METHOD TO CLASSIFY SKIN LESIONS USING DERMOSCOPIC IMAGES

Hemanth Nadipineni[*]
Department of Computer Science
Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram
Chennai, Tamil Nadu, India
`ced16i024@iiitdm.ac.in`
`hemanthnadipineni@yahoo.in`

August 24, 2020

ABSTRACT

Skin cancer is the most common cancer in the existing world constituting one-third of the cancer cases. Benign skin cancers are not fatal, can be cured with proper medication. But it is not the same as the malignant skin cancers. In the case of malignant melanoma, in its peak stage, the maximum life expectancy is less than or equal to 5 years. But, it can be cured if detected in early stages. Though there are numerous clinical procedures, the accuracy of diagnosis falls between 49% to 81% and is time-consuming. So, dermoscopy has been brought into the picture. It helped in increasing the accuracy of diagnosis but could not demolish the error-prone behaviour. A quick and less error-prone solution is needed to diagnose this majorly growing skin cancer. This project deals with the usage of deep learning in skin lesion classification. In this project, an automated model for skin lesion classification using dermoscopic images has been developed with CNN(Convolution Neural Networks) as a training model. Convolution neural networks are known for capturing features of an image. So, they are preferred in analyzing medical images to find the characteristics that drive the model towards success. Techniques like data augmentation for tackling class imbalance, segmentation for focusing on the region of interest and 10-fold cross-validation to make the model robust have been brought into the picture. This project also includes usage of certain preprocessing techniques like brightening the images using piece-wise linear transformation function, grayscale conversion of the image, resize the image. This project throws a set of valuable insights on how the accuracy of the model hikes with the bringing of new input strategies, preprocessing techniques. The best accuracy this model could achieve is 0.886

Keywords Skin Lesion Classification · Deep Learning · Convolutional Neural Network

---

1 Introduction

Skin lesion refers to the abnormality on the skin. Skin lesions may be of type cancerous, allergic, etc. Among these, cancer-causing skin lesions are hazardous to health. Some forms of these cancerous lesions are deadly. Melanoma is considered to be the one with a high mortality rate of 8% among the cancerous lesions. The occurrence rate of melanoma is increasing day by day [1].

Skin cancers are classified majorly into two types. They are

- Melanoma Skin Cancers (MSC),
- Non-Melanoma Skin Cancers (NMSC).

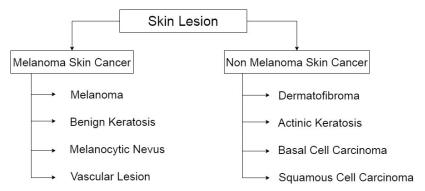In total, there are 8 types of cancerous skin lesions and are classified as follows :

Figure 1: Classification of skin lesions

Clinically, skin cancers are diagnosed based on ABCDE rule [2]. Where,

- A - Asymmetry
- B - Border irregularity
- C - Color of lesion
- D - Diameter of lesion
- E - Enlarging lesion.

This above diagnosis can be done through the naked eye. But most skin cancers mimic each other with ABCDE properties. So, there is a possibility of making an error. And there are other clinical procedures like biopsy, etc and are error-prone. On average, the accuracy of the specialists ranges between 49% to 81%, with one-third of melanomas are wrongly predicted as benign lesions [3, 4, 5, 6, 7]. So dermoscopy came into existence.

Dermoscopy is one of the most important techniques to examine skin lesions and can capture high-resolution images of the skin escaping the interruption of surface reflections. Dermoscopy helped in making more accurate decisions compared to traditional clinical procedures [8].

Dermoscopy increased the sensitivity, specificity but could not eradicate the probability of making errors and all the above clinical procedures are time-consuming, requires human efforts. Early detection of melanoma can save the lives of people. So, a technique is to be created which is fast and less error-prone.

On the other hand, machine learning, especially deep learning-based algorithms have become a methodology of choice for analyzing medical images. A deep convolution algorithm can be more objective, accurate, and reproducible when it has been well trained. So, this related study lead to the conclusion that developing an automated model for skin cancer classification using dermoscopic images can aid us in taking quicker, error-less decisions.

