Multimodal and Uncertainty-Aware Deep Learning for Skin Cancer Detection: A Review

Sonia Arora[1] *, Shashi Kant Gupta*[2,3]
[1,2] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia;
[3] Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, Índia.
sonia.phd21@gmail.com , shashigupta@lincoln.edu.my
ORCID: [1] https://orcid.org/0009-0004-1833-3951, [2] https://orcid.org/0000-0001-6587-5607

Abstract: Skin cancer is one of the most prevalent and increasing cancers globally, and melanoma is the most aggressive type of skin cancer. While deep learning has been able to approach dermatologist-level accuracy in diagnosis using only an image, current methods mostly focus on image-only predictions, overlooking other diagnostic features used to inform decisions. This review paper examines recent developments in multimodal and uncertainty quantification deep learning techniques for skin cancer diagnosis between 2018 and 2025. The authors discuss image-only, multimodal fusion (dermoscopic image and meta-data), uncertainty quantification and explainable artificial intelligence (XAI) methods. The results show that multimodal models achieve better performance compared with image-only models by incorporating additional meta-data, such as age, lesion site and family history of melanoma, which boost performance and sensitivity to melanoma. Sophisticated multimodal fusion approaches such as cross-attention and gating also help. Predictive uncertainty, such as Bayesian neural networks and Monte Carlo dropout, helps with confidence in predictions, and explainability helps with interpretability. But there are issues with class imbalance, inconsistencies in the evaluation protocol, patient-based validation and diversity of data. This article addresses these problems and suggests future research for clinically relevant AI models.

Keywords: Skin cancer detection; Multimodal deep learning; Uncertainty-aware learning; Explainable artificial intelligence; Medical image analysis.

1. Introduction

One of the most prevalent types of cancer in the world and a significant health burden on the populace in general is skin cancer. Exposure to ultraviolet light, environmental factors, an ageing population and increased awareness of the disease have led to the increased rates of both non-melanoma and melanoma skin cancers over the last few decades. This is because early diagnosis will result in effective treatment and an increased rate of survival but a late diagnosis will result in disease progression and death. Due to its metastatic nature, Melanoma is the most severe form of skin cancer, and the cause of most skin cancer deaths [1], [2].

Traditional methods of diagnosis are visual inspection, dermoscopy and biopsy. Dermoscopy helps in detection of underlying structures and increases the diagnostic accuracy; however, these are subject to variability. Also, in most places and during biopsy procedures, there is a shortage of dermatologists, which are also invasive, time consuming and costly [3]. These challenges have spurred the development of computer-aided diagnostic systems with artificial intelligence. Deep learning, and specifically convolutional neural networks (CNNs) have shown promising results in the area of medical image processing in recent years. CNNs have been used to classify, segment and predict the malignancy of skin lesions with similar accuracy as that of dermatologists in controlled environments [4], [5]. Models that have been shown to be highly accurate and stable on benchmark datasets include ResNet, DenseNet, EfficientNet and multi-model ensembles [6], [7], [8]. But the existing models are predominantly image models and fail to take into account vital clinical data. Dermatologists use image analysis and contextual information such as age, sex, location, family history and symptoms are used to diagnose skin lesions. Removal of metadata interferes with the capability of image-only models to simulate human clinical reasoning and it may result in less accurate diagnosis [4]. In addition, imbalance of classes is a severe problem, and in the public datasets, there are much more benign lesions than melanoma. This can bias the performance of models on the dominant class and limit the chances of identifying rare yet significant clinical cases [5]. The other challenge is the uncertainties of model forecasts. Even when the deep learning model makes predictions based on uncertain or out-of-distribution inputs, the predictions made by the deep learning model are over-confident, and may be misleading to the clinicians. To this end, uncertainty-conscious learning algorithms such as Bayesian neural networks, Monte Carlo dropout and deep ensembles have been proposed to quantify the uncertainty in predictions and to make the predictions better [6]. Also, suboptimal validation methods, including image-level data splitting, may result in data leakage and inflated performance metrics, highlighting the importance of patient-level and multicenter validation.

To address these challenges, multimodal deep learning approaches that combine lesion images with structured data have shown great potential. Multimodal approaches better simulate clinical practice by integrating complementary information sources, leading to better diagnostic performance. Sophisticated fusion techniques, like cross-attention and gated fusion, also improve interaction between modalities, facilitating context-sensitive feature extraction. At the same time, explainable artificial intelligence methods, such as Grad-CAM, saliency maps, and SHAP, are gaining recognition to enhance the transparency and interpretability of models.

