**Arash Mahmoudian**
**Project: Skin Cancer Diagnosis**

## Introduction:

D is a disease that occurs when abnormal skin cells grow out of control. It's the most common cancer in the United States. Skin cancer is caused by damage to skin cells, which can be caused by:

- Ultraviolet (UV) radiation from the sun, tanning beds, or sunlamps
- Unrepaired DNA damage

Skin cancer can develop on skin exposed to the sun, but it can also occur on areas of skin not normally exposed to sunlight.

A dermatologist can diagnose skin cancer with a skin biopsy. A skin biopsy is the only way to know for sure if someone has skin cancer. In this procedure, the dermatologist removes a sample of suspicious skin for lab testing. The biopsy can determine if someone has skin cancer and what type of skin cancer they have. Because of the high occurrence of skin cancers AI/ML models of cancer diagnosis are becoming more powerful tools for dermatologists.

For the first milestone of this course a skin tumors image dataset holding 10015 images has been selected. This dataset is available in Harvard.edu Dataverse as described below:

1- Number of cases: 10015.
2- Number diagnosis classes: 7
3- Patients ages ranges from 1 to 85
4- Patients sex: Male: %54.0, Female: %45.5, and Unknown: %0.5
**5-** Image resolution is width: 600 pixels and height: 450 pixels.

## Objective:

This study intended to offer a practicable deep learning-based method for the skin cancer analysis in lesion images, to help physicians in diagnosis.

## Analysis:

Figure [1] shows 18 skin cancer images. These cancerous parts of skin vary by shape and color.

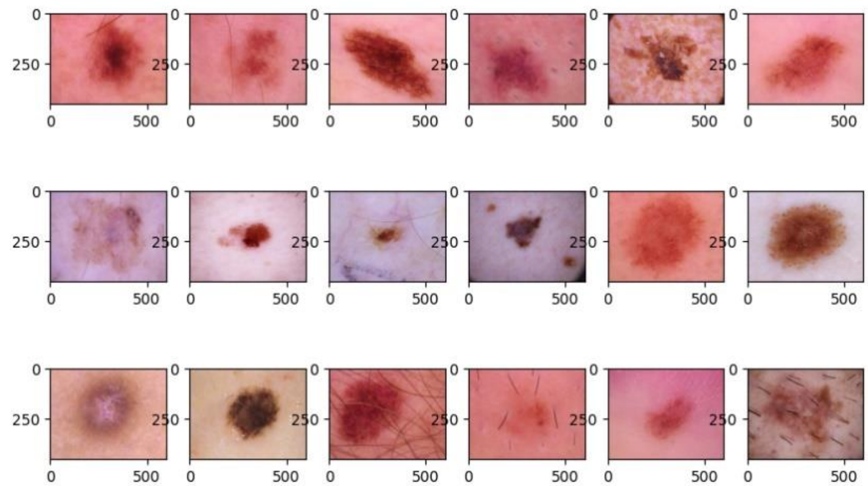


Figure [1]

Patients age distribution histogram shows age is ranging from 1 to 86 illustrating samples from almost all ages. From age distribution point of view the dataset is a decent depiction of reality.

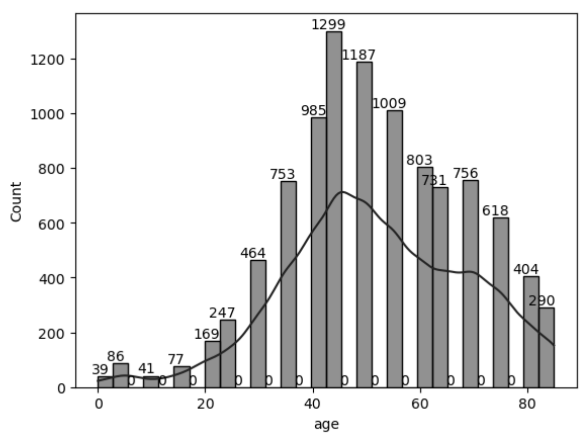



Figure [2]

Figure [3] shows the distribution of skin disease diagnosis classifications. Most of cases have been diagnosed as “**nv**”, and “**mel**” & “**bkl**” take second and third places of diagnosed diseases.

