Predicting the Phase of Cataract Surgery with Deep Learning

Joshua Bierbrier[1], Rebecca Hisey[1], Bining Long[1], Adrienne Duimering[1], Christine Law[1], Gabor Fichtinger[1], Matthew Holden[2]

[1]Queen's University, Kingston, Ontario, Canada, [2]Carleton University, Ottawa, Ontario, Canada

Introduction. Cataract surgery, a widely performed procedure, involves replacing the native lens with an artificial one. Grading resident surgeons in this skill is challenging and labour-intensive. Automated approaches to assess surgical skills have been explored[1]. Our overarching aim is to provide phase-specific skill assessments during cataract surgery. The first step, towards this aim, is to recognize the phases of cataract surgery. Deep learning approaches are used in several fields of surgical phase recognition[2]. Therefore, our goal is to classify the phases of cataract surgery using deep learning. We trained a hierarchical Long Short-Term Memory (LSTM) model that uses features extracted from the ResNet50 Convolutional Neural Network (CNN) to classify the phase for each frame of a surgical video.

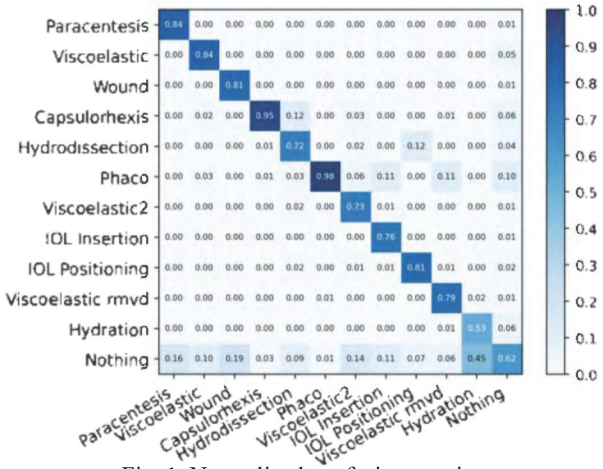

Fig. 1. Normalized confusion matrix.

Methods. Thirty recordings of cataract surgeries performed at Kingston Health Sciences Centres were obtained from surgical microscopes. Fifteen surgeries were performed by resident ophthalmologists and fifteen from staff ophthalmologists. The ground truth data was created by an expert manually labelling each video frame as belonging to one of twelve surgical phases. The model was trained in two steps. The pre-trained CNN ResNet50 architecture was first trained to predict the phase for each frame of the surgical videos. After discarding the output layer of the ResNet50 model, the final layer was used as the input feature vector for a temporal model. A hierarchical LSTM model provided the final prediction, one of twelve classes, for each frame of each video. Class balancing with replacement was performed given the unequal number of frames in each phase. The dataset was divided into six folds, with a split of 20 training, 5 validation, and 5 test videos. The outcomes of interest included accuracy (the percent of correctly predicted frames), F-score (a measure of precision and recall), and Jaccard Index. Results were averaged across test videos.

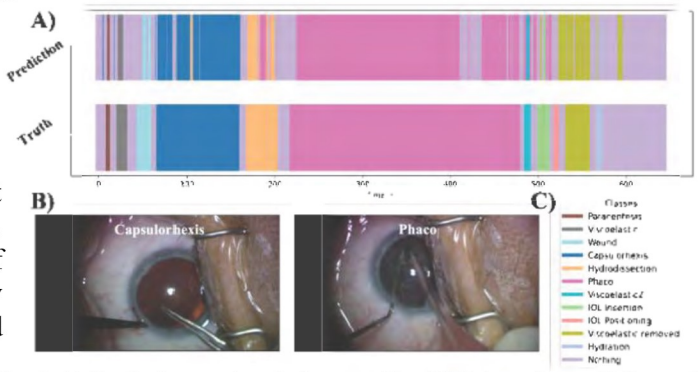

Fig. 2. A) Predictions and truth for one video B) Select phases C) Legend

Results. The model achieved an accuracy of 0.83±0.06, an F-score of 0.73±0.10, and a Jaccard Index of 0.62±0.10 (Fig. 1). Fig. 2A demonstrates the model's predictions for a single video. The model performed best on the Phacoemulsification and Capsulorhexis phases with F-scores of 0.95±0.05 and 0.89±0.09, respectively (Fig. 2B). Performance was poorest on Inserted Lens Positioning (F-score:0.49±0.41) and Hydration (F-score:0.50±0.28).

Conclusion. Phase recognition in cataract surgery is a challenging task due to factors like tool similarity across phases, differing surgeon technique and experience level, and ocular movements. The model performed well overall. It performed best on Phacoemulsification, the lengthiest and most technically challenging surgical phase in which the native lens is emulsified and resorbed. There are two instruments with large movement vectors in this phase, which makes is easier to differentiate from other phases. Capsulorhexis, which the model also performed well on, is similarly distinct. The model struggled with shorter phases with instruments that resemble those used in other phases, such as Hydration and Lens Positioning. Interestingly, there are instances where the model detects "Nothing" during other surgical phases when no instruments are in the field of view. Future work involves training the model on a larger dataset, comparing different architectures, and evaluating the performance against inter-rater variability. Eventually, the model can be used in surgical training by providing feedback and objective performance scores.

References. 1 Levin, M. *et al. J Surg Educ* 76, 1629–1639 (2019)

2 Demir, K. C. *et al. IEEE J Biomedical and Health Informatics* 27, 5405–5417 (2023)