A Computational Framework for Intraoperative Pupil Analysis in Cataract Surgery

Binh Duong Giap, PhD,[1] *Karthik Srinivasan, MD, MS,*[2] *Ossama Mahmoud, MD,*[1,3] *Dena Ballouz, MD,*[1]
Jefferson Lustre, BS,[1] *Keely Likosky, BS,*[1] *Shahzad I. Mian, MD,*[1] *Bradford L. Tannen, MD, JD,*[1]
Nambi Nallasamy, MD[1,4]

Purpose: Pupillary instability is a known risk factor for complications in cataract surgery. This study aims to develop and validate an innovative and reliable computational framework for the automated assessment of pupil morphologic changes during the various phases of cataract surgery.

Design: Retrospective surgical video analysis.

Subjects: Two hundred forty complete surgical video recordings, among which 190 surgeries were conducted without the use of pupil expansion devices (PEDs) and 50 were performed with the use of a PED.

Methods: The proposed framework consists of 3 stages: feature extraction, deep learning (DL)-based anatomy recognition, and obstruction (OB) detection/compensation. In the first stage, surgical video frames undergo noise reduction using a tensor-based wavelet feature extraction method. In the second stage, DL-based segmentation models are trained and employed to segment the pupil, limbus, and palpebral fissure. In the third stage, obstructed visualization of the pupil is detected and compensated for using a DL-based algorithm. A dataset of 5700 intraoperative video frames across 190 cataract surgeries in the BigCat database was collected for validating algorithm performance.

Main Outcome Measures: The pupil analysis framework was assessed on the basis of segmentation performance for both obstructed and unobstructed pupils. Classification performance of models utilizing the segmented pupil time series to predict surgeon use of a PED was also assessed.

Results: An architecture based on the Feature Pyramid Network model with Visual Geometry Group 16 backbone integrated with the adaptive wavelet tensor feature extraction feature extraction method demonstrated the highest performance in anatomy segmentation, with Dice coefficient of 96.52%. Incorporation of an OB compensation algorithm improved performance further (Dice 96.82%). Downstream analysis of framework output enabled the development of a Support Vector Machine—based classifier that could predict surgeon usage of a PED prior to its placement with 96.67% accuracy and area under the curve of 99.44%.

Conclusions: The experimental results demonstrate that the proposed framework (1) provides high accuracy in pupil analysis compared with human-annotated ground truth, (2) substantially outperforms isolated use of a DL segmentation model, and (3) can enable downstream analytics with clinically valuable predictive capacity.

Financial Disclosures: Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2025;5:100597 © 2024 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (*http://creativecommons.org/licenses/by-nc-nd/4.0/*).*

Cataract surgery is one of the most commonly performed surgeries worldwide and is essential to addressing preventable blindness. In 2015, there were more than 20 million surgeries performed worldwide, of which 3.6 million cases were in the United States and 4.2 million cases were in the European Union.[1] Performing cataract surgery requires access to the crystalline lens, which in turn requires adequate dilation and stability of the pupil. The pupil is typically pharmacologically dilated for cataract surgery through the use of medications administered preoperatively and/or intraoperatively. Most active surgical maneuvers in cataract surgery take place within the boundaries of the pupil because the cataract is anatomically located behind the iris, and access to the cataract requires passage of instruments through the pupil. Medications, operating microscope illumination, and surgical maneuvers can alter the shape, size, and appearance of the pupil during different surgical phases.

A typical cataract surgery can be broken into 11 active surgical phases: Paracentesis, Medication and Viscoelastic Insertion, Main Wound, Capsulorrhexis Initiation, Capsulorrhexis Formation, Hydrodissection, Phacoemulsification, Cortical Removal, Lens Insertion, Viscoelastic Removal, and Wound Closure.[2] These surgical steps involve different instrumentation and result in varying appearances of the pupil intraoperatively.

In the past decades, many studies have been proposed to investigate changes in the pupil related to cataract surgery.[3-9] Ordiñaga-Monreal et al recently investigated pupil diameters of 109 randomized eyes preoperatively and 3 months postoperatively using pupillometer software of the Topolyzer Vario.[5] This group found that pupil size was reduced after cataract surgery but also saw that the reduction was larger in men than in women. Ba-Ali et al sought to evaluate the postoperative trajectory of pupil changes for patients undergoing cataract surgery.[7] Maximal pupil diameter reduction was observed 3 weeks postoperatively but subsequently recovered 3 months after surgery. The relation of cataract surgery to pupil size was also investigated by Rickmann et al.[8] In that study, the pupil size of healthy participants was measured with the infrared-video PupilX pupillometer at different illumination levels before and after cataract surgery. This group also saw that pupil diameter decreased after cataract surgery but increased back to preoperative levels 4 weeks after cataract surgery.

Prior work examining pupillary changes related to cataract surgery has primarily focused on changes between *preoperative* and *postoperative* measurements of the pupil, ignoring *intraoperative* changes. Furthermore, most studies of surgery-related pupillary changes have relied on specialized pupillometry hardware not suitable for intraoperative use. Even when intraoperative pupil measurements have been obtained, they have been obtained manually and for very few time points.[9]

Previous studies demonstrated that an adequately dilated pupil is a prerequisite for safe cataract extraction.[10-13] Vision-threatening complications of cataract surgery have been shown to be associated with pupillary instability.[14] Intraoperative floppy iris syndrome (IFIS) is associated with an increased risk of severe complications (odds ratio = 2.82), including posterior capsular rupture.[15] Although IFIS was first reported in association with use of alpha-antagonist drugs such as tamsulosin,[16,17] 77% of IFIS cases are not associated with alpha-antagonists, making it difficult for surgeons to prepare for pupillary instability. Large-scale studies involving the automated intraoperative tracking of pupillary morphology along with patient clinical history and medication lists could aid in the identification of additional medications, systemic conditions, and intraoperative findings that may indicate increased risk for IFIS, without the need for manual measurements that disrupt surgical flow and increase surgical time.

Pupillary dynamics are important beyond cataract surgery, and they affect the execution of vitreoretinal and corneal surgeries as well. In Descemet membrane endothelial keratoplasty (DMEK), for example, the timing of pupillary dilation and miosis plays an important role. Early dilation assists in retroillumination of the endothelium (facilitating descemetorrhexis), whereas later miosis is vital for protecting graft endothelial cells. Thus, although the context of the present study is cataract surgery, it is likely that applications for intraoperative pupil analysis exist for other domains of ophthalmic surgery as well. For example, a pupil analysis system could be applied in DMEK triple procedures to study regimens of mydriatic and miotic medications to achieve the optimal combination of dilation for phacoemulsification and Descemet stripping and subsequent miosis for DMEK graft injection and positioning. Similarly, a pupil analysis system could be utilized to identify patients undergoing retinal surgery who may require pupil expansion devices (PEDs) to maintain adequate visualization throughout surgery prior to decompensation of the surgical view.[18,19]

In this study, we propose and validate a novel computational framework to track and analyze changes in pupil morphology *during* cataract surgery. The proposed framework consists of three primary parts: feature extraction, anatomy segmentation, and obstruction (OB) detection/compensation. This approach is designed to address the primary difficulties encountered in a standard system for recognizing pupils. These challenges include (1) the potential for interference or OB caused by surgical instruments, eyelids, drapes, and similar structures; (2) the cropping of the pupil due to decentration in the camera sensor's field of view; and (3) variations in magnification. It is hoped that this framework will enable the large-scale study of pupillary changes in not only cataract surgery, but ophthalmic surgery in general. Such studies will be necessary to identify clinical and intraoperative risk factors for pupillary instability as well as to develop intraoperative early warning systems for surgeons.

