



A CLINICALLY GROUNDED AND UNCERTAINTY-AWARE EXPLAINABLE ARTIFICIAL INTELLIGENCE DETAILED ANALYSIS FOR SKIN LESION DIAGNOSIS

Vijay Kumar *, Pooja Koshti and Shubham Dwivedi

School of Computer Technology, Sanjeev Agrawal Global Educational University, Bhopal, Madhya Pradesh, India.

* Corresponding Author

ORCID Details

Vijay Kumar: <https://orcid.org/0009-0009-0246-1736>

Pooja Koshti: <https://orcid.org/0009-0003-1611-0781>

Shubham Dwivedi: <https://orcid.org/0009-0008-5846-4605>

Open Access Research Journal of Engineering and Technology, 2026, 11(01), 085–098

Publication history: Received on 08 August 2026; revised on 13 September 2026; accepted on 16 September 2026

Article DOI: <https://doi.org/10.53022/oarjet.2026.11.1.0066>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Artificial intelligence (AI), specifically machine learning and deep learning, has demonstrated remarkable potential for the automated diagnosis of skin lesions using dermoscopic images. While deep learning models can achieve high classification performance, significant barriers to clinical adoption remain, including limited interpretability, uncertainty in model predictions, dataset bias, and limited clinical validation. This review examines the development of AI-based skin lesion diagnosis placing special emphasis on explainable AI (XAI), clinically meaningful concept representation, uncertainty estimation, and dermatologist-centred evaluation.

The reviewed literature suggests that traditional pixel-level interpretation methods can identify image regions associated with model predictions but fail to effectively communicate clinically meaningful evidence. In contrast, clinically grounded approaches can link visual information to recognizable dermatological concepts such as asymmetry, border characteristics colour variation, diameter-related features, and relevant dermoscopic structures. This paper synthesizes these developments and highlights critical research gaps regarding concept-level interpretation, uncertainty-aware interpretation, longitudinal assessment, dataset generalization, and clinical validation.

Based on this synthesis, a conceptual framework is proposed that integrates dermoscopic image analysis, lesion segmentation, deep learning-based prediction, clinical concept recognition, uncertainty estimation, clinical reasoning, and structured explanation. This framework distinguishes the reliability of the prediction from the reliability of the explanation and characterizes concepts as clearly detected undetected uncertain or non assessable particular attention is paid to progression, which cannot be reliably determined from a single dermoscopic image without longitudinal evidence. The proposed framework aims to support transparent and clinically meaningful AI-assisted skin lesion assessment while maintaining the dermatologist as the ultimate clinical decision-maker. Future research should involve quantitative validation of the framework through dermatologist-centred studies and evaluate the quality of explanations, consistency, uncertainty calibration, and clinical utility.

Keywords: Artificial Intelligence; Explainable AI; Skin Lesion Diagnosis; Melanoma; Dermoscopy; Deep Learning; Clinical Concepts; Uncertainty Estimation; Clinical Decision Support

1 INTRODUCTION

Artificial Intelligence has moved from systems that follow rules to systems that learn from data and can find complex patterns in big sets of information. In medicine Artificial Intelligence is being studied more and more for tasks such as diagnosis detecting disease analyzing medical images and supporting clinical decisions [1, 2].

Skin cancer is a challenge in health care and melanoma is one of the most aggressive types of skin cancer because it can spread quickly and grow new tumors [1]. A quick and precise assessment is therefore crucial. In recent years, dermoscopic image analysis has become an active area of machine learning and deep learning research, and many studies have investigated different computational methods for detecting and classifying melanoma [1, 2].

Dermoscopy is an imaging method that makes details within skin lesions easier to visualize and can help clinicians distinguish benign from suspicious lesions [1]. However, image interpretation depends on clinical expertise and may vary between observers. This creates an opportunity for Artificial Intelligence systems to assist clinicians by identifying important visual patterns in dermoscopic images.

Deep learning models, including convolutional neural networks (CNNs), can automatically learn detailed features such as color texture shape and other visual patterns [2]. A common Artificial Intelligence diagnostic pipeline usually includes taking the picture cleaning it up separating the lesion pulling out features building the model and checking its performance with numbers like accuracy sensitivity specificity precision recall F1-score and the area under the receiver operating curve (AUC) [1, 2].

A high prediction score does not automatically mean that the system is trustworthy in a clinical setting. Deep learning systems are often called black-box models because clinicians cannot easily understand how they make decisions [6, 7]. A model might give a confident answer while not giving enough detail about the clinical evidence behind that answer.

