RESEARCH Open Access

Automated detection of steps in videos of strabismus surgery using deep learning

Ce Zheng[1], Wen Li[2], Siying Wang[2], Haiyun Ye[2], Kai Xu[2], Wangyi Fang[2], Yanli Dong[2], Zilei Wang[2] and Tong Qiao[2*]

Abstract

Background Learning to perform strabismus surgery is an essential aspect of ophthalmologists' surgical training. Automated classification strategy for surgical steps can improve the effectiveness of training curricula and the efficient evaluation of residents' performance. To this end, we aimed to develop and validate a deep learning (DL) model for automated detecting strabismus surgery steps in the videos.

Methods In this study, we gathered 479 strabismus surgery videos from Shanghai Children's Hospital, affiliated to Shanghai Jiao Tong University School of Medicine, spanning July 2017 to October 2021. The videos were manually cut into 3345 clips of the eight strabismus surgical steps based on the International Council of Ophthalmology's Ophthalmology Surgical Competency Assessment Rubrics (ICO-OSCAR: strabismus). The videos dataset was randomly split by eye-level into a training (60%), validation (20%) and testing dataset (20%). We evaluated two hybrid DL algorithms: a Recurrent Neural Network (RNN) based and a Transformer-based model. The evaluation metrics included: accuracy, area under the receiver operating characteristic curve, precision, recall and F1-score.

Results DL models identified the steps in video clips of strabismus surgery achieved macro-average AUC of 1.00 (95% CI 1.00–1.00) with Transformer-based model and 0.98 (95% CI 0.97-1.00) with RNN-based model, respectively. The Transformer-based model yielded a higher accuracy compared with RNN-based models (0.96 vs. 0.83, $p < 0.001$). In detecting different steps of strabismus surgery, the predictive ability of the Transformer-based model was better than that of the RNN. Precision ranged between 0.90 and 1 for the Transformer-based model and 0.75 to 0.94 for the RNN-based model. The f1-score ranged between 0.93 and 1 for the Transformer-based model and 0.78 to 0.92 for the RNN-based model.

Conclusion The DL models can automate identify video steps of strabismus surgery with high accuracy and Transformer-based algorithms show excellent performance when modeling spatiotemporal features of video frames.

Keywords Deep learning, Strabismus surgery, Automated detection, Surgical videos

*Correspondence:
Tong Qiao
qiaojoel@163.com
[1] Department of Ophthalmology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China
[2] Department of Ophthalmology, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Lu Ding Road # 355, Shanghai 200000, China

Background

Strabismus, defined as any binocular misalignment, affects 0.8–6.0% of children and can lead to amblyopia, visual impairment, or even hampered visual system's development if left untreated [1–4]. Strabismus surgery, successfully realigns the eyes by adjusting eye muscle tension or position [5, 6], thereby alleviating double vision and improving quality of life. Competence in strabismus surgery is crucial for ophthalmology residents worldwide, as recommended from the American Board of Ophthalmology (ABO) in the United States [7, 8]. In the United Kingdom [9], the ophthalmic specialty training curriculum requires trainees to have completed 20 surgical strabismus procedures by completion of training. In China, however, ophthalmology residents reportedly perform less surgery during their training [10] than their counterparts in developed countries. Studies suggest that 50 cases might be necessary for an ophthalmologist to reach proficiency in strabismus surgery [11]. Given these concerns, it is imperative to explore ways to enhance the effectiveness of ophthalmology training programs, ultimately better preparing trainees for performing strabismus surgery.

To drive ophthalmic surgical trainees along the surgical learning curve in a competency-based setup, surgical educators must develop curricula with systematic skill and competency assessments to delegate appropriate responsibilities while ensuring patient safety. However, there is no universally accepted standard for strabismus surgery competency assessment worldwide. Typically, surgical skill is evaluated through the procedure's constituent steps or phases (e.g., conjunctival incision, exposure of muscle) using videography [12]. Residency training curricula assess surgical steps of intraoperative technical skill based on structured or unstructured rating scales [13–15], which are time-consuming, subjective, and highly variable. Therefore, an automated classification strategy for surgical steps is crucial to enhance trainees' learning in surgical curricula and serve as an reliable tool for evaluating resident physicians' performance.

Artificial intelligence (AI), particularly deep learning (DL), has the potential in automated ophthalmic surgical phase recognition. For cataract surgery, Charrière et al. proposed a statistical-based model for real-time analysis cataract surgery videos [16], while Primus et al [17]. adopted DL (recurrent neural network (RNNs)) to automatically assign cataract video frames to operation phases. More recently, our group reported a novel Transformed-based DL algorithm of cataract phase-specific augmented reality (AR) guidance system. Our results demonstrating its superior performance compared to related works [18].

