Article

Aspects of Lighting and Color in Classifying Malignant Skin Cancer with Deep Learning

Alan R. F. Santos *, **Kelson R. T. Aires** and **Rodrigo M. S. Veras**

Department of Computing, Federal University of Piauí, Teresina CEP 64049-550, PI, Brazil;
kelson@ufpi.edu.br (K.R.T.A.); rveras@ufpi.edu.br (R.M.S.V.)
* Correspondence: arfs@ufpi.edu.br

Abstract: Malignant skin cancers are common in emerging countries, with excessive sun exposure and genetic predispositions being the main causes. Variations in lighting and color, resulting from the diversity of devices and lighting conditions during image capture, pose a challenge for automated diagnosis through digital images. Deep learning techniques emerge as promising solutions to improve the accuracy of identifying malignant skin lesions. This work aims to investigate the impact of lighting and color correction methods on automated skin cancer diagnosis using deep learning architectures, focusing on the relevance of these characteristics for accuracy in identifying malignant skin cancer. The developed methodology includes steps for hair removal, lighting, and color correction, defining the region of interest, and classification using deep neural network architectures. We employed deep learning techniques such as LCDPNet, LLNeRF, and DSN for lighting and color correction, which still need to be tested in this context. The results emphasize the importance of image preprocessing, especially in lighting and color adjustments, where the best results show an accuracy increase of between 3% and 4%. We observed that different deep neural network architectures react variably to lighting and color corrections. Some architectures are more sensitive to variations in these characteristics, while others are more robust. Advanced lighting and color correction can thus significantly improve the accuracy of malignant skin cancer diagnosis.

Keywords: skin cancer; lighting; color constancy; deep learning

1. Introduction

Tropical countries exhibit a high incidence of malignant skin cancer, with carcinoma and melanoma being the most common types [1]. These neoplasms have the potential to develop throughout the body, with a higher likelihood in areas most exposed to the sun [2]. Risk factors include fair skin, family history, and the presence of atypical lesions. Melanoma is the most aggressive type of skin cancer due to its metastatic capacity, but carcinomas can also reach this stage [3].

Color is an essential aspect in the diagnosis of malignant lesions. Variations in the hues of a lesion can reflect abnormalities, aiding in the diagnosis of skin cancer [4]. Shades of brown, black, or other atypical colors may indicate the presence of malignancies [5]. Early identification of changes in lesion color is an important factor for early diagnosis and effective treatment [6].

The diagnosis of skin cancer is fast; the problem lies in the delayed search for medical care by patients [7]. In early stages, malignant lesions can be confused with benign lesions due to the lack of clear symptoms. An increasingly adopted solution in dermatology is the visual monitoring of lesions, aided by computerized systems for analysis of digital images [8]. The use of deep learning has shown promising results comparable to the diagnostic accuracy of dermatologists [9].

The quality of the images is a relevant feature in developing deep neural network models and requires proper treatment [10]. Public image databases training these networks

do not follow a standard capture protocol regarding lighting [11]. Another factor is the variability of sensors, such as dermatoscopes, digital cameras, and smartphones [12]. The lack of standardization can compromise the consistency in the representation of colors and the reliability in the automated classification of neoplasms in digital images [5].

Considering the importance of accurate diagnosis and the challenges posed by variations in lighting and color in skin images, we propose a study that investigates the relevance of these aspects in the classification of skin cancer with deep learning. The goal is to understand the impact of lighting and color variations on classification models employing deep neural networks. This research develops an approach that combines existing techniques but prioritizes innovations in correcting lighting and color variations and the correlations of these variations in accurate classification with deep learning models.

The justification for investigating this work lies in the advances in deep neural networks. The progress of these techniques has allowed for increased accuracy in detecting malignant lesions in digital images. Therefore, it is essential to investigate the need to propose such techniques for the correction and treatment of lighting and color variations in images of skin lesions. Furthermore, in addition to proposing corrections for such variations, it is also necessary to understand how the interaction between lighting and color in classification with deep learning affects the precise identification of malignant lesions.

The main contribution of this research is to develop an approach that synergistically combines techniques based on the clinical optics of identifying malignant skin cancer. This approach prioritizes aspects of lighting and color, considering the divergence in the capture and treatment of skin lesion images. Moreover, this approach adapts, applies, and optimizes deep learning techniques never used before to deal with lighting and color bias. Although well-known deep learning models in the literature are used, the contributions lie in the reliable comparative analysis produced by these models.

Related Work

The relevance of color in automated skin cancer diagnosis has been highlighted in several studies, surpassing texture characteristics in the effectiveness of image classification [4,5]. Variation in lighting directly affects the representation of colors, resulting in a wide range of prototypic appearances in images of malignant lesions. This variability negatively affects the precision of detection results [10,13].

Some analyses indicate that color constancy significantly increases the accuracy in diagnosing malignant lesions in images [4]. In this regard, a comparative evaluation between approaches that use local and global characteristics demonstrated the superiority of color characteristics over textural ones in identifying melanomas. Among the evaluated strategies, the approach focusing on local characteristics proved more effective, underscoring the importance of color adjustments to increase the accuracy of automatic diagnosis [14]. This evidence suggests that lighting correction techniques generate positive effects in maintaining color constancy [10].

