Deep Learning based Melanoma Detection from Skin Lesion Images

Harshil Dadlani
Amisha Somaiya
Aiswarya Janardanan
Ahmed Alhumoud
Harsha Vardhan

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2023

Reading Committee:

Professor Alexander Mamishev, Chair

Cynthia Yu

Jared Haar

Program Authorized to Offer Degree:

Electrical Engineering

University of Washington

Abstract

Deep Learning based Melanoma Detection from Skin Lesion Images

Harshil Dadlani
Amisha Somaiya
Aiswarya Janardanan
Ahmed Abdulsalam Alhumoud
Harsha Vardhan

Clinical diagnosis using Deep Learning has recently gained momentum as these methodologies aid the accurate and early detection of several diseases and thus help in saving lives. These predictive modeling methods are especially promising for melanoma detection due to the specific patterns found in the color, texture, and shape of melanoma lesions. Traditional methods are heavily based on 2D RGB images of skin lesions. The proposed algorithm combines the depth dimension representative of the skin's surface rugosity with the original RGB lesion images. This is further integrated with a feature extraction framework using wavelet scattering, generating a data processing pipeline that greatly improves model learning. Using combinatorial cues from RGB images and depth maps, the proposed method achieves a 1.53X higher Balanced Accuracy Coefficient score (BAC) compared to the baseline.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to all those who have contributed to completing our thesis at the University of Washington. First and foremost, we would like to thank our advisor, Professor Alexander Mamishev, for his invaluable guidance, support, and encouragement throughout the research process. His expertise, insights, and unwavering dedication have been instrumental in shaping our research direction and producing a high-quality thesis.

We would also like to extend our appreciation to the faculty members of the Electrical Department for their valuable feedback, comments, and suggestions. Their constructive criticism and insights have greatly enhanced the quality and rigor of our research.

We would like to thank the staff and administrators of the Electrical Engineering Department for their assistance, support, and resources that have enabled us to carry out our research effectively. We would also like to thank our fellow graduate students and colleagues for their helpful discussions, ideas, and feedback.

Chapter 1. INTRODUCTION

Melanoma is one of the most prevalent and lethal forms of skin cancer [1]. Since melanoma is a cancer of melanocytes, pigment producing cells on the skin, it is also one of the most visually manifesting cancers. Identification of melanoma from skin lesion images has the advantage of being non-intrusive and allows for early diagnosis over biopsy analysis. Early detection of melanoma is an important step to prevent further complications or death from the disease. Due to its non-invasive nature, computer vision-based melanoma classification could make melanoma identification accessible and affordable, saving many lives in the process. The release of the SKINL2 dataset [2], which contains RGB images of skin lesions and their spatially corresponding depth image, allows for new research that integrates the use of topographic features as part of the classification pipeline. The potential of using depth image data to improve melanoma diagnosis is significant because topological research on melanoma has shown that the lesions have differing groove patterns from non-cancerous skin [3]. Primary attributes that distinguish Melanoma from other lesions are based on asymmetry of the lesion, irregularity of the border, multi-color shade, and diameter of more than 6mm as proposed in ImageNet [4]. Deep learning-based pattern recognition techniques show potential in identifying and learning features corresponding to irregularity, color, and groove patterns.

Chapter 2. BACKGROUND

2.1 Magnitude of the Problem

Cancer is a major global health issue, with numerous fatalities recorded each year [1]. Amongst the various forms of cancer, skin cancer is one of the most frequently diagnosed types

[2]. Among the different types of skin cancer, melanoma is considered the most dangerous and aggressive, with a higher mortality rate than other forms of cancer worldwide [2]. The situation is particularly alarming in the United States, where melanoma has seen a staggering increase of nearly 60% between 1993 and 2019, highlighting the severity of the problem [3]. This increase in the incidence of melanoma emphasizes the need for continued research and effective treatments to address this growing public health concern. In the United States, the American Cancer Society (ACS) estimated approximately 106,110 cases of melanoma and 7,180 deaths due to melanoma in 2021 [5]. Regional variations in melanoma incidence are noteworthy, with higher rates seen in locations with more sun exposure [6]. For instance, Australia is known for having the highest melanoma per capita in the world, with a crude incidence rate of 8.7 times higher than the global norm [7]. More than 28,000 Australians receive in situ melanoma diagnoses yearly, and 17,000 receive invasive melanoma diagnoses [7]. However, if detected early, patients have an improved prognosis [7]. The similarity between the malignant "melanoma" and benign "nonmelanoma" skin cancer lesions makes it difficult to classify different cancer kinds [8]. Numerous techniques in image processing aid dermatologists in the early detection of skin cancer [8].

2.2 DATA CHARACTERISTICS

Medical patterns of hues, veils, pigmented nets, globs, and ramifications are the main characteristics of data used for melanoma detection. [9, 10] Other morphological structures used include ABCD (Asymmetrical form, Border anomaly, Color discrepancy, and Diameter) [9-11] To classify images in the past, dermoscopic images strongly depended on the extraction of handcrafted characteristics. [4, 12]. Non-professional dermoscopic images have a predictive value of 75-80% for Melanoma, but interpretation takes time and is highly subjective, depending on the

experience of the dermatologist; hence DL approaches could overcome this difficulty and aid early detection.

