SKIN DISEASE PREDICTION USING MACHINE LEARNING

[1] Ms. I. Kavitha, [2] Mr. D. Dharish Prasath, [3] Ms. A. Faziya, [4] Mr. S. Jai Ganesh, [5] Ms. M. KalaiArasi
[1] Assistant Professor, Department of Information Technology
[2, 3, 4, 5] UG Students, Department of Information Technology
[1, 2, 3, 4, 5] SRM Valliammai Engineering College, Kattankulathur, Tamil Nadu, India.
Corresponding author: kavithai.it@srmvalliammai.ac.in

***Abstract*– Melanoma, a severe skin cancer, is strongly linked to UV-induced DNA damage and genetic susceptibility. Ecoskin AI integrates U-Net for accurate lesion segmentation and Vision Transformer (ViT) for global pattern recognition, combining dermoscopic images, UV exposure data, and genetic signatures. This fusion improves melanoma detection, risk prediction, and interpretability through explainable AI. Results show higher accuracy and sensitivity than single-model methods, offering a scalable tool for early diagnosis and personalized prevention.**

Keywords: Melanoma, UV exposure, genetics, U-Net, Vision Transformer, AI detection.

I. INTRODUCTION

Skin problems are very common and can affect people of any age. Some conditions are simple and temporary, while others can be serious and dangerous if not treated early. Diseases like Melanoma and Basal Cell Carcinoma are types of skin cancer that require early detection to save lives. Many other skin conditions, such as infections or allergies, also need proper diagnosis to avoid complications.

Usually, dermatologists examine the skin visually and may use special tools or tests to confirm the disease. However, diagnosis can sometimes take time and depends heavily on the doctor’s experience. In rural or remote areas, people may not have easy access to skin specialists, which can delay treatment.

This is where Machine Learning plays an important role. Machine Learning is a part of artificial intelligence that allows computers to learn from data and make predictions. In skin disease prediction, the system is trained using many images of different skin conditions. It learns patterns such as color changes, texture differences, and lesion shapes. Once trained, the model can analyze a new skin image and predict the possible disease.

The process usually includes image preprocessing to improve quality, feature extraction to identify important characteristics, and classification using algorithms like Support Vector Machine (SVM), Random Forest, or Convolutional Neural Networks (CNN). These models help in identifying whether a skin lesion is harmful or not.

Overall, skin disease prediction using machine learning can support doctors by providing faster and more accurate results. It helps in early detection, reduces manual effort, and improves healthcare accessibility. By combining medical knowledge with advanced technology, this system can contribute to better diagnosis and timely treatment for patients.

II. RELATED WORKS

Melanoma detection and prevention have been widely studied using artificial intelligence, bioinformatics, and genetic analysis. Several researchers have explored the integration of deep learning and environmental factors to improve diagnostic accuracy. This section highlights key related works that form the foundation for the proposed Ecoskin AI framework:

The relationship between UV radiation, DNA damage, and melanoma development has been extensively studied, and recent advances in artificial intelligence (AI) have enabled deeper exploration of these mechanisms. Several related works form the scientific foundation for the proposed ECOSKIN AI system.

Brash et al. [1] – This study highlighted the role of UV-induced DNA mutations, particularly C→T and CC→TT transitions, as the mutational hallmarks of melanoma. Their findings established UV radiation as a primary carcinogenic factor and provided the basis for linking environmental exposure with genetic damage.

Alexandrov et al. [2] – The authors systematically identified and cataloged mutational signatures across different cancers, including UV-induced melanoma. Their work reinforced the connection between environmental mutagens and distinct genomic alterations, thereby supporting the inclusion of mutation pattern analysis in melanoma risk assessment.

Esteva et al. [3] – In their landmark study published in Nature, deep learning models were shown to classify skin cancers with dermatologist-level accuracy. This work demonstrated the potential of AI-driven image analysis in clinical dermatology and paved the way for integrating CNN-based frameworks in melanoma detection.

