# Performance optimization on skin lesion image classification using pre-trained models

MSc Research Project
MSc in Data Analytics

Krishnachander Rajendran
Student ID: x18122779

School of Computing
National College of Ireland

Supervisor: Dr. Bahman Honari

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

| | | | |
|---|---|---|---|
| **Student Name:** | Krishnachander Rajendran | | |
| **Student ID:** | X18122779 | | |
| **Programme:** | Data Analytics | **Year:** | 2019-2020 |
| **Module:** | MSc Research Project | | |
| **Supervisor:** | Dr. Bahman Honari | | |
| **Submission Due Date:** | 03/02/2020 | | |
| **Project Title:** | Performance optimization on skin lesion image classification using pre-trained models | | |
| **Word Count:** | 8134 | **Page Count:** | 23 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** ……………………………………………………………………………………………………………

**Date:** ……………………………………………………………………………………………………………

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project,** both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Performance optimization on skin lesion image classification using pre-trained models

Krishnachander Rajendran
X18122779

**Abstract**

*Skin disease classification through CNN has become more sophisticated with the inception of high resolution training image datasets. The automatic classification of skin diseases act as the much needed alternative for the traditional methods such as biopsy and cutaneous examination. The requirement of feature extraction is also not necessary with the CNN as opposed to the automated skin classifiers that were built in the past. This shift in focus is toward CNN enabled the introduction of pre-trained models to be used on image classification as part of transfer learning. Such an approach is followed in this research with the use of MobileNet model and DenseNet201 models along with a custom built CNN. All three models are pitted against each other for performance evaluation with the main criteria being the optimizer being used. Adam optimizer and Stochastic Gradient Descent (SGD) are used to explore the hypothesis that SGD optimizer performs better than Adam optimizer. The accuracy obtained from the research provides evidence that the MobileNet model with the SGD optimizer performs better (91.68 %) than the other models in terms of categorical accuracy.*

# 1 Introduction

Melanoma, since the 1970s, has seen an increase of many fold in the coming decades (Ruiz et al., 2016). In Australia, Melanoma is the major type of cancer that exists (Green et al., 1991). The prevalence of Melanoma has been increasingly observed in countries of many white-skinned populations around the globe (Green et al., 1991). Improvements in diagnosis of skin lesions make use of dermatoscopic techniques as opposed to the manual examination that includes human elemental error (Tschandl et al., 2018). In 2011, a WHO report showed that 132,000 cases of melanoma skin cancer occurred each year (Ruiz et al., 2016), whereas the number of non-melanoma cases accounted for over 3 million each year globally (Chaturvedi et al., 2019). Ultra-violet (UV) rays are the primary cause for 90% and 86% of non-melanoma and melanoma respectively.

The likelihood of reducing mortality rate due to skin cancer greatly increases with early detection screening and diagnosis of skin cancer (Rebecca et al., 2017). A 99% survival rate of 5 years is observed when melanoma patients were found to be treated before the cancer reached the lymph nodes[1]. Several steps need to be considered for performing cutaneous examinations in order to detect the type of skin cancer (Kopf et al., 1995). There are diagnostic techniques

[1]https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures2016.html;%202016.

that are extremely painful and pose discomfort for several days after the day of diagnosis (Grapengiesser et al., 2002). The use of ice packs and analgesics that are of non-narcotic nature are often recommended (Allison et al., 2006).

In 1994, dermatoscopic images were trained on an artificial neural network for identifying melanomatic cancer from melanocytic nevi (Binder et al., 1995). Dermascopy, however, is another technique that uses visual inspection by magnifying skin and removing surface reflection (Codella et al., 2018). Non-invasive method such as this allows visual inspection over structures, lesions and outgrowths present on the epidermal region.

Research works have been published verifying the diagnostic accuracy of dermascopy to range from 75% to 84%, given proper training was first established (Codella et al., 2018). Since advanced melanoma is considered to be very difficult to cure, detecting melanoma in its early stages is vital in reducing the mortality (Rubegni et al., 2002).

The literature is structured in the following way: Section 2 speaks about the related work performed in the area of skin lesion classification and image classification closely related with skin diseases. Section 3 provides the approach lying behind the research methodology. Section 4 follows with design specification and section 5 elaborates the implementation of models present in the research work. The evaluation metrics are discussed along with the results of the experiments with a number of models in section 6 followed by section 7 where the conclusion of the research and possible future works are discussed.

### Research Question

Is there a significant impact by the introduction of the stochastic gradient descent to substitute Adam optimizer in training models used for the multi-label classification task of diagnosing skin cancer types?

# 2 Related Work

In the early 1990s, Neural Networks had been newly introduced in the field of medical diagnosis as a form of pattern classifiers (Ercal et al., 1994). Artificial intelligence applications also incorporated neural networks for better performances (Ercal et al., 1994). The belief that images of skin tumours can be used in order to automate a diagnosis system using their physical features and color information was exercised by (Ercal et al., 1994). The majority of the early detection and treatment focus on asymmetry, border irregularity, color and diameter for the differentiation among tumours. Feedforward networks were used for the classification of digital images of tumour into smaller number of categories. Skin lesion images were studied by using spectroscopic oblique-incidence reflectometry (Garcia-Uribe et al., 2004). The set of images having been split into 2 groups were combined with a Bayesian classifier for identifying benign (group 1) from the cancerous (group 2). The spatio-spectral features were combined to form a genetic algorithm that produces the best performance with respect to the classification accuracy.

Some studies have focused on granularity as the main parameter as it is most closely associated with malignant melanoma diagnosis. Optical coherence tomography (OCT) is one of the first few non-invasive diagnosis for skin cancer. It uses infra-red light, photonics and fiber optics. The presence of dark globules and white dots or streaks in OCT corresponded to the presence of BCCs and AKs respectively. However the differences obtained from OCT morphology failed to be sufficient enough to separate the BCC cases from that of AK without machine intervention (Jørgensen et al., 2008).

## 2.1 Use of traditional systems in skin image classification

Researchers in Spain have proposed a solution of detecting the skin cancer cells based on ABCD (Asymmetry, Border Irregularity, Color and Diameter) protocol where the dimensions of specific area in a mole is obtained and analyzed using epiluminescence methods (Zapirain et al., 2007). The image processing here not only does external measurements but also internal measurements within the mole for accurate diagnostics. The External contour measuring involves image filtering and equalizing, and a high pass filter to improve the contrast and edges of the mole (Zapirain et al., 2007). For the Internal contours, RGB decomposition is performed, followed by extraction of internal contours. The area, parameter, and perimeter are calculated for the internal moles that assist in diagnosing if the lesion is benign or malignant. This process is faster, and the image analysis takes less than 1500 ms (Zapirain et al., 2007). The author has used multiple Image enhancement techniques that are essential to compute the parameters of the lesions of interest (Zapirain et al., 2007). A model that performs internal image pre-processing before classification will further improve the analysis rate, for quicker diagnosis (Zapirain et al., 2007).

