

# ARTIFICIAL INTELLIGENCE IN MULTIMODAL OPHTHALMIC IMAGING FOR THE EARLY DETECTION OF UVEAL MELANOMA: CURRENT EVIDENCE AND FUTURE DIRECTIONS

**Dominik Poszwa**<sup>1</sup>  , **Zuzanna Dynowska**<sup>1</sup> ,  
**Mikołaj Bluszcz**<sup>2</sup> , **Zuzanna Muszkiet**<sup>3</sup> ,  
**Aleksandra Rysak**<sup>2</sup> , **Adam Janota**<sup>3</sup> ,  
**Patryk Roczniak**<sup>4</sup> , **Michał Bar**<sup>2</sup> ,  
**Zuzanna Galicka** 

<sup>1</sup> Heliodor Świącicki Clinical Hospital in Poznań, Poland

<sup>2</sup> Medical University of Silesia, Katowice, Poland

<sup>3</sup> NZOZ "Bredent", Bytom, Poland

<sup>4</sup> Municipal Hospitals Complex in Chorzów, Poland



[download article \(pdf\)](#)

 [dominikposzwa99@gmail.com](mailto:dominikposzwa99@gmail.com)

## ABSTRACT

### BACKGROUND

Uveal melanoma is the most common primary intraocular malignancy in adults and remains an important cause of metastatic mortality despite advances in ocular imaging and local therapy. Early detection is clinically important because it enables timely referral to ocular oncology services and supports eye preserving treatment strategies. However, small melanomas may resemble benign choroidal nevi, and their differentiation often requires interpretation of multimodal ophthalmic imaging. Artificial intelligence, particularly deep learning, has recently expanded in ophthalmic image analysis and offers potential tools for automated detection, classification, and risk assessment of melanocytic lesions.

### AIMS

To synthesize current scientific evidence on the use of artificial intelligence for analysis of multimodal ophthalmic imaging in the early detection of uveal melanoma and its differentiation from choroidal nevi, and to identify major methodological and clinical factors influencing translation into clinical practice.

### METHODS

This narrative review was conducted using a structured literature search of PubMed MEDLINE and cross referenced sources to identify relevant publications from 2017 to 2026. Search terms included uveal melanoma, choroidal nevus, artificial intelligence, deep learning, machine learning, and ophthalmic imaging modalities such as fundus photography, ultra widefield imaging, optical coherence tomography, OCT angiography, ultrasonography, and magnetic resonance

imaging. The evidence was analyzed using thematic narrative synthesis, and reporting quality and potential bias were interpreted in the context of contemporary methodological guidance for artificial intelligence studies.

## RESULTS

Available studies show that deep learning algorithms can differentiate choroidal nevi from small uveal melanomas using fundus photographs and that multimodal models integrating ultra widefield fundus imaging with ocular ultrasonography may improve classification performance compared with single modality approaches. Artificial intelligence models have also been applied to the detection of risk features of choroidal nevi and to radiomics analysis of magnetic resonance imaging for prognostic assessment. However, the current evidence base remains limited by relatively small datasets, heterogeneous diagnostic reference standards, class imbalance, and limited external validation.

## CONCLUSIONS

Artificial intelligence applied to multimodal ophthalmic imaging demonstrates potential as a decision support and triage tool for earlier and more consistent detection of uveal melanoma and high risk melanocytic lesions. Further research should focus on multicenter datasets, standardized diagnostic labeling, rigorous external validation, and prospective clinical evaluation aligned with contemporary reporting and risk of bias frameworks.

Keywords: Uveal melanoma; choroidal nevus; artificial intelligence; deep learning; multimodal imaging; fundus photography; ultrasonography; optical coherence tomography; magnetic resonance imaging; radiomics.

## INTRODUCTION

Uveal melanoma (UM) is the most common primary intraocular malignancy in adults and arises from melanocytes of the uveal tract (choroid, ciliary body, iris) [1,2]. Although many patients can be treated successfully with eye-preserving local therapies, approximately half develop metastatic disease, most frequently to the liver, with historically poor survival once metastasis occurs [1,2]. The biology of UM differs substantially from cutaneous melanoma, including characteristic driver mutations (e.g., GNAQ/GNA11) and distinct molecular subtypes, which influence prognosis and metastatic risk [3–6]. Recent systemic therapy advances, including tebentafusp for HLA-A\*02:01-positive metastatic UM, provide the first proven overall survival benefit in a phase 3 setting, but do not eliminate the high unmet need across the disease continuum [7–9].

From a clinical standpoint, “early detection” in UM has two pragmatic meanings. First, it encompasses earlier identification and referral of malignant lesions when they are small, potentially enabling globe-sparing treatment, reducing complications, and shortening time to definitive management. Second, it includes earlier discrimination of small melanomas from benign choroidal nevi, which are common and often incidentally detected. Differentiating a small melanoma from a nevus is a classic diagnostic challenge requiring synthesis across multimodal imaging (color fundus photography, ultra-widefield imaging, autofluorescence, OCT, OCT angiography, and ultrasonography), supplemented by clinical features and longitudinal change [10–15].

In real-world practice, the diagnostic pathway is heterogeneous. In many systems, initial detection occurs in general ophthalmology or optometry, where access to comprehensive ocular oncology expertise and multimodal imaging may be limited. Referral decisions may therefore depend on subjective assessment, variable imaging quality, and inconsistent follow-up intervals. Diagnostic delay is clinically meaningful: delayed ocular oncology evaluation can lead to treatment at larger tumor size, with potential impacts on vision outcomes and metastatic risk [12,13]. These realities motivate interest in scalable decision-support and triage tools.

