The effect of hair removal and filtering on melanoma detection: a comparative deep learning study with AlexNet CNN

Angélica Quishpe-Usca[1,*], Stefany Cuenca-Dominguez[1], Araceli Arias-Viñansaca[1], Karen Bosmediano-Angos[1], Fernando Villalba-Meneses[1], Lenin Ramírez-Cando[1], Andrés Tirado-Espín[2], Carolina Cadena-Morejón[2], Diego Almeida-Galárraga[1] and Cesar Guevara[3,*]

[1] School of Biological Sciences and Engineering, Yachay Tech University, San Miguel de Urcuquí, Imbabura, Ecuador
[2] School of Mathematical and Computational Sciences, Yachay Tech University, San Miguel de Urcuquí, Imbabura, Ecuador
[3] Quantitative Methods Department, CUNEF Universidad, Madrid, Madrid, Spain
[*] These authors contributed equally to this work.

ABSTRACT

Melanoma is the most aggressive and prevalent form of skin cancer globally, with a higher incidence in men and individuals with fair skin. Early detection of melanoma is essential for the successful treatment and prevention of metastasis. In this context, deep learning methods, distinguished by their ability to perform automated and detailed analysis, extracting melanoma-specific features, have emerged. These approaches excel in performing large-scale analysis, optimizing time, and providing accurate diagnoses, contributing to timely treatments compared to conventional diagnostic methods. The present study offers a methodology to assess the effectiveness of an AlexNet-based convolutional neural network (CNN) in identifying early-stage melanomas. The model is trained on a balanced dataset of 10,605 dermoscopic images, and on modified datasets where hair, a potential obstructive factor, was detected and removed allowing for an assessment of how hair removal affects the model's overall performance. To perform hair removal, we propose a morphological algorithm combined with different filtering techniques for comparison: Fourier, Wavelet, average blur, and low-pass filters. The model is evaluated through 10-fold cross-validation and the metrics of accuracy, recall, precision, and the F1 score. The results demonstrate that the proposed model performs the best for the dataset where we implemented both a Wavelet filter and hair removal algorithm. It has an accuracy of 91.30%, a recall of 87%, a precision of 95.19%, and an F1 score of 90.91%.

INTRODUCTION

Skin cancers are the most commonly diagnosed group worldwide (*Apalla et al., 2017*). They are generally classified into two major categories: melanoma and nonmelanoma skin cancer. Malignant melanomas may rarely occur in the mouth, intestine, or eye but

more commonly in the skin. Environmental factors as well as genetics play a role in the development of melanoma. Exposure to ultraviolet (UV) radiation from the sun and tanning lamps is taught to be the leading cause of melanoma (*Garibyan & Fisher, 2010*; *Sun et al., 2020*). Although it can affect people of all skin types, melanoma is notably more prevalent among males and individuals with fair skin complexion (*Howlader et al., 2014*). The highest incidence of melanoma is found in Australia, New Zealand, and North America, but it is also becoming increasingly prevalent in Africa, Asia, and South America (*Nikolaou & Stratigos, 2014*). According to the International Agency for Research on Cancer (IARC), melanoma is the most aggressive and prevalent type of skin cancer in the world, with an estimated 325 000 new cases identified worldwide and 57 000 deaths in 2020. If this tendency persists, it is projected that by 2,040, new cases will increase by more than 50% and deaths by 68% (*Arnold et al., 2022*).

Melanoma is considered the most dangerous type of skin cancer because it grows quickly and can spread to any organ. If not treated early, it might result in death or other severe, incurable conditions. Also, visible areas of the skin might be affected, resulting in psychological distress, depression, and emotional turmoil (*Dinnes et al., 2018*; *Vojvodic et al., 2019*). Early detection is crucial for successful melanoma treatment (*Xavier et al., 2016*). Visual examination by medical professionals is typically the initial technique employed in melanoma detection. It involves a comprehensive inspection of abnormal skin lesions, to categorize them as either malignant (melanoma) or benign, based on the ABCDE criteria (asymmetry, border, color, diameter, and evolving) (*Johr, 2002*; *Aljanabi et al., 2020*). It is important to note that the evidence to support the visual inspection's accuracy is insufficiently reported. Various authors agree that its accuracy might be influenced by factors such as the experience of physicians and the diverse appearance of the lesions (*Schwartz et al., 2002*; *See, 2012*; *Dinnes et al., 2018*; *Bansal, Garg & Soni, 2022*). Moreover, it requires significant investments of time, personal, and financial resources, making it a labor-intensive, time-consuming, and error-prone process (*Vestergaard et al., 2008*; *Yu et al., 2016*; *Dabeer, Khan & Islam, 2019*).

Deep-learning methods have emerged as a noninvasive technique for melanoma detection offering a promising approach to overcome the limitations associated with visual inspection. Nowadays, a great amount of work in the detection of skin cancer is concentrated on applying machine-learning techniques to dermoscopic images to improve diagnostic accuracy (*Dildar et al., 2021*). This technique allows efficient automation of skin lesion assessment based on the creation of statistical algorithms for learning purposes (*Char, Shah & Magnus, 2018*; *Fu'adah et al., 2020*). Deep learning can be conceptualized as a hierarchical process for eliciting features, inspired by the workings of the human brain (*Winkler et al., 2019*). These strategies analyze unstructured patterns in dermoscopic images by extracting and organizing deep and precise features that are not easily observed by simple visual inspection. These algorithms have the ability to evaluate large data sets at the pixel level in order to extract melanoma-specific features (*Hagerty et al., 2019*). The usefulness of the method lies in carefully examining every detail of the images and accurately associating it with melanoma, learning to classify by patterns and distinctive features (*Winkler et al., 2019*). This approach not only helps to reduce errors but also

allows a greater amount of data to be analyzed in less time, providing a timely diagnosis and treatment.

Skin lesion classification has recently been addressed in studies using Transfer Learning and Vision Transformers (ViTs) (*Dosovitskiy et al., 2020*; *Xin et al., 2022*). Transformers are a deep learning architecture based on multi-head self-attention mechanisms, originally proposed for sequence-to-sequence tasks in natural language processing (NLP), but now adapted for computer vision tasks such as image classification (*Usman, Zia & Tariq, 2022*). These transformers, known as vision transformers, have recently been applied to classification tasks of biological and epidemiological images of skin lesions. The performance of ViTs for multiclass classification of skin lesions has been compared with convolutional neural networks (CNNs), showing that when trained from scratch, they achieve lower performance (*Maurício, Domingues & Bernardino, 2023*). *Cirrincione et al. (2023)* have also proposed integrated models that combine transformer encoders with masked labels and visual features of CNNs for melanoma classification in ISIC datasets. However, CNNs exhibit higher computational efficiency on standard hardware and have been more researched in the medical imaging domain. In addition, other approaches for feature extraction in image classification studies include local binary pattern (LBP), Data-Efficient Image Transformers (DeiTs), and Convolutional Vision Transformers (CvT) (*Chen et al., 2021*).

CNNs are widely utilized in the field of deep learning for melanoma detection. The underlying principle behind CNNs lies in their architecture, which employs three major types of layers to teach the network how to identify and categorize patterns in images (*Yamashita et al., 2018*). Several convolutional layers are responsible for detecting features in the input image, such as edges and textures. They are usually followed by pooling layers, which aim to shrink large-size feature maps to create smaller feature maps to recognize some general patterns that are only perceptible in resized images (*Nasr-Esfahani et al., 2016*; *Alzubaidi et al., 2021*; *Zhang, 2021*). The third layer, named the fully connected layer, maps the extracted features into a final output, which is the classification. In comparison with traditional machine learning algorithms, CNNs are specifically designed to handle recognition tasks in image spatial structure. One of the main advantages of CNNs is their computational efficiency. Its weight-sharing feature reduces the number of trainable network parameters simplifying the training process and avoiding overfitting.

The VGG16 architecture, created by the Visual Geometry Group at Oxford University, marks a significant evolution in CNNs. Introduced in 2013, this model is renowned for its depth, consisting of 16 layers that include 13 convolutional layers (*Ibrahim et al., 2023*). A key innovation of VGG16 is its use of small 3×3 convolutional filters, a departure from larger filters used in earlier models. This design choice enhances its ability to process spatial features in images, aiding in more accurate feature extraction. VGG16's architecture is elegantly simple yet highly effective, making it a popular choice for various applications, especially in image classification tasks like skin cancer detection. Its structure involves multiple layers of convolution, each followed by non-linear activation functions, which contribute to the model's efficiency in learning complex patterns (*Hasan et al., 2021*). For skin cancer classification, VGG16 is often adapted through transfer learning, where

its pre-trained layers are fine-tuned to identify specific features of skin lesions. This methodology leverages the network's in-depth learning from vast datasets to provide precise classifications between malignant and benign skin conditions.

