ISSN: 3152-8686

Vol3; Issue2, May 2026

JOURNAL OF
TAZEEZ FOR PUBLIC HEALTH
AN OFFICIAL JOURNAL OF SAUDI HEALTH PROMOTION AND EDUCATION ASSOCIATION

Systematic review

Artificial intelligence-assisted diagnosis and surgical planning in oculoplastic disorders: a systematic review

Meshal Abdulrahman Ogran[1], Ahmed Abou El Fotouh[2], Bandar Mohammed Asiri[3], Turki Bijad Alotaibi[4], Faisal Nasser M. Alahmari[5]

1. Senior Registrar, Ophthalmology, Aseer Central Hospital, Abha, Saudi Arabia
2. Ophthalmology Consultant, Armed Forces Hospitals Southern Region (AFHSR), Saudi Arabia
3. Ophthalmology Resident, Aseer Central Hospital, Abha, Saudi Arabia
4. Ophthalmology Resident, Armed Forces Hospitals Southern Region (AFHSR), Saudi Arabia
5. General Practitioner, Aseer Central Hospital, Abha, Saudi Arabia

Abstract

Background: Artificial intelligence (AI) is increasingly used to oculoplastic disorders, where diagnosis and surgical planning rely on imaging and objective periocular measurements. Our review summarizes studies on AI-assisted diagnosis, quantification, and surgical planning in oculoplastic. **Methods:** A systematic review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles. We conduct a literature search on PubMed (MEDLINE), Scopus, and Web of Science were searched from inception to the final search date for original human studies which evaluate AI for oculoplastic diagnosis, assessment, or surgical planning. Eligible studies reported quantitative performance metrics. **Results:** Twelve studies were included, computed tomography based thyroid eye disease models showed high discrimination for disease severity and compressive optic neuropathy screening (area under the receiver operating characteristic curve (AUC) 0.99; independent test AUC 0.92). Automated orbital segmentation achieved Dice values 0.90. Ptosis detection models achieved strong performance on high quality images (AUC 0.92) but degraded on low quality images. Automated eyelid measurement errors were sub millimeter, and postoperative appearance prediction systems reported eye overlap ratios 0.85 to 0.87. Eyelid tumor models using clinical images or histopathology shoe high diagnostic accuracy (AUCs 0.92 to 1.00). **Conclusions:** AI shows strong potential for oculoplastic diagnosis and planning, but external validation, and clinical impact evaluation were the important gaps.

Keywords: Artificial intelligence; deep learning; machine learning; oculoplastic; thyroid eye disease; blepharoptosis

Published: May 3, 2026

https://doi.org/10.65759/p5a2y968

Introduction

Oculoplastic disorders including eyelid malposition's, blepharoptosis, periocular tumors, and thyroid eye disease (TED), are common causes of visual dysfunction, ocular surface symptoms, cosmetic concern, and impaired quality of life. Clinical assessment and surgical planning in oculoplastics are highly image dependent, relying on external examination, standardized facial photography, cross sectional imaging, and histopathology for tumor diagnosis. This setting is well suited to Artificial intelligence (AI), machine learning and deep learning, which have shown performance in extracting clinically relevant patterns from medical images and other high-dimensional ophthalmic data. Recent reviews describe rapid growth of AI in ophthalmology, driven by larger datasets, and wider availability of digital imaging workflows (Jin et al. 2022; Srivastava et al. 2022).

AI support individualized care by quantifying morphology, and assist decision making, these capabilities are relevant to oculoplastics, where objective measurements and prediction of postoperative appearance impact surgical approach, and shared decision making (Khan et al. 2025; Jin et al. 2022). Methodological quality is important barrier to implementation, and the Checklist for AI in Medical Imaging (CLAIM) (Fig 1) developed to improve transparency in dataset description, and evaluation, enabling better appraisal and reproducibility. (Mongan et al. 2020; Tejani et al. 2024), reporting guidance for predictive models and clinical trials has evolved for AI interventions, including TRIPOD+AI for prediction model reporting and Consolidated Standards of Reporting Trials–Artificial Intelligence (CONSORT-AI)/ Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence (SPIRIT-AI) extensions for trials and protocols. (Collins et al. 2024; Ibrahim et al. 2021).