1- **nv**: Melanocytic nevi,
2- **mel**: dermatofibroma
3- **bkl**: Benign keratosis-like lesions
4- **bcc**: Basal cell carcinoma
5- **akiec**: Actinic keratoses
6- **vasc**: Vascular lesions
7- **df**: Dermatofibroma

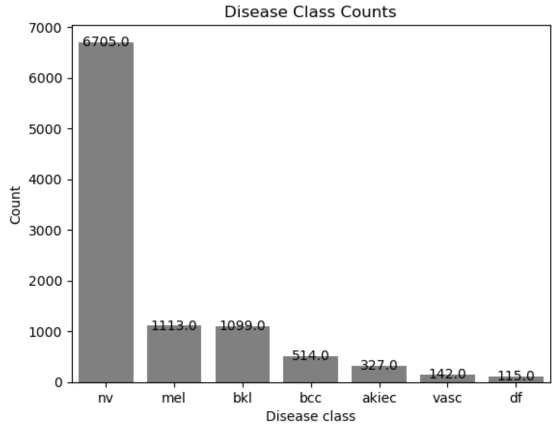



Figure [3]

Figure [4] shows that the number of males patients were more than female patients in the dataset. This needs investigation to determine if males do get more skin cancer than females or it was just a sampling issue.

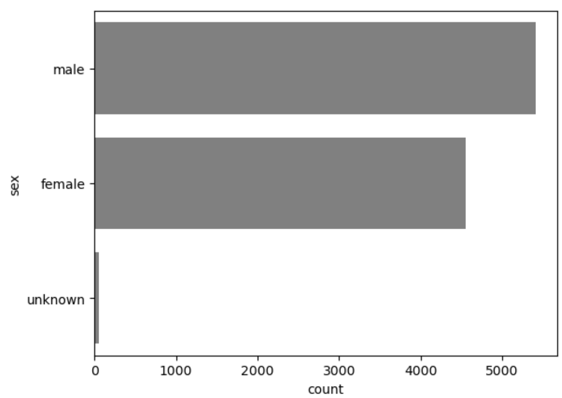



Figure [4]

**Methods:**

To classify the disease, type the cancerous images were passes through a sequential Keras CNN deep learning model in 20 epochs. Out of these 10015 images %80 was used to train the model, and the remaining %20 was used for the validation and evaluation purpose. Original images resolution was (600*450) pixels which was reduced to (32* 32) pixels to speed up the training process. Convolution size was set to (3*3) and activation to “relu”. Additionally, padding was also used to allow for more space for the convolution filter to cover in the image. Moreover, Batch normalization is used between each layer of neural network which essentially sets the pixels in all feature maps in a convolution layer to a new mean and a new standard deviation. This makes training of artificial neural networks faster and more stable through normalization of the layers' inputs by re-centering and re-scaling. Next, max pooling is also used for object recognition tasks, as it helps to identify the most distinctive features of an object, such as its edges and corners.

**ReLU** stands for Rectified Linear Unit. It is a type of activation function that is commonly used in neural networks. ReLU is a non-linear function that has the following formula:

$$\text{RELU: } f(x) = \max(0, x)$$

**Model Evaluation:**

After passing images through the CNN neural network maximum achieved training accuracy was %83.77 whereas the maximum achieved validation accuracy was %76.86. Similarly, the minimum achieved training loss was 0.45 and validation loss was 0.65. Figure [5] shows model performance for both training and validation after each epoch.

