Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep Learning in the Diagnosis of Choroidal Melanoma: A Systematic Review and Meta-Analysis

Muhammad Islampanah
muhammad.iriran@gmail.com
Mashhad University of Medical Sciences
Mahdi Naseri Takaloo
Mashhad University of Medical Sciences
Elnaz Farhangi
Mashhad University of Medical Sciences
Eileen Farhangi
Zahedan University of Medical Sciences
Shayan Yousefi
Mashhad University of Medical Sciences
Fereshteh Naderi
Mashhad University of Medical Sciences
Zeynab Ghasemi
Mashhad University of Medical Sciences
Niloofar Nikpour
Mashhad University of Medical Sciences
Sarina Hassanpour Dargah
Kerman University of Medical Sciences
Ghazal Asadi
Mashhad University of Medical Sciences
Armin Doostparast
Mashhad University of Medical Sciences

Systematic Review

Keywords: deep learning, choroidal melanoma, uveal melanoma, diagnosis, systematic review

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10935535/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose: Uveal melanoma (UM), particularly its most common subtype, choroidal melanoma (CM), is the leading intraocular malignancy in adults. Early diagnosis is crucial for improved prognosis, but accurately identifying small or early lesions remains challenging due to limitations in imaging. Artificial intelligence, especially deep learning (DL), shows promise for enhancing diagnosis and risk assessment of choroidal lesions. This study summarized DL models' diagnostic performance, clinical usefulness, limitations, pooled estimates, and research needs for detecting CM.

Methods: A systematic review and meta-analysis were conducted following PRISMA guidelines. Databases searched included PubMed, Scopus, Web of Science, IEEE Xplore, and Embase up to August 3rd, 2026. Our analysis comprised research studies that used DL to diagnose CM among other conditions. Quality assessment was performed using the QUADAS-AI tool.

Results: Ten studies met the inclusion criteria for the qualitative synthesis. Six of the ten studies were included in the meta-analysis. The pooled sensitivity was 0.92 (95% CI, 0.89–0.94). Individual study estimates of sensitivity ranged from 0.80 to 0.96. The pooled specificity was 0.97 (95% CI, 0.79–1.00). Across the included studies, individual specificity estimates ranged from 0.76 to 1.00. The leave-one-out sensitivity analysis showed that the pooled sensitivity and specificity estimates were consistent, regardless of which studies were excluded.

Conclusion: Our study strongly supports the use of DL in diagnosing CM while highlighting areas for further research. These findings could guide future studies and clinical practices in diagnosing and managing CM.

Key message

- Early diagnosis of choroidal melanoma is crucial for improved prognosis.
- Deep learning showed promise for diagnosing choroidal melanoma.
- Pooled sensitivity was 0.92 and specificity was 0.97 for choroidal melanoma.
- Leave-one-out analysis showed consistent pooled diagnostic estimates.

1. Introduction

Uveal melanoma (UM) is the most common primary intraocular malignancy in adults. Choroidal melanoma (CM) is the most prevalent subtype, making up around 90% of the cases [1, 2]. Annually, UM affects 4–7 individuals per million in Canada and the United States [3]. Approximately 50% of the cases eventually lead to distant metastasis [4]. Studies demonstrate that once distant metastases have occurred, the prognosis is poor and effective treatment is less likely. Therefore, a timely diagnosis at an early stage is crucial and is associated with improved prognosis and longer survival, and a late diagnosis can cause blindness and mortality [5].

Despite advances in ophthalmic imaging, the diagnosis of CM remains challenging, especially for small or early-stage choroidal lesions [6]. This diagnostic challenge is notable because aggressive behavior has been identified even in small posterior UM [7]. This diagnostic uncertainty may cause inappropriate referral decisions, including both unnecessary referrals for benign lesions and potential under-recognition of malignant lesions [6].

The current diagnosis of CM relies on clinical examination and multimodal ophthalmic imaging, including fundus photography, optical coherence tomography (OCT), ultrasonography, and other complementary imaging modalities [5]. These imaging techniques are used to characterize choroidal nevi and to identify features associated with malignant transformation [5]. However, the interpretation of the imaging findings in small choroidal lesions remains challenging, and clinical features may not always provide sufficient discrimination between nevi and melanoma [6, 8]. Intraocular biopsy can provide histopathologic and molecular information in selected patients with uncertain diagnosis, but its invasive nature limits its routine use as a diagnostic approach [9]. These challenges have led to growing interest in artificial intelligence (AI)-based approaches for improving diagnosis and risk assessment of choroidal lesions [10].

AI has been increasingly explored in medicine, particularly in disease screening and diagnosis. Machine learning (ML) is a subset of AI that enables computer systems to learn patterns from datasets, while deep learning (DL) is a branch of ML that uses multilayer neural networks to process data and recognize disease-associated patterns. The strong reliance on image-based diagnosis makes ophthalmology a suitable field for the application of AI-based methods [11, 12]. Several original studies have explored the potential of DL for detecting CM. Hoffmann et al. evaluated a DL assistant for distinguishing benign from malignant choroidal lesions using fundus photographs and reported an accuracy of 84.8% for multi-class classification and 90.9% for binary classification [13]. Similarly, Sabazade et al. evaluated the MelAInoma algorithm for supporting referral decisions for small pigmented choroidal lesions, reporting a sensitivity of 80%, a specificity of 90%, and an accuracy of 88% [6]. These findings demonstrate the potential of DL for assessing choroidal melanocyte lesions.

However, the rapidly expanding body of evidence on DL-based detection of CM and the growing number of recent studies warrant an up-to-date, comprehensive evaluation of the available data. Therefore, this systematic review and meta-analysis aimed to summarize the diagnostic performance, clinical applicability, and limitations of DL models for detecting CM, provide pooled estimates of diagnostic performance, and identify areas requiring further research.