Because of this limitation, increasing attention has been given to Explainable Artificial Intelligence. Common XAI methods such as Grad-CAM and LIME can provide visual or feature-based clues that explain model predictions [6, 9]. But pointing out an area in an image does not always mean it shows a clinically important skin feature. For instance a heatmap may show where the model looked. It may not say if the lesion has asymmetry an uneven border mixed color or another important feature.

Therefore, it is necessary to consider explanations at the clinical concept level rather than simply stating, "This image region affected the prediction." A system grounded in clinical knowledge might instead state, "The lesion shows marked asymmetry, an irregular border, and heterogeneous colour, which increase suspicion." This approach links the computational analysis more directly to the features considered by clinicians [7, 8].

Another key issue is uncertainty. A model can produce a probability even when the available image does not provide clear or sufficient evidence. Estimating uncertainty has become a part of trustworthy Artificial Intelligence especially in safe areas like health care [4]. It is important to remember that confidence in the diagnosis and confidence in the explanation that supports it are not always the same.

This review therefore looks at how Artificial Intelligence skin lesion diagnosis explainability clinical concepts and uncertainty are connected. The paper also points out problems in current methods and offers a new framework that is both clinically grounded and aware of uncertainty for skin lesion assessment.

1.1 Review objective

The goal of this review is to examine how AI and XAI are currently being used for skin lesion assessment. It identifies limitations in current approaches. It then proposes a framework based on clinical needs. The framework includes uncertainty representation and dermatologist involvement in evaluation.

The review focuses on the difference between AI prediction capability and the clarity of the evidence used to support those predictions. It also considers situations in which important clinical characteristics cannot be reliably determined from an image, particularly when assessing how a lesion has changed over time.

1.2 Research question

The central research question is:

- Can AI-based skin lesion diagnosis provide clinically meaningful and uncertainty-aware explanations that can be evaluated and understood by dermatologists
- This question shifts the focus from prediction accuracy alone toward the broader requirements of transparent and clinically useful AI-assisted assessment.

2 AI IN SKIN LESION DIAGNOSIS

2.1 Automated skin lesion classification

Automated classification is one of the most established applications of AI in dermatology. Deep learning models can be trained using dermoscopic images to distinguish melanoma from benign lesions or to perform multiclass skin lesion classification [1, 2].

Existing studies have investigated CNN-based architectures and related deep learning approaches using datasets such as HAM10000 and ISIC [1, 2]. These studies demonstrate the ability of AI systems to learn discriminative visual patterns from dermoscopic images.

The conventional objective of such systems is generally expressed as:

Dermoscopic Image → Diagnostic Class

For Ex.

Input Image → Melanoma

or

Input Image → Benign Lesion

Although this approach is useful, the classification result alone does not necessarily communicate why the model reached its decision. Therefore, diagnostic classification should be considered one component of a broader clinically interpretable AI system.

2.2 Lesion segmentation

Before extracting lesion-related characteristics, localization of the lesion can be important. Segmentation separates the lesion from surrounding skin and may reduce the influence of irrelevant image regions [1, 2, 3].

A general diagnostic workflow can therefore be represented as

Dermoscopic Image → Preprocessing → Lesion Segmentation → Feature/Concept Extraction → Classification → Explanation

Segmentation is particularly relevant to a concept-based framework because clinical characteristics such as asymmetry and border irregularity depend on an accurate representation of the lesion boundary.

However segmentation itself should not be interpreted as clinical reasoning rather it provides a computational representation that can support subsequent feature and concept analysis.

2.3 Clinical concept identification

A major limitation of purely image-level classification is that the learned representation may not correspond directly to concepts recognizable by clinicians.

For melanoma assessment, the ABCDE rule provides a commonly recognized framework involving:

- A - Asymmetry
- B - Border irregularity
- C - Color variation
- D - Diameter
- E - Evolution

The reviewed literature identifies clinically meaningful characteristics as important components of melanoma assessment [1] Concept-based approaches have also investigated the connection between visual information and higher-level clinical concepts [8].

Rather than treating ABCDE as only a traditional checklist, an AI system can investigate whether these concepts can be computationally estimated or represented.

A possible mapping is shown below:

Table 1: Representation of Clinical Concepts and Their Evidence Availability

Clinical Concept	Possible AI Representation	Evidence Availability
Asymmetry	Difference between lesion halves	Generally assessable from image
Border	Boundary irregularity or complexity	Generally assessable
Color	Number and distribution of color regions	Generally assessable
Diameter	Estimated lesion size	Conditional on scale information
Evolution	Temporal change	Requires longitudinal images

This representation highlights an important principle: not every clinical concept is equally observable from every input.

