Melanoma Prediction among Benign Pigmented Lesions from an Image using a Convolutional Neural Network

Ivan Kaipov, Darya Belavus

***Abstract*—**This research article presents a study on the use of deep learning models for melanoma prediction in skin lesions. We trained a VGG19 model on the ISIC dataset and achieved an accuracy of 88.4% on the test set. However, there were false negative cases, which we addressed by applying a threshold of 38% to the probability of an image being classified as melanoma. Using this threshold, we achieved a sensitivity of 100% for melanoma diagnosis with no false negative cases, while still maintaining a relatively high specificity. Our results highlight the potential of deep learning models for improving melanoma prediction in clinical practice, and the importance of incorporating regular skin cancer screenings and early detection measures, even among expert dermatologists.

***Keywords*—**Benign lesions, Convolutional neural network (CNN), feature extraction, medical image analysis, melanoma prediction.

I. Introduction

More than 3.3 million new cases of skin cancer are diagnosed each year worldwide. There are several types of skin cancer: squamous cell, basal cell, adenocarcinoma and melanoma. The deadliest skin cancer is melanoma, also known as the Black Death. According to the World Health Organization (WHO), melanoma accounts for approximately 1% of all skin cancer but causes the majority of skin cancer deaths. In 2020, an estimated 325 thousand new cases of melanoma were diagnosed worldwide, and 61 thousand people died from the disease [1]. The total cost of diagnosing melanoma can vary widely depending on many factors, including the specific tests and procedures used. However, according to a 2018 study published in the Journal of the American Academy of Dermatology, the average cost of diagnosing and managing a new case of melanoma in the United States was estimated to be approximately 2300-2500 US dollars [2]. At the same time, when diagnosing melanoma of the first stage, the probability of patient survival in the next five years is 90%. The probability of survival of the patient in the next five years with the fourth stage of cancer is 20%. Thus, early diagnosis of melanoma is the most important task for patient survival.

The purpose of this paper is to propose a method for predicting the presence of melanoma in the image of a pigmented formation on human skin. At the same time, the proposed prediction method should eliminate false negative predictions when the pigmented formation is melanoma, but the result of the forecast is a benign formation, and minimize false positive predictions when the benign formation is predicted as a melanoma.

This publication considers the problem of classification by image of three classes of skin formations: melanoma, nevus, seborrheic keratosis. Seborrheic keratosis is a very common non-cancerous (benign) skin growth that appears as a waxy, raised, or rough texture lesion that ranges in color from light tan to black. These growths are typically round or oval in shape and can vary is size from a few millimeters to several centimeters in diameter. They are most commonly found on the face, neck, chest, and back, but can appear anywhere on the body (Fig. 1). Seborrheic keratoses are usually harmless and do not require treatment.

I. K. Kaipov, Belarusian State University, Minsk, Belarus (e-mail: kaipov1995@gmail.com)
D. P. Belavus, Vitebsk State Order of Peoples' Friendship Medical University, Vitebsk, Belarus (e-mail: d.belous@vgmu.vitebsk.by).

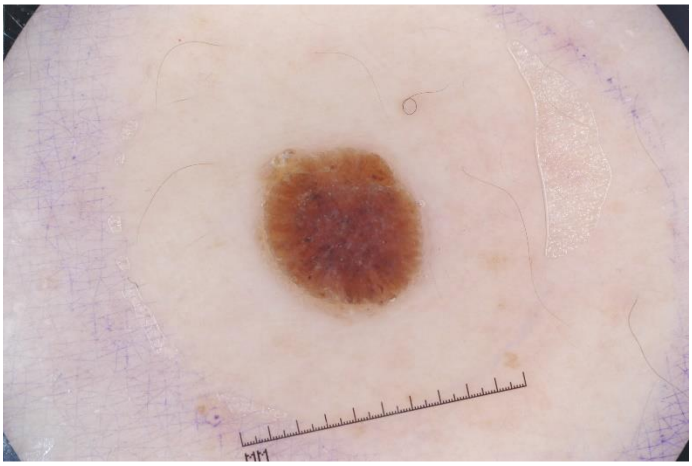

Fig. 1 Seborrheic keratoses

A nevus, also known as a mole, is a common type of benign (non-cancerous) skin growth that is made up of pigmented cells called melanocytes. Moles can appear anywhere on the body, and they usually brown, black, or flesh-colored (Fig. 2). They can range in size from very small to larger than a pencil eraser. Most people have moles, and they often appear in childhood or adolescence. Moles can continue to appear throughout a person's life and it is normal to have anywhere from 10 to 40 moles by adulthood. Most moles are harmless.

