Dermatology Practical & Conceptual

Artificial Intelligence–Assisted Diagnosis of Atopic Eczema in Darker Skin Types: A CNN-Based Study

Atinuke Arinola Ajani[1], Ishaya Gambo[2], Oluwasegun Faniyi[2]

1 Department of Dermatology and Venereology, Obafemi Awolowo University, Ile-Ife, Nigeria

2 Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Key words: Atopic dermatitis, Darkly pigmented skin, Artificial intelligence, Machine learning, Convolutional neural networks (CNNs)

Citation: Ajani AA, Gambo I, Faniyi O. Artificial Intelligence–Assisted Diagnosis of Atopic Eczema in Darker Skin Types: A CNN-Based Study. *Dermatol Pract Concept*. 2026;16(1):5897. DOI: https://doi.org/10.5826/dpc.1601a5897

Accepted: August 15, 2025; **Published:** January 2026

Funding: None.

Competing Interests: None.

Authorship: All authors have contributed significantly to this publication.

Corresponding Author: Ajani Atinuke Arinola, Consultant Dermatologist & Genitourinary Physician, Department of Dermatology and Venereology, Obafemi Awolowo University, Ile-Ife, Nigeria. ORCID ID: 0000-0002-1979-0773. E-mail: daraolu@yahoo.com

ABSTRACT

Introduction: Atopic dermatitis (AD) presents significant diagnostic challenges in darker skin types due to the masking of classical signs like erythema by pigmentation and the lack of diagnostic criteria tailored for this population. The use of machine learning (ML) as a noninvasive diagnostic tool in this context remains underexplored.

Objectives: This study aimed to develop and evaluate the accuracy of an ML model in diagnosing AD in darker skin types and distinguishing it from healthy skin.

Methods: A cross-sectional study using images of AD cases in dark-skinned patients was conducted. An EfficientNet-based convolutional neural network (CNN) model was trained on 1403 labelled images (687 atopic and 716 healthy skin images). Performance metrics including accuracy, precision, recall, and F1 score were assessed. Ethical approval was obtained, and datasets were anonymized for confidentiality.

Results: The dataset comprised images from 50 AD patients (age range 1–93 years, median 18.0, IQR: 7.0–32.0) and 60 healthy volunteers (age range 18–50 years, median 27.5, IQR: 23.0–33.5), with a slight female predominance in both groups (56.0% and 53.3%, respectively). The model achieved an accuracy of 91.1%, precision of 83.7%, recall of 97.0%, and an F1 score of 89.9%, correctly classifying 103/106 AD images and 129/149 healthy skin images. External application programming interfaces (APIs) were integrated into the design to further enhance accuracy and generalizability.

Conclusion: This study highlights the potential of ML as a diagnostic tool for AD in darker skin types. Further optimization and large-scale validation could enhance diagnostic accuracy, improving clinical outcomes for populations with darker skin. Iterative refinement and user feedback will ensure its continued efficacy and reliability in clinical practice.

Introduction

Atopic dermatitis (AD) is a chronic relapsing inflammatory skin condition with significant variability in clinical phenotype and course. The burden of AD is multifaceted, affecting individuals on physical, psychological, social, and economic levels. The condition can significantly impair a person's quality of life as the persistent itching and discomfort disrupt sleep and daily activities[1]. Early diagnosis and timely treatment are crucial to reducing the psychological impact of the disease and improving overall patient outcomes.

Diagnosing AD in individuals with darker skin tones, however, presents unique challenges. The characteristic erythema observed in lighter skin tones is often masked by pigmentation. Additionally, atypical variants, such as lichenified or prurigo-like lesions, are more common in this population [2]. Furthermore, the overlap of AD features with other dermatoses, coupled with the absence of standardized diagnostic criteria that account for the unique characteristics in darker skin types, further complicate accurate diagnosis.

Machine learning (ML) has emerged as a transformative tool in dermatology, harnessing pattern recognition in clinical images to enhance diagnostic precision. The algorithms employed in ML models enable rapid and accurate processing of extensive data, potentially reducing diagnostic timelines and improving patient outcomes[3]. While ML has been successfully applied to several dermatological conditions, including AD in light-skinned populations[4–8], its potential to address diagnostic challenges in AD in darker skin types remains largely unexplored.

