Review Article

Clinical application of artificial intelligence in corneal transplantation

Nima Rastegar Rad [1]

[1] *Fayyazbakhsh Hospital, Ophthalmology Department, Sazeman Tamin Ejtemaie, Tehran, Iran*

ABSTRACT

Background: Artificial intelligence (AI) is increasingly being incorporated into ophthalmic practice and shows promise in improving diagnostic accuracy, patient selection, surgical planning, and postoperative care. In corneal transplantation, AI-based tools may help clinicians make more informed decisions throughout the patient journey. This narrative review examines the current evidence on the use of AI in corneal transplantation, with a focus on its clinical applications, existing challenges, and future potential for routine practice.

Methods: A literature review was conducted to identify studies evaluating the use of AI in corneal transplantation. Searches were conducted in PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar databases for articles published up to 20 August 2025. The search strategy combined terms related to "artificial intelligence", "machine learning", and "deep learning" with "corneal transplantation" and "keratoplasty".

Results: Twenty-two publications met the inclusion criteria and were represented by 20 synthesis entries after two pairs of companion reports were linked. Most studies used imaging modalities such as anterior segment optical coherence tomography (AS-OCT), specular microscopy, or surgical videos, while others incorporated structured clinical information or multimodal datasets. Descemet membrane endothelial keratoplasty (DMEK) was the most commonly investigated procedure. AI approaches predominantly utilized supervised methodologies, spanning deep learning networks and traditional machine learning algorithms alongside a limited application of unsupervised and survival-based models. Common objectives included predicting graft outcomes, detecting and quantifying graft detachment, identifying factors associated with graft failure, and assisting surgical decision-making. Classification and image segmentation were the predominant analytical tasks. AI models evidenced favorable predictive and analytical performance across a broad range of clinical applications, highlighting their potential to support preoperative risk stratification, intraoperative decision-making, postoperative monitoring, and long-term graft outcome prediction.

Conclusions: The current literature suggests that AI has the potential to become an important clinical support tool in corneal transplantation. By integrating imaging, biometric, and clinical data, AI models may help predict graft survival, identify patients at higher risk of complications, and facilitate more personalized management strategies. Although the reported results are encouraging, several barriers remain, including limited external validation, lack of standardized datasets, and concerns about model interpretability. Future work should focus on developing robust, explainable, and generalizable AI systems through multicenter collaboration and multimodal data integration to support their safe and effective adoption in clinical practice.

KEYWORDS

machine learning, AI (artificial intelligence), learning, deep, learning, machine, corneas, corneal transplantations, penetrating keratoplasty, lamellar keratoplasty, descemet stripping endothelial keratoplasty

Correspondences: Nima Rastegar Rad, Fayyazbakhsh Hospital, Ophthalmology Department, Sazeman Tamin Ejtemaie, Tehran, Iran, Email: nima.rastegarrad@atiehhospital.ir, rrnima88@yahoo.com. ORCID iD: https://orcid.org/0009-0003-5413-4899.

How to cite this article: Rastegar Rad N. Clinical application of artificial intelligence in corneal transplantation. Med Hypothesis Discov Innov Optom. 2026 Summer; 7(2): 80-96. DOI: https://doi.org/10.51329/mehdioptometry249

Received: 29 May 2026; Accepted: 19 July 2026

INTRODUCTION

Corneal transplantation (keratoplasty) is the most common transplant procedure globally, with disorders affecting the corneal endothelium being the primary reason for surgical intervention [1–4]. Worldwide demand for donor human corneas is steadily increasing, driven by an ageing population, broader access to treatment, and continual advances in surgical techniques. However, donor tissue availability remains insufficient to meet this growing demand [5–7].

Despite its clinical benefits, corneal transplantation can result in significant postoperative complications, including graft rejection, infectious keratitis, fixed dilated pupil, cataract formation, iris atrophy, suture-related complications, and secondary glaucoma [8–12]. Preventing these complications [13–15] is critical toward preserving vision and quality of life, reducing the need for costly secondary interventions, and ensuring optimal use of limited donor corneal tissue.

Keratoplasty can be broadly categorized into penetrating keratoplasty (PK) and lamellar keratoplasty, the latter encompassing anterior lamellar keratoplasty (ALK), deep anterior lamellar keratoplasty (DALK), Descemet's stripping automated endothelial keratoplasty (DSAEK), and Descemet's membrane endothelial keratoplasty (DMEK) [16]. These techniques have undergone substantial refinement, driven by advances in microsurgical instrumentation, imaging technologies, and the understanding of corneal pathophysiology. Innovations in donor tissue preparation, femtosecond laser-assisted surgery, and postoperative care have improved wound stability and precision while reducing suture requirements, resulting in better visual outcomes, fewer complications, and faster recovery [17–19].

As corneal transplantation has evolved, so too has the volume and complexity of data captured by transplant registries. These datasets often contain numerous clinical, surgical, and longitudinal variables, making them increasingly difficult to analyze using conventional statistical methods alone, including Kaplan-Meier survival analysis and Cox regression models [20, 21]. Artificial intelligence (AI) offers a promising alternative by extracting complex patterns from high-dimensional datasets and generating predictive models for surgical outcomes and graft survival. These tools enable more accurate risk stratification, personalized clinical decision-making, and earlier identification of patients at risk of complications, ultimately improving transplant outcomes [20–23].

The expanding use of AI in corneal transplantation has generated growing interest in its potential clinical value [20–23]. This narrative review provides an overview and critical appraisal of current AI applications in corneal transplantation and identifies priorities for future research, model validation, and clinical implementation.

METHODS

A literature search was conducted in PubMed/MEDLINE, Scopus, Web of Science, and supplementary databases, including Google Scholar, for articles published up to 20 August 2025. The search combined terms related to "artificial intelligence", "machine learning", and "deep learning" with "corneal transplantation" and "keratoplasty". Titles and abstracts of retrieved records were initially screened for studies that met the inclusion criteria. When eligibility could not be determined from the title and abstract alone, the full text was reviewed. Following study selection, all eligible articles underwent detailed full-text assessment.

Articles were eligible if they were original human research studies (prospective, retrospective, cohort, case-control, or randomized) that evaluated applications of AI in corneal transplantation); included patients of any age undergoing, being assessed for, or followed after keratoplasty procedures (including PK, DALK, DSAEK, DMEK, Descemet stripping endothelial keratoplasty [DSEK], and other lamellar keratoplasty techniques); involved donor corneal graft assessment and transplantation-related workflows; used AI approaches (including machine learning, deep learning, computer vision, radiomics, and automated image analysis) to analyze clinical, imaging, surgical, or multimodal datasets (including endothelial cell images, specular microscopy, corneal topography, optical coherence tomography [OCT] scans, surgical videos, and patient records); applied AI to diagnosis, risk stratification, surgical planning, intraoperative guidance, surgical phase recognition, robotic assistance, postoperative monitoring, automated image interpretation, quantitative assessment of graft or endothelial health, graft quality evaluation, endothelial function assessment, or prediction of graft-related outcomes and complications; compared AI-based approaches with conventional clinical assessment, manual image interpretation, existing software tools, expert evaluation, or standard statistical methods; reported clinically relevant outcomes, quantitative measurement performance, agreement analyses, or model performance metrics (such as area under the curve [AUC], sensitivity, specificity, accuracy, Dice similarity coefficient (DSC), mean absolute error, graft survival, and complication rates); and provided sufficient methodological detail on the AI algorithms used. In view of the rapidly advancing nature of AI research, conference proceedings and abstracts reporting original human data were considered when sufficient methodological details and outcome measures were available. For the two included conference reports, relevant information was extracted from their published abstracts and interpreted alongside the corresponding full-text companion publications; they were not treated as independent study populations.

Excluded were studies focused on ophthalmic conditions unrelated to corneal transplantation (such as laser-assisted in situ keratomileusis, corneal cross-linking, or cataract surgery without a transplantation-related objective) or which relied solely on conventional statistical methods without incorporating clinical AI; studies restricted to molecular biology or transcriptomic workflows that lacked a direct application to clinical predictive AI modeling; studies using automated machine learning platforms without adequate reporting of model architecture, analytical workflow, or feature interpretability; non-English publications, review articles, editorials, commentaries, letters, short communications; laboratory-based, *ex vivo*, or animal studies; case reports and small case series lacking quantitative analysis; and studies that did not report clinically relevant outcomes, quantitative performance metrics, or sufficient methodological details regarding the AI approach.

A structured qualitative assessment of study characteristics, methodological quality, and findings was performed. Data extracted from the included studies comprised author, year of publication, study design characteristics (prospective, retrospective, registry-based, randomized controlled trial [RCT], or cohort), country and setting, patient demographics, surgical procedures, dataset characteristics, data type (imaging, registry, multimodal, video, or clinical records), AI approaches, machine learning category (supervised, unsupervised, or deep learning), analytical methods, algorithms, model architecture, level of analysis, and data-preprocessing techniques. Information on imaging modalities, software platforms, study objectives, model performance, clinical applications, and key AI outcomes, including classification or clustering tasks, performance metrics, results summaries, and measures of clinical relevance, was also recorded. In addition, methodological quality and risk-of-bias assessments were extracted and summarized.

Methodological quality was evaluated using tools selected according to the objective and design of each study. The Prediction model Risk Of Bias Assessment Tool for Artificial Intelligence (PROBAST+AI) [24] domains were applied to assess model-development quality, risk of bias in model-performance evaluation, and applicability, as appropriate. The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [25] was applied to diagnostic-accuracy studies in which an AI method was evaluated as an index test against a reference standard. For technical image-analysis studies evaluating segmentation, delineation, or measurement performance rather than diagnostic accuracy, relevant QUADAS-2 domains were adapted to assess dataset selection, model evaluation, reference annotations, and data flow. Overall methodological judgments were synthesized descriptively rather than calculated as numerical scores.

RESULTS

The literature search identified 1,092 records, including 1,087 retrieved from electronic databases and five identified through manual searches. After exclusion of 930 non-relevant records and removal of 49 duplicates, 113 reports were assessed for eligibility. Following application of the eligibility criteria, 91 publications were excluded, leaving 22 publications for inclusion, including two conference abstracts. These 22 publications represented 20 synthesis entries in Tables 1–3 because two pairs of companion reports were each treated as a single study unit.

Reports based on the same or overlapping study populations were identified during data extraction. Muijzer et al. and Wisse et al. [22, 26], and the two reports by Joseph et al. [27, 28], were treated as companion reports and synthesized as two study units. Ang et al. [20] and Ng Yin Ling et al. [21] originated from the same Singapore corneal-transplant registry programme but evaluated distinct indication and procedure cohorts and were retained separately. The former included DSAEK and PK performed for endothelial diseases, whereas the latter included PK and DALK performed for optical indications and specifically excluded endothelial diseases. Heslinga et al. [29, 30] analysed the same RCT–derived imaging cohort of 80 AS-OCT scans from 68 participants but addressed distinct technical objectives; post-DMEK pachymetry [29] and graft-detachment quantification [30]. They were therefore retained as separate task-specific analyses but were not considered independent patient populations.

Table 1 summarizes the characteristics of the included studies extracted from the methods sections:

- The included evidence was geographically diverse, covering the United States, Europe (the Netherlands, Germany, France, and Denmark), East Asia (Japan, the Republic of Korea, and China), Turkiye, Iran, India, and Singapore. Of the 20 synthesis entries, 13 used fewer than 500 samples, three used more than 3000 samples, and four used between 500 and 3000 samples. The analytical units varied among studies and included participants, eyes, grafts, scans, images, and surgical videos; these categories therefore describe dataset size rather than numbers of independent patients. Demographic characteristics of study participants, where reported, are summarized.
- Twelve synthesis entries primarily used imaging-based data, including optical coherence tomography (OCT) imaging (n = 8), microscopic endothelial imaging (n = 3), and surgical video recordings (n = 1). Three entries relied primarily on registry-based data, whereas five used clinical or multimodal datasets combining clinical, imaging, donor, laboratory, or operative variables. Among the latter, one study combined Scheimpflug corneal tomography with clinical, questionnaire-derived, and laboratory variables.
- DMEK was the most frequently investigated surgical procedure. Other studies examined DALK, DSAEK, PK, ultrathin DSAEK, or mixed keratoplasty populations. One study focused on donor-corneal graft quality assessment rather than postoperative transplant recipients.
- AI methodologies varied considerably across studies. Supervised approaches predominated and included both deep-learning architectures and conventional machine-learning models. One study used unsupervised learning. Several studies evaluated more than one analytical method; therefore, the methodological categories were not mutually exclusive.
- Data analysis approaches were heterogeneous, and several studies implemented multiple analytical methods, resulting in overlapping categories. Overall, analytical strategies included classification and outcome prediction, image segmentation

and quantitative image analysis, prognostic modeling and survival analysis, as well as unsupervised clustering and feature-based exploratory analyses.

- ❖ Studies employed a diverse array of analytical algorithms; deep learning architectures, including convolutional neural networks (CNNs) and U-Net; classical machine learning models like random forests and logistic regression; and statistical and survival modeling techniques, including Cox regression and least absolute shrinkage and selection operator (L1-regularized logistic regression [LASSO]). Unsupervised clustering and dimension reduction methods were used in a few studies, while others applied evaluation frameworks such as receiver operating characteristic (ROC) analysis, cross-validation, and analysis of variance (ANOVA).
- ❖ Analytical focus varied across studies. Some analyses were conducted at the patient level or eye level, integrating clinical and imaging data to model outcomes or predict risk. Others focused on image-based techniques, including segmentation, classification, and spatial quantification of acquired scans. A subset of studies targeted intraoperative procedures or endothelial cellular structures, or explored global data patterns.
- ❖ OCT was the most common imaging instrument; other instruments were specular microscopy, phase-contrast microscope, and alternative imaging modalities such as Pentacam and surgical video capture. Not all studies specified the imaging instrument used.
- ❖ Analytical objectives included predicting graft outcomes and complications (successful big-bubble formation, graft detachment [GD], rejection, failure, postoperative rebubbling), estimating graft survival, identifying associated risk factors and latent patterns, automating detection and corneal mapping (objective detection of minimal corneal edema, interface delineation), applying deep learning for surgical guidance and endothelial cell analysis, and predicting related clinical conditions such as postoperative ocular hypertension.
- ❖ Performance evaluation approaches encompassed classification metrics (AUC, F1 score, accuracy, sensitivity, and specificity), regression metrics (R-squared and explained variance), survival analysis metrics (C-index and hazard ratios), image segmentation metrics (DSC, Jaccard index, and Class Intersection-over-Union [C-IoU]), statistical agreement assessments (Bland–Altman plots and Pearson correlation coefficients with their corresponding confidence intervals), and ANOVA for multi-group significance testing.
- ❖ AI models were applied across multiple stages of the corneal transplantation pathway. Several studies focused on preoperative risk stratification, enabling identification of high-risk grafts, personalized prognostic assessment, prediction of intraoperative procedural success, timely clinical intervention, long-term graft survival forecasting, donor selection, longitudinal postoperative follow-up, and optimization of surgical decision-making. These approaches have the potential to improve patient selection, guide perioperative management, and reduce postoperative complications. Other studies enhanced surgical planning by supporting intraoperative decision-making and robotic-assisted keratoplasty. AI-based systems were also developed to predict procedural outcomes and assist surgeons during complex keratoplasty procedures.
- ❖ Postoperative applications were particularly prominent. AI models facilitated early detection and monitoring of complications, including GD, graft rejection, postoperative ocular hypertension, and corneal edema. Several studies focused on automated quantification and localization of GD, objective assessment of endothelial cell health, and monitoring of corneal recovery following endothelial keratoplasty. A number of studies investigated personalized outcome prediction, enabling early identification of patients at increased risk of adverse events and supporting preventive interventions, such as intensified surveillance or corticosteroid therapy in patients at risk of graft rejection. AI-based analysis of corneal imaging facilitated objective advance detection of minimal or subclinical corneal edema. Additionally, AI enabled objective assessment of endothelial parameters, donor cornea quality evaluation, and support for graft-release decisions.
- ❖ Various studies addressed similar analytical objectives. These included prediction of graft survival or failure; prediction of graft detachment or rebubbling after DMEK; segmentation and quantification of graft detachment on OCT; automated endothelial-cell analysis; and early prediction of graft rejection. Due to overlapping populations and outcomes, themes were synthesized descriptively rather than as mutually exclusive categories.
- ❖ Additional studies addressed more specialized applications, including postoperative pachymetry assessment after DMEK, prediction of favorable versus unfavorable DALK outcomes, prediction of successful big-bubble formation during DALK, pre-DMEK objective detection of preoperative corneal edema, and prediction of post-DMEK ocular hypertension. These latter applications were less frequently investigated but illustrate the expanding scope of AI across the corneal transplantation pathway.
- ❖ Binary classification was the most frequently applied modeling strategy, being primarily used to predict outcomes such as successful big-bubble formation, rebubbling, graft rejection, GD, graft failure, and postoperative ocular hypertension. Multi-class classification was less common and was mainly applied to surgical phase recognition and GD grading. Unsupervised clustering was used in one study to stratify keratoplasty risk and characterize corneal ectasia severity patterns. Ensemble approaches combining multiple machine learning classifiers were employed in a small number of studies to improve predictive performance. Regression-based estimation of endothelial parameters was applied in two studies for quantitative assessment of endothelial cell morphology and density. Instead of performing clustering or classification directly, several studies focused instead on image segmentation, quantitative image analysis, survival prediction, or variable importance modeling without discrete outcome grouping.

