

# **Skin Cancer Detection Using Deep Learning**

**by**

Alon Meshulam

Ori Ekshtein

Ariel Dvori

## **Abstract:**

Skin cancer is one of the most prevalent forms of cancer. Early and accurate diagnosis is critical for effective treatment. In this work, we develop and evaluate deep learning models for classifying skin lesions as benign or malignant from dermoscopic images. Three models are implemented and compared: a neural network (NN), logistic regression, and a convolutional neural network (CNN). The NN and CNN achieve the highest accuracy of 88% and 91% respectively on a dataset of 13,900 images. The CNN outperforms the other models, benefiting from convolutional layers that extract spatial features. These results demonstrate the potential of deep learning for automated skin cancer diagnosis.

# **1 Introduction**

Skin cancer represents a significant public health problem, with over 5 million cases diagnosed annually in the U.S. alone [1]. Melanoma is the deadliest form, accounting for most skin cancer fatalities [2]. Early detection dramatically improves prognosis, with 5-year survival rates exceeding 95% for localized melanoma compared to only 23% for late stage [3]. However, diagnosis can be challenging even for dermatologists, with reported diagnostic accuracy around 75-84% [4]. This has motivated research into computer-aided diagnosis using dermoscopic images to improve diagnostic accuracy.

Recent advances in deep learning provide new opportunities for automated skin lesion analysis. Convolutional neural networks (CNNs) in particular have achieved impressive results for medical image analysis across domains [5]. CNNs contain convolutional layers to automatically learn spatial hierarchies of visual features optimized for the task. In this work, we develop and evaluate CNN models for classifying skin lesions as benign or malignant from dermoscopic images. We also implement and compare two other models: a standard neural network (NN) and logistic regression.

# **2 Related Work**

A number of prior studies have investigated deep learning for skin cancer classification in dermoscopic images. Esteva et al. [7] classified lesions using a pre-trained Inception v3 CNN architecture, achieving accuracy comparable to dermatologists. Other work has developed custom CNN models tailored to this task [8,9]. Haenssle et al. [9] compared a CNN to 58 dermatologists on melanoma classification and found it outperformed all of them. These studies demonstrate the promise of deep learning for this application. Our work aims to establish baselines with simple model architectures to build upon.

# **3 Background**

Common skin cancers include basal cell carcinoma, squamous cell carcinoma, and melanoma. Melanoma accounts for only 1% of cases but a majority of deaths [2]. Diagnosis often begins with visual examination for asymmetry, irregular bor-

ders, color variegation, large diameter, and evolution over time (ABCDE criteria) [6]. Dermoscopy provides additional morphological information invisible to the naked eye. Computational analysis of dermoscopic images can aid diagnosis by discovering subtle, hard-to-perceive patterns.

Deep learning has shown considerable promise for medical image analysis. CNNs and related architectures provide state-of-the-art performance on tasks such as disease detection, localization, segmentation, and classification [5]. Key advantages include the automated learning of problem-specific features through multiple convolutional layers versus hand-crafted feature extraction. By training end-to-end on labeled datasets, CNNs can discover intricate visual patterns for accurate diagnosis.

## 4 Methods

We utilize a public dataset of 13,900 dermoscopic images from Kaggle [10]. The data comprises uniformly sized 224 x 224 pixel RGB images with expert labels of benign (9371 images) or malignant (4529 images).

Three models are implemented in Python with Keras using TensorFlow:

**Neural Network (NN):** The NN contains an input layer to flatten the 224 x 224 x 3 input images into vectors, 3 fully connected hidden layers of 64, 32, and 32 units with ReLU activation, and an output layer with sigmoid activation for binary classification.

**Logistic Regression (a.k.a SoftMax):** Logistic regression is applied to the flattened input images with sigmoid output activation. This simple linear model establishes a baseline.

**Convolutional Neural Network (CNN):** The CNN comprises 4 Conv2D layers with 32 filters and 3x3 kernels, each followed by 2x2 max pooling layers. This feeds into a flatten layer and single sigmoid output node. ReLU activation is used.

All models are trained for 25 epochs with Adam optimization, binary cross-entropy loss, and a 128 batch size. Standard data augmentation (rescale only) is applied during training. Key metrics tracked are accuracy, loss, precision, recall, and confusion matrix on the validation set.

Model	Accuracy	Precision	Recall	Loss
Logistic Regression	71%	93%	46%	0.287
Neural Network	88%	87%	89%	0.118
Convolutional NN	91%	87%	95%	0.097

Table 1: Model learning curves for training and validation.