Asymmetry border and color characteristics may be investigated from a single image Diameter may require an appropriate scale or calibration information. Evolution however is inherently temporal and cannot be reliably established from a single image.

Concept-level representation therefore provides a potential bridge between image features and clinical explanation [8, 1].

2.4 Explainable and uncertainty-aware AI

2.4.1 Conventional explainable AI

Conventional XAI methods attempt to provide information about the relationship between an input and a model prediction. Grad-CAM, for example, can highlight image regions associated with a CNN prediction [9]. LIME provides local explanations by approximating model behaviour around a particular prediction [9]. Broader XAI research has also established taxonomies and principles for interpretable and responsible AI systems [6, 5].

These approaches are valuable because they provide information beyond the final classification label. Nevertheless, a highlighted region does not automatically correspond to a clinically meaningful explanation.

Example

- Melanoma: 94% probability
- provides a prediction but limited clinical evidence.
- A heatmap can improve the situation by showing an influential region but a clinician may still need to interpret that region independently.

A concept-based explanation could instead provide:

- Prediction- Suspicious lesion
- Clinical concepts- High asymmetry irregular border heterogeneous pigmentation
- Dermoscopic evidence: Relevant structural patterns detected
- Explanation confidence: Moderate

This type of output attempts to translate image evidence into concepts that are more directly understandable within a clinical context.

2.5 From pixel-level explanation to concept-level explanation

The difference between pixel-level and concept-level explanations can be summarized as

Pixel-level XAI

Image → Important Region → Prediction

Concept-based XAI

Image → Clinical Observation → Clinical Concept → Prediction

The second representation is potentially more aligned with clinical reasoning because the explanation is expressed in terms of recognizable characteristics.

Existing research has emphasized the importance of explainability and concept-based interpretation in skin lesion diagnosis [7, 8] However, clinical concepts themselves are not always perfectly objective. Determining whether a border is irregular, for example may involve expert interpretation [8, 1].

Therefore, concept extraction should not automatically be treated as unquestionable ground truth. Where possible, concept-level outputs should be evaluated against dermatologist annotations and assessed for consistency and clinical relevance.

2.6 Uncertainty-aware skin lesion diagnosis

Uncertainty is particularly important in medical AI because a confident model output does not necessarily guarantee that the underlying evidence is sufficient. Research on uncertainty in deep neural networks has identified different forms and approaches for representing uncertainty in model predictions [4].

For skin lesion diagnosis, uncertainty should be considered at more than one level.

2.7 Prediction confidence

Prediction confidence represents the model's confidence in its diagnostic output.

For example

Melanoma probability = 0.91

may be reported as a high-confidence prediction.

However, numerical confidence should not automatically be interpreted as clinical certainty. Model calibration and reliability must also be considered when confidence values are communicated.

2.8 Explanation confidence

Explanation confidence represents the confidence that a detected clinical concept provides reliable supporting evidence for the prediction.

For example, a system might produce:

Prediction confidence High

Explanation confidence Moderate

This situation is possible because a model can be confident about a classification while being less certain about the specific clinical characteristics associated with that classification. Therefore,

Prediction confidence is not equal to Explanation confidence

This distinction represents an important direction for future research in clinically grounded XAI.

2.9 Evidence availability

The framework should also distinguish between evidence that is absent and evidence that is unavailable.

Clinical concepts could therefore be represented as:

Detected Not detected Uncertain Not assessable

This distinction is important because:

Not detected is not equal to Not assessable.

For example if a single image does not provide longitudinal information the correct output for evolution should not be

Evolution Not detected

because that could incorrectly imply that no evolution occurred.

A more appropriate output is:

Evolution: Not assessable - insufficient longitudinal evidence.

This follows the principle of distinguishing absence of evidence from evidence of absence.

2.10 Challenges and research gaps

BLACK-BOX NATURE OF DEEP LEARNING

Deep learning models can achieve high predictive performance, but their internal decision-making processes remain difficult to interpret [6, 7]. This creates a gap between computational performance and clinical understanding.

The problem is not simply whether the model is accurate. A clinically useful system must also communicate evidence in a form that clinicians can evaluate.

2.11 Limitations of pixel-level explanations

Grad-CAM and LIME provide useful mechanisms for examining model behavior [9]. However,, a visual heatmap does not necessarily correspond to a recognized dermatological concept.

