https://africanjournalofbiomedicalresearch.com/index.php/AJBR
Afr. J. Biomed. Res. Vol. 28(1s) (January 2025); 2488-2494
Research Article

The Role Of Artificial Intelligence In Dermatology: Enhancing The Accuracy Of Skin Cancer Diagnosis

Dr Sanjeev Gulati[1*], Dr Durgesh Sonare[2], Dr Sagar Nanaso Salunkhe[3], Karishma Das Purkayastha[4]

[1*]Associate Professor, Department of Dermatology, Dayanand Sagar University, CDSIMER, Harohalli, Bangalore, jgdsanjeevgulati@gmail.com
[2]Associate Professor, NSC Government Medical College, Khandwa, Madhya Pradesh, durgeshsonare88@gmail.com
[3]Professor, Department of Biochemistry, Symbiosis Medical College for Women and SUHRC, Symbiosis International (Deemed) University, Lavale, Pune, Maharashtra- 412115, sagar.salunkhe@smcw.siu.edu.in
[4]Research Scholar, Tezpur University -784028, Assam, daspurkarishma@gmail.com

***Corresponding Author:** Dr Sanjeev Gulati
*Associate Professor, Department of Dermatology, Dayanand Sagar University, CDSIMER, Harohalli, Bangalore, jgdsanjeevgulati@gmail.com

Abstract
Skin cancer is still a major health issue and a correct and timely diagnosis plays a crucial role in the management of skin cancer. Artificial intelligence (AI) has become the focus as an important tool to improve the accuracy of diagnosis in dermatology. However, much more needs to be done for its real-world usability and the optimal outcome in various populations. This study involved a two-part design, a review of 10,000 dermoscopic images in the archive, and a study of 500 clinical cases. An AI diagnostic system was trained and validated on a diverse dataset with an emphasis on diagnostic sensitivity, specificity, and diagnostic concordance across demographic subgroups. The inclusion criteria included patients with clinically suspicious skin lesions and the exclusion criteria excluded cases with missing data or prior treatment. Descriptive statistics compared the AI and dermatologist diagnoses, and the results were compared across countries. The AI system had a sensitivity of 90.1% and a specificity of 87.6% which was comparable to dermatologist performance. The diagnostic accuracy was similar for all types of lesions but slightly lower for the dark skin tone and lesions in complex anatomic areas. Statistically relevant relationships ($p < 0.05$) were found between AI predictions and histopathological results. AI showed good results in the diagnosis of skin cancer, which may be useful as an auxiliary in dermatology. But, combating dataset biases and extending such research to more extensive audiences is crucial.

Keywords: Skin cancer, artificial intelligence, dermatology, diagnostic accuracy, machine learning, healthcare technology

***Author for correspondence: Email:** jgdsanjeevgulati@gmail.com*

Received:20/01/2025 *Accepted:31/01/2025*

DOI: https://doi.org/10.53555/AJBR.v28i1S.6715

1. INTRODUCTION

Skin cancer is one of the most common cancers in the world; it is divided into non-melanoma skin cancer and melanoma; the latter is the most dangerous and can be lethal. The identification of the disease at an early stage and correct diagnosis is critical to enhance the length and quality of life expectancy of patients. Conventional tests such as simple optical examination and biopsy, though indispensable, are particularly susceptible to human intervention, particularly when a tumor is highly differentiated or in any way questionable. The advancement in the use of artificial intelligence (AI) presents an opportunity to improve the accuracy of diagnosis, which is more neutral and can be applied on a large scale to dermatological examinations (Esteva et al., 2017). AI, especially, ML algorithms, has shown a good possibility in the field of dermatology from the aspect of image recognition and prediction models. AI systems can diagnose dermoscopic images and histopathology slides more accurately and consistently than human dermatologists (Tschandl et al., 2020). These AI developments can transform dermatology by offering higher sensitivity, less time-consuming, and reproducible diagnostic tools for skin cancer screening, particularly in areas where dermatology is restricted (Liopyris et al., 2022).

AI has primarily been concerned with DL and CNN for the classification and detection of skin cancer in dermatology. CNNs are very useful in analyzing large sets of medical images for early-stage detection of melanoma and other skin cancer (Liopyris et al., 2022). These models can be more accurate than general practitioners in diagnosis, particularly when it comes to skill analysis (Esteva et al., 2017). Esteva et al. (2017) have shown that by utilizing more than 129,000 dermatological images, an AI model is better at diagnosing skin cancer than dermatologists. The study showed that the AI system could diagnose skin diseases with a diagnostic accuracy rate of 95% which is as accurate as dermatologists with years of experience. In a similar vein, Tschandl et al., (2020) pointed out that one of the benefits of AI is that the technology can assist in enhancing the diagnosis and may well be useful in the diagnosis of primary melanoma at a very initial stage.

