Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 10
October 2023

Early Prediction of Melanoma Using Image Processing and Conventional Neural Network

Fiyad Alebrahaimi
Department of Computer Science, Faculty of Computers & Information Technology, University of Tabuk, Saudi Arabia.
Saud Albalawi
Department of Computer Science, Faculty of Computers & Information Technology, University of Tabuk, Saudi Arabia.
Fahad Aldawsari
Department of Computer Science, Faculty of Computers & Information Technology, University of Tabuk, Saudi Arabia.
Omar Almuyaba
Department of Computer Science, Faculty of Computers & Information Technology, University of Tabuk, Saudi Arabia.
Zaid bassfar
Department of Computer Science, Faculty of Computers & Information Technology, University of Tabuk, Saudi Arabia.
Aadel Alatwi
Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Saudi Arabia
Tareq Alhmiedat
Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Saudi Arabia
Mohammed Alotaibi
Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Saudi Arabia

***Corresponding Author:** Mohammed Alotaibi
*Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Saudi Arabia

Abstract
Skin cancer is quite a common type of cancer. Its incidence is higher in Caucasians, and melanoma is the most lethal type. To increase patient prognosis, developing tools to assist in early diagnosis is necessary. In this work, we develop an easy and accessible mobile app to assist melanoma detection. The app is linked to a classification model for classifying the images as melanoma or non-melanoma, motivated by the performance of convolutional neural networks in computer vision trained on images collected from smartphones and clinical lesion information. The significance of a melanoma detection application is in its capacity to facilitate early detection, which is done by collecting a comprehensive and representative dataset of skin lesion photos. These photos encompass a wide range of diversity and balance, encompassing both melanoma and non-melanoma instances. Moreover, it provides annotations on the photos by adding labels that indicate whether each lesion is classified as melanoma or non-melanoma. Since the occurrence of melanoma is much smaller than other skin lesions, most of the datasets for this problem are imbalanced which has been demonstrated by the findings of this study. However, achieving data set balance is a crucial step in ensuring optimal performance of the detection model for both melanoma and non-melanoma instances. Nevertheless, it is of utmost importance to uphold an accurate portrayal of the frequency of melanoma in practical situations and to further enhance the model's efficacy by engaging in partnerships with dermatologists and persistently gathering data. In order to come up with a balanced dataset, an approach has been presented that is based on evolutionary algorithm and it entails enhancing the data distribution. The findings suggested that evolutionary algorithms have the potential to aid in the selection of the most pertinent features from photos, hence enhancing the accuracy of the detection model.

Keywords: Malignant melanoma, image classification, computer vision, CNN, skin lesion analysis

1. Introduction

Skin cancer develops when abnormal skin cells grow out of control, causing the skin to quickly proliferate and develop cancerous tumors. The number of new aggressive melanoma cases significantly increased by 54% between 2008 and 2019 [1]. This highlights the urgent need for improved diagnostic tools to aid in the early detection of melanoma. The cells (melanocytes) that make melanin, the pigment responsible for the skin's color, grow into melanoma, the most dangerous type of skin cancer. The actual reason all melanomas occur is unknown, although exposure to ultraviolet (UV) radiation from sunshine, tanning beds, or tanning lamps increases the risk of getting the disease. One can lower their chances of developing melanoma by limiting their exposure to UV light [2]. People under 40, particularly women, appear to be at increased risk for melanoma. It is possible to prevent the spread of skin cancer by being aware of the warning signals of the condition.

Early detection is vital to the successful treatment of melanoma [2]. Deep learning algorithms, particularly in computer vision, have demonstrated promising results in identifying melanoma from medical images [22].

Computer-aided diagnosis (CAD) systems have shown promise in helping to diagnose melanoma, but there are still certain difficulties that computer-aided diagnosis (CAD) systems face today despite the significant advancements that CAD has made since the invention of computers [3]. The procedures of a CAD system, including input data collecting, pre-processing, processing, and system evaluations, present certain obstacles due to various algorithmic limits. Since algorithms are typically created to choose a single likely diagnosis, people with several concurrent conditions sometimes receive less than ideal outcomes [4]. Today, electronic health records (EHRs) are the primary source of input data for CAD. Any CAD system must include effective design, implementation, and analysis for EHRs. Additionally, there are no established evaluation criteria for CAD systems. The FDA's permission for commercial usage may be tough to obtain because of this reality. Furthermore, research for evaluating their algorithms for clinical application has scarcely been confirmed, even though many improvements in CAD systems have been demonstrated [3, 5]. The use of dermoscopic images is constrained by developing this method. This means that unless the device has a particular dermoscopic link to it, this method cannot be used, for example, in smartphone apps. In this situation, the models must be expanded to incorporate clinical images. However, in this instance, a significant public archive, such as the International Skin Imaging Collaboration (ISIC), is not accessible [6].

The issue of healthcare professionals implementing modern CAD systems in clinical practice is related to other difficulties. For instance, the use of CAD may be discouraged by several unfavorable researches. Additionally, improper system outcomes are brought about by health workers' lack of training in the use of CAD. Several challenges were identified in diagnosing melanoma. These include: (i) Melanoma diagnosis often relies on visual inspection by dermatologists. However, this method can be subjective and prone to interobserver variability, meaning different doctors may reach different conclusions when examining the same lesion [7,8]; (ii) the visual inspection approach can lead to both false positives (identifying non-cancerous lesions as melanoma) and false negatives (missing melanomas that are present) [9]; (iii) over diagnosis occurs when benign lesions are unnecessarily treated as melanoma, leading to unnecessary stress and medical interventions for patients [10]; (iv) in many regions, access to dermatologists, who are experts in melanoma diagnosis, can be limited, leading to delayed diagnosis and treatment [11]; (v) melanoma is a highly heterogeneous disease, and not all lesions follow classic visual patterns. Some melanomas may present atypically and can be challenging to diagnose visually [12]; (vi) while there have been advancements in using technologies like dermoscopy and CAD tools, these methods may not be widely available or accessible in all healthcare settings; (vii) incomplete or inaccurate patient information can hinder the diagnostic process [13]; comprehensive diagnostic methods, such as mole mapping and genetic testing, can be expensive, limiting their accessibility to all patients [14].

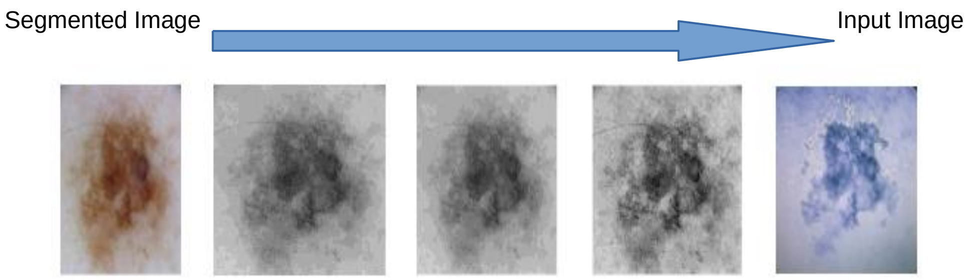
To address these challenges, the proposed approach is to design a mobile app for both IOS and Android devices containing a CAD system for diagnosing skin cancer and specialty melanoma, determining the location of a lesion, and estimating illness likelihood. The CAD system receives digital images as input. To improve the accuracy of the diagnosis, the first phase involves image preparation, which allows for the reduction of ill effects and different artifacts, such as hair, that may be present in images. The lesion is then detected using an image segmentation technique. Once the lesion has been identified, several chromatic and morphological characteristics can be quantified and used for classification. By providing a reliable and accessible tool for the diagnosis of melanoma, this mobile app has the potential to improve the early detection of skin cancer and ultimately save lives.

