This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Hemorrhage First: Rethinking Surgical Safety Assessment via Video-based Blood Loss Estimation

Pengfei Li[1,2][0009−0002−7527−0219]*, Jérôme A. Kurylec[1,2][0009−0000−8999−9110], Rao Rohan Girish[1,3][0009−0001−5366−9459], Aatharshan Kannathasan[1,2],[0009−0000−4553−2471], Philip C. Müller[2][0000−0001−8317−3449], Beat P. Müller[1,2][0000−0002−8552−8538], Simon Pezold[1,2][0000−0002−4998−8983], and Joël L. Lavanchy[1,2][0000−0003−0248−4996]

[1] Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland
[2] Department of Surgery, Clarunis – University Digestive Health Care Center Basel, Basel, Switzerland
[3] Biozentrum, University of Basel, Basel, Switzerland

Abstract. Bleeding is one of the most frequent and clinically significant events affecting surgical safety. In both intraoperative risk control and postoperative assessment, blood loss is closely associated with patient safety and mortality, surgical quality, and treatment decisions. Traditional methods for blood loss assessment, such as suction canister volume, only provide a global, postoperative estimate and thus fail to capture temporal characteristics and local risks of bleeding. To address these limitations, we propose two video analysis tasks related to surgical bleeding. The first task is clip-level bleeding severity classification. We constructed a newly annotated dataset containing 66 videos of robotic liver resections under a slightly modified version of the Validated Intraoperative Bleeding (VIBe) scale. This task aims to classify the bleeding severity of short surgical video clips, enabling fine-grained analysis of intraoperative bleeding conditions. The second task is surgery-level blood loss estimation for the entire surgery. Inspired by the blood loss component of the Surgical Apgar Score, we divide the total blood loss of a surgery into different severity levels and use clip-level bleeding predictions to estimate the overall blood loss category of the full surgical video. We evaluate several mainstream video recognition models on the proposed tasks and further introduce Bleeding Severity MViT, a MViT-based model designed for bleeding severity classification. Our model achieves the best performance among the evaluated baselines, particularly in recognizing more severe bleeding cases. We also explore a two-stage framework combining a binary bleeding classifier with a bleeding severity classifier, which demonstrates the potential to outperform end-to-end models. Finally, we apply our model to unlabeled surgical videos and validate its effectiveness for surgery-level blood loss estimation with an accuracy of 56.1%.

Keywords: Surgical safety · Bleeding classification · Minimally invasive surgery.

* Corresponding author (peng.li@stud.unibas.ch).

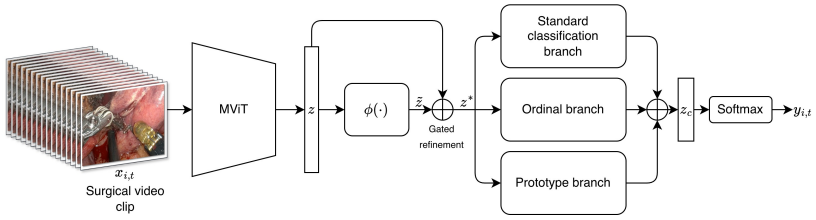


Fig. 1. Overview of the Bleeding Severity MViT framework.

1 Introduction

Intraoperative bleeding is one of the most frequent and clinically meaningful events affecting surgical safety. In minimally invasive surgery, bleeding not only reflects potential tissue or vessel injury, but may also influence subsequent decisions on hemostasis and procedural strategy. More importantly, severe bleeding is associated with an increased risk of mortality [2]. Current clinical practice usually estimates blood loss by visually reading suction canisters, subtracting irrigation fluid, weighing blood-soaked gauzes, or relying on surgeons' end-of-case estimation [7, 17]. Although these methods provide useful estimates, their retrospective and global nature limits their ability to capture the relationship between intraoperative bleeding dynamics and final blood loss volume.

Meanwhile, current surgical video analysis tasks have not fully explored the fine-grained bleeding severity assessment. Large-scale datasets such as MultiBypass140 [12] mainly provide annotations for surgical phase and step recognition. Later studies further analyzed intraoperative adverse events in MultiBypass140 and showed that bleeding accounts for the majority of recorded adverse events, highlighting its clinical importance [11, 16]. Although bleeding has been increasingly recognized as an important indicator for evaluating surgical safety, current bleeding-related studies remain largely limited to binary bleeding detection and bleeding source localization [8, 9, 1]. Standardized datasets and learning tasks for clinically meaningful bleeding severity assessment remain limited.