This project deals with developing an automated model for skin lesion classification involves the creation of different models and usage of different strategies. The dataset is class imbalanced, so data augmentation is done using affine transformations(rotate, scale, etc) to generate sufficient numbers of images to tackle class imbalance. Skin lesion images have been segmented to the lesion part so that the model is made to concentrate more on the lesion region and exclude the surrounding healthy skin. 10-fold cross-validation has been brought into the picture. Cross-validation makes the model robust. The accuracy will be the mean of the accuracy of all the folds of data.

1.1 Motivation

- Manual interpretation demands high time consumption and is highly prone to mistakes.
- Early detection and treatment can save more than 95% of people.
- Automated skin lesion classification model makes the work easier and faster.

1.2 Objectives of the work

- To propose a deep learning model to automatically classify skin lesions.
- To validate the proposed model for robustness.

2 Literature Work

Initially, the major work was done on developing an automated model for diagnosing lesions based on ABCDE properties. In these models [9, 10, 11], they converted these ABCDE properties of each image into numerical figures, generated a cumulative score by adding all the numerical figures of each property and diagnose the images based on the cumulative score. There is another developed on a seven-point checked checklist [12], considered to be an improvisation of ABCDE properties.

Later, the work was done on the classification of skin lesion images using Raman spectroscopy. Raman spectroscopy is a noninvasive optical procedure fit for estimating vibrational methods of biomolecules inside practical tissues. In this study, they [13, 14] tested the application of an integrated Raman spectroscopy real-time framework for the in vivo diagnosis of skin cancer

In recent times, the major work was carried out on developing models based on CNN variants (ResNet, EfficientNets, SENet, so on) for classifying skin lesions using dermoscopic images [15, 16, 17]. They preprocessed the images to eliminate the not interesting regions like eliminating dark circles in the images around the lesion region, contrast balancing, etc. They used data augmentation, random cropping strategies for tackling class imbalance. 5-fold cross-validation, Ensembling strategy for making the model robust.

Let's have a broader look at these three models as they include state-of-the-art CNN architectures, intriguing input strategies, and modern robust training techniques.
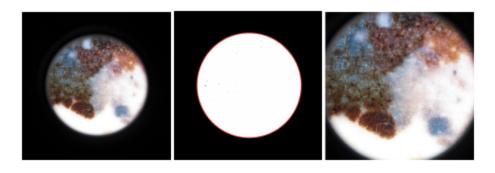
The present models are evaluated on ISIC 2019 dataset

2.1 Input and Expected Output

The input dataset is ISIC 2019 dataset and it consists of dermoscopic images related to '8' different skin lesions. The expected output of the images is a one-hot vector of size '8'

2.2 Preprocessing

There are certain images in which lesion is surrounded by black color. So this part of the image with pixel value '0' place no role in learning the characteristics of the images. So, the darker parts of the image area to be removed. This helps in two ways - i) Helps us concentrate more on a region of interest. ii) Decreasing the size of the image lessens the computation time.

This task is achieved in the following steps: i)Take the image which is surrounded completely by black color. ii) Find a threshold which highlights(make pixel value as '1') the lesion region and darkens(make pixel value as '0') the surrounded black region when the value greater than equal to the threshold is put as '1' and rest as '0'. iii) Find the major axis endpoints and minor axis endpoints of the highlighted regions. extract the rectangular region whose length is equal to the length of the major axis, breadth is equal to the length of the minor axis.

The color contrast of all the images differ a lot so apply Minkowski norm with p = 6 to make the contrast of all the images to as similar as possible

$$d(x, y) = (\sum_{i=0}^{n-1} |x_i - y_i|^p)^{1/p}$$

3 Deep Learning Models

3.1 Architecture

These papers involved state-of-the-art deep CNN architectures for training. They are i) EfficientNets: This family of models includes eight different models that are structurally similar and are pretrained on the ImageNet dataset. ii)SENet154, iii) ResNet.

The input sizes of all these networks are different. The depth of a network increases with the increase in the size of the input image.

The count of images of each class is not the same. When a model is trained with this unequal count dataset, the model gets biased in classifying the images. The model may end up predicting the given image as the class with high image count as it learned more characteristics related to that class. So this data imbalance should be tackled to generalize your model. To augment the data.

3.2 Data Augmentation

The data augmentation is done through affine transformations and it involves: i) Random brightness, ii) contrast changes, iii) random flipping, iv) random rotation, v) random scaling, vi) random shear, vii) CutOut - This involves creating holes of different sizes on the images i.e. technically making a random portion of image inactive.