2. Materials and Methods

2.1. Search and selection

The systematic review was conducted in accordance with the PRISMA guidelines [14]. The study was registered in the “International Prospective Register of Systematic Reviews” (PROSPERO) in 2026

(CRD420261467940). A researcher (M.I.) conducted a thorough search across PubMed, Scopus, Web of Science, IEEE Xplore, and Embase up to August 3rd, 2026, using the terms "choroidal melanoma" AND "deep learning" and their synonyms. The detailed search strategies and all variants are described in Table S1. The initial strategy was created using PubMed's Medical Subject Headings (MeSH) and then applied consistently across all databases. There were no restrictions on publication date or language. Additionally, the reference lists of included articles, previous reviews, and citations from the most recent studies were manually checked to find any relevant additional research.

2.2. Inclusion and exclusion criteria

Studies were eligible if they: (1) applied at least one DL model to diagnose CM; (2) at least one group of participants had CM; and (3) participants were human, or the samples were taken from humans.

Studies were excluded if they: (1) evaluated solely conventional machine-learning methods without incorporating DL components; (2) participants did not have CM; (3) participants had CM but the analysis results were combined with those of participants with other conditions, lacking a subgroup analysis for CM patients; (4) were reviews, protocols, conference abstracts with insufficient information, or animal studies; or (5) had no available full text in English.

2.3. Eligibility and quality assessment

Two individuals, M.I. and M.N., initially reviewed the studies by assessing titles and abstracts. Subsequently, both M.I. and M.N. evaluated the full texts to determine their relevance. In cases of disagreement, a third reviewer, El.F., made the final inclusion decision. The quality of the selected studies was independently appraised by El.F. and Ei.F. utilizing the Quality Assessment of Diagnostic Accuracy Studies Using AI (QUADAS-AI) tool (Fig. 1) [15]. When disagreements occurred, a third reviewer (S.Y.) made the final decision regarding inclusion.

2.4. Data extraction

Two independent reviewers (S.Y., and G.A.) extracted the following data: title, first author, publication year, dataset/center used to obtain data, inclusion criteria, exclusion criteria, aim, tools used to diagnose lesions, CM stage, nevi stage, status of the other groups except the nevi, drusen, number of total participants, number of patients with CM, nevi, and other conditions, type of DL model, accuracy, specificity, sensitivity (recall), precision, F1 score, area under the curve (AUC), and whether the study used a publicly-available dataset. Each study was independently extracted by two reviewers. When disagreements occurred, a third reviewer (M.I) made the final decision regarding inclusion. Whenever available, the diagnostic criteria or assessment tools used to establish CM status, such as fundus photography, OCT, slit-lamp biomicroscopy, A/B-scan ultrasonography, or biopsy, were extracted. Studies were not excluded solely because of poorly described diagnostic tools; this was considered in the risk-of-bias assessment. The data were stored in a Microsoft Excel spreadsheet.

2.5. Data synthesis

A narrative synthesis was performed on data from the selected studies, using both text and tables to detail their features and findings. A meta-analysis employing a random-effects model was conducted to estimate the pooled sensitivity and specificity for studies that provided a confusion matrix. We also conducted a leave-one-out sensitivity analysis to identify potential outlier studies. If the confusion matrices were not supplied by the study, but there existed sufficient data to compute them (including sensitivity, specificity, and sample size), the Diagnostic Test Calculator [16, 17] was employed to create the confusion matrix. The Diagnostic Test Calculator utilizes the Wilson score and Haldane-Anscombe correction (for zero cells) to compute the counts of true positives, true negatives, false positives, and false negatives based on sensitivity, specificity, and sample size. A confusion matrix generator was used to generate the pooled confusion matrix [18]. All other statistical analyses were conducted using the metadta package in Stata 17 (Stata Corporation, College Station, TX, USA).

3. Results

3.1. Baseline characteristics

3.1.1. Deep learning models

The study selection process is presented in Fig. 2. One hundred and one studies were retrieved for full-text assessment, of which 86 (File S1) were excluded because they did not include patients with CM or did not provide a subgroup analysis for patients with CM. Three studies were excluded because they did not use DL, and two were removed because their aim was not to diagnose CM, although they did use DL and the patients had CM. Ten studies evaluating DL approaches for diagnosing CM were included (Tables 1 & S2) [19–28]. Six studies provide sufficient data for the meta-analysis of sensitivity and specificity [19, 23, 24, 26–28]. The models and architectures varied across studies. MelAInoma was evaluated in two studies [19, 20], while other approaches included the Multimodal Medical Concept Bottleneck Model (MMCBM) [21], YOLO11 [22], U-Net [23], RETFound [24, 27], multilayer perceptron (MLP) [26], convolutional neural networks (CNNs) [25], ResNeXt-based Multicon [26], and HyperTumorEyeNet based on ResNet50 [28].

3.1.2. Sample size, age, gender, and datasets

The studies used datasets from several clinical centers and publicly available resources. The reported sample sizes varied substantially, from 12 participants in the study using tear-derived protein corona analysis [25] to 750 participants [21]. The reported mean overall age ranged from 38.75 ± 7.24 years in the study by Rakhshandeh et al. [25] to 61 ± 15 years in the study by Sabazade et al. [19]. Three studies used a dataset from St. Erik Eye Hospital [19, 20, 23], two used data from Liverpool Ocular Oncology Centre [24, 27], one study did not report its dataset [25] and the others used the data from Choroid Tri-Modal Imaging (CTI) dataset [21], an ocular oncology clinic [22], Ocular Oncology Service of University of California San Francisco [26], and Ocular Oncology Department, Charité University Hospital Berlin [28]. Only Wu et al. used a publicly available dataset [21].

