Implementing Vision Transformers in Dermatological Practice: A Web Application for Melanoma Screening

Daniele Sirico[1,†], Giuseppe Accardo[1,†] and Valentina Esposito[1,*]

[1]*Data Jam srl, Centro Direzionale Isola F8, Via F. Lauria, Naples, 80143, Italy*

Abstract

In this study, we introduce a pioneering web-based application designed to enhance melanoma detection accuracy through the innovative use of a Vision Transformer (ViT) model. Leveraging the power of advanced deep learning architectures, our application processes dermatological images to identify potential melanomas with a precision previously unattainable in conventional screening methods. The development process involved finetuning the ViT model on a diverse dataset of dermatoscopic images, ensuring robustness and reliability across a wide range of images. The web application is intuitively designed, allowing for easy access and use by dermatologists and potentially by the general public for preliminary screening purposes. This research not only underscores the viability of ViT models in medical imaging but also offers a practical tool for early melanoma detection, thereby contributing to better clinical outcomes and facilitating early treatment interventions.

Keywords

AI, Melanoma, Computer Vision, Vision Trasformer

1. Introduction

The early diagnosis of melanoma, a highly aggressive form of skin cancer, is crucial for improving patient outcomes [1, 2, 3]. When detected at an early stage, melanoma can often be treated effectively, significantly reducing mortality rates [1]. However, the challenge lies in the timely and accurate identification of potential melanomas among a vast array of skin lesions, which requires a high level of expertise and experience. In this context, the application of artificial intelligence (AI) in medical imaging has emerged as a groundbreaking advancement [4, 5]. AI, particularly deep learning models, has shown remarkable success in enhancing the accuracy and efficiency of diagnostic processes in various medical fields [6, 7], including dermatology [8].

The integration of AI into medical imaging for melanoma detection allows for the analysis of dermatological images with a level of detail and precision that surpasses human capability [9]. This not only aids dermatologists in making more informed decisions but also has the potential to democratize access to high-quality diagnostic services, especially in under-resourced areas. Furthermore, the advent of user-friendly web applications for medical purposes represents a significant leap forward. Applications built on platforms like Streamlit offer an accessible interface for both medical professionals and the general public, facilitating widespread screening and awareness.

Such web applications can transform smartphones and personal computers into powerful tools for preliminary screening, empowering individuals to seek professional advice at the earliest suspicion of melanoma. This approach to healthcare leverages the ubiquity of internet-connected devices to bridge the gap between advanced diagnostic technologies and end-users [10]. The ease of use and accessibility of these applications are critical factors in their adoption and effectiveness, enabling timely intervention and potentially saving lives. In sum, the convergence of AI in medical imaging and user-friendly web applications marks a pivotal moment in the fight against melanoma, offering new horizons for early detection and treatment.

To date, the ABCDE method is the standard approach used by medical professionals for the diagnosis of melanoma, emphasizing the evaluation of lesions based on Asymmetry, Border irregularity, Color variation, Diameter larger than 6mm, and Evolution. Despite its widespread adoption and utility in raising awareness, this method's subjective nature can lead to variability in diagnostic accuracy, potentially overlooking early-stage melanomas or prompting unnecessary biopsies of benign lesions [1]. In response, Artificial Intelligence, especially through deep learning algorithms, presents a significant advancement by providing an objective and precise analysis far surpassing the traditional methods. By incorporating AI-driven diagnostic capabilities into a user-friendly web application, the project aims not only to enhance diagnostic precision but also to make sophisticated melanoma screening tools accessible to a wider population. This initiative marks a crucial step forward in improving early detection rates and patient outcomes by

Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy
* * *Corresponding author.
† These authors contributed equally.
dg.sirico@almaviva.it (D. Sirico); gi.accardo@almaviva.it (G. Accardo); v.esposito@almaviva.it (V. Esposito)
0000-0002-2760-9209 (D. Sirico)

bridging the gap between traditional diagnostic methods and the potential of modern technology [11, 12].

In this study, we have developed a comprehensive dataset derived from the ISIC (International Skin Imaging Collaboration) Challenge datasets of 2019-2020. Utilizing this dataset, we have fine-tuned the pre-trained Vision Transformer (ViT) Large model provided by Google [13], with the specific objective of classifying dermatological images into two distinct categories: "Melanoma" and "Non-Melanoma". The model has been meticulously adapted to the unique characteristics of dermatological imagery. While the ViT model inherently possesses a broad capability for image recognition owing to its extensive initial training on diverse data, our fine-tuning process has significantly enhanced its accuracy and sensitivity to the specific features of skin lesion images. This targeted refinement improves the model's diagnostic precision, making it a highly effective tool for distinguishing between melanoma and non-melanoma cases. By leveraging the cutting-edge ViT architecture and tailoring it to the nuances of dermatological conditions, we aim to advance the field of medical imaging and offer a more accurate, AI-driven approach to melanoma detection.