Consequently, an important research question is how pixel-level evidence can be connected to higher-level clinical concepts [6, 8].

The challenge is therefore: Image region → What clinically meaningful characteristic does this region represent?

2.12 Clinical concept ambiguity

Clinical characteristics may not always have perfectly objective boundaries. Border irregularity and color variation can involve expert interpretation [8, 1].

For this reason, dermatologist annotations should be treated as an important reference for evaluating concept extraction rather than assuming that every concept can be represented as a perfectly objective binary label.

2.13 Evolution cannot be determined reliably from one image

Evolution is inherently temporal.

A single dermoscopic image cannot directly establish whether a lesion has changed in size, shape, colour, border, or internal structure

Therefore, systems receiving only a single image should explicitly represent evolution as

Not assessable

or

Insufficient longitudinal evidence.

Longitudinal analysis can instead use

Image 1 → Image 2 → Image 3 → Change Detection → Evolution Assessment

This represents an important direction for future AI research.

2.14 Prediction confidence versus explanation confidence

A model may be highly confident about the diagnostic class while being uncertain about the clinical concept used to explain that prediction.

This creates a research gap because conventional model evaluation frequently emphasizes classification performance rather than the reliability of the explanation itself. Future studies should therefore evaluate both:

How accurate is the prediction and

How reliable and clinically meaningful is the explanation?

2.15 Dataset bias and generalization

Dataset diversity and generalization remain important challenges in AI-based skin lesion diagnosis [1]. A model trained on one dataset should not automatically be assumed to perform equally well on images collected using different devices, populations, imaging conditions, or clinical environments.

External validation is therefore important for determining whether a model's performance and explanations remain stable beyond the dataset used for development.

2.16 Clinical validation

Technical accuracy alone does not establish clinical usefulness [7]. A clinically grounded explanation should eventually be evaluated by dermatologists for-

- Correctness
- Clinical relevance
- Understandability
- Consistency
- Usefulness
- Agreement with expert assessment

Clinical evaluation should therefore become an explicit component of future XAI research rather than treating explainability as merely a visualization feature.

2.17 Research gap synthesis

The reviewed literature suggests that several important gaps remain at the intersection of AI, XAI, uncertainty and clinical dermatology.

Table 2: Synthesis of Research Gaps and Suggested Directions

Challenge	Existing Situation	Identified Gap	Suggested Direction
Diagnostic prediction	Strong classification performance reported	Prediction alone provides limited clinical evidence	Add clinical concept representation
Black-box models	Deep learning learns complex image patterns	Internal decision process is difficult to interpret	Clinically grounded explanations
Pixel-level XAI	Heatmaps and local explanations available [9]	Image regions may not map directly to clinical concepts	Concept-level interpretation
Clinical concepts	ABCDE and dermoscopic characteristics are clinically relevant [1]	Automated representation remains challenging	Dermatologist-validated concept extraction
Prediction confidence	Model confidence can be reported	Confidence may not reflect explanation reliability	Separate prediction and explanation confidence
Evidence availability	Models often produce outputs for all inputs	Some concepts cannot be assessed from available evidence	Detected / Not detected / Uncertain / Not assessable
Evolution	Single-image diagnosis is common	Evolution requires temporal evidence	Longitudinal lesion analysis
Dataset diversity	Public dermoscopic datasets support model development [1, 2]	Generalization across populations and environments remains uncertain	External and multi-source validation
Clinical validation	Technical metrics dominate many evaluations	Clinical usefulness may remain insufficiently assessed	Dermatologist-centered evaluation

This synthesis suggests that the next stage of research should move beyond the question:

Can AI classify skin lesions accurately?

Can AI classify skin lesions accurately while identifying clinically meaningful evidence, communicating uncertainty, and providing explanations that clinicians can evaluate?

3 FUTURE DIRECTIONS AND CLINICAL IMPLICATIONS

3.1 From prediction to clinical evidence

Future dermatology AI systems should not be limited to generating a diagnostic label. A more clinically meaningful pathway can be represented as:

Image → Observation → Clinical Concepts → Evidence → Prediction → Uncertainty

This structure creates a more transparent connection between the input image and the diagnostic output

The goal is not to claim that AI has reproduced the complete reasoning process of a dermatologist. Instead, the objective is to provide clinically interpretable evidence associated with the prediction.