3.2. Performance metrics

3.2.1. Accuracy

Accuracy was reported for several models. In the Wu et al. study [21], the MMCBM achieved an accuracy of 93.6%, compared with 93.1% for the black-box. RETFound achieved 83% and 82% for binary and tertiary classification, respectively [24]. Another study that used RETFound, achieved an accuracy of 98% [27]. In the Rakhshandeh et al. study [25], mean accuracy was 97.6 ± 0.8% for VGG16, 94.8 ± 2.4% for ResNet50, 95.3 ± 1.8% for Xception, and 94.3 ± 0.4% for the DNN. The Multicon model achieved an accuracy of 97.23%, compared with 86.00% for the MLP [26]. For HyperTumorEyeNet, accuracy was 84.8% for trinary classification using Optos and Clarus images, 90.9% for binary classification using Optos and Clarus images, 86.5% for trinary classification using Clarus images, and 95.8% for binary classification using Clarus images [28].

3.2.2. Sensitivity and Specificity

The sensitivity ranged from 73% [24] to 100% [23, 26], while the specificity ranged from 74% [23] to 100% [26, 27]. Six studies evaluating the use of DL for the diagnosis of CM were included in the pooled analysis of diagnostic performance [19, 23, 24, 26–28] (Figs. 3 & 4). The pooled sensitivity was 0.92 (95% CI, 0.89–0.94). Individual study estimates of sensitivity ranged from 0.80 to 0.96. The pooled specificity was 0.97 (95% CI, 0.79–1.00). Across the included studies, individual specificity estimates ranged from 0.76 to 1.00. Figure 5 shows the pooled confusion matrix.

3.2.3. Precision and F1 Score

Precision and F1 score were reported for selected models. For the MMCBM, precision was 91.5%, and the corresponding value for the black-box was 90.5% [21]. For the U-Net model, the F1 score was 0.72 in the main test cohort and 0.71 in the external test cohort [23]. In the study by Rakhshandeh et al., the precision ranged from 92.2 ± 0.3% to 97.7 ± 0.6, while the F1 score ranged from 0.945 ± 0.003 to 0.976 ± 0.008 [25]. In the study by Hoffmann et al., the highest precision was 97%, and the highest F1 score was 0.96 [28].

3.2.4. Area Under the Curve (AUC)

AUC values were reported for several models. MelAInoma achieved an AUC of 0.80 in the Sabazade et al. study and 0.92 in the Stripling et al. study [19, 20]. The U-Net model achieved an AUC of 0.88 in both the main and external test cohorts [23]. In the Rakhshandeh et al. study, AUC values were 0.997 ± 0.002 for VGG16, 0.985 ± 0.012 for ResNet50, 0.992 ± 0.004 for Xception, and 0.988 ± 0.001 for the DNN [25]. Multicon model had an AUC of 1.00, compared with 0.86 for the MLP [26]. For HyperTumorEyeNet, AUC values ranged from 0.96 to 1.00 across the reported binary and trinary classification settings [28].

3.3. Leave-one-out sensitivity analysis

The leave-one-out sensitivity analysis demonstrated that the pooled sensitivity and specificity estimates remained consistent across all omitted studies (Table 2). When studies by Jackson et al. were removed, the sensitivity reached its lowest at 0.88 [24, 27], while after excluding other studies, it remained at 0.92. The specificity was 0.93 after we omitted the study by Jackson et al. [27], and 0.95 after the omission of Abdeljaber et al. [26]. The specificity remained at 0.98 after excluding other studies.

4. Discussion

4.1. Principal Findings

In our systematic review and diagnostic test accuracy meta-analysis, among 10 included studies evaluating DL approaches for CM diagnosis, six provided sufficient data for pooled diagnostic analysis. The DL models and architectures were heterogeneous, including MelAInoma, MMCBM, YOLO11, U-Net, RETFound, MLP, CNNs, Multicon, and HyperTumorEyeNet. Across the included studies, reported accuracy varied substantially, with several models achieving accuracies above 90%. The pooled sensitivity was 0.92 (95% CI, 0.89–0.94) and pooled specificity was 0.97 (95% CI, 0.79–1.00), with individual study estimates ranging from 0.80 to 0.96 for sensitivity and 0.76 to 1.00 for specificity. Reported AUCs ranged from 0.80 to 1.00, while precision and F1 scores were also high in several models. Leave-one-out analysis showed that the pooled estimates remained generally consistent, with sensitivity ranging from 0.88 to 0.92 and specificity from 0.93 to 0.98 after excluding individual studies.

The findings of our study indicate that DL models can differentiate CM from other conditions, such as nevi, with strong overall discriminatory performance. However, the variation in reporting made it difficult to formally evaluate algorithm performance, as only six out of ten studies provided sufficient data for a diagnostic test meta-analysis. Furthermore, differences in DL models, diagnostic tools for CM, sample sizes, and CM prevalence across studies make direct comparisons between studies and algorithms very difficult.

Despite these challenges, AI models hold significant potential for early melanoma detection, decreasing physician workload, and standardizing diagnostic accuracy. Nonetheless, it's crucial to highlight that AI should complement, not replace, clinical judgment. Physicians must actively participate in AI training, validation, and real-world deployment to build trust in its clinical use [29].

4.2. Impact of Imaging Modality and Diagnostic Task on Model Performance

Fundus photography is the dominant input modality in current DL [30]. This is clinically plausible because fundus images capture lesion color, margins, and surface features effectively, but fundus-only models cannot directly assess tumor thickness or three-dimensional characteristics, which are central to the conventional melanoma workup [30].

This limitation helps explain why multimodal strategies are increasingly emphasized. The MMCBM reported 94.5% accuracy, exceeding its black-box comparator, suggesting that combining imaging and clinical concepts can improve both performance and interpretability [21]. Likewise, an attention-based multimodal model integrating ultra-widefield fundus images and ocular ultrasound achieved 94% accuracy and AUC 0.9606, outperforming single-modality configurations by leveraging complementary surface and structural information [31].

Recent reviews and modality-specific studies support the use of multimodal imaging for CM because fundus photography characterizes lesion surface features and serial growth, OCT detects subretinal fluid and high-resolution retinal/choroidal structural changes, and ultrasonography assesses tumor thickness, shape, and internal reflectivity [32, 33]. A recent imaging-based meta-analysis also found that ultrasound-based models had higher pooled specificity than fundus-only models and a higher HSROC AUC [34].