Barata et al. [15] provides a comprehensive overview of feature extraction, focusing on four classes: manual, dictionary-based, deep learning, and clinically inspired. This study highlights the evolution of techniques and improvements in the accuracy of detecting malignant skin cancer. The transition from manual techniques to advanced deep learning methods is relevant. However, this study also points out challenges such as the low quality of images in databases described in the literature and the efficiency of generalization across different skin types and lighting conditions, corroborating with [16].

Color standardization plays a significant role in the segmentation of skin lesion [17,18]. Statistical techniques such as Gray World (GW) [19], Shades of Gray (SoG) [20], and maxRGB [21] stand out among the common approaches to adjust this aspect. These methods seek to uniformly correct the lighting of images through adjustments based on statistical models, which can directly impact the uniformity of colors [16]. However, they may result in the loss of important information for the accurate analysis of skin cancer [22].

New approaches for correcting lighting and color aspects based on deep learning have emerged in the literature. These methodologies can predict and optimize these characteristics in digital images, gaining prominence after multiple validations on public image databases [23]. A significant example of this innovation is the Illumination Equalization using the Counter Exponential Transform (IECET) method, developed by [22]. This method applies counter-exponential transformations to the luminosity component in the HSV color space while maintaining the hue and saturation components. It also uses a regressor based on the ResNet architecture to calibrate the parameters, ensuring the homogeneity and adequacy of lighting in images of skin lesions.

The deep learning technique generative adversarial network (GAN) has been applied in new color normalization processes for medical images, being adaptable to identify color constancy through uniform lighting adjustment. This methodology benefits from both supervised and unsupervised learning. A notable method is the Dermo Color Constancy–Generative Adversarial Network (DermoCC–GAN), developed by [16], which manages to heuristically preprocess images of skin lesions, significantly improving lighting and color constancy. Dermatologists have positively evaluated the correction results achieved by this method [13].

Lighting calibration, aimed at improving accuracy in color representation, is already a common practice in histopathological images [24]. Applying deep learning architectures has proven promising in correcting lighting and color aspects, and improving tumor and cancer detection [25]. Cong et al. [26] proposed a notable study employing GAN with a generator network of dual decoders to ensure consistency between outputs. Furthermore, it also utilizes histogram losses. The technique adapts to color variations while keeping the structural information intact. The results of this study demonstrate an improvement in classification performance.

While statistical methods can improve the quality of images, they face significant limitations for skin lesion images [26]. Deep learning techniques, employed to standardize lighting and normalize colors, demonstrate greater promise in dermatological contexts. They apply to histopathological and other medical images [16]. The main challenges of these techniques include the variability of resulting color tones and the balance between normalization and the preservation of details [24].

The investigation of lighting correction techniques is still relevant for improving accuracy in the diagnosis of skin cancer. Studies show that lighting variations can distort colors in images. Maintaining color consistency has been shown to increase the accuracy of results, highlighting the importance of choosing and applying appropriate adjustments. While traditional methods, such as Gray World and Shades of Grays, are used in images of skin lesions, potentially resulting in losses of color consistency, deep learning techniques like IECET and DermoCC-GAN present promising solutions. These solutions may be a viable alternative for improving classification rates of malignant lesions.

As deep learning is also widely used in classifying skin cancer images, it is important to investigate how color variations are handled in such techniques [6]. Lighting can affect the perception of colors in lesions, and subtle differences in hues may lead to incorrect diagnoses. Understanding how colors are handled in deep learning could allow adjustments in architectures to identify color nuances, improving early detection of lesions in malignant conditions.

2. Materials and Methods

The research proposed in this work investigates the evaluation of lighting correction and color constancy using three deep learning methods. This correction is assessed and analyzed in classification with well-known deep architectures in the literature. The developed methodology aims to assess the impact of these characteristics on learning and classification in deep architectures for malignant skin cancer diagnosis. Figure 1 describes the steps to execute the experiments.
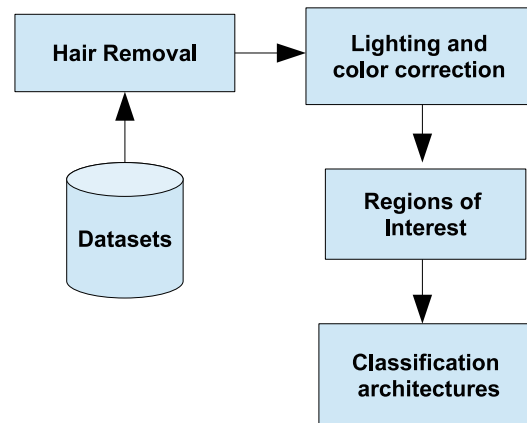
Figure 1. Steps of the suggested experimental methodology.

This work developed a methodology that was executed in four main stages. The first stage removes excess hair present in the images. The second adjusts the lighting and color aspects using deep learning techniques. The third defines the region of interest, cropping the image at the best bounding box coordinates. The last stage trains and classifies the resulting images using deep learning architectures. The following subsections will describe the methods applied at each stage, following a logical and coherent order.