2. Dataset Preparation

In this work, we have constructed a dataset starting from the data of the ISIC Challenges of 2019 [14, 15, 16] and 2020 [17], with which we fine-tuned Google's pre-trained Vision Transformer (ViT) Large model, aiming to distinguish between two classes: "Melanoma" and "No Melanoma". The model has been specifically adapted to the characteristics of dermatological images. Although the network already possesses a broad capacity for recognition due to its prior training, fine-tuning allows us to refine its precision and sensitivity to the peculiarities of skin lesions, thereby improving the accuracy of diagnoses.

For the dataset preparation, with the objective of achieving balanced classes between Melanoma and No Melanoma, images were extracted from ISIC 2019 and ISIC 2020. The resulting dataset comprises 8,223 training images and 1,450 test images, witnessing a significant increase in the sample size, especially in the number of images within the Melanoma class of interest (3,890 No Melanoma, 4,333 Melanoma). This dataset was synthesized from images sourced from ISIC 2019 and ISIC 2020. Specifically, the "Melanoma" class benefited from contributions from both datasets, while the "No-Melanoma" class was formed using images exclusively from ISIC 2019. This balanced approach ensures a more equitable distribution of classes, enhancing the model's ability to learn and accurately differentiate between Melanoma and No Melanoma, which is crucial for the model's performance in real-world diagnostic applications.

In addition to the dataset preparation from ISIC 2019 and ISIC 2020, the fine-tuned model was subsequently tested on an entirely different dataset, MEDNODE [18], to assess its generalization capability and performance in real-world scenarios. This step was crucial for validating the effectiveness of our fine-tuning process and ensuring that the model could accurately classify melanoma across diverse datasets. The MEDNODE dataset, distinct in its composition and image characteristics, provided a challenging environment to evaluate the model's robustness and adaptability. This further testing underscores the model's potential for application in a wide range of clinical settings, demonstrating its ability to maintain high levels of accuracy and sensitivity in detecting melanoma, even when confronted with data significantly different from that on which it was trained. This cross-dataset validation is a critical aspect of our research, confirming the model's utility as a reliable tool in the early detection of melanoma, potentially revolutionizing dermatological diagnostics.

3. Fine Tuning and Validation

3.1. Fine Tuning

The concept of transfer learning, applied to the context of melanoma detection using Google's pre-trained Vision Transformer (ViT) Large model, leverages a model developed for general vision tasks and adapts it to the specific challenge of identifying melanoma in dermatological images. This approach benefits from the original model's learning capabilities, significantly reducing the time and resources needed for training from scratch. Thanks to its prior training, the model can be optimized to recognize the specific features of melanoma with greater efficiency.

During the fine-tuning process, the already pre-trained ViT Large model from Google is further optimized for melanoma detection. In this phase, the model is specifically adapted to the characteristics of dermatological images. Although the network already has a broad recognition capability thanks to its previous training, fine-tuning allows for the refinement of its precision and sensitivity to the peculiarities of skin lesions, thus enhancing diagnostic accuracy.

The development environment utilized for training the model features the following specifications in AWS Environment: p3.2xlarge, Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30GHz, 64GB RAM, Graphics Card: Tesla V100-SXM2 with 16GB of VRAM.

For the inference phase, no specific dedicated hardware with GPU is required. This flexibility in hardware requirements for inference ensures that the fine-tuned model can be deployed in a wide range of environments,

Table 1
Performance Evaluation of Skin Lesion Classification on the ISIC Test Set: This table presents a detailed breakdown of classification metrics for distinguishing between melanoma and non-melanoma skin lesions. Each class is evaluated based on precision, recall, F1-score, and support, highlighting the model's ability to accurately identify and classify each condition within the dataset.

Class	Precision	Recall	F1-Score	Support
No-Melanoma	0.86	0.83	0.85	677
Melanoma	0.86	0.88	0.87	773

making it accessible for clinical use without the need for high-performance computing resources.

3.2. Validation on ISIC Test Set

The results obtained from the model on the test set, as detailed in Table 1, underscore its adeptness in distinguishing between melanoma and non-melanoma lesions, showcasing substantial precision, recall, and F1-score metrics across both categories. For the "No Melanoma" class, precision was marked at 0.86, with a recall of 0.83 and an F1-score of 0.85, across a support of 677 cases. This high level of accuracy in identifying non-melanoma instances indicates a strong balance between the precision and recall, reflecting the model's efficiency in minimizing false positives while effectively recognizing true negatives.