Objectives

This study aimed to fill this gap by developing and validating an ML algorithm specifically trained to diagnose AD and differentiate it from healthy skin in individuals with darker skin types.

Methods

Study Design

A cross-sectional observational design was employed using clinical images of patients with clinically confirmed AD sourced from the Dermatology Clinics of the Obafemi Awolowo University Teaching Hospitals Complex (OAUTHC), Ile-Ife, Nigeria. The dataset encompassed a diverse range of clinical phenotypes and severities of confirmed AD cases. Additionally, images of normal skin obtained from volunteers with similar skin phototypes were included to serve as a control group.

Inclusion Criteria

Pretreatment clinical images obtained from patients with AD diagnosed by experienced dermatologists using the Hanifin and Rajka criteria[9]. High-quality clinical images and images of normal skin obtained from healthy controls.

Exclusion Criteria:

- Poor quality or incomplete images
- Images obtained from individuals younger than six months
- Images of AD lesions depicting comorbid dermatological conditions such as eczema herpeticum, tinea corporis, and so on.
- Images of AD in treatment-experienced patients.

Data Collection and Annotation

High-resolution images capturing the clinical features of treatment-naïve AD lesions were obtained from image repositories of the Department of Dermatology & Venereology, OAUTHC, Ile-Ife, Nigeria. An experienced dermatologist meticulously reviewed and selected representative images of AD while providing clinical context for each case. To facilitate comparison and evaluation, images of healthy volunteers without skin lesions were included. Figure 1 shows samples of images used in the study.

The selected images were annotated by the dermatologist to indicate whether they depicted AD or healthy skin. The dataset was purposefully selected to encompass a diverse spectrum of AD severity, anatomical distribution, sex, and age groups to enhance the model's generalizability. Patient demographics, including age and sex as well as other relevant clinical details, were extracted from the patients' medical records and are presented in Table 1.

Data Preprocessing

Eligible clinical images were subjected to data preprocessing to enhance model robustness and reduce biases related to lighting and image quality. The preprocessing steps included:

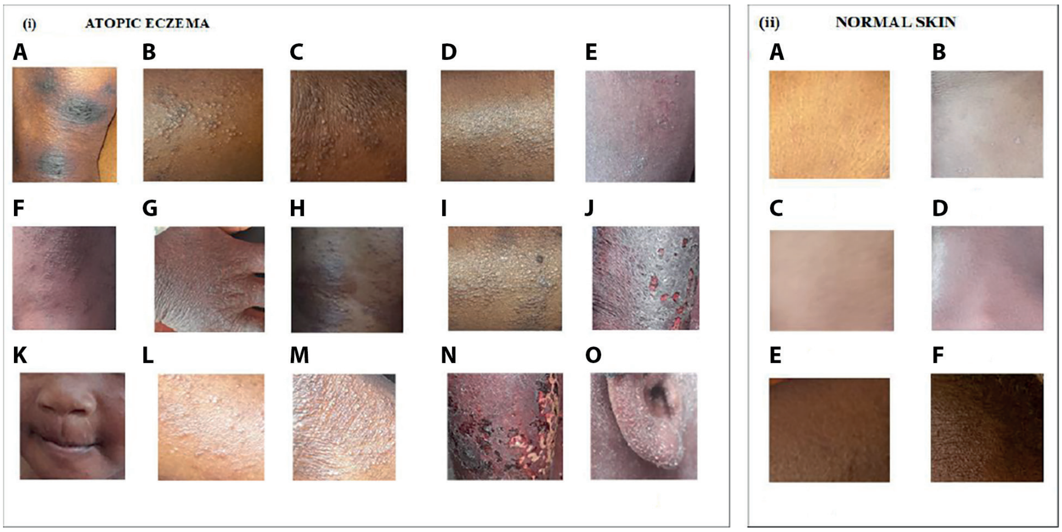

Figure 1. Representative training set images illustrating the clinical diversity captured for model development. (i) Atopic dermatitis spectrum: A–E: Distinct morphological variants: A) nummular variant, B) papular variant, C) subacute eczema, D) classical morphology, E) erythrodermic form. F–J: Age-related presentations: F) infantile, G) childhood, H) adolescent, I) adult-onset, J) elderly. K–O: Severity grades: K–L) mild, M) moderate, N–O) severe disease. (ii) Healthy skin controls: A-F: Images depicting a range of clinical phototypes in darker skin types to enhance model generalizability across diverse populations.