Despite advances in AI-assisted ophthalmic surgical phase recognition, strabismus surgery has received limited attention. This study aimed to develop and validate a novel DL algorithm for automated detecting strabismus surgery steps in the videos without manual intervention. We further compare the performance of the proposed DL algorithm with that of previous RNN-based DL algorithm.

Methods

Study design and datasets

In this retrospective cohort study, we collected videos of 496 eyes of strabismus surgery by five pediatric ophthalmologists from the Department of Ophthalmology, Shanghai Children's Hospital (SCH), between July 2017 and October 2021. The institutional review board (IRB) of SCH approved this study (identifier, 2021R065-F01), and a waiver of informed consent was granted due to the retrospective cohort of deidentified videos captured for training purposes. All methods followed the tenets set forth in the Declaration of Helsinki, and all videos were deidentified according to Health Insurance Portability and Accountability Act (HIPAA) [19].

All videos were captured by an integrated digital video-capture system with Zeiss surgical microscopy. Based on the International Council of Ophthalmology's Ophthalmology Surgical Competency Assessment Rubrics (ICO-OSCAR: strabismus) [20], we analyze eight steps in strabismus surgery: (1) conjunctival incision & Tenon's dissection, (2) hooking rectus muscle, (3) exposure of rectus muscle, (4) placement of suture in muscle, (5) disinsertion of the rectus muscle, (6) use of caliper/scleral ruler, (7) reattachment of muscle (intrascleral needle pass), and (8) conjunctival closure (when appropriate). Using ICO-OSCAR's definitions for the eight steps, three pediatric ophthalmologists (XXZ, WYF, XYL and YH) manually identified the start and end of the different surgical steps, and cut the videos into clips. The dataset included 3345 videos of strabismus surgery procedures performed from SCH. We randomly divided the raw video dataset by eye-level into training (for updating model parameters), validation (for hyperparameter tuning and model selection), and testing dataset (for assessing model performance) at a 6:2:2 ratio. The flowchart of the current study was showed in Fig. 1.

Video processing and development of the DL algorithms

To analyze surgical video, we initially captured frames using OpenCV (Open Source Computer Vision Library. 2015). As videos consist of numerous sequential frames, a single surgical clip may contain thousands of frames that demand expensive hardware. To address this issue, we downsampled the surgical clips to 100 frames using an average method. To enhance our dataset, we performed data augmentation techniques including random

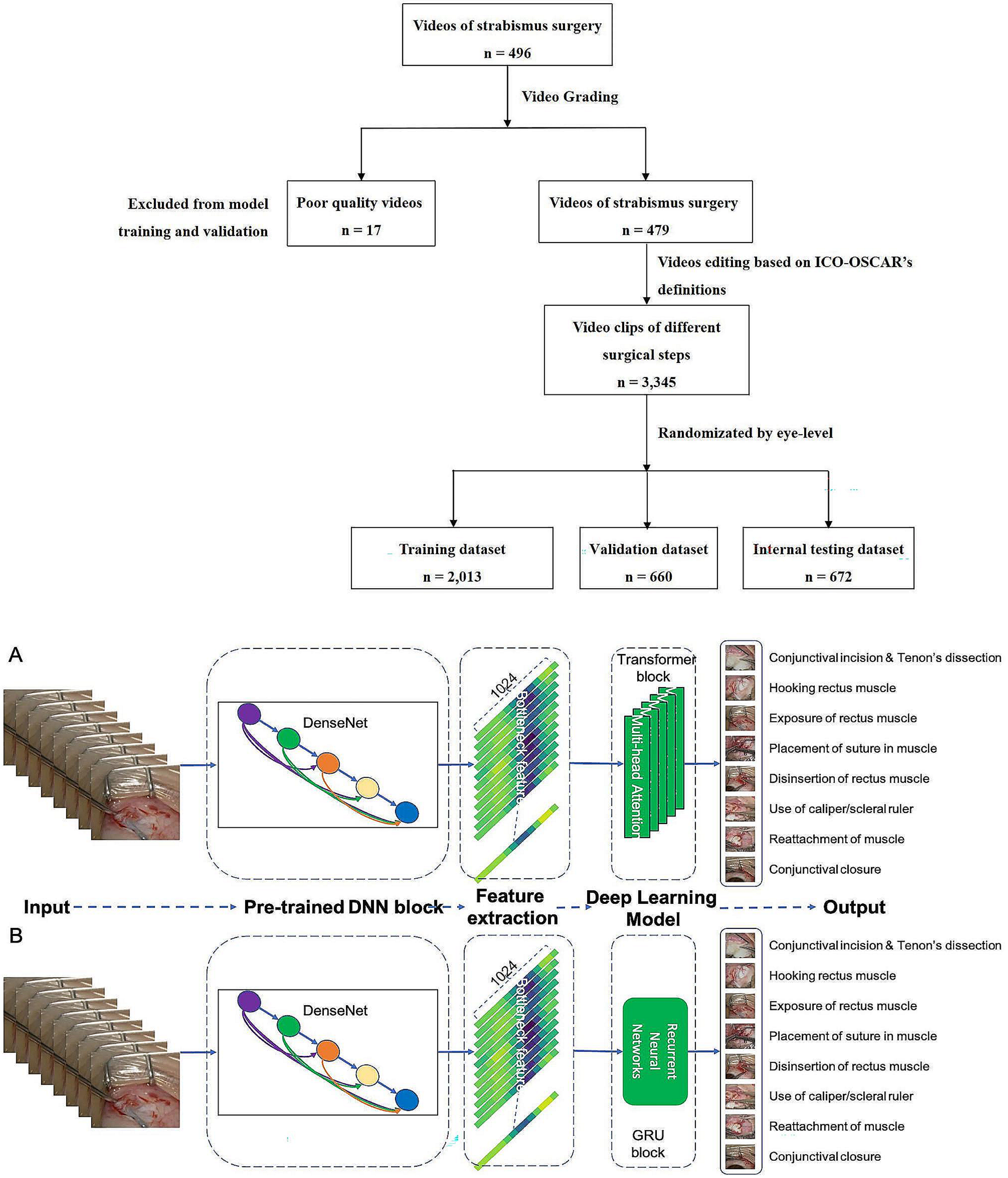
Videos of strabismus surgery
n = 496
Video Grading
Excluded from model training and validation
Poor quality videos
n = 17
Videos of strabismus surgery
n = 479
Videos editing based on ICO-OSCAR's definitions
Video clips of different surgical steps
n = 3,345
Randomizated by eye-level
Training dataset
n = 2,013
Validation dataset
n = 660
Internal testing dataset
n = 672
A
DenseNet
1024
Bottleneck feature
Transformer block
Multi-head Attention
Conjunctival incision & Tenon's dissection
Hooking rectus muscle
Exposure of rectus muscle
Placement of suture in muscle
Disinsertion of rectus muscle
Use of caliper/scleral ruler
Reattachment of muscle
Conjunctival closure
Input
Pre-trained DNN block
Feature extraction
Deep Learning Model
Output
B
DenseNet
1024
Bottleneck feature
Recurrent Neural Networks
GRU block
Conjunctival incision & Tenon's dissection
Hooking rectus muscle
Exposure of rectus muscle
Placement of suture in muscle
Disinsertion of rectus muscle
Use of caliper/scleral ruler
Reattachment of muscle
Conjunctival closure