Scoping work in TED indicate heterogeneity in AI objectives, and validation approaches, underscoring the need for synthesis and identification of gaps (Chng et al. 2023). A focused synthesis of AI assisted diagnosis and AI supported surgical planning in oculoplastic conditions is needed, so our systematic review aims to summarize AI applications for diagnosis, assessment, and surgical planning in oculoplastic disorders, and map current studies, and barriers to clinical translation. (Jin et al. 2022)

Methods

This systematic review was conducted according to the PRISMA guidelines (Fig 2). A comprehensive electronic literature search performed in PubMed (MEDLINE), Scopus, and Web of Science (Core Collection) to identify original studies about AI for diagnosis, assessment, or surgical planning in oculoplastic disorders. Searches were run from database inception to the final search date, with no restrictions on country. Only articles available in English and with accessible full text were considered for inclusion. We included studies involving patients with oculoplastic conditions in which an AI model was developed and evaluated using clinical photographs, imaging, or histopathology, and reported quantitative performance outcomes. Reviews, editorials, protocols, animal studies, and conference abstracts without full methods will be excluded.

Articles identified through a structured search of the literature complemented by citation screening of included papers. After removal of duplicates, titles and abstracts were screened for the eligibility criteria, followed by full text review to confirm inclusion. Disagreements were resolved through discussion, and reasons for exclusion at full text were documented.

Data were extracted using a standardized form capturing author, year, country, setting, study design, target disorder, data modality, AI task, model architecture, dataset size and splits, reference standard, and validation approach. We also extracted intended use and any direct comparisons with clinician performance. Extracted fields were cross checked for consistency before synthesis.

Primary outcomes were discrimination and diagnostic performance for classification tasks (area under the receiver operating characteristic curve, accuracy, sensitivity, specificity) and technical performance for segmentation tasks. When multiple models were evaluated in one study, we extracted the best performing model as defined by the authors on the relevant test set, while also documenting performance in internal and prospective validation datasets when available.

Given clinical and methodological heterogeneity, different disorders, modalities, architectures, reference standards, and outcome metrics; results were analyzed narratively rather than pooled in meta-analysis. Studies were grouped by clinical application. Findings were summarized in Table 1 and 2, characteristics of included studies and main findings and key performance results.

Risk of bias and applicability concerns were appraised using design-appropriate criteria for diagnostic accuracy and prediction modelling, focusing on participant selection, reference standards, handling of missing data, overfitting, and the presence or absence of external validation. We also noted reporting elements relevant to clinical translation, including prospective testing and performance under variable image quality.