3.2 Integration of multimodal clinical information

Dermoscopic images may not provide all information required for comprehensive lesion assessment. Future systems could potentially integrate

- Dermoscopic images
- Patient age
- Lesion location
- Clinical history
- Previous lesion images
- Histopathological information

Such multimodal integration could provide a richer clinical representation; however, it should ideally be introduced after establishing a reliable image-based explainability framework [10].

3.3 Longitudinal lesion analysis

Evolution is particularly challenging because it requires temporal information.

Future research can investigate:

- Image 1 → Image 2 → Image 3 → Temporal Change Detection → Evolution Explanation
- Such systems could investigate changes in:
 - Lesion diameter:-
 - Shape Border Color
 - Internal structures

This would allow evolution to be represented using actual temporal evidence rather than being estimated from a single image.

3.4 External clinical validation

A clinically meaningful system should eventually be evaluated on external datasets and, where possible, on real-world clinical data.

External evaluation is particularly important because differences in populations, imaging devices, acquisition conditions, and clinical environments can affect model performance and explanation behaviour [1].

3.5 Human - ai collaboration

The intended role of the proposed framework is not to autonomously replace dermatologists.

Instead, the desired model is

Dermatologist + AI Prediction + Clinical Evidence + Explanation + Uncertainty

The dermatologist should remain responsible for the final clinical assessment while AI provides computational assistance and structured evidence.

This human - AI collaboration is particularly important in medical applications where inappropriate confidence or misleading explanations could influence clinical decisions.

3.6 Toward trustworthy medical AI

Trustworthy medical AI requires more than high classification accuracy. Important properties include:

- Accuracy
- Robustness
- Explainability Uncertainty awareness
- Generalizability
- Clinical validity
- Transparency

Therefore explainability should not be treated merely as an additional visualization placed after model development it should be considered as an integral part of the overall diagnostic system.

3.7 Proposed clinically grounded and uncertainty-aware framework

Based on the reviewed literature and identified research gaps this paper proposes a conceptual framework for clinically grounded and uncertainty-aware skin lesion diagnosis.

