ARIEL UNIVERSITY
COMPUTER SCIENCE

Project submission March 3, 2024

Skin Cancer Detection Using Deep Learning

by

Alon Meshulam

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 80% and 88% 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 112 x 112 x 3 input images into vectors, 3 fully connected hidden layers of 128, 64, and 64 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 convolutional, normalization, and pooling layers structured into repeating blocks for hierarchical feature extraction. Specifically, the network contains two convolutional blocks (input block and block with L layers, in our case L = 2), each consisting of a Conv2D layer followed by batch normalization and max pooling.

The input layer takes 112x112x3 RGB dermoscopic images. The first convolutional block applies 32 filters of size 3x3x3 through a Conv2D layer to extract

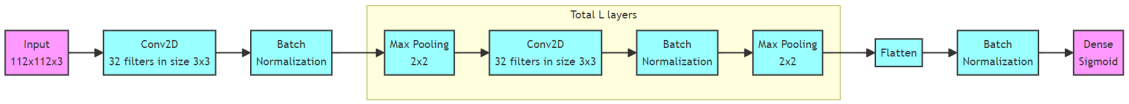

Figure 1: The architectures of the used CNN model

low-level visual features. Batch normalization then normalizes the output to stabilize training. Max pooling reduces spatial dimensions while retaining important activations.

The second convolutional block repeats this process, learning higher-level abstract features with another Conv2D layer. Batch normalization and max pooling are again applied. After two such blocks, the feature maps are flattened into a 1D vector.

An additional batch normalization calibrates the features before feeding into a final Dense layer with sigmoid activation for binary benign/malignant classification. The model compiles with binary cross-entropy loss and the Adam optimizer.

This structured stacking of paired convolution and pooling blocks interspersed with normalization trains robust hierarchical feature extractors. The small 3x3 filter size maintains efficiency while still learning spatially-aware patterns. Overall, the architecture balances model depth, parameter size, and ease of optimization for accurate skin lesion analysis.

All models are trained for 1 epochs with Adam optimization (with 0.0001 learning-rate), 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.

5 Results

Table 1 summarizes the performance of the three models. The CNN achieves the best overall accuracy of 88%, followed by the NN at 80%. The simpler logistic regression model performs noticeably worse with 69% 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.

Model	Accuracy	Precision	Recall	Loss
Logistic Regression	69%	62%	97%	0.312
Neural Network	80%	91%	67%	0.198
Convolutional NN	88%	91%	84%	0.112

Table 1: Model learning curves for training and validation.

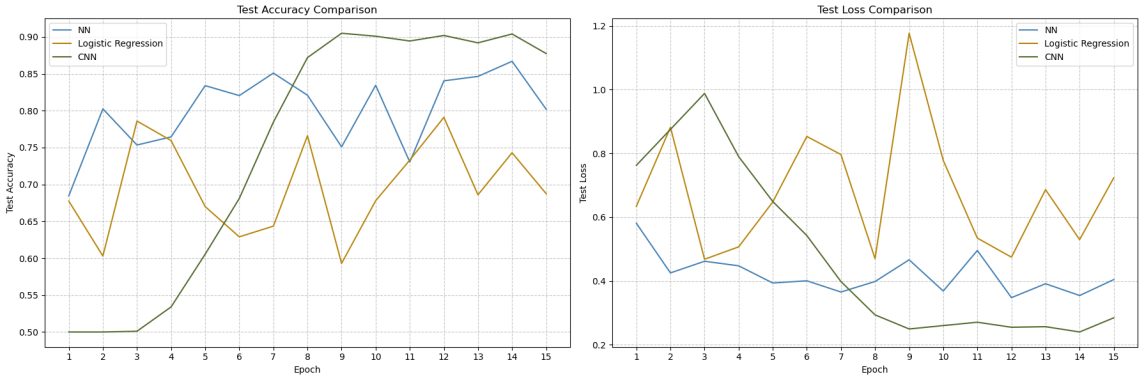

Figure 2: Test Accuracy and Loss Comparison

Interestingly, while the NN achieves lower overall accuracy than the CNN, their precision are the same. The CNN has higher recall (sensitivity) of 84% compared to 67% for the NN. The CNN prioritizes detecting more malignant cases at the expense of more false positives. Depending on the application and time consuming, the NN may represent a better operating point.

The learning curves in Figure 2 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 learning curves show the CNN improving persistently from 50% accuracy to 90%, while the NN and logistic regression are more unstable. The CNN also converges faster within 5 epochs. The plots verify the superior optimization of the CNN via the continual decrease in loss versus fluctuations for the NN and logistic regression.

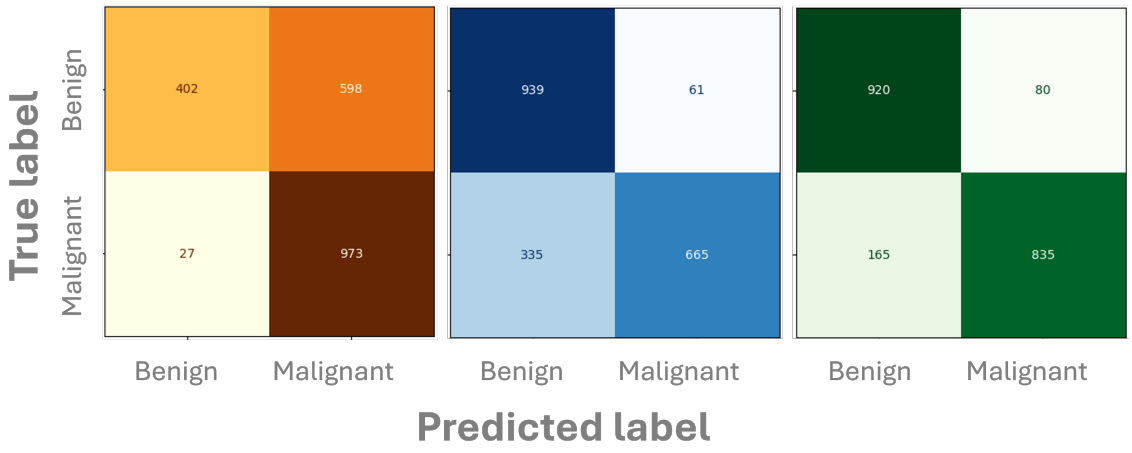

Figure 3: Confusion matrices for the logistic regression, NN, and CNN models

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 939 benign and 665 malignant cases, while mispredicting 61 benign as malignant and 335 malignant as benign. Logistic regression substantially underpredicts malignancies, with 599 false negatives. The CNN achieves 835 true malignant predictions but 80 false malignant. The confusion matrices thus reveal nuances in the types of errors beyond the overall accuracy.

6 Conclusions

In this work, we developed deep neural networks for classifying skin lesions from dermoscopic images. A CNN architecture achieved 88% 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.