Fig. 1 Flowchart of the Current Study and Hybrid DL Algorithm Diagrams. (**A**): a Convolutional Neural Network (CNN) and a Transformer module (RNN-based model) (**B**): a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN) consisting of Gated Recurrent Unit (GRU) layers (RNN-based model)

cropping and adjustments to saturation, brightness, and contrast.

We evaluated two hybrid DL algorithms (Fig. 1). The first DL algorithm has a CNN and a RNN architecture. To learn spatiotemporal surgical features for the entire step, we use a pre-trained network (DenseNet model) [21] to extract features from the video frames. The DenseNet model is a CNN model, which is a type of DL algorithm that processes images, and is frequently used for diseases classification tasks. We then concatenated these features to train multilayer RNN networks, learning spatiotemporal patterns that discriminate across steps (Fig. 1B). RNN is another type of DL algorithm that process data that comes in a sequence, such as words, sentences, audio, or video [22]. The RNN can learn from the whole sequence of data, not just from individual pieces. For compiling the RNN-based model, we use an Adam optimizer and a minibatch gradient descent of size 32.

The second algorithm we evaluated was a Transformer-based model [23]. Transformer have recently emerged as state-of-the-art DL architectures as described previously by various research groups [24]. Briefly, Transformer architectures are based on a self-attention mechanism that learns the relationships between elements of a sequence. As visual data follows a typical structure (e.g., spatial and temporal coherence), Transformer models and their variants have been successfully used for image recognition [25], object detection [26], and video understanding [27, 28]. Similar to the above RNN-based model, we also adopted a pre-trained DenseNet network for feature extraction. Since videos are ordered sequences of frames, we embed the positions of the frames present inside videos with an Embedding layer (positional encoding) and added these positional embeddings to the precomputed DenseNet feature maps (Fig. 1A). We then applied Transformer with multi-heads networks (number of heads=6) for video classification. For compiling the Transformer-based model, we use an Adam optimizer (an initial learning rate of 0.001, beta 1 of 0.9, beta 2 of 0.999) and a minibatch gradient descent of size 32. Early stopping was applied when the validation loss did not decrease for ten epochs.

Experimental setup

We implement the two DL algorithms with the Tensorflow framework (Google, TensorFlow-metal PluggableDevice, version 2.6.0, MacOSX-arm64) and Keras API (version 2.6.0). Our hardware included a MacBook Air (macOS Monterey 12.0.1 operation system) with an Apple M1 chip (7-core GPU and 16-core Neural Engine) and a 16GB RAM system.