2.1. Datasets

The experiments conducted in this work use the following public databases of dermatological images: International Skin Imaging Collaboration (ISIC 2020) [27], Dermatological and Surgical Assistance Program of the Federal University of Espírito Santo (PAD-UFES-20) [28], Diverse Dermatology Images (DDI) [29], and PH^2 [30]. These databases include clinical and dermatoscopy images covering various categories of skin lesions. For this study, we focus exclusively on two main categories: malignant and benign lesions. The malignant lesions include melanoma, basal, and squamous cell carcinoma. All other types of lesions are considered benign. Table 1 details the distribution adopted in this study, presenting the number of images per class.

Table 1. Quantitative images per database for both malignant and benign classes.

Datasets	Benign	Malignant	Total
ISIC 2020	32.542	584	33.126
PAD-UFES	1.215	1.094	2.298
DDI	644	219	625
PH^2	160	40	200
			36.280

2.2. Hair Removal

The excess of hair can compromise the accuracy of diagnosis prediction. The methodology developed to alleviate this problem treats hair as curvilinear structures. Morphological closing operations in the directions of 0°, 45°, and 90° are performed to identify these structures. One noteworthy aspect is that this technique allows for the detection and removal of hair without significantly affecting the characteristics of the lesions.

The hair removal process begins with the image's decomposition into channels, followed by the binarization of each component. Subsequently, morphological filters are applied, considering the rotation angles to identify segments of pixels forming curves. The final phase involves combining these components to create a specific hair mask. This mask is then applied to the original image, effectively removing pixels representing hair. A bilinear interpolation technique is used to restore areas affected by this removal. Figure 2 illustrates a practical example of this method.
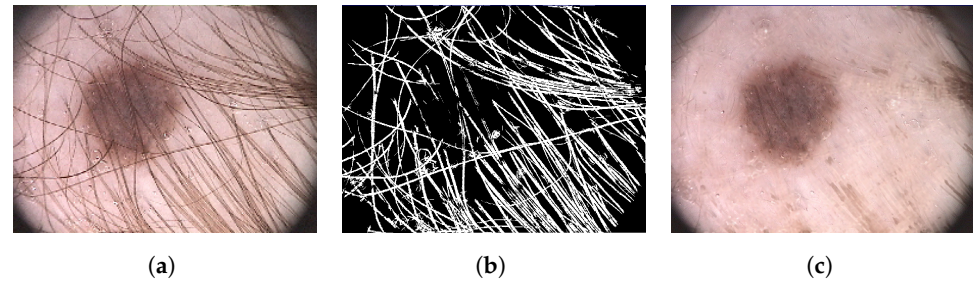
Figure 2. Removal of excess hair. The first image (**a**) is the input to the method. The second image (**b**) is the mask produced by combining the components. The third image (**c**) is the corrected image with bilinear interpolation.

2.3. Region of Interest

The process of reducing the image to highlight only the area of the lesion begins with segmentation. This phase aims to establish bounding box coordinates to delineate the lesion areas instead of enhancing edge definitions. These coordinates crop the input image, reducing it to the size containing only the lesion content, eliminating parts of healthy skin and noise.

This work utilizes the DeepLabV3+ segmentation technique developed by [31]. This architecture, designed for semantic segmentation tasks, employs atrous convolution to capture contextual information at multiple scales without losing the original image resolution. The technique modifies the receptive field of convolution filters, enhancing the ability to learn spatial dependencies more efficiently. Figure 3 illustrates the applied process.
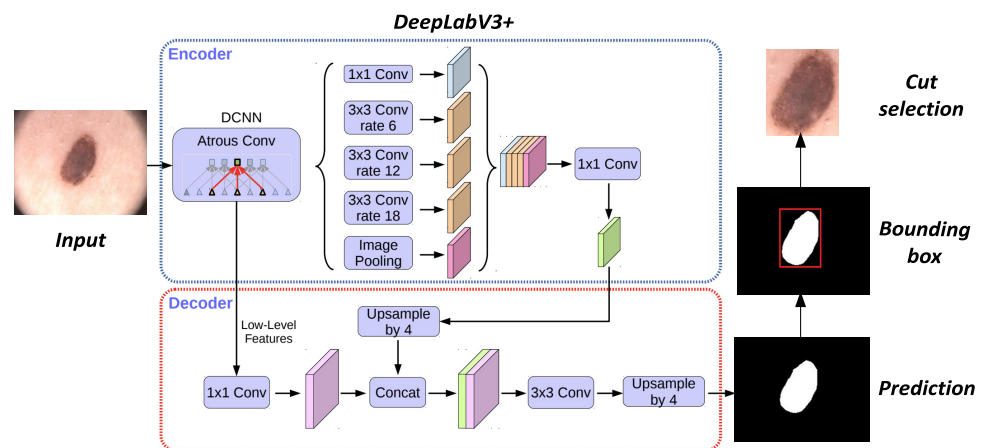

Figure 3. Definition of the region of interest. The Encoder processes the input image with atrous convolution at multiple rates. The Decoder integrates the captured features to create a segmentation mask. Then, a bounding box is defined over the lesion to isolate the region of interest cropped from the original image.

DeepLabV3+ uses Atrous Spatial Pyramid Pooling (ASPP) to capture contexts at multiple scales, employing atrous convolution with different dilation rates. The architecture adopts an encoder–decoder design, where the encoder extracts relevant features from the image, and the decoder refines these features to generate the final high-resolution segmentation. The version of the DeepLabV3+ architecture applied in this study, as defined in [32], demonstrated satisfactory test performance.