2. **Related work**

Over the past decade, machine learning techniques for skin cancer diagnosis have shifted from traditional machine learning approaches using engineered features from skin lesion images, to deep learning and multimodal diagnosis. Previous research mainly relied on engineered features (texture, asymmetry, border irregularity and pigmentation) from dermoscopic images and classification with support vector machines, random forests or shallow neural networks (NNs). Such systems have performed reasonably well but their accuracy was heavily reliant on expert-driven features, and are not as flexible to recognise a broad spectrum of skin lesion subtypes [1]. The advent of deep learning, with convolutional neural networks (CNNs), have revolutionised the automated diagnosis of skin lesions, by automatically extracting features from images. Popular CNN deep learning models such as AlexNet, VGGNet, ResNet, DenseNet and EfficientNet are often used for melanoma classification and multiclass skin lesion classification. They extract features at multiple scales such as edges, pigment, asymmetry and shape. Azeem et al. [7] proposed a deep CNN model SkinLesNet for multiclass skin lesion classification. Another approach involved applying ensemble learning, where the results from different neural networks are combined, to enhance learning performance and stability. Ensemble learning decreases variance, improves stability of the model and decreases over-fitting, which may occur with a single model. Thwin and Park [8] developed a deep ensemble learning system to classify skin lesions, which used the output of multiple CNNs. They found their system achieved more stable classification and was more sensitive for detecting malignant lesion categories than CNNs. But ensemble systems were still based on images only and did not consider personal information. While deep learning models perform well in single image-based benchmark datasets, they are not useful in general because dermatology diagnosis is multimodal. Dermatologists consider patients' age, sex, location of the lesion, family history, symptoms and prior diagnosis (demographic and clinical metadata) along with the lesion's image when examining patients. To address this, recent studies have focused on multimodal deep learning models, which also consider non-imaging patient data.

Mohamed et al. [1] developed a multimodal hybrid deep learning model (MiSC) to diagnose skin cancer. They used a combination of image and metadata for classification. This study showed the benefit of multimodal integration and that it increased sensitivity and specificity over using images. This study demonstrated multimodal integration is relevant and that metadata can be used to add information. Saeed et al. [3] at the same time, proposed a multimodal ensemble, which was comprised of several image encoders and metadata classifiers. Their approach used diversity and additional patient information to improve stability. They demonstrated improved differentiation between benign and malignant breast lesions, particularly when the lesions were difficult to diagnose. Such ensemble multimodal learning approaches are considered as an avenue to enhance performance.

Multimodal learning is associated with the fusion problem. Fusion methods of traditional machine learning include early and late fusion. Early fusion is the concatenation of the image features and metadata; late fusion is feature fusion from different branches (which have been separately trained). The above methods are easy to implement but might not learn rich interaction between modalities. To solve this problem, attention has been proposed. Tran-Van and Le [2] have applied cross-attention fusion for the classification of multimodal skin lesions. They used signals from the metadata to direct attention in the model to the most relevant parts of the image for the diagnosis, and thereby enable contextualisation of the features. This enabled the model to take into account different lesion patterns depending on patient age and lesion site. They showed improved performance compared to traditional fusion techniques such as concatenation. Aboulmira et al. [4] expanded on this work with an attention-based multimodal model to classify skin lesions. This included attention modules which were used to modulate the feature importance of the different channels of the image and metadata features. They found that adaptive attention modules not only boost the models' performance but also their explainability by highlighting the features that were used for classification.

