Classification of Pigmented Skin Lesions using Dermatoscopic Dataset and Convolution Neural Network (CNN): Theoretical Studies and Experimental Results Generation

I. Theoretical

"The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore, the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind."

James Clerk Maxwell (1850)

Abbreviations

ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolution Neural Network
CRF	Conditional Random Field
GDPR	General Data Protection Regulation
LSTM	Long Short-Term Memory
NN	Neural Network
RNN	Recurrent neural Network
FLOPS	Floating Point Operations per Second

DCNN	Diffusion Convolutional Neural Networks
GD	Gradient Descent
BP	Back Propagation
LR	Learning Rate

ABSTRACT

Skin lesions are a widely known problem caused due to unhealthy lifestyle, skin reactions, prolonged exposure to direct sunlight, hereditary, etc. But such diseases are often ignored due to the scarcity of dermatologists and inaccessibility to the pathology lab for diagnostics. These diseases remain undiagnosed with the naked eye. To examine such diseases, dermatologists use a technique called Dermatoscopy which uses surface microscopy to examine skin surface. Dermatoscopy outputs closely rendered images of the skin epidermis layer for diagnosis. With improved mathematical models using neural networks, LSTMS, these skin lesions can be classified with high accuracy and reach the sections of the population that don't have access to pathology lab facilities. Furthermore, Convolutional neural network (CNN) models have been widely used for skin disease diagnosis and they have the potential to aid dermatologists to improve clinical decision making. But still, there are certain challenges in employing CNNbased models such as limited availability of less diverse and less biased publicly available datasets towards melanomas.

This paper devises a novel solution for the classification of pigmented skin lesions using image augmentation. Moreover, the CNN technique handles any imbalances in the dataset with high accuracy. The overall approach is explained as follows. Firstly, the dataset is pre-processed to remove the noise, and impulse features by applying bilateral filtering and gray scaling. Secondly, image augmentation techniques such as rescaling, shearing, shifting (horizontal, vertical, right, left), flipping, padding, and filling are used to create more diverse features in the existing dataset for better accuracy and prevention of overfitting. Thirdly, CNNs are trained on processed and augmented datasets to verify whether the model can generate desirable output with moderate complexity. Lastly, Dropout layers are added to reduce overfitting and two optimization algorithms, namely, EarlyStopping and ReduceLROnPlateau are added to achieve optimum computational efficiency and to update the model's weights and biases to a more

optimum point by reducing the learning rate. Finally, a specific Sparse Categorical Cross Entropy (SCCE) loss function is introduced. The proposed approach has shown relatively high accuracy than multiple ensembling models on different skin lesions dermoscopic datasets. The model promises to obtain optimum results while keeping the computational power low.

INTRODUCTION

Skin lesions are a serious malignancy with sever increase in patients. Lesions are detected visually by dermatologists with the help of Dermatoscopic imaging. Dermatoscopy is a diagnostic technique that is used to examine the surface of skin using surface microscopy for examination of skin lesions that are unrecognized with the naked eye. These skin lesions can be hereditary, allergic reactions, systematic disorder in the skin due to sunlight exposure or due to some diseases like cancer. Depending on the cause, these skin lesions can be of various types such as Melanocytic Nevi (NV), Melanoma (Mel), Vascular Lesions (Vasc), Dermatofibroma (DV), Basal Cell Carcinoma (bcc), Benign Keratosis-like Lesions (BKL), Actinic Keratoses (akiec). But due to shortage of dermatologists and inaccessibility of pathology lab facilities, patients don’t get right kind of diagnostics leading to improper treatment methods. Artificial neural network can be employed to recognise such skin lesions. Dermatoscopic images provide diverse demographic feature to train artificial neural network to classify and diagnose pigmented skin lesions with high level of consistency and accuracy. This can be applied to an automated system to help reduce the lack of dermatologists’ availability.

The deep learning approach of using artificial neural network rely on Dermatoscopic images sourced from ISIC Dermatoscopic dataset which classify the image into seven broad categories based on diagnostics provided by the dermatologists by international standard. But such datasets are biased particularly towards melanomas and melanocytic lesions due to lack of diversity of data for other lesions. Here, we are concerned about the classification of seven types of skin lesions using a highly optimised Convolution Neural Network (CNN) model into broader classes.

Despite various researches and model, there were hinderances for improving the diagnostic efficiency of such model. First, the various datasets available were biased towards particularly two types of lesions Melanocytic and Melanomas. For example, HAM10000 dataset (HAM), one of the most commonly used Dermatoscopic research dataset, contains only 10015 Dermatoscopic images of which more than 60% images are of nevi or melanomas. This hinders models training capabilities due to lack of diversity in features. Even models tend to overfit if not optimized to exact parameters. Secondly, the most of the public datasets have a very small sample size. To classify and recognise such lesions, large and diverse dataset provide a good base on which model can be trained. To ensure this, the Dermatoscopic image dataset undergo data pre-processing technique of data augmentation for enhancing the dataset and its feature diversity.

Considering the consistency, accuracy, and losses in diagnostics, a Convolution neural network (CNN) model is preferred model of study to train and validate the result for efficient diagnostics. However, for classification problems with large sized image dataset, state-of-theart models from Alex Net to RegNet or similar models with high parameter capacity can always be used. But this doesn't hold true in case of small sized datasets as increased parameters can cause change from under-fitting state to over-fitting state.

The complexity of the model can be more accurate but at the same time this will have a large number of parameters associated with each layer. Thus, converging all the parameters to optimum will take more computation power and iterations leading to a lot more complexity and high chances of overfitting or even under-fitting if not trained with proper parameter estimation. There are many complex models available that uses less-complex parameters for fine tuning.

LITERATURE REVIEW

CNNs have been a major breakthrough in the field of image processing and image classification with lower Flops and higher accuracy than the tradition model and machine learning algorithms. In [2] Gao, Zhimin, et al. proposed a framework using CNNs that showed high accuracy in visual recognition task in the paper "*HEp-2 cell image classification with deep convolutional neural networks*". This model has good adaptability for different datasets including cell (dermatoscopic) image recognition dataset with complex and varying testing conditions. It also showed the data augmentation techniques for cell images and masking effectiveness for image classification.