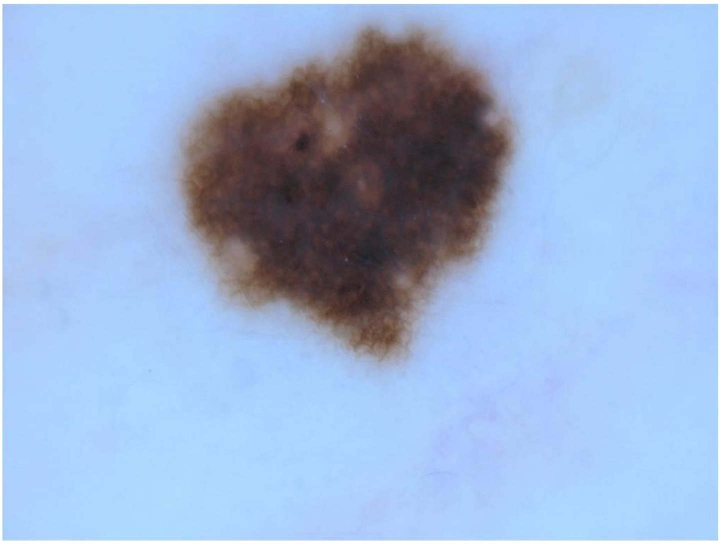

Fig. 2 Nevus

In the same time, signs of melanoma include a mole or spot that is asymmetrical, has an irregular border, is multi-colored or has uneven coloring, has a diameter larger than a pencil eraser, or is changing in size, shape, or color (Fig. 3). It may also be itchy, bleeding, or painful.

It can be difficult to distinguish melanoma from a nevus of seborrheic keratosis, especially in the early stages. This is because melanoma can sometimes look similar to a normal nevus or a harmless skin growth.

Traditionally, dermatologists have relied on their expertise and visual inspection to diagnose skin lesions. However, this process can be time-consuming, expensive and subjective, and there can be variation among dermatologists in their ability to accurately diagnose skin lesions. In addition, some skin lesions may be difficult to diagnose even for experienced dermatologists.

Using a convolutional neural network (CNN) to classify skin lesions based on images can help overcome some of these challenges. CNNs are a type of deep learning algorithm that can automatically extract features from images and use them to make predictions. By training a CNN on a large dataset of skin lesion images that have been labeled as melanomas, nevi, or seborrheic keratoses, the algorithm can learn to recognize patterns and features that are characteristic of each type of lesion.

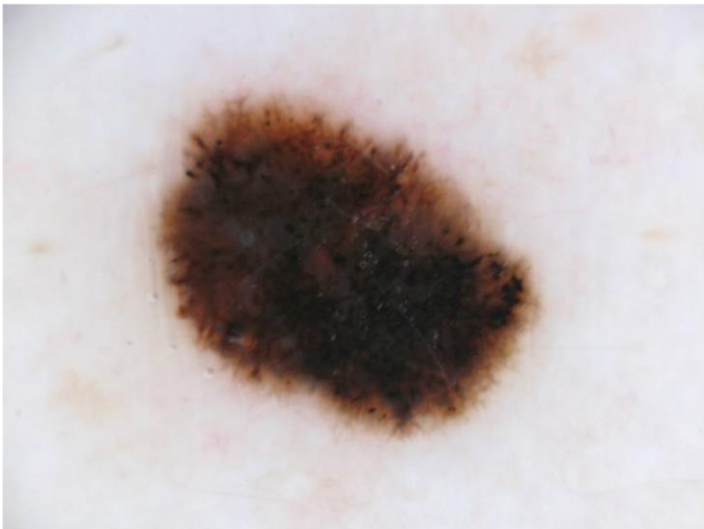

Fig. 3 Melanoma

Once the CNN has been trained, it can be used to quickly and accurately classify new skin lesion images. This can help dermatologists make more accurate and consistent diagnoses, and it can be also help improve access to diagnosis in areas where there may be a shortage of dermatologists or other healthcare providers.