Tschandl et al. [4] – The authors used deep learning on large-scale dermoscopic image datasets (e.g., ISIC archive) and showed that AI systems could outperform average dermatologists in melanoma recognition. Their findings validated the feasibility of deploying AI-assisted lesion classification in real-world settings.

Xie et al. [5] – This research applied a U-Net based architecture for automated skin lesion segmentation, reporting significant improvements in Dice scores and IoU metrics compared to traditional approaches. Their work highlighted the robustness of U-Net in handling irregular lesion boundaries, making it an ideal choice for ECOSKIN AI.

Yu et al. [6] – The authors proposed a multimodal AI framework that integrated dermoscopic images with patient metadata such as skin type and age. Their findings revealed that combining clinical and non-visual data significantly enhanced predictive accuracy, aligning with ECOSKIN AI's objective of fusing image, UV, and genomic information.

Mar et al. [7] – Their genomic investigations identified key driver mutations (BRAF, NRAS, TP53) in melanoma, and showed how these mutations interact with UV-induced DNA signatures. This underscores the importance of integrating genetic biomarkers into AI-driven melanoma risk prediction systems.

III. EXISTING WORKS

The existing system for skin disease detection mainly relies on manual clinical examination by dermatologists. Diagnosis is performed through visual inspection and dermoscopic analysis of skin lesions. In many cases, biopsy and laboratory testing are required to confirm serious diseases such as skin cancer. Traditional methods depend heavily on the doctor's experience and expertise. The process is time-consuming and may lead to misdiagnosis in early stages. In rural or remote areas, lack of specialists further delays proper diagnosis and treatment.

Conventional computer-based systems use basic image processing techniques and simple machine learning algorithms. These systems require manual feature extraction such as color, texture, and shape analysis. Accuracy is often limited due to poor image quality and small datasets. Many models suffer from overfitting and cannot generalize well to new data. The existing approaches do not provide real-time, highly accurate predictions. Hence, there is a need for an advanced machine learning-based automated system for reliable skin disease prediction.

Existing computer-based systems use basic image processing and traditional machine learning techniques that require manual feature extraction, which limits accuracy and scalability. These systems also struggle with large datasets, varying image quality, and imbalanced data. As a result, the current system lacks automation, consistency, fast processing, and high accuracy, creating the need for an advanced machine learning-based skin disease prediction system.

IV. PROPOSED SYSTEM

The proposed system is designed to provide a comprehensive framework for the early detection and risk assessment of melanoma by integrating UV exposure data, genetic pattern analysis, and advanced AI-driven segmentation techniques. It leverages the power of the U-Net algorithm for precise skin lesion segmentation, combines it with genomic markers and environmental UV exposure indices, and delivers a user-friendly risk evaluation platform for clinicians and patients.

Core Components:

1. Image Segmentation (U-Net Module):
The U-Net algorithm is employed to automatically segment skin lesions from dermoscopic images with high accuracy. It captures lesion boundaries, irregular shapes, and color variations, providing clear lesion masks for further analysis.

2. Genetic Pattern Analysis:
The system integrates genomic data (mutations such as BRAF, NRAS, and p53) and UV-induced mutational signatures (C→T, CC→TT transitions). These features are extracted and analyzed to assess genetic predisposition and potential vulnerability to UV-related skin damage.

3. UV Exposure Tracking:
UV data is collected through environmental UV index records, wearable sensors, or patient self-reports. Cumulative exposure levels are calculated, normalized by skin type, and linked with lesion risk factors to provide a holistic view of environmental influence.

4. Risk Prediction & Analytics:
The segmented lesion features, genetic patterns, and UV exposure history are fused into a predictive AI model. The system generates a personalized melanoma risk score, highlights key contributing factors, and provides trend analytics for ongoing monitoring.

5. User-Friendly Interface:

The platform includes a visually intuitive dashboard where clinicians and users can view segmented lesion images, UV timelines, and genetic risk summaries. Graphical charts, heatmaps, and probability scores enhance interpretability. Real-time alerts notify users of high-risk patterns, excessive UV exposure, or suspicious lesion changes.

