Melanoma Detection: An automated approach

Submitted by

Aishwarya Rajan, 13CO109
Ankit Tripathi, 13CO113
Tejaswi Ravu, 13CO248

Under the Guidance of
Dr. Jeny Rajan
Dept. of Computer Science and Engineering,
NITK, Surathkal
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF

BACHELOR OF TECHNOLOGY
IN
COMPUTER ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL
SRINIVASNAGAR-575 025 KARNATAKA, INDIA
2016-2017

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

DECLARATION

by the B.Tech. Student

I/We hereby *declare* that the Project Work Report entitled

.....

.....

which is being submitted to the **National Institute of Technology Karnataka, Surathkal** for the award of the Degree of **Bachelor of Technology** in .....

.....

is a *bonafide report of the work carried out by me/us.* The material contained in this Project Work Report has not been submitted to any University or Institution for the award of any degree.

Register Number, Name & Signature of the Student(s):

1) 13CO113 - Ankit Tripathi

2) 13CO109 - Aishwarya Rajan

3) 13CO248 - Tejaswi Ravu

Department of .....

Place: NITK, SURATHKAL

Date:

CERTIFICATE

This is to certify that the B.Tech. Project Work Report entitled

.....

..... submitted by :

Sl.No. Register Number & Name of Student(s)

1) 13CO113 - Ankit Tripathi

2) 13CO109 - Aishwarya Rajan

3) 13CO248 - Tejaswi Ravu

as the record of the work carried out by him/her/them, is accepted as the B.Tech. Project Work Report submission in partial fulfillment of the requirements for the award of degree of **Bachelor of Technology** in .....

Guide(s)
(Name and
Signature with Date)

Chairman - DUGC
(Signature with Date and Seal)

Abstract

This work is a report on the topic **Automated Melanoma detection using deep convolutional networks.** The first chapter discusses about the disease, malignant melanoma, statistics, clinical diagnosis methods. We also address the Indian scenario and the motivation to take up this project. The third subpart of Introduction gives insight to automated methods that have been explores by various researchers. This section is the base point of our work.

In second chapter, we describe the Problem definition and an overview of our field of work. Third chapter explores various state of art methodologies in this area. We have covered two areas of study that was taken up, Melanoma detection and Deep learning architectures.

Chapter 4 discusses about the data we have used in our project and chapter 5 introduces our methodology, explaining our architecture. The results are presented in chapter 6 along with graphs and discussions.

CONTENTS

List of Tables

Table of Figures

Chapter 1: Introduction

Melanoma is among the most curable cancers when detected early, but ends as one of the most lethal due to delays in diagnosis and treatment, killing one person every hour. The prognosis quickly worsens as the disease progresses, due to its tendency to rapidly metastasize. Melanoma incidence is soaring. In the 1930s, 1 in 1 500 USA residents developed the disease; in the 2010s that incidence jumped to 1 in 59 [1]. Medical systems, public and private, are struggling to keep up with those rates, since the number of specialists able to diagnose the disease has not grown proportionally. Melanoma is difficult to diagnose reliably, requiring extensively trained specialists. Improving the diagnosis of melanoma is an urgent need. Automated screening appears as an alternative to alleviate the problem [2–14]. Those techniques are still too inaccurate to give definitive diagnosis, but may soon become accurate enough to assist medical personnel in the triage of patients. Focusing specialists' time on the patients at risk, while safely discharging the others, can greatly increase the quality and cost-effectiveness of care. Such triage can greatly benefit (rural, isolated, or poor) communities that cannot keep a specialist full time; and also help any primary care provider in the difficult decision of referring or not each patient to specialists.

In order to combat the rising mortality of melanoma, early detection is critical. Currently, highly trained experts and professional equipment are necessary for accurate and early detection of melanoma. Dermoscopy is a specialized method of high-resolution imaging of the skin that reduces skin surface reflectance, allowing clinicians to visualize deeper underlying structures. Using this device, specially trained clinicians have demonstrated a diagnostic accuracy as high as 75-84% [7]. However, recognition performance drops significantly when the clinicians are not adequately trained [8, 9]. Dermoscopy is a technique often used for imaging skin lesions. It operates on basis of the play between light and skin analogous to the nature of light and water. By treating the skin so that it absorbs the light rather than reflect off in the same manner as light rays on the ocean surface, the hidden layers of the skin become visible to the naked eye. The image is taken by covering the selected area, encompassing the entirety of the lesion, with a liquid to remove any glare or reflectiveness. The skin consequently creates transparency in the cornified layer of skin allowing the visualization of deeper layers in the epidermis. This is done to

view key features of the pigmented layers such as colours, artifacts, symmetry, border, vascularity, and texture.

Images obtained from dermoscopy are then fed to different algorithms for diagnosis. The popular ones include the ABDC rule, 7 point checklist, Menzies method, C.A.S.H., BLINCK algorithm. All of these algorithms approach classification as a process that occurs after identifying the same criterion. The criteria can be divided into global and local views. All the different patches have a global feature like a particular pattern (e.g. globular, parallel, reticular) which is holistic view determined by looking at the placement of local features which vary from images like vascularity, pigmentation structure, dots, etc. These are the most fundamental observations in determining the diagnosis of lesions as melanoma or non-melanoma.

While the incidences of melanoma have greatly increased, there has not been a similar rise in the number of specialized experts capable of classifying the growth. It has been a huge burden on the medical education institutions and hospitals. To allay this struggle, semi-automated and automated melanoma classification methods exist. The difficulty with automation lies in the lack of accuracy in classification but this may change. If the identification of patients with benign tumors succeeds, this allows experts more time to focus on terminal cases. It is also worth noting medical imaging is improving and there are now datasets of greater size though some techniques like deep neural networks work but require even larger quantities of data than it is receiving now.

Each type of skin cancer has a different appearance. A dermatologist may be able to make a diagnosis simply by examining the skin. Dermoscopy, a technique involving a hand-held instrument that magnifies the skin surface, can help to differentiate skin cancer from other skin disorders; in some cases, a physician will take a histological section of the skin to confirm the diagnosis. When caught early, skin cancer is often treated with a very high success rate—so early diagnosis is very important for both melanoma and nonmelanoma skin cancers.

Indian Scenario

Melanoma scenario in India was studied by All India Institute of Medical Sciences and here is their review.

Methods and Materials : All melanoma patients treated at the Department of Radiotherapy, All India Institute of Medical Sciences, New Delhi, India, from 1995 to 2007 were studied retrospectively. The endpoints were loco-regional recurrence, distant recurrence, recurrence-free survival (RFS), and duration of follow-up (DOFU). RFS and DOFU were analyzed with respect to the factors like age, sex, tissue of origin, site of disease, number of nodes, lymphadenopathy, ulceration, stage, and operability to find out any association.

