

# Skin Cancer Detection, Classification and Modeling

## DTSA 5506: Data Mining Project

Abstract- One of the most common yet highly curable cancers in the United States is skin cancer. Due to the high rates of treatability, harnessing the latest advancements in deep learning algorithms, in particular convolutional neural networks (CNN) for image classification in early detection, is crucial. Being able to distinguish the unique patterns and characteristics of the three most common skin cancers (Basal carcinoma, Squamous cell carcinoma and Melanoma) is what could make this technology a good fit. Many groups have already had great success in predicting the presence of cancer accurately in skin lesion images. Challenges still exist in the lack of large datasets to train these models and the ability of models to distinguish cancer types more accurate than current clinical methods. In this project I explored CNN models to classify skin cancer images using a newly released larger dataset from the International Skin Imaging Collaboration (ISIC). I evaluated my models using Area Under the Curve and Accuracy metrics. I was able to obtain decent results in the detection model with an AUC value of 0.92. I was unsuccessful in creating a model that could classify skin cancer type with high accuracy. My best model only achieved a little over 50 percent accuracy on a validation dataset.

### 1. INTRODUCTION

Skin cancer is the most common cancer in the United States with over 6 million people being treated per year. It most commonly occurs from cellular damage to the skin due to ultraviolet radiation from the sun. When left undetected, it can spread through the body and become more deadly.

The three main types of skin cancer are Basal cell carcinoma, Squamous cell carcinoma and Melanoma. Basal cell carcinoma is the most common type of skin cancer. It forms in the lower epidermis. It is often round and flesh-colored. Squamous cell carcinoma is the second-most common type of skin cancer and forms on the outside layer of skin. It often appears as a red bump or scaled patchy skin that does not heal. The final type of common skin cancer is Melanoma. It forms in the bottom layer of the epidermis. It looks like a dark patch on the skin or on a mole. Melanoma is the most deadly form of skin cancer because it has a tendency to rapidly spread to other tissue before it can be detected.

Fortunately, skin cancer is very curable if detected early. For example, when detected early, the five year survival rate of Melanoma is 99 percent [4]. Detection of skin cancer is first done with a visual inspection, usually in a doctor's office. Visual inspection by a physician has about a 60 percent accuracy rate [1]. The use of dermoscopy with a

visual physical diagnosis has an accuracy rate of 89 percent [1].

With advancements in deep learning models for image classification, this new technology could serve a beneficial purpose in the fight to save lives and prevent cancer. These models could lend themselves well to skin cancer detection and classification because of their complex ability to learn and track spacial features and patterns. The three different types of common skin cancers present with distinct appearances and patterns from each other and from benign skin lesions.

Having an accurate way to detect skin cancer from an image could be especially useful for people with limited time or access to healthcare. An image classification tool could enable a patient to send a smartphone picture of a suspected lesion to a healthcare provider for quicker detection, reducing the need for an in-person appointment and potentially lowering the cost of initial diagnosis.

In this data mining project I first used preprocessing techniques to understand the data and prepare it for modeling. I then explored how machine learning algorithms for image classification can be implemented to help solve the problem of detecting skin cancer and classifying types of skin cancer. I implemented two models. First, I developed a model that detected whether a skin lesion image was cancerous or not. My second model attempted to classify the skin cancer in the image. I used Area Under the Curve as a performance metric in the binary case (cancer vs not-cancer) and Accuracy as the performance metric in the classification case.

### 2. RELATED WORK

Many groups have modeled skin cancer detection with deep learning algorithms and have had good success in creating accurate models mostly using convolutional neural networks to detect cancerous lesions. Abbas and Gul used NASNet to achieve 97.7% accuracy in detection. Adegun et al. used a fully connected CNN to achieve 97% accuracy in detection. Khan et al. used ResNet to achieve 98.7 % accuracy in detection. In 2022, Maniraj and Maran used VGG architecture to achieve 93.3% accuracy in detection and Rashid et al. used MobileNet V2 to achieve 98.2% accuracy in detection [1].

