SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 12459-12466
Open Access. Online & Print

DOI: 10.5281/zenodo.20475525
www.sci-cult.com

AUTOMATED CLASSIFICATION MODEL OF SKIN PATHOLOGIES BASED ON VISION TRANSFORMER USING DEEP LEARNING APPLIED TO CLINICAL DERMATOLOGY

Leydi Johana Polo Amador[1*], Sergio Suárez Barajas[2], Yezid Yair García Zambrano[3], Jhon Fredys Linares Amador[4]

[1*]Senior, Unidades Tecnológicas de Santander – UTS ORCID: https://orcid.org/0000-0002-5335-2814
lpolo@correo.uts.edu.co
[2]Asociado, Unidades Tecnológicas de Santander – UTS ORCID: https://orcid.org/ 0000-0002-5074-5997
ssuarez@correo.uts.edu.co
[3]Asociado, Unidades Tecnológicas de Santander – UTS ORCID: https://orcid.org/ 0000-0002-1164-2688
yygarcia@correo.uts.edu.co
[4]Asociado, Unidades Tecnológicas de Santander – UTS ORCID: https://orcid.org/0000-0002-7998-3972
jlinares@correo.uts.edu.co

Abstract - Early diagnosis of dermatological pathologies continues to be a challenge in health systems, especially in contexts with limited access to specialists. In this framework, the present study develops and validates an automated classification model of skin diseases based on artificial intelligence, using a Vision Transformer (ViT) architecture and computer vision techniques for the categorization of 14 classes of dermatological lesions. The methodology adopted an experimental approach structured in five stages: planning, data collection and processing, model development, validation and deployment of the system. Standardized public datasets were used, specifically marmal88/skin_cancer and ahmed-ai/skin-lesions-classification-dataset, which were integrated and preprocessed for model training. Optimization strategies such as mixed-precision training (fp16) and class imbalance management techniques were incorporated in order to improve performance in multiclass scenarios. The proposed model achieved an accuracy of 92.9% and a weighted F1 score of 92.85%, evidencing a robust performance in the classification of skin lesions, validated by standard metrics such as the confusion matrix. In addition, the system includes a visual interface developed in Python, aimed at facilitating its use by non-specialized personnel. The results obtained demonstrate that the use of Vision Transformers in dermatological diagnosis constitutes a scalable and effective solution, particularly in environments with limited access to dermatologists, such as rural areas. This work contributes to the development of digital health tools based on artificial intelligence, contributing to the reduction of diagnostic subjectivity and the strengthening of techno-social strategies in the clinical field.

Keywords: *Deep learning; Vision Transformer; Dermatological diagnosis; Medical Image Classification; Digital Health*

Introduction

Dermatological diseases constitute a relevant challenge for health systems, especially in contexts with limited access to specialists. The early diagnosis of pathologies such as melanoma depends largely on visual evaluation, which introduces variability and subjectivity among professionals (Haenssle et al., 2018). In this scenario, artificial intelligence has demonstrated high potential through the use of convolutional neural networks and Vision Transformers, improving diagnostic accuracy and efficiency (Martorell et al., 2022; Flosdorf et al., 2024). Recent studies report performances of more than 85% in the classification of skin lesions, evidencing their ability to match or exceed, in certain cases, the traditional clinical evaluation (Salas Padilla, 2024; Ghanta et al., 2023).

However, challenges persist such as the imbalance of classes in datasets and the need for interpretable and accessible systems for clinical environments (Martínez Yong, 2023; Ghanta et al., 2023). In countries such as Colombia, where the availability of dermatologists is limited in rural areas, these technologies represent an opportunity to strengthen diagnosis in primary care (Lacouture Fierro, 2025; Zhang et al., 2025). In this context, the present study develops a predictive model based on Vision Transformers to classify 14 types of skin lesions from dermatological images, complemented by an intuitive interface in Python. The research takes an experimental approach using standardized public datasets, organizing the process into phases of data preparation, development, validation, and deployment. The model achieved an accuracy of 92.9% and a weighted F1-score of 92.85%, evidencing a robust performance. This work provides a scalable solution that reduces diagnostic subjectivity and facilitates its use in contexts with limited resources, contributing to the advancement of digital health and the generation of techno-social solutions applied to the clinical field.