Conversely, the "Melanoma" class demonstrated precision and recall scores of 0.86 and 0.88, respectively, achieving an F1-score of 0.87 over 773 instances. These metrics highlight the model's capability in reliably detecting melanoma lesions, with the elevated recall indicating a particular strength in reducing false negatives—crucial for melanoma screening where the cost of missing a positive diagnosis is exceedingly high.

The consistency in precision across both classes, paired with the balanced F1-scores, attests to the model's robustness, suggesting its potential as a dependable tool in the diagnostic toolkit. Such performance, detailed in Table 1, affirms the advanced AI models, like the Vision Transformer's, significant role in dermatological diagnostics.

Further insight into the model's performance is provided by Figure 1, which depicts the confusion matrix and the ROC curve for the model's predictions. The confusion matrix visually illustrates the model's accuracy in classifying the test cases, offering a clear depiction of the true positive and negative rates, alongside the instances of false positives and negatives. The ROC curve, accompanying this matrix, further elucidates the model's diagnostic ability across different thresholds, showcasing its exceptional capability to balance sensitivity and specificity effectively. Together, Table 1 and Figure 1 offer a comprehensive overview of the model's diagnostic performance, highlighting its potential to revolutionize melanoma screening and diagnosis through the integration of cutting-edge AI technologies into clinical practices, thereby enhancing diagnostic accuracy and facilitating improved patient care.

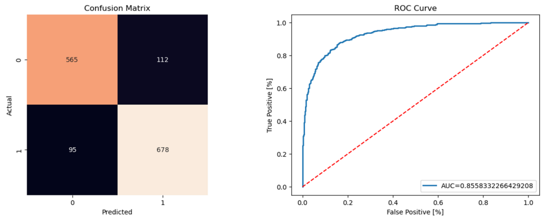

Figure 1: Diagnostic Performance of the Vision Transformer Model for Melanoma Detection on ISIC test set. On the left, the confusion matrix illustrates the model's accuracy in classifying 1: 'Melanoma' and 0: 'No Melanoma' cases, with the number of true positives, true negatives, false positives, and false negatives. On the right, the ROC (Receiver Operating Characteristic) curve.

3.3. Validation on MEDNODE Dataset

The model's performance on the MEDNODE dataset, as detailed in Table 2, reflects its diagnostic precision in distinguishing between melanoma and non-melanoma cases. With a precision of 0.83 and a recall of 0.98 for "No Melanoma," the model demonstrates a high capability in correctly identifying non-melanoma cases, as evidenced by an F1-score of 0.90 across 100 instances. This high recall rate is crucial, indicating the model's strength in minimizing the risk of false negatives in non-melanoma diagnoses, which is vital for avoiding unnecessary further testing and anxiety for patients.

Conversely, for the "Melanoma" category, the precision stands at an impressive 0.96, showing the model's reliability in its melanoma predictions. However, the recall of 0.71 highlights a potential area for improvement in capturing all true melanoma cases, with an F1-score of 0.82 across 70 instances reflecting the balance between precision and the need to improve recall.

Figure 2 provides further insight into these results through a visual representation. The left side of the figure features the confusion matrix, illustrating the model's

Table 2
Performance Evaluation of Skin Lesion Classification on the MEDNODE Dataset: This table presents a detailed breakdown of classification metrics for distinguishing between melanoma and non-melanoma skin lesions. Each class is evaluated based on precision, recall, F1-score, and support, highlighting the model's ability to accurately identify and classify each condition within the dataset.

Class	Precision	Recall	F1-Score	Support
No-Melanoma	0.83	0.98	0.90	100
Melanoma	0.96	0.71	0.82	70

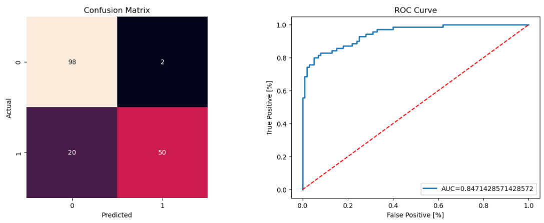

Figure 2: Diagnostic Performance of the Vision Transformer Model for Melanoma Detection on MEDNODE Dataset. On the left, the confusion matrix illustrates the model's accuracy in classifying 1: 'Melanoma' and 0: 'No Melanoma' cases, with the number of true positives, true negatives, false positives, and false negatives. On the right, the ROC (Receiver Operating Characteristic) curve.

accuracy in classifying cases into "Melanoma" and "No Melanoma" categories, revealing the true positive, true negative, false positive, and false negative counts. The right side of the figure displays the ROC curve, showcasing the model's ability to differentiate between the two classes at various threshold levels, thus highlighting the trade-off between sensitivity (true positive rate) and specificity (true negative rate).