The reported variations in performance and accuracy across different segmentation tasks and datasets can be attributed to the design variations among different architectures (*Perez, Avila & Valle, 2019*). Numerous studies have shown that CNNs are highly effective for classification, localization, detection, segmentation, and registration in dermoscopic image analysis. These investigations have led to significant advancements in the development and refinement of CNN-based algorithms specifically designed to accurately and efficiently detect melanoma.

Table 1 summarizes previous studies on CNN-based skin cancer classification, providing insights into their efficiency, effectiveness, strengths, and weaknesses.

In this context, studies have evaluated strategies to measure model performance, highlighting the accuracy (ACC) metric emerges as a global metric that provides a comprehensive measure of classifier accuracy. Balanced accuracy (BACC) is the average between the sensitivity and the specificity and is prioritized for unbiased evaluation (*Brodersen et al., 2010*). Specificity (SP) for negative cases and sensitivity (SE) for positive cases are analyzed. Precision (PR) excels in the accurate identification of positive cases. These criteria are meticulously chosen to provide a comprehensive and fair assessment of the models' ability to classify efficiently, offering a complete picture of their performance.

Segmentation of dermoscopic images using CNNs can be significantly influenced by various factors, such as low contrast between the lesion and healthy skin, color variations, the presence of air bubbles, and the obstruction caused by hair covering the lesions. Hairs and their shadows on the skin may occlude relevant information and can cause segmentation algorithms to commit errors in estimating texture measures (*Oliveira et al., 2016*). Consequently, hair removal might be considered a crucial preprocessing step in training CNNs for accurate segmentation.

Traditional approaches have treated hair removal as a hair detection and occlusion inpainting problem, employing filter-based methods (*Nguyen, Lee & Atkins, 2010*; *Eltayef, Li & Liu, 2017*; *Li et al., 2021*). Recent advancements utilize multi-scale curvilinear matched filters for hair detection, followed by region-growing algorithms and linear discriminant analysis for precise localization. Modern data-driven approaches employ deep neural network auto-encoders trained on finely annotated datasets (*Bardou et al., 2022*). However, constructing paired hair-containing and hair-removed images for training these algorithms remains challenging due to their high cost and limited availability. While the effect of hair in dermoscopic images on CNN performance has been studied, it has yet not been sufficiently reported. Some studies have highlighted the significant improvement in classification performance achieved by removing hair from dermoscopic images (*Kim & Hong, 2021*; *Li et al., 2021*). Various types of filters such as Gaussian, Average, Median and Wiener have also been used to reduce noise and smooth dermoscopic images of medical images combined with machine learning algorithms (*Bektaş et al., 2018*).

The main objective of this study is to assess the impact of hair removal combined with different filtering techniques on the performance of a CNN utilizing the AlexNet

Table 1 References of skin cancer segmentation with typical CNN frameworks in the literature. Only the metrics for the best-performing architectures or datasets are presented in studies that involve the comparison of multiple architectures or datasets.

Ref/Year	Dataset	Architecture	Highlights	Limitations	Performance
(*Ameri, 2020*)/2020	HAM10000	AlexNet	The suggested approach eliminates the need for complex procedures of lesion segmentation and feature extraction by taking an unprocessed image as input and learning directly from the raw data.	Only 3,400 images from the dataset were used due to the need for an equal number of benign and malignant images for training.	ACC = 0.84, SE = 0.81, SP = 0.88
(*Yao et al., 2022*)/2022	ISIC-2017, ISIC-2018, ISIC-2019, 7-PT	RegNetY-3.2GF	The study proposes a novel Multi-weighted New Loss method to address the issue of class imbalance and improve accuracy in detecting key classes such as melanoma. RegNetY performed the best on the ISIC2018 dataset.	Almost all publicly accessible skin disease image datasets suffer a problem of severe data imbalance that might affect the performance of CNNs.	BACC = 0.858
(*Perez, Avila & Valle, 2019*)/2019	ISIC-2017	Inception-ResNet-v2, MobileNetV2 , PNAS-Net, ResNet , SENet, Xception, VGG16, VGG19, and DenseNet	The authors systematically assessed the factors that impact the selection of a CNN structure by examining 13 different factors across nine models.	The article's dataset has limitations as it is smaller in size compared to other studies and its exclusive focus on classifying melanoma.	Top-1 ACC = 0.827
(*Javid et al., 2023*)/2023	A recompilation of ISIC datasets	ResNet 50, EfficientNet B6, Inception V3, and Xception	The results obtained from each individual model are inputted into a meta-learner in order to combine and utilize the outputs from these models to make a final prediction.	–	ACC = 0.935, SE = 0.9, PR = 0.96, F1 score = 0.92
(*Alwakid et al., 2022*)/2022	HAM10000	Modified version of Resnet-50	The proposed method suggests utilizing DL to precisely extract a lesion zone. The approach involves enhancing the image quality using ESRGAN and then using segmentation to isolate Regions of Interest (ROI).	To showcase the effectiveness of the proposed technique, it is necessary to conduct more experiments on a sizable and intricate dataset that encompasses potential cancer cases.	ACC = 0.86, SE = 0.86, PR = 0.84, F1 score = 0.86

Table 1 (*continued*)

Ref/Year	Dataset	Architecture	Highlights	Limitations	Performance
(*Raza et al., 2022*)/2022	Clinical repositories in Korea	InceptionV3, Xception, InceptionResnetV2, DenseNet121, VGG16	A novel stacked ensemble framework has been introduced in the study, specifically designed to augment generalizability and bolster robustness in the context of acral lentiginous melanoma classification.	–	ACC = 0.979, SE = 0.978, PR = 0.98, F1 score = 0.98
(*Gupta & Mesram, 2022*)/2022	ISIC 2016–17	AlexNet, DenseNet-121	The study suggests a mixed CNN model that involves merging a pre-trained AlexNet CNN model with an optimized pre-trained DenseNet-121 CNN model.	Numerous healthcare institutions possess substantial patient data; however, they face challenges in making this information accessible to the public due to privacy concerns.	ACC = 0.9065, SE = 0.91, PR = 0.9065, F1 score = 0.91
(*Esteva et al., 2017*)/2017	Clinical repositories	GoogleNet, Inception v3	The model is adapted to be used on mobile devices. It is predicted that by the year 2021, there will be approximately 6.3 billion smartphone subscriptions globally.	Further investigations are necessary to evaluate how this method performs in a clinical setting.	ACC = 0.721

Notes.

Alwakid et al. (2022)
Ameri (2020)
Esteva et al. (2017)
Gupta & Mesram (2022)
Javid et al. (2023)
Perez, Avila & Valle (2019)
Raza et al. (2022)
Yao et al. (2022)

architecture for early melanoma detection. We use a dataset that is the collection of dermoscopic images from multiple freely available International Skin Imaging Collaboration (ISIC) and HAM 10,000 datasets. Additionally, modified datasets were obtained by detecting and subsequently removing hair by a morphological algorithm. Different filtering techniques are used to enhance image quality and remove noise and are applied before hair removal for further comparison. The main contributions of this research are: (i) combining a hair removal algorithm with four different filtering techniques (Fourier, Wavelet, Average blur, and Low-pass) ; (ii) conducting a comparison, in terms of accuracy, recall, precision, and F1 score, for each filter combination to determine its effect on the CNN performance for melanoma classification ; (iii) using a balanced dataset comprising dermoscopic images from various multi-class skin lesion datasets. However, a notable limitation is the similarity in metric results, prompting the need for further validation through one-factor analysis of variance.

Related work

In recent research, significant progress has been observed in the use of CNNs and other artificial intelligence techniques for skin cancer classification. These approaches have shown great potential to help specialists reduce the time and resources required for diagnosis, which, in turn, allows for more appropriate treatment. A study by *Leiter, Keim & Garbe (2020)* compared expert opinion with artificial neural networks, and found that the software established a sensitivity of 95% and specificity of 88%, results comparable to those reported by dermatologists.