PROBLEM:

Dermatological diseases are a major public health problem due to their high prevalence and the need for timely diagnosis to prevent the progression of potentially serious pathologies, such as melanoma. In clinical practice, the identification of these conditions depends largely on the experience of the specialist, which introduces diagnostic variability and limits the capacity to respond at primary care levels. This situation is aggravated by the insufficient dermatological training of general practitioners, which can lead to delays in early detection and, consequently, negatively affect the patient's prognosis.

In the Colombian context, the problem is intensified due to the limited availability of dermatologists, especially in rural areas, where approximately one specialist is reported for every 100,000 inhabitants (Ministry of Health and Social Protection, 2022). This gap in access to specialized services contributes to disparities in care and hinders the implementation of effective early diagnosis strategies. At the clinical level, traditional methods based on visual inspection have inherent limitations, as they are subject to the subjectivity of the assessor and interobserver discrepancies, particularly in lesions with similar characteristics (Haenssle et al., 2018; Dildar et al., 2021).

In response to these limitations, artificial intelligence, and specifically deep learning, has emerged as a promising alternative for automating dermatological diagnosis. Recent studies have shown that models based on convolutional neural networks can achieve levels of accuracy of more than 90% in the detection of skin cancer, even matching or surpassing the performance of specialists in controlled scenarios (Dildar et al., 2021). However, the implementation of these technologies faces technical and operational challenges, such as the imbalance of classes in datasets, the need for accessible interfaces for non-expert users, and compliance with regulatory frameworks in the handling of sensitive data, as established by Law 1581 of 2012 in Colombia.

In this context, there is a need to develop predictive models based on artificial intelligence that integrate advanced computer vision and deep learning techniques, in order to improve accuracy, reduce subjectivity and facilitate access to diagnosis in resource-limited environments. From the above, the following research question arises: how can artificial intelligence improve the timely and accurate diagnosis of dermatological diseases compared to traditional methods?

OBJECTIVES

General Objective

To develop a predictive model based on deep learning for the automated classification of dermatological diseases from images, in order to improve diagnostic accuracy and support clinical decision-making.

Specific Objectives

1. Collect and pre-process a representative dermatological dataset, ensuring its quality, standardization, and suitability for model training.
2. Design and implement a deep learning model based on Vision Transformers for the classification of skin pathologies from images.
3. Evaluate model performance using classification metrics and develop a user interface that allows its use and interpretation in clinical contexts.

State of the art

At the international level, the use of artificial intelligence (AI) in the field of health has grown significantly, especially in disease diagnosis and prevention processes. Various studies agree that the true potential of these technologies does not lie in replacing health professionals, but in complementing their work. In this sense, Martorell et al. (2022) highlight that the effective application of AI in medicine is achieved when there is an integration between human clinical knowledge and the analytical capabilities of intelligent systems. Similarly, Martínez Yong (2023) points out that techniques such as machine learning and deep learning make it possible to identify complex patterns in medical images that are not easily perceptible to the naked eye. However, he also warns that one of the main challenges is the limited training in these technologies by health personnel.

In the specific field of dermatology, multiple studies have demonstrated the effectiveness of models based on deep learning. For example, Salas Padilla (2024) developed a platform based on convolutional neural networks (CNNs) for the detection of common skin conditions, reaching an accuracy of 88.01%. These types of results confirm the ability of CNNs to recognize complex visual patterns in dermatological images. Along the same lines, Sánchez (2024) and Fernández Morales (2024) proposed predictive models for the early detection of skin cancer using images from the ISIC Challenge, incorporating techniques such as data augmentation, oversampling and hyperparameter optimization. Their findings highlight the importance of model assembly as a strategy to improve overall performance.

More recent research has explored more advanced architectures and optimization techniques. Molina Centeno (2025) developed a model focused on the detection of different types of skin lesions, while Molina Pérez and Vivas Toro (2025) implemented models based on DenseNet121, integrating techniques such as fine-tuning, focal loss and Mixup, achieving an accuracy of 86.47% and an average AUC of 0.96. These results demonstrate not only the predictive capacity of current models, but also the importance of combining advanced image processing techniques to improve discrimination between classes.

