narra j

Original Article

Enhancing dermoscopic pigmented skin lesion classification: A refined approach using the pre-trained Inception-V3 architecture

Erwin S. Nugroho[1,2*], Igi Ardiyanto[1] and Hanung A. Nugroho[1*]

[1]Departement of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia; [2]Department of Informatics, Politeknik Caltex Riau, Pekanbaru, Indonesia

*Corresponding authors: erwinsn@pcr.ac.id (ESN) and adinugroho@ugm.ac.id (HAN)

Abstract

Skin cancer is one of the most prevalent cancers worldwide, with early diagnosis being critical for improving survival rates. Dermoscopy, a non-invasive imaging tool, is widely used for identifying pigmented skin lesions. However, its accuracy is heavily dependent on expert interpretation, which introduces variability and limits accessibility in resource-constrained settings. This highlighted the need for automated solutions to enhance diagnostic consistency and aid in early detection. The aim of this study was to develop a refined machine-learning framework for classifying pigmented skin lesions using dermoscopy images. We employed an enhanced Inception-V3 model, a state-of-the-art convolutional neural network, integrated with a simplified soft-attention mechanism, advanced data augmentation techniques, and Bayesian hyperparameter tuning. These innovations improved the model's ability to accurately focus on and identify relevant lesion features, marking a significant advancement in the field. Using the ISIC-2019 dataset, a publicly available resource containing dermoscopy images classified into eight diagnostic categories, we implemented preprocessing steps such as resizing, cleaning, and data balancing. Additionally, ImageNet transfer learning and Bayesian optimization were applied to refine the model. The inclusion of a soft-attention mechanism further enhanced the model's capacity to identify patterns within lesion images. Our model exhibited outstanding performance on the ISIC-2019 dataset, achieving a sensitivity of 98.5%, specificity of 99.62%, precision of 97.42%, accuracy of 97.38%, an F1 score of 97.34%, and an area under the curve (AUC) of 0.99. These metrics underscored the model's superior capability in accurate and reliable classification of pigmented skin lesions, surpassing current benchmarks and demonstrating significant advancements over existing methodologies.

Keywords: Medical image processing, dermoscopy, pigmented skin lesion, convolutional neural network, Inception-V3

Introduction

The classification of pigmented skin lesions (PSLs) over dermoscopy techniques has become a significant subject in dermatology and skin oncology in recent years [1,2]. Early and accurate identification of skin lesions, such as melanoma, can significantly reduce mortality rates, given the high global prevalence of skin cancer [3]. A report published by the American Cancer Society stated an increased trend in skin cancer prevalence over the past 30 years, with an average annual percent change (AAPC) of 1.8% [4]. Meanwhile, according to the Australian Institute of Health

https://narraj.org/

Received: December 2, 2024 - Accepted: March 19, 2025 - Published online: April 21, 2025

and Welfare, the AAPC rate is 1.7% in the last 25 years [5]. However, challenges remain in the consistent and objective interpretation of dermoscopic images, as it often requires specialized expertise and can be subjective.

Artificial intelligence, particularly deep learning, is pivotal in advancing skin lesion research, offering exceptional diagnostic performance, especially with dermoscopy images. Key architectures and AI methods are driving improvements in early detection and reliable diagnosis, shaping the future of this field [6,7]. Its main advantage is the ability to process raw data directly, reducing the need for complex pre-processing. This allows computer-aided diagnostic (CAD) systems to detect abnormalities and disease markers more accurately, which sometimes can surpass medical experts. Research in the field of PSLs primarily concentrates on three key areas: segmentation [8-12], feature extraction [13-17], and classification [18-23]. Furthermore, the field of interpretable machine learning (IML), also known as explainable artificial intelligence (XAI), is rapidly growing, intending to tackle ethical challenges in the healthcare industry [24].

Recent studies [18,19,25-27] on dermoscopic image classification have increasingly focused on using deep learning models, particularly with the ISIC-2019 dataset for multiclass classification tasks. The ISIC-2019 dataset is a public dataset that was curated by the International Skin Imaging Collaboration (ISIC) and released in 2019 for the purpose of facilitating research in the field. This dataset is extensively utilized in the domains of dermatology and artificial intelligence research, particularly for developing and evaluating algorithms related to the analysis of skin lesions, with a particular focus on pigmented skin lesions. These approaches aim to address challenges such as class imbalance, low contrast between skin lesions and surrounding areas, and the presence of artifacts. A study [28] proposed a deep convolutional neural network (CNN) model specifically designed for multiclass skin cancer classification using the ISIC-2019 dataset. Their model achieved significant performance, with an accuracy of 97.1%, and outperformed other transfer learning-based models like VGG16 and DenseNet. Alsahafi *et al.* [19] introduced Skin-Net, a deep residual network leveraging multilevel feature extraction and cross-channel correlation. Tested on both ISIC-2019 and ISIC-2020 datasets, their model effectively handled the challenge of dataset imbalance and achieved improved accuracy in multiclass classification. Another study proposed a Swin Transformer model for skin lesion classification, leveraging the strengths of both CNNs and transformers [18]. Using the ISIC-2019 dataset, their method outperformed many traditional CNN models, achieving high sensitivity and specificity [18]. Cauvery *et al.* [26] implemented a transfer learning-based convolutional neural network model that classified dermoscopic images into multiple categories using the ISIC-2019 dataset and their ensemble approach achieved a balanced accuracy of 81.2%. Another study also contributed to this area by using transfer learning with the GoogleNet architecture to classify eight skin lesion classes from the ISIC-2019 dataset [27]. The model achieved a classification accuracy of 94.92%, further demonstrating the efficiency of CNNs in multiclass skin lesion classification [27].

The Inception-V3 architecture is a highly advanced deep learning model that has demonstrated its ability to identify critical features of dermoscopic images [29,30]. However, there is still a room for improvement, particularly in adding specific model methods to specialized datasets. The aim of this study was to explore and enhance the efficacy of the Inception-V3 architecture for classifying PSLs through dermoscopy. The research aimed to adapt the Inception-V3 model using techniques such as augmentation, transfer learning, fine-tuning, and specific methods to align it with the unique nuances of dermoscopic images of skin lesions.

This study utilized the Inception-V3 model with the ISIC-2019 dataset to classify PSLs. The key achievement of this study was the development of an improved model that enhanced the accuracy of classification tasks. This study makes several key contributions: (a) introduced a simplified soft-attention mechanism that significantly enhances the model's focus on relevant lesion features, improving both interpretability and classification accuracy; and (b) integrated Bayesian hyperparameter tuning optimizes the model performance with reduced computational complexity, leading to superior results compared to existing methods.

While data augmentation, Bayesian tuning, and attention mechanisms have been explored in other machine-learning fields, the innovation of this study lies in how these techniques were uniquely applied and optimized for dermoscopic image analysis in the context of skin lesion

classification. Specifically, our approach integrated a simplified soft-attention mechanism that enhanced the model's ability to focus on lesion-relevant features without significantly increasing computational complexity. This simplification improved both classification accuracy and interpretability, which is crucial in a clinical setting where transparent decision-making is important. Additionally, our application of Bayesian hyperparameter tuning was tailored to balance the specific challenges of dermoscopic image datasets, such as class imbalance and feature variability. This combination of techniques, though not entirely new in isolation, represents a novel and effective solution when applied in unison to the specific problem of skin lesion classification, providing superior results over existing methodologies.