Table 1. Included studies – Summary of methods

Author (Year), Country	Input data	Age	M/F, n or %	Type of surgery	Data type	ML type	Data analysis type	Analytical algorithm	Model architecture	Analytical level	Data preprocessing	Imaging instrument	Analysis software/aim of analysis	Performance evaluation	Clinical application
Yousefi et al. (2020), USA/Japan [31]	3318 baseline corneal OCT examinations from 3162 subjects and a separate dataset of 333 eyes that subsequently underwent keratoplasty	69.7 ± 16.1 y	41% / 59%	PK, LKP, DALK, DSAEK, DMEK	Corneal OCT-derived topography, pachymetry, and elevation parameters (424 features)	Unsupervised learning	Dimensionality reduction, clustering, validation, and post-hoc analysis	PCA, tSNE, and density-based clustering	NA (unsupervised clustering framework)	Patient-level analysis; global corneal data structure and surgical risk stratification	Feature selection, PCA-based dimensionality reduction, tSNE manifold learning, outlier removal	SS-1000 CASIA OCT Imaging System (Tomey, Japan)	R version 3.5/ Identify latent patterns in corneal OCT data and predict future keratoplasty risk without predefined labels.	Five clusters were identified; the normalized likelihood of future surgery was calculated for each cluster.	Risk stratification for future keratoplasty and identification of high-risk patients to support monitoring and surgical decision-making.
Hayashi et al. (2023), Japan/Germany [32]	AS-OCT B-scans from 312 post-DMEK eyes (138 training, 174 testing); mean 53.28 B-scans per eye	Mean: 66.4–71.7 (range 17–90)	139/161	DMEK	Anterior segment OCT B-scans	Supervised deep learning	Image classification and RB prediction	EfficientNetB0 with transfer learning	EfficientNetB0 (ImageNet pretrained)	Image-level classification and eye-level RB prediction	Image resizing (224 × 224), normalization, data augmentation.	Spectralis spectral-domain OCT (Heidelberg Engineering, Germany)	Python (TensorFlow/Keras, Optuna, Scikit-learn, SciPy) / Predict GD and postoperative RB	AUC; sensitivity; specificity; CIs	Postoperative decision support for RB after DMEK based on OCT
Kundu et al. (2025) India/Netherlands [33]	250 DALK eyes (a 5-year follow-up for advanced keratoconus) with clinical, questionnaire, laboratory, and Pentacam HR data	28 ± 3.9 y	135/115	DALK	Multimodal: clinical, questionnaire, Pentacam tomography, and laboratory data	Supervised	Classification (favorable or unfavorable outcomes)	RF classifier (10 trees)	RF	Eye-level	Pentacam quality screening, manual scan verification, feature ranking, leave-one-out cross-validation	Pentacam HR (OCULUS, Germany)	Orange3 (version 3.25.0), MedCalc 19 (MedCalc Software, Ostend, Belgium) / Predict long-term DALK outcomes using preoperative risk factors and tomography	AUC; sensitivity; specificity; classification accuracy; precision; recall; F1-score.	Preoperative risk stratification and personalized prognosis before DALK
Kim et al. (2025), Republic of Korea [34]	124 eyes from 119 patients underwent DMEK (106 eyes from 101 patients analyzed) demographic data, medical/surgical history, IOP, donor graft size, lens status, glaucoma history, and AS-OCT biometric data	59.9 ± 14.1 y	Female: 49.1% overall; 52.0% (normal) vs. 41.9% (postoperative OHT)	DMEK	Clinical/electronic medical records data + imaging (AS-OCT biometric measurements)	Supervised	Classification and survival analysis	XGBoost, RF, CatBoost, logistic regression; PCA-derived feature engineering; Cox proportional hazards regression	Ensemble model of XGBoost + RF + CatBoost + logistic regression	Eye-level	Mean imputation for missing values, synthetic minority oversampling technique, PCA-derived ARA-comb variable, feature importance evaluation, feature dropping, stratified 5-fold cross-validation	AS-OCT (CASIA; Tomey, Nagoya, Japan)	SPSS Statistics (version 26; IBM, Armonk, NY, USA)/Predict post-DMEK OHT and identify preoperative risk factors	Accuracy, precision, recall, AUC-ROC, AUC-PR; Kaplan–Meier and log-rank test	Preoperative risk stratification and personalized postoperative monitoring and treatment planning to reduce complication.
Ang et al. (2022), Singapore [20]	1335 consecutive patients/eyes (946 DSAEK, 389 PK) with FED or BK from the Singapore Cornea Transplant Registry; donor characteristics, recipient characteristics; clinical outcomes, complications, visual acuity, endothelial cell counts, and op. data	Total: 68.3 ± 11.4 y; PK: 67.4 ± 12.0 y; DSAEK: 68.7 + 11.1 y	Total: 635 (47.6%)/700 (52.4%); PK: 191 (49.1%)/198 (50.9%); DSAEK: 444 (46.9%)/502 (53.1%)	DSAEK and PK	Clinical registry data	Supervised ML + survival modeling	Survival prediction, variable importance ranking, prognostic modeling	RSF, Cox proportional hazards regression, Kaplan–Meier survival analysis	RSF model (10 000 trees) with VIMP-guided nested Cox model	Eye-level survival analysis	Variable selection using VIMP ranking; RSF parameter tuning; bootstrap/ OOB validation; Schoenfeld testing; penalized spline assessment of non-linearity	Non-contact specular microscope (Konan Medical Corp, Hyogo, Japan)	R (version 4.0.2, R Foundation for Statistical Computing) with the randomForestSRC package/ Identify factors associated with 10-year graft survival and GF after DSAEK and PK	OOB Harrell's C-index; hazard ratios (HRs) with 95% CIs from Cox regression; Kaplan–Meier survival analysis and log-rank testing.	Long-term graft survival prediction and GF risk stratification
Ng Yin Ling et al. (2025), Singapore [21]	1687 consecutive Singapore Corneal Transplant Registry patients undergoing PK (n = 1163) or DALK (n = 524) for optical indications, excluding endothelial diseases. Fifty-five candidate variables were selected from more than 680 prospectively collected donor, recipient, clinical, operative, and postoperative variables. The internal development set included 1343 complete records, and the held-out test set included 344 records with missing data (less than 15% per record, which was imputed for analysis).	51.8 ± 21.2 y (PK: 57.5 ± 19.4 y, DALK: 39.0 ± 19.5 y)	Overall: 913/774 (54.1%/45.9%); PK: 625 (53.7%)/538 (46.3%), DALK: 288 (55.0%)/ 236 (45.0%)	PK and DALK	Prospectively collected clinical registry data	Supervised machine learning and survival modelling	Long-term graft-survival prediction, variable-importance analysis, and risk-score development	Random survival forest, nested Cox proportional-hazards regression, AutoScore, and Kaplan–Meier analysis	RSF comprising 1000 survival trees with VIMP and minimal-depth ranking, followed by Cox-based variable selection and an AutoScore risk-scoring model	Patient-level survival analysis	Complete-case development dataset; mean/mode imputation in the held-out test set using development-set values; RSF hyperparameter tuning using out-of-bag error; stepwise nested Cox-model selection; categorization of continuous predictors; sensitivity analysis excluding intraoperative variables	Not applicable	R packages RandomForestSRC v3.1.0 and AutoScore v1.0.0. Aim: to identify predictors of 10-year PK/DALK graft survival and develop an interpretable clinical risk score.	C-index and time-integrated AUC with bootstrapped 95% CIs; Kaplan–Meier survival and log-rank testing; HRs with 95% CIs	Preoperative long-term graft-survival risk stratification and patient counselling after PK or DALK
Feizi et al. (2024), Iran [23]	1214 consecutive eyes of 985 keratoconus patients undergoing primary keratoplasty between 1994 and 2021, including 574 PK and 640 DALK eyes. Nineteen recipient, donor, operative, and postoperative variables were evaluated. Graft rejection occurred in 341 eyes (28.1%).	29.4 ± 8.7 y (range, 12–63)	682/303 patients (69.2%/30.8%)	PK and DALK	Retrospective interventional clinical, donor, operative, and postoperative data	Supervised machine learning and survival modelling	Binary graft-rejection prediction, time-to-event analysis, model comparison, and variable-importance assessment	Artificial neural network, support vector machine, gradient boosting, extra-trees classifier, and random survival forest	Five separately trained ML models; gradient boosting used 100 rounds and the extra-trees classifier used 100 estimators; complete ANN and SVM architectures were not reported	Eye-level analysis; both eyes were included for bilaterally transplanted patients	80%/20% development/test split (971/243 eyes) and K-fold cross-validation; participant-level separation and missing-data procedures were not clearly reported	Not applicable; donor endothelial cell density was used as a structured variable, but no images were analyzed by the models	Python v3.10 and SPSS v25. Aim: to compare ML models and identify factors associated with graft rejection after PK or DALK for keratoconus.	C-statistic, Brier score, test and cross-validation accuracy, precision, MSE, and RMSE	Identification of potentially modifiable factors associated with graft rejection; findings may inform postoperative surveillance, corticosteroid management, and suture management but require further validation
Hayashi et al. (2021), Japan/Germany [35]	46 eyes from 46 patients; preoperative SD-OCT images and Scheimpflug tomography (Pentacam HR)-derived biometric data	48.1 ± 17.7 y	35 (76.1%)/11 (23.9%)	DALK	SD-OCT images + corneal biometric data (Pentacam HR)	Supervised deep learning	Classification (SBB vs. FBB)	VGG-16 with Score-CAM visualization	VGG-16 (fine-tuned)	Eye-level	Image resizing (1024 × 242→224 × 224), normalization, data augmentation, K-fold cross-validation (K=5),	SD-OCT (Heidelberg Engineering, Heidelberg, Germany) and Pentacam HR (Oculus GmbH, Wetzlar, Germany)	Keras (a Python TensorFlow Application Programming Interface) [API]); JMP Pro (version 14.0.0; SAS Institute, Cary, NC, USA) for analysis/Predict SBB formation before DALK and compare predictions with preoperative biometric parameters	AUC, ROC analysis, determination success rate, sensitivity, specificity, 95% CI	Preoperative DALK planning and big-bubble success prediction
Hayashi et al. (2020), Japan [36]	992 AS-OCT images (496 images of 31 eyes from 24 patients required RB and 496 images of 31 eyes from 29 patients did not require RB [non-RB]) from 62 eyes of 53 patients; postoperative day-5 images after DMEK.	RB: 74.3 ± 5.5 y; non-RB: 74.2 ± 7.0 y	RB: 19/5; non-RB: 21/8	DMEK	AS-OCT imaging	Supervised deep learning	Classification (RB vs non-RB)	VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, Xception, DenseNet121, DenseNet169, DenseNet201; ensemble modeling	Nine deep neural network (DNN) architectures; best ensemble: VGG16, VGG19, DenseNet121, DenseNet169, DenseNet201	Image-level classification with eye-level RB determination	Image resizing (1600×857→256×256), normalization, k-fold cross-validation (k=4), image augmentation (contrast, gamma correction, smoothing, histogram flattening, Gaussian noise, salt-and-pepper noise)	Swept-source AS-OCT (Casia SS-1000; Tomey, Nagoya, Japan)	Keras, TensorFlow, Python scikit-learn, SciPy, statsmodel libraries./ Evaluate the efficacy of deep learning for determining the need for RB after DMEK	AUC, sensitivity, specificity, ROC/Youden analysis, and ensemble scoring (AUC, sensitivity, and specificity). Performance was evaluated for both the best single model and the best ensemble model.	Automated postoperative RB assessment and management after DMEK

