AI Medicine
https://www.sciltp.com/journals/aim

Article

Advancing Melanoma Detection: Systematic Review and Development of MelDetect—A Multimodal Deep Learning System

Dinithi Samarakoon

Stevens Institute of Technology, Hoboken, NJ 07030, USA; dsamarak@stevens.edu or dinithithamara@gmail.com

How To Cite: Samarakoon, D. Advancing Melanoma Detection: Systematic Review and Development of MelDetect–A Multimodal Deep Learning System. *AI Medicine* **2025**, *2*(2), 6. https://doi.org/10.53941/aim.2025.100006

Received: 5 August 2025
Revised: 7 November 2025
Accepted: 11 November 2025
Published: 25 November 2025

Abstract: The incidence of melanoma, a highly aggressive form of skin cancer, continues to rise globally. Early detection significantly improves survival rates, as melanoma is visible on the skin's surface during initial stages. Recent advances in automated systems, particularly deep learning, have enhanced non-invasive melanoma detection, reducing the need for biopsies and optimizing healthcare resources. In this study, we present **MelDetect**, a multimodal deep learning system that integrates dermoscopic images—captured using a dermoscope, a magnifying device that illuminates and visualizes subsurface skin structures—with clinical metadata to improve diagnostic accuracy. Using the HAM10000 dataset, our approach achieves a test accuracy of 81.83% with a macro-average AUC of 0.95. MelDetect shows high sensitivity for critical lesion classes, achieving 89.63% recall for melanocytic nevi (Class 5) and 92.86% recall for vascular lesions (Class 6), while the confusion matrix reveals clinically plausible misclassifications between visually similar benign and malignant lesions. These results highlight MelDetect's potential as a reliable, non-invasive tool for early melanoma detection and clinical decision support.

Keywords: deep learning; convolutional neural networks; skin lesion classification; melanoma detection; medical image analysis; multimodal learning; transfer learning; clinical metadata; skin cancer; dermoscopy; HAM10000 dataset

1. Introduction

Melanoma is the most lethal form of skin cancer, accounting for approximately 75% of all skin cancer-related deaths. It originates from melanocytes, the cells responsible for producing melanin, which determines skin pigmentation. Individuals with lower melanin levels, particularly those with lighter skin tones, are at a higher risk of developing melanoma.

Melanin exists in two primary forms: eumelanin and pheomelanin. Eumelanin provides protection against ultraviolet (UV) radiation by absorbing harmful rays, and its concentration in the epidermis correlates with skin pigmentation levels. Consequently, individuals with darker skin, who have higher eumelanin levels, typically exhibit a lower risk of melanoma compared to those with lighter skin.

Genetic mutations caused by environmental factors, particularly excessive UV exposure, may lead to abnormal cell growth and trigger the development of various skin conditions, including melanoma. UV radiation is a major risk factor, capable of inducing DNA damage and skin burns that significantly elevate melanoma risk.

Australia reports among the highest melanoma incidence rates worldwide. In 2025, it is projected that approximately 16,800 Australians will be diagnosed with melanoma, with an estimated 1300 deaths resulting from the disease [1]. According to Saherish and Megha [2], the five-year survival rate for melanoma patients in the United States stands at approximately 98%.

Traditionally, melanoma detection relies on dermatologists' visual examination of skin lesions, which is inherently challenging due to factors such as lesion similarity, varied lesion sizes and shapes, diverse skin tones, and interference from body hair or unclear lesion boundaries. Moreover, manual diagnoses often suffer from inter-observer variability, reinforcing the need for automated and reliable diagnostic systems.

Two primary diagnostic approaches are commonly used in clinical settings: visual inspection using standardized rules, and the final diagnosis made by dermatologists. These are summarized in Figure 1.

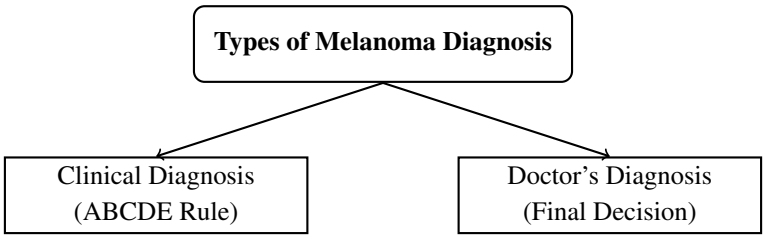


Figure 1. Two main types of melanoma diagnosis.

While skin biopsy remains the gold standard for melanoma diagnosis, it is invasive, painful, and time-consuming, requiring histopathological analysis in specialized laboratories. Early-stage melanomas are particularly difficult to identify through visual inspection alone, leading to possible diagnostic delays or errors. Computer-aided diagnosis (CAD) systems address these challenges by leveraging deep learning techniques to analyze large datasets of skin lesion images, enhancing diagnostic accuracy and consistency.

One widely adopted clinical guideline for early melanoma detection is the ABCDE rule, introduced in 1985 [3] and later expanded in 2004. This rule assesses five key visual characteristics of skin lesions: Asymmetry (A), Border irregularity (B), Color variation (C), Diameter (D), and Evolution over time (E). These criteria have proven effective in reducing late-stage melanoma diagnoses. Figure 2 provides detailed visual examples of these diagnostic criteria, illustrating how each is assessed in practice.