Overall, using a CNN to classify skin lesions by image has the potential to improve accuracy, speed, cost, and accessibility of diagnosis, which can ultimately lead to better patient outcomes.

II. IMAGE PREPROCESSING

A. Dataset

To train a convolutional neural network, an ISIC melanoma dataset is used. The ISIC melanoma dataset is widely used dataset of skin lesion images that is commonly used for research in skin lesion classification and melanoma detection. The dataset contains over 23 thousand images of skin lesions, including melanomas, nevi and seborrheic keratoses.

The images in the ISIC melanoma dataset are collected from various sources, including clinical settings, research studies, and social media platforms, and are annotated by dermatologists to indicate the diagnosis of the lesion. The high quality and diversity of the images in the dataset make it valuable resource for researches and healthcare professionals working on skin lesion diagnosis and melanoma detection.

The ISIC melanoma dataset contains over 2000 melanoma images, 14000 nevus images and 3000 seborrheic keratosis images. Each image in the ISIC melanoma dataset is annotated by one or more expert dermatologists to indicate the diagnosis of the lesion. The images are high quality and have a resolution of 768x1024 pixels. The images were captured using various imaging devices, including dermatoscopes, digital cameras and smartphones.

The dataset was divided into training, validation and test sets in the proportions presented in Table 1. Stratified sampling is used to ensure that each set contains a representative sample of each class. Stratified sampling helps to ensure that model is trained and evaluated on a representative sample of the population and that it does not overfit to certain classes.

TABLE 1
DIVIDING IMAGES INTO SETS

	Training set	Validation set	Test set
Melanoma	2318	462	531
Nevus	16232	3246	3728
Seborrheic keratosis	4638	918	1053

B. Image Preprocessing

The first step is data preparation. Images from dataset were resized to size 224x224 pixels. The pixel values were rescaled between 0 and 1. Also, the pixel values were normalized by subtracting the mean and dividing by the standard deviation. This helps to ensure that the model is not biased towards certain pixel values. To ensure that the skin lesion is in the center of the image, center cropping were applied. Data augmentation parameters are:

- Rotation range – 90 degrees;
- Zoom range – 1.1;

Input images were converted to grayscale to apply histogram equalization to enhance the contrast of the images. Lesion segmentation is performed on the enhanced image using a thresholding technique [3]. The proposed approach is able to improve the accuracy of lesion segmentation and feature extraction. Another common technique for image enhancement is Gaussian smoothing [4]. There are another techniques for image enhancement, in [5] provides an overview of image processing techniques for melanoma detection, including color space conversion, image filtering and feature extraction. Another research article proposes a deep learning framework for image recognition, which has been applied to melanoma detection with good results [6]. This survey paper [7] provides a comprehensive review of image processing techniques for lesion border detection in dermoscopy images, including preprocessing techniques such as color correction, noise reduction, and image enhancement. This article [8] proposes efficient and effective preprocessing techniques for skin lesion segmentation, including color normalization, contrast stretching, and morphological operations. Another review article provides an overview of preprocessing techniques for skin lesion detection and segmentation, including noise reduction, histogram equalization, and contrast enhancement [9].

In conclusion, effective image preprocessing techniques can significantly improve the accuracy and reliability of melanoma detection using machine learning models. By standardizing and enhancing images in the dataset, we can reduce the impact of variations in lighting, skin color, and other factors that can affect image quality. Applying data augmentation techniques can also help to increase the size of the dataset and improve the robustness of the model. Moreover, the use of advanced techniques such as histogram equalization and Gaussian smoothing can further enhance the quality of images, making it easier for the machine learning model to detect melanoma accurately.

Overall, by implementing a thorough image preprocessing pipeline, it’s possible increase the sensitivity and specificity of the melanoma detection model, leading to earlier detection and better treatment outcomes. This can have a significant impact on the lives of those at risk of melanoma and contribute to the wider goal of reducing the incidence of this deadly form of skin cancer.