Several researchers have proposed specific approaches for skin cancer classification using CNN and other deep-learning techniques. Among these approaches, *Yanchatuña et al. (2021)* used a combination of CNNs and support vector machines (SVMs) to detect and classify skin cancer, obtaining an average accuracy between 80.67% and 90%, with an outstanding performance of 90.34% for the AlexNet plus SVM model. *Popescu, El-Khatib & Ichim (2022)* developed a skin lesion classification system involving multiple CNNs trained on the HAM10000 dataset, capable of predicting seven skin lesions, including melanoma. *Ameri (2020)* presented a deep CNN using transfer learning with AlexNet, eliminating the need for complex segmentation and feature extraction procedures by automatically learning useful features from raw images.

Other researchers explored the use of pre-trained networks (such as AlexNet and VGG16) in learning transfer and as feature extractors (*Gulati & Bhogal, 2019*). VGG16 with transfer learning was found to outperform other techniques in terms of accuracy and efficiency. This is attributed to its deep network structure and the ability to adapt to specific tasks through the use of pre-trained weights, resulting in better performance in the classification of skin lesions, as has been documented in various investigations (*Anand et al., 2022*). *Alwakid et al. (2022)* proposed an approach that incorporates ESRGAN, segmentation techniques, and a CNN along with a modified version of Resnet-50 for accurate classification. Similarly, hybrid models were presented that combine CNNs with SVM classifiers (*Keerthana et al., 2023*) or dimension reduction techniques such as PCA (*Olayah et al., 2023*). *Naeem et al. (2022)* developed the SCDNet model, which combines

VGG16 with CNN and compares its accuracy with pre-trained classifiers in the medical domain.

In addition, comparisons of various CNN architectures were performed, where GoogleNet proved to be the most accurate (74.91% and 76.08%) in both training and test sets (*Aljohani & Turki, 2022*). *Acosta et al. (2021)* proposed a two-stage approach using CNN-based masks and regions and a ResNet152 structure to classify lesions as "benign" or "malignant". *Nida et al. (2019)* introduced an automated melanoma region segmentation method based on a deep region-based convolutional neural network (RCNN) that accurately detects multiple affected regions. *Gouda & Amudha (2020)* presented LCNet, a model that requires no preprocessing or computation of specific features. *Tahir et al. (2023)* compared their novel DSCC_Net model with other deep benchmark networks, demonstrating improved accuracy.

Finally, *Bansal, Garg & Soni (2022)* introduced methods to remove hair in dermatoscopic images and used integrated features extracted using manual techniques and deep learning models to improve melanoma detection. Experimental results highlighted improvements in accuracy compared to approaches that do not apply preprocessing or use manual or deep learning features separately.

MATERIALS & METHODS

This study proposes a method for training and testing a CNN with AlexNet architecture for early melanoma detection. Four filtering techniques are implemented before a morphological hair removal algorithm for further comparison. First, dermoscopic images of skin lesions are obtained and preprocessed from the so-called "MSCD10000 dataset" (*Javid et al., 2023*). Then, the different layers of the CNN architecture are designed and finally, the model performance is evaluated using 10-fold cross-validation and four metrics. The software used in the current study is Python.

MSCD10000 dataset

MSCD10000 dataset stands for Melanoma Skin Cancer Dataset of 10,000 Images and consists of 10,605 dermoscopic images scaled to 300x300 pixels (*Javid et al., 2023*). These images were collected and resized from various publicly available ISIC (ISIC 2019, ISIC 2018) and HAM 10,000 datasets to create a balanced dataset with both benign and malignant classes. The dataset was downloaded from Kaggle at the following link: https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images. It is organized into two main categories, *train* and *test*. Within each category, the dataset is further divided into *benign* and *malignant* classes. Table 2 provides details on the distribution of dermoscopic images across these categories.

Hair removal and filtering

Before the hair removal process on dermoscopic images, we used four different filtering techniques to reduce noise and improve image quality: Fourier, Wavelet, Average blur, and Low-pass. Each filter was integrated using *Python 3.x*, *cv2* libraries, and *OpenCV*, with our codebase adeptly adjusted to accommodate the unique requirements of each filter type. To

Table 2 Dataset distribution.

	Train		Test		Total (%)
	Number	Percentage (%)	Number	Percentage (%)	
Malignant	4,605	43.42	500	4.71	5,105 (48.14%)
Benign	5,000	47.14	500	4.71	5,500 (51.86%)
Total	9,605	90.57	1 000	9.43	

illustrate, when implementing the low-pass filter, we applied a Gaussian blur with a 3×3 kernel size. *PyWavelets* for the implementation of Wavelet filters was used. The choice of these libraries allowed for efficient and reproducible processing of the images.

The process of filter implementation and hair removal is illustrated in Fig. 1. First, the images were converted to grayscale. They were then processed with a black top-hat filter and thresholding, followed by an inpainting method based on cv2.INPAINT_TELEA to restore the image (*Telea, 2004*). This process was performed using a 17x17 size core for the MORPH BLACKHAT operation, highlighting dark spots, which correspond mainly to hair, in the dermoscopic images (*Ashraf et al., 2022*).

After applying the filtering techniques, the grayscale images were subjected to a morphological operation for hair detection. We use a threshold value of 10 for binarization of the images, classifying them into pixels that represent hair (greater than 10) and background pixels (less than or equal to 10). The images were stored in separate folders labeled "with hair" and "without hair."

The computational complexity of the hair removal process was considerable. On average, processing each image folder took between 2:30 and 3 h, with a memory requirement of approximately 6 GB. This time includes the application of the filters, the grayscale conversion, the morphological operation, the thresholding, and the filling method (inpainting). These factors were essential to ensure the viability of the method in both research and clinical practice.

Image pre-processing

The recommended optimal size for input images to ensure compatibility with the CNN architecture's input layer is 224 × 224 × 3 (*Krizhevsky, Sutskever & Hinton, 2012*). The resized images are converted into a PyTorch tensor and then, the tensor values for each image are normalized sequentially with the mean and standard deviation values for each color channel. That is, scaling the values so that they have a mean of 0 and a standard deviation of 1 (*LeCun et al., 1998*; *Glorot & Bengio, 2010*; *Ioffe & Szegedy, 2015*).

Furthermore, data augmentation is used to address overfitting issues and improve the model's capacity to generalize to new data by increasing the number of training samples for each class through random transformations not only for imbalanced datasets but also for datasets of any size (*Perez et al., 2018*; *Ali et al., 2022*). We applied different transformations, which included random horizontal flipping with a 50% probability, random rotations of up to 10 degrees, random adjustments in brightness, contrast, saturation, and hue within a maximum range of 0.2, as well as color channel normalization and conversion to the torch. Tensor format to effectively utilize the images as inputs.
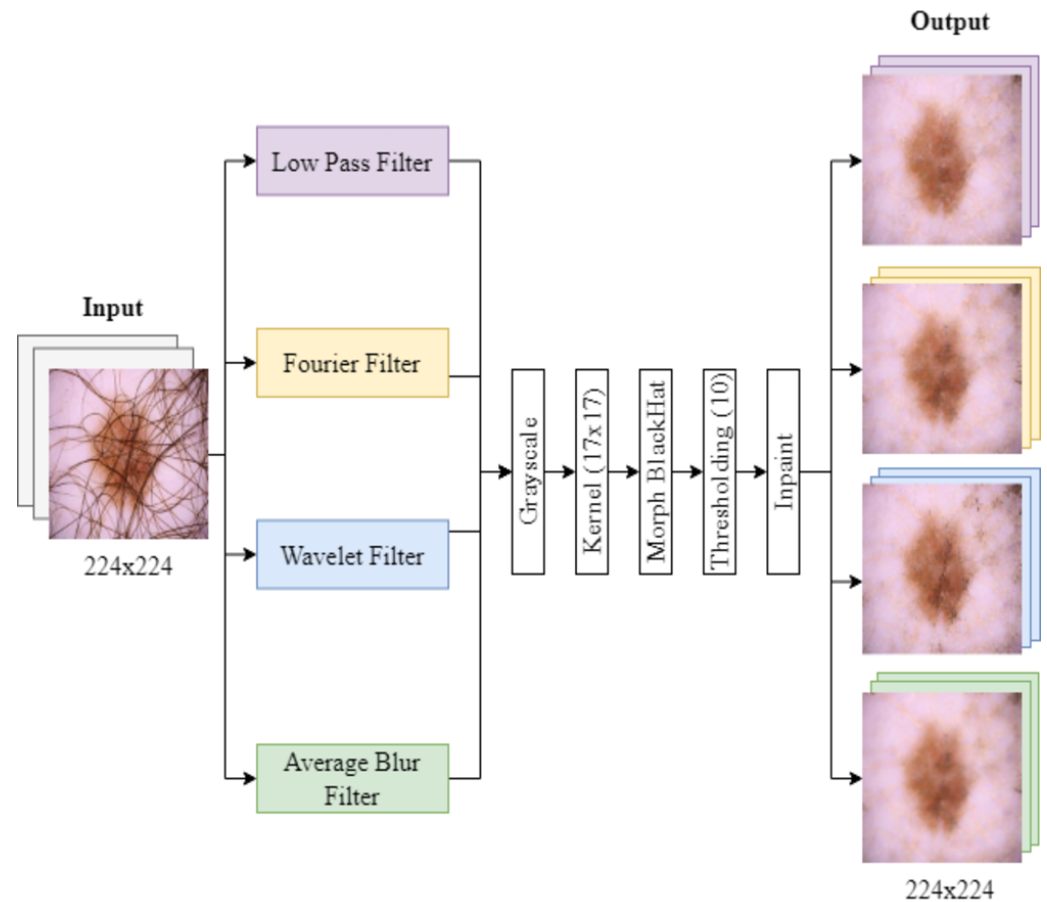
Figure 1 **Filtering and hair removal algorithm.** Images are first turned grayscale, then processed with a black top-hat filter and thresholding, and lastly, an inpainting method restores the image.