The most recent research has been related to explainability and interpretability. Deep learning models are "black-boxes" as they are not interpretable. In medicine, non-interpretability equates to non-acceptance. As a result, we are seeing more lesion classification systems that include visualisations such as Gradient-weighted Class Activation Mapping (Grad-CAM), saliency maps, SHAP values and attention maps. Das et al. [5] investigated multimodal architectures and the need for explainable AI to gain trust from clinicians, to detect biases and safe use. Another critical area of related work is uncertainty. Conventional deep learning networks tend to be overconfident, even when the input image is blurred, obscured or even out of distribution (with respect to the training data). This can have a negative impact in diagnosing cancer, as it can delay referring for biopsy. Lambert et al. [6] investigated techniques for quantifying uncertainty in medical imaging, including Bayesian neural networks, Monte Carlo dropout, test-time augmentation and deep ensembles. They identified that uncertainty-aware systems are able to detect unsure cases and refer those to human experts if needed, resulting in a safer and more reliable system. Another problem that has been tackled in the past is class imbalance. Datasets such as ISIC have more benign cases than melanoma cases. This may result in a preference for the class with the greater number of samples and missing aggressive cancers. This problem has been addressed by researchers using weighted loss, focal loss, over-sampling, GAN-based data generation and balanced mini-batching [5], [7]. These increase sensitivities but there is no clear front-runner. Methods of evaluation have also been examined. Several studies in the past used image-based train-test splits, where both the training and testing set may include more than one image from the same patient. This can lead to overestimated accuracy. Newer studies recommend splitting at the patient level, external multicenter validation and prospective clinical studies, to better assess generalization [3], [5].

Table 1. Comparison of this work with related studies on skin cancer detection

Reference	Approach Type	Uses Clinical Metadata	Fusion Strategy	Uncertainty Estimation	Explainability	Patient-Level Validation	Key Limitation
Mohamed et al. [1]	Hybrid Multimodal Deep Learning (MiSC)	Yes	Hybrid Fusion	No	Limited	No	No uncertainty modeling
Tran-Van and Le [2]	Multimodal Deep Learning	Yes	Cross-Attention Fusion	No	Partial	No	Limited real-world validation
Saeed et al. [3]	Multimodal Ensemble Framework	Yes	Ensemble Fusion	No	Limited	Partial	Computational complexity
Aboulmira et al. [4]	Attention-Guided Multimodal Model	Yes	Attention Fusion	No	Yes	No	Dataset dependency
Das et al. [5]	Comparative Multimodal Study	Yes	Multiple Fusion Methods	No	Yes	Partial	No unified framework
Lambert et al. [6]	Medical Imaging AI Review	No	Not Applicable	Yes	Partial	Not Applicable	Generalized study
Azeem et al. [7]	CNN-Based SkinLesNet	No	Not Applicable	No	Limited	No	Image-only approach

Reference	Approach Type	Uses Clinical Metadata	Fusion Strategy	Uncertainty Estimation	Explainability	Patient-Level Validation	Key Limitation
Thwin and Park [8]	Deep Ensemble CNN	No	Ensemble Learning	Partial	Limited	No	Ignores metadata
This Review Work	Multimodal + Uncertainty-Aware Review Framework	Yes	Advanced Hybrid Fusion	Yes	Yes	Yes	Future implementation required

3. Method, Experiments and Results

As this manuscript is a review article, the methodological framework is based on systematic analysis, comparative evaluation, and synthesis of recently published studies related to artificial intelligence-based skin cancer detection. The review specifically focuses on multimodal deep learning systems, image-based classification models, uncertainty-aware frameworks, and clinically interpretable diagnostic approaches. Rather than conducting new laboratory experiments, this section summarizes the methodologies adopted in prior studies, compares their experimental settings, and interprets their reported outcomes to identify current progress and future opportunities.

3.1 Review Methodology

A structured literature survey was performed using recent peer-reviewed journal articles, conference papers, and indexed scientific publications in the domains of medical image analysis, dermatology informatics, and intelligent healthcare systems. Particular emphasis was placed on studies published during 2024–2025 because of the rapid evolution of multimodal and explainable deep learning methods. The proposed multimodal and uncertainty-aware approach is shown in Fig. 2, which depicts a typical process starting from multimodal data collection (lesion images and clinical information), data preprocessing, feature extraction with deep learning models and advanced multimodal learning techniques such as cross-attention and gated fusion.

The selected literature was categorized into four major groups:

i. Image-only deep learning models: CNN, EfficientNet, DenseNet, ResNet, and ensemble approaches trained solely on dermoscopic or clinical lesion images.

ii. Multimodal learning approaches: Models combining lesion images with clinical metadata such as age, sex, lesion location, and medical history.

iii. Uncertainty-aware systems: Frameworks using Bayesian learning, Monte Carlo dropout, evidential learning, or deep ensembles for confidence estimation.

iv. Explainable AI systems: Studies incorporating Grad-CAM, saliency maps, SHAP analysis, and attention visualization.