The additional data of '995' healthy dermoscopic images from the 7-point dataset are also added to the dataset. The data augmentation technique helps in creating additional so that training dataset can be balanced and training the model with wide varieties of skin lesion image variations.

3.3 Input Strategy

The complete training dataset is segregated into five folds. Each fold consists of images related to one particular class. Two ways of passing inputs have been followed: i)Same-sized cropping: Patches of equal sizes are taken from the preprocessed images and are trained. ii) Random-sized cropping: Patches of random sizes are taken from the preprocesses images and are scaled to the required size are trained.

3.3.1 Training

All these models are trained using 'Adam' optimizers, all the other parameters like batch_size and learning are adjusted based on the GPU memory requirements of each architecture.

Internal evaluation is based on the mean values of sensitivities.

$$S = 1/C \sum_{i=1}^{C} TP_i/Tp_i + FN_i$$

where,
TP = True positives, FN = Fales negatives, C = number of classes.

3.4 Ensembling

They create large ensemble out of all the trained models. An optimal subset of models is selected based on cross-validation performance. Top configurations of our choice can be selected and used for prediction.

3.5 Literature Table

Title	Authors	Dataset	Target Classes
Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule	R. Kasmi, K. Mokrani	EDRA [9]	Melanoma, benign skin lesions
Melanoma recognition framework based on expert definition of abcd for dermoscopic images	Q. Abbas, M. Emre Celebi, I. F. Garcia, W. Ahmad.	EDRA [10]	Melanoma
Multi-Category Skin Lesion Diagnosis Using Dermoscopy Images and Deep CNN	Steven Zhou1, Yixin Zhuang, Rusong Meng	ISIC 2019	Melanoma, Melanocytic Nevus Basal Cell Carcinoma, Actinic Keratosis, Benign Keratosis, Dibrofitoma, Vascular Lesion, Squamous cell Carcinoma
Automatic diagnosis of melanoma: a software system based on the 7-point check-list	G. Di Leo, A. Paolillo, P. Sommella, G. Fabbrocini	public [12]	Melanoma
In vivo nonmelanoma skin cancer diagnosis using raman microspectroscopy	C. A. Lieber, S. K. Majumder, D. L. Ellis, D. D. Billheimer, A. Mahadevan-Jansen	IRB(patients) [13]	Melanoma, Melanocytic Nevus Basal Cell Carcinoma, Actinic Keratosis, Benign Keratosis, Dibrofitoma, Vascular Lesion, Squamous cell Carcinoma
Real-time raman spectroscopy for in vivo skin cancer diagnosis	H. Lui, J. Zhao, D. McLean,, H. Zeng	Vancouver General Hospital Skin Care Centre(patients) [14]	Melanoma, Melanocytic Nevus Basal Cell Carcinoma, Actinic Keratosis, Benign Keratosis, Dibrofitoma, Vascular Lesion, Squamous cell Carcinoma
Skin lesion classification using ensembles of multi-resolution efficientnets with meta data	N. Gessert, M. Nielsen, M. Shaikh, R. Werner, A. Schlaefe	ISIC 2019 [15]	Melanoma, Melanocytic Nevus Basal Cell Carcinoma, Actinic Keratosis, Benign Keratosis, Dibrofitoma, Vascular Lesion, Squamous cell Carcinoma

4 Proposed method

This project involves development of three models -

- Model - 1
- Model - 2 :
    - CNN Model - 1 - One Path
    - CNN Model - 2 - Dual Path

where, model - 2 involves two CNN Models. The 'Model-1' takes the original images as input, 'Model-2' with 'CNN Model - 1' takes image formed with the multiplication of original image and segmented mask as an input ans 'Model - 2' with 'CNN Model - 2' takes both original image and segmented mask as inputs. There is the usage of different input strategies, preprocessing techniques, and training techniques. The detailed description is as follows :

4.1 Input and Expected Output

The input is ISIC 2019 dataset and the expected output is 'one-hot vector' of size '8'as explained above.
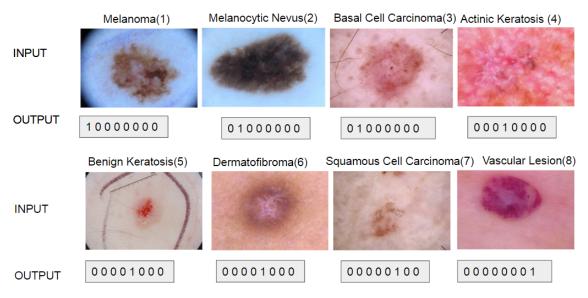

Figure 2: Input and expected ouput of the model developed

The dataset is an accumulation of 3 different datasets.