Together, Table 2 and Figure 2 offer a comprehensive overview of the model's diagnostic performance on the MEDNODE dataset. They underline the model's strengths in identifying non-melanoma cases with high accuracy while also pointing out the necessity for further refinement to enhance its sensitivity to melanoma detection. This balance between precision and recall, especially for a condition as critical as melanoma, is paramount in developing AI diagnostic tools that can effectively assist clinicians in making accurate and timely diagnoses, thereby improving patient care and outcomes.

In essence, the model's strong performance on the MEDNODE dataset, a collection distinct from the training set provided by the ISIC challenges, enhances its credibility as a robust diagnostic tool. It affirms the potential for AI-driven models, specifically those based on advanced architectures like the Vision Transformer, to revolutionize melanoma detection. By effectively bridging the gap between different imaging sources and maintaining high diagnostic accuracy, the model paves the way for broader clinical adoption, offering a promising tool for early melanoma detection and thereby improving patient outcomes through timely and accurate diagnoses.

4. Streamlit WebApp

In recent years, data visualization has become increasingly crucial in the comprehension and communication of information, especially in the realm of medical image analysis. Streamlit has emerged as a powerful and flexible tool for creating interactive web applications, particularly excelling in the manipulation and analysis of visual data. This chapter delves into the pivotal role of Streamlit in developing a web-based application aimed at melanoma detection through the analysis of dermatological images.

Leveraging the capabilities of Streamlit, the objective was to construct an interactive application directly linked to the melanoma recognition system based on the image models trained and discussed in previous chapters. This application provides an intuitive user interface, allowing users to upload and analyze dermatological images to assess the potential presence of melanoma. Furthermore, the application displays the results from the melanoma recognition system, enhancing the user's understanding and interaction with the diagnostic process. The web application can be segmented into four main sections:

Why: This section elucidates the importance of early melanoma detection, providing users with background information on the significance of timely diagnosis and how advancements in AI and medical imaging have facilitated this process.

How: Here, the application explains the underlying technology and algorithms that power the melanoma detection process, offering insights into the workings of the AI model and the role of deep learning in analyzing dermatological images.

Image Upload: Users are presented with a straightforward mechanism to upload dermatological images. This functionality underscores the application's user-friendly design, ensuring that users can easily navigate the process of submitting images for analysis.

Prediction: Upon image submission, this segment of the application presents the AI model's predictions. It visualizes the diagnostic results, including the probability of melanoma presence. This section exemplifies the critical role of data visualization in making complex AI analyses accessible and understandable to users, facilitating an informed interpretation of the results.

The emphasis on data visualization and ease of use within the Streamlit application not only democratizes access to advanced melanoma detection tools but also significantly enhances the user experience. By translating sophisticated AI diagnostics into intuitive visual outputs, the application bridges the gap between complex medical data and actionable insights, making it an invaluable tool in the early detection of melanoma.

5. Conclusion

In conclusion, this article has presented a comprehensive exploration of an innovative web application developed for the early detection of melanoma, leveraging the advanced capabilities of Vision Transformers (ViTs) and the intuitive platform of Streamlit. The initial stages of this work involved the meticulous construction of a dataset from the ISIC Challenges of 2019 and 2020, followed by the fine-tuning of a pre-trained ViT Large model from Google, with the dual objectives of distinguishing between "Melanoma" and "No Melanoma" cases and adapting the model to the unique characteristics of dermatological images. This process was underscored by the strategic preparation of the dataset to ensure balanced classes, thereby enhancing the model's learning and predictive accuracy.

Subsequent testing on the distinct MEDNODE dataset confirmed the model's robustness and adaptability, demonstrating significant diagnostic precision across varying conditions. The application of transfer learning techniques further exemplified the utility of leveraging existing AI models for specialized tasks, reducing both the time and resources required for model development from scratch. The deployment environment, characterized by high-performance computing resources, facilitated the model's training and validation phases, while the streamlined requirements for the inference phase underscored the model's practical applicability in diverse clinical settings.