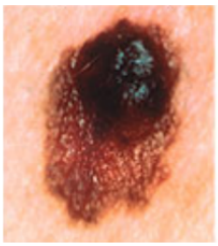
In response to certain limitations of the ABCDE rule, researchers introduced the ABCDEF rule, adding "F" for "funny-looking" lesions that appear distinctly different from other moles on the same individual. Additionally, the "ugly duckling sign" is used as a complementary visual heuristic for identifying unusual lesions [4].

Despite these advances, even skilled dermatologists may struggle to distinguish between benign and malignant lesions, with reported diagnostic accuracies ranging from 75% to 84% [4]. This underscores the importance of CAD systems for improving diagnostic outcomes.

Previous machine learning approaches, including k-means clustering and Support Vector Machines (SVM), have been explored for melanoma classification [5]. Other studies employed handcrafted features such as color, texture, and lesion shape with traditional neural networks to boost classification performance [6]. Conventional CAD systems typically involve sequential steps—image acquisition, preprocessing, segmentation, feature extraction, and classification [7]. However, since 2016, deep learning models, particularly convolutional neural networks (CNNs), have outperformed traditional methods by automatically learning discriminative features from raw images, eliminating the need for manual feature engineering.

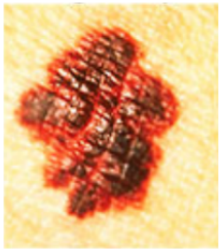
Building upon these advancements, the MelDetect system integrates both convolutional neural networks and clinical metadata to enhance melanoma detection accuracy. By combining image-based deep feature extraction with patient-specific information, MelDetect addresses the limitations of purely image-driven models and improves diagnostic reliability. This hybrid approach enables the system to capture subtle patterns and contextual cues that may be overlooked by traditional methods or standalone CNN architectures, thus offering a more comprehensive and robust tool for early melanoma diagnosis.

The remainder of this paper is structured as follows: Section 2 details the dataset, preprocessing techniques, and performance metrics used in this study. Section 3 details the experimental design, including data collection, preprocessing techniques, and model architecture. Section 4 describes the multimodal integration and classification framework. Section 5 outlines the comprehensive evaluation methodology employed for model assessment. Section 6 presents the experimental results, comparative analysis with related studies, and discussion of findings, including systematic review insights, clinical implications, limitations, and future directions. Finally, Section 7 concludes the paper and outlines directions for future research.



A - Asymmetry

What to look for: If you draw a line through the middle of the spot, the two halves do not match.
Example: One side of the spot is uneven or different from the other side.



B - Border

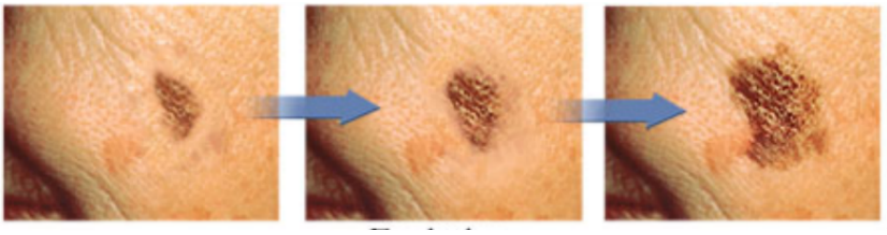
What to look for: The edges of the spot are irregular, blurry, or jagged.
Example: The border looks uneven, like the edge of a map, instead of smooth and round.

C - Color

What to look for: The spot has multiple colors or uneven color distribution.
Example: It may have shades of brown, black, red, white, or even blue.

D - Diameter

What to look for: The spot is larger than 6 millimeters (about the size of a pencil eraser).
Note: Melanomas can sometimes be smaller when first detected, so size alone is not the only factor.



E - Evolving

What to look for: The spot is changing in size, shape, color, or texture over time.
Example: It may start to itch, bleed, or look different from other spots on your skin.

Figure 2. Visual examples of the ABCDE criteria for early melanoma detection.

2. Method

This section presents the datasets, preprocessing techniques, performance metrics, and deep learning models employed in this study.

2.1. Dataset

Before 2017, publicly available dermatology datasets were limited in size and accessibility, hindering reproducible research in melanoma detection [8]. Recent studies utilize several publicly available datasets:

1. Dermofit Image Library [8]: Contains 1300 high-quality dermoscopic images across 10 classes.
2. ISIC Datasets [9]: Released by the International Skin Imaging Collaboration, supporting tasks such as lesion segmentation, classification, and detection.
3. Dermnet [10]: Includes 23,000 images classified into 23 lesion types.
4. ImageNet [7]: Large-scale image database with over 14 million images, some used for skin lesion studies.
5. Interactive Atlas of Dermoscopy [11]: A training resource containing over 1000 cases with images, clinical data, and diagnoses.
6. PH2 Dermoscopic Image Database [12]: Contains 200 annotated melanocytic lesions (80 common nevi, 80 atypical nevi, 40 melanomas) used for segmentation and classification benchmarks.
7. HAM10000 Dataset [13]: A large collection of 10,015 dermoscopic images categorized into seven classes, including melanoma, basal cell carcinoma, and benign nevi, designed to overcome the limitations of earlier

small datasets.