Full-size 🖼 DOI: 10.7717/peerjcs.1953/fig-1

AlexNet architecture

A visual representation of the AlexNet architecture used in this study is provided in Fig. 2. It was trained with hyperparameters carefully chosen to balance effective model convergence and computational efficiency. The learning rate was set to 0.001 to control parameter updates, while a momentum of 0.9 accelerated convergence by incorporating past gradients. A weight decay of 0.001 added regularization to prevent overfitting. The Cross-Entropy Loss function, suitable for multi-class or binary classification, was employed to measure prediction accuracy. A batch size of 64 determined the number of samples processed in each iteration, balancing memory constraints and training speed. The model underwent 30 epochs.

The initial layout encompasses five convolutional layers, each using specialized filters to identify specific features, such as contours and textures, present in the original image. These convolutional layers alternate with maximum aggregation layers, designed to condense the feature maps and simplify the recognition of more holistic patterns, particularly those that might remain unnoticed in larger images. The first, second, third, and fifth convolutional
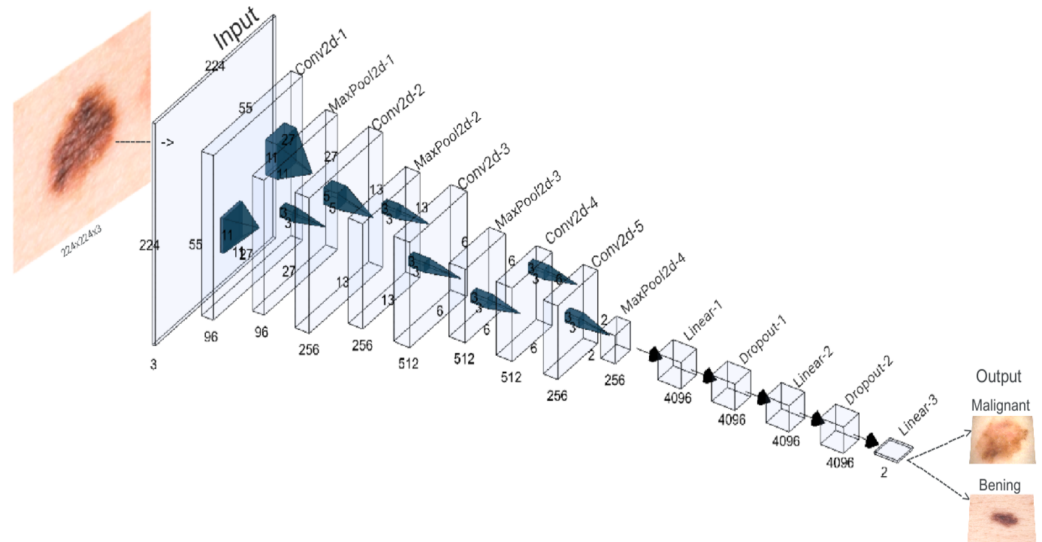
Figure 2 AlexNet architecture of the proposed CNN. A new adaptive layer is added between the convolutional and fully connected layers to modulate the binary output size.

Full-size ⬜ DOI: 10.7717/peerjcs.1953/fig-2

layers are seconded by maximum aggregation layers. An additional classifier module, consisting of three fully connected layers, is introduced to map the extracted features and thus generate the final binary classification.

In a variation on the original AlexNet structure, a new adaptive layer is incorporated between the convolutional and fully connected layers. This dynamically adjustable layer modulates the size of the output tensor to fit a standardized 6x6 dimension. This adjustment leads to an optimization of the model's performance, allowing it to efficiently adapt to the diversity of features present in the medical images under analysis.

Performance evaluation

We employed the K-fold cross-validation technique for evaluating the average performance of the model in terms of the average training and test loss. The *train* directory of each dataset, original and modified, is divided into 10 folds of equal size. The choice of the value for *k* was determined based on the dataset's size. The model undergoes 10 training iterations, each consisting of 30 epochs, with one fold serving as the validation set in each iteration and the remaining 9 folds used for training.

In addition, a quantitative evaluation of the CNN performance was carried out using four common metrics: accuracy, sensitivity, precision, and F1 score. True negatives (TN) and true positives (TP) refer to the accurate classification of negative and positive instances, respectively. On the other hand, false negatives (FN) and false positives (FP) refer to the incorrect classification of positive and negative instances, respectively.

$$Accuracy = \frac{True\ detected\ melanoma\ cases\ (TP + TN)}{All\ cases}$$

$$Recall = \frac{True\ detected\ melanoma\ cases\ (TP)}{All\ melanoma\ cases\ (TP + FN)}$$

$$Precision = \frac{True\ detected\ melanoma\ cases\ (TP)}{True\ and\ false\ detected\ melanoma\ cases\ (TP + FP)}$$

$$F1\ score = \frac{Precision * Recall}{Precision + Recall}.$$

RESULTS AND DISCUSSION

In this section, we describe the effectiveness of the proposed CNN for detecting melanoma. We first evaluate the impact of hair removal on its performance through a comparison of the results for the model trained in the original and the modified datasets, where filtering techniques and hair removal were applied. Subsequently, we compare the outcomes of the four metrics obtained from the model trained on the best-performing dataset to those achieved by existing studies in the field.

Hair removal results

To ensure optimal model selection, we performed model evaluation every 30 epochs for each fold. Figure 3 shows the training and testing (validation) loss *vs.* epochs for the different datasets (original and modified) presenting a detailed analysis of the evolution of the loss throughout the training. The model curves are not smooth, instead, a lot of fluctuations are observed. These fluctuations could be attributed to several factors, with batch size being one of them. Notably, prior research suggests that a batch size of 32 often leads to optimal results in neural network training and as batch size increases, the convergence can be negatively impacted (*Keskar et al., 2017*). However, it is observed that the training loss is reduced to 0.1 for all of the model loss curves, which can be considered a good sign for correct segmentation results as other studies conclude (*Ishida et al., 2021*; *Guefrechi et al., 2021*).

The loss curve for the original images shows an increase in both training and testing loss at 30 epochs, which is not observed in the rest of the curves. Furthermore, even when data augmentation was used, some overfitting was observed when using the Fourier filter before hair removal, which can cause a decrease in generalization ability.