Results : Seventy-two patients were found evaluable with 40 males and 32 females (median age 46.5 years). Eye was the commonest primary site with visual disturbance as the commonest symptom. Overall, 87% of the lesions were single, with most of the nonocular lesions presenting in the advanced stage. During the disease course, regional lymphadenopathy and distant metastases were seen in 33% and 32% of cases, respectively. Highest incidence of lymphadenopathy was seen in skin lesions and in primaries from trunk and extremities. Of all treated patients, 47% achieved complete response, 18% partial response, and others had either stable or progressive disease. The median DOFU was 6.2 months. RFS was studied only in curatively treated cases with a median of 10 months. Operability at presentation was the only prognostic factor influencing DOFU.

Conclusion : Malignant melanoma is an uncommon disease in India carrying a lot of morbidity due to late presentation. Its management is still not clear regarding the optimum use and schedule of treatment modalities. More prospective studies in the future are required to come to a definite conclusion.

Melanoma treatment has not been in focus in medical research in India. Yet, a breakthrough in automated detection of Malignant melanoma is necessary as it’s curable when treated early.

Automated Detection

While in the United States there are over 10,000 dermatologists, in other areas of the world the supply of expertise is limited. For example, in Australia, the number of registered dermatologists in 2004 was approximately 340 [10], and in New Zealand, there were 16 [11]. Restricted access to expert consultation leads to additional challenges in providing adequate levels of care to the

populations that are at risk.

In order to address the limited supply of experts, there has been effort in the research community to develop automated image analysis systems to detect disease from dermoscopy images. Such technology could be used as a diagnostic tool by primary care physicians and staff for regular screening, or by clinicians who are otherwise not trained to interpret dermoscopy images.

More recent work has begun to examine the efficacy of the state-of-the-art deep learning approaches to image recognition within the dermatology and dermoscopy application domain [32,33]. Representations learned from the natural photo domain were leveraged, in conjunction with unsupervised and hand-coded features, to achieve state-of-the-art performance in a data of over 2,000 dermoscopy images [32]. However, the work was limited to lesion images that had been manually pre-segmented: images were already cropped around the lesion of interest. In 2016, the International Skin Imaging Collaboration (ISIC) organized an international effort to aggregate a dataset of dermoscopic images from multiple institutions for the purposes of developing and evaluating clinical and automated techniques for the diagnosis of melanoma [34].

A snapshot of the dataset that contained the most complete set of annotations was selected to host a melanoma recognition challenge at the 2016 International Symposium on Biomedical Imaging (ISBI 2016). The challenge was titled “Skin Lesion Analysis toward Melanoma Detection” [35]. In total, 38 individual participants contributed 79 submissions across 3 image analysis tasks, including 43 submissions toward disease classification. This was the first publicly organized large-scale standardized evaluation of algorithms for the detection of melanoma. Top performing techniques involved deep learning approaches, including Deep Residual Networks for classification [36], and fully convolutional networks for segmentation [37,38].

Chapter 2: Problem Definition

The project seeks to classify lesions as benign or malign lumps. The lesions can be further divided into the regularity of features according to physician diagnosis: melanoma, atypical, and benign. Melanoma is the malign lesion and benign is self-evident, however the atypical lesion is an edge case in which classification is problematic and difficult to predict. As pigmented lesions occurring on the surface of the skin, melanoma is amenable to early detection by expert visual inspection. It is also amenable to automated detection with image analysis. Given the widespread availability of high-resolution cameras, algorithms that can improve our ability to screen and detect troublesome lesions can be of great value. As a result, many centers have begun their own research efforts on automated analysis. However, a centralized, coordinated, and comparative effort across institutions has yet to be implemented.

Dermoscopy is an imaging technique that eliminates the surface reflection of skin. By removing surface reflection, visualization of deeper levels of skin is enhanced. Prior research has shown that when used by expert dermatologists, dermoscopy provides improved diagnostic accuracy, in comparison to standard photography. As inexpensive consumer dermatoscope attachments for smart phones are beginning to reach the market, the opportunity for automated dermoscopic assessment algorithms to positively influence patient care increases.

Earlier approaches were inspired mainly by the intuitive appeal of the text search algorithm and likewise modified for image processing and classification. This definition of the problem however may limit the processing ability of the system to human dimensions.

Present day solutions focus on using a highly-automated process, utilizing machine-centric techniques like the state of art, Deep convolutional networks. With the advanced GPU technology, and increased amount of data, deep networks have become a feasible option. Image net competitions have seen a performance on par with humans' ability for Image classification.

The scope of the problem focuses on building an appropriate deep learning architecture that works well for Melanoma classification. The architecture has to be then optimized based on

experimentation. The desired methodologies require high volumes of data for training and testing purposes. We primarily use the International Skin Imaging Collaboration (ISIC) dataset and supplement this with other independent university datasets.

Chapter 3: Literature Review

The section gives an overview of general trends in the literature and examines the most successful models.

The best model in terms of performance was by Lequan Yu, Hao Chen, etc [50]. It features a two-step, deep convolution neural network with specific schemes monitoring the performance metrics to manage data constraints in training and max information gain for each added layer in the network. A fully connected convolution residual network (FCRN) working with a context framework that is general enough to solve other segmentation tasks. It's trained on the ISBI 2016 dataset to achieve an accuracy of 93.1% with segmentation.
In the literature, we can see three major stages for melanoma detection, namely, lesion segmentation, feature segmentation and classification.

Melanoma Detection

The entire process of melanoma image detection is below.

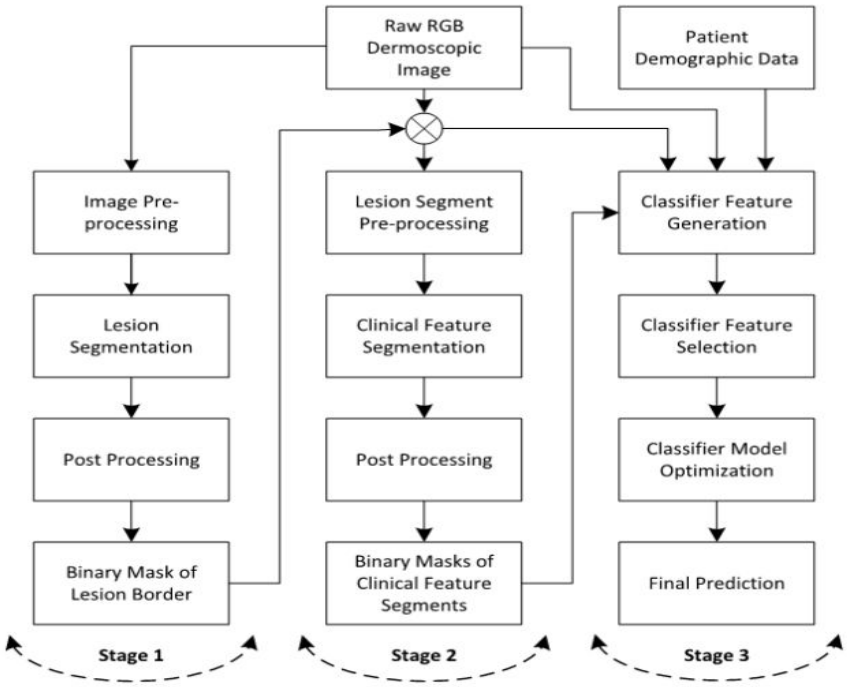

Fig 2: General block diagram of computerised melanoma detection

In the early reliable works, we can see three major stages for melanoma detection, namely, lesion segmentation, feature segmentation and classification(Figure 1).