Although current algorithms do well with detecting skin cancer, they do not perform as well with classifying types of skin cancer as compared with dermoscopy [1]. One problem is that past research has relied on smaller datasets to build and train models. These models may not translate well to large scale applications. Early models used dermoscopic images from the PH2 dataset which contains only 200 skin lesion images.

Other downsides with past models include efficiency and ability to use models in real-time. Another major downside is that most models were trained on lighter skin colors and are even less accurate with darker skin colors. To help advance the potential of deep learning models to more robustly classify skin cancers, the International Skin Imaging Collaboration (ISIC) continues to publish larger datasets.

### 3. PROPOSED WORK

In my project, I have used a new dataset provided by the International Skin Image Collaboration group on Kaggle in 2024. The training dataset has over 400,000 images of skin lesion samples. In addition to this, they have also provided a .csv file with annotation information indicating which images are cancerous, as well as the type of skin cancer and

Feature	Description
isic_id	Unique Identifier
target	Target Classifier 1: Malignant 0: Benign
iddx_3	Third level lesion diagnosis
age_approx	Patient Age
sex	Patient Sex
clin_size_long_diam_mm	Maximum diameter of the lesion in mm
tbp_lv_H	Hue inside the lesion
tbp_lv_Hext	Hue outside lesion
tbp_lv_areaMM2	Area of lesion ( $\text{mm}^2$ )
tbp_lv_deltaLNorm	Contrast between the lesion and the surrounding skin
tbp_lv_eccentricity	Eccentricity
tbp_lv_norm_border	Border irregularity (0-10 scale)
tbp_lv_norm_color	Eccentricity
tbp_lv_perimeterMM	Perimeter of lesion (mm)

Figure 1: Example Features of Annotated File

other information. In total, there were 55 features in the dataset from the .csv file. Some example features are summarized in Figure 1.

I performed an exploratory data analysis to gain knowledge of the dataset. As expected, there was a very large imbalance in the dataset with most of the images classified as

“benign”. I calculated an imbalance ratio of over one thousand. This imbalance makes sense, since it is more common for lesions to not be cancerous, and the different types of cancers occurring at different rates. I used the Python Imbalanced Learning Package to created a “majority class” under-sampled dataset.

I loaded the images into a new dataset as an array of pixel-values and performed pre-processing. The images were different sizes so I resized and scaled. I used the Keras Image Data Generator class for image augmentation techniques. Image augmentation can be used to better train the model by creating different representations (rotations, shifts, flip, etc.) of the image.