2.4. Lighting and Color Correction

In this work, lighting correction and color adjustment use three deep learning strategies based on local color distributions (LCDs), neural radiance fields (NeRFs), and deep symmetric network (DSN). Each technique was selected to tackle specific correction chal-

lenges: the LCDs apply color adjustments based on context, the NeRFs offer detailed modeling of complex scenes, and the DSN focuses on detecting and correcting symmetries. The following subsections describe each technique, exploring their fundamental principles and characteristics.

2.4.1. LCDPNet

The Local Color Distribution Prior Network (LCDPNet), developed by [33], represents a relevant lighting and color correction approach. This deep learning technique focuses on LCDs to address underexposure and overexposure issues in digital images. A distinctive feature of this technique is the use of a module called Local Color Distribution Embedded (LCDE), which extracts multiple LCDs by scales. This module aims to model correlations between different regions of the image.

The Dual-Illumination Estimation mechanism addresses learning under nonuniform lighting conditions, enhancing the identification of underlit and overlit areas. It decomposes the image into a lighting map and a reflection map. This decomposition increases pixel values, causing inefficiency in correcting overillumination, as described by [33]. As a solution, the LCDPNet implements a lighting prediction mechanism, treating overillumination as a case of underillumination but applying the logic to an inverted image.

The LCDPNet focuses on detecting local color distributions, which can show significant variations between different regions due to lighting. It seeks to identify and analyze distinctive features in small areas of images, proving effective when an image contains regions with varying lighting and color conditions. A positive aspect of this architecture is its training with the MIT AdobeFivek dataset, curated by experts, which includes both underlit and overlit images, as detailed in [33].

2.4.2. LLNeRF

The Low-Light Neural Radiance Field (LLNeRF), developed by [34], is a technique for enhancing digital images captured in low-light conditions. This technique uses a deep learning approach to model complex scenes, allowing for the reconstruction of the scene's new views, geometry, and reflection characteristics. NeRF can create realistic images using a sampling process along dimensions, applying traditional rendering methods.

Unlike other lighting correction techniques, LLNeRF applies principles of radiance to develop a lighting map, learning complex adjustments that are inaccessible through direct mapping between underlit and overlit images. LLNeRF is an unsupervised technique for light enhancement. A custom loss function guarantees the mitigation of overfitting and the assurance of color fidelity in the produced image [34].

LLNeRF is optimized during the training process to prevent degradation of image aspects. At the end of this phase, the architecture can learn to light maps, produce images with optimized lighting and color, and recover clear details, distinct contrast, and natural colors [34]. Both LCDPNet and LLNeRF utilize the same customized image set from the MIT AdobeFiveK database for training.

2.4.3. DSN

The deep symmetric network (DSN), proposed by [35], stands out in deep learning by suggesting an approach that uses recurrent learning with attention. This technique employs symmetric neural networks to correct underlighting. A distinctive feature of DSN is learning bidirectional features to avoid color distortions and efficiently recover image content, utilizing a residual attention module.

DSN introduces a recurrent residual-attention module (RRAM) designed to progressively improve focus on specific image areas through sequential steps to achieve fine adjustments in color features. This strategy allows for incremental adjustment, avoiding overloading the model with all corrections at once. The mechanism of RRAM, known as residual soft channel attention, focuses on structural information regardless of the color

channel. The DSN model used in this work was produced with the AdobeFive5k and LOL datasets.

Figure 4 demonstrates how the LCDPNet, LLNeRF, and DSN techniques apply lighting and color corrections, exemplifying improvements in lesion visualization. These representations highlight the ability of these techniques to improve visual fidelity and enhancement under various lighting conditions.
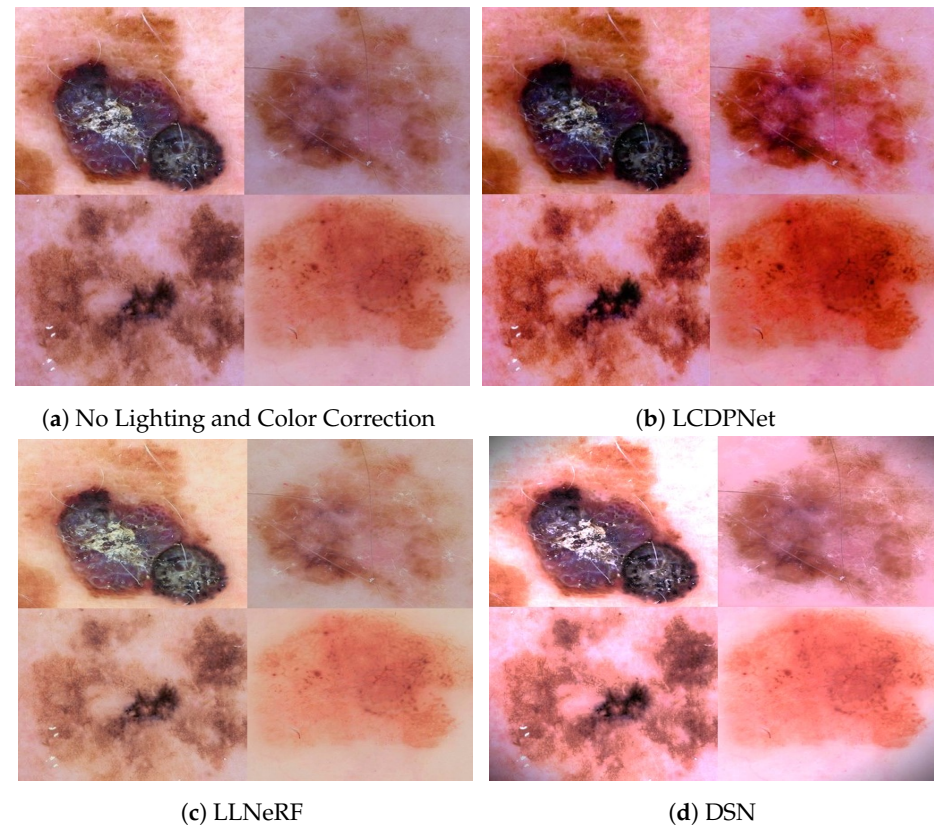

