Age-specific Detection of Skin Cancer Using Tree-Based and Deep Learning Models

Tianhong Shen, Weiguo Jiang
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
{t54shen,w33jiang}@uwaterloo.ca

Abstract

Skin cancer is the abnormal development of skin cells and primarily occurs on skin that is overly exposed to the sun. Most skin cancer can be cured easily at its early stages. However, the initial development of it looks no different than a regular mole on the skin, and diagnosis of malignant skin cancer requires visiting the hospital. Therefore, people tend to overlook these small changes on their skin and potentially miss the opportunity for early treatment. In this paper, we present both a tree-based approach and a deep learning approach to predict whether a given dermoscopic image is benign or malignant. A classical machine learning model will be used as a baseline for comparison. In particular, we are interested in fitting our models based on different age groups to assist potential patients better.

1 Introduction

According to Skin Cancer Foundation, skin cancer is "the most common cancer in the United States and worldwide", with "more people diagnosed with skin cancer each year in the U.S. than all other cancers combined" (Skin Cancer Foundation, 2023). At the same time, recent reports have called attention to a growing shortage of dermatologists per capita (Kimball and Resneck, 2008). Of all skin cancers, melanoma is the most dangerous category. Dermatology experts believe that in 2040, there will be half a million melanoma patients in the world, a 62% increase compared with that in 2018. Based on these statistics, skin cancer can be considered a global epidemic. To prevent the adverse consequences of skin cancer, dermatologists emphasize the importance of early treatment which can greatly improve the survival probability, especially for melanoma (See Figure 1) (Cassidy et al., 2022).

In this paper, we aim to detect the existence of malignant skin cancer given a dermoscopic image. The data are collected from the International Skin Imaging Collaboration (ISIC) gallery. Two approaches will be used, namely, extreme gradient boosting (XGBoost) which is tree-based and convolutional neural network (CNN) that is a deep learning algorithm. We will use the result from logistic regression as a baseline for this binary classification problem. A positive label of 1 means that the image is malignant and a negative label of -1 represents benign. Additionally, to better serve different age groups, we augment the data with age information, hoping to achieve a better performance than a generic classifier. Various metrics are employed to measure the performance of models against the baseline.

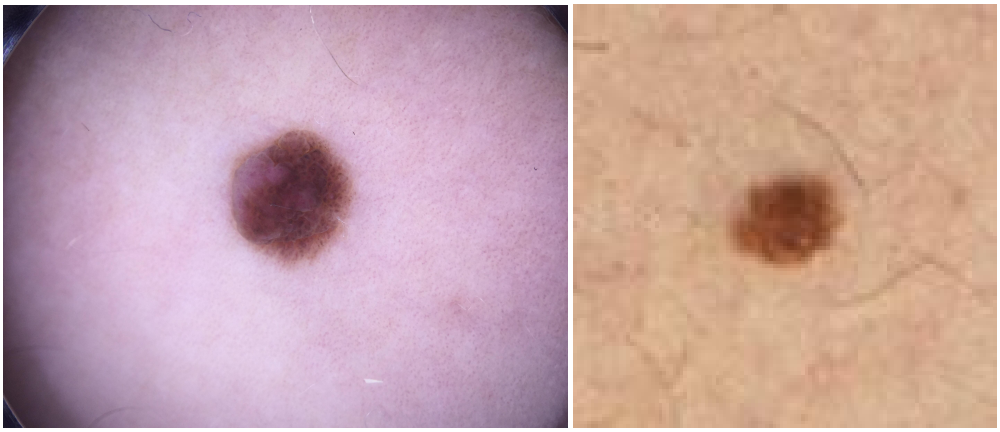

Figure 1: Benign vs Malignant. The left dermoscopic image is benign, and the right dermoscopic image is a malignant skin cancer, melanoma to be specific. It's hard to tell which one looks malignant based on appearance, and this is one of the main reasons why most people with malignant cancer tend to overlook it at the early stage.