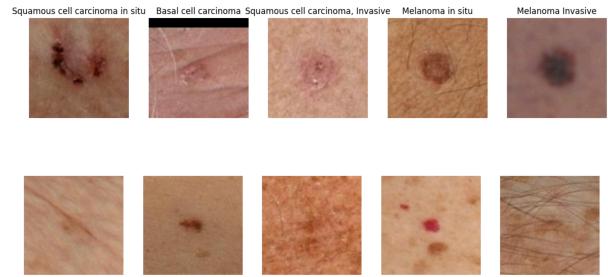


Figure 2: Example Images of Malignant (top) and Benign (bottom) skin lesions

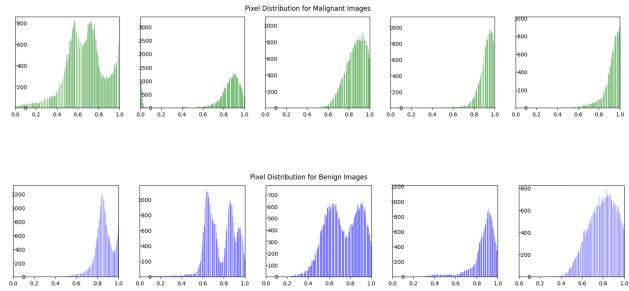
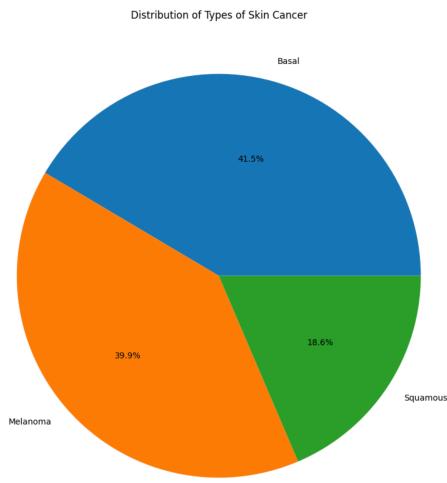


Figure 3: Histogram of Pixel Intensity for Malignant (top) and Benign (bottom) skin lesions

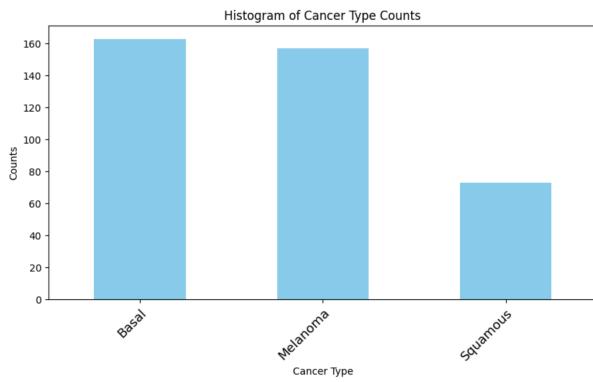
I viewed some example images (Figure 2) along with pixel histograms (Figure 3). The pixel histograms show the number of pixels for each intensity value. It was a little difficult to see trends or make generalizations from these images and distributions.

I created a new dataset of only malignant images and viewed the distribution of the different types of cancer represented in these images. This is shown in Figures 4 and

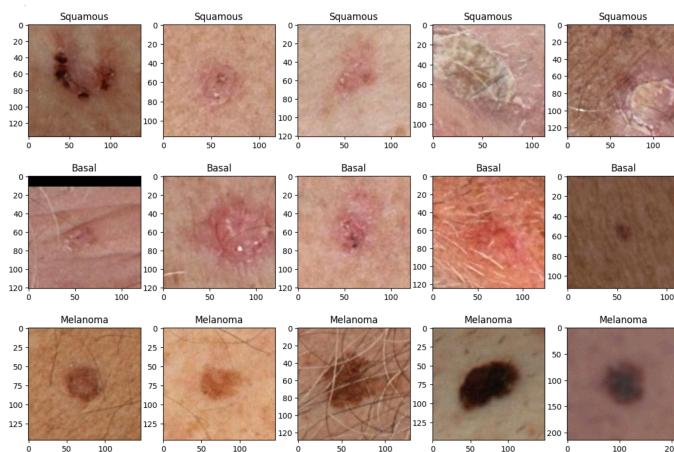


**Figure 4: Pi Chart of Cancer Types in Dataset**

5. I also viewed some more example images to compare the appearances of the different types of skin cancer. Figure 6 shows these examples. Looking at the images, it's easier to see differences between Melanoma and the other two. The



**Figure 5: Distribution of Cancer Types in Dataset**



**Figure 6: Example images of Squamous, Basal and Melanoma Skin Cancer**

images classified as Squamous and Basal appear more similar.