The model performance depends partly on the clinical difficulty of the classification task. Sabazade et al. evaluated small melanomas versus nevi and found AUC 0.88 in both internal and external test cohorts, with sensitivity dropping from 100% in one internal test set to 80% in the external clinic cohort, despite maintaining external validation performance overall [23]. By contrast, multimodal can produce higher discrimination, but this may reflect easier separation between larger melanomas and smaller nevi rather than superior clinical subtlety [31].

4.3. Clinical Relevance

A major practical implication is that DL could improve triage and referral, particularly where ocular oncology expertise is scarce. The MelAInoma work showed that the algorithm outperformed ophthalmologists, opticians, and optometrists for melanoma detection and outperformed MOLES and TFSOM-UHHD [23, 35]. In optometric triage, adding MelAInoma raised overall accuracy to 88%, quadrupled the odds of correct melanoma referral, and reduced false-positive referrals tenfold, although this evidence came from a small image set containing only five melanomas [19].

Sequential multimodal imaging with color photography, autofluorescence, and OCT detected progression in nearly all monitored melanocytic lesions, even without ultrasonography [36], supporting image-based surveillance frameworks used by the DL models in our study. At the same time, photographic triage systems can generate many false positives when thresholds are decreased to avoid missed melanomas, so algorithm operating points must be selected according to local clinical priorities and referral capacity [19].

Another clinically important observation is that explainability is beginning to emerge, but remains limited. Interpretable or concept-based models and SHAP-assisted classifiers suggest that more transparent decision support is feasible, which may help clinician trust and adoption [35, 37]. However, most studies still evaluate performance more than explanation, and limited use of explainability has been identified as a persistent barrier to clinical [29].

4.4. Limitations and Future Directions

The main limitation of the present evidence base is generalizability. Most included studies were retrospective and single-center, and many datasets were small relative to the complexity of the models being trained. Even large image collections were sometimes derived from one institution, which improves internal consistency but does not guarantee robustness across devices, populations, or referral settings.

Several studies relied on specialist multimodal clinical diagnosis rather than histopathology, which reflects real-world ophthalmic practice but introduces the possibility that some indeterminate lesions were mislabeled [28]. A multicenter fundus study partly addressed this by requiring long follow-up for nevi and using comprehensive ocular oncology examination, including ultrasound, OCT, and biopsy when needed, but such rigor was not universal across the literature [23].

Methodological reporting also remained inconsistent. A large meta-analysis of diagnostic DL across medical imaging found substantial heterogeneity in methodology, terminology, and outcomes, warning that such variation can overestimate diagnostic performance and arguing for stronger AI-specific reporting standards [38]. That concern applies directly here, where studies reported differing thresholds, task formulations, and performance metrics, and not all provided full confusion-matrix data needed for robust evidence synthesis.

Future work should therefore prioritize prospective multicenter validation, standardized reporting, and models trained on diverse multimodal data. Particular attention should be given to small or indeterminate lesions, where the real unmet clinical need lies and where current models are most vulnerable to optimistic performance estimates.

Conclusion

Overall, the results of this systematic review and diagnostic test accuracy meta-analysis suggest that DL has substantial promise for diagnosing CM, particularly as a triage and decision-support tool for differentiating melanoma from nevus. Nevertheless, the standard clinical implementation will rely less on further improvements in accuracy and more on thorough external validation, multimodal integration, transparent model conduct, and evidence of benefit within real-world ocular oncology pathways.

Declarations

Ethics approval and consent to participate

N/A

Consent to publish

N/A

Acknowledgments

N/A

Funding

N/A

Clinical Trial

N/A

Competing interests

The authors declare no conflicts of interest.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author(s) used Grammarly (available from https://app.grammarly.com) in order to improve the grammatical accuracy and fluency of the text. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Authors' contributions

Muhammad Islampanah: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review and editing, Supervision, Project administration

Mahdi Naseri Takaloo: Data Curation, Writing – original draft

Elnaz Farhangi: Data Curation, Writing – original draft

Eileen Farhangi: Data Curation, Writing – original draft

Shayan Yousefi: Data Curation, Writing – original draft

Fereshteh Naderi: Writing - Review & Editing

Zeynab Ghasemi: Writing - Review & Editing

Niloofar Nikpour: Writing – original draft, Writing – review and editing

Sarina Hassanpour Dargah: Writing - Review & Editing

Ghazal Asadi: Data Curation, Writing – original draft

Armin Doostparast: Writing – review and editing

References

1. Jackson M, Kalirai H, Hussain RN*, et al.* Differentiating Choroidal Melanomas and Nevi Using a Self-Supervised Deep Learning Model Applied to Clinical Fundoscopy Images. Ophthalmol Sci 2025;5(2):100647
2. Cao X, Zeng J, Ou Y*, et al.* A prognostic nomogram for the cancer-specific survival rate of choroidal melanoma using the Surveillance, Epidemiology, and End Results database. Front Med (Lausanne) 2024;11:1392336
3. Neto RB, Burnier MNN, Costa MAN*, et al.* Vascular clefts in choroidal melanoma: Correlation with histopathological prognostic factors. The Pan-American Journal of Ophthalmology 2026;8
4. Păsărică MA, Curcă PF, Dragosloveanu CDM*, et al.* Pathological and Molecular Diagnosis of Uveal Melanoma. Diagnostics (Basel) 2024;14(9)
5. Shields CL, Lally SE, Dalvin LA*, et al.* White Paper on Ophthalmic Imaging for Choroidal Nevus Identification and Transformation into Melanoma. Transl Vis Sci Technol 2021;10(2):24
6. Sabazade S, Lumia Michalski MA, Wittskog S*, et al.* Challenges for Opticians in Evaluating Small Pigmented Choroidal Lesions: Potential Support From the MelAInoma Deep Learning Algorithm. Transl Vis Sci Technol 2025;14(6):29
7. Stålhammar G, Kalirai H, Coupland SE*, et al.* Metastatic potential of small posterior uveal melanomas. Acta Ophthalmol 2026;104(2):e191-e203
8. Stripling K, Coudé Adam H, Holmström M*, et al.* Comparison of MOLES and MelAInoma for Differentiating Small Choroidal Melanomas from Nevi. Cancers (Basel) 2026;18(5)
9. Branisteanu DC, Bogdanici CM, Branisteanu DE*, et al.* Uveal melanoma diagnosis and current treatment options (Review). Exp Ther Med 2021;22(6):1428
10. Karamanli K-E, Maliagkani E, Petrou P*, et al.* (2025) Artificial Intelligence in Decoding Ocular Enigmas: A Literature Review of Choroidal Nevus and Choroidal Melanoma Assessment. *Applied Sciences,* 10.3390/app15073565.
11. Grewal PS, Oloumi F, Rubin U*, et al.* Deep learning in ophthalmology: a review. Can J Ophthalmol 2018;53(4):309-313
12. Wang S, He X, Jian Z*, et al.* Advances and prospects of multi-modal ophthalmic artificial intelligence based on deep learning: a review. Eye Vis (Lond) 2024;11(1):38
13. Hoffmann L, Runkel CB, Künzel S*, et al.* Using Deep Learning to Distinguish Highly Malignant Uveal Melanoma from Benign Choroidal Nevi. J Clin Med 2024;13(14)
14. Page MJ, Moher D, Bossuyt PM*, et al.* PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. bmj 2021;372