Peng Yao, Shuwei Shen, et al. in their paper [1] "*Single Model Deep Learning on Imbalanced*

Small Datasets for Skin Lesion Classification" proposed a transfer learning model solution using DCNN's with different computation efficiency using Flops of different DCNN model. This paper compared performance efficiency (based on Flops-> Floating Point Operations per Second) and accuracy of various model like ResNet, VGG16, RegNet-X, RegNet-Y, DenseNet, EfficientNet, etc on HAM-10015 dataset and implementing various augmentation parameters for achieving desired accuracy and performance focusing mainly on the reducing the computations to be used in an application-based product.

[10] explained the dataset validity and robustness by providing the insights of the international certification that HAM dataset is being tested. It provided the size, features, and demography in the dataset as well as providing the outlook of the model implemented in the challenge by other researchers. Alongside [5] Skin Lesion Analysis towards melanoma Detection: Challenge at 2017 International Symposium on Biomedical Imaging (ISBI) provided the main objective, problems and criteria for evaluation of the model in the HAM challenge dataset which later became a diverse dataset used by various researchers to conduct perform tests on model for underlying images.

Various above model proposed on dermatoscopic dataset image classification tend to extract the features as critical as possible in the image but [3] Rosendahl, C., Tschandl, P., et al. in Diagnostic accuracy of Dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. J Am Acad Dermatol analysed the discriminatory power and reliability of dermoscopic criteria used for the classification in different models. It created a check list for accuracy of different diseases like melanoma and visualizing interclass correlations that could lead to false positive or negatives. It tested 6 different models which have comparable accuracies but varying level of complexities in diagnostics criteria.

Another break through proposed by [6] Jia Deng, et al. Dept. of Computer Science, Princeton University, USA in paper ImageNet: A large-scale hierarchical image database, showed the practical accuracy by performing the CNN technique on a high diverse ImageNet dataset on which is the base of most of the image recognition, and classification model for transfer learning. This model is trained on state-of-the art dataset. It demonstrated the hierarchy in the image dataset and showed the hierarchy and correlation among different layers while training. On such a dataset, it achieved an accuracy of 99.8% and formed an important transfer learning model to be used for pre-trained weights and biases.

In [9] Han et al. proposed a clinical image classification model for 12 different skin diseases. They finetuned ResNet model with combination of Asan dataset, MED-NODE dataset, and atlas site images a comprising of 19,398 training images.

[10] presented the comparison of CNN with the international group of 58 dermatologist for the classification of the skin cancer. Most dermatologists were outperformed by the CNN. “Authors concluded that, irrespective of any physicians' experience, they may benefit from assistance by a CNN's image classification.” Google's Inception v4 CNN architecture was trained and validated using dermoscopic images and corresponding diagnoses. The existing algorithms used for the detection and classification of skin cancer disease uses machine learning and neural network algorithms

METHODOLOGY

The main idea is to first normalize the dataset and tackle the problem of imbalance in the dataset using pre-processing techniques before training. For pre-processing dataset undergo several filters, boosting, and segmentation for noise removal, filter reduction for improved computation for optimum training, and uniformity in image dimension. Then the normalized data is augmentation after duplicating the existing dataset diseases (less duplication of biased disease i.e., melanomas) for handling imbalances. Finally, the processed data is given as input to the deep learning model for feature extraction and training. After training, several optimization algorithms like Callbacks, optimizers for better gradient updation, loss function updation, etc are applied based on the log from the first training report. Then model has again been trained with the updated parameters and log report generated again for checking the performance again. This process continued until the model losses and accuracies didn’t converged comparable to the values of state-of-the art model. After, that the model is further checked for computation power by checking the Flops (Floating Point Operations Per Second) of the trained model.

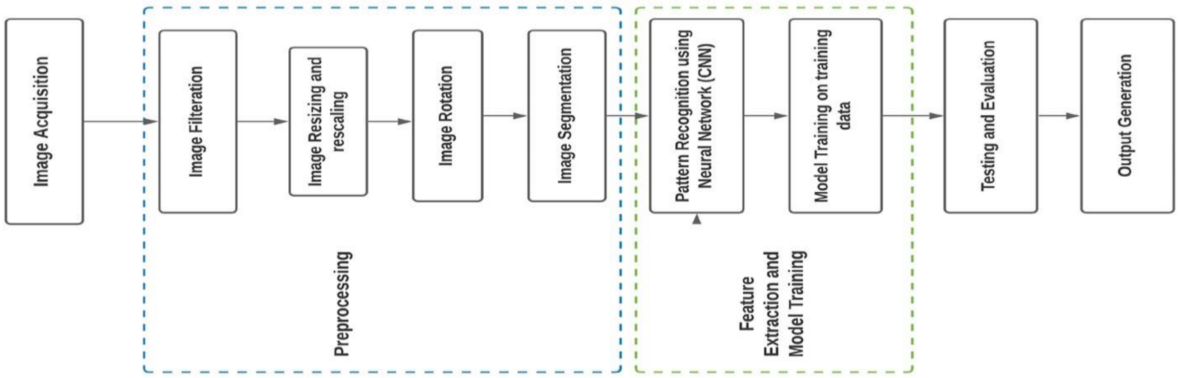
Figure 1 shows a summarized insight of the various stages, to get a proper understanding of the project pipeline, the whole project is divided into three subsections.

- First part illustrates the distribution of the most used dataset along with their advantages and disadvantages.
- Second part deals with the data pre-processing (Image) techniques describing gray scaling, bilateral filtering and data augmentation technique.
- Third part provides in-depth explanation of the proposed CNN model and its parameters along with the evaluation strategy and various optimization functions

integrated for better convergence. Along with this, the loss function used for evaluation and its properties are stated.

Figure 1

Various stages and process involved in the project divided into 3 subsections a) Data Acquisition b) Data pre-processing c) Feature extraction and model tuning



The process involved in developing complete insight and development include following steps:

1. Dataset Description (Data Acquisition)
2. Data Pre-Processing (Image)
    a. Grayscale
    b. Resizing (done while apply deep scan on images)
    c. Denoising using Bilateral Filtering
    d. Data Augmentation
3. Model
    a. Basic NN (Neural Network) model's approach
    b. CNN (Convolution Neural Network)
    c. Optimization parameters

   d. Loss function
   e. Training parameters, accuracy, and losses
4. Results
   a. Performance
   b. Comparisons with other state-of-the art model trained on same dataset

Dataset Description

Dataset under study is HAM-10015 dataset for Skin Cancer which is available on Kaggle [12,13,14]. It contains 10015 labelled images of pigmented skin lesions classified into seven common skin lesions. The images present in the dataset need to be augmented to create diversification to become useful in improving performance for tasks like segmentation, feature extraction, deep learning, and transfer learning, etc.

