# Automated Skin Lesion Classification: A Comprehensive Analysis of Transfer Learning and Machine Learning Techniques

By
Muhammad Zain Amin
Md Imran Hossain
Taiabur Rahman

# Skin Lesion

- **Skin lesion** refers to an abnormal growth or appearance on the skin that deviates from the surrounding healthy skin.

- *According to the World Health Organization (WHO) estimates, Globally in 2020, over 1.5 million cases of skin cancers were diagnosed and over 120,000 skin cancer associated deaths were reported.*

**Causes and Risk factors:**

- Age factor
- Exposure to radiation
- Ultra-violet from the sun
- Substance abuse, chemicals, environment pollutants

# Treatment and Detection of Skin Cancer

- To reduce mortality growth rate due to skin cancer, early detection and treatment is essential.

- **But the problem is that:-**

  *“Even an experienced dermatologists analysing the skin lesions manually may reach to wrong decision in grading”.*

- **Problem Solution:-**

  *“Automatically getting the most significant information from skin lesion images, we can use machine learning and deep learning technique for detection and classification of skin lesions types*”.

# Project Objective

- **To develop a computer aided diagnosis (CAD) system** in order to detect skin lesions using deep learning and machine learning techniques.

- The **ISIC2017 challenge dataset** is used to classify the lesions categories such as benign, melanoma, and seborrheic keratosis.

- A **two-step hierarchal classification pipeline** is developed: the first stage is “benign vs. others” and the second stage is “melanoma vs. seborrheic keratosis”.

# ISIC 2017 Skin Lesion Dataset

- The **ISIC 2017 dataset** has been widely used for training and evaluating machine learning and deep learning models in skin classification tasks.

- The **ISIC 2017 dataset** has high resolution and different sizes (from 540x722 to 4000x6000 pixels) images.

ISIC2017 Data Info

| Lesion Types | Number of Images | | |
|---|---|---|---|
| | Train | Validation | Test |
| **Benign** | 1372 | 78 | 393 |
| **Melanoma** | 374 | 30 | 117 |
| **Seborrheic keratosis** | 254 | 42 | 90 |
| **Total** | **2000** | **150** | **600** |

* https://challenge.isic-archive.com/data/#2017

# ISIC 2017 Skin Lesion Dataset

Training Ground Truth Data

| image id | melanoma | seborrheic keratosis |
| --- | --- | --- |
| ISIC_0000000 | 0 | 0 |
| ISIC_0000001 | 0 | 0 |
| ISIC_0000002 | 1 | 0 |
| ISIC_0000003 | 0 | 0 |
| ISIC_0000004 | 1 | 0 |

Validation Ground Truth Data

| image id | melanoma | seborrheic keratosis |
| --- | --- | --- |
| ISIC_0001769 | 0 | 0 |
| ISIC_0001852 | 0 | 0 |
| ISIC_0001871 | 0 | 0 |
| ISIC_0003462 | 0 | 0 |
| ISIC_0003539 | 0 | 0 |