Lesion Segmentation:

In a dermoscopy image, a skin lesion is a bounded region that is distinguishable from the normal surrounding skin by virtue of different color or texture. This single area is considered to be the region of interest for further processing.[19]. Lesion segmentation deals with separating lesion region from the normal skin region (non-lesion). This is a very important step in the analysis of dermoscopy images since it allows the identification of various global morphological features specific to the lesion and at the same time provides a confined region for segmentation of various local clinical features at a later stage.

There are various broadly categorized image segmentation methods including but not limited to methods employing histogram thresholding [1] [2] [3], clustering [4] [5] [6] [7] [8], localized and distributed region identification, active contours (snakes) [9] [10], edge detection [11] [12], fuzzy logic, supervised learning, graph theory [13], and probabilistic modeling [18] [15]. These methods can be applied individually or in combination to achieve maximum accuracy [14] [16]

[17]. Post-processing is essential for accurate lesion segmentation and feature detection and generation.

Feature Segmentation:

Malignancy of a lesion can be detected by observing various clinical features from a given dermoscopy image. A feature could be spotted or local or spanning depending on the extent and frequency of occurrence. Thus, feature segmentation has to be performed over multiple small segments in and around the lesion area.

Pigment network [20], atypical and typical networks, streaks [21] [22] [23], regression structures [24], starburst pattern, dots and globules [25], blotches [26], blue-white veil [27] [28] [29], pink shades [30], white areas [31], milia-like cysts, vascular structures [32], etc. are some of the 6 commonly employed features in predicting melanoma [33] [34] [35]. Most of these feature structures can also be broadly categorized as having different patterns such as reticular, globular, cobblestone, homogeneous, starburst, parallel, multicomponent, lacunar and unspecific [33] [36] [37] [38]. While using these features for identifying melanoma, benign dermoscopic features in melanomas can be equally important in automatic identification [39].

As in lesion segmentation, feature segmentation also includes pre-processing, and post processing steps. Color, texture, shape, structure, relative size, location in the lesion etc. are some of the main attributes used in clinical feature segmentation [40]. In addition to their presence, distribution of a feature in the lesion area provides further diagnostic information.

All types of segmentation algorithms used for lesion segmentation can also be used for feature segmentation. However, the final feature segmentation output will have multiple distributed segments of different shapes and sizes based on the feature being segmented. Choosing the right combination of color channels for applying the segmentation algorithm is equally important as choosing the right segmentation approach. The post-processing in this case is also crucial and should be selected carefully based on the kind of filtering required to achieve best results. It should be kept in mind that features used in the ultimate classification of the lesion will be generated later using the feature segments.

Feature generation and classification:

The final process is to classify using machine learning algorithms based on the extracted features. It is a binary classification problem, that is to classify as malignant or benign. The clinical features can also be used along with the features extracted after lesion segmentation and feature segmentation.

Lesion-related morphological features such as estimated diameter (estimated because the images may be acquired at different magnification levels), symmetry, irregularity, eccentricity etc. can be calculated from the lesion border. Color and texture features related to the lesion can also be calculated from the lesion area; these can be referred to as global features. Various color channels can be used for this purpose. Because of the significance of various color distribution and texture around the lesion, clustering methods can also be used to divide the lesion into various regions and then color and texture features can be calculated from those regions separately. The lesion area may also be divided concentrically into various peripheral and central regions from which global features can be extracted

A number of classifiers are available for experimentation and they can be chosen based on their performance; the final model can be tuned for better performance [41]. Some of the very commonly used classifiers are artificial neural networks (ANNs) [42] [43], support vector machines (SVMs) [44], logistic regression [45], decision trees [46], ensemble learners [47] [48], k-nearest neighbors (kNN) [49], Bayesian classifiers, deep learning algorithms [50], discriminant analysis [51], etc. There are numerous other classifiers that can also be explored for classification. The evaluation of the classifier result is based on the overall accuracy, sensitivity and specificity of the system on the test set. In this domain, it is very important not to miss melanoma while being able to correctly identify benign lesions as much as possible. In other words, the ultimate goal is to target highest sensitivity while optimizing to increase specificity, thereby increasing the overall accuracy.

In the most recent works, there has been a radical shift from the procedure followed by researchers as mentioned above. Deep convolutional networks, Automated melanoma screening to date has been reviewed broadly by Masood and Ai Al-Jaumaily and extended by Fornaciali and Carvalho [52] on assessing models and creating more accountability in current research. The review is based on the review of Fornaciari and Carvalho for advanced image

classification and hybrid configurations in existing methods. Current research can be divided into image classification and melanoma screening process.

Deconstructing the image pixels into meaningful information has been done using the Bag-of-Visual-Words model (BoVWs) for a long time. It is based on the text information approach by dividing the image in patches with extracted visual words analogous to the atomic word units extracted from a text document. The extracted units form a visual codebook which the image is then searched for and a frequency histogram is produced.

The BoVWs approach is a selective method of 3 layers focusing respectively on the low-level local features, mid-level global features (for the codebook), and final classification.

It starts at the lowest level, gathering data from small patches in the form of local feature extractions and works its way up quantizing data to create a generalized global feature based on the local details. The global features are then used to classify the image based on the characteristics of the labeled dataset.

It is notable this method breaks from the text search analogy as it lacks the inherent grammar structure of words. While an image does not follow grammar per se, its location relative to another patch could be significant however all the geometric data is lost in the process of gaining its robustness and generality of application. The majority of current research shows strong bias toward conservative techniques.

Deep neural networks (DNN) have been a longstanding idea however it is which the advent of multi-core GPUs and better datasets that they have become more relevant. Its architecture consists of a complex, many-layered convolution network that uses aggressive unsupervised learning to teach itself the global features of the input. Its drawbacks include the high training time, need for powerful GPUs, robustness of datasets, and fitting process. However, even with these constraints, the DNN is the more competitive model. Extant work dominated by traditional techniques with few papers like that of Fornaciari and Abiding focusing on complex BoVWs.