ISIC 2018. ISIC 2018 is a widely used "HAM (Human Against Machine) 10015 challenge dataset containing 7 common types of pigmented skin lesions". It is one of the largest, diverse and commonly used publicly available research dataset containing 10015 images of skin lesions classified into 7 broad categories namely, melanoma (MEL), basal cell (BCC), benign keratosis (BKL), melanocytic nevus (NV), actinic keratosis (AK), dermatofibroma (DF) and vascular lesions (VASC). The distribution of each image in the dataset is shown in Table 1.

ISIC 2019. "ISIC 2019[1] dataset is made up of HAM dataset [12], MSK dataset [16] and BCN_20000 dataset [15]. It contains 25331 images come from MEL, NV, BCC, AK, BKL, DF, VASC and squamous cell carcinoma (SCC). The labels for 8238 test images are also unpublished, and it is also necessary to upload the predicted results to the ISIC website for performance evaluation."

7-PT Dataset. 7-PT Dataset is a less diverse dataset containing 1011 cases with a meta dataset file for label mapping. It contains data of only 5 category of skin lesions including BCC (42), NV (575), MEL (252), MISC (97) and SK (45).

Also, the dataset also has 1512 unlabelled images for testing. The accuracy is calculated by uploading the model results to the ISIC website for testing of performance. All the images' labels in dataset labels are confirmed through histopathology testing and dermatologists at international standards.

TABLE 1

DISTRIBUTION OF HAM 10015 TRANINING DATSET.

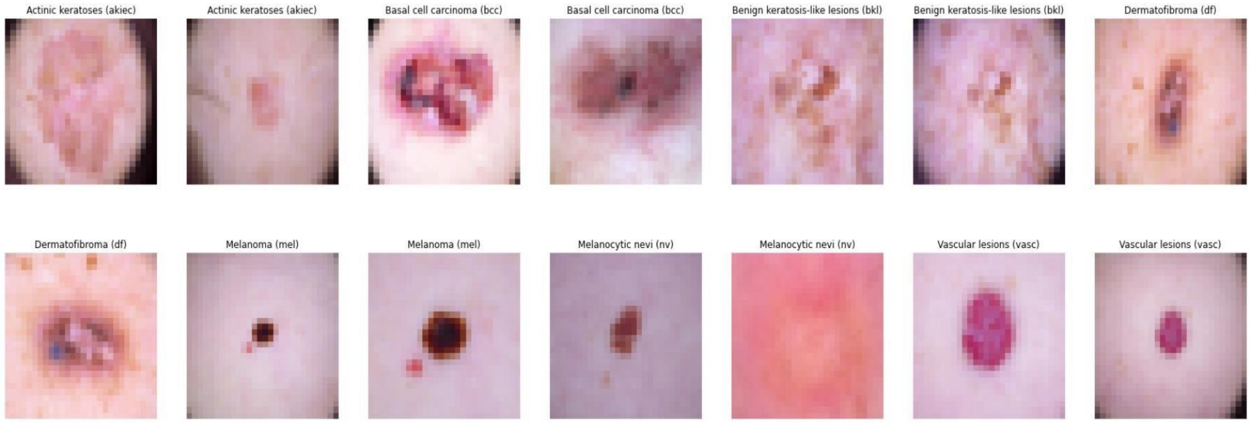
CLASSIFIED LESIONS ARE MELANOMA (MEL), MELANOCYTIC NEVUS (NV), BASAL CELL CARCINOMA (BCC), ACTINIC KERATOSIS (AK), BENIGN KERATOSIS (BKL), DERMATOFIBROMA (DF), VASCULAR LESIONS (VASC)

Dataset	**MEL**	**NV**	**BCC**	**AK**	**BKL**	**DF**	**VASC**
HAM 100015	1113	6705	514	327	1099	115	142
7-PT Dataset	252	575	42	-	-	-	97

Figure 2

Images of all HAM 10015 TRANINING DATSET.

CLASSIFIED LESIONS ARE MELANOMA (MEL), MELANOCYTIC NEVUS (NV), BASAL CELL CARCINOMA (BCC), ACTINIC KERATOSIS (AK), BENIGN KERATOSIS (BKL), DERMATOFIBROMA (DF), VASCULAR LESIONS (VASC)


Actinic keratoses (akiec)
Actinic keratoses (akiec)
Basal cell carcinoma (bcc)
Basal cell carcinoma (bcc)
Benign keratosis-like lesions (bkl)
Benign keratosis-like lesions (bkl)
Dermatofibroma (df)
Dermatofibroma (df)
Melanoma (mel)
Melanoma (mel)
Melanocytic nevi (nv)
Melanocytic nevi (nv)
Vascular lesions (vasc)
Vascular lesions (vasc)

Data Pre-Processing (Image)

Data pre-processing (image) is a technique used to discard or reduce unbalance in the data extracted from various sources in order to achieve a better insight of the problems and employ different machine learning algorithms to achieve desired output. Image pre-processing is the method to use image processing techniques to resize, denoise, filter, segment, padding and rescaling image data to prepare it suitable for training using deep learning model and other mathematical models.

Image pre-processing is an essential technique, especially when working with image dataset and particularly with noisy and microscopic images, to prepare image dataset in order to derive some insightful data understanding and derive some desirable output. This paper use three main type of image pre-processing technique such as Grayscale conversion, Bilateral filtering, and Data augmentation.

Figure 3. Distribution of different skin lesions showing the biasness or imbalance in the dataset towards melanocytic nevi comprising of more than 50% of the images.