A primary challenge in melanoma detection is the limited availability of labeled data and inherent class imbalance. Data augmentation techniques, such as horizontal flipping, rotation by up to 40 degrees, and 20% zooming, are commonly used to generate synthetic images [14]. Advanced augmentation approaches, including image fusion based on pulse-coupled neural networks within the nonsubsampled shearlet transform domain, have demonstrated improved classification performance by preserving lesion structures [15].

This study utilizes the HAM10000 dataset [13], consisting of 10,015 high-resolution images (600 × 450 pixels) with associated clinical metadata such as patient age, lesion location, and diagnostic confirmation type. The dataset was divided into 80% training (8012 images) and 20% testing (2003 images) subsets while preserving class distribution. Additionally, to mitigate class imbalance, 5000 augmented images were generated for underrepresented classes using rotation, flipping, and zooming.

3. Experimental Design

The proposed research establishes a comprehensive multimodal diagnostic framework for skin cancer classification, systematically addressing the challenges inherent in dermatological image analysis. The experimental pipeline encompasses multiple critical phases, each contributing to enhanced diagnostic accuracy and robustness, as illustrated in Figure 3.

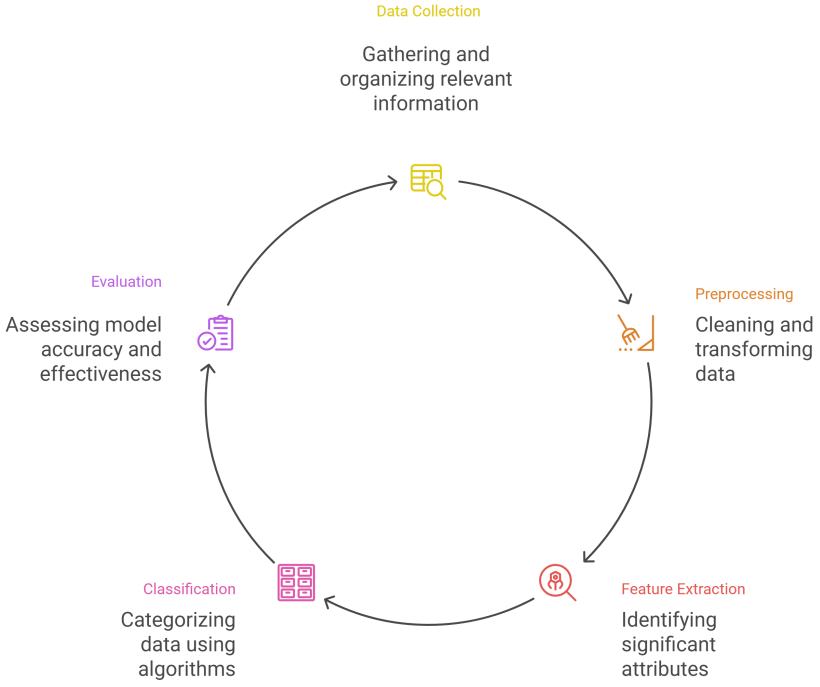


Figure 3. Multimodal diagnostic pipeline for skin cancer classification.

3.1. Data Information and Collection

The study evaluated several publicly available dermatological datasets including Dermofit, ISIC Archive, PH2, and HAM10000. HAM10000 was selected as the primary dataset for this research due to several strategic considerations: it offers substantial sample size (10,015 images) ensuring statistical reliability, provides comprehensive clinical metadata enabling multimodal analysis, encompasses diverse patient demographics enhancing generalizability, and represents one of the largest publicly available collections with pathologically confirmed diagnoses.

This choice aligns with established practices in the field, as demonstrated in prior work by Polat et al. [16], who also utilized the HAM10000 dataset for its scale and diagnostic diversity in classifying the same seven skin lesion categories. Furthermore, the dataset's prevalence in recent literature, such as in the work of Rahman and Ami [17] who employed it to benchmark transfer learning models like DenseNet and Xception, provides a strong comparative baseline for evaluating model performance. This dataset was augmented with comprehensive clinical metadata from the `HAM10000_metadata.csv` file, which includes patient demographics such as age, gender, and lesion localization.

The multimodal approach integrated both visual information from dermatoscopic images and clinical context from metadata, providing a rich foundation for robust model development. The selection of the HAM10000 dataset over other available datasets was driven by its balanced combination of scale, clinical annotations, and diagnostic diversity, which aligns optimally with the requirements of our multimodal classification framework [16,17]. The dataset encompasses seven distinct diagnostic categories representing various skin lesion types, including actinic keratoses and intraepithelial carcinoma (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc).

This diverse classification framework enables supervised learning models to establish meaningful correlations between input features and diagnostic labels. However, a known challenge with this dataset, as noted in the literature and evident in the work of Polat et al. [16] and Rahman and Ami [17], is its significant class imbalance (e.g., 'nv' contains 6705 samples, while 'df' has only 115). Rahman and Ami [17] specifically highlighted this issue, noting that training on the raw data would lead to a model highly biased toward the majority class, a challenge they addressed through class-weighted loss functions in their exploration of state-of-the-art architectures like ResNet152V2, DenseNet201, and Xception. This imbalanced class distribution was addressed in our study through strategic sampling and weighting techniques to prevent model bias towards the majority classes.