Artificial intelligence (AI) has rapidly matured in ophthalmology, driven by large image datasets, standardized imaging modalities, and a track record of clinically validated deep learning systems for retinal disease screening and referral [16–21]. Parallel advances in medical imaging AI, explainability methods, and deployment considerations support the feasibility of AI-based tools for ocular oncology, including UM and its mimickers [22–28]. Importantly, the promise of AI in UM hinges on multimodal imaging: many clinically discriminative features (e.g., subretinal fluid, orange pigment surrogates, thickness/reflectivity on ultrasound) are not fully captured by a single modality, and reference standard diagnosis often relies on multimodal consensus by ocular oncologists [10–15]. Thus, multimodal AI is an intuitively appropriate strategy.

## RELEVANCE

Uveal melanoma remains a clinically significant oncologic problem in ophthalmology because it combines relatively low incidence with a high risk of metastatic spread and disease related mortality [1,2]. The biological behavior of this tumor differs from that of cutaneous melanoma and is associated with characteristic molecular alterations, including mutations in GNAQ and GNA11 and distinct molecular subtypes that influence metastatic potential and clinical prognosis [3–6]. Although local treatments allow preservation of the eye in many patients, survival outcomes remain

strongly dependent on the development of metastatic disease.

An important clinical challenge is the early identification of malignant lesions among the large number of benign choroidal nevi detected during routine ophthalmic examinations. Small melanomas may initially present with subtle clinical features and therefore require careful assessment using several complementary imaging techniques. Modern ophthalmic diagnostics relies on multimodal imaging, including color fundus photography, ultra widefield imaging, optical coherence tomography, OCT angiography, and ultrasonography, which together provide information about lesion morphology, retinal alterations, vascular changes, and tumor thickness [10–15].

In routine clinical practice, the first detection of suspicious choroidal lesions often occurs outside specialized ocular oncology centers, particularly in general ophthalmology or optometry settings. Limited access to multimodal imaging and subspecialty expertise may lead to variability in clinical interpretation and delayed referral for expert evaluation. In this context, the development of artificial intelligence methods for medical image analysis has attracted increasing attention. Recent advances in deep learning for ophthalmic imaging create opportunities for automated image interpretation systems that could assist clinicians in identifying suspicious lesions and supporting referral decisions in patients with suspected uveal melanoma [16–21].

## SCIENTIFIC NOVELTY

The scientific novelty of this article lies in the structured synthesis of contemporary research on the application of artificial intelligence for the analysis of multimodal ophthalmic imaging in the early detection of uveal melanoma and its differentiation from choroidal nevi. The article integrates evidence across several imaging modalities, including color fundus photography, ultra widefield imaging, optical coherence tomography, OCT angiography, ultrasonography, and magnetic resonance imaging, which represent complementary sources of diagnostic information in clinical ophthalmology [10–15,39–41]. The review analyzes existing deep learning models for the classification of choroidal lesions, including algorithms based on fundus photographs, multimodal models combining ultra widefield imaging and ultrasound, and models designed to identify risk factors for malignant transformation of choroidal nevi [35–38]. The article also discusses emerging research directions, including the use of magnetic resonance imaging based radiomics for prognostic modeling in uveal melanoma [39]. In this way, the review provides an integrated overview of current research on artificial intelligence in ocular oncology and highlights important methodological limitations of existing studies.

## AIM

The aim of this review article is to summarize and analyze current scientific evidence on the application of artificial intelligence methods for the analysis of multimodal ophthalmic imaging for the early detection of uveal melanoma and its differentiation from benign choroidal nevi [35–38].

### Research objectives

1. To analyze existing ophthalmic imaging modalities and clinical diagnostic features used in the development of artificial intelligence algorithms for the detection of uveal melanoma [10–15].
2. To evaluate the capabilities of deep learning models in distinguishing choroidal nevi from early uveal melanoma based on fundus photographs [35,36].
3. To examine multimodal artificial intelligence models integrating several imaging modalities, including ultra widefield imaging and ultrasonography [37].
4. To analyze approaches aimed at identifying risk factors for malignant transformation of choroidal nevi using artificial intelligence algorithms [38].
5. To review the use of radiomics and machine learning methods based on magnetic resonance imaging for prognostic assessment of the disease [39].
6. To analyze the main methodological limitations of existing studies, including limited datasets, variability in diagnostic reference standards, and insufficient external validation of models [31,32].
7. To identify future research directions and requirements for methodological transparency and reporting quality in artificial intelligence research in medicine [31,32,34].

## METHODS

### REVIEW DESIGN AND REPORTING FRAMEWORK

This work is a narrative review with a structured literature search. The search and selection of studies followed established principles of transparent evidence identification and were informed by PRISMA 2020 recommendations

where applicable [29]. Because the included studies show substantial heterogeneity in AI model design, datasets, imaging modalities, and reported outcomes, quantitative meta analysis was not appropriate. The evidence is therefore synthesized thematically with emphasis on clinical relevance and methodological interpretability.

Eligibility criteria

Inclusion criteria

1. Peer reviewed original research or high quality review articles published between 2017 and 2026 addressing artificial intelligence, machine learning, or deep learning applied to ophthalmic imaging in uveal melanoma, choroidal melanoma, or differentiation of choroidal nevi from melanoma.
2. Studies using one or more ophthalmic imaging modalities with explicit description of the model input modality, including fundus photography, ultra widefield imaging, OCT or OCT angiography, ultrasonography, or MRI.
3. Studies reporting outcomes related to detection or classification, identification of risk factors for malignant transformation, segmentation tasks, prognostic modeling, or clinically relevant triage endpoints.
4. Studies providing a clear reference standard, such as diagnosis established by ocular oncology specialists based on multimodal imaging and or longitudinal follow up, histopathology when available, or clinical consensus.
5. Exclusion criteria
6. Artificial intelligence studies focused on non ophthalmic melanoma or not specific to uveal melanoma or choroidal melanocytic lesions.
7. Case reports without development or evaluation of artificial intelligence models, unless used to contextualize the role of a specific imaging modality.
8. Conference abstracts lacking full methodological description or complete results.