15. Guni A, Sounderajah V, Whiting P*, et al.* Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies Using AI (QUADAS-AI): Protocol for a Qualitative Study. JMIR Res Protoc 2024;13:e58202
16. *Diagnostic Test Calculator*. https://conductscience.com/tools/diagnostic-test-calculator?srsltid=AfmBOoqOe1kMeXR5VZcxLPB3NgG0HHm3ZUsg6IfsARc22wM0Dby5FhQO.
17. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association 1927;22(158):209-212
18. Perri D, Simonetti M, Gervasi O. Synthetic Data Generation to Speed-Up the Object Recognition Pipeline. Electronics 2022;11(1):2
19. Sabazade S, Lumia Michalski MA, Wittskog S*, et al.* Challenges for Opticians in Evaluating Small Pigmented Choroidal Lesions: Potential Support From the MelAInoma Deep Learning Algorithm. Translational Vision Science & Technology 2025;14(6):29-29
20. Stripling K, Coudé Adam H, Holmström M*, et al.* Comparison of MOLES and MelAInoma for Differentiating Small Choroidal Melanomas from Nevi. Cancers 2026;18(5):818
21. Wu Y, Liu Y, Yang Y*, et al.* A concept-based interpretable model for the diagnosis of choroid neoplasias using multimodal data. Nature Communications 2025;16(1):3504
22. Qin H, Yuan S, Bayhaqi YA*, et al.* Detection and Border Estimation of Choroidal Melanoma in OCT Scans Using Deep Learning Method. Investigative Ophthalmology & Visual Science 2025;66(8):1733-1733
23. Sabazade S, Lumia Michalski MA, Bartoszek J*, et al.* Development and Validation of a Deep Learning Algorithm for Differentiation of Choroidal Nevi from Small Melanoma in Fundus Photographs. Ophthalmology Science 2025;5(1)
24. Jackson M, Kalirai H, Hussain RN*, et al.* Differentiating Choroidal Melanomas and Nevi Using a Self-Supervised Deep Learning Model Applied to Clinical Fundoscopy Images. Ophthalmology Science 2025;5(2)
25. Rakhshandeh H, Nasiraei A, Riazi-Esfahani H*, et al.* Non-invasive detection of choroidal melanoma via tear-derived protein corona on gold nanoparticles: a machine learning approach. Scientific Reports 2025;15(1):32829
26. Abdeljaber L, Khan L, Zamani M*, et al.* A Novel Application of a Semi-Supervised Learning Approach to Identify Choroidal Tumors in Fundus Photographs. In: *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)*. *2021*, p. 132-139.
27. Jackson M, Kalirai H, Hussain R*, et al.* Using Deep Learning to differentiate between primary choroidal melanomas and metastases to the choroid. Investigative Ophthalmology & Visual Science 2025;66(8):965-965
28. Hoffmann L, Runkel CB, Künzel S*, et al.* Using Deep Learning to Distinguish Highly Malignant Uveal Melanoma from Benign Choroidal Nevi. Journal of Clinical Medicine 2024;13(14):4141
29. Karamanli K-E, Maliagkani E, Petrou P*, et al.* Artificial Intelligence in Decoding Ocular Enigmas: A Literature Review of Choroidal Nevus and Choroidal Melanoma Assessment. Applied Sciences

2025;15(7):3565

30. Dadzie AK, Iddir SP, Ganesh S*, et al.* Artificial intelligence in the diagnosis of uveal melanoma: advances and applications. Experimental Biology and Medicine 2025;Volume 250 - 2025
31. Dadzie AK, Iddir SP, Abtahi M*, et al.* Attention-Based multimodal deep learning for uveal melanoma classification using ultra-widefield fundus images and ocular ultrasound. Ophthalmology Science 2025:100985
32. Troisi M, Vitiello L, Lixi F*, et al.* Clinical Applications of Optical Coherence Tomography and Optical Coherence Tomography Angiography in Uveal Melanoma: A Narrative Review. Diagnostics 2025;15(19):2421
33. Verbeek S, Dalvin LA. Advances in multimodal imaging for diagnosis of pigmented ocular fundus lesions. Canadian Journal of Ophthalmology 2024;59(4):218-233
34. Bravo-Gonzalez A, Dominguez-Ruiz P, Buitrago-Gonzalez MJ*, et al.* Imaging-based machine learning for the diagnosis and prognosis of uveal melanoma: a systematic review and meta analysis. Graefe's Archive for Clinical and Experimental Ophthalmology 2026:1-14
35. Kanavos T, Birbas E, Francis JH*, et al.* Artificial Intelligence–Driven Differentiation Between Uveal Melanoma and Nevus Based on Fundus Photographs: A Systematic Review and Meta-Analysis. Translational Vision Science & Technology 2026;15(1):34-34
36. Roelofs KA, O'Day R, Al Harby L*, et al.* Detecting Progression of Melanocytic Choroidal Tumors by Sequential Imaging: Is Ultrasonography Necessary? Cancers 2020;12(7):1856
37. Shakeri E, Mohammed E, Crump T*, et al.* Deep learning-based detection and classification of uveal melanoma using convolutional neural networks and shap analysis. In: *2023 IEEE 24th international conference on information reuse and integration for data science (IRI)*. *2023*, p. 215-220. IEEE.
38. Aggarwal R, Sounderajah V, Martin G*, et al.* Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ digital medicine 2021;4(1):65

Tables

Tables 1 and 2 are available in the Supplementary Files section.