3.7.1 Framework architecture

The proposed framework consists of the following stages-

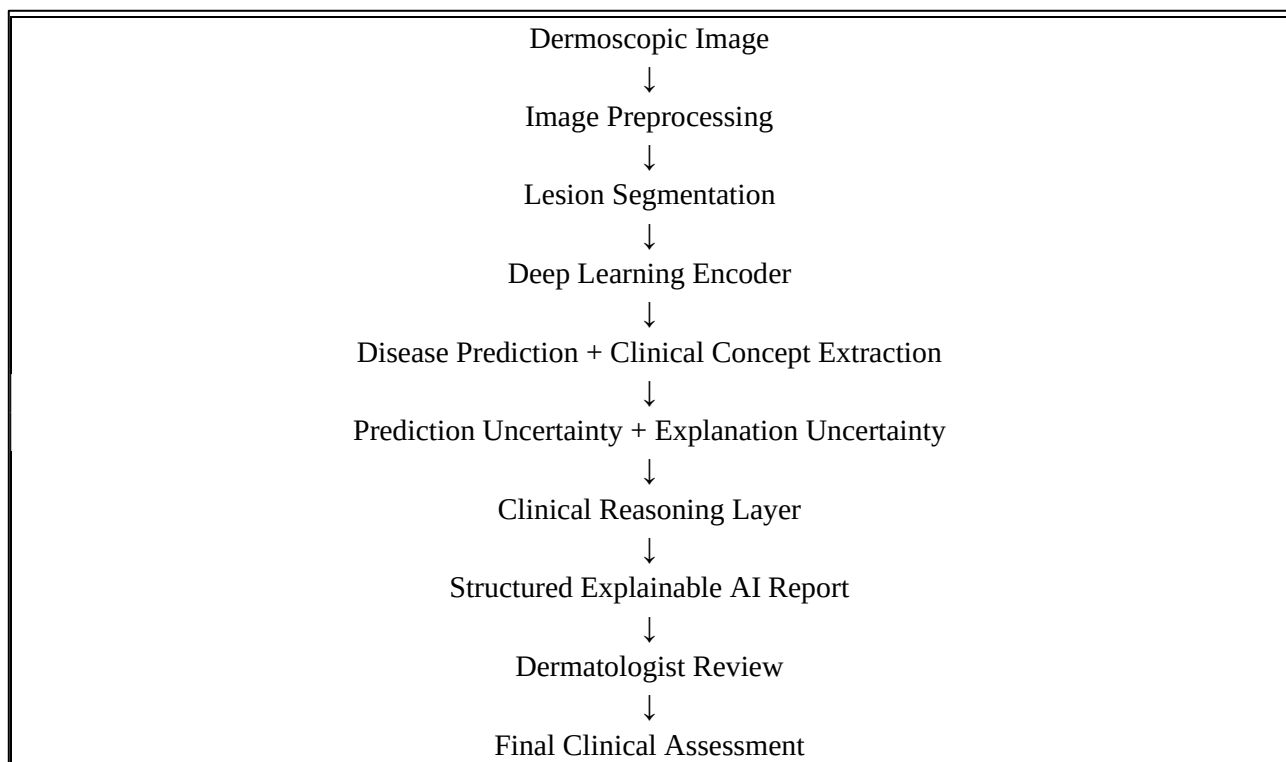


Figure 1: Proposed clinically grounded and uncertainty-aware framework

The framework intentionally separates disease prediction from clinical concept extraction. This enables the system to examine both the diagnostic class and the observable evidence associated with that class.

3.7.2 Clinical concept layer

- The clinical concept layer can include
- A – Asymmetry
- Comparison of different regions or halves of the lesion
- B – Border
- Characteristics related to boundary irregularity or complexity
- C – Color
- Distribution and heterogeneity of visible colors.

- D – Diameter
- Estimated size where an appropriate scale or calibration is available
- E – Evolution
- Temporal change assessable only when appropriate longitudinal evidence is available
- Additional dermoscopic features may also be incorporated where clinically relevant

3.7.3 Evidence-state representation

Each concept should not necessarily be forced into a binary output.

A four-state representation is proposed-

Table 3: Evidence-State Representation for Clinical Concepts

State	Meaning
Detected	Evidence supports the concept
Not detected	Available evidence does not support the concept
Uncertain	Evidence is ambiguous
Not assessable	Required evidence is unavailable

This structure is particularly important for evolution and diameter related characteristics.

4 DUAL UNCERTAINTY REPRESENTATION

The framework distinguishes:

- Prediction Uncertainty
- Uncertainty is associated with the diagnostic classification
- Explanation Uncertainty
- Uncertainty associated with the reliability of clinical concepts or evidence used to explain that prediction

The two should be reported separately.

For example

- Prediction Suspicious lesion
- Prediction confidence High
- Clinical concepts Asymmetry - Detected; Border - Irregular; Color – Heterogeneous
- Explanation confidence Moderate
- Evolution Not assessable
- Reason: No longitudinal images available

This provides a more transparent representation of what the model knows and what it cannot establish.

4.1 CLINICAL REASONING LAYER

The clinical reasoning layer is conceptual rather than a claim of human-like reasoning.

Its purpose is to organize:

Model prediction + clinical concepts + evidence availability + uncertainty into a structured explanation.

Example

High asymmetry + irregular border + heterogeneous pigmentation + supporting dermoscopic features → increased lesion suspicion

The explanation should communicate evidence associated with the prediction without claiming that the AI has recovered a definitive causal reasoning process.

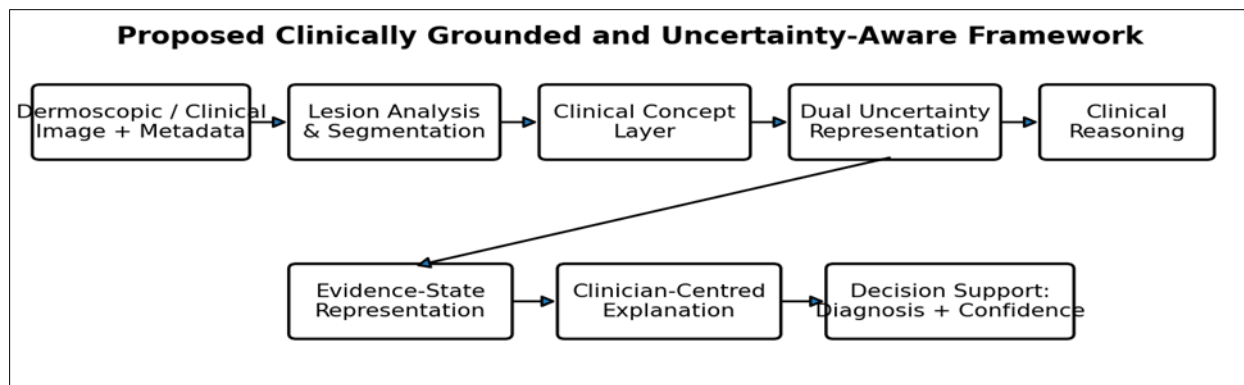


Figure 2: Framework architecture for clinically grounded and uncertainty-aware skin lesion diagnosis