## INFORMATION SOURCES AND SEARCH STRATEGY

We searched PubMed/MEDLINE for relevant literature. In addition, the reference lists of relevant articles and authoritative reviews were screened to identify additional eligible studies. Search terms combined descriptors related to uveal or choroidal melanoma with artificial intelligence terms and imaging modalities. Representative search strings included: ("uveal melanoma" OR "choroidal melanoma" OR "choroidal nevus") AND ("artificial intelligence" OR "deep learning" OR "machine learning") AND (fundus OR "ultra-widefield" OR OCT OR "optical coherence tomography" OR ultrasonography OR ultrasound OR MRI OR radiomics OR multimodal).

The search aimed to identify representative studies illustrating the main approaches to the application of artificial intelligence in ophthalmic imaging for uveal melanoma detection and risk assessment.

## STUDY SELECTION AND DATA EXTRACTION

Titles and abstracts were screened for relevance. Full texts were then reviewed to confirm eligibility. Extracted items included study design, population size, imaging modalities, AI architecture (when reported), reference standard definition, validation strategy (internal vs external), performance metrics (AUC, sensitivity, specificity), and key limitations.

## RISK OF BIAS, APPLICABILITY, AND REPORTING QUALITY CONSIDERATIONS

Given the nature of AI prediction models, we structured critical appraisal around contemporary reporting and bias guidance for AI and prediction modeling. We used TRIPOD+AI for reporting expectations in prediction model studies [31], and PROBAST+AI principles to frame risk-of-bias and applicability concerns [32]. For narrative synthesis quality, we considered SANRA domains as an organizing tool for completeness and transparency [30,33]. For clinical evaluation studies involving AI interventions, we referenced CONSORT-AI requirements conceptually [34]. These tools were used qualitatively to guide critique rather than as a formal scoring exercise, because many included studies were development-stage rather than prospective clinical trials.

# RESULTS

## EVIDENCE LANDSCAPE AND STUDY ARCHETYPES

The literature on AI for UM detection and classification has expanded notably in the past 2–3 years, shifting from proof-of-concept single-modality models to more clinically aligned multimodal approaches. The evidence can be grouped into five archetypes:

- A) Fundus photography deep learning for nevus vs melanoma classification (single modality) [35,36].
- B) Multimodal deep learning integrating ultra-widefield fundus imaging and ocular ultrasound [37].
- C) Deep learning for identifying melanoma risk factors or high-risk nevus characteristics from fundus images, potentially reducing dependence on multimodal equipment at first contact [38].
- D) MRI-based radiomics or machine learning models for prognostic endpoints, complementing ocular imaging with systemic risk stratification [39].
- E) Imaging modality reviews and clinical frameworks supporting multimodal interpretation and potential AI targets (e.g., OCT/OCTA in UM) [40].

We additionally considered foundational UM biology and clinical risk factor literature to contextualize AI targets and the clinical plausibility of imaging biomarkers [1–6,10–15].

### CLINICAL DETECTION PATHWAY AND WHY MULTIMODAL IMAGING MATTERS FOR AI

UM is often detected as an intraocular mass on clinical examination or incidental imaging. In practice, early lesions may be small, minimally elevated, or partially obscured by media opacities, and the diagnostic question frequently becomes whether a lesion is a benign nevus warranting observation or a melanoma requiring urgent referral and treatment.

Multimodal imaging is essential because different modalities capture distinct components of the lesion and its microenvironment:

- Fundus photography and ultra-widefield imaging: provide lesion color, borders, pigmentation patterns, and associated findings such as drusen or RPE changes; ultra-widefield expands peripheral visualization.
- Ultrasonography: provides thickness and internal reflectivity, critical for assessing elevation and tumor acoustic profile, and is a mainstay for melanoma characterization and treatment planning.
- OCT: detects subretinal fluid, retinal changes, and photoreceptor disruption; enhanced depth imaging may help characterize choroidal architecture in small lesions [14].
- Fundus autofluorescence: may serve as an imaging correlate of orange pigment (lipofuscin), a classic risk feature.
- OCT angiography: may reveal vascular alterations or treatment-related microvascular changes, though its role is evolving [40,41].
- MRI: can support characterization, staging considerations, and radiomics-based prognostic modeling [15,39].

Clinically, the diagnostic “reference standard” for many early lesions is not histopathology (which is rarely obtained in small tumors), but expert consensus supported by multimodal imaging and longitudinal behavior (growth, development of risk features) [10–15]. This has immediate implications for AI: models must be trained with reference standards that reflect real-world diagnostic ground truth, and validation must anticipate that “truth” is often a composite construct rather than a single test.

The key ophthalmic imaging modalities relevant to early uveal melanoma detection and the corresponding AI-relevant biomarkers and analytical tasks are summarized in Table 1.