Figures

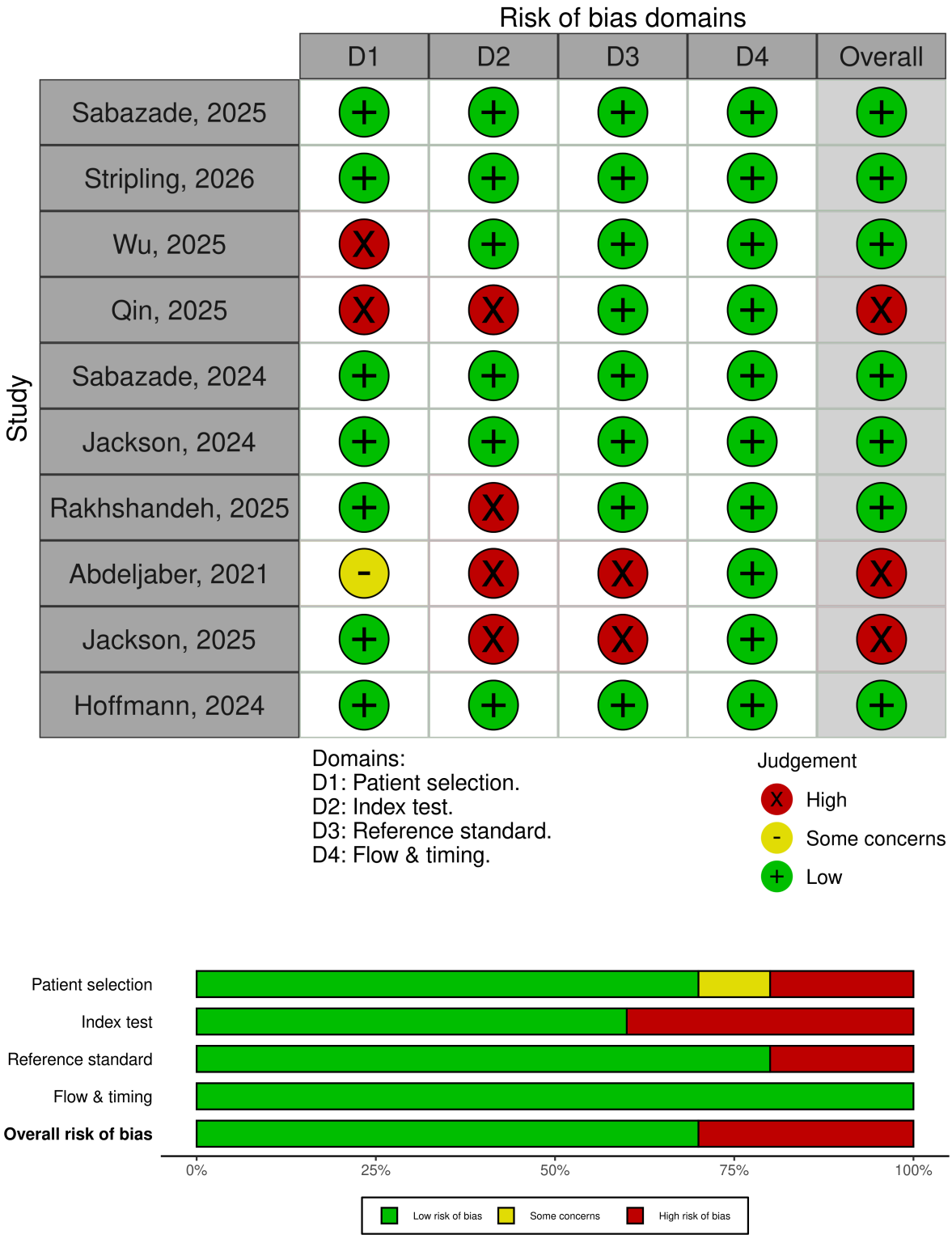


Figure 1

Quality assessment of included studies using the QUADAS-AI tool. D1 represents the patient selection domain, and D2, D3, and D4 represent the index test, reference standard, and workflow domains, respectively. Risk of bias is categorized as low, high, or some concerns.