The Codella [50] paper featured an innovative hybrid approach to classification. Using the ISIC dataset as input, they fed a pre-trained Caffe model using deep learning and sparse coding for

feature extraction. They normalized the features using the sigmoid function and incorporated sparse-coding based BoVW and average pooling for manipulation of color and grayscale images. In parallel, deep neural network was used to extract features. From the FC-6 and FC-8 layers of the Caffe net network, features were extracted. These features were fed into SVM trainers for binary classification. They fused or combined all of these models to get the maximal results of 93% for regular cases and 72% for atypical versus melanoma. Some concerns include the use a network optimized for real-world imaging, stochastic learning which weights initial samples more than later data (or a high annealing learning rate), and the need for large volumes of data.

Chapter 4: Proposed Method

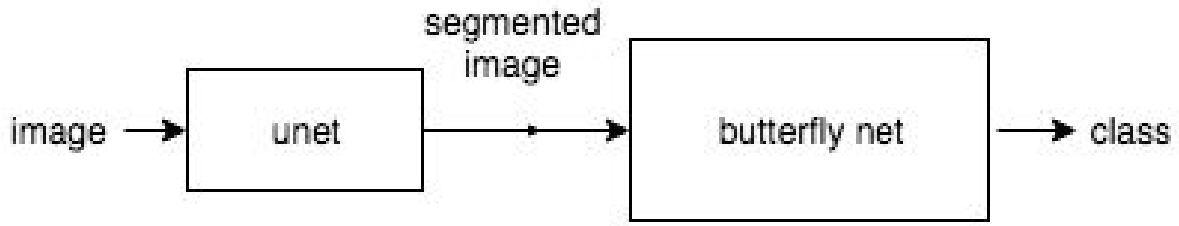

Fig 3: Architecture of the proposed method

Our methodology includes two phases, segmentation of lesion and classification. Through the experiments we found that segmentation of training images gives better results when compared to a raw dermoscopy image. Segmented images helps the machine focus on the lesion region as it cuts out unnecessary information because in the segmented images the classifier doesn't have to learn the properties of the skin which is not cancer. Hence, the classifier learns only the properties of the lesion, cutting out the noise.

Segmentation

Segmentation is done by using a deep learning based classifier which learns the properties of skin and the lesion irrespective of the type of lesion and the amount of data used for training is significantly increased due to no discrimination between cancer and non-cancer images in this stage.

Unet is a fully convolution network which feeds output of the starting layers to the ending layers to highlight the important part of the image for the final classification in the final layers of the model and hence, it performs well for bio-medical images. Unet uses upsampling of the layers which is a type of deconvolution.The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two

3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers. Unet works well even with a small dataset and hence, it is found to perform well for the ISBI Skin Cancer Challenge 2017 dataset.

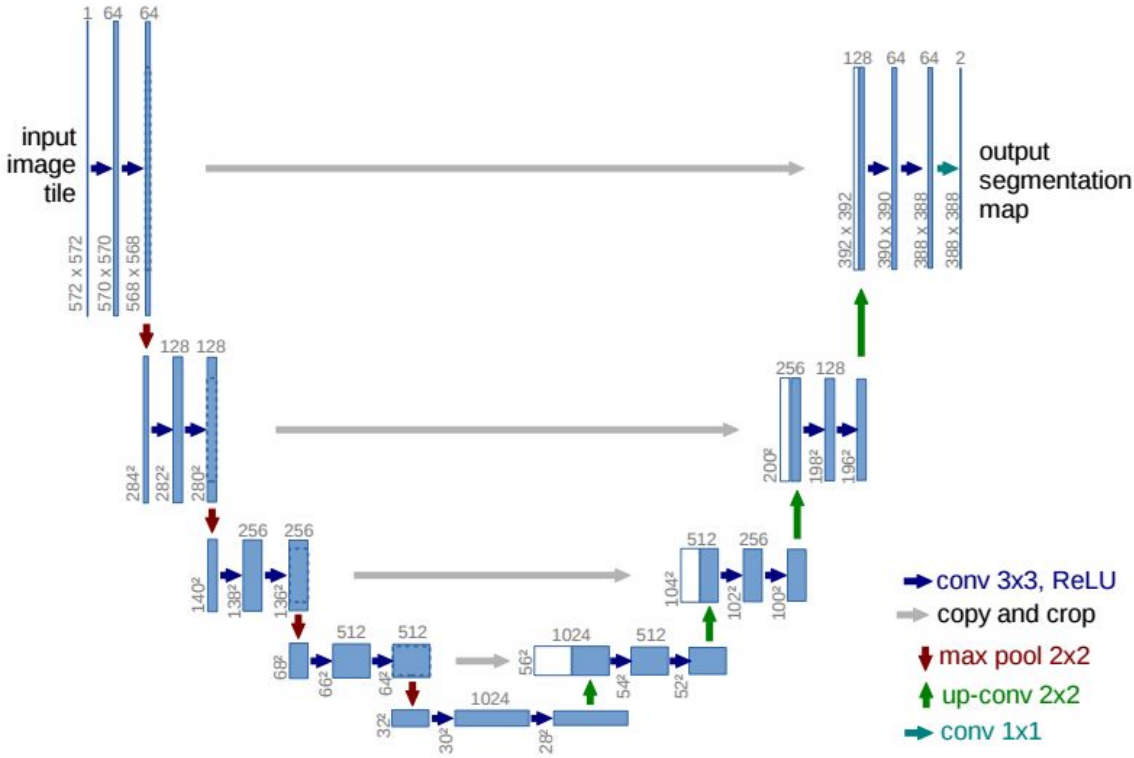

Fig 4: U-net Architecture for segmentation of dermoscopy[54]

Classification

Butterfly model is an advancement of Residual Network of Residual Network(RoR). A residual network is a model where the residual part is combined with input to that part. A network built by linear combination of such blocks is Resnet. Resnet is found to solve the problem of vanishing gradient which is the problem of minimal changes in the weights of the layer farther away from the result layer. This problem is solved due to shortcut connections which reduce the expected depth of the layers in the residual network. Degradation is the problem which shows the unusual trend in the linear deep learning models where the deeper models perform poorly compared to shallow models. The expected reason for this the expected depth. Minimal Feature Reuse is a

problem referring to the low information we can visualize due to you many layers piled one after another during forward propagation. This problem is not effectively solved by the Resnet. Residual Network of Residual network solves this problem by adding more inputs so that the information is refreshed after every few layers highlighting the most important parts in the image. However, the problem with RoR is that it saturates after a point and adding more layers doesn't improve accuracy.

