Skin Cancer Detection and Classification

Shivani Choudhary
Computer Sciences
Undergraduate
choudhary26@wisc.edu

Michael Chen
Computer Sciences
Undergraduate
mchen429@wisc.edu

Jay Sangani
ECE
Undergraduate
jsangani@wisc.edu

Abstract

This project aims to use a convolutional neural network for the classification of images of skin lesions. The images in the set will fall under and will be sorted into 10 categories: actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis (seborrheic keratosis), squamous cell carcinoma, and a control group of noncancerous lesions. The model will be trained and tested on a combination of two datasets from the International Skin Imaging Collaboration (ISIC), with images classified as benign or malignant, along with the type of cancer (if malignant).

1. Introduction

Skin cancer affects 20% of the US population - by the age of 70, 1 in 5 Americans will suffer it [1]. Patients often visit a dermatologist with a skin lesion like a lump, spot, or mole. It is then up to the dermatologist's clinical expertise [2] to decide whether further testing like a biopsy is needed. While a human in the loop (in the best case, an experienced dermatologist) is vital, the use of an algorithm to diagnose skin cancers can not only improve patient outcomes, but democratize healthcare for those without access to specialized care.

Backing up a clinical diagnosis with algorithmic classification can increase the accuracy of skin cancer detection, and perhaps even catch cancers earlier, which makes a tremendous difference in treatment. In rural, underdeveloped, or underprivileged settings, an algorithmic classifier could make an enormous difference - perhaps even between life and death [2] - since specialists are hard to come by for these communities.

1.1 Related Work

Other studies have been done in this vein, using datasets like the HAM10000 dataset[1]. In preprocessing the data, the images were resized, normalized, and augmented. The model had several layers - four Conv2D layers, MaxPooling, Flatten, and Dense layers. It was trained using the Adam optimizer and categorical cross-entropy loss over 50 epochs. Convolutional neural networks outperformed existing models, yielding an accuracy of 97.78%, while other models score below 95%.

2. Method

To achieve our goals we eventually went with the ResNet50 CNN architecture. Compared to most CNN which is already fit for our data, ResNet50 has much better performance and accuracy. It also uses residual calculation where it will skip certain layers and prevent vanishing gradients. For a project that requires low error rates the ResNet50 is perfect. The architecture can be made with torch or with a pretrained ResNet50 from torch or another package. We decided to use pretrained ResNet50 due to our limited dataset and its higher accuracy. To maximize our experiment we can always use other models and compare to see which one plays out the best. Starting with ResNet50 will at least give a way to make sure we can build an acceptable model first and then find other ways to predict this data.

3. Experiments and Results

Evaluation/findings to be completed after training and testing.

Data was preprocessed with Resize((244, 244)) to fit ResNet input size, ToTensor(), and Normalize().

Our dataset is a combination of the following two datasets. This dataset contains two folders(Test and Train) which both contain images from the other datasets sorted into two folders(Malignant

or Benign). Vascular lesion images from dataset 2 are excluded from our combined dataset and model because it can either be malignant or benign which is not labeled in the images.

Dataset 1: Malignant vs. Benign

This dataset contains data from the ISIC(The International Skin Imaging Collaboration) archive. It consists of two folders of 1800 images each, respectively featuring either malignant or benign skin moles.

Dataset 2: Skin Cancer ISIC

This dataset contains data from the ISIC(The International Skin Imaging Collaboration) archive. It consists of 2357 images of malignant and benign skin diseases as follows:

- actinic keratosis(benign)
- basal cell carcinoma(malignant)
- dermatofibroma(benign)
- melanoma(malignant)
- nevus(benign)
- seborrheic keratosis(benign)
    - also known as pigmented benign keratosis
- squamous cell carcinoma(malignant)
- vascular lesion(mixed)

4. Discussion

To be completed after training and testing

5. Conclusion

To be completed after training and testing

References

1. "Skin Cancer Facts & Statistics." *The Skin Cancer Foundation*, The Skin Cancer Foundation, 27 Jan. 2025, www.skincancer.org/skin-cancer-information/skin-cancer-facts/.
2. Musthafa, M Mohamed, et al. "Enhanced Skin Cancer Diagnosis Using Optimized CNN Architecture and Checkpoints for Automated Dermatological Lesion Classification - BMC Medical Imaging." *BioMed Central*, BioMed Central, 2 Aug. 2024, bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-024-01356-8.
3. Mayo Clinic Staff. "Skin Cancer." *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 8 Apr. 2025, www.mayoclinic.org/diseases-conditions/skin-cancer/diagnosis-treatment/drc-20377608.

Individual Contribution

Michael - Assembled the combined dataset. Worked on the model(preprocessing, configuring the pretrained resnet50 model). Wrote section 3.

Jay - Worked on researching the best CNN for our data and cause. Landed and resnet50, as well as divided up data to train and test and their labels based on their folder.

Shivani - worked on the model, researched other studies in the area of skin cancer classification, wrote the abstract and introduction