MoleMonitor: Analysis of Skin Lesion Classification Through Deep Learning Models

André Branco[1], Francisco Silva[1], João Silveira[1], Manuel Pinto[1]

[1]Universidade Nova de Lisboa.

Contributing authors: afm.branco@campus.fct.unl.pt; fpg.silva@campus.fct.unl.pt; jpg.silveira@campus.fct.unl.pt; msb.pinto@campus.fct.unl.pt;

Abstract

Skin cancer is among the most common and life-threatening diseases worldwide, where early detection significantly improves patient outcomes. This study explores the application of deep learning for automated skin lesion classification, aiming to enhance diagnostic accuracy and reliability. We investigate the performance of a baseline CNN alongside advanced architectures, including ResNet50, a capsule network, and Inception-ResNet-v2 (IRv2) with and without self-attention mechanisms. Results indicate that while the baseline CNN yields moderate performance, incorporating deeper architectures and attention mechanisms significantly improves classification accuracy. We address open challenges such as class imbalance, intra-class variability and generalization to diverse population.

1 Introduction

Skin cancer is one of the most commonly diagnosed cancers worldwide, with over 1.5 million new cases estimated in 2022, leading to nearly 60,000 deaths globally [1, 2] . Early detection significantly improves survival rates, but distinguishing between benign and malignant lesions remains challenging due to their similar visual characteristics. Traditional diagnostic methods, including visual examination and dermoscopic imaging, rely on clinician expertise and are subject to variability.

Dermatologists follow a structured diagnostic process that typically begins with a visual inspection, followed by dermoscopic examination, and, when necessary, histopathological analysis through biopsy. Despite their expertise, studies have shown that dermatologists achieve an average diagnostic accuracy of approximately 74.1% in distinguishing between benign and malignant lesions [3]. This variability stems from factors such as lesion heterogeneity, clinician experience, and the presence of visual artifacts in dermoscopic images.

In this project, we followed a three-phase approach using a dataset of high-resolution dermoscopic images from publicly available skin cancer databases. First, **Exploratory Data Analysis (EDA)** is performed to assess dataset biases, class distributions, and common image artifacts. Next, **Model Development** involves training and comparing multiple deep learning architectures, including standard convolutional neural networks (CNNs), pre-trained models, hybrid approaches, and attention-enhanced networks [4]. Finally, **Evaluation** is conducted using standard performance metrics such as accuracy, precision, recall, and F1-score, comparing the effectiveness of different model types in improving classification performance and lesion differentiation [5]. The goal of this project is to develop a deep learning model capable of achieving diagnostic performance comparable to, or potentially exceeding, that of trained professionals. However, rather than replacing clinical expertise, the intent is to provide an assistive tool that improves diagnostic consistency and supports dermatologists in making more informed decisions.

2 Literature Review

2.1 Common Datasets for Skin Cancer Classification

Effective skin cancer classification relies on publicly available datasets that serve as benchmarks for evaluating classification methods. Notable datasets are represented in Table 1.

2.2 Common Approaches

Various approaches have been explored to classify skin lesions, ranging from classical machine learning techniques to modern deep learning architectures.

Classic Machine Learning:

Early classification models relied on handcrafted feature extraction methods combined with classifiers such as Support Vector Machines (SVMs), Random Forests (RFs), and k-Nearest Neighbors (k-NNs). While computationally efficient and interpretable, these approaches often struggled with feature generalization and complex lesion variations [4].

Convolutional Neural Networks:

CNNs have transformed skin cancer classification by directly learning hierarchical feature representations from raw images. Architectures such as VGG16, ResNet, and Inception have been widely fine-tuned for lesion classification. However, CNNs often lack interpretability, posing challenges for clinical validation [4].

Transfer Learning:

To address dataset size limitations, transfer learning has been extensively used, leveraging pre-trained models such as ResNet, InceptionV3, and EfficientNet. Fine-tuning these models on skin lesion datasets has significantly improved generalization and reduced the need for extensive labeled data [4].

Attention-Based Models:

Attention mechanisms, including Vision Transformers (ViTs) and attention-augmented CNNs, allow models to emphasize relevant lesion regions while suppressing background noise. Despite their potential, these models require substantial computational resources [4].

Hybrid and Multimodal Approaches:

Combining CNNs with traditional machine learning classifiers or integrating clinical metadata with image-based features has been explored to improve classification robustness [4].