Statistical analysis

According to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [29], we evaluated the performance of DL algorithms to classify strabismus surgery steps using the following metrics: accuracy, precision, recall and F1-score with a 2-sided 95% confidence interval (CI). Their formulas for calculation are as follows:

$$\mathrm{Accuracy} = \frac{True\ Positive + True\ Negative}{All} \tag{1}$$

$$\mathrm{Precision} = \frac{True\ Positive}{TruePositive + False\ Positive} \tag{2}$$

$$\mathrm{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

$$\mathrm{F1-score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

We used the area under the receiver operating characteristic (ROC) curve (AUC) to describe the ability of each DL algorithm to discriminate steps of strabismus surgery. For multiclass classification of surgical steps, we used one vs. all technique [30] to estimate steps-specific metrics and their 95% CIs.

We calculate the 95% confidence intervals (CIs) for our model's performance metrics using the Wilson score interval method. These confidence intervals furnish a statistical measure of certainty regarding the estimates of model performance. Specifically, a narrow confidence interval denotes high confidence in the performance estimate, while a wider interval indicates greater uncertainty. The calculation of the Wilson score interval is as follows:

$$\mathrm{CI} = \hat{\mathrm{p}} \pm \mathrm{z} * \mathrm{sqrt}((\hat{\mathrm{p}}(1-\hat{\mathrm{p}})/\mathrm{n}) + (\mathrm{z}^2/(4\mathrm{n}^2)))$$
$$\hat{\mathrm{p}} = (1/(1+\mathrm{z}^2/\mathrm{n})) * (\mathrm{P} + \mathrm{z}^2/(2\mathrm{n}))$$

Where $\hat{p}$ represents the model's performance metrics, z is the z-score corresponding to the standard normal distribution (for a 95% confidence interval, z=1.96)), and n denotes the sample size [31].

All statistical analyses were carried out using Python's statistical programming language (ver. 3.8.1, Python Software Foundation, Beaverton, US) and sklearn library (ver.1.0.2) [32].

Results

Of the total 496 strabismus surgical videos, 19 (3.8%) were excluded due to poor video quality or off-center of the surgical area, leaving the total dataset with 479 videos from 249 patients. Among the participants who underwent strabismus surgery, the average age was 6.23±2.91 years, and 121 of them (48.59%) were female.

Table 1 Summary metrics of algorithm performance for surgical steps classification in validation dataset

Metrics	Transformer-based model	CNN-RNN-based model
Accuracy (95% CI)	0.96 (0.94 to 0.98)	0.83 (0.79 to 0.87)
Macro-AUC (95%CI)	1.00 (1.00 to 1.00)	0.98 (0.97 to 1.00)

Supplementary Table 1 illustrates the number of video clips of each step in our dataset. There were 2,013, 660 and 672 video clips of surgical steps in the training, validation and testing dataset. Data augmentation further enriched the sample size of the training dataset to 10,065. After training for 100 epochs (the model showed no improvement in both accuracy and cross-entropy loss, Supplementary Fig. 1), DL models for identifying the steps in video clips of strabismus surgery achieved macro-average AUC of 1.00 (95% CI 1.00–1.00) with Transformer-based model and 0.98 (95% CI 0.97-1.00) with RNN-based model, respectively (Table 1; Fig. 2). Transformer-based model yielded a higher accuracy compared with RNN-based models (0.96 vs. 0.83, $p<0.001$).

Table 2 demonstrated step-level metrics, including accuracy, sensitivity, specificity, and precision of two DL algorithms across steps. In detecting different steps of strabismus surgery, the predictive ability of the Transformer-based model was better than that of the RNN-based model (Fig. 3). Precision ranged between 0.90 and 1 for the Transformer-based model and 0.75 to 0.94 for the RNN-based model. The f1-score ranged between 0.93 and 1 for the Transformer-based model and 0.78 to 0.92 for the RNN-based model.

