**Title:**

**Skin Cancer Prediction Using Deep Learning Technique**

**ABSTRACT**

In the United States, there are over 5 million cases of skin cancer each year, making it the most common type of cancer. There is a high risk of skin cancer for some people, while it is lower for others. Sun exposure or tanning beds are the most preventable causes of skin cancer. To implement the research, I collected skin cancer data from Kaggle [34]. Data frames are created after data is collected in Kaggle. The image path and image name are contained in the data frame. The data was then split into two parts, and I processed each part separately. There are two parts to the application, one for training purposes and one for testing purposes. There are two separate parts to the training part, and they are called training data and validation data. As part of my work, I have created data generators. In the process of creating a data generator, data augmentation technologies are being used. A customized mobile network architecture was developed by me. To conduct training properly, customized mobile nets need to be provided with training and validation data. After each epoch, the model will try to evaluate the validation data based on how much the model learned in that epoch. Therefore, I provided validation data during the training phase. Reports on classification and confusion matrix are used to evaluate model evolution. SGD, Adam, and Adagrad Optimizer were used to test the model. Approximately 86% of Optimizer SGD data are accurate, and 90% of Adam Optimizer SGD data are accurate according to the estimate. A study conducted by Adagrad found that the optimizer was 87% efficient. My research has been 90% accurate with customized mobile networks.

## AIM OF THIS PROJECT

A major goal of this project is to construct a model that can predict skin cancer accurately.

**Objectives:**

**This research has several main objectives, among them:**

* Literature review is the first step in my research. Any research should start here.
* As part of the literature review, I will try to collect information about the research findings, such as how much accuracy the authors obtained, what new techniques they used, and how they applied the new techniques.
* There is a Keras API which can be used to download the mobile Net
* Develop a mobile net model that can be customized
* As part of my data processing plan, I will balance, augment, normalize, and split data.
* The training data will be used to train the model. A classification report as well as a confusion matrix are used to evaluate the model.

## LITERATURE REVIEW

For the detection of mobile skin cancer, [13] a study was conducted using Faster R-CNN incorporated with MobileNet v2. To be able to take advantage of the smartphone's camera and its technology, a mobile device was used. Using the Faster R-CNN model for skin cancer detection, the Faster R-CNN model achieved better accuracy than MobileNet v2 on the training and Jupyter notebook tests with the highest values of 87.2%. Meanwhile, MobileNet v2 obtained the same high accuracy as Faster R-CNN when using the Android app for skin cancer detection with 86.3% accuracy.

Diagnoses are made using the concept of image processing and deep learning. The number of images has also been enriched as a result of the use of different techniques for image augmentation. Lastly, to significantly improve the precision of the classification tasks, we employ the transfer learning approach to further improve the accuracy. By using the proposed CNN (Convolution Neural network) method, we find that the precision, recall, F1-score, and accuracy of the predictors are approximately 0.76, 0.78, 0.76, and 79.45 per cent, respectively.

Using the Fourier identifiers of the edge of the lesion from dermoscopic images after the dermoscopic images have been divided, Kamasak et al. were able to classify dermoscopic images. 84.333% of melanoma diagnoses were accurate.

The use of a multi-scale convolutional network, a per-image normalization, and pooling over an augmented feature space increase accuracy consistently. [16] We achieve higher accuracy compared to the current state-of-the-art of 85.8% over 5 classes (compared to 75.1%) with noticeable improvements over underrepresented classes (e.g., 60% compared to 15.6%). This method achieves an accuracy of 81.8% on a 10-class dataset of 1300 images captured from a standard (non-dermoscopic) camera, exceeding the previously published accuracy of 67%.

Doctors' abilities and knowledge play a major role in determining skin conditions. It is therefore necessary to create a system for identifying skin cancer. [17] This identification system uses convolutional neural networks (CNNs) that detect images and patterns. CNNs work through three layers, including convolutions, pooling, and fully connected layers. Based on the dermoscopy images of the HAM10000 skin cancer dataset, this identification system was developed. 80% and 78% of the skin cancer identification system are accurate after training and testing, respectively.

Synthetic skin cancer images are created using a Generative Adversarial Network (GAN) to compensate for the lack of training data for CNN algorithms. The classification performance of the designed trained CNN without the synthetic images is approximately 53%, but when the synthetic images are added to the primary database, the performance of the model increases to 71% [18].

Using 10,0000 images of skin cancer, we [19] pre-trained a model based on a Convolutional Neural Network. In the case of a mobile phone device, the model will be deployed, and inference will take place locally where the test data remains, i.e., whenever a new test image is given, all computations will take place locally, including a comparison with the old image. By using this approach, latency can be reduced, bandwidth can be saved, and privacy can be enhanced. There is a 75.2% accuracy rate obtained in this study.