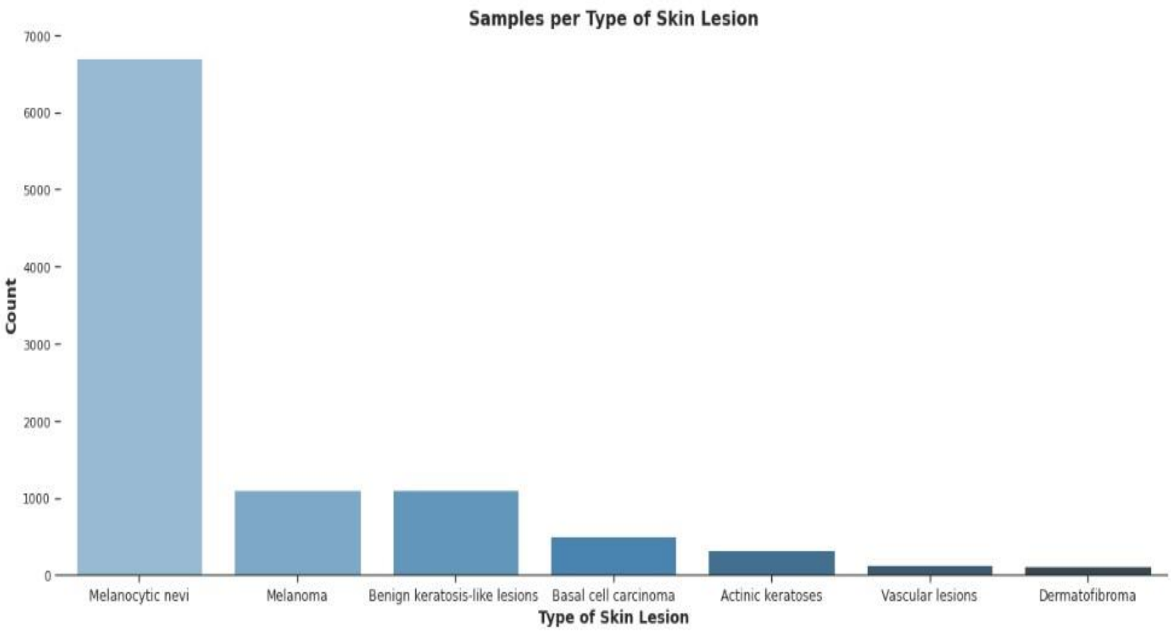


Figure 3 clearly show that there is an imbalance in the dataset where ‘melanocytic nevi’ comprises of about fifty percent of the total dataset which could led to improper training and neglection of minority skin lesion due to lack of diversity. Thus, dataset has to be pre-processed before training the model. To tackle this, we have applied pre-processing techniques to enhance the learnability of the network and even performed Data Augmentation to avoid overfitting. For this, several copies of the original images in dataset are created and then applied data augmentation techniques. Also, the image size of reduced to (28x28x3) at the time of deep scanning. The contrast of the skin lesions in the dataset images has also been increased using filtering before application of pre-processing techniques, applied at the time of deep scanning of images from the directory, so that the dermatoscopic images with smaller dimensions doesn’t cause improper extraction with training.

Data Pre-processing involves following processes:

1. **Grayscale Conversion**
2. **Denoising using Bilateral Filtering**
3. **Data Augmentation**
    a. **Rescaling, and re-sizing**
    b. **Shifting**
        i. **Horizontal and Vertical**
        ii. **Left and Right**
    c. **Rotation**
    d. **Shearing**
    e. **Padding**

Grayscale Conversion

Grayscale is the technique of converting image from any intensity filters i.e., RGB, BGR, HLS etc, to shades of gray (black and white) where each pixel value is an 8-bit integer in the range 0-255, 0 representing black and 255 white. Using Grayscaling technique the pixel value are converted to a single filter of gray while preserving the features of the natural image.

The images in HAM 10015 dataset are present in three filters RGB, RGB refers to Red Green Blue. Each of these images have pixels intensity different in all the R,G,B filters. Grayscaling is implemented on a single pixel by applying a weighted average to its pixel value in R,G,B filters.

$$Y[I]linear = \frac{1}{3}\,[RLinear + Glinear + Blinear]$$

The above equation is of Average Weighted method,

Y[I] -> the grayscale pixel value

R, G, B -> the filter values of the pixels

$$Y[I]weighted = 0.299 * Rfilter + 0.587 * Gfilter + 0.114 * Bfilter$$

The above equation is of Luminous Weighted method,

Y[I] -> the grayscale pixel value

R, G, B -> the filter values of the pixels

Grayscaling is performed in the images to reduce the computational complexity when working with neural networks. This is because, suppose if a CNN layer performs weighted convolution on all the filters, then it will have to perform much greater number of mathematical computations which will not be optimum. Thus, to tackle this, the images are converted to a single filter of gray using Grayscaling which helps to reduce computational complexity to much more optimum value. However, Grayscaling should only be applied to the images that cannot be classified with the naked eye like X-ray diagnosis, skin epidermis layer, etc.

Grayscaling is performed first to make other pre-processing techniques computationally more efficient by reducing some features while conversion from coloured to grayscale. These image needs to be further processed to reduce denoise and undergo augmentation.

Denoising using Bilateral Filtering

Filtering is most fundamental approach in image processing that processes and replace each pixel in image by performing some mathematical functions on neighbouring pixel values. This helps to smoothen the images by creating the neighbouring pixel values normalized to a certain range. The images gathered contains noise, impulsive pixel intensities and unnormalized pixel values which tend to unset the model while training and results in high validation and training losses. To overcome such imbalances, our filter of choice is *bilateral filter*.

$$_{BF}[I]p = \frac{1}{Wp}\sum G * \sigma s(|p-q|) * G * \sigma r(|Ip - Iq|) *_{Iq}$$

In the above equation,

W_p -> the normalization constant/factor

σs -> size of neighbourhood

σr-> minimum amplitude of the edge

σs represents the blurring and sharpening factor of edges in the image. Higher the value higher the blurring.

σr represents the kernel size chosen for convolution.

Bilateral filter is a smoothening, edge-preserving and non-linear filter that replaces the value of each pixel with weighted average of intensity values of neighbouring pixels. One of the advantages of this filter is non-blurring and no feature loss property. As most of the filters tends to blur-out neighbouring pixels of a feature irrespective of it to be a noise or an edge which bilateral filter takes into consideration.

Data Augmentation

Data augmentation is method used to overcome the problem of limited and less diverse training data by performing augmentation techniques like flipping, scaling, rotation, shearing, and shifting on data randomly to generate diversity and improve training number of image dataset. As an optimized augmentation design required intensive augmentation experience and testing to capture various feature, our paper uses self-developed augmentation engine on which dataset is augmented. The parameters used came after running multiple tests on the varying values of the parameters. The optimum values selected are listed in the table below:

Table 2. *Image Augmentation parameters and values*

S.N.	Technique	Values
1.	Rescale	1./255
2.	Rotation Range	10
3.	Horizontal Flip	True
4.	Vertical Flip	True
5.	Width Shift Range	0.2
6.	Height Shift Range	0.2
7.	Shear Range	0.2
8.	Fill Mode	Nearest

All the augments are randomly applied to the images. Rescale is used to change the range of the data to make it easier by converting into smaller values for better processing. Rotation rotates the image in both clockwise and anti-clockwise direction randomly. Image flipping is also applied both horizontal and vertical. Width and height shift range are specified with fill mode set to nearest to keep consistency in the images pixel value thus avoiding noise generation.