Inspired by the Validated Intraoperative Bleeding (VIBe) scale [13], we design our annotation scheme by considering both the criteria in VIBe and the practical characteristics of real surgical scenarios, where severe bleeding is promptly controlled by surgeons and highest-severity samples are inherently scarce. Based on these considerations, we construct a clip-level bleeding severity annotation scheme tailored to surgical video analysis. Rather than treating bleeding as a binary event, our scheme distinguishes visually different bleeding states according to their clinical relevance. This design better reflects surgical decision-making by indicating whether bleeding can be ignored, monitored, or requires immediate hemostatic intervention. Moreover, clip-level severity annotations enable surgery-level blood loss estimation by linking the temporal distribution of bleeding severity to overall blood loss.

Table 1. Modified Validated Intraoperative Bleeding (VIBe) scale used in this study.

Class	Visual presentation	Anatomic appearance	Qualitative description
No bleeding	No bleeding.	No bleeding.	No bleeding.
Low	Oozing or intermittent bleeding.	Capillary-like bleeding.	Mild.
Medium	Continuous flow.	Venule- or arteriolar-like bleeding.	Moderate.
High	Spurting, overwhelming, inaccessible, or gushing bleeding.	Noncentral or central venous-/arterial-like bleeding.	Severe or life-threatening bleeding requiring urgent hemostatic intervention.

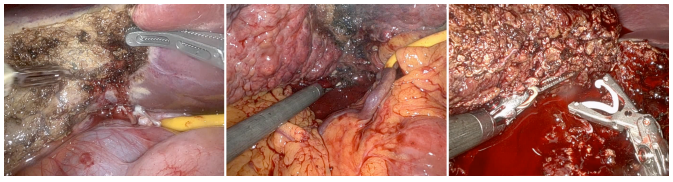
The contribution of our work is threefold: 1) We introduce a clinically motivated surgery-level blood loss estimation task. 2) We provide a new dataset of 66 videos of robotic liver resections with ground-truth total blood loss, of which 25 videos are additionally annotated with frame-level bleeding severity. 3) We propose Bleeding Severity MViT, a new method for clip-level bleeding classification and surgery-level blood loss estimation.

2 Methods and Materials

2.1 Dataset

Based on the proposed blood loss assessment tasks, we constructed a dedicated dataset of robotic liver resection videos from 66 patients at University Digestive Health Care Center in Basel, Switzerland, with corresponding surgery-level blood loss ranges. The videos varied substantially in duration, with a length of 139.26 ± 88.61 min (mean $\pm$ SD). Following common practice in previous surgical video studies [19, 10], the original videos were recorded at 25 fps and downsampled to 1 fps to improve processing efficiency and reduce temporal redundancy. All patient data were anonymized. The institutional review board waived the requirement for informed consent.

Clip-level Bleeding Classification. We selected 25 videos from the dataset for frame-level bleeding severity annotation. Two surgeons from our university hospital annotated the videos by jointly considering the visual characteristics of



(a) Low bleeding (b) Medium bleeding (c) High bleeding

Fig. 2. Representative examples of different intraoperative bleeding severity levels.

Table 2. Dataset statistics.

(a) Clip-level bleeding severity classification						
Split	Patients	No bleeding	Low	Medium	High	Total
Train	17	4428	2235	1096	61	7820
Val	4	995	419	237	24	1675
Test	4	1117	479	304	38	1938
Total	25	6540	3133	1637	123	11433
(b) Surgery-level estimated blood loss distribution						
Split		$\leq$ 100 mL	101–600 mL		$>$ 600 mL	Total
Train		5	9		4	18
Val		2	3		2	7
Test		17	21		3	41
Total		24	33		9	66

In (a), all values except Patients are clip counts; in (b), all values are patient counts.

real surgical videos and the original VIBe scale. Unlike the controlled scenarios used in the original VIBe study [13], naturally occurring surgical videos rarely contain extremely severe bleeding (VIBe grade 4), since surgeons actively avoid procedures that may cause substantial harm to patients. After reviewing our surgical videos, no uncontrollable or life-threatening bleeding events were observed. Therefore, we merged the original *Severe* and *Life-threatening* categories (VIBe grades 3 and 4) into a single *High* bleeding category. The original *Mild* and *Moderate* categories were preserved in their ordinal relationship and mapped to *Low* and *Medium*, respectively. The resulting annotation standard used in this work is summarized in Table 1. Representative examples of the severity levels are shown in Fig. 2.