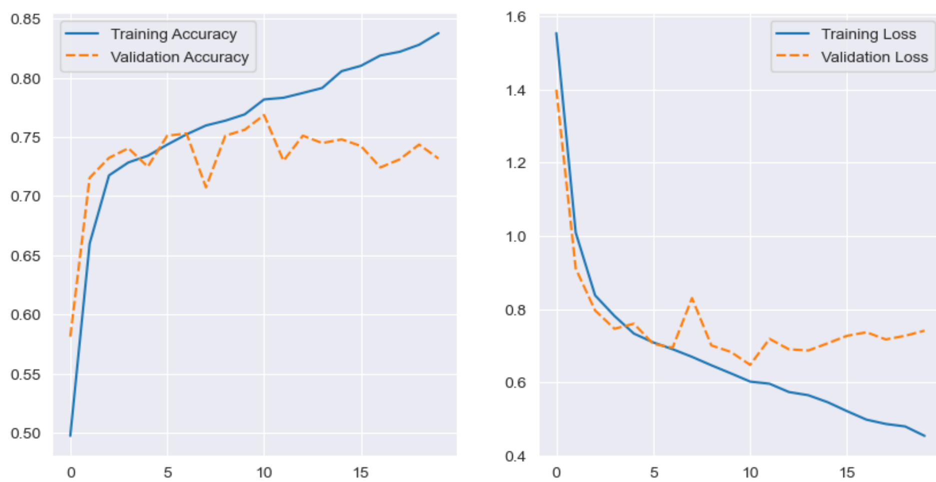
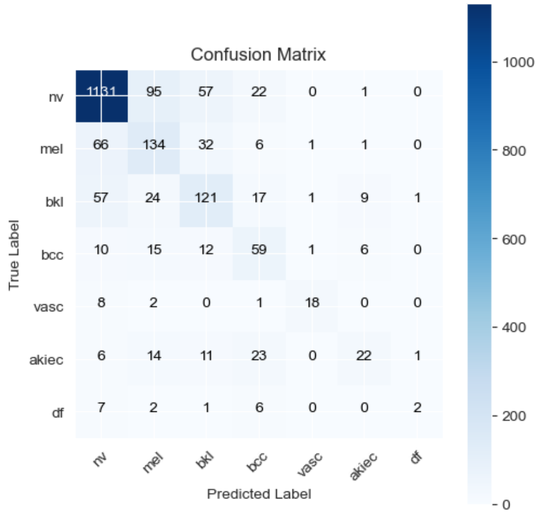



Figure [5]

After Epoch number 10 even though the training accuracy shows improvements, the model gets overfitted and no improvement on validation accuracy [%76.86] can be observed. Likewise, the validation loss weights show no decrease after 10th epoch. Furthermore, model evaluation shows that overall test accuracy the CNN model has is %74.2 and performs well on “nv” class (%87) but poorly on the remaining classes. In addition, confusion matrix Figure [6] shows that the model has the best accuracy of %86 for disease “**nv**” and the worst %11 for the disease “**df**”. Overall, the model performs an acceptable range for only **vasc** and **nv** classes (more than %60), and for the rest not an acceptable range.




| Class | Accuracy |
| --- | --- |
| nv | 87% |
| mel | 56% |
| bkl | 53% |
| bcc | 57% |
| vasc | 62% |
| akiec | 29% |
| df | 11% |

Figure [6]

This difference arises from insufficient image samples for various disease classes in the dataset. A supplementary method was designed to overcome this problem by generating augmented images considering the following specifications in Figure 7.

```
datagen = ImageDataGenerator(rotation_range=45,
                             shear_range=0.2,
                             zoom_range = 0.2,
                             width_shift_range=0.2,
                             height_shift_range=0.2,
                             horizontal_flip=True,
                             fill_mode="reflect")
```

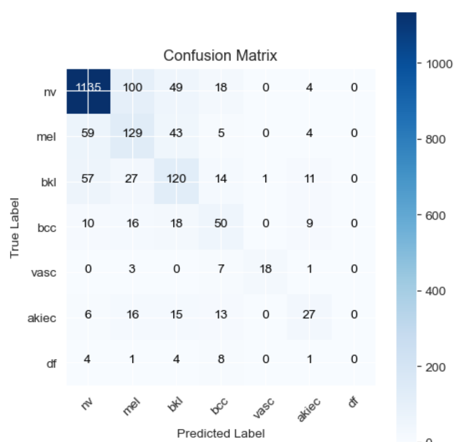
Figure [7]

**Model (Augmented images) Evaluation:**

The maximum achieved training accuracy was % 74.55 whereas the maximum achieved validation accuracy was % 73.83. Similarly, the minimum achieved training loss was 0.67 and validation loss was 0.68. Figure [8] show the model performance for both training and validation after each epoch.