Neural Network Model (Convolution Neural Network CNN)

ANN (Artificial Neural Network) or Neural Network are a series of layers (each having some weights and biases) that are interconnected to the subsequent layers trying to learn and recognize the relationship within the data forming a network like in human brain. Each deep learning model contains various neural network layers called hidden layers that perform the task related to recognition, clustering, regression, association, etc. All the processing is done within these hidden layers on the input provided by the input layer and output generated by the output layer. In neural network, each layer has neurons which have some weight associated with it.

Every neural network architecture has following layer:

1. Input Layer
2. Hidden Layer
   a. CNN (Convolution Layer)
   b. LSTM (Long-Short Term Memory)
   c. Dense Layer (Fully Connected Layer)
3. Output Layer

For learning first, the input is fed into the model from the input layer. The input layer is used to provide the data of desired input shape or type to the hidden layer. This input is connected to hidden layer for feature extraction and finding relationship. Second, the hidden layer processes the input data by using mathematical algorithms. There can be any number of hidden layers depending on the use case. After processing by the hidden layer, the output is generated and send to the output layer. The output layer produces the output which can then be used further.

Also, while training, the output of the model if checked for losses using cost/loss function for updating weights and biases. This is process is called Back-Propagation. It uses gradient descent algorithm to fine-tune the model parameters to minimize the losses.

Back-propagation updation rule for model weights and biases:

$$w(new) = w(old) - \eta \frac{\partial fn}{\partial w}$$

In the above equation,

$w_{(new)}$ -> updated weight $w_{(old)}$

-> previous state weight

η -> learning rate

fn-> cost/loss function

$\frac{\partial fn}{\partial w}$-> gradient of the cost/loss function w.r.t weight

Some of the state-of-the art image classification model are:

1. AlexNet
2. VGG-16
3. ResNet
4. DCNN (RegNet, EfficeintNet, Regnet-Y)
5. GoogleNet
6. VGG-19

All these model have high accuracy and performace but have high flops value (Floation Point Operation per Second). Thus, this makes them unusable in light applications thus find use in applications where computation efficiency is shadowed by accuracy. This is paper we proposed a simple yet highly accurate and computationally efficient model leveraging power of CNN for skin disease classification.

Convolution Neural Network (CNN) Model

Convolutional neural networks (CNN) are a type of ANN (Artificial Neural Network) used primarily in image processing and recognition task. They are state-of-the art ANN that processes every pixel in the image to extraction feature and store them for pattern recognition and segmentation tasks. In CNN, the feature is generated using the summation of weighted convolution of image pixel values with the neurons weight matrix as represented by the equation in the equation given below.

Equation used for performing kernel convolution

$$F[m,n] = (f * h)[m,n] = \sum_{i}\sum_{j} h[i,j]f[m-i,n-j]$$

Where,

F[m,n] -> the feature map after convolution $f[m,n]$ -> input image matrix of (mxn) dimension $h[m,n]$ -> kernel/weight matrix of (mxn) dimension for feature extraction

This produces a feature map of reduced dimension to that of the image containing the sensitive information of pixel value that the layer corresponds to. This feature matrix is passed thorough an activation function to introduce the non-linearity in the data.

Equation of activation function a) (ReLU) b) Sigmoid

$$F(x) = \begin{cases} 0, for\ x \leq 0 \\ x, for\ x > 0 \end{cases}$$

$$F(x) = \frac{1}{1 + e^{-x}}, for\ x \in R$$

This is further passed to special type of CNN layer called Maxpooling layer. This layer takes the maximum value from the activation matrix based on the kernel size provided. Thus, the output has the pixel values of the most dominant feature in the image. There are numerous filters in each CNN layer each extracting different dominant feature.

Equation used for performing MaxPooling2D

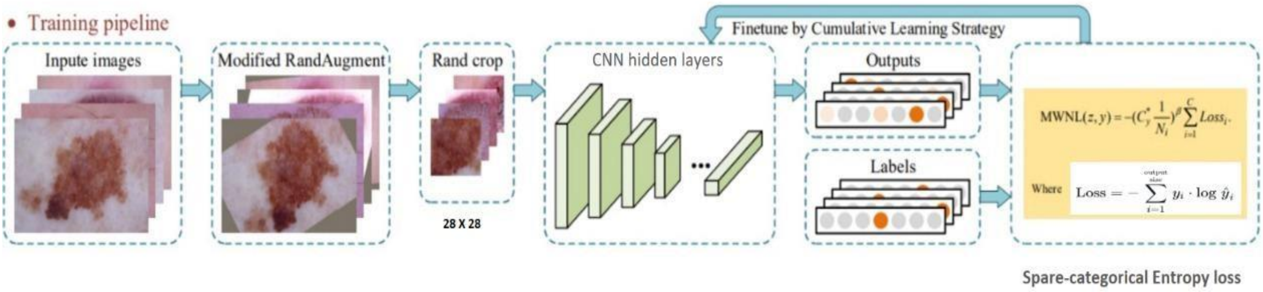
$$Fmax[i,j] = \max\,[A[i,j], A[i+1][j+1], A[i][j+1], A[i+1][j]]$$

Where,

Fmax[i,j] -> the output matrix element at (i,j) position

$A[i,j]$ -> the activation matrix

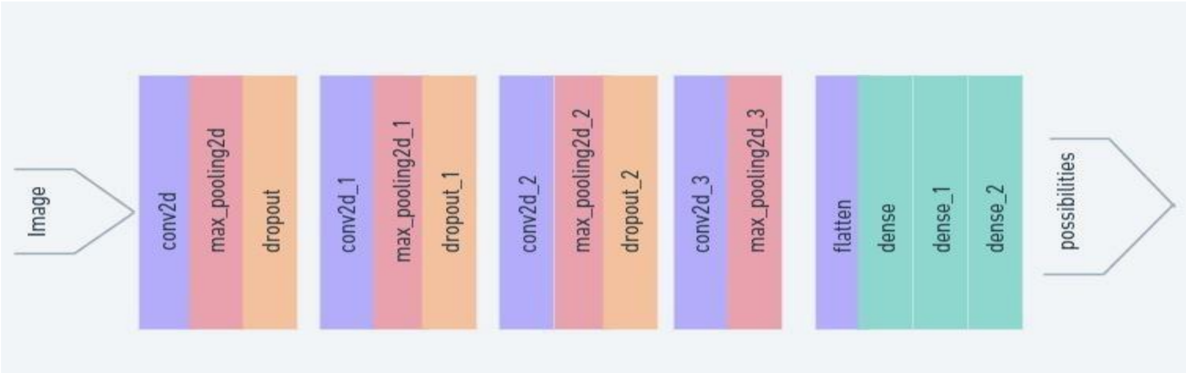
Figure 4. *Training pipeline of the model with finetuning using cumulative learning strategy for loss function*