2.3   CURRENT METHODS

In [13], the author describes a way to fuse handcrafted and pre-trained CNN features for the purpose of classification. The handcrafted features are extracted using established image segmentation and feature extraction methods. In [14], the authors propose a method using state-of-the-art (SOTA) object detection techniques like YOLO to fine-tune a classifier for melanoma detection by inferring bounding box regions around malignant lesions. One of the most used datasets is the ISIC dataset [15] which contains RGB images of skin lesions. They obtained a 0.79 BAC (Balanced Accuracy Score) on the ISIC dataset with a non-linear SVM model and expert pathologists. None of these methods explore using the depth images captured using a plenoptic light-field camera as an additional feature for the purpose of classification.

Chapter 3. RELATED WORK

Prior to the release of the SKINL2 dataset, no project utilized melanoma topographical data for melanoma classification. Previous works on melanoma classification with computer vision techniques used biopsy or skin-surface images of skin lesions as input. One of the most used datasets is the ISIC dataset [6] which contains RGB images of skin lesions. Gutman obtained a 0.79 BAC score on the ISIC dataset with a non-linear SVM model and expert pathologists [6]. Unlike the ISIC skin lesion dataset, the images in the SKINL2 dataset were not captured with uniform imaging techniques and contained extraneous non-skin features like dense hair, body parts, and background objects. However, combining the depth and RGB images in the SKINL2

dataset could provide additional features to further improve performance at the melanoma classification task.

Chapter 4. DATASET
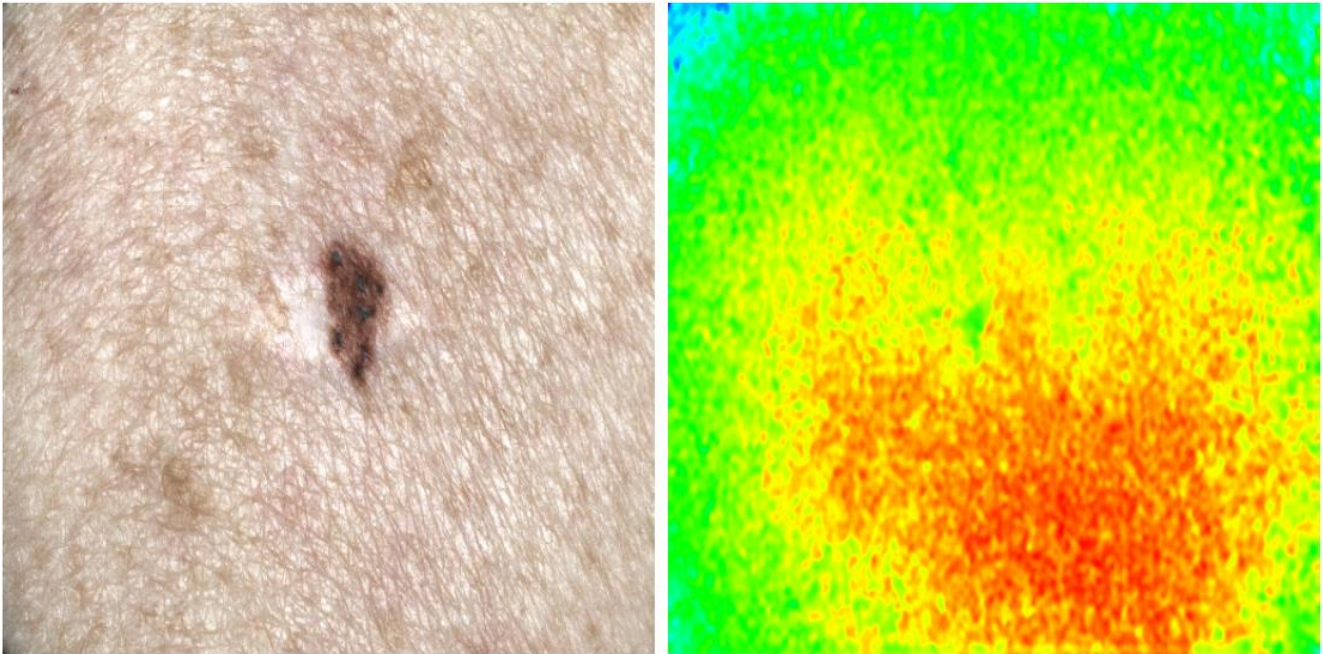
Figure 4-1. RGB (left) and corresponding depth-map (right) for a melanoma sample

The SKINL2 V1 dataset consists of 250 non-melanoma RGB images and depth-maps for different skin conditions like melanocytic nevus, psoriasis, benign keratosis, melanoma, angioma, and others. There are 15 images of melanoma samples, making the dataset highly imbalanced. The images were generated from a light-field imagery of the skin lesions, captured with a Raytrix R42 camera. Each RGB and depth image comprises 3858×2682 pixels. As the light-field imagery captures the incident angle of photons into the sensor, the camera can generate a topographical map of the skin lesion. The captured depth maps are outputted as TIFF images and then converted to color-mapped PNG images. Both the TIFF and PNG depth-maps are stored in the dataset but only the PNG format is used for our project.

Figure 4-1 shows an example of an RGB image and its corresponding depth-map for a melanoma sample from the SKINL2 V1 dataset. The RGB image (on the left) shows the color image of the skin lesion, while the depth-map (on the right) represents the topographical map of the skin lesion. The depth-map is a grayscale image that encodes the depth information of the skin lesion, with lighter regions indicating higher elevations and darker regions indicating lower elevations.