2. Related Work

Melanoma is one of the deadliest types of skin cancer because it spreads fast throughout the body. With the increasing prevalence and lethality of melanoma, it is critical to develop CAD support systems to assist physicians in diagnosing skin cancer. Many studies have been conducted over the past two decades on the rapid and reliable identification of melanoma using dermoscopic images, with diagnostic accuracy ranging from 70% to 95% [15]. Thus, it is crucial that methods to help medical professionals identify skin cancer in its early stages in light of this significant uptick are launched and advanced. Various deep learning and machine learning techniques have been employed in the diagnosis of melanoma, encompassing convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks [1-3]. The utilization of extensive training datasets has been shown to enhance performance; nevertheless, it is important to acknowledge that there are also associated limitations [29]. For example, many models depend on features that are manually constructed, necessitate a substantial amount of training data, and possess demanding computational needs.

The objective of our strategy, which is based on the Xception architecture, is to enhance accuracy while simultaneously reducing computational complexity in comparison to previous research endeavors. The work presented in this study makes a noteworthy contribution by demonstrating the model's ability to balance memory and precision. However, it is important to consider the potential impact of false negative rates. Our objective is to elucidate the distinctive advantages of our methodology and its potential impact on the field of melanoma diagnosis by a comparative analysis with existing methodologies. Adopting an ensemble of deep models rather than a single technique is another development in this area. The major objective of this approach is to increase the accuracy and dependability of the forecasts. For example, malignant melanomas, the deadliest form of skin cancer, were found using an ensemble of several deep models, including deep residual networks (DRNs) and convolutional neural networks (CNNs) [16].

By leveraging very deep CNNs, a proposed two-stage framework involves constructing a very deep, fully convolutional residual network (FCRN) to segment skin lesions and employing a very deep residual network to precisely distinguish melanomas from non-melanoma lesions. The method in [9] uses low-level hand-crafted features or shallower CNN architectures with substantially deeper networks (more than 50 layers) that acquire richer and more discriminative features for more accurate recognition [9]. The very deep residual network was employed in a two-stage framework. In the first stage, a highly deep, FCRN was used to accurately segment the skin lesions by incorporating multi-scale feature representations, as shown in Figure 1. The segmentation results were then passed on to the second stage, where a very deep residual network was used to distinguish between melanomas and non-melanoma lesions based on the segmentation results. The network was trained to recognize the features that differentiate melanomas from non-melanoma lesions, using the segmentation results as input [17].



Segmentation process using FCRN technique [17].

The use of a very deep residual network for melanoma recognition provides several advantages, including the extraction of rich and discriminative features, the ability to overcome the problem of vanishing gradients, and the capturing of complex relationships between features. These advantages are important for the accurate recognition of melanoma in dermoscopic images. The two-stage framework has demonstrated superior performance compared to other systems [17].

The FCRN approach significantly improves the accuracy of skin lesion segmentation compared to other methods and can be trained with a relatively small amount of data. Additionally, the proposed two-stage strategy with deep CNNs has the potential to revolutionize automated melanoma identification and improve early detection rates [9]. A DRN comprises a set of residual blocks, each consisting of a few stacked layers (e.g., convolutional layers, rectified linear unit layers, and batch normalization layers). Given the l-th e residual block B_l, they denote the input and output of B_l asH_(l-1) andH_l, respectively, and employ H_l (x) to represent the underlying mapping of these stacked layers traditionally [17].

A combination of skin detection and rapid hierarchical segmentation was proposed to identify a skin lesion in a computationally efficient manner. Two simple segmentation algorithms were combined to produce a rough representation of the lesion. Fine segmentation was applied using the coarse segmentation result as an input to create the lesion's outline. Four feature categories were chosen from the final segmented region to precisely define the lesion's color, shape, border, and texture. A classifier was constructed for each feature category, and their output was merged to produce the final results [10]. Rapid hierarchical segmentation and skin detection were combined by the researchers in [18] to find the skin lesion, which may have been computationally less expensive than sophisticated segmentation techniques. The skin image was down sampled first, and then the system combined two simple segmentation approaches: Otsu's method and the minimum spanning tree (MST) method were used to obtain the first segmentation results with the down sampled version to produce a rough representation of the lesion.

The lesion's outline is then produced by applying a fine segmentation using the coarse segmentation result as the input, choosing four feature categories that precisely describe the color, shape, border, and texture of the lesion from the final segmented region. A classifier is built for each feature category to categorize the skin lesion, and the results are combined to produce the final results [18]. At first, they constructed a "skin detection" model utilizing a technique based on a skin color model to identify the skin pixels, delivering 32 skin and non-skin color maps of size 64 x 64 x 64 for an RGB color picture with no pre-processing. They then used criteria to integrate the findings of different segmentation algorithms with modest computing needs. After obtaining the skin region area, the image was down sampled, and rules were applied to integrate the findings of both approaches—Otsu's method and the MST method [18].

However, in [19], an automated classification method for a cutaneous lesion in digital dermatoscopic images was used to detect the presence of melanoma. This method comprises two fundamental stages. In Stage 1, a bounding box is cropped around only the skin lesion in the input image using Mask R-CNN. Mask R-CNN is a deep learning technique that combines object detection and instance segmentation to identify and locate objects in an image. This stage is essential for accurately isolating the skin lesion from the rest of the image, which can be challenging in dermatoscopic images. In Stage 2, the cropped bounding box is classified using ResNet152, a deep neural network architecture that has been shown to in image classification tasks. ResNet152 is trained on a large dataset of skin lesion images and can accurately classify skin lesions as benign or malignant, including identifying the presence of melanoma [19].

A deep learning framework was proposed to simultaneously perform segmentation and coarse classification of skin lesions. The framework consists of two FCRNs and a lesion index calculation unit for refining the classification results. A simple CNN was also proposed for dermoscopic feature extraction. The proposed frameworks were evaluated on the ISIC 2017 dataset, and promising accuracies were achieved [20].

To pre-process the original training set of 2000 skin lesion images with varying resolutions, data augmentation was performed, and the images were rotated to establish a class-balanced dataset. The FCRN-88 was used for lesion segmentation and classification, and the lesion feature network was developed for dermoscopic feature extraction. Superpixel extraction was performed to locate the positions of dermoscopic features, and data augmentation was used to balance the number of images of different categories in the extracted patch dataset [20].

The system used a multi-stage and multi-scale approach and a SoftMax classifier for pixel-wise classification of melanoma lesions. A new method called lesion classifier was introduced to classify skin lesions into melanoma and non-melanoma based on the results of pixel-wise classification. Experiments on two public benchmark skin lesion datasets, ISBI 2017 and PH2, demonstrated that the proposed method outperformed some state-of-the-art methods, achieving an accuracy and dice coefficient of 95% and 92%, respectively, on the ISIC 2017 dataset and an accuracy and dice coefficient of 95% and 93%, respectively, on the PH2 dataset [21].

The proposed deep convolutional encoder-decoder architecture is interconnected through a series of skip pathways to enhance feature learning ability and feature extraction. The network is designed to handle various sizes of skin lesion images and is moderate in size. The multi-stage and multi-scale approaches are used to enhance the learning of complex features. The lesion classifier is a computationally efficient predictive method that distinguishes melanoma lesions from non-melanoma images using the output of the SoftMax modules. It is particularly effective for analyzing challenging skin lesions with fuzzy boundaries and heterogeneous textures for melanoma detection [21].

3. Materials and Methods

Figure 1. Shows that the methodology for melanoma detection begins with collecting a dataset of images containing the objects to be detected. The images are then pre-processed to normalize them and ensure consistent sizes, which prepares them for the model training. An object detection algorithm is then selected based on the type of objects and the complexity of the detection task. A CNN model is then constructed and trained using the pre-processed images and their object labels as ground truth. Various hyper para meters of the model, such as learning rate and number of epochs, are tuned to optimize the model's performance. The trained model is then evaluated on a test set of held-out images to calculate metrics, such as accuracy and loss, to assess how well it detects the objects. Based on the evaluation results, the model is improved by retraining with additional data or changing the architecture. Once the model achieves satisfactory detection results, it is deployed for detecting objects in new unseen images in the real world.