CNN involves three stages of neural layers to assemble its structures:

1) **Convolution Layer:** The main layer is convolutional layer. In this layer the result of the output layer is gotten from the input by filtering in specific condition. This layer is constructed by the neurons which is in the shape of cubical blocks.
2) **Max-pooling layer:** Pooling layer executes the operation after convolution layer. These layers are utilized to minimize the size of the neurons. These are small rectangular grids that acquires small portion of convolutional layer and filters it to give a result from that block. The most commonly used layer is max pooling that fetch that maximum pixel from the block.
3) **Fully Connected layers:** The final layer of a convolutional neural network(CNN) is a fully connected layer that is formed from the attachment of all preceding neurons. It reduces the spatial information as it is fully connected like in artificial neural network. It contains neurons beginning at input neurons to the output neurons.

Figure 5. *Architecture of implemented CNN model consisting of all the layers*



The *figure 5* shows the model flow and feature extraction via various layers implemented in the model.

1) **Input Layer and Conv2d sets:** This set contains three layers,
    a. **Input**
    b. **CNN**
    c. **MaxPooling**
    d. **Dropout**

    The model contains 4 sets of such type for handling feature extraction and imbalances in the data.

2) **Flatten:** This layer works for dimension conversion i.e., from n-dimensional to 2dimentional so that the fully connected neural network can process the output from the upper layers.
3) **Dense Layers:** The final layer are fully connected layer or dense layers that are responsible for producing the probabilistic distribution using SoftMax activation function.

Equation of activation function SoftMax

$$s(xi) = \frac{e_{xi}}{\sum_{i=1}^{n} e_{xi}}$$

Table 3. *Detailed Architectural summary of implemented CNN model with output and kernel*

Layers	Output Size	Kernel size
Input	28 x 28 x 3	
Convolution	28 x 28 x 16	3 x 3
ReLu Activation	28 x 28 x 16	
MaxPool2D	14 x 14 x 16	2 x 2 + (Stride = 2)
Dropout	14 x 14 x 16	
Convolution	14 x 14 x 32	3 x 3
ReLu Activation	14 x 14 x 32	
MaxPool2D	7 x 7 x 32	2 x 2 + (Stride = 2)
Dropout	7 x 7 x 32	
Convolution	7 x 7 x 64	3 x 3
ReLu Activation	7 x 7 x 64	
MaxPool2D	4 x 4 x 64	2 x 2 + (Stride = 2)
Dropout	4 x 4 x 64	
Convolution	4 x 4 x 128	3 x 3
ReLu Activation	4 x 4 x 128	
MaxPool2D	2 x 2 x 128	2 x 2 + (Stride = 2)
Flatten	1 x 512	
Dense	1 x 64	
ReLu Activation	1 x 64	
Dense	1 x 32	
ReLu Activation	1 x 32	
Dense	1 x 7	
Sigmoid Activation	1 x 7	

Optimization Parameters (Callbacks, Regularization, Optimizers)

Optimization parameters are the algorithms or function used to optimize the response of the model and provide enhanced convergence power to the loss function towards the global minimum. The optimization functions like EarlyStopping helps in optimization of the computation efficiency and prevent overfitting by stopping the training if the loss doesn't change much for a given patience value of the epochs. The proposed model make use of three optimization parameters while training for better model training, weight and biases updation, and better convergence of the loss function.

The parameters used are:

1) **ReduceLROnPlateau:** This is a callback function that reduces the LR (Learning Rate) when the losses or metrics stops improving. Metrics referred can be 'validation loss metrics', 'accuracy metrics', 'loss metrics', etc. Model parameters are better optimized by reducing the LR by a factor of 2-10 decimal points if the monitoring metrics stops to improve. This updates the model parameter to even smaller floatingpoint values which can't be done by specifying a general LR. ReduceLROnPlateau takes 3 parameters are input:

Parameters	**Values**
Monitor	validation loss
Patience	5
Factor	0.1

2) **EarlyStopping:** Early stopping is a regularization technique that tend to stop the model training if the monitored metrics stops improving. This is done to tackle problem of overfitting as too many epochs can lead to overfitting and even ensure optimum computation efficiency by stopping the training early rather than going for arbitrary number of epochs. The EarlyStopping parameters specified are:

Parameters	**Values**
Monitor	validation loss

Patience	10
Mode	auto

3) **Adam Optimizer:** Adam refers to Adaptive Moment Estimation Optimizer. This algorithm is used to optimize gradient descent technique using Adaptive Gradient Algorithm (AdaGrad) and Root Mean Squared Propagation (RMSProp).

   a. AdaGrad maintains the pre-parametrized LR and helps to converge the gradient descent algorithm towards the minima using exponential weighted averages.

   b. RMSProp uses adaptive pre-parametrized LR by taking the averages of the recent gradient as weights for updation. Thus, it tackles the problem of quick shifting and oscillatory nature of the gradient.

The Adam Optimizers parameters specified are:

Parameters	**Values**
LR (Learning Rate)	0.001
Epsilon	10^-7
Amsgrad	False
Beta_1	0.9 (default)
Beta_2	0.999

Loss Function (Sparse Categorical Cross Entropy Loss)

Right choice of loss/cost function is necessary for better optimization and convergence of weight and biases to global minimum. In classification of model, there are mainly two type of loss function that are used.

1. **Categorical Cross Entropy Loss:** Cross Entropy loss is by default function for multiclass classification in which each class is mapped to an integer value in the range from 0(number of classes -1). The equation of cross entropy loss calculates sample loss using weighted difference of the "actual and predicted probability distributions for all classes. The score is minimized and a perfect cross-entropy value is 0."

$$Loss = - \sum_{i=1}^{output\ size} y\ i * \log yip$$

where:

yi -> the actual target output yip -> the model predicted probabilistic output

But one of the limitations of cross entropy loss it required the target output to be onehot encoded. Thus, the dataset target values must be one-hot encoded. To tackle this, we used sparse categorical cross entropy loss.