In the Colombian context, relevant contributions have also been made in this area. Altuve Mouthon and Colón Muñoz (2021) proposed a melanoma detection model based on deep learning, evidencing improvements in diagnostic accuracy. Similarly, Flórez Fuentes et al. (2022) achieved results close to 100% in controlled scenarios, which demonstrates the high potential of these models when trained with well-structured and labeled datasets.

More recently, Santiago Rodríguez et al. (2024), from the Universidad de los Andes, developed a solution that integrates computer vision and machine learning in a mobile application for the early diagnosis of skin cancer. This approach facilitates remote access to image analysis, strengthening care in contexts with limited access to specialists. Along the same lines, Duque Ortiz and León Chávez (2024) identified in their review that AI models can reach levels of accuracy close to 99% in the detection of dermatological diseases. For his part, Molano Muñoz (2024) evaluated architectures such as MobileNet, Xception, and ResNet101V2, highlighting their effectiveness in identifying malignant lesions and their potential for applications in preventive diagnostics.

Together, these studies show significant advances in the application of artificial intelligence in dermatology. However, they also highlight the need to continue developing more robust, interpretable and accessible models, which can be used in real environments and contribute to improving the timely diagnosis of skin diseases.

THEORETICAL FRAMEWORK:

Convolutional neural networks (CNNs) operate by using filters or *kernels* that move over the image, performing product point operations between the filter and local regions of the image. This process allows you to generate feature maps that capture relevant information in a hierarchical manner. From a mathematical perspective, two-dimensional convolution is expressed as:

Figure 1. CNN Math in Deep Learning

$$(f * g)(i,j) = \sum_m \sum_n f(m,n)\, g(i-m, j-n)$$

where fff represents the input image and ggg the convolutional filter. This operation facilitates the progressive extraction of features, from simple patterns such as borders and textures in the first layers, to more complex representations at higher levels. In the dermatological field, CNNs have been widely used for the classification of skin lesions, reaching accuracies of more than 90% in datasets such as ISIC with architectures such as Inception-v3 (Esteva et al., 2017). However, its main limitation lies in its local nature, which makes it difficult to capture global dependencies in the image, a critical aspect in the analysis of lesions where the spatial distribution of color and shape is determinant.

In order to overcome these limitations, Transformers emerged, initially proposed by Vaswani et al. (2017) in the field of natural language processing. These architectures are based on the *self-attention* mechanism, which allows the model to assign dynamic weights to different parts of the input, capturing global relationships efficiently. This mechanism is based on the calculation of three matrices: Query (Q), Key (K) and Value (V), obtained by linear transformations of the input. The function of attention is defined as:

Figure 2. Mathematical Foundations of Self-Attention in Transformers

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where dkd_kdk corresponds to the dimension of the keys. The *softmax* function normalizes attention weights, while the dk\sqrt{d_k}dk scale factor contributes to numerical stability during training. This mechanism allows each element of the sequence to interact with all the others, capturing long-range dependencies. Although its computational complexity is O(n2)O(n^2)O(n2), the architecture allows for high parallelization, which makes it efficient in GPU environments.

From this base, Dosovitskiy et al. (2020) introduced the Vision Transformer (ViT), adapting the Transformers paradigm to the domain of computer vision. In this approach, images are divided into fixed-size patches (e.g., 16×16 pixels), which are transformed into vectors and treated as token sequences. These tokens are processed by encoder-like layers, where self-attention allows global relationships between different regions of the image to be modeled. To preserve spatial information, positional *embeddings* are incorporated , and a special token [CLS] is used for the final classification.

The ViT model has demonstrated comparable and even superior performance to CNNs in image classification tasks, especially when trained on large volumes of data. Its ability to integrate local and global information makes it an especially suitable alternative for dermatological image analysis, where both fine details and the overall structure of the lesion are critical for an accurate diagnosis.

Research Design

The present study adopts a quantitative approach **of experimental and applied type**, aimed at the development and validation of an automated classification system of dermatological pathologies based on deep learning. The research is based on the analysis of data derived from labeled dermatological images, using performance metrics such as accuracy and F1-weighted *F1 score* to evaluate the effectiveness of the model.

