The Role of Segmentation in Skin Lesion Classification using an ISIC dataset

1st Alen Bhandari
Department of Computer Science of NTNU
Norwegian University of Science and Technology (NTNU)
Gjøvik, Norway
alenbh@stud.ntnu.no

2nd Shaikat Deb Nath
Department of Computer Science of NTNU
Norwegian University of Science and Technology (NTNU)
Gjøvik, Norway
shaikatn@stud.ntnu.no

***Abstract*—Malignanat Melanoma is one of the most dangerous types of skin cancers. So, early detection of this cancer can help the patient recover better. By analysing dermatoscopic images and detecting abnormal skin lesions can help detecting this cancer at the initial phase. Researches have been done to implement several image classification algorithms on skin lesion images. However, there's not much research have been done on investigating the impact of segmentation on skin lesion classification. In our research, we tried to implement several classification algorithms both on segmented and non-segmented images to check the effect of segmentation on the classification. We also tried to explore the performance metrics of different classification algorithms.Our research has shown that, for small dataset the traditional CNN works better when the image is segmented and for the images without segmented the transfer learning algorithm works better compared to other classification algorithms.**

***Index Terms*—Segmentation, U-Net, Classification**

I. Introduction

One of the most common cancer types is skin cancer specially in white population worldwide [1]. Malignant melanoma is the deadliest among several types of skin cancers [2]. One of the most crucial parts of treating skin cancer patient is accurately diagnosing skin lesions to distinguish between benign and malignant skin lesions [3]. The part of the skin that has an unusual growth or appearance compared to the skin around it is called a skin lesion [4]. But, just implementing classification of skin lesion images sometimes can give poor accuracy due to the variance in quality of the images. To enhance the classification we can implement segmentation technique on skin lesions. In image segmentation, the digital image is broken down into various Image segments to help reducing the complexity of the image to make further processing or analysis of the image simpler [5]. Implementing skin lesion segmentation technique is also quite challenging as there can be low contrast between the lesion and its surrounding background, irregular border shapes, fuzzy borders, and fragmentation [6]. However, lesion segmentation will affect the image classification. To evaluate the role of segmentation in lesion classification, we are introducing this project. The basic pipeline that we want to follow in this project is implementing lesion segmentaion techniques, then ,after performing segmentation, skin lesion features can be extracted from the lesion area. Intensity-based features, shape- based features, and textural-based features are among the most used features for skin lesion classification [7]. The extracted features can then used for training classifiers.

For image segmentation, we have chosen Active contour [7] and Chan-Vese [8] as they are two of the most popular image segmentation models.Active contour utilizes the forces and energy constraints of an image for separation of region of interests in that particular image [5].The Chan-Vesen model is a combination between the Mumford-Shah [9] model for segmentation and active contour models [8]. This model performs really good in noisy images [8].We have chosen another deep learning network named U-Net [10] [11] for image segmentation. The reason behind choosing this model is it's efficiency in localising the area of abnormaility and distinguishing borders by doing classification on every pixel [12].So, after implementing U-Net for segmentation, we can compare the model with the best among active contour and chanvessan model. As, the exisiting studies in this research domain has lack of proper comparison between U-Net and traditional segmentation techniques, we are motivated to implement and compare U-Net with other approaches.

To help evaluate the image segmentation models we have used Intersection Over Union(IoU) and Dice Coefficient matrics. IoU matrics quantifies the percent overlap between the target mask and the prediction output [13] whereas the Dice Coefficient matrics [14] is very positively corelated with IoU with a minor difference in the calculations.

For classification, we have followed transfer learning approach as we are using a pre-trained classification model by Google named EfficientNet [15] as it's the most recently developed model with better prediction accuracy compared to exisiting CNNs [16] and Deep Neural Network. To evaluate our classification models we have used recall [10] and F1 score.

So, we can formulate our research questions as follows:

- Does segmentation plays a vital role in improving the accuracy of the classification models?
- Among traditional CNNS,traditional Deep Neural Networks(DNN) and Transfer Learning Models which model performs better for the Skin Lesions dataset?
- Excluding the time constraint, which one is more efficient: Deep Neural Network or Traditional Image algo-

rithms?