Chapter 5. THE PROPOSED APPROACH

Figure 5-1 provides an overview of the image processing pipeline. It's necessary to perform image segmentation to mask out and extract the lesion region from the healthy skin, as the classifier's performance could be hindered due to noisy features generated by the surrounding skin region. Once the segmented images have been generated, the images are resized to different sizes for classification iterations. The images are normalized to make the data distributions across image samples to have a similar scale, this speeds up the convergence of the optimizer used during the classification stage. A baseline classifier is created using segmented images. To improve upon the baseline results, a feature extraction framework using wavelet scattering is used for images of different sizes and types, like RGB and RGBZ, before the final classification step.
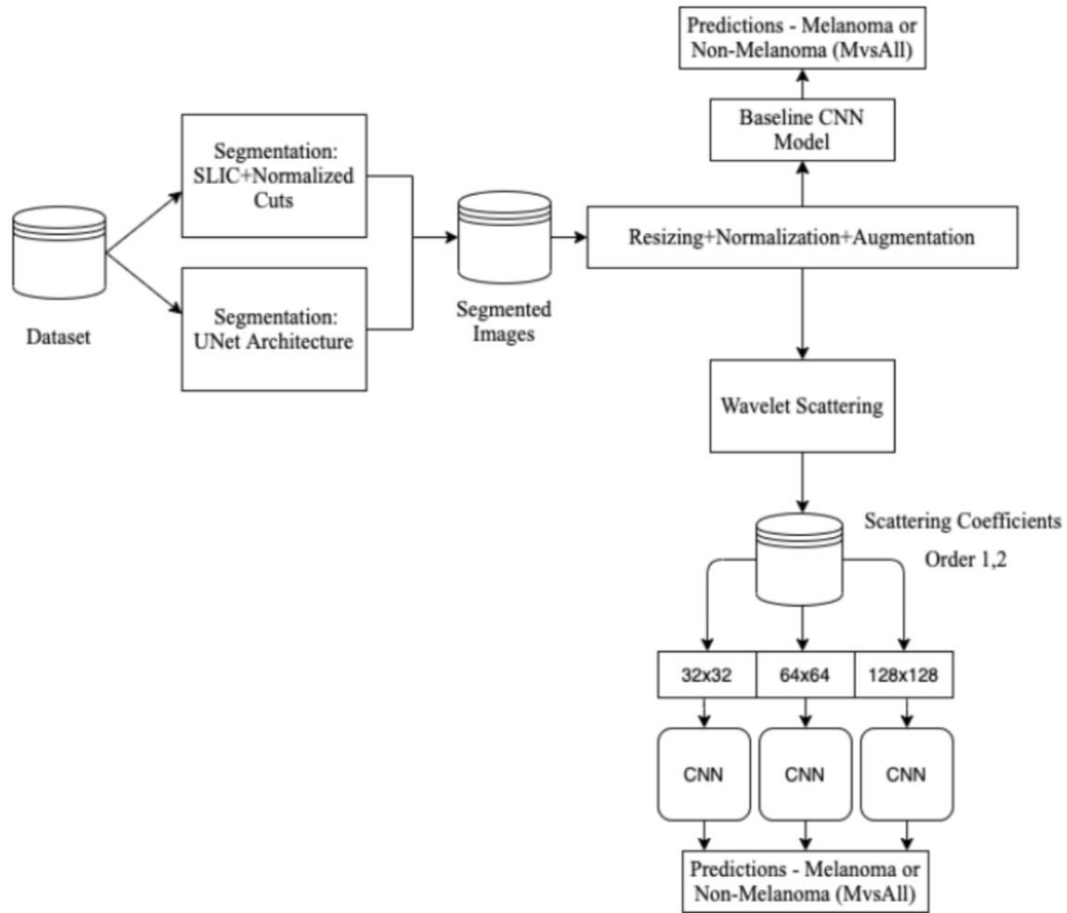
Figure 5-1. Image processing pipeline

5.1   SEGMENTATION

The images are resized to (512, 512) before segmentation to reduce the computational time for generating results. It is important to segment out the healthy skin region and hair from the images to prevent the CNN classifier from using these outlier features for classification. The image segmentation was carried out based on the process described in [7], called Normalized Cuts, where the grouping of similar pixels is carried out using SLIC, as proposed in [8]; this generates a low-dimension representation of the image since each super pixel region is now represented by a single