Methods

Dataset and data preprocessing

The ISIC-2019 dataset [31-33] was chosen for this study due to its large number of lesion images and classes. Dermoscopy images in eight disease classes, along with metadata, were included (**Table 1**). However, it should be noted that the dataset is imbalanced, with almost half of the images being melanocytic nevus.

The data preprocessing stage of the study was comprised of three primary steps. First, duplicate images were detected based on the metadata of each image in the dataset. Second, the data was cleaned to ensure that each image had a lesion identity document (ID) in its metadata. Third, the images were resized to meet the input requirements of the model. Detection of duplicates is important for maintaining dataset diversity and preventing data leakage that could artificially inflate model performance. Data cleaning ensures that the dataset is properly labeled and structured for training while resizing standardized all images to a uniform size, aligning with the model's input dimensions and improving computational efficiency during training and inference. We also created a testing data subset by selecting 100 images from each class, resulting in a total of 800 images for testing. The testing data subset did not overlap with the training or validation data subsets. The training set used the remaining data and was augmented to increase the data volume and achieve balance among different disease classes.

During the augmentation phase, we used the Image Data Generator, a utility provided by deep learning libraries, to perform real-time data augmentation during training. This involved applying specific parameters to enhance the diversity and variability of the training dataset. These parameters included rotations up to 180 degrees, shifts in height and width by 10% (0.1), zoom modifications of 10%, and flips both horizontally and vertically. Brightness levels were also modified to range from 90% to 110% of the original image brightness. For images necessitating fill due to the applied transformations, we opted for the 'nearest' fill mode. The augmentation process utilized a batch size of 20 to generate a total of 9200 images to align with the largest image count observed in the melanocytic nevus category. Details of this preprocessing and augmentation method are presented in **Table 1**.

Table 1. Processing, cleaning, division, and enhancement of the International Skin Imaging Collaboration (ISIC) 2019 dataset

Disease classes	Baseline data	No having an ID	Testing subset	Learning subset	Augmented training subset
(1) Melanocytic nevus (NV)	12,875	3,647	100	9,128	9,128
(2) Melanoma (MEL)	4,522	495	100	3,927	9,204
(3) Basal cell carcinoma (BCC)	3,323	138	100	3,085	9,220
(4) Benign keratosis (BKL)	2,624	436	100	2,088	9,202
(5) Actinic keratosis (AK)	867	36	100	731	9,060
(6) Squamous cell carcinoma (SCC)	628	24	100	504	8,606
(7) Dermatofibroma (DF)	239	11	100	128	8,436
(8) Vascular lesion (VASC)	253	39	100	114	7,918
Total	25,331	4,826	800	19,705	70,774

We could express the process mathematically to better understand how image augmentation impacted the dataset. Let Ni represent the number of original images in class i, and Mi the total

number of augmented images after the augmentation process. The goal of augmentation was to synthetically expand the dataset by employing random transformations such as rotation, shifting, zooming, and flipping, which we denoted as transformation operators Tk, where k=1, 2, ..., K. Each original image xi undergoes these transformations, creating new augmented images $xi(aug)$, which can be modeled as: $xi(aug)=Tk(Tj(...T1(xi)...))$ (1) where the transformation parameters (e.g., angles for rotation or scaling factors for zooming) were randomly selected from predefined ranges.

The final count of augmented images for a given class is determined by the target value, M. If a class contains N original images and the augmentation generator produced images in batches of size B, the total number of augmented images, M, is obtained by repeatedly generating batches until the target M is met: $M = N + \sum_{j=1}^{I} B_j$ (2) where Bj was the batch size at the j-th iteration, and I was the total number of iterations.

The total number of augmented images Mi for each class was determined by a target value, aiming to achieve class balance in the dataset. For example, in the present study, the target number of images per class was set to approximately 9200, except for the class melanocytic nevus where no augmentation was performed. The final number of augmented images did not always represent an integer multiple of the original dataset size due to the stochastic (random) nature of the augmentation process and the batch-wise way augmentation was applied.

This discrepancy arose due to the augmentation process that did not strictly adhere to integer multiples but rather aimed to meet the desired target number through iterative batch processing. Each original image underwent transformations based on random parameters, resulting in variability in the number of images generated. By modeling the augmentation process in this way, we provided a structured understanding of how the variability in image count arose from the augmentations, helping future researchers appreciate the underlying mechanics in medical imaging datasets.

Model architecture

The Inception-V3 [34] architecture was a significant advancement in the field of deep learning, particularly in the area of image recognition and classification. It is the third iteration of the Inception architecture, first introduced by Google. Key aspects of Inception-V3 include complex architecture: Inception-V3 is known for its complex architecture, which is a deeper and wider version of the original Inception model. It contains numerous 'Inception modules' that allow it to efficiently process information at various scales and complexities. Improved performance: this model improved upon its predecessors in terms of accuracy and efficiency, particularly in large-scale image recognition tasks.

It achieved higher accuracy with lower computational cost compared to earlier versions. Factorization into smaller convolutions: Inception-V3 breaks down larger convolutions into smaller, more manageable operations. For example, a 5×5 convolution can be broken down into two 3×3 convolutions. This reduces the number of parameters, helping in controlling overfitting and reducing computational requirements. Grid size reduction: Inception-V3 introduces an efficient grid size reduction technique that avoids a representational bottleneck. This capacity was reached by not using pooling to reduce grid size but instead using convolution with a stride. Batch normalization: the design heavily incorporates batch normalization, a procedure that normalizes the input to a layer by adapting and scaling its activations. This approach accelerated the training process and offers a degree of regularization, diminishing the necessity for dropout. Applications: Inception-V3 has been widely used for image classification, object detection, and facial recognition tasks. Its efficiency made it suitable for both academic research and real-world applications, including those running on devices with limited computational resources.

Inception-V3 represented a significant step in the evolution of CNNs, offering a balance between computational effectiveness and model performance, and has been a base for further innovations in the field. The primary consideration for utilizing the Inception-V3 architecture in this research stems from its superior performance in PSL classification [29,30]. This model outperforms other architectures in accurately classifying PSLs, making it the optimal choice for this study.

Model optimization

The optimization process combined Bayesian tuning for hyperparameter optimization with the integration of an attention module. This combination enhanced the model's ability to focus on critical features while systematically identifying the best parameter settings to maximize performance. This dual approach ensures both efficiency and accuracy in model training.

Bayesian tuning

Bayesian optimization was a probabilistic model-based optimization method aimed at finding the optimal hyperparameter set $\theta*$ that maximizes or minimizes an objective function $f(\theta)$, in this case, the validation accuracy. Given a dataset D, the model learns the function f(θ), where θ represents the hyperparameters (e.g., learning rate, dropout rate).