Karaca et al. (2025), Turkey [37]	Donor, graft, surgical, and clinical variables from 209 DMEK eyes	Mean donor age: 60.66 ± 5.59 y	Not reported	DMEK	Clinical/tabular donor and surgical data	Supervised	Binary classification (GF vs. no GF)	k-Nearest Neighbor (kNN), SVM, RF, RUSBoosted tree, ANN, ANOVA	Shallow single-layer neural network (100 hidden units)	Eye-level prediction	Input normalization; holdout validation (75% training, 25% testing)	Not applicable	Not reported/Predict early GF after DMEK using donor and graft characteristics	Accuracy, sensitivity, specificity, precision, recall, F1-score, AUC-ROC	Prediction of GF risk and donor selection support for DMEK
Muijzer et al. (2022), with an earlier conference report by Wisse et al. (2021), Netherlands [22, 26]	Nationwide Dutch Cornea Transplant Registry (NOTR) cohort of 3647 posterior lamellar keratoplasties performed at 16 centres during 2015–2018, comprising 2651 DSEK and 996 DMEK procedures. Ninety-one donor-, recipient-, surgical-, and practice-pattern predictors were evaluated. After correction for the DMEK learning curve, 3464 procedures were included in the full publication's ML analysis.	Mean donor age: 70 ± 9 y	Donors: 61.4% male; female percentage not explicitly reported	DSEK, including ultrathin DSEK, and DMEK	Nationwide registry-based clinical, surgical, donor, recipient, and centre-specific practice-pattern data	Supervised	Binary classification and predictor identification for GD/RB versus no GD/RB; the earlier conference report additionally evaluated GF versus no GF	Full publication: L1-regularized LASSO, CTA, and RFC. Earlier conference report: LASSO and CTA	LASSO, CTA, and RFC with 1000 trees	Procedure-/transplant-level analysis	Recoding of 91 predictors; exclusion of the first 20 DMEK procedures per centre to reduce learning-curve effects; 70%/30% training/test split; ROSE-based oversampling for LASSO and CTA; class weighting for RFC; and fivefold cross-validation for tree pruning.	Not applicable	R (version 4.0.5) and Python version 3.8 with scikit-learn (version 0.24.1). Aim: to identify donor, recipient, surgical, and practice-pattern factors associated with postoperative GD/RB; the earlier report additionally evaluated GF.	Full publication: AUC, sensitivity, and specificity in an independent test set. Earlier conference report: explained variance for RB and GF	Risk stratification, optimization of surgical practice patterns, identification of potentially modifiable factors, and hypothesis generation for reducing postoperative GD, RB, and GF
Glatz et al. (2021), Germany [38]	AS-OCT scans consisting of 256 cross-sectional images per 360° scan; training set: 27 scans (6912 images), validation set: 20 scans (5120 manually labeled images), application set: 107 eyes of 103 participants.	69 y (IQR 62–78) in the application cohort; training set 69 y (62–79), validation set 73 y (58–78)	Application set: 70 F (68%); training set: 14 F (54%); validation set: 12 F (71%)	DMEK	High-resolution AS-OCT imaging data	Deep learning (supervised image segmentation)	Image segmentation and quantification of GD (pixel-wise segmentation)	UNet++ image-segmentation network	CNN (UNet++)	Pixel-level segmentation with scan-level and eye-level detachment quantification	Median filtering, cropping, contrast enhancement, resizing, image augmentation (shifting, flipping, scaling, rotation, shearing), post-processing with polar transformation, Gaussian filtering, artifact removal	AS-OCT (Casia-1, Tomey GmbH)	TensorFlow (version 1.4.1), Keras (version 2.2.2), ImageJ/Fiji; (NIH).custom software for annotation and extraction/ Develop and validate a deep-learning model to quantify DMEK GD and generate 2-dimensional maps of detachment area and 3-dimensional maps of detachment volume	Youden index, Sorensene–DSC, sensitivity, specificity, R-squared	Automated quantification and visualization of DMEK GD to support monitoring and RB (intervention) decisions.
Joseph et al. (2023), with an earlier conference report by Joseph et al. (2021), USA [27, 28]	Full publication: 925 central EC images from 44 post-DMEK eyes, comprising 22 eyes with subsequent graft rejection and 22 age- and sex-matched control eyes. Images from the last and second-to-last time points before rejection were analysed. The earlier conference report described a preliminary analysis of 171 EC images acquired 1–12 months before graft-rejection diagnosis.	Not reported; rejection and control groups in the full publication were age-matched	Not reported; rejection and control groups in the full publication were sex-matched	Full publication: DMEK. Earlier conference report: full- and partial-thickness keratoplasty, without procedure-specific numbers	Specular-microscopy EC images with image-derived morphometric, intensity, shape, spatial-arrangement, cell-graph, and longitudinal features	Supervised ML with deep-learning-assisted image segmentation	Binary classification of future graft rejection vs. non-rejection/control eyes; analysis of single-time-point and longitudinal delta-radiomic features	Earlier conference report: U-Net-assisted segmentation, mRMR feature selection, and RFC. Full publication: U-Net-assisted segmentation, multistep feature selection, RF, and LR classification	U-Net segmentation model; RF classifier with 100 classification trees; and binomial LR classifier	Image-level feature extraction with patient-eye-level classification and prediction	Image cropping, pixel-resolution normalization, illumination-gradient correction, U-Net segmentation followed by guided manual correction, extraction of image-derived features, removal of highly correlated features using Spearman correlation, z-score normalization, and patient-level separation of training and held-out test sets. The full study used 10 iterations of threefold cross-validation; the earlier report used fivefold cross-validation and mRMR feature selection.	Full publication: Topcon SP3000 non-contact specular microscope (Topcon Europe Medical, Capelle aan den IJssel, Netherlands). Imaging instrument not specified in the earlier conference report	Analysis software not reported. Aim: to develop ML classifiers using EC images to identify postkeratoplasty patients at risk of future graft rejection before clinical manifestation.	Earlier conference report: AUC, sensitivity, and specificity using fivefold cross-validation. Full publication: classification accuracy from repeated threefold cross-validation and a patient-eye-level held-out test set, with comparison of novel image-derived features against traditional morphometric measures	Early identification of postkeratoplasty patients at risk of graft rejection to support closer surveillance, earlier intervention, and individualized treatment planning before clinically apparent rejection.
Heslinga et al. (2021), Netherlands [29]	960 AS-OCT B-scans from 50 patients (derived from 80 AS-OCT scans of 68 patients); training set=752 images, validation set=208 images, test set=320 images (AS-OCT B-scan imaging data obtained from a randomized controlled trial.)	Not reported	Not reported	DMEK	AS-OCT B-scan images	Supervised deep Learning	Supervised corneal interface delineation/segmentation and pachymetry measurement	Patch-based CNN, U-Net, CNN with dimension reduction	Patch-based CNN (5 trainable layers); U-Net with 4 downsampling/upsampling blocks, batch normalization and residual layers; CNN with dimension reduction using downsampling/upsampling paths and funneling subnetworks	Pixel-level interface delineation/segmentation and image-level pachymetry mapping	Deep-learning scleral spur localization; horizontal alignment; cropping centered on corneal apex; patch extraction (patch CNN); splitting images into superior/inferior halves; rotation, translation, noise augmentation, Gaussian blur, contrast augmentation (model dependent)	Swept-source AS-OCT (CASIA2, Tomey Corp., Nagoya, Japan)	Keras with TensorFlow backend; Adam optimizer/ Automatically delineate corneal interfaces and accurately measure corneal thickness in post-DMEK AS-OCT scans; construct pachymetry and differential pachymetry maps	MAE of corneal thickness measurements within 3-, 6-, and 9-mm diameters; five repeated training runs with different random seeds; comparison against three independent manual annotation sets; inter-observer variability assessment; and one-way ANOVA for model performance comparison.	Automated pachymetry and differential pachymetry mapping for monitoring restoration of endothelial function and postoperative recovery after DMEK
Feng et al. (2024), China [39]	Kera-Dataset consisting of 28 videos of actual human PK surgeries; 1000 annotated frames used for segmentation (900 training, 100 validation images)	Not reported	Not reported	PK	Intraoperative surgical microscope video images / video frames	Deep learning (for segmentation) combined with rule-based spatio-temporal task estimation framework	Multi-object segmentation, geometric feature extraction, surgical phase recognition, and instrument path planning/prediction	ST-ITEF; pixel-wise segmentation; geometric feature extraction; surgical phase estimation, instrument path planning	The state-of-the-art (SOTA) deep CNN models: DeepLabV3, FCN8s, U-Net, and UNet++ evaluated. Optimized (filtered) DeepLabV3 selected as best-performing segmentation model	Pixel-level segmentation; object-level geometric feature extraction; phase-level recognition; tool-path level prediction	Image filtration (post-processing), image resizing (384×640), random rotation, flipping, color jittering, normalization; object tracking and geometric feature extraction	Leica M844 surgical microscope (Leica Microsystems, Wetzlar, Germany) with 3CCD color video camera (Sony Group Corp., Tokyo, Japan) for hospital-acquired videos; additional videos obtained from online video sources	PyTorch; experiments run on two NVIDIA GeForce GTX 3090 GPUs/ Recognize surgical phases and predict surgical instrument paths during PK to provide cognition guidance for surgical robotics	C-PA; C-IoU for segmentation; MPE for instrument localization; stitch-point error; phase-level Jaccard for surgical phase recognition; accuracy, precision, recall, and Jaccard for phase-level performanc.	Computer-assisted cognition guidance, surgical phase recognition, and instrument path planning for robotic-assisted PK
Dreesbach et al. (2024), Germany [40]	9375 consecutive archived phase-contrast microscopic images from 6168 corneal grafts collected between 2014–2023	Not reported	Not reported	Corneal transplantation / keratoplasty graft quality assessment and release for ECD.	Phase-contrast microscopic images of donor corneal endothelium	Supervised deep learning	Automated endothelial cell segmentation, cell detection, and ECD quantification	Deep learning–based automated endothelial cell segmentation and counting; comparison with the Rhine-Tec Endothelial Analysis System.	U-Net architecture	Image-level endothelial cell segmentation and graft-level ECD estimation	Manual annotation of 100 endothelial images for training; conversion of images into three-channel probability maps (red = cell body, green = cell boundary, blue = non-evaluable area), followed by binary image generation for cell-center calculation and automatic exclusion of non-evaluable image regions.	Nikon ECLIPSE TE 2000-S phase-contrast microscope; images digitized at 640 × 480 pixels using a video digitizer	R (for statistical analyses)/ To compare a fully automated deep-learning method for ECD measurement with the conventional Rhine-Tec Endothelial Analysis System and evaluate agreement and validity between the methods.	Mean difference analysis; correlation analysis; Bland–Altman agreement analysis; confusion-matrix agreement assessment; and subgroup analyses for medium 1 and medium 2.	Objective, automated quality control of donor corneas ECD and support for graft-release decisions in keratoplasty
Vigueras-Guillén et al. (2020), Netherlands [41]	Main dataset comprised 383 post-UT-DSAEK specular microscopy images from 41 eyes of 41 patients. Training was assisted by a secondary dataset consisting of 400 manually annotated specular microscopy images from a Baerveldt glaucoma device implantation study (trial registration: NL4823).	73 ± 7 y (average age at time of surgery surgery)	Not reported	UT-DSAEK	Specular microscopy images of the corneal endothelium; 8-bit grayscale images, 240 × 528 pixels	Deep learning (supervised CNN-based segmentation)	Automated endothelial cell segmentation, region of interest (ROI) detection, and estimation of ECD, CV, and HEX	CNN-Edge for endothelial cell edge detection; CNN-ROI for region-of-interest detection; watershed-based postprocessing; parameter computation (ECD, CV, HEX)	Dense U-Net architecture with batch renormalization, exponential linear unit activations, average pooling, and dropout layers	Image-level endothelial cell segmentation and ECD/morphometric parameter estimation	Manual segmentation/annotation using image manipulation program GIMP (version 2.10); five-fold cross-validation; left-right and up-down flipping data augmentation; ROI annotation; edge probability image generation; smoothing based on Fourier-derived cell size; watershed segmentation; exclusion of border cells and cells with <75% ROI coverage	Noncontact specular microscope Topcon SP-1P (Topcon Co., Tokyo, Japan) with IMAGEnet i-base v1.32 software	Python (version 3.7), TensorFlow (version 2.10); Matlab 2018a (MathWorks, Natick, MA)/ To develop a fully automatic deep learning method for estimating ECD, CV, and HEX from specular microscopy images and evaluate its use in 1-year follow-up after UT-DSAEK	Success rate and MAE; Wilcoxon test, ANOVA, Bland–Altman agreement analysis, Pearson correlation, paired t-tests, Shapiro–Wilk normality test, and Levene's test for homogeneity.	Enables reliable automated assessment of ECD, CV, and HEX after UT-DSAEK and supports clinical monitoring of endothelial healing and graft status following corneal transplantation

Bitton et al. (2022), France [42]	1992 AS-OCT images from 290 eyes of 255 patients (50 normal eyes and 240 pre-DMEK eyes) from a corneal graft registry	Normal: 42.1 ± 19.5 y FECD: 68.8 ± 10.3 y Post-cataract decompensation: 75.3 ± 9.8 y Post-ACIOL/Artisan: 71.2 ± 14.9 y Total pre-DMEK: 71.1 ± 11.0 y	Not reported	DMEK for FECD and other causes of endothelial dysfunction	AS-corneal OCT images (radial scans)	Supervised deep learning using U-Net-based segmentation models	Pixel-wise corneal edema detection/classification and edema quantification using EF	Pixel-level classification of OCT images into normal or edema; EF calculated from edema-labelled pixels; ROC-based threshold determination using DCCT, threshold optimization for edema detection	U-Net models: (1) epithelial detection, (2) stromal edema detection on epithelium-removed image, (3) edema detection on original image with epithelial prediction retained; final prediction combines stromal and epithelial outputs	Pixel-level analysis with patient-level EF aggregation and classification	9-mm scans cropped to central 1020 pixels to match 6-mm scans; epithelium removal before stromal analysis; binary mask generation; largest connected component retained to remove isolated artifacts; no other preprocessing applied	Avanti OCT (RTVue, Optovue, Fremont, CA); Scheimpflug Pentacam used for comparative analysis in a subgroup	Python 3.6, PyTorch; EasyMedStat (version 3.4); Stata; Seaborn, and Matplotlip libraries in Python 3.6/ To evaluate a deep-learning approach for objective detection of minimal corneal edema before DMEK surgery and support surgical decision-making	ROC analysis (AUC, sensitivity, specificity, and optimal threshold determination) for edema detection; Pearson correlation analysis; and comparative performance assessment across different DCCT-based edema thresholds and patient subgroups.	Facilitates objective detection of subclinical corneal edema before DMEK, improving OCT interpretation and supporting surgical decision-making, including selection of DMEK, triple procedures, or cataract surgery alone
Heslinga et al. (2020); Netherlands/Denmark [30]	1280 AS-OCT B-scans from 80 AS-OCT scans of 68 participants (as part of a randomized controlled trial); each scan contained 16 radial B-scans (2133 × 1466 pixels)	Training/validation set: 70.2 ± 7.47 y Test set: 73.3 ± 6.10 y	Training/validation: 27/23 Test set: 6/12	DMEK	Human AS-OCT clinical imaging data acquired immediately after surgery or postoperative day 7	Supervised deep learning	Scleral spur localization, GD segmentation, biomarker extraction, and 2D detachment-map generation	Regression-based scleral spur localization, ellipse fitting, semantic segmentation of detached graft regions, skeletonization-based detachment length measurement, horizontal projection mapping, Dice-score evaluation	Four-step deep learning pipeline: ResNet-50-based scleral spur localization model, ellipse-fit refinement, U-Net segmentation model for detached graft areas, and biomarker extraction.	B-scan/image level, pixel-level segmentation, and scan-level detachment mapping	AS-OCT B-scans are converted to grayscale and resized to 512×352 for localization. They are centered via scleral spur estimates, cropped to 1920×768, split into inferior and mirrored superior halves, and down-sampled to 480×384. Data is augmented using random rotation, translation, and scleral-spur perturbations, with annotation masks generated from expert point markings.	Swept-source AS-OCT (CASIA2; Tomey Corp., Nagoya, Japan)	Models were implemented in Keras using a TensorFlow backend (statistical software not separately specified)/ Automatically locate and quantify GD after DMEK and generate objective detachment maps to support postoperative management	Scleral spur localization error assessment; GD length estimation agreement analysis; segmentation performance evaluation using DSC; comparison with human expert annotations and inter-rater performance.	Automated, instant localization and quantification of GD, generation of detachment maps, support for RB/repeat-DMEK decisions, and standardization of postoperative DMEK assessment
O'Brien et al. (2021), USA [43]	Data from the Cornea Preservation Time Study (CPTS): 1090 participants representing 1330 eyes undergoing DSAEK; 50 donor, recipient, eye bank, and intraoperative variables evaluated (40 clinical sites, 70 surgeons, 23 eye banks).	Median age 70 y (range 42–90 y)	663 women (60.8%); remaining participants male	DSAEK	Multicenter randomized clinical trial/cohort data; donor, recipient, eye bank, operative, and follow-up survival data	Supervised ML (RSF)	Time-to-event survival prediction, variable importance ranking, adjusted graft survival estimation, RMST analysis	RSF, variable hunting using VIMP ranking, RMST estimation, subsampling-based confidence intervals (CIs)	50 candidate variables considered. RSF ensemble of survival trees; final model contained 5 predictors including DSAEK intraoperative complication, surgeon, eye bank, donor diabetes history, and preservation to surgery (in days)e	Eye-level graft survival outcome (GF time-to-event)	Structured CPTS database variables; variable selection using VIMP-based variable hunting; missing donor-thickness variables noted but no imputation procedure reported. Preservation time forced into model selection, subsampling procedure for CIs	Not applicable (no imaging data analyzed)	SAS, version 9.4 (SAS Institute Inc) for data preparation; R, version 3.6.3 (R Foundation for Statistical Computing) with randomForestSRC package for analysis/Reanalyze intraoperative complications associated with DSAEK GF using RSF and obtain adjusted graft survival estimates	Model discrimination assessment using Harrell's C-statistic; variable importance ranking with RSF; adjusted survival analysis using RMST comparisons and Cis estimation.	Risk stratification and identification of predictors of DSAEK GF to support clinical decision-making and future ophthalmic surgical studies

Abbreviations: M, male; F, female; ML, machine learning; OCT, optical coherence tomography; y, years; PK, penetrating keratoplasty; LKP, lamellar keratoplasty; DALK, deep anterior lamellar keratoplasty; DSAEK, Descemet stripping automated endothelial keratoplasty; DMEK, Descemet membrane endothelial keratoplasty; PCA, principal component analysis; t-SNE, t-distributed stochastic neighbor embedding; AS-OCT, anterior segment OCT; RB, rebubbling; GD, graft detachment; AUC, area under the curve; CI, confidence interval; RF, Random forest; IOP, intraocular pressure; OHT, ocular hypertension; AUC-ROC, area under the receiver operating characteristic curve; AUC-PR, area under the precision-recall curve; FECD, Fuchs endothelial corneal dystrophy; BK, bullous keratopathy; OOB, out-of-bag; RSF, random survival forest; VIMP, variable importance; GF, graft failure; SD-OCT, spectral-domain OCT; SBB, successful big bubble; FBB, failed big bubble; VGG, visual geometry group; SVM, support vector machine; ANN, artificial neural network; ANOVA, analysis of variance; DSC, Dice similarity coefficient; LASSO, L1-regularized logistic regression; CTA, classification tree analysis; ROSE, random oversampling of examples; RFC, random forest classification; IQR, interquartile range; CNN, convolutional neural network; EC, endothelial cell; mRMR, minimal redundancy maximal relevance; LR, logistic regression; MAE, mean absolute error; ST-ITEF, spatio-temporal intraoperative task estimating framework; C-PA, Class Pixel Accuracy; C-IoU, Class Intersection-over-Union; MPE, mean position error; ECD, endothelial cell density; UT-DSAEK, ultrathin DSAEK; CV, coefficient of variation; HEX, hexagonality; EF, edema fraction; DCCT, differential central corneal thickness; 2D, two-dimensional; RMST, restricted mean survival time. Note: ARA-comb, an integrated variable of angle recess area (ARA) and iridotrabecular contact (ITC). Note: References 29 and 30 used the same RCT-derived cohort of 80 AS-OCT scans from 68 participants for distinct AI tasks; they represent separate analyses but not independent patient populations. References 20 and 21 originated from the same Singapore registry programme but evaluated distinct indication and procedure cohorts and were treated as separate synthesis entries.