V. ARCHITECTURE DIAGRAM

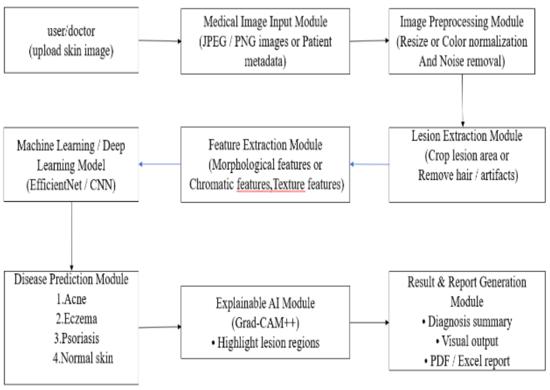

FIG 1: ARCHITECTURE DIAGRAM

VI. MODULES DESCRIPTION

1. Medical Image Input Module:

This module is responsible for acquiring dermatological images from patients using devices such as dermatoscopes, digital cameras, or clinical imaging systems. It ensures that high-quality skin images are captured and stored in standardized formats for further processing.

A. Image Acquisition & Standardization:

High-Resolution Capture: Ensures clear visualization of lesion details such as color, border, and texture.Standard File Formats: Stores images in formats like JPEG or PNG for compatibility and processing efficiency.Patient Metadata Integration: Links images with relevant patient details such as age, gender, and lesion location.

B. Data Management & Security:

Stopatient images securely to maintain privacy and confidentiality.Data Organization: Maintains structured datasets for easy retrieval and training purposes.Protects patient identity by removing sensitive personal information. Ensures safe backup and recovery of medical image records.

2.Preprocessing & Segmentation Module (U-Net):

This module enhances the quality of input skin images and accurately separates the lesion region from surrounding healthy skin. Image preprocessing improves clarity, while U-Net segmentation ensures precise boundary detection for further analysis.

A. Image Preprocessing Techniques:

Converts images into a uniform size suitable for modelinput. Color Normalization: Adjusts brightness and contrast for consistent analysis.Noise Removal: Removes unwanted artifacts such as hair, shadows, and background noise.Image Enhancement: Improves lesion visibility using filtering techniques.

B. Lesion Segmentation Process:

Generates binary masks highlighting lesion areas.Identifies clear lesion edges and irregular borders. Detects uneven lesion structures for accurate analysis.Separates infected skin area from healthy tissue for focused feature extraction.

3. Lesion Extraction Module:

This module focuses on isolating and extracting only the segmented lesion area for detailed analysis. By removing unnecessary background information, it ensures that only the skin region affected is considered for feature extraction and disease classification.

A. Lesion Cropping & Isolation:

Extracts only the segmented lesion area from the original image.Eliminates surrounding healthy skin to reduce interference. Ensures accurate focus on the pathological tissue. Uses segmentation masks to define exact lesion boundaries.

B. Artifact Removal & Refinement:

Eliminates hair strands that may affect feature analysis. Reduces lighting variations and uneven illumination. Removes minor distortions or unwanted pixels. Enhances lesion clarity for accurate feature extraction.

4. Disease Prediction Module (EfficientNet / CNN):

This module is responsible for classifying the extracted lesion into specific skin diseases using deep learning models such as EfficientNet or CNN. It analyzes morphological, chromatic, and texture features to accurately predict the skin condition and generate confidence scores.

A.Multi-ClassDiseaseClassification:

Uses EfficientNet or CNN to classify skin diseases.Evaluates asymmetry, border irregularity, color variation, and texture patterns. Identifies acne, eczema, psoriasis, normal skin, and serious conditions like Melanoma. Provides confidence values for each predicted class.

B. Performance Evaluation & Optimization:

Evaluates model performance using accuracy metrics.Ensures reliable detection of different disease types.Uses training and testing datasets for performance verification. Minimizes false positives and false negatives through xd34ewfd optimization techniques.