Bayesian optimization constructed a surrogate model $M(\theta)$ over the objective function and uses an acquisition function $a(\theta|M)$ to decide the next hyperparameter point to evaluate. The surrogate model $M(\theta)$ is modeled as a Gaussian process (GP): $M(\theta) \sim \text{GP}(\mu(\theta), k(\theta, \theta'))$ (3) where $\mu(\theta)$ is the mean function, and $k(\theta, \theta')$ is the covariance (kernel) function. The acquisition function $a(\theta|M)$ balances exploration (probing new regions of the hyperparameter space) and exploitation (focusing on areas with known good results). A common acquisition function is the Expected Improvement (EI): $EI(\theta) = \mathbb{E}(\max(0, f(\theta) - f(\theta+)))$ (4) where $f(\theta+)$ is the best-observed value of the objective function. The next hyperparameter set $\theta n+1$ is chosen by maximizing the Expected Improvement.

Algorithm 1 Bayesian tuning (D, p, E_{init}, n, E_{stop}, T): (a) Select a subset of the dataset, D_{subset}, by sampling from D with a proportion p. (b) Split D_{subset} into training data (D_{train}) and validation data (D_{val}). (c) Define the search space for the hyperparameters θ. (d) Set up the objective function $f(\theta)$ based on validation accuracy. (e) Initialize the number of epochs E to E_{init} and the iteration counter T to 0. (f) While T is less than n, perform the following steps: Train the model using the hyperparameters θ for E epochs. Apply early stopping if required. Optimize the hyperparameters θ through Bayesian optimization. Update E_{opt} to reflect the highest validation accuracy obtained so far. Increase the iteration counter T by 1. (g) Upon completing all iterations, train the model using the best hyperparameters $\theta*$ for E_{opt} epochs. (h) Keep the trained model with the optimal hyperparameters $\theta*$.

The present study carried out hyper-parameter fine-tuning through Bayesian optimization to identify the best parameter settings. The process explored hyperparameters such as the learning rate (ranging between 1×10^{-6} and 1×10^{-2}), dropout rate (ranging from 0.1 to 0.5), and optimizer (fixed to adaptive moment estimation (ADAM)). The document described this optimization process in Algorithm 1, requiring inputs like the dataset D, a subset percentage p, the initial epoch count Einit, trial number n, and the epoch at which to stop Estop. For Bayesian optimization, the research utilized 25% of the expanded dataset for training, sets the amount of testing to 5, and targets validation accuracy as the performance metric. The outcome of this optimization yielded optimal settings, including a dropout rate of 0.3, a learning rate of 0.0001, and the selection of ADAM for optimization.

The purpose of Bayesian tuning in modeling was to optimize model parameters using a Bayesian probabilistic approach. Its main objectives included enhancing model performance, mitigating overfitting or underfitting, optimizing parameters efficiently, improving computational efficiency, and adapting to specific data characteristics. By systematically exploring parameter space, Bayesian tuning aimed to find the best-performing parameter combinations while considering trade-offs and adapting to data intricacies.

Attention module

The attention module, also known as the attention layer, consisted of a series of layers that implemented the attention mechanism to focus on influential features within a feature map. Attention mechanisms allowed the model to appoint varying degrees of importance to different parts of the input, enabling it to selectively attend to the most relevant features or elements. By incorporating attention layers, a model can better capture long-range dependencies, improve its understanding of context, and enhance performance in tasks such as machine translation, text summarization, image captioning, and sequence prediction. Ultimately, attention layers

contributed to the model's interpretability, robustness, and overall effectiveness in processing sequential or set-based data.

The attention mechanism used in this study is a simplified soft-attention module designed to focus on the most relevant features in dermoscopic images. The attention mechanism assigns a weight αi to each feature i of the input, computed as follows: $\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)}$ (5) where ei is the importance score for feature i, and αi is the attention weight. The importance score ei is typically computed through learned compatibility functions.

To maintain simplicity in the model, we have introduced a novel and simplified spatial attention mechanism consisting of a Conv2D layer, Global Average Pooling, and a Density layer, as shown in Algorithm 2 and **Table 2**. This attention module is a simplification of spatial attention introduced by Woo *et al.* [35]. We modulated the feature map output from the base model using a convolutional and dense layer-based approach. Let $F \in \mathbb{R}H \times W \times C$ represent the feature map from the base model, where H and W are the spatial dimensions, and C is the number of channels. The attention mechanism is defined as follows: Apply a 1×1 convolution to the feature map: $F' = \text{Conv1×1}(F)$ (6). Perform global average pooling on F' to obtain a vector $v \in \mathbb{R}C$: $v = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F'_{ij}$ (7). Pass v through two fully connected (dense) layers: $d1 = \text{ReLU}\ (W1v + b1)$ (8) $d2 = \text{Sigmoid}\ (W2d1 + b2)$ (9) where $W1 \in \mathbb{R}C \times 256$, $W2 \in \mathbb{R}256 \times C$, and $b1$, $b2$ are bias terms. The output from the second dense layer $d2$ produces the attention weights $\alpha \in \mathbb{R}C$, which are multiplied elementwise with the original feature map F: $F_{att} = F \odot \alpha$ (11) where $\odot$ denotes element-wise multiplication.

Table 2. Simplified spatial attention layer structure

Step	Layer	Number of units/ filters	Activation	Output dimensions
1	1×1 convolution	2048	ReLU	H×W×2048
2	Global average pooling	-	-	1×1×2048
3	Reshape	-	-	1×1×2048
4	Dense 1	256	ReLU	1×1×256
5	Dense 2	2048	Sigmoid	1×1×2048

Algorithm 2 simplified spatial attention with some steps. The simplified spatial attention algorithm started by adding an attention mechanism to the base model. This step involved extracting the base model's output, performing a 1×1 convolution operation with 2048 filters and ReLU activation, applying global average pooling to the convolution result, reshaping the pooled output into a 1×1×2048 format, passing it through a dense layer with 256 units and ReLU activation, and finally adding another dense layer with 2048 units and sigmoid activation. The attention mechanism was then incorporated into the base model by applying it to the base model's output, reducing its dimension using global pooling, passing it through a dense layer with 1024 units and ReLU activation, and adding a SoftMax layer for predictions. The complete model was constructed by using the base model's input as the starting point and combining it with the output of the prediction layer to form the final model.

The simplified soft-attention mechanism used in the present study was designed to address the limitations of traditional attention mechanisms by focusing on computational efficiency and interpretability. Unlike traditional models that compute attention weights across all input regions, our approach emphasized channel-wise importance using a streamlined structure. This involved a 1×1 convolutional layer, global average pooling, and two dense layers to generate attention weights. By avoiding the complexity of multi-head or spatial attention, this mechanism reduced computational overhead while maintaining the ability to focus on key features relevant to dermoscopic image analysis.

This simplified design was particularly suited for dermoscopic images, which required highlighting critical patterns like pigmentation, texture, and lesion borders. The mechanism enhanced interpretability by clearly identifying focused regions, a critical feature in clinical applications. Additionally, its efficiency and simplicity mitigate overfitting risks, especially in imbalanced datasets like ISIC-2019, ensuring robust performance without excessive computational demands. This balance of precision, efficiency, and interpretability made the mechanism highly effective for medical image classification tasks.