II. Background

Here we are trying to explore some existing works on this research domain:

A. Existing Systems

1) Evaluation of effect of lesion segmentatioin of image classification: In this project [7] the role of using skin lesion segmentation masks has been explicitly investigated on the performance of dermatoscopic image classification. In this research a baseline classifier without using any segmentation masks was developed . Then the classification performances were investigated by using manually or automatically created segmentation masks both in training and test phases. The system architecture is shown in Figure1

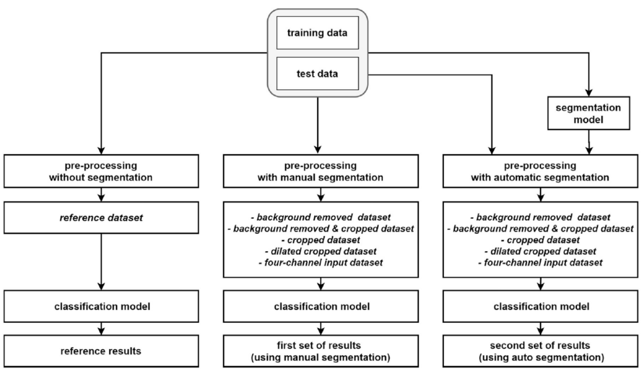

Fig. 1. System architecture for Evaluation of effect of lesion segmentatioin of image classification

The ISIC 2017 challenge dataset has been used in this work.From the dataset, 2150 training and validation images has been used in the training phase and evaluated the classification performance on the 600 test images.In the pre-processing part, gray world colour constancy algorithm [17] [18] has been implemented to handle various lightening conditions.Then a standard pre-processing technique of subtracting the mean RGB intensity value of the ImageNet dataset [19] from the images'(both training and test images) RGB channels has also been implemented.A baseline single classification model was developed to be used as a reference model in all experiments. EfficientNet family (Efficient- NetB0) [20] has been used for the pre-trained model selection. For the fine-tuning of the model, FC layer of the pre-trained network was removed and a global average pooling layer was used to connect it to two blocks of batch FC layers,normalisation layers and dropout layers. Four-channel input images (RGB channels plus the mask channel) were used , and a 3 ×3 ×3 convolutional layer was added to convert the 4 channel inputs to 3 channel data and then connected it to the utilised Efficient- NetB0 model.For training, Adam optimisation method [21] was used and the network was trained for 70 epochs.

The overall result of the research shows that, there's no significant improvement in the MM classification while using segmentation masks, then using segmentation masks for dilated cropping, SK classification performance was significantly improved though there's no significant difference of using manual or automatic creation of segmentation masks.

2) Skin lesion classification using CNN without segmentation: In this research work [22], the authors proposed a new prediction model as shown in Figure 2 which can classify skin lesions into benign or malignant lesions based on a novel regularizer technique. It's a binary classifier and the average accuracy of the model is 97.49%Ṙegularizer helps control the complexity of the classifier.Here, in this work the novel regularizer is implemented based on the standard deviation of the weight matrix of the classifier. The dataset was taken from ISIC archives [23]. It is a benchmark dataset used by Pomponiu [24] and Harangi [25] which contains images with benign and malignant lesions. In total, there are 23906 images of various classes available. The dataset was divided into three equal parts containing almost 8000 images of benign and malignant categories of approximately 600x600 resolutions.

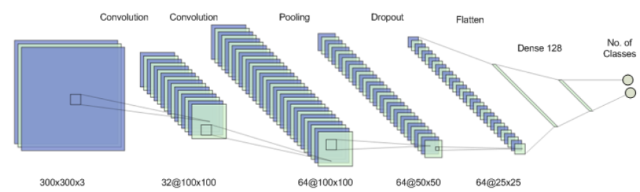

Fig. 2. System architecture for skin lesion classification using CNN without segmentation

The CNN architecture shown in Fig. 2 contains two convolution layers followed by a pooling layer and dropout. After dropout, the 2-D outputs are flattened in a 1-D array and fully connected with the next layer having 128 neurons. In the last layer, there is one output neuron per class. In each convolution layer, the novel regularizer is embedded. The reason of embedding the novel regularizer is to control the values of filter or kernel matrix values.

The performance of CNN in terms of AUC-ROC(Area Under the ROC Curve-Receiver Operating Characteristic Curve) with an embedded novel regularizer is tested on multiple use cases. The area under the curve (AUC) achieved for nevus against melanoma lesion, seborrheic keratosis versus basal cell carcinoma lesion, seborrheic keratosis versus melanoma lesion, solar lentigo versus melanoma lesion is 0.77, 0.93, 0.85, and 0.86, respectively. The proposed model significantly outperformed the existing classification algorithms.