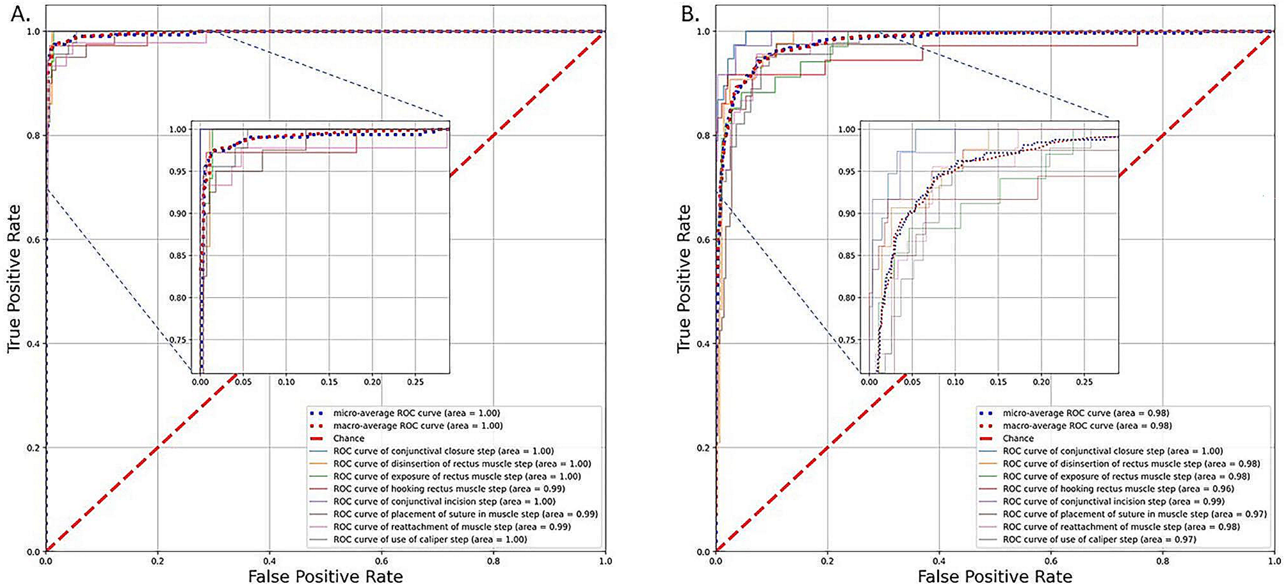

Fig. 2 Performance of two DL model in detection of surgical steps in the testing set. (A) ROC curve for detecting different steps of strabismus surgery with the Transformer-based model. (B) ROC curve for detecting different steps of strabismus surgery with the RNN-based model

Table 2 Accuracy, sensitivity, specificity, and precision for algorithms across different steps of strabismus surgery

Algorithm and Metrics	Conjunctival closure	Disinsertion of muscle	Exposure of rectus muscle	Hooking muscle	Conjunctival incision	Placement of suture	Reattachment of muscle	Use of caliper
Transformer-based model								
Accuracy	1	1	0.91 (0.88 to 0.94)	0.97	1	0.95	0.89	0.96
Precision	1	0.93	0.97	0.95	1	0.9	0.98	0.96
Recall	1	1	0.91	0.97	1	0.95	0.89	0.96
F1-score	1	0.97	0.94	0.96	1	0.93	0.93	0.96
CNN-RNN-based model								
Accuracy	0.89	0.84	0.85	0.83	0.92	0.78	0.84	0.73
Precision	0.85	0.86	0.78	0.94	0.92	0.79	0.75	0.82
Recall	0.89	0.84	0.85	0.83	0.92	0.78	0.84	0.73
F1-score	0.87	0.85	0.82	0.88	0.92	0.78	0.79	0.78

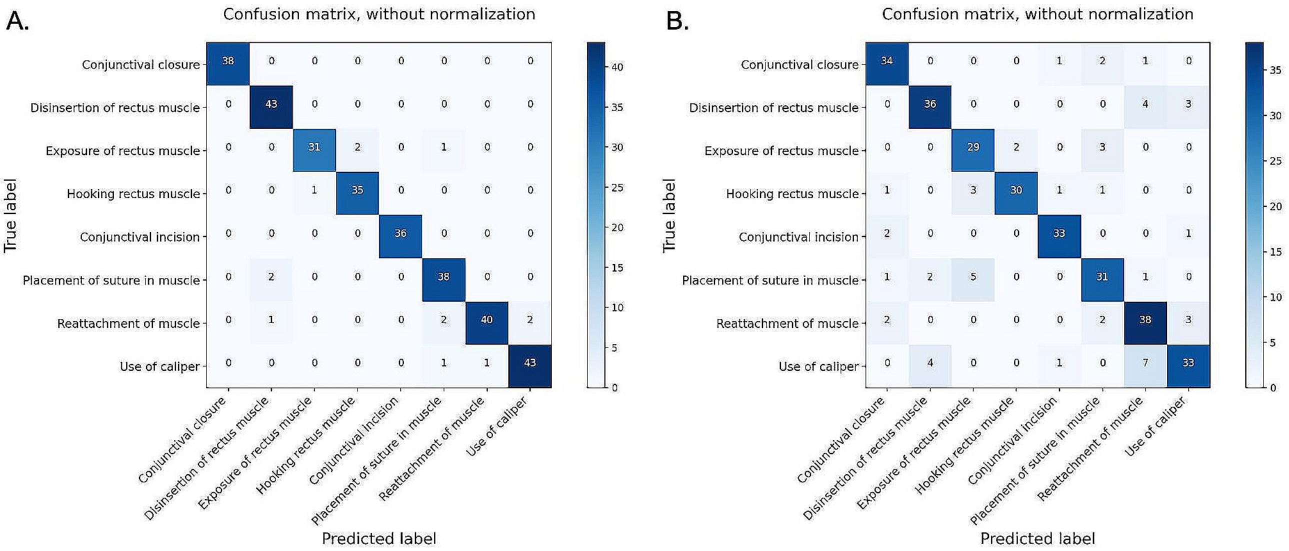

Fig. 3 Confusion matrices of two DL model in detection of surgical steps in the testing set. (A)Confusion matrix for detecting different steps of strabismus surgery with the Transformer-based model. (B)Confusion matrix for detecting different steps of strabismus surgery with the RNN-based model