2 Related work

As with any kind of machine learning task involving images, pre-processing plays a big role in determining how good the performance of models can be. Dermatologists have their own set of rules for making a diagnosis. In this paper, two commonly used image-processing techniques for detecting skin cancers are used, namely ABCD and GLCM.

Asymmetry, Boundary irregularity, Color, Diameter (ABCD) method is the standard method for any dermatological applications. These four attributes of a skin lesion characterize the symptoms and aid in the detection of skin cancer. Asymmetry quantifies how two halves of an image differ; boundary irregularity measures the unevenness of the image; color is obtained by averaging the intensity of the three channels; and the diameter of the skin lesion is measured. The method of finding these parameters is termed the ABCD method (Monika et al., 2020).

Grey Level Co-occurrence Matrices (GLCM) is "one of the earliest methods for texture feature extraction", introduced by Haralick and his colleagues in 1973. Each entry (i, j) of the matrix represents the number of times a pixel of intensity i is adjacent to a pixel of intensity j (V. et al., 2012). This method is applied extensively in real-world problems, and it is still one of the most powerful feature extraction techniques in the field of texture analysis.

International Skin Imaging Collaboration (ISIC) hosted challenges over the years. Most of them focus on classifying the categories of skin cancer, such as melanoma, basal cell carcinoma (BCC), and Squamous cell carcinoma (SCC) (ISIC Archive, a). This led to the publication of several papers including the classification of malignant skin cancer into 8 different categories by M. Krishna Monika, et al.; on ISIC 2019 Challenge dataset, skin lesion detection towards melanoma by Noel C. F. Codella, et al.; on ISIC 2017 Challenge dataset, and B. Cassidy, et al created models to distinguish melanoma or non-melanoma (V. et al., 2012; Codella et al., 2017; Cassidy et al., 2022). Though our research on papers may not be exhaustive. However, after searching on common publication websites like arXiv, Google Scholar, and ResearchGate, we found that no work attempted to give a generic prediction on whether the dermoscopic scan is benign or malignant, let alone specializing with respect to ages.

3 Execution plan

3.1 Data collection

The dataset used in this paper is collected from ISIC's gallery. All images are RGB with 3 channels and in jpg format. However, not all of them have the same dimension (i.e., length and width). All images are manually labeled with many attributes including gender, benign or malignant, and approximate age. Several filters are available to assist researchers in retrieving customized data

(ISIC Archive, b). Since we are specializing in detecting malignant skin cancers with respect to age, we mainly utilize the BENIGN OR MALIGNANT filter and APPROXIMATE AGE filter for data retrieval. We will randomly select pictures in the gallery, with approximately 98% of benign and 2% of malignant as our training dataset. For the test dataset, we would increase the proportion of malignant to better challenge our models.

3.2 Data pre-processing

Firstly, we will try to remove bias in images, such as duplicates. After removing it, we can minimize the impact of noise on model accuracy. Next, the main feature engineering methods we will use are ABCD and GCLM. They are all used to extract numeric features from images since logistic regression and XGboost cannot be applied to images directly. Based on papers we have read, we found that both ABCD and GCLM are the main aspects used to distinguish benign or malignant.

3.3 Methodology

In general, we will apply XGboost and Convolutional neural network to these images. Meanwhile, we will use the result from logistic regression as the baseline, so that we can make comparisons between models.

3.3.1 Logistic regression (baseline)

Logistic regression is the fundamental model in dealing with binary classification. Packages can be applied directly to find the optimized parameters. However, due to its inherent inability to extract patterns and interpret features from raw images, logistic regression performs poorly on high-dimensional, hierarchical data.

3.3.2 Extreme gradient boosting

Extreme gradient boosting, also known as XGboost, is one of the most powerful models in machine learning. It was invented by T. Chen and his colleagues in 2016, and it showed a strong performance on tabulated and structured data (Chen and Guestrin, 2016). In this paper, we have a high expectation on it.