A CNN algorithm can be used to classify skin lesion photos based on repeated results. Fine-tuning was performed using Resnet50, InceptionV3, and Inception Resnet. A contribution and innovation of this study are ESRGAN's use as a pre-processing step, in addition to the designed CNN, Resnet50, InceptionV3, and Inception Resnet. [20] Comparable results were obtained with our designed model. Results showed that the suggested strategy works with the ISIC 2018 skin lesion dataset. InceptionV3 (85.8%), Inception Resnet (84%) and Resnet50 (83.7%) were better than CNN at 83.2% accuracy.

According to Chakravorty et al. [21], a colour histogram divergence and structural similarity metric can be used to classify skin lesion distributions. The accuracy of their classification method is 83%, which is a very good result.

[22] The proposed N-Net network and IRRCNN models are evaluated on ISIC-2018 benchmark datasets for skin cancer segmentation and classification. Segmentation tasks are performed better with the R2U-Net than with the Recurrent Residual U-Net. ISIC2018 dataset shows a testing accuracy of 87% for dermoscopic skin cancer classification.

In their work, Celebi et al. [23] propose a method for dealing with colour schemes in Dermo-scope images. As a result, they used a clustering algorithm called K-means to achieve their goal. To achieve a sensitivity of 62%, it is necessary to reduce the colour of colour reduced colour images.

# PROCESS

The literature review related to this research has been completed to the best of my knowledge so far. Additionally, I found skin cancer data from Kaggle during the study.

**The setup processes**

Python Jupyter Notebook environment has been created. My plan is to use the colab environment when I am dealing with a large volume of data, due to large volume of data.

**Description of the dataset**

In order to get the data set, I took it from the Kaggle website. There are mainly two directories in the data set. The first one is a training session, and the second one is a test session. Every single directory consists of two further subdirectories. There are benign and malignant subdirectories within the train directory. There are also benign and malignant subdirectories within the test directory. There are 500 images of skin cancer in the test directories for benign and malignant. These images are all related to skin cancer. 5000 files are in the benign train directory and 4605 files are in the malignant train directory.

**Data frames arrange data**

As a first step, I have imported all the necessary libraries. The data set data is read from the library after it has been imported. There is an arrangement of the data in the form of data frames.

I have split the data into two parts, one of them being img\_path and the other being the target.

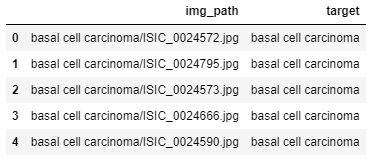


Figure 1: Distribution of the data

**Data balancing**

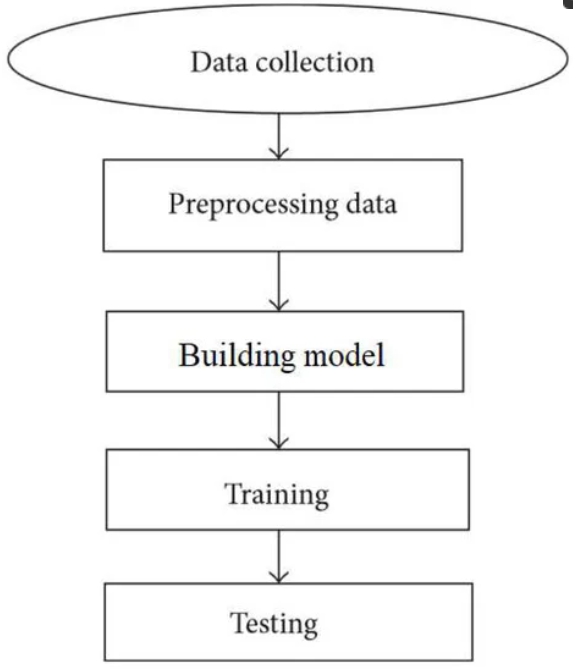
It has been found that the data set is in an imbalanced form, so I am using augmentation techniques to balance the data.

**Further work**

In order to decide on whether to transfer or customize the learning, I will do some research and then make a decision.

This research will be organized as follows:

* Analyzing exploratory data and preprocessing the data.
* A preprocessing step that addresses data imbalance.
* Data classification based on deep learning.
* The model has been evaluated.
* The analysis is compared to existing results of the state of the art in this field.
* Documentation to be completed



## Conclusion

Excessive exposure to UV radiation from the sun appears to be the main risk factor that leads to skin cancer in this country, with 86% of skin cancer cases being attributed to excessive sun exposure. Exposure to artificial sources of UV radiation is the second most common source of UV exposure, which includes tanning beds and/or lamps used indoors. People are typically diagnosed with this disorder when they are 65 years of age or older. Even young people, who are under the age of 30, can suffer from melanoma, even when they are unaware of it. Cancer of the young (especially women) is one of the most common cancers. I have been using customized mobile networks for the entire research I have been doing. With the help of customized mobile networks, I have been able to achieve 90% accuracy in my research.

The model was tested with SGD and Adam Optimizer. It is estimated that Optimizer SGD is 86% accurate, and Adam Optimizer SGD is 90% accurate. There is an 87% efficiency rate for the Adagrad optimizer.