(**a**) No Lighting and Color Correction   (**b**) LCDPNet

(**c**) LLNeRF   (**d**) DSN

Figure 4. Suggested lighting and color correction methods: The subfigure (**a**) represents the uncorrected images, the subfigure (**b**) represents the images corrected with the LCDPNet method, the subfigure (**c**) shows the images corrected with the LLNeRF method, and the subfigure (**d**) represents the images corrected with the DSN method.

2.5. Classification Architectures

The efficiency and effectiveness of classification do not depend on the size of the deep learning architecture but on the individual characteristics of each [10]. We recommend using three distinct architectures: Inception-V3, Xception, ResNet-50, and DenseNet-121. These architectures were selected for their excellent results, reported in the literature.

We identify several notable differences upon comparing the InceptionV3, Xception, ResNet50, and DenseNet121 architectures. InceptionV3 enhances efficiency with modules that perform convolutions of various sizes. Xception utilizes depthwise separable convolutions to optimize efficiency. ResNet50 eases the training of deep networks with its residual connections that help mitigate gradient vanishing. Meanwhile, DenseNet121 amplifies feature reuse through a densely connected structure. Overall, all prove effective in detecting skin cancer from images, each with specific architectural innovations.

2.6. Experiments and Evaluation Metrics

The methodology developed in the proposal for this work was implemented in Python (v3.11.7) using TensorFlow (v2.15.1) and PyTorch (v2.2.1). The experiments were conducted on a computer with an Intel Core i9-11900KF processor, 32 GB of RAM, and an GIGA-BYTE NVIDIA 3090 24 GB GDDR6 GPU. It is important to highlight that lighting and

color correction techniques are applied before cropping the images and using them in the training process.

In this work, we adopted the experimental systematic of k-fold cross-validation to divide the processed images into sets. We allocated 80% of the images for training and 20% for testing, setting aside 20% of the training set specifically for validation. This process was repeated k times, ensuring each division was used exactly once. We recommend setting k = 5 for this procedure [36].

In our experiments, we adopted a binary classification approach. In this context, we categorized images of melanoma, basal cell carcinoma, and squamous cell carcinoma as malignant, while all other images were classified as benign. The imbalance between malignant and benign classes poses a significant challenge, potentially diminishing the effectiveness of classification models. We resort to the test-time augmentation (TTA) [37] technique to mitigate the imbalance between classes and achieve a more effective balance. This approach is applied to training, validation, and test sets, focusing on the class of malignant lesions.

The classification architecture training was conducted through transfer learning, followed by fine-tuning. The suggested architectures had already been pretrained with the ImageNet dataset. Furthermore, we developed a hyperparameter calibration to improve the performance of the models. This calibration includes adjustments to the learning rate, batch size, and optimization function without restrictions on these parameters.

We suggest using accuracy (*Acc*) and precision (*P*) metrics to assess the architectures' classification performance. Accuracy represents the proportion of correct predictions, covering the total number of samples and both true positives (*TP*) and true negatives (*TN*) [36]. On the other hand, precision measures the proportion of true positives among all instances classified as positive, including both true and false positives (*FP*). The following formulas describe these metrics, respectively.

$$Acc = \frac{TP + TN}{Total} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

The area under the curve (*AUC*) is also an important metric for evaluating the predictive capacity of models derived from the receiver operating characteristic (*ROC*). This metric assesses the model's ability to distinguish between the defined classes, regardless of the classification threshold. A single value represents the area under the curve, shaped by the relationship between true positives (*TP*) and false positives (*FP*). The *AUC* can range from 0 to 1, where values close to 1 indicate an effective model, while values below 0.5 suggest a random prediction model [36].

Finally, the standard deviation acts as another crucial statistical measure to evaluate the dispersion of classification results among the folds. This metric reveals how close to or far away from each other the individual results of each fold are. A standard deviation value relative to zero signals minimal dispersion among the results, indicating that the variations in performance observed across different folds are low [36].

3. Results and Discussions

Following the experimental methodology, we developed four distinct scenarios to test the performance of classification models under different conditions. These scenarios include testing the models' performance without any lighting and color correction, applying LCDPNet, using LLNeRF for corrections, and applying the DSN technique. For comparison purposes, we used the models' performance in classifying images without the intervention of any correction techniques as a reference. Table 2 displays the results achieved in all suggested scenarios.

Table 2. Classification results with all developed scenarios.