2.3 Open Challenges and Future Directions

Despite significant advancements, skin cancer classification models face several challenges.

Many datasets exhibit class imbalances, leading to biased models that struggle with rare lesion types. Additionally, the lack of annotated data poses a significant challenge, as dermatologist-annotated datasets are costly and time-consuming to generate, limiting the availability of large-scale training data. Another issue is the inter-class similarity and intra-class variability, where different lesion types may share similar visual characteristics, while significant variations can exist within the same class, making classification more difficult.

Furthermore, the black-box nature of deep learning models, such as CNNs and ViTs, presents challenges for clinical validation and adoption, as their lack of interpretability makes

Table 1: Common Datasets for Skin Cancer Classification (* indicates malignant Labels)

Datasets	Nº of Images	Nº Labels	Labels
HAM10000	10,015	7	akiec*, bcc*, blk, df, mel*, nv, vasc
ISIC Challenge Datasets 2016	1,279	2	Benign, Malignant*
ISIC Challenge Datasets 2020	44,000	5	mel*, nv, SK, Lentigo NOS, Unknown
PH2	200	3	CN, AN, mel*
Dermofit	1,300	10	akiec*, bcc*, nv, SK, SCC*, IC*, PG, HMG, df, mel*

akiec = Actinic Keratoses; bcc = Basal Cell Carcinoma; blk = Benign Keratosis-like lesions; df = Dermatofibroma; mel = Malignant Melanoma; nv = Melanocytic Nevus; vasc = Vascular Lesions; SK = Seborrheic Keratosis; SCC = Squamous Cell Carcinoma; IC = Intraepithelial Carcinoma; PG = Pyogenic Granuloma; HMG = Haemangioma; CN = Common Nevus; AN = Atypical Nevus

it difficult for medical professionals to trust and verify their decisions. Lastly, the computational requirements of state-of-the-art deep learning models are substantial, making deployment in resource-constrained environments difficult and limiting accessibility in many real-world applications.

Future research should prioritize dataset diversity, develop explainable AI techniques, and explore efficient architectures to reduce computational overhead.

3 Data Acquisition and Preprocessing

3.1 Data Acquisition

In this study, we utilized the HAM10000 [6] dataset, which comprises 10,015 dermatoscopic images representing seven distinct types of skin lesions. Each image is stored in JPG format with a resolution of 600 $\times$ 450 pixels. Accompanying each image is metadata that includes patient age, gender, lesion location, and unique identifiers for both the image and the lesion. The dataset is divided into two parts for download and also includes segmentation masks for lesion boundaries, facilitating tasks like lesion segmentation.

To assess the generalization capability of our model, we also employed the ISIC 2018 dataset during the testing phase. This dataset is a large-scale collection of dermoscopic images used primarily for challenges in lesion segmentation.

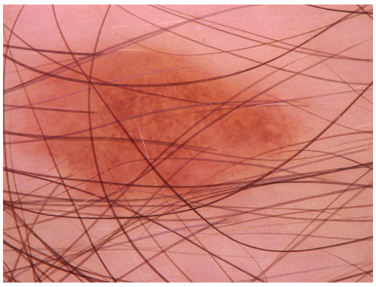
3.2 Preprocessing

To prepare the images for analysis, each was resized to a uniform dimension to maintain consistency across the dataset. Artifacts such as hair and markings, which could interfere with lesion analysis, were removed using the DullRazor algorithm [7] as shown in Figure 1. Following artifact removal, image normalization was performed to standardize color distribution and intensity values, reducing variability due to differing acquisition conditions.

4 Exploratory Data Analysis - EDA

The HAM10000 dataset is composed by seven distinct skin lesion categories. However, the distribution of these categories is notably imbalanced, with melanocytic nevi (NV) constituting approximately 67% of the dataset, while other classes like dermatofibroma (DF) represent less than 1%. This imbalance poses challenges for training machine learning models, as they may become biased toward the more prevalent classes. As detailed in Table 2 with examples in Figure 2.