Table 1. Multimodal imaging modalities and AI-relevant biomarkers for early uveal melanoma detection

Imaging modality	Key lesion features detectable	AI-relevant biomarkers	Potential AI tasks	Clinical relevance	References
Color fundus photography	Lesion pigmentation, borders, drusen, RPE alterations	Shape, color heterogeneity, border irregularity	Image classification (nevus vs melanoma), lesion detection	First-line imaging and screening	[10–13,35,36]
Ultra-widefield (UWF) imaging	Peripheral lesion visualization,	Peripheral morphology, size estimation	Classification, lesion localization	Improved visualization of	[37]

	lesion extent			peripheral tumors	
Fundus autofluorescence (FAF)	Lipofuscin accumulation (“orange pigment”)	Autofluorescence intensity patterns	Risk-feature detection	Indicator of malignant potential	[10–15]
Optical coherence tomography (OCT)	Subretinal fluid, retinal layer disruption	Fluid segmentation, photoreceptor alterations	Segmentation, biomarker detection	Detection of secondary retinal changes	[14,40]
Optical coherence tomography angiography (OCTA)	Microvascular changes, vascular density alterations	Flow signal patterns, vascular networks	Vascular feature analysis, longitudinal monitoring	Monitoring tumor-related vascular changes	[40,41]
Ocular ultrasonography	Tumor thickness, internal reflectivity	Thickness measurements, acoustic reflectivity	Classification, segmentation, growth monitoring	Core modality for melanoma confirmation	[10–13,37]
Magnetic resonance imaging (MRI)	Tumor extent, extraocular extension	Radiomic texture features	Prognostic modeling	Systemic staging and prognostic assessment	[15,39]

### FUNDUS PHOTOGRAPH DEEP LEARNING: DIFFERENTIATING NEVI FROM SMALL MELANOMA

Several recent studies demonstrate that deep learning can meaningfully differentiate choroidal nevi from small melanomas using fundus photographs alone.

A multicenter Ophthalmology Science study developed and validated a deep learning algorithm to differentiate choroidal nevi from small melanoma using wide field and standard field fundus photographs, with external clinic testing and comparison against ophthalmologist performance [38]. Key strengths include multicenter image sources and external testing, as well as clinically meaningful endpoints aligned with referral decisions. The study underscores that AI can achieve high sensitivity at moderate specificity, suggesting a potential role as a triage tool where false negatives must be minimized.

A Journal of Clinical Medicine study evaluated deep learning to distinguish highly malignant UM from benign choroidal nevi using fundus photographs, emphasizing human-machine interaction (HMI) and exploring a workflow where the model supports clinician decision-making rather than replacing it [35]. This is clinically relevant because early deployment in ocular oncology is likely to occur as decision-support, where clinicians retain responsibility and the system is designed for interpretability and error checking.

These studies collectively suggest that fundus-based AI can:

1. provide consistent pre-stratification of suspicious lesions, and
2. potentially reduce reliance on immediate multimodal imaging access at first contact, but they also highlight that fundus images alone may be insufficient for nuanced risk features that depend on lesion thickness, reflectivity, or subtle subretinal fluid. Therefore, fundus-only AI is best conceptualized as a triage layer rather than a definitive diagnostic device.

### MULTIMODAL DEEP LEARNING: INTEGRATING ULTRA-WIDEFIELD FUNDUS AND ULTRASOUND

A major advance in clinical realism is the emergence of multimodal AI models that integrate fundus photography with ultrasonography, reflecting how ocular oncologists evaluate lesions. An attention-based multimodal deep learning study in Ophthalmology Science integrated ultra-widefield fundus images and B-scan ultrasound to classify UM vs choroidal nevi [37]. The design aligns strongly with real-world diagnostics: ultrasound contributes thickness and reflectivity cues, while fundus imaging provides surface-level morphological context and lesion extent.

Key implications:

- Complementarity: single-modality performance is constrained by modality-specific blind spots; integration can reduce ambiguity where one modality is equivocal.
- Architecture choice: attention mechanisms can weight modality contributions dynamically, which may improve robustness when one modality is of lower quality (e.g., suboptimal ultrasound gain or fundus artifacts).
- Clinical relevance: multimodal models may better support referral and treatment planning decisions than fundus-only systems.

However, the clinical translation of multimodal AI raises additional implementation challenges:

- Data synchronization and standardization: matching ultrasound slices (transverse/longitudinal) to lesion location and fundus orientation requires protocol consistency.
- Device heterogeneity: ultrasound appearance varies with probe frequency, gain settings, and operator technique; fundus imaging varies across cameras and color profiles.
- Workflow integration: multimodal AI presupposes both modalities are available; this may be realistic in secondary/tertiary ophthalmology but less so in optometric screening settings.

Thus, multimodal AI may be best positioned initially at the level of ophthalmology clinics and ocular oncology referral networks, where both imaging streams exist and quality control can be enforced.

## AI TO IDENTIFY HIGH-RISK NEVUS FEATURES FROM FUNDUS IMAGES

A complementary approach is to predict clinically meaningful risk factors using limited input modalities, reducing dependence on multimodal equipment at early stages. An AJO International study developed a deep learning model to classify choroidal melanoma risk factors based on color fundus photographs, explicitly motivated by the limited availability of multimodal imaging in resource-limited settings [38]. This approach is clinically attractive because it targets actionable intermediate phenotypes (e.g., risk features associated with malignant transformation), rather than forcing a binary "nevus vs melanoma" decision on a single modality.

Conceptually, this aligns with how ocular oncologists reason:

- The decision to refer or treat is rarely based on a single feature; instead, it is driven by composite risk profiles (symptoms, fluid, pigment, thickness, growth).
- AI systems that output a risk-factor panel may be more interpretable and clinically acceptable than a single classification score, because clinicians can understand which elements are driving the risk assessment.

This approach also supports modular multimodal extension: a system could output fundus-derived risk features and later incorporate ultrasound or OCT-derived risk features when available, gradually improving certainty.