3) Skin lesion classification using hybrid deep neural networks: In this research work [3], an unique and fully automated classification model was proposed with optimised deep features from a number of popular CNNs and from different abstraction levels. The system architecture is shown in Figure 3.Deep models such as AlexNet, VGG16 and ResNet-18, were used as deep feature generator. The extracted features were used to train support vector machine classifiers. In a final stage, the classifier outputs are fused to obtain a classification.the proposed method is shown to achieve very good classification performance after it was Evaluated on the 150 validation images from the ISIC 2017 classification challenge. obtaining an AOC of 83.83% for melanoma classification and of 97.55% for seborrheic keratosis classification.

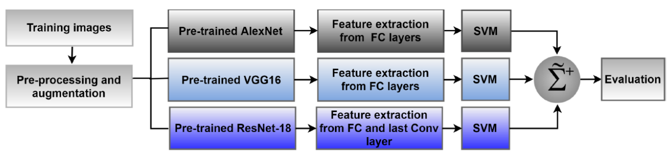

Fig. 3. System architecture for skin lesion classification using hybrid deep neural networks

The training, validation and test images of the ISIC 2016 competition [26] was used and the training set of the ISIC 2017 competition was used for training the classifiers. In total, 2037 colour dermoscopic skin images were used of various sizes (from 1022 x 767 to 6748 x 4499 pixels), photographic angles and lighting conditions and different artefacts. A separate set of 150 skin images were used as validation images to evaluate the results of the proposed method.

To ensure better generalisation ability incase of testing on other dermoscopic skin lesion datasets,the pre-processing steps being kept as minimal as possible .The three steps pre-processing contains normalising the images by subtracting the mean RGB value of the ImageNet dataset as suggested in [19] ,then the images are resized using bicubic interpolation to be fed to the networks (227x227 and 224x224). Finally, the training set was augmented by rotating the images by 0, 90, 180 and 270 degree and then further applying horizontal flipping. The deep feature extractor contains AlexNet [27], a variation of VGGNet named VGG16 [28], and a variation of ResNet named ResNet-18 [29]. To prevent overfitting the shallowest variations of VGGNet and ResNet were used. The features are mainly extracted from the last fully connected (FC) layers of the pre-trained AlexNet and pre-trained VGG16. For ResNet-18, since it has only one FC layer, the features have also been extracted from the last convolutional layer of the pre-trained model. The above features along with the corresponding labels (i.e. skin lesion type) are then used to train multi-class non-linear support vector machine (SVM) classifiers. To evaluate the classification results, the svm scores are mapped to probabilities using logistic regression [30]. The research shows that, fusing the deep features from various layers of a single network or from various pre-trained CNNs lead to better classification performance.

B. Existing Techniques

1) Segmentation:

a) Active Contour Model: Active contour is defined as one of the active models for the segmentation process. Contours are the boundaries which help defining the region of interest in an image. A contour is a collection of points that have been interpolated via polynomial, linear or splines interpolation procedure based on the curve definition in the image [31].Simply put, this model helps defining and identifying both smooth and uneven shapes in images [31]. Active contour model is mainly used in various medical image processing applications where desired regions in the images need to be extracted.

b) Chan-Vese Model: The Chan-Vese model is based on an energy minimization problem which leads to an easier way to solve problem by being reformulated in the level set formulation [32]. Unlike other traditional segmentation techniques, Chan-Vese doesn't rely heavily on edge detection [33]. As a result, this model can detect objects whose boundaries are not so smooth or well defined by gradient [8].

c) U-Net: U-Net is a convolutional neural network and it was developed for biomedical image segmentation [11].U-Net was created in 2015 [10] and It's a modification of Fully convolutional networks for semantic segmentation [34]. The model consists of a contracting path and an expansive path which are structured in a way that the whole model forms a U-shaped architecture. In the structure,the upsampling part contains a large number of feature channels through which the context information is passed towards higher resolution layers [11].the pixels in the border region of the image is being predicted by extrapolation of missing context via mirroring the input image [11].