Methods

BigCat Dataset Collection

One hundred ninety high-definition video recordings of cataract surgeries performed by surgeons at University of Michigan's Kellogg Eye Center were collected in the period from 2020 to 2023. The study was approved (HUM00160950) by the Michigan Medicine IRB (IRBMED) in May 2019. To obtain the high-quality videos in the dataset, Zeiss high-definition 1-chip imaging sensors, which were integrated into Zeiss Lumera 700 microscopes, were used for recording the cataract surgeries at 1920 × 1080 resolution and a frame rate of 30 frames per second (FPS). In addition, horizontal white-to-white distances were measured in millimeters preoperatively for all eyes in the dataset using Lenstar LS 900 optical biometers (Haag-Streit, EyeSuite software V.i9.1.0.0). Periods of inactivity prior to surgery and after completion of surgery were trimmed. Frames were extracted from the video stream at a frequency of 15 FPS. This strategic down-sampling was implemented to reduce the overall time required for subsequent analyses while retaining essential temporal information. The extracted frames were resized to dimensions of 480 × 270 pixels using bilinear interpolation to maintain sufficient video frame quality while minimizing time consumption.

To generate a dataset for the development and testing of models for anatomy recognition and analysis, the surgical videos were then processed to extract 23 random frames from each of 11 surgical phases. Ground truth surgical phase annotations were performed by trained human annotators for all frames within all 190 videos. Randomized selection of frames by phase was performed to ensure that pupil obscuration by a variety of surgical instruments associated with different phases of surgery was adequately represented in the dataset. In addition, because the Medication and Viscoelastic Insertion, Phacoemulsification, and Viscoelastic Removal phases have highly varied appearances (due to transient pupil distortion from viscoelastic instillation, nucleus disassembly, and eye

rotation, respectively), an additional 2 frames per phase were randomly selected for these 3 phases. We further expanded the dataset by incorporating 2 additional frames from each video during which no active surgical maneuvers were being performed, termed a "No Activity" phase. The resulting dataset consisted of 5700 frames that were resized to 480 × 270 pixels and stored in 24-bit Portable Network Graphics format without editing. The number of images corresponding to each surgical phase in the collected dataset is shown in Table 1.

Ground truth manual segmentations of anatomical components of the eye—the palpebral fissure, limbus, and pupil—were performed manually on all images in the dataset by trained human annotators. This annotation process was conducted using MAT-LAB version R2022a (The MathWorks) in conjunction with a Wacom One drawing tablet (Wacom Co, Ltd) to ensure the accuracy of the annotations. The videos within the dataset were randomly divided into training, validation, and testing subsets consisting of 60%, 20%, and 20% of the dataset, respectively. As a result, the training set consisted of 3420 images from 114 surgical videos, the validation set comprised 1140 images from 38 videos, and the remaining 1140 images from 38 videos were allocated for the testing set. This data splitting strategy aims to ensure a balanced distribution of data for training, validation, and testing, resulting in improved model reliability.

Data Preprocessing and Feature Extraction

Significant variations in pupil appearance can occur during cataract surgery due to the presence of surgical instruments, hydration of the crystalline lens material, nuclear disassembly, and eventual replacement of the crystalline lens with the intraocular lens implant. Accordingly, even DL models can benefit from the utilization of tailored feature extraction methods for the task of semantic segmentation.[20] In order to overcome these challenges, we propose the incorporation of an image preprocessing method in this phase to extract meaningful image features and attempt to improve downstream segmentation performance. A schematic of the overall framework is depicted in Figure 1.

In this study, we employ the adaptive wavelet tensor feature extraction (AWTFE) method,[21] described previously by our research group, for feature extraction. The primary objectives in employing the AWTFE method at this stage are (1) to eliminate irrelevant information in the context of segmentation tasks and (2) to extract and enhance the distinctive features of the pupil, limbus, and palpebral fissure within the image. Given the high variance within captured video frames, the AWTFE method emphasizes object boundaries, textures, shapes, and other

distinctive attributes relevant to the segmentation of the anatomical landmarks of interest. Sample output of the AWTFE method is depicted in Figure 2.

The AWTFE method was originally proposed for pupil region feature extraction to improve the accuracy of DL-based pupil segmentation models. Specifically, based on tensor theory and the wavelet transform, we first represent the correlations among spatial information, color channels, and wavelet subbands of a video frame by constructing a third-order tensor. We then utilize higher-order singular value decomposition to adaptively eliminate redundant information and estimate pupil feature information. Using the AWTFE method, features relevant to the pupil region are identified, significantly improving the performance of DL-based segmentation models. In a previous study,[19] we conducted additional experiments to demonstrate that the AWTFE method can be extended to effectively extract and highlight features of other regions, such as the iris. The impact of the AWTFE method on anatomy segmentation performance is described in the following section.

DL-based Anatomy Recognition

In the proposed framework, the accuracy of the analysis results heavily relies on anatomy segmentation. In this stage of the framework, the pupil, limbus, and palpebral fissure, are segmented within the video frames. Segmentation of the limbus and palpebral fissure was necessary for OB detection and compensation, as described in the next section. Three state-of-the-art DL-based segmentation models (UNet,[22] LinkNet,[23] and Feature Pyramid Network [FPN][24]) and 4 distinct convolutional backbone networks (Visual Geometry Group 16 [VGG16],[25] ResNet50,[26] DenseNet169,[27] and MobileNet[28]) were considered to select an optimal combination for the anatomy segmentation task. Accordingly, 12 model-backbone combinations were studied in total. Backbone networks were each pretrained on the ImageNet dataset.[29] Two instances of each model−backbone combination were considered, 1 utilizing the AWTFE method for preprocessing of images, and 1 using raw input images. Accordingly, 24 models were studied in total. The implementations of the segmentation models can be found in https://github.com/qubvel/segmentation_models.[30]

The training set sizes were equivalent for all 24 models studied, with or without utilizing the AWTFE method. Training was performed with batch size of 32. All images were downsampled to 224 × 224 pixels for the input to the segmentation models. The images were preprocessed by subtracting the mean Red, Green, and Blue values, which were computed on the training set, from each pixel. The Adam algorithm with an initial learning rate of 0.0001 was used as the optimizer for the segmentation models. The learning rate was then decreased by 0.1 when the validation loss stopped improving. Each model was trained for a maximum of 200 epochs, corresponding to 21 400 training iterations. Early stopping was implemented to avoid overfitting. To reduce overfitting and enhance the generalizability of the model, data augmentation was utilized for training all models in this study. In particular, random cropping, flipping, and changing of Red, Green, and Blue color channel intensity were applied for the data augmentation. The trained weights achieving the lowest validation loss during the training process were saved and utilized for validation.

In this study, the performance of the segmentation models was evaluated with precision and recall as the primary metrics, which are determined as follows:

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

Table 1. Number of Images Included from BigCat by Surgical Phase

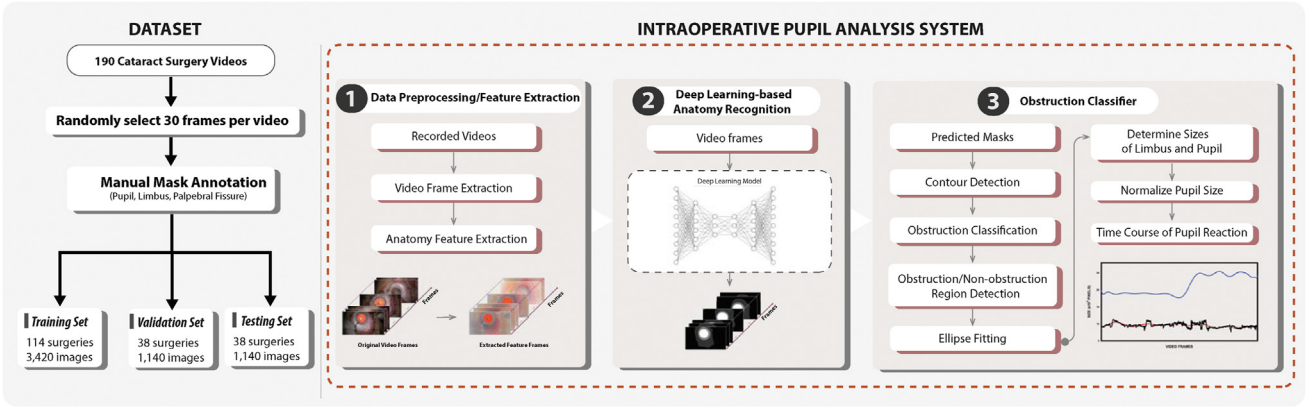| Surgical Phases	Number of Images
No Activity	380
Paracentesis	380
Medication and Viscoelastic Insertion	760
Main Wound	380
Capsulorrhexis Initiation	380
Capsulorrhexis Formation	380
Hydrodissection	380
Phacoemulsification	760
Cortical Removal	380
Lens Insertion	380
Viscoelastic Removal	760
Wound Closure	380

Fig. 1. The proposed intraoperative pupil analysis framework for cataract surgery.