## OCT AND OCT ANGIOGRAPHY: EVOLVING ROLES AND AI-RELEVANT TARGETS

OCT-based characterization of small choroidal melanocytic lesions has been explored for years, including enhanced depth imaging OCT to visualize small melanomas and related choroidal structure [14]. More recently, narrative syntheses of OCT and OCTA applications in UM highlight potential diagnostic and monitoring roles, including detection of subretinal fluid, photoreceptor disruption, and treatment-related vascular changes [40]. A case-based OCTA report in plaque brachytherapy-related radiation retinopathy illustrates how OCTA captures longitudinal microvascular compromise after UM treatment, which may be relevant for AI monitoring and complication prediction rather than primary detection [41].

For AI, OCT/OCTA offers:

- high-dimensional volumetric data (OCT cubes) with subtle biomarkers;
- opportunities for segmentation (retinal layers, fluid) and quantification;
- potential for longitudinal modeling (trajectory-based risk and treatment response).

Yet the evidence for AI specifically targeting OCT/OCTA in UM remains less mature than fundus and ultrasound models. Key barriers include smaller labeled datasets, variability in scan protocols, and the need for robust annotation (fluid, RPE changes, choroidal boundaries). The near-term opportunity is likely in adjunctive tasks: fluid detection, tumor-related retinal change quantification, and prediction of radiation retinopathy severity rather than primary lesion

classification.

## MRI RADIOMICS AND MACHINE LEARNING FOR PROGNOSIS

**MRI-based radiomics extends AI beyond detection to prognostic modeling.** A Journal of Computer Assisted Tomography study developed an MRI-based radiomics model to predict disease-free survival in pretreatment UM, reporting an initial validation approach [39]. This is relevant because systemic risk stratification is increasingly important in UM care, especially as systemic treatments and surveillance strategies evolve.

Radiomics models face distinct challenges:

- Small cohorts and overfitting risk: radiomics feature spaces are large relative to sample size, increasing optimism bias unless rigorous validation is used.
- Harmonization: MRI protocols vary widely across centers; radiomics features can be sensitive to acquisition parameters.
- Clinical integration: prognostic outputs must map to clear decisions (surveillance intensity, adjuvant trial enrollment) and must be evaluated for net benefit.

Nevertheless, the direction is consistent with broader oncology trends: multimodal AI may ultimately integrate ocular imaging, systemic imaging, and molecular predictors into unified risk models.

## SUMMARY OF THE CLINICAL PERFORMANCE SIGNAL

Across the most clinically proximal evidence, there is a consistent “performance signal” suggesting that deep learning can:

- differentiate nevi from melanoma on fundus photography with clinically useful discrimination [35,36];
- improve classification by combining fundus and ultrasound in multimodal models [38];
- predict risk features or risk factor patterns from limited data, potentially enabling earlier triage [39].

However, across studies, common constraints limit certainty about generalizability:

- limited external validation across independent health systems;
- variability in reference standards (expert consensus vs follow-up vs histology);
- spectrum bias (datasets enriched for clear-cut cases rather than borderline lesions);
- inconsistent reporting of calibration, decision thresholds, and clinical utility measures.
- Representative studies illustrating the main artificial intelligence approaches for differentiation of uveal melanoma and choroidal nevi across different imaging modalities are summarized in Table 2.

Table 2. Summary of key AI studies for uveal melanoma and choroidal nevus differentiation (2019–2026)

Study	Imaging modality	Dataset / study population	AI approach	Reference standard	Validation	Key findings	References
Sabazade et al., 2024	Fundus photography	Multicenter dataset of choroidal nevi and small melanoma	Deep learning CNN	Expert diagnosis with multimodal imaging and follow-up	External validation	High discrimination between nevi and melanoma	[33]
Hoffmann et al., 2024	Fundus photography	Fundus images of benign and malignant lesions	Deep learning classifier with human-machine interaction	Clinical diagnosis by ocular oncology specialists	Internal validation	Accurate differentiation with clinically relevant sensitivity	[34]

Dadzie et al., 2025	Ultra-widefield imaging + ultrasound	Multimodal imaging dataset of choroidal lesions	Attention-based multimodal deep learning	Multimodal expert diagnosis	Independent test set	Multimodal models outperform single-modality models	[35]
Dalvin et al., 2025	Color fundus photography	Dataset of melanocytic lesions with known risk features	Deep learning risk-factor detection model	Clinical risk criteria for malignant transformation	Internal validation	Prediction of melanoma risk features	[36]
Su et al., 2023	MRI radiomics	Pretreatment MRI of UM patients	Radiomics-based machine learning model	Clinical outcomes and follow-up	Internal validation	Prediction of disease-free survival	[38]
Shields et al., 2019	Fundus + clinical imaging	Large cohort of choroidal nevi with longitudinal follow-up	Quantitative imaging analysis	Longitudinal malignant transformation	Observational dataset	Identification of imaging risk features	[37]

## DISCUSSION

### INTERPRETING THE EVIDENCE: WHAT AI CAN REALISTICALLY DO TODAY

The most plausible near-term clinical role for AI in early UM detection is triage and decision support. This is because:

1. The reference standard is often expert multimodal consensus and follow-up, not pathology. AI should therefore support earlier referral rather than claim definitive diagnosis.
2. Sensitivity is paramount: missing a melanoma has high clinical cost, while over-referral can be managed with specialist evaluation.
3. Many settings lack ocular oncology expertise; AI could standardize assessment and reduce diagnostic delay.

Fundus-only models can function as an “early warning system,” flagging lesions that merit prompt multimodal imaging or specialist referral [35,36]. Multimodal models integrating ultra-widefield fundus and ultrasound can support more specialized clinics where both modalities are available, potentially improving accuracy and consistency in borderline cases [39]. Risk-factor detection models can provide interpretable outputs aligned with clinician reasoning and could be used to structure referrals and follow-up schedules [38].

