory Journal*, vol. 3, no. 1, p. 33324, 2016.

[3] D. R. STATHER, P. MACEACHERN, A. CHEE, E. DUMOULIN, and A. TREMBLAY, "Trainee impact on advanced diagnostic bronchoscopy: An analysis of 607 consecutive procedures in an interventional pulmonary practice," *Respirology*, vol. 18, no. 1, pp. 179–184, 2013.

[4] A. Ernst, M. M. Wahidi, C. A. Read, J. D. Buckley, D. J. Addrizzo-Harris, P. L. Shah, F. J. F. Herth, A. de Hoyos Parra, J. Ornelas, L. Yarmus, and G. A. Silvestri, "Adult Bronchoscopy Training Current State and Suggestions for the Future: CHEST Expert Panel Report," *CHEST*, vol. 148, no. 2, pp. 321–332, August 2015.

[5] M. M. Wahidi, G. A. Silvestri, R. D. Coakley, J. S. Ferguson, R. W. Shepherd, L. Moses, J. Conforti, L. G. Que, K. J. Anstrom, F. McGuire, H. Colt, and G. H. Downie, "A prospective multicenter study of competency metrics and educational interventions in the learning of bronchoscopy among new pulmonary fellows," *Chest*, vol. 137, no. 5, p. 1040–1049, May 2010.

[6] E. C. Gerretsen, A. Chen, J. T. Annema, M. Groenier, E. H. van der Heijden, W. N. van Mook, and F. W. Smeenk, "Effectiveness of flexible bronchoscopy simulation-based training: A systematic review," *CHEST*, vol. 164, no. 4, pp. 952–962, 2023.

[7] Z. Deng, M. Khadem, K. Dhaliwal, A. D. L. Marshall, and D. Hanley, "Automated skill evaluation of bronchoscopy operators using geometric feature analysis," *Journal of Medical Robotics Research*, vol. 10, no. 03n04, p. 2550007, 2025.

[8] N. Voduc, N. Dudek, C. M. Parker, K. B. Sharma, and T. J. Wood, "Development and validation of a bronchoscopy competence assessment tool in a clinical setting," *Annals of the American Thoracic Society*, vol. 13, no. 4, pp. 495–501, 2016.

[9] H. Colt, *Flexible Bronchoscopy Inspection, BAL, BX*, Bronchoscopy International, Irvine, CA, 2020.

Fig. 7. Refinement of the logistic regression classifier over three iterations. Initially, clip-level labels inherit trajectory-level annotations, producing noisy boundaries. Iterative refinement removes outliers and reweights ambiguous clips, resulting in a cleaner separation of expert and novice classes.

[10] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh *et al.*, "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling," in *MICCAI workshop: M2cai*, vol. 3, no. 2014, 2014, p. 3.

[11] N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, and G. D. Hager, "A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2025–2041, 2017.

[12] M. J. Fard, S. Ameri, R. Darin Ellis, R. B. Chinnam, A. K. Pandya, and M. D. Klein, "Automated robot-assisted surgical skill evaluation: Predictive analytics approach," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 1, p. e1850, 2018.

[13] Z. Li, L. Gu, W. Wang, R. Nakamura, and Y. Sato, "Surgical skill assessment via video semantic aggregation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 410–420.

[14] J. Atoum, G. L. Johnston, N. Simaan, and J. Y. Wu, "Multi-modal gesture recognition from video and surgical tool pose information via motion invariants," in *2025 International Symposium on Medical Robotics (ISMR)*. IEEE, 2025, pp. 150–156.

[15] J. Rosell, A. Pérez, P. Cabras, and A. Rosell, "Motion planning for the virtual bronchoscopy," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2932–2937.

[16] J. Deng, P. Li, K. Dhaliwal, C. X. Lu, and M. Khadem, "Feature-based visual odometry for bronchoscopy: A dataset and benchmark," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 6557–6564.

[17] Q. Qiao, G. Borghesan, J. De Schutter, and E. Vander Poorten, "Force from shape—estimating the location and magnitude of the external force on flexible instruments," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1826–1833, 2021.

[18] S. De, J. Rosen, A. Dagan, B. Hannaford, P. Swanson, and M. Sinanan, "Assessment of tissue damage due to mechanical stresses," *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1159–1171, 2007.

[19] E. Mackute, F. X. Zhang, K. Dhaliwal, and M. Khadem, "Navigational bronchoscopy in critical care via end-to-end pose regression," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2025*, ser. Lecture Notes in Computer Science, vol. 15968. Springer, 2025, pp. 404–414.

[20] F. X. Zhang, E. Mackute, M. Kasaei, K. Dhaliwal, R. Thomson, and M. Khadem, "BREA-Depth: Bronchoscopy realistic airway-geometric depth estimation," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2024*. Springer, 2024, pp. 676–686.

[21] E. Mackutė, A. Abdalla, S. Dickson, K. Dhaliwal, and M. Khadem, "On challenges of monocular pose estimation for endoluminal navigation," *Journal of Medical Robotics Research*, vol. 09, no. 03n04, p. 2440009, 2024.

[22] S. Miyawaki, S. Choi, E. A. Hoffman, and C.-L. Lin, "A 4dct imaging-based breathing lung model with relative hysteresis," *Journal of Computational Physics*, vol. 326, pp. 76–90, 2016.

[23] Z. Deng, M. Khadem, K. Dhaliwal, A. D. L. Marshall, and D. Hanley, "Automated skill evaluation of bronchoscopy operators using geometric feature analysis," *Journal of Medical Robotics Research*, 2025.

[24] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Information-theoretic model predictive control: Theory and applications to autonomous driving," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1603–1622, 2018.

[25] J. Alvarez-Padilla, J. Z. Zhang, S. Kwok, J. M. Dolan, and Z. Manchester, "Real-time whole-body control of legged robots with model-predictive path integral control," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 14 721–14 727.