In Saudi paper, a hybrid technique using artificial neural network and k-nearest neighbour have been used to classify skin cancer using images (Elgamal, 2013). This work involves a three-stage process where image features are extracted using wavelet transformation to extract wavelet coefficients, followed by dimensionality reduction using Principle Component analysis and supervised machine learning algorithms for classification. It can be observed that this hybrid technique was able to achieve 100% sensitivity and 97.5% accuracy. However, the effect of various other features in disease classification are not explained and the processing time taken for image pre-processing, feature extraction and classification have not been explained in detail as timely diagnosis is critical when it comes to skin cancer treatment (Elgamal, 2013).

In Bombay, the author explains a novel approach that involves GLCM to extract features and then using a Support Vector Machine for illness classification (Ansari and Sarode, 2017). This work does image pre-processing in three levels beginning with Grayscale conversion, Noise remove and Image enhancement. Maximum entropy thresholding is used for image segmentation where the foreground and background are separated for feature extraction (Ansari and Sarode, 2017). In this paper, features such as Mean, Homogeneity, Energy, and contrast are extracted for further classification. The Support Vector Machine model can accurately classify cancerous images with an accuracy of about 95%. The effects of various kernels like radial basis, polynomial, and Gaussian within the Support Vector machine has not been evaluated in this paper (Ansari and Sarode, 2017).

Another research work discusses breast cancer detection using a fuzzy based noise removal filter and background removal. This technique then extracts 8 different features followed by classification using SVM and MLP (Jaffar et al., 2009). In Pre-processing, noise removal is carried out, followed by Background removal and Histogram equalization for image enhancement (Jaffar et al., 2009). Texture is basically described by grey level primitives and spatial organization of the greay level primitives (Haralick et al., 1973).

In Sri Lanka, three different classes of skin diseases are identified using captured images (Amarathunga et al., 2015). Firstly, the captured images are enhanced by eliminating noises using Median filtering and Gaussian smoothing (Amarathunga et al., 2015). After this, the author discusses about histogram equalisation and image segmentation to separate the region of Interest so the machine learning algorithm can be much effective in classification (Amarathunga et al., 2015). This paper uses algorithms such as Naïve Bayes, AdaBoost, J48 (Decision Tree), Multi-layer perceptron, and BayesNet. It was observed that multi-layer Perceptron has the highest accuracy in classifying the correct disease, followed by J48 classifier (Amarathunga et al., 2015).

## 2.2 Use of ANN in Skin disease classification

Aswin et al, (2014) explored on the need for early detection system for Melanoma condition. As opposed to the conventional method like clinical biopsy, the computer aided system proves to be painless, cuts down the time constraint and is more economical. Digital image processing along with the application of artificial intelligence is used for the classification task. Noise removal and border refinement were included as part of the digital image processing along with noise reduction using mean filtering (Aswin et al., 2014). A simple neural network architecture is constructed by using color segmentation and feature extraction using Gray level co-occurrence matrix (GLCM) and Red-Blue-Green (RBG) color features for the final classifier.

In 2016, there was a study carried out with the blob detection also known as background subtraction which primarily removes the foreground in the image for further processing, this combined with edge detection and masking made up the image segmentation process. In the classification stage of the artificial neural network architecture, the feed forward architecture is present with three layers attached. The output from the network is of singular nature and informs about the presence or absence of cancer (Kanimozhi and Murthi, 2016).

SLP-ANN also was used along with the input nodes for the automated diagnosis system (Rubegni et al., 2002). A feature selection of the best parameters is carried out for greater power of generalization (Rubegni et al., 2002).

## 2.3 Use of MLP in Skin cancer classification

Several techniques were introduced for improving accuracy in models required to train and test skin cancer diagnosis (Sheha et al., 2012). In 1987, the dermoscope, also known as epiluminescence microscope (ELM) was mentioned by as one of the very first non-invasive diagnostic techniques (Litaker, 1992).

Tissue counter analysis (TCA), in 2003, was used for the partitioning of image into equal set of square shaped elements for the calculation of grey level histogram and co-occurrence matrix. A few different types of classifiers of ANN such as Area multi-layered perceptron, a Bayesian classifier and k-nearest neighbour algorithm were proposed in 2011 (Ruiz et al.,2011)). This study consisted of 3 stages of execution with detection, description, and decision. The first stage, detection allows the image pre-processing of the input along with the selection of the region of interest. Feature extraction with form, border and colour occur at the description stage along with the designated feature selection, which is the connected architecture of K-nearest neighbour method, forward and backward network. This combination along with the multi-layer perceptron contributes to the voting collaborative system for performing the classification task to evaluate the diagnosis.

A sample of 102 dermascope images were used for the performance evaluation of multiple decision support systems (Sheha et al., 2012). Feature selection enabled the selection of only the important attributes from the feature extraction to provide for better detection accuracy (Sheha et al., 2012). The results carried out by Sheha in 2012 concludes with the finding that co-occurrence matrix along with ANN is a reliable technique for dermascopic image detection. The research paper proposed for the feature extraction using gray level Co-occurrence matrix (GLCM) and Using Multilayer perceptron classifier (MLP) to classify between Melanocytic Nevi and Malignant melanoma as part of a binary classification task. Both the techniques, traditional and automatic MLP were involved in building the classifier to obtain a performance evaluation between the two (Sheha et al., 2012).

## 2.4 CNN in skin image classification

In 2016, research work was carried out in transforming a convolutional neural network that is pre-trained on single resolution to a convolutional neural network that can accommodate multiple resolution as input (Kawahara and Hamarneh, 2016). Auxiliary loss functions and end-to-end optimization were introduced to build the CNN architecture (Kawahara and Hamarneh, 2016). For the pre-trained architecture, the AlexNet model was used (Pham et al., 2018) along with parameters that excluded ImageNet layer for the initial stage of the network in order to accommodate untrained layers to be brought in during the latter stage of the network (Kawahara and Hamarneh, 2016). On passing an image into the model, the resulting output is of size 1x1x4096. The next stage is where the untrained layer is introduced with a much lower resolution of the order 1x1x256. These two layers are vital in bringing down the necessary training parameters. The auxiliary loss functions along with untrained parameters are provided as additional inputs to the initial stage of the network (Kawahara and Hamarneh, 2016). Since the scales in objects are kept as a constant, the various segments of the CNN develop the ability to train alongside properties that are exclusively related to resolution. Hence, the architecture created by Kawahara and Hamarneh in 2016 is found to be compatible with skin image classification.

VGG19 and ResNet50 models were used as part of the transfer learning in comparison with the same models with additional SVM working as a hybrid model were compared by (Kwasigroch et al., 2017). After the necessary resizing of images and augmenting them to meet the requirements of the training dataset for the models, the learning methods to implement the architecture were applied. These networks were trained by using minibatch gradient descent along with Nesterov momentum. There is evidence observed from this study, a contradiction to the general belief of ResNet50 performing better than most models, whereas in this case ResNet50 model performs much lower than that of VGG19 and VGG19-SVM models (Kwasigroch et al., 2017).