Discussion

In this study, we investigated the performance of the DL algorithms in identifying steps of strabismus surgery from video clips. Our results revealed that the Transformer-based model achieved robust performance (AUC=1.00; and accuracy=96%) in classifying different surgical steps. To the best of our knowledge, no other DL system has been developed for recognizing the steps in ophthalmic operation other than cataract surgery. Therefore, we believe that our method provides a unique tool for objective, step-specific assessments of strabismus surgery.

Our study differs from previous cataract surgery studies in several key aspects. First, previous studies primarily relied on CV or machine learning techniques to model instrument labels [33–35], either alone or in combination with video images. In contrast, we adopted a hybrid DL architecture that leveraged Transformer-based models for video classification, building upon our previous work. This approach eliminates the need for instrument segmentation and reduces misalignment and misclassification errors. As limited studies have focused on strabismus surgery, we cannot directly compare our result with state of the art in this field. However, we recently proposed similarly Transformer-based models for cataract phase recognition, which outperforms several strong baselines in surgical phase recognition [36].

Not surprisingly, the Transformer-based model exhibits superior predictive performance compared to the RNN-based model in the current study. RNNs are algorithms for processing sequential data such as natural languages, sound, and time-series data [37]. However, RNNs suffer from gradient explosion/vanishing [38], which makes it challenging to process over long sequences. On the other hand, Transformers are new neural network architectures unveiled by Google AI in 2017 [23]. Utilizing the self-attention mechanism, Transformer-based models capitalize on parallel processing, making the training faster and building a better model in less time. Transformers have outperformed both CNNs and RNNs across a wide range of research areas [39–41]. Our results also confirm that the Transformer-based model demonstrates excellent performance in handling long sequences data. These findings confirm our hypothesis that Transformers are optimal choice for analyzing ophthalmic surgical sequences.

Our study holds the potential in many clinical settings. Ophthalmic surgery, such as strabismus surgery, learning curves for residents are closely tied to feedback-based teaching guidance. A DL algorithm could potentially mitigate surgical errors and guide surgeons, particularly novices, by providing real-time reminders of the next step and warnings for incorrect actions during surgery. It is also possible to develop a real-time supervion and objective surgical evaluation system to improve strabismus surgical outcomes. Wong, et al., recently reported a CNN-based system named DeepSurgery for the evaluation and supervision of cataract surgical procedures [42]. Furthermore, as AI technology advances, intelligent robots equipped with DL algorithms could become invaluable assistants in improving surgical precision and safety. In our previous work, we developed a novel phase-specific augmented reality (AR) [36] guidance system that provides ophthalmologists with customized visual cues based on the recognized surgical phase. This DL algorithm holds the potential to accelerate the development of such intelligent surgical robots, ultimately paving the way for precision medicine in ophthalmology.

There are several limitations to this study. First, our datasets were collected from a single center. There was no external validation dataset to confirm the performance of DL models. Diverse multicenter surgical videos are needed to validate the generalization of our DL model in future studies. Second, we used the dataset with multiple records per surgeon. Further study involving independent datasets with surgeries by different surgeons and clinical contexts is necessary to verify the classification error. Third, we did not include complicated cases or surgery with complications in this pilot study. Therefore, further study is required to assess the generalizability of DL algorithm in complex strabismus surgery cases. Finally, our DL model is developed based on video clips manually labeled and pre-segmented. The real-life applications will require algorithms to detect segment boundaries and assign steps' labels.

Conclusion

We demonstrate that DL models can automatically identify strabismus surgery steps with high accuracy based on surgical videos. Furthermore, Transformer algorithms show excellent performance when modeling spatiotemporal features of video frames. Further studies to determine the generalizability of the DL model in real life and its usefulness and potential application in surgical education seem to be warranted.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s12886-024-03504-8.

Supplementary Material 1

Acknowledgements

We wish to thank Xiaoxiao Zhang, Xinyue Li and Ying Huang for helping us with processing the large number of videos.