In developing a melanoma detection mobile app utilizing computer vision, several specific considerations must be considered. One such consideration is model optimization, as mobile devices typically have limited computational resources compared to desktop computers or servers. Therefore, optimizing the trained model for deployment on mobile devices is crucial and can involve techniques such as model compression, quantization, or pruning to reduce the model size and computational complexity. Another important consideration is app design and development. This involves designing and developing the app interface, integrating the optimized model with the app code, and testing the app on various mobile devices to ensure compatibility and performance. Finally, user privacy and security are also crucial considerations when dealing with sensitive medical data, such as images of skin lesions. Techniques such as data encryption, secure communication protocols, and user authentication and authorization can be utilized to ensure user privacy and security.

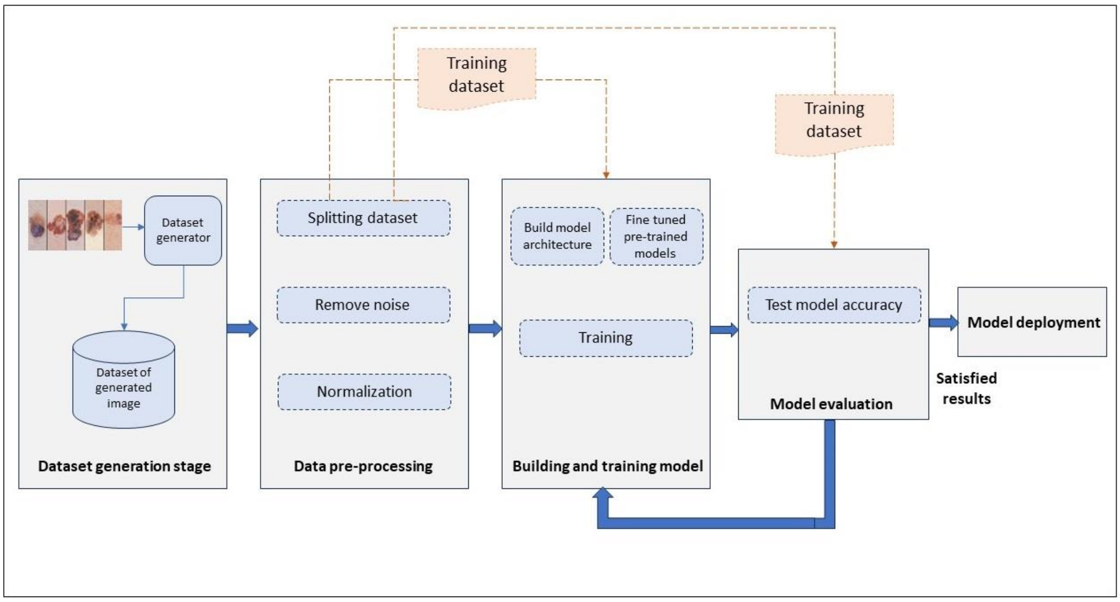


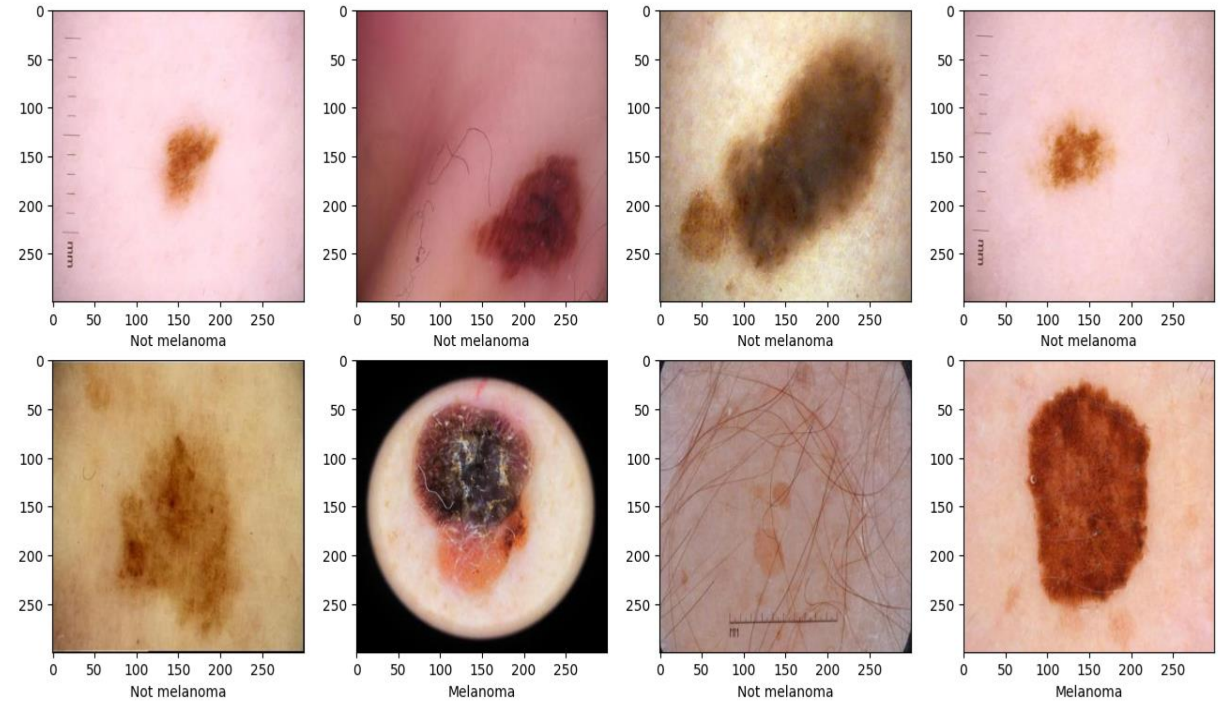
Figure 1. Melanoma detection mobile app methodology

3.1. Dataset

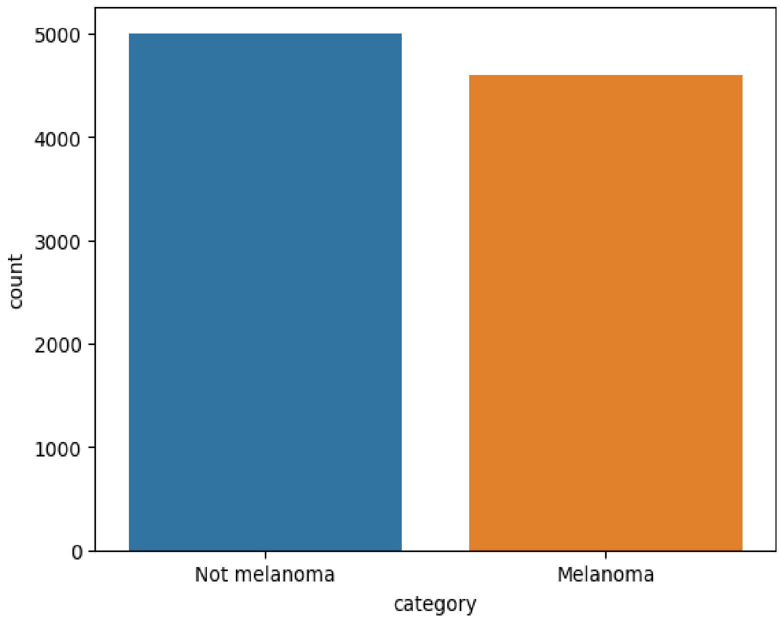
The efficacy of melanoma analysis using neural networks is contingent upon the appropriate selection of input data. The dataset was selected with great care from reference [22] because to its meticulously organized structure, equitable representation, and extensive collection of skin lesion photographs accompanied by relevant metadata. The primary objective is to leverage the accurate diagnostic capabilities of neural networks for melanoma detection. While acknowledging the inherent limitations of individual datasets and models, we posit that the amalgamation employed in this study offers a robust foundation for future enhancements. However, we are always seeking additional datasets and model architectures that have the potential to enhance the accuracy of predictions. Our choice underscores the promise of this area of investigation, and we hold a positive outlook that advancements in the discipline will further enhance the significance and efficacy of our approach. This dataset contains 10,015 images of skin lesions, including 4,605 images that are confirmed melanoma cases and 5,000 images that are non-melanoma cases. The images were collected from various sources, including the HAM10000 and ISIC datasets. The HAM10000 dataset is a large collection of multi-source of pigmented sckin lesions images. These images were collected over a span of 20 years from the Dermatology Department at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia [39]. Conversely, the ISIC dataset contains of a series of sub-datasets, which were collected over a 5-year period, with data collection commencing in 2016 [40-44].