In terms of demographics, the HAM10000 dataset encompasses a diverse range of patient ages, spanning from infancy to 85 years peaking around 35–50 and 60–75 years, as illustrated in Figure 3a. Gender distribution is relatively balanced, with 54.1% male and 45.5% female participants, as shown in Figure 3b. Notably, the dataset lacks annotations regarding skin



(a) Before Transformation

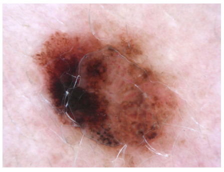
(b) After Transformation

Fig. 1: Comparison of Before and After Transformation

Table 2: Class Distribution in the HAM10000 Dataset

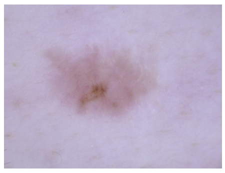
Lesion Type	**nv**	**mel**	**bkl**	**bcc**	**akiec**	**vasc**	**df**
Count	6,705	1,113	1,099	514	327	142	115

(a) Melanocytic Nevus (nv)

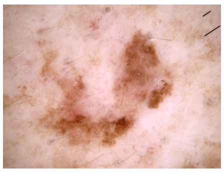


(b) Melanoma (mel)

(c) Seborrheic Keratosis (bkl)

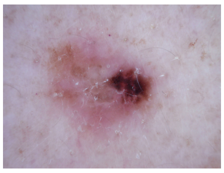


(d) Basal Cell Carcinoma (bcc)



(e) Actinic Keratosis (akiec)

(f) Vascular Malformation (vasc)



(g) Dermatofibroma (df)

Fig. 2: Examples of each dermatological lesion.

color, a recognized limitation that can lead to biases in model performance. Models trained without diverse skin tone data may not generalize well to all populations [8].

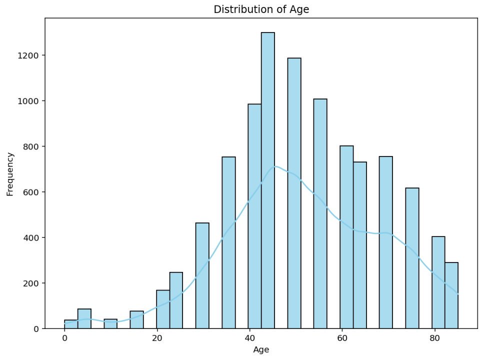
5 Methodology

In this project, we compared advanced machine learning techniques to develop a robust skin lesion classification system. Our approach encompassed several categories:

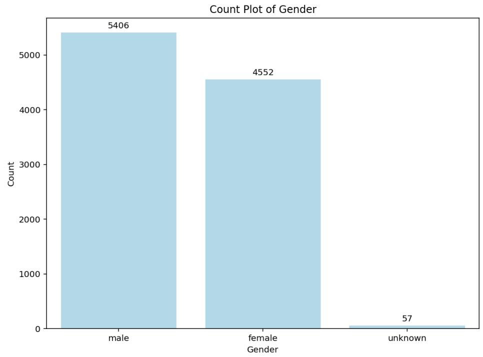
5.1 Approach and Techniques

5.1.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have revolutionized image analysis by their ability to automatically and learn spatial hierarchies of features through backpropagation [4]. In the context of skin lesion classification, CNNs can identify intricate patterns and textures indicative of various dermatological conditions [4]. We designed a baseline CNN model through iterative experimentation, adjusting parameters such as the number of layers, filter sizes, and activation functions to enhance performance metrics.



(a) Age Distribution in the HAM10000 Dataset



(b) Gender Distribution in the HAM10000 Dataset

Fig. 3: Demographic Distributions in the HAM10000 Dataset

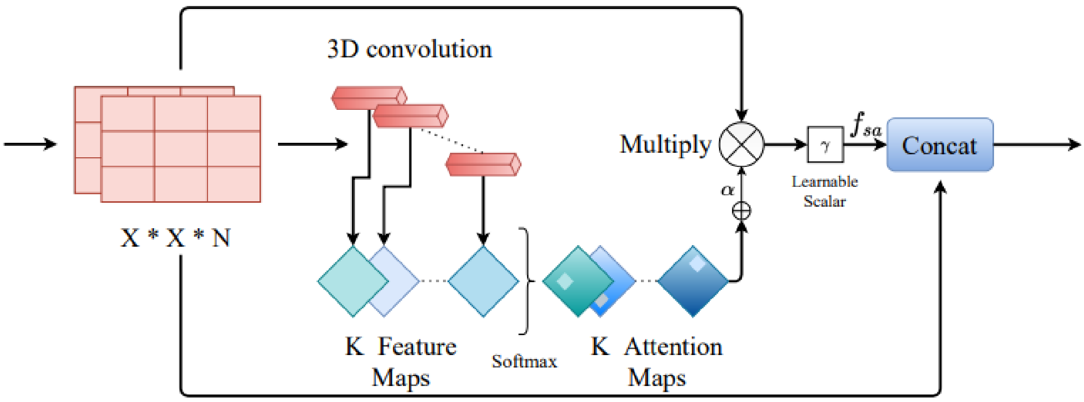


Fig. 4: Soft attention unit proposed by Datta et al [9]