C. Hair detection

Hair present in the images of the ISIC melanoma dataset interfere with the accurate detection of melanoma. This is because the model may focus on the hair rather than the skin, leading to false positives or false negatives predictions. For hair segmentation was used method proposed by [10].

The hair segmentation method involves using a convolutional neural network (CNN) to classify each pixel in the image as either hair or non-hair. The CNN was trained on a separate dataset of hair and non-hair images, and then applied to the skin lesion images to segment out the hair.

The neural network architecture used for hair segmentation is U-net. In addition to the basic U-net architecture, a hair detection module was added to the model. This module consisted of convolutional and pooling layers that extracted features from the input image, followed by a fully connected layer that predicted whether each pixel was part of hair or not. This prediction was then used to mask the input image before passing it through the main CNN for melanoma prediction.

III. Model Architecture

The success of deep learning in various fields has made it an increasingly popular approach for medical image analysis. In particular, convolutional neural networks (CNNs) have been shown to be effective for various medical image analysis tasks, including skin lesion analysis for melanoma detection. The architecture of the CNN is crucial for its success in this task, as it determines the complexity and expressiveness of the model. In this chapter, we will discuss the various CNN architectures that have been proposed for skin lesion analysis, with a particular focus on the VGG19 architecture. Also, will discuss the modifications that can be made to the VGG19 architecture to better fit the requirements of the ISIC melanoma dataset.

VGG19 is a popular deep learning architecture that has been widely used for image classification tasks. It is known for its simplicity and high accuracy, making it a preferred choice for various applications. However, it is not the only architecture available for image classification tasks, and there are other popular alternatives that have their own strengths and weaknesses. The VGG19 architecture achieved state-of-the-art accuracy on the ImageNet dataset, which contains over a million labeled images from 1000 different object categories.

ResNet-50 is another widely used deep learning architecture that is known for its ability to avoid the vanishing gradient problem, which can be a common issue in deep networks. ResNet uses skip connections to enable easier training of deeper networks, allowing it to achieve higher accuracy on more complex datasets.

Another popular architecture is Inception-v3, which is known for its ability to handle different scales of features in an image. It uses a series of convolutional layers with varying kernel sizes to capture both local and global features, making it useful for tasks such as object detection and segmentation.

There are also newer architectures, such as EfficientNet, which use a combination of scaling and compound scaling techniques to achieve high accuracy with fewer parameters. EfficientNet has been shown to outperform other architectures on large-scale image classification tasks, making it a promising option for future research.

There are several factors to consider when choosing a deep learning model architecture for a specific task such as melanoma detection. One of the most important factors is the model's accuracy and ability to perform well on the task at hand.

In the field of medical images classification, various deep learning architectures have been proposed, and comparative studies have been conducted to evaluate their performance. One such study [11] compared the performance of VGG19, Inception-v3, Inception-ResNet-v2, ResNet-50 and DenseNet-169. The study found that VGG19 achieved the highest accuracy among the tested models, with an accuracy 91.1%, followed by Inception-v3 with 90.0%, Inception-ResNet-v2 with 89.7%, ResNet-50 with 88.8%, and DenseNet-169 with 88.1%. The study concluded that VGG19 is an effective model for medical images classification, especially when a high level of accuracy is required. Therefore, based on the results of this study, it can be argued that VGG19 is a suitable model architecture for melanoma detection.

VGG19 also has a relatively simple architecture compared to other models and has been shown to work well on a wide range of image recognition tasks. The model is composed of series of convolutional layers with small filters followed by max pooling layers, which help to reduce the spatial dimensions of the feature maps. The fully connected layers at the end of the model allow for classification based on the extracted features.

Another advantage of VGG19 is its availability and ease of use. The pre-trained model is readily available in popular deep learning libraries such as Keras and TensorFlow, allowing researches and developers to quickly implement and fine-tune model.

Therefore, based on its high accuracy, simple architecture, and ease of use, VGG19 is a strong choice for melanoma detection.