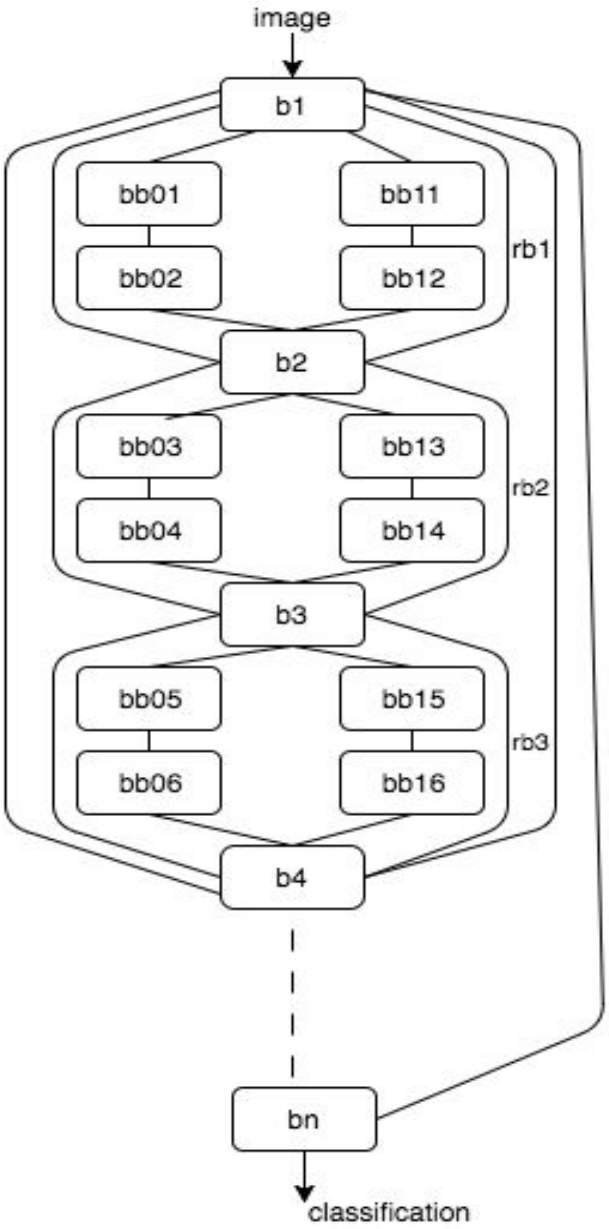

Fig 5: Architecture of Butterfly Model

Our butterfly model can be imaged as ensemble of classifiers which receive the same data and process them separately and then add them so that the new data for the next block is ready. The freedom to separately evaluate input and then binding by adding the result of the output of the classifiers at each level is the key concept that we have evaluated.

The model is also a way to increase the number of learning parameters at the same time not making the global memory of GPUs overflow. The two ensembles at the same level makes sure that depth is not affected but the number of learning parameters are increased. The merge operation at the end of every residual block regularises the size of the intermediate output.

Chapter 5: Experiments and Results

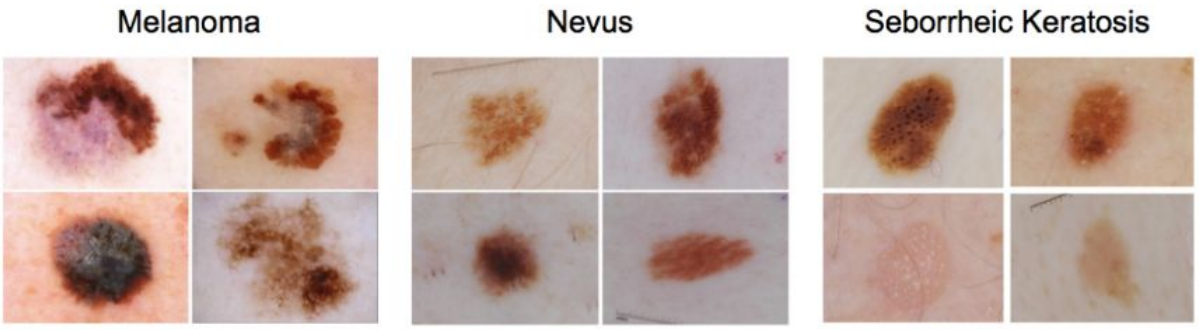

Fig 6: ISIC Skin Cancer Challenge 2017 dataset

In recent years, there have been a number of computerized diagnosis techniques developed which use digital image processing in an attempt to classify benign and malignant skin lesions from dermoscopy images[4]. To accelerate the innovation in this area the International Skin Imaging Collaboration (ISIC) 1 has initiated the ISIC Challenge. This challenge provides a set of 2,000 publicly available dermoscopic images to participants to apply computer vision and machine learning techniques to three different tasks: lesion segmentation, dermoscopic feature extraction, and lesion classification. Our approach aims to address the challenge of lesion classification. The lesion classification portion of the challenge is composed of two independent binary image classification tasks that involve three unique diagnoses of skin lesions (melanoma, nevus, and seborrheic keratosis). The first binary classification task involves distinguishing (a) melanoma vs. (b) nevus and seborrheic keratosis. In the second binary classification task, the classifier needs to distinguish between (a) seborrheic keratosis and (b) nevus and melanoma. Melanoma, as described earlier, is a malignant skin tumor, derived from melanocytes (melanocytic), Nevus is benign skin tumor, derived from melanocytes (melanocytic) while Seborrheic keratosis is also a benign skin tumor, derived from keratinocytes (non-melanocytic). Examples from the dataset of these three skin lesions are shown in Figure 1. Deep neural network models, particularly deep convolutional neural networks and its variants, are currently the best performing image classification models for a wide variety of tasks and applications. They have made significant impact in multiple problem domains, including computer-aided diagnosis using medical images. A recent paper published in Nature [5] is the testament to the success of deep convolutional neural networks for detecting melanoma from dermoscopy images, achieving an AUC of 0.94 on this task. A critical factor that contributed to the high level

of performance of that particular model was the availability of a large dataset, containing more than 120,000 images (60x larger than available for the ISIC Challenge). For this challenge, our deep-learning model is similar to what was used by [5], however since we only have access to a much smaller dataset we apply a number of optimization and post-processing techniques to improve the performance of our classifier.

Codella - using CNN and SVM (Transfer Learning):

(same approach in “Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms.”)

Dataset : International Skin Imaging Collaboration (ISIC) dataset

1. Uses pre-trained Caffenet model from the ILSVRC
    a. Concept detector layer, FC8 (1000 dimensions)
    b. Fully connected layer, FC6 (4096 dimensions)
2. The preprocessing step involves resizing the image and Subtract the model’s input mean image to “centralize”
3. Feature normalization : Sigmoid
4. Classifier : Non-linear SVM using a histogram intersection kernel
5. SVM score averaged FUSION.

Codella - Sparse Coding and SVM:

1. Dataset : International Skin Imaging Collaboration (ISIC) dataset
2. Unsupervised methods - learns a dictionary of sparse codes
3. SPAMS sparse coding dictionary learning algorithm- based on stochastic approximations.
4. Images are rescaled to 128x128 pixel dimensions before extraction of 8x8 patches, to learn dictionaries of 1024 elements.
5. Two dictionaries are constructed in color (RGB) and grayscale color spaces.
6. Classifier : Non-linear SVM using a histogram intersection kernel.