2. **Sparse Categorical Cross Entropy Loss:** In case of a large number of classes, onehot encoding is not feasible and thus categorical entropy loss can't be used. To overcome this, sparse categorical cross entropy loss is used which output just a single probability where truth labels are integer encoded.

$$Loss = - \sum_{i=1}^{output\ size} y\ i * \log yip$$

where:

yi -> the actual target output

yip -> the model predicted probabilistic output

Its loss function is same as categorical cross entropy but the difference is in the output form.

Training Parameters, Accuracy and Losses

The initial training, *figure 6 and table 4*, after pre-processing showed clear signs of overfitting and biasness of a single parameter used for proper estimation and result. The testing pipeline used:

1) Validation Loss: for evaluation of overfitting and underfitting

2) Training Accuracy: curve for steady settling for consecutive epochs representing stable model

3) Precision, Recall, and F1-score: for detection of false negative and false positive.

Figure 6. *a) Model Accuracy curves b) Model Loss curves after first iteration without tuning and parameter updation*

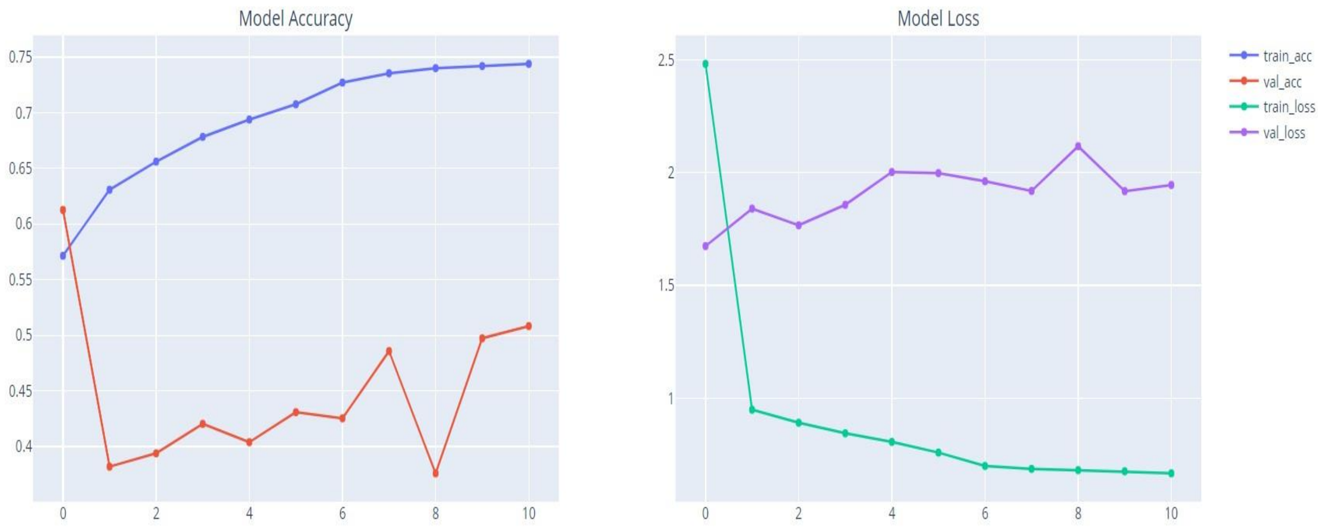


Table 4. Testing Results before optimization

Sr. No	Class of Lesion	Precision	Recall	f1-score
1	Actinic Keratoses	0.00	0.00	0.00
2	Basal cell carcinoma	0.00	0.00	0.00
3	Benign keratosis	0.15	0.01	0.02
4	Dermatofibroma	0.00	0.00	0.00
5	Melanoma	0.21	0.23	0.22
6	Melanocytic nevi	0.67	0.89	0.77
7	Vascular skin lesions	0.00	0.00	0.00

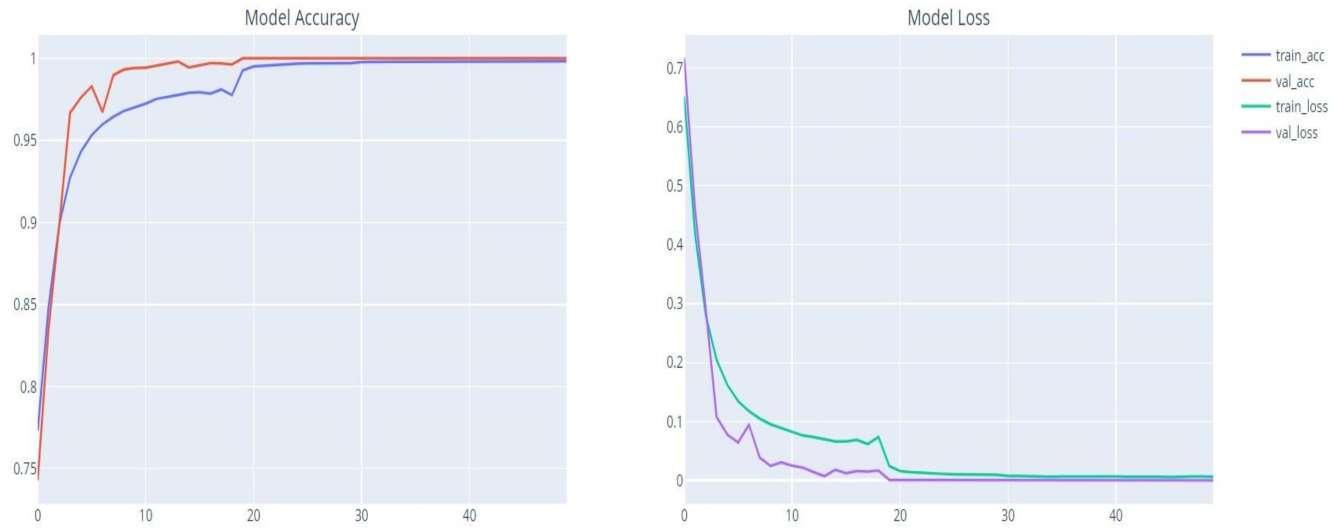
Table 5. Testing Results before optimization CNN (average value)

Sr. No	Entity	Precision	Recall	f1-score
1	Macro Average	0.15	0.16	0.14
2	Weighted Average	0.49	0.61	0.53

The final training, *figure 7*, after pre-processing and optimization showed that the optimization parameters used were able to converge the curves to the desired value. The validation results parameters are:

1) Validation Loss: 4.318x10^-5
2) Training Accuracy: 99.8%
3) Precision, Recall, and F1-score: 1.00, 1.00, 1.00 respectively

Figure 7. *a) Model Accuracy curves b) Model Loss curves after final fine-tuning and parameter updation*



Parameters tuned:

1) LR (Learning Rate): 0.001
2) Loss Function: Sparse Categorical Cross Entropy Loss with cumulative learning and updation tuning strategy
3) Optimizer: Adam (Optimized with beta_1 = 0.9, beta_2 = 0.999 and Epsilon = 10^-7 values)
4) Epochs: Initially 250 epochs
    a. EarlyStopping: Terminated at 52 Epochs with patience 10 5)

Activation function:

  a. ReLu for CNN
  b. Tanh for Dense
  c. SoftMax for output

6) Argmax: Separate function for selection of maximum argument by using cumulative learning results

Results

The training and validation curves of the model settled at the steady value with zero fluctuations clearly states that the problem of overfitting and underfitting has been successfully tackled with excellent training accuracy. At the same time, the loss curves also show a significant dropping from the previous training present in the original dataset. The training accuracy came out to be remarkably *99.8%* with validation losses to converge to the *power of -5 with base 10 (4.318x10^-5)* with reduced utilization of computation power at 0.00677 G Flops.

The *table 4 and table5,* showing the testing results shows the f1-score, recall score, and precision to be 1.00 showing that the model feature extraction with defined layer showed remarkably good results.

Table 6. Testing Results of self-made models

Sr. No	Class of Lesion	Precision	Recall	f1-score
1	Actinic Keratoses	1.00	1.00	1.00
2	Basal cell carcinoma	1.00	1.00	1.00
3	Benign keratosis	1.00	1.00	1.00
4	Dermatofibroma	1.00	1.00	1.00
5	Melanoma	1.00	1.00	1.00
6	Melanocytic nevi	1.00	1.00	1.00
7	Vascular skin lesions	1.00	1.00	1.00

Table 7. Testing Results of self-made CNN (average value)

Sr. No	Entity	Precision	Recall	f1-score
1	Macro Average	1.00	1.00	1.00
2	Weighted Average	1.00	1.00	1.00

The *table 6* compares the proposed models training accuracy, validation losses, time for training, epochs before termination, and Flops with various other classification model trained on the same HAM_10015 challenge dataset.

Table 8. Comparison Summary of other Machine Learning models

Sr. No.	Parameters Models	Training Accuracy	Validation Losses	Epochs (Early Stopping)/ Time/ Flops (Floating Point Operation per Second)
1	Proposed Model	99.8%	4.318 x 10^-5	50 / 5 hrs 23 min (0.00677G Flops)
2	XGBoost (eXtreme Gradient Boosting)	73.2%	0.842	39 / 3 hrs 47 min
3	DCNN's	97.4%	0.782	47 / 7 hrs 39 min (0.85 Flops)
4	Attentions Based Model (Vgg16 derived) IRV2+GCAM	92%	3x10^-3	123 / 29 hrs (23.6 Flops)
5	Random Forest	65.9%	0.73	4 hrs 37 min
6	Support Vector Classifier	65.86%	0.043	4 hrs 02 min

Thus, it clearly states that the model shows comparable or even surpass the accuracies of some the complex models like VGG16, DCNN's, and ensemble models like Random Forest.

CONCLUSION

The paper proposes a moderately complex and low computation resource method for classification of pigmented skin lesions consisting of modified CNN model comprising of Dropout layers for reducing overfitting, sparce categorical entropy loss function for better tuning of the weights and finely tuned optimization algorithms like EarlyStopping , ReduceOnPlateau for tacking problem of computational efficiency. This proposed method achieved low computation resources with high performance on an unbalanced and small dataset by creation of diversity in the skin disease type with the image processing and data augmentation techniques, transforming the original dataset to a more complex form for achieving a lot more feature that could not be possible with original dataset. Also, CNN model trained on HAM_10000 dermoscopic image dataset shows that a moderately complex with optimum tuned parameter, like loss function, optimizer, learning rate, selection layers, model can outperform even the larger ones. Image processing and Image augmentation technique helps not only to reduce and structure images but also to balance the imbalances in the data and provide varied features for improved extraction for improved accuracy. Thus, the model achieved excellent classification accuracy on the dataset surpassing the accuracies of ensembling models. Also, the low computation requirement makes the model best suited for the web, and application deployment requiring high accuracies with low computation processing and low memory on the cloud deployment.

References

1. Peng Yao, Shuwei Shen, Mengjuan Xu, Peng Liu, Fan Zhang, Jinyu Xing, Pengfei Shao, Benjamin Kaffenberger, and Ronald X. Xu: Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification Preprint at https://arxiv.org/abs/2102.01284
2. Gao, Zhimin, et al. "HEp-2 cell image classification with deep convolutional neural networks." IEEE journal of biomedical and health informatics 21.2 (2016): 416-428
3. Rosendahl, C., Tschandl, P., Cameron, A. & Kittler, H. Diagnostic accuracy of dermatoscopy for melanocytic and nonmelanocytic pigmented lesions. J Am Acad Dermatol 64, 1068–1073 (2011).
4. Binder, M. et al. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. Br J Dermatol 130, 460–465 (1994).
5. Codella, N. C. F. et al. 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). Preprint at https://arxiv.org/abs/1710.05006 (2017).
6. Deng, J. et al. ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248–255 (2009).
7. Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S. & Rozeira, J. PH2 - A dermoscopic image database for research and benchmarking, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, 2013, pp. 5437–5440 (2013).
8. Argenziano, G. et al. Interactive Atlas of Dermoscopy (Edra Medical Publishing and New Media: Milan, 2000).
9. Han, Seung Seog, et al. "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm." Journal of Investigative Dermatology 138.7 (2018): 1529-1538
10. Haenssle, Holger A., et al. "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists." Annals of Oncology 29.8 (2018): 1836-1842
11. Dreiseitl, S., Binder, M., Hable, K. & Kittler, H. Computer versus human diagnosis of melanoma: evaluation of the feasibility of an automated diagnostic system in a prospective clinical trial. Melanoma Res 19, 180–184 (2009).
12. The HAM10000 dataset, a large collection of multi-sources dermatoscopic images of common pigmented skin lesions https://nature.com/articles/sdata2018161
13. ISIC Archive https://www.isic-archive.com

14. Harvard Dataverse https://www.dataverse.harvard.edu/dataset