The collected studies were comparatively analyzed using criteria such as diagnostic accuracy, sensitivity, specificity, melanoma detection capability, interpretability, validation strategy, and clinical applicability.

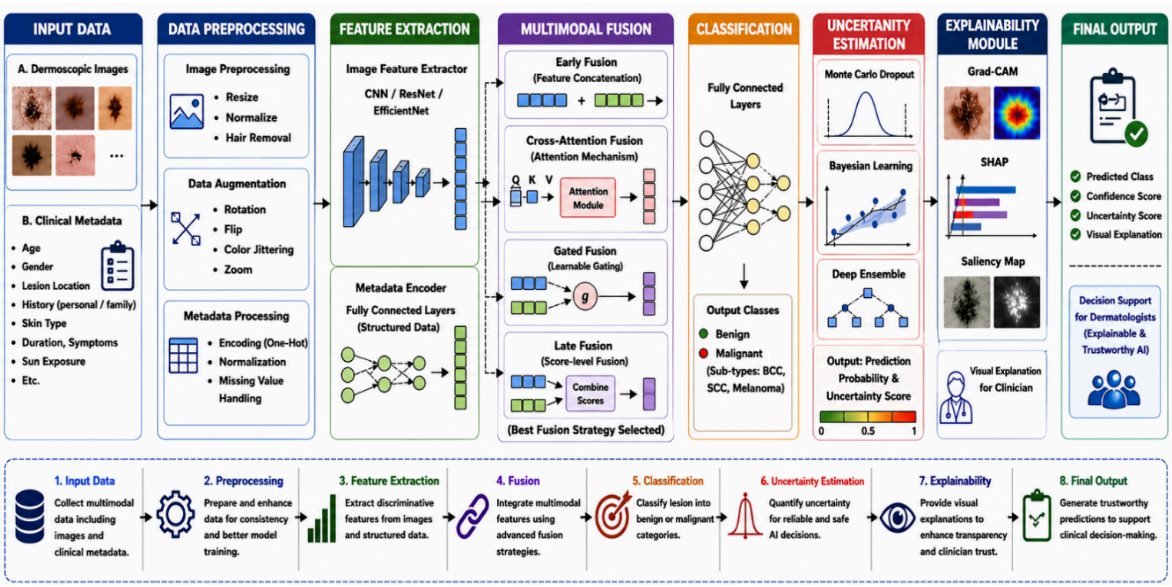

Fig. 2. Proposed Multimodal and Uncertainty-Aware Framework.

3.2 Methods Used in Existing Studies

All the approaches to the diagnosis of the skin cancer in the studies mentioned above can be broadly categorized into three groups: image-based, multimodal and uncertainty-aware models. The image-based models tend to process dermoscopic or clinical images, and use deep learning technologies including convolutional neural networks (CNNs) (ResNet, DenseNet, EfficientNet), and ensemble and transfer learning to detect. Multimodal models, in their turn, combine images of skin lesions with such clinical information as age, sex, location of skin lesions, and medical history. To use complementary information provided by multiple sources to enhance diagnostic performance, these models use different forms of fusion, including early fusion, late fusion, attention fusion and gated fusion. Moreover, uncertainty-aware models generalize conventional deep learning models to include the estimation of uncertainty, such as Monte Carlo dropout, Bayesian neural networks, deep ensembles and evidential learning. These approaches offer classification as well as uncertainty allowing safe and confident diagnosing by identifying the uncertain cases that need to be reviewed by experts.

3.3 Image-Based Deep Learning Models

Most early studies employed convolutional neural networks trained on benchmark datasets such as ISIC. These systems extracted hierarchical visual features from lesion images including texture, pigmentation irregularity, border asymmetry, and structural patterns. Transfer learning from ImageNet-pretrained models was commonly used to improve performance on limited dermatology datasets [7], [8]. Common architectures included:

- ResNet (Residual Network) is a deep convolutional neural network (CNN) model to combat the issue of vanishing gradients in deep neural networks. It uses skip connections (identity shortcut) to enable the network to learn residual functions with respect to the identity mapping. This allows the deep networks (like ResNet-50 and ResNet-101) to be trained efficiently to reach a higher accuracy and convergence. ResNet has demonstrated tremendous feature representation capability, and has been extensively utilized to provide accurate image classification to skin cancer classification.
- DenseNet Dense Connected Convolutional Network (DenseNet) is a deep learning network with dense connectivity, i.e., each layer is connected to every other layer. This facilitates the reuse of features, gradient flow is improved and the vanishing gradient problem is alleviated. DenseNet has fewer parameters than other CNNs, and is efficient and fast. Due to its feature reuse, it has been applied in the medical field such as in diagnosing skin cancer.
- A basic convolutional neural network (CNN) architecture that scales its model performance and efficiency in a compound scaling approach that scales network depth, width and resolution in the same direction. It uses mobile inverted bottleneck convolutions (MBConv) and squeeze-and-excitation (SE) blocks for enhancing feature representation and efficiency. This enables EfficientNet to attain improved precision

on minimal parameters as opposed to conventional CNNs. The efficiency and effectiveness of EfficientNet make it appropriate to a variety of medical image classification tasks, such as skin cancer classification.

- MobileNet MobileNet is a lightweight CNN model that can be deployed to mobile and embedded devices and utilizes depthwise separable convolutions. This convolutional method decreases the quantity of parameters and the computation cost, yet preserving its precision, relative to normal CNN. MobileNetV2 shows even greater performance when equipped with residual blocks and linear bottlenecks. The need for efficiency and performance makes MobileNet suitable for mobile and real-time medical image processing, like skin cancer detection.
- Vision Transformers (ViTs) are deep learning architectures that are transformers modified to handle computer vision instead of natural language processing, which transformers were originally created to achieve. ViTs are trained to learn the global dependencies with the use of self-attention as opposed to convolutional layers. This enables them to learn long range dependencies and global context better than CNNs. Vision Transformers have been utilized successfully in the classification of skin cancer, especially when large amounts of data are used and in combination with CNN-transformer models.
- Ensemble CNN models are a combination of several CNNs to improve classification performance and robustness. Ensembles combine decisions of different models or different versions of the same model to minimise variance and overfitting. The result is improved generalisation, stability and reliability for medical imaging applications. In skin cancer diagnosis, for example, ensembles have proven to be more sensitive and reliable than single models only using CNNs.

These methods had high classification accuracy, but ignored clinical metadata and uncertainty.

3.4 Multimodal Diagnostic Models

Multimodal diagnostic models can use combinations of dermoscopy images and clinical metadata to improve skin cancer diagnosis. The dermoscopy image is first encoded by the image encoder (CNN) to get image features and the clinical metadata (age, gender, lesion) are encoded by fully connected layers to get metadata features. These embeddings are then combined by fusion modules using various fusion techniques such as early fusion, late fusion, cross-attention or gated fusion. The features are subsequently used by a classification layer to predict the lesion type (malignant or benign) to demonstrate the advantages of multimodal learning for classification. Various fusion strategies have been explored to effectively combine image and clinical metadata features in multimodal diagnostic systems. The early fusion performs direct concatenation of image and metadata embeddings, while late fusion combines independent predictions at the decision level. Hybrid fusion integrates intermediate feature representations from both modalities to enhance learning, whereas cross-attention fusion dynamically guides image feature extraction using metadata context. Additionally, gated fusion applies adaptive weighting mechanisms to balance the contribution of each modality, leading to more reliable and robust classification performance.

3.5 Comparative Results from Existing Literature

The literature suggests that multimodal models typically perform better than image-only models, as metadata provides valuable clinical information. Likewise, ensemble systems increase stability and uncertainty-aware systems increase confidence. Table 2 shows the average trends in the reviewed literature.