ISSN: 3152-8686

Fig 1: Checklist for AI in medical imaging

Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update

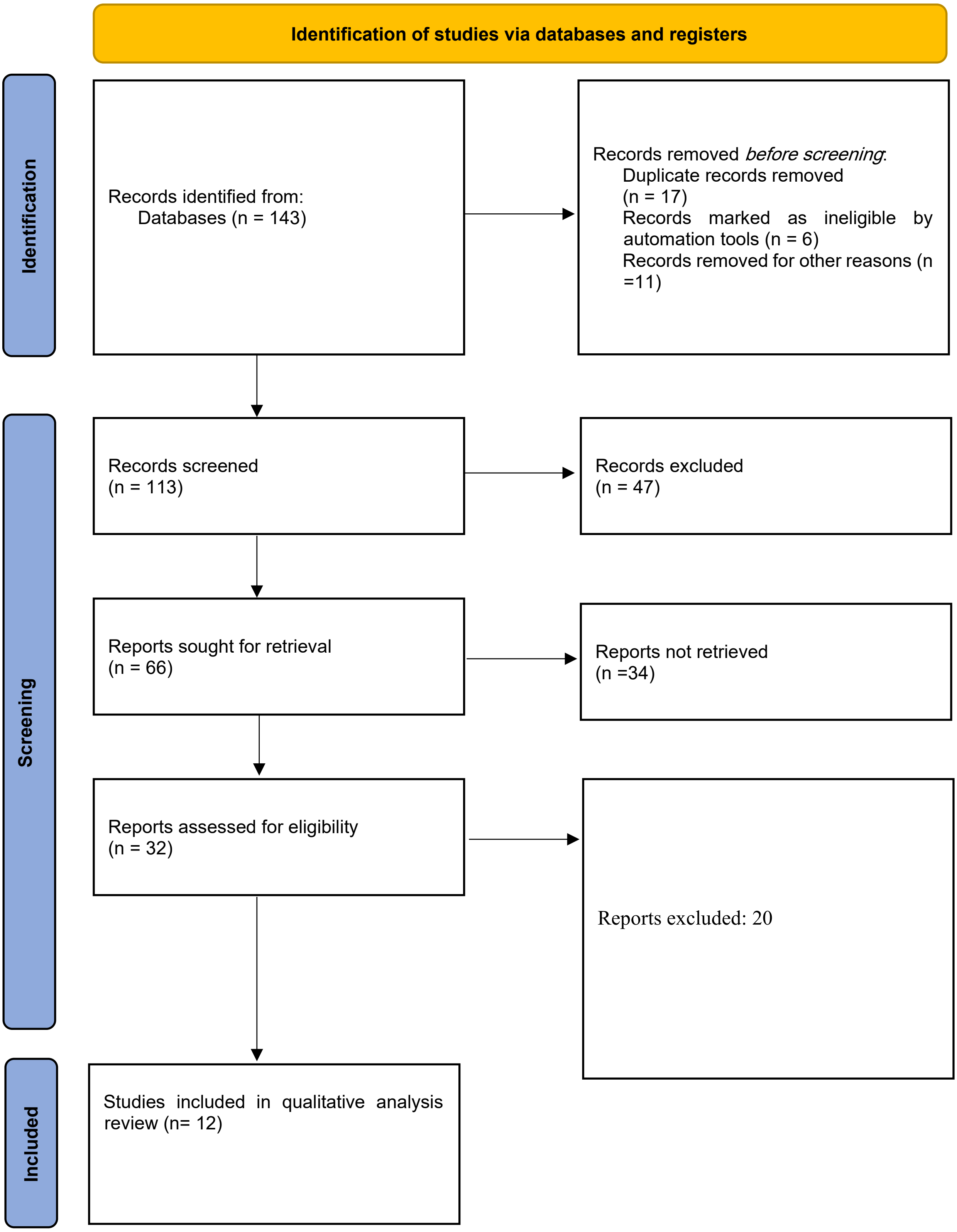
Section/Topic	No.	Item
TITLE/ABSTRACT		
	1	Identification as a study of AI methodology, specifying the category of technology used (eg, deep learning)
ABSTRACT		
	2	Summary of study design, methods, results, and conclusions
INTRODUCTION		
	3	Scientific and/or clinical background, including the intended use and role of the AI approach
	4	Study aims, objectives, and hypotheses
METHODS		
Study Design	5	Prospective or retrospective study
	6	Study goal
Data	7	Data sources
	8	Inclusion and exclusion criteria
	9	Data preprocessing
	10	Selection of data subsets
	11	De-identification methods
	12	How missing data were handled
	13	Image acquisition protocol
Reference Standard	14	Definition of method(s) used to obtain reference standard
	15	Rationale for choosing the reference standard
	16	Source of reference standard annotations
	17	Annotation of test set
	18	Measures of inter- and intrarater variability of features described by the annotators
Data Partitions	19	How data were assigned to partitions
	20	Level at which partitions are disjoint
Testing Data	21	Intended sample size
Model	22	Detailed description of model
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters
Training	25	Details of training approach
	26	Method of selecting the final model
	27	Ensembling techniques
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability
	32	Evaluation on internal data
	33	Testing on external data
	34	Clinical trial registration
RESULTS		
Data	35	Numbers of patients or examinations included and excluded
	36	Demographic and clinical characteristics of cases in each partition
Model Performance	37	Performance metrics and measures of statistical uncertainty
	38	Estimates of diagnostic performance and their precision
	39	Failure analysis of incorrectly classified cases
DISCUSSION		
	40	Study limitations
	41	Implications for practice, including intended use and/or clinical role
OTHER INFORMATION		
	42	Provide a reference to the full study protocol or to additional technical details
	43	Statement about the availability of software, trained model, and/or data
	44	Sources of funding and other support; role of funders

