# TITLE PAGE

**AN ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS FOR DERMOSCOPIC IMAGES CLASSIFICATION**

A Thesis

Presented to the

School of Information Technology

Mapua Institute of Technology

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Computer Science

by

Camila, Dexter B.

Reyes, Alexander J.

Vega, Asaph F.

May 2020

# RECOMMENDATION SHEET

This is to certify that I have supervised the preparation of and read the thesis proposal prepared by **Asaph F. Vega, Alexander Reyes, and Dexter B. Camila** entitled **AN ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS FOR DERMOSCOPIC IMAGES CLASSIFICATION** and that the said paper has been submitted for final examination by the Oral Examination Committee.

|  |
| --- |
| **Prof. Joel C. De Goma MSc.** |
| Thesis Adviser |

Note:

This page will be replaced with the Acceptance Form, Revisions List, and Conformity of Revisions once Accepted.

# **TABLE OF CONTENTS**

[TITLE PAGE i](#_Toc19741698)

[RECOMMENDATION SHEET ii](#_Toc19741699)

[TABLE OF CONTENTS iii](#_Toc19741700)

[LIST OF FIGURES v](#_Toc19741701)

[LIST OF TABLES vi](#_Toc19741702)

[Chapter 1 1](#_Toc19741703)

[1.1. Context of the Study 1](#_Toc19741704)

[1.1. The Opportunity 3](#_Toc19741705)

[1.2. Research Objectives 3](#_Toc19741706)

[1.3. Research Questions 4](#_Toc19741707)

[1.4. Significance of the Study 4](#_Toc19741708)

[1.5. Scope and Limitations (or Delimitations) 5](#_Toc19741709)

[Chapter 2 6](#_Toc19741710)

[2.1. Introduction 6](#_Toc19741711)

[2.2. Skin Lesion Segmentation 6](#_Toc19741712)

[2.3. Skin Lesion Classification 8](#_Toc19741713)

[2.4. Survey for Feature Segmentation 11](#_Toc19741714)

[2.5. Survey for Image Classification 13](#_Toc19741715)

[2.6. Summary 15](#_Toc19741716)

[2.7. References 16](#_Toc19741717)

[Chapter 3 23](#_Toc19741718)

[3.1. Image Segmentation 24](#_Toc19741719)

[3.2. Convolutional Neural Networks 25](#_Toc19741720)

[Chapter 4 27](#_Toc19741721)

[4.1. Data Gathering 28](#_Toc19741722)

[4.3. Data Pre-processing 29](#_Toc19741723)

[4.3. Skin Lesion Segmentation 29](#_Toc19741724)

[4.4. Applying Transfer Learning 30](#_Toc19741725)

[4.5. Classification 32](#_Toc19741726)

[APPENDICES 33](#_Toc19741727)

[REFERENCES 34](#_Toc19741728)

# LIST OF FIGURES

Table 2.1 12

Table 2.2 14

25

Equation 2.1 25

Equation 2.2 26

Figure 3.2 Semantic segmentation networks [3] using deep deconvolutional layers 26

Figure 4.1 Training methodology 28

Figure 4.2 Results after resizing 28

Figure 4.2 Normalization 29

Figure 4.1 Training methodology 36

Equation 4.1 36

Figure 5.1 40

VGG16 Training Results 41

Figure 5.2 41

Figure 5.3 41

# LIST OF TABLES

Table 2.1 12

Table 2.2 14

# Chapter 1

**IN.TRO.DUCTION**

The American Cancer Society states that Melanoma is widely considered the deadliest of skin cancers and the incidence rates continues to rise worldwide. Skin cancer can occur anywhere on the body, even on areas not exposed to the sun such as the inside of the mouth or genitals, but most of the time, it will develop on sun-exposed areas. This includes the scalp, face, lips, ears, neck, chest, arms, hands, and on the legs [27]. While Merkel cell carcinoma has a higher chance to lead to a fatality, melanoma is the cause of mo.re dea.ths than a.ny other ty.pe of ski.n ca.n.ce.r overall [1]. This cancer is po.ten.tia.lly de.te.cta.ble using dermoscopy at a ve.ry ea.rly stage while it is still mana.geable to cure. Dermoscopy is defined as the method of capturing an illuminated, magnified image of skin lesions [2]. The advancement of machine learning and image analysis have presented the ability to identify distinct malignant melanoma from benign incidents allowing the possibility for early detection. By using the pre-trained ensemble of VGG16 boosted by the ResNet-152, this study will present a 7-class classification model for the most common skin diseases including Melanoma and Basal Cell Carcinoma.

## Context of the Study

Mo.st recent methods in the field of skin lesion classification re.ly on ha.nd-crafted

Feat.ures, su.ch as A.B.C.D.E rule (the acronym stands for Asymmetry, Border, Color, Dermoscopic structure a.nd Evolving) [43], 3-point checklist [44], 7-.poi.nt chec.klist [45], Men.zies met.hod [46], and C.A.S.H (Col.or, Arch.itec.ture, Sym.met.ry, and Homo.geneity) [47]. Physicians often rely on person.al experience and evaluate ea.ch patient’s lesions on a case-by-case basis by considering the patient’s local lesion patterns [48]. The accuracy of diagnosis for melanoma detection without computer-based assistance is reported to be between 65% and 80% [49].

Segmentation is typically used as a preprocessing method before the classification stage. Its purpose is to remove potentially irrelevant information from the input theoretically simplifying the classification process. Further research should be done in order to determine the benefits or lack thereof among dermoscopic image segmentation. The purpose of this study is to better understand the impact on the overall performance of the skin lesion classification architecture when the segmentation of the images is applied as a pre-process. This can be done by isolating and comparing the results of the original and segmented dermoscopic images. We hope the results will shed some light on whether the data surrounding the skin lesion is detrimental to the skin lesion classification process or are instead advantageous by providing relevant context.

According to the study of Mishra and Celebi, automating the dermoscopy method allows for faster and more frequent analy.sis of sk.in lesi.ons increasing the possibility of early detection [3]. While the efficiency of dermoscopy methods when evaluated by a dermatologist or licensed operator is relatively free of discrepancy with regards to skin tone, current deep learning image analysis techniques do not share the same results [4][5]. This can be addressed by improving the skin lesion segmentation process. However, there are several reasons skin lesion segmentation is a challenging task. Distinguishing between normal skin and lesional skin when contrast is poor is the primary reason. Mishra and Celebi proposed pre- and post-processing steps to maximize skin lesion segmentation accuracy not yet implemented by the related studies described in this paper.

Bisla et al. addressed the problem of images with less than optimal conditions and contaminants that affect lesion segmentation accuracy [51]. To resolve this data purification problem, Bisla et al. utilized traditional data processing methods to find and remove hair and ruler on the images. They extended the hair-removal algorithm [6] by overlaying the processed image with the segmented lesion obtained from their segmentation network. They utilized a U-Net architecture [7] that had been successfully applied to the problem of medical image segmentation and won the ISBI Cell Tracking Challenge [8] in 2015. Applying the techniques described here could potentially increase lesion segmentation accuracy and will be further discussed in chapter 3.