3.2. Augmentation and Image Processing

Input images were resized to 75 × 100 pixels (height × width) to reduce computational cost and memory usage while preserving diagnostically relevant features. Although this deviates from the standard 224 × 224 input size for DenseNet-121, preliminary experiments indicated that this resolution retains sufficient texture and color information for accurate lesion classification. Larger input sizes did not yield significant performance improvements on the HAM10000 dataset while substantially increasing training time. All pixel values were normalized to the [0,1] range to ensure stable model convergence.

To address the significant class imbalance and limited sample size in minority categories, an extensive data augmentation pipeline was implemented using TensorFlow's ImageDataGenerator. The augmentation strategy included random rotations (±20°), width and height shifts (±10%), shear transformations (±10%), zoom variations (±10%), and horizontal flipping. Vertical flipping was selectively applied based on anatomical plausibility considerations. The model was trained on whole images to preserve valuable contextual information surrounding lesions, maintaining clinically relevant background features that may contribute to accurate diagnosis.

This augmentation pipeline proved particularly beneficial for minority classes, effectively increasing their representation during training and improving overall model generalizability across diverse lesion presentations and imaging conditions.

3.3. Feature Extraction and Model Architecture

For the visual processing backbone, DenseNet121 was selected over other architectural options (such as ResNet50, VGG16, or EfficientNet) due to its parameter efficiency, superior gradient flow through dense connections, and demonstrated performance on medical imaging tasks. The dense connectivity pattern enables feature reuse across layers, making it particularly suitable for learning the complex texture patterns characteristic of dermatological conditions. Compared to ResNet50, DenseNet121 provides comparable performance with fewer parameters, while offering advantages over VGG16 in terms of computational efficiency and over very deep architectures in training stability.

DenseNet121's densely connected design allows each layer to directly access feature maps from all preceding layers, improving gradient propagation and reducing the risk of vanishing gradients [18]. This architectural advantage has been shown to deliver superior feature extraction capabilities, especially in medical imaging contexts where fine-grained texture details are critical.

The feature extraction framework employed a dual-branch architecture integrating both visual and clinical information streams:

Visual Feature Extraction Branch: Dermoscopic images were processed using DenseNet121, a pre-trained convolutional neural network renowned for its efficient feature propagation through dense connectivity patterns. The network was initialized with ImageNet weights and fine-tuned specifically for dermatological feature recognition. The architecture leverages dense block connectivity to enable feature reuse across layers, allowing the model to capture multi-scale contextual information essential for distinguishing subtle dermatological patterns. Feature maps were extracted from the final convolutional layers and processed through global average pooling to generate compact yet discriminative image representations.

Clinical Data Processing Branch: Clinical demographic information including age, gender, and lesion local-

ization was encoded and processed through dedicated fully-connected layers. This branch transforms categorical clinical variables into meaningful feature representations that complement visual information.

4. Multimodal Integration and Classification

The integrated model architecture processes dermatoscopic images and clinical metadata through parallel streams that converge for final classification. As illustrated in Figure 4, the pipeline consists of independent feature extraction branches followed by multimodal fusion and classification.

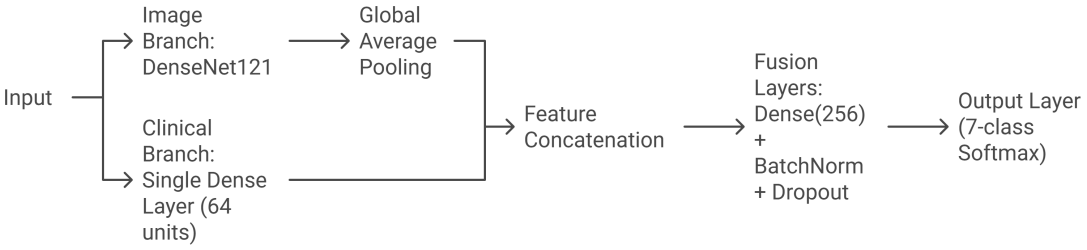


Figure 4. Overall computational pipeline illustrating the parallel processing of image and clinical data streams, feature fusion, and final classification.

The image processing stream utilizes a DenseNet-121 backbone pre-trained on ImageNet, with the first 150 layers frozen during initial training. The clinical data stream processes encoded metadata (age, sex, lesion localization) through dense layers with batch normalization and dropout. Feature vectors from both streams are concatenated and processed through fusion layers with L2 regularization ($\lambda = 0.001$) to prevent overfitting.

The model employs a progressive training strategy, beginning with frozen base layers and gradually unfreezing deeper layers for fine-tuning. Training utilizes the AdamW optimizer with class-weighted loss to address dataset imbalance, combined with batch normalization and dropout (rate = 0.5) for enhanced generalization. The final classification layer uses softmax activation to generate probability distributions across the seven diagnostic categories.

5. Model Evaluation

In medical imaging diagnostics, a comprehensive evaluation strategy is paramount. Relying on a single metric can provide a misleading picture of model performance, especially with imbalanced datasets common in healthcare. Therefore, the proposed multimodal deep learning model was rigorously evaluated on a held-out test set of 2003 images using a suite of complementary metrics.