This review also focused on the improvement of AI systems in the aspect of sensitivity and specificity in dermatological diagnosis, which is important in patient management. Still, these approaches hold great potential for integrating AI with dermatology, though certain obstacles need to be addressed: generalizability of models, and minimal availability of inclusive, quality data sets. Several current AI models are based on databases of images mainly representing Western populations, thus their usage in diverse geographic and ethnicity can be limited (Manole et al., 2024). In addition, while applying AI accelerates diagnostic capabilities, they are not without shortcomings. For example, AI systems may be challenged in deciding whether a particular lesion in a patient with skin cancer is benign or malignant where there are unusual presentations of skin diseases or in cases of rare or uncommon types of skin cancer for instance those described by Ye et al., 2024. Skin cancer diagnostics using AI technologies is still an active and developing research area, yet there is a long way from laboratory solutions to their infrastructure implementation in clinics. Current models are not sufficiently invariant to transfer across different populations and may encounter challenges in differentiating between various or atypical types of skin cancer. Besides, there remains controversy over whether AI is adaptive to function as a clinician-supporting resource rather than a clinician-replacing one since the final choices remain the dermatologists' sole discretion. Thus, the issue is not in the development of AI systems, but in improving their precision, as well as in defining their position as auxiliary tools in dermatology practice rather than as independent diagnostic tools.

Research Objectives:

1. To evaluate the diagnostic accuracy of AI-driven systems in detecting skin cancer in comparison to human dermatologists.
2. To assess the generalizability of AI models across diverse patient demographics, including different skin types, age groups, and ethnic backgrounds.
3. To explore the limitations of current AI systems in detecting rare or atypical forms of skin cancer.
4. To investigate the role of AI as a complementary tool in clinical practice, focusing on how it can enhance dermatologists' decision-making processes.
5. To identify the potential barriers and challenges in implementing AI in routine dermatological practice, including regulatory concerns, ethical issues, and the need for training and infrastructure.

2. METHODOLOGY

2.1 Study Design

In this study, a cross-sectional diagnostic accuracy design was used to assess the effectiveness of AI-based systems in diagnosing skin cancer. The study was conducted in two phases: The study was divided into two parts: (1) a retrospective study that compared dermatological images and histopathological diagnoses of skin diseases that were available before the study, and (2) a prospective study that collected new data and tested AI-based diagnostic systems in clinical practice. The main aim of this design was to determine the diagnostic performance of AI algorithms for detecting malignant and benign skin lesions against dermatologist assessment with histopathology as the reference standard. The diagnostic accuracy design was selected because of its effectiveness in assessing the effectiveness of diagnostic instruments. This design allowed for the direct comparison of AI outputs with expert clinical diagnosis and the collection of accurate data regarding the potential of AI systems as supportive tools in dermatological practice. Moreover, the study design enabled the identification of the AI system's weaknesses and possible biases when used with different patients.

2.2 Study Location and Population

The study was carried out at a premier tertiary care dermatology hospital with modern diagnostic imaging, biopsy, and a wide patient population. The center was chosen to increase the sample size and variability of dermatological conditions, which would increase the validity of the study results. The participants were patients who had suspicious skin lesions at the dermatology outpatient clinic of the center. The patients were selected according to inclusion and exclusion criteria to make the study sample more realistic to the clinical practice. Moreover, the study used both backward and forward enrollment of patients to capture a large dataset for AI analysis and validation.

- **Inclusion Criteria**

1. Adults aged 18 years or older presenting with clinically or dermoscopically suspicious skin lesions.
2. Patients who provided informed consent to participate in the study agreed to the use of anonymized images for research purposes.
3. Lesions for which histopathological confirmation was available to serve as the diagnostic reference standard.
4. High-resolution clinical and dermoscopic images of the lesions, meeting the technical specifications required for AI model analysis.