The importance of data augmentation was observed by (Han et al., 2018) when the performance deep CNN model improved. The ISBI dataset of 2000 melanoma and non-melanoma images is used for the melanoma classification task undertaken in the study performed better with more trained images. At the time, this collection of images were the largest publicly available dataset. Random forest and Support Vector Machine (SVM) are compared against each other with and without data augmentation (Han et al., 2018). The training of images were observed to be better with the models that used the augmented image dataset (Han et al., 2018).

# 3 Research Methodology

## 3.1 Data description and selection

This paper makes use of the dataset prepared by Tschandl et al (2018). Dataset consists of 10015 images and was downloaded from the dataverse website of Harvard university. The motive behind developing such a dataset was to enable neural networks to better train on skin pigmented lesions for automated diagnosis (Tschandl et al., 2018). A wide range of population along with multiple modalities were included in obtaining the HAM10000 (Tschandl et al., 2018). A seven type classification of images make up the HAM10000 dataset, which are Actinic Keratoses (akiec), Basal cell carcinoma (bcc), Benign keratosis (bkl), Dermatofibroma (df), Melanocytic nevi (nv), Melanoma (mel), Vascular skin lesions (vasc) (Tschandl et al., 2018). The small size along with the insufficient diversity present in datasets pertaining to dermatoscopic images has diminished the flow research on neural networks for building an automated diagnosis system of pigmented skin lesions. Figure 1 represents the Actinic Keratoses also known as Solar Keratoses and Intraepithelial Carcinoma which is otherwise called as Bowen's disease. This type of disease is considered non-invasive and does not require surgery (Tschandl et al., 2018), Figure 2 which is of the type Basal cell carcinoma, is a variety of epithelial skin cancer which occurs commonly. It may grow to be dangerous if untreated. The size and shape may vary to be flat, nodular, pigmented or even cystic (Han et al., 2018). Figure 3 is from the type benign keratosis. This type of skin disease includes seborrheic keratoses which are also called as seborrheic keratosis, solar lentigo which is nothing but a flat shaped seborrheic keratosis, and also keratoses that are of lichen-planus type which is a solar lentigo that occurs with inflammation and regression (Zaballos et al., 2010).

**Fig 1: Skin lesions with actinic keratoses, basal cell carcinoma, and benign keratosis**

Figure 4, belongs to the dermatofibroma class which is a benign skin lesion which is generally caused because of inflammatory reaction to minor scale trauma. Reticular lines on the edges with fibrosis in the centre that is denoted by a white patch is the most common visual representation of the dermatofibroma (Zaballos et al., 2008) Figure 5, is of the melanocytic nevi class of skin lesions and are supposedly benign neoplasms of melanocytes. They occur in multiple variants and all such variants are represented in the HAM10000. The degree with which each variant from this class may differ can be very high from a dermatoscopic perspective. There is a striking difference to this class from that of melanoma with reference to the symmetry in the colour distribution and the overall structure (Rosendahl et al., 2012).

**Fig 2: Skin lesions with Dermatofibroma, and Melanocytic Nevi**

Figure 6, belongs to the major type of skin lesions, Melanoma. This type of neoplasm occurring due to melanocytes are malignant in nature and occur in a myriad of variants. A simple surgical intervention may cure the disease if detected in an early stage. All variants except for the subungual, ocular and mucosal melanoma were included in the dataset (6, 7) Figure 7, belongs to the vascular skin lesions that has a wide range of variants from cherry angiomas to angiokeratomas (Zaballos et al., 2007), and pyogenic granulomas (Zaballos et al., 2010). This category of skin lesions also include hemorrhage. The number of images in the dataset correspond to a repetitive set of images taken at different angles and magnification in order to provide further data augmentation (Tschandl et al., 2018).

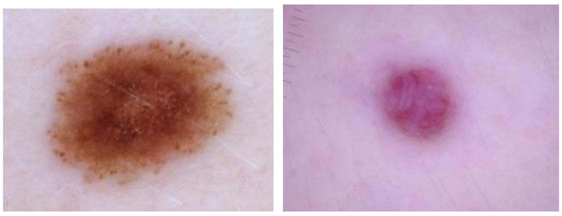

**Fig 3: Skin lesions with Melanoma, and Vascular**

## 3.2 Methodology

A knowledge discovery in databases (KDD) methodology is kept as the fundamental structure for the approach in this project with the clear understanding that the aim of this study lies in classifying the type of skin lesion from the image dataset (Fig 4).

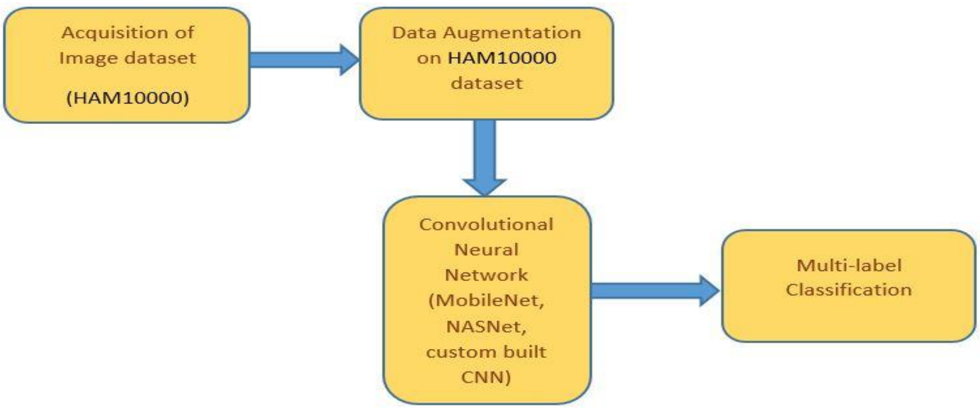


**Fig 4: KDD methodology with CNN implementation**

## 3.3 Data selection

Selecting the compatible dataset is the first step in proceeding with the research methodology. The chosen dataset of Human-against-machine with 10000 training images (HAM10000) fulfils the purpose of providing enough images in order to build the Convolutional neural network architecture.

### 3.4 Data pre-processing and transformation

After selecting the right dataset, the data needs to be cleaned and provided changes so that it is perfectly compatible in bringing out the best performance of the model architecture that are built. Exploratory data analysis is performed on the metadata to understand the dataset and the distribution of the classes of skin lesion images. Duplicate images are identified, and those that do possess duplicates are removed. Data augmentation is performed over the clean set of images by using imagedatagenerator function. This function builds the dataset to a much less skewed representation of the seven classes for the sake of the compatibility with the model.

### 3.5 Selection of models and optimization

In reference to the literature review and previous research works, there is ample evidence of transfer learning models on Convolutional Neural Network working better than artificial neural networks and complex statistical inferences on a small dataset of images.

### 3.6 Evaluation of models

The performance metrics of the models present differentiated by multiple optimizers such as SGD and Adam is compared in solving the multi-label classification task. The data visualization of the performance metrics is performed for comparison.