In the present study, the primary innovation lied in the integration of a simplified soft-attention mechanism with the Inception-V3 architecture, designed specifically for enhancing the classification of pigmented skin lesions. Unlike traditional attention mechanisms, which tend to increase model complexity, our approach introduced a novel, streamlined mechanism that focuses the model's attention on the most relevant lesion features without adding significant computational overhead. This simplicity not only improved the model's interpretability but also reduced the risk of overfitting, particularly on smaller or imbalanced datasets like ISIC-2019.

Moreover, by refining the attention mechanism to focus on spatially significant areas of dermoscopic images, the model can better isolate critical features for lesion classification. This enhancement contributes to improved diagnostic performance, as reflected in the high sensitivity, specificity, and accuracy metrics achieved. Our method not only outperforms traditional attention-based models but also demonstrates that such improvements can be made without extensive computational costs, making this approach accessible for real-time medical diagnostics.

In addition to the attention mechanism, our use of Bayesian optimization to fine-tune hyperparameters and advanced data augmentation techniques further enhances the model's performance. These innovations collectively provide a substantial advancement in medical image analysis, offering a more interpretable and efficient solution for pigmented skin lesion classification compared to existing methodologies.

Experiment scenarios

The comprehensive workflow employed in the experiment for classifying pigmented skin lesions from the ISIC-2019 dataset is depicted in **Figure 1**. The process began with the data pre-processing stage, which involved steps such as identical detection, data cleaning, and image resizing. After pre-processing, the data is augmented, balanced, and split using 5-fold cross-validation to ensure consistent model performance during training.

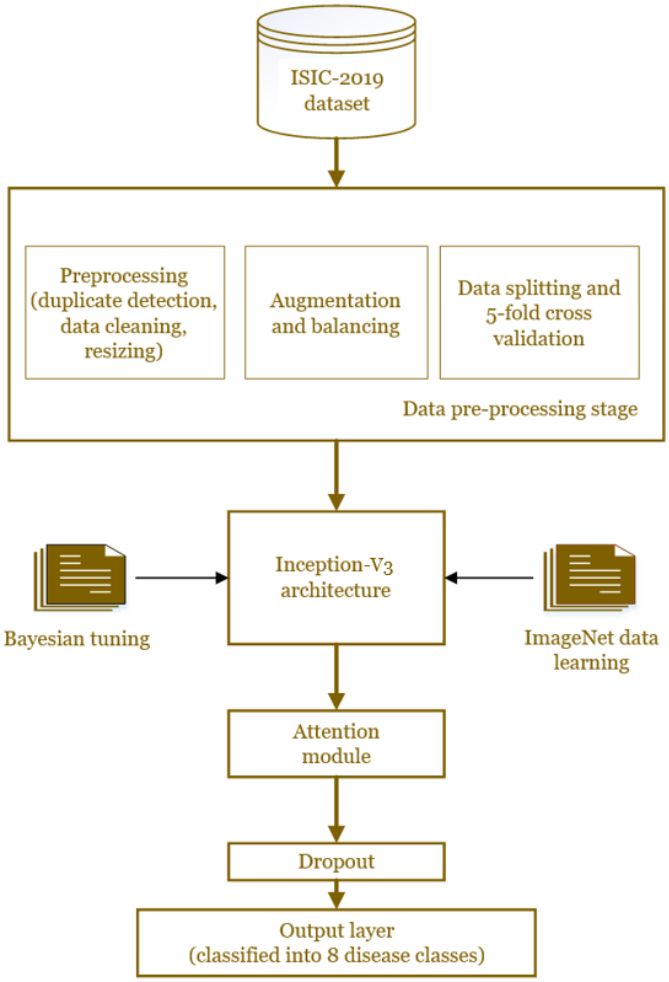

Figure 1. Experimental workflow for pigmented skin lesion classification.

The prepared data was then input into the fine-tuned Inception-V3 architecture, pre-trained on ImageNet and optimized using Bayesian tuning. To enhance feature learning, an attention module is integrated, followed by a dropout layer to mitigate overfitting. Ultimately, the model outputs classifications into one of eight disease categories. The initial model parameters included the ADAM optimization algorithm, a learning rate of 0.1, a dropout rate of 0.1, a batch size of 20, and 50 epochs. This model's performance serves as the baseline for subsequent evaluations.

The experiments were conducted on a high-performance workstation equipped with an Intel Core i9-10900K processor (10 cores, 20 threads, base clock speed of 3.7 GHz, and boost clock up to 5.3 GHz), 128 GB of DDR4 RAM (3200 MHz), and an NVIDIA GeForce RTX 3080 GPU with 11 GB of GDDR6X memory. This configuration provided robust computational power for the training and evaluation of deep learning models. The workstation operated on Ubuntu 18.04 LTS, a Linux distribution known for its stability in machine learning and scientific research tasks.

The software environment included JupyterHub for experiment management and interactive development, Python 3.8.10 as the primary programming language, and essential deep learning libraries such as TensorFlow 2.6 and PyTorch 1.10. CUDA Toolkit 11.3 and cuDNN 8.2 were installed to optimize GPU computations. Additionally, package management was handled through Conda 4.10.3 to ensure a reproducible environment. For data preprocessing and analysis, libraries like NumPy 1.21.0, Pandas 1.3.0, and Matplotlib 3.4.3 were utilized.

Four testing scenarios were implemented: (1) Using the model without incorporating augmented data. (2) Implementing the model with data augmentation applied. (3) Deploying the model with both data augmentation and Bayesian hyperparameter optimization. (4) Running the model with augmented data, Bayesian hyperparameter optimization, and the addition of an attention mechanism.

The Inception-V3 architecture contains approximately 23 million parameters, and each forward pass requires 5 billion FLOPs, which can increase inference time. However, the inclusion of Bayesian optimization and attention mechanisms reduced overfitting and improved the model's performance, making the slight increase in complexity justifiable.

Evaluation

Evaluating a learning algorithm with test data is crucial to assess its effectiveness. The assessment starts with the confusion matrix, a key tool for analyzing performance. Essential metrics for classification tasks, including sensitivity (SEN), specificity (SPE), accuracy (ACC), precision (PREC), and the area under the curve (AUC), play a vital role in measuring the algorithm's ability to differentiate between classes. These metrics deliver valuable insights into the precision and reliability of the algorithm.

Results

Deep learning, specifically CNNs, has demonstrated impressive capabilities in image classification. In this study, we utilized the Inception-V3 pre-trained model to classify images into eight distinct categories using the ISIC-2019 dataset. We implemented several stages of treatment to enhance the performance of our model.

Training models

Model training was a crucial stage in which the model was introduced to feature variations in each class, enabling it to generalize to new images. The success of the model was highly dependent on this stage. Four-line graphs representing the training and validation accuracy of an Inception-V3 model under four different conditions (**Figure 2**): without augmentation, with augmentation, with augmentation and Bayesian tuning and with augmentation, Bayesian tuning and attention module. With augmentation, Bayesian tuning and attention module had the best performance among the four scenarios (**Figure 2**). The validation accuracy was not only stable but also closer to the training accuracy, suggesting a well-generalizing model.