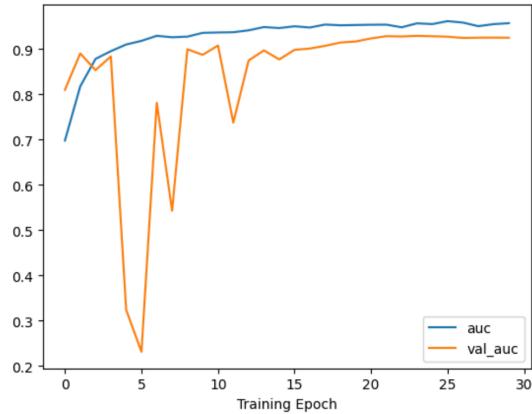
I created my convolutional neural network models using the Tensorflow Keras library package. I used a convolutional neural network due to its ability to learn spatial features and patterns in images. My model consists of:

- An input layer corresponding to the pixel dimensions of the input images.
- Three convolutional layers used with filters to learn patterns and visual elements
- Three Max Pooling layers to reduce dimensionality while keeping important information
- Two fully connected layers with a rectified linear unit activation function and flattening and batch normalization
- A dropout layer to prevent overfitting
- And finally a fully connected layer for the output with size 3 for the three distinct skin cancer classes. Here I tried using sigmoid and linear activation function.

I used binary and categorical cross-entropy as loss functions with adaptive moment estimation optimization. I created automatic learning rate adjustments.

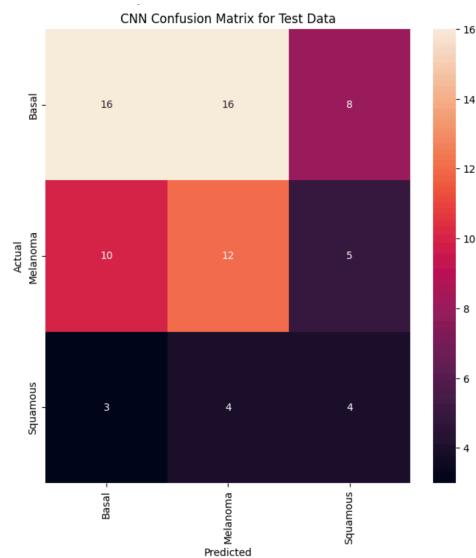
#### 4. EVALUATION

I split the data into training and testing datasets. The training set was used to train the model; the test dataset was used to validate model performance and evaluation. My first model was a binary classifier to detect whether the image was cancerous or not. To evaluate my first model, I used Area Under the Curve (AUC) to determine how well the model classified the images in a binary situation (cancerous vs not cancerous). AUC is a graphical representation and measure of the probability that the model can distinguish between two targets, commonly used in healthcare applications to establish diagnoses (for example, “disease present” or “not present”). Systems for diagnosis are required to be highly sensitive with AUC values above 80%. In my initial approach I used the resampled dataset with a 1:1 ratio of benign:malignant images. With this distribution, my model achieved an AUC score of 0.82. I resampled the data a second time with a 5:1 split and my model achieved an AUC score of 0.92 in detecting cancer in the test image dataset. Figure 6 shows this model's performances of AUC as a function of training epoch number. I set learning rate adjustments in the model which can be seen as the initial instability, but the model starts to converge to a validation AUC value of 0.92 at about 20 training epochs.



**Figure 7: Resampled Binary Classifying Model AUC by Training Epoch**

For my second model, in attempting to classify the type of cancer in the image I used Accuracy to determine how well the model performed at classifying the different types of skin cancer. Accuracy describes how well a model performed across all classes. Unfortunately, I was unable to produce a model that obtained high accuracy rates. My best model was only able to achieve about 50% accuracy. Figure 8 shows results for this model for over 100 training epochs. I examined a confusion matrix (Figure 9) to understand how images were getting classified.