Table 1. Baseline demographic characteristics of study participants.

Variable	Atopic Dermatitis (N=50)	Healthy Skin (N=60)
Sex, N (%)		
Female	28 (56.0)	32 (53.3)
Male	22 (44.0)	28 (46.7)
Age (years)		
Mean ± SD	23.4 ± 22.54	28.7 ± 7.0
Median (IQR)	18.0 ((7.0–32.0))	27.5 (23.0–33.5)
Range (Min–Max)	1–93	18–50
Age group, n (%)		
Pediatric (<18 years)	25 (50.0)	0 (0.0)
Young adults (18–35 years)	15(30.0)	48 (80.0)
Older adults (>35 years)	10(20.0)	12 (20.0)

1. Image cleaning: removing artefacts and noise to enhance image quality
2. Standardization: resizing images and normalizing pixel values to ensure consistency across the dataset.
3. Augmentation: transformation of images through techniques such as rotation, flipping, and brightening adjustments to increase the dataset's robustness and diversity.
4. Feature extraction: employing techniques such as edge detection and color-based analysis to highlight key diagnostic features.

Data Categorization

The processed images were divided into three subsets for model development and evaluation:

i. Training set: used for model development
ii. Validation set: used for fine-tuning and hyperparameter optimization
iii. Test set: employed for performance evaluation and quality control.

Machine Learning Model Development

An EfficientNet-based convolutional neural networks (CNN) model was employed in this study. Multiple machine learning algorithms, including diverse CNNs, decision trees, random forests, and support vector machines (SVMs), were initially evaluated for their suitability to dermatological image classification. EfficientNet-B0 was selected as the CNN backbone due to its optimal balance between accuracy and computational efficiency, achieved through compound scaling, which enables high performance with fewer parameters and reduced computational cost. These characteristics make it particularly suited to processing heterogeneous dermatology images in resource-limited and mobile health settings. Validation datasets were used to fine-tune hyperparameters and select the optimal model based on accuracy, sensitivity, and specificity.

Model Training

A supervised learning approach was employed, using EfficientNet-based CNN trained on labelled images of AD

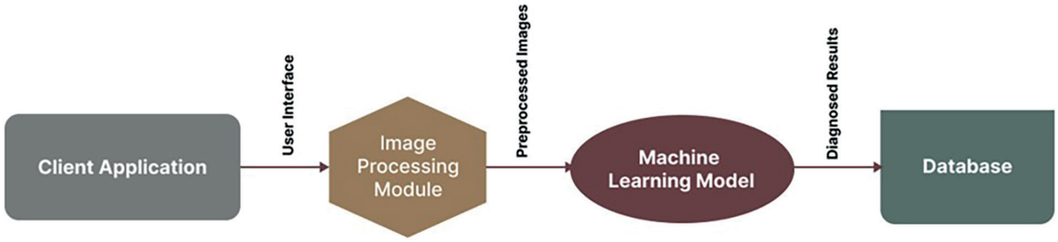

Figure 2. System architecture for the atopic dermatitis diagnostic platform: The workflow begins with the client application, which serves as the user interface for capturing and submitting skin images. These images pass through the image processing module, where they are resized and cropped to the required dimensions for analysis. The ML model then processes the preprocessed images to classify them as either atopic dermatitis or healthy skin. The diagnostic results are subsequently stored in the database, enabling retrieval and review by healthcare providers.

and healthy skin. Training progress was monitored using learning curves to identify patterns of underfitting or overfitting. Techniques such as early stopping and regularization were employed to improve the model's robustness and mitigate overfitting. This iterative refinement process aimed to optimize the model's performance.

Testing and Validation

Following training, the model was evaluated using an independent test dataset to ensure generalizability. The performance metrics, including accuracy, sensitivity, specificity, recall, and F1 score, were calculated. This evaluation provided key insights into the model's diagnostic accuracy and its potential clinical utility for diagnosing atopic eczema.

Statistical Analysis

The model's performance was compared against dermatologists' assessment using receiver operating characteristic (ROC) curve analysis. Key metrics, including area under the curve (AUC), confusion matrix, and Kappa statistics, were used to assess agreement between the model and clinical experts.