METHODS	ACCURACY	SPECIFICITY
Sparse Coding	72.7%	81%
Deep Learning [caffenet]	77.6%	80%

Table 1: Accuracy and Specificity of existing skin cancer detection methods.

Results

The following are the graphs of the loss function of training and validation phase.

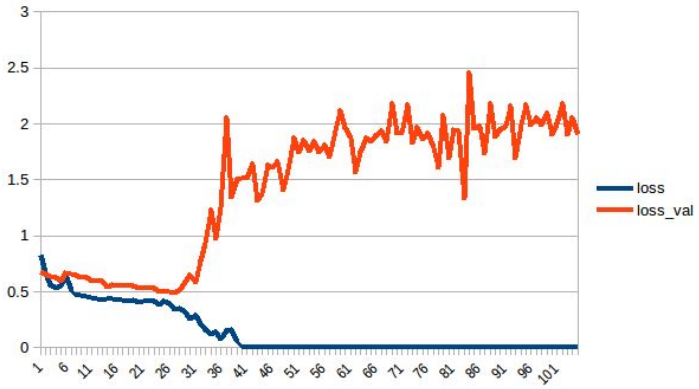

Fig 7: Loss function of Resnet-18

In Resnet loss function graph, we can see that after 26 epochs the val loss starts increasing which tells that beyond that training would not improve the results. If the number of layers are increased the performance is expected to improve but the attainment of the convergence of performance at very low epoch suggest that the features learnt from the small dataset like ours is very limited in the case of Resnet.

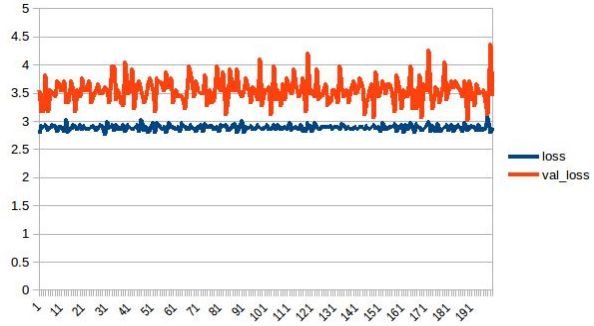

Fig 8: Loss function of RoR-20

In RoR, the performance is slightly improved when compared to resnet because the extra shortcut connection highlight the most important part of the image well and hence, the further layers of the model act like the layers of the initial part because they get the lower level abstraction of the original layer.

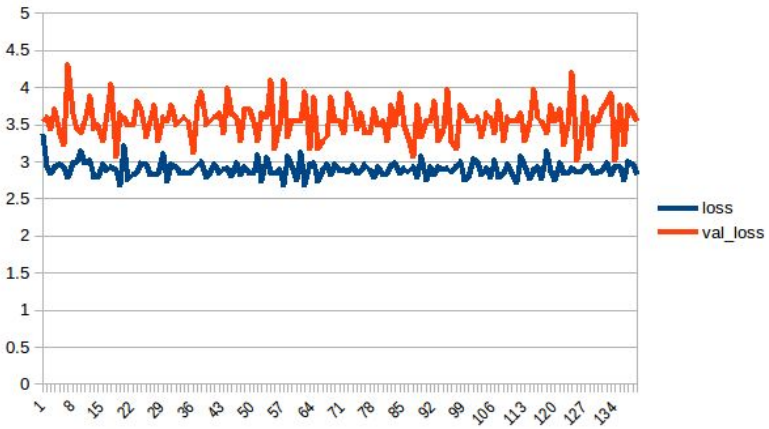

Fig 9: Loss function of RoR20-SD

In RoR-SD, we drop some layers during training time so that the expected length of the model during training is lower than during the validation time. This helps in reducing the training time because the amount of computations that have to be performed is less when compared to a resnet model of the same size. Moreover, it helps in giving each layer a chance to have less amount of abstraction on the original image and hence it can extract more features and classify better at the original place. It also makes it more robust and hence, it is found to perform better than the RoR without stochastic Depth.

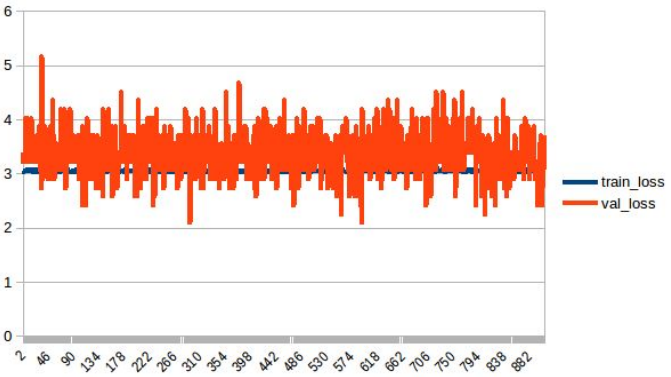

Fig 10: Loss function of RoR-SD-Weighted

We designed the RoR-SD with weighted dropping of layers so that we have a fine gradient with respect to the probability of dropping. It was an experiment which we performed so that we can probe the effect of weighted dropping of the layers and see its implications. We found that there is scope for better techniques of dropping layers.

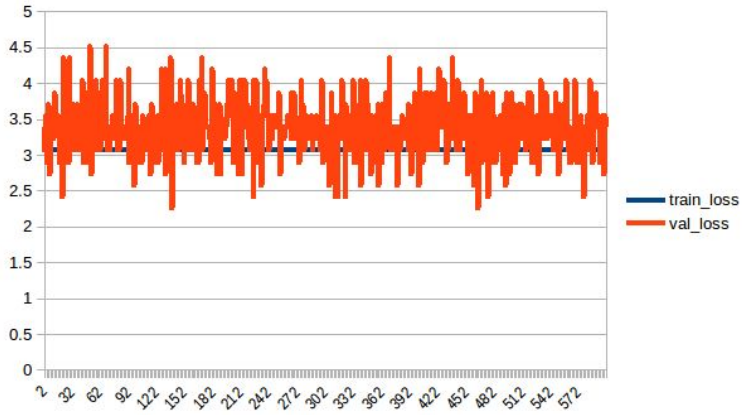

Fig 11: Loss function of Butterfly-100

Butterfly model was designed keeping in mind the properties of RoR, Resnet and the implications of symmetry. The training loss was slightly higher than RoR or RoR-SD due to the fact that number of layers in the model are nearly double of the number of layers in either of the above mentioned. But the pulsating behaviour of the val loss was peculiar since it had scope for very low validation loss. We suspect that the learning rate is high and hence, it is pulsating. But on the brighter side it has scope for further tuning up of the parameters to achieve lower validation loss.