2) Segmentation Metrics:

a) Intersection over Union (IoU): This metric quantifies the percent overlap between the target mask and the prediction output [13]. In other words, it's the ratio of the overlap area between the ground truth and the predicted segmentation to the union area of the ground truth and the predicted segmentation [14]. It's also known as Jaccard Index. Computationally, this metric measures the number of pixels common between the prediction and targetmasks divided by the total number of pixels present across both masks [13].

b) Dice Coefficient: It's very much simillar to the Intersection over union(IoU) as both of them will provide same qualitative evaluation and comparison among different models. In other words, they are positively corelated [14]. Like IoU, this metric also ranges from value 0 to 1, where 1 depicts the most simillarity between truth and predicted and 0 depicts the least. Mathematically, it's the ratio between twice the area of overlap to the total number of pixels in both images [14].

3) Classification:

a) EfficientNet: EfficientNet features a new scaling method called compound scaling [16].This compound scaling method scales network depth, width and resolution uniformly with fixed scaling coefficients [35].The benefit of this compound scaling is for bigger input image, the network increases the receptive field and more channels to capture more fine-grained patterns on the image [35]. Moreover, EfficientNet is the most recently developed model and also have achieved both higher accuracy and better efficiency over existing Convolutional Neural Network architecures [16].

b) ResNet: Residual Network (ResNet) is a specific type of neural network that was introduced in 2015 [36].It features

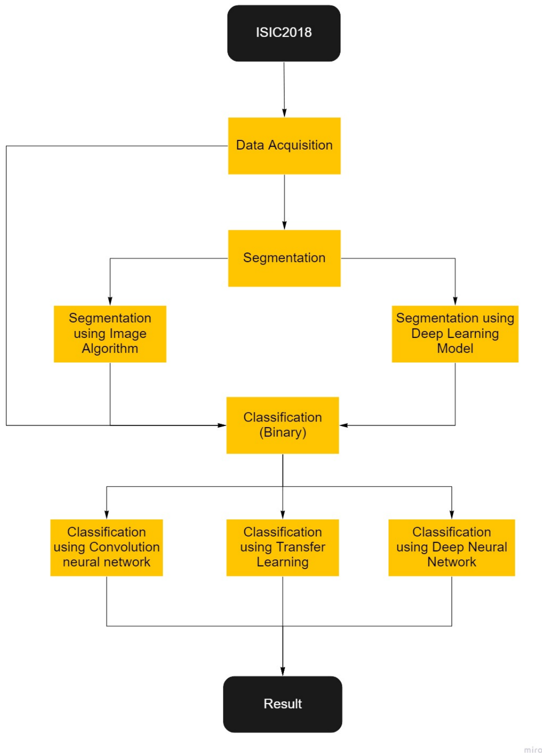

Fig. 4. Architecture

ease of the training of networks that are substantially deeper compared to previous models [36]. Instead of learning unreferenced functions, The layers are reformulated as learning residual functions with reference to the layer inputs [36]. The skip connections in ResNet allows alternate shortcut path for the gradient to flow through and thus solve the problem of vanishing gradient in deep neural networks [37]. In ResNet the higher layer will perform as good as the lower layer becuase skip connections allows the model to learn the identity functions [37].

4) Classification matrics: We have used recall and F1 Measure to evaluate the classification models:

a) Recall: Recall measures the number of positive class predictions made from all the positive examples in the dataset [38]. We have implemented recall for our binary classification. For binary classification recall is measured as the number of true positives divided by the total number of true positives and false negatives [38].

b) F1 Measure: F-Measure gives a single score that balances both the concerns of precision and recall in one number [38].Mathematically, The F1 score can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0 [39].

III. Method

The proposed system consists of an image analysis pipeline suggested in [40] which consists of:

1) **Data Acquisition**
2) **Segmentation**
3) **Evaluation**
4) **Classification**
5) **Model Evaluation**

A. Data Acquisition:

The project incorporated both classification and segmentation due to which we needed the segmentation dataset along with its ground truth and classification dataset along with its diagnosis. For this project the dataset that we decided to choose was ISIC2018 which was made open source for a Challenge hosted by an International Skin Imaging Collaboration(ISIC) [41]. The dataset consists of two sets of data:

1) **Segmentation Challenge:** 2594 images and 12970 corresponding ground truth response mask
2) **Classification Challenge:** 10015 entries grouping each lesion by image and diagnosis confirm type