Ethical Considerations

This research, with protocol number ERC/2023/07/05, received ethical approval from the Ethics and Research Committee of the OAUTHC, with the following registration numbers: International IRB/IEC/0004553 and national registration NHREC/17/03/2021. Informed consent was pre-obtained from all patients and/or their guardians during routine dermatology clinics as part of the standard departmental protocol for clinical image collection and documentation, while volunteers with healthy skin were required to provide written informed consent prior to their inclusion in the study. All clinical images were anonymized to ensure confidentiality.

Result

A total of 1403 clinical images were used to train the model, comprising 687 photographs of AD lesions obtained from 50 patients and 716 images of healthy skin (Figure 1). The AD cohort showed a slight female preponderance (56.0%), with ages ranging from 1 to 93 years (median 18.0 years, IQR: 7.0–32). The healthy skin cohort (N=60) had a similar sex distribution (53.3% female), with ages ranging from 18 to 50 years (median 27.5 years, IQR: 23.0–33.5). Detailed demographic characteristics are presented in Table 1.

System Architecture

The diagnostic platform comprises three core components: a client-server interface for healthcare providers, an image-processing module, and a machine learning model (Figure 2). The client application enables the secure upload of clinical history and skin images, which are automatically preprocessed (resizing, normalization, and noise reduction) by the image processing module to optimize input quality. A dedicated database stores de-identified images and metadata for efficient retrieval and scalability. External application programming interfaces (APIs) and dermatology-specific datasets were incorporated into support diagnostic performance, although their impact on accuracy was not evaluated as it was not within this study's scope.

System Functionality

Following image upload, preprocessed files were analyzed by the CNN to generate a diagnostic probability for AD, expressed as a percentage confidence score (Figure 3). The output was relayed to the provider via the client interface alongside ethe patients record for clinical integration. The workflow, including image acquisition, preprocessing model evaluation, and result display, are summarized in Figure S1.

Final Output: Diagnostic Page

Based on the clinical information provided and the uploaded image, the model generates a probability, indicating the likelihood that the lesion represents AD. The prediction is expressed as a percentage, reflecting the confidence level of the model in diagnosing AD, as illustrated in Figure S2.

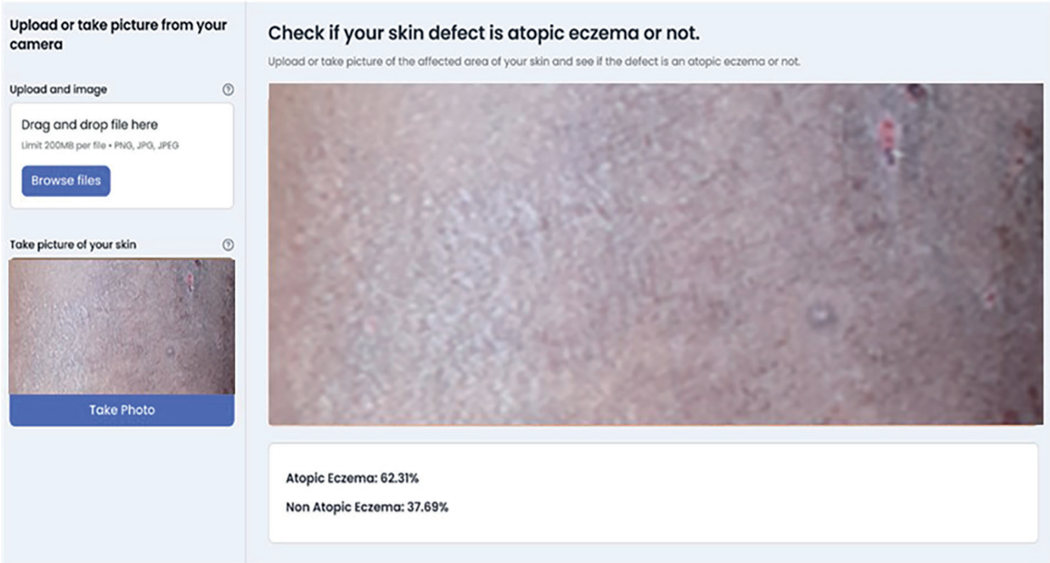

Figure 3. Model diagnostic probability outputs for sample input images, showing predicted probabilities for a test image: Atopic Eczema (62.31%) and Non-Atopic Eczema (37.69%). These values represent the model's confidence in classifying the presented lesion.