In RoR, RoR-SD, RoR-SD weighted training loss is fairly constant but a sudden drop and further constant loss is expected but that didn't happen. We tried different training optimizers but the trend remained the same. We suspect the fact that dataset is small which didn't lead to the drop in the training loss. However the val loss has been spiking during the training to give good results at some epochs.

Automated Melanoma	ROR-SD	Butterfly Ground Truth	Butterfly Unet
85.5%	82%	87%	86%

Table 2: Accuracy measurement of Deep Learning Methods on Skin Cancer Dataset

In the accuracy table we found that the results obtained are best for the butterfly model. The difference in accuracy obtained by using ground truth based segmentation and Unet based segmentation is very less and hence, the butterfly model is robust to small inaccuracies in segmentation stage.

Chapter 6: Conclusion and Future Work

Our project dealt with Melanoma detection from skin lesion images through deep learning. This automated process has two phases, segmentation and classification. The segmentation is done by using a deep learning based classifier U-net. The u-net segmented images are sent to a new model that we designed, butterfly net.

The butterfly net model derives from residual networks and have proven to show better results than any of the existing residual architectures. The best results were obtained from a 50 layer deep architecture with two wings, essentially learning like a 100 layer architecture.
We obtained an 86% accuracy while classifying U-net segmented images and 87% accuracy while classifying an already segmented images. These results are better than the existing, state of art model that produces 85% accuracy.

Our future work includes mechanisms to train the butterfly model with more data. Presently, we have used 2000 images from the ISBI Challenge 2017. There is scope of getting at least twice the amount of present data from multiple channels. Also, we would like to work out possibilities of enhancements through fine tuning the learning parameters to obtain a robust model.

References

[1] R. Garnavi, M. Aldeen, M. E. Celebi, G. Varigos and S. Finch, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 105-115, 2011.
[2] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," Skin Research and Technology, vol. 19, no. 1, pp. e252-e258, 2013.
[3] D. D. Gómez, C. Butakoff, B. K. Ersbøll and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," IEEE Transactions on Biomedical Imaging, vol. 55, no. 1, pp. 157- 161, 2008.
[4] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," IEEE Transactions on Medical Imaging, vol. 18, no. 2, pp. 164-171, 1999.
[5] R. Melli, C. Grana and R. Cucchiara, "Comparison of color clustering algorithms for segmentation of dermatological images," in SPIE Medical Imaging, San Diego, 2006, pp. 61443S-61443S.
[6] M. Mete, S. Kockara and K. Aydin, "Fast density-based lesion detection in dermoscopy images," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 128-136, 2011.
[7] H. Zhou, M. Chen, R. Gass, L. Ferris, L. Drogowski and J. M. Rehg, "Spatially constrained segmentation of dermoscopy images," in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, 2008, pp. 800-803.
[8] H. Zhou, G. Schaefer, A. H. Sadka and M. E. Celebi, "Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images," IEEE Journal Selected Topics in Signal Processing, vol. 3, no. 1, pp. 26-34, 2009.
[9] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," Skin Research and Technology, vol. 11, no. 1, pp. 17-26, 2005.
[10] H. Zhou, X. Li, G. Schaefer, M. E. Celebi and P. Miller, "Mean shift based gradient vector flow for image segmentation," Computer Vision and Image Understanding, vol. 117, no. 9, pp. 1004-1016, 2013.
[11] Q. Abbas, M. E. Celebi, Irene Fondón García and M. Rashid, "Lesion border detection in dermoscopy images using dynamic programming," Skin Research and Technology, vol. 17, no. 1, pp. 91-100, 2011.
[12] M. I. Rajab, M. S. Woolfson and S. P. Morgan, "Application of region-based segmentation and neural network edge detection to skin lesions," Computerized Medical Imaging and Graphics, vol. 28, no. 1, pp. 61-68, 2004.
[13] X. Yuan, N. Situ and G. Zouridakis, "A narrow band graph partitioning method for skin lesion segmentation," Pattern Recognition, vol. 42, no. 6, pp. 1017-1028, 2009. 12
[14] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters and e. al, "Border detection in dermoscopy images using statistical region merging," Skin Research and Technology, vol. 14, no. 3, pp. 347-353, 2008.
[15] A. Wong, J. Scharcanski and P. Fieguth, "Automatic skin lesion segmentation via iterative stochastic region merging," IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 6, pp. 929-936, 2011.

[16] M. E. Celebi, G. Schaefer, H. Iyatomi and W. V. Stoecker, "Lesion Border Detection in Dermoscopy Images," Computerized Medical Imaging and Graphics, vol. 33, no. 2, pp. 148-153, 2009.
[17] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," IEEE Journal of Selected Topics in SIgnal Processing, vol. 3, no. 1, pp. 35-45, 2009.
[18] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," in Dermoscopy Image Analysis, M. E. Celebi, T. Mendonca and J. S. Marques, Eds., Boca Raton, CRC Press, 2015, pp. 97-129.
[19] Mishra, Nabin K., and M. Emre Celebi. "An overview of melanoma detection in dermoscopy images using image processing and machine learning." *arXiv preprint arXiv:1601.07843* (2016).
[20] M. Sadeghi, M. Razmara, T. K. Lee and M. S. Atkins, "A novel method for detection of pigment network in dermoscopic images using graphs," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 137-143, 2011.
[21] M. Sadeghi, T. K. Lee, D. McLean, H. Lui and M. S. Atkins, "Oriented pattern analysis for streak detection in dermoscopy images," in Medical Image Computing and Computer-Assisted Intervention (MICCAI 2012), vol. 7510, Nice, Springer Berlin Heidelberg, 2012, pp. 298-306.
[22] M. Sadeghi, T. K. Lee, D. McLean, H. Lui and M. S. Atkins, "Detection and analysis of irregular streaks in dermoscopic images of skin lesions," IEEE Transactions on Medical Imaging, vol. 32, no. 5, pp. 849-861, 2013.
[23] H. Mirzaalian, T. K. Lee and G. Hamarneh, "Learning features for streak detection in dermoscopic color images using localized radial flux of principal intensity curvature," in IEEE Workshop on In Mathematical Methods in Biomedical Image Analysis (MMBIA 2012), Breckenridge, 2012, pp. 97- 101.
[24] G. D. Leo, G. Fabbrocini, A. Paolillo, O. Rescigno and P. Sommella, "Towards an automatic diagnosis system for skin lesions: estimation of blue-whitish veil and regression structures," in 6th International Multi-Conference on Systems, Signals and Devices, 2009. SSD'09., Djerba, 2009, pp. 1- 6. 13
[25] S. Yoshino, T. Tanaka, M. Tanaka and H. Oka, "Application of morphology for detection of dots in tumor," in SICE 2004 Annual Conference, Sapporo, 2004, pp. 591-594.
[26] W. V. Stoecker, K. Gupta, R. J. Stanley, R. H. Moss and B. Shrestha, "Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color," Skin Research and Technology, vol. 11, no. 3, pp. 179-184, 2005.
[27] J. L. G. Arroyo, B. G. Zapirain and A. M. Zorrilla, "Blue-white veil and dark-red patch of pigment pattern recognition in dermoscopic images using machine-learning techniques," in IEEE International Symposium on In Signal Processing and Information Technology (ISSPIT), 2011, Bilbao, 2011, pp. 196-201.
[28] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano and H. P. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," Computerized Medical Imaging and Graphics, vol. 32, no. 8, pp. 670-677, 2008.
[29] M. E. Celebi, H. A. Kingravi, Y. A. Aslandogan and W. V. Stoecker, "Detection of blue-white veil areas in dermoscopy images using machine learning techniques," in SPIE Medical Imaging, San Diego, 2006, pp. 1861-1868.
[30] R. Kaur, P. P. Albano, J. G. Cole, J. Hagerty, R. W. LeAnder, R. H. Moss and W. V. Stoecker, "Real- time supervised detection of pink areas in dermoscopic images of melanoma:

importance of color shades, texture and location," Skin Research and Technology, vol. 21, no. 4, pp. 466-473, 2015.
[31] A. Dalal, R. H. Moss, R. J. Stanley, W. V. Stoecker, K. Gupta, D. A. Calcara and J. Xu, "Concentric decile segmentation of white and hypopigmented areas in dermoscopy images of skin lesions allows discrimination of malignant melanoma," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 148-154, 2011.
[32] G. Argenziano, I. Zalaudek, R. Corona, F. Sera, L. Cicale, G. Petrillo, E. Ruocco, R. HofmannWellenhof and H. P. Soyer, "Vascular structures in skin tumors: a dermoscopy study," Archives of dermatology, vol. 140, no. 12, pp. 1485-1489, 2004.
[33] J. Malvehy, S. Puig, G. Argenziano, A. A. Marghoob and H. P. Soyer, "Dermoscopy report: proposal for standardization: results of a consensus meeting of the International Dermoscopy Society," Journal of the American Academy of Dermatology, vol. 57, no. 1, pp. 84-95, 2007.
[34] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara and R. Hofmann-Wellenhof, "Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet," Journal of the American Academy of Dermatology, vol. 48, no. 9, pp. 679-693, 2003.
[35] R. P. Braun, H. S. Rabinovitz, M. Oliviero, A. W. Kopf and J.-H. Saurat, "Dermoscopy of pigmented skin lesions," Journal of the American Academy of Dermatology, vol. 52, no. 1, pp. 109-121, 2005
[36] A. Saez, C. Serrano and B. Acha, "Global pattern classification in dermoscopic images," in Dermoscopy Image Analysis, M. E. Celebi, T. Mendonca and J. S. Marques, Eds., Boca Raton, CRC Press, 2015, pp. 183-209.
[37] C. Serrano and B. Acha, "Pattern analysis of dermoscopic images based on Markov random fields," Pattern Recognition, vol. 42, no. 6, pp. 1052-1057, 2009.
[38] J. L. G. Arroyo and B. G. Zapirain, "Comparison of image processing techniques for reticular pattern recognition in melanoma detection," in Dermoscopy Image Analysis, M. E. Celebi, T. Mendonca and J. S. Marques, Eds., Boca Raton, CRC Press, 2015, pp. 131-181.
[39] A. D. Stefani, C. Massone, H. P. Soyer, I. Zalaudek, G. Argenziano, E. Arzberger, G. P. Lozzi, S. Chimenti and R. Hofmann-Wellenhof, "Benign dermoscopic features in melanoma," Journal of the European Academy of Dermatology and Venereology, vol. 28, no. 6, pp. 799-804, 2014.
[40] C. Barata, M. Ruela, M. Francisco, T. Mendonça and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," IEEE Systems Journal, vol. 8, no. 3, pp. 965-979, 2014.
[41] E. Alpaydin, Introduction to machine learning, Cambridge: MIT Press, 2014.
[42] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," IEEE Transactions on Biomedical Engineering, vol. 41, no. 9, pp. 837-845, 1994.
[43] R. Husemann, S. Tölg, W. V. Seelen, P. Altmeyer, P. J. Frosch, M. Stücker, K. Hoffmann and S. ElGammal, "Computerised diagnosis of skin cancer using neural networks," in Skin Cancer and UV Radiation, Springer Berlin Heidelberg, 1997, pp. 1052-1063.
[44] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker and R. H. Moss, "A methodological approach to the classification of dermoscopy images," Computerized Medical Imaging and Graphics, vol. 31, no. 6, pp. 362-373, 2007.
[45] A. Blum, H. Luedtke, U. Ellwanger, R. Schwabe, G. Rassner and C. Garbe, "Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer

algorithm based on analysis of 837 melanocytic lesions," Computerized Medical Imaging and Graphics, vol. 31, no. 6, pp. 362-373, 2007.
[46] P. Kahofer, R. Hofmann-Wellenhof and J. Smolle, "Tissue counter analysis of dermatoscopic images of melanocytic skin tumours: preliminary findings," Melanoma research, vol. 12, no. 1, pp. 71-75, 2002.
[47] G. Schaefer, B. Krawczyk, M. E. Celebi and H. Iyatomi, "An ensemble classification approach for melanoma diagnosis," Memetic Computing, vol. 6, no. 4, pp. 233-240, 2014. 15
[48] M. Abedini, Q. Chen, N. C. F. Codella, R. Garnavi and X. Sun, "Accurate and Scalable System for Automatic Detection of Malignant Melanoma," in Dermoscopy Image Analysis, M. E. Celebi, T. Mendonça and J. S. Marques, Eds., Boca Raton, CRC Press, 2015, pp. 293-343.
[49] L. Ballerini, R. B. Fisher, B. Aldridge and J. Rees, "A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions," in Color Medical Image Analysis, M. E. Celebi and G. Schaefer, Eds., Netherlands, Springer, 2013, pp. 63-86.
[50] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern and J. R. Smith, "Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images," in Machine Learning in Medical Imaging, Munich, Springer, 2015, pp. 118-126.
[51] A. Green, N. Martin, G. McKenzie, J. Pfitzner, F. Quintarelli, B. W. Thomas, M. O'Rourke and N. Knight, "Computer image analysis of pigmented skin lesions," Melanoma research, vol. 1, no. 4, pp. 231-236, 1991.
[52] Fornaciali, Michel, et al. "Towards Automated Melanoma Screening: Proper Computer Vision & Reliable Results." arXiv preprint arXiv:1604.04024 (2016).
[53] Yu, Lequan, et al. "Automated melanoma recognition in dermoscopy images via very deep residual networks." IEEE Transactions on Medical Imaging 36.4 (2017): 994-1004.
[54] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.