color, as shown in Figure 5-2 (top right). The super pixels are then used as nodes to construct an undirected graph of adjacent super pixels, with weighted edges. These weights are determined based on the affinity of each super pixel with its adjacent counterpart. The super pixels having high affinity are then retained to form a mask, as shown in Figure 5-2 (bottom left). We present a more automated approach for segmentation than the self-annotating approach used in lazy snapping as done by [9]. However, while implementing SLIC, it's important to choose the number of super pixel segments carefully since choosing a large number would generate multiple adjacent regions with strong affinity, and we end up having healthy skin regions to be included in the segmentation mask region. The procedure resulted in state-of-the-art results for approximately 85% of the images. The remaining images were segmented using a UNet architecture as proposed in [10] since it's effective at segmentation with limited training data. The segmented images from the SLIC + Normalized Cuts method were used as ground truth for the UNet model. The UNet architecture has multiple skip connections to retain spatial features through the deep layers, allowing for a better model convergence [4]. The mean of the UNet output is applied as a threshold to the output to generate the final mask for lesion segmentation. The output of this step comprises segmented lesions for the RGB and Depth (Z) images.
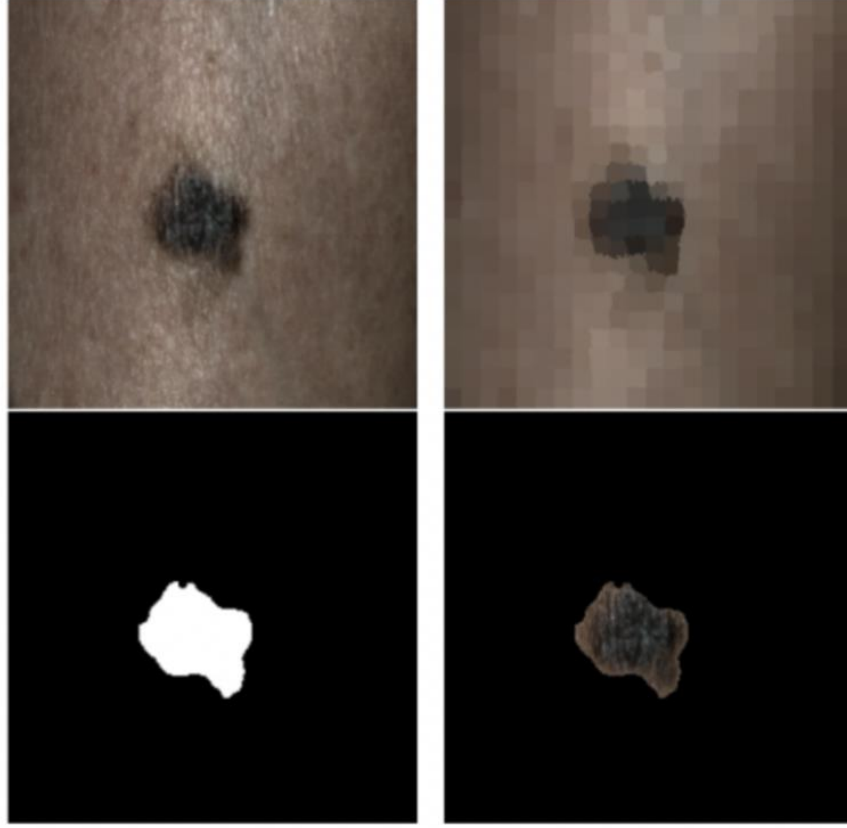
Figure 5-2. Original Image (top-left); the super pixels are formed by over segmenting the image using SLIC (top-right); the lesion mask is generated using normalized cuts (bottom-left); filtered out lesion region (bottom-right)

5.2   RESIZING, NORMALIZATION AND AUGMENTATION

The RGB and Depth (Z) images are resized to different dimensions (32×32), (64×64), and (128×128). This is done to compare the different image sizes on the classification results. Table **5-1** explains the comparisons between normalization strategies used for the RGB and Depth (Z) components. In the case of RGB, the normalization transforms each component of every image to be in the range [-2, 2]. We followed a standard technique that has been used in [4] and subtracted the values (0.485, 0.456, 0.406) and divided them by the values (0.229, 0.224, 0.225) from the respective components (R, G, B). In the case of the depth component, a similar operation is applied where we subtract 6.26 and divide by 3.03 from each component to ensure the range of the pixel

values are set between [-2,2]. The augmentation of the images has been performed by randomly rotating each image by an angle between (−180, 180) before the start of a new batch during classification to ensure that the CNN model generalizes well to new samples and does not overfit.

Table 5-1 Comparison of normalization transforms for RGB and RGBZ

Image Type	Image size	Subtracted Values	Divided Values
RGB	32 x 32	(0.485, 0.456, 0.406)	(0.229, 0.224, 0.225)
RGB	64 x 64	(0.485, 0.456, 0.406)	(0.229, 0.224, 0.225)
RGB	128 x 128	(0.485, 0.456, 0.406)	(0.229, 0.224, 0.225)
RGBZ	32 x 32	6.26	3.03
RGBZ	64 x 64	6.26	3.03
RGBZ	128 x 128	6.26	3.03

5.3   WAVELET SCATTERING

At the current stage of the pipeline, we have the segmented RGB images and the corresponding depth images (Z) resized and normalized. The small number of training examples prevents CNN from learning generalized high-level features in the initial layers. The wavelet scattering transform provides a substitute for the initial layers of a CNN by generating unique features invariant to translation, rotation, and scaling [12]. One key difference between scattering transforms, and deep neural networks are that in the case of scattering transforms, the filters are defined initially rather than being learned [13]. This enables the usage of shallower models, an important consideration given the size of the dataset. The transformation allows the generation of 1st and 2nd order sets of coefficients to be used for performing classification. The only difference is that the 2nd-order set of coefficients provides a denser feature stack than the 1st-order

coefficient. The filter banks corresponding to each scattering coefficient are convolved with x[n], with x[n] being the pre-processed segmented RGB or depth (Z) image. The family of wavelet coefficients in the filter banks is generated by translating and scaling the mother wavelet using equation (5.1); here, a is the scaling factor, and b is the translation parameter. Here, Ψ corresponds to the mother wavelet, which in this case is the Morlet wavelet.