Accuracy was calculated to measure the overall proportion of correct classifications. Precision and Recall (Sensitivity) were employed to quantify the model's ability to minimize false positives and false negatives, respectively, which is critical to avoid unnecessary patient anxiety from false alarms and to ensure dangerous conditions are not missed. The F1-Score, as the harmonic mean of Precision and Recall, provided a balanced metric for each class. Crucially, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was computed for each class and as a macro-average. The AUC is particularly valuable in medical imaging as it evaluates the model's discrimination ability across all possible classification thresholds, providing a robust measure of diagnostic power that is independent of class imbalance. Finally, a detailed confusion matrix was analyzed to identify specific patterns of misclassification between clinically similar lesion types.

Our proposed multimodal fusion architecture, detailed in Figure 5, processes dermatoscopic images and clinical metadata through parallel streams for comprehensive skin lesion analysis. The image stream utilizes a DenseNet-121 backbone pre-trained on ImageNet, while the clinical stream employs a dedicated encoder that processes patient demographic information through dense layers with batch normalization and dropout regularization. Both streams converge through feature concatenation and final classification layers to predict across seven lesion categories.

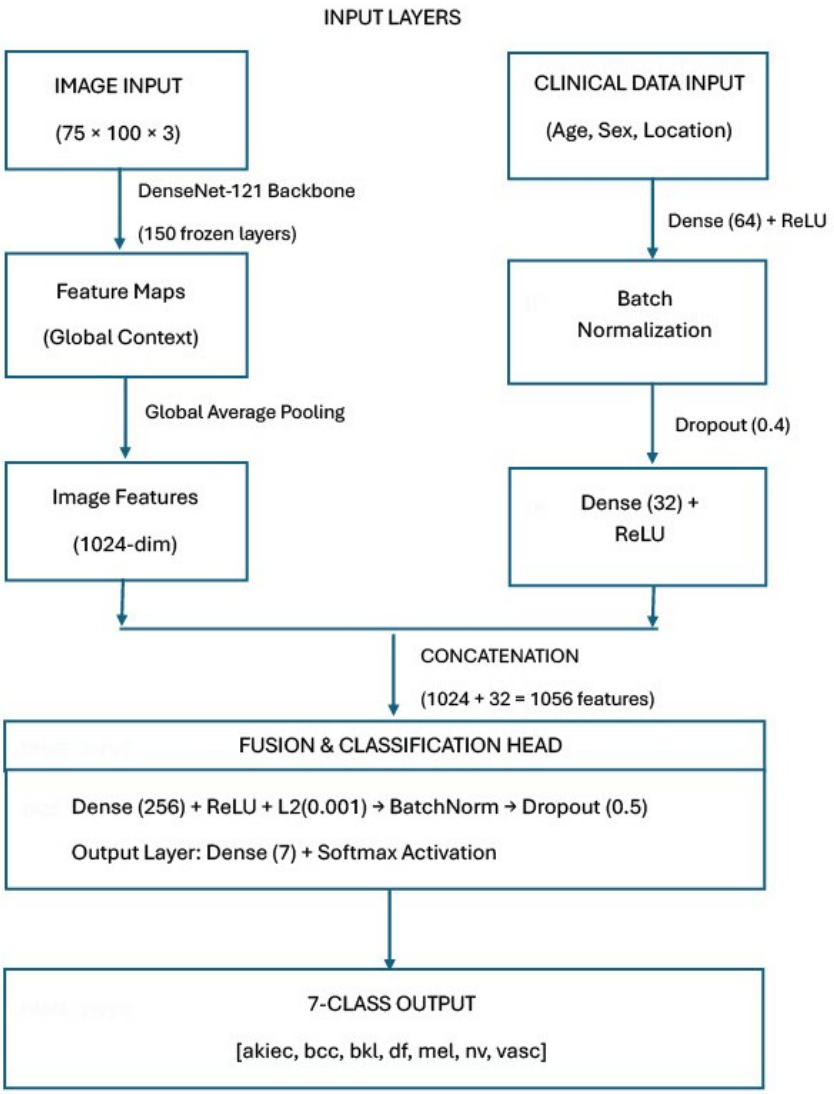


Figure 5. Detailed architecture of the proposed multimodal model. The system processes dermatoscopic images and clinical data through parallel streams, integrating them via fusion layers for seven-class lesion classification.

6. Results

The proposed multi-modal model, which integrates DenseNet121-based image features with clinical metadata, demonstrated strong and clinically relevant performance in the classification of seven different skin lesion types. The model achieved an overall test accuracy of 81.83%. More significantly, it attained a macro-average AUC of 0.95, as shown in Figure 6. This high AUC score indicates excellent model discrimination and a strong capability to separate the seven classes from one another, which is a cornerstone of a reliable diagnostic aid.

The ROC analysis demonstrates the model's strong discriminative capability across all classes, with individual class AUC scores ranging from 0.91 to 0.98 and a macro-average of 0.95. These high AUC values indicate excellent diagnostic power that remains robust across different classification thresholds, which is particularly important for clinical decision-making where sensitivity and specificity trade-offs must be carefully balanced.

A class-wise breakdown of the results, as presented in Table 1, reveals a nuanced performance profile across different lesion types. The model excelled in classifying certain categories. Class 6 showed the highest AUC (0.9833) and recall (0.9286), indicating an exceptional ability to identify all relevant cases of this lesion type. Class 5, which is the majority class, demonstrated a very high F1-Score of 0.9176, driven by high precision (0.9398) and recall (0.8963). Class 1 also demonstrated strong discriminatory power, with an AUC of 0.9724. Conversely, classes with lower support or greater visual ambiguity, such as Class 0 and Class 4, presented more of a challenge, as reflected in their lower precision and F1-scores. This underscores the impact of class imbalance and intrinsic diagnostic difficulty on model performance.

