Proceedings on Automation in Medical Engineering
Proc AUTOMED, 2026, Vol. 3, No 1, Article ID 2482, 2 Pages

Infinite Science Publishing

Modeling surgeon-specific instrument usage profiles to improve instrument anticipation

L. Wagner[1*], L. Bernhard[1], Alexander Geiger[1], Dennis Schneider[1], A.Knoll[2] and D. Wilhelm[1]

[1] *Department of Surgery, Research Group MITI, TUM University Hospital rechts der Isar, Munich, Germany*
[2] *Chair of Robotics, Artificial Intelligence (AI) and Embedded Systems, Technical University of Munich, Munich, Germany*
[*] *Corresponding author, email: lars.wagner@tum.de*

Abstract: Anticipating surgical instrument usage is essential for robotic assistance systems to facilitate smooth intraoperative workflows. This study investigates whether modeling surgeon-specific profiles improves the prediction of laparoscopic instrument sequences. We evaluate a profile-aware GRU against a standard GRU and a hidden Markov model. The results demonstrate that incorporating surgeon-specific profiles enhances predictive performance, highlighting the importance of personalization for effective robotic assistance in surgery.

I. Introduction

Robotic assistance systems will be increasingly introduced into operating rooms to assist with routine surgical tasks. This is driven by a growing shortage of personnel and the potential to support clinical workflows more efficiently. Although several prototypes have been developed in the scientific community for roles such as camera guidance, instrument delivery, and setup assistance, replicating the nuanced behavior of experienced human scrub nurses remains a serious challenge. Experienced human assistants can anticipate the needs of the surgical team based on subtle cues, procedural knowledge, and familiarity with individual surgeon preferences. Emulating this anticipatory behavior is a key step toward enabling robotic systems to provide meaningful intraoperative support instead of just executing tasks reactively. Our previous work [1] has shown that predicting the surgeon's need for laparoscopic instruments through real-time video analysis is feasible, laying the groundwork for a smooth robotic workflow support with a decreased need for verbal communication. As of now, existing approaches focus on procedural or visual cues alone and do not account for variability in instrument usage among surgeons. In related clinical decision support research, Du et al. [2] proposed a method for modeling surgeon-specific preferences across multiple decision dimensions, illustrating that personalized models can improve interpretability and trust. Extending these concepts to robotic systems is a promising approach. By incorporating surgeon-specific patterns, robotic assistants could increase predictive accuracy and enhance surgeon acceptance in real-world settings. In this work, we investigate whether modeling surgeon-specific profiles improves the anticipation of laparoscopic instrument sequences. By clustering surgeons based on their instrument usage profiles and training models on cluster identity, we aim to enhance their predictive accuracy.

II. Material and methods

We collected sequences of laparoscopic instrument usage from 20 minimally invasive cholecystectomies performed by nine surgeons. To capture individual patterns, we extracted normalized frequency vectors of instrument usage for each surgeon, separately for the left and right working trocars. We then clustered these vectors using k-means clustering (k = 3), yielding latent groups of surgeons with similar instrument usage patterns. The objective is to anticipate the next surgical instrument used during a laparoscopic procedure based on the preceding sequence. Given a partial instrument sequence $\mathcal{X}_{\leq t} = \{x_1, \dots x_t\}$ and the surgeon's cluster unique identifier (ID) $c \in \{0,1,2\}$, the model learns to predict the subsequent instrument x_{t+1}. Formally, the task is to learn the function

$$f\colon (\mathcal{X}_{\leq t}, c) \rightarrow x_{t+1}. \quad (1)$$

Each sequence is split into multiple input-target pairs for supervised learning. We use a Gated Recurrent Unit (GRU) [3], commonly used in recommendation systems [4], to model the sequential anticipation of laparoscopic instruments. The model includes two embedding layers, one for instrument tokens and one for the cluster IDs. The instrument embedding is concatenated with the embedding of the cluster ID and passed to the GRU. We compare this approach against a GRU without incorporating any cluster information and a cluster-specific Hidden Markov Model (HMM), where a separate model is trained for each cluster. In the HMM setting, predictions for the next instrument are derived from the emission distribution of the most probable hidden state after observing the preceding sequence. We report top-1 accuracy, weighted F1 score, and mean reciprocal rank (MRR) as the key evaluation metrics. We also include top-k accuracy curves to evaluate model performance at various thresholds. We evaluate the models using 5-fold cross-validation.

III. Results and discussion

Figure 1 shows a t-SNE projection of the learned instrument frequency features. Each point represents a surgeon, and dashed lines connect the left and right trocar embeddings of the same surgeon, enabling comparison of instrument usage across hands. Most surgeons exhibit similar patterns on both sides, clustering closely together in the embedded space. However, surgeons with IDs 2 and 8 show a cluster change between their left and right trocars, suggesting different profiles or usage patterns depending on the trocar side.