For bleeding frames, we annotated the bleeding region using bounding boxes and assigned a severity label. To align with the VIBe scale and standardize video processing, annotations were converted into 16-second clips, with each clip assigned a single mutually exclusive severity label. Shorter events were centered within a clip, while longer events were split into consecutive non-overlapping clips. Clips were generated without temporal overlap, and no frames were duplicated across adjacent clips.

The data were split at the patient level to prevent patient overlap across splits, while considering surgical video duration and bleeding-category distribution to ensure representative training, validation, and test sets. The number of *No bleeding* clips was reduced to mitigate class imbalance. The final split statistics are reported in Table 2(**a**).

Surgery-level Blood Loss Estimation. For surgery-level blood loss estimation, we adopted the estimated blood loss (EBL) criterion from the 10-point Surgical Apgar Score [6]. In this scoring scale, total blood loss is assessed using an ordinal scale, where 4 points indicate no blood loss, 3 points correspond to an EBL of $\leq$ 100 mL, 2 points correspond to 101–600 mL, 1 point corresponds to $>$ 600mL. Following this clinically motivated grading scheme, we categorized the overall blood loss of each surgery according to its recorded postoperative EBL range. The distribution of the 66 surgical cases is summarized in Table 2(**b**).

2.2 Problem Definition

Let V_i denote the i-th surgical video, which is divided into a sequence of non-overlapping 16-second clips: $V_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,T_i}\}$, where T_i is the number of clips in the i-th video. Each clip $x_{i,t}$ is assigned a bleeding severity label $y_{i,t} \in \mathcal{C} = \{0, 1, 2, 3\}$, corresponding to *no bleeding*, *low bleeding*, *medium bleeding*, and *high bleeding*, respectively.

In the first stage, we train a clip-level classifier f_θ to predict the bleeding severity of each 16-second clip. The predicted clip-level labels are then aggregated into a surgery-level severity count vector: $\mathbf{n}_i = [n_i^0, n_i^1, n_i^2, n_i^3]$, where n_i^c denotes the number of clips predicted as severity class c. In the second stage, the count corresponding to clips predicted as *no bleeding*, n_i^0 is excluded, and the resulting count vector is used as input to another classifier g_ϕ to estimate the overall surgery-level blood loss range: $e_i = g_\phi(\mathbf{n}_i)$, where $e_i \in \mathcal{B} = \{0, 1, 2\}$ denotes the surgery-level estimated blood loss category, corresponding to $\leq$ 100 mL, 101–600 mL, and $>$ 600 mL, respectively.

2.3 Bleeding Severity MViT

For clip-level classification, we adopt MViT-v2-S [14] as the backbone. MViTv2 is a vision transformer [3] based architecture that further improves MViT [4] by introducing decomposed relative positional embeddings and residual pooling connections, leading to stronger visual representation learning for video recognition.

In real-world surgical videos, the detection of dynamic blood flow is often hindered by artifacts such as surgical instruments, smoke, and static blood pools. To robustly identify genuine blood flow dynamics, we propose the following design, an overview of which is given in Figure 1.

Gated Feature Refinement. Considering the complexity and visual variability of surgical scenes, we introduce a gated feature refinement module to emphasize bleeding-relevant information. Given the backbone feature $\mathbf{z}$, a non-linear refinement function $\phi(\cdot)$ first produces a refined representation $\tilde{\mathbf{z}}$. A lightweight bottleneck gating branch then generates channel-wise weights $\mathbf{a}$ to selectively modulate the refined feature. The final enhanced feature is obtained by: $\mathbf{z}^* = \mathbf{z} + \mathbf{a} \odot \tilde{\mathbf{z}}$.

Complementary Prediction Branches. Unlike general classification tasks, bleeding severity has an inherent ordinal structure, progressing from no bleeding to low, medium, and high bleeding. Therefore, it is important to model how severe the bleeding condition is within each clip. To better capture this severity-aware information, we design a three-branch architecture:

Standard classification branch: This branch directly predicts the class logits $\mathbf{z}_c^{\text{cls}} \in \mathbb{R}^K$, where $K = 4$ corresponds to no, low, medium, and high bleeding.