$$\Psi_{a,b} = |a|^{\frac{1}{2}} \Psi \left(\frac{t-a}{b}\right) \tag{5.1}$$

These convolutions are performed in the frequency domain to reduce the overall time complexity. For coefficients of order 1 and 2, 17 and 81 scattering coefficients are generated for each image channel, respectively [9]. The tensors are formed by stacking these coefficients along the first dimension to ensure that features corresponding to the same spatial location are grouped together. For example, for an RGB image of size (64, 64) with coefficient order 1, the dimensions of the tensors would be $(17{\times}3, \frac{64}{4}, \frac{64}{4})$, computed using equation (5.2), here 17 scattering coefficients are obtained from the 1st order coefficient, and the multiplier 3 corresponds to the 3 channels in RGB. The division by 4 is the reduction in the size of the images when rotation filters are used to ensure that the results have rotation invariance, as explained in [9]. The procedure is performed for the RGB and depth (Z) images separately, and the tensors are stacked along the 1st dimension for the RGBZ images. The representation for the scattering tensors in terms of their dimensions would be as follows, $KC \times \hat{N}$, where KC corresponds to the total number of channels post-stacking and $\hat{N}$ represents the (N, N) of each scattering coefficient in the tensor, see Figure 5-3. Some of the extracted scattering coefficients for a melanoma image are shown in Figure 5-4. One final thing to note is that complete rotational invariance is not being induced because the maximum coefficient order we are using is 2 for the scattering coefficients to reduce the computational time for generating coefficients.

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - f}{s} \right\rfloor + 1 \tag{5.2}$$

Where, the size of input channels, f = filter size, p = padding size, s = stride
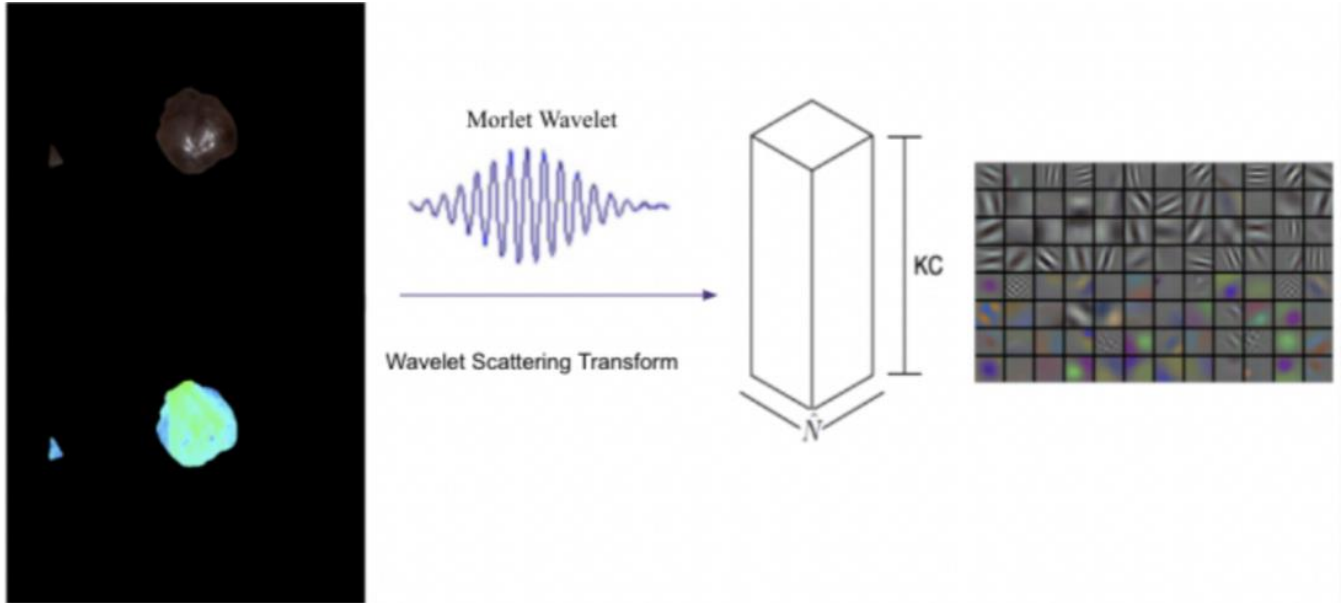
Figure 5-3. The stacked scattering coefficients generated have the dimensions $KC \times \widehat{N}$

Figure 5-4. Scattering coefficients corresponding to a melanoma image. The features of texture patterns have been extracted out by wavelet scattering.

5.4 CLASSIFICATION

The classification stage in the pipeline operates for 2 different tensor groups; the first group comprises the segmented images (3, N, N), where N corresponds to the width and height of the image and can also be represented as (C × $\widehat{N}$). Here C corresponds to the number of channels, and $\widehat{N}$ represents the dimensions of each image. These groups of images are used to create the baseline classifier. The second group comprises scattering coefficients of the segmented images (KC ×$\widehat{N}$), which are obtained from the previous stage of the pipeline. The group of scattering coefficients for the RGB and RGBZ components are used to improve the results over the baseline. Each image in the tensor group comprises a set of features represented by a $\widehat{V}$. These tensor groups are passed to the CNN model to classify the images into melanoma versus non-melanoma. The CNN architecture can be broadly divided into three parts:

i) A **fitting** part is where these features are convolved by a kernel to reduce the dimension of the features to fit them to a particular dimension.

ii) An **expansion** part contains two repeating working blocks that process the data with convolutions and expand the data into a higher dimension.

iii) A **discrimination** part where these higher dimensional data are passed into a fully connected network to output a melanoma or non-melanoma classification.