Models	*Acc* (%)	*P* (%)	*AUC* (%)
	No Lighting and Color Correction		
InceptionV3	89.3	91.4	91.8
Xception	87.8	83.6	85.6
ResNet50	89.9	90.1	82.6
DenseNet121	82.6	84.5	81.6
	LCDPNet		
InceptionV3	90.1	93.2	90.6
Xception	86.5	84.2	85.3
ResNet50	91.3	92.3	84,3
DenseNet121	81.2	83.6	84.6
	LLNeRF		
InceptionV3	91.8	90.3	89.9
Xception	89.3	86.2	88.5
ResNet50	92.9	90.6	90.3
DenseNet121	84.6	86.2	89.3
	DSN		
InceptionV3	88.2	89.1	89.6
Xception	86.8	81.6	84.3
ResNet50	93.1	91.1	83.1
DenseNet121	86.5	87.9	83.1

When analyzing Table 2, we noticed that the InceptionV3, even without adjustments for lighting and color, achieves good performance, with a precision of 91.4%. The ResNet50 and DenseNet121 perform similarly, but the ResNet50 stands out in accuracy and *AUC*. On the other hand, the Xception exhibits lower metrics under the same conditions compared to the different models.

When applying the LDCPNet correction technique, we observed a growth trend in the results for most of the analyzed models compared to the scenario without correction. This technique increases the precision of the ResNet50 and InceptionV3 models to 93.2% and 92.3%, respectively. However, we observed a decrease in the *AUC* of the ResNet50, which might indicate a loss in its ability to make accurate predictions. On the other hand, the DenseNet121 does not show significant improvements in the metrics, suggesting that the correction did not benefit the model.

Applying the LLNeRF method for correction improved metrics across all evaluated models compared to the scenario without correction. We noted an increase in the accuracy of InceptionV3 from 89.3% to 91.8%. However, precision and *AUC* saw a slight decrease. In contrast, Xception and DenseNet121 displayed significant increases in all metrics. Highlighting all results after this correction, the ResNet50 was the model that showed the most significant performance gains.

When using the DSN correction technique, we observed varied results. InceptionV3 decreased accuracy, from 91.4% to 89.1%, and in *AUC*, from 91.8% to 89.6%. Xception also experienced a significant reduction in accuracy, dropping from 85.6% to 81.6%. Conversely, ResNet50 and DenseNet121 improved metrics with increased accuracy rates and *AUC*. Notably, the accuracy of ResNet50 increased from 89.9% to 93.1%.

To make the variations between the analyzed models without lighting correction and the evaluated correction techniques clearer, Table 3 presents the specific difference values and weighted standard deviation. This approach allows for a quantitative comparison of the impact of each correction technique.

Table 3. Differences between noncorrection and application of lighting and color correction methods.

Models	*P*(%)		
	LCDPNet	**LLNeRF**	**DSN**
InceptionV3	1.8 ± 0.046	-1.1 ± 0.046	-2.3 ± 0.040
Xception	0.6 ± 0.094	2.6 ± 0.018	-2.0 ± 0.071
ResNet50	2.2 ± 0.050	0.5 ± 0.027	1.0 ± 0.036
DenseNet121	-0.9 ± 0.051	1.7 ± 0.048	3.4 ± 0.077

The lighting and color correction techniques LCDPNet, LLNeRF, and DSN enhanced the effectiveness of the models in the majority of the experiments. We observed an increase in precision rates in detecting cancer, achieving an average gain of approximately 3.4%. However, it is important to highlight that each model exhibited distinct behavior when dealing with these specific characteristics.

Comparison of Results

We compared similar approaches to highlight the influence of lighting and color correction on deep learning classification models with contextual and semantic information. It is important to point out that while these studies did not follow the same steps as this study, they share the use of lighting and color correction techniques and classification via deep neural networks. Table 4 presents the main results, facilitating the comparison with the methodology of this work.

Table 4. Performance comparison with similar works.

Works	Acc (%)	P (%)	Comments
Galdran et al. [5]	72.3	69	Result achieved with ResNet50. Data augmentation focusing on dual estimation of lighting.
Nahata et al. [38]	85	86	Better results with ResNet50. Correction using statistical methods, focusing on global lighting adjustments.
Salvi et al. [16]	79.2	-	Results achieved with DenseNet121. It uses GAN for lighting and color correction, highlighting the potential of this technique in segmentation and classification. Precision data not specified.
Proposed approach	91.3	92.3	Results were determined by ResNet50, using the LCDPNet correction technique.
	86.5	87.9	Better results with DenseNet121. Application of the DSN correction technique.

While the primary goal of the evaluated studies is to improve classification accuracy through lighting and color correction, comparisons among the studies described in the table face limitations. These limitations stem from differences in the image datasets used in the experiments, the correction methods applied, the classification architectures, and the evaluation metrics. The observed results can be attributed not only to the correction techniques but also to the intrinsic characteristics of the solution.

Table 4 highlights the potential of the approach we developed in this study. The metrics, especially precision, demonstrate our superiority in classification rates. Despite the limitations in comparisons, we observe that statistical corrections for lighting and color may not lead to good results in classification with deep neural networks. The lighting and color correction techniques we applied in this work effectively increase precision in classifying malignant lesions, thus enhancing the performance of deep neural networks in this application.