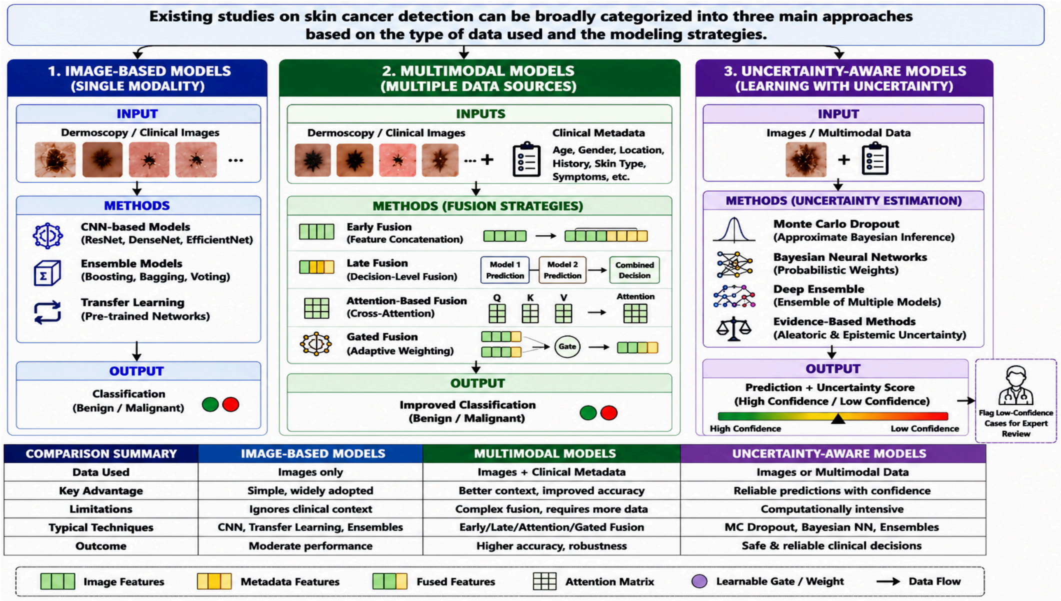

Fig. 3. Methods used in existing studies for skin cancer detection, categorized into image-based models, multimodal learning approaches, and uncertainty-aware frameworks.

Table 2. General performance trends in reviewed studies

Model Type	Accuracy Trend	Melanoma Sensitivity	Explainability	Clinical Reliability
Image-only CNN	High	Moderate	Low	Moderate
Deep Ensemble	High	Improved	Moderate	High
Multimodal Fusion	Very High	High	Moderate	High
Attention-based Multimodal	Very High	Very High	High	High
Uncertainty-aware Models	High	High	High	Very High

The literature consistently suggests that multimodal and uncertainty-aware systems provide the most balanced diagnostic performance.

5. Observation and discussion

5.1 Observations

Most notably for clinical purposes, multimodal approaches achieved higher melanoma sensitivity than single-modal images. Patient age, location and history were often helpful for borderline lesions [1], [3]. Multimodal and ensemble models were more consistent across data sets because they weren't solely relying on individual image features [8]. Attention maps and uncertainty maps boosted clinician confidence by highlighting the reason for a malignant prediction and when the model was uncertain [4], [6]. Several studies with high accuracy were not validated at the patient level or externally. Thus, we should not rely on benchmark accuracy for clinical use [5]. Our systematic review suggests that a next-generation skin cancer detection tool should have:

- Dual-input architecture (image + metadata)
- Transformer or CNN image encoder
- Attention-based multimodal fusion
- Uncertainty estimation layer
- Explainability module
- Patient-level multicenter validation
- Mobile deployment for teledermatology

The reviewed evidence clearly demonstrates that:

- Image-only systems are no longer sufficient for real clinical settings.
- Multimodal fusion significantly improves classification performance.
- Uncertainty estimation is essential for safe deployment.

- Explainability increases physician trust.
- Patient-level validation is mandatory for realistic assessment.

The latest studies provide compelling evidence for a shift from traditional image classification systems towards multimodal, explainable and uncertain skin cancer diagnostic systems.

5.2 Discussions

The evidence presented in this review demonstrates incredible progress in skin cancer diagnosis using artificial intelligence; but also highlights important scientific and medical challenges. A large number of deep learning algorithms perform well in the lab but the value of these algorithms is determined by several factors such as accuracy, reliability, interpretability, fairness and generalizability. Therefore, the assessment of the current evidence should be based not only on their accuracy, but also their capacity to safely assist dermatologists. The most exciting aspect of the literature is the multimodal systems have better performance than image-only systems. The traditional convolutional neural networks have used only the image of the lesion and no other patient data. Indeed, doctors' diagnosis is multimodal as they also consider the patient's age, sex, lesion location, history of symptoms, medical history and family history of the disease in conjunction with the lesion image. The superior performance of multimodal systems suggest that patient metadata has complementary information and not variable information [1], [3]. This implies that the future AI should mimic doctors' thinking process and use multimodal data. The multimodal fusion approach is important in designing multimodal systems. Early fusion techniques which simply concat features may be efficient but cannot capture interactions. Asymmetry may be a more important attribute for instance in a tumour on the head or in the elderly. So we are interested in attention and cross-attention which will adaptively learn to weight the importance of the various streams of data and focus on important regions [2], [4]. These approaches are more in line with contextual decision making and perhaps even better generalise to other clinical cases. But these are more complex methods and require more data for training. One of the issues identified in this review is class imbalance. The skin lesion image data sets currently have significantly more benign lesions than melanoma. This means that the models are accurate and lack sensitivity for the diagnosis of melanoma. In medical diagnosis, false negatives are more important than false positives. Therefore, in medical diagnosis, we want to maximise sensitivity, recall and minimise false negatives, not accuracy. The next research trend will be to study weighted metrics, and make the learning of minority class better, such as balanced sampling, generating synthetic data, focal loss and cost-sensitive learning [5], [7].