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

where *TP*, *FP*, *TN*, and *FN* denote the true positive, false positive, true negative, and false negative rates, respectively. In addition, the performance of the models was also evaluated by using the Intersection Over Union and Dice Coefficient, which are determined as follows:

$$IoU = \frac{TP}{TP + FP + FN}, \quad (3)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}. \quad (4)$$

Obstruction Detection and Compensation

The quantitative assessment of intraoperative pupil morphologic changes in cataract surgery presents challenges despite the ability to utilize DL models for semantic segmentation of the pupil region. In particular, 3 primary challenges are (1) the potential for interference or OB caused by surgical instruments, eyelids, drapes, and similar structures, (2) the cropping of the pupil due to decentration
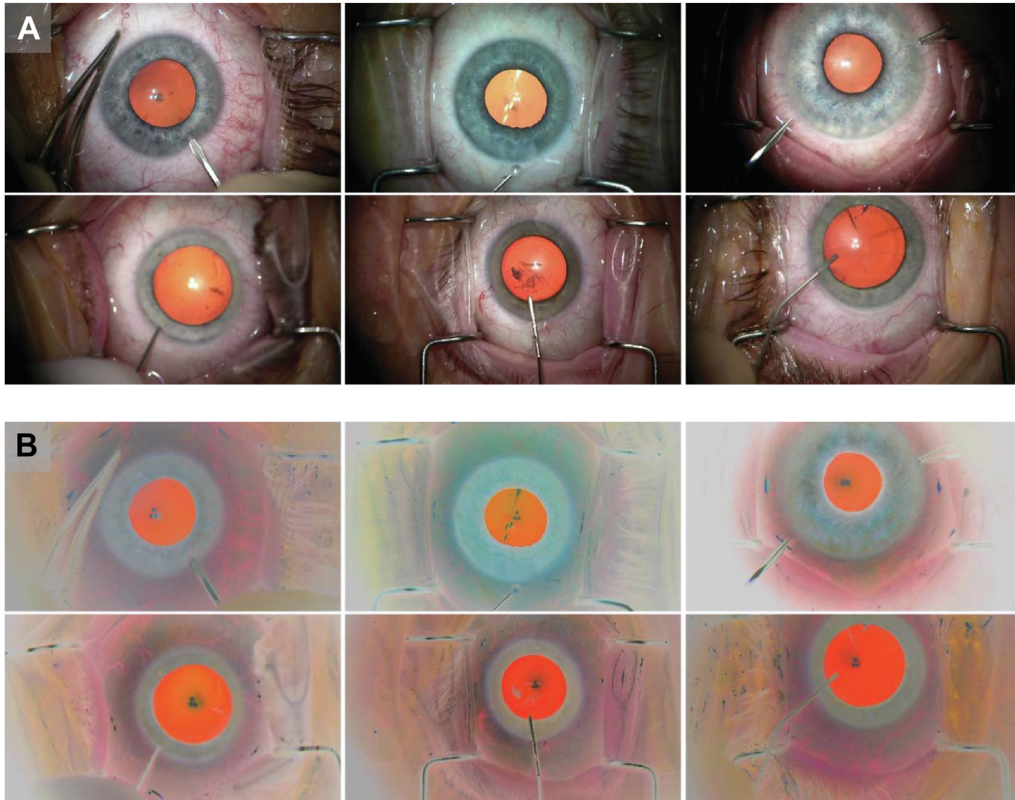

Fig. 2. Anatomic feature extraction using the AWTFE method. **A,** Original cataract surgery images in the dataset. **B,** Corresponding feature-extracted images generated by the AWTFE method. AWTFE = adaptive wavelet tensor feature extraction.

in the camera sensor's field of view, and (3) variations in magnification, as shown in Figure 3A, B. To address those challenges, our proposed framework introduces an Obstruction Detection and Compensation algorithm. This algorithm relies on automated segmentations of the pupil, palpebral fissure, and limbus generated during the segmentation phase to identify and compensate for OBs. Furthermore, to ensure accuracy across diverse magnification settings, the computed pupil size is ultimately normalized by utilizing the predicted limbus size.

The OB classifier utilizes a pair of masks generated by the DL models for the palpebral fissure and pupil, denoted as M_q and M_p, respectively, to ascertain the presence of OB within a given frame. Initially, contours of both the palpebral fissure and the pupil are determined by binarizing the respective masks and applying edge computation through a Canny filter,[31] as illustrated in Figure 4. Subsequently, points $q_i(x_q, y_q)$ and $p_j(x_p, y_p)$ along the contours of the palpebral fissure and pupil are vectorized to enable the calculation of the Euclidean distance, $D_{i,j}$, between these 2 vectors as follows:

$$D_{i,j} = \sqrt{(x_q \text{-} x_p)^2 + (y_q \text{-} y_p)^2} \qquad (5)$$

The smallest computed Euclidean distance value is then compared against an OB threshold, τ. If the distance is larger than τ, the pupil region is devoid of OB, and the pupil size can be simply determined via the original pupil mask M_p generated by the DL model. In contrast, if the pupil is identified as an obstructed region based on the OB threshold τ, the subsequent steps involve determining obstructed and unobstructed segments. This is accomplished by calculating the minimum distance between each pupil point $p_j(x_p, y_p)$ and all palpebral fissure points $q_i(x_q, y_q)$. If the minimum value is larger than τ, $p_j(x_p, y_p)$ is treated as an unobstructed point, and vice versa. Subsequently, an ellipse, which is considered an estimate of the pupil region, is fitted using the unobstructed points.[32] By considering solely the unobstructed pupil points, the estimated ellipse eliminates noise caused by obstructed points on the pupillary boundary. To this end, the estimated pupil size P can be established by computing the area of the fitted ellipse. The results of this process are shown in Figure 3C.

The magnification of images within a recorded video may vary due to operating microscope adjustments made by surgeons during surgery. To address this challenge, the pupil size, which was previously estimated as shown, was normalized using the corresponding size of the (OB-compensated) limbus region. In contrast to the pupil, the size of the limbus region is physically fixed and cannot be significantly altered under normal circumstances during phacoemulsification. However, the size of the segmented limbus can also be affected by inaccuracies resulting from surgical OBs. Therefore, to ensure the accuracy of the system, the actual size of the limbus, denoted as L, is first estimated by the OB compensation algorithm described here using the predicted masks of the palpebral fissure and limbus. The size of the pupil region at k^{th} frame P_k is then normalized with respect to the size of the corresponding limbus region L_k as follows:

$$P_k^{norm} = P_k \times \frac{L_0}{L_k}, \qquad (6)$$

where L_0 is determined as the limbus region computed directly from the predicted mask of the first frame by the DL model. This ensures that the computed size of the pupil region at k^{th} frame P_k is adjusted for any magnification changes made intraoperatively. To enable evaluation of the pupil size in absolute terms, the size of P_k^{norm} was finally converted to millimeters using each eye's horizontal white-to-white distance, which was measured preoperatively for all eyes in the dataset.