Table 2. Included studies – Summary of results

Author (Year), Country	Results summary	Clustering/classification	Performance metrics	Clinical relevance
Yousefi et al. (2020), USA/Japan [31]	Unsupervised ML identified five non-overlapping clusters of eyes from OCT-derived corneal parameters. Post hoc mapping of eyes that later underwent keratoplasty showed different likelihoods of future surgery across clusters. Cluster 3 showed the highest likelihood of future keratoplasty.	Principal component analysis reduction followed by t-distributed stochastic neighbor embedding (t-SNE) mapping and density-based unsupervised clustering identified five clusters corresponding to distinct corneal parameters and ectasia severity patterns. Post hoc mapping estimated likelihood of future keratoplasty across clusters.	Five clusters identified; 18 significant principal components retained > 70% of total variance. Normalized likelihood of future surgery varied by cluster: overall surgery likelihood ranges 2.9%–45.3% (3.1%, 2.9%, 45.3%, 29.5%, 19.2%). Endothelial keratoplasty likelihood reached 68.7% in cluster 3 (cluster 1 = 4.3%, cluster 2 = 5.2%, cluster 4 = 21.8%, cluster 5 = 0%).	AI-assisted identification of patients at higher risk for future keratoplasty using corneal shape, thickness, and elevation parameters; supports monitoring, risk stratification, and surgical decision-making.
Hayashi et al. (2023), Japan/Germany [32]	An EfficientNet-based AI algorithm trained on OCT images from one surgeon accurately predicted the need for post-DMEK RB from postoperative OCT images when applied to patients operated on by a different surgeon, showing good transferability across surgeons.	Binary classification of eyes into RB vs. non-RB based on postoperative OCT-derived GD patterns, with external validation on an independent second surgeon's dataset.	AUC 0.875 (95% CI 0.880–0.929), optimal cutoff (YI) 0.214, sensitivity 78.9% (67.6–87.7%), specificity 78.6% (69.5–86.1%).	Provides objective postoperative decision support for identifying patients likely to require post-DMEK RB and may assist inexperienced surgeons in management decisions about GD evaluation, need for RB, and postoperative treatment guidance, thus facilitating standardized management and providing objective treatment guidance.
Kundu et al. (2025) India/Netherlands [33]	AI accurately classified DALK outcomes as favorable or unfavorable using preoperative clinical risk factors and Pentacam-derived tomography parameters. The model achieved correct classification rates of 92.2% and 89.4% for favorable and unfavorable outcomes, respectively.	RF classification of favorable vs. unfavorable long-term DALK outcomes; AI-based ranking of influential risk factors.	AUC 0.957, sensitivity 98%, specificity 86%, accuracy 91.4%, precision 91.3%, recall 0.91, F1 score 0.92. The model correctly classified 92.2% of favorable and 89.4% of unfavorable eyes.	Supports pre-DALK preoperative risk stratification and identifies factors (systemic allergy, serum IgE level, eye rubbing, vitamin D level, posterior corneal defocus, ocular allergy, posterior corneal spherical aberration, central keratoconus index, maximum anterior corneal curvature, and minimum corneal thickness) associated with long-term graft outcomes.
Kim et al. (2025), Republic of Korea [34]	The outcome was defined as intraocular pressure ≥ 22 mmHg or an increase > 10 mmHg from baseline preoperative values. An ensemble ML model predicted post-DMEK ocular hypertension (OHT) using preoperative clinical and AS-OCT biometric features. The model identified a clinically predictable high-/low-risk subgroup (61/106 eyes; 57.5%) and stratified patients into high- and low-risk groups.	Ensemble classification combining XGBoost, RF, CatBoost, and logistic regression models; feature importance analysis and Kaplan-Meier/Cox survival analyses were also performed.	Accuracy 80.2%, precision 60.0%, recall 59.7%, AUC-ROC 82.3%, AUC-PR 68.0%, Kaplan-Meier log-rank P-value 0.0089.	Preoperative identification of patients at increased risk for post-DMEK OHT, supporting individualized postoperative monitoring and treatment planning.
Ang et al. (2022), Singapore [20]	RSF analysis identified key predictors of 10-year graft failure/survival in 1,335 consecutive keratoplasty cases (946 DSAEK and 389 PK), generated an optimal Cox regression model for risk prediction, and confirmed superior long-term outcomes of DSAEK compared with PK.	RSF modeling with variable importance (VIMP) ranking and pairwise interaction analysis, followed by multivariable Cox proportional hazards regression for survival prediction and risk stratification.	Best out-of-bag (OOB) Harrell C-index 0.701; bullous keratopathy associated with GF (hazard ratio 2.838, 95% CI 1.89–4.26, P < 0.001); PK associated with GF vs. DSAEK (hazard ratio 1.643, 95% CI 1.19–2.27, P 0.002); 10-year graft survival outcome for DSAEK 73.6% vs PK 50.9% (log-rank P < 0.001). Lower rejection, epitheliopathy, and wound dehiscence rates supporting DSAEK outperformance of PK.	Identifies major risk factors for long-term GF, supports surgical decision-making between DSAEK and PK, and informs prediction of graft survival and postoperative complication risk assessment.
Ng Yin Ling et al. (2025), Singapore [21]	RSF-guided analysis identified five principal predictors of 10-year graft survival: recipient sex, preoperative visual acuity, surgical procedure, surgical indication, and active inflammation. These variables were converted into an interpretable AutoScore risk score. Ten-year graft survival was 92.3% after DALK and 56.6% after PK. At the endpoint, survival was 73.6% in the low-risk group and 39.0% in the high-risk group.	No unsupervised clustering was performed. RSF and nested Cox models were used for time-to-event prediction and variable selection. AutoScore classified patients into low- and high-risk groups.	Internal test set for cumulative 10-year survival: ML C-index 0.72 (95% CI 0.66–0.77) and iAUC 0.73 (95% CI 0.67–0.78). For survival beyond five years, the ML model achieved a C-index of 0.76 (95% CI 0.62–0.89) and iAUC of 0.75 (95% CI 0.58–0.90), compared with 0.71 and 0.69, respectively, for traditional modelling. Low- versus high-risk survival differed in both development and test sets for cumulative 10-year survival (P < 0.001); however, beyond-five-year risk groups differed significantly in the development set but not in the smaller test set.	The interpretable risk score may support preoperative counselling and long-term risk stratification after PK or DALK. External validation is required before clinical implementation.
Feizi et al. (2024), Iran [23]	Five supervised ML models demonstrated similar, modest discrimination for graft-rejection prediction. Keratoplasty technique was identified as a predictor by all models. Other frequently selected factors included donor age, transplantation in the fellow eye, donor endothelial cell density, corticosteroid duration, time to complete suture removal, and suture-related complications (identified by four of five models), with patient age also implicated in three models. Overall, 341 of 1214 eyes (28.1%) experienced at least one rejection episode.	Binary classification of graft rejection versus no rejection using ANN, SVM, gradient boosting, and extra-trees models, together with RSF time-to-event modelling. No clustering was performed.	C-statistics: ANN 0.671, SVM 0.664, gradient boosting 0.675, extra trees 0.674, and RSF 0.672. Brier scores were 0.174, 0.183, 0.182, 0.183, and 0.194, respectively. Test accuracies were 0.763, 0.763, 0.758, 0.759, and 0.742; cross-validation accuracies were 0.794, 0.786, 0.787, 0.792, and 0.777, respectively. Precision ranged from 0.525 to 0.569.	The models identified potentially modifiable postoperative factors associated with graft rejection. However, the findings are primarily hypothesis-generating because several predictors were measured after surgery or close to the rejection event, and external validation was not undertaken.
Hayashi et al. (2021), Japan/Germany [35]	A deep learning model using preoperative spectral domain OCT images predicted successful vs. failed big bubble (SBB vs. FBB) formation during DALK. The study showed the feasibility of successful automated preoperative prediction of big bubble.	Binary classification of SBB vs. FBB formation using a VGG-16 CNN with softmax output; Score-CAM method used to visualize image regions influencing predictions.	AUC 0.746 (95% CI 0.603–0.889); determined success rate 78.3% (18/23 eyes; 95% CI 56.3–92.5%) for SBB and 69.6% (16/23 eyes; 95% CI 47.1–86.8%) for FBB.	Supports preoperative prediction of SBB formation during DALK and may aid surgical planning and patient identification; more likely to achieve anatomical surgical success.
Hayashi et al. (2020), Japan [36]	Deep learning analysis of postoperative day-5 AS-OCT images accurately identified eyes requiring post-DEMK RB, showing the feasibility of automated RB decision support.	Binary classification of RB vs. non-RB eyes using AS-OCT images; nine CNN architectures were evaluated, VGG19 and ensemble models showed the best performance.	Best single model (VGG19): AUC 0.964, sensitivity 0.967, specificity 0.915. Best ensemble model (VGG16 + VGG19 + DenseNet121 + DenseNet169 + DenseNet201): AUC 0.956, sensitivity 0.913, specificity 0.921.	Supports early identification of eyes likely to require post-DEMK RB and may assist postoperative management and intervention decisions.

Karaca et al. (2025), Turkey [37]	Five ML models (support vector machine, RF, RUSBoosted Tree, k-nearest neighbor (kNN) and ANNs [1-ANN, 2-ANN, 3-ANN]) were trained to predict early GF within 12 months post-DMEK using donor, graft, and perioperative characteristics. The ML methods could effectively predict GF; the single-hidden-layer ANN (1-ANN) achieved best overall performance.	Binary classification (GF vs. no GF) using five models. Feature importance was assessed using ANOVA.	Performance metrics for best model: 1-ANN (accuracy 0.96, sensitivity 0.82, specificity 1.00, F1 score 0.90, AUC 0.99). Additional findings: death-to-preservation time ANOVA importance score 37.702, ICU duration ANOVA importance score 24.856. Death-to-preservation time and ICU duration were significant predictors for GF ($P < 0.05$).	The ML models can predict post-DMEK GF and may support preoperative risk stratification, donor tissue selection, personalized surgical planning, and clinical decision-support for DMEK patients. Identification of patients at higher risk of GF for closer monitoring and intervention.
Muijzer et al. (2022), with an earlier conference report by Wisse et al. (2021), Netherlands [22, 26]	ML models identified procedure-, donor-, recipient-, and practice-pattern factors associated with postoperative GD/RB following DMEK or DSEK. DMEK was consistently associated with increased GD risk, while other important factors included previous GF, insertion device, SF6 use, graft size, supine positioning, preoperative laser iridotomy, and donor-tissue preparation, although their effects varied across models. The earlier conference report from the same registry cohort additionally identified factors associated with GF.	Supervised prediction/classification of GD/RB vs. no GD/RB and GF vs. no GF. The full publication used L1-LASSO, CTA, and RF classification. The earlier conference report used LASSO and CTA after ROSE-based class balancing.	Full publication: LASSO, AUC 0.70, sensitivity 0.70, and specificity 0.65; CTA, AUC 0.65, sensitivity 0.65, and specificity 0.62; RF classification, AUC 0.72, sensitivity 0.68, and specificity 0.62. Earlier conference report: explained variance with LASSO/CTA was 0.68/0.66 for RB and 0.66/0.77 for GF. RB incidence was 17.4% after DMEK and 7.1% after DSEK; GF incidence was 8.4% and 3.8%, respectively.	Identification of predictive and protective factors associated with GD, RB, and GF may support risk stratification, optimization of surgical practice patterns, development of strategies to reduce postoperative complications, and hypothesis generation for future prospective studies.
Glatz et al. (2021), Germany [38]	A deep learning segmentation model (UNet++) was developed to automatically detect, quantify, and visualize post-DMEK GD from AS-OCT images and generate 3D detachment maps. The model was trained on 27 scans (6912 labeled images), validated on 20 scans (5120 labeled images), and applied to 107 eyes from a prospective dataset. The network was more sensitive than the cornea specialists' slit-lamp–based assessment for significant GDs.	The study performed semantic image segmentation using UNet++ to classify and quantify GD from 256 radial AS-OCT slices per scan. These pixel-level segmentations were aggregated to generate 2D color-coded area maps and 3D volumetric maps of the detachment. Model validation was performed against manual expert annotations (ground truth) and clinical slit-lamp assessments. Last, the application cohort was stratified and categorized clinically as: no detachment, < 1/3 detachment, and ≥ 1/3 detachment.	Training set (sensitivity 0.998, specificity 0.997, YI 0.994, DSC 0.767); area-detachment map performance (YI 0.994, DSC 0.836, sensitivity 0.970, specificity 0.965); overall performance (R-squared 0.96). Validation set (YI 0.850, DSC 0.729, sensitivity 0.854, specificity 0.996); area-detachment map performance (YI 0.908, DSC 0.833, sensitivity 0.955, specificity 0.953); overall performance (R-squared 0.90). Application cohort (107 eyes): mean difference vs. clinician assessment 8.2% (95% CI 6.2–10.2). Overall agreement vs. reannotated consensus: R-squared 0.85.	The model detected and quantified DMEK GD with high precision, outperforming standard slit-lamp examinations which significantly underestimated true detachment areas (median 8.5% vs. 0.0%). This clinical underestimation was most pronounced in flat/planar detachments and eyes already showing clinical indications for intervention. Automated 3D topological mapping addresses human visual limitations with complex radial data, directly supporting objective postoperative assessment, localized monitoring, and timely RB decision-making.
Joseph et al. (2023), with an earlier conference report by Joseph et al. (2021), USA [27, 28]	The earlier conference report used a RF classifier trained on more than 190 morphometric features extracted from 171 endothelial cell images to predict graft rejection 1–12 months before clinical diagnosis. Cell borders were segmented using a U-Net–based approach followed by manual correction, with mRMR feature selection and 5-fold cross-validation. The subsequent full publication expanded the analysis to 432 endothelial image features, including cell-graph spatial arrangement, intensity, and shape features, and evaluated longitudinal images from 44 post-DMEK eyes. RF- and LR-based models predicted graft rejection 1–24 months before clinical manifestation and generally outperformed models based on traditional morphometric measures, including ECD, CV, and percentage of hexagonal cells.	Binary classification of future graft rejection vs. non-rejection/control eyes. The earlier conference report used U-Net-assisted endothelial cell segmentation, mRMR feature selection, an RF classifier, and fivefold cross-validation. The full publication evaluated RF and LR classifiers using features from the second-to-last and last imaging time points and delta-radiomics, with repeated threefold cross-validation and a patient-level held-out test set. No clustering methods were used.	Earlier conference report: AUC 0.87 ± 0.03, sensitivity 0.86 ± 0.12, and specificity 0.86 ± 0.10 using fivefold cross-validation. Full publication—held-out test set: at T1 (second-to-last imaging timepoint), best RF accuracy was 0.71 (3 or 5 features) and best LR accuracy was 0.64 (5 or 7 features and traditional); at T2 (last imaging timepoint before rejection), best RF and LR accuracies were 0.83 (5, 7, and 10 features) and 0.85 (3, 5, and 7 features), respectively; for delta-radiomics, best RF and LR accuracies were 0.77 (5 features) and 0.50 (traditional), respectively. Novel-feature models generally outperformed traditional morphometric models; RF at T1 vs. traditional morphometrics, $P < 0.001$, and LR at T2 vs. traditional morphometrics, $P < 0.05$. Frequently selected predictors included minimum-spanning-tree edge-length standard deviation, mean cell circularity, image-pixel intensities of 163.2–173.4, and average cell circularity.	Enables preclinical identification of patients at risk of graft rejection before clinically apparent manifestations, potentially supporting closer surveillance, earlier corticosteroid intervention, and individualized treatment planning to reduce endothelial damage, GF, repeat keratoplasty, vision loss, patient burden, and healthcare costs.
Heslinga et al. (2021), Netherlands [29]	Three deep-learning approaches (patch-based CNN, U-Net, and CNN with dimension reduction) successfully automated delineation of anterior and posterior corneal interfaces in post-DMEK AS-OCT scans, enabling accurate corneal thickness measurement and generation of pachymetry and differential pachymetry maps. Across models, mean absolute thickness measurement errors for the central 9-mm cornea ranged 13.98–15.50 µm, corresponding to < 3% error relative to average corneal thickness.	No clustering performed. No diagnostic/prognostic classification performed. AI task: deep-learning segmentation/delineation of corneal interfaces (anterior and posterior boundaries) in AS-OCT images using 3 deep learning approaches.	No report of accuracy, sensitivity, specificity, AUC, or F1 score. Primary outcome was segmentation/delineation performance measured through corneal thickness MAE (µm); corneal thickness MAE: 13.84–15.50 µm across models and diameters; 9-mm MAE 13.98–15.50 µm; performance comparable to or better than interobserver variability. Independent test set (320 B-scans): corneal thickness for 3 mm MAE (patch-based CNN 14.40 ± 0.69 µm, U-Net 13.94 ± 0.38 µm, CNN with dimension reduction 13.94 ± 0.25 µm); 6 mm MAE (patch-based CNN 14.80 ± 0.51 µm, U-Net 13.84 ± 0.22 µm, CNN with dimension reduction 14.17 ± 0.22 µm); and 9 mm MAE (patch-based CNN 15.50 ± 0.59 µm, U-Net 13.98 ± 0.15 µm, CNN with dimension reduction 14.40 ± 0.15 µm). Additional findings: errors corresponded to < 3% of average corneal thickness. Interobserver MAE range 13.66–23.71 µm, indicating model performance comparable to or better than observer agreement. Built-in OCT software produced clinically relevant thickness inaccuracies in 90/320 B-scans (28%), with severe errors in 37/320 B-scans (12%); 16/20 AS-OCT scans (80%) contained at least a clinically relevant thickness inaccuracy when using the built-in software.	Automated deep-learning pachymetry enables accurate objective monitoring of post-DMEK corneal thickness restoration, including generation of differential pachymetry maps to track regional changes over time. These tools may improve understanding of endothelial recovery, graft attachment, postoperative monitoring, and post-DMEK surgery management through automated pachymetry mapping.
Feng et al. (2024), China [39]	ST-ITEF is a three-stage framework combining multi-object segmentation, geometric feature extraction, surgical phase recognition, and instrument path prediction for PK. The framework uses multi-object tracking of surgical instruments and corneal structures to recognize surgical phases and predict surgical instrument paths. The optimized DeepLabV3 with image filtration achieved the best segmentation performance and supported subsequent phase recognition and path planning.	No clustering analysis performed. Classification task: surgical phase recognition/classification of 10 predefined PK surgical phases (P0–P9). Additional AI task: multi-object semantic segmentation and instrument path prediction/planning.	Several performance measures reported. 1) Multi-object segmentation (best model: filtered DeepLabV3); best segmentation C-IoU (recipient 78.7%, graft 91.2%, forceps 75.0%, holder 88.1%, needle 49.5%, thread 6.3%). 2) Surgical phase recognition, mean phase Jaccard index 55.58% (best among compared methods). 3) With filtration vs. without filtration: (phase recognition accuracy 94.50% vs. 90.85%, precision 75.91% vs. 62.97%; recall 70.60% vs. 60.17%, Jaccard 55.58% vs. 43.88%). 4) Instrument localization (F-DeepLabV3): needle tip MPE 0.27 mm; needle center MPE 0.41 mm. 5) Stitch-point determination: mean stitch-point error 0.15 ± 0.06 mm using F-DeepLabV3.	The ST-ITEF framework has potential for computer-assisted cognition guidance, surgical navigation, intraoperative task estimation, and guidance of surgical robots during keratoplasty. By recognizing surgical phases, tracking instruments and corneal structures, and predicting tool paths, the framework may support ophthalmologists and robotic systems during keratoplasty.
Dreesbach et al. (2024), Germany [40]	The authors developed and evaluated a fully automated deep learning method based on U-Net architecture for ECD measurement. The system automatically detected all visible endothelial cells within the selected image without a fixed counting frame and was compared against the conventional Rhine-Tec Endothelial Analysis System using 9375 archived phase-contrast microscopy images. The two methods yielded similar mean ECD measurements with significant correlation and good overall agreement.	No clustering analysis or classification task performed. AI task: automated endothelial cell segmentation/detection and quantitative ECD measurement, followed by statistical comparison with the Rhine-Tec fixed-frame method.	1) Primary comparison results: mean difference (Rhine-Tec vs. deep learning) −23 cells/mm^2 (95% CI −29 to −17 cells/mm^2); statistically significant correlation coefficient + 0.748 ($P < 0.001$); Bland-Altman limits of agreement −602 to +557 cells/mm^2, with deviation clustering around the clinically relevant ECD range of 2000–2500 cells/mm^2. 2) Subgroup analyses: correlation coefficients 0.716 (medium 1) and 0.775 (medium 2); mean differences −35.9 cells/mm^2 (95% CI −44.72 to −27.04) and −8.79 cells/mm^2 (95% CI −16.81 to −0.78), respectively. 3) Agreement analysis based on confusion matrices: overall agreement 83.4% (4226 images; medium 1) and 86.1% (3526 images; medium 2). 4) Bland-Altman analysis yielded deviation clustering in the clinically relevant ECD range of 1500–2500 cells/mm^2, with generally higher cell-density estimates obtained using the Rhine-Tec system.	Deep learning method provides a valid and potentially more objective approach to ECD quantification for corneal graft evaluation than the fixed-frame Rhine-Tec method. By evaluating all visible endothelial cells rather than a small frame of 15–40 cells, the approach may improve objectivity and reproducibility in graft quality assessment. However, the system does not replace physician-based overall endothelial quality assessment, which remains essential for graft release decisions.
Vigueras-Guillén et al. (2020), Netherlands [41]	Fully automatic deep learning pipeline, trained on UT-DSAEK and Baerveldt datasets, consisting of CNN-edge for endothelial cell edge detection, CNN-region of interest (CNN-ROI) for automatic region-of-interest selection, and watershed-based postprocessing for final segmentation. The system automatically estimated ECD, CV, and HEX from post UT-DSAEK specular microscopy images and was compared with Topcon software and manual segmentation. The method successfully generated ECD/CV estimates in 98.4% of images and HEX estimates in 99.7% of eligible images.	No clustering analysis or classification task performed. AI task: automated image segmentation and quantitative parameter estimation (ECD, CV, HEX) from specular microscopy images up to 1 year after UT-DSAEK.	1) Detection performance: average cells detected per image for DL 115 and Topcon 30 ($P < 0.0001$). 2) Success rates: ECD/CV estimation for DL 98.4% and Topcon 71.5%; HEX estimation for DL 99.7% and Topcon 30.5%. 3) Overall MAE for DL (ECD 2.5%, CV 5.7%, HEX 5.7%) was significantly better than Topcon (ECD 7.5%, CV 17.5%, HEX 18.3%) ($P < 0.0001$). 4) MAE (selected images): ECD 30.0 ± 38.5 cells/mm^2; CV 1.5 ± 2.0%; HEX 3.0 ± 3.2%. DL estimates were significantly better than Topcon ($P < 0.0001$), with no statistically significant difference from manual gold-standard assessments ($P > 0.05$).	The method provides reliable and accurate estimations of endothelial parameters even in challenging post-keratoplasty images, and enables clinical assessment of CV and HEX, parameters that are generally not used clinically because of poor reliability of existing automatic methods. CV and HEX may be useful to study postoperative healing and may become useful biomarkers and potential predictors of graft health and survival in future studies.
Bitton et al. (2022), France [42]	The authors evaluated a previously developed deep learning pipeline based on three U-Net models that performs pixel-level automatic detection of corneal edema on AS-OCT images. Each pixel is classified as either edema or normal tissue; an EF is calculated as the proportion of corneal pixels classified as edema. The model was validated in 1992 OCT images from normal eyes and eyes undergoing DMEK for endothelial dysfunction (290 eyes: 50 normal and 240 pre-DMEK).	Pixel-wise binary classification (edema vs normal tissue); no clustering analysis performed. The model classifies OCT pixels as edema or normal and generates edema probability maps and computes the EF.	Performance metrics for detection of clinically significant edema defined by a differential central corneal thickness ≥ 20 µm: in overall cohort (AUC 0.97, sensitivity 94.5%, specificity 94.3%); in FECD + normal eyes (AUC 0.96, sensitivity 94.5%, specificity 92.0%); in non-FECD + normal eyes (AUC 0.99, sensitivity 96.0%, specificity 97.9%). Optimal EF threshold for overall cohort and eyes with FECD was 0.143 and for other eyes 0.171. Additional findings: mean EF in normal eyes 0.06 ± 0.2 and pre-DMEK eyes 0.80 ± 0.3 ($P < 0.001$). Correlation of EF with Scheimpflug edema classification in eyes with FECD: r 0.58, $P < 0.001$.	The model can objectively detect minimal/subclinical corneal edema before DMEK surgery and may assist clinicians in OCT interpretation, surgical planning, and decision-making about endothelial keratoplasty. The authors specifically suggest potential use of the model in identifying subclinical edema, determining candidacy for DMEK, choosing between cataract surgery alone vs. combined procedures, monitoring patients with FECD, and supporting preoperative decision-making.
Heslinga et al. (2020); Netherlands/Denmark [30]	A deep learning pipeline was developed that automatically localizes the scleral spur in AS-OCT scans, centers the scans, segments detached graft regions post-DMEK, and constructs two-dimensional GD maps from multiple radial AS-OCT B-scans. The pipeline consisted of a ResNet-50-based scleral spur localization model and a U-Net segmentation model that automatically identifies and quantifies GD in AS-OCT scans and generates detachment maps post-DMEK.	Semantic segmentation of GD regions in AS-OCT B-scans; no clustering analysis. The output is quantitative localization and mapping of detached graft areas rather than simple binary classification.	1) Scleral spur localization: mean localization error 0.155 mm (4.97 pixels); inter-rater difference 0.090 mm; 95% of localization errors within 0.275 mm (8.79 pixels). 2) GD quantification: 69% of model detachment-length estimates were within 10 pixels (~150 µm) of expert annotations; inter-expert agreement was 77% within 10 pixels 3) Detachment projection accuracy: A DSC of 0.896 ± 0.149 (model) vs. 0.880 ± 0.172 (inter-rater performance) when empty masks were excluded; including all B-scans, DSC were 0.906 ± 0.190 (model) and 0.916 ± 0.160 (inter-rater performance).	The system can automatically and objectively quantify post-DMEK GD, generate accurate detachment maps comparable to human experts, support and standardize postoperative management and clinical decision-making, and assist with decisions about graft monitoring, RB, and repeat DMEK procedures.
O'Brien et al. (2021), USA [43]	RSF analysis identified DSAEK intraoperative complications as a major predictive factor for GF in the final model, ranking it among the most important predictors after surgeon and eye bank effects. The model evaluated 50 donor, recipient, eye bank, and intraoperative variables, and selected a final model containing five predictors.	No clustering analysis performed. RSF for survival prediction and variable importance ranking were used to model GF risk. The study used GF prediction using donor, recipient, eye bank, surgeon, and intraoperative variables, plus VIMP ranking and adjusted survival modeling.	DSAEK intraoperative complication ranked third most predictive factor of GF in the final RSF model (after surgeon and eye bank). Estimated difference in restricted mean survival time (RMST) over 47 months was −227 days (99% CI: −352 to −70 days), Harrell C-statistic for the final RSF model 0.71.	RSF analysis can improve identification of factors predictive of GF and provide adjusted graft survival estimates. The findings support the importance of intraoperative complications as predictors of poorer graft survival and may help guide future keratoplasty research and clinical decision-making about keratoplasty outcomes.