Codella et al. proposed the Deep Residual Network (DRN). They report state-of-the-art performance results using ConvNets to extract image descriptors by using a pre-trained model from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 [48]. They also investigate a more recent network structure called Deep Residual Network (DRN) [8].

Yuheng et al. discuss several of the most popular segmentation methods and applications [52]. The results of the study suggest that the necessity of segmentation in skin lesions remains unclear as to whether the skin data around the skin lesion affects the accuracy of the classification process.

## The Opportunity

The original authors of the U-Net architecture conducted experiments showing effective results and ranked first when applied to its original purpose single cell segmentation. Since its inception, the U-Net architecture has been successfully applied to various applications in the biomedical field. By incorporating the pre- and post-processing lesion segmentation methods proposed by Mishra and Celebi [3], the implementation of the U-Net architecture to skin.. lesion segmentation by Codella et al. [7], and the techniques known as Transfer learning put forth by the studies of Kawahara et al. [6], an opportunity for a comparative evaluation of classification without delegated segmentation (entire image as input) with the two approaches of segmentation followed by classification in order to assess which approach achieves better classification results presents itself.

## Research Objectives

The emergence of the deep learning paradigm and the advance in computational power have enabled the development of the automated medical image analysis that present remarkable performance and promising results. The primary objective of this study is to apply the most cutting-edge, state-of-the-art machine learning models to skin lesion segmentation and classification through transfer learning from existing models. This can be accomplished by fulfilling the following objectives:

* **Analyze** the aforementioned algorithms for prediction of skin lesion segmentations from dermoscopic images in the form of binary masks;
* **Modify** current methods for skin lesion disease detection by delegating segmentation and classification into separate CNN architectures;
* **Compare** the skin lesion segmentation and classification delegation effects on the accuracy of the overall architecture with results of classification alone.

## Research Questions

The following research questions were derived from the aforementioned objectives of this study:

* What affects would applying the Transfer Learning paradigm to the skin lesion segmentation model have on the training time of the skin lesion classification model?
* What impact will delegating the segmentation task to a separate, dedicated CNNs have on the performance of the overall architecture?
* What affects do segmented images of skin lesions have compared to the current techniques in which classification is conducted on the entire image?

## Significance of the Study

The accuracy of diagnosis for melanoma detection without computer-based assistance

is reported to be between 65% and 80% [49]. The focus of the study is on the problem of skin lesion classification and present a machine learning, deep neural network based solution for the problem of dermoscopic image classification containing a skin lesion as malignant or benign. The proposed solution is built around the VGG16 and U-Net convolutional neural network architectures and applies the transfer learning paradigm from top performing experiments.

With this study, we attempted to improve the methodology for classifying skin lesions by implementing the state-of-the-art techniques of various studies discussed in this paper.

## Scope and Limitations (or Delimitations)

The overall objective of this study is to improve existing methods of skin lesion classification by applying Transfer Learning of top performing architectures and related state-of-the-art techniques discussed in this paper. The inclusion, but not modification, of skin lesion segmentation and classification methods discussed in this paper will be implemented. Furthermore, the attempt to improve skin lesion detection accuracy will be primarily focused on whether the process of segmentation yields any comparative effects.

# Chapter 2

**REVIEW OF RELATED LITERATURE**

## Introduction

Dermatologists usually rely on using skin lesion-specific image cues to differentiate harmful lesions from benign lesions and to support their diagnosis. Dermoscopy (commonly known terms are epiluminescence micro.scopy or dermatoscopy) is a procedure that magnifies the inspected regions and removes the surface reflection of the region as well to get a much clearer image of the skin that improves the diagnostic accuracy [2]. The improvement of machine learning and image analysis have bestowed the power of detecting cancerous skin lesions from benign lesions permitting the likelihood of early detection.

## Skin Lesion Segmentation

As stated in the study of Mishra and Celebi, automating the dermoscopy technique by applying image processing and machine learning algorithms permits for faster and more of frequent analysis of skin lesions increasing the chance of early detection [4]. While the efficiency of dermoscopy methods when evaluated by a dermatologist or licensed operator is free of discrepancy with regards to skin tone, current deep learning image analysis techniques do not share the same results [4]. An operator has the ability to adjust the lighting, focus, and other parameters to produce the best possible image before analyzing whereas neural networks must be trained with predetermined images and thus cannot adjust any parameters when processing different skin types. Mishra and Celebi proposed pre- and post-processing steps for skin lesion segmentation, feature segmentation, and classification to maximize classification accuracy not yet implemented by the related studies described in this paper. The proposed model starts with a combination of pre-processing steps for lesion segmentation that plays an important role segmenting skin lesion. The proposed pre-processing steps includes elimination of variable lighting effects [7] [8] [9] [10], changing the data into a precise color region. [11], choosing correct channel of color. [12], enhancement of the chosen color channel. [13], increasing contrast. [14] [15], data purification (removing hair), normalization of color variation, image smoothing. [16] [17] [18] [19] [20], removing the vignette effect. [21], and localizing the lesion. [22]. The model has feature segmentation which is different from lesion segmentation. This method segments the feature of the lesions according to its pattern. In every feature segmentation can have multiple segments around the lesion region. Feature segmentation also includes pre-processing steps which are feature dependent, segmentation and post-processing steps. This feature segmentation procedure uses a completely distinctive approach that is obtain to manage artifacts that oftentimes obstructs the region of interest. The model used feature generation and classification to dictate if the lesion is malignant or harmless. The attributes used in feature segmentation was used to generate clinical feature. The attributes were used in the classifier to be able to differentiate melanomas from harmless lesions. It was concluded that automating dermoscopy by implementing image processing techniques and machine learning algorithms makes it faster and have more frequent analysis of skin lesions increasing the chance of its early detection and treatment [4].

Codella et al. proposed a system that employs hand-crafted feature extraction techniques including local binary patterns (LBP) [29]. These features have been used in previous research that have achieved best performance in dermoscopic images and other medical image datasets [32]. Image descriptors were extracted using a deep learning framework by using pre-trained models from Image Large Scale Visual Recognition Challenge (ILSVRC). Much more recent network structure to win the ImageNet 2015 which eases the optimization and convergence of extremely deep networks was also investigated [33]. They have discovered that object's important characteristics contribute complimentary data to the classifier which improved performance [32]. This new state of the art performance showed improvements in the area under receiver by a factor of 2.9 [29]. They also investigate a more recent network structure called Deep Residual Network (DRN) [29].