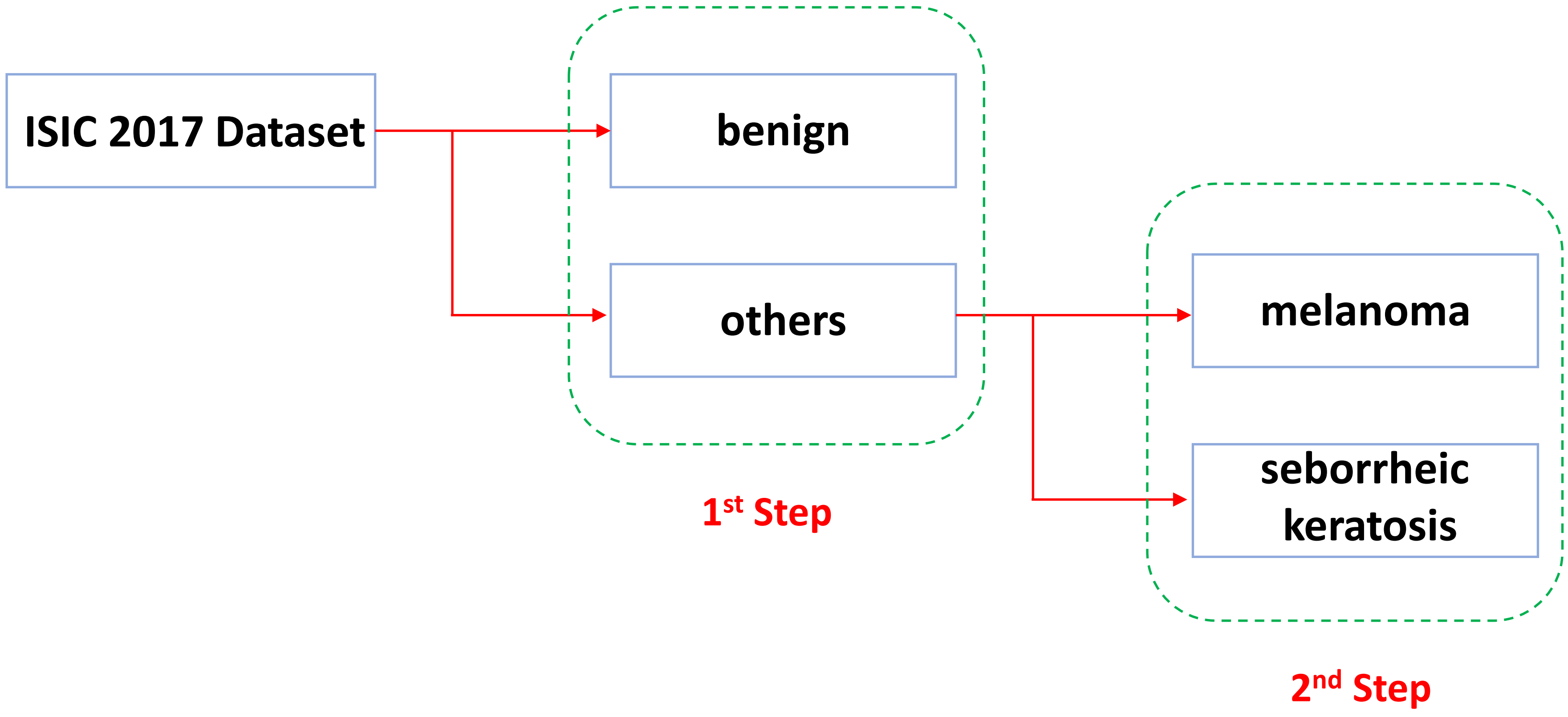
Testing Ground Truth Data

| image id | melanoma | seborrheic keratosis |
| --- | --- | --- |
| ISIC_0012086 | 0 | 1 |
| ISIC_0012092 | 0 | 0 |
| ISIC_0012095 | 0 | 0 |
| ISIC_0012134 | 0 | 1 |
| ISIC_0012136 | 0 | 1 |

* https://challenge.isic-archive.com/data/#2017

# Dataset Handling

- **Two-step hierarchal** data handling:



# Dataset Handling

Given Training Ground Truth Data

| image id | melanoma | seborrheic keratosis |
|---|---|---|
| ISIC_0000000 | 0 | 0 |
| ISIC_0000001 | 0 | 0 |
| ISIC_0000002 | 1 | 0 |
| ISIC_0000003 | 0 | 0 |
| ISIC_0000004 | 1 | 0 |

Modified Training Ground Truth Data

| image id | **melanoma** | **seborrheic keratosis** | **benign** | **others** |
|---|---|---|---|---|
| ISIC_0000000 | **0** | **0** | 1 | **0** |
| ISIC_0000001 | 0 | 0 | **1** | **0** |
| ISIC_0000002 | **1** | **0** | **0** | 1 |
| ISIC_0000003 | 0 | 0 | **1** | **0** |
| ISIC_0000004 | 1 | 0 | **0** | **1** |