In order to assess the effectiveness of this approach, we evaluated the performance of the proposed framework in its capability to classify obstructed frames and estimate the actual size of the pupil using 1140 images selected randomly from 38 videos in the test set. The images were manually classified into pupil *OB* and nonobstruction (*NOB*) classes by 2 trained annotators. Accordingly, 22.45% of the testing set (256 images) belonged to the *OB* class and 77.55% of the testing set (884 images) were in *NOB* class. Furthermore, the actual pupil region, including the obstructed region not visible to the camera in the images of the OB class, was manually estimated and annotated to be used as the ground truth for the experiment.

Downstream Analysis Example: Prediction of Pupil Expansion Device Use

In order to examine the utility of the output of the pupil analysis framework described here, we investigated whether the timecourse of pupillary area was predictive of surgeon usage of a PED in a separate dataset.

The use of a PED by an experienced surgeon is an indication that the surgeon detects pupillary characteristics that may impair successful completion of the surgery. A PED is preferably placed prior to the creation of the anterior capsulotomy so as to avoid inadvertently capturing the capsulotomy edge with the PED. Accordingly, we investigated whether the pupil area time series generated by our framework for the surgical phases prior to the initiation of the capsulorrhexis (excluding PED placement in those cases that involved PED placement) could be used to predict whether a PED would later be placed.

We first constructed a dataset that included the mean pupil sizes during 3 early surgical phases of cataract surgery: Paracentesis, Medication and Viscoelastic Insertion, and Main Wound, across 50 cases with PED placement and 50 cases without PED placement. It is important to note that all 100 videos utilized for the analysis were completely separate from the training set of the BigCat dataset. In order to identify the surgical phase within PED videos, we employed the CatStep surgical phase classification model.[25] It is noted that the CatStep model utilized was not previously trained on PED videos. Hence, we conducted manual validation of its phase classification results to ensure the accuracy of phase boundaries in the PED videos. The videos in the dataset were randomly divided into training (70%) and testing (30%) sets, maintaining the balance of classes in the training and testing sets. We investigated the capacity of classification models, including Support Vector Machine (SVM),[33] K-Nearest Neighbors,[34] Random Forest,[35] Decision Tree,[36] Naïve Bayes,[37] and Logistic Regression,[38] to predict PED usage based on the averaged pupil size time series. All experiments in this study were implemented on a workstation with a 24-Core Xeon Intel CPU, with 128 GB RAM and 4 NVIDIA RTX 2080 Ti GPUs running Ubuntu.

Results

In this section, the performance of the proposed framework is presented in detail for each of its components. Additionally, we present an analysis of pupillary changes by phase of surgery. To demonstrate the potential for downstream analyses utilizing the pupil analysis framework, we also evaluate the performance of algorithms for predicting PED use based on the pupillary size timecourse derived from the framework.
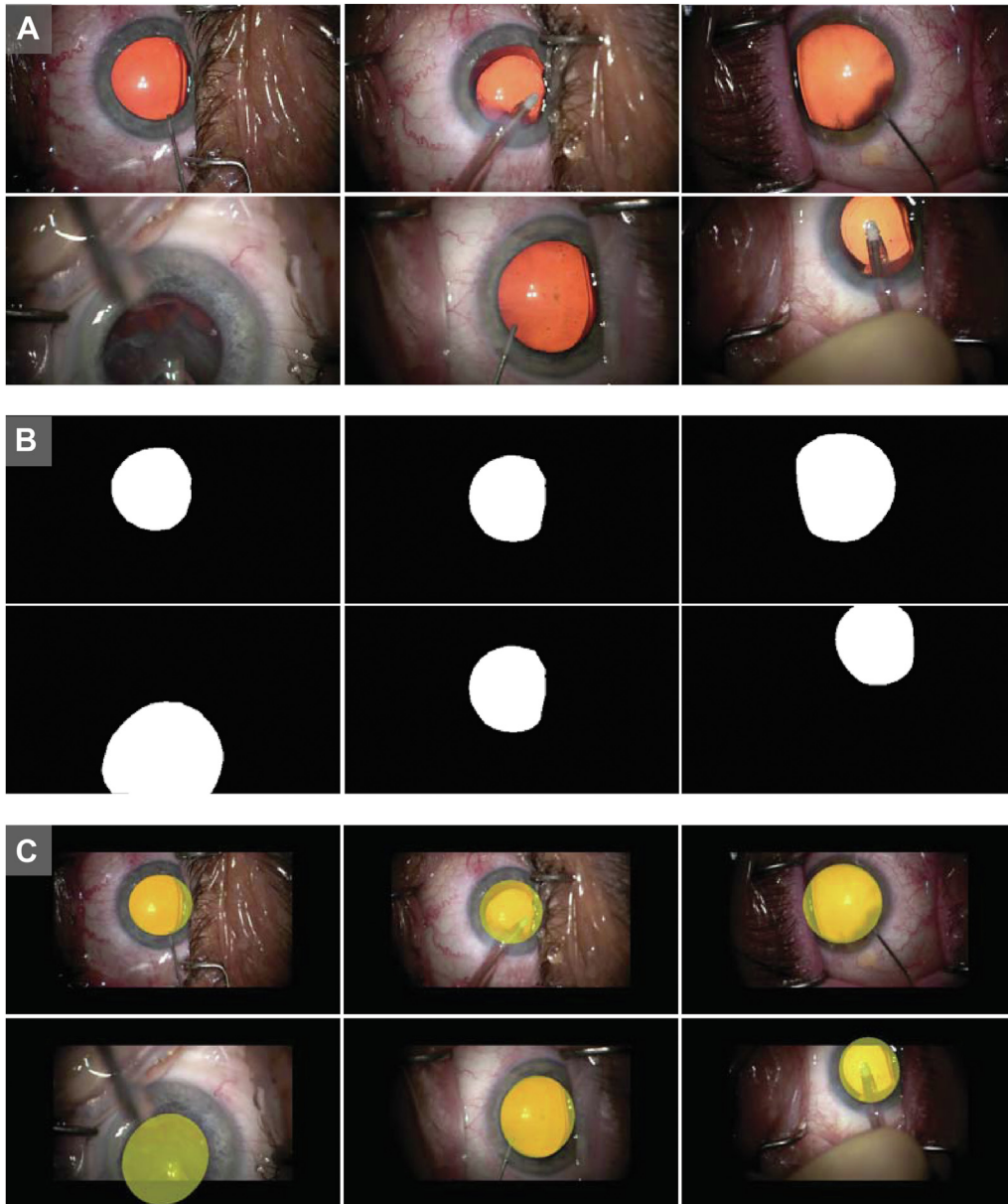
Fig. 3. Sample obstruction compensation results. **A,** Obstruction and decentration video frames. **B,** Pupil segmentation masks generated by the deep learning model. **C,** Overlay images of pupil masks generated by the obstruction compensation algorithm.

Feature Extraction and Anatomy Segmentation

The anatomy segmentation performance of the DL models considered is detailed in Table 2. The FPN architecture combined with the VGG16 backbone demonstrated the highest mean Dice coefficient (95.03%) across the 3 anatomic segmentation classes (pupil, limbus, and palpebral fissure) compared with all other model−backbone combinations.

The performance of every model-backbone combination was improved through the addition of feature extraction using the AWTFE method ($P < 0.0001$). Mean Dice coefficients across the 3 segmentation classes of all investigated models were improved by up to 2.23%. The FPN-VGG16 architecture with AWTFE feature extraction outperformed all other models considered, with a Dice coefficient of 96.52%. The FPN-VGG16+AWTFE model outperformed the original FPN-VGG16 network by 1.49%. Accordingly, the FPN-VGG16+AWTFE model was selected for incorporation into the pupil analysis framework.