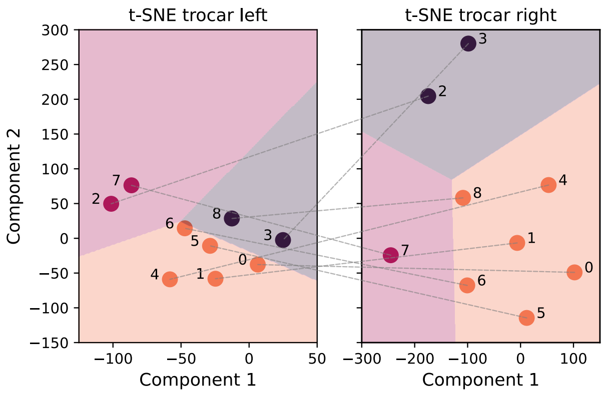

Figure 1: t-SNE embeddings of surgeon-specific instrument usage patterns for the left and right trocar. Colors denote the three clusters obtained via k-means. Dashed lines connect the same surgeon across both embeddings, illustrating how individual surgeons may shift between clusters depending on trocar side.

Table 1 reports the average performance metrics across 5 folds for each model configuration. The cluster-aware GRU model achieves the best results for both trocar sides. While the standard GRU benefits from learning sequential patterns, adding cluster-specific embeddings further improves the anticipation performance. In contrast, the HMM struggles to capture sequential dynamics and performs poorly, particularly for the left working trocar, despite being trained cluster-specifically.

Figure 2 illustrates the top-k accuracy curve for the right working trocar, demonstrating the increase in prediction performance with higher k values. The cluster-aware GRU exhibits a steeper and higher top-k accuracy curve than the baselines, which further confirms the advantages of incorporating surgeon specific instrument usage patterns in surgical instrument anticipation models.

Table 1: Selected metric scores. The averaged metrics over 5 folds are reported (%) with the corresponding standard deviation (±).

Model	Accuracy@1	wF1	MRR
Right working trocar			
HMM	29.14 ± 5.32	25.17 ± 5.10	47.59 ± 4.98
GRU w/o cl. ID	46.67 ± 4.60	42.53 ± 3.55	64.91 ± 3.47
GRU with cl. ID	51.12 ± 7.82	47.86 ± 7.68	68.07 ± 4.74
Left working trocar			
HMM	10.31 ± 7.12	9.62 ± 6.34	29.64 ± 6.37
GRU w/o cl. ID	55.33 ± 6.63	47.06 ± 7.73	72.03 ± 4.33
GRU with cl. ID	57.81 ± 3.25	51.73 ± 2.80	71.50 ± 3.77

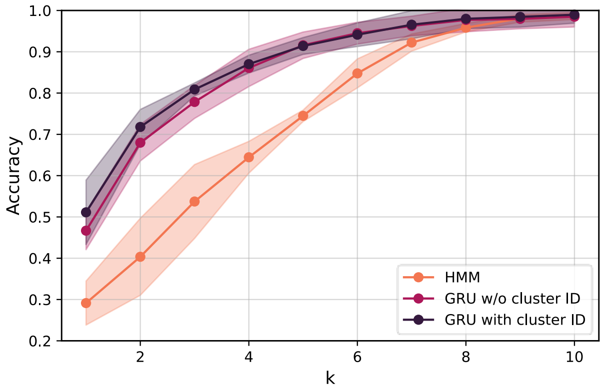

Figure 2: Top-k accuracy curves for the prediction task of the instruments in the right trocar. The GRU model with cluster ID outperforms both the GRU without cluster ID and the HMM baseline across all k values, indicating the predictive benefit of incorporating surgeon-specific behavioral clusters.

The results suggest that generalized instrument usage patterns offer valuable structural insights that help sequential models, such as GRUs, predict the next instrument in a surgical procedure. With larger datasets, it is conceivable that more fine-grained usage patterns could be extracted for individual surgeons, enabling more accurate modeling of personal preferences. When combined with real-time data streams in the operating room, this information could support future robotic systems that adapt instrument provisioning or assistance workflows to the operating surgeon's individual needs.

IV. Conclusions

These findings underscore the potential benefits of incorporating surgeon-specific patterns into predictive models of intraoperative instrument usage. These models could support context-aware robotic assistance systems by accurately anticipating surgeon preferences during surgical procedures.

Author's statement

Research funding: The author state funding support by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (StMWi) (Grant No. DIK0372). Conflict of interest: Authors state no conflict of interest. Ethical approval: The research related to human use complies with all relevant national regulations and institutional policies and was performed in accordance with the tenets of the Declaration of Helsinki. Ethical approval for the use of data was granted by the Ethics Committee of the TUM University Hospital rechts der Isar (No. 337/21 S).

References

[1] L. Wagner *et al.*, “Robotic scrub nurse to anticipate surgical instruments based on real-time laparoscopic video analysis,” *Communications Medicine*, vol. 4, no. 1, p. 156, 2024.

[2] Z. Du *et al.*, “Interpretable personalized surgical recommendation with joint consideration of multiple decisional dimensions,” *NPJ Digital Medicine*, vol. 8, no. 1, p. 168, 2025.

[3] K. Cho *et al.*, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[4] B. Hidasi *et al.*, “Session-based recommendations with recurrent neural networks,” in *Proc. Int. Conf. Learning Representations (ICLR)*, San Juan, Puerto Rico, May 2016.