Each step taken to enhance the model's training process contributed to a more robust and generalizable model, as indicated in **Figure 2**. The augmentation added variability to the training data, the Bayesian tuning optimized hyperparameters, and the attention mechanism allowed the model to focus on the most informative parts of the input data, which together achieved better

performance. In the first scenario, where no data augmentation was used, the model's performance was inconsistent. It excelled at recognizing some classes (like melanocytic nevus, which stood for a particular data type, identified correctly 100% of the time), but it struggled with others, often confusing one class for another (like dermatofibroma and vascular lesion, which were other data types). When data augmentation was introduced in the second scenario, which meant the model was trained with a more diverse set of data, there was a noticeable improvement. This technique helped the model not to be fooled by minor variations in the data, leading to better recognition across most classes. It wasn't perfect, but it was a step up from the first scenario.

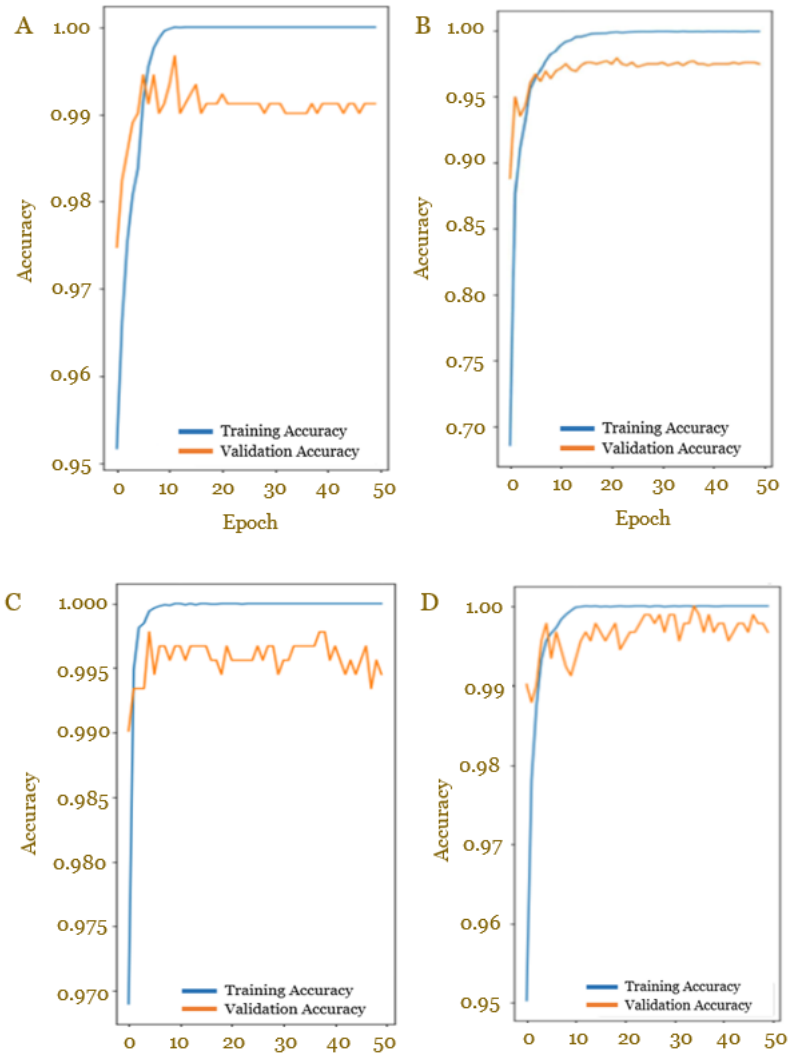

Figure 2. Performance training and validation accuracy of Inception-V3 with four different conditions. (A) Without augmentation, the graph showed very high training accuracy that quickly reaches nearly 100% and remains stable. However, the validation accuracy was volatile, suggesting the model may be overfitting due to the lack of input data variability. (B) With augmentation, the introduction of data augmentation improved the stability of the validation accuracy, reducing overfitting. There's still a gap between training and validation accuracy, but it's less pronounced. (C) With augmentation and Bayesian tuning, the addition of Bayesian hyperparameter tuning along with data augmentation showed an even more stable validation accuracy with less fluctuation, and the gap between training and validation accuracy is further narrowed. (D) With augmentation, Bayesian tuning and attention module indicates the best performance among the four scenarios. The validation accuracy was not only stable but also closer to the training accuracy, suggesting a well-generalizing model.

The third scenario added Bayesian tuning on the top of data augmentation (**Figure 2**). Bayesian tuning was a statistical method that helped the model make better decisions by considering the certainty of its predictions. This led to a significant jump in accuracy, with the model now perfectly identifying several classes (like actinic keratosis, dermatofibroma,

melanocytic nevus, and vascular lesion) that it previously had trouble with. Finally, the fourth scenario built on all the previous improvements and added an attention mechanism (**Figure 2**). This mechanism allowed the model to 'focus' on the most important parts of the data when deciding. It was like giving the model a way to zoom in on what mattered most, which resulted in the highest accuracy levels across all classes. In summary, each step of enhancement—data augmentation, Bayesian tuning, and the attention mechanism—brought the model closer to perfect performance. By the end of this progression, the model made very few mistakes, demonstrating that these techniques were quite powerful in teaching AI to correctly classify complex data.

Testing models

Upon completion of the model training phase, subsequent validation was performed utilizing a designated dataset for testing purposes. This dataset was curated by selecting images from the onset of each category at intervals of 100 images. Evaluation metrics, containing sensitivity, specificity, precision, accuracy, F1 Score, and AUC, are computed for each model based on the information extracted from the confusion matrix, facilitating a complete assessment of model performance. The confusion matrices of pre-trained Inception-V3 for PSL classification with four scenarios are presented in **Figure 3**.

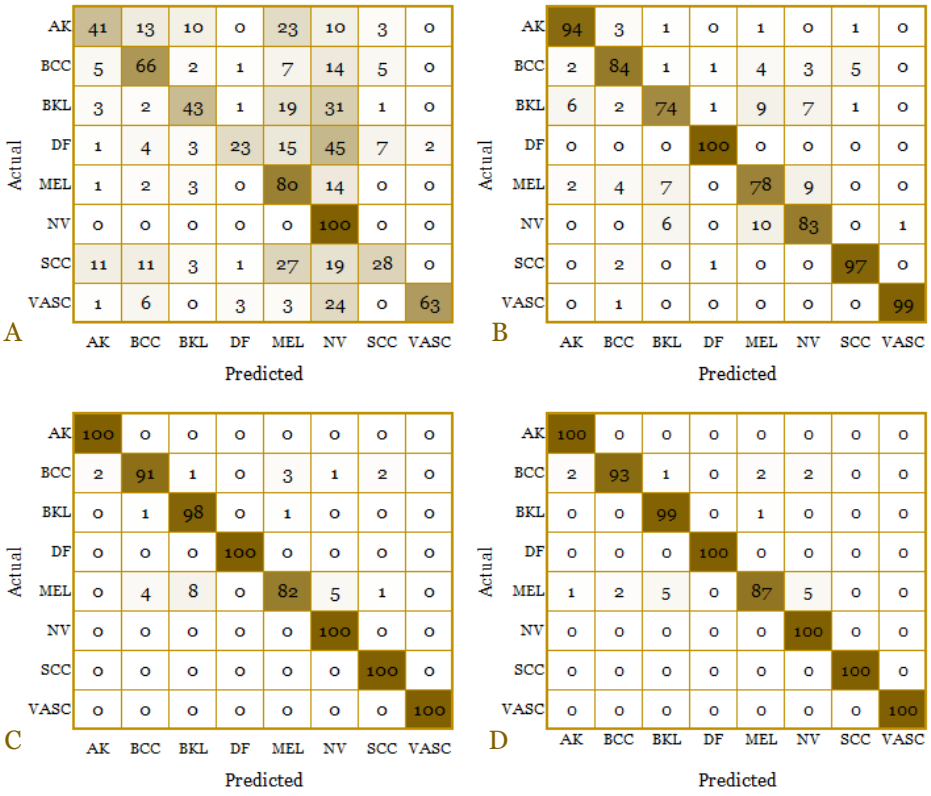

Figure 3. Confusion matrix of Inception-V3: (A) without data augmentation, (B) with data augmentation, (C) with data augmentation and Bayesian tuning, and (D) with data augmentation, Bayesian tuning and attention module.