Obstruction Detection and Compensation Performance

In order to optimize the performance of the obstruction detection component of the proposed framework, we first examined the impact of threshold values, τ_P, on the accuracy of the obstruction classifier for pupil on the 1140
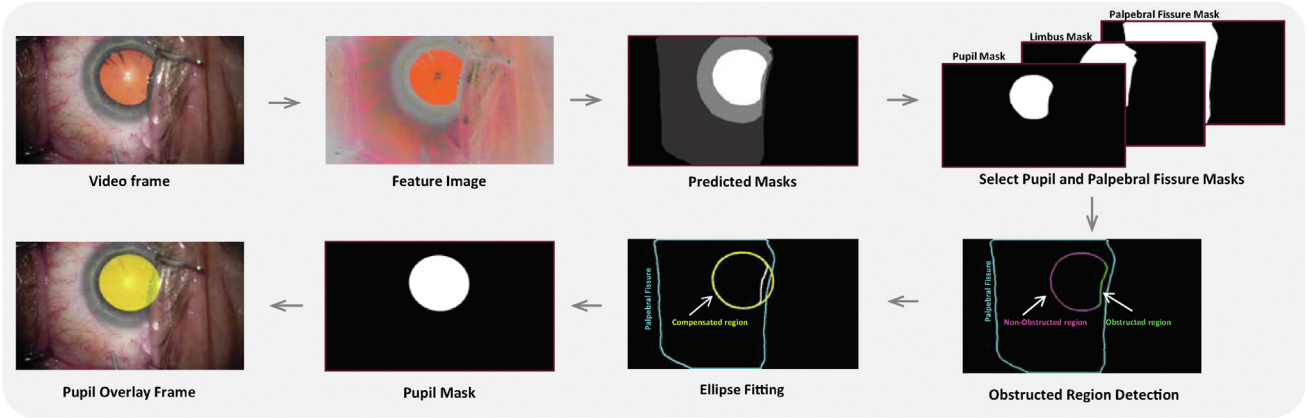
Fig. 4. Schematic depicting the obstruction classifier embedded in the proposed pupil analysis framework.

images in the validation set. As shown in Figure 5, the highest classification performance (*Dice*) was obtained with $\tau = 8$. With this value of τ, the edge detection−based classifier achieved a Dice coefficient of 79.12%. The classification performance gradually decreased with increasing values of τ. By using the proposed framework with $\tau = 8$, with obstruction compensation applied as described in the Methods section, a Dice coefficient of 96.82% for pupil segmentation was achieved across the 1140 test set images.

Utilization of obstruction detection led to an increase in pupil segmentation performance by 0.51%, as shown in

Figure 6A. The Dice coefficients for all 11 active phases of surgery were significantly higher with obstruction detection and compensation than without obstruction detection ($P = 0.0002$).

Of the 1140 images in the test set, 256 (22.45%) had obstructions (*OB* class), whereas the remaining 884 did not (*NOB* class). To evaluate more directly the impact of the obstruction detection and compensation system when obstructions were present, performance was assessed further on the *OB* subset. The framework with obstruction detection and compensation achieved a Dice coefficient of 93.12%, whereas the framework without obstruction detection

Table 2. Segmentation Performance of Deep Learning Models With and Without the AWTFE method on the Validation set

Network Architecture	Backbone Network	Precision (%)	Recall (%)	IoU (%)	Dice (%)
UNet	MobileNet	94.09	94.69	89.52	94.15
UNet + AWTFE		**95.86**	**95.89**	**92.28**	**95.66**
LinkNet		93.47	93.90	88.27	93.14
LinkNet + AWTFE		**95.52**	**95.49**	**91.63**	**95.27**
FPN		93.53	93.97	88.37	93.46
FPN + AWTFE		**95.67**	**96.33**	**92.45**	**95.79**
UNet	DenseNet169	94.72	94.92	90.28	94.63
UNet + AWTFE		**96.47**	**96.57**	**93.45**	**96.39**
LinkNet		94.43	94.60	89.74	94.31
LinkNet + AWTFE		**96.08**	**96.44**	**92.99**	**96.13**
FPN		94.20	94.78	89.68	94.28
FPN + AWTFE		**96.39**	**96.49**	**93.26**	**96.31**
UNet	ResNet50	94.86	95.12	90.57	94.77
UNet + AWTFE		**96.64**	**96.49**	**93.49**	**96.41**
LinkNet		94.35	94.68	89.71	94.28
LinkNet + AWTFE		**95.99**	**96.63**	**93.04**	**96.14**
FPN		94.61	94.79	90.06	94.47
FPN + AWTFE		**96.36**	**95.99**	**92.79**	**95.98**
UNet	VGG16	94.96	95.04	90.58	94.79
UNet + AWTFE		**96.63**	**96.03**	**93.11**	**96.18**
LinkNet		94.25	94.65	89.63	94.19
LinkNet + AWTFE		**96.33**	**96.07**	**92.85**	**95.99**
FPN		94.99	95.45	91.00	95.03
FPN + AWTFE		**96.41**	**96.87**	**93.66**	**96.52**

Bold fonts indicate the better performance across models.
AWTFE = adaptive wavelet tensor feature extraction; FPN = Feature Pyramid Network; IoU = Intersection Over Union; VGG16 = Visual Geometry Group 16.
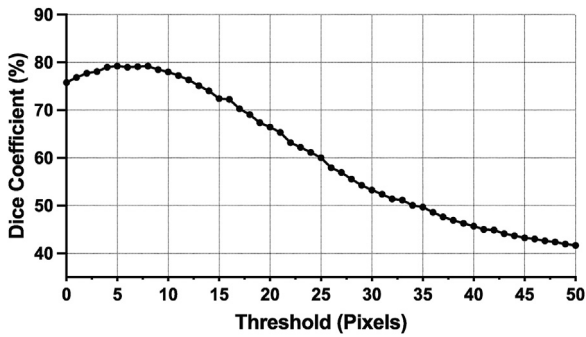
Fig. 5. Classification performance of the obstruction classifier on 1140 images of 38 surgical videos.

yielded only 90.88% ($P = 0.0014$). Phase-wise performance differences in the OB subset are depicted in Figure 6B.

The processing time required by the proposed analysis system for an obstructed frame was approximately 19.10 milliseconds (ms) from start to finish. Of the 19.10 ms, the obstruction detection and compensation component required only 4.30 ms. These findings reveal that the proposed system is well-suited for real-time intraoperative applications, with a throughput of over 52 FPS while maintaining the accuracy described here.

Phase-Based Pupil Reaction Analysis

To analyze phase-based changes in pupil size, we randomly selected 15 videos from the testing set (none determined to have IFIS and none requiring a PED) and executed the framework on the entire videos. The pupil size timecourses for 4 surgical cases analyzed using the proposed framework are depicted in Figure 7. The mean pupil sizes within the 11 active surgical phases relative to the initial pupil size determined at the beginning of each surgery are shown in

Figure 8. An increase in pupil size is seen following the Medication Injection (buffered lidocaine and epinephrine) and Viscoelastic Insertion. During this phase, the pupil size increased by an average of 9.64% compared with the initial pupil size.

The Main Wound phase was consistently followed by a reduction in pupil size, a pattern evident in all trajectories plotted in Figure 7. This phase shows the highest mean pupil size, which then gradually decreased in subsequent phases, including Capsulorrhexis Initiation, Capsulorrhexis Formation, and Hydrodissection, as detailed in Figure 8. The reduction in pupil size may be related to loss of viscoelastic through the main wound during these phases. Ultimately, the pupil size post−Wound Closure closely approximated the preoperative pupil size (higher by just 0.99% relative to the initial size).

PED Use Prediction Performance

In order to evaluate the utility of the pupil size timecourse output by the proposed framework for downstream analyses, attention was then turned to a dataset of 50 videos with PED placement and 50 videos without PED placement (as described in the Methods section). In comparison with standard surgeries performed without PED, pupil size during the initial phases (Paracentesis, Medication and Viscoelastic Insertion, and Main Wound) of PED surgeries were not significantly different ($P = 0.375$).