In the entire model, each of the convolution steps is followed by a batch normalization (see equation (5.3) and ReLU activation. Unless mentioned, we used a stride of 1 pixel and kernel size of 3x3 for all the convolution operations in the model.

$$y_i = \alpha \left(\frac{x_i - batch\_mean}{\sqrt{batch\_variance^2 + \epsilon}} \right) + \beta \tag{5.3}$$

where α, β = learnable parameters, ϵ = error for numerical stability, x_i = images in batch
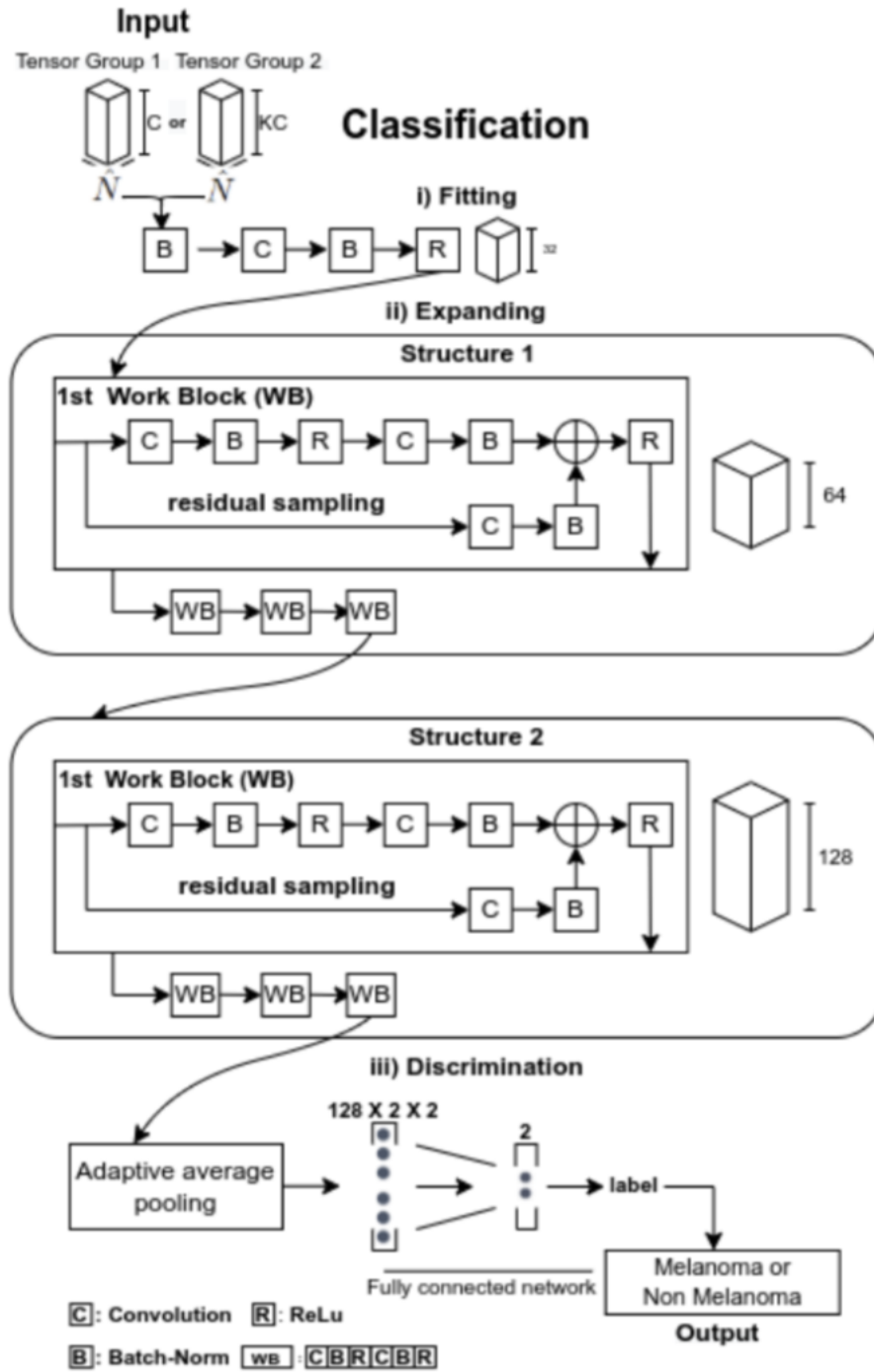

Figure 5-5. CNN architecture

5.4.1   FITTING

The fitting portion includes a batch normalization followed by a convolution operation by 32 kernels. In this layer, the features $\widehat{V}$ have a size of $C \times \widehat{N}$ or $KC \times \widehat{N}$, which vary with the tensor group type. Then, the tensors are transformed to a fixed size of $32 \times \widehat{N}$ after fitting. This part of the network has a total of $C \times 288 + 64$ or $K \times C \times 288 + 64$ trainable parameters.

5.4.2   EXPANSION

The outputs of the previous part are passed as inputs into this part of the network. The features entering this part would have a shape of $N \times N \times C$ or $N \times N \times KC$. The expansion part of the network comprises two structures, where each structure is made up of four working blocks. The differences between these structures are the number of kernels used for the convolutions in each of the 6 structures, which are 64 kernels in the first and 128 kernels in the second. Each working block performs two convolution operations, except for the first working block in that structure. The first working block in each structure does another convolution operation, which is a residual sampling layer that takes the same input as that of the first convolution layer of that working block. The residual layer performs convolution, followed by batch normalization, without a ReLU. This residue is then stacked up with the output of the second convolution batch normalization output, and a final ReLU operation is performed in the working block. However, the kernel size of the residue operation is 1×1. The two sections of the network structure have 279,860 and 1,116,416 trainable learning parameters, respectively.