5. Explainability & Visualization Module (Grad-CAM++):

This module enhances transparency by providing visual explanations for the model's predictions. It helps doctors understand which regions of the lesion image influenced the final classification result, thereby increasing trust in the AI system.

A. Visual Explanation Generation:

Generates heatmaps highlighting important lesion regions. Identifies areas that contributed most to the prediction. Shows how texture, color, and shape influenced the result. Superimposes heatmaps on the original image for clarity. Makes AI predictions easier for doctors to understand.Increases confidence in automated diagnosis Supports reasoning similar to dermatologist examination. Helps identify incorrect predictions for model improvement.

6. Diagnosis Summarization Module:

This module combines the prediction results, confidence scores, visual explanations, and patient metadata into a structured and easy-to-understand summary. It provides a clear overview of the suspected skin condition to support clinical decision-making.

A. Prediction Aggregation:

Collects confidence scores from the disease prediction model. Identifies the most probable skin disease category. Displays likelihood values for all predicted classes. Organizes results in a clear and readable format.

B. Clinical Decision Support:

Provides a brief explanation of the predicted condition.Shows reliability of prediction to assist doctors. Helps clinicians decide whether further tests like biopsy are needed. Presents results in a simple format for better understanding.

7.Result Export & Reporting Module:

This module is responsible for saving and exporting the final diagnosis results, segmented images, and visual explanations. It ensures proper documentation and integration with hospital information systems for future reference and clinical use.

A: Report Generation:

Generates detailed diagnostic reports including predicted disease and confidence score. Attaches segmented images and Grad-CAM++ heatmaps in the report. Provides downloadable reports in PDF or Excel format. Displays final diagnosis in a clear and organized manner.

B. Data Storage & Integration:

Secure Data Saving: Stores prediction results and patient information securely.Electronic Health Record (EHR) Integration.Supports hospital record management systems.Maintains history of predictions for future verification. Ensures proper record maintenance and data backup.

OUTPUT:

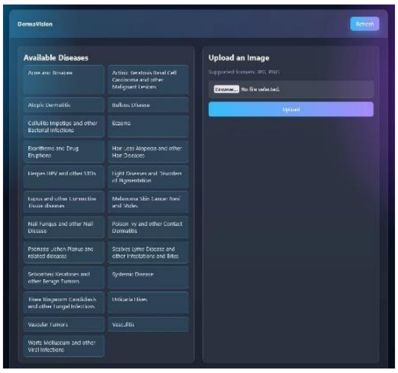

FIG 2: UPLOAD IMAGE

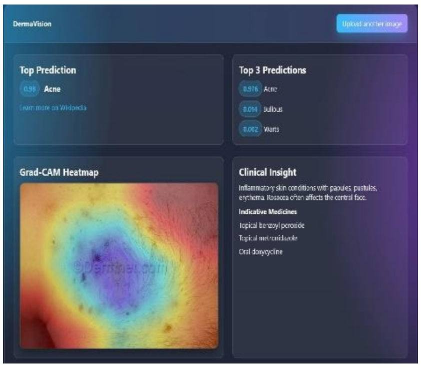

FIG 3: OUTPUT

Stage	Survival rate
0 (in situ)	99.9% 5-year survival; 98.9% 10-year survival
I/II	89 to 95% 5-year survival
II	45 to 79% 5-year survival
III	24 to 70% 5-year survival
IV	7 to 19% 5-year survival

FIG 4: STAGE LEVELS

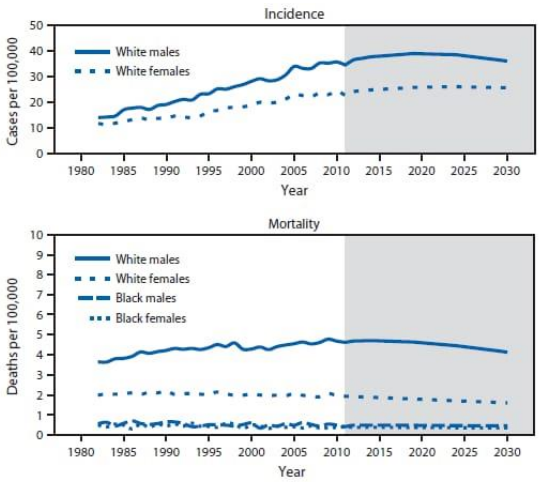

FIG 5: MELANOMA INCIDENCE

VII.RESULT & DISCUSSION:

The DermaVision system successfully analyzed the uploaded skin image using an advanced deep learning model trained on 22 different skin disease categories and generated a detailed prediction report based on visual pattern recognition, texture analysis, and color distribution features extracted from the image. After processing the input, the model identified Acne as the top predicted condition with a confidence score of 33%, indicating that among all learned categories, Acne has the highest probability based on the detected characteristics in the image. In addition to the primary result, the system also displayed the top three predictions to ensure transparency and provide a clearer understanding of the model's confidence distribution, where Lichen appeared as the second likely condition with a probability of 10.2% and Vasculitis as the third with 8.3%, showing that while Acne is the leading prediction, there are overlapping visual similarities with other inflammatory skin conditions. This probability-based ranking approach helps users understand that AI predictions are not absolute diagnoses but statistical estimations derived from learned patterns in training data. To enhance interpretability and reduce the black-box nature of artificial intelligence, the system generated a Grad-CAM (Gradient-weighted Class Activation Mapping) heatmap overlay on the original image, visually highlighting the regions that most influenced the model's decision. In the heatmap, warmer colors such as red and yellow indicate high attention areas, suggesting that the model focused primarily on inflamed, reddened, or lesion-dense regions of the skin, while cooler blue areas represent regions with minimal influence on the final classification. This visual explanation strengthens user trust by demonstrating that the model is concentrating on medically relevant areas rather than irrelevant background details. The clinical insight section further supports the prediction by explaining that Acne is an inflammatory skin disorder commonly characterized by papules, pustules, nodules, and erythema, often affecting areas such as the face, chest, shoulders, and back due to clogged hair follicles and excess oil production. The system also provides indicative treatment options including topical benzoyl peroxide, topical metronidazole, and oral doxycycline, which are commonly prescribed medications aimed at reducing bacterial growth and inflammation; however, these suggestions are presented strictly for educational purposes and not as direct medical prescriptions. A clearly stated disclaimer reminds users that the application functions solely as an AI-based decision-support and screening tool and should not replace consultation with a certified dermatologist or healthcare professional. The structured interface, probability scores, visual heatmap explanation, and clinical mapping together create a comprehensive and user-friendly digital dermatology support system that demonstrates the practical integration of artificial intelligence, computer vision, and medical knowledge. Overall, while the system suggests Acne as the most probable condition based on the analyzed features, final diagnosis, confirmation tests, and treatment decisions should always be performed by a qualified medical expert to ensure safe and accurate healthcare outcomes.

VIII. REFERENCES

1. Zhang X., Liu Y., Ouyang G., Chen W., Xu A., Hara T., Zhou X., Wu D., "DermViT: Diagnosis-Guided Vision Transformer for Robust and Efficient Skin Lesion Classification," Bioengineering, vol. 12, no. 4, article 421, 2025.

2. Amirreza Mahbod, Rupert Ecker, Ramona Woitek, "Fusion of Foundation and Vision Transformer Model Features for Dermatoscopic Image Classification," arXiv preprint, May 2025.

3. Abadh K Chaurasia, Patrick W Toohey, Helen C Harris, Alex W Hewitt, "Multiresolution vision transformer model for skin cancer subtype classification using histopathology slides," medRxiv preprint, Jan 2025.

4. Vivek N., Ramesh K., "Deep Learning Multi-Modal Melanoma Detection: Algorithm Development and Validation," JMIR AI, vol. 4, e66561, 2025.