Ordinal branch: This branch explicitly models the ordered relationship among severity levels. Given the severity label space $\mathcal{C} = \{0, 1, 2, 3\}$, the ordinal branch predicts the probabilities of the clip severity exceeding each ordinal threshold: $q_0 = P(y > 0), \quad q_1 = P(y > 1), \quad q_2 = P(y > 2)$, which indicates whether the severity is higher than no, low, and medium bleeding, respectively. We then compute the expected severity score as

$$s = \sum_{k=0}^{2} q_k.$$

To convert this scalar ordinal estimate into class-wise logits, we measure the distance between s and each severity level c. The ordinal prior logits are therefore defined as $\mathbf{z}_c^{\text{ord}} = -\tau(c - s)^2$ with $c \in \{0, 1, 2, 3\}$, where τ is a temperature parameter controlling the sharpness of the ordinal prior distribution.

Prototype branch: This branch learns a prototype for each bleeding severity level and compares each video feature with these prototypes. Given the enhanced feature $\mathbf{z}^*$, we first project it into a lower-dimensional prototype space: $\mathbf{u} = \psi(\mathbf{z}^*) \in \mathbb{R}^D$, where $\psi(\cdot)$ denotes the projection network and D is the prototype dimension. We define a set of learnable class prototypes as $\mathbf{P} = \{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ with $\mathbf{p}_c \in \mathbb{R}^D$, where each prototype represents one bleeding severity class.

To obtain the prototype-based logits, both the projected feature and the class prototypes are ℓ_2-normalized. We then compute their cosine similarities as class-wise scores:

$$\mathbf{z}_c^{\text{proto}} = \frac{\mathbf{u}}{\|\mathbf{u}\|_2} \cdot \frac{\mathbf{p}_c}{\|\mathbf{p}_c\|_2}, \quad c \in \{0, 1, 2, 3\}.$$

The resulting prototype logits $\mathbf{z}^{\text{proto}} \in \mathbb{R}^K$ provide a similarity-based prediction, encouraging the model to associate each clip with the learned representation of each severity level.

Final prediction: The final prediction $y_{i,t}$ is computed as:

$$\mathbf{z}_c = \mathbf{z}_c^{\text{cls}} + \lambda_o \mathbf{z}_c^{\text{ord}} + \lambda_p \mathbf{z}_c^{\text{proto}},$$
$$y_{i,t} = \arg\max_k \ \text{softmax}(\mathbf{z}_c)_k.$$

where $\mathbf{z}_c$ is the final logits obtained by fusing the 3 branches' outputs and λ_o and λ_p control the contributions of the ordinal and prototype branches, respectively.

2.4 Surgery-Level Blood Loss Estimation

At the surgery level, each complete video V_i was represented by a severity count vector $\mathbf{n}_i$, containing the numbers of clips predicted as *low*, *medium*, and *high* bleeding (recall Sect. 2.2). For the 25 clip-annotated patients, the severity count vectors were constructed from the ground-truth clip-level annotations and used to train a multilayer perceptron (MLP) classifier g_ϕ to predict the final surgery-level blood-loss category $e_i = g_\phi(\mathbf{n}_i)$ and evaluated it on the remaining 41 patients.

3 Experimental Setup

Implementation. We used MViT-v2-S as the backbone. Each video clip contained 16 frames resized to 224×224. The model was trained for 25 epochs using the AdamW optimizer with a batch size of 16, and a dropout rate of 0.3. To mitigate class imbalance, we applied both weighted random sampling and class-weighted loss during training.

We compared our method with several representative video classification baselines, including R3D-18 [18], R(2+1)D-18 [18], SlowFast-R50 [5], Swin3D-S [15], and MViT-v2-S.

In addition to end-to-end four-class classification, we also evaluated a two-stage strategy. This strategy used the same backbone architectures as the corresponding end-to-end models, but with different output heads. In the first stage, the original four-class classification head was replaced by a binary head to distinguish *bleeding* from *non-bleeding* clips. In the second stage, only clips predicted as *bleeding* were passed to a severity classifier, whose output head was modified to predict three bleeding severity classes: *Low*, *Medium*, and *High*. For the baseline models, we used standard cross-entropy loss. For BS-MViT, we adopted a task-guided loss motivated by two key properties of bleeding severity classification: class imbalance and ordinal label structure. We used focal loss with $\gamma = 2.0$ to emphasize hard and under-represented samples, and added an ordinal distance regularization term to penalize predictions according to their severity distance from the ground truth. The ordinal distance loss is defined as $\mathcal{L}_{\text{ord}} = \text{SmoothL1}(\bar{y}, y)$, where $\bar{y}$ denotes the expected predicted severity level and y is the ground-truth severity label. The final training objective is:

$$\mathcal{L} = \mathcal{L}_{\text{focal}} + \alpha \mathcal{L}_{\text{ord}},$$

where $\alpha = 0.2$.