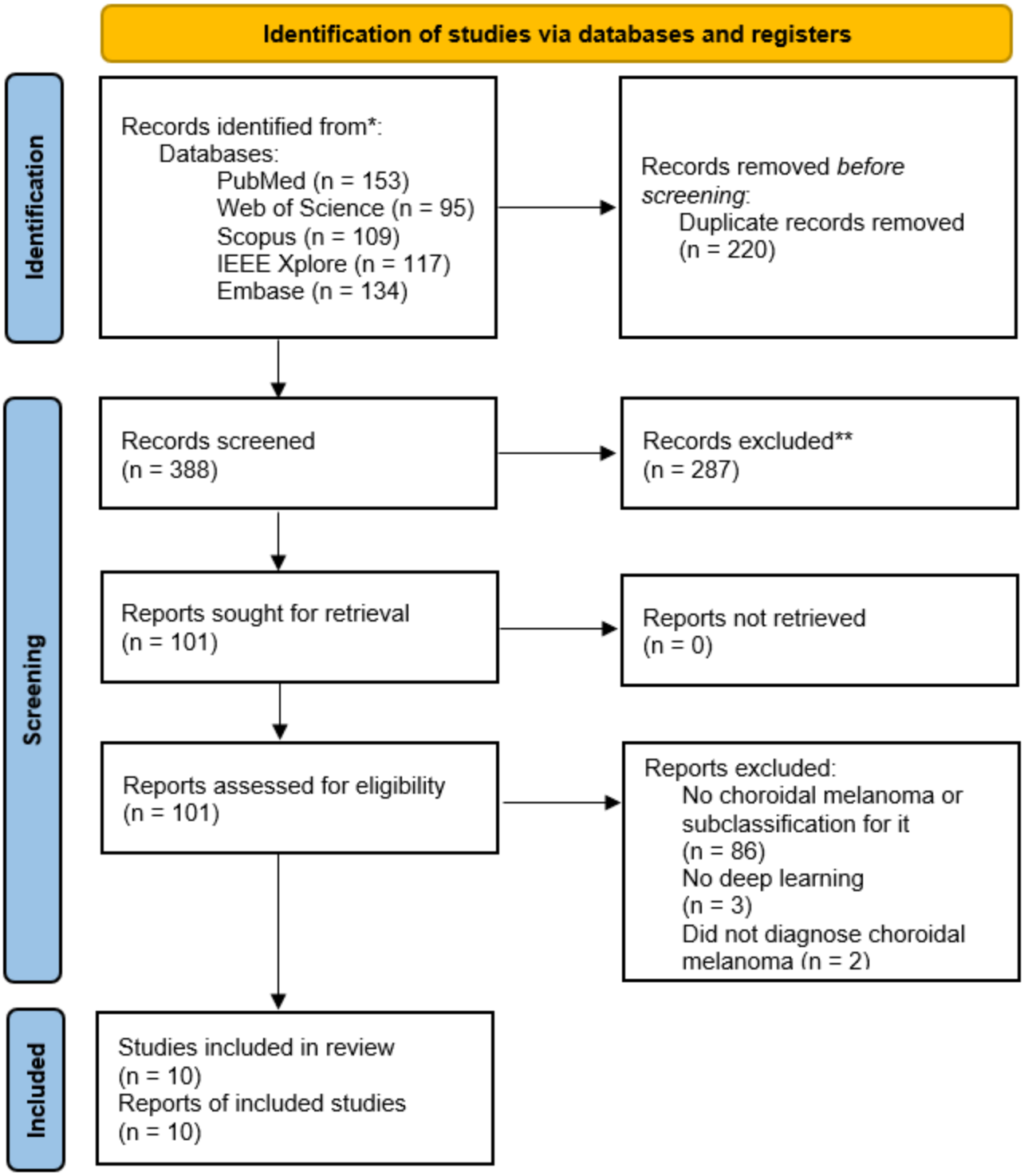


Figure 2

PRISMA flow diagram showing the study selection process, covering identification, screening, eligibility, and inclusion stages, along with record counts at each step and reasons for exclusion.