Kawahara et al. delve into the concept of engaging a pre-trained ConvNet as a feature extractor to differentiate amongst ten categories of non-dermoscopic skin images [42]. The paper showed how filters from a C.N.N pre-trained on natural images generalize to classifying 10 different classes of non-dermoscopic skin images, performing better than prior researches’ results. They hypothesized that subtracting the mean RGB pixel values computed over every individual image (per-image-mean) can improve the discriminant values within the ensuing feature vector. The per-image-mean gave improvements to classification’s accuracy [42].

## Skin Lesion Classification

Li et al. illustrated the effectiveness of deep convolutional networks and lower level image feature descriptors for skin lesion analysis even with limited and unbalanced training data. An image classifier and an interpretation method were used in this study. Convolutional Neural Networks was used as the image classifier since it has hierarchical feature learning ability that is broadly used in image classification and recognition [26]. The Convolutional Neural Networks based techniques performed better than traditional techniques particularly in the recent ImageNet challenges [27]. Two CCN image classification architecture was modified and fine-tuned: ResNet50 and VGG for encoding image features. LightGBM that boosts tree-based algorithm was also used to combine different CNN model features. The interpretation method used in the study was based on Zingraf et al.’s study [28] which visualizes how the CNNs respond to a specific corrupted input image in order to justify a specific classification created by the network. The method corrupts the pixels in a sliding window that scans the focused region, consequently, analyzes the distinction of prediction outcome. The results showed the effectiveness of deep convolutional networks and low-level image feature descriptors. As this study was conducted for the Open Challenge of Skin Lesion Analysis Towards Melanoma Detection (ISIC 2018), the dataset was limited and contained samples of predominantly lighter skin tones.

Haofu Liao proposed to create a universal skin disease classification using deep convolutional neural network. Although, the human engineered feature extraction is not advisable when pertaining to the universal skin disease classification system as well as the hand-crafted features in which it is purposed for a limited number of skin disease. To make the study feasible he used feature learning so that the machine will be trained to decide on which feature is best to be use [34]. Convolutional neural network became well-known in object classification and feature learning. High performance GPU allows the training of large-scale datasets, according to various researchers from Image..Net Lar.ge Scale Visual Recognition Challenge (ILSVRC) portrays that advanced Convolutional Neural Network Architectures are able to exceed humans in object classification [35]. It has been shown that in several cases transfer learning may be accustomed efficiently train a deep CNN [37, 38]. In transfer learning, rather than coaching the network from indiscriminately initialized parameters, individuals takes a pre-trained network and fine-tunes its weights by continued the backpropagation. In their approach, they did transfer learning by fine-tuning ImageNet [35] pre-trained models with Caffe [36], a deep learning framework that supports communicative and efficient deep CNN training. As they concluded that building a universal skin disease classification system using deep CNN is feasible by tackling the problem by fine-tuning ImageNet pre-trained models.

The aforementioned studies produced promising results when classifying skin lesions. The work of Kawahara et al. explores the idea of using a pre-trained CNN as a feature extractor to distinguish among 10 classes of non-dermoscopic skin images. Liao describes an attempt to construct a universal skin disease classification by applying transfer learning on a deep CNN and fine-tuned its weights. Codella et al. report state-of-the-art performance using CNNs to extract image descriptors by using a pre-trained model from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. They also investigate a more recent structure called Deep Residual Network (DRN).

## Survey for Feature Segmentation

Table 2.1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STUDY | DATASET | DISEASES | METHODS | ACCURACY | | SENSITIVITY | SPECIFICITY |
| Li et al. [26] | 10,015 images from ISIC 2018 | 327 images of Actine Keratosis, 514 images of basal cell carcinoma, 115 images of dermatofibroma, 1,113 images of melanoma, 6,705 images of nevus, 1,099 images of pigmented benign keratosis, 142 images of vascular lesion | VGG + ResNet50 | 0.85 | | N/A | N/A |
| Haofu Liao [34] | 23,000 images from Dermnet | Acne, Rosacea, Malignant Lesions, Atopic Dermatitis, Bullous Disease, Bacterial Infections, Eczema, Exanthems, Drug Eruptions, Hair Diseases, STDs, Pigmentation Disorders, Connective Tissue Diseases, Melanoma, Nevi & Moles, Nail Diseases, Contact Dermatitis, Psoriasis & Lichen Planus, Infestations & Bites, Benign Tumors, Systemic Disease, Fungal Infections, Urticaria, Vascular Tumors, Vasculitis, and Viral Infections | VGG16 | Top-1 | Top-5 | N/A | N/A |
| 72.7% | 91.0% |
| VGG19 | 73.1% | 90.9% | N/A | N/A |
| GoogleNet | 71.8% | 90.7% | N/A | N/A |
| 1,300 images from OLE | Rosacea, Actinic Keratosissis, Basal Cell Carcinoma, Squamous Cell Carcinoma, Atopic Dermatitis, Verruca, Nummular Eczema, Lupus Erythematosus, Melanoma, Melanocytic Nevus, Contact Dermatitis, Lichen Planus, Pityriasis Rosea, Psoriasis, Seborrheic Keratosis, Tinea Corposis, Tinea Versicolor, Urticaria, and Herpes | VGG19 | 24.8% | 61.7% | N/A | N/A |
| VGG19 Improved | 31.1% | 69.5% | N/A | N/A |
| Barata et al. [39] | 176 images from Hospital Pedro Hispano – Matosinhos | Not Specified | Global Features | Not Specified | | 96% | 80% |
| Local Features | 100% | 75% |

## Survey for Image Classification

Table 2.2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STUDY | DATASET | DISEASES | METHODS | ACCURACY | | SENSITIVITY | SPECIFICITY |
| Li et al. [26] | 10,015 images from ISIC 2018 | 327 images of Actine Keratosis, 514 images of basal cell carcinoma, 115 images of dermatofibroma, 1,113 images of melanoma, 6,705 images of nevus, 1,099 images of pigmented benign keratosis, 142 images of vascular lesion | VGG + ResNet50 | 0.85 | | N/A | N/A |
| Haofu Liao [34] | 23,000 images from Dermnet | Acne, Rosacea, Malignant Lesions, Atopic Dermatitis, Bullous Disease, Bacterial Infections, Eczema, Exanthems, Drug Eruptions, Hair Diseases, STDs, Pigmentation Disorders, Connective Tissue Diseases, Melanoma, Nevi & Moles, Nail Diseases, Contact Dermatitis, Psoriasis & Lichen Planus, Infestations & Bites, Benign Tumors, Systemic Disease, Fungal Infections, Urticaria, Vascular Tumors, Vasculitis, and Viral Infections | VGG16 | Top-1 | Top-5 | N/A | N/A |
| 72.7% | 91.0% |
| VGG19 | 73.1% | 90.9% | N/A | N/A |
| GoogleNet | 71.8% | 90.7% | N/A | N/A |
| 1,300 images from OLE | Rosacea, Actinic Keratosissis, Basal Cell Carcinoma, Squamous Cell Carcinoma, Atopic Dermatitis, Verruca, Nummular Eczema, Lupus Erythematosus, Melanoma, Melanocytic Nevus, Contact Dermatitis, Lichen Planus, Pityriasis Rosea, Psoriasis, Seborrheic Keratosis, Tinea Corposis, Tinea Versicolor, Urticaria, and Herpes | VGG19 | 24.8% | 61.7% | N/A | N/A |
| VGG19 Improved | 31.1% | 69.5% | N/A | N/A |
| Barata et al. [39] | 176 images from Hospital Pedro Hispano – Matosinhos | Not Specified | Global Features | Not Specified | | 96% | 80% |
| Local Features | 100% | 75% |