Indicate page and/or line number for each checklist item that is present.

03/05/2026

ISSN: 3152-8686

Fig 2: PRISMA consort chart



ISSN: 3152-8686

Results

Twelve original studies were included and grouped into three clinical application areas, blepharoptosis diagnosis, TED assessment (n=2), surgical planning prediction (n=6), and eyelid tumor detection or triage (n=4). Most studies were retrospective development studies using curated image datasets, with minority incorporating prospective or external validation cohorts.

Two studies applied deep learning to orbital computed tomography (CT). One Ophthalmology Science study used 1,187 annotated axial CT images from 141 participants in two academic centers (normal, mild TED, and severe TED with compressive optic neuropathy) and get 89.5% accuracy for three class classification; binary AUCs ranged from 0.982 to 0.993 for TED severity discrimination and the reported AUC for optic neuropathy screening was 0.922 in an independent test set. A second study trained a convolutional neural network (CNN) to segment extraocular muscles and orbital fat on CT (n=281), reporting Dice coefficients of 0.902 (muscles) and 0.921 (fat); using derived imaging features, the TED classification model achieved 83.8% accuracy and AUC 0.929.

Four studies addressed ptosis detection and two focused on postoperative appearance prediction, a convolutional neural network approach for ptosis screening based on facial photographs reported a best performing model with sensitivity 90.1% and specificity 82.4% on the study dataset. Another VGG based Blepharoptosis-CNN trained on periocular images reported AUC 0.918 with sensitivity 98% and specificity 88% on high-definition images, and performance dropped on lower quality images (AUC 0.700). A deep learning-based system estimated eyelid parameters with mean absolute errors of 0.369 mm (MRD1) and 0.609 mm (MRD2). For surgical planning management, an XGBoost based clinical decision model for ptosis surgery used 103 entries from 152 eyes and achieved accuracies of 0.848 to 0.857 with AUCs up to 0.833 depending on the input feature set. Postoperative appearance prediction was addressed by two generative approaches: a generative adversarial network (GAN) based system trained on 407 paired pre and postoperative photographs reported mean overlap ratios of 0.85 between predicted and actual eyes and favorable subjective similarity scores. A diffusion model approach trained on 39 preoperative images reported an overlap ratio of 0.87 between generated and real postoperative eyes.

Four studies developed AI tools to classify eyelid tumors, two used clinical photographs, a journal of Big Data study reported prospective test performance with AUC 0.966 and accuracy 0.889 on 36 patients. A NPJ Digital Medicine study developed a mobile application; in external validation, the combined model show AUC 0.917 and accuracy 0.921. Histopathology-based classification was evaluated using whole-slide images, achieving AUC 0.998 and accuracy 0.983 on internal testing; prospective testing and clinician assistance analyses were also reported. A YOLOv7 based detection pipeline with a Vision Transformer classifier found AUC 0.945 internally and AUC 0.915 externally for benign malignant discrimination. Limitations included single center datasets, image quality issues, limited external validation, and sparse reporting of clinical impact overall.