The overall CNN performance is determined by comparing the results of the four metrics between the original dataset and the modified datasets. The results in Table 3 reveal that the model performs the best among three of the four metrics with images where we have applied the Wavelet filter for noise reduction combined with hair removal. Here, the model achieves an accuracy of 91.30%, and a recall of 87%, while the low-pass filter yields a slightly higher recall of 89.80%. Additionally, when utilizing the Wavelet filter, the model has a precision of 95.19% and an F1 score of 90.91%.
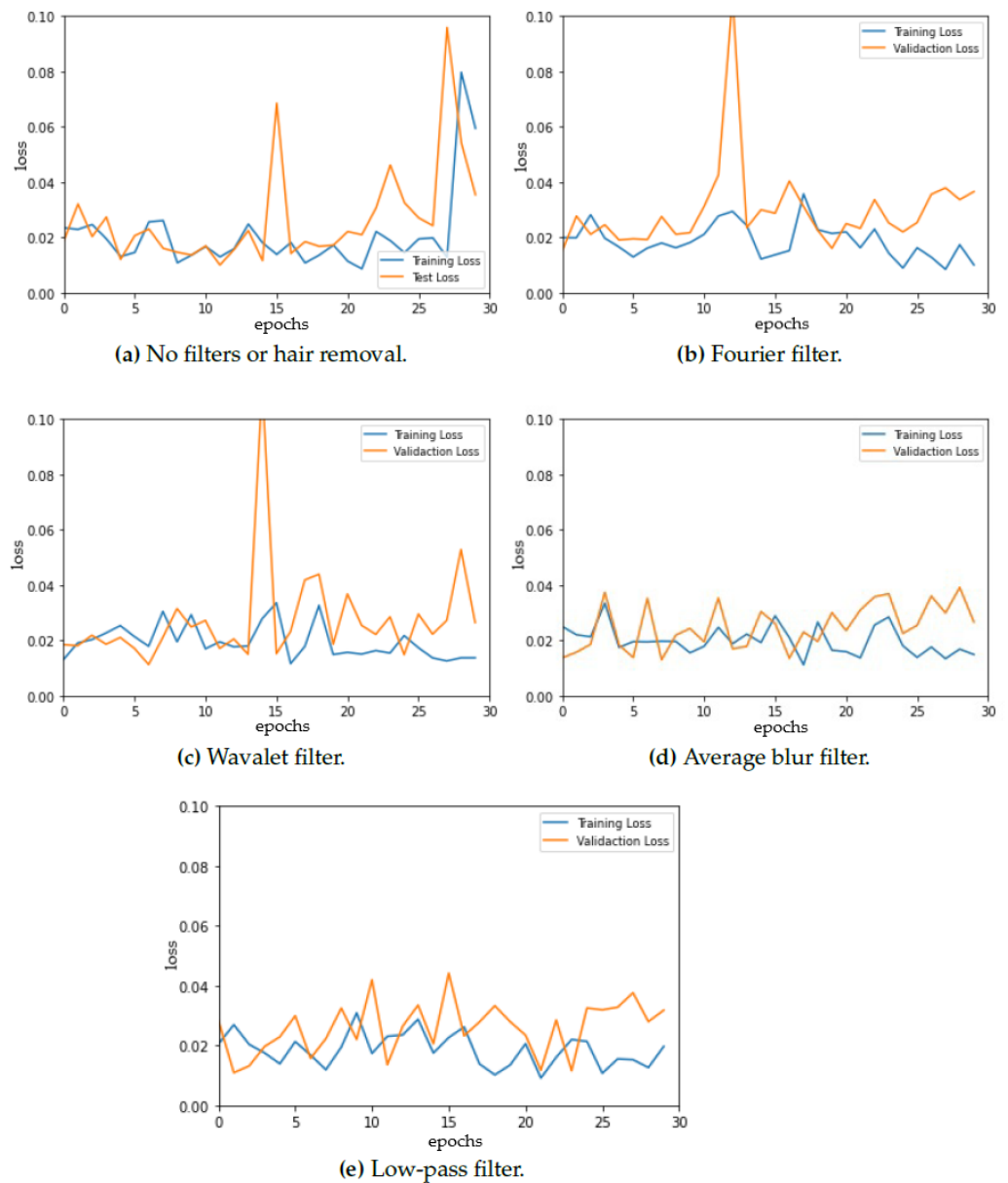
(a) No filters or hair removal.

(b) Fourier filter.

(c) Wavalet filter.

(d) Average blur filter.

(e) Low-pass filter.

Figure 3 Model loss for (A) original dermoscopic images *vs.* modified images that combine (B) Fourier filter, (C) Wavelet filter, (D) average blur, and (E) low-pass filter with hair removal.

Full-size ☑ DOI: 10.7717/peerjcs.1953/fig-3

The results indicate that hair removal improves the model's accuracy and precision when it is combined with the Wavelet filter allowing the model to not only make correct predictions overall but also minimize false positives. Wavelet transform has already been used to equalize the noise created by fine hairs in dermoscopic images as well as to increase the algorithm's detection speed in hybrid models combining deep learning and machine learning (*Suiçmez et al., 2023*). Some of the highest accuracy and sensitivity scores in the literature have been obtained when using Wavelet filters (*Narasimhan & Elamaran, 2016*).

Table 3 Trained model results on the original images from the dataset and the images where different filtering techniques were applied and hair was removed.

Metrics	Filtering technique combined with hair removal				
	None	Fourier	Wavelet	Average blur	Low-pass
Accuracy (AC)	91.10%	90.90%	91.30%	90.80%	90.00%
Recall (SE)	87.60%	88.20%	87.00%	87.40%	89.80%
Precision (PR)	94.19%	93.23%	95.19%	93.78%	90.16%
F1 score (F1 sc)	90.78%	90.65%	90.91%	90.48%	89.98%

On the other hand, the model trained on the original images exhibits the second-highest accuracy and precision of 91.10% and 94.19%, respectively. However, the second highest sensibility is reached when using the Fourier filter combined with hair removal.

These results are supported by the graphical representation of a box plot (Fig. 4), which visually points out the significant difference between the filters as a function of their performance. In Fig. 4, it is important to note that the Y-axis of the plot does not represent a specific metric, but rather the combination of four different metrics, thus providing a comprehensive view of each filter's performance. The length of the boxes in the figure indicates the variability in the performance of each filter: those with longer boxes exhibit higher variability, while those with shorter boxes exhibit lower variability. This observation is supported by examining the confidence intervals, which are considerably wider for the Wavelet filter compared to the other filters. However, it is pertinent to note that, despite this superiority, the confidence intervals for the Wavelet filter are also the widest, indicating considerable variability in its performance. This observation raises the possibility that, although the Wavelet filter excels in general terms, its performance may fluctuate significantly in different scenarios, which is why we consider that further investigation into the robustness and generalizability of the filter's performance in diverse settings is needed.

Furthermore, it is important to note that the significance of this improvement in relation to variations in the initialization of network parameters or other potential factors is not extensively discussed since hyper-tuning techniques were not applied to the network used in this study. It could be suggested to continually evaluate and potentially fine-tune these hyperparameters and initialization methods based on the characteristics of the melanoma dataset to ensure robust training and accurate classification results.

Overall, these outcomes strongly suggest the model's competence in effectively distinguishing between benign and malignant skin lesions, as evidenced by its robust performance metrics. CNNs offer significant advantages for physicians in clinical settings since they can process large volumes of data quickly and consistently, continuously learn new information, and offer objective assessments (Krishnan et al., 2023). For instance, the study by Haenssle et al. (2018) compared a CNN's diagnostic performance with that of a large international group of 58 dermatologists, revealing the CNN's superior performance. The CNN exhibited a higher specificity (82.5%) compared to dermatologists at both level-I (71.3%, $P < 0.01$) and level-II (75.7%, $P < 0.01$) sensitivities of 86.6% and 88.9%, respectively. Nonetheless, experts maintain an edge in handling complex or ambiguous skin lesion cases.
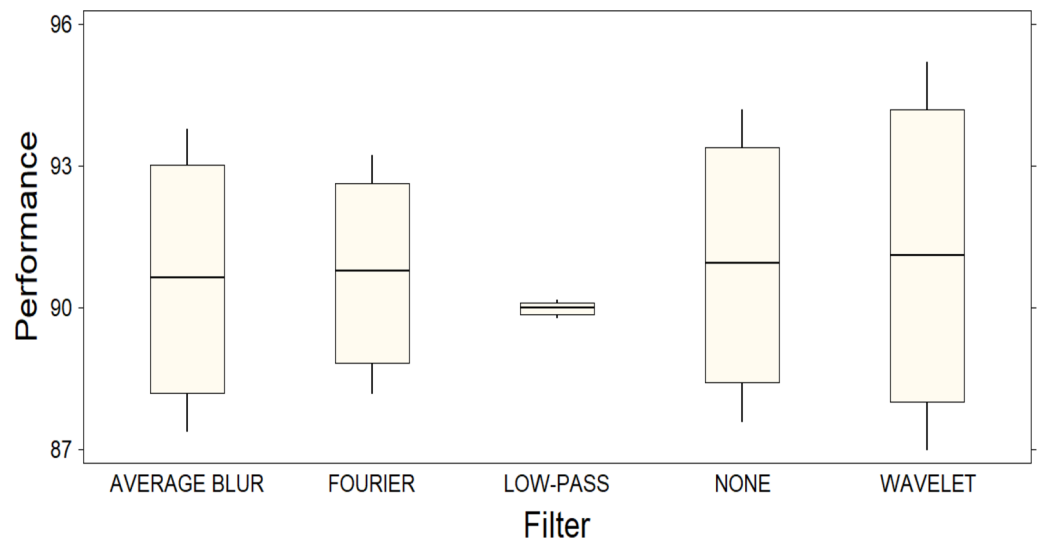
Figure 4 Box plot representing the distribution of performance among different filters. A significant difference between the filters is observed, highlighting the superior performance of the Wavelet filter compared to the others.