## Summary

The techniques discussed here have potential to improve both convolutional neural networks for skin lesion segmentation and classification. The pre-processing methods proposed by Mishra and Celebi include removing variable lighting effects, normalizing color variation, and deleting contaminants affecting the subject (i.e. data purification). These processes are to be applied to the data before entering the lesion segmentation stage. Post-processing steps were then applied to the output of the lesion segmentation stage before applying a binary mask of the lesion border. The process was then reiterated replacing the lesion segmentation stage with Clinical Feature segmentation in the next iteration and finishing with binary masks of clinical feature segments. Clinical features are specific structures in the lesion which separates them from benign lesions. A final iteration was then conducted involving feature generation, feature selection, and finally, model optimization before the final prediction.

The aforementioned studies produced promising results when classifying skin lesions. The work of Kawahara et al. explores the idea of using a pre-trained CNN as a feature extractor to distinguish among 10 classes of non-dermoscopic skin images. Liao describes an attempt to construct a universal skin disease classification by applying transfer learning on a deep CNN and fine-tuned its weights. Codella et al. report state-of-the-art performance using CNNs to extract image descriptors by using a pre-trained model from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. They also investigate a more recent structure called Deep Residual Network (DRN).

## References

[1] Key Statistics for Melanoma Skin Cancer. American Cancer Society.https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. Accessed: 2019-04-19.

[2] Dermoscopy for the Family Physician. Encyclopedia of Mental Disorders. https://www.aafp.org/afp/2013/1001/p441.html. Accessed: 2019-04-19.

[3] Philippines Economic Update April 2017: Http://www.workldbank.org/en/news/feature/2017/05/04/philippines-economic-update-april-2017. Accessed: 2019 – 04 – 29.

[4] Mishra, N., Celebi, M. 2016.An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning. https://arxiv.org/pdf/1601.07843.pdf. Accessed: 2019-04-20.

[5] Sun Protection Tips for Tropical Getaways. https://www.skincancer.org/prevention/tropical. Accessed: 2019-04-29.

[6] P. S. Saugeon, J. Guillod, and J. P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27(1):65 – 78, 2003.

[7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, MICCAI, 2015.

[8] EEE International Symposium on Biomedical Imaging (ISBI) Cell Tracking Challenge. http://www. celltrackingchallenge.net/index.html.

[9] C. Grana, G. Pellacani and S. Seidenari, "Practical color calibration fordermoscopy, applied to a digital epiluminescence microscope," Skin Research and Technology, vol. 11, no. 4, pp. 242-247, 2005.

[10] J. Quintana, R. Garcia and L. Neumann, "A novel method for color correction in epiluminescence microscopy," Computerized Medical Imaging and Graphics, vol. 35, no. 7, pp. 646-652, 2011.

[11] R. C. Gonzalez and R. E. Woods, Digital Image Processing, New Jersey: Pearson Education, 2008.

[12] 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.

[13] G. Schaefer, M. I. Rajab, M. E. Celebi and H. Iyatomi, "Colour and contrast enhancement for improved skin lesion segmentation," Computerized Medical Imaging and Graphics, vol. 35, no. 2, pp. 99-104, 2011.

[14] M. E. Celebi, H. Iyatomi and G. Schaefer, "Contrast enhancement in dermoscopy images by maximizing a histogram bimodality measure," in 6th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2601-2604.

[15] Q. Abbas, I. F. Garcia, M. E. Celebi, W. Ahmad and Q. Mushtaq, "A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images," Skin Research and Technology, vol. 19, no. 1, pp. 490-497, 2013.

[16] Q. Abbas, I. F. Garcia, M. E. Celebi and W. Ahmad, "A Feature‐Preserving Hair Removal Algorithm for Dermoscopy Images," Skin Research and Technology, vol. 19, no. 1, pp. 27-36, 2013.

[17] Q. Abbas, M. E. Celebi and I. F. García, "Hair removal methods: a comparative study for dermoscopy images," Biomedical Signal Processing and Control, vol. 6, no. 4, pp. 395-404, 2011.

[18] N. H. Nguyen, T. K. Lee and M. S. Atkins, "Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel," in SPIE Medical Imaging, San Diego, 2010, pp. 76234N-76234N.

[19] F.-Y. Xie, S.-Y.Qin, Z.-G.Jiang and R.-S.Meng, "PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma," Computerized Medical Imaging and Graphics, vol. 33, no. 4, pp. 275-282, 2009.

[20] T. Lee, V. Ng, R. Gallagher, A. Coldman and D. McLean, "Dullrazor®: A software approach to hair removal from images," Computers in Biology and Medicine, vol. 27, no. 6, pp. 533-543, 1997.

[21] F.-Y. Xie, Y. Lu, A. Bovik, Z. Jiang and R. Meng, "Application-Driven No-Reference Quality Assessment for Dermoscopy Images with Multiple Distortions," IEEE Transactions on Biomedical Engineering, vol. PP, no. 99, p. 1, 2015.

[22] M. E. Celebi, H. Iyatomi, G. Schaefer and W. V. Stoecker, "Approximate lesion localization in dermoscopy images," Skin research and technology, vol. 15, no. 3, pp. 314-322, 2009

[23] 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.

[24] 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.

[25] R. Kasmi, K. Mokrani, R. K. Rader, J. G. Cole and W. V. Stoecker, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," Skin Research and Technology, vol. 0, no. 0, pp. 1-15, 2015.

[26] Li, X., Wu, J., Chen, E., Jiang, H. 2018.What evidence does deep learning model use to classify Skin Lesions? https://arxiv.org/pdf/1811.01051.pdf. Accessed: 2019-04-29.

[27] Jia Deng, Olga Russakovsky, Jonathan Krause, Michael S Bernstein, Alex Berg, and Li Fei-Fei, “Scalable multi-label annotation,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2014, pp. 3099–3102.

[28] Luisa M Zintgraf, Taco S Cohen, Tameem Adel, and Max Welling, “Visualizing deep neural network decisions: Prediction difference analysis,” arXiv preprint arXiv:1702.04595, 2017.

[29] Codella, N. C.F., Nguyen, Q. B., Pankanti, S., Gutman, D., Helba, B., Halpern, A., Smith, J.R. Deep learning ensembles for melanoma recognition in dermoscopy images. https://arxiv.org/ftp/arxiv/papers/1610/1610.04662.pdf. Accessed: 2019-05-19.