Without data augmentation, the model exhibited significant misclassification across several classes, reflecting suboptimal precision and recall (**Figure 3A**). The inclusion of data augmentation markedly enhanced performance, reducing misclassification, particularly in classes such as actinic keratosis, squamous cell carcinoma and vascular lesion (**Figure 3B**). Further refinement using Bayesian hyperparameter tuning resulted in near-perfect classification for most classes, with several, such as dermatofibroma, melanocytic nevus, squamous cell carcinoma and vascular lesion, achieving 100% accuracy (**Figure 3C**). Finally, the addition of an attention module achieved the highest overall performance, with nearly flawless classification across all classes (**Figure 3D**). These results highlighted the effectiveness of integrating data

augmentation, Bayesian tuning, and attention mechanisms in improving the model's precision, recall, and robustness.

The results of the performance matrix calculation for each scenario are presented in **Table 3**. Our data indicated that the pre-trained Inception-V3 model's performance significantly improved across all evaluation metrics—sensitivity, specificity, precision, accuracy, F1 score, and AUC—when data augmentation techniques, Bayesian tuning, and the attention mechanism were applied.

Table 3. Performance comparison of pre-trained Inception-V3 across four different scenarios

Model scenarios	SEN (%)	SPE (%)	PRE (%)	ACC (%)	F1 score (%)	AUC
Without augmentation	55.50	90.40	65.06	55.50	54.10	0.500
With augmentation	88.63	98.23	88.53	88.63	88.53	0.940
With augmentation and Bayesian tuning	96.38	99.47	96.40	96.38	96.29	0.979
With augmentation, Bayesian tuning and attention module	98.50	99.62	97.42	97.38	97.34	0.985

ACC: accuracy; AUC: area under the curve; PRE: precision; SEN: sensitivity; SPE: specificity

Initially, without augmentation, the model struggled to identify true positive cases, as evidenced by low sensitivity (55.50%) and accuracy (55.50%), indicating poor differentiation capabilities in varied or imbalanced data conditions. While the model showed relative strength in identifying true negatives (specificity at 90.40%), this also suggested a possible bias towards the majority class. Low precision (65.06%) and F1 score (54.10%), alongside an AUC of 0.500, indicated a lack of predictive accuracy and discriminative power, essentially no better than random guessing. The introduction of data augmentation marked a substantial improvement across all metrics, particularly in sensitivity and accuracy (both at 88.63%), demonstrating the significant benefits of augmenting data to enhance model generalization. This was further enhanced by Bayesian tuning, which optimized hyperparameters, resulting in even higher performance in sensitivity (96.38%), accuracy (96.38%), and a remarkable AUC of 0.979, reflecting excellent classification and discriminative abilities. The final implementation of the attention mechanism elevated performance to new heights, with sensitivity reaching 98.50% and accuracy at 97.38%, underscoring the model's improved ability to recognize positive cases and its overall high accuracy. Exceptionally high specificity (99.62%), precision (97.42%), F1 score (97.34%), and an AUC of 0.985 demonstrated outstanding model performance, with superb capability in distinguishing between different classes. In conclusion, the progression from no data augmentation to the incorporation of data augmentation techniques, Bayesian tuning, and the attention mechanism significantly enhanced every aspect of performance. This underscored the importance of data augmentation in enhancing data variability and representation, hyperparameter tuning for model optimization, and the attention mechanism for focusing on important features—collectively boosting accuracy, sensitivity, specificity, and the discriminative ability of the model.

The training time required by the model across various scenarios is summarized in **Table 4**. First, the significant reduction in training time from 427 minutes to 178 minutes when data augmentation was applied suggested the effectiveness of augmentation in enhancing model training efficiency. Data augmentation improved the diversity of the training dataset through transformations like rotation, scaling, and flipping. This approach helped prevent overfitting, promoted faster convergence, and improved the model's ability to generalize effectively. The introduction of Bayesian tuning further reduced the training time to 98 minutes, indicating the efficiency of this optimization technique in fine-tuning hyperparameters. Bayesian optimization searched for the optimal set of hyperparameters more systematically and efficiently than random or grid search, by building a probabilistic model of the function mapping hyperparameter values to the objective evaluated on a validation set. This method allowed for a more directed search for the optimal hyperparameters, resulting in faster convergence and reduced training time. Incorporating an attention mechanism into the model, alongside data augmentation and Bayesian tuning, extended the training time slightly to 125 minutes, reflecting the additional computational effort required to focus on the most relevant features of the data input. This

increase was due to the added computational complexity introduced by the attention mechanism. However, the attention mechanism enhanced model performance by enabling a focus on the most relevant features in the input data. Consequently, the trade-off between the marginally longer training time and the potential gained in accuracy and performance was justified.

Table 4. Duration of model training periods

Training scenarios	Duration (min)
Without augmentation	427
With augmentation	178
With augmentation and Bayesian tuning	98
With augmentation, Bayesian tuning and attention module	125

Discussion

The ISIC-2019 dataset that was utilized in the present study is one of the most comprehensive public datasets for dermoscopic images, featuring eight distinct classes of pigmented skin lesions. While its large size and detailed annotations make it a valuable resource for machine learning research, several limitations and biases could affect the generalizability of the results. First, the dataset was heavily imbalanced, with certain classes, such as melanocytic nevus, constituting nearly half of the total images. This imbalance could lead models to overfit on majority classes, even with techniques like data augmentation and weighted loss functions. Second, the dataset may not fully represent the diversity of real-world skin lesion characteristics, such as variations in skin type, lesion morphology, and imaging conditions. Most images were derived from controlled clinical environments, which may not reflect the heterogeneity of data encountered in non-clinical or resource-limited settings. These factors could potentially limit the applicability of the model to broader populations. Additionally, artifacts such as hair, ruler markings, or uneven illumination present in some images might inadvertently introduce noise, impacting model performance. Addressing these biases through strategies like domain adaptation or additional real-world data collection is essential to ensure robust and generalizable applications in clinical practice.