Six classification models were trained using pupil size timecourses from the dataset of 50 PED and 50 non-PED videos. Only the timecourses from the Paracentesis, Medication and Viscoelastic Insertion, and Main Wound phases were included for this analysis in order to simulate surgeon decision-making. In our experiment, all classification algorithms were trained on the training set consisting of the pupil size timecourses of 70 surgeries and validated on the testing set, which included 30 surgeries. The performance of the classification models on the testing set is shown in
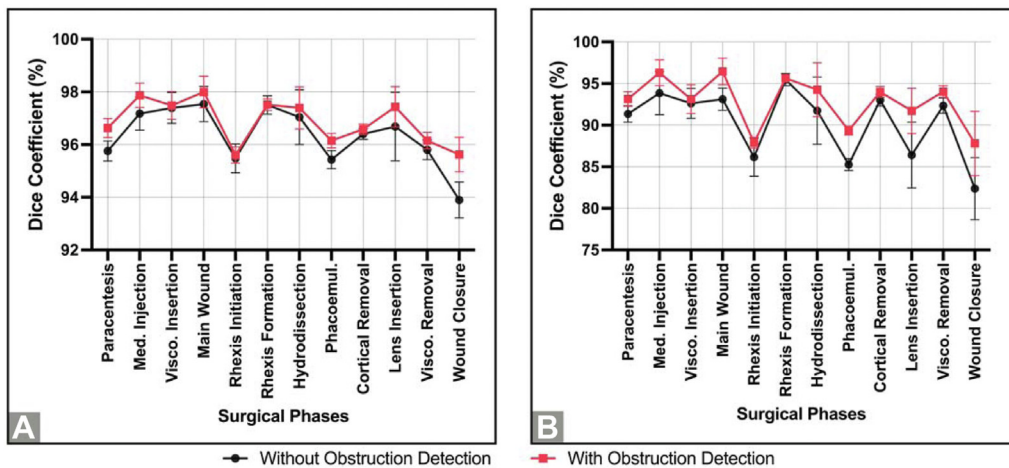

Fig. 6. Dice coefficients by surgical phase for the proposed framework with and without obstruction detection. **A,** The results on all 1140 images in the testing set. **B,** The results on all 256 obstruction images in the testing set. Med. = Medication; Rhexis Formation = Capsulorrhexis Formation; Rhexis Initiation = Capsulorrhexis Initiation; Phacoemul. = Phacoemulsification; Visco. = Viscoelastic.
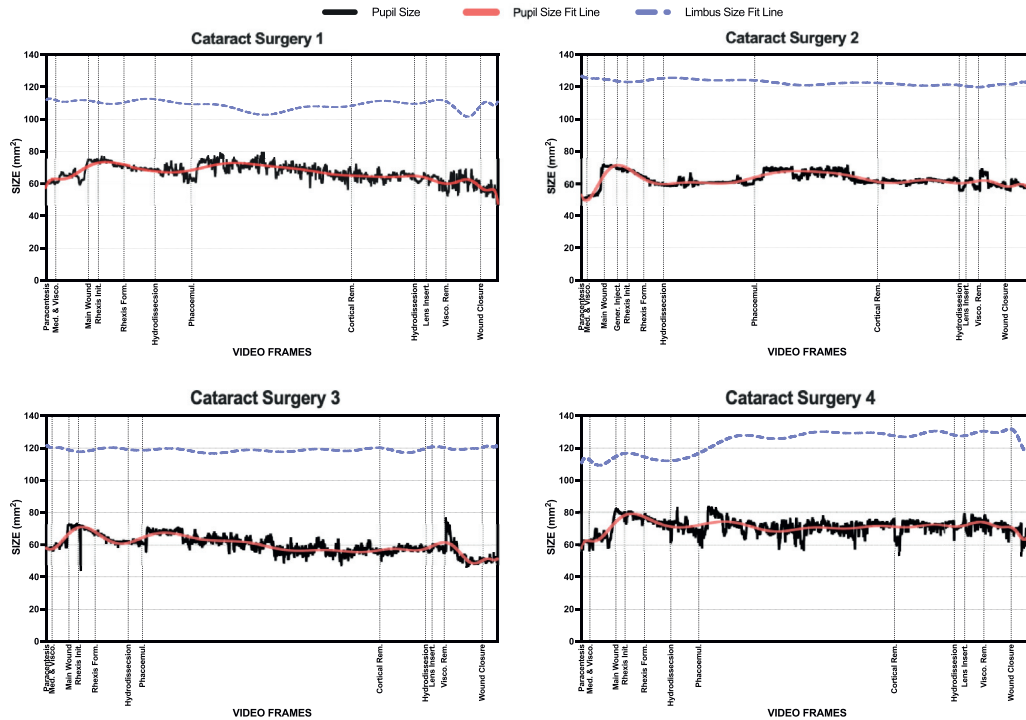
Fig. 7. Pupil and limbus area output from the proposed framework for 4 cataract surgery cases with phases of surgery indicated. Cortical Rem. = Cortical Removal; Med. & Visco. = Medication and Viscoelastic Insertion; Rhexis Form. = Capsulorrhexis Formation; Rhexis Init. = Capsulorrhexis Initiation; Phacoemul. = Phacoemulsification; Visco. Rem. = Viscoelastic Removal.

Table 3 and Figure 9. Among the 6 machine learning models, the Random Forest and SVM models achieved the highest classification accuracy at 96.67% and the highest area under the curve (AUC) of 99.33% and 99.44%, respectively. These experimental results demonstrate that the intraoperative pupil size across early surgical phases can be effectively utilized to predict surgeon PED usage in cataract surgery.

We then employed the mentioned machine learning models trained with preoperative pupil size data alone

instead of using pupil size timecourses. Preoperative pupil size was determined at the initial frame of each surgical video using our proposed system. Our primary objective was to evaluate the effectiveness of using the timecourse generated by our proposed system compared with the traditional approach of relying on preoperative pupil size measurements performed by surgeons in clinic or at the beginning of surgery. The models trained on preoperative pupil size alone achieved lower performance compared with those trained on pupil size timecourses from the early
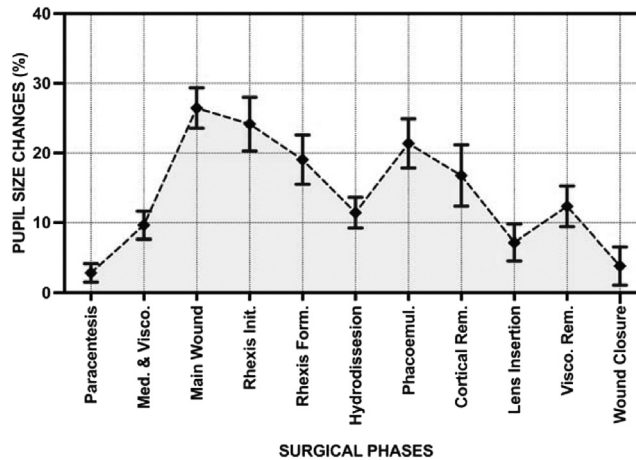

Fig. 8. Mean pupil size by surgical phase relative to preoperative pupil size across 15 surgical videos. Cortical Rem. = Cortical Removal; Med. & Visco. = Medication and Viscoelastic Insertion; Rhexis Form. = Capsulorrhexis Formation; Rhexis Init. = Capsulorrhexis Initiation; Phacoemul. = Phacoemulsification; Visco. Rem. = Viscoelastic Removal.

Table 3. Pupil Expansion Device Usage Prediction Performance of Machine Learning Models

Models	Preoperative Pupil Size		Pupil Size Timecourse	
	Accuracy (%)	*AUC (%)*	*Accuracy (%)*	*AUC (%)*
Naïve Bayes	83.33	87.00	**93.33**	**93.33**
KNN	76.67	87.33	**83.33**	**89.44**
Logistic Regression	83.33	87.22	**90.00**	**97.74**
Decision Tree	73.33	73.33	**93.33**	**93.33**
SVM	83.33	87.00	**96.67**	**99.44**
Random Forest	73.33	86.56	**96.67**	**99.33**

Bold fonts indicate the better performance.
AUC = area under the curve; KNN = K-Nearest Neighbors; SVM = Support Vector Machine.

surgical phases. Notably, Logistic Regression achieved the highest performance among the models trained on preoperative pupil sizes with an accuracy of 83.33% and an AUC of 87.22%, which are significantly lower than the highest performance of the machine learning model achieved on the pupil size timecourses by 13.34% and 12.22%, respectively. These results indicate that prediction of PED usage can be more performed more accurately using timecourses generated by proposed intraoperative pupil analysis framework than using preoperative pupil size alone.