The images are in JPG format and have a resolution of 224 x 224 pixels. The dataset also includes additional metadata for each image, including the age and sex of the patient, the location of the lesion on the body, and the diagnosis of the lesion, which can be used to train the neural network The dataset is well-organized and easy to use. The images are divided into two folders, one for melanoma images and one for non-melanoma images, and the dataset is divided into train and test folders. The train folder contains 9,605 images, and the test folder contains 1,000 images. Each image is named with the patient's ID and the type of lesion, making it easy to keep track of the images during training and testing. The dataset also includes a CSV file that contains the patient's demographics and clinical information, which can be used for further analysis. The images in the dataset are clear and well-labeled. The images were taken by various cameras, and the lighting conditions. This makes the dataset more challenging to train a model on, but it also makes the model more robust to variations in lighting conditions [22].

As shown in Figure 2, the images dataset was found to be balanced, with an equal distribution of both melanoma and benign skin lesion images. A balanced dataset allows the model to learn the features of both classes equally, resulting in a more accurate and robust model. With a balanced dataset, the model is exposed to an equal number of positive and negative cases during training, which improves its ability to correctly classify new images. Additionally, a balanced dataset can help reduce false positives and false negatives, which are both critical in diagnosing melanoma [22].



(a)



(b)

Figure 2. (a) Data sample (b) Dataset classes balancing

3.1. Data Pre-Processing

3.1.1. Data Augmentation

Image augmentation, as evolutionary algorithm techniques, offered a few further benefits in this work. Images from the two different classes can occasionally be slightly different from one another. In these situations, taking several different perspectives of the same image will be helpful. Various augmentation parameters can be used to perform various transformations on the images. These transformations include random rotations, shearing, zooming, horizontal flipping, and shifting. During model training, generating batches of augmented image data is a crucial step that can improve the model's performance and robustness. To achieve this, the Image Data Generator class is utilized, which takes various parameters to perform data augmentation on the images.

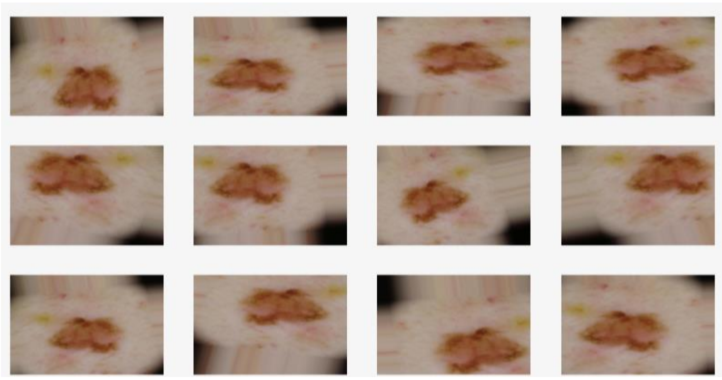


Figure 3. Augmented image

3.1.2. Data Normalization

Image data normalization is a process of adjusting the values of the pixels in an image so that they have a mean of 0 and a standard deviation of 1. This helps to ensure that the model does not learn to rely on any particular features of the images, such as brightness or contrast. It can improve the training process and help the model converge faster. It also makes the model more robust to changes in lighting conditions and color variations in the input images. The most common normalization technique is to scale the pixel values to the range [0, 1] so each pixel value is divided by the maximum pixel value that can be represented in the image's color space.

3.2. Model Building

In this study, various modeling methods were utilized to develop accurate models for computer vision tasks. The methods used included transfer learning with pre-trained models, freezing layers of the pre-trained models, and adding custom layers on top of the pre-trained models. There are two options for creating a classification model: One is to build the model from scratch (layer by layer), and the other is to transfer learning using pre-trained networks, as shown in Figure 4.

Several models were developed for the custom classification task, utilizing pre-trained models as the base models and adding custom layers on top. The base model layers were frozen, and only the top layers were trained. One of the models used the EfficientNetB3 pre-trained model as the base model and added a few dense layers on top. The base model layers were frozen except for the last bottleneck layer. The model was compiled and trained for 24 epochs. Another model used the InceptionV3 pre-trained model as the base model and added the GlobalAveragePooling2D and Dense layers on top. The base model layers were frozen, and only the top layers were trained. The model was also compiled and trained for 24 epochs.

The ResNet50 pre-trained model, which has ability to handle deep networks effectively and alleviate the vanishing gradient problem [31], was used as the base model for another model, with Flattening, Dense, and Dropout layers added on top for fine-tuning. The base model layers were frozen, and the top layers were trained. The model was compiled and trained for 24 epochs. Similarly, the Xception pre-trained model, which is known for its exceptional feature extraction ability [31], was used as the base model for transfer learning. Flattening, Dense, and Dropout layers were added on top, and the base model layers were frozen. The model was also compiled and trained for 24 epochs.

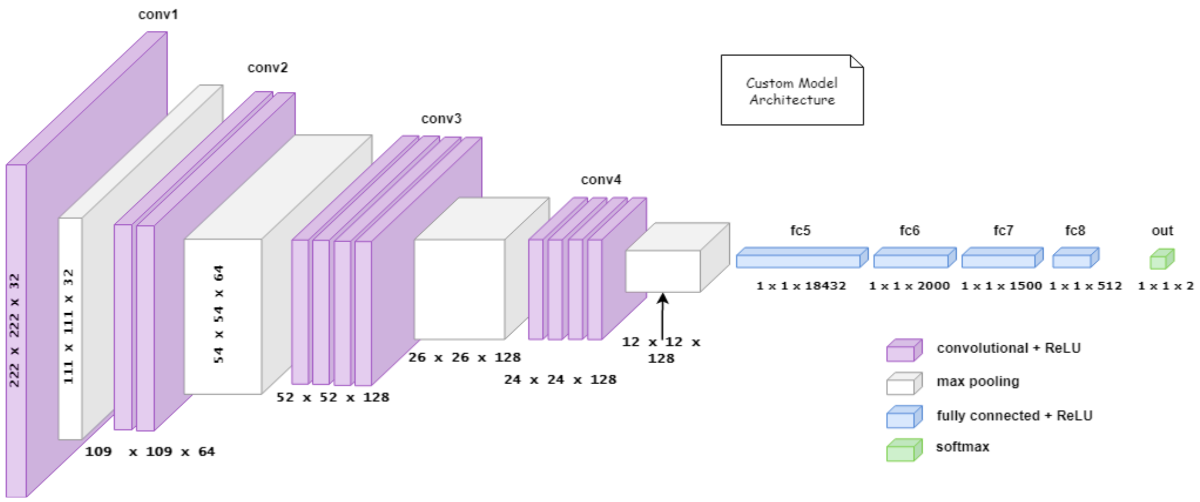


Figure 4. Custom model architecture

3.3. Model Training

The training model was used for machine language model training. The machine language model training process involves feeding machine learning algorithms with data to help the algorithms identify and learn good values for all attributes involved. The training model runs the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is then used to modify the model. To train a model, the training dataset, which is a sample of the whole dataset used in the learning stage of the model, must be used.