In terms of model complexity and performance trade-offs, the enhancements we made to the Inception-V3 architecture, including data augmentation, Bayesian hyperparameter tuning, and the introduction of a simplified attention mechanism, inevitably resulted in a slight increase in computational complexity. Specifically, the attention mechanism, while improving interpretability and focusing on relevant features, adds an additional layer of computation. This marginal increase in inference time, as evidenced by the slight rise in training duration from 98 minutes (with Bayesian tuning) to 125 minutes (with attention), is a necessary trade-off for the substantial improvements in classification accuracy and robustness. The overall computational requirements were still manageable, particularly when considering the high performance achieved—exemplified by a sensitivity of 98.5% and an AUC of 0.99—making this approach suitable for real-time applications. However, in resource-constrained environments, further optimization techniques could be explored to mitigate the impact on inference time without compromising model accuracy.

The improved classification accuracy achieved by our model has significant clinical implications for dermatological practice. With high sensitivity and specificity, the model enhanced the early detection of malignant lesions like melanoma, enabling timely intervention and reducing mortality rates. Its ability to minimize false positives can reduce unnecessary biopsies, optimizing clinical resources and improving patient experiences. Moreover, this model can support practitioners in resource-limited settings or integrate into telemedicine platforms to provide accessible dermatological care.

The inclusion of interpretable features also fostered trust and facilitates its adoption in routine workflows, ultimately contributing to better patient outcomes and more efficient dermatological practices. In summary, the analysis of the training times for the pre-trained Inception-V3 model across different scenarios highlighted the effectiveness of data augmentation and Bayesian tuning in improving training efficiency. While the addition of an attention module slightly increases training time, it offered potential benefits in enhancing model performance.

These insights could guide researchers and practitioners in optimizing their deep-learning models for both efficiency and effectiveness.

A comprehensive comparison of our modified Inception-V3 model's performance against prior studies on the classification of PSLs is provided in **Table 5**. The values in bold represent the highest achieved figures in each performance metric. Our evaluation is based on the ISIC-2019 dataset, encompassing eight classes of skin lesions. Our modified Inception-V3 model achieved the highest performance metrics across the board compared to other models. This included the highest sensitivity at 98.50%, specificity at 99.62%, precision at 97.42%, accuracy at 97.38%, F1 score at 97.34%, and AUC at 0.99. These results highlighted the effectiveness of our modifications in enhancing the model's diagnostic accuracy.

Specificity and sensitivity are critical for medical diagnosis to minimize false positives and false negatives. Our model's specificity (99.62%) and sensitivity (98.50%) outperform all other models, indicating a balanced and reliable performance in distinguishing between different types of pigmented skin lesions. Given the high-performance metrics, our modified Inception-V3 model was particularly well-suited for clinical applications where accuracy is paramount. Its ability to maintain high precision and sensitivity can aid dermatologists in making more accurate diagnoses, potentially leading to better patient outcomes.

Table 5. Performance evaluation against prior studies. Values in bold represent the highest achieved figures

Study	Year	Model	Performance metrics					
			SEN (%)	SPE (%)	PRE (%)	ACC (%)	F1 score (%)	AUC
[36]	2020	DenseNet-201	66.45	97.85	91.61	97.35	-	-
[27]	2020	GoogleNet	79.80	97.00	80.36	94.92	80.07	-
[26]	2021	Ensemble CNN	62.00	98.00	73.00	81.00	56.00	-
[37]	2023	Clinical Inspired	53.80	97.40	-	64.00	-	0.91
[17]	2023	CLCM-net	84.80	-	85.30	91.73	85.05	-
[18]	2023	Swin transformer	82.30	97.90	-	97.20	-	-
[19]	2023	Skin-net	70.78	96.78	72.56	94.65	71.33	-
[28]	2024	DCNN	97.12	99.61	97.09	97.11	97.08	0.99
Ours	2024	Modified Inception-V3	98.50	99.62	97.42	97.38	97.34	0.99

ACC: accuracy; AUC: area under the curve; CLCM: consecutive layerwise weight constraint MaxNorm model; CNN: convolutional neural network; DCNN: deep convolutional neural network; PRE: precision; SEN: sensitivity; SPE: specificity

The consistently high scores across multiple metrics for our modified Inception-V3 suggested that the enhancements that were made to the base model have effectively improved its robustness and generalizability for the classification task. This included better handling of the complexities and variations in the ISIC-2019 dataset. This superiority is attributed to several key innovations integrated into our approach. First, the introduction of a simplified soft-attention mechanism enhanced the model's capability to focus on spatially significant features within dermoscopic images, enabling better identification of critical lesion characteristics. Unlike traditional attention mechanisms, our simplified design maintained computational efficiency while improving model interpretability—an essential factor for clinical applicability. Second, the application of Bayesian hyperparameter optimization ensured that the model parameters were finely tuned to the specific challenges posed by the ISIC-2019 dataset, such as class imbalance and feature variability. This optimization led to a more robust and generalizable model performance, which was evidenced by its consistently high sensitivity (98.50%) and specificity (99.62%). Compared to previous methods like DCNN or Skin-Net, which primarily rely on deep convolutional layers, our approach effectively combined augmentation techniques and hyperparameter tuning to address dataset complexities. Moreover, the advanced data augmentation techniques employed in this study contribute significantly to the improved generalization of the model. By increasing the diversity and balance of training data, our model outperforms traditional architectures like DenseNet-201 and ensemble CNNs, which exhibited lower sensitivity and accuracy in handling complex multiclass classification tasks. Together, these innovations result in a notable leap in performance metrics, positioning our method as a leading approach in the classification of pigmented skin lesions.

Despite the promising results that were achieved by the Inception-V3-based model in classifying dermoscopic pigmented skin lesions, several limitations must be acknowledged. The first limitation is related to the dataset's generalizability. The model was trained primarily on the ISIC-2019 dataset, which may not fully represent the diversity of real-world skin types, lesion variations, or imaging conditions. This dataset-specific nature could limit the model's performance when applied to more diverse populations or imaging setups. Another significant limitation is the issue of class imbalance. Although techniques such as weighted loss functions and data augmentation were employed, the dataset's inherent class imbalance may still impact the model's sensitivity, particularly for underrepresented lesion types. The model's interpretability also presents constraints. While the integrated soft-attention mechanism provides insights into the model's focus areas, more advanced interpretability tools are needed to further clarify the decision-making process for clinicians. Additionally, the model's computational requirements pose challenges. Although the computational needs are optimized for performance, they may still create barriers in resource-constrained environments. Further work is necessary to improve efficiency without compromising accuracy. Lastly, deployment challenges must be considered. Variability in imaging devices and conditions could affect model consistency across different clinical environments, necessitating calibration for diverse dermoscopic equipment. Moreover, integrating the model into clinical workflows, such as electronic health record (EHR) systems, requires significant effort to ensure usability and adoption. Ethical and regulatory considerations, including data privacy and approval from governing bodies like the Food and Drug Administration (FDA) or European Conformity (EC) marking, must also be addressed before widespread deployment.

Conclusion

The present study demonstrated a notable enhancement in classification accuracy in comparison to the conventional pre-trained Inception-V3 model, exhibiting superior performance to other recent research endeavors within the domain of dermatology. The study employed a comprehensive evaluation of the Inception-V3 architecture for the classification of dermoscopic images of pigmented skin lesions, integrating techniques such as data augmentation, transfer learning, Bayesian optimization, and a soft attention mechanism. These enhancements resulted in a notable improvement in classification accuracy, thereby underscoring the pivotal role of sophisticated deep-learning techniques in medical image analysis.