Testing and Validation

The performance of the developed diagnostic model was evaluated using a set of standard machine learning metrics, including accuracy, precision, recall, and F1 score. These metrics collectively provided a comprehensive assessment of the model's diagnostic proficiency.

The model was tested using 255 clinical images (106 AD lesions) and 149 normal skin and achieved an overall accuracy of 91%, precision of 84%, recall of 97%, and F1 score of 90%, as shown in Table 2. These metrics underscore the model's robust ability to accurately diagnose AD. Further analysis, depicted in Table 3, utilized a confusion matrix to provide a meticulous examination of the model's predictions. Among the 106 instances of AD, the model correctly identified 103 cases, with three false positives. Similarly, out of 149 instances of normal skin, the model accurately identified 129 cases alongside 20 false positives.

Table 2. Performance metrics of the machine learning model for diagnosing atopic dermatitis and normal skin.

	Precision	Recall	F1 Score
Atopic Eczema	0.84	0.97	0.90
Normal Skin	0.98	0.87	0.92
Accuracy			0.91

Discussion

This study evaluated the application of CNNs EfficientNet-B0 to classify skin images of dark-skinned individuals into AD and normal-skin categories. ML algorithms, particularly supervised learning models such as support vector machine, random forest (RF), decision trees, and artificial neural networks (ANNs), have been widely utilized in diagnosing and monitoring skin conditions like AD. However, most applications have focused on populations in developed regions [4–8].

CNNs are a type of deep learning architecture particularly suited to image processing tasks. Unlike traditional ML models that require preprocessing steps like segmentation and handcrafted feature extraction, CNNs utilize convolutional filters to process raw image data directly, enabling end-to-end learning[10]. This allows for improved generalizability in complex image-based diagnostic tasks, as demonstrated in previous studies [4–6,11–13].

In our study, the CNN model achieved robust performance, with sensitivity of 97% for AD, a negative predictive value of 97.7% for normal skin, and an F1 score of 89.9%. These results align closely with findings by Maulana et al., where ResNet50 achieved a precision of 90.0%, specificity of 96.6%, and an F1 score of 89.9% in classifying AD severity in an Indonesian population[14]. However, our study uniquely incorporated the clinical diagnostic criteria for AD (Hanifin and Rajka criteria)[9] and focused on images of darker skin types, thereby addressing a significant gap in dermatological AI research.

While the application of CNN-based models to diagnosing AD remains limited, other studies have reported similarly promising results [15–18]. For example, Son et al. combined CNNs for feature extraction with SVMs to classify various skin conditions including AD, achieving a specificity of 95.1%, sensitivity of 61.4%, and F1 score of 60.8% through

Table 3. Confusion matrix depicting the performance of the machine learning model in diagnosing atopic eczema and normal skin.

Dermatologist Diagnostic annotation	ML model Predicated Atopic Eczema	ML model Predicted Normal Skin	Total
Atopic eczema	103 (TP)	3 (FN)	106
Normal skin	20 (FP)	129 (TN)	149
Total	123	132	255

Abbreviations: ML model: machine learning model; TP: true positive; FN: false negative; FP: false positive; TN: true negative.

refined contextual segmentation[19]. Similarly, Wu et al.[18] employed CNNs to categorize inflammatory skin conditions, including AD, reporting an accuracy of 92.57%, sensitivity of 94.56%, and specificity of 94.41% for AD and eczema.

However, despite these robust performances, the reliance on erythema as a diagnostic feature in these studies presents significant limitations for darker skin types. In darker skin types, erythema is often masked, potentially reducing the diagnostic accuracy and increasing the risk of misdiagnosis in this population. To address this, our model was trained on images representative of darker skin tones and integrated clinical data such as age, medical history, and validated diagnostic criteria. This approach provides a more comprehensive assessment and minimizes potential biases, thereby enhancing diagnostic accuracy across diverse populations.