Table 4 Quantitative comparison of the results for the different metrics. All values were rounded to 3 decimal places.

Reference	Model	AC	SE	PR	F sc
(*Ameri, 2020*)	AlexNet	0.840	0.810	–	–
(*Javid et al., 2023*)	Some	0.935	0.900	0.960	0.920
(*Alwakid et al., 2022*)	Modified Resnet-50	0.860	0.860	0.840	0.860
(*Raza et al., 2022*)	Some	0.979	0.978	0.980	0.980
(*Gupta & Mesram, 2022*)	Alexnet, DenseNet-121	0.907	0.910	0.906	0.910
Proposed 1	Modified AlexNet	0.913	0.870	0.952	0.909
Proposed 2	Modified AlexNet	0.910	0.876	0.942	0.908

Notes.
Alwakid et al. (2022)
Ameri (2020)
Gupta & Mesram (2022)
Javid et al. (2023)
Raza et al. (2022)

Comparison with related work

We validated our model's performance against two setups: Wavelet filter with hair removal (Proposed 1) and original images without hair removal (Proposed 2). These results are contrasted with some well-established recent studies in the field shown in Table 1. As can be seen in Table 4, the proposed model was optimal in terms of the four metrics: accuracy (AC), recall (SE), precision (PR), and F1 score (Fsc). It has one of the highest accuracy and precision values.

Single-factor analysis of variance (ANOVA) was applied to investigate the existence of statistically significant differences in performance among various neural network models.

Table 5 Results of the single-factor analysis of variance (ANOVA) based on the data presented in Table 4.

	DF*	SS**	MS***	F****	P*****
Models	6	0.91	0.15	4.55	0.0042
Error	21	0.70	0.03	–	–
Total	27	1.61	–	–	–

Notes.
*DF (Degrees of Freedom).
**SS (Sum of Squares).
***MS (Mean Square).
****F (F-statistic).
*****P (*p*-value).

This method will provide a comprehensive perspective on whether at least one model is statistically different from the others with respect to the performance metrics under evaluation. The results, obtained through Statistix 10 software and presented in the table below, reveal an F-statistic of 4.55 with an associated *p*-value of 0.0042 as shown in Table 5.

The associated *p*-value being below the critical threshold of 0.05, leads to the rejection of the null hypothesis. Consequently, the existence of at least one significant difference in performance between the various neural network models is established.

Overall, the proposed model exhibits competitive performance when compared to the referenced studies. It achieves similar levels of accuracy, recall, precision, and F1 scores, emphasizing its efficacy in detecting melanoma in dermoscopic images. However, it is important to acknowledge that there is still room for improvement to develop and compare new filters for noise removal that might enhance hair removal outcomes in the field of melanoma detection.

CONCLUSIONS

This study focused on improving the performance of a CNN with AlexNet architecture for early melanoma detection by integrating hair removal and various filtering techniques. The model showed improvements in accuracy and precision when Wavelet filtering combined with hair removal was applied. Thus, filtering and hair removal improve image quality and enhance the model's ability to correctly classify images as melanoma. Finally, the proposed model on images from the original and modified dataset was evaluated and compared with other models in the field, demonstrating high levels of accuracy, precision, recall, and F1 score. In percentage terms, when evaluating the proposed model on the original dataset it achieved an accuracy of 91.10% and 91.30% on the modified images.

Overall, this study contributes to the development of a reliable and accurate system for melanoma detection using CNN. Future research directions could involve enhancing filtering techniques, exploring combinations of different filters, and developing adaptive filters to evaluate their impact on CNN performance concerning image quality and noise reduction. Additionally, the focus will shift towards implementing these systems in medical environments using convolutional networks, with the potential to significantly reduce mortality rates, enhance diagnostic precision, and streamline healthcare processes. The

innovative applications of this technology extend to vital domains such as radiology, digital pathology, and telemedicine, promising notable advancements in medical diagnostics.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Angélica Quishpe-Usca conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Stefany Cuenca-Dominguez conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Araceli Arias-Viñansaca conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Karen Bosmediano-Angos conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Fernando Villalba-Meneses conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Lenin Ramírez-Cando conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Andrés Tirado-Espín conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Carolina Cadena-Morejón conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Diego Almeida-Galárraga conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Cesar Guevara conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at Mendeley: Quishpe-Usca, Angélica; Cuenca-Dominguez, Stefany; Arias-Viñansaca, Araceli; Bosmediado-Angos, Karen; Villalba-Meneses, Fernando; Ramírez, Lenin; Tirado-Espín, Andrés; Cadena-Morejón, Carolina; Guevara, Cesar; Almeida-Galárraga, Diego (2024), ''CNN for Melanoma Detection Data'', Mendeley Data, V3, doi: 10.17632/ggh6g39ps2.3.

The original data is available at the Kaggle:

Available at https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.1953#supplemental-information.

REFERENCES

Acosta MFJ, Tovar LYC, Garcia-Zapirain MB, Percybrooks WS. 2021. Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Medical Imaging* **21**:6 DOI 10.1186/s12880-020-00534-8.

Ali K, Shaikh ZA, Khan AA, Laghari AA. 2022. Multiclass skin cancer classification using EfficientNets—a first step towards preventing skin cancer. *Neuroscience Informatics* **2**:100034 DOI 10.1016/j.neuri.2021.100034.

Aljanabi M, Enad MH, Chyad RM, Jumaa FA, Mosheer AD, Sali Altohafi A. 2020. A review ABCDE evaluated the model for decision by dermatologists for skin lesions using bee colony. *IOP Conference Series: Materials Science and Engineering* **745**:12098 DOI 10.1088/1757-899X/745/1/012098.

Aljohani K, Turki T. 2022. Automatic classification of melanoma skin cancer with deep convolutional neural networks. *AI* **3**:512–525 DOI 10.3390/ai3020029.

Alwakid G, Gouda W, Humayun M, Sama NU. 2022. Melanoma detection using deep learning-based classifications. *Healthcare* **10(12)**:2481 DOI 10.3390/healthcare10122481.

Alzubaidi L, Zhang J, Humaidi AJ, Duan Y, Santamaría J, Fadhel MA, Farhan L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* **8**:53 DOI 10.1186/s40537-021-00444-8.

Ameri A. 2020. A deep learning approach to skin cancer detection in dermoscopy images. *Journal of Biomedical Physics and Engineering* **10**:801–806 DOI 10.31661/jbpe.v0i0.2004-1107.

Anand V, Gupta S, Altameem A, Nayak SR, Poonia RC, Saudagar AKJ. 2022. An enhanced transfer learning based classification for diagnosis of skin cancer. *Diagnostics* **12**:1628 DOI 10.3390/diagnostics12071628.

Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Ioannides D. 2017. Epidemiological trends in skin cancer. *Dermatology Practical & Conceptual* **7**:1–6 DOI 10.5826/dpc.0702a01.

Arnold M, Singh D, Laversanne M, Vignat J, Vaccarella S, Meheus F, Cust AE, De Vries E, Whiteman DC, Bray F. 2022. Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA Dermatology* **158**:495–503 DOI 10.1001/jamadermatol.2022.0160.

Ashraf H, Waris M, Ghafoor M, Gilani S, Niazi I. 2022. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Scientific Reports* **12**:3948 DOI 10.1038/s41598-022-07885-y.