Three callbacks were used during the training process to monitor the model's performance and make adjustments accordingly. The first callback, Reduce LROn Plateau, was used to reduce the learning rate when the validation accuracy plateaued. This helps the model converge more effectively and can prevent overfitting. The callback monitored the validation accuracy and reduced the learning rate by a factor of 0.2 if the validation accuracy did not improve for five epochs. The minimum learning rate was set to 0.001. The second callback, Early Stopping, was used to stop the training process early if the validation accuracy did not improve by a certain amount over a certain number of epochs. This helps to prevent overfitting and saves time during the training process. The callback stopped the training process if the validation accuracy did not improve by at least 0.01 over eight epochs. The third callback, Model Check point, was used to save the weights of the model with the best validation accuracy during training.

This allows the model to be loaded and used later for further training or inference. The callback saved the weights to a file with the specified name when the validation accuracy improved.

The Adamax optimizer with the custom model and the Adam optimizer with the rest. The Adamax optimizer is chosen for its strong

performance on large-scale, sparse datasets, while the Adam optimizer is a widely used optimizer for gradient-based optimization and is effective for a variety of deep-learning tasks.

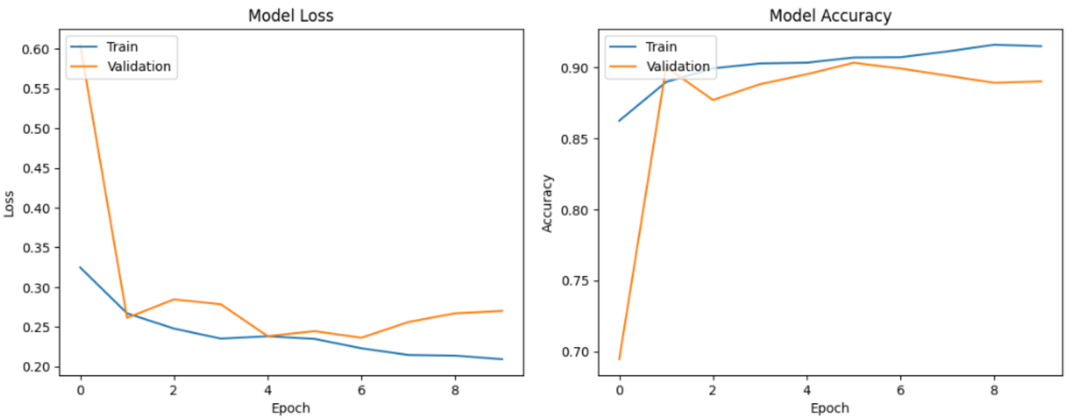
4. Results

4.1. Customized Model

The results of the customized model indicate an accuracy of 89% for both the training and validation sets, with a loss of 0.2706. However, the classification report (Table I) and confusion matrix (Figure 5) demonstrate inadequate performance in detecting melanoma skin lesions, with a precision of only 0.50 and a recall of 0 for the melanoma class. As melanoma skin cancer is the minority class in the balanced dataset, the accuracy of 50% is not useful. These findings suggest that training an accurate and reliable model from scratch with the available dataset may not be achievable. To improve performance, a more complex model architecture and a larger dataset may be necessary.

Table 1. Customized model results in the testing dataset

	Precision	Recall	R1-score	Support
0	0.50	1.00	0.67	500
1	0.00	0.00	0.00	500
				500
Accuracy			0.50	1000
Macro avg	0.25	0.50	0.33	1000
Weighted avg	0.25	0.50	0.33	1000



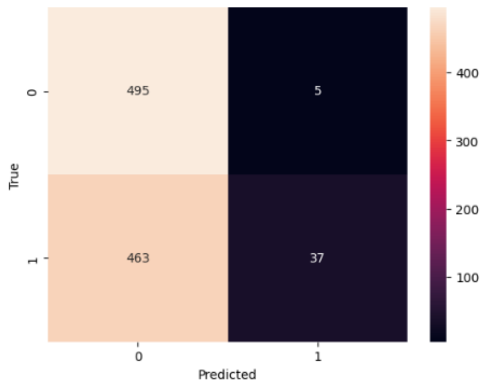


Figure 5: Customized Model Results In The Training And Testing Dataset
Figure 6. Customized model results in the training and testing dataset.

4.2. EfficientNetB3 Model.

The use of a pre-trained EfficientNetB3 model is a promising approach for melanoma detection. However, the results of this study suggest that the model's performance may not be satisfactory. While the model achieved a high accuracy rate of 91.6% for training and validation data, the confusion matrix and classification report revealed significant misclassifications, particularly for the positive class.

The model misclassified 30 negative images as positive and 210 positive images as negative, which is a significant error rate for melanoma detection. While the overall accuracy of 76% suggests that the model has some utility, it still falls short of being a reliable diagnostic tool. The precision and recall values for the positive class are particularly concerning, indicating that the model may not adequately identify melanoma cases, which is the main objective of this study.

Table 2. EfficientNetB3-based model results in the testing dataset.

	Precision	Recall	R1-score	Support
0	0.69	0.94	0.80	500
1	0.91	0.58	0.71	500
				500
Accuracy			0.76	1000
Macro avg	0.80	0.76	0.75	1000
Weighted avg	0.80	0.76	0.75	1000