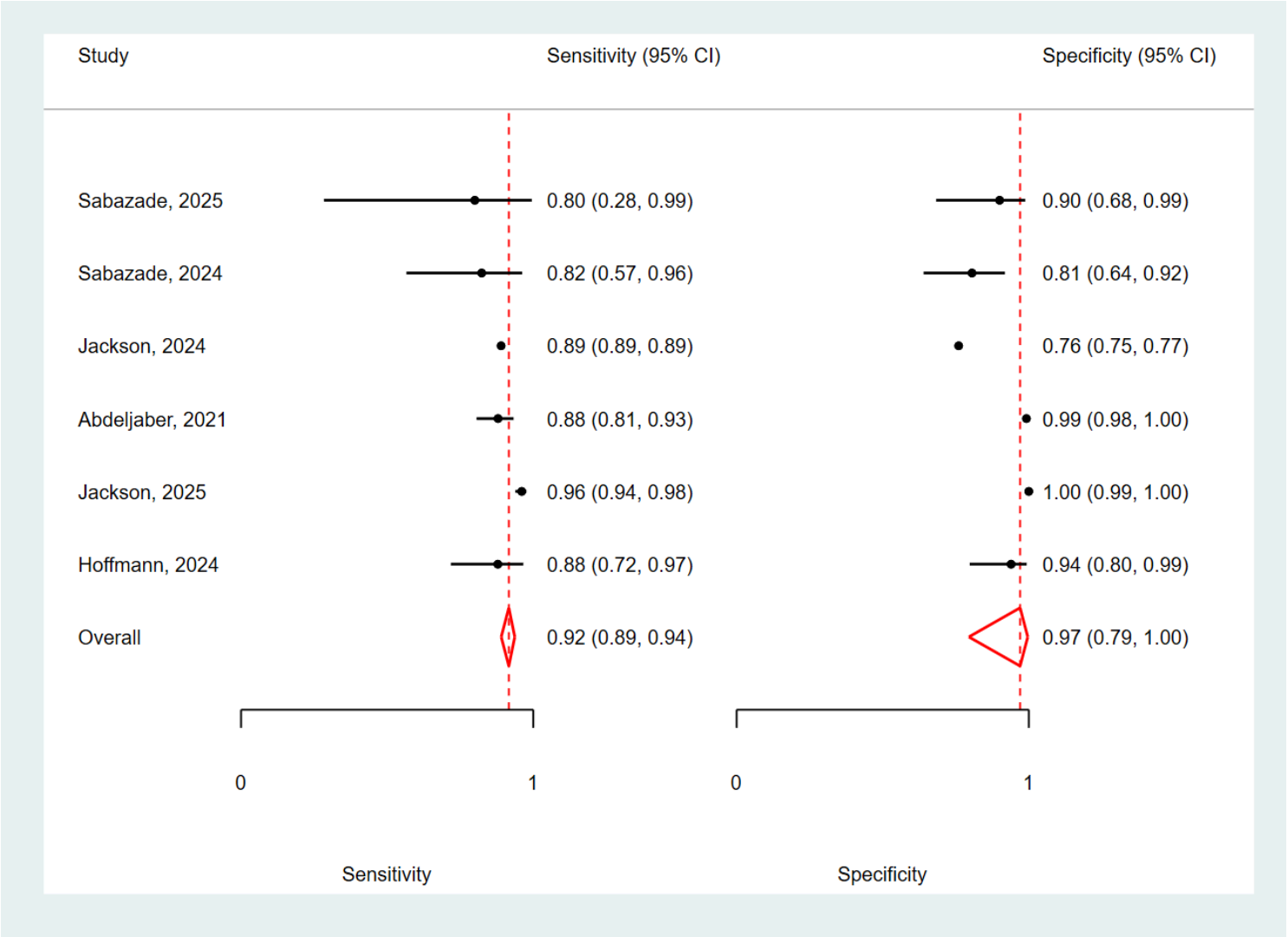


Figure 3

Sensitivity and specificity meta-analysis. Between-study heterogeneity was generally low to moderate across the diagnostic accuracy parameters. For sensitivity, τ^2 was 0.10, and I^2 was 13.19%, indicating low heterogeneity. For specificity, τ^2 was 6.03 with an I^2 of 32.76%, consistent with moderate heterogeneity across studies. In the generalized (joint) model, τ^2 was 0.00, and I^2 was 0.00%, indicating low overall heterogeneity in the bivariate diagnostic accuracy framework.

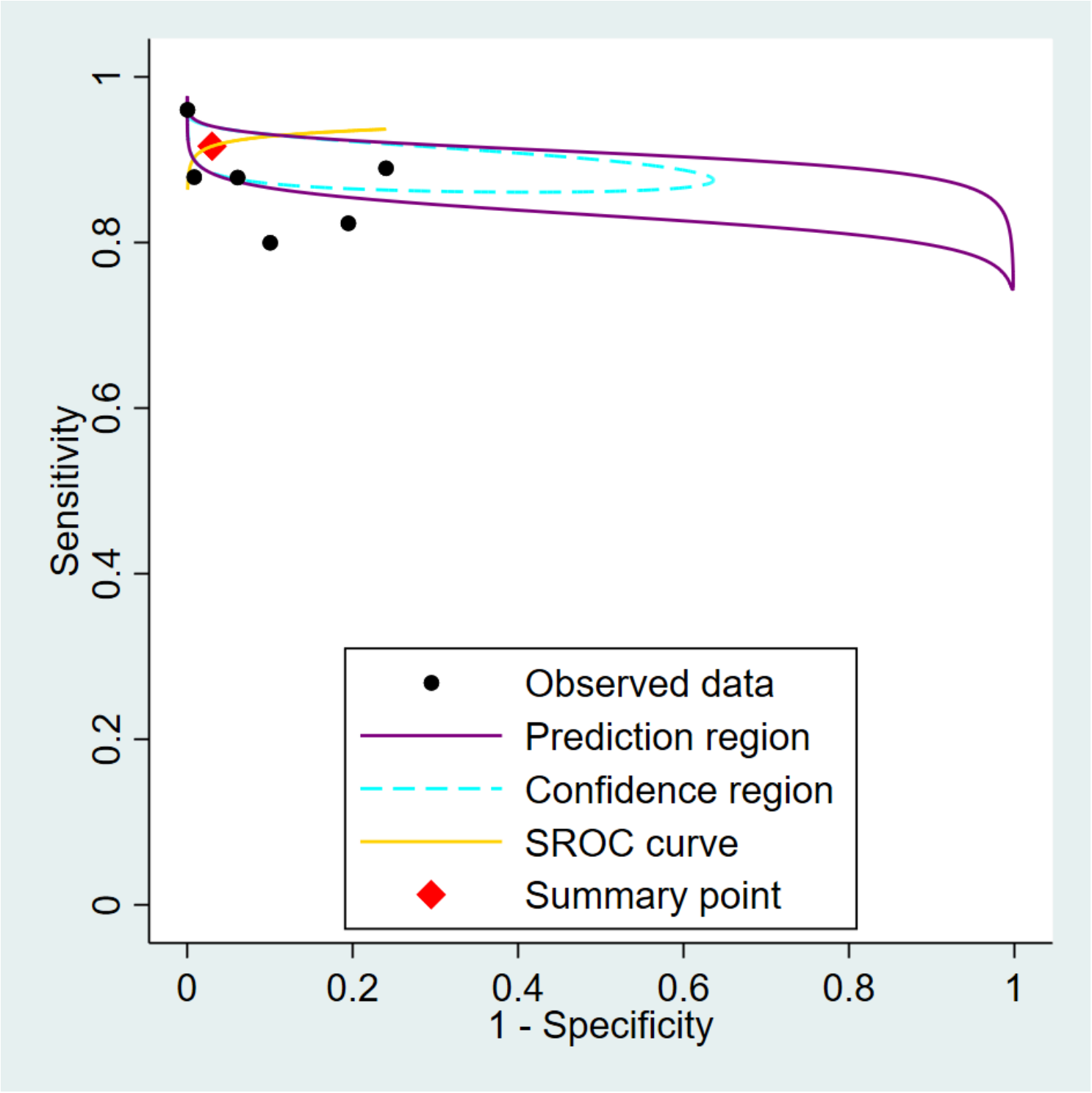


Figure 4

Summary receiver operating characteristic curve. SROC: summary receiver operating characteristic.

Prediction \ Actual	CM	non-CM	Total
CM	17061 58.94%	2097 7.25%	19158 89.05% 10.95%
non-CM	2076 7.17%	7710 26.64%	9786 78.79% 21.21%
Total	19137 89.15% 10.85%	9807 78.62% 21.38%	24771 / 28944 85.58% 14.42%

Figure 5

The pooled confusion matrix.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- TableS1.docx
- TableS2.xlsx
- FileS1.docx