From an applied perspective, the objective is to design a technological solution that responds to a specific need in the clinical field: to improve the accuracy and timeliness of dermatological diagnosis through artificial intelligence techniques. In this sense, the study not only focuses on the construction of the model, but also on its validation and potential implementation in real contexts.

The development of the system follows an experimental approach, based on the construction, training and iterative evaluation of deep learning models, specifically using the Vision Transformer (ViT) architecture. The process runs in a GPU-accelerated computational environment, which allows performance to be optimized in both image processing and model training, ensuring efficiency and scalability.

Scientific method

The research is based on an iterative and incremental methodological approach, which integrates principles of the scientific method with agile development practices. This approach allows a progressive evolution of the system, facilitating the continuous validation of the results and the improvement of the model based on empirical evidence.

From an epistemological point of view, an inductive approach is adopted, starting from the analysis of dermatological datasets to identify patterns and build a model capable of classifying 14 types of skin lesions. In addition, an analytical approach is employed, whereby the development process is broken down into structured phases that include data preparation, model design, training, evaluation, and validation.

This methodological combination makes it possible to guarantee coherence between the problem posed, the proposed solution and the results obtained, ensuring both the scientific validity and the practical applicability of the developed system.

Software Development Lifecycle

The development of the system was structured under an **iterative and incremental approach**, inspired by agile methodologies such as Scrum, which allowed a progressive evolution of the model and its continuous validation. This approach facilitated the adaptation of the system to the requirements of the problem and the optimization of performance in each phase of the process.

The development life cycle was organized into the following stages:

1. **Requirements analysis:**
 Functional and non-functional requirements of the system were identified and defined, including accurate image classification in 14 dermatological classes, data imbalance management, and performance optimization in GPU-accelerated environments. This phase made it possible to establish the technical and operational bases of the model.
2. **System design:**
 The overall architecture of the system was defined, integrating the Vision Transformer (ViT)-based model, image preprocessing processes, and user interface design. In this stage, the data flows and the interaction between the different components of the system were established.
3. **Implementation:**
 The system was developed in a modular way,

incorporating components for data loading and processing, model training, evaluation of results and deployment. This structure allowed for improved scalability and facilitated future modifications or expansions.

4. **Testing:**

The validation of the system was carried out through unit, functional and performance tests, with the aim of guaranteeing the correct execution of each component and the stability of the model in different scenarios.

5. **Deployment:**

The integration and packaging of the trained model was carried out together with the graphical interface, allowing its use in local environments and potentially in clinical contexts. This phase ensured the operational availability of the system.

6. **Maintenance:**

A continuous improvement process was established based on the feedback of the system, including the optimization of the model, the incorporation of new data and the updating of the interface, in order to guarantee its sustainability and evolution over time.

Data collection techniques

Data collection was aimed at guaranteeing the quality, representativeness, and relevance of the information used in the development of the classification model. To do this, open and standardized data sources were used, widely used in computer vision research applied to dermatology.

In particular, two datasets available on the Hugging Face platform were used:

1. *marmal88/skin_cancer*, which contains pre-labeled dermatological images.
2. *ahmed-ai/skin-lesions-classification-dataset*, which includes images classified into multiple categories of skin lesions.

In order to ensure consistency in the training process, both datasets were integrated through a process of unification and standardization of labels, allowing the consolidation of a common scheme of 14 diagnostic classes, including melanoma, basal cell carcinoma and actinic keratosis, among others.

This integration process was implemented at the software level, guaranteeing the semantic consistency of the classes and facilitating the training of the model in a homogeneous environment. In addition, the use of multiple data sources contributed to improving the diversity of samples and the generalizability of the model.

Technical elements of development:

The system was implemented in Python, using PyTorch and the Transformers library (Hugging Face), with GPU acceleration (CUDA). NumPy and Scikit-learn were used for data processing and evaluation.

Optimization techniques such as mixed-precision (fp16) and GPU memory management were applied to improve performance and reduce resource consumption.

The modular architecture of the system allows scalability, facilitating the integration of new datasets and extension to other diagnostic classes in future developments.