- Ham_1000 (600 x 450) pixels.
- BCN_2000 (1024 x 1024) pixels.
- MSK (1024 x 1024) pixels.

The total count of images is 25331. The count of images for each class is unequal which leads to class imbalance. Each class images' resolutions differ from other class images.

MLN MCN BCC AK BK DF VL SCC
4522 12875 3323 867 2624 239 253 628

Challenges faced: i) Class imbalance, ii) varying contrast of images, iii) Improper images(lesion color is similar to the skin color), iv) Different resolutions for images.

4.2 Classification Model - 1

The 'Model - 1' involves simple preprocessing of the input image, then pass the preprocessed image as an input to the CNN model, train the model, and start classifying the images.
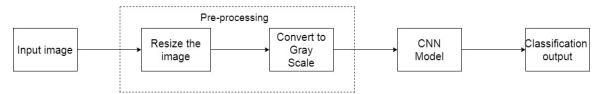

Figure 3: Architecture of Classifcation 'Model - 1'

This model is used to study the effectiveness of CNN in capturing features of an image for classification. The major concentration is on CNN and it's variants as they are good at capturing neighbourhood properties in an image which is essential in differentiating one skin cancer image from another image.

4.2.1 Preprocessing

All the input images are of various sizes so to pass these images as input to the model developed they should be of the same size. So, in this phase, all the images are scaled down to the size 512 x 512 pixels. These images are further converted into grayscale. This is from our intuition that skin lesion classification does not depend on the color of the lesion.
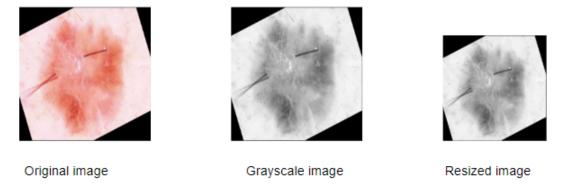

Figure 4: shows how an input image is preprocessed to the required form

4.2.2 CNN Architecture of Classification 'Model - 1'
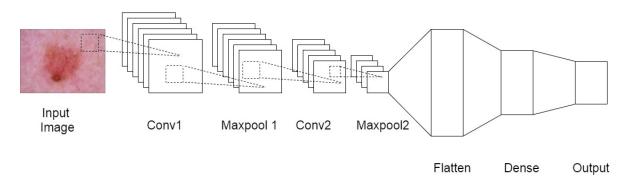
This is the architecture of the developed CNN Model :

Figure 5: CNN architecture of Classification 'Model - 1'

The size of the input image is '512 x 512' pixels and the kernel size is '3 x 3'. In the convolution layer - 1, '64' filters are used which results in the creation of '64' feature maps and size of the output image shrinks to '510 x 510' pixels. Then this output is passed as an input to the max pool layer - 1 with stride '2' for the filters, which results in the creation of '64' feature maps whereas the size of the output images shrinks to the size '255 x 255'. Similarly, the number of feature maps at convolution layer - 2 is '32', the output image size is '253 x 253'. The number of feature maps at max pool layer - 2 is 32, where the size of the output image is '121 x 121'. At flattening stage, the count of neurons is 121 x 121 x 32 = '468512'. The number of a neuron at Dense - 32 layer is '32'. The number of neurons at the output layer is '8'. The model is trained on 90 percent of images with optimizer 'Adams' for '20' epochs. The activation function at

convolution and dense layers are 'Relu'.

$$x = max(0, x)$$

The activation function at output layer is 'softmax'.

$$z_i = e^{x_i} / \sum_{j=1}^{K} e^{x_j}$$

The loss function used is 'categorical cross_entropy'

$$L(\hat{y}, y) = -1/N \sum_{i}^{N} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]$$

The accuracy is 0.204. After going through past work done in the field and clinical procedure for the diagnosis, there are some important issues where model-1 least stressed.