5.1.2 Transfer Learning with Soft Attention

Transfer learning involves leveraging pre-trained models that have been trained on large datasets like ImageNet to address similar tasks. This approach is particularly beneficial when dealing with limited medical image data, as it allows the model to utilize existing knowledge, thereby accelerating the training process and improving performance [9]. In Figure 4 we can see a diagram of a Soft Attention Unit. In our project, we fine-tuned pre-trained architectures, including ResNet50 and Inception-ResNet-V2 (IRv2), on our specific dataset. To further improve the model's focus on critical regions of the images, we integrated a Soft Attention mechanism into the IRv2 model. This mechanism enables the network to highlight important features while suppressing irrelevant information.

5.1.3 Hybrid Models: EfficientNet and Vision Transformers (ViT) with SVM Classifier

Hybrid models combine the strengths of different architectures to capture both local and global features effectively [4, 10]. In our approach, we integrated EfficientNet, known for its efficient scaling and strong performance, with ResNet50 and with Vision Transformers (ViT), which excel at modeling long-range dependencies in images. This combination aimed to leverage EfficientNet's convolutional feature extraction and ViT's attention-based modeling to improve classification accuracy. Additionally, we employed a Support Vector Machine (SVM) classifier at the output layer to enhance decision boundaries.

5.2 Alternative Methods Considered

5.2.1 VGG Networks

VGG networks are renowned for their deep architectures with uniform layer configurations, which have demonstrated strong performance in various image classification tasks. However, their substantial depth and high parameter count can lead to increased computational demands and potential overfitting, especially when applied to limited datasets [4, 11]. While VGG networks offer a straightforward design and have been successful in numerous applications, we considered these factors and opted for architectures that provided a more favorable balance between complexity and performance for our specific dataset.

5.2.2 Capsule Networks

Capsule networks aim to address certain limitations of traditional CNNs by preserving the hierarchical relationships between features, potentially offering improved handling of spatial information and pose variations [12]. Despite their theoretical advantages, capsule networks are relatively novel and can be complex to implement and train. Their efficacy in large-scale medical image classification tasks is still under investigation, and they may require more computational resources compared to traditional CNNs. Given these considerations, we decided to focus on more established architectures that aligned with our project's objectives and constraints.

6 Model Development

6.1 Common approaches between models

We used the HAM10000 dataset for training and evaluation, with additional data from the ISIC2018 dataset for external testing. The HAM10000 data was split into train (80%), validation (10%) and test (10%). To reduce class imbalance, we applied an oversampling strategy in which samples from underrepresented classes were replicated.

Data augmentation was used to artificially increase the diversity of the training set and help the model to generalize better. We applied transformations including geometric manipulations such as horizontal and vertical flips, random rotations (including rotations by 90 degrees), and combined shift-scale-rotate operations to emulate variations in image orientation and scale. Additionally, we introduced color and contrast variations through random brightness, contrast adjustments, and hue-saturation modifications, which helped the model become robust against changes in lighting and color discrepancies. Finally, all images were resized to 224x224 pixels and normalized using ImageNet statistics to ensure consistency with the pre-trained models. This augmentation strategy not only expanded our effective dataset size, but also improved the model's resilience to diverse imaging conditions in real-world applications.

For running the code we mainly used Kaggle Notebooks with GPU P100 accelerator.

6.2 Resnet50

Residual Networks (ResNets) are a class of deep neural networks designed to address challenges in training very deep architectures, particularly the vanishing gradient problem. As networks deepen, training becomes difficult due to diminishing gradient magnitudes during backpropagation. ResNets mitigate this by introducing residual connections which allow the network to bypass one or more layers. In this setup, each layer's input is added directly to its output, enabling the network to learn residual functions modeling the difference between the input and the desired output, rather than the entire transformation. This approach facilitates training of deeper networks by ensuring that gradients flow more effectively during backpropagation.