It is noteworthy that our model exhibited slightly lower specificity compared to Wu et al.[18], whose models were trained on larger and more diverse datasets. This highlights the critical role of extensive, well-labelled data in optimizing machine learning performance. CNNs, like other machine learning models, rely heavily on the size and diversity of training datasets, with repeated training and finetuning shown to improve accuracy and classification performance [17]. Despite these challenges, CNNs have consistently demonstrated superior accuracy and lower error rates compared to traditional ML algorithms such as logistic regression in diagnosing a variety of skin conditions [4,20]. As such, CNNs remain a promising tool for the development of accurate and efficient image-based diagnostic systems.

Overall, our study demonstrates the robust performance of the model, particularly in minimizing false negatives for atopic eczema, making it a valuable tool for real-world clinical application. However, the observed lower specificity indicates occasional misclassification of normal skin as atopic eczema, underscoring the need for further optimization. Expanding the training dataset to include other skin conditions with similar morphological features to AD could enhance the model's discriminatory accuracy and overall reliability.

These findings highlight the potential of CNN-based models as effective diagnostic tools in dermatology, particularly for underrepresented populations with darker skin tones. Future efforts should prioritize the development of larger, more diverse datasets and the refinement of model architectures to further improve precision and ensure applicability across varied clinical settings. By addressing these areas, CNNs can play a transformative role in promoting equitable and accurate dermatological care.

Study Limitations

While promising, this study has several limitations. The slightly lower specificity of our model underscores the need for larger and more diverse datasets to further optimize its performance. The model occasionally misclassified normal skin as AD, suggesting the need to expand the training dataset to include a broader range of skin types. The exclusion of common disease mimickers such as psoriasis, contact dermatitis, and seborrheic dermatitis is another key limitation of this study. While this approach minimized diagnostic ambiguity during model training, it limits applicability in real-world settings, where distinguishing between similar conditions is essential. Incorporating such conditions in future datasets will be critical to improving diagnostic robustness. Additionally, the reliance on a single dataset from a specific region may limit the generalizability of the findings to other populations.

Finally, while clinical data integration improved diagnostic accuracy, the model's performance in real-world clinical settings, where data variability, including image quality and noise, are prevalent remains to be validated. Future studies should focus on external validation using multicenter datasets and exploring advanced model architectures to enhance specificity and overall reliability.

Conclusion

This study highlights the potential of CNN models as a highly effective diagnostic tool for AD in clinical practice. By focusing on images of darker skin types, our approach addresses a significant gap in dermatological AI research, fostering more inclusive and accurate diagnostic capabilities and promoting equitable AI applications in dermatology.

The integration of deep neural networks with clinical data enhances the system's diagnostic accuracy, mitigating the biases inherent in traditional image-based methods. The system's user-friendly graphical user interface (GUI) streamlines clinician interactions, facilitating the management of patient information, input of clinical data, and efficient classification of skin images.

As medical knowledge and technology evolve, the system remains adaptable, with ongoing improvements driven by user feedback and data insights. Its scalability and flexibility ensure seamless adoption across diverse clinical settings, ultimately enhancing diagnostic efficiency and improving clinical outcomes.

Acknowledgements: The authors acknowledge the contributions of the OAUTHC dermatology team for their assistance in data collection.

Declaration of the use of Generative AI and AI assisted technologies in the research writing process: During the preparation of this work, the authors used ChatGTP as a grammar editing tool to enhance the readability and linguistic quality of the manuscript. Following its use, the authors thoroughly reviewed and edited the content and hereby assume full responsibility for the authenticity and integrity of the publication.