Figure [8]

After epoch number 5 even though the training accuracy shows improvements, but the model gets overfitted and no improvement on validation accuracy [%60.71] can be observed. Likewise, the validation loss weights show no decrease after epoch number 5. Furthermore, the confusion matrix Figure [9] shows that overall test accuracy of the model is %60 and has the best accuracy of %100 for disease class of “**bcc**” and the worst of %0 for the disease class of “**vasc**”, “**akiec**”, and “**df**”. Overall, the model doesn’t seem to be reliable when augmented images are passed through.




| Class | Accuracy |
|---|---|
| nv | 87% |
| mel | 54% |
| bkl | 50% |
| bcc | 27% |
| vasc | 62% |
| akiec | 35% |
| df | 0% |

Figure [9]

**Model comparison:**

Performance comparison for the given two approaches shows that the augmented images approach not only doesn’t improve the accuracy but also performing worse. As compared to original images the accuracy drops from %74.2 to %73.83. This indicates that the image augmentation was not helpful in this case. In Addition, to overall model accuracy problem, the augmented method looks extremely unreliable for the class of df.

**Conclusion:**

Skin cancer is diagnosed by a dermatologist using skin biopsy. A skin biopsy is the only way to know for sure if someone has skin cancer. This approach is time consuming and costly. To speed this process and reduce the expenses a neural network was trained using cancerous images dataset to help the printaniers diagnosing possible diseases. This dataset only has skin samples (images) from seven different cancerous diseases and there are no healthy skin images into the dataset. Therefore, the classification was between seven disease categories and not healthy or cancerous determination. There are enough samples from all age ranges varying from early ages to old ages, plus samples for both sex genders (Female/Man). This indicates that the outcome of the study can be extrapolated to board as there are enough samples from both genders and ages. After passing images through a CNN deep learning model overall maximum test accuracy of %74.2 was achieved, however the accuracy level varies on various categories depending on their available samples in the dataset. The more sample of a class the more accuracy level can be achieved.

Since there were a smaller number of samples for some classes a secondary approach of image augmentation was designed to overcome this issue. In this method classes with smaller number of samples were selected to generate augmented images for them. All images we passed through the same CNN model and the result were not promising as the overall maximum achieved test accuracy was reduced from %74.2 to %73.83. The worst part of this exercise was that the model accuracy for classe of "**df**" was zero indicating that the secondary approach was not reliable for cancer classification.

Finally, it is suggested to gather more samples especially for the classes that there are not enough samples to extrapolate the conclusion and reduce the dataset skewness. This empowers the model to be more robust and increase its accuracy.

**Assumption:**

The images gathered from all races and regardless of a patient's race is the cancerous image shape doesn't change.

**Limitations:**

There are a smaller number of sample/images for some cancerous skin diseases this makes the dataset be skewed and biased.

**Challenges:**

Images resolution was reduced from (600* 450) pixels to (32* 32) pixels to speed up the models training. This in turn also reduces the image quality and consequently the model accuracy. Examining different image resolution from (28*28), (32*32), and (64*64) we even though was time consuming, but no major improvement on model accuracy was observed using higher resolution (64*64).

**Future Uses/Additional Applications:**

With a robust and accurate model, a skin disease can be diagnosed faster with high precision. This can also help the dermatologist to skip the biopsy part and help patients and health insurances to reduce their medical costs.

**Recommendations:**

Besides gathering more cancerous samples from all forms it is also recommended to gather more patients' attributes like their race, geographical living location. These features can be added to images and be fed to the DL models. Also, images augmentation approaches need to be examined investigated why it is worsening our CNN model.