## 4 Design Specification

**Convolutional Neural Network:** Convolutional networks have obtained great performances with large-scale image and video recognition (Krizhevsky et al., 2012). The reason behind this trait is partly because of the availability of public image repositories (Deng et al., 2009). Several parameters have been used in building various Convolutional Network Architecture pertaining to the requirement of the task. One of the main obstacles that frequently hamper the performance of convolutional neural network models is the feature selection. A multiple range of techniques are available for the task of image classification, and among them, one of the top techniques that provide a stable and strong performance in comparison to the majority of techniques available. The performance evaluation among the models is taken using the classification ability. The fine-tuning, such as increasing or decreasing the width and breadth of the data used in the models, can assist to the betterment of the performance. The correct assumptions are carried out on the image attributes such as pixels and statistics (Krizhevsky et al., 2012). The last layer of the CNN architecture is equipped with the soft-max activation layer of a seven-way type in order to classify the images into seven different types of skin lesions. The underlying principle behind CNN performance lies in the multi-level representation meaning the higher level features provide better information on the data.

**Convolutional layer:** The very basic unit in the construction of a CNN architecture is the Convolutional layer. Convolutional layer possesses an important trait of having rich spatial information embedded on itself. The Height (H) of feature map, Width (W) of feature map, and Depth (D) number of dimensions in feature map provide the visual structure for each region. These layers therefore are also referred to as 2-D arrays of D dimensional local features. In order to classify with ease, the HxW arrays are seen as spatial units (Liu et al., 2015).

**Fully connected layer:** The last section of the CNN is one or more fully connected layers fully connected layers in contrast to the convolutional layers, do not possess spatial information embedded on them. They simply take in the convolutional layers' activations as input and convert them into feature vectors that are a combined representation of the complete image (Liu et al., 2015). This transformation into feature vectors results in the decay of spatial information which basically refers to the inability to recover spatial information from activation functions (Liu et al., 2015).

**Max-pooling layer:** Max pooling layer helps in down-sampling by bringing down the dimensions of feature maps. This layer therefore helps combat overfitting of the sample data

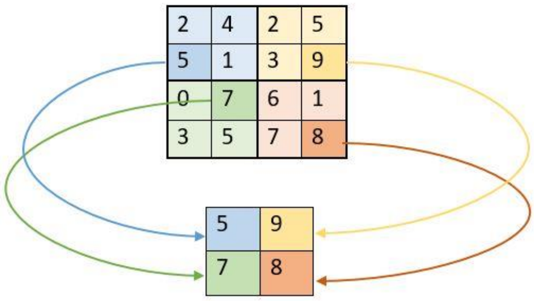


**Fig 5: Max-Pooling sampling**

The Max pooling layer works in such a way that it provides the highest value from among a set of values that are present in the pooling window (Fig 5) (Guo et al., 2017). This form of selecting the maximum value provides greater robustness in feature extraction along with rightly selecting the features by improving the generalization capability (Guo et al., 2017).

**Relu activation layer and dropout:** ReLu activation layer (Fig 6) is used in introducing non-linearity into the network architecture. The principal behind ReLU is that, non-negative values alone are kept in memory and converts all the negative values to zero. The dropout layer is used for shutting down neurons of random seed. This is done to improve the ability of the model to generalize.

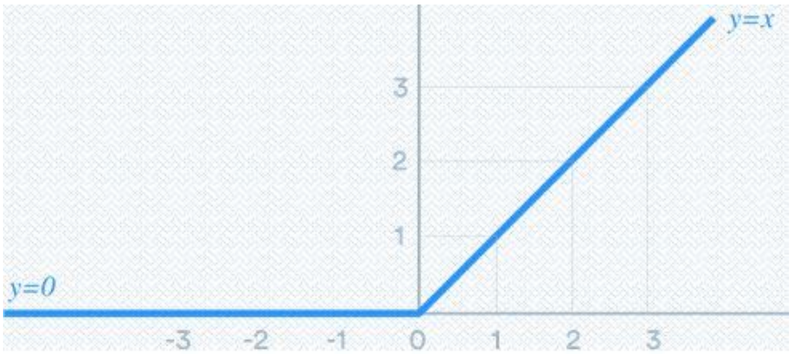

**Fig 6: ReLu activation function**

**Optimization:** The performance of neural networks' classification can be improved using a few changes in the parameters that are being used. One such parameter is the optimizer. Several optimizers are used for the betterment of the models' performances, such as RMSProp, Adagard, Stochastic Gradient Descent (SGD), Adam etc. This research paper provides performance comparisons between the two optimizers Adam and SGD (Stochastic gradient descent). The reliability of a modal is measured by comparing between the output that is expected and the output that is obtained through the model, this comparison is termed as cross-entropy. The cross-entropy decreases in value along with the decrease in the difference between the expected and the obtained output.

The motive behind the optimizer lies in reducing the cross-entropy. The objective of the optimizer is to bring down the cost-function J (x, y) and reach the minimum local by altering the parameters of y and x. During the initial run, random values are assigned to y and x but the gradient descent optimizer captures the right values for x and y with respect to the lowest cost function (Taqi et al., 2018). Among the various optimizers, Adam optimizer has shown the greatest performance in Tensorflow platform for CNN because of its self-learning ability to determine the learning rate best for each parameter within the model (Erhan et al., 2010).

**Learning rate:** is an important hyper-parameter that provides great impact on the learning of a neural network. Optimal learning rate is a necessity and anything lesser than that results in poor convergence between the layers present in the CNN leading to higher computational time. The same is true for higher learning rates as there is a chance of overlooking the minimum and never managing to converge (Amiri et al., 2018). Beginning with a low learning rate and increasing marginally for each run often results in greater performance (Smith, 2017).

**Early stopping** is a hyper-parameter that introduces efficient regularization when the SGD optimizer is involved and is considered to be a good measure (Perronnin et al., 2012). Early stopping provides fast training of the model and great generalization power (Georgieva and Jordanov, 2009). The early stopping keeps track of the progress of classification in each class using a probabilistic model and calculates if there is a need for more computation for further determining ability of the class predicted (Sznitman et al., 2014).

**MobileNet:** Intricacy that is dependent on depth is used in a mobile-net based model, this provides a unique filter to each of the response channels. The construction of the MobileNet model is based on the depth wise distinguishable intricacies. This parameter is a form of complexity that is factorized using depth (Kaiser et al., 2017), the 1x1 complexity is referred to as pointwise complexity. The pointwise complexity gives a 1x1 complexity to combine with the outputs along with the depth wise complexity. Where as in contrast to this, the standard complexity filters both and add inputs to generate a new output set in a single step (Kaiser et al., 2017).The complexity that is distinguishable depthwise, splits this into two. Each of the 2 newly split layers is used for filtering and linking respectively. This procedure of splitting layers provides a great impact in reducing the computational time and also the model size.