Abbreviations: ML, machine learning; AS-OCT, anterior segment optical coherence tomography; AI, artificial intelligence; RB, rebubbling; DMEK, Descemet membrane endothelial keratoplasty; GD, graft detachment; AUC, area under the curve; CI, confidence interval; DALK, deep anterior lamellar keratoplasty; RF, random forest; RSF, random survival forest; DSAEK, Descemet stripping automated endothelial keratoplasty; PK, penetrating keratoplasty; GF, graft failure; SBB, successful big bubble; FBB, failed big bubble; VGG, Visual Geometry Group; CNN, convolutional neural network; ANN, artificial neural network; ANOVA, analysis of variance; YI, Youden index; DSC, Dice similarity coefficient; LASSO, logistic regression using least absolute shrinkage; CTA, classification tree analysis; 2D, two-dimensional; 3D, three-dimensional; mRMR, minimal redundancy maximal relevance; ROSE, random oversampling of examples; LR, logistic regression; ECD, endothelial cell density; CV, coefficient of variation; MAE, mean absolute error; µm, micrometer; ST-ITEF, spatio-temporal intraoperative task estimating framework; C-IoU, Class Intersection-over-Union; MPE, mean position error; UT-DSAEK, ultrathin DSAEK; HEX, hexagonality; EF, edema fraction; FECD, Fuchs endothelial corneal dystrophy. Note: References 29 and 30 used the same RCT-derived cohort of 80 AS-OCT scans from 68 participants for distinct AI tasks; they represent separate analyses but not independent patient populations. References 20 and 21 originated from the same Singapore registry programme but evaluated distinct indication and procedure cohorts and were treated as separate synthesis entries.

Table 3. Included studies – Summary of methodological quality

Author (Year), Country	Type of study	Tool	Domains assessed	Rating	Overall methodological/applicability judgement
Yousefi et al. (2020), USA/Japan [31]	Cohort study using unsupervised machine learning and post-hoc analysis to estimate the likelihood of future keratoplasty from baseline corneal OCT parameters.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis.	**Participants and data sources:** high concern because the postoperative keratoplasty dataset was selected retrospectively and only 168 of 333 surgical eyes had eligible preoperative baseline measurements. **Predictors:** some concerns because 424 OCT-derived parameters came from a single imaging system and eyes with missing ESI values were excluded. **Outcome:** some concerns because heterogeneous keratoplasty procedures were combined and the clinical criteria leading to surgery were not standardized. **Analysis:** high concern, the risk estimates were derived from exploratory post-hoc clustering, approximately 10% of observations were excluded as outliers, and no independent validation, calibration, or conventional discrimination assessment was performed.	High methodological concern; applicability concerns: moderate to high
Hayashi et al. (2023), Japan/Germany [32]	Retrospective prognostic deep-learning model-development and transferability study using prospectively collected post-DMEK AS-OCT data from two surgeons.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis.	**Participants and data sources:** low concern because consecutive DMEK cases were obtained from a prospective database and a separate surgeon's cases were used for testing. **Predictors:** low concern because AS-OCT images were obtained using a consistent protocol. **Outcome:** some concerns because the reference outcome was the treating surgeon's decision to perform rebubbling, for which no universally standardized threshold exists. **Analysis:** some concerns because transferability was evaluated between two surgeons working at the same institution, the test-set threshold was selected using the Youden index, and calibration was not reported.	Moderate methodological concern; applicability concerns: moderate
Kundu et al. (2025) India/Netherlands [33]	Retrospective prognostic machine-learning model-development study predicting five-year DALK outcomes in advanced keratoconus.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because the study was conducted at a single tertiary centre, included only Indian–Asian patients who completed at least five years of follow-up, and excluded eyes with intraoperative complications. **Predictors:** some concerns because questionnaire-derived variables, including allergy and eye rubbing, were partly subjective. **Outcome:** high concern because the favorable/unfavorable classification was a composite of several clinical and visual outcomes, and graft-clarity grading was subjective. **Analysis:** high concern because approximately 40 candidate predictors were evaluated in 250 eyes, validation was limited to leave-one-out cross-validation, independent external validation and calibration were not reported, and it was unclear whether feature ranking was repeated within each validation cycle.	High methodological concern; applicability concerns: moderate to high
Kim et al. (2025), Republic of Korea [34]	Retrospective prognostic machine-learning model-development and internal-evaluation study predicting postoperative ocular hypertension after DMEK.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because the study used a small, single-centre cohort with variable follow-up and only 31 OHT events. **Predictors:** some concerns because missing values were replaced by mean imputation and feature selection involved repeated feature-importance assessment and feature dropping. **Outcome:** low concern because postoperative OHT was defined using prespecified IOP thresholds and excluded elevations occurring within 30 days of surgery or rebubbling. **Analysis:** high concern because the models lacked external validation and calibration, and ensemble performance was reported only for 61 eyes with concordant predictions across all four models, which may overestimate performance in the complete cohort.	High methodological concern; applicability concerns: moderate
Ang et al. (2022), Singapore [20]	Prospective registry-based prognostic model-development study using random-survival-forest–guided Cox regression to evaluate 10-year graft survival.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources—low risk:** The study included 1,335 consecutive patients undergoing primary DSAEK or penetrating keratoplasty for Fuchs endothelial dystrophy or bullous keratopathy. Data were collected prospectively through the Singapore Corneal Transplant Registry, with clearly reported eligibility criteria, standardized surgical protocols, and involvement of multiple surgeons. **Predictors—low risk:** Forty-nine candidate donor, recipient, clinical, and operative predictors were selected through a literature review and recorded before graft failure occurred. Predictor definitions and measurement methods were generally reported. **Outcome—unclear risk:** Graft failure was clearly defined as irreversible loss of graft clarity compromising vision for at least three consecutive months and was prospectively recorded. However, the authors reported substantially reduced follow-up at ten years, without evaluating whether censoring or loss to follow-up was related to patient prognosis. **Analysis—high risk:** Model tuning, variable ranking, selection of the best predictor combination, and performance evaluation were conducted using the same dataset. Although out-of-bag estimation and bootstrap sampling were used, there was no independent or external validation, and the selection process was not nested within a separate validation procedure. The random-survival-forest analysis screened 49 predictors and pairwise interactions, after which the selected variables were entered into Cox regression using the same data. Consequently, the reported hazard ratios, confidence intervals, and P values may be affected by post-selection bias and model optimism. Calibration was not reported, and performance was limited to an out-of-bag C-index of 0.701. Fifty-two patients with missing preoperative visual acuity were excluded from the final model rather than addressed using imputation. Penetrating keratoplasty and DSAEK were performed during partly different calendar periods because DSAEK was introduced in 2006. Therefore, procedure type may be associated with changes in patient selection, surgical experience, and postoperative management over time. The predominantly Chinese, single-center Asian cohort, historical surgical techniques, and restriction to primary grafts for two endothelial diseases limit applicability to other populations and contemporary practice.	High methodological concern; applicability concerns: moderate
Ng Yin Ling et al. (2025), Singapore [21]	Registry-based prognostic model-development and internal-evaluation study using prospectively collected data.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis.	**Participants and data sources—some concerns:** The study included 1,687 consecutive patients from a prospectively maintained registry with clearly reported eligibility criteria. However, it was conducted within one Singapore programme, 871 patients were lost to follow-up after five years, and the internal test set consisted specifically of records containing missing data rather than a randomly or temporally selected representative sample. **Predictors—some concerns:** Fifty-five candidate variables were selected through literature review from over 680 available variables, and the final five-variable score used information available prior graft failure. Substantial data-driven reduction and stepwise selection were performed, and the prediction time and intended use of the score were not completely formalized. **Outcome—low concern:** Graft failure was prospectively recorded and defined as irreversible loss of optical clarity sufficient to compromise vision for at least three consecutive months. Nevertheless, follow-up intensity varied according to clinical severity. **Analysis—high concern:** The test set was determined by missing-data status, potentially making it systematically different from the complete-case development set. Model tuning, variable ranking, nested Cox-model selection, and risk-score development involved several data-driven stages. The report indicates that candidate models were compared using test-set performance, potentially compromising the independence of the test set. Calibration was not adequately reported, the risk threshold was data-derived, and there was no external validation. The beyond-five-year test sample included only 97 patients, and separation of its risk groups was not statistically significant.	High methodological concern, driven mainly by analysis; applicability concerns: moderate.
Feizi et al. (2024), Iran [23]	Retrospective prognostic model-development and internal-evaluation study predicting graft rejection after PK or DALK for keratoconus.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis.	**Participants and data sources—some concerns:** The study included 1,214 consecutive eyes, but eligibility required at least one year of follow-up and complete suture removal. Both eyes of bilaterally transplanted patients were included. The 27-year study period encompassed substantial changes in practice, with PK performed predominantly before 2005 and DALK thereafter. **Predictors—high concern:** Several predictors were postoperative and may not have been available at a clearly defined prediction time. Particularly, corticosteroid discontinuation "at the time of graft rejection," suture-related complications within three months before rejection, and other postoperative events create risks of temporal leakage, reverse association, and unequal predictor ascertainment between eyes with and without rejection. Procedure type was also strongly associated with calendar period. **Outcome—some concerns:** The first rejection episode was clinically classified, but fully reproducible diagnostic criteria, masking, and independent outcome adjudication were not reported. Different rejection phenotypes after PK and DALK were combined into one outcome. **Analysis—high concern:** The 80%/20% split appeared to be performed at the eye level, and participant-level separation was not reported despite inclusion of bilateral eyes. Multiple algorithms were evaluated using the same internal test data, model tuning and feature-selection procedures were incompletely described, and external validation was absent. Most classifiers did not clearly account for variable follow-up and censoring, and calibration was represented only by Brier scores without calibration plots or calibration-in-the-large.	High methodological concern; applicability concerns: high.