Discussion

In the present study, we have proposed and validated a computational framework for intraoperative analysis of pupil morphology during cataract surgery. The system involves 3 primary stages: feature extraction, anatomy segmentation, and obstruction detection/compensation. In the feature extraction stage, we employed the AWTFE method, previously developed by our group, to generate feature-rich versions of video frames for the effective segmentation of relevant anatomical structures. When combined with state-of-the-art DL-based segmentation models, the AWTFE method significantly improved segmentation performance for all models considered. The FPN model with VGG16 backbone and AWTFE feature extraction (FPN-VGG16+AWTFE) was found to have the highest performance in validation and was ultimately chosen for

incorporation into the framework. The final model achieved a Dice coefficient of 96.52% when evaluated on the held-out testing set.

In other previous studies, the pupil region recognized by DL models is used directly for analysis without any post-processing (obstruction detection and compensation).[39,40] Consequently, this leads to unreliable results, as the pupil and limbus regions are frequently obstructed by other objects during surgery. In order to ensure the reliability of the system in delivering precise intraoperative analysis of pupillary changes, we introduced a novel obstruction detection/compensation component. This innovation addresses typical issues arising from obstructions and varying image scales in cataract surgery videos. Our findings demonstrate that when employing the proposed obstruction detection and compensation algorithm, the system can achieve an overall Dice coefficient of 96.82% for pupil segmentation and can improve segmentation performance by 2.24% for frames in which obstructions were present. Because pupil-obstructed frames comprised 22.45% of the randomly selected test set, obstruction detection and compensation are likely to be essential for ensuring the reliability of downstream time-series analyses of pupillary metrics.

The run-time assessment revealed that the proposed analysis system can fully process an obstructed frame in under 19.10 ms, achieving throughput of over 52 FPS. Because cataract surgeries involve a mix of obstructed and unobstructed frames, 52 FPS represents a lower bound on throughput for the proposed system running on the described hardware. The high throughput (>30 FPS) and high accuracy of the proposed system indicate its potential for future real-time intraoperative use as part of a broader decision support or analysis system.

We utilized the proposed system to examine pupil trajectories through the various phases of cataract surgery. Findings such as (1) the increase in pupil size after medication and viscoelastic injection, (2) the reduction in pupil size during the course of phacoemulsification, and (3) the reduction in pupil size after viscoelastic removal align with typical surgeon experience.

To demonstrate the potential for downstream analysis of the pupil morphology timecourse generated by the proposed framework, we performed additional experiments to assess
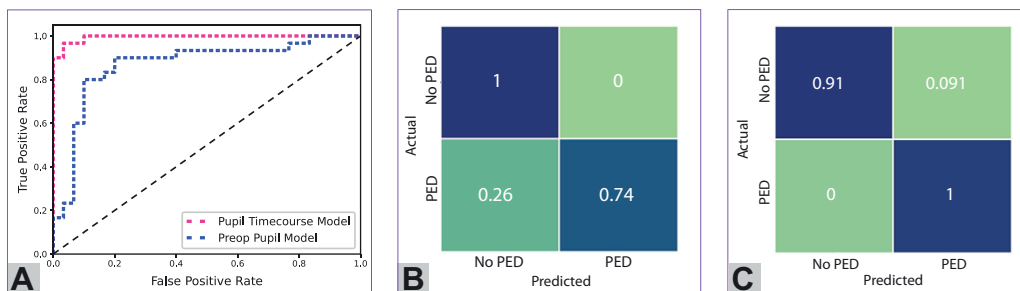
Fig. 9. The performance comparison for prediction of pupil expansion device usage of the classification models using the pupil timecourses and the preoperative pupil size. **A,** The Receiver Characteristic Curves of the pupil timecourse model and the preoperative pupil model (logistic regression). **B,** The confusion matrix of the preoperative pupil model. **C,** The confusion matrix of the pupil timecourse model. PED = pupil expansion device.

the ability to predict PED usage using output from the framework. Using only pupil morphology data in the first 3 phases of surgery (Paracentesis, Medication and Viscoelastic Insertion, and Main Wound), it was possible to create an SVM classifier with 96.67% accuracy and 99.44% AUC in predicting the eventual use of a PED during surgery. The performance of this approach was significantly higher than what could be achieved with preoperative pupil size data alone. The final pupil timecourse-based SVM could serve as part of a decision support system for trainees or early stage attending surgeons and demonstrates the potential for building upon the output of our proposed framework.

Limitations of the study include the utilization of surgical videos from a single institution. However, the AWTFE method has previously been validated on the external Cataract Dataset for Image Segmentation dataset,[41] and the BigCat dataset utilized here is the largest database of deeply annotated surgical video in the world,[42] comprising over 4 million frames in total. Furthermore, the proposed framework is seen as a starting point for intraoperative pupil analysis, as additional challenges remain. These challenges include compensation for ocular rotation, anterior chamber depth, and corneal power, which will be addressed in future studies.

Given the demonstrated performance contributions of each component of the computational framework proposed here, it appears this framework can serve as a reliable foundation for further analyses of intraoperative pupillary changes. We believe that the proposed framework can serve as a research tool for the large-scale examination of intraoperative risk factors for pupillary instability, which is known to be associated with increased rates of complications of cataract surgery. Although the current work has focused on cataract surgery, the framework is likely to be applicable to other pupil-sensitive forms of surgery, including vitreoretinal surgery and corneal surgery. Future work will attempt to validate this framework for other forms of ophthalmic surgery and explore additional downstream analyses and decision support scenarios.

Footnotes and Disclosures

[1] Kellogg Eye Center, Department of Ophthalmology & Visual Sciences, University of Michigan, 1000 Wall Street, Ann Arbor, Michigan, 48105.

[2] Department of Vitreo Retinal, Aravind Eye Hospital, Chennai, Tamil Nadu, 600077, India.

[3] Wayne State University School of Medicine, 540 E Canfield Street, Detroit, Michigan, 48201.

[4] Department of Computational Medicine & Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, Michigan, 48109.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The authors have made the following disclosures:

S.I.M.: Grants − KOWA; Royalties − UptoDate; Travel expenses − Eversight Eye Bank; Patents planned, issued or pending − US 63/621,208; Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid − Eye Bank Association of America, Cornea Society, Michigan Society of Eye Physicians and Surgeons, Eversight Eye bank.

B.L.T.: Patents planned, issued or pending − US 63/621,208; Others − Bausch and Lomb Americas Inc. (Food and Beverage).

N.N.: Patents planned, issued or pending − US 63/621,208; Participation on a Data Safety Monitoring Board or Advisory Board − Recordati Rare Diseases SARL.

HUMAN SUBJECTS: No human subjects were included in this study. This study received full ethics approval by the Michigan Medicine Institutional Review Board (HUM00160950), and it was determined that informed consent was not required because of its retrospective nature and the anonymized data utilized in this study. The study was carried out in accordance with the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Giap, Srinivasan, Mahmoud, Ballouz, Mian, Tannen, Nallasamy

Data collection: Giap, Srinivasan, Mahmoud, Ballouz, Lustre, Likosky, Mian, Tannen, Nallasamy

Analysis and interpretation: Giap, Srinivasan, Mahmoud, Ballouz, Lustre, Likosky, Mian, Tannen, Nallasamy

Obtained funding: Tannen, Nallasamy

Overall responsibility: Giap, Ballouz, Mian, Tannen, Nallasamy

Abbreviations and Acronyms:

AUC = area under the curve; **AWTFE** = adaptive wavelet tensor feature extraction; **DL** = deep learning; **DMEK** = Descemet membrane endothelial keratoplasty; **FPN** = Feature Pyramid Network; **FPS** = frames per second; **IFIS** = intraoperative floppy iris syndrome; **ms** = milliseconds; **NOB** = nonobstruction; **OB** = obstruction; **PED** = pupil expansion device; **SVM** = Support Vector Machine; **VGG16** = Visual Geometry Group 16.