ISSN: 3152-8686

Table 1. Characteristics of the included studies

Study	aim	Design	Data modality	AI task	Model	Dataset size	Validation
Lee et al., 2024	TED screening & severity classification	Retrospective model development (2 centers)	Orbital CT	3-class + binary classification	Deep learning (VGG-16)	141 participants; 1,187 annotated images	Hold out testing + independent test for CON
Alkhadrawi et al., 2024	TED diagnosis using orbital tissue segmentation	Retrospective	Orbital CT	Segmentation + classification	CNN segmentation; diagnostic model using extracted features	281 participants	Internal test set
Hung et al., 2021	Blepharoptosis identification	Retrospective	Facial photographs	Binary classification	CNN	434 photos after exclusions	Cross-validation, hold-out testing
Abascal-Azanza et al., 2023	Automated ptosis screening	Retrospective	Periocular images	Binary classification	VGG-16-based CNN	Training: 6,180 images after augmentation	Internal test + low-quality field test
Lou et al., 2021	Automated eyelid morphology quantification in blepharoptosis	Retrospective	Facial images	Measurement	Deep learning computer-aided system	3,130 images	Internal testing, validation
Sun et al., 2022	Postoperative appearance prediction after ptosis correction	Retrospective	Paired pre post-op facial photos	Image-to-image prediction	GAN (pix2pix) + segmentation modules	407 paired images; eyelid, iris segmentation trained on 25,000 images	Internal testing
Huang et al., 2024	Postoperative appearance prediction in ptosis	Retrospective	Preoperative facial photos	Image generation prediction	Conditional diffusion model	39 preoperative images	Internal testing
Song et al., 2021	Clinical decision modelling for ptosis surgery	Retrospective follow-up cohort	Clinical features	Classification, decision support	XGBoost	103 entries from 152 eyes	Internal testing; 2D vs 3D feature sets
Hui et al., 2022	Eyelid tumor recognition & triage	Model development +	Clinical photos	Multi-class classification	Deep learning ensemble	309 photos dev; prospective: 36 photos	Prospective test set

03/05/2026

Study	aim	Design	Data modality	AI task	Model	Dataset size	Validation
		prospective test					
Luo et al., 2022	Differential diagnosis of eyelid basal cell carcinoma (BCC) and sebaceous gland carcinoma	Retrospective + prospective evaluation	Whole-slide histopathology images	Binary classification	Deep learning on whole-slide image (WSIs)	116 BCC; 129 sebaceous gland carcinoma WSIs, sections	Internal test + prospective validation
Jiang et al., 2024	Benign and malignant eyelid tumor discrimination	Retrospective development	Clinical images	Detection + classification	You Only Look Once version 7 (YOLOv7)+ ViT	1,307 images; internal test 571; external test 235	External test reported
Hui et al., 2025	Mobile app for eyelid tumor diagnosis	Retrospective development + external validation	Smartphone clinical images	Classification	Deep learning ensemble in app	1,195 eyes from 616 patients; external validation 38 patients	External validation

Table 2. Main findings of the included studies

Study	Main performance results	Main finding
Lee et al., 2024	Accuracy 89.5% (3 class); AUC 0.982 (mild), 0.993 (severe), 0.985 (TED); convolutional neural network AUC 0.922	CT-based deep learning achieved high discrimination for TED severity and screening for compressive optic neuropathy.
Alkhadrawi et al., 2024	Dice 0.902 extraocular muscle and 0.921 (fat); TED classification accuracy 0.838; AUC 0.929	Automated segmentation of orbital tissues supported accurate TED classification from CT-derived features.
Hung et al., 2021	Sensitivity 90.1%; specificity 82.4%; area under the receiver operating characteristic curve 0.953	CNNs could identify blepharoptosis from facial photographs with clinically relevant sensitivity, specificity.
Abascal-Azanza et al., 2023	high-definition images: AUC 0.918; sensitivity 98%; specificity 88%. Low-quality images: AUC 0.700	Performance was strong on high-quality images but degraded substantially with reduced image quality.
Lou et al., 2021	Mean absolute error: margin reflex distance 1 (MRD1) 0.369 mm; MRD2 0.609 mm; scleral show 0.240 mm; margin reflex distance 2 (MRD2) errors 4.80%, 9.16%	Automated eyelid measurements approximated clinician assessment and may support objective pre/postoperative evaluation.