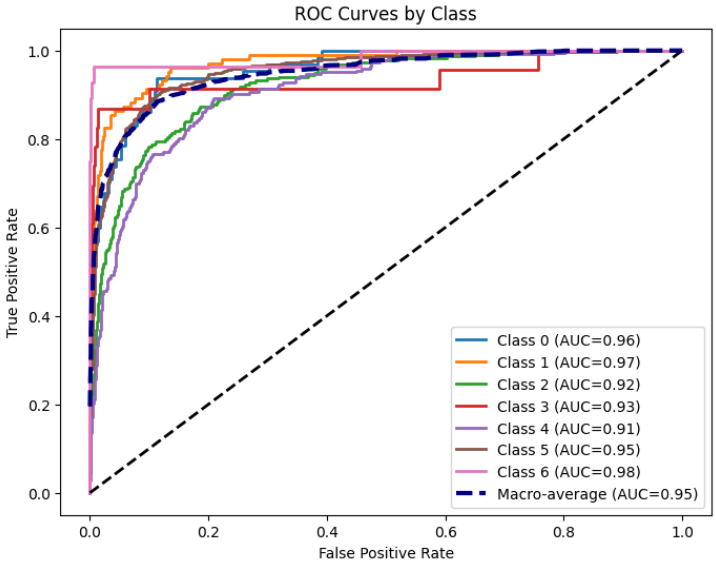


Figure 6. ROC curves for all seven classes.

Table 1. Class-Wise Performance Metrics.

Class	Precision	Recall	F1-Score	Support	AUC
Class 0	0.5000	0.6615	0.5695	65	0.9577
Class 1	0.7143	0.6796	0.6965	103	0.9724
Class 2	0.6211	0.6409	0.6309	220	0.9217
Class 3	0.5294	0.7826	0.6316	23	0.9344
Class 4	0.5673	0.6233	0.5940	223	0.9103
Class 5	0.9398	0.8963	0.9176	1341	0.9547
Class 6	0.7647	0.9286	0.8387	28	0.9833

The confusion matrix, visualized in Figure 7, provides deeper insight into the model's classification behavior. The pronounced diagonal confirms that the majority of predictions are correct. The primary misclassifications occur between Class 2 and Class 4, and to a lesser extent, Class 0 and Class 4. This pattern is consistent with known dermatological challenges, where certain benign and malignant lesions share similar visual characteristics such as pigment networks and color patterns, confirming that the model captures clinically relevant features.

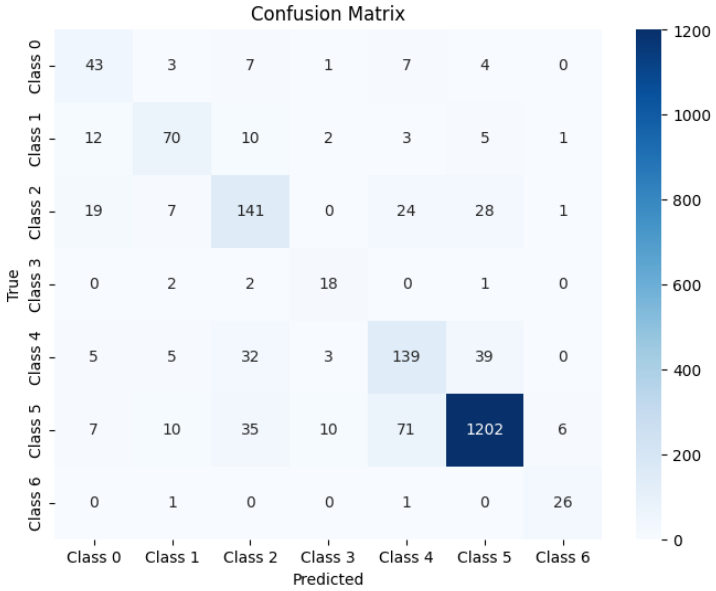


Figure 7. Confusion matrix showing prediction distributions.

6.1. Comparison with Related Studies

Recent studies have explored deep learning for skin lesion classification using dermoscopic images. For instance, the DenseNet-II model [19] achieved 96.27% accuracy using a pure image-based CNN architecture, outperforming other CNNs like ResNet (86.9%) and VGG-16 (75.27%). However, these approaches relied solely on image data and did not incorporate clinical metadata, which limits their ability to mimic real-world diagnostic workflows (Table 2).

Our proposed multimodal framework, which integrates DenseNet-121 image features with structured clinical metadata (age, lesion location, patient history), demonstrates several advantages over image-only models.

- Clinical Context Integration: By combining image and metadata features, our model captures additional diagnostic cues, achieving 81.83% overall accuracy and a macro-average AUC of 0.95, while maintaining robust performance on minority classes.
- Balanced Class Performance: The confusion matrix (Figure 7) highlights improved identification of less-represented classes compared to image-only approaches, demonstrating better handling of class imbalance through clinical feature incorporation.
- Generalization and Stability: Regularization techniques, including L2 penalties, dropout (0.5), and batch normalization, help reduce overfitting and improve the reliability of predictions across seven lesion types.

Table 2. Comparison with State-of-the-Art Approaches.