RESULTS

Developing Results

Python modules were developed with PyTorch and Hugging Face:

1. Upload and unification: load_and_process_datasets function for combining datasets and mapping labels.
2. Preprocessing: ViTImageProcessor to standardize images at 224x224 pixels.
3. Model: Custom WeightedTrainer with weights calculated by class_counts.
4. Interface: Developed with libraries such as Gradio or Tkinter for image loading and result display.

The code was managed with Git, ensuring versioning. Approximate duration: 3 weeks.

***Table 1** Project development environment*

Component	Tool	Description
Language	Python	Support for machine learning with PyTorch.
Framework	Transformers (Hugging Face)	Pre-trained models (ViT) and processing tools.
Backend	PyTorch	GPU computing with CUDA, fp16 support.
Datasets	Hugging Face Datasets	Load of dermatological datasets.
Bookstores	NumPy, Scikit-learn	Data processing and metrics.
Version Control	Git	Source code management.

Source: Authorship

Transformers (Hugging Face)

Loading the Dermatological ImageTransformers (Hugging Face)

Upload an image: Click the Select Image button and choose a file in JPG, PNG, or GIF format from your device. The image will appear in the interface for review, allowing you to confirm that it is correct before starting the analysis. You can change the image if necessary.

Figure 3. Selection of the dermatological image

Source: Authorship

Figure 4 Previewing the Image in the Interface

Source: Authorship

The selected image is base64-encoded on the frontend and sent to the backend via a POST request to the /predict endpoint. On the server, it is decoded and processed with ViTImageProcessor, which resizes the image to 224×224 pixels and converts it into an RGB tensor. Subsequently, pixel normalization (mean and standard deviation of 0.5) is applied to adjust it to the requirements of the ViT model. Preprocessing is optimized with small batches (batch_size=16) to reduce GPU memory consumption. On the frontend, the image is visualized using a component <img> managed with React, ensuring a correct user experience. In addition, the file size is validated, restricting uploads greater than 10 MB to avoid overload in transmission.

The interface incorporates a load indicator (spinner) to signal processing in progress. Asynchronous communication is managed by axios/fetch, with error handling for *timeouts* and connection failures. The request includes headings that specify the format of the image, ensuring its correct interpretation in the backend.

At the end of the analysis, the interface presents the predicted class, probability distribution for the 14 categories, and image metadata (resolution and format), along with technical information about the model. In addition, a brief clinical description and a risk level (low, medium, high) are included to facilitate the interpretation of the result by the user.

Conclusions

The development of the predictive model based on artificial intelligence allowed the construction of a robust system for the automated classification of 14 dermatological pathologies, reaching an accuracy of 92.9% and a weighted F1-score of 92.85%. These results demonstrate the effectiveness of the Vision Transformer (ViT) architecture to capture complex patterns and global dependencies in medical images, overcoming limitations associated with the subjectivity of traditional clinical diagnosis.

The model's performance was strengthened by the integration of standardized public datasets and the application of preprocessing techniques, such as normalization, data augmentation, and label unification. Likewise, the implementation of optimization strategies, including weighted loss functions and mixed-precision training (fp16), made it possible to mitigate class imbalance and optimize the use of computational resources, making the system viable even in environments with hardware limitations.

During the process, relevant challenges were identified, such as variability in labels and high resource consumption in training. These were addressed through standardization processes and optimization techniques that allowed an efficient evaluation, with average inference times close to 1.5 seconds per image. The analysis of results showed superior performance in well-represented classes, although opportunities for improvement are evident in less frequent categories, suggesting the need to strengthen the data balance in future iterations.

A significant contribution of the project is the integration of an intuitive visual interface that facilitates interaction with the model, allowing the loading and analysis of images in real time. This feature expands its applicability in clinical contexts, especially in regions with limited access to specialists, such as rural areas of Colombia, contributing to reducing gaps in timely diagnosis.