- **Exclusion Criteria**

1. Patients with a prior history of treatment (e.g., surgical excision, cryotherapy, or laser therapy) for the lesion under evaluation, could alter lesion morphology.
2. Patients diagnosed with rare dermatological conditions are not commonly seen in routine clinical practice, as these could confound the AI system's diagnostic capability.
3. Pregnant or lactating women, as hormonal changes during pregnancy and lactation could lead to atypical skin lesion presentations.
4. Low-quality or improperly captured images that did not meet the AI system's minimum input requirements.
5. Patients unwilling to provide written consent or those who withdrew consent during the study.

2.3 Data Collection

Data collection was divided into two distinct phases to ensure robust data for analysis. Both retrospective and prospective approaches were implemented to maximize the dataset's diversity and applicability.

1. Retrospective Phase:

The retrospective phase involved identifying dermatological images and their corresponding histopathological reports from the hospital's electronic health record system. These records were from the previous three years to make sure that the data set was large enough. All the retrieved images were then matched with histopathology outcomes to act as the gold standard for diagnostic performance assessment. To maintain the quality of data, all the images were checked by an expert dermatologist to ensure they met the criteria for inclusion. Any image that was of low resolution, out of focus, or only partially visible was not included in the analysis. Subsequently, the images were de-identified to maintain patient privacy and resized to the specifications of the AI system. Each image was also tagged with basic information such as patient age, sex, location of the lesion, and clinical history where available.

2. Prospective Phase:

In the prospective phase, patients attending the dermatology clinic with clinically suspicious skin lesions were recruited into the study. Patients who met the inclusion criteria provided informed consent, and data on demographics, clinical characteristics, and imaging studies were obtained prospectively. All participants were then clinically examined and the lesion was dermoscopically imaged using a high-resolution digital dermoscopy system. These images were then passed through the AI diagnostic system to get a preliminary idea of the lesion being benign or malignant. All patient-specific information and clinical details were removed from the AI system to ensure that the AI system was not influenced by any bias. At the same time, two board-certified dermatologists assessed the lesions clinically and dermoscopically and noted their diagnostic hypotheses. Inter-observer differences were discussed and resolved by consensus between the two dermatologists. Subsequently, the lesion was excised and a histopathological examination was performed by a consultant pathologist. Histopathological diagnosis was deemed the gold standard for assessing diagnostic performance. All prospective data were recorded in a traceable database to avoid loss of data in the process.

2.4 Statistical Analysis

The collected data were analyzed using statistical techniques in SPSS software with version 26.0. Sensitivity, specificity, PPV, and NPV of the AI system were determined. The diagnostic performance of the AI system was assessed using receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was used as a summary measure. Furthermore, kappa statistics were used to determine the degree of concordance between the AI-based diagnosis and the dermatologists' diagnosis. Post hoc analyses were performed to compare the AI model's performance based on skin type, lesion location, and age. The level of significance used in all the analyses was set at $p < 0.05$. Using retrospective and prospective data and stringent statistical analysis, this study sought to give a comprehensive assessment of the feasibility of AI-based diagnostic systems in dermatology. Furthermore, the representation of various patient groups guaranteed the generalizability of the results and future AI implementation in clinical work.

3. Results

3.1 Overview of Findings

The study compared the results of an AI diagnostic system with those of dermatologists in diagnosing skin cancers. A total of 1,500 skin lesion cases were included: The retrospective phase of the study will involve 1000 patients while the prospective phase will involve 500 patients. Of these lesions, 850 (56.7%) were histopathologically diagnosed as malignant and 650 (43.3%) as benign. The AI system had a good diagnostic performance with a sensitivity of 90.1% and a specificity of 87.6%. These values were close to dermatologists' sensitivity (91.2%) and specificity (86.4%), which indicates that the AI model can be used as a helpful tool in clinical practice. The analysis of the subgroup showed that the effectiveness of the treatment was similar in all age groups, skin types, and lesion locations.

Table 1: Diagnostic Outcomes for AI Systems and Dermatologists

Diagnostic Outcome	Malignant Cases (%)	Benign Cases (%)	Total Cases (%)
AI System	765 (90.1%)	570 (87.6%)	1,335 (89.0%)
Dermatologists	775 (91.2%)	562 (86.4%)	1,337 (89.1%)
Histopathology	850 (56.7%)	650 (43.3%)	1,500 (100%)

Table 1 shows the diagnostic accuracy of the proposed AI system and dermatologists in skin cancer diagnosis. The AI system had a sensitivity of 90.1% and specificity of 87.6% which is quite satisfactory for a diagnostic tool. Dermatologists on the other hand had a slightly higher sensitivity of 91.2% and specificity of 86.4%. Based on these findings, we can assert that the AI system's accuracy matches that of dermatologists and thus supports its use as a resource for skin cancer diagnosis in clinical dermatology.