Hayashi et al. (2021), Japan/Germany [35]	Retrospective prognostic deep-learning model-development study predicting successful big-bubble formation during DALK	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because only 46 eyes from one institution and two surgeons were included, and eyes converted to penetrating keratoplasty following intraoperative Descemet membrane perforation were excluded. **Predictors:** low concern because preoperative SD-OCT images were obtained before surgery using a consistent imaging approach. **Outcome:** low concern because successful or failed big-bubble formation was determined directly during surgery. **Analysis:** high concern because of the small sample, absence of external validation and calibration, selection of VGG-16 after comparison with other models without complete reporting, and unclear separation of images from the same eye across the five cross-validation folds.	High methodological concern; applicability concerns: moderate
Hayashi et al. (2020), Japan [36]	Retrospective prognostic deep-learning model-development and internal-validation study predicting the need for rebubbling after DMEK.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because the study included only 62 eyes, and the non-rebubbling group was randomly selected to match the number of rebubbled eyes rather than represent the natural clinical population. **Predictors:** low concern because postoperative AS-OCT images were obtained using a consistent protocol, and images from the same patient were kept within the same cross-validation fold. **Outcome:** some concerns because the decision to perform rebubbling followed criteria applied by a single surgeon, although the image classifications were masked and assessed by three corneal specialists. **Analysis:** high concern because validation was internal, external validation and calibration were absent, and the best individual and ensemble models were selected after comparing numerous models using the same cross-validation results.	High methodological concern; applicability concerns: moderate
Karaca et al. (2025), Turkey [37]	Retrospective prognostic machine-learning model-development and internal-evaluation study predicting early graft failure after DMEK.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because the study was conducted at one centre, included only patients completing at least 12 months of follow-up, and excluded eyes with intraoperative complications or important coexisting ocular conditions. **Predictors:** high concern because some variables, including tamponade used for rebubbling, were obtained after the primary DMEK procedure and may not have been available at the intended prediction time. **Outcome:** high concern because an explicit and reproducible definition of graft failure was not provided. **Analysis:** high concern because only 34 graft failures occurred, evaluation used a single 75%/25% holdout split, several models were compared using the same small test set, regularization and calibration were absent, and no external validation was performed.	High methodological concern; applicability concerns: high
Muijzer et al. (2022), with an earlier conference report by Wisse et al. (2021), Netherlands [22, 26]	Nationwide registry-based prognostic model-development study using supervised machine learning; retrospective analysis of prospectively collected registry data. The conference report arose from the same cohort and was not assessed independently.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** low concern. **Predictors:** some concerns because centre-level practice-pattern data were linked to individual procedures and some missing categorical values were coded as "unknown." **Outcome:** some concerns because GD was operationally defined by the occurrence of rebubbling rather than independent anatomical assessment of detachment. **Analysis:** high concern because 91 candidate predictors were evaluated, external validation was absent, evaluation relied on a single 70%/30% split, and calibration was not reported.	High methodological concern; applicability concerns: moderate
Glatz et al. (2021), Germany [38]	Deep-learning image-segmentation model-development and validation study using data from a prospective DMEK cohort.	Adapted QUADAS-2	Patient selection, index test, reference standard, flow and timing	**Patient selection—high risk:** Participants came from a prospective cohort, but the independent validation set was deliberately selected according to slit-lamp–estimated graft detachment in a 3:7:10 ratio for no detachment, less than one-third detachment, and at least one-third detachment. This enriched sampling may not represent the distribution encountered in routine clinical practice. **Index test—low risk:** The best-performing neural network was selected using the training data, locked, and subsequently evaluated in a separate validation dataset that was not used for model development. The 107-eye application cohort was also separate from the training and validation cohorts. **Reference standard—unclear risk:** Manual AS-OCT annotations followed a consensus labelling protocol and were re-reviewed, with excellent interobserver agreement in 19 representative images. However, each remaining image was generally annotated by only one of five investigators, and independent duplicate annotation was not performed for the complete validation dataset. **Flow and timing—low risk:** Training, validation, and application cohorts were kept separate. In the application cohort, AS-OCT imaging and slit-lamp assessment were performed on the same day, and all eligible eyes with both assessments were included. The single-centre cohort, one AS-OCT system, and enriched validation sample limit generalizability, although these concerns primarily affect applicability rather than model-evaluation integrity.	High methodological concern; applicability concerns: moderate
Joseph et al. (2023), with an earlier conference report by Joseph et al. (2021), USA [27, 28]	Retrospective prognostic machine-learning model-development and internal-evaluation study using post-DMEK endothelial cell images. The earlier conference report was a preliminary report from the same research programme and was not assessed independently.	PROBAST+AI	Participants and data sources; predictors; outcome; analysis	**Participants and data sources:** some concerns because the study included a small, selected, balanced cohort of 22 rejection and 22 control eyes. **Predictors:** some concerns because numerous image-derived features were evaluated and segmentation included guided manual correction. **Outcome:** some concerns because details of clinical rejection ascertainment and timing were limited. **Analysis:** high concern because 432 candidate features were evaluated in only 44 eyes, the held-out test set was small, external validation was absent, and calibration was not reported.	High methodological concern; applicability concerns: moderate
Heslinga et al. (2021), Netherlands [29]	Deep-learning segmentation and measurement-validation study using prospectively collected imaging data from a randomized controlled trial.	Adapted QUADAS-2	Patient selection; index test; reference standard; flow and timing	**Patient selection—low risk:** AS-OCT scans were obtained from participants in a randomized controlled trial with clearly reported eligibility criteria. Repeat DMEK procedures and patients with previous keratoplasty were excluded according to predefined criteria. **Index test—low risk:** The dataset was divided at the participant level into 752 training, 208 validation, and 320 independent test images, preventing images from the same participant from appearing across development and test sets. Model performance was also examined using repeated training and additional participant-level data splits. **Reference standard—low risk:** Training and validation images were annotated by one of three observers under the supervision of a cornea specialist. All test images were independently annotated by all three observers, and interobserver variability was explicitly evaluated. **Flow and timing—low risk:** All 320 B-scans in the independent test set were evaluated, and model-derived measurements were compared with all three sets of manual annotations. No participant overlap occurred between the training, validation, and test sets. The absence of external-centre and external-device validation limits generalizability but does not substantially increase the internal risk of bias.	Low methodological concern; applicability concerns: moderate
Feng et al. (2024), China [39]	Deep-learning surgical-video segmentation, phase-recognition, and instrument-path estimation study with internal validation.	Adapted QUADAS-2	Patient selection, index test, reference standard, flow and timing.	**Patient selection—high risk:** Only four videos came from the authors' hospital; another 24 were purposively selected from an online video source. The videos were selected partly because they contained all six required objects, rather than through consecutive or random sampling. **Index test—high risk:** The 1,000 labelled frames were divided into 900 training and 100 validation images, but there was no separate independent test set. The same validation set was used to compare and select the segmentation models. **Reference standard—low risk:** Objects were manually labelled by students and subsequently validated by experienced ophthalmologists. **Flow and timing—unclear risk:** The authors did not clearly state whether splitting was performed at the video/patient level. Therefore, temporally adjacent frames from the same surgical video may have appeared in both sets. The proposed instrument-path guidance was not tested by controlling a surgical robot or assessing clinical outcomes; agreement with surgeons' movements was only a proxy for successful navigation.	High methodological concern; applicability concerns: high
Dreesbach et al. (2024), Germany [40]	Retrospective, single-centre technical method-comparison and measurement-validation study using archived endothelial images from consecutive donor corneas.	Adapted QUADAS-2	Dataset selection, model evaluation, reference annotations, and data flow.	**Dataset selection—low risk:** The study included 9375 consecutive archived images from 6168 donor corneas, providing a large real-world sample. **Model evaluation—unclear risk:** The model had been trained on 100 manually annotated images from the same eye-bank environment. It was not clearly stated whether these images were excluded from the present dataset, and no external-centre or external-device validation was undertaken. **Reference annotations—high risk:** The Rhine-Tec system was not a true gold standard. It assessed only approximately 15–40 manually corrected cells within a subjectively positioned frame. The authors explicitly acknowledged the absence of a true gold standard and identified systematic deviations around the clinically important threshold of 2000 cells/mm². **Data flow—low risk:** The same archived images were evaluated by both methods, and analyses were reported separately for the two storage media. A further limitation is that some corneas contributed images from both media, but the analysis did not clearly account for within-cornea correlation.	High methodological concern; applicability concerns: moderate

Vigueras-Guillén et al. (2020), Netherlands [41]	Prospective, single-centre technical segmentation and morphometric-validation study using post-UT-DSAEK specular microscopy images.	Adapted QUADAS-2	Dataset selection, model evaluation, reference annotations, and data flow.	**Dataset selection—low risk:** Images were obtained prospectively from 41 adult patients enrolled in a registered clinical study, with clearly reported eligibility criteria and standardized postoperative imaging at 1, 3, 6, and 12 months. **Model evaluation—unclear risk:** Five-fold cross-validation was performed with all images from the same eye retained within one fold, appropriately preventing eye-level data leakage. However, the network was refined and evaluated within the available datasets, without testing in a fully independent external-centre or external-device dataset. **Reference annotations—unclear risk:** All 383 UT-DSAEK images were manually segmented and treated as the gold standard. However, the number and expertise of annotators, blinding procedures, duplicate annotation, and interobserver or intraobserver reliability were not reported. **Data flow—low risk:** Every eye was confined to one cross-validation subset, and each subset served as the test set once. Images without successful parameter estimation were assigned a 100% error in the comparative analysis, reducing the risk of selectively excluding model failures. The use of a single center, one microscope model, and a restricted post-UT-DSAEK population limits generalizability. In addition, selecting the image containing the highest number of visible cells for the longitudinal clinical analysis may represent better-than-routine image conditions.	Moderate methodological concern; applicability concerns: moderate
Bitton et al. (2022), France [42]	Retrospective, single-centre diagnostic-accuracy study evaluating a previously developed deep-learning model for detecting corneal edema on OCT images.	QUADAS-2	Patient selection, index test, reference standard, flow and timing.	**Patient selection—high risk:** Pre-DMEK cases were retrospectively selected from a surgical registry and restricted to patients with successful surgery, defined by improved visual acuity and reduced corneal thickness at six months. The comparison group consisted of strictly selected healthy controls who were substantially younger than the pre-DMEK patients. This two-group design may exaggerate diagnostic performance and does not represent the intended clinical population of diagnostically uncertain patients. **Index test—high risk:** Although the deep-learning pipeline had been developed previously and its training patients were excluded from this cohort, the edema-fraction threshold was derived and evaluated using the same dataset. Therefore, the reported sensitivity and specificity were not independently validated. **Reference standard—high risk:** No accepted gold standard for subclinical corneal edema was available. Differential central corneal thickness at six months was used as a surrogate reference standard, and the 20-µm cutoff was selected because it produced the highest AUC for the AI-derived edema fraction. The reference definition was therefore partly determined by the performance of the index test. Differential thickness could also be affected by Descemet membrane thickness and imprecise registration of preoperative and postoperative measurements. **Flow and timing—high risk:** The reference assessment was performed six months after the index OCT examination, and only patients with successful DMEK outcomes were included. Healthy controls were assigned a differential central corneal thickness of zero without undergoing the same preoperative–postoperative reference procedure, resulting in differential verification between cases and controls. In addition, 290 eyes were obtained from 255 patients, but the statistical analysis did not clearly account for correlation between fellow eyes. The single-centre, single-device, retrospective design and absence of prospective external validation further restrict generalizability.	High methodological concern; applicability concerns: high
Heslinga et al. (2020); Netherlands/Denmark [30]	Technical deep-learning development and validation study for AS-OCT localization, graft-detachment segmentation, and quantitative mapping, using imaging data from a randomized controlled trial.	Adapted QUADAS-2	Dataset selection, model evaluation, reference annotations, and data flow.	**Dataset selection—low risk:** AS-OCT scans were obtained from participants in a randomized controlled trial with clearly reported eligibility criteria. The dataset comprised 80 scans from 68 participants, including scans obtained immediately after DMEK and at postoperative day 7. **Model evaluation—low risk:** The dataset was randomly divided at the participant level into 960 training/validation B-scans from 50 participants and an independent test set of 320 B-scans from 18 participants. Model optimization was performed using the development data, while final performance was evaluated only in the held-out test set. **Reference annotations—unclear risk:** All scans were manually annotated by one DMEK expert, and a second DMEK expert annotated the test set to evaluate interobserver agreement. However, the first expert's annotations served as the principal reference, no consensus or adjudication process was reported, and the authors acknowledged that some graft sections were inherently difficult to distinguish and annotate. **Data flow—low risk:** All 320 B-scans in the independent test set were evaluated using the finalized models. B-scans without graft detachment were included, and performance was reported both with and without empty masks. The participant-level division prevented images from the same participant appearing in both development and test sets. The relatively small single-centre test sample, use of one AS-OCT device, and absence of external validation limit generalizability. The model was also not evaluated for its effect on rebubbling decisions or clinical outcomes.	Moderate methodological concern; applicability concerns: moderate
O'Brien et al. (2021), USA [43]	Post hoc prognostic prediction-model development study using prospectively collected data from a multicentre randomized clinical trial.	PROBAST+AI	Participants and data sources, predictors, outcome, and analysis.	**Participants and data sources—low risk:** Data were collected prospectively within a registered, multicentre, double-masked randomized clinical trial involving 1,330 eyes of 1,090 participants, 40 clinical sites, 70 surgeons, and 23 eye banks. Eligibility criteria and follow-up schedules were clearly reported. **Predictors—low risk:** The 50 candidate donor, recipient, eye-bank, and intraoperative variables were prospectively recorded and available before the occurrence of graft failure. Definitions of the principal predictors, including intraoperative complications, were reported. **Outcome—low risk:** Graft failure was defined using prespecified clinical criteria, including regrafting, persistent postoperative corneal cloudiness, and late graft failure. Follow-up examinations were performed according to a standardized schedule. **Analysis—high risk:** Only 81 graft failures occurred despite consideration of 50 candidate predictors. Variable selection and model development were conducted using the same dataset, without independent external validation. The model reported discrimination using a C-statistic of 0.71 but did not report calibration or clinically useful individual risk estimates. The authors acknowledged the low event rate and potential model optimism. Additional analytical concerns include the absence of clear adjustment for correlation between fellow eyes in the 240 bilateral participants, incomplete reporting of missing-data handling, and post hoc selection of 47 months rather than the original three-year outcome horizon. Surgeon and eye-bank identifiers were among the most important predictors. These centre-specific variables may capture local practice patterns and are unlikely to be directly applicable to new surgeons, eye banks, or healthcare systems.	High methodological concern; applicability concerns: moderate

Abbreviations: USA, United States of America; AI, artificial intelligence; PROBAST+AI [24], the Prediction model Risk Of Bias Assessment Tool for Artificial Intelligence; QUADAS-2 [25], Quality Assessment of Diagnostic Accuracy Studies–2; RCT, randomized controlled trial. Note: References 29 and 30 were appraised separately because they evaluated distinct AI models and technical objectives, although they used the same RCT-derived imaging cohort; The overall judgments provide a descriptive synthesis of methodological and applicability concerns across the relevant domains. They should not be interpreted as numerical quality scores or, for adapted QUADAS-2 assessments, as validated overall QUADAS-2 ratings. References 20 and 21 originated from the same Singapore registry programme but evaluated distinct indication and procedure cohorts and were treated as separate synthesis entries.

Table 2 summarizes the results reported by the included studies. Most studies reported strong model performance. Classification studies frequently achieved high discrimination, with AUC, accuracy, sensitivity, specificity, and F1 score often exceeding 0.85, although performance varied across applications. Segmentation-based models demonstrated robust performance, with DSC coefficients generally between 0.73 and 0.90 and localization errors below 0.2 mm in some studies. Survival-analysis methods and metrics, including Harrell's C-index, time-integrated AUC, hazard ratios, Kaplan–Meier analysis, and random survival forests, were used in five synthesis entries. Agreement analyses, including Bland-Altman plots and correlation coefficients, were reported in studies comparing AI-derived measurements with conventional or manual assessments. Several investigations also conducted feature-importance analyses or model comparisons, identifying key predictors such as death-to-preservation time, intensive care unit (ICU) stay duration, donor and surgical characteristics, and endothelial morphometric features. Although the reviewed studies demonstrated promising performance across diverse clinical applications, substantial heterogeneity in outcome definitions, performance metrics, and reporting standards hindered direct comparison of model performance across studies (Table 2).

Table 3 summarizes the methodological quality and risk-of-bias assessments of the included studies. Methodological quality varied considerably across studies. Common concerns included limited sample sizes, absence of external validation, model selection and evaluation within the same dataset, incomplete reporting of calibration, imperfect reference standards, and possible data leakage or non-independence of observations. Methodological strengths included prospective or multicentre data collection, participant-level separation of development and test datasets, independent test sets, and expert reference annotations. Accordingly, the reported model-performance estimates should be interpreted in relation to the design-specific limitations of each study.

The evidence base is heterogeneous and predominantly exploratory. Although several studies reported promising performance, frequent reliance on small or single-centre datasets, internal validation, incomplete reporting, and limited independent external validation restricts confidence in model generalizability and clinical readiness.

DISCUSSION

This review of 22 publications, representing 20 synthesized study units, illustrates the growing role of AI, particularly machine-learning and deep-learning approaches, in corneal transplantation. AI models showed promising applications across the transplantation pathway, including preoperative risk stratification, prediction of surgical and graft-related outcomes, postoperative monitoring, automated image analysis, and donor graft assessment. The reviewed studies utilized diverse data sources, including anterior segment OCT, specular and phase-contrast microscopy, surgical videos, clinical variables, and donor-recipient registry data. Despite substantial methodological heterogeneity, several common themes emerged. These include challenges related to data diversity and dataset scale, the expanding use of multimodal imaging and clinical data, the increasing adoption of advanced AI algorithms for prediction and segmentation tasks, and the growing clinical utility of AI for surgical planning, prognostic assessment, and postoperative management. The findings reveal the considerable potential of AI in corneal transplantation while underscoring the need for greater data standardization, larger multicenter datasets, and rigorous external validation to support clinical translation.

The reviewed evidence displayed substantial variation in dataset size and structure. Of the 20 synthesis entries, 13 used fewer than 500 samples [28–30, 32–39, 41, 42], three used more than 3000 samples [22, 26, 31, 40], and four used between 500 and 3000 samples [20, 21, 23, 43]. These figures refer to the primary analytical units, such as participants, eyes, grafts, scans, images, or surgical videos, rather than to unique patients. In particular, references 29 and 30 analyzed the same 80 AS-OCT scans from 68 RCT participants for different technical objectives. This variability has important implications for model development, validation, and clinical translation. Smaller datasets may increase the risk of overfitting and limit generalizability [44–46], whereas larger datasets may support more robust model development and validation [47, 48]. However, inconsistent reporting of dataset characteristics complicated comparisons across studies.