The VGG19 model architecture for melanoma detection with the number of weights in each layer (Table 2), batch size is 16:

TABLE 2
VGG19 MODEL ARCHITECTURE

Layer (type)	Output Shape	Number of Weights
Input	(16, 224, 224, 3)	0
Block 1 Conv 1	(16, 224, 224, 64)	1792
Block 1 Conv 2	(16, 224, 224, 64)	36928
Block 1 Max Pool	(16, 112, 112, 64)	0
Block 2 Conv 1	(16, 112, 112, 128)	73856
Block 2 Conv 2	(16, 112, 112, 128)	147584
Block 2 Max Pool	(16, 56, 56, 128)	0
Block 3 Conv 1	(16, 56, 56, 256)	295168
Block 3 Conv 2	(16, 56, 56, 256)	590080
Block 3 Conv 3	(16, 56, 56, 256)	590080
Block 3 Max Pool	(16, 28, 28, 256)	0
Block 4 Conv 1	(16, 28, 28, 512)	1180160
Block 4 Conv 2	(16, 28, 28, 512)	2359808
Block 4 Conv 3	(16, 28, 28. 512)	2359808
Block 4 Max Pool	(16, 14, 14, 512)	0
Block 5 Conv 1	(16, 14, 14, 512)	2359808
Block 5 Conv 2	(16, 14, 14, 512)	2359808
Block 5 Conv 3	(16, 14, 14, 512)	2359808
Block 5 Max Pool	(16, 7, 7, 512)	0
Flatten	(16, 25088)	0
Dense 1 with Dropout	(16, 4096)	102764544
Dense 2 with Dropout	(16, 4096)	16781312
Output	(16, 3)	12291
Total number of weights	-	134272835

The following hyperparameters proved to be the best in accuracy (Table 3):

TABLE 3
VGG19 HYPERPARAMETER VALUES

Hyperparameter	Value	Description
Learning rate	0.0001	The step size at which the model’s weights are updated during training
Loss function	Categorical cross-entropy	The loss function used to measure the difference between predicted and actual class probabilities
Optimizer	Adam	The optimization algorithm used to update the model’s weights during training
Batch size	16	The number of training examples used to each iteration of training
Epochs	50	The number of times the entire training dataset is passed through the model during training
Early stopping	Yes, 5 epochs	Stopping the training process when the validation loss stops improving
Dropout rate	0.5	The proportion of nodes in the network that are randomly dropped out during training to prevent overfitting
Weight initialization	He normal	The method used to initialize the weights in

		the network

IV. RESULTS

There were a total of 5312 images in test set, of which 531 were melanoma, 3728 were seborrheic keratosis, and 1053 were nevi. The output gives the probability that the image belongs to one of the three classes. The image was assigned to the class with the highest probability. The results are presented in Table 4.

TABLE 4
RESULTS WITH ASSIGNMENT TO THE CLASS IN THE HIGHEST PROBABILITY

	Melanoma	**Seborrheic keratosis**	**Nevus**
Melanoma	502	18	11
Seborrheic keratosis	77	3478	173
Nevus	134	204	715

The overall accuracy is 88.4% but there are false negative cases in the detection of melanoma. In order to eliminate false negative cases, the image threshold has been optimized to not classify melanoma images as seborrheic keratosis or nevus. Threshold level is 38%.

When applying a new level of threshold, false negative cases were excluded, but the number of false positive cases increased (Table 5).

TABLE 5
RESULTS WITH OPTIMIZED THRESHOLD

	Melanoma	**Seborrheic keratosis**	**Nevus**
Melanoma	531	0	0
Seborrheic keratosis	172	3316	169
Nevus	247	193	684

With the applied threshold level of 38%, the model has achieved a significant improvement in correctly identifying melanoma cases, with all 531 melanoma cases being correctly identified as such (true positives) and none being missed (false negatives). The overall accuracy is 84.4% with optimized threshold. However, this has come at the cost of an increase in false positives for seborrheic keratosis and nevi, which may lead to unnecessary biopsies of further tests in those cases.

A commonly cited study published in the Journal of the American Academy of Dermatology in 2018 [12] evaluated the diagnostic accuracy of 157 dermatologists on a set of 370 images of skin lesions, including 111 melanomas. The dermatologists had a mean accuracy of 86.0%. Dermatologists have had false negatives cases with melanoma. According to a study published in JAMA Dermatology in 2018 [13], dermatologists missed a total of 30 out of 370 melanomas in clinical images, resulting in a accuracy of 92.2%. This means that even expert dermatologists can miss some cases of melanoma.