ResNet50, a specific implementation within the ResNet family, is a deep convolutional neural network comprising 50 layers. Its architecture is organized into five stages, each containing convolutional layers followed by identity and convolutional blocks. These blocks consist of multiple convolutional layers and a shortcut path that bypasses one or more layers, allowing the network to learn residual functions, as represented in Figure 5.

Fig. 5: Block diagram of Resnet50 network by ElGhany et al [13]

To adapt the pretrained ResNet50 model for our skin lesion classification task, we made several modifications to tailor it to our specific dataset and requirements. We began by removing the final fully connected (FC) layer of the original ResNet50, which is designed for 1,000 classes, and replaced it with a custom classifier suited to our number of skin lesion categories. This custom classifier comprises an adaptive average pooling layer to handle variable input sizes, a flattening layer to convert the pooled features into a one-dimensional vector, a dropout layer with a probability of 0.3 to mitigate overfitting, and a fully connected layer that outputs the desired number of classes.

6.3 Inception-ResNet-v2

Inception-ResNet-v2 is a deep convolutional neural network that combines two effective architectures: Inception modules and residual connections.

The term "Inception" refers to the network's ability to capture multi-scale features through its unique module design, which processes input data with multiple convolutional filters of different sizes simultaneously. The "ResNet" part of the name highlights the incorporation of residual connections, which facilitate the training of deeper networks by allowing the network to learn residual functions.

To adapt the Inception-ResNet-v2 for our skin lesion classification task, we made specific architectural modifications. We started with a pretrained Inception-ResNet-v2 model and removed its original classification layer similar to ResNet-50. In its place, we added a custom classifier which includes an adaptive average pooling layer to handle variable input sizes, a flattening layer to convert the pooled features into a one-dimensional vector, a dropout layer with a probability of 0.5 to reduce overfitting, and a fully connected layer that outputs the number of skin lesion categories in our study.

Soft attention is a mechanism in deep learning that helps models focus on the most relevant parts of an input, enhancing their ability to capture important features. Unlike hard attention, which selects specific regions and can be non-differentiable, soft attention assigns continuous weights to all parts of the input, allowing the model to emphasize significant areas while still considering the entire input.

In our project, we integrated a soft attention mechanism into the Inception-ResNet-v2 architecture to improve skin lesion classification. By incorporating soft attention, the model can dynamically highlight critical regions of skin lesion images, such as irregular borders or color variations, which are essential for accurate diagnosis. This enhancement allows the network to focus more on informative features, potentially leading to better performance in distinguishing between different types of skin lesions.

6.4 Fusion Model + SVM Classifier

After exploring individual models, we developed a fusion model leveraging the strengths of multiple architectures. This approach integrates EfficientNet-B3, ResNet50, and Vision Transformer (ViT) models, aiming to capture a comprehensive range of features from dermoscopic images.

EfficientNet-B3 is known for its balanced depth, width, and resolution, achieved through a compound scaling method. This architecture efficiently captures fine-grained details, making it adept at recognizing subtle patterns in skin lesions. However, its performance can be limited when handling complex spatial relationships due to its convolutional nature.

ResNet50 addresses training challenges in deep networks by introducing residual connections, which help mitigate the vanishing gradient problem. Its 50-layer architecture excels at extracting hierarchical features, beneficial for identifying varied lesion characteristics. Despite its depth, ResNet50 may not fully capture global contextual information, as it primarily focuses on localized features.

Vision Transformer (ViT) offers a different perspective by treating images as sequences of patches, employing self-attention mechanisms to model long-range dependencies. This capability allows ViT to understand global context, which is crucial for distinguishing between lesions with similar local features but different overall structures. Nonetheless, ViTs often require large datasets for effective training and can be computationally intensive.

By combining these models, the fusion approach aims to harness their complementary strengths: EfficientNet-B3's efficiency in capturing fine details, ResNet50's hierarchical feature extraction, and ViT's global context understanding. This ensemble is designed to provide a more robust and holistic analysis of skin lesions, potentially leading to improved classification accuracy.

To further enhance the model's performance, we incorporated a Support Vector Machine (SVM) classifier. Deep learning models excel at automatic feature extraction, but integrating an SVM can offer several benefits. SVMs are effective in high-dimensional spaces and can provide a clear margin of separation between classes, which is advantageous when dealing with complex and overlapping data distributions. By training an SVM on the features extracted from the fusion model, we aim to improve the decision boundaries, leading to more precise classifications.