4. Conclusions and Future Work

Deep learning achieves good results in skin cancer classification, but this methodology also attracts interest in solving problems related to image acquisition. Techniques developed

for this purpose can automatically adjust lighting and color in digital images. Observing and adjusting these characteristics can significantly improve classification results in the automatic detection of skin cancer.

This study emphasizes the importance of lighting and color in classifying malignant skin cancer using deep neural network architectures. We demonstrated that applying advanced correction techniques, such as LCDPNet, LLNeRF, and DSN, can increase the precision and accuracy of these networks, achieving improvements of between 3 and 4% in precision. The models that used images corrected with the LCDPNet and LLNeRF techniques showed improvements in most classification metrics. Although these techniques are not specific to images of skin lesions, they enhanced the precision in diagnosing malignant skin cancer.

We also noted that the architectures used for classification achieved different results by applying lighting and color correction techniques. The Xception and DenseNet121 models proved more robust, while InceptionV3 and ResNet50 were more sensitive to these changes. One explanation is that color characteristics can change after convolution operations, causing the architecture to focus more on structural information and reducing the emphasis on colors. Factors such as the size of the kernels, the number of layers, and pooling operations can increase or decrease the importance of color in optimizing a loss function. Proper preprocessing of lighting and color aspects and a fine-tuned classification model can improve the identification of malignant skin cancer.

For future work, we recommend using other lighting and color correction methods, such as methods based on self-supervised learning, attention mechanisms, and style transfer, to test the sensitivity of other, more modern deep architectures. Another suggestion is to check and compare this sensitivity with other conventional techniques, such as histograms, white balance, color temperature, and chromatics. New methodologies of this type would provide valuable information about the capabilities of these lighting and color correction techniques in skin cancer images. Furthermore, analysis of advanced correction methods for these features would drive the customization of deep architectures to deal with dynamic variations, improving learning capabilities and accuracy rates in cancer detection.

Author Contributions: K.R.T.A. and R.M.S.V. contributed to the guidance and supervision of the research. A.R.F.S. and K.R.T.A. developed the methodology. A.R.F.S. implemented the methodology and conducted the experiments. A.R.F.S. wrote the manuscript. K.R.T.A. and R.M.S.V. validated the paper. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Cortez, J.L.; Vasquez, J.; Wei, M.L. The impact of demographics, socioeconomics, and health care access on melanoma outcomes. *J. Am. Acad. Dermatol.* **2021**, *84*, 1677–1683. [CrossRef]
3. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **2021**, *149*, 778–789. [CrossRef]
4. Barata, C.; Marques, J.S.; Celebi, M.E. Improving dermoscopy image analysis using color constancy. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 3527–3531.
5. Galdran, A.; Alvarez-Gila, A.; Meyer, M.I.; Saratxaga, C.; Araújo, T.; Garrote, E.; Aresta, G.; Costa, P.; Mendonça, A.M.; Campilho, A. Data-driven color augmentation techniques for deep skin image analysis. *arXiv* **2017**, arXiv:1703.03702.