## 5 Results

Table 1 summarizes the performance of the three models. The CNN achieves the best overall accuracy of 91%, followed closely by the NN at 88%. The simpler logistic regression model performs noticeably worse with 71% accuracy. The superiority of the CNN and NN validate the importance of learning intricate features through multiple layers rather than relying on linear decision boundaries.

Interestingly, while the NN achieves slightly lower overall accuracy than the CNN, its precision and recall are more balanced. The CNN has higher recall (sensitivity) of 95% but lower precision of 87% compared to 87% for the NN. The CNN prioritizes detecting more malignant cases at the expense of more false positives. Depending on the application, the NN may represent a better operating point.

The learning curves in Figure 1 show how the models improve with more training. The NN and CNN obtain only 60-70% accuracy after the first epoch but continuously improve until converging. The CNN converges faster within the first 5 epochs. The gap between training and validation accuracy indicates some overfitting, especially for the NN. Overall, the results verify the capability of deep learning for this task.

The confusion matrices in Figure 2 provide insight into the types of errors made by each model. The NN achieves a relatively balanced number of false positives and false negatives. Meanwhile, logistic regression exhibits a large number of false negatives, misclassifying many malignant cases as benign. The CNN has fewer false negatives but more false positives compared to the NN.

Quantitatively, the NN correctly predicts 866 benign and 898 malignant cases, while mispredicting 134 benign as malignant and 102 malignant as benign. Logistic regression substantially underpredicts malignancies, with 539 false negatives. The CNN achieves 949 true malignant predictions but 144 false malignant. The confusion matrices thus reveal nuances in the types of errors beyond the overall

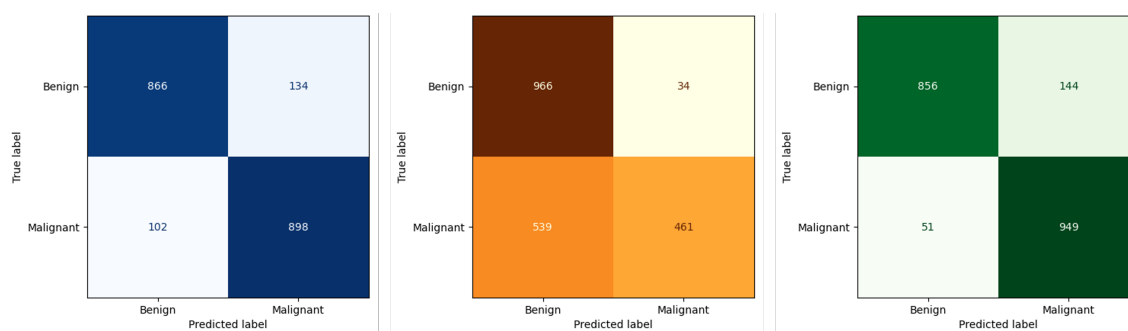


Figure 1: Confusion matrices for the NN, logistic regression, and CNN models

accuracy.

## 6 Conclusions

In this work, we developed deep neural networks for classifying skin lesions from dermoscopic images. A CNN architecture achieved 91% accuracy, outperforming both logistic regression and a standard NN model. Our experiments confirm the ability of deep learning and CNNs to extract informative features from medical imaging data. The models still have difficulty with some challenging cases, as reflected in the precision-recall tradeoff. Future work should expand the CNN with more parameters and training data. Overall, this represents a promising application area for deep learning in healthcare.

## 7 References

- [1] Rogers et al. Incidence Estimate of Nonmelanoma Skin Cancer in the United States, 2006. *Arch Dermatol* 2010.
- [2] American Cancer Society. Cancer Facts & Figures 2020. Atlanta: ACS; 2020.
- [3] American Cancer Society. Survival rates for melanoma skin cancer.
- [4] Kittler et al. Diagnostic accuracy of dermatologists and non-dermatologists. *British Journal of Dermatology* 2002.

- [5] Shen et al. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017.
- [6] Abbasi et al. Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. *JAMA* 2004.
- [7] Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017.
- [8] Matsunaga et al. ImageNet classification with deep convolutional neural networks for dermatologist level classification of skin cancer. *ArXiv* 2017.
- [9] Haenssle et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 2018.
- [10] Bhavesh Mittal. Melanoma Cancer Dataset, Kaggle.