The Streamlit-based web application represents a significant stride towards democratizing access to advanced diagnostic tools. By offering an intuitive interface for uploading and analyzing dermatological images, coupled with real-time presentation of diagnostic results, the application emphasizes the critical role of data visualization in enhancing user engagement and understanding. Each segment of the application, from providing background on melanoma detection to visualizing AI-generated predictions, is designed to make the complex process of AI-driven diagnostics accessible to a broad audience.

In essence, this work illustrates the synergy between cutting-edge AI technology and user-centric application design in addressing the critical healthcare challenge of early melanoma detection. By marrying the technical prowess of Vision Transformers with the accessibility and clarity provided by Streamlit, this initiative paves the way for future advancements in the field of medical imaging and diagnosis. It stands as a testament to the potential of AI to not only revolutionize diagnostic processes but also to empower individuals with the tools and knowledge necessary for early detection and intervention, ultimately contributing to better healthcare outcomes and the broader goal of reducing melanoma-related mortality.

References

[1] H. Tsao, J. M. Olazagasti, K. M. Cordoro, J. D. Brewer, S. C. Taylor, J. S. Bordeaux, M.-M. Chren, A. J. Sober, C. Tegeler, R. Bhushan, et al., Early detection of melanoma: reviewing the abcdes, Journal of the American Academy of Dermatology 72 (2015) 717–723.

[2] L. E. Davis, S. C. Shalin, A. J. Tackett, Current state of melanoma diagnosis and treatment, Cancer biology & therapy 20 (2019) 1366–1379.

[3] M. Rastrelli, S. Tropea, C. R. Rossi, M. Alaibac, Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification, In vivo 28 (2014) 1005–1011.

[4] H. Bhatt, V. Shah, K. Shah, R. Shah, M. Shah, State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review, Intelligent Medicine 3 (2023) 180–190.

[5] D. Pirone, A. Montella, D. Sirico, M. Mugnano, D. Del Giudice, I. Kurelac, M. Tirelli, A. Iolascon, V. Bianco, P. Memmolo, et al., Phenotyping neuroblastoma cells through intelligent scrutiny of stain-free biomarkers in holographic flow cytometry, APL bioengineering 7 (2023).

[6] D. Pirone, D. Sirico, L. Miccio, V. Bianco, M. Mugnano, P. Ferraro, P. Memmolo, Speeding up reconstruction of 3d tomograms in holographic flow cytometry via deep learning, Lab on a Chip 22 (2022) 793–804.

[7] H. Polo Friz, V. Esposito, G. Marano, L. Primitz, A. Bovio, G. Delgrossi, M. Bombelli, G. Grignaffini, G. Monza, P. Boracchi, Machine learning and lace index for predicting 30-day readmissions after heart failure hospitalization in elderly patients, Internal and Emergency Medicine 17 (2022) 1727–1737.

[8] S. Jain, N. Pise, et al., Computer aided melanoma

skin cancer detection using image processing, Procedia Computer Science 48 (2015) 735–740.
[9] A. Bosco, S. Capuozzo, B. Celano, M. Gravina, S. Marrone, M. P. Maurelli, V. Moscato, G. Pontillo, M. Postiglione, A. M. Rinaldi, et al., Ai in healthcare: Activities of the university of naples federico ii node of the cini-aiis lab (2023).
[10] J. Lorca-Cabrera, C. Grau, R. Martí-Arques, L. Raigal-Aran, A. Falcó-Pegueroles, N. Albacar-Riobóo, Effectiveness of health web-based and mobile app-based interventions designed to improve informal caregiver's well-being and quality of life: a systematic review, International Journal of Medical Informatics 134 (2020) 104003.
[11] R. Indraswari, R. Rokhana, W. Herulambang, Melanoma image classification based on mobilenetv2 network, Procedia computer science 197 (2022) 198–207.
[12] G. Cirrincione, S. Cannata, G. Cicceri, F. Prinzi, T. Currieri, M. Lovino, C. Militello, E. Pasero, S. Vitabile, Transformer-based approach to melanoma detection, Sensors 23 (2023) 5677.
[13] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, P. Vajda, Visual transformers: Token-based image representation and processing for computer vision (2020). arXiv:2006.03677.
[14] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, et al., Bcn20000: Dermoscopic lesions in the wild, arXiv preprint arXiv:1908.02288 (2019).
[15] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Scientific data 5 (2018) 1–9.
[16] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al., Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 168–172.
[17] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman, et al., A patient-centric dataset of images and metadata for identifying melanomas using clinical context, Scientific data 8 (2021) 34.
[18] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, N. Petkov, Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images, Expert systems with applications 42 (2015) 6578–6585.