### REFERENCE STANDARD COMPLEXITY AND LABEL NOISE

A central challenge in ocular oncology AI is defining the ground truth. For many small lesions, histopathology is unavailable; instead, “truth” is inferred from:

- expert diagnosis supported by multimodal imaging,
- longitudinal stability (nevus) or growth/risk evolution (melanoma),
- response to treatment and subsequent behavior.

This creates label noise and temporal leakage risks. For example, if a lesion is labeled as “nevus” because it did not progress over 5 years, the model may inadvertently learn site-specific follow-up patterns or imaging artifacts associated with surveillance rather than biology. Conversely, lesions treated early may not have prolonged follow-up, and their labels may rely on pre-treatment expert judgment.

Practical mitigation strategies include:

- explicit label definitions and time windows;
- sensitivity analyses using alternative label criteria;

- separating development and evaluation cohorts by institution and time;
- incorporating longitudinal modeling explicitly rather than forcing static labels.

## **DATASET LIMITATIONS: SCARCITY, IMBALANCE, AND SPECTRUM BIAS**

UM is relatively rare compared with common retinal diseases. This scarcity drives several issues:

- small datasets amplify overfitting risk, especially with deep learning;
- class imbalance (many nevi, fewer melanomas) can inflate accuracy while hiding poor sensitivity;
- spectrum bias arises when datasets over-represent “classic” melanomas and exclude ambiguous borderline lesions, which are precisely the cases where clinical decision support is most needed.

Multimodal datasets are even harder to assemble, as they require paired imaging across modalities and consistent clinical labeling [37]. This highlights the importance of multicenter consortia, standardized protocols, and prospective dataset building in ocular oncology.

## **EXTERNAL VALIDATION, DOMAIN SHIFT, AND GENERALIZABILITY**

Generalizability is the major barrier between promising results and clinical deployment. Domain shift can be severe in ophthalmic imaging:

- fundus images vary by camera type, illumination, color balance, and field of view;
- ultra-widefield systems have characteristic peripheral distortions;
- ultrasound depends on operator skill, gain, frequency, and probe positioning;
- OCT/OCTA vary by manufacturer and segmentation algorithms;
- MRI varies by scanner, sequence parameters, and reconstruction.

Medical imaging AI literature emphasizes that models trained on one domain may degrade on unseen domains unless robust validation and domain generalization strategies are used [28]. In ocular oncology, this argues for:

- multi-institution training data;
- explicit external validation across institutions and devices;
- reporting of performance stratified by device type;
- domain generalization and harmonization methods evaluated prospectively.

## **EXPLAINABILITY AND CLINICAL ACCEPTANCE**

For high-stakes oncology decisions, clinicians require interpretability. While saliency methods such as Grad-CAM provide visual heatmaps that can be used to assess whether the model attends to plausible lesion regions, these methods can be misleading if not carefully validated [25]. Complementary explainability frameworks (e.g., SHAP for structured models) can support interpretation when the model includes engineered features or multimodal fusion components [26].

In practice, explainability should be used as:

- a development tool to detect shortcut learning (e.g., the model relies on image borders, labels, or camera artifacts);
- a clinical communication tool to build trust, while acknowledging limitations;
- part of a structured failure analysis strategy.

A key future direction is “clinically interpretable outputs” rather than purely post-hoc explanations. Risk-factor detection models exemplify this approach by outputting clinically recognized features instead of opaque scores [38].

## **REPORTING QUALITY AND RISK-OF-BIAS CONSIDERATIONS**

Many AI studies historically suffer from incomplete reporting, unclear validation, and optimistic performance estimates. Modern reporting guidance directly addresses these gaps. TRIPOD+AI provides updated guidance for reporting prediction models using regression or machine learning, emphasizing transparency in data handling, model development, validation, and intended use [31]. PROBAST+AI provides a structured approach to assess risk of bias and

applicability, including domains related to participants, predictors, outcomes, and analysis [32]. CONSORT-AI highlights what is required when AI interventions are evaluated in trials, including the need to describe human–AI interaction and handling of inputs/outputs [34]. These frameworks are increasingly essential if ocular oncology AI is to transition from retrospective development to prospective clinical validation.

Specific recurrent risk-of-bias patterns in UM imaging AI include:

- non-representative case selection (enriched cohorts, referral bias);
- unclear reference standards and inconsistent follow-up for “nevus” labels;
- inadequate separation of training and testing at the patient and site level;
- absence of external validation;
- lack of calibration reporting and decision-threshold justification.

The field should move toward pre-registered analysis plans, standardized outcome definitions, and prospective evaluation, even if initial deployment is limited to triage settings.

## **CLINICAL UTILITY: BEYOND AUC**

High AUC does not guarantee clinical benefit. For triage, the most relevant metrics are:

- sensitivity/negative predictive value at a clinically chosen threshold,
- false positive rates and referral burden,
- performance in borderline and low-quality imaging cases,
- calibration and risk stratification accuracy,
- decision-curve analysis and net benefit (where appropriate).

Given the potential harms of delayed melanoma diagnosis, AI tools should likely prioritize sensitivity. However, high false positives may overwhelm specialist services and erode trust. Therefore, workflow design matters: AI may be most useful when it triggers a structured next step (repeat imaging, ultrasound acquisition, expedited referral) rather than a single binary “malignant” label.

## **INTEGRATION WITH EVOLVING SYSTEMIC THERAPY AND SURVEILLANCE**

As systemic therapies improve (e.g., tebentafusp demonstrating survival benefit in metastatic UM) [9], the value of early detection and accurate risk stratification may increase: earlier diagnosis enables timely local treatment and may support earlier systemic surveillance and trial enrollment for high-risk molecular or imaging phenotypes. Multimodal AI could eventually integrate ocular imaging (lesion phenotype), systemic imaging (e.g., liver surveillance), and molecular prognostic markers into a unified risk framework, consistent with broader oncology directions.