Both missed melanoma and unnecessary biopsies of benign lesions have serious consequences, so it is important to minimize both. However, missed melanoma is more concerning as it can lead to delays in diagnosis and treatment, which can result in poorer outcomes and even death. On the other hand, unnecessary biopsies of benign lesions can cause anxiety, scarring, and other complications, but are generally less serious than missed melanoma. Therefore, it is important to prioritize the detection of melanoma while minimizing unnecessary biopsies.

V. CONCLUSION

In conclusion, this research article explored the application of deep learning techniques for melanoma detection using the ISIC melanoma dataset. Specifically, we implemented the VGG19 architecture and achieved an accuracy of 88.4% on the test set. However, we identified the issue of false negative results, which can have serious consequences in the diagnosis of melanoma. To address this, we applied a threshold of 38% probability for melanoma detection, which resulted in an accuracy of 84.4% with no false negative cases.

We also compared our results with the diagnostic accuracy of dermatologists on skin lesions and found that our model is comparable with the mean accuracy of dermatologists in a commonly cited study. However, it is important to note that even expert dermatologists can miss cases of melanoma, emphasizing the need for continued efforts in skin cancer screenings and early detection.

Overall, this study demonstrates the potential of deep learning techniques for melanoma detection and highlights the importance of developing accurate and reliable methods for early diagnosis. Further research could explore the use of other deep learning architectures and datasets to improve the accuracy and generalizability of melanoma detection models.

REFERENCES

[1] World Health Organization. Available at https://gco.iarc.fr/today/data/factsheets/cancers/16-Melanoma-of-skin-fact-sheet.pdf (accessed 2022, Nov).

[2] Alam, M., et. al., "The economic burden of melanoma: Direct and indirect costs in the United States", *Journal of the American Academy of Dermatology*, vol. 78, no. 3, 2018, pp. 501-511.e11.

[3] S. Saha, S. Mukherjee, “A Systematic Approach for Automated Melanoma Detection Using Infrared and Histogram Equalization”, *International Journal of Advanced Computer Science and Applications*, Dec. 2017.

[4] P. Chouhan, A. Singh, and S. Biswas, “Skin Lesion Segmentation Using Deep Learning with Gaussian Filter”, *in International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, Feb. 2019, pp. 97-102.

[5] S. S. Mirhashemi, N. Nodehi and M. R. Gity, “Image Processing Techniques for Melanoma Detection”, *in IEEE International Conference on Computational Intelligence and Computer Research*, Enathi, 2011, pp. 1-6.

[6] He, K., et al. “Deep Residual Learning for Image Recognition”, *in IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016, pp. 770-778.

[7] M. Emre Celebi, H. Iyatomi and Y. Ohno, “A State-of-the-Art Survey on Lesion Border Detection in Dermoscopy Images”, Dermoscopy Image Analysis, 2015, pp. 129-159.

[8] A. Tsantili, D. G. Tsalikakis and I. Pitas, “Efficient and Effective Preprocessing Techniques for Skin Lesion Segmentation”, *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, Sept. 2011, pp. 801–813.

[9] R. Alomari, R. Alkhuzai and K. Alahmad, “Preprocessing Techniques for Skin Lesion Detection and Segmentation: A Review”, *International Journal of Advanced Computer Science and Application*, vol. 10, 2019, pp. 284-291.

[10] Tschandl, P., et. al., “Semantic segmentation with hair detection using convolutional neural networks for skin lesion analysis”, *in International Symposium on Biomedical Imaging*, pp. 496-500.

[11] Esteva, A., et al., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, 2014, pp. 115-118.

[12] Lallas, A., et. al., “Accuracy of dermatologists on the diagnosis of skin cancer and melanoma: is it possible to improve training?”, *Journal of the American Academy of Dermatology*, 2018, vol. 79, no. 5, pp. 856–858.

[13] Lott, JP., et. al., “Diagnostic Inaccuracy of Smartphone Applications for Melanoma Detection”, *JAMA Dermatol*, 2018, vol. 154, no. 4, pp. 422-427.