References

1. Fasseeh AN, Elezbawy B, Korra N, et al. Burden of Atopic Dermatitis in Adults and Adolescents: A Systematic Literature Review. *Dermatol Ther (Heidelb)*. 2022;12(12):2653-2668. DOI:10.1007/s13555-022-00819-6. PMID: 36197589
2. Chiricozzi A, Maurelli M, Calabrese L, Peris K, Girolomoni G. Overview of Atopic Dermatitis in Different Ethnic Groups. *J Clin Med*. 2023;12(7):2701. DOI:10.3390/jcm12072701. PMID: 37048783
3. Premaladha J, Ravichandran KS. Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms. *J Med Syst*. 2016;40(4):96. DOI:10.1007/s10916-016-0460-24. PMID: 26872778
4. Bhadula S, Sharma S, Juyal P, Kulshrestha C. Machine Learning Algorithms based Skin Disease Detection. *International Journal of Innovative Technology and Exploring Engineering*. 2019;9(2):4044–9. DOI: 10.35940/ijitee.B7686.129219.
5. Rajeshwari J, Sughasiny M. Modified PNN classifier for diagnosing skin cancer severity condition using SMO optimization technique. *AIMS Electronics and Electrical Engineering*, 2023, 7(1): 75-99. DOI: 10.3934/electreng.2023005
6. Mazhar T, Haq I, Ditta A, et al. The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer. *Healthcare (Basel)*. 2023;11(3):415. DOI:10.3390/healthcare11030415. PMID: 36766989
7. Ahammed M, Mamun M Al, Uddin MS. A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*. 2022;2. DOI:10.1016/j.health.2022.100122
8. Maul L V., Meienberger N, Kaufmann L. Role of artificial intelligence in assessing the extent and progression of dermatoses. *Hautarzt*. 2020;71(9):677-685. DOI:10.1007/s00105-020-04657-5. PMID: 32710130
9. Eichenfield LF, Tom WL, Berger TG, et al. Guidelines of care for the management of atopic dermatitis: section 2. Management and treatment of atopic dermatitis with topical therapies. *J Am Acad Dermatol*. 2014;71(1):116-132. DOI:10.1016/j.jaad.2014.03.023. PMID: 24813302
10. Tschandl P, Rosendahl C, Akay BN, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol*. 2019;155(1):58-65. DOI:10.1001/jamadermatol.2018.4378. PMID: 30484822
11. Alsaade FW, Aldhyani THH, Al-Adhaileh MH. Developing a Recognition System for Diagnosing Melanoma Skin Lesions Using Artificial Intelligence Algorithms. *Comput Math Methods Med*. 2021;2021:9998379. DOI:10.1155/2021/9998379. PMID: 34055044
12. Li H, Pan Y, Zhao J, Zhang L. Skin disease diagnosis with deep learning: A review. *Neurocomputing*. 2021;464. DOI:10.48550/arXiv.2011.05627
13. Wiala A, Ranjan R, Schnidar H, Rappersberger K, Posch C. Automated classification of hidradenitis suppurativa disease severity by convolutional neural network analyses using calibrated clinical images. *J Eur Acad Dermatol Venereol*. 2024;38(3):576-582. DOI:10.1111/jdv.19639. PMID: 38013510
14. Maulana A, Noviandy TR, Suhendra R, et al. Evaluation of atopic dermatitis severity using artificial intelligence. *Narra J*. 2023;3(3):e511. DOI:10.52225/narra.v3i3.511. PMID: 38450339
15. Suaini NHA, Yap GC, Bui DPT, et al. Atopic dermatitis trajectories to age 8 years in the GUSTO cohort. *Clin Exp Allergy*. 2021;51(9):1195-1206. DOI:10.1111/cea.13993. PMID: 34310791
16. Fajar R, Nirmawati RA, Huahua RL, Syailendra Pertiwi EM, Panjaro BL. Machine learning-based differential diagnosis of psoriasis and atopic dermatitis using clinical data and medical images. *Journal of Investigative Dermatology*. 2023;143(11). DOI:10.1016/j.jid.2023.09.051
17. Rasheed A, Umar AI, Shirazi SH, Khan Z, Nawaz S, Shahzad M. Automatic eczema classification in clinical images based on hybrid deep neural network. *Comput Biol Med*. 2022;147:105807. DOI:10.1016/j.compbiomed.2022.105807. PMID: 35809409
18. Wu H, Yin H, Chen H, et al. A deep learning, image-based approach for automated diagnosis for inflammatory skin diseases. *Ann Transl Med*. 2020;8(9):581. DOI:10.21037/atm.2020.04.39. PMID: 32566608
19. Son HM, Jeon W, Kim J, et al. AI-based localization and classification of skin disease with erythema. *Sci Rep*. 2021;11(1):5350. DOI:10.1038/s41598-021-84593-z. PMID: 33674636
20. Guetari R, Ayari H, Sakly H. Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches. *Knowl Inf Syst*. DOI:10.1007/s10115-023-01894-7. PMID: 37361377