## **PRACTICAL ROADMAP FOR TRANSLATION**

A feasible translation roadmap includes:

Phase 1: Retrospective multicenter validation

- Expand training datasets across devices/institutions;
- define consistent reference standards;
- validate on external cohorts with pre-specified thresholds.

Phase 2: Prospective silent-mode evaluation

- Deploy model in clinics without affecting decisions;
- measure performance on real-world cases, including borderline lesions;
- perform failure analysis and refine.

Phase 3: Clinical impact evaluation

- Randomized or stepped-wedge trials of AI-assisted triage pathways, with endpoints such as time-to-referral, time-to-treatment, diagnostic accuracy, and patient-centered outcomes, reported using CONSORT-AI principles

[34].

#### Phase 4: Post-deployment monitoring

- continuous performance auditing for domain drift;
- governance for model updates;
- integration with regulatory and ethical requirements.

## CONCLUSION

AI for early uveal melanoma detection is evolving toward clinically relevant multimodal approaches. Based on the analyzed literature, the following conclusions can be drawn.

1. Multimodal ophthalmic imaging remains essential for accurate characterization of uveal melanocytic lesions. Fundus and ultra widefield imaging provide morphological features, optical coherence tomography detects subretinal fluid and retinal changes, ultrasonography assesses tumor thickness and internal reflectivity, while fundus autofluorescence and OCT angiography provide complementary information on lipofuscin related signals and vascular alterations.
2. Deep learning based on fundus photography differentiates choroidal nevi from small uveal melanomas with clinically meaningful discrimination, but its role is limited to triage and referral support, particularly where access to ocular oncology expertise or multimodal imaging is restricted.
3. Multimodal artificial intelligence models, especially combining fundus imaging with ultrasonography, better reflect clinical workflows and improve classification performance, while systems detecting clinically relevant risk features provide more interpretable outputs for risk stratification and referral decisions.
4. MRI based radiomics and machine learning extend artificial intelligence beyond detection and contribute to prognostic assessment in uveal melanoma.
5. The evidence base remains limited by small datasets, heterogeneous reference standards, incomplete external validation, and risks of spectrum bias and domain shift.
6. Further progress requires multicenter datasets, standardized diagnostic labeling with longitudinal follow up, rigorous external validation, and prospective clinical evaluation, with transparent reporting and systematic bias assessment according to contemporary AI frameworks.

Under these conditions, artificial intelligence supported multimodal imaging may improve referral accuracy, reduce diagnostic delay, and support early detection and risk stratification in uveal melanoma.

## DISCLOSURES

### AUTHORS' CONTRIBUTIONS

#### USE OF AI

The authors acknowledge the use of ChatGPT (OpenAI) for assistance in drafting and language editing of early manuscript versions. All sections were subsequently reviewed, revised and approved by the authors, who take full responsibility for the scientific content and conclusions.

#### CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## REFERENCES

1. Jager MJ, Shields CL, Cebulla CM, Abdel-Rahman MH, Grossniklaus HE, Stern MH, et al. Uveal melanoma. Nat Rev Dis Primers. 2020;6(1):24. [doi:10.1038/s41572-020-0158-0](https://doi.org/10.1038/s41572-020-0158-0)
2. Kaliki S, Shields CL. Uveal melanoma: relatively rare but deadly cancer. Eye (Lond). 2017;31(2):241-257. <https://doi.org/10.1038/eye.2016.275>
3. Robertson AG, Shih J, Yau C, Gibb EA, Oba J, Mungall KL, et al. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. Cancer Cell. 2017;32(2):204-220.e15. <https://doi.org/10.1016/j.ccell.2017.07.003>