Evaluation. We report overall accuracy, per-class accuracy, macro-F1 score, and weighted-F1 score. Since the bleeding severity dataset is class-imbalanced, macro-F1 is used as the primary metric to evaluate the balanced performance across different severity levels.

4 Results and Discussions

Clip level: Table 3 (**a**) summarizes the clip-level bleeding severity classification results. BS-MViT with ordinal loss achieves the highest accuracy on the *high bleeding* class (0.184), while maintaining competitive performance on the *low* (0.499) and *medium* (0.668) classes. However, its relatively weak performance on the *no bleeding* class suggests that it can be further improved by combining it with a model that is strong at distinguishing non-bleeding clips but less effective for fine-grained bleeding severity classification. This observation motivates our two-stage framework that combines the complementary strengths of different

Table 3. Classification results at clip level and surgery level.

(a) Clip-level bleeding severity classification results (Asterisk (*) indicates use of ordinal loss).

Method	Overall Acc.	Class Accuracy				Macro F1	Weighted F1
		No Bl.	Low	Medium	High		
R3D	0.609	0.722	0.509	0.421	0.079	0.439	0.613
R(2+1)D	0.611	**0.826**	0.376	0.266	0.026	0.387	0.591
SlowFast	0.549	0.557	0.516	0.641	0.000	0.394	0.561
Swin3D	0.584	0.686	0.336	**0.671**	0.000	0.398	0.578
MViT	0.539	0.637	0.476	0.322	0.158	0.395	0.552
MViT*	0.581	0.599	**0.555**	0.625	0.026	0.433	0.591
BS-MViT*	0.510	0.483	0.499	0.668	**0.184**	0.416	0.533
R3D + BS-MViT*	0.598	0.684	0.403	0.641	**0.184**	0.458	0.606
R(2+1)D + BS-MViT*	**0.633**	0.818	0.282	0.559	**0.184**	**0.464**	**0.615**

(b) Surgery-level blood loss classification results.

Method	Overall Acc.	Class Accuracy			Macro F1	Weighted F1
		$\leq$ 100 mL	101–600 mL	$>$ 600 mL		
BS-MViT	0.488	0.412	**0.571**	**0.333**	**0.413**	0.501
R(2+1)D + BS-MViT	**0.561**	**0.706**	0.524	0.000	0.398	**0.553**

models: R3D or R(2+1)D is first used for binary bleeding detection, and BS-MViT with ordinal loss is then applied to classify the bleeding clips. Among all methods, R(2+1)D + BS-MViT with ordinal loss achieves the best overall performance, with overall accuracy 0.633, Macro F1 score 0.464, and Weighted F1 score 0.615.

Surgery level: We apply BS-MViT and R(2+1)D + BS-MViT to the remaining 41 videos without clip-level annotations for surgery-level blood loss estimation. As shown in Table 3 (**b**), the two methods achieve overall accuracy of 0.488 and 0.561, respectively. It should be noted that the test set includes only three patients with surgery-level blood loss >600 mL, which may limit the reliability of the performance estimates for this category.

Limitations: Several limitations remain. First, bleeding may be partially or fully occluded by instruments, gauze, smoke, or anatomical structures, which can cause videos with few annotated bleeding clips to still have high final blood loss, and which highlights the limitation of purely vision-based estimation. Second, bleeding severity is dynamic and may change continuously during surgery, leading to ambiguous boundaries between severity levels. This is reflected in an inter-rater comparison: 120 randomly selected annotated clips were re-annotated by another surgeon, and the two annotators achieved an overall agreement of 0.583 and a quadratic weighted kappa of 0.484, which highlights the inherent difficulty of the presented task. Third, irrigation may introduce uncertainty because water mixed with blood can resemble active bleeding. Fourth, bleeding duration varies across real surgical scenarios. Although we used 16-second clips to align with the VIBe scale, actual bleeding events are not fixed in duration. At the surgery

level, the relatively small cohort and substantial class imbalance further limit the robustness of the results.

5 Conclusion

This study represents an initial step toward fine-grained and clinically interpretable bleeding analysis in surgical videos. Future work will focus on expanding the dataset, improving the recognition of rare high-severity bleeding events, establishing more fine-grained clinical criteria, and exploring more robust temporal modeling for procedure-level blood loss estimation.

Acknowledgements We would like to thank Julia Ruppel for her assistance in obtaining the necessary patient data.

Disclosure of Interests. JLL was funded by the Swiss National Science Foundation (P5R5PM 21766), the Novartis Foundation for medical-biological Research (23C162), and by the Vontobel Foundation (0867/2024).