Component	AI Output
Diagnosis	Melanoma / Benign / Other
Prediction Confidence	High / Moderate / Low
Asymmetry	Detected / Not detected / Uncertain
Border	Regular / Irregular / Uncertain
Color	Homogeneous / Heterogeneous / Uncertain
Diameter	Estimated where scale permits
Evolution	Assessable / Not assessable
Dermoscopic Features	Relevant detected patterns
Clinical Evidence	Structured concept summary
Explanation Confidence	High / Moderate / Low
Recommendation	Further expert assessment when appropriate

4.2 Proposed ai output

The final system could generate a structured report such as:

- Is the prediction correct, is the supporting evidence meaningful, and is the explanation useful to a clinician?
- Importantly these outputs represent the proposed framework not experimental results
- The purpose of this structure is to provide the clinician with three distinct forms of information
- What the model predicts
- What evidence supports the prediction
- How certain the model is about both the prediction and the explanation

4.3 Dermatologist-centred evaluation

A major requirement for future validation is assessment by dermatologists rather than evaluation solely through machine learning metrics.

Potential evaluation dimensions include

- Concept Correctness
- Does the AI correctly identify the relevant clinical characteristic
- Explanation Relevance
- Does the explanation contain clinically meaningful evidence
- Explanation Consistency
- Does the explanation remain stable for similar cases
- Explanation Confidence
- Does the reported confidence correspond to the reliability of the explanation

- Clinical Usefulness
- Does the explanation help the dermatologist evaluate the AI output
- Clinical Agreement
- Does the explanation correspond reasonably with expert assessment
- Traditional machine-learning metrics such as accuracy precision recall F1- score and AUC should therefore be complemented by concept - level and clinician-centered measures.
- This shifts evaluation from
- Is the prediction correct

Limitations of the proposed framework

The proposed framework is conceptual and has several limitations:

- First, clinical concepts such as border irregularity and color variation may involve expert interpretation and may not have perfectly objective definitions [8, 1].
- Second, the ability to estimate diameter depends on the availability of appropriate scale or calibration information.
- Third, evolution cannot be reliably assessed from a single dermoscopic image longitudinal images are required to directly investigate temporal change.
- Fourth, a clinically grounded explanation does not necessarily represent the complete internal mechanism of a deep learning model. The proposed explanation should therefore be understood as a structured representation of clinically relevant evidence rather than a definitive description of causal model reasoning.
- Fifth, dataset bias and limited diversity may affect both diagnostic predictions and explanations [1]. Therefore external validation across different populations devices and imaging environments will be necessary
- Finally, the proposed framework has not yet been experimentally validated quantitative model evaluation uncertainty calibration concept - level assessment and dermatologist - centered evaluation remain important future research activities

4.3.1 Expected research contributions

The proposed review and conceptual framework provide several potential research contributions:

4.3.2 Clinically grounded xai

The framework emphasizes recognizable dermatological concepts rather than relying exclusively on pixel-level heatmaps [6, 8].

4.3.3 Concept level interpretation

The approach proposes mapping visual information from dermoscopic images to clinically meaningful characteristics such as asymmetry border color diameter and relevant dermoscopic features [8, 1].

4.3.4 Uncertainty-aware explanation

The framework distinguishes diagnostic prediction confidence from explanation confidence and explicitly identifies concepts that cannot be assessed [4].

4.3.5 Structured clinical reporting

The framework transforms model outputs into a structured representation containing prediction clinical concepts evidence availability uncertainty and explanation.

4.3.6 Dermatologist centred validation

The framework identifies expert evaluation as an important future validation component rather than relying exclusively on conventional machine learning performance metrics [7].

5 CONCLUSION

Artificial intelligence and deep learning have demonstrated significant potential for the automated diagnosis of skin lesions and melanoma from dermoscopic images [1, 2]. However, high predictive performance alone does not guarantee interpretability reliability generalizability or clinical utility.

This review examines the development of AI based skin lesion diagnosis placing particular emphasis on interpretability clinical concepts uncertainty and clinical validation the analysis reveals that while conventional pixel level explanations can provide useful information regarding a model's focus they may not always convey evidence in a manner that directly aligns with clinical interpretation [6, 9].