Survival-analysis methods and metrics, including Harrell's C-index, time-integrated AUC, hazard ratios, Kaplan–Meier analysis, and random survival forests, were used in five synthesis entries [20, 21, 23, 34, 43]. Large registry-based analyses predominantly employed conventional machine-learning and survival-modelling approaches [20–22, 26]. In contrast, several smaller imaging-based studies utilized advanced deep-learning architectures, including CNNs and U-Net variants [27–30, 35, 36, 38, 39, 41, 42]. Despite their limited sample sizes, these studies generally exhibited favorable performance across a range of clinical tasks, including prognosis, risk stratification, GD assessment, endothelial cell analysis, edema detection, and surgical guidance. Several reported excellent discriminations, with AUC values exceeding 0.95 for predicting favorable DALK outcomes, rebubbling after DMEK, graft failure, and subclinical corneal edema [33, 36, 37, 42]. Other models achieved good predictive performance, with ranges of AUC values 0.87–0.88 and balanced sensitivity and specificity 78.6–86% [27, 32]; more challenging tasks, such as predicting big-bubble formation and future graft rejection, showed moderate performance [28, 35]. Segmentation-based approaches yielded strong agreement with expert annotations, with a DSC approaching or exceeding 0.90 for graft detachment mapping and localization [30, 38]. Similarly, automated endothelial cell analysis systems achieved performance comparable to manual reference assessments while outperforming conventional software-based approaches [41].

Despite the promising results reported across many studies, several important methodological limitations warrant consideration. A substantial proportion of the reviewed studies relied on small cohorts, single-center datasets, or institution-specific data sources [29, 35], with validation often restricted to internal testing strategies such as cross-validation rather than independent external cohorts [27, 28, 30, 32, 35–38, 42]. These constraints heighten the risk of overfitting and may limit the generalizability of AI models [44–46] across diverse patient populations, surgical techniques, imaging platforms, and healthcare settings [35]. Consequently, the exceptionally high performance reported in some studies, including AUC values approaching

0.99 and sensitivities or specificities exceeding 95% [36, 37, 42], should be interpreted in the context of dataset characteristics and validation methodology rather than algorithmic sophistication alone. Several investigators explicitly highlighted the need for larger multicenter datasets and broader data integration to improve model robustness and clinical applicability of current AI models [28, 35, 36]. The findings underscore the importance of multicenter collaboration, standardized data collection protocols, harmonized reporting frameworks, and external validation across heterogeneous populations. Addressing these challenges will be essential to ensure the reliable, generalizable, and equitable translation of AI technologies into routine corneal transplantation practice.

With respect to imaging modalities and data sources, the reviewed studies displayed substantial diversity in the datasets used to develop AI models for corneal transplantation. Both retrospective [22, 27–37, 39, 40, 42] and prospective [20, 26, 38, 41, 43] datasets were utilized, ranging from small single-center studies [28, 32–34, 36–38, 41, 42] to large multicenter registries and national databases [20, 22, 26, 31, 43]. This diversity reflects the broad spectrum of clinical applications addressed, including postoperative monitoring, complication detection, surgical outcome assessment, and graft survival prediction. The predominance of imaging data, particularly AS-OCT [29–32, 35, 36, 38, 42], Scheimpflug corneal tomography [33], and specular microscopy [20, 28, 41], highlights the central role of imaging biomarkers in AI-driven corneal transplantation research. AS-OCT was the most frequently used modality [29–32, 35, 36, 38, 42], enabling automated segmentation, graft-detachment quantification, pachymetric analysis, and volumetric assessment following endothelial keratoplasty [29, 30, 34, 36, 38, 42]. Specular microscopy facilitated quantitative evaluation of endothelial cell morphology and density, supporting the identification of endothelial dysfunction, graft failure, and rejection-related changes [20, 28, 41]. Several studies further integrated multimodal datasets combining clinical, imaging, and laboratory information [33, 34, 37, 43], underscoring the potential of data integration to enhance predictive performance and personalized risk assessment. Slit-lamp examination served as an important clinical reference for evaluating GD, clinical signs of keratoconus, and postoperative outcomes [33, 36, 38], whereas intraoperative imaging and surgical video data enabled the development of computer-vision systems for surgical phase recognition, instrument tracking, and robotic guidance during keratoplasty procedures [39]. These findings highlight the need for standardized data harmonization and multimodal integration frameworks to fully realize the potential of AI across heterogeneous clinical settings.

The integration of imaging and clinical data has enabled the development of increasingly sophisticated AI frameworks in corneal transplantation. Deep CNNs were trained on AS-OCT images [29, 30, 36, 38, 42] and endothelial microscopy images [28, 41] to detect structural abnormalities, quantify graft-related changes, and assess endothelial health. Classical machine-learning algorithms were primarily applied to structured clinical, laboratory, and registry-based data [20–23, 26, 28, 33, 34, 37, 43]. Several studies adopted multimodal strategies that combined imaging biomarkers with demographic, clinical, surgical, and laboratory variables, improving predictive performance while enhancing model interpretability and clinical relevance [33, 34, 37, 43]. These findings highlight the value of data fusion approaches for capturing the multifactorial nature of graft survival and postoperative outcomes.

The reviewed studies display the growing application of AI, particularly machine learning and deep learning, across a broad spectrum of corneal transplantation tasks. Advanced DL architectures, including visual geometry group-16 (VGG-16), VGG19, DenseNet, ResNet-50, EfficientNet, U-Net, UNet++, and DeepLabV3, were employed to analyze AS-OCT scans, endothelial microscopy images, and surgical video data [29, 30, 35, 36, 38, 39, 41, 42]. Classical machine learning algorithms like random forest, logistic regression, support vector machines, XGBoost, CatBoost, k-nearest neighbors, artificial neural networks, and random survival forests, were primarily applied to clinical, laboratory, and registry-based datasets [20, 22, 28, 33, 34, 37, 43]. Ensemble learning approaches were also used to improve predictive performance [34, 36], while unsupervised clustering methods were explored to identify latent keratoconus features and stratify keratoplasty risk [31].

Analytical objectives were diverse and spanned multiple stages of the corneal transplantation pathway. A major focus was the detection, localization, and quantification of GD following endothelial keratoplasty, where deep learning models enabled automated measurement of detachment extent and supported timely decisions regarding rebubbling and other postoperative interventions [26, 30, 32, 36, 38]. Predictive models addressed graft survival or failure [20, 21, 37, 43], graft rejection [23, 27, 28], graft-detachment-related rebubbling [22, 26, 32, 36], and postoperative ocular hypertension [34]. Other applications included future keratoplasty risk [31], DALK outcome and big-bubble prediction [33, 35], GD segmentation and quantification [30, 38], post-DMEK pachymetry [29], endothelial assessment [40, 41], edema detection [42], and surgical-video analysis [39]. Results expose the capacity of AI to integrate imaging, biometric, surgical, and clinical data, enabling earlier identification of high-risk patients, more personalized decision-making, and improved management throughout the corneal transplantation pathway.

Supervised learning and deep learning dominated the methodological landscape, with CNN-based architectures and ensemble models displaying strong performance in classification, prediction, and segmentation tasks [28, 30, 32, 36, 38, 41, 42]. Although unsupervised learning was represented by only a single study [31], survival modeling approaches—including Cox regression, Kaplan-Meier analysis, and random survival forests—were successfully applied in several investigations to evaluate long-term graft outcomes [20, 34, 43]. These findings illustrate both the versatility of AI methodologies and the expanding scope of their clinical applications in corneal transplantation.

AI applications were particularly prominent in three areas: prognostic prediction [20, 29, 33], postoperative monitoring [28–31, 34, 37, 38, 41, 42], and risk-factor identification [20– 23, 26, 27, 34, 37, 43]. Risk-factor identification was investigated by Ang et al. [20], Ng Yin Ling et al. [21], Feizi et al. [23], Kim et al. [34], Karaca et al. [37], the Muijzer and Wisse companion reports [22, 26], and O'Brien et al. [43]. These studies identified potential associations involving surgical procedure, indication, donor and recipient characteristics, inflammation, previous graft failure, anterior-chamber parameters, postoperative treatment, and operative factors. However, several findings require cautious interpretation because of internal validation, data-driven variable selection, or predictors measured after surgery. Several studies exposed the value of AI for predicting future corneal interventions and graft-related outcomes [20, 22, 26, 28, 32, 34, 37, 38, 43]. Using unsupervised clustering of OCT-derived corneal parameters, Yousefi et al. [31] identified distinct corneal features associated with markedly different probabilities of subsequent keratoplasty, illustrating how AI can uncover clinically meaningful patterns not readily apparent through conventional analysis [31]. Beyond risk stratification, the visualization of these clusters provided an intuitive framework for clinical interpretation and patient counseling. These findings further support the growing evidence that combining multiple corneal parameters—including topography, pachymetry, and elevation parameters—enhances disease characterization and prognostic assessment [49–55].

Similarly, studies by Joseph [27] and Karaca [37] evidenced the prognostic value of machine learning for predicting graft rejection and graft failure before clinical manifestation [27, 37]. Novel image-derived endothelial features outperformed conventional endothelial parameters in identifying eyes at risk of rejection [27]. In parallel, donor-related variables such as death-to-preservation time and intensive care unit stay emerged as important predictors of graft failure, emphasizing the influence of donor quality and perioperative factors on transplant success [37]. Together, these findings suggest that AI-based models may facilitate individualized surveillance strategies and earlier preventive interventions.

Postoperative monitoring represented another major area of AI application. Several studies showed that deep learning models can accurately detect, localize, and quantify GD following endothelial keratoplasty, often identifying subtle detachments that may be overlooked during routine clinical assessment [30, 32, 38]. Muijzer et al. [22] evidenced the utility of machine learning for identifying predictors of GD after posterior lamellar keratoplasty and estimating postoperative risk using large-scale registry data [22]. Their findings highlight the potential of AI-driven risk stratification to support personalized postoperative management, optimize surgical decision-making, and identify patients who may benefit from closer surveillance after DMEK or DSEK [22]. Glatz et al. [38] showed that automated three-dimensional graft-detachment mapping from AS-OCT images could support rebubbling decisions and improve assessment of graft attachment. Likewise, Heslinga et al. [29] illustrated how AI-generated pachymetry maps provide a more comprehensive assessment of endothelial recovery than conventional central corneal thickness measurements, enabling localized monitoring of graft function and objective monitoring of corneal thickness restoration after DMEK. Hayashi et al. [32] further showed that AI models trained on OCT data could predict the need for rebubbling across two surgeons, highlighting the potential for broader clinical deployment of standardized decision-support systems, thus facilitating standardized management and providing objective treatment guidance. Bitton et al. [42] developed an OCT-derived biomarker for the objective assessment of corneal edema, offering a quantitative approach for detecting subtle or subclinical endothelial dysfunction before the onset of clinically evident decompensation. By enabling automated identification of minimal corneal edema prior to DMEK, their model may enhance the interpretation of OCT findings, support preoperative evaluation, and assist clinicians in selecting the most appropriate management strategy, including decisions regarding endothelial keratoplasty and related surgical interventions [42]. Heslinga et al. [30] showed that deep learning-based analysis of AS-OCT images can accurately localize and quantify GD after DMEK, achieving performance comparable to human experts: the automated generation of GD maps may facilitate objective postoperative monitoring and support timely decisions regarding rebubbling and further surgical management [30]. These studies indicate that AI can enhance pre- and postoperative surveillance through objective, reproducible, and quantitative assessment of cornea. AI applications also extended to intraoperative guidance and surgical workflow optimization. Feng et al. [39] showed that computer vision techniques could accurately recognize surgical phases, track instruments, and support robotic guidance during keratoplasty procedures [39]. These developments suggest a future role for AI not only in outcome prediction but also in real-time surgical assistance and automation.

The identification of risk factors for graft failure and postoperative complications was another recurring theme. Studies by Ang et al. [20], Wisse et al. [26], Kim et al. [34], Karaca et al. [37], and O'Brien et al. [43] evidenced the advantages of machine learning approaches over conventional statistical methods for analyzing complex, multidimensional datasets. Across these studies, factors such as surgical technique, donor characteristics, previous graft failure, and anterior chamber parameters emerged as potential predictors of graft failure, GD, and postoperative ocular hypertension [20, 26, 34, 37, 43]. By capturing nonlinear relationships and interactions among variables, machine learning models provide opportunities for more accurate risk stratification and individualized postoperative management.

Beyond prognostic applications, AI also improved the assessment of endothelial cell health and graft quality. Studies by Dreesbach et al. [40] and Vigueras-Guillen et al. [41] illustrated how automated image-analysis systems can provide reliable estimates of endothelial cell density, coefficient of variation, and hexagonality even when image quality is suboptimal [40, 41]. Such approaches may improve the objectivity and reproducibility of graft evaluation, support donor-tissue quality control, and assist with clinical decision-making about graft suitability and release.

The findings of this review highlight the growing potential of AI as a complementary clinical tool across the corneal transplantation pathway, encompassing preoperative risk stratification, surgical planning, postoperative monitoring, and long-term outcome prediction. A major strength of this review is the comprehensive synthesis of diverse AI applications across multiple keratoplasty procedures, imaging modalities, and analytical approaches, providing an integrated overview of the current state of the field. Nevertheless, several limitations should be acknowledged. Despite encouraging performance, the included studies exhibited substantial heterogeneity in study design, sample size, data sources, outcome measures, AI methodologies, and reporting practices, precluding direct comparison of model performance and limiting quantitative synthesis. In addition, many studies relied on single-center datasets and lacked external validation, raising concerns about generalizability and real-world applicability. Future research should prioritize multicenter collaborations, standardized reporting and evaluation frameworks, prospective validation studies, and the development of explainable and externally validated AI models. The integration of large-scale multimodal datasets combining imaging, clinical, surgical, and biomarker information may further enhance model robustness and facilitate the safe, reliable, and equitable translation of AI into routine corneal transplantation practice.

CONCLUSIONS

AI is emerging as a valuable adjunct across the corneal transplantation pathway, with applications spanning preoperative risk stratification, surgical planning, postoperative monitoring, and long-term outcome prediction. The studies included in this review evidence that AI models can effectively integrate imaging, clinical, biometric, and donor-related data to identify patients at risk of graft failure, rejection, GD, and other postoperative complications while supporting personalized clinical decision-making. Furthermore, AI-based approaches have shown promise in automated image analysis, endothelial assessment, surgical guidance, and graft survival prediction. Despite these encouraging advances, several barriers must be addressed before widespread clinical implementation can be achieved. The current evidence base is characterized by substantial heterogeneity in datasets, study designs, outcome measures, and reporting standards, while external validation and prospective multicenter evaluation remain limited. Future research should prioritize the development of robust, explainable, and generalizable AI models through large-scale collaborative datasets, standardized evaluation frameworks, and multimodal data integration. Addressing these challenges will be critical to ensure the safe, reliable, and equitable translation of AI technologies into routine corneal transplantation practice and to realize their full potential for improving patient outcomes and advancing precision transplantation.

ETHICAL DECLARATIONS

Ethical approval: Not required.
Conflict of interests: None.

FUNDING

None.

ACKNOWLEDGMENTS

The author gratefully acknowledges Catherine (Cathy) Moor from Austin, Texas, for verifying the data extraction and quality assessment.

REFERENCES

1. Gain P, Jullienne R, He Z, Aldossary M, Acquart S, Cognasse F, Thuret G. Global Survey of Corneal Transplantation and Eye Banking. JAMA Ophthalmol. 2016 Feb;134(2):167-73. doi: 10.1001/jamaophthalmol.2015.4776. PMID: 26633035.
2. Ong HS, Ang M, Mehta JS. Evolution of therapies for the corneal endothelium: past, present and future approaches. Br J Ophthalmol. 2021 Apr;105(4):454-467. doi: 10.1136/bjophthalmol-2020-316149. Epub 2020 Jul 24. Erratum in: Br J Ophthalmol. 2023 May;107(5):e1. doi: 10.1136/bjophthalmol-2020-316149.corr1. PMID: 32709756; PMCID: PMC8005807.
3. Maghsoudlou P, Sood G, Gurnani B, Akhondi H. Cornea Transplantation. 2024 Feb 24. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan–. PMID: 30969512.
4. Mishra SK, Joshi A, Ginu PM, Sati A, Kumar SV. Corneal transplantation: A walk to vision. Med J Armed Forces India. 2023 Nov-Dec;79(6):645-650. doi: 10.1016/j.mjafi.2023.08.010. Epub 2023 Sep 13. PMID: 37981925; PMCID: PMC10654358.
5. Reeves SW, Ellwein LB, Kim T, Constantine R, Lee PP. Keratoconus in the Medicare population. Cornea. 2009 Jan;28(1):40-2. doi: 10.1097/ICO.0b013e3181839b06. PMID: 19092403; PMCID: PMC2723840.
6. Dorado-Cortez O, Poinard S, Epinat M, Collange F, Ninotta S, Goin P, Perrot JL, Gain P, Pollock G, Thuret G. The burden of medical contraindications to corneal donation: Time for review. PLOS Glob Public Health. 2024 Dec 31;4(12):e0003537. doi: 10.1371/journal.pgph.0003537. PMID: 39739771; PMCID: PMC11687701.
7. Liu S, Wong YL, Walkden A. Current Perspectives on Corneal Transplantation. Clin Ophthalmol. 2022 Mar 4;16:631-646. doi: 10.2147/OPTH.S289359. PMID: 35282172; PMCID: PMC8904759.
8. Szkodny D, Wróblewska-Czajka E, Wylęgała A, Nandzik M, Wylęgała E. Incidence of Complications Related to Corneal Graft in a Group of 758 Patients. J Clin Med. 2022 Dec 28;12(1):220. doi: 10.3390/jcm12010220. PMID: 36615021; PMCID: PMC9821265.
9. Zhang LJ, Dana R, Lorch AC, Elze T, Miller JW, Dohlman TH, Oke I; IRIS Registry Analytic Center Consortium. Complications After Pediatric Penetrating Keratoplasty: An IRIS Registry Study. Cornea. 2024 Dec 1;43(12):1555-1559. doi: 10.1097/ICO.0000000000003656. Epub 2024 Aug 13. PMID: 39137435; PMCID: PMC11537838.
10. Minasian M, Ayliffe W. Fixed dilated pupil following deep lamellar keratoplasty (Urrets-Zavalia syndrome). Br J Ophthalmol. 2002 Jan;86(1):115-6. doi: 10.1136/bjo.86.1.115. PMID: 11801517; PMCID: PMC1770988.
11. Zemba M, Stamate AC. Glaucoma after penetrating keratoplasty. Rom J Ophthalmol. 2017 Jul-Sep;61(3):159-165. doi: 10.22336/rjo.2017.30. PMID: 29450392; PMCID: PMC5710032.
12. Price MO, Price DA, Fairchild KM, Price FW Jr. Rate and risk factors for cataract formation and extraction after Descemet stripping endothelial keratoplasty. Br J Ophthalmol. 2010 Nov;94(11):1468-71. doi: 10.1136/bjo.2009.175174. Epub 2010 May 27. PMID: 20508038; PMCID: PMC2976606.
13. Dirisamer M, van Dijk K, Dapena I, Ham L, Oganes O, Frank LE, Melles GR. Prevention and management of graft detachment in descemet membrane endothelial keratoplasty. Arch Ophthalmol. 2012 Mar;130(3):280-91. doi: 10.1001/archophthalmol.2011.343. Epub 2011 Nov 14. PMID: 22084160.
14. Ghafarian S, Koo EH. Long-Term Graft Survival and Visual Outcomes in Deep Anterior Lamellar Keratoplasty (DALK): A Narrative Review. J Ophthalmol. 2026 Feb 23;2026:8120487. doi: 10.1155/joph/8120487. PMID: 41737287; PMCID: PMC12927906.