Since, all the experiments were conducted on a local machine due to which we had to reduce the image number to 1400 images which belong to the benign and malignant class along with downscaling the image original size. The folder organization of the project is shown in 5

```
Database
   |-- Classification
           |--Segmentation Classification
               |--ISIC2018
                   |--Test
                       |-- images
                   |--Train
                       |--Melanoma
                           |-- images
                       |-- Non_melanoma
                           |-- images
                   |--Validation
                       |--Melanoma
                           |-- images
                       |-- Non_melanoma
                           |-- images
           |-- Non_segmentation_classification
               |--Similar Structure to Segmentation Classification
   |-- Segmentation
       |--ISIC2018_Task1-2_Training_Input
           |--mask
       |--ISIC2018_Task1-2_Validation_Input
           |--mask
       |--ISIC2018_Task1_Training_GroundTruth
           |--mask
       |--ISIC2018_Task1_Validation_GroundTruth
           |--mask
```

Fig. 5. Folder organization

B. Segmentation

In the Image Processing pipeline Segmentation plays an important role. Due to its ability of extracting area of interest,object detection which can be used for the further processing which can be recognition or description. For the image processing task it is considered as a practice for the classification of image pixels [42]. In this project we are implementing multiple segmentation models along with their evaluation. At first we are evaluating the traditional Image processing algorithms such as Active contour and the Chanvessen algorithm. After, that we are comparing the algorithms with the state of the art model like UNET for the segmentation.

In our Proposed system pipeline the role of Segmentation consist of 2 parts:

1) Comparing the Traditional segmentation Algorithm along with the State-of the-art model i.e. UNET
2) Testing the importance of segmentation for the classification task(Binary Classification)

C. Classification

The process of categorizing and labeling groups of pixels or vectors within an image based on particular criterias is called Image classification [43]. There are two types of classification, one is supervised and another is unsupervised image classification. Unsupervised classification is done without the ground truth training data, where the algorithm will detect specific characteristics in image and classify them. Whereas in supervised classification, the images are assigned to pre-selected categories obtained from training data. In our project we are doing supervised classification as we already know what characteristics of image belong to what specific class/label. We are doing binary classification to classify images in melanoma and non-melanoma categories. We are using traditional convolutional neural network and transfer learning model EfficientNetB0 and Deep Neural Network ResNet.

1) Convolutional Neural Network: Convolutional Neural Network is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other [44]. We are using CNN because the pre-processing required in a CNN is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, CNN have the ability to learn these filters/characteristics.

2) Transfer Learning: Transfer learning is specific type of machine learning methodology that focuses on storing knowledge obtained while solving one problem and implementing it to a different but relevant problem [45]. Here we are using EfficientNet-b0 which is a convolutional neural network that is trained on more than a million images from the ImageNet database [45]. The network has the capability of classifing images into 1000 object categories(i.e keyboard, mouse, pencil, and many animals). As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of (224x224) [45].

3) Deep Neural Network: The residual neural network (ResNet) is a type of artificial neural network which was built on constructs similar to pyramidal cells in the cerebral cortex [46]. ResNet implement this by utilizing skip connections or simply put making shortcuts to jump over some layers [46]. The reason for using skip connections is to avoid the problem of vanishing gradients problem, where including more layers to a deep model leads to larger training error [46].

D. Model Evaluation

1) segmentation: We have used IoU and Dice Coefficient matrics for evaluation of the segmentation models.

a) Intersection over union(IoU): Mathematically, IoU is the ratio of the overlap area between the ground truth and the predicted segmentation to the union area of the ground truth and the predicted segmentation [14]. It's also known as Jaccard Index. Computationally, this metric measures the number of pixels common between the prediction and target masks divided by the total number of pixels present across both masks [13].

b) Dice Coefficient: Dice Coefficient is very much similar to the Intersection over union(IoU) as both of them will provide same qualitative evaluation and comparison among different models. In other words, they are positively corelated [14]. Like IoU, this metric also ranges from value 0 to 1, where 1 depicts the most simillarity between truth and predicted and 0 depicts the least. Mathematically, it's the ratio between twice the area of overlap to the total number of pixels in both images [14].

2) Classification: We have used recall and F1 Measure to evaluate the classification models:

3) Recall: Recall is calculated as the number of positive class predictions obtained from all positive examples in the dataset [38]. Here in our research, We have implemented recall for our binary classification. For binary classification recall is calculated as ratio between the number of true positives to the total number of true positives and false negatives [38].

4) F1 Measure: F-Measure provides a single score that balances both the concerns of precision and recall in one number [38].We have used F1 score to have a combination score of the precision and recall of the model.