Bansal P, Garg R, Soni P. 2022. Detection of melanoma in dermoscopic images by integrating features extracted using handcrafted and deep learning models. *Computers & Industrial Engineering* **168**:108060 DOI 10.1016/j.cie.2022.108060.

Bardou D, Bouaziz H, Lv L, Zhang T. 2022. Hair removal in dermoscopy images using variational autoencoders. *Skin Research and Technology* **28**:445–454 DOI 10.1111/srt.13145.

Bektaş B, Emre ÍE, Kartal E, Gulsecen S. 2018. Classification of mammography images by machine learning techniques. In: *2018 3rd International conference on computer science and engineering (UBMK)*. Piscataway: IEEE, 580–585 DOI 10.1109/UBMK.2018.8566380.

Brodersen KH, Ong CS, Stephan KE, Buhmann JM. 2010. The balanced accuracy and its posterior distribution. In: *In 2010 20th international conference on pattern recognition*. Piscataway: IEEE, 3121–3124 DOI 10.1109/ICPR.2010.764.

Char DS, Shah NH, Magnus D. 2018. Implementing machine learning in health care — addressing ethical challenges. *New England Journal of Medicine* **378**(11):981–983 DOI 10.1056/nejmp1714229.

Chen L, Li S, Bai Q, Yang J, Jiang S, Miao Y. 2021. Review of image classification algorithms based on convolutional neural networks. *Remote Sensing* **13**(22):4712 DOI 10.3390/rs13224712.

Cirrincione G, Cannata S, Cicceri G, Prinzi F, Currieri T, Lovino M, Militello C, Pasero E, Vitabile S. 2023. Transformer-based approach to melanoma detection. *Sensors* **23**(12):5677 DOI 10.3390/s23125677.

Dabeer S, Khan MM, Islam S. 2019. Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked* **16**:100231 DOI 10.1016/j.imu.2019.100231.

Dildar M, Akram S, Irfan M, Khan HU, Ramzan M, Mahmood AR, Alsaiari SA, Saeed AHM, Alraddadi MO, Mahnashi MH. 2021. Skin cancer detection: a review using

deep learning techniques. *International Journal of Environmental Research and Public Health* **18**:5479 DOI 10.3390/ijerph18105479.

Dinnes J, Deeks JJ, Grainge MJ, Chuchu N, di Ruffano LF, Matin RN, Thomson DR, Wong KY, Aldridge RB, Abbott R, Fawzy M, Bayliss SE, Takwoingi Y, Davenport C, Godfrey K, Walter FM, Williams HC. 2018. Visual inspection for diagnosing cutaneous melanoma in adults. *Cochrane Database of Systematic Reviews* **2018(12)**:CD013194 DOI 10.1002/14651858.CD013194.

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. 2020. An image is worth 16x16 words: transformers for image recognition at scale. ArXiv arXiv:2010.11929.

Eltayef K, Li Y, Liu X. 2017. Detection of melanoma skin cancer in dermoscopy images. *Journal of Physics: Conference Series* **787**:12034 DOI 10.1088/1742-6596/787/1/012034.

Esteva A, Kuprel B, Novoa R, Ko J, Swetter SM, Blau HM, Thrun S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542(7639)**:115–118 DOI 10.1038/nature21056.

Fu'adah YN, Pratiwi NC, Pramudito MA, Ibrahim N. 2020. Convolutional neural network (CNN) for automatic skin cancer classification system. In: *IOP conference series: materials science and engineering.* IOP Publishing Ltd DOI 10.1088/1757-899X/982/1/012005.

Garibyan L, Fisher DE. 2010. How sunlight causes melanoma. *Current Oncology Reports* **12**:319–326 DOI 10.1007/s11912-010-0119-y.

Glorot X, Bengio Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics.* 249–256.

Gouda N, Amudha J. 2020. Skin cancer classification using ResNet. In: *2020 IEEE 5th International conference on computing communication and automation (ICCCA).* Piscataway: IEEE, 536–541 DOI 10.1109/ICCCA49541.2020.9250855.

Guefrechi S, Jabra M, Ammar A, Koubaa A, Hamam H. 2021. Deep learning based detection of COVID-19 from chest X-ray images. *Multimedia Tools and Applications* **80**:31803–31820 DOI 10.1007/s11042-021-11192-5.

Gulati S, Bhogal RK. 2019. Detection of Malignant Melanoma using deep learning. In: Singh M, Gupta P, Tyagi V, Flusser J, Ören T, Kashyap R, eds. *Advances in computing and data sciences. ICACDS 2019. Communications in computer and information science,* Cham: Springer DOI 10.1007/978-981-13-9939-8_28.

Gupta P, Mesram S. 2022. AlexNet and DenseNet-121-based hybrid CNN architecture for skin cancer prediction from dermoscopic images. *International Journal for Research in Applied Science and Engineering Technology* **10**:540–548 DOI 10.22214/ijraset.2022.45325.

Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Kalloo A, Hassen ABH, Thomas L, Enk A, Uhlmann L, Reader study level-I and level II Groups, Alt C, Arenbergerova M, Bakos R, Baltzer A, Bertlich I, Blum A, Bokor-Billmann T, Bowling J, Braghiroli N, Braun R, Buder-Bakhaya K, Buhl T, Cabo H, Cabrijan L,

Cevic N, Classen A, Deltgen D, Fink C, Georgieva I, Hakim-Meibodi L-E, Hanner S, Hartmann F, Hartmann J, Haus G, Hoxha E, Karls R, Koga H, Kreusch J, Lallas A, Majenka P, Marghoob A, Massone C, Mekokishvili L, Mestel D, Meyer V, Neuberger A, Nielsen K, Oliviero M, Pampena R, Paoli J, Pawlik E, Rao B, Rendon A, Russo T, Sadek A, Samhaber K, Schneiderbauer R, Schweizer A, Toberer F, Trennheuser L, Vlahova L, Walkd A, Winkler J, Wölbing P, Zalaudek I. 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology* **29(8)**:1836–1842 DOI 10.1093/annonc/mdy166.

Hagerty JR, Stanley RJ, Almubarak HA, Lama N, Kasmi R, Guo P, Drugge RJ, Rabinovitz HS, Oliviero M, Stoecker WV. 2019. Deep learning and hand-crafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. *IEEE Journal of Biomedical and Health Informatics* **23(4)**:1385–1391 DOI 10.1109/JBHI.2019.2891049.

Hasan MR, Fatemi MI, Khan MM, Kaur M, Zaguia A. 2021. Comparative analysis of skin cancer (Benign *vs.* Malignant) detection using convolutional neural networks. *Journal of Healthcare Engineering* **2021**:5895156 DOI 10.1155/2021/5895156.

Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. 2014. SEER Cancer Statistics Review. Bethesda: National Cancer Institute *Available at* https://seer.cancer.gov/archive/csr/1975_2011/index.html.

Ibrahim A, Elbasheir M, Badawi S, Mohammed A, Alalmin A. 2023. Skin cancer classification using transfer learning by VGG16 architecture (case study on Kaggle dataset). *Journal of Intelligent Learning Systems and Applications* **15**:67–75 DOI 10.4236/jilsa.2023.153005.

Ioffe S, Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *ICML'15: Proceedings of the 32nd International conference on machine learning - Volume 37*. 448–456.

Ishida T, Yamane I, Sakai T, Niu G, Sugiyama M. 2021. Do we need zero training loss after achieving zero training error? In: *ICML'20 Proceedings of the 37th international conference on machine learning. Volume 428*. 4604–4614.

Javid MH, Jadoon W, Ali H, Ali MD. 2023. Design and analysis of an improved deep ensemble learning model for melanoma skin cancer classification. In: *2023 4th International conference on advancements in computational sciences (ICACS)*. Piscataway: IEEE, 1–6 DOI 10.1109/ICACS55311.2023.10089716.

Johr RH. 2002. Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. *Clinics in Dermatology* **20**:240–247 DOI 10.1016/s0738-081x(02)00236-5.

Keerthana D, Venugopal V, Nath MK, Mishra M. 2023. Hybrid convolutional neural networks with SVM classifier for classification of skin cancer. *Biomedical Engineering Advances* **5**:100069 DOI 10.1016/j.bea.2022.100069.

Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M, Tang PTP. 2017. On large-batch training for deep learning: generalization gap and sharp minima. ArXiv arXiv:1609.04836.