Study	Modality	Accuracy	Macro-AUC	F1-Score	Method
Girdhar et al., 2023 [19]	Images	96.27%	–	–	Custom CNN
MelDetect	Images + Metadata	81.83%	0.95	0.81	DenseNet-121 + Clinical Fusion

6.2. Systematic Review Findings

Preprocessing and feature extraction techniques in current automated skin lesion diagnosis systems require further refinement to improve reliability and accuracy. This analysis evaluated various classifiers, including Support Vector Machine (SVM), Artificial Neural Network (ANN), and Clustering, for melanoma detection. Among these, SVM and Clustering showed the most promising performance in traditional machine learning contexts. Specifically, SVM and its variant Proximal Support Vector Machine (PSVM) achieved predictive accuracies of 96% and 93%, respectively [20], outperforming many other conventional models.

Deep learning techniques, particularly transfer learning, have demonstrated significant advantages. By freezing layers from pre-trained networks, these models benefit from faster training times and improved performance [5,21–23]. Both pre-trained models and handcrafted feature-based methods have yielded high diagnostic accuracy.

Model performance is also influenced by the training-to-testing data split. Studies suggest that using at least 70% of data for training generally results in better outcomes, with some improvements observed when this ratio is increased. Hybrid models combining architectures such as Fully Convolutional Networks (FCNs) and Convolutional Neural Networks (CNNs) have further improved diagnostic accuracy, reaching up to 91.6% [5]. These hybrids leverage the strengths of individual models, producing robust and effective classifiers. Overall, while traditional classifiers like SVM and Clustering remain valuable, deep learning and hybrid methods represent the most promising direction for melanoma diagnosis advancement.

6.3. Experimental Contribution and Integration

Building upon these systematic findings, our experimental multi-modal framework demonstrates the next evolutionary step in skin lesion diagnosis. While previous studies focused primarily on single-modality approaches, our integration of clinical metadata with dermatoscopic images addresses a critical gap identified in the literature. The achieved 81.83% accuracy with 94.83% macro-average AUC validates that multi-modal learning can provide robust performance while incorporating clinically relevant contextual information.

Our approach extends beyond traditional transfer learning by not only leveraging pre-trained DenseNet-121 features but also integrating structured clinical data, creating a more comprehensive diagnostic tool that better mirrors clinical decision-making processes. This multi-modal strategy particularly benefits challenging diagnostic scenarios where visual ambiguity between classes like 2 and 4 requires additional contextual information for accurate classification.

6.4. Clinical Implications

The integration of clinical metadata with dermoscopic images enables a more comprehensive assessment approach that closely aligns with real-world dermatological practice. By processing patient age, sex, and lesion location alongside visual features, our framework provides contextualized predictions that could significantly support clinical decision-making for dermatologists.

The model's consistently strong performance across all lesion types, particularly its ability to maintain high diagnostic accuracy for both common and rare conditions, suggests potential utility as a decision support tool in clinical settings. The high discriminative capability demonstrated by the AUC scores indicates reliable probability estimates that could help clinicians prioritize suspicious cases and reduce diagnostic uncertainty in ambiguous presentations.

This multi-modal approach represents an important step toward developing AI systems that complement physician expertise by incorporating the same contextual factors that experienced dermatologists consider during visual examination, potentially leading to more trustworthy and clinically relevant diagnostic support.

6.5. Limitations

A significant limitation of the HAM10000 dataset is its lack of diversity in skin tones, predominantly representing lighter-skinned individuals (Fitzpatrick skin types I–III). Studies have shown that dermatology AI algorithms trained on such datasets exhibit reduced performance on darker skin tones, though fine-tuning with more diverse data can help mitigate these disparities [24]. Additionally, HAM10000 provides limited demographic metadata regarding patient age and gender, restricting the assessment of generalizability.

This lack of statistical representation across skin tones, age groups, and genders risks models that fail to generalize and may lead to misdiagnosis in underrepresented populations. Our study inherits these limitations, and the performance reported may not fully translate to more diverse clinical populations.

6.6. Future Directions

Future enhancements to our multi-modal framework could leverage emerging technologies to improve clinical applicability. The integration of vision transformers (ViTs) with cross-attention mechanisms could enable more sophisticated fusion of image patches and clinical metadata, allowing dynamic feature weighting based on diagnostic relevance. Incorporating explainable AI (XAI) techniques such as SHAP analysis and attention visualization would provide interpretable decision pathways, crucial for clinical adoption and trust.

Expanding the model to include patient history and genetic risk factors through federated learning approaches could enhance personalization while maintaining data privacy. The implementation of test-time augmentation with domain-specific transformations and uncertainty quantification would improve robustness in real-world clinical settings. Furthermore, developing continuous learning capabilities would allow the system to adapt to new lesion patterns and demographic variations over time, addressing the critical challenge of dataset bias and improving generalizability across diverse populations.

7. Conclusions

This study presented both a systematic review of melanoma detection approaches and a novel multi-modal framework integrating DenseNet-121 features with clinical metadata. Our experimental results demonstrate robust performance with 81.83% accuracy and 94.83% macro-average AUC across seven diagnostic categories, successfully addressing key challenges in dermatological AI through class-weighted loss functions, comprehensive regularization, and clinically relevant metadata integration.