IV. Overview and Results

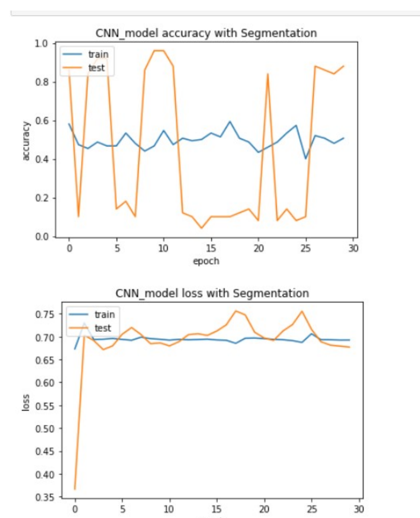

Fig. 6. CNN with segmentation

Our CNN architecture consists stacked conv2D and maxpooling2D layers with rectified linear unit as activation function and the input shape is (180,180,3). The total no of trainable parameters were 6,201,692 The used optimizer is Adam. To train the model we have used 1400 images and to validate we have used 193 images. For 10 epochs the final test accuracy is almost 46.66% with segmentation and without segmentation the test accuracy is almost 83%.

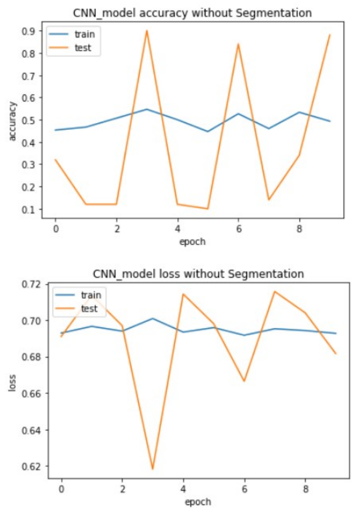

Fig. 7. CNN without segmentation

In the EfficientNetB0 architecture, the input shape is (224,224,3). The weights are taken from pre-trained model on imagenet. The fully-connected layer is included in the top of the network.

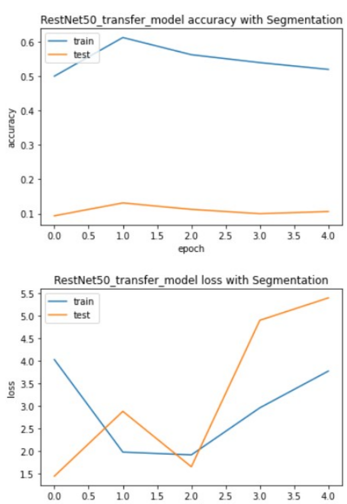

Fig. 8. EfficientNet With Segmentation

In the EfficientNetB0, Adam has been used as the optimizer. Training the model for 10 epochs give us the result of 53.94% approximately with segmentation and without segmentation the test accuracy is almost 11.24%.

For ResNet, we have constructed residual blocks with stacked conv2D layers with rectified linear unit(relu) as activation function with customized adam optimizer. We have

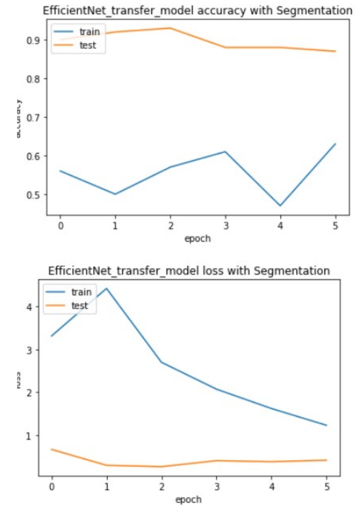

Fig. 9. EfficientNet Without Segmentation

used 1400 training images to train the model and 193 images as validation set of images. For 10 epochs we got accuracy rate of 44.70% approximately with segmentation and without segmentation the test accuracy is almost 9.30%.

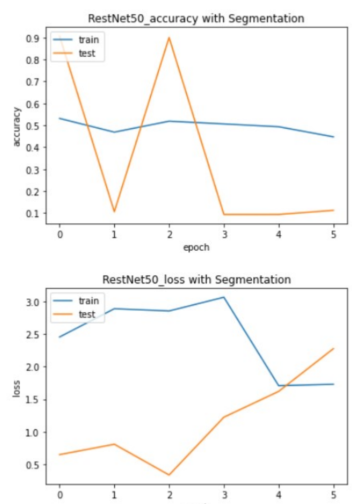

Fig. 10. ResNet50 With Segmentation

	With Segmentation			
Models	Accuracy		F1Score	
	Train	Test	Train	Test
CNN	46.66%	87.90%	74.90%	83.90%
ResNet50	44.70%	11.24%	49.34%	9.30%
EfficientNetB0	53.94%	10.00%	56.25%	11.24%