**benign = 1** ; if melanoma and seborrheic keratosis both are **0**

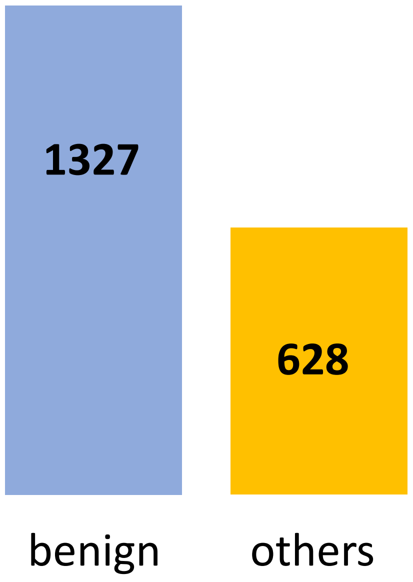
**Others = 1** ; if either melanoma and seborrheic keratosis is **1**

* We used the similar technique for both the validation and test data set

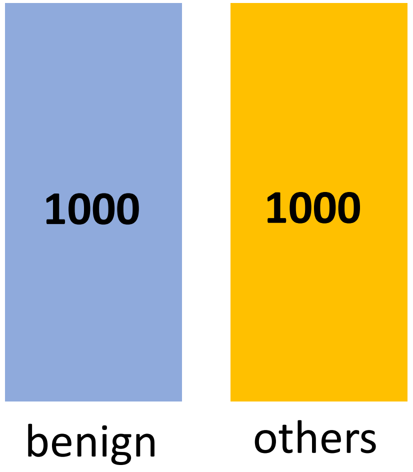
# Data Balancing

- Training dataset is **highly imbalanced** [ benign 1372 vs others 628].

- To perform the **first step of the hierarchal classification**, we have downsampled the Benign cases to **1000** samples from **1372** samples. Others class consisting of melanoma and seborrheic keratosis cases have been upsampled to **1000** samples from **628** samples.




- To perform the **second step of the hierarchal classification**, we have removed the benign cases from the training set in order to perform the melanoma vs. seborrheic keratosis classification.