7 Results

In this section, we present the performance outcomes of our skin lesion classification project. We evaluated several models—including ResNet-50, Inception-ResNet-v2, and our proposed fusion model (which integrates EfficientNet-B3, ResNet50, and Vision Transformer [ViT])—on both the HAM10000 and ISIC2018 test sets. Additionally, we assessed the impact of training a Support Vector Machine (SVM) classifier on the features extracted from the fusion model.

Table 3: Summary of Skin Lesion Classification Results

Model	**Dataset**	**Accuracy**	**Macro F1**	**Weighted F1**
Inception-ResNet-v2 + SA	HAM10000	90%	84%	90%
Fusion (pre-SVM)	HAM10000	90%	82%	90%
Fusion (post-SVM)	HAM10000	91%	84%	91%
ResNet-50	ISIC2018	80%	71%	81%
Inception-ResNet-v2 + SA	ISIC2018	82%	71%	81%
Fusion (pre-SVM)	ISIC2018	82%	71%	82%
Fusion (post-SVM)	ISIC2018	83%	73%	83%

The performance of each model in our skin lesion classification task can be attributed to their distinct architectural characteristics:

ResNet-50:

Known for its simplicity and effectiveness, ResNet-50 achieved 80% accuracy on the ISIC2018 dataset. Its residual connections facilitate training deeper networks, but the lower macro F1 score (71%) suggests it may struggle with classifying less common lesions, potentially due to its depth leading to overfitting on dominant classes.

Inception-ResNet-v2 + Soft Attention:

This model combines Inception modules with residual connections, enabling it to capture both deep and wide features. The integration of a soft attention mechanism further enhances performance by focusing on the most relevant regions of the input image, effectively boosting important features and suppressing noise. This architectural design achieved an accuracy of 90% on the HAM10000 dataset, reflecting its strength in handling complex patterns. However, the slightly lower macro F1 score (84%) suggests challenges in classifying less represented classes, possibly due to the dataset's imbalance.

Fusion Model (pre-SVM):

By integrating EfficientNet-B3, ResNet50, and Vision Transformer (ViT), the fusion model leverages diverse feature extraction methods. Achieving 90% accuracy on HAM10000, it benefits from the combined strengths of its components. Yet, the macro F1 score of 82% indicates that without a specialized classifier, the model may not fully capitalize on the ensemble's potential to distinguish between all classes effectively.

Fusion Model (post-SVM):

Applying an SVM classifier to features extracted by the fusion model resulted in improved performance, with accuracy rising to 91% and macro F1 score to 84% on HAM10000. The SVM enhances class separation, particularly benefiting the classification of underrepresented classes by focusing on maximizing margins between them.

7.1 Comparison with State-of-the-Art Benchmark Models

In our study, we evaluated the performance of the Inception-ResNet-v2 model with Soft Attention on the HAM10000 dataset. To contextualize our results, we compared them with the findings from Datta et al. (2021), who also investigated the impact of Soft Attention mechanisms on skin lesion classification using the same dataset. The comparison is presented in Table 4.

Table 4: Comparison of Inception-ResNet-v2 + Soft Attention Performance on HAM10000 Dataset

Study	**Dataset**	**Accuracy**	**Precision**	**Macro F1**
Datta et al. (2021)	HAM10000	93.7%	93.7%	Not Reported
Our Project	HAM10000	90%	90%	84%

Datta et al [9]. presents one of the most well-documented state-of-the-art models for skin lesion classification, utilizing Inception-ResNet-v2 with a Soft Attention mechanism. Their model achieved an accuracy and weighted precision of 93.7% on the HAM10000 dataset. In comparison, our study attained an accuracy of 90%. The difference in reported metrics highlights a key limitation in their evaluation: while weighted precision accounts for class imbalance by favoring dominant classes, it does not provide insight into the model's ability to classify underrepresented lesion types. By contrast, our macro F1 score ensures a more balanced assessment of performance across all classes, making it particularly relevant in highly imbalanced datasets like HAM10000.

Another crucial limitation of Datta et al.'s study is the absence of generalization testing on external datasets. Their results, though strong on HAM10000, lack validation on datasets like ISIC2018, leaving uncertainty regarding their model's robustness beyond the training dataset. In contrast, we evaluated our model on both HAM10000 and ISIC2018, allowing for a deeper analysis of its generalization capability. This broader evaluation is essential for real-world applicability, where models must perform reliably across diverse datasets rather than being optimized solely for a single benchmark.