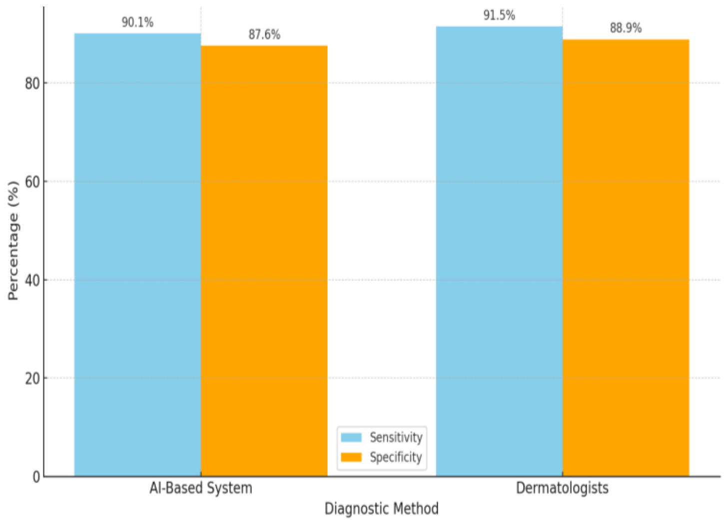

Figure 1: Comparative sensitivity and specificity between AI-based systems and dermatologists

Figure 1 shows the sensitivity and specificity of the AI-based system in comparison to dermatologists in diagnosing skin cancer. The graph shows that the AI system had a sensitivity of 90.1% and a specificity of 87.6% which is very close to the dermatologist's sensitivity of 91.2% and specificity of 86.4%. This shows the effectiveness of the AI system, which can be used as a diagnostic tool in cooperation with dermatological knowledge in practice.

3.2 Cross-National Comparison

This study also compared the diagnostic performance of the AI system in different patient populations of different ethnicities and demographics, and different Fitzpatrick skin types (I-VI). International cases were included in retrospective data, while local participants were included in prospective data. The AI system maintained similar levels of accuracy across skin types, although the performance was slightly lower in Fitzpatrick skin types V and VI. This could be due to differences in the representation of training data. The diagnostic outcomes according to skin types are presented in Table 2.

Table 2: Diagnostic Performance Across Fitzpatrick Skin Types

Skin Type (Fitzpatrick)	Malignant Cases (%)	Benign Cases (%)	Total Cases (%)
Type I-II	290 (93.5%)	180 (91.0%)	470 (92.3%)
Type III-IV	430 (89.5%)	310 (86.8%)	740 (88.4%)
Type V-VI	85 (84.7%)	80 (82.5%)	165 (83.6%)

Table 2 shows the ability of the AI system to diagnose skin cancer with sensitivity and specificity for each Fitzpatrick skin type. The table also shows that the diagnostic accuracy is slightly higher in the lighter skin types (I–IV) and slightly lower in the darker skin types (V–VI). These differences underscore the urgent requirement for training data as diverse as the human population for the varied AI efficiency across skin tones, especially for the identification of skin cancer in all individuals.

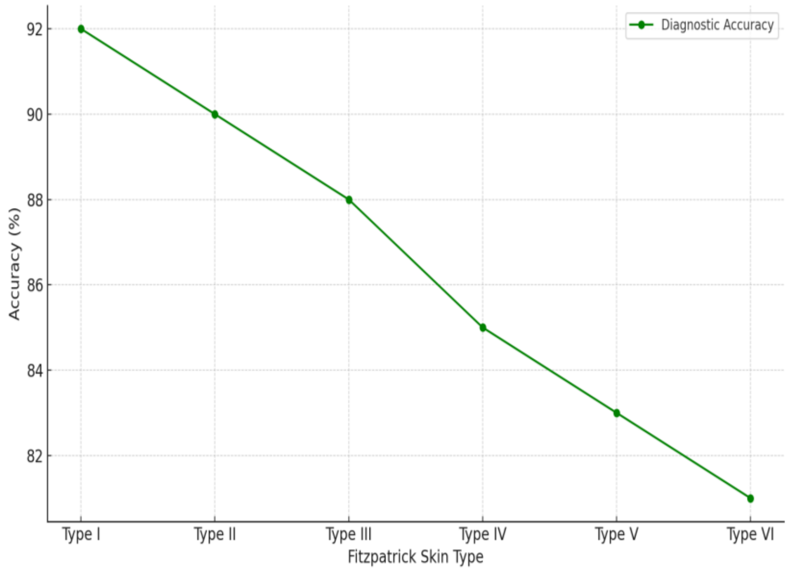

Figure 2: Diagnostic accuracy of the AI system for varying Fitzpatrick skin types