03/05/2026

Study	Main performance results	Main finding
Sun et al., 2022	Mean eye overlap ratio 0.85; by procedure 0.87/0.83/0.82; subjective similarity scores 0.80	GAN-based system predicted postoperative appearance with high structural similarity on paired photo evaluation.
Huang et al., 2024	Overlap ratio 0.87 between predicted and actual eyes (n=39)	Diffusion-model generation produced close postoperative-eye similarity in a small proof-of-concept dataset.
Song et al., 2021	Accuracy 0.848–0.857; AUC up to 0.833 depending on 2D vs 3D inputs	An XGBoost decision model using clinical and measurement features showed moderate–good discrimination for surgical decision support.
Hui et al., 2022	Prospective test: AUC 0.966; accuracy 0.889	Deep learning achieved high prospective performance for eyelid tumor recognition using clinical photos.
Luo et al., 2022	Internal test: AUC 0.998; accuracy 0.983. Prospective: AUC 0.970; accuracy 0.940	WSI-based model accurately differentiated BCC from sebaceous carcinoma and was evaluated prospectively.
Jiang et al., 2024	Benign malignant discrimination: internal AUC 0.945; external AUC 0.915	A YOLO+ Vision Transformer pipeline maintained high AUC on an external dataset for malignant risk discrimination.
Hui et al., 2025	External validation: combined model AUC 0.917; accuracy 0.921	A smartphone-based application showed strong external validation performance for eyelid tumor diagnosis.

Discussion

Our systematic review indicates that AI applications for oculoplastic disorders are expanding in two clinically relevant domains; image-based diagnosis, triage and support for surgical planning or outcome estimation. The broader oculoplastics literature describes AI as suited to this subspecialty because decision making depends on standardized imaging, objective morphologic features, and repeatable measurements over time (Ing et al. 2025).

Despite promising performance signals reported in many ophthalmic AI studies, translation into routine oculoplastics practice were limited. Provider-level articles suggest clinicians anticipate near term impact but report low current uptake and gaps in AI literacy and implementation readiness (Fazekas et al. 2025). This mismatch, high research activity vs. slow deployment, indicate the need for implementation science with algorithm development.

In medical imaging external validation is uncommon and performance decreases when evaluated on truly independent datasets, underscoring dataset shift and hidden confounding as major threats to clinical reliability (Yu et al. 2022). For oculoplastics where image acquisition conditions, population mix, and disease spectrum vary between clinics, the risk of degraded performance is substantial unless multicenter evaluation is routine.

Reporting completeness determines how findings can be interpreted. Regarding imaging AI studies, CLAIM provides a practical checklist to standardize reporting of data handling, ground truth labeling, model development, and evaluation, details that affect bias and reproducibility (Mongan et al. 2020). For prediction models TRIPOD + AI extends transparent reporting requirements and show clearer terminology and evaluation standards, which is essential when oculoplastic tools are framed as risk scores or prognostic aids (Collins et al. 2024).

Equity and safety considerations must be treated as core outcomes, not optional extras. Bias can enter at every stage of the ophthalmic AI lifecycle, producing subgroup performance gaps that may not be visible in aggregate metrics (Nakayama et al. 2024). Medical imaging literature also stresses that fairness requires definition and measurement, and that mitigation strategies should be planned prospectively (Ricci Lara et al. 2022). Emerging ophthalmology specific methods show feasibility of reducing sex and age-related bias while maintaining accuracy (Tan et al. 2024).

Prospective evaluation and post deployment monitoring are critical for oculoplastic AI tools that could impact diagnosis and surgical decisions. CONSORT-AI and SPIRIT-AI provide reporting requirements for clinical trials and protocols of AI interventions, respectively, and should guide future oculoplastics studies toward clinically meaningful endpoints and transparent human–AI interaction descriptions (Liu et al. 2020; Rivera et al. 2020). Because calibration and performance can drift over time as clinical environments change, continuous monitoring frameworks are increasingly important (Davis et al. 2020).