6. Goyal, M.; Knackstedt, T.; Yan, S.; Hassanpour, S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Comput. Biol. Med.* **2020**, *127*, 104065. [CrossRef] [PubMed]
7. Jones, O.; Matin, R.; Schaar, M.V.; Bhayankaram, K.; Ranmuthu, C.; Islam, M.; Behiyat, D.; Boscott, R.; Calanzani, N.; Emery, J.; et al. Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: A systematic review. *Lancet Digit. Health* **2022**, *4*, 466–476. [CrossRef] [PubMed]
8. Esteva, A.; Kuprel, B.; Novoa, R.; Ko, J.; Swetter, S.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]
9. Esteva, A.; Chou, K.; Yeung, S.; Naik, N.; Madani, A.; Mottaghi, A.; Liu, Y.; Topol, E.; Dean, J.; Socher, R. Deep learning-enabled medical computer vision. *NPJ Digit. Med.* **2021**, *4*, 5. [CrossRef]
10. Adegun, A.; Viriri, S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artif. Intell. Rev.* **2021**, *54*, 811–841. [CrossRef]
11. Taghanaki, S.A.; Abhishek, K.; Cohen, J.P.; Cohen-Adad, J.; Hamarneh, G. Deep semantic segmentation of natural and medical images: A review. *Artif. Intell. Rev.* **2021**, *54*, 137–178. [CrossRef]
12. Dildar, M.; Akram, S.; Irfan, M.; Khan, H.; Ramzan, M.; Mahmood, A.; Alsaiari, S.; Saeed, A.; Alraddadi, A.; Mahnashi, M. Skin cancer detection: A review using deep learning techniques. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5479. [CrossRef] [PubMed]
13. Branciforti, F.; Meiburger, K.M.; Zavattaro, E.; Veronese, F.; Tarantino, V.; Mazzoletti, V.; Cristo, N.D.; Savoia, P.; Salvi, M. Impact of artificial intelligence-based color constancy on dermoscopical assessment of skin lesions: A comparative study. *Ski. Res. Technol.* **2023**, *29*, e13508. [CrossRef] [PubMed]
14. Barata, C.; Ruela, M.; Francisco, M.; Mendonça, T.; Marques, J.S. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *Syst. J.* **2013**, *8*, 965–979. [CrossRef]
15. Barata, C.; Celebi, M.E.; Marques, J.S. A survey of feature extraction in dermoscopy image analysis of skin cancer. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1096–1109. [CrossRef] [PubMed]
16. Salvi, M.; Branciforti, F.; Veronese, F.; Zavattaro, E.; Tarantino, V.; Savoia, P.; Meiburger, K.M. DermoCC-GAN: A new approach for standardizing dermatological images using generative adversarial networks. *Comput. Methods Programs Biomed.* **2022**, *225*, 107040. [CrossRef] [PubMed]
17. Ng, J.H.; Goyal, M.; Hewitt, B.; Yap, M.H. The effect of color constancy algorithms on semantic segmentation of skin lesions. *Med. Imaging 2019 Biomed. Appl. Mol. Struct. Funct. Imaging* **2019**, *10953*, 138–145.
18. Mahbod, A.; Schaefer, G.; Wang, C.; Dorffner, G.; Ecker, R.; Ellinger, I. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Comput. Methods Programs Biomed.* **2020**, *193*, 105475. [CrossRef] [PubMed]
19. Buchsbaum, G.A. A spatial processor model for object colour perception. *J. Frankl. Inst.* **1980**, *310*, 1–26. [CrossRef]
20. Finlayson, G.D.; Trezzi, E. Shades of gray and colour constancy. *Color Imaging Conf.* **2004**, *2004*, 37–41. [CrossRef]
21. Funt, B.; Shi, L. The effect of exposure on MaxRGB color constancy. *Hum. Vis. Electron. Imaging* **2010**, *7527*, 282–288.
22. Venugopal, V.; Nath, M.K.; Joseph, J.; Das, M.V. A deep learning-based illumination transform for devignetting photographs of dermatological lesions. *Image Vis. Comput.* **2024**, *142*, 104909. [CrossRef]
23. Salvi, M.; Branciforti, F.; Molinari, F.; Meiburger, K.M. Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm. *Expert Syst. Appl.* **2024**, *245*, 123105. [CrossRef]
24. Tosta, T.A.A.; Freitas, A.D.; de Faria, P.R.; Neves, L.A.; Martins, A.S.; Nascimento, M.Z.D. A stain color normalization with robust dictionary learning for breast cancer histological images processing. *Biomed. Signal Process. Control.* **2023**, *85*, 104978. [CrossRef]
25. Zhou, J.; Wu, Z.; Jiang, Z.; Huang, K.; Guo, K.; Zhao, S. Background selection schema on deep learning-based classification of dermatological disease. *Comput. Biol. Med.* **2022**, *149*, 105966. [CrossRef] [PubMed]
26. Cong, C.; Liu, S.; Di Ieva, A.; Pagnucco, M.; Berkovsky, S.; Song, Y. Colour adaptive generative networks for stain normalisation of histopathology images. *Med. Image Anal.* **2022**, *82*, 102580. [CrossRef] [PubMed]
27. International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset. International Skin Imaging Collaboration. 2020. Licence: Creative Commons Attribution Non Commercial 4.0 International. Available online: https://challenge2020.isic-archive.com/ (accessed on 12 June 2023).
28. Pacheco, A.; Lima, G.; Salomao, A.; Krohling, B.; Biral, I.; Angelo, G.; Alves, F., Jr.; Esgario, J.; Simora, A.; Castro, P.; et al. PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data Brief* **2020**, *32*, 106221. [CrossRef] [PubMed]
29. Daneshjou, R.; Vodrahalli, K.; Novoa, R.A.; Jenkins, M.; Liang, W.; Rotemberg, V.; Ko, J.; Swetter, S.M.; Bailey, E.E.; Gevaert, O.; et al.. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci. Adv.* **2022**, *8*, eabq6147. [CrossRef] [PubMed]
30. Mendonça, T.; Ferreira, P.M.; Marques, J.S.; Marçal, A.R.S.; Rozeira, E.J. PH 2—A dermoscopic image database for research and benchmarking. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5437–5440.
31. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, E.H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.

32. da Silva, J.V.M.; Aires, K.R.T.; Santos, A.R.F.D.; Veras, R.d.S.; Neto, L.D.S.B.; Sousa, L.P.D.; Filho, E.F.D.C.I. Segmentação Semântica do Câncer de Pele Utilizando Aprendizado Profundo. In Proceedings of the Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), Sao Paulo, SP, Brazil, 27–30 June 2023; pp. 304–315.

33. Wang, H.; Xu, K.; Lau, E.R.W.H. Local color distributions prior for image enhancement. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 343–359.

34. Wang, H.; Xu, X.; Xu, K.; Lau, R. Lighting up NeRF via Unsupervised Decomposition and Enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 12632–12641.

35. Zhao, L.; Lu, S.-P.; Chen, T.; Yang, Z.; Shamir, A. Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 1–17 October 2021; pp. 12075–12084.

36. Rainio, O.; Teuho, J.; Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **2024**, *14*, 6086. [CrossRef]

37. Kimura, M. Understanding test-time augmentation. In Proceedings of the International Conference on Neural Information Processing, Sanur, Bali, Indonesia, 8–12 December 2021; pp. 558–569.

38. Nahata, H.; Singh, S.P. Deep learning solutions for skin cancer detection and diagnosis. *Mach. Learn. Health Care Perspect. Mach. Learn. Healthc.* **2020**, *13*, 159–182.