[30] N. Codella, J. Connell, S. Pankanti, M. Merler, J.R. Smith. “Automated medical image modality recognition by fusion of visual and text information.” In: MICCAI 2014, Part II.LNCS, vol. 8674, pp. 487–495. Springer, Heidelberg (2014)

[31] C. Zhu, C. Bichot, and L. Chen, “Multi-scale color local binary patterns for visual object classes recognition.” In: 20th IAPR International Conference on Pattern Recognition (ICPR),pp. 3065–3068. IEEE Press, New York (2010)

[32] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, J. Smith. “Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images”. Proceedings of the 6thInternational Workshop on Machine Learning in Medical Imaging (MLMI), LNCS 9352, pp. 118-126. 2015.

[33] ImageNet Large Scale Visual Recognition Challenge 2015 (ILSVRC2015). http://image-net.org/challenges/LSVRC/2015/results. Accessed: 2019-05-23

[34] Haofu Liao. 1970. A Deep Learning Approach to Universal Skin Disease Classification. (January 1970). Retrieved May 23, 2019 from https://www.semanticscholar.org/paper/A-Deep-Learning-Approach-to-Universal-Skin-Disease-Liao/af34fc0aebff011b56ede8f46ca0787cfb1324ac

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.

[36] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.

[37] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. CoRR, abs/1403.6382, 2014.

[38]J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? CoRR, abs/1411.1792, 2014

[39] Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J. 2018. Two Systems for the Detection of Melanomas in Dermoscopy Images Using Texture and Color Features.http://welcome.isr.tecnico.ulisboa.pt/wp-content/uploads/2015/05/3430\_14\_sj.pdf. Accessed: 2019-05-23.

[40] T. Leung and J. Malik, “Representing and recognizing the visual appearance of materials using three-dimensional textons,” Int. J. Comput. Vis., vol. 43, no. 1, pp. 29–44, Jun. 2001.

[41] R. Fergus, P. Perona, and A. Zisserman, “Object class recognition by unsupervised scale-invariant learning,” in Proc

[42] Kawahara, J., BenTaieb, A., Harmaneh, G. 2016. Deep features to classify skin lesions. https://www.triage.com/publications/isbi2016b.pdf. Accessed: 2019-05-23.

[43] F. Nachbar, W. Stolz, T. Merkle, A.B. Cognetta, T. Vogt, M. Landthalerv, P. Bilek, O. Braun-Falco, and G. Plewig. The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. In Journal of the American Academy of Dermatology, 30(4), pp.551-559., 1994.

[44] H.P. Soyer. Three-point checklist of dermoscopy. In Dermatology, 208(1), pp.27-31., 2004.

[45] G. Argenziano. Seven-point checklist of dermoscopy revisited. In British Journal of Dermatology 164, no. 4 (2011): 785-790., 2011.

[46] S.W. Menzies. A method for the diagnosis of primary cutaneous melanoma using surface microscopy. In Dermatologic clinics, 19(2), pp.299-305., 2001.

[47] J.S. Henning, S.W. Dusza, S.Q. Wang, A.A. Marghoob, H.S. Rabinovitz, D. Polsky, and A.W. Kopf. The cash (color, architecture, symmetry, and homogeneity) algorithm for der- moscopy. In Journal of the American Academy of Dermatology, 56(1), pp.45-52., 2007.

[48] J. Gachon, P. Beaulieu, J.F. Sei, J. Gouvernet, J.P. Claudel, M. Lemaitre, M.A. Richard, and J.J. Grob. First prospective study of the recognition process of melanoma in dermatological practice. In Archives of Dermatology, 141(4), 2005, 434-8, 2005.

[49] G. Argenziano and H.P. Soyer. Dermoscopy of pigmented skin lesions–a valuable tool for early diagnosis of melanoma. In The Lancet Oncology, 2(7), 2001, 443-9., 2001

[50] Blum A, Luedtke H, Ellwanger U, Schwabe R, Rassner G, Garbe C: Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. Br J Dermatol 2004, 151: 1029–1038. 10.1111/j.1365-2133.2004.06210.x

[51] D. Bisla, A. Choromanska, R. Berman, D. Polsky, J. Stein, Towards Automated Melanoma Detection with Deep Learning: Data Purication and Augmentation, in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ISIC Skin Image Analysis Workshop, 2019

[52] S. Yuheng, Y. Hao. Image Segmentation Algorithms Overview. https://arxiv.org/abs/1707.02051. Accessed 10-14-2019.

# Chapter 3

**THEORETICAL FRAMEWORK**

There are two main components of this study that will discussed here. During the past decades many approaches have been proposed to automatically generate image representations that can provide support to tasks like image classification, object detection, recognition or semantic segmentation. Most of them have relied on hand-engineered low-level descriptors. But since the publication of AlexNet in 2012, state-of-the-art methods in computer vision mostly rely on learning representations using deep convolutional neural networks.

Most current methods in the field of skin lesion classification still rely on hand-crafted features. Typically, after the feature extraction based on these descriptions, machine learning methods such as k-nearest neighbors (kNN), Artificial Neural Networks (ANNs), logistic regression, decision trees and support vector machines (SVMs) have been explored to solve the classification task with moderate success [10].

Examples of related work using hand-crafted features and popular classifiers include:

• Codella et al. [11] use hand-crafted feature extraction techniques including color histogram, edge histogram, and a multi-scale variant of color local binary patterns (LBP).

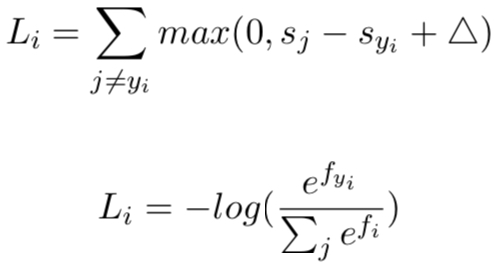
• Barata et al. [12] explore two different methods for the detection of melanoma in dermoscopy images based on global and local features. They conclude that color features perform much better than texture features alone.

More recently, the emergence of a machine learning paradigm known as deep learning has enabled the development of medical image analysis systems that can display remarkable accuracy, to the point of raising concerns about the future of the human radiologist [13][14].

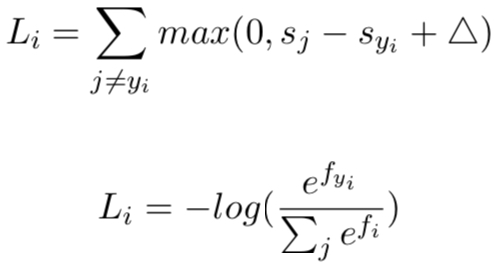
The next sections provide an overview of the deep learning techniques for image segmentation and image classification used in this study.