**DesneNet201:** DenseNet models (Fig 7) have numerous benefits and few such benefits lie in reducing the problem of vanishing-gradient, feature propagation enhancement, enabling reuse of model, and keeping the number of trainable parameters to low (Huang et al., 2017).

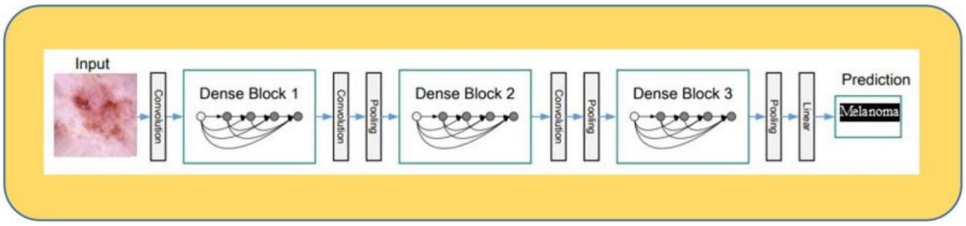


**Fig 7: DenseNet neural network architecture**

The DenseNet201 is also capable of differentiating between the information that gets added and the ones that are preserved (Huang et al., 2017).

# 5 Implementation

## Technologies used:


**Fig 8: Technologies used in this research work**

## Image pre-processing and augmentation:

**Image resizing**: The prerequisite for using the image dataset as input to the CNN architecture with MobileNet model, DenseNet201 model and the custom built CNN model is in maintaining the shape of the images the same as the input shape of the aforementioned models. The images of size 450x600 are resized to 224x224x3 for the models using transfer learning methods and 232x232x3 for the custom built CNN.

**Image filtering:** The filtering of images is done over the whole image dataset. The first step for this is in identifying the number of duplicate images present in the population dataset. Once the number of duplicates is identified by creating a function using python language. From this population dataset, a validation set is created by removing the redundant images present in the population set.

**Image augmentation:** The performance of neural network models with respect to classification accuracy is directly proportional to the number of training images but not always is such the case. The technique that allows small training set to fit into neural network models as an appropriate input dataset is data augmentation. The data augmentation technique involves several parameters that can be used, some of the parameters used here in this study are rotation shift range, width shift range, height shift range, range of magnification, flipping vertically, flipping horizontally, range of brightness, range of colour and fill mode setting to the nearest. This technique also allows the model to stay clear of being overtrained as it brings in minor changes and distortions to the images (Sladojevic et al., 2016).

## Custom built CNN architecture:

| Blocks | Description |
|---|---|
| First block | Convolutional layer with ReLu activation layer and of size (3x3),<br>Number of Kernels = 32 |
| Second block | Convolutional layer with ReLu activation layer and of size (3x3),<br>Number of Kernels = 64<br>Max pooling layer of size (2x2)<br>Drop out = 0.25 |
| Third block | Convolutional layer with ReLu activation layer and of size (3x3),<br>Number of Kernels = 64<br>Max pooling layer of size (2x2)<br>Dropout = 0.25 |
| Dense Layer | Number of Nodes = 128 |
| Output | Softmax Activation with 7 outputs |

**Table 1: Layers of CNN architecture**

The Anaconda platform was used in implementing the research through the Jupyter notebook IDE. The hardware requirements of the system used was of 16 GB RAM, with a graphic card of 6GB in NVIDIA

# 6 Evaluation

This part of the report explains the performance metrics of the models implemented through this research study. The evaluation metrics are compared with respect to accuracy and losses in the models among the various CNN architectures.

## 6.1 Case Study 1

The CNN architecture with Adam optimizer utilizing transfer learning with the MobileNet model used 39,507 images. This set of images were split into 38,569 images for training and 938 images for validation. The number of epochs were chosen to be 30.

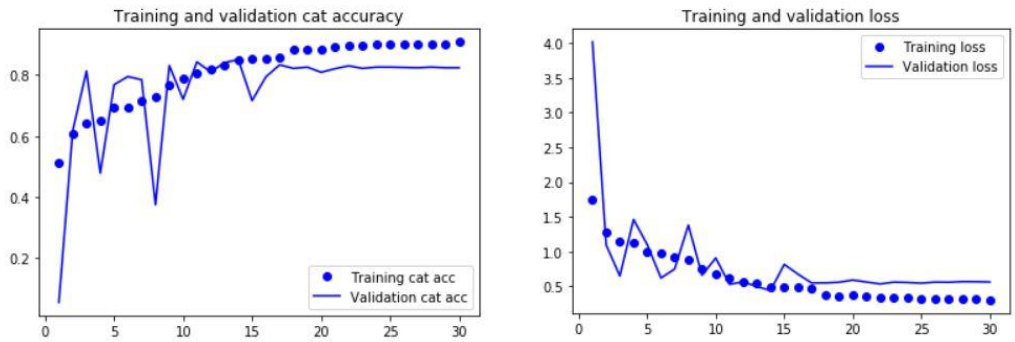


**Fig 9: Accuracy and Loss vs Epochs – MobileNet model with Adam**

The training accuracy for categorical and top_3 categories goes to maximum of 90.28% and 99.18% respectively (Fig 9). The validation accuracy for categorical and top_3 categories were observed to be 85.07% and 97.12% respectively. These minor differences in values of the training and validation results indicate the high power of classification the model is able to achieve. Provided below are the training and validation categorical accuracies against the number of epochs. Along with it is the graph for the training and validation loss that is experienced by the model.

With the increase in the number of epochs, the model appears to be able to reduce both the training and validation loss. This ability allows the model to train well with each epoch run along with the increasing exposure to images, increased robustness is observed in classifying each image to its respective classes. The validation loss reduces from 4.0 to 0.6 and the training loss from 1.8 to 0.4.

## 6.2 Case Study 2

The MobileNet model is used again with a difference in optimization parameter. The Stochastic Gradient Descent optimizer is used to bring out the performance difference in the model as compared to the Adam optimizer. The conditions were set out to be similar to the first case with the exception of learning rate being fixed at 0.001 with 30 epochs. The following graphs represent the performance of the model with SGD. The training accuracy raises from 85% reaching up to a maximum of 91.68% (Fig 10). The converse of the above is observed in the next graph where the training loss and validation loss reach to a minimum of 0.28 each. This performance exceeds that of the loss observed in the first case with the Adam optimizer.