**Figure 9: Confusion Matrix for Classifier Model**

## 5. DISCUSSION

The following was my approximate project timeline:

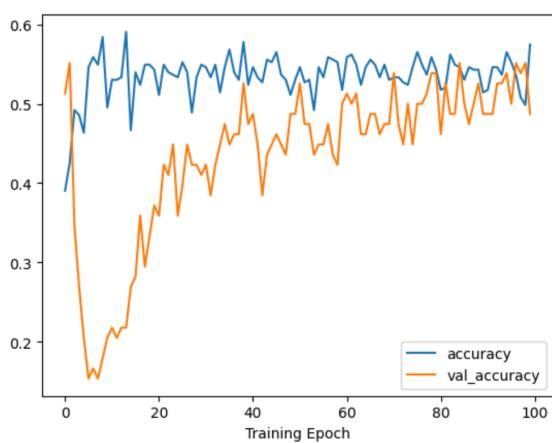
### Week 1 (November 25-29):

- Download image data - 2 hrs.
- Clean process and explore the image data (Exploratory Data Analysis) - 4hrs.
- Create models and use these models to make predictions - 20 hrs.
- Model Parameter Tuning - 4 hrs.

### Week 2 (December 2 - 6):

- Finalize Conclusions - 2 hrs.
- Write Report - 8 hrs.
- Create Presentation - 4 hrs.
- Practice and Record Presentation- 2 hrs.

From my research, past models have performed very well with the binary aspect but are less accurate with classification. This was seen in my project as well. The model I built to detect whether an image was cancerous or not, performed very well. It achieved an AUC value of 0.92 on test images. My model for classifying skin cancer type was around 50% accurate on test data. After observing sample images, I hypothesized that it would be able to classify the Melanoma cases correctly but have trouble with the other two type of skin lesions. The Melanoma skin lesions appeared darker and more distinct. However, looking



**Figure 8: Classifier Model Accuracy by Epoch**

at the confusion matrix, this trend does not appear to be the case. In fact, the Basal skin lesions samples had a higher percentage of correct classification. This makes me think the model could be overfitting and there could be errors in the model setup. A major problem was that the dataset was very small, which makes it hard to produce an accurate model.

Some future work may involve researching other approaches, changing the algorithm or architect used. Another idea could be to use synthetic image generating to create more training images.

After examining the dataset from the annotated .csv file, I also thought it may be interesting to create a model from this information using supervised learning approaches such as Random Forest Classifiers. This file contained a bunch of information such as lesion dimensions and characteristic, as well as patient age. These features could be used to train a classifying model.

## 6. CONCLUSION

Skin cancer is a very common type of cancer in the United States and worldwide. However, with early detection, it is highly treatable. A key to successful early detection lies in the development of reliable tools to efficiently and accurately detect skin cancer. Auto-detection tools will be more and more useful in the future with an aging population since most skin cancer cases are detected in patients 60 years and older. The advancement of computer vision and image classification makes this an ideal application of the technology.

In this project I built two deep learning CNN models with recently released data from ISIC in order to detect and classify skin cancer. My model for detecting cancer in the images performed well, however, I was unable to obtain high results with my classifying model.

This project provided a meaningful way to go through the data mining pipeline process and also apply the advancements of deep learning in image classification in order to gain insight on the ability of these algorithms to be useful tools in the diagnosis and prevention of skin cancer.

## REFERENCES

1. Naqvi M, Gilani SQ, Syed T, Marques O, Kim HC. Skin Cancer Detection Using Deep Learning-A Review. *Diagnostics (Basel)*. 2023 May 30;13(11):1911. doi: 10.3390/diagnostics13111911. PMID: 37296763; PMCID: PMC10252190.
2. ISIC 2024 - Skin Cancer Detection with 3D-TBP. (2024). @Kaggle. <https://www.kaggle.com/competitions/isic-2024-challenge/overview>
3. Skin Cancer Detection using TensorFlow. (2022, November 20). GeeksforGeeks. <https://www.geeksforgeeks.org/skin-cancer-detection-using-tensorflow/>
4. Skin Cancer Foundation. (2024, February). *Skin Cancer Facts & Statistics - The Skin Cancer Foundation*. The Skin Cancer Foundation; The Skin Cancer Foundation. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>