Conclusion

AI applications in oculoplastic improve image-based diagnosis, objective measurement, and surgical planning. In the TED, blepharoptosis, and eyelid tumor studies, many models achieved high discrimination and good quantitative agreement with clinical assessments, and some tools were tested prospectively. Most studies were retrospective and derived from single center or curated datasets, with performance sensitive to image quality and dataset shift.

List of abbreviations

AI, Artificial intelligence; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; MEDLINE, Medical Literature Analysis and Retrieval System Online; AUC, Area under the receiver operating characteristic curve; TED, Thyroid eye disease; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; TRIPOD+AI, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis–Artificial Intelligence; CONSORT-AI, Consolidated Standards of Reporting Trials–Artificial Intelligence; SPIRIT-AI, Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence; CT, Computed tomography; CNN, Convolutional neural network; VGG, Visual Geometry Group; VGG-16, Visual Geometry Group 16-layer network; CON, Compressive optic

03/05/2026

neuropathy; MRD1, Margin reflex distance 1; MRD2, Margin reflex distance 2; GAN, Generative adversarial network; XGBoost, Extreme Gradient Boosting; NPJ, Nature Partner Journals; WSI, Whole-slide image; WSIs, Whole-slide images; BCC, Basal cell carcinoma; YOLO, You Only Look Once; YOLOv7, You Only Look Once version 7; ViT, Vision Transformer; 2D, Two-dimensional; 3D, Three-dimensional; n, Number/sample size; mm, Millimeter; Fig, Figure; vs, versus; post-op, postoperative; dev, development.

References

Abascal Azanza C, Barrio-Barrio J, Ramos Cejudo J, Ybarra Arróspide B, Devoto MH. Development and validation of a convolutional neural network to identify blepharoptosis. Sci Rep. 2023;13:17585. doi:10.1038/s41598-023-44686-3.

Alkhadrawi AM, Lin LY, Langarica SA, Kim K, Ha SK, Lee NG, et al. Deep-learning based automated segmentation and quantitative volumetric analysis of orbital muscle and fat for diagnosis of thyroid eye disease. Invest Ophthalmol Vis Sci. 2024;65(5):6. doi:10.1167/iovs.65.5.6.

Chng C-L, Zheng K, Kwee AK, Lee M-HH, Ting D, Wong CP, et al. Application of artificial intelligence in the assessment of thyroid eye disease (TED) – a scoping review. Front Endocrinol (Lausanne). 2023 Dec 20;14:1300196. doi:10.3389/fendo.2023.1300196. PMCID: PMC10761414.

Collins GS, Moons KGM, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. 2024. PMCID: PMC11019967.

Davis SE, et al. Detection of Calibration Drift in Clinical Prediction Models to Inform Model Updating. 2020. PMCID: PMC8627243.

Fazekas B, et al. Artificial Intelligence in Oculoplastics: A Survey-Based Study on Provider Perspective. Cureus. 2025. PMCID: PMC12032587.

Guo X, et al. MedShift: Automated Identification of Shift Data for Medical Image Dataset Curation. 2023. PMCID: PMC10513895.

Huang S, Xie J, Yang B, Gao Q, Ye J. PtosisDiffusion: a training-free workflow for precisely predicting post-operative appearance in blepharoptosis patients based on diffusion models. Front Cell Dev Biol. 2024;12:1459336. doi:10.3389/fcell.2024.1459336.

Hui S, Dong L, Zhang K, Nie Z, Jiang X, Li H, et al. Noninvasive identification of benign and malignant eyelid tumors using clinical images via deep learning system. J Big Data. 2022;9:84. doi:10.1186/s40537-022-00634-y.

Hui S, Xie J, Dong L, Wei L, Dai W, Li D. Deep learning-based mobile application for efficient eyelid tumor recognition in clinical images. NPJ Digit Med. 2025;8:185. doi:10.1038/s41746-025-01539-9.

Hung JY, Perera C, Chen KW, Myung D, Chiu HK, Fuh CS, et al. A deep learning approach to identify blepharoptosis by convolutional neural networks. Int J Med Inform. 2021 Apr;148:104402. doi:10.1016/j.ijmedinf.2021.104402.