5. Galib Muhammad Shahriar Himel, Md Masudul Islam, Kh Abdullah Al-Aff, Shams Ibne Karim, Md Kabir Uddin Sikder, "Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermatoscopybased Non-invasive Digital System," arXiv preprint, Jan 2025.

6. Nogales M., Acha B., Alarcón F., Pereyra J., Serrano C., "Robust Melanoma Thickness Prediction via Deep Transfer Learning enhanced by XAI Techniques," arXiv preprint, 2025.
7. Carolin Flosdorf, Justin Engelker, Igor Keller, Nicolas Mohr, "Skin Cancer Detection utilizing Deep Learning: Classification of Skin Lesion Images using a Vision Transformer," arXiv preprint, Jul 2024.

8. Efat AH, Hasan SMM, Uddin MP, Mamun MA, "SkinViT: A transformer based method for Melanoma and Nonmelanoma classification," PLoS One, vol. 19, issue 10, Oct 2024.

9. Wies C., Schneider L., Haggenmüller S., Bucher T-C., Hobelsberger S., Heppt M. V., et al., "Evaluating deep learning-based melanoma classification using immunohistochemistry and routine histology: A three center study," PLoS ONE, vol. 19 (1), e0297146, 2024.

10. Himel MS, Islam M, Karim SI, Sikder KU., "Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system," Int J Biomed Imaging, 2024;2024:1.

11. Desale RP, Patil PS., "An efficient multiclass classification of skin cancer using optimized vision transformer," Med Biol Eng Comput, vol. 62, pp. 773–789, 2024.

12. Hu B., Zhou P., Yu H., Dai Y., Wang M., Tan S., Sun Y., "LeaNet: Lightweight Ushaped architecture for high-performance skin cancer image segmentation," Comput Biol Med, vol. 169, article 107919, 2024.

13. Mateen M., Hayat S., Arshad F., Gu Y.-H., Al-Antari M. A., "Hybrid Deep Learning Framework for Melanoma Diagnosis Using Dermoscopic Medical Images," Diagnostics, vol. 14, no. 19, p. 2242, 2024.

14. Abbas Q., Daadaa Y., Rashid U., Ibrahim ME., "Assist-Dermo: A lightweight separable vision transformer model for multiclass skin lesion classification," Diagnostics, vol. 13, no. 15, article 2531, 2025.

15. Mirikharaji Z., Abhishek K., Bissoto A., Barata C., Avila S., Valle E., Celebi ME., Hamarneh G., "A survey on deep learning for skin lesion segmentation," Med Image Anal, vol. 88, article 102863, 2025.

16. Muhammad Asad Arshed, Shahzad Mumtaz, Muhammad Ibrahim, Saeed Ahmed, Muhammad Tahir, Muhammad Shafi, "Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models," Information, vol. 14, no. 7, article 415, Jul 2025.

17. Sethanan K., Pitakaso R., Srichok T., Khonjun S., Thannipat P., Wanram S., Boonmee C., Gonwirat S., Enkvetchakul P., Kaewta C., Nanthasamroeng N., "Double AMIS-ensemble deep learning for skin cancer classification," Expert Syst Appl, vol. 234, article 121047, 2025.

18. Goceri E., "Classification of skin cancer using adjustable and fully convolutional capsule layers," Biomed Signal Process Control, vol. 85, article 104949, 2025.

19. Tabrizchi H., Parvizpour S., Razmara J., "An Improved VGG Model for Skin Cancer Detection," Neural Process Lett, vol. 55, pp. 3715–3732, 2024.

20. Tembhurne JV., Hebbar N., Patil HY., Diwan T., "Skin cancer detection using ensemble of machine learning and deep learning techniques," Multimed Tools Appl, vol. 82, pp. 27501–27524, 2025.

21. Xu X., Fan H., Wang M., Cheng X., Chen C., "A Method for Classification of Skin Cancer Based on VisionTransformer," ICNC-FSKD 2022, vol. 153 LNAI Lect. Notes Data Eng. Commun. Technol., pp. 210–217, Jan 2024.