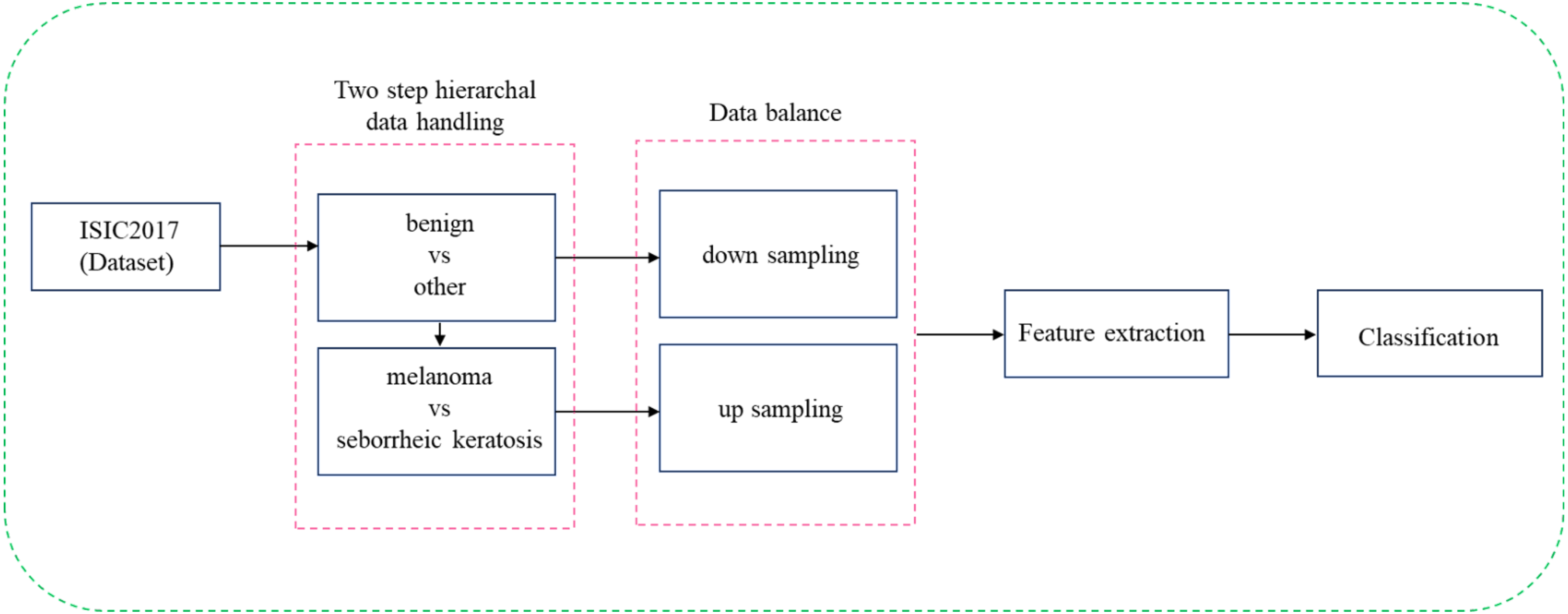
# Data Preprocessing

Before feeding the images to the network, few steps are necessary: For each skin lesion image the following preprocessing pipeline is applied:

- **Resizing** the images to 128x128.
- **Normalization** of dataset images. [Each pixel of images is divided by 255. By dividing by 255, the pixel values are scaled to the range of 0 to 1. This is a common technique for normalizing pixel values in images, as it brings them to a standardized range]

# Proposed Computer Aided Diagnosis (CAD) System

- Our CAD system is developed based on a two-step binary hierarchal classification. Several models based on Hybrid Architecture and Transfer Learning Techniques have been evaluated.



# Transfer Learning Approach

Training a model from scratch requires a lot of data, and even if such data are available, it would require a lot of time.

- We download several **pretrained CNN models** (e.g., VGG16, ResNet50, Xception, Inception, EfficientNetV2) from Keras Application and **retrained** mentioned models using ISIC 2017 dataset for the classification challenge.

**Step1:**

- We used the pretrained model as **feature extractor**.
- **Freeze** all network parameters (to keep parameters fixed during retraining).
- Remove the last layer (**classification** layer).
- Replace final layer (with a **sigmoid function** for binary classification)

- **Sigmoid Function,** $s = \frac{1}{1 + e^{-x}}$ **= S [0,1]**

# Transfer Learning Approach

**Step 2:**

- We have fine tuned the pretrained model on our Skin lesion dataset.
- We retrained the pretrained ***imagenet*** models using our own data (ISIC 2017) without weight initialization.
- Its not what we trained from scratch since the network weights are those of the pretrained nets.

# Experimental Environment for the Training

The experiment mainly uses deep learning models. The hyperparameters uniformly set for these models are shown in Table.

(i) The goal of the **optimizer** is to update the parameters of a model iteratively in order to minimize the loss function.

(ii) The **learning rate** determines the step size taken during optimization. It controls how much the weights of the model are updated based on the calculated gradients.

| Parameters | Value |
|---|---|
| Batch Size | 16 |
| Epochs | 10 |
| Learning Rate | 0.001, 0.005 [varies] |
| Optimizers | Adam, SGD |

# Experimental Environment for the Training

(iii) In deep learning, **Batch size** refers to the number of training examples or data points that are processed together in a single forward and backward pass during the training of a neural network. In this case, our entire train data will be divided into 16 batches.

(iv) In deep learning, an **epoch** refers to a complete iteration through the entire dataset during the training process of a neural network. In our project, we used 10 epoch for training the model.

# Hybrid Approach

- Extracting features using pretrained CNN models and execute classification task using machine learning classifier models.

**Working Steps:**

- We used the pretrained CNN models (e.g., VGG16, ResNet50, Xception, Inception, EfficientNetV2) from Keras Application as **feature extractor**.
- **Freeze** all network parameters (to keep parameters fixed during retraining).
- **Remove** the last layer and train traditional Machine Learning Classifiers (e.g,. Random Forest, XGBoost, SVM, Logistic Regression) on top of the CNN features.