15. Azevedo Magalhaes O, Shalaby Bardan A, Zarei-Ghanavati M, Liu C. Literature review and suggested protocol for prevention and treatment of corneal graft rejection. Eye (Lond). 2020 Mar;34(3):442-450. doi: 10.1038/s41433-019-0517-9. Epub 2019 Jul 22. PMID: 31332293; PMCID: PMC7042259.

16. Espandar L, Carlson AN. Lamellar keratoplasty: a literature review. J Ophthalmol. 2013;2013:894319. doi: 10.1155/2013/894319. Epub 2013 Oct 7. PMID: 24223301; PMCID: PMC3816057.

17. Malta JB, Soong HK, Shtein R, Banitt M, Musch DC, Sugar A, Mian SI. Femtosecond laser-assisted keratoplasty: laboratory studies in eye bank eyes. Curr Eye Res. 2009 Jan;34(1):18-25. doi: 10.1080/02713680802535255. PMID: 19172466.

18. Aggarwal S, Kumari M, Bhatnagar N. Advancements in keratoplasty: Exploring newer techniques and imaging modalities for enhanced surgical outcomes. Saudi J Ophthalmol. 2024 Mar 1;39(4):337-348. doi: 10.4103/sjopt.sjopt_234_23. PMID: 41367843; PMCID: PMC12685274.

19. Al-Shami K, Almurabi S, Pop A, Vincent C, Popescu A, Pop IC, Bader NS, Faour H, Barhoush R, Karaja S. Emerging technologies in corneal transplantation: Innovations, challenges, and global implications. Transplantation Reports. 2025 Sep 1;10(3):100180. doi: 10.1016/j.tpr.2025.100180.

20. Ang M, He F, Lang S, Sabanayagam C, Cheng CY, Arundhati A, Mehta JS. Machine Learning to Analyze Factors Associated With Ten-Year Graft Survival of Keratoplasty for Cornea Endothelial Disease. Front Med (Lausanne). 2022 Jun 2;9:831352. doi: 10.3389/fmed.2022.831352. PMID: 35721073; PMCID: PMC9200960.

21. Ng Yin Ling C, He F, Lang S, Sabanayagam C, Cheng CY, Arundhati A, Mehta JS, Ang M. Interpretable Machine Learning-Based Risk Score for Predicting 10-Year Corneal Graft Survival After Penetrating Keratoplasty and Deep Anterior Lamellar Keratoplasty in Asian Eyes. Cornea. 2024 Jul 24;44(5):539-549. doi: 10.1097/ICO.0000000000003641. PMID: 39046776.

22. Muijzer MB, Hoven CMW, Frank LE, Vink G, Wisse RPL; Netherlands Corneal Transplant Network (NCTN). A machine learning approach to explore predictors of graft detachment following posterior lamellar keratoplasty: a nationwide registry study. Sci Rep. 2022 Oct 21;12(1):17705. doi: 10.1038/s41598-022-22223-y. PMID: 36271020; PMCID: PMC9586999.

23. Feizi S, Javadi MA, Bayat K, Arzaghi M, Rahdar A, Ahmadi MJ. Machine learning methods to identify risk factors for corneal graft rejection in keratoconus. Sci Rep. 2024 Nov 25;14(1):29131. doi: 10.1038/s41598-024-80967-1. PMID: 39587303; PMCID: PMC11589830.

24. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, Beam AL, Van Calster B, Celi LA, Denaxas S, Denniston AK, Ghassemi M, Heinze G, Kengne AP, Maier-Hein L, Liu X, Logullo P, McCradden MD, Liu N, Oakden-Rayner L, Singh K, Ting DS, Wynants L, Yang B, Reitsma JB, Riley RD, Collins GS, van Smeden M. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025 Mar 24;388:e082505. doi: 10.1136/bmj-2024-082505. PMID: 40127903; PMCID: PMC11931409.

25. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009. PMID: 22007046.

26. Wisse RP, Muijzer MB, Hoven CM, Van Luijk CM, Vink G, Frank L. A machine learning approach to identify predictors of adverse events in posterior lamellar keratoplasty: a nationwide registry study. Acta Ophthalmologica. 2021 Mar 1;99(S266):6-. Abstracts of the Netherlands Ophthalmological Society (NOG) Annual Congress, 24–26 March 2021, Maastricht, The Netherlands. doi: 10.1111/aos.14800.

27. Joseph NM, Benetz BA, Menegay H, Oellerich S, Baydoun L, Melles G, Lass JH, Wilson D. Early detection of at-risk keratoplasties and prediction of future corneal graft rejection from pre-diagnosis endothelial cell images. In Medical Imaging 2021: Computer-Aided Diagnosis 2021 Feb 18 (Vol. 11597, pp. 543-552). SPIE. doi: 10.1117/12.2582171.

28. Joseph N, Benetz BA, Chirra P, Menegay H, Oellerich S, Baydoun L, Melles GRJ, Lass JH, Wilson DL. Machine Learning Analysis of Postkeratoplasty Endothelial Cell Images for the Prediction of Future Graft Rejection. Transl Vis Sci Technol. 2023 Feb 1;12(2):22. doi: 10.1167/tvst.12.2.22. PMID: 36790821; PMCID: PMC9940770.

29. Heslinga FG, Lucassen RT, van den Berg MA, van der Hoek L, Pluim JPW, Cabrerizo J, Alberti M, Veta M. Corneal pachymetry by AS-OCT after Descemet's membrane endothelial keratoplasty. Sci Rep. 2021 Jul 7;11(1):13976. doi: 10.1038/s41598-021-93186-9. PMID: 34234179; PMCID: PMC8263705.

30. Heslinga FG, Alberti M, Pluim JPW, Cabrerizo J, Veta M. Quantifying Graft Detachment after Descemet's Membrane Endothelial Keratoplasty with Deep Convolutional Neural Networks. Transl Vis Sci Technol. 2020 Aug 21;9(2):48. doi: 10.1167/tvst.9.2.48. PMID: 32884855; PMCID: PMC7445365.

31. Yousefi S, Takahashi H, Hayashi T, Tampo H, Inoda S, Arai Y, Tabuchi H, Asbell P. Predicting the likelihood of need for future keratoplasty intervention using artificial intelligence. Ocul Surf. 2020 Apr;18(2):320-325. doi: 10.1016/j.jtos.2020.02.008. Epub 2020 Feb 28. PMID: 32120005.

32. Hayashi T, Iliasian RM, Matthaei M, Schrittenlocher S, Masumoto H, Tanabe M, Tabuchi H, Siggel R, Bachmann B, Cursiefen C, Siebelmann S. Transferability of an Artificial Intelligence Algorithm Predicting Rebubblings After Descemet Membrane Endothelial Keratoplasty. Cornea. 2023 May 1;42(5):544-548. doi: 10.1097/ICO.0000000000003049. Epub 2022 Apr 21. PMID: 35543586.

33. Kundu G, D'Souza S, Modak D, Balaraj S, Shetty R, Nuijts RMMA, Narasimhan R, Roy AS. Using Artificial Intelligence for an Efficient Prediction of Outcomes of Deep Anterior Lamellar Keratoplasty (DALK) in Advanced Keratoconus. Transl Vis Sci Technol. 2025 May 1;14(5):30. doi: 10.1167/tvst.14.5.30. PMID: 40434371; PMCID: PMC12126132.

34. Kim MS, Kim H, Lee HK, Kim CY, Choi W. Artificial Intelligence in Predicting Ocular Hypertension After Descemet Membrane Endothelial Keratoplasty. Invest Ophthalmol Vis Sci. 2025 Jan 2;66(1):61. doi: 10.1167/iovs.66.1.61. PMID: 39869086; PMCID: PMC11771522.

35. Hayashi T, Masumoto H, Tabuchi H, Ishitobi N, Tanabe M, Grün M, Bachmann B, Cursiefen C, Siebelmann S. A deep learning approach for successful big-bubble formation prediction in deep anterior lamellar keratoplasty. Sci Rep. 2021 Sep 17;11(1):18559. doi: 10.1038/s41598-021-98157-8. PMID: 34535722; PMCID: PMC8448733.
36. Hayashi T, Tabuchi H, Masumoto H, Morita S, Oyakawa I, Inoda S, Kato N, Takahashi H. A Deep Learning Approach in Rebubbling After Descemet's Membrane Endothelial Keratoplasty. Eye Contact Lens. 2020 Mar;46(2):121-126. doi: 10.1097/ICL.0000000000000634. PMID: 31425350.
37. Karaca EE, Bulut Ustael A, Keçeli AS, Kaya A, Uçan A, Evren Kemer O. Predicting Success in Descemet Membrane Endothelial Keratoplasty Using Machine Learning. Cornea. 2025 Feb 1;44(2):189-195. doi: 10.1097/ICO.0000000000003599. Epub 2024 Jun 20. PMID: 38913970.
38. Glatz A, Böhringer D, Zander DB, Grewing V, Fritz M, Müller C, Bixler S, Reinhard T, Wacker K. Three-Dimensional Map of Descemet Membrane Endothelial Keratoplasty Detachment: Development and Application of a Deep Learning Model. Ophthalmol Sci. 2021 Sep 29;1(4):100067. doi: 10.1016/j.xops.2021.100067. PMID: 36275190; PMCID: PMC9562373.
39. Feng X, Zhang X, Shi X, Li L, Wang S. ST-ITEF: Spatio-Temporal Intraoperative Task Estimating Framework to recognize surgical phase and predict instrument path based on multi-object tracking in keratoplasty. Med Image Anal. 2024 Jan;91:103026. doi: 10.1016/j.media.2023.103026. Epub 2023 Nov 13. PMID: 37976868.
40. Dreesbach M, Böhringer D, Betancor PK, Glegola M, Maier PC, Reinhard T, Heinzelmann S. Quality Control in the Corneal Bank with Artificial Intelligence: Comparison of a New Deep Learning-based Approach with Conventional Endothelial Cell Counting by the "Rhine-Tec Endothelial Analysis System". Klin Monbl Augenheilkd. 2024 Jun;241(6):734-740. English, German. doi: 10.1055/a-2299-8117. Epub 2024 Apr 4. PMID: 38574759.
41. Vigueras-Guillén JP, van Rooij J, Engel A, Lemij HG, van Vliet LJ, Vermeer KA. Deep Learning for Assessing the Corneal Endothelium from Specular Microscopy Images up to 1 Year after Ultrathin-DSAEK Surgery. Transl Vis Sci Technol. 2020 Aug 21;9(2):49. doi: 10.1167/tvst.9.2.49. PMID: 32884856; PMCID: PMC7445361.
42. Bitton K, Zéboulon P, Ghazal W, Rizk M, Elahi S, Gatinel D. Deep Learning Model for the Detection of Corneal Edema Before Descemet Membrane Endothelial Keratoplasty on Optical Coherence Tomography Images. Transl Vis Sci Technol. 2022 Dec 1;11(12):19. doi: 10.1167/tvst.11.12.19. PMID: 36583911; PMCID: PMC9807180.
43. O'Brien RC, Ishwaran H, Szczotka-Flynn LB, Lass JH; Cornea Preservation Time Study (CPTS) Group. Random Survival Forests Analysis of Intraoperative Complications as Predictors of Descemet Stripping Automated Endothelial Keratoplasty Graft Failure in the Cornea Preservation Time Study. JAMA Ophthalmol. 2021 Feb 1;139(2):191-197. doi: 10.1001/jamaophthalmol.2020.5743. PMID: 33355637; PMCID: PMC7758826.
44. Charilaou P, Battat R. Machine learning models and over-fitting considerations. World J Gastroenterol. 2022 Feb 7;28(5):605-607. doi: 10.3748/wjg.v28.i5.605. PMID: 35316964; PMCID: PMC8905023.
45. Goetz L, Seedat N, Vandersluis R, van der Schaar M. Generalization-a key challenge for responsible AI in patient-facing clinical applications. NPJ Digit Med. 2024 May 21;7(1):126. doi: 10.1038/s41746-024-01127-3. PMID: 38773304; PMCID: PMC11109198.
46. Montesinos López OA, Montesinos López A, Crossa J (2022). 'Multivariate Statistical Machine Learning Methods for Genomic Prediction. Chapter 4, Overfitting, Model Tuning, and Evaluation of Prediction Performance'. Cham (CH): Springer; 2022. Available at: https://www.ncbi.nlm.nih.gov/books/NBK583970/ doi: 10.1007/978-3-030-89010-0_4 (Accessed: 03 March, 2026).
47. Eche T, Schwartz LH, Mokrane FZ, Dercle L. Toward Generalizability in the Deployment of Artificial Intelligence in Radiology: Role of Computation Stress Testing to Overcome Underspecification. Radiol Artif Intell. 2021 Oct 27;3(6):e210097. doi: 10.1148/ryai.2021210097. PMID: 34870222; PMCID: PMC8637230.
48. Wang SY, Pershing S, Lee AY; AAO Taskforce on AI and AAO Medical Information Technology Committee. Big data requirements for artificial intelligence. Curr Opin Ophthalmol. 2020 Sep;31(5):318-323. doi: 10.1097/ICU.0000000000000676. PMID: 32657996; PMCID: PMC8164167.
49. Toprak I, Yaylalı V, Yildirim C. A combination of topographic and pachymetric parameters in keratoconus diagnosis. Cont Lens Anterior Eye. 2015 Oct;38(5):357-62. doi: 10.1016/j.clae.2015.04.001. Epub 2015 Apr 27. PMID: 25936634.
50. Miháltz K, Kovács I, Takács A, Nagy ZZ. Evaluation of keratometric, pachymetric, and elevation parameters of keratoconic corneas with pentacam. Cornea. 2009 Oct;28(9):976-80. doi: 10.1097/ICO.0b013e31819e34de. PMID: 19724217.
51. Razak N, Mohd Ali B, Halim WHWA. Evaluation of Corneal Tomographic, Biomechanical and Pachymetric Characteristics in Patients With Keratoconus, Their First-Degree Relatives, and Normal Individuals. Cornea. 2026 May 1;45(5):541-549. doi: 10.1097/ICO.0000000000004058. Epub 2025 Dec 2. PMID: 41330423; PMCID: PMC13034747.
52. Ambrósio R Jr, Salomão MQ, Barros L, da Fonseca Filho JBR, Guedes J, Neto A, Machado AP, Lopes BT, Sena N Jr, Esporcatte LPG. Multimodal diagnostics for keratoconus and ectatic corneal diseases: a paradigm shift. Eye Vis (Lond). 2023 Nov 3;10(1):45. doi: 10.1186/s40662-023-00363-0. PMID: 37919821; PMCID: PMC10623885.
53. Piñero DP, Alió JL, Alesón A, Escaf Vergara M, Miranda M. Corneal volume, pachymetry, and correlation of anterior and posterior corneal shape in subclinical and different stages of clinical keratoconus. J Cataract Refract Surg. 2010 May;36(5):814-25. doi: 10.1016/j.jcrs.2009.11.012. PMID: 20457375.
54. Saad A, Gatinel D. Topographic and tomographic properties of forme fruste keratoconus corneas. Invest Ophthalmol Vis Sci. 2010 Nov;51(11):5546-55. doi: 10.1167/iovs.10-5369. Epub 2010 Jun 16. PMID: 20554609.
55. Ambrósio R Jr, Caiado AL, Guerra FP, Louzada R, Sinha RA, Luz A, Dupps WJ, Belin MW. Novel pachymetric parameters based on corneal tomography for diagnosing keratoconus. J Refract Surg. 2011 Oct;27(10):753-8. doi: 10.3928/1081597X-20110721-01. Epub 2011 Jul 29. PMID: 21800785.