8 Discussion

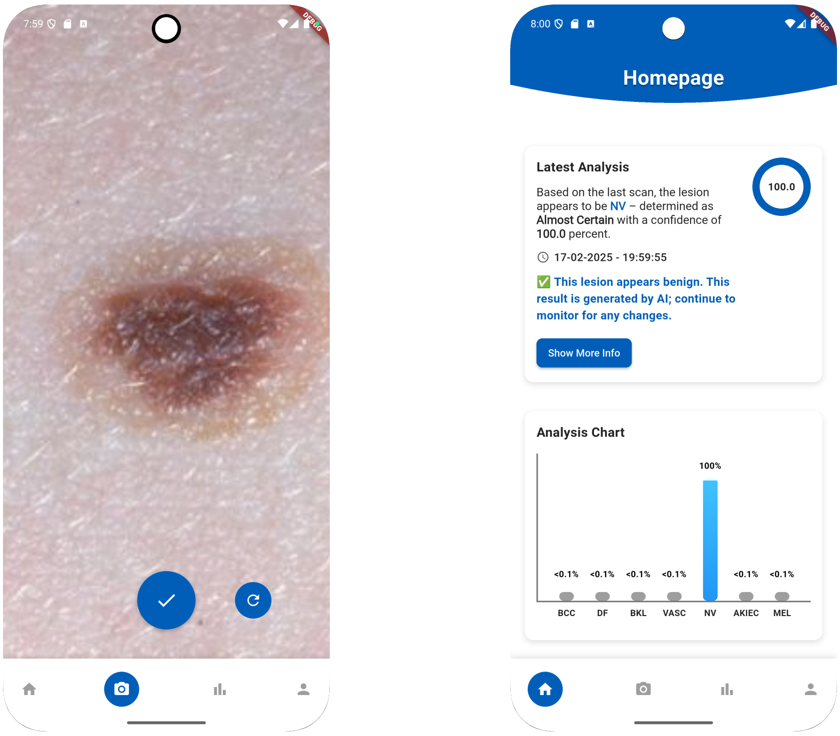
Our results indicate that advanced deep learning architectures have significant potential to improve skin lesion classification, aligning well with the project's objective of developing an assistive diagnostic tool that matches or surpasses clinician performance. Inception-ResNet-v2, achieved accuracies of 82% on the ISIC2018 test set and performed even better on the HAM10000 test set. The fusion model that integrates EfficientNet-B3, ResNet-50, and Vision Transformer (ViT) demonstrated a slight increase in performance. In particular, when the fused features were refined through a Support Vector Machine (SVM) classifier, the model achieved an accuracy of 91% on the HAM10000 test set and 83% on the ISIC2018 test set.

Despite these promising outcomes, several limitations and challenges remain. The class imbalance inherent in the datasets, despite our oversampling strategy, continues to pose a challenge, especially for less represented lesion types. Moreover, while the use of pre-trained models from ImageNet accelerates convergence and improves performance, these models are not originally tailored for medical imaging, which might limit their ability to capture domain-specific features fully. The relatively limited size and diversity of publicly available datasets may also restrict the generalization of the models to broader clinical populations.

8.1 Google Vertex

In our project to deploy a scalable machine learning model for skin lesion classification, we chose Google Cloud's Vertex AI platform.

9 API Documentation



(a) A melanocytic nevus (nv) (b) Analysis Results

Fig. 6: Screenshots of the mobile application

We created a custom Docker image containing our Flask API, which serves our IRv2+SA model. The Docker image, tagged as 'gcr.io/sic molemonitoring/mymodel:latest', was uploaded to Google Container Registry. We then integrated this image into Vertex AI as a custom container, specifying the '/predict' route for predictions and a '/health' route for health checks, and deployed the model to a dedicated endpoint.

We tested the endpoint using cURL and integrated it with our Flutter application by implementing image compression, base64 encoding, and constructing the necessary JSON payload for HTTP POST requests. In production, we plan to manage authentication securely, utilizing a backend service to handle token refreshes.

Access to Vertex AI endpoints requires a valid OAuth 2.0 token in the Authorization header. Initially, tokens were obtained via 'gcloud auth print-access-token', acknowledging their short-lived nature.