A clinically grounded approach can instead connect image information with recognizable concepts such as asymmetry border characteristics color variation diameter related characteristics and relevant dermoscopic structures [8, 1] The review further highlights the importance of explicitly representing evidence availability and uncertainty. One particularly important distinction is

Prediction confidence is not equal to Explanation confidence

A model may be confident about its diagnostic prediction while remaining uncertain about the specific clinical evidence associated with that prediction. Future AI systems should therefore evaluate and communicate these forms of uncertainty separately.

The review also emphasizes that certain clinical characteristics cannot be reliably inferred from a single image. In particular, evolution requires longitudinal evidence. A responsible AI system should therefore communicate

Evolution Not assessable insufficient longitudinal evidence

rather than generating an unsupported conclusion.

Based on these observations, a conceptual framework was proposed integrating lesion segmentation deep learning prediction clinical concept identification dual uncertainty representation structured clinical explanation and dermatologist review.

The ultimate objective is not to replace the dermatologist. Rather, the intended direction is:

AI Prediction + Clinical Evidence + Explanation + Uncertainty + Dermatologist Assessment

Such an approach may help move skin lesion AI from a system that simply predicts a diagnostic label toward a system that communicates what it predicts what evidence supports the prediction and when the available evidence is insufficient.

Future research should focus on quantitative validation calibration of uncertainty concept level evaluation external generalization longitudinal lesion analysis and dermatologist centred assessment of explanation quality and clinical usefulness.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] H. Naseri and A. A. Safaei, "Diagnosis and prognosis of melanoma from dermoscopy images using machine learning and deep learning: A systematic literature review," *BMC Cancer*, vol. 25, p. 75, 2025, doi: 10.1186/s12885-024-13423-y.
- [2] Z. Ye, D. Zhang, Y. Zhao, M. Chen, H. Wang, S. Seery, Y. Qu, P. Xue, and Y. Jiang, "Deep learning algorithms for melanoma detection using dermoscopic images: A systematic review and meta-analysis," *Artif. Intell. Med.*, vol. 155, p. 102934, 2024, doi: 10.1016/j.artmed.2024.102934.
- [3] H. Kutlu and C. Çolak, "UncerKAN-Mamba: A clinically robust, transparent, and explainable AI framework for low-latency skin lesion segmentation with deterministic single-pass uncertainty estimation," *Mach. Learn. Knowl. Extr.*, vol. 8, no. 6, p. 153, 2026, doi: 10.3390/make8060153.
- [4] M. Z. Sajid, M. F. Hamid, Z. Yang, M. Alhefdi, S. Alsenan, and Y. Nam, "Morphology-guided attention networks for explainable skin cancer detection under clinical uncertainty," *Front. Oncol.*, vol. 16, p. 1804329, 2026, doi: 10.3389/fonc.2026.1804329.

- [5] J. Whewell, R. Peters, and J. Kulon, "Explainable artificial intelligence for skin lesion classification: A comprehensive review of methods and challenges," *Technologies*, vol. 14, no. 7, p. 391, 2026, doi: 10.3390/technologies14070391.
- [6] R. K. Yadav, A. Nandi, V. Kumar, R. Moriwal, and M. Kumar, "Fuzzy cognitive explainable AI framework integrating ResNet-50 and causal clinical concept reasoning for skin lesion detection," *Sci. Rep.*, 2026, doi: 10.1038/s41598-026-60659-8.
- [7] T. Chanda et al., "Dermatologist-like explainable AI enhances melanoma diagnosis accuracy: Eye-tracking study," *Nat. Commun.*, vol. 16, p. 4739, 2025, doi: 10.1038/s41467-025-59532-5.
- [8] C. Patrício, L. F. Teixeira, and J. C. Neves, "A two-step concept-based approach for enhanced interpretability and trust in skin lesion diagnosis," *Comput. Struct. Biotechnol. J.*, vol. 28, pp. 71–79, 2025, doi: 10.1016/j.csbj.2025.02.013.
- [9] M. Arshad, M. A. Khan, N. A. Almujaal, A. Alasiry, M. Marzougui, and Y. Nam, "Multiclass skin lesion classification and localization from dermoscopic images using a novel network-level fused deep architecture and explainable artificial intelligence," *BMC Med. Inform. Decis. Mak.*, vol. 25, p. 215, 2025, doi: 10.1186/s12911-025-03051-2.
- [10] P. Murali and D. H. Mazumder, "DermaScanAI: An explainable hybrid deep learning framework for automated skin lesion classification using dual attention and metadata fusion," *Sci. Rep.*, vol. 16, p. 15130, 2026, doi: 10.1038/s41598-026-46011-0