4. Van Raamsdonk CD, Griewank KG, Crosby MB, Garrido MC, Vemula S, Wiesner T, et al. Mutations in GNA11 in uveal melanoma. *N Engl J Med*. 2010;363(23):2191-2199. <https://doi.org/10.1056/NEJMoa1000584>
5. Van Raamsdonk CD, Bezrookove V, Green G, Bauer J, Gaugler L, O'Brien JM, et al. Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature*. 2009;457(7229):599-602. <https://doi.org/10.1038/nature07586>
6. Harbour JW, Onken MD, Roberson ED, Duan S, Cao L, Worley LA, et al. Frequent mutation of BAP1 in metastasizing uveal melanomas. *Science*. 2010;330(6009):1410-1413. <https://doi.org/10.1126/science.1194472>
7. Rodriguez-Vidal C, Fernandez-Diaz D, Fernandez-Marta B, Lago-Baameiro N, Pardo M, Silva P, et al. Treatment of metastatic uveal melanoma: systematic review. *Cancers (Basel)*. 2020;12(9):2557. <https://doi.org/10.3390/cancers12092557>
8. Mallone F, Sacchetti M, Lambiase A, Moramarco A. Molecular insights and emerging strategies for treatment of metastatic uveal melanoma. *Cancers (Basel)*. 2020;12(10):2761. <https://doi.org/10.3390/cancers12102761>
9. Nathan P, Hassel JC, Rutkowski P, Baurain JF, Butler MO, Schlaak M, et al. Overall survival benefit with tebentafusp in metastatic uveal melanoma. *N Engl J Med*. 2021;385(13):1196-1206. <https://doi.org/10.1056/NEJMoa2103485>
10. Shields CL, Cater J, Shields JA, Singh AD, Santos MC, Carvalho C. Combination of clinical factors predictive of growth of small choroidal melanocytic tumors. *Arch Ophthalmol*. 2000;118(3):360-364. <https://doi.org/10.1001/archophth.118.3.360>
11. Shields CL, Furuta M, Thangappan A, Nagori S, Mashayekhi A, Lally DR, et al. Metastasis of uveal melanoma millimeter-by-millimeter in 8033 consecutive eyes. *Arch Ophthalmol*. 2009;127(8):989-998. <https://doi.org/10.1001/archophthalmol.2009.208>
12. Damato EM, Damato BE. Detection and time to treatment of uveal melanoma in the United Kingdom: an evaluation of 2384 patients. *Ophthalmology*. 2012;119(8):1582-1589. <https://doi.org/10.1016/j.ophtha.2012.01.048>
13. Singh AD, Kalyani P, Topham A. Estimating the risk of malignant transformation of a choroidal nevus. *Ophthalmology*. 2005;112(10):1784-1789. <https://doi.org/10.1016/j.ophtha.2005.06.011>
14. Shields CL, Materin MA, Mehrotra N, Mashayekhi A, Shields JA. Optical coherence tomography of choroidal melanoma in 120 consecutive cases: enhanced depth imaging and comparison with choroidal nevus. *Arch Ophthalmol*. 2012;130(4):448-455. <https://doi.org/10.1001/archophthalmol.2012.1135>
15. Yonekawa Y, Kim IK. Epidemiology and management of uveal melanoma. In: Gombos DS, ed. *Ocular Oncology*. <https://doi.org/10.1016/j.hoc.2012.08.004>
16. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410. <https://doi.org/10.1001/jama.2016.17216>
17. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342-1350. <https://doi.org/10.1038/s41591-018-0107-6>
18. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39. <https://doi.org/10.1038/s41746-018-0040-6>
19. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167-175. <https://doi.org/10.1136/bjophthalmol-2018-313173>
20. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017;37(2):505-515. <https://doi.org/10.1148/rg.2017160130>
21. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. <https://doi.org/10.1016/j.media.2017.07.005>
22. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys*. 2019;29(2):102-127. <https://doi.org/10.1016/j.zemedi.2018.11.002>
23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis (ICCV)*. 2017:618-626. <https://doi.org/10.1109/ICCV.2017.74>
24. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67. <https://doi.org/10.1038/s42256-019-0138-9>

25. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56. <https://doi.org/10.1038/s41591-018-0300-7>
26. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25(1):30-36. <https://doi.org/10.1038/s41591-018-0307-0>
27. Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE Trans Med Imaging.* 2020;39(7):2531-2540. <https://doi.org/10.1109/TMI.2020.2973595>
28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *J Clin Epidemiol.* 2021;134:178-189. <https://doi.org/10.1016/j.jclinepi.2021.03.001>
29. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. *Res Integr Peer Rev.* 2019;4:5. <https://doi.org/10.1186/s41073-019-0064-8>
30. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26(9):1364-1374. <https://doi.org/10.1038/s41591-020-1034-x>
31. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>
32. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025;388:e082505. <https://doi.org/10.1136/bmj-2024-082505>
33. Sabazade S, Lumia Michalski MA, Bartoszek J, Fili M, Holmström M, Ståhlhammar G. Development and validation of a deep learning algorithm for differentiation of choroidal nevi from small melanoma in fundus photographs. *Ophthalmol Sci.* 2024;5(1):100613. <https://doi.org/10.1016/j.xops.2024.100613>
34. Hoffmann L, Runkel CB, Künzel S, Kabiri P, Rübsam A, Bonaventura T, et al. Using deep learning to distinguish highly malignant uveal melanoma from benign choroidal nevi. *J Clin Med.* 2024;13(14):4141. <https://doi.org/10.3390/jcm13144141>
35. Dadzie AK, Iddir SP, Abtahi M, Ebrahimi B, Rahimi M, Ganesh S, et al. Attention-based multimodal deep learning for uveal melanoma classification using ultra-widfield fundus images and ocular ultrasound. *Ophthalmol Sci.* 2025;6(2):100985. <https://doi.org/10.1016/j.xops.2025.100985>
36. Dalvin LA, Shields CL, Ancona-Lezama D, Yu MD, Di Nicola M, Williams BK Jr, et al. Development of a deep learning model to classify choroidal melanoma risk factors based on color fundus photographs. *AJO Int.* 2025;2(4):100167. <https://doi.org/10.1016/j.ajoint.2025.100167>
37. Shields CL, Dalvin LA, Ancona-Lezama D, Yu MD, Di Nicola M, Williams BK Jr, et al. Choroidal nevus imaging features in 3,806 cases and risk factors for transformation into melanoma in 2,355 cases: The 2020 Taylor R. Smith and Victor T. Curtin Lecture. *Retina.* 2019;39(10):1840-1851. <https://doi.org/10.1097/IAE.0000000000002440>
38. Su Y, Xu X, Wang F, Zuo P, Chen Q, Wei W, et al. Prognostic value of the radiomics-based model in the disease-free survival of pretreatment uveal melanoma: an initial result. *J Comput Assist Tomogr.* 2023;47(1):151-159. <https://doi.org/10.1097/RCT.0000000000001384>
39. Troisi M, Vitiello L, Lixi F, Timofte Zorila MM, Abbinante G, Pellegrino A, et al. Clinical applications of optical coherence tomography and optical coherence tomography angiography in uveal melanoma: a narrative review. *Diagnostics (Basel).* 2025;15(19):2421. <https://doi.org/10.3390/diagnostics15192421>
40. Binkley EM, Boldt HC, Grumbach IM. Longitudinal optical coherence tomography angiography (OCT-A) in a patient with radiation retinopathy following plaque brachytherapy for uveal melanoma. *Am J Ophthalmol Case Rep.* 2022;26:101508. <https://doi.org/10.1016/j.ajoc.2022.101508>
41. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118. <https://doi.org/10.1038/nature21056>

[back](#)