9.1 Flutter App & Android Studio

The mobile application was built using Flutter, taking advantage of its flexibility and cross-platform capabilities. Developed in Android Studio, the app serves as the user interface, allowing users to capture or upload images of skin lesions as in Figure 6a. These images are then sent to an endpoint in Vertex AI via an HTTP request, including an access token for authentication. The server processes the image using the trained model and returns the classification results represented in the app as shown in Figure 6b. This architecture ensures a lightweight and responsive application while leveraging the power of a dedicated back-end for analysis.

10 Conclusion

Our most effective model—a fusion of EfficientNet-B3, ResNet-50, and Vision Transformer architectures refined by a Support Vector Machine (SVM) classifier—achieved an accuracy of 91% on the HAM10000 dataset and 83% on the ISIC2018 dataset. These results are comparable to those reported in previous studies. Developing this model required overcoming challenges inherent in skin lesion classification, such as managing class imbalances within datasets and ensuring model generalization across diverse populations. Addressing these issues involved careful data augmentation, preprocessing, and model tuning to enhance performance and applicability. Having met our initial objectives, we now look toward future work to further improve our project.

11 Future Work

For future work, one key direction is to further fine-tune the model and improve the dataset. This involves incorporating additional, diverse data sources to not only to address class imbalance but also to improve the model's generalization. Given that lesions within the same category can exhibit substantial variability in appearance, expanding the dataset to include more varied examples will enable the model to better capture the full spectrum of lesion characteristics, leading to more robust and accurate classifications.

Another important avenue is to establish collaborations with dermatology professionals. Engaging with clinicians will provide critical insights into the practical challenges they face and the specific needs of their diagnostic processes. Such partnerships can help refine the model to deliver clinically relevant outputs and ensure that the tool is both useful and user-friendly in a real-world setting. By integrating expert feedback, the project can be steered toward addressing the nuances of skin lesion diagnosis more effectively, ultimately enhancing its impact as a decision support system in dermatology. [4]

References

[1] International Agency for Research on Cancer: Skin Cancer. Accessed: 2025-02-17 (n.d.). https://www.iarc.who.int/cancer-type/skin-cancer/

[2] World Cancer Research Fund: Skin Cancer Statistics. Accessed: 2025-02-17 (n.d.). https://www.wcrf.org/preventing-cancer/cancer-statistics/skin-cancer-statistics/

[3] Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., Liao, W.: Machine learning in dermatology: current applications, opportunities, and limitations. Dermatology and therapy **10**, 365–386 (2020)

[4] Ray, A., Sarkar, S., Schwenker, F., Sarkar, R.: Decoding skin cancer classification: perspectives, insights, and advances through researchers' lens. Scientific Reports **14**(1), 30542 (2024) https://doi.org/10.1038/s41598-024-81961-3

[5] Hicks, S.A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M.A., Halvorsen, P., Parasa, S.: On evaluation metrics for medical applications of artificial intelligence. Scientific reports **12**(1), 5979 (2022)

[6] Harvard Dataverse: Skin Cancer Classification Dataset. Accessed: 2025-02-17 (2023). https://doi.org/10.7910/DVN/DBW86T . https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

[7] BlueDokk: DullRazor Algorithm Repository. Accessed: 2025-02-17 (n.d.). https://github.com/BlueDokk/Dullrazor-algorithm/tree/main

[8] Kim, Y., Kobic, A., Vidal, N.: Distribution of race and fitzpatrick skin types in data sets for deep learning in dermatology: A systematic review. J Am Acad Dermatol **87**(2), 460–461 (2022) https://doi.org/10.1016/j.jaad.2021.10.010 . Epub 2021 Oct 20

[9] Datta, S.K., Shaikh, M.A., Srihari, S.N., Gao, M.: Soft attention improves skin cancer classification performance. In: Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4, pp. 13–23 (2021). Springer

[10] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., Ellinge, I.: Skin lesion classification using hybrid deep neural networks. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1229–1233 (2019). IEEE

[11] Tabrizchi, H., Parvizpour, S., Razmara, J.: An improved vgg model for skin cancer detection. Neural Processing Letters **55**(4), 3715–3732 (2023)

[12] Goceri, E.: Capsule neural networks in classification of skin lesions. In: International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing, pp. 29–36 (2021)

[13] ElGhany, S.A., Ibraheem, M.R., Alruwaili, M., Elmogy, M.: Diagnosis of various skin cancer lesions based on fine-tuned resnet50 deep network. Computers, Materials & Continua **68**(1) (2021)