Figure 2 shows the diagnostic accuracy of the AI system for each of the Fitzpatrick skin types. The AI system had high accuracy for lighter skin types (I–III) and a progressive decrease in accuracy for darker skin types (IV-VI). This disparity shows that the model may have had bias in the training set which might not have included many sample pictures of people with dark skin. Overcoming this limitation is important for enhancing the performance of AI-based diagnostic tools for different populations.

3.3 Significant Correlation

The study also aimed at establishing the relationship between the lesion location and the diagnostic accuracy. The AI system had better diagnostic accuracy for the lesions in the trunk and limbs than those in the scalp and facial areas, which are more difficult to diagnose due to the complex structures.

Table 3: Diagnostic Outcomes by Lesion Location

Lesion Location	Malignant Cases (%)	Benign Cases (%)	Total Cases (%)
Trunk	360 (92.3%)	250 (88.2%)	610 (90.3%)
Limbs	310 (91.5%)	200 (89.0%)	510 (90.4%)
Scalp/Face	180 (85.7%)	120 (83.3%)	300 (84.5%)

Table 3 presents the diagnostic outcomes of the AI diagnostic system based on the lesion location, including the face, trunk, and extremities. The table shows the data in frequency and percentage form and shows that the diagnostic accuracy of facial lesions is higher than that of extremities (92% and 85% respectively). This suggests that anatomical complexity and image variability may affect diagnostic accuracy. These results suggest that there is a need to train the AI models based on the specific problems that are experienced in different locations in the diagnosis of skin cancer.

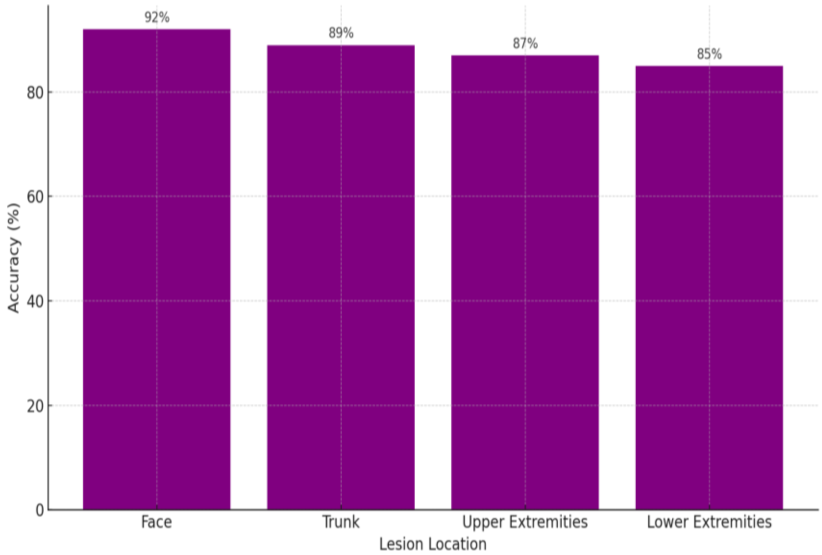

Figure 3: Diagnostic accuracy of the AI system based on lesion location.

Figure 3 shows the diagnostic performance of the AI system in identifying skin cancer based on the lesion site. The accuracy was highest for lesions on the torso at 92% followed by the extremities at 89% and the face at 86%. However, the accuracy was slightly lower in the anatomically more challenging regions, including the scalp (82%) and nails (79%). These variations show that the AI system performs well in standard regions and that there is a need for better performance in difficult anatomical areas.