# Convolutional Neural Networks

Convolutional Neural Networks (also known as CNNs or ConvNets) are inspired in the behavior of biological systems through artificial neurons with learnable weights and biases. The layered architecture that Neural Networks performs based on matrix multiplications enables its application for image classification tasks. For this reason, ConvNets architectures assume that the input are images that have to be transformed into an output holding the class score predicted. The loss function is used to measure how well the predicted scores agrees with the ground truth labels in the input data. Most common loss functions are the Multiclass Support Vector Machine (SVM) (Equation 2.2) and the Softmax (Equation 2.3).



Equation 2.1



Equation 2.2

# Image Segmentation

Typically, segmentation is used as a preprocessing method in the classification process to remove potentially non-relevant information from the classification process [21]. ConvNets are typically used on image classification tasks: in the case of supervised learning, the input are images sorted by class, and the output to a certain image is a single class label. However, in the biomedical image processing field, a localization is often also required in addition to a global scale label, i.e. the network assigns a class label to each pixel. This is the main idea of a semantic segmentation model [15] using ConvNets.

Novel proposals for semantic segmentation [16] using convolutional neural networks introduce the idea of deconvolutional networks on the top of common ConvNets. This backward-strided convolution is convenient to generate a segmentation map of the input through a sequence of deconvolution operations. These layers (shown in Figure 2.1) compute a dense pixel-wise class probability map by consecutive operations of unpooling, deconvolution, and rectification.

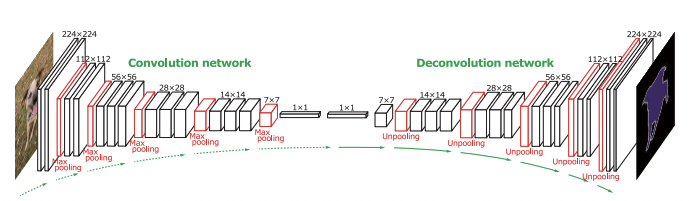


Figure 3.2 Semantic segmentation networks [3] using deep deconvolutional layers

1. Transfer Learning

Training an entire convolutional neural network in the medical imaging field is not always possible, due to the fact that datasets are often not large enough. Alternatively, random initialization of weights is replaced by a pre-trained network on large datasets, i.e. ImageNet [18], that contains 1.2 million images labeled with 1,000 classes. This technique is known as Transfer Learning [19] and it is very common in machine learning scenarios. The main objective is the improvement of learning in a new task by transferring knowledge from a related task that has already been learned.

In practice, Transfer Learning from a pre-trained ConvNet is typically used in these two different ways [20]:

• **Fixed Feature Extraction**. Use the output of one intermediate layer to train a machine learning classifier.

• **Fine-tuning**. Replace the last fully connected layer of the network with a new classification task and use the training data from the new domain to update the weights of some layers. In general, the earlier layers of the ConvNets are not updated and only the deeper ones are updated to the new task.\

1. Conceptual Framework

The conceptual framework of this study was derived from a few prominent, yet relative studies of Ronneberger [7] and Li et al. [26]. In the former study, the U-Net architecture was used to segment biomedical images for cell tracking and yielded outstanding results. This study utilized the same techniques in segmenting skin lesions from dermoscopic images. The latter study encompassed the process of classifying skin lesions efficiently using the VGG16 + ResNet50 architecture.

A close up of a device

Description automatically generated

Figure 4.1 Training methodology

A picture containing pink, sitting, looking, food

Description automatically generatedThe training methodology depicted above was inspired by the studies of the two aforementioned researchers. Using the popular HAM10000 7-class skin lesion dataset, a map was created to easily identify each image skin type during the development process and to simplify the preprocessing methods. This was followed by a resampling technique called downsampling which removed some of the images belonging majority class to minimize the disparity amongst the classes and therefore reducing the possibility of overfitting. After downsampling, the images were then uniformly resized (**Figure 4.2**). It

Figure 4.2 Results after resizing

was specifically done in this order to reduce the amount of computing resources required. Then, a normalization process was applied to 90% of the images after removing any null values. The remaining 10% of the images were label encoded (**Figure 4.3**). The normalization process included mean subtraction and standard deviation via the Numpy library. As the images was originally collected from different sources, the normalization process was important because it reduced discrepancies across the images. The dataset was then shuffled before doing a 70/30 split for training data and validation respectively. A screenshot of a cell phone

Description automatically generatedFinally, the training and validation sets were reshaped to 3 dimensions of 96 by 128 by 3.

Figure 4.2 Normalization

Transfer Learning was the next step in this study. The dataset was then used to fine-tune the VGG16 and ResNet50 models. Both pretrained models were loaded from the Keras API. Through trial and error, it was determined that adding a Dropout layer to the end of the VGG16 network helped reduce overfitting. In the initial tests of the models followed the same configuration used in the research of Li et al [26]. Due to hardware restrictions, the batch size was increased to 10 and the number of epochs was set to 100 (down from 500). The initial setting of 500 appeared to be redundant in our experiment as the loss rate began to fluctuate when approaching 100 epochs while the accuracy rate remained more or less the same (see the following figure).

A screenshot of a cell phone

Description automatically generated

A screenshot of a map

Description automatically generatedThe results after training the classification models still showed signs of overfitting as the validation accuracy rates digressed.

In order to combat this, class weights were computed and passed into the bias parameter of each classification model. This drastically improved the validation accuracy rates as well as the loss rates.

A screenshot of a cell phone

Description automatically generated

Once each of the classification models had been trained and optimized, the ensemble was ready to be created. The ensemble used the Average of the outputs and was compiled using the SGD optimizer. The loss function applied was Categorical Cross Entropy.

The U-Net model was trained using the ISIC 2018 skin lesion segmentation dataset which included binary masks for each image. This, however, introduced inconsistencies in the ensemble as it only contains cases of melanoma and was perhaps the cause of the low confidence rates in the final results. Despite this, the segmentation model was able to perform well on most of the classes in the HAM10000 dataset. This may be due to the diversity within the images of melanoma in the ISIC 2018 dataset.

A picture containing food

Description automatically generated

The network used followed the same network structure as described in the study of Ronneberger [7]. One significant modification was made, however, as the dimensions of the input layer was not compatible with the HAM10000 dataset and the output of the classifier models. The original study had also set the epochs to 500 which, as with the classifier models, was not feasible considering the hardware limitations of this study. Epochs again were set to 100 for the same reasons. Accuracy and loss rate results were more than acceptable and only a few points less than what Ronneberger [7] had achieved.

Applying the masks was a straightforward process using the Bitwise AND operation:

A screenshot of a cell phone

Description automatically generated

# Chapter 4

**METHODOLOGY**

This is a study in the Deep Learning field about the impact and effects of removing skin image background by applying image segmentation methods for a subsequent classification of skin lesions. The goal of this study is to better understand the effects on performance results when segmenting a dermoscopic image by isolating and comparing results from both unaltered and U-Net segmented skin lesion images. Specifically, it hopes to better understand whether the values outside the lesion are detrimental to lesion classification or are instead beneficial to lesion classification by providing contextual information relevant to each lesion.