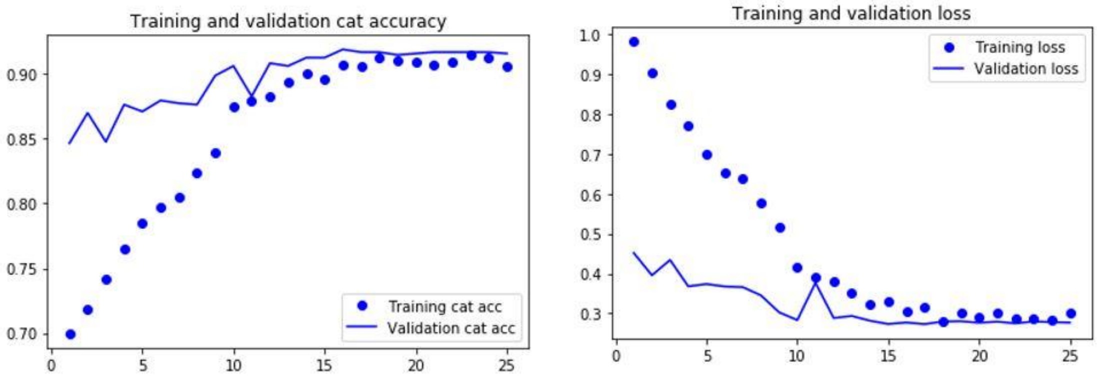


**Fig 10: Accuracy and Loss vs Epochs – MobileNet model with SGD**

## 6.3 Case Study 3

DenseNet201 model with the Adam optimizer performs very poorly as the validation loss and the validation accuracy both increase with decrease in training loss and validation loss suggesting the weakness of the model in classifying the skin lesion images to its appropriate class. The validation accuracy is measured at 64.3% whereas 78% training accuracy is observed for the DenseNet201 model (Fig 11).

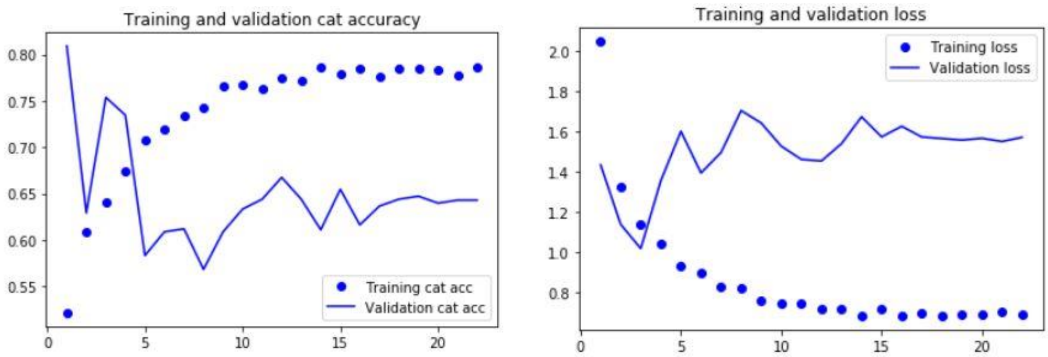


**Fig 11: Accuracy and Loss vs Epochs – DenseNet201 model with Adam**

## 6.4 Case Study 4

DenseNet201 model with the SGD optimizer at 0.0006 learning rate produces results where the validation loss appears to be much higher than the validation loss however the training accuracy and validation accuracy are at 84.5% and 82% respectively reflecting on the model's decency in training with the image data (Fig 12).

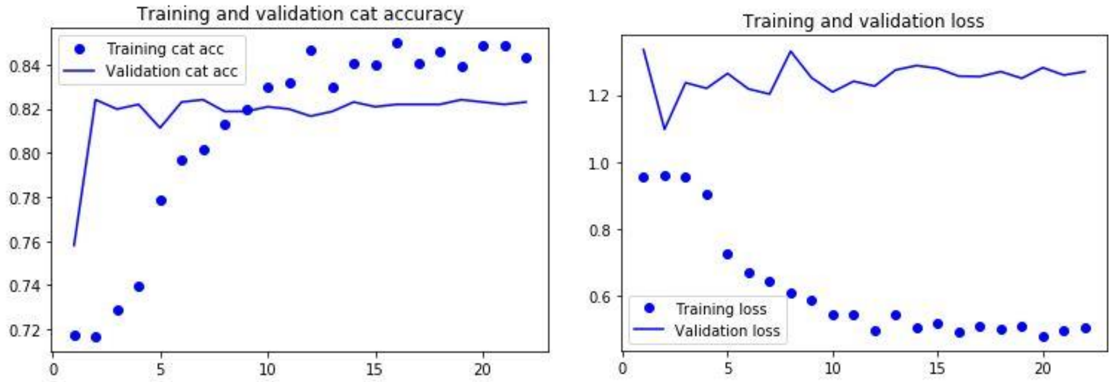


**Fig 12: Accuracy and Loss vs Epochs – DenseNet201 model with SGD**

## 6.5 Case Study 5

The custom built CNN model utilizes the Adam optimizer. The training accuracy and validation accuracy kept increasing with each epoch and there was only little evidence of a plateau to suggest stability of the model on the graph and so the number of epochs were increased to 50 to observe further the training of the model. The training and validation accuracies reached a maximum of 78% and 76% respectively (Fig 13). The training and validation losses reached 0.75 and 0.76 respectively (Fig 13).

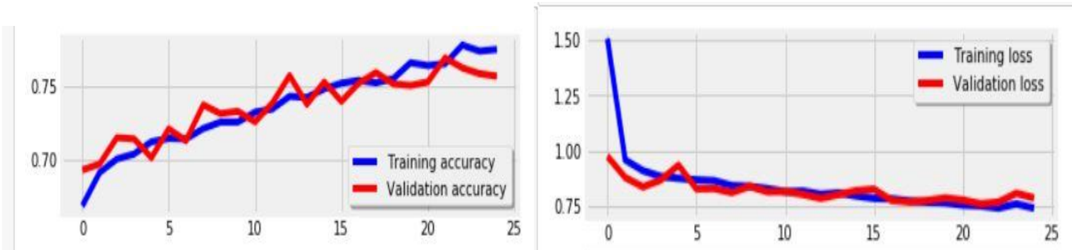


**Fig 13: Accuracy and Loss vs Epochs (25) – Custom CNN with Adam**

The graph below shows the model implemented at 50 epochs. Though the training accuracies improve to over 80%, the validation accuracy remains to be at 78% (Fig 14). The same is observed with the losses as the training loss appears to go down but the validation loss is maintained at a constant level of 0.76 (Fig 14).

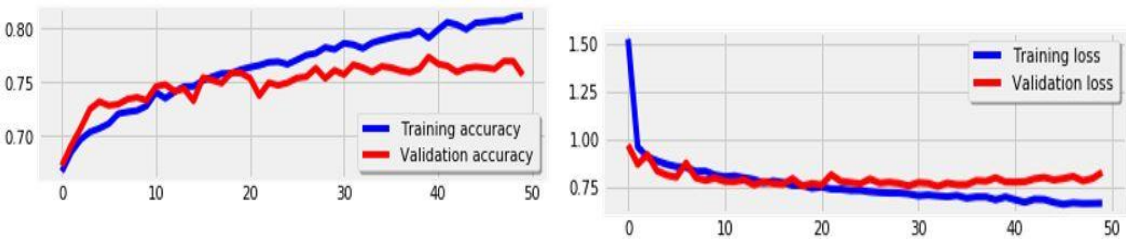


**Fig 14: Accuracy and Loss vs Epochs (50) – Custom CNN with Adam**

## 6.6 Case Study 6

The same as in Case Study 5, the model CNN is kept the same but the Adam optimizer is substituted for the SGD optimizer at 25 epochs. The value of the Training accuracy is observed to reach a maximum of 70.5% with the validation accuracy touching 69% (Fig 15). The training loss and validation loss kept decreasing to finally sit at 2.4 and 2.39 respectively.