Fig. 11. Classification Metrices with segmentation

As stated earlier, we have evaluated our segmentation models. The result of active contour is with dice value 0.34, jaccard_index value 0.232 whereas for Chan-vesse the dice value was 0.339 and jaccard_index value 0.239 and finally

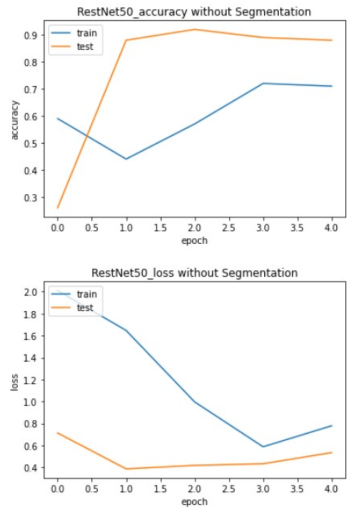

Fig. 12. ResNet50 Without Segmentation

Models	Without Segmentation			
	Accuracy		F1Score	
	Train	Test	Train	Test
CNN	49.33%	87.99%	53.33%	34.00%
ResNet50	71.00%	88.00%	72.00%	88.99%
EfficientNetB0	63.00%	87.00%	56.99%	93.00%

Fig. 13. Classification Metrices without segmentation

the dice value for the Unet model was 0.5307 with the jaccard_index value 0.3998.

Segmentation Metrices		
Models	Metrices	
	Dice	IOU
Active Contour	10.60%	7.10%
Chan-Vese	33.90%	23.90%
Unet	53.07%	39.98%

Fig. 14. Segmentation metrices

V. Discussion

1) Segmentation Model:: Despite traditional algorithm like active contour and Chan-vese being well known in the Image Processing pipeline the Active contour was no were near compared to the State of the art Deep Neural Network i.e Unet. Chan-vese was also outperformed by the Unet model but it performed more better than the active contour.

Hence, In case of segmentation based on our experimental setup, results and findings we conclude that Unet is more favourable for segmentation.

2) Binary Classification without Segmentation:: As, we know that transfer learning performs more accurately compared to the traditional CNN and Deep Neural Network(ResNet50) due to it's pretrained weights. Based on our experimental setup as we can refer to table 13 in the Result Section, the F1 score of the Efficient Net(Transfer Learning) was 93% despite having less accuracy compared to ResNet50. And as we know, the F1 score is actually a way of combining the precision and recall of the model and the higher F1 score the better accuracy is, with 0% being the worst possible and 100% being the best.

3) Binary Classification with Segmentation:: The result was quite surprising in this experimental setup which was carried out taking the segmented images were we got results completely opposite from the result in the experimental setup where we didn't segment the images. In this setup the traditional CNN performed with better accuracy in the test result which was 87.9% and with the highest f1 score of 83.9%.

Therefore, based on the experimental setup,results and findings we can conclude that in case of smaller dataset the traditional CNNs perform better for binary classification compared to other models.

VI. Conclusion

Hence, in this experiment we realised something new that was traditional CNN performing better compared to the renowned transfer learning and Deep Neural Networks. some reasons that we can infer from the from the experiment is:

1) the dataset was smaller,
2) the epochs used for training the model was just 10 since the experiment was conducted in my local machine,
3) and, the models were implemented with default values and default input shapes

Finally, we end this project by stating the future work of the project:

1) Simulating the experiment in a bigger setup with higher number of images and higher number of epochs,
2) Fine tuning the models by finding the best hyper=parameters for my dataset.

References

[1] U. M. B. Leiter-Stöppke, T. K. Eigentler, and C. Garbe, "Epidemiology of skin cancer," 2014.

[2] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[3] ——, "Skin lesion classification using hybrid deep neural networks," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[4] K. Holland, "What's causing this skin lesion?" 2021, [Online; accessed 11-November-2021]. [Online]. Available: https://www.healthline.com/health/skin-lesions

[5] M. Tyagi, "Image segmentation: Part 1." [Online]. Available: https://towardsdatascience.com/image-segmentation-part-1-9f3db1ac1c50

[6] M. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 33, no. 2, p. 148–153, 2009.

[7] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The effects of skin lesion segmentation on the performance of dermatoscopic image classification," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105725, 2020.