To achieve the project goal, two successful and well-known Convolutional Neural Networks architectures in the image semantic segmentation and image classification tasks have been adopted. The order of execution is:

1. **Skin lesion segmentation**. The first task performed an automatic prediction of lesion segmentation from dermoscopic images taking the form of binary masks using the U-Net architecture.

2. **Skin lesion classification**. The second task performed the automatic classification amongst 7 classes of skin diseases. In order to find the best classification model, this task was divided into two subtasks according to different type of input skin images:

(a) **Unaltered lesion classification**. The basic model performed the classification over the original skin RGB images contained in the ISIC dataset.

(b) **U-Net Segmented lesion classification**. The third, and most complex model performed the classification over the automatically segmented images from the U-Net segmentation model.

## Data Gathering

Datasets from preceding research [23] were gathered in order to maintain consistency amongst the results across the studies mentioned in this paper (ISIC 2018, HAM10000). Images in the “Human Against Machine with 10000 training images” dataset (HAM10000) include a representative collection of all important diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc) totalling in 7 classes of skin diseases.

More than 50% of lesions were confirmed through histopathology (histo), the ground truth for the rest of the cases is either follow-up examination (follow\_up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal). As the supervised learning technique was used in this study, all of the aforementioned datasets contain labels for each image, while the ISIC 2018 dataset includes binary masks and metadata such as patient biometrics.

## Data Pre-processing

This project took advantage of ConvNets properties regarding input preprocessing: few previous processing techniques are needed. Although some basic preprocessing forms were performed:

• **Mean subtraction**. In order to center the cloud of RGB values from input data around zero along every dimension of the image, a mean subtraction was applied across the image features.

• **Image normalization**. By dividing each RGB dimension of input images by its standard deviation, a normalization was obtained from its original 0- and 255-pixel values to 1 and 0 normalized values. This preprocessing technique avoids further issues caused by poor contrast images.

• **Image cropping and resizing**. Input images were preprocessed to be accepted by the architecture though cropping the image to the same aspect ratio as needed and resizing the original image to 96x128 pixels for the U-Net, VGG-16, ResNet50, and the 12-layer CNN used to augment the ensemble.

The dataset had been shown to be imbalanced with images of Melanocytic Nevi

coming in at over 6,000; meaning that the remaining classes only account for about 70% of the available images in the dataset. To combat this, resampling methods were employed. Downsampling appeared to be the most logical solution given that a single class is overwhelmingly represented. The images belonging to the majority class were shuffled then reduced to 1,500 to more closely match the next largest class.

## Skin Lesion Segmentation

The overview of the feature extraction task followed the main structure described in section 3.1. Training and Validation data from the Segmentation dataset were processed with a Python script in order to load the images and masks and convert them into NumPy binary format files (.npy) that allowed a faster loading during the learning and prediction algorithm. Data were previously organized into train, train masks and test folders.

The proposed model is the U-Net network trained from scratch, which means that weights were randomly initialized and optimized with backpropagation. The network output was an array containing the 379 test binary masks, which can be converted into JPG images of 96 to 128 pixels dimension. This means a posterior post-processing technique was needed to enlarge the images.

# Applying Transfer Learning

The same training methodology was proposed to face the two tasks of this project. The only difference between the segmentation and classification tasks are the models used. As previously commented in Section 3.3, the U-Net architecture is the architecture proposed for the segmentation task, while the VGG-16 is associated with a classification model. Following the diagram of Figure 4.1, the training data was trained through the learning algorithm defined by each model, which applied the stochastic gradient descent (SGD).

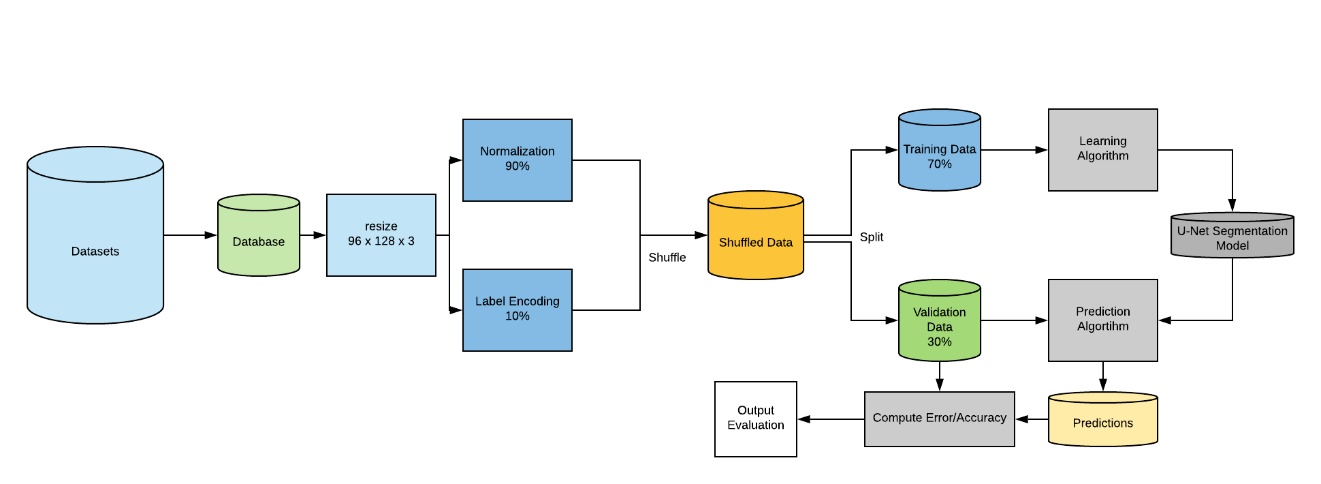


Figure 4.1 Training methodology

During training, some parameters were considered to be altered in order to get the best performance of the network proposed regarding the problem to be solved. Typical ConvNets parameters are the following:

• **Batch size**. The batch size was attributed to the number of training images in one forward or backward pass. It is important to highlight that the higher the batch size, the more memory will be needed.

• **Iterations**. The number of iterations was the number forward or backward of passes: each pass using a batch size number of images.

• **Epoch**. The number of epochs measured how many times every image has been seen during training (i.e. one epoch means that every image has been seen once). It can be also understood as a one forward pass and one backward pass of all the training examples. It is numerically computed as:



Equation 4.1

• **Loss function**. Loss function (also called cost function) evaluated the penalty between the prediction and the ground truth label in every batch.

• **Learning rate**. The learning rate parameter defined the step size for which the weights of a model were updated regarding the stochastic gradient descent.

• **Decay**. The weight decay was an additional weight update parameter that induced the weights to exponentially decay to zero once the update process is over.