# Binary Cross Entropy

In addition, we have used the binary cross entropy loss function for our classification purpose. It measures the dissimilarity between the predicted probability distribution and the true binary labels. The formula can be expressed as:

$$\textit{Binary Cross Entropy} = -(y * \log \sigma(p) + (1 - y) * \log(1 - \sigma(p))$$

where,

**p** = prediction of the model

**y** = ground truth

**σ** = sigmoid function.

# Best Result for the Hybrid Model

| Feature Extractor | Classification Model | Benign vs Other | | | Melanoma vs Seborrheic Keratosis | | | BMA |
|---|---|---|---|---|---|---|---|---|
| | | Accuracy | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | |
| **Xception** | **Random Forest** | **0.70** | **0.82** | **0.47** | **0.76** | **0.74** | **0.80** | **0.79** |
| | Naïve Bayes | 0.56 | 0.48 | 0.72 | 0.71 | 0.74 | 0.69 | 0.63 |
| | Support Vector Machine | 0.68 | 0.73 | 0.59 | 0.73 | 0.72 | 0.74 | 0.74 |
| | Logistic Regression | 0.70 | 0.75 | 0.62 | 0.72 | 0.70 | 0.74 | 0.73 |
| | Decision Tree | 0.60 | 0.65 | 0.50 | 0.70 | 0.66 | 0.76 | 0.69 |
| | K Nearest Neighbors | 0.56 | 0.46 | 0.75 | 0.70 | 0.82 | 0.54 | 0.60 |
| | Gradient Boosting Machines | 0.70 | 0.72 | 0.67 | 0.72 | 0.68 | 0.77 | 0.72 |
| | XGBoost | 0.70 | 0.77 | 0.57 | 0.70 | 0.66 | 0.76 | 0.73 |

* Performance matrix of all the models are available in the report.

$$BMA = \frac{1}{M}\sum_{i=1}^{M}\frac{TP_i}{N}$$

**M** = Number of class.
**TP** = Ture positive.
**N** = Total number of sample.

# Best Results for the Transfer Learning Models

| Model | Optimizer | Learning Rate | Benign vs Other | | | Melanoma vs Seborrheic Keratosis | | | BMA |
|---|---|---|---|---|---|---|---|---|---|
| | | | Accuracy | Sensitivity | Specificity | Accuracy | Sensitivity | Specificity | |
| **VGG16** | Adam | 0.005 | 0.72 | 0.90 | 0.38 | 0.70 | 0.71 | 0.68 | **0.76** |
| | SGD | 0.005 | 0.72 | 0.83 | 0.52 | 0.70 | 0.66 | 0.76 | 0.75 |
| **InceptionResNetv2** | Adam | 0.005 | 0.77 | 0.83 | 0.64 | 0.67 | 0.53 | 0.86 | **0.74** |
| | SGD | 0.005 | 0.77 | 0.79 | 0.75 | 0.74 | 0.75 | 0.73 | **0.76** |
| **DenseNet201** | Adam | 0.005 | 0.69 | 0.75 | 0.58 | 0.75 | 0.71 | 0.82 | **0.76** |
| | SGD | 0.001 | 0.66 | 0.62 | 0.73 | 0.70 | 0.76 | 0.62 | 0.67 |

* Performance matrix of all the models are available in the report.

$$BMA = \frac{1}{M}\sum_{i=1}^{M}\frac{TP_i}{N}$$

**M** = Number of class.
**TP** = Ture positive.
**N** = Total number of sample.

# Conclusion and Future Scope

- The results show a potential of improvement in this task using **hybrid technique**.
- Possible future work includes **more optimization** and **preprocessing** in the feature engineering step of machine learning, further **fine tuning** of the model parameters in deep learning.
- It is clear that some models performed outstanding with benign and other classes and some models performed better with melanoma and seborrheic keratosis classes.
- Therefore, a new model with the **combination of the best performed model** with benign and other classes and melanoma and seborrheic keratosis classed can be proposed in future.

Thank you