[8] T. Chan and L. Vese, "An active contour model without edges," *Scale-Space Theories in Computer Vision*, p. 141–151, 1999.

[9] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, pp. 577–685, 1989.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, p. 234–241, 2015.

[11] Wikipedia, “U-net,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://en.wikipedia.org/wiki/U-Net

[12] Zhang, Jeremy, “UNet — Line by Line Explanation,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5

[13] J. Jordan, “Evaluating image segmentation models,” 2021, [Online; accessed 25-November-2021]. [Online]. Available: https://www.jeremyjordan.me/evaluating-image-segmentation-models/

[14] E. Tiu, “Metrics to evaluate your semantic segmentation model,” 2021, [Online; accessed 25-November-2021]. [Online]. Available: https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2

[15] Tan, Mingxing, “EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html

[16] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” 2020.

[17] A. Fidalgo Barata, E. Celebi, and J. Marques, “Improving dermoscopy image classification using color constancy,” *IEEE Journal of Biomedical and Health Informatics*, p. 1–1, 2014.

[18] A. Sharib, “Color constancy toolbox matlab central file exchange,” 2021, [Online; accessed 12-November-2021]. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/52633-color-constancy-toolbox

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[20] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946

[21] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2015.

[22] M. A. Albahar, “Skin lesion classification using convolutional neural network with novel regularizer,” *IEEE Access*, vol. 7, p. 38306–38313, 2019.

[23] “The international skin imaging collaboration (isic),” 2021, [Online; accessed 12-November-2021]. [Online]. Available: https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery

[24] V. Pomponiu, H. Nejati, and N.-M. Cheung, “Deepmole: Deep neural networks for skin mole lesion classification,” *2016 IEEE International Conference on Image Processing (ICIP)*, 2016.

[25] B. Harangi, “Skin lesion classification with ensembles of deep convolutional neural networks,” *Journal of Biomedical Informatics*, vol. 86, p. 25–32, 2018.

[26] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and et al., “Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic),” *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, p. 84–90, 2017.

[28] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[30] J. C. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” in *ADVANCES IN LARGE MARGIN CLASSIFIERS*. MIT Press, 1999, pp. 61–74.

[31] D. Bhatt, “Active contours – a method for image segmentation in computer vision,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/09/active-contours-a-method-for-image-segmentation-in-computer-vision/

[32] R. Cohen, “The chan-vese algorithm,” 2011.

[33] P. Getreuer, “Chan-vese segmentation,” *Image Processing On Line*, vol. 2, p. 214–224, 2012.

[34] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, p. 640–651, 2017.

[35] paperswithcode.com, “EfficientNet,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://paperswithcode.com/method/efficientnet

[36] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[37] H. Mujtaba, “Introduction to resnet or residual network,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://www.mygreatlearning.com/blog/resnet/

[38] J. Brownlee, “How to calculate precision, recall, and f-measure for imbalanced classification,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/

[39] scikit-learn.org, “sklearn.metrics.f1$_{s}core,'' https : //scikit - learn.org/stable/modules/generated/sklearn.metrics.f1_{s}core.html$, 2021, [
$December - 2021$].

[40] K. Yu, N. Kirchgessner, C. Grieder, A. Walter, and A. Hund, “An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping,” *Plant Methods*, vol. 13, no. 1, 2017.

[41] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, “Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC),” *arXiv:1902.03368 [cs]*, Mar. 2019, arXiv: 1902.03368. [Online]. Available: http://arxiv.org/abs/1902.03368

[42] W. Cai, S. Chen, and D. Zhang, “Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation,” *Pattern Recognition*, vol. 40, no. 3, p. 825–838, 2007.

[43] M. Shinozuka and B. Mansouri, “4 - synthetic aperture radar and remote sensing technologies for structural health monitoring of civil infrastructure systems,” in *Structural Health Monitoring of Civil Infrastructure Systems*, ser. Woodhead Publishing Series in Civil and Structural Engineering, V. M. Karbhari and F. Ansari, Eds. Woodhead Publishing, 2009, pp. 113–151. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781845693923500049

[44] S. Saha, “A comprehensive guide to convolutional neural networks — the eli5 way,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[45] Wikipedia, “Transfer learning,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://en.wikipedia.org/wiki/Transfer_learning

[46] ——, “Residual neural network,” 2021, [Online; accessed 1-December-2021]. [Online]. Available: https://en.wikipedia.org/wiki/Residual_neural_network