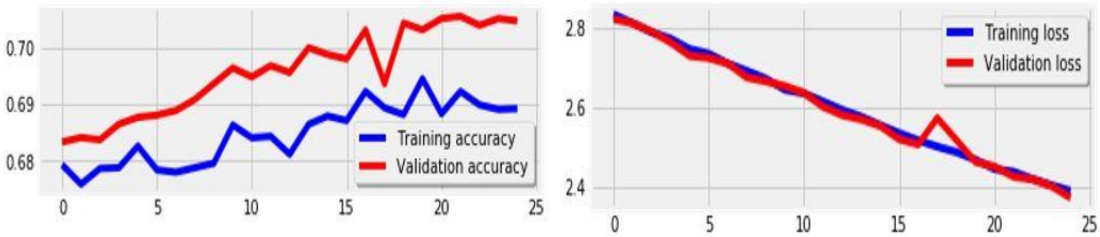


**Fig 15: Accuracy and Loss vs Epochs (25) – Custom CNN with SGD**

The model at 50 epochs provided similar results as it did with 25 epochs. The training and validation accuracies were observed to be close with each other suggesting the model's effectiveness is very good. The training loss however reached the same point as it did with the model at 25 epochs (Fig 16). The increase in number of epochs did not provide enough change to the models' training on the image dataset.

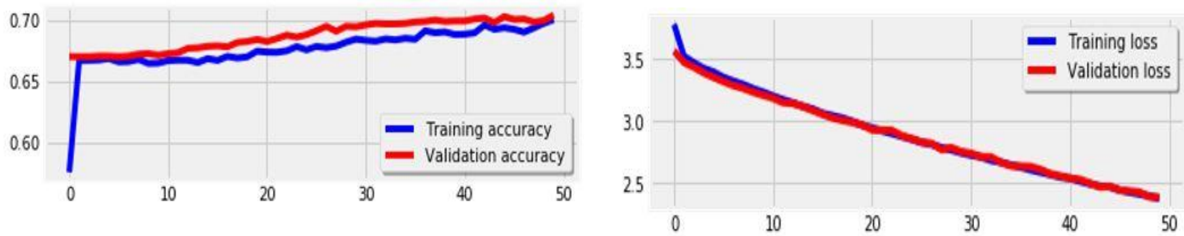


**Fig 16: Accuracy and Loss vs Epochs (50) – Custom CNN with SGD**

## 6.7 Discussion

When deep learning is exercised, a major setback lies in the lack of training data to allow the model to learn patterns. As the pre-requisites for deep learning were not satisfied, it was difficult to perform on skin image classification and research works in the past have stuck with a few hundred images at the most (3). 326 digital images were used to obtain a classification accuracy of 80% using an ANN with a feed forward network (3). When the HAM10000 dataset was used along with deep learning the top_3 accuracy reached a maximum of 95.34% (4).

For building an image classifier to meet requirements in by-passing the problem, several neural network architectures were experimented and the Convolutional Neural Network prevailed as the top performing architecture with respect to image classification in particular. In this research, three models with different optimizers were used, two of the three models followed the transfer learning method where the models were pre-trained with millions of images. The MobileNet model and DenseNet201 models possessed 3,214,151 and 20,013,928 trainable parameters respectively. The custom built CNN contained 352,263 trainable parameters. All three models were repeatedly implemented with Adam and SGD optimizers.

The SGD optimizers performed better than Adam optimizer on the CNN used with transfer learning through the MobileNet models and DenseNet201 models. Adam optimizer performed better than its counterpart, SGD in the case of the custom built CNN.

# 7 Conclusion and Future Work

From the plethora of image classification research works pertaining to skin diseases, the previous methods that made use of simple statistical techniques and simple neural networks worked well with a limited supply of training data. The spatial orientation and lighting observed in images also provide misinformation to the model's training, therefore further emphasizing on quality of training data for the above methods. The neural networks however, experienced discrepancies when used with methods from past researches in the area feature selection and feature extraction prior to the model being trained. These challenges and the popularity of deep neural networks coincide in time. The challenge spoken about was simply removed from the equation by deep neural network architecture as it did not require feature selection and feature extraction to play a separate part prior to the model's training.

Several models designed to classify skin lesion images to what disease it belongs to are in requirement of heave computational infrastructure and to be connected online through cloud platforms for any kind of real time analysis in the health industry. A prior check up on the machines classifying before being consulted by the dermatologist provides a lot of saving in time. Recent researches provide evidence of a new technology called CapsuleNet that tackles the problems of identifying feature relationship, requiring large sets of training data and feature detectors that are of scalar nature, all of which are experienced in CNN. However, the capsule networks do need an image dataset with high quality resolution in order of 1000s.

# References


Allison, R.R., Sibata, C.H., Downie, G.H. and Cuenca, R.E., 2006. A clinical review of PDT for cutaneous malignancies. *Photodiagnosis and Photodynamic Therapy*, *3*(4), pp.214-226.

Amarathunga, A.A.L.C., Ellawala, E.P.W.C., Abeysekara, G.N. and Amalraj, C.R.J., 2015. Expert system for diagnosis of skin diseases. *International Journal of Scientific & Technology Research*, *4*(01), pp.174-178.

Amiri, Z., Hassanpour, H., Khan, N.M. and Khan, M.H.M., 2018, August. Improving the Performance of Multilayer Backpropagation Neural Networks with Adaptive Leaning Rate. In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-4). IEEE.

Ansari, U.B. and Sarode, T., 2017. Skin cancer detection using image processing. *Int Res J Eng Technol*, *4*(4), pp.2875-2881.

Aswin, R.B., Jaleel, J.A. and Salim, S., 2014, July. Hybrid genetic algorithm—Artificial neural network classifier for skin cancer detection. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)* (pp. 1304-1309). IEEE.

Binder, M., Steiner, A., Schwarz, M., Knollmayer, S., Wolff, K. and Pehamberger, H., 1994. Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study. *British Journal of Dermatology*, *130*(4), pp.460-465.

Cancer facts & figures 2016. Atlanta, American Cancer Society 2016. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures2016.html; 2016. [Accessed: 31-Mar-2019]

Chaturvedi, S.S., Gupta, K. and Prasad, P., 2019. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet. *arXiv preprint arXiv:1907.03220*.

Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H. and Halpern, A., 2018, April. 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). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168-172). IEEE.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

Elgamal, M., 2013. Automatic skin cancer images classification. *IJACSA) International Journal of Advanced Computer Science and Applications*, *4*(3), pp.287-294.