3.3.3 Convolutional neural network

Convolutional neural network (CNN) is an advanced image recognition method in machine learning, introduced by A. Krizhevsky, et al in 2012 (Krizhevsky et al., 2012). CNN can be directly applied to raw images, which is different from the two models we just mentioned. Till this day, it is still one of the most popular deep learning models for image processing tasks such as object detection, image classification, and facial recognition.

4 Evaluation

We will use several criteria to measure the performance of each model. Accuracy is the most common criterion which measures the proportion of images that are correctly classified.

Since most samples in our training set are benign, the accuracy is high if the model always predicts benign. Thus, we also introduce the F-measure, which ignores the true negative (TN) and places emphasis on how the model performs on positive instances. The formula is

$$F = \frac{2 \times TP}{2 \times TP + FP + FN}$$

where TP stands for true positive, FP refers to false positive, and FN means false negative.

At the same time, sensitivity and specificity are reasonable criteria to include as well. Sensitivity measures the proportion of malignant skin cancer detected, with the formula

$$\textit{sensitivity} = \frac{TP}{TP + FN}$$

Specificity calculates the proportion of benign identified, with the formula

$$\textit{specificity} = \frac{TN}{TN + FP}$$

Furthermore, Receiver Operating Characteristic (ROC) and Area under Curve (AUC) can be derived based on sensitivity and specificity.

To prevent the risk of misclassification, we will also give a confidence score which represents the probability of benign for each image. Thus, we can use logarithmic loss to measure the overall performance of each model. Its formula is

$$\textit{logarithmic loss} = \frac{1}{N} \times \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} \log(p_{ij})$$

where N represents the total number of images, and M stands for the total number of classes (benign and malignant). $y_{ij} = 0$ if image i belongs to class j; otherwise, $y_{ij} = 1$. Moreover, p_{ij} refers to the probability of class j for image i.

5 Expected outcome

In this paper, results from the logistic regression are the baseline for the other models. In general, XGboost and CNN are expected to have a better if not dominant performance over logistic regression. With its superior performance on tabulated data, we believe that XGboost will work well if the pre-processing steps effectively capture key features. We also have a high expectation on CNN due to its ability to reduce the high dimensionality of images via convolutional layers. Based on papers we have surveyed, we hope to achieve prediction accuracy of over 80% on testing data. One worry we have about CNN is whether our dataset is large enough for it to extract all meaningful features. Due to the number of layers and parameters that a CNN has, we fear it may underfit.

References

Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., and Yap, M. H. (2022). Analysis of the isic image datasets: Usage, benchmarks and recommendations. *Medical Image Analysis*, 75:102305.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754.

Codella, N. C. F., Gutman, D. A., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N. K., Kittler, H., and Halpern, A. (2017). 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). *CoRR*, abs/1710.05006.

ISIC Archive. Isic archive challenges. `https://challenge.isic-archive.com/challenges/`. Accessed: 2024-10-08.

ISIC Archive. Isic archive gallery. `https://gallery.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery?filter=%5B%5D`. Accessed: 2024-10-08.

Kimball, A. B. and Resneck, J. S. J. (2008). The us dermatology workforce: a specialty remains in shortage. *Journal of the American Academy of Dermatology*, 59(5):741–745.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90.

Monika, M. K., Arun Vignesh, N., Usha Kumari, C., Kumar, M., and Lydia, E. L. (2020). Skin cancer detection and classification using machine learning. *Materials Today: Proceedings*, 33:4266–4270. International Conference on Nanotechnology: Ideas, Innovation and Industries.

Skin Cancer Foundation (2023). Skin cancer facts. Accessed: 2023-10-08.

V., B. S., Unnikrishnan, A., and Balakrishnan, K. (2012). Gray level co-occurrence matrices: Generalisation and some new features. *CoRR*, abs/1205.4831.