Ing E, Bondok M. Oculoplastics and Augmented Intelligence: A Literature Review. J Clin Med. 2025 Sep 28;14(19):6875. doi:10.3390/jcm14196875. PMCID: PMC12524889.

Jiang J, Liu H, He L, Pei M, Lin T, Yang H, et al. HM_ADET: a hybrid model for automatic detection of eyelid tumors based on photographic images. Biomed Eng Online. 2024;23:25. doi:10.1186/s12938-024-01221-3.

Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. Adv Ophthalmol Pract Res. 2022 Aug 24;2(3):100078. doi:10.1016/j.aopr.2022.100078. PMCID: PMC10577833.

Khan AYR, Malik MB, Naveed M, Hussain M, Kamran H, Hussain A, et al. Artificial Intelligence in Ophthalmology: Practical Applications, Subspecialty Evidence and Real-World Deployment. Cureus. 2025 Nov 5;17(11):e96121. doi:10.7759/cureus.96121. PMCID: PMC12587209.

Li Z, Wang L, Zhou H, Jiang J, Ye J, Lou L, et al. Artificial intelligence in ophthalmology: The path to the real-world clinic. Cell Rep Med. 2023 Jun 28;4(7):101095. doi:10.1016/j.xcrm.2023.101095. PMCID: PMC10394169.

Lin LY, Zhou P, Shi M, Lu JE, Jeon S, Kim D, et al. A deep learning model for screening computed tomography imaging for thyroid eye disease and compressive optic neuropathy. Ophthalmol Sci. 2024;4:100412. doi:10.1016/j.xops.2023.100412.

Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, on behalf of CONSORT-AI Group. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. Lancet Digit Health. 2020 Sep 9;2(10):e537–e548. doi:10.1016/S2589-7500(20)30218-1. PMCID: PMC8183333.

Lou L, Cao J, Wang Y, Gao Z, Jin K, Xu Z, et al. Deep learning-based image analysis for automated measurement of eyelid morphology before and after blepharoptosis surgery. Ann Med. 2021;53(1):2278-2285. doi:10.1080/07853890.2021.2009127.

Luo Y, Zhang J, Yang Y, Rao Y, Chen X, Shi T, et al. Deep learning-based fully automated differential diagnosis of eyelid basal cell and sebaceous carcinoma using whole slide images. Quant Imaging Med Surg. 2022;12(8):4166-4175. doi:10.21037/qims-22-98.

Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiology: AI. 2020. PMCID: PMC8017414.

Nakayama LF, et al. Unmasking biases and navigating pitfalls in the ophthalmic artificial intelligence lifecycle: A narrative review. 2024. PMCID: PMC11460710.

Ricci Lara MA, Echeveste R, Ferrante E. Addressing fairness in artificial intelligence for medical imaging. 2022. PMCID: PMC9357063.

Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ, on behalf of SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Lancet Digit Health. 2020 Sep 9;2(10):e549–e560. doi:10.1016/S2589-7500(20)30219-3. PMCID: PMC8212701.

Song X, Tong W, Lei C, Huang J, Fan X, Zhai G, et al. A clinical decision model based on machine learning for ptosis. BMC Ophthalmol. 2021;21:169. doi:10.1186/s12886-021-01923-5.

Srivastava O, Tennant M, Grewal P, Rubin U, Seamone M. Artificial intelligence and machine learning in ophthalmology: A review. Indian J Ophthalmol. 2022 Dec 30;71(1):11–17. doi:10.4103/ijo.IJO_1569_22. PMCID: PMC10155540.

Sun Y, Huang X, Zhang Q, Lee SY, Wang Y, Jin K, et al. A fully automatic postoperative appearance prediction system for blepharoptosis surgery with image-based deep learning. Ophthalmol Sci. 2022;2:100169. doi:10.1016/j.xops.2022.100169.

Tan W, et al. Fairer AI in ophthalmology via implicit fairness learning for mitigating sexism and ageism. 2024. PMCID: PMC11150422.

Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. Radiology. 2022. PMCID: PMC9152694.