Ercal, F., Chawla, A., Stoecker, W.V., Lee, H.C. and Moss, R.H., 1994. Neural network diagnosis of malignant melanoma from color images. *IEEE Transactions on biomedical engineering*, *41*(9), pp.837-845.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P. and Bengio, S., 2010. Why does unsupervised pre-training help deep learning?. *Journal of Machine Learning Research*, *11*(Feb), pp.625-660.

Garcia-Uribe, A., Kehtarnavaz, N., Marquez, G., Prieto, V., Duvic, M. and Wang, L.V., 2004. Skin cancer detection by spectroscopic oblique-incidence reflectometry: classification and physiological origins. *Applied Optics*, *43*(13), pp.2643-2650.

Georgieva, A. and Jordanov, I., 2009. Intelligent visual recognition and classification of cork tiles with neural networks. *IEEE Transactions on neural networks*, *20*(4), pp.675-685.

Grapengiesser, S., Gudmundsson, F., Larkö, O., Ericson, M., Rosen, A. and Wennberg, A.M., 2002. Pain caused by photodynamic therapy of skin cancer. *Clinical and experimental dermatology*, *27*(6), pp.493-497.

Green, A., Martin, N., McKenzie, G., Pfitzner, J., Quintarelli, F., Thomas, B.W., O'Rourke, M. and Knight, N., 1991. Computer image analysis of pigmented skin lesions. *Melanoma Res*, *1*(4), pp.231-236.

Guo, T., Dong, J., Li, H. and Gao, Y., 2017, March. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(* (pp. 721-724). IEEE.

Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I. and Chang, S.E., 2018. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, *138*(7), pp.1529-1538.

Haralick, R.M., Shanmugam, K. and Dinstein, I.H., 1973. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), pp.610-621.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Jaffar, M.A., Ahmed, B., Hussain, A., Naveed, N., Jabeen, F. and Mirza, A.M., 2009, December. Multi domain Features based Classification of Mammogram Images using SVM and MLP. In *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)* (pp. 1301-1304). IEEE.

Jørgensen, T.M., Tycho, A., Mogensen, M., Bjerring, P. and Jemec, G.B., 2008. Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography. *Skin Research and Technology*, *14*(3), pp.364-369.

Kaiser, L., Gomez, A.N. and Chollet, F., 2017. Depthwise separable convolutions for neural machine translation. *arXiv preprint arXiv:1706.03059*.

Kanimozhi, T. and Murthi, A., 2016. Computer aided melanoma skin cancer detection using artificial neural network classifier. *Singaporean Journal of Scientific Research (SJSR), Journal of Selected Areas in Microelectronics (JSAM)*, *8*(2), pp.35-42.

Kawahara, J. and Hamarneh, G., 2016, October. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In *International Workshop on Machine Learning in Medical Imaging* (pp. 164-171). Springer, Cham.

Kopf, A.W., Salopek, T.G., Slade, J., Marghoob, A.A. and Bart, R.S., 1995. Techniques of cutaneous examination for the detection of skin cancer. *Cancer*, *75*(S2), pp.684-690.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Kwasigroch, A., Mikołajczyk, A. and Grochowski, M., 2017, August. Deep neural networks approach to skin lesions classification—A comparative analysis. In *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)* (pp. 1069-1074). IEEE.

Litaker, M.S., 1992. Clinical predictors of malignant pigmented lesions. *J Dermatol Surg Oncol*, *18*, pp.22-6.

Liu, L., Shen, C. and van den Hengel, A., 2015. The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4749-4757).

Perronnin, F., Akata, Z., Harchaoui, Z. and Schmid, C., 2012, June. Towards good practice in large-scale learning for image classification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3482-3489). IEEE.

Pham, T.C., Luong, C.M., Visani, M. and Hoang, V.D., 2018, March. Deep CNN and data augmentation for skin lesion classification. In *Asian Conference on Intelligent Information and Database Systems* (pp. 573-582). Springer, Cham.

Rebecca, S.M.P.H., Siegel, M.P., Kimberly, D., Miller, M.P.H. and Ahmedin Jemal, D.V.M., 2017. Cancer statistics. *Ca Cancer J Clin*, *67*(27), pp.7-30.

Rosendahl, C., Cameron, A., McColl, I. and Wilkinson, D., 2012. Dermatoscopy in routine practice:'Chaos and clues'. *Australian family physician*, *41*(7), p.482.

Rubegni, P., Cevenini, G., Burroni, M., Perotti, R., Dell'Eva, G., Sbano, P., Miracco, C., Luzi, P., Tosi, P., Barbini, P. and Andreassi, L., 2002. Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, *101*(6), pp.576-580.

Ruiz, D., Berenguer, V., Soriano, A. and SáNchez, B., 2011. A decision support system for the diagnosis of melanoma: A comparative approach. *Expert Systems with Applications*, *38*(12), pp.15217-15223.

Sheha, M.A., Mabrouk, M.S. and Sharawy, A., 2012. Automatic detection of melanoma skin cancer using texture analysis. *International Journal of Computer Applications*, *42*(20), pp.22-26.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience, 2016*.

Smith, L.N., 2017, March. Cyclical learning rates for training neural networks. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 464-472). IEEE.

Sznitman, R., Becker, C. and Fua, P., 2014, September. Fast part-based classification for instrument detection in minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 692-699). Springer, Cham.

Taqi, A.M., Awad, A., Al-Azzo, F. and Milanova, M., 2018, April. The impact of multi-optimizers and data augmentation on TensorFlow convolutional neural network performance. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 140-145). IEEE.

Tschandl, P., Rosendahl, C. and Kittler, H., 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, *5*, p.180161.

Zaballos, P., Carulla, M., Ozdemir, F., Zalaudek, I., Bañuls, J., Llambrich, A., Puig, S., Argenziano, G. and Malvehy, J., 2010. Dermoscopy of pyogenic granuloma: a morphological study. *British Journal of Dermatology*, *163*(6), pp.1229-1237.

Zaballos, P., Daufí, C., Puig, S., Argenziano, G., Moreno-Ramírez, D., Cabo, H., Marghoob, A.A., Llambrich, A., Zalaudek, I. and Malvehy, J., 2007. Dermoscopy of solitary angiokeratomas: a morphological study. *Archives of dermatology*, *143*(3), pp.318-325.

Zaballos, P., Puig, S., Llambrich, A. and Malvehy, J., 2008. Dermoscopy of dermatofibromas: a prospective morphological study of 412 cases. *Archives of dermatology*, *144*(1), pp.75-83.

Zaballos, P., Salsench, E., Serrano, P., Cuellar, F., Puig, S. and Malvehy, J., 2010. Studying regression of seborrheic keratosis in lichenoid keratosis with sequential dermoscopy imaging. *Dermatology*, *220*(2), pp.103-109.

Zapirain, B.G., Zorrilla, A.M., Oleagordia, I.R., Nuñez, G. and Abtane, A., 2009, December. Skin cancer parameterisation algorithm based on epiluminiscence image processing. In *2009 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* (pp. 236-241). IEEE.