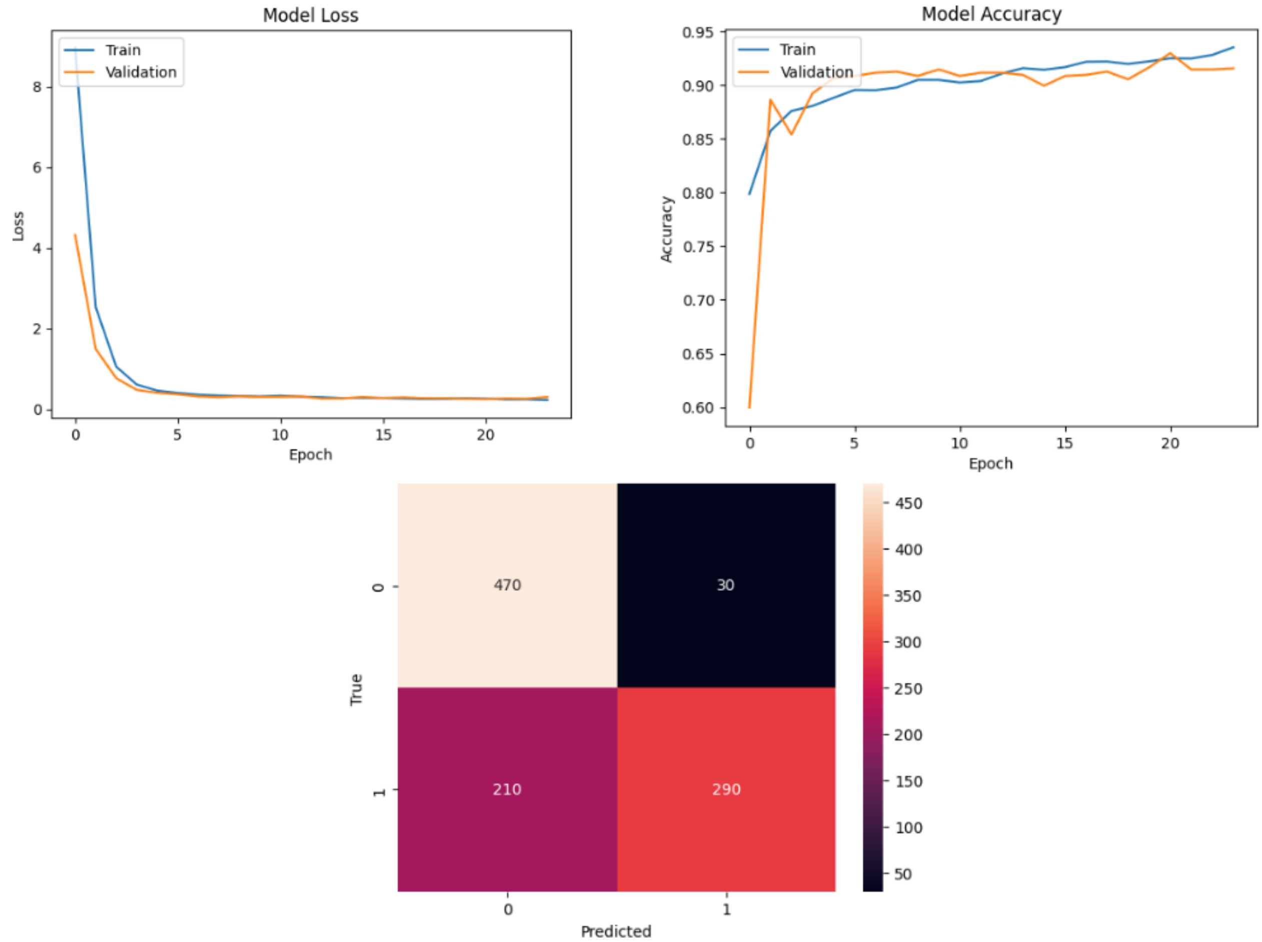


Figure 7. InceptionV3--based model results in the training and testing dataset

4.3. InceptionV3-based Model

The InceptionV3-based model results indicate that an accuracy of 89.8% was achieved for both the training and validation sets by the InceptionV3-based model, with a loss of 0.2519. The confusion matrix (Figure 7) demonstrates that 456 images were accurately classified as negative for melanoma, and 372 images were accurately classified as positive for melanoma. On the other hand, 44 negative images were misclassified as positive, and 128 positive images were misclassified as negative. The classification report shows that the negative class had a precision of 0.78 and a recall of 0.91, while the positive class had a precision of 0.89 and a recall of 0.74. The model's overall accuracy was 83%, indicating reasonable performance for melanoma detection. These results indicate that the InceptionV3-based model has reasonable performance for melanoma detection, with an accuracy rate of 83%. The model achieved high precision and recall values for the negative class, indicating that it can accurately identify benign skin lesions. However, the precision and recall values for the positive class suggest that the model may not adequately identify melanoma cases.

Table 3. Inception-based model results in the testing dataset

	Precision	**Recall**	**R1-score**	**Support**
0	0.78	0.94	0.84	500
1	0.89	0.58	0.81	500
				500
Accuracy			0.83	1000
Macro avg	0.84	0.83	0.83	1000
Weighted avg	0.84	0.83	0.83	1000

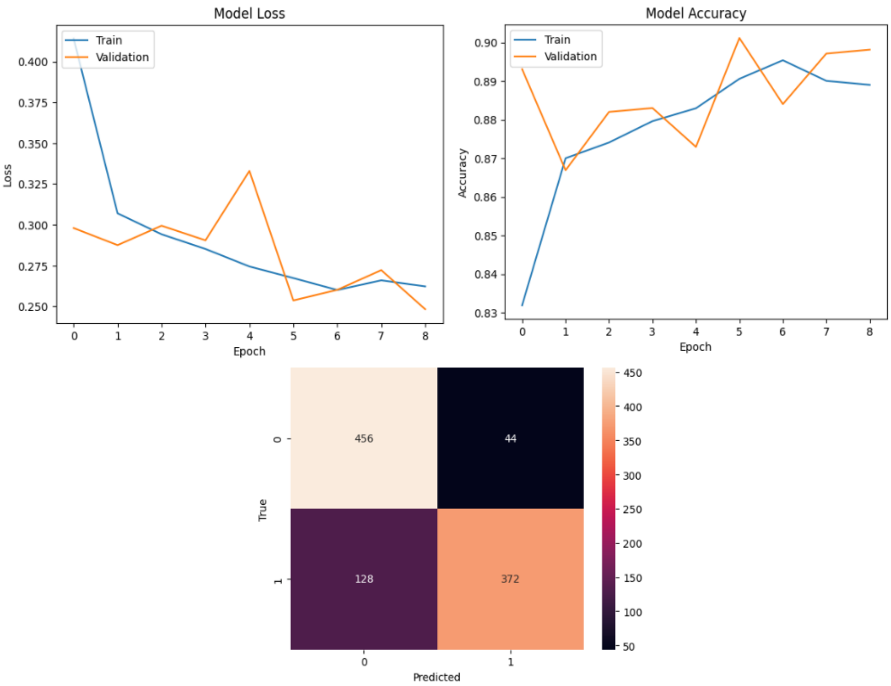


Figure 8. Inception-based model results in the training and testing dataset

4.4. ResNet50-based Model

The ResNet50-based model achieved a low accuracy rate of 50% for both the training and validation sets, with a loss of 0.6934. The confusion matrix indicates that the model misclassified all images as positive for melanoma, as shown in the confusion matrix in Figure 8. The classification report (Table 4) shows that the model has a precision of 0.50 and a recall of 1.00 for the positive class but a precision of 0.00 and a recall of 0.00 for the negative class. The model's overall accuracy is 50%, indicating that it is no better than random guessing. The ResNet50-based model performed poorly for melanoma detection, achieving an accuracy rate of only 50%. The model misclassified all images as positive for melanoma, indicating that it failed to learn any meaningful patterns in the data. This suggests that the model's architecture is unsuitable for this task.

Table 4. ResNet50-based model results in the testing dataset.

	Precision	**Recall**	**R1-score**	**Support**
0	0.00	0.00	0.00	500
1	0.50	1.00	0.67	500
				500
Accuracy			0.50	1000
Macro avg	0.25	0.76	0.33	1000
Weighted avg	0.25	0.76	0.33	1000

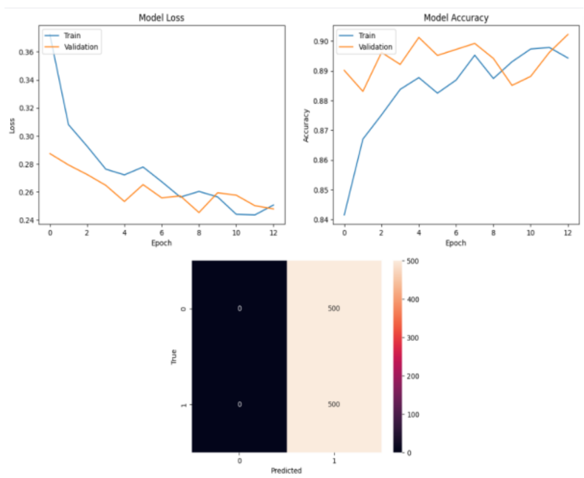


Figure 9. ResNet50-based model results in the training and testing dataset.

4.5. Xception-based Model

The Xception-based model had an accuracy rate of 89.5% for both the training and validation datasets. Significantly, the algorithm accurately categorized a total of 400 photographs as melanoma-positive and 452 images as melanoma-negative. However, it also misclassified 48 images as negative when they were actually positive and 100 images as positive when they were actually negative. The melanoma detection model showed potential, with an overall accuracy rate of 85%. Nevertheless, the considerable proportion of false negatives, which involves erroneously categorizing 100 valid photographs as negative, presents a potential hazard to public health and might potentially result in the delay of critical medical interventions.

Although a significant portion of our study focused on deep learning models, it is essential to underscore the intricate trade-off between precision and computational intricacy. The findings of the Xception model highlight the trade-off between achieving high accuracy and perhaps sacrificing real-world applicability due to increased complexity. The concept of seeking equilibrium is seen in The Trade-off Hypothesis (Skehan, 2009), which posits that optimal performance occurs when there is a harmonious balance between accuracy and manageable complexity.