References

1. Bose, R., Nwoye, C.I., Lazo, J.F., Lavanchy, J.L., Padoy, N.: Feature mixing approach for detecting intraoperative adverse events in laparoscopic roux-en-y gastric bypass surgery. In: Gee, J.C., et al. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Lecture Notes in Computer Science, vol. 15970, pp. 178–188. Springer, Cham (2026)
2. DeMatteo, R., Blumgart, L., D'Angelica, M., Fong, Y., Jarnagin, W., Ruo, L., Gonen, M., Liau, K., Shia, J., Katz, S.: Operative blood loss independently predicts recurrence and survival after resection of hepatocellular carcinoma. Annals of Surgery **249**(4), 617–623 (2009)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale (2021)
4. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., Feichtenhofer, C.: Multiscale vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021)
5. Feichtenhofer, C., Fan, H., Malik, J., He, K.: Slowfast networks for video recognition (2019)
6. Gawande, A.A., Kwaan, M.R., Regenbogen, S.E., Lipsitz, S.A., Zinner, M.J.: An apgar score for surgery. Journal of the American College of Surgeons **204**(2), 201–208 (2007)
7. Gerdessen, L., Neef, V., Raimann, F.J., Zacharowski, K., Piekarski, F.: The visually estimated blood volume in scaled canisters based on a simulation study. BMC Anesthesiology **21**(1), 54 (2021)
8. Horita, K., Hida, K., Itatani, Y., Fujita, H., Hidaka, Y., Yamamoto, G., Ito, M., Obama, K.: Real-time detection of active bleeding in laparoscopic colectomy using artificial intelligence. Surgical Endoscopy **38**(6), 3461–3469 (2024)
9. Hua, S., Gao, J., Wang, Z., Yeerkenbieke, P., Li, J., Wang, J., He, G., Jiang, J., Lu, Y., Yu, Q., Han, X., Liao, Q., Wu, W.: Automatic bleeding detection in laparoscopic surgery based on a faster region-based convolutional neural network. Annals of Translational Medicine **10**(10), 546 (2022)
10. Kirtac, K., Aydin, N., Lavanchy, J.L., Beldi, G., Smit, M., Woods, M.S., Aspart, F.: Surgical phase recognition: From public datasets to real-world data. Applied Sciences **12**(17), 8746 (2022)
11. Lavanchy, J.L., Alapatt, D., Sestini, L., Kraljević, M., Nett, P.C., Mutter, D., Müller-Stich, B.P., Padoy, N.: Analyzing the impact of surgical technique on intraoperative adverse events in laparoscopic Roux-en-Y gastric bypass surgery by video-based assessment. Surgical Endoscopy **39**(3), 2026–2036 (2025)
12. Lavanchy, J.L., Ramesh, S., Dall'Alba, D., et al.: Challenges in multi-centric generalization: phase and step recognition in Roux-en-Y gastric bypass surgery. International Journal of Computer Assisted Radiology and Surgery **19**, 2249–2257 (2024)
13. Lewis, K.M., Li, Q., Jones, D.S., Corrales, J.D., Du, H., Spiess, P.E., Lo Menzo, E., DeAnda, A.J.: Development and validation of an intraoperative bleeding severity scale for use in clinical studies of hemostatic agents. Surgery **161**(3), 771–781 (2017)
14. Li, Y., Wu, C., Fan, H., Mangalam, K., Xiong, B., Malik, J., Feichtenhofer, C.: Improved multiscale vision transformers for classification and detection. CoRR **abs/2112.01526** (2021)

15. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., Hu, H.: Video swin transformer (2021)
16. Mishra, S., Xu, J., Ang, J., Mazomenos, E.B., Ang, B.T., Jin, Y.: Dbt-bleed: Dual-branch temporal modeling with key-frame selection for surgical bleeding detection. arXiv preprint arXiv:2606.22829 (2026)
17. Perri, G., Sparrelid, E., Siriwardena, A.K., Marchegiani, G., E-AHPBA Consensus Group on Blood Loss in Hepatopancreatobiliary Surgery: Estimation of intraoperative blood loss in hepatopancreatobiliary surgery: a delphi consensus process of the european–african hepato-pancreato-biliary association (E-AHPBA). BJS **111**(10), znae256 (2024)
18. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition (2018)
19. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M., Padoy, N.: Endonet: A deep architecture for recognition tasks on laparoscopic videos. IEEE Transactions on Medical Imaging **36**(1), 86–97 (2017)