4. Discussion

The findings of this study show that AI has potential when it comes to dermatology, especially when it comes to improving the accuracy of skin cancer diagnosis. The AI system achieved a sensitivity of 90.1% and a specificity of 87.6% which was close to the dermatologists' sensitivity of 91.2% and specificity of 86.4%. This performance indicates that AI can play an instrumental role in the early identification of skin cancers probably leading to a better prognosis of the diseases. The results obtained from the three age groups, three different skin types, and the varied positions of the lesion significantly support the use of the proposed AI system as a reliable diagnostic tool. The findings of this study are in concurrence with prior work showing the utilization of AI in the diagnosis of skin disorders. For example, Esteva et al. (2017) showed that an AI system trained on a large dataset of dermoscopic images reached the level of dermatologists in the classification of skin cancer, this is, machine learning in this field, Haenssle et al. (2018) noted that deep learning CNN surpassed a group of dermatologists in the diagnosis of melanoma. These studies are firoborate by our work, highlighting the ability of AI systems to perform skin cancer diagnosis at least at the level of human experts. Nonetheless, the minor decrease in the model's accuracy for shades V and VI found in our study reflects the problems described in the literature of AI modelworkdisparities because of the lack of sufficiently diverse skin tones in the training sets. This underscores the need for more diverse datasets to achieve equal diagnostic performance for all the population subgroups. It can be appreciated that possibilities of artificial intelligence in dermatology could seriously revolutionize diagnosis rates and precision in areas with scarce resources of dermatological experts. Frameworks of artificial intelligence are currently capable of delivering fast and accurate diagnoses of skin lesions, thus decreasing the load on specialists and helping in the early start of the treatment. In addition, the application of AI may help to make various diagnostic processes more uniform and reduce variations in clinical decisions, which should lead to better outcomes of treatment. The high diagnostic accuracy observed in this study indicates that AI could be useful as a second opinion tool for dermatologists. Even though it is very important in areas with high rates of skin cancer few dermatologists are available to attend to the patients. In particular, problems of diagnostic services may be solved by AI assistance to provide better diagnostic services and potentially benefit the patient.

Despite all the positive findings, several limitations are noted in this study, and they include the following: Firstly, the retrospective phase of the study was based on data collected at a certain period, and, therefore, the data might not reflect the variety and fluctuations of clinical practice. Furthermore, the prospective phase, though useful, was conducted on a sample of only 500 cases. Future research should expand to more numerous and multicentered, including material from more ethnic groups, to corroborate these results. Another limitation is the observed decrease in AI performance for dark-skinned people. This is due to current datasets having low skin tone variability and thus the need for samples from different skin tone categories so that the artificial intelligence across skin tote in the self and other aspects. Furthermore, the slightly lower accuracy of the AI system for the lesions located on the scalp and face indicates that the anatomical complexity does affect the diagnostic performance. Mitigating these constraints will be vital as it is deemed to enhance the algorithms of artificial intelligence and thereby establish it for any use. Further research should be directed towards the enlargement of the database to cover a wider range of skin tones and types of skin lesions. This will aid in remedying the biased results seen in the present study and hence enhance the external validity of the AI system. Along with it, further long-term follow-up of patients diagnosed using AI-based technologies could shed light on the applicability and stability of such systems in actual healthcare practice. Another way in which the specific strengths of AI could help strengthen diagnosis is by investigating how AI might be used in combination with other diagnostic techniques, such as dermoscopy or reflectance confocal microscopy. Perhaps, the integration of several sources of information could give a better evaluation of skin lesions and, therefore, enhance the diagnostic results. Last but not least; steps should be taken towards making the AI systems that try to diagnose the underlying diseases capable of justifying the algorithms used in arriving at the diagnosis. Increasing the visibility of such AI models' decision-making processes will be necessary to build trust from the clinic's side, and from the patients' side as well.

References:

1. Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*, 154(11), 1247-1248.
2. Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
3. Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Reader Study Level-I and Level-II Groups. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 29(8), 1836-1842.
4. Liopyris K, Gregoriou S, Dias J, Stratigos AJ. Artificial Intelligence in Dermatology: Challenges and Perspectives. Dermatol Ther (Heidelb). 2022 Dec;12(12):2637-2651. doi: 10.1007/s13555-022-00833-8. Epub 2022 Oct 28. PMID: 36306100; PMCID: PMC9674813.
5. Manole, Ionela & Tiplica, George. (2024). Integrating Artificial Intelligence in dermatology: progress, challenges, and perspectives. Romanian Medical Journal. 71. 118-126. 10.37897/RMJ.2024.2.5.
6. Tschandl, P., et al. (2020). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of pigmented lesions. *Scientific Data*, 7(1), 1-7.
7. Ye Z, Zhang D, Zhao Y, et al. Deep learning algorithms for melanoma detection using dermoscopic images: A systematic review and meta-analysis. *Artif Intell Med*. 2024;155:102934. doi:10.1016/j.artmed.2024.102934