**Questions:**

1- **Can this model classify black and white images?**
Yes, a Convolutional Neural Network (CNN) model can classify black and white (grayscale) images. CNNs are capable of processing both grayscale and color images, and they can be trained to perform image classification tasks on images of various color modes.

2- **Any methods other than CNN for image processing?**
Certainly, there are several methods and techniques for image processing and analysis beyond Convolutional Neural Networks (CNNs). The choice of method depends on the specific task and requirements. Here are some alternative approaches:

- Recurrent Neural Networks (RNNs): Used for tasks involving sequences of images or time-series data.
- Generative Adversarial Networks (GANs): Used for image generation and image-to-image translation.
- Transformers: Recently, transformers have shown promise in image analysis tasks, particularly with vision transformer (ViT) models.

3- **What additional patients' attributes can be added to the as supplemental features?**
- Age
- Sex/Gender

- Skin Type
- Location of Lesion
- Personal and Family History
- Sun Exposure History
- Environmental Factors

4- **How reliable is the model?**

The reliability of a skin cancer classifier model, such as one based on a Convolutional Neural Network (CNN), can vary depending on several factors:

- Quality and Quantity of Data
- Model Architecture
- Generalization
- Validation and Testing
- Ethical Considerations

5- **Can Deep Learning Models Replace Human Expertise?**

While CNNs can aid in the diagnosis of skin cancer, they should not replace the expertise of dermatologists and medical professionals. They should be seen as tools to assist medical experts rather than replace them.

6- **What other method can be used (CNN + Decision Tree)?**

Yes, Convolutional Neural Networks (CNNs) can be used in conjunction with decision trees or other machine learning algorithms in a variety of ways, depending on the specific problem you're trying to solve. This combination is often referred to as an "ensemble" method, where multiple models work together to improve overall performance. Here are some common ways to use CNNs along with decision trees:

- Feature Extraction with CNN + Decision Trees
- Cascade of Classifiers
- Ensemble Learning
- Hybrid Models

7- **How long does it take to get a CNN model trained?**

The time it takes to train a Convolutional Neural Network (CNN) model can vary significantly depending on several factors. Here are some key considerations that influence the training time:

- Model Complexity
- Dataset Size
- Hardware
- Batch Size
- Learning Rate and Optimization Algorithm
- Convergence Criteria

8- **How much improvement can be achieved by adding a new dense layer?**

There is no specific answer for this question.

**9- What is the role of drop out layer?**

- The dropout layer is a regularization technique used in neural networks, including Convolutional Neural Networks (CNNs). Its primary role is to prevent overfitting and improve the generalization performance of the model. Overfitting occurs when a neural network learns to perform exceptionally well on the training data but fails to generalize effectively to unseen data, such as validation or test data.

**10- Why sometimes doesn't image augmentation improve the accuracy?**

Image augmentation doesn't always improve accuracy for several reasons:

- Sufficient Original Data
- Inappropriate Augmentation
- Overfitting
- Validation Set
- Data Imbalance

**Ethical Consideration:**

Ethics endorsement of the datasets has been attained from the Medical University of Vienna (Protocol-No. 1804/2017). Patients' identities remained confidential, for this there is only a patient's index and no relevant info about their identity. Additionally, because there is a chance that trained model be used by practitioners or someone who is relying on it to get an accurate result, the model needs to be fed by actual and real data to avoid any misleading result.

**Reference:**

Tschandl, P. (2018). *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. Https://Dataverse.Harvard.edu. https://doi.org/10.7910/DVN/DBW86T/XJZSQ6

Ahmadi Mehr, R. A., & Ameri, A. (2022). Skin Cancer Detection Based on Deep Learning. *Biomed Phys Eng*. https://doi.org/10.31661/jbpe.v0i0.2207-1517

Appendix:

1- https://www.ibm.com/topics/deep-learning
2- https://keras.io/
3- https://www.skincancer.org/skin-cancer-information/
4- https://www.ibm.com/topics/neural-networks