Kim D, Hong B-W. 2021. Unsupervised feature elimination via generative adversarial networks: application to hair removal in melanoma classification. *IEEE Access* **9**:42610–42620 DOI 10.1109/ACCESS.2021.3065701.

Krishnan G, Singh S, Pathania M, Gosavi S, Abhishek S, Parchani A, Dhar M. 2023. Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm. *Frontiers in Artificial Intelligence* **6**:1227091.

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. *Commnications of the ACM* **60(6)**:84–90 DOI 10.1145/3065386.

LeCun Y, Bottou L, Bengio Y, Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**:2278–2324 DOI 10.1109/5.726791.

Leiter U, Keim U, Garbe C. 2020. Epidemiology of skin cancer: update 2019. In: Richrath J, ed. *Sunlight, Vitamin D and skin cancer.* Cham: Springer, 123–139 DOI 10.1007/978-3-030-46227-7_6.

Li W, Joseph Raj AN, Tjahjadi T, Zhuang Z. 2021. Digital hair removal by deep learning for skin lesion segmentation. *Pattern Recognition* **117**:107994 DOI 10.1016/j.patcog.2021.107994.

Maurício J, Domingues I, Bernardino J. 2023. Comparing vision transformers and convolutional neural networks for image classification: a literature review. *Applied Sciences* **13(9)**:5521 DOI 10.3390/app13095521.

Naeem A, Anees T, Fiza M, Naqvi RA, Lee SW. 2022. SCDNet: a deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. *Sensors* **22(15)**:5652 DOI 10.3390/s22155652.

Narasimhan K, Elamaran V. 2016. Wavelet-based energy features for diagnosis of melanoma from dermoscopic images. *International Journal of Biomedical Engineering and Technology* **20**:243–252 DOI 10.1504/IJBET.2016.075427.

Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, Najarian K. 2016. Melanoma detection by analysis of clinical images using convolutional neural network. In: *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC).* Piscataway: IEEE, 1373–1376 DOI 10.1109/embc.2016.7590963.

Nguyen NH, Lee TK, Atkins MS. 2010. Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel. In: *Proceedings of the society of photo-optical instrumentation engineers (SPIE) medical imaging 2010: image processing.* DOI 10.1117/12.844572.

Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT. 2019. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. *International Journal of Medical Informatics* **124**:37–48 DOI 10.1016/j.ijmedinf.2019.01.005.

Nikolaou V, Stratigos AJ. 2014. Emerging trends in the epidemiology of melanoma. *British Journal of Dermatology* **170**:11–19 DOI 10.1111/bjd.12492.

Olayah F, Senan EM, Ahmed IA, Awaji B. 2023. AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features. *Diagnostics* **13(7)**:1314 DOI 10.3390/diagnostics13071314.

Oliveira RB, Mercedes Filho E, Ma Z, Papa JP, Pereira AS, Tavares JMRS. 2016. Computational methods for the image segmentation of pigmented skin lesions: a review. *Computer Methods and Programs in Biomedicine* **131**:127–141 DOI 10.1016/j.cmpb.2016.03.032.

Perez F, Avila S, Valle E. 2019. Solo or Ensemble? Choosing a CNN architecture for melanoma classification. In: *2019 IEEE/CVF Conference on computer vision and pattern recognition workshops (CVPRW)*. Piscataway: IEEE, 2775–2783 DOI 10.1109/CVPRW.2019.00336.

Perez F, Vasconcelos C, Avila S, Valle E. 2018. Data augmentation for skin lesion analysis. In: *OR 2.0 Context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis. Lecture notes in computer science.* Cham: Springer DOI 10.1007/978-3-030-01201-4_33.

Popescu D, El-Khatib M, Ichim L. 2022. Skin lesion classification using collective intelligence of multiple neural networks. *Sensors* **22(12)**:4399 DOI 10.3390/s22124399.

Raza R, Zulfiqar F, Tariq S, Anwar GB, Sargano AB, Habib Z. 2022. Melanoma classification from dermoscopy images using ensemble of convolutional neural networks. *Mathematics* **10(1)**:26 DOI 10.3390/math10010026.

Schwartz JL, Wang TS, Hamilton TA, Lowe L, Sondak VK, Johnson TM. 2002. Thin primary cutaneous melanomas: associated detection patterns, lesion characteristics, and patient characteristics. *Cancer* **95**:1562–1568 DOI 10.1002/cncr.10880.

Suiçmez Ç, Tolga Kahraman H, Suiçmez A, Yılmaz C, Balcı F. 2023. Detection of melanoma with hybrid learning method by removing hair from dermoscopic images using image processing techniques and wavelet transform. *Biomedical Signal Processing and Control* **84**:104729 DOI 10.1016/j.bspc.2023.104729.

See JE. 2012. Visual inspection : a review of the literature. Technical Report. Sandia National Laboratories, USDOE National Nuclear Security Administration (NNSA). Washington D.C.: Department of Energy, Office of Science and Technical Information DOI 10.2172/1055636.

Sun X, Zhang N, Yin C, Zhu B, Li X. 2020. Ultraviolet radiation and melanomagenesis: from mechanism to immunotherapy. *Frontiers in Oncology* **10**:951 DOI 10.3389/fonc.2020.00951.

Tahir M, Naeem A, Malik H, Tanveer J, Naqvi RA, Lee SW. 2023. DSCC_Net: Multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images. *Cancer* **15(7)**:2179 DOI 10.3390/cancers15072179.

Telea A. 2004. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools* **9**:23–34 DOI 10.1080/10867651.2004.10487596.

Usman M, Zia T, Tariq A. 2022. Analyzing transfer learning of vision transformers for interpreting chest radiography. *Journal of Digital Imaging* **35**(6):1445–1462 DOI 10.1007/s10278-022-00666-z.

Vestergaard ME, Macaskill P, Holt PE, Menzies SW. 2008. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *British Journal of Dermatology* **159**:669–676 DOI 10.1111/j.1365-2133.2008.08713.x.

Vojvodic A, Vlaskovic-Jovicevic T, Vojvodic P, Vojvodic J, Goldust M, Peric-Hajzler Z, Matovic D, Sijan G, Stepic N, Tirant M. 2019. Psychological impact of melanoma, how to detect, support and help. *Open Access Macedonian Journal of Medical Sciences* **7**(18):3043–3045.

Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, Lallas A, Blum A, Stolz W, Haenssle HA. 2019. Association between surgical skin markings in images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatology* **155**(10):1135–1141 DOI 10.1001/jamadermatol.2019.1735.

Xavier MHSB, Drummond-Lage AP, Cyntia Baeta, Rocha L, Almeida AM, Wainstein AJA. 2016. Delay in cutaneous melanoma diagnosis: sequence analyses from suspicion to diagnosis in 211 patients. *Medicine* **95**:e4396 DOI 10.1097/MD.0000000000004396.

Xin C, Liu Z, Zhao K, Miao L, Ma Y, Zhu X , et al, Chen H. 2022. An improved transformer network for skin cancer classification. *Computers in Biology and Medicine* **149**:105939 DOI 10.1016/j.compbiomed.2022.105939.

Yamashita R, Nishio M, Do RKL, Togashi K. 2018. Convolutional neural networks: an overview and application in radiology. *Insights Into Imaging* **9**:611–629 DOI 10.1007/s13244-018-0639-9.

Yanchatuña OP, Pereira JP, Pila KO, Vásquez PA, Veintimilla KS, Villalba-Meneses GF, Alvarado-Cando O, Almeida-Galárraga D. 2021. Skin lesion detection and classification using convolutional neural network for deep feature extraction and support vector machine. *International Journal on Advanced Science, Engineering and Information Technology* **11**(3):1260–1267 DOI 10.18517/ijaseit.11.3.13679.

Yao P, Shen S, Xu M, Liu P, Zhang F, Xing J, Shao P, Kaffenberger B, Xu RX. 2022. Single model deep learning on imbalanced small datasets for skin lesion classification. *IEEE Transactions on Medical Imaging* **41**(5):1242–1254 DOI 10.1109/TMI.2021.3136682.

Yu L, Chen H, Dou Q, Qin J, Heng P-A. 2016. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging* **36**:994–1004 DOI 10.1109/TMI.2016.2642839.

Zhang R. 2021. Melanoma detection using convolutional neural network. In: *2021 IEEE International conference on consumer electronics and computer engineering (ICCECE)*. Piscataway: IEEE, 75–78 DOI 10.1109/ICCECE51280.2021.9342142.