Author contributions

C.Z. contributed to the conception and design, and was involved in drafting and revising the manuscript. C.Z., W.L., S.Y.W., H.Y.Y., K.X., W.Y.F., Y.L.D., and Z.L.W. contributed to acquisition of data. C.Z., W.L., and S.Y.W. were involved in analysis and interpretation of data. T.Q. supervised the study and contributed as co-corresponding authors. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (81371010), Hospital Funded Clinical Research, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (21XJMR02) and Hospital Management Research Program of Institute of Hospital Development Strategy, China Hospital Development Institute, Shanghai Jiao Tong University (HDSI-2022-A-001).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review board (IRB) of Shanghai Children's Hospital, School of medicine, Shanghai Jiao Tong University (identifier, 2021R065-F01); the informed consent was waived. All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

Dr Zheng has 1 patent related to artificial intelligence applied to strabismus but no financial disclosures involving the work under consideration for publication (during the time involving the work, from initial conception and planning to present), nor any other relevant financial activities outside the submitted work (during the three years prior to submission), and no other relationships or activities that readers could perceive to have influenced, or that give the appearance of potentially influencing what is written in the submitted work (based on all relationships that were present during the three years prior to submission). Other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Received: 29 November 2023 / Accepted: 28 May 2024
Published online: 10 June 2024

References

1. Chia A, Roy L, Seenyen L. Comitant horizontal strabismus: an Asian perspective. Br J Ophthalmol. 2007;91(10):1337–40. https://doi.org/10.1136/bjo.2007.116905
2. Wallace DK, Christiansen SP, Sprunger DT, et al. Esotropia and Exotropia Preferred Practice Pattern®. Ophthalmology. 2018;125(1):P143–83. https://doi.org/10.1016/j.ophtha.2017.10.007
3. Mojon-Azzi SM, Kunz A, Mojon DS. Strabismus and discrimination in children: are children with strabismus invited to fewer birthday parties? Br J Ophthalmol. 2011;95(4):473–6. https://doi.org/10.1136/bjo.2010.185793
4. Chia A, Dirani M, Chan YH, et al. Prevalence of amblyopia and strabismus in young Singaporean Chinese children. Invest Ophthalmol Vis Sci. 2010;51(7):3411–7. https://doi.org/10.1167/iovs.09-4461
5. Rosenbaum AL. The goal of adult strabismus surgery is not cosmetic. Arch Ophthalmol. 1999;117(2):250. https://doi.org/10.1001/archopht.117.2.250
6. Gunton KB. Impact of strabismus surgery on health-related quality of life in adults. Curr Opin Ophthalmol. 2014;25(5):406–10. https://doi.org/10.1097/ICU.0000000000000087
7. Lee AG, Carter KD. Managing the new mandate in resident education: a blueprint for translating a national mandate into local compliance. Ophthalmology. 2004;111(10):1807–12. https://doi.org/10.1016/j.ophtha.2004.04.021
8. Lee AG. The new competencies and their impact on resident training in ophthalmology. Surv Ophthalmol. 2003;48(6):651–62. https://doi.org/10.1016/j.survophthal.2003.08.009
9. Dean SG, Mchugh J, Bowes O, Spencer F. Ophthalmology specialist trainee survey in the United Kingdom. The Royal College of Ophthalmologists. Accessed 07 Aug 2023. https://curriculum.rcophth.ac.uk/curriculum/ost/faqs/
10. Young AL, Jhanji V, Liang Y et al. A survey of perceived training differences between ophthalmology residents in Hong Kong and China. BMC Med Educ. 2015;15:158. Published 2015 Sep 28. https://doi.org/10.1186/s12909-015-0440-0
11. Kim Y, Kim YG, Kim HJ, et al. Learning curves for strabismus surgery in two ophthalmologists. Indian J Ophthalmol. 2015;63(11):821–4. https://doi.org/10.4103/0301-4738.171962
12. Cremers SL, Ciolino JB, Ferrufino-Ponce ZK, Henderson BA. Objective Assessment of skills in intraocular surgery (OASIS). Ophthalmology. 2005;112(7):1236–41. https://doi.org/10.1016/j.ophtha.2005.01.045
13. Puri S, Sikder S. Cataract surgical skill assessment tools. J Cataract Refract Surg. 2014;40(4):657–65. https://doi.org/10.1016/j.jcrs.2014.01.027
14. Pilling RF, Bradbury JA, Reddy AR. Strabismus surgical skills assessment tool: development of a surgical assessment tool for strabismus surgery