Table 5. Xception-based model results in the testing dataset.

	Precision	Recall	R1-score	Support
0	0.82	0.90	0.86	500
1	0.89	0.80	0.84	500
				500
Accuracy			0.85	1000
Macro avg	0.86	0.85	0.85	1000
Weighted avg	0.86	0.85	0.85	1000

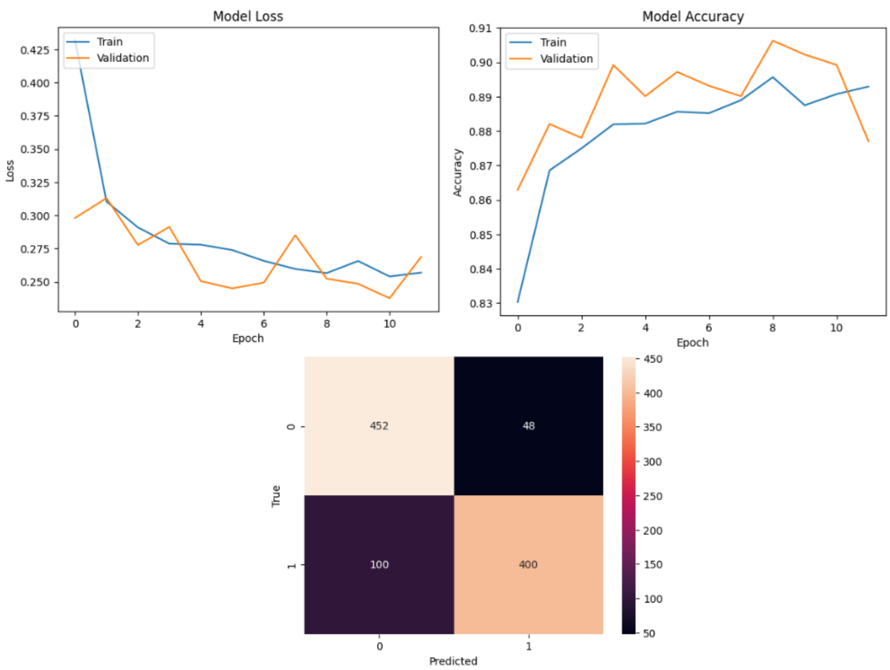


Figure 10. Xception-based model results in the training and testing dataset.

5. Discussion

The outcomes from the five different models utilized for diagnosing melanoma using image processing exhibit varying degrees of performance. First, the customized model demonstrates an accuracy of 50%, indicating substantial difficulty in correctly identifying melanoma cases. This is evident from its low precision of 0.50, suggesting that it frequently misclassifies non-melanoma cases as melanoma. Furthermore, the recall for melanoma is 0, signifying that it fails to identify any true melanoma cases. These results underscore the significant room for improvement in the model's architecture and training approach.

In contrast, the EfficientNetB3-based model exhibits relatively better performance with an accuracy of 76%. However, its precision of 0.69 and a recall of 0.58 for melanoma indicate that, while it is superior to the customized model, it still struggles to accurately classify melanoma cases. The Inception-based model demonstrates promise with an accuracy of 83% and well-balanced precision and recall values for both melanoma and non-melanoma classes, suggesting its potential for melanoma detection. A recent study [32] has identified that Inception V3 model worked best in deep learning algorithms (the accuracy was 93.74%, the sensitivity was 94.36%, and the specificity was 85.64%) further supporting the efficiency of Inception-based model. Nevertheless, further optimization is required to enhance its accuracy for practical deployment. The Xception-based model stands out with an accuracy rate of 85%, achieving the highest accuracy among the models. It demonstrates robust precision and recall values for both melanoma and non-melanoma classes, indicating its potential for effective melanoma detection. These results highlight the Xception-based model as the most promising among the five, offering potential for further refinement and practical application in melanoma diagnosis. Xception-based model reflected good accuracy compared to other baseline models in previous studies [33-37]; but was observed to be less effective in terms of accuracy when compared with deep learning models [32,38]. In contrast, the customized and ResNet50 models require substantial improvements to be viable in real-world scenarios, while the EfficientNetB3 model shows potential but still faces challenges in correctly identifying melanoma cases.

6. Conclusion and Future Work

The research emphasizes the possibility of employing the Xception-based model for CNN-based early melanoma prediction, which is a significant field that is increasingly gaining attention across several domains. The model demonstrates commendable performance, particularly in its ability to accurately identify benign skin lesions, with an accuracy rate of 85%. However, the current accuracy and recall metrics utilized in melanoma detection highlight potential vulnerabilities that necessitate more investigation. It is crucial to acknowledge the inherent limitations of our strategic approach. Despite the considerable capabilities of neural networks, they are not without their limits. Occasionally, deep learning systems may encounter challenges in accurately capturing the intricate patterns associated with melanoma. One may encounter many challenges, such as dependence on the variety and quality of the dataset, the potential issue of overfitting, and the interpretability of the model. Furthermore, it is important to consider that external factors, such as variations in picture resolutions or lighting conditions, might potentially impact the final results. In light of these considerations, next research endeavors will prioritize the enhancement of the Xception-based model, either by modifying the model's structure or by including more diverse datasets. The investigation of attention processes and ensemble techniques may offer a potential avenue for enhancing the sensitivity and specificity of the model in the context of melanoma diagnosis.

References

[1] Pillay, Verosha, et al. "Melanoma skin cancer classification using transfer learning." *Advances in Computational Collective Intelligence*, 2020, pp. 287–297, https://doi.org/10.1007/978-3-030-63119-2_24.

[2] Mayo Clinic. "Melanoma - Symptoms and causes." Mayo Clinic. https://www.mayoclinic.org/diseasesconditions/Melanoma/symptomscauses/syc20374884#:%7E:text=Melanoma%2C%20the%20most%20serious%20type,in%20your%20nose%20or%20throat (accessed Sep. 12, 2022).

[3]. Yanase, Juri, and Evangelos Triantaphyllou. "The seven key challenges for the future of computer-aided diagnosis in medicine." *International journal of medical informatics* vol. 129 (2019): 413-422. doi:10.1016/j.ijmedinf.2019.06.017

[4]. Wadhwa, Raoul R et al. "The accuracy of computer-based diagnostic tools for the identification of concurrent genetic disorders." *American journal of medical genetics*. Part A vol. 176,12 (2018): 2704-2709. doi:10.1002/ajmg.a.40651

[5] Bron, Esther E et al. "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge." *NeuroImage* vol. 111 (2015): 562-79. doi:10.1016/j.neuroimage.2015.01.048

[6] Wu, Xiongwei, et al. "Recent advances in Deep Learning for Object Detection." *Neurocomputing*, vol. 396, 2020, pp. 39–64, https://doi.org/10.1016/j.neucom.2020.01.085.

[7] Elmore, Joann G, et al. "Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: Observer accuracy and reproducibility study." *BMJ*, 2017, https://doi.org/10.1136/bmj.j2813.

[8] Jaarsma-Coes, Myriam G., et al. "Inter-observer variability in MR-based target volume delineation of uveal melanoma." *Advances in Radiation Oncology*, vol. 8, no. 3, 2023, p. 101149, https://doi.org/10.1016/j.adro.2022.101149.

[9] Carli, Paolo, et al. "The problem of false-positive diagnosis in melanoma screening." *Melanoma Research*, vol. 13, no. 2, 2003, pp. 179–182, https://doi.org/10.1097/00008390-200304000-00011.

[10] Kutzner, Heinz, et al. "Overdiagnosis of melanoma – causes, consequences and solutions." JDDG: *Journal Der Deutschen Dermatologischen Gesellschaft*, vol. 18, no. 11, 2020, pp. 1236–1243, https://doi.org/10.1111/ddg.14233.