The following can be considered as the drawbacks of the above model: i) Not focussing more on regions of interest, ii) item Color plays a major role in differentiation, iii) No tackling for class imbalance.

4.3 Classification Model-2

This model includes segmentation which helps us in extracting the region of interest and helps us in concentrating more on the region of interest. The input for the model is the same as the input for the model-1.
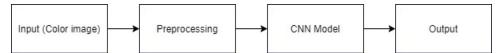

Figure 6: Architecture of Classification Model - 2

4.3.1 Preprocessing

The required lesion region of interest is segmented out from the given dermoscopic images
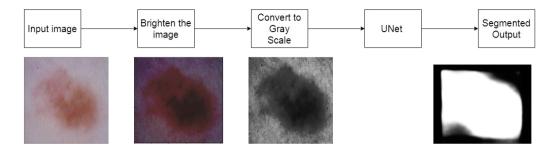

Figure 7: shows how an input image is preprocessed to the required state

Steps involved are :

- **Brighten the image :** Color of some class images is so light that there is no distinction between the colors of skin and the lesion. This task is achieved using the 'piece-wise linear transformation' function.
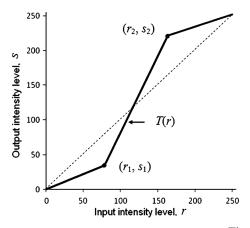

Figure 8: Piece-wise linear function

9

Where,

- – (r1,s1) , (r2,s2) as parameters stretches intensity levels,
- – Decreases intensity of dark pixels,
- – Increases intensity of bright pixels.

Logic for piece-wise linear transformation :

```
def pixelVal(pix, r1, s1, r2, s2):
if (0 <= pix and pix <= r1):
return (s1 / r1)*pix
elif (r1 < pix and pix <= r2):
return ((s2 - s1)/(r2 - r1)) * (pix - r1) + s1
else:
return ((255 - s2)/(255 - r2)) * (pix - r2) + s2
```

- **Grayscale the image :** Convert image to grayscale, this makes the lesion region to stand out with high brightness.

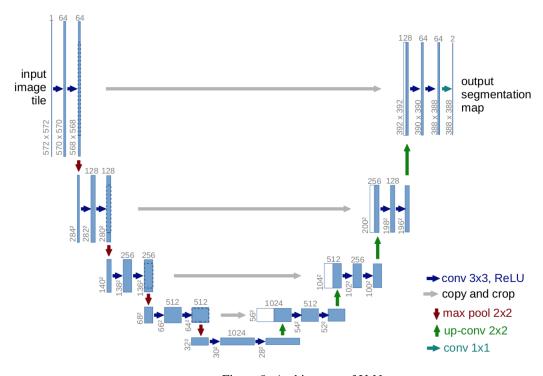- **U-net Architecture :** U-net,which is a CNN variant is the state-of-the-art architecture for segmentation.
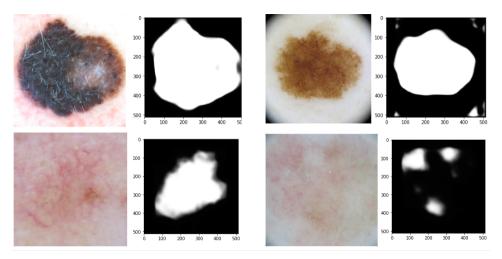

Figure 9: Architecture of U-Net

Sample Outputs :

Figure 10: Input images and their predicted segmentation masks

4.4 Tackling Data imbalance

The count of images for each class is -

MLN	MCN	BCC	AK	BK	DF	VL	SCC
4522	12875	3323	867	2624	239	253	628

The least size is '239'. So picking 239 images from each class, training the model with a dataset of size '239 * 8' images will not be robust. So, additional data is to be created, and for that to happen 'Data augmentation' should be brought into the picture. The list of the classes with count lesser than 2000: AK, DF, VL, SCC

So additional data has been created for the above-mentioned classes using affine transformations. Affine transformations include rotation, cropping, scaling, flipping(horizontal and vertical), shearing, changing contrast.

4.5 K-Fold Cross Validation

K-fold cross-validation technique makes the model robust. This model is made robust through 10-fold cross-validation in which 1-fold is for validation and 9-other folds are for training.