This study offers valuable insights into the application of machine learning in dermatology, contributing to advancements in the diagnosis and treatment of skin cancer. Furthermore, it paves the way for future research in skin cancer diagnostics. The findings not only underscore the effectiveness of the proposed approach in refining model performance but also highlight its potential to develop more accurate and reliable diagnostic tools for dermatologists, enabling early and precise detection of skin cancer.

Ethics approval

Not required.

Acknowledgments

None

Competing interests

The authors confirm that they have no conflicts of interest to disclose.

Funding

This work was funded by the Indonesia Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan* (LPDP)).

Underlying data

The derived data supporting this study's results can be obtained from the corresponding author upon request.

Declaration of artificial intelligence use

This study used artificial intelligence tools and methodologies in manuscript writing support: AI-based language models, sush as ChatGPT, Quillbot, and DeepL were employed to language refinement.

How to cite

Nugroho ES, Ardiyanto I, Nugroho HA. Enhancing dermoscopic pigmented skin lesion classification: A refined approach using the pre-trained Inception-V3 architecture. Narra J 2025; 5(2): e1852 - http://doi.org/10.52225/narra.v5i2.1852.

References

1. Rembielak A, Tagliaferri L. Non-melanoma skin cancer. Boca Raton: CRC Press; 2023.
2. Rosendahl C, Marozava A. Dermatoscopy and skin cancer: A handbook for hunters of skin cancer and melanoma, updated edition 2023. Banbury: Scion Publishing; 2023.
3. Kassani SH, Kassani PH. A comparative study of deep learning architectures on melanoma detection. Tissue Cell 2019;58(4):76-83.
4. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin 2024;74(1):12-49.
5. Australian Institute of Health and Welfare. Cancer data in Australia (web report). Available from: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/about. Accessed: 10 October 2024.
6. Magalhaes C, Mendes J, Vardasca R. Systematic review of deep learning techniques in skin cancer detection. BioMed Inform 2024;4(4):2251-2270.
7. Vardasca R, Mendes JG, Magalhaes C. Skin cancer image classification using artificial intelligence strategies: A systematic review. J Imaging 2024;10(11):265.
8. Li H, He X, Zhou F, *et al*. Dense deconvolutional network for skin lesion segmentation. IEEE J Biomed Health Inform 2019;23(2):527-537.
9. Jahanifar M, Zamani TN, Mohammadzadeh AB, *et al*. Supervised saliency map driven segmentation of lesions in dermoscopic images. IEEE J Biomed Health Inform 2019;23(2):509-518.
10. Hatamizadeh A, Hoogi A, Sengupta D. Deep active lesion segmentation. In: Suk HI, Liu M, Yan P, editors. Machine learning in medical imaging. Heidelberg: Springer-Verlag; 2019.
11. Wei Z, Song H, Chen L, *et al*. Attention-based denseunet network with adversarial training for skin lesion segmentation. IEEE Access 2019;7:136616-136629.
12. Tu W, Liu X, Hu W, *et al*. Dense-residual network with adversarial learning for skin lesion segmentation. IEEE Access 2019;7:77037-77051.
13. Maurya A, Stanley RJ, Lama N, *et al*. A deep learning approach to detect blood vessels in basal cell carcinoma. Skin Res Technol 2022;28(4):571-576.
14. Afza F, Khan MA, Sharif M, *et al*. Microscopic skin laceration segmentation and classification: A framework of statistical normal distribution and optimal feature selection. Microsc Res Tech 2019;82(9):1471-1488.
15. Chatterjee S, Dey D, Munshi S, *et al*. Extraction of features from cross correlation in space and frequency domains for classification of skin lesions. Biomed Signal Process Control 2019;53:101581.
16. Ashfaq M, Minallah N, Ullah Z, *et al*. Performance analysis of low-level and high-level intuitive features for melanoma detection. Electronics 2019;8(6):672.
17. Gopikha S, Balamurugan M. Regularised layerwise weight norm based skin lesion features extraction and classification. Comput Syst Sci Eng 2023;44(3):2727-2742.
18. Ayas S. Multiclass skin lesion classification in dermoscopic images using swin transformer model. Neural Comput Appl 2023;35(9):6713-6722.
19. Alsahafi YS, Kassem MA, Hosny KM. Skin-Net: A novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier. J Big Data 2023;10(1):105.
20. Arora G, Dubey AK, Jaffery ZA, *et al*. A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data. Neural Comput Appl 2023;35(11):7989-8015.
21. Omeroglu AN, Mohammed HMA, Oral EA, *et al*. A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification. Eng Appl Artif Intell 2023;120(10):105897.

22. Benyahia S, Meftah B, Lézoray O. Multi-features extraction based on deep learning for skin lesion classification. Tissue Cell 2022;74(11):101701.
23. Nigar N, Umar M, Shahzad MK, *et al*. A Deep learning approach based on explainable artificial intelligence for skin lesion classification. IEEE Access 2022;10(11):113715-113725.
24. Hasan MK, Ahamad MA, Yap CH, *et al*. A survey, review, and future trends of skin lesion segmentation and classification. Comput Biol Med 2023;155(11):106624.
25. Houssein EH, Abdelkareem DA, Hu G, *et al*. An effective multiclass skin cancer classification approach based on deep convolutional neural network. Clust Comput 2024;27:12799-12819.
26. Cauvery, Siddalingaswamy, Pathan S, *et al*. A multiclass skin lesion classification approach using transfer learning based convolutional neural network. In: 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII). Chennai; 2021.
27. Kassem MA, Hosny KM, Fouad MM. Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access 2020;8:114822-114832.
28. Houssein EH, Abdelkareem DA, Hu G, *et al*. An effective multiclass skin cancer classification approach based on deep convolutional neural network. Clust Comput 2024;27:12799-12819.
29. Nugroho ES, Ardiyanto I, Nugroho HA. Boosting the performance of pretrained CNN architecture on dermoscopic pigmented skin lesion classification. Skin Res Technol 2023;29(11):1-10.
30. Nugroho ES, Ardiyanto I, Nugroho HA. Comparative performance of pre-trained CNN architectures on dermoscopic pigmented skin lesions classification. In: 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS). Yogyakarta; 2023.
31. Codella N, Rotemberg V, Tschandl P, *et al*. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging Collaboration (ISIC). arXiv 2019:1902.03368v2.
32. Combalia M, Codella NCF, Rotemberg V, *et al*. BCN20000: Dermoscopic lesions in the wild. arXiv 2019:1908.02288v2
33. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5(1):180161.
34. Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas; 2016.
35. Woo S, Park J, Lee J young, *et al*. CBAM: Convolutional block attention module. In: 15th European Conference on Computer Vision (ECCV). Munich; 2018.
36. Molina-Molina EO, Solorza-Calderón S, álvarez-Borrego J. Classification of dermoscopy skin lesion color-images using fractal-deep learning features. Appl Sci Switz 2020;10(17):5954.
37. Liu Z, Xiong R, Jiang T. CI-Net: Clinical-inspired network for automated skin lesion recognition. IEEE Trans Med Imaging 2023;42(3):619-632.