In general terms, the developed system represents a scalable and applicable solution in the field of digital health, aligned with the need to strengthen diagnostic support tools. However, limitations related to validation in real clinical environments and performance in unbalanced classes are identified, which raises the future work of expanding the dataset, optimizing the model and its integration into telemedicine platforms. In this sense, the research consolidates the potential of artificial intelligence as a complementary tool in dermatology, promoting advances towards more accessible, accurate and ethically responsible systems

BIBLIOGRAPHY

1. Altuve Mouthon, A. D., & Colón Muñoz, E. E. (2021). *Modelo de predicción de detección de melanomas a partir de características en imágenes de lesiones cutáneas* [Trabajo de grado, Universidad del Sinú].
2. Carreño Gómez, D. A., & Portilla González, G. A. (2023). Desarrollo de una interfaz de clasificación de enfermedades dermatológicas basadas en inteligencia artificial.
3. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health, 18*(10), 5479. https://doi.org/10.3390/ijerph18105479
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv*. https://doi.org/10.48550/arXiv.2010.11929
5. Duque Ortiz, J. D., & León Chávez, Á. F. (2024). *Herramientas de inteligencia artificial en diagnóstico médico* [Trabajo de especialización].
6. Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*, 115–118. https://doi.org/10.1038/nature21056
7. Fernández Morales, E. (2024). *Análisis de cáncer de piel mediante inteligencia artificial* [Trabajo académico].
8. Flórez Fuentes, A. S., Guzmán Cabrera, R., & Vargas Rodríguez, E. (2022). Identificación automática de cáncer de piel aplicando machine learning. *Revista Colombiana de Tecnologías de Avanzada, 2*(40), 1–6. https://doi.org/10.24054/rcta.v2i40.2350

9. Flosdorf, C., et al. (2024). Skin cancer detection using deep learning with Vision Transformers. *arXiv*. https://arxiv.org/abs/2407.18554
10. Ghanta, S. K., et al. (2023). LesionAid: Vision Transformers-based skin lesion classification. *arXiv*. https://arxiv.org/abs/2302.01104
11. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
12. Haenssle, H. A., Fink, C., Schneiderbauer, R., et al. (2018). Diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *Annals of Oncology, 29*(8), 1836–1842.
13. Lacouture Fierro, J. D. (2025). *Detección temprana de melanoma mediante aprendizaje profundo* [Tesis de maestría].
14. Leal, N. E., & Rojas, I. A. (2024). Sistema de detección de melanoma mediante inteligencia artificial. *Repositorio UNAB*.
15. Maglogiannis, I., & Doukas, C. (2009). Overview of computer vision systems for skin lesion characterization. *IEEE Transactions on Information Technology in Biomedicine, 13*(5), 721–733.
16. Martínez Yong, D. (2023). Inteligencia artificial aplicada a la dermatología.
17. Martorell, A., Martin-Gorgojo, A., Ríos-Viñuela, E., et al. (2022). Inteligencia artificial en dermatología: ¿amenaza u oportunidad? *Actas Dermo-Sifiliográficas, 113*(1), 30–46.
18. Molano Muñoz, L. T. (2024). *Clasificación de cáncer cutáneo mediante técnicas de análisis de datos* [Trabajo de grado].
19. Molina Centeno, M. A. (2025). *Modelo de inteligencia artificial para detección de lesiones cutáneas* [Trabajo de titulación].
20. Molina Pérez, A. M., & Vivas Toro, S. C. (2025). Inteligencia artificial en dermatología moderna. *Revista Piel, 2*(1), 1–24.
21. Salas Padilla, A. E. (2024). *Plataforma de IA para detección de enfermedades de la piel* [Trabajo de grado].
22. Sánchez, A. (2024). Modelo predictivo para detección de cáncer de piel con redes neuronales.
23. Santiago Rodríguez, Y. C., Romero Sarmiento, W. M., Leyton López, F. V., & Cristancho Cruz, S. (2024). Diagnóstico temprano de cáncer de piel mediante IA.
24. Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Grad-CAM: Visual explanations from deep networks. *ICCV*. https://doi.org/10.1109/ICCV.2017.74
25. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *NeurIPS*. https://doi.org/10.48550/arXiv.1706.03762
26. Xin, C., Liu, Z., Zhao, K., et al. (2022). Transformer networks for skin cancer classification. *Computers in Biology and Medicine, 149*, 105939.
27. Zhang, X., Liu, Y., Ouyang, G., et al. (2025). DermViT: Vision transformer for skin lesion classification. *Bioengineering, 12*(4), 421.