So, the plan is to train the model on the dataset of 16000 images(2000 images from each class). The model is to be trained for 10 times by giving a chance for each fold to act as a validation data set. Find the best model at each time of training and average the accuracies. The mean accuracy will be the accuracy of the model.

4.6 CNN Model

The CNN model section explains about two developed CNN models. Of which one is one pathed and another one is dual pathed.

4.6.1 CNN Model - 1(One path)

This CNN model -1 takes a multiplied image as an input. The multiplied image is created by multiplying the original image with its corresponding segmentation output.
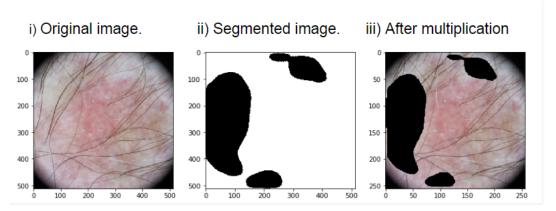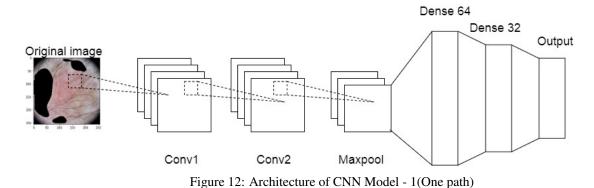
11

Figure 11: Resultant mutliplied image

The architecture of CNN Model - 1 that makes use of this multiplied image.

Figure 12: Architecture of CNN Model - 1(One path)

The size of the input image is '256 x 256 x 3' pixels and the kernel size is '3 x 3'. In the convolution layer - 1, '32' filters are used which results in the creation of '32' feature maps and size of the output image shrinks to '254 x 254' pixels. Then this output is passed as an input to the convolution layer - 2, which results in the creation of '64' feature maps whereas the size of the output images shrinks to the size '252 x 252'. this output is passed as an input to the max pooling layer with stride - '2', which results in the creation of '64' feature maps whereas the size of the output images shrinks to the size '126 x 126'. At flattening stage, the count of neurons is 126 x 126 x 64 = '1,016,064'. The number of neurons at Dense - 64 layer is '64'. The number of neurons at Dense - 32 layer is '32'. The number of neurons at the output layer is '8'.

4.6.2 CNN Model - 2(Two path)

This model takes both the original image and segmented images as inputs, there are parallel similar architectures for processing these inputs into the next levels. These images are combined at the flattening stage and are connected to the dense levels.
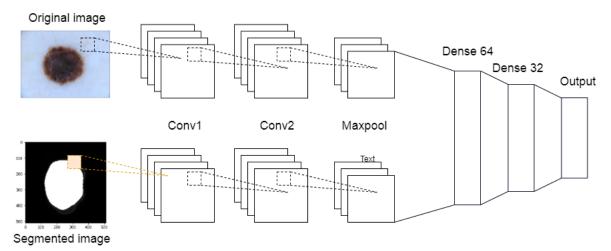
12

Figure 13: Architecture of CNN Model - 2(Two Path)

The output is a one-hot vector which points to the predicted class with value '1'.

The size of the input image passed to (path - 1) is '256 x 256 x 3' pixels and the kernel size is '3 x 3'. In the convolution layer - 1, '32' filters are used which results in the creation of '32' feature maps and size of the output image shrinks to '254 x 254' pixels. Then this output is passed as an input to the convolution layer - 2, which results in the creation of '64' feature maps whereas the size of the output images shrinks to the size '252 x 252'. this output is passed as an input to the max pooling layer with stride - '2', which results in the creation of '64' feature maps whereas the size of the output images shrinks to the size '126 x 126'. At flattening stage, the count of neurons is 126 x 126 x 64 = '1,016,064'.

At the flattening stage 'path - 1' which takes the original image as input produces '1,016,064' neurons. So, till the flattening stage architecture is the same as 'path - 1'. So the number of neurons contributed by 'path - 2' to the flattening stage is also '1,016, 064'. Therefore, the total count of neurons at the flattening stage is '2,032,128'.

The number of the neuron at Dense - 64 layer is '64'. The number of the neuron at Dense - 32 layer is '32'. The number of neurons at the output layer is '8'.