[11] Duniphin, Darlla. “Limited access to dermatology specialty care: barriers and teledermatology.” *Dermatology Practical & Conceptual,* 2023, https://doi.org/10.5826/dpc.1301a31.
[12] McGuire, Lindsay K., et al. “Melanoma of the lentigo maligna subtype.” *Plastic and Reconstructive Surgery*, vol. 129, no. 2, 2012, https://doi.org/10.1097/prs.0b013e31823aeb72.
[13] Sinz, Christoph, et al. “Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin.” *Journal of the American Academy of Dermatology*, vol. 77, no. 6, 2017, pp. 1100–1109, https://doi.org/10.1016/j.jaad.2017.07.022.
[14] Eftekhari, Aziz, et al. “Current analytical approaches in diagnosis of melanoma.” *TrAC Trends in Analytical Chemistry*, vol. 116, 2019, pp. 122–135, https://doi.org/10.1016/j.trac.2019.05.004.
[15] Razmjooy, N., et al. “A computer-aided diagnosis system for malignant melanomas.” *Neural Computing and Applications*, vol. 23, no. 7–8, 2012, pp. 2059–2071, https://doi.org/10.1007/s00521-012-1149-1.
[16] Codella, N. C., et al. “Deep Learning Ensembles for melanoma recognition in Dermoscopy Images.” *IBM Journal of Research and Development*, vol. 61, no. 4/5, 2017, https://doi.org/10.1147/jrd.2017.2708299.
[17] Yu, Lequan, et al. “Automated melanoma recognition in dermoscopy images via very deep residual networks.” *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, 2017, pp. 994–1004, https://doi.org/10.1109/tmi.2016.2642839.
[18]. Do, Thanh-Toan, et al. “Accessible melanoma detection using smartphones and mobile image analysis.” *IEEE Transactions on Multimedia*, vol. 20, no. 10, 2018, pp. 2849–2864, https://doi.org/10.1109/tmm.2018.2814346.
[19] Li, Yuexiang, and Linlin Shen. “Skin lesion analysis towards melanoma detection using Deep Learning Network.” *Sensors*, vol. 18, no. 2, 2018, p. 556, https://doi.org/10.3390/s18020556.
[20] Adegun, Adekanmi A., and Serestina Viriri. “Deep learning-based system for automatic melanoma detection.” *IEEE Access*, vol. 8, 2020, pp. 7160–7172, https://doi.org/10.1109/access.2019.2962812.
[21] Jojoa Acosta, Mario Fernando, et al. “Melanoma diagnosis using Deep Learning techniques on dermatoscopic images.” *BMC Medical Imaging,* vol. 21, no. 1, 2021, https://doi.org/10.1186/s12880-020-00534-8.
[22] American Cancer Society. “Cancer facts and figures 2020.” American Cancer Society. https://www.aimatMelanoma.org/about-Melanoma/Melanoma-stats-facts (accessed Nov. 29, 2021)
[23] Zhang, Long et al. “Optimization of the Convolutional Neural Networks for Automatic Detection of Skin Cancer.” *Open medicine (Warsaw, Poland)* vol. 15 27-37. 13 Jan. 2020, doi:10.1515/med-2020-0006
[24] Esteva, Andre et al. “Dermatologist-level classification of skin cancer with deep neural networks.” *Nature* vol. 542,7639 (2017): 115-118. doi:10.1038/nature21056
[25] Ericsson Inc. “Ericsson mobility report, 2019.” Ericsson. https://www.ericsson.com/en/mobilityreport/reports/june-2019 (accessed Nov. 29, 2021).
[26] Perez, Fabio, et al. “Solo or ensemble? choosing a CNN Architecture for Melanoma Classification.” 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019,* https://doi.org/10.1109/cvprw.2019.00336.
[27] ADDI. “Automatic computer-based diagnosis system for dermoscopy images.” ADDI. https://www.fc.up.pt/addi/ph2%20database.html (accessed Nov. 29, 2021).
[28] H. Javed. “Melanoma skin cancer dataset of 10000 Images.” Kaggle. https://www.kaggle.com/datasets/hasnainjaved/melanoma-skin-cancer-dataset-of-10000-images
[29] Mukhamediev, Ravil I., et al. “Review of Artificial Intelligence and Machine Learning Technologies: Classification, restrictions, opportunities and challenges.” *Mathematics*, vol. 10, no. 15, 2022, p. 2552, https://doi.org/10.3390/math10152552.
[30] Gayatri, Erapaneni, and S. L. Aarthy. “Challenges and imperatives of deep learning approaches for detection of melanoma: A Review.” *International Journal of Image and Graphics*, vol. 23, no. 03, 2022, https://doi.org/10.1142/s0219467822400125.
[31] Panthakkan, Alavikunhu, et al. “Concatenated Xception-Resnet50 — a novel hybrid approach for accurate skin cancer prediction.” *Computers in Biology and Medicine*, vol. 150, 2022, p. 106170, https://doi.org/10.1016/j.compbiomed.2022.106170.
[32] Cui, Xiaoyu, et al. “Assessing the effectiveness of artificial intelligence methods for melanoma: A retrospective review.” *Journal of the American Academy of Dermatology*, vol. 81, no. 5, 2019, pp. 1176–1180, https://doi.org/10.1016/j.jaad.2019.06.042.
[33] Phillips, Michael et al. “Detection of Malignant Melanoma Using Artificial Intelligence: An Observational Study of Diagnostic Accuracy.” *Dermatology practical & conceptual* vol. 10,1

e2020011. 31 Dec. 2019, doi:10.5826/dpc.1001a11

[34] Dascalu, A., et al. “Non-melanoma skin cancer diagnosis: A comparison between dermoscopic and smartphone images by unified visual and Sonification Deep Learning Algorithms.” Journal of *Cancer Research and Clinical Oncology*, vol. 148, no. 9, 2021, pp. 2497–2505, https://doi.org/10.1007/s00432-021-03809x.

[35] Phillips, Michael, et al. “Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions.” *JAMA Network Open*, vol. 2, no. 10, 2019, https://doi.org/10.1001/jamanetworkopen.2019.13436.

[36] Kim, Chan-Il, et al. “Computer-aided diagnosis algorithm for classification of malignant melanoma using Deep Neural Networks.” *Sensors*, vol. 21, no. 16, 2021, p. 5551, https://doi.org/10.3390/s21165551.

[37] Rajpara, S.M., et al. “Systematic review of Dermoscopy and Digital Dermoscopy/ artificial intelligence for the diagnosis of melanoma.” *British Journal of Dermatology,* vol. 161, no. 3, 2009, pp. 591–604, https://doi.org/10.1111/j.1365-2133.2009.09093.x.

[38] De Logu, Francesco, et al. “Recognition of cutaneous melanoma on digitized histopathological slides via artificial intelligence algorithm.” *Frontiers in Oncology*, vol. 10, 2020, https://doi.org/10.3389/ fonc.2020.01559.

[39] Tschandl, P., Rosendahl, C. & Kittler, H. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.” *Sci Data,* vol. 5, 2018, https://doi.org/10.1038/sdata.2018.161

[40] Gutman, David; Codella, Noel C. F.; Celebi, Emre; Helba, Brian; Marchetti, Michael; Mishra, Nabin; Halpern, Allan. "Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)". eprint arXiv:1605.01397. 2016.

[41] Codella N, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza S, Kalloo A, Liopyris K, Mishra N, Kittler H, Halpern A. "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)". arXiv: 1710.05006 [cs.CV]

[42] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368

[43] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288.

[44] Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., Codella, N., Combalia, M., Dusza, S., Guitera, P., Gutman, D., Halpern, A., Helba, B., Kittler, H., Kose, K., Langer, S., Lioprys, K., Malvehy, J., Musthaq, S., Nanda, J., Reiter, O., Shih, G., Stratigos, A., Tschandl, P., Weber, J. & Soyer, P. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Sci Data 8, 34 (2021). https://doi.org/10.1038/s41597-021-00815-z