5.4.3   DISCRIMINATION

The features that enter the discrimination part, after passing through the expansion, would have a shape of $128 \times \widehat{N}$. Since the features entering this part of the network could have a size of

$128 \times \hat{N}$, they will have a variable size in two of the three dimensions due to $\hat{N}$. Therefore, we use an average pooling layer on the features so that the features compiled for different input sizes remain comparable. Then, these features are passed into a fully connected layer to reduce the dimensions to $512 \times 2 \times 2$. This discrimination part of the network has 1026 total trainable parameters. The network is trained using Adam's optimizer with a learning rate of 0.0001. As the dataset is highly imbalanced in favor of the non-melanoma class, during training, we used a SoftMax cross-entropy loss function. This ensures that the error on the smaller class is weighted more than the error on the larger class, thereby attempting to balance the dataset.

Chapter 6. EXPERIMENTAL ASSESSMENT

Accuracy is not always a good metric to evaluate the results in medical imaging. There could be cases where the model could detect all the true negatives correctly and not the true positives. For example, if the validation data set has 2 positives and 14 negatives, the model could detect all 14 negatives but could fail to detect the 2 positives, but the accuracy could be 87.5 %, which is the wrong metric to evaluate the model. Thus, we use a balanced accuracy score (BAC), the average value between sensitivity and specificity.

$$\text{BAC} = \frac{sensitivity + specificity}{2} = \frac{1}{2}\left(\frac{truepositives}{positives} + \frac{truenegatives}{negatives}\right) \qquad (6.1)$$

Sensitivity (also known as the true positive rate) expresses a model's effectiveness at correctly identifying the positive cases as positive. Specificity (also known as the true negative rate) expresses a model's effectiveness at correctly identifying negative cases as negative. Balanced accuracy is simply the arithmetic mean of these two metrics. Balanced accuracy accounts for the imbalance in the dataset and is the most appropriate metric to present the results from our model.

Chapter 7. RESULTS

We use the average Balanced Accuracy score of the 10-fold cross-validation with 7 epochs to evaluate our models. For the baseline model, as shown in Table 7-1, which utilized the RGB segmented images, we used the architecture provided in the reference and achieved a Balanced Accuracy Score (BAC) of 0.45.

Table 7-1 Baseline BAC scores

Method	Image size	BAC
Baseline	32×32	0.34
	64×64	0.41
	128×128	0.45

The model developed using the data from tensor group-2, post-wavelet scattering, outperforms the baseline, which was developed using tensor group-1, for every combination of image types. Figure 7-1 shows the comparison of results between tensor group 1 and tensor group 2. The RGBZ inputs to our model outperform the baseline model; however, the performance of RGBZ inputs is worse than that of RGB inputs. This shows that the depth dimension (Z) did not contribute to improving the overall results. We also found no significant difference in the BAC scores on varying input image sizes Table 7-2 shows the Balanced Accuracy score for RGB input for various image sizes. The maximum BAC that was attained was 0.69 for 64x64 RGB input with a scattering coefficient order of 1, which is a 53.3% improvement in performance over the best baseline results. Table 7-3 shows the Balanced Accuracy score for RGBZ input for various image sizes. The maximum BAC attained was 0.61 for 32x32 RGBZ input with a scattering coefficient order of 2, which is just a 35.56% improvement in the performance over the best baseline results. Thus, RGB has a 17.74% improvement in performance over RGBZ.
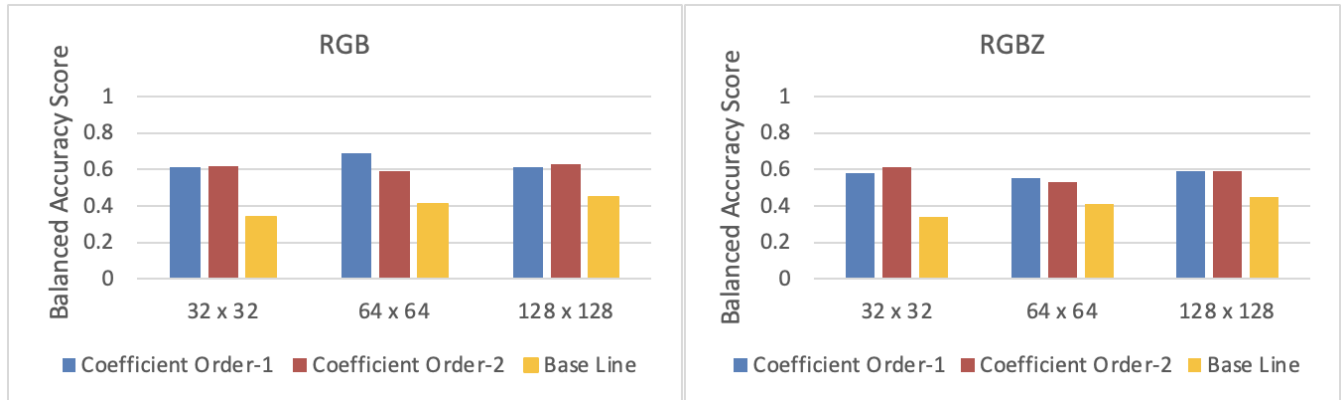
Table 7-2 Balanced Accuracy Score (BAC) of RGB

Order / Image size	Coefficient Order-1	Coefficient Order-2
32 x 32	0.61	0.62
64 x 64	0.69	0.59
128 x 128	0.61	0.63