The model is trained on 9-folds of images each time whereas the other one-fold will acts as a validation data set. The process has to be repeated until each acts as a validation set exactly once. The model is trained with optimizer 'Adams' for '2' epochs. The batch_size is 75. The initialization of weights is Keras default initialization i.e. Xavier uniform initialization. The activation function at convolution and dense layers are 'Relu'.

$$x = max(0, x)$$

The activation function at output layer is 'softmax'.

$$z_i = e^{x_i} / \sum_{j=1}^{K} e^{x_j}$$

The loss function used is 'categorical cross_entropy'

$$L(\hat{p}, p) = -1/N \sum_{i}^{N} [p_i \log \hat{p}_i + (1 - p_i) \log(1 - \hat{p}_i)]$$

The accuracy of the model is mean of the accuracies of all the k-models.

$$Acc = (\sum_{i=1}^{K} Acc_i)/K$$

13

5 Results

The following table describes the accuracy of the models with different ways of attempt.

Ways of attempt	Model - 1	Model - 2 with one path -CNN model	Model-2 with two path CNN model
Without Segmentation	0.204	N.A	N.A
With Segmentation	0.412	0.561	0.576
With Segmentation, Augmented data	0.483	0.681	0.732
With Segmentation, Augmented data 10-fold cross validation	0.583	0.814	0.886

Accuracy measure = TP/(TP + FN)
Where,
True positive = predicting the true class correctly,
False negative = predicting the true class incorrectly.

When K-fold cross-validation is used accuracy measure is not just the accuracy of one model but the mean of the accuracies of the K-models.

So, the accuracy of the model with K-fold cross-validation is

$$Acc = (\sum_{i=1}^{K} Acc_i)/K$$

This table shows the comparison of this project with the state-of-the-art models in terms of accuracy.

Model	Accuracy
Skin lesion classification using loss balancing and ensemble of multi-resolution efficient nets [15]	0.926
Multi-Category Skin Lesion Diagnosis using dermoscopic images and deep CNN models [16]	0.917
Classification Model - 2 with two path CNN model (The present project)	0.886
ISIC 2019 Skin lesion analysis towards melanoma detection [18]	0.851
Skin lesion analysis towards melanoma detection using siamese network	0.832

5.1 Inferences

The accuracy of 'Model - 1' is less than in all ways compared to 'Model - 2 with one path CNN model' and 'Model - 2 with two path CNN model'. The explanation for this kind of behaviour can be made from the number of neurons involved in the training of the model. The number of neurons at the flattening layer for i) 'Model - 1' is '4,68,512', ii) 'Model - 2 with one path CNN model' is 1,016,064. iii) 'Model - 2 with two path CNN model' is 2,032,128. But there is a significant difference in the number of neurons between 'Model - 2 with one path CNN model' and 'Model - 2 with two path CNN model' whereas there is no significant difference in the accuracy. So, there is threshold between '4,68,512' - '1,016,064' that surges up the accuracy but there is no threshold between '1,016,064' - '2,032,128'.

There is an improvement in the accuracy of injecting segmentation into the model. This is happening because when the model is trained with an original image without segmentation, the model tends to learn about the unwanted

characteristics of lesion classification. From the [fig 3.1] we can observe that size of the skin lesion region is small compared to the image size. So, to concentrate on the region of interest, the unwanted portion is to be eliminated. So when the image the original is multiplied with the segmented mask, the pixel values of an unwanted portion is set to '0'. So they don't involve in training.

There is an improvement in accuracy when data augmentation is involved. This is because the model is to made to learn on multiple skin variations. So this helped model to learn variety characteristics related to skin lesions. Therefore, it conveys that model with data augmentation tend to learn more distinguishing characteristics features rather like a model without data augmentation.

Involving segmentation, data augmentation, and other preprocessing techniques can improve the accuracy of the model. But that model cannot give significant results with the testing data until it is robust. The model trained can become biased towards a certain class. So, to develop a generalized model, it should be trained with a balanced data set.

6 Conclusion and Future Work

6.1 Conclusion

In this project of developing an automated model for skin lesion classification, major work has done on improving the accuracy of the model by bringing in new techniques into the model.