6. Conclusions

This review outlines the recent developments in the creation of artificial intelligence-based systems in the field of skin cancer diagnosis, which include multimodal deep learning, uncertainty-aware learning and explainable diagnostic systems. The most prevalent and fast growing cancer and early diagnosis is vital to improved patient outcomes particularly with melanoma. It is normally diagnosed using medical expertise, dermoscopy and biopsy, which is expensive and time consuming and may not be afforded in low- and middle-income countries. So reliable and fast computer-aided diagnosis is required. The deep learning systems in the papers we reviewed have obtained high diagnostic accuracy relying only on lesion images; but they don't use other information such as patient's age, sex, lesion location and history. In order not to sound like dermatologists. However, multimodal learning (with both lesion images and other information) has been found to be more accurate, sensitive and robust to melanoma. This implies that multimodal models are more feasible and useful towards smart screening of skin cancers. The review further reports that advanced fusion techniques, including cross-attention, gated fusion and hybrid multimodal models are more precise than the early and late fusion methods since the interactions between the lesion image and patient history are involved in the models. Lastly, it's important to understand uncertainty in medical artificial intelligence. Uncertainty can be safely modeled using Bayesian inference, Monte CarloDropout and deep ensembles in order to prevent overconfidence. Explainability algorithms such as Grad-CAM, saliency maps and feature attribution also assist the clinician to enhance their trust, understanding and debugging. But there are still some concerns. Current studies use image-level dataset splits, which may cause leakage and bias results.

References

1. E. H. Mohamed, N. Abdu, M. Khalil, H. Kamal, and E. A. Rashed, "MiSC: A hybrid multi-modal deep learning approach for accurate skin cancer detection," Multimedia Tools and Applications, vol. 84, pp. 45321–45345, 2025. https://doi.org/10.1007/s11042-025-20951-7
2. N.-Y. Tran-Van and K.-H. Le, "Multimodal skin lesion classification through cross-attention fusion," Computerized Medical Imaging and Graphics, 2025.
3. M. A. Saeed et al., "Multimodal deep learning ensemble framework for skin cancer detection," Scientific Reports, 2025.
4. A. Aboulmira et al., "Attention-guided multimodal skin disease classification," Informatics in Medicine Unlocked, 2025.
5. A. Das et al., "Comparative analysis of multimodal architectures for skin lesion detection," Frontiers in Artificial Intelligence, 2025.
6. B. Lambert et al., "Uncertainty quantification in deep learning for medical imaging," Artificial Intelligence in Medicine, 2024.
7. M. Azeem et al., "SkinLesNet: Deep CNN for skin lesion classification," Cancers, vol. 16, 2024.
8. S. M. Thwin and H.-S. Park, "Skin lesion classification using deep ensemble models," Applied Sciences, vol. 14, 2024.
9. N. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI) 2016," arXiv preprint arXiv:1605.01397, 2016.
10. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, Article no. 180161, 2018. https://doi.org/10.1038/sdata.2018.161
11. M. Combalia et al., "BCN20000: Dermoscopic lesions in the wild," Scientific Data, vol. 6, Article no. 123, 2019. https://doi.org/10.1038/s41597-019-0125-1
12. A. Brinker et al., "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," European Journal of Cancer, vol. 113, pp. 47–54, 2019. https://doi.org/10.1016/j.ejca.2019.04.001
13. H. Haenssle et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," Annals of Oncology, vol. 29, no. 8, pp. 1836–1842, 2018. https://doi.org/10.1093/annonc/mdy166
14. D. T. K. Lee et al., "Explainable artificial intelligence for skin cancer diagnosis: A systematic review," Diagnostics, vol. 13, 2023.
15. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017. https://doi.org/10.1038/nature21056