Table 7-3 Balanced Accuracy Score (BAC) of RGBZ

Order Image Size	Coefficient Order-1	Coefficient Order-2
32 x 32	0.58	0.61
64 x 64	0.55	0.53
128 x 128	0.59	0.59

Figure 7-1. Comparison of results between tensor group 1 and tensor group 2

Chapter 8. DISCUSSION OF RESULTS

In this paper, we demonstrate an automated pipeline for skin lesion data preprocessing, training, and validation while evaluating the benefit of lesion depth images for melanoma classification. Overall, the developed model outperforms the baseline, showcasing the importance of wavelet scattering as a tool for feature extraction. We expected the inclusion of the Depth (Z) dimension in the RGBZ case to outperform the RGB results. However, that wasn't the case. On further manual inspection, we saw that the PNG depth images did not align with their corresponding RGB images spatially, and the PNG depth images also had lower topographic granularity as compared to TIFF depth images. This resulted in the inclusion of noisy information as part of the depth (Z) dimension. As mentioned previously, there were issues with the dataset.

The image capturing settings (i.e., brightness, locality, etc.) were not consistent throughout the photos, as compared to images in datasets that were used by papers mentioned in the related work section. In the proposed method, we initially tried k-means to segment out the image, but there was an overlap of clusters of lesions with healthy skin regions for images with multiple lesions. During the classification stage, to make up for the small size of the dataset, we generated 72 random augmentations of each tensor as part of the training data, as mentioned in [9], and shuffled the dataset to ensure distribution across batches. This resulted in 16200 (225×72) additional images in the training set while keeping the size of the validation set the same. However, this resulted in overfitting for many of the iterations, and validation loss remained constant throughout. We changed the optimizer to SGD and tuned the momentum to prevent the model from stagnating at a local minimum, and added L2 regularization to improve generalization, but the changes did not improve results. Finally, we also experimented by including only the depth (Z) 8 tensors as inputs, but due to the presence of noisy components, the results were only slightly better or equivalent to the baseline and were not worth reporting.

Chapter 9. FUTURE WORK

To properly evaluate our model, we propose using the light-field imaging technique to develop a dataset with consistent imaging settings so the model can focus on learning differences in the skin lesion rather than adapting to different image settings. Additionally, as discussed in the previous section, using the TIFF format for depth images could lead to improvements in the RGBZ results. As opposed to building and training a deep convolution neural network from scratch, we would also like to refine pre-trained medical imaging networks to classify melanoma [14]. This

could improve results as pre-trained networks are optimized to identify key features commonly seen during image classification.

BIBLIOGRAPHY

[1]   W. H. Organization. "Cancer." World Heath Organization. https://www.who.int/news-room/fact-sheets/detail/cancer (accessed 03 Feb 2022, 2022).

[2]   R. Ashraf *et al.*, "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," *IEEE Access,* vol. 8, pp. 147858-147871, 2020, doi: 10.1109/ACCESS.2020.3014701.

[3]   N. C. Institute. "Cancer Stat Facts: Melanoma of the Skin." https://seer.cancer.gov/statfacts/html/melan.html

[4]   P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data,* vol. 5, no. 1, pp. 1-9, 2018.

[5]   R. R. Maniyar *et al.*, "Interacting Genetic Lesions of Melanoma in the Tumor Microenvironment: Defining a Viable Therapy," *Tumor Microenvironment: Cellular, Metabolic and Immunologic Interactions,* pp. 123-143, 2021.

[6]   U. Leiter, U. Keim, and C. Garbe, "Epidemiology of skin cancer: update 2019," *Sunlight, Vitamin D and Skin Cancer,* pp. 123-139, 2020.

[7]   I. Economics, "Melanoma Institute Australia and Melanoma Patients Australia," *State of the nation. A report into melanoma–a national health priority. Final report. Sydney,* 2022.

[8]   Y. Filali, H. El Khoukhi, M. A. Sabri, and A. Aarab, "Analysis and classification of skin cancer based on deep learning approach," 2022: IEEE, pp. 1-6.

[9]   G. Alwakid, W. Gouda, M. Humayun, and N. U. Sama, "Melanoma Detection Using Deep Learning-Based Classifications," 2022, vol. 10: MDPI, 12 ed., p. 2481.

[10]   R. Thamizhamuthu and D. Manjula, "Skin melanoma classification system using deep learning," *CMC-COMPUTERS MATERIALS & CONTINUA,* vol. 68, no. 1, pp. 1147-1160, 2021.

[11]   G. Reshma *et al.*, "Deep Learning-Based Skin Lesion Diagnosis Model Using Dermoscopic Images," *Intelligent Automation & Soft Computing,* vol. 31, no. 1, 2022.

[12]   M. Efimenko, A. Ignatev, and K. Koshechkin, "Review of medical image recognition technologies to detect melanomas using neural networks," *BMC bioinformatics,* vol. 21, no. 11, pp. 1-7, 2020.

[13]   Y. Filali, H. El Khoukhi, M. A. Sabri, and A. Aarab, "Efficient fusion of handcrafted and pre-trained CNNs features to classify melanoma skin cancer," *Multimedia Tools and Applications,* vol. 79, no. 41, pp. 31219-31238, 2020/11/01 2020, doi: 10.1007/s11042-020-09637-4.

[14]   S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour," *IEEE Access,* vol. 8, pp. 198403-198414, 2020.

[15]   D. Gutman *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," *arXiv preprint arXiv:1605.01397,* 2016.