It is clear from the results, that there is an increment in the accuracy of bringing new strategies into the picture. Though the CNN model is two-layered, it is inferring that the effective preprocessing, input, and training strategies can bring a significant change in the model's accuracy. The segmentation model 'U-net' worked well.

6.2 Future Work

The developed 'model-2 with CNN mode-1' which is of depth two layers produced significant results. But bringing in the deep CNN models will be very helpful in learning the distinguishing characteristics of the skin lesions from dermoscopic images.

Work extensions -

- Bringing in deep CNN variants (EfficientNets, SENet, ResNet, so on)
- Involving ensembling methods can create a significant imapact in improving the accuracy.

References

[1] Alfred W Kopf, Darrell S Rigel, and Robert J Friedman. The rising incidence and mortality rate of malignant melanoma. *The Journal of dermatologic surgery and oncology*, 8(9):760–761, 1982.

[2] Robert J Friedman, Darrell S Rigel, and Alfred W Kopf. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA: a cancer journal for clinicians*, 35(3):130–151, 1985.

[3] Andrew R MacKenzie-Wood, Gerry W Milton, and John W de Launey. Melanoma: accuracy of clinical diagnosis. *Australasian journal of dermatology*, 39(1):31–33, 1998.

[4] Caron M Grin, Alfred W Kopf, Bruce Welkovich, Robert S Bart, and Marcia J Levenstein. Accuracy in the clinical diagnosis of malignant melanoma. *Archives of dermatology*, 126(6):763–766, 1990.

[5] Maria Miller and A Bernard Ackerman. How accurate are dermatologists in the diagnosis of melanoma? degree of accuracy and implications. *Archives of dermatology*, 128(4):559–560, 1992.

[6] CA Morton and RM Mackie. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *The British journal of dermatology*, 138(2):283–287, 1998.

[7] Bernt Lindelöf and Mari-Anne Hedblad. Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic. *The Journal of dermatology*, 21(7):461–464, 1994.

[8] Giuseppe Argenziano, Susana Puig, Zalaudek Iris, Francesco Sera, Rosamaria Corona, Mercè Alsina, Filomena Barbato, Cristina Carrera, Gerardo Ferrara, Antonio Guilabert, et al. Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer. *Journal of Clinical Oncology*, 24(12):1877–1882, 2006.

[9] Reda Kasmi and Karim Mokrani. Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule. *IET Image Processing*, 10(6):448–455, 2016.

[10] Qaisar Abbas, M Emre Celebi, Irene Fondón Garcia, and Waqar Ahmad. Melanoma recognition framework based on expert definition of abcd for dermoscopic images. *Skin Research and Technology*, 19(1):e93–e102, 2013.

[11] Nilkamal S Ramteke and Shweta V Jain. Abcd rule based automatic computer-aided skin cancer detection using matlab. *International Journal of Computer Technology and Applications*, 4(4):691, 2013.

[12] Giuseppe Di Leo, Alfredo Paolillo, Paolo Sommella, and G Fabbrocini. Automatic diagnosis of melanoma: a software system based on the 7-point check-list. In *2010 43rd Hawaii international conference on system sciences*, pages 1–10. IEEE, 2010.

[13] Chad A Lieber, Shovan K Majumder, Darrel L Ellis, D Dean Billheimer, and Anita Mahadevan-Jansen. In vivo nonmelanoma skin cancer diagnosis using raman microspectroscopy. *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery*, 40(7):461–467, 2008.

[14] Harvey Lui, Jianhua Zhao, David McLean, and Haishan Zeng. Real-time raman spectroscopy for in vivo skin cancer diagnosis. *Cancer research*, 72(10):2491–2500, 2012.

[15] Nils Gessert, Maximilian Nielsen, Mohsin Shaikh, René Werner, and Alexander Schlaefer. Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. *MethodsX*, page 100864, 2020.

[16] Quoc V. Le Mingxing Tan. Efficientnet: Rethinking model scaling for convolutional neural networks. *ICML 2019*, 2019.

[17] Giuseppe Argenziano, HP Soyer, V De Giorgi, Domenico Piccolo, Paolo Carli, and Mario Delfino. Interactive atlas of dermoscopy (book and cd-rom). 2000.

[18] Linlin Shen Yuexiang Li. Skin lesion analysis towards melanoma detection using deep learning network. *arXiv preprint arXiv:1703.00577*, 2017.