training. Am J Ophthalmol. 2010;150(2):275–e2782. https://doi.org/10.1016/j.ajo.2010.03.007
15. Gauba V, Tsangaris P, Tossounis C, Mitra A, McLean C, Saleh GM. Human reliability analysis of cataract surgery. Arch Ophthalmol. 2008;126(2):173–7. https://doi.org/10.1001/archophthalmol.2007.47
16. Charrière K, Quellec G, Lamard M, et al. Real-time analysis of cataract surgery videos using statistical models. Multimed Tools Appl Published Online. 2016. https://doi.org/10.48550/ARXIV.1610.05465
17. Primus MJ, Putzgruber-Adamitsch D, Taschwer M et al. Frame-Based Classification of Operation Phases in Cataract Surgery Videos. In: Conference on Multimedia Modeling. 2018. https://api.semanticscholar.org/CorpusID:40964125
18. Tu P, Ye H, Young J, Xie M, Zheng C, Chen X. Efficient Spatiotemporal Learning of Microscopic Video for Augmented reality-guided phacoemulsification cataract surgery. In: Medical Image Computing and Computer assisted intervention – MICCAI 2023. Springer Nat Switz. 2023;682–92. https://doi.org/10.1007/978-3-031-43990-2_64
19. Office for Civil Rights. Guidance Regarding Methods for Deidentification of Protected Health Informa- tion in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. U.S. Department of Health and Human Services. Accessed 07 Aug 2023. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
20. Golnik KC, Motley WW, Atilla H, et al. The ophthalmology surgical competency assessment rubric for strabismus surgery. J AAPOS. 2012;16(4):318–21. https://doi.org/10.1016/j.jaapos.2012.04.005
21. Huang G, Liu Z, Van Der Maaten L. K. Q. Weinberger. Densely Connected Convolutional Networks.; 2017. https://doi.org/10.1109/CVPR.2017.243
22. Dey R, Salem FM. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks.; 2017. https://doi.org/10.1109/MWSCAS.2017.8053243
23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Published Online. 2017. https://doi.org/10.48550/ARXIV.1706.03762
24. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Published Online. 2018. https://doi.org/10.48550/ARXIV.1810.04805
25. Dosovitskiy A, Beyer L, Kolesnikov A et al. An image is worth 16x16 words: transformers for image recognition at scale. Published online 2020. https://doi.org/10.48550/ARXIV.2010.11929
26. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. Published Online. 2020. https://doi.org/10.48550/ARXIV.2005.12872
27. Sun C, Myers A, Vondrick C, Murphy K, Schmid C. VideoBERT: a joint model for video and language representation learning. Published online 2019. https://doi.org/10.48550/ARXIV.1904.01766
28. Naseer M, Ranasinghe K, Khan S, Hayat M, Khan FS, Yang MH. Intriguing properties of vision transformers. Published Online. 2021. https://doi.org/10.48550/ARXIV.2105.10497
29. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799. https://doi.org/10.1136/bmjopen-2016-012799. Published 2016 Nov 14.
30. Hand DJ, Till RJ. A simple generalisation of the Area under the ROC curve for multiple class classification problems. Mach Learn. 2001;45:171–86. https://doi.org/10.1023/A:1010920819831
31. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22(158):209. https://doi.org/10.2307/2276774
32. Scikit-learn. Machine Learning in Python, Pedregosa JMLR 12, pp. 2825–2830, 201.
33. Yu F, Silva Croso G, Kim TS, et al. Assessment of Automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. JAMA Netw Open. 2019;2(4):e191860. https://doi.org/10.1001/jamanetworkopen.2019.1860. Published 2019.
34. Odysseas MJ, Efthymia M. A review on invariant manifolds and targeted energy transfer. Orient J Phys Sci Published Online 2018. https://api.semanticscholar.org/CorpusID:127250740
35. Martiano D, Charrière K, Lamard M, Cochener B. Indexing of cataract surgery video by content based video retrieval. Acta Ophthalmol. 2014;92. https://api.semanticscholar.org/CorpusID:75763827
36. Tu P, Ye H, Shi H, et al. Phase-specific augmented reality guidance for microscopic cataract surgery using long-short spatiotemporal aggregation transformer. Published Online. 2023. https://doi.org/10.48550/ARXIV.2309.05209
37. Murugan P. Learning the sequential temporal information with recurrent neural networks. Published online 2018. https://doi.org/10.48550/ARXIV.1807.02857
38. Kanai S, Fujiwara Y, Iwamura S. Preventing Gradient Explosions in Gated Recurrent Units. In: Neural Information Processing Systems. 2017. https://api.semanticscholar.org/CorpusID:13825036
39. Ahmad K, Alam F, Qadir J, et al. JMIR Form Res. 2022;6(5):e36238. https://doi.org/10.2196/36238. Published 2022 May 11. Global User-Level Perception of COVID-19 Contact Tracing Applications: Data-Driven Approach Using Natural Language Processing.
40. Karimi D, Vasylechko SD, Gholipour A. Convolution-Free Medical Image Segmentation using Transformers. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021. https://api.semanticscholar.org/CorpusID:232069052
41. Zhao C, Shuai R, Ma L, Liu W, Wu M. Improving cervical cancer classification with imbalanced datasets combining taming transformers with T2T-ViT. Multimed Tools Appl. 2022;81(17):24265–300. https://doi.org/10.1007/s11042-022-12670-0
42. Wang T, Xia J, Li R, et al. Intelligent cataract surgery supervision and evaluation via deep learning. Int J Surg. 2022;104:106740. https://doi.org/10.1016/j.ijsu.2022.106740

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.