Correspondence:

Nambi Nallasamy, MD, Kellogg Eye Center, University of Michigan, 1000 Wall Street, Ann Arbor, MI, 48105. E-mail: nnallasa@med.umich.edu.

References

1. Grzybowski A. Recent developments in cataract surgery. *Ann Transl Med*. 2020;8:1540−1545.

2. Mahmoud O, Zhang H, Matton N, et al. CatStep: automated cataract surgical phases classification and boundary

segmentation leveraging inflated 3D-CNN architectures and BigCat. *Ophthalmol Sci*. 2024;4:100405.

3. Kim H, Kim HJ, Joo CK. Change of pupil diameter after cataract surgery or after-cataract surgery. *J Korean Ophthalmol*. 2005;46:51−56.

4. Hayashi K, Hayashi H. Pupil size before and after phacoemulsification in nondiabetic and diabetic patients. *J Cataract Refract Surg*. 2004;30:2543−2550.

5. Ordiñaga-Monreal E, Castanera-Gratacós D, Castanera F, et al. Pupil size differences between female and male patients after cataract surgery. *J Optom*. 2022;15:179−185.

6. Zeng Z, Giap BD, Kahana E, et al. Evaluation of methods for detection and semantic segmentation of the anterior capsulotomy in cataract surgery video. *Clin Ophthalmol*. 2024;18:647−657.

7. Ba-Ali S, Lund-Andersen H, Brøndsted AE. Cataract surgery affects the pupil size and pupil constrictions, but not the late post-illumination pupil response. *Acta Ophthalmol*. 2017;95: e252−e253.

8. Rickmann A, Waizel M, Szurman P, Boden KT. Relation of pupil size and cataract surgery using PupilX. *Int J Ophthalmol Clin Res*. 2016;3.

9. Ong-Tone L, Bell A. Pupil size with and without adrenaline with diclofenac use before cataract surgery. *J Cataract Refract Surg*. 2009;35:1396−1400.

10. Lumme P, Laatikainen LT. Risk factors for intraoperative and early postoperative complications in extracapsular cataract surgery. *Eur J Ophthalmol*. 1994;4:151−158.

11. Guzek JP, Holm M, Cotter JB, et al. Risk factors for intraoperative complications in 1000 extracapsular cataract cases. *Ophthalmology*. 1987;94:461−466.

12. Vasavada A, Singh R. Phacoemulsification in eyes with a small pupil. *J Cataract Refract Surg*. 2000;26:1210−1218.

13. Kanellopoulos AJ, Asimellis G. Clear-cornea cataract surgery: pupil size and shape changes, along with anterior chamber volume and depth changes. A Scheimpflug imaging study. *Clin Ophthalmol*. 2014;8:2141−2150.

14. Malyugin B. Cataract surgery in small pupils. *Indian J Ophthalmol*. 2017;65:1323−1328.

15. Herranz Cabarcos A, Pifarré Benítez R, Martínez Palmer A. Impact of intraoperative floppy IRIS syndrome in cataract surgery by phacoemulsification: analysis of 622 cases. *Arch Soc Esp Oftalmol*. 2023;98:78−82.

16. Chang DF, Campbell JR. Intraoperative floppy iris syndrome associated with tamsulosin. *J Cataract Refract Surg*. 2005;31: 664−673.

17. Pärssinen O, Leppänen E, Keski-Rahkonen P, Mauriala T, Dugué B, Lehtonen M. Influence of tamsulosin on the iris and its implications for cataract surgery. *Invest Ophthalmol Vis Sci*. 2006;47:3766−3771.

18. Isac MMS, Ting DSJ, Patel T. Spontaneous pupillary recovery of urrets-zavalia syndrome following descemet's membrane endothelial keratoplasty. *Med Hypothesis Discov Innov Ophthalmol*. 2019;8:7−10.

19. Yilmaz I, Perente I, Saracoglu B, et al. Changes in pupil size following panretinal retinal photocoagulation: conventional laser vs pattern scan laser (PASCAL). *Eye*. 2016;30: 1359−1364.

20. Giap BD, Srinivasan K, Mahmoud O, et al. Tensor-based feature extraction for pupil recognition in cataract surgery. *Annu Int Conf IEEE Eng Med Biol Soc*. 2023;2023:1−4.

21. Giap BD, Srinivasan K, Mahmoud O, et al. Adaptive tensor-based feature extraction for pupil segmentation in cataract surgery. *IEEE J Biomed Health Inform*. 2024;28: 1599−1610.

22. Ronneberger O, Fischer P, Brox T. *U-net: convolutional networks for biomedical image segmentation. Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer; 2015: 234−241.

23. Chaurasia A, Culurciello E. *LinkNet: exploiting encoder representations for efficient semantic segmentation. IEEE Visual Communications and Image Processing*. St. Petersburg, FL: IEEE; 2017:1−4.

24. Lin TY, Dollár P, Girshick R, et al. *Feature pyramid networks for object detection. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. Honolulu, HI: IEEE; 2017: 936−944.

25. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2014. https://doi.org/10.48550/arXiv.1409.1556.

26. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition. , Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*. Las Vegas, NV: IEEE; 2016:770−778.

27. Huang G, Liu Z, Pleiss G, et al. Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell*. 2022;44:8704−8716.

28. Howard AG, Zhu M, Chen B, et al. MobiNets: efficient convolutional neural networks for mobile vision applications. *arXiv*. 2017. https://doi.org/10.48550/arXiv.1704.04861.

29. Deng J, Dong W, Socher R, et al. *ImageNet: a large-scale hierarchical image database. Proc. IEEE Conf. Comput. Vis. Pattern Recognit*. Miami, FL: IEEE; 2009:248−255.

30. Iakubovskii P. Segmentation models [Online] https://github.com/qubvel/segmentation_models. Accessed January 15, 2023.

31. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell*. 1986;8:679−698.

32. Fitzgibbon AW, Fisher RB. *A buyer's guide to conic fitting. Proceeding of the 6th British Conference on Machine Vision*. Birmingham, UK: BMVA Press; 1995:513−522.

33. Cortes C, Vapnik V. Support-vector network. *Mach Learn*. 1995;20:273−297.

34. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theor*. 1967;13:21−27.

35. Breiman L. Random forests. *Mach Learn*. 2001;45:5−32.

36. Quinlan R. *C4.5: programs for machine learning*. Boston, MA: Morgan Kaufmann Publishers; 1993.

37. Mitchell TM. *Machine learning*. New York City, NY: McGraw-Hill, Inc.; 1997.

38. Hosmer Jr DW, Lemeshow S. *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons; 2000.

39. Sokolova N, Schoeffmann K, Taschwer M, et al. Automatic detection of pupil reactions in cataract surgery videos. *PLoS One*. 2021;16:e0258390.

40. Yeh HH, Jain AM, Fox O, et al. PhacoTrainer: deep learning for cataract surgical videos to track surgical tools. *Transl Vis Sci Technol*. 2023;12:23.

41. Grammatikopoulou M, Flouty E, Kadkhodamohammadi A, et al. CaDIS: cataract dataset for surgical RGB-image segmentation. *Med Image Anal*. 2021;71:102053.

42. Matton N, Qalieh A, Zhang Y, et al. Analysis of cataract surgery instrument identification performance of convolutional and recurrent neural network ensembles leveraging BigCat. *Transl Vis Sci Technol*. 2022;11:1.