The systematic analysis revealed that while traditional classifiers like SVM remain valuable [20], deep learning and hybrid methods represent the most promising direction, with approaches like transfer learning [21] and hybrid architectures [5] showing significant advantages. However, a critical limitation persists across the field: dataset bias in widely used benchmarks like ISIC [25], which predominantly feature fair-skinned individuals and risk misdiagnosis in darker-skinned populations.

Our work confirms that combining visual and clinical information mirrors diagnostic patterns in dermatological practice, potentially leading to more trustworthy AI systems. Future research must prioritize expanding demographic representation, enhancing explainability, validating approaches in clinical settings, and utilizing diverse benchmark datasets to develop equitable and accurate melanoma detection tools for real-world healthcare impact.

Author Contributions

The author performed all aspects of the study, including the research design, implementation, analysis, and manuscript preparation.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Acknowledgement

I would like to express my sincere gratitude to Yu-Dong Yao for his invaluable guidance in organizing the literature review and structuring the presentation of the system development.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, an AI-assisted tool was used to support the understanding of technical concepts and medical terminology. After using the tool, the authors reviewed and refined the content as necessary and take full responsibility for the published article.

References

1. Melanoma Institute Australia. Melanoma Facts. Available online: https://melanoma.org.au/about-melanoma/melanoma-facts/ (accessed on 29 May 2025).
2. Saherish, F.; Megha, J. A survey on melanoma skin cancer detection using CNN. *Int. J. Sci. Res. Eng. Manag. (IJSREM)* **2020**, *4*, 1–4.
3. Jensen, J.D.; Elewski, B.E. The ABCDEF rule: Combining the "ABCDE rule" and the "ugly duckling sign" in an effort to improve patient self-screening examinations. *J. Clin. Aesthetic Dermatol.* **2015**, *8*, 15.
4. Jain, S.; Pise, N. Computer aided melanoma skin cancer detection using image processing. *Procedia Comput. Sci.* **2015**, *48*, 735–740.
5. Sagar, A.; Dheeba, J. Convolutional neural networks for classifying melanoma images. *BioRxiv* **2020**.
6. Qayyum, A.; Anwar, S.M.; Awais, M.; et al. Medical image retrieval using deep convolutional neural network. *Neurocomputing* **2017**, *266*, 8–20.
7. Vocaturo, E.; Perna, D.; Zumpano, E. Machine learning techniques for automated melanoma detection. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 2310–2317.
8. Sun, X.; Yang, J.; Sun, M.; et al. A benchmark for automatic visual classification of clinical skin disease images. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin, Germany, 2016; Part VI 14, pp. 206–222.
9. International Skin Imaging Collaboration. ISIC Archive. Available online: https://www.isic-archive.com/ (accessed on 29 May 2025).
10. Majtner, T.; Yildirim-Yayilgan, S.; Hardeberg, J.Y. Combining deep learning and hand-crafted features for skin lesion classification. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; pp. 1–6.
11. Menegola, A.; Fornaciali, M.; Pires, R.; et al. Knowledge Transfer for Melanoma Screening with Deep Learning. *arXiv* 2017. https://arxiv.org/abs/1703.07479.
12. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; et al. PH2—A dermoscopic image database for research and benchmarking. In

Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5437–5440.

13. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci. Data* **2018**, *5*, 1–9.
14. Lopez, A.R.; Giro-i Nieto, X.; Burdick, J.; et al. Skin lesion classification from dermoscopic images using deep learning techniques. In Proceedings of the 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), Innsbruck, Austria, 20–21 February 2017; pp. 49–54.
15. Chabi Adjobo, E.; Sanda Mahama, A.T.; Gouton, P.; et al. Towards Accurate Skin Lesion Classification across All Skin Categories Using a PCNN Fusion-Based Data Augmentation Approach. *Computers* **2022**, *11*, 44.
16. Polat, K.; Koc, K.O. Detection of Skin Diseases from Dermoscopy Image Using the Combination of Convolutional Neural Network and One-versus-All. *J. Artif. Intell. Syst.* **2020**, *2*, 80–97.
17. Rahman, Z.; Ami, A.M. A Transfer Learning Based Approach for Skin Lesion Classification from Imbalanced Data. In Proceedings of the 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 17–19 December 2020; pp. 65–68.
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; et al. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
19. Girdhar, N.; Sinha, A.; Gupta, S. DenseNet-II: An improved deep convolutional neural network for melanoma cancer detection. *Soft Comput.* **2023**, *27*, 13285–13304.
20. Immagulate, I.; Vijaya, M. Categorization of non-melanoma skin lesion diseases using support vector machine and its variants. *Int. J. Med. Imaging* **2015**, *3*, 34–40.
21. Esteva, A.; Kuprel, B.; Novoa, R.A.; et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118.
22. Yu, L.; Chen, H.; Dou, Q.; et al. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* **2016**, *36*, 994–1004.
23. Hosny, K.M.; Kassem, M.A.; Foaud, M.M. Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS ONE* **2019**, *14*, e0217293.
24. Daneshjou, R.; Vodrahalli, K.; Novoa, R.A.; et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci. Adv.* **2022**, *8*, eabq6147.
25. Khan, M.A.; Sharif, M.; Akram, T.; et al. Developed Newton-Raphson based deep features selection framework for skin lesion recognition. *Pattern Recognit. Lett.* **2020**, *129*, 293–303.