• **Optimizer**. Keras framework provided optimizers [58] in order to find the most optimal set of hyperparameters for the model.

# Classification

The main issue of the classification task was to avoid overfitting caused by the small number of images of skin lesion in most dermatology datasets. In order to solve this problem, the objective of the proposed model was to firstly extract features from images with the original reduced network VGG-16 (with 138m parameters) over the VGG-19 (with 144m parameters) and secondly load those extracted representations on a fine-tuned VGG-16 architecture. A pre-trained VGG-16 model trained with the ImageNet dataset implemented in Keras can be found on Github[[1]](#footnote-1). The classification model was then boosted with a ResNet-50 model pre-trained on the ImageNet dataset and augmented by a 12-layer CNN trained by the same dataset. Some recent work done by Kornblith et al. [26] showed that ResNets are the leading feature extractors in terms of performance. The pre-trained ResNet model used in this study can also be found on Kaggle[[2]](#footnote-2). The standard k-fold cross-validation was done by partitioning the data into subsets (folds) then iteratively trained (excluding the last fold) while the remaining “holdout fold” served as the test set.

# Validation

In order to validate the segmentation model, the ISIC 2018 dataset were split between training data and validation data (70% and 30% respectively). The overview of the complete training process was illustrated in Figure 4.1. The K-fold cross-validation resampling procedure was implemented in the training process of the classification models (VGG-16 + ResNet50) using the HAM10000 dataset. The procedure were as follows:

1. Shuffle the dataset randomly
2. Split the dataset into k groups
3. For each unique group:
   1. Take the group as a hold out or test data set
   2. Take the remaining groups as a training data set
   3. Fit a model on the training set and evaluate it on the test set
   4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores

Once the classification model has learned the weights with the training data, the output was compared with the annotated dataset in order to determine if the test data had been correctly classified. A final model evaluation were performed by comparing the predictions with the ground truth data and results reviewed by a licensed dermatologist, Dra. Patricia Tinio of Makati Medical Center.

# Chapter 5

**DISCUSSION OF THE RESULTS**

Over the course of the development of the models used in the study, several key changes had been made to optimize the results. Reviewing the evaluation of the models revealed significant obstacles in skin lesion image classification.

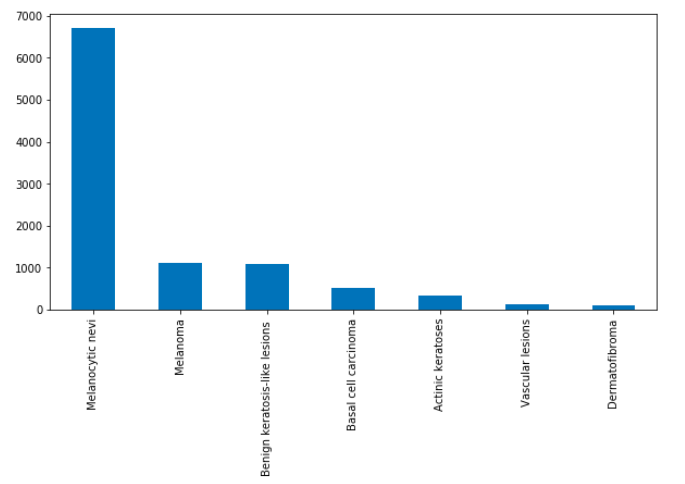
Attempts to address overfitting in the VGG16 model was accomplished by adding a Dropout layer with a fraction rate of 0.2. This reduced the validation loss drastically which successfully prevented overfitting during training. But due to the imbalanced nature of the dataset, overfitting across the minority classes still occurred (see **Figure 5.1**). The solution for this problem was resampling the dataset. The downsampling method was specifically chosen to accomplish this because one majority class accounted for 66% of all the images in the dataset. Oversampling would create too many duplicates across the minority classes, so manipulating a single class was the more reasonable direction. Downsampling resulted in removing a sizeable portion of the dataset which did cause underfitting issues. However, adding class weights to each of the models improved the results. This was calculated by simply getting the inverse percentage of each class count and passing the resulting array into each of the models. Despite these modifications, the results of the ensemble are unfavorable.

Figure 5.

The results of each of the model evaluations are as follows:

A picture containing screenshot

Description automatically generatedA picture containing screenshot

Description automatically generatedVGG16 Training Results

Figure 5.2

**A screenshot of a social media post

Description automatically generatedA screenshot of a social media post

Description automatically generatedResNet50** **Training** **Results**

Figure 5.3

# Both models had a few callback methods passed in as arguments. Checkpoint (saves the progress to allow training to resume at a later point), earlyStopping (stops the training process if a decreasing trend is detected in accuracy or increasing trend in loss), and the learning rate reduction method. Combined, these callback methods help get the best possible results from the training process and contributes towards reducing overfitting. Adding class weights increased the accuracy of the VGG16 by 12.7 points and reduced the loss by 14.8 points and resulted in an adequately converged classifier. The ResNet50 model, however, still did not improve it validation loss. As it is an ensemble in which the average of the outputs of the models is the output, any attempts to further optimizing the model would likely yield negligible improvements.

The general trend of the training progress shows a sufficient convergence across both classification models. The results of the U-Net segmentation models yielded better outcomes:

~~~~~~~~~~~~~~~Stats after 100 epoch~~~~~~~~~~~~~~~~~~~

-------------On Train Set--------------------------

315/315 [==============================] - 4s 12ms/step

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

IOU: | 96.97 |

Dice Coef: | 84.72 |

Precision: | 96.96 |

Recall: | 96.50 |

Accuracy: | 97.94 |

Loss: | 11.75 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

-------------On Test Set--------------------------

50/50 [==============================] - 1s 17ms/step

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

IOU: | 93.41 |

Dice Coef: | 79.02 |

Precision: | 90.64 |

Recall: | 89.27 |

Accuracy: | 93.86 |

Loss: | 19.56 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

-------------On validation Set---------------------

135/135 [==============================] - 1s 7ms/step

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

IOU: | 94.64 |

Dice Coef: | 81.06 |

Precision: | 92.38 |

Recall: | 92.30 |

Accuracy: | 95.11 |

Loss: | 17.03 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A screenshot of a cell phone

Description automatically generated This was a direct implementation of the architecture proposed by Codella et al. [7], but instead trained with the ISIC 2018 Skin Lesion dataset. There was no need for any modifications as it already performed optimally with a 95.11% accuracy on the validation set and 17.03% loss.

The confusion matrix above reflects the class imbalance in the dataset. Unsurprisingly, the Melanocytic Nevi majority class is the most accurately predicted class due to its overwhelming presence in the dataset even after apply downsampling. The bar graph below represents the percentages of incorrect predictions derived from the computations of the confusion matrix. It reaffirms the Melanocytic Nevi class is the most accurately predicted class. Unfortunately, it also indicates that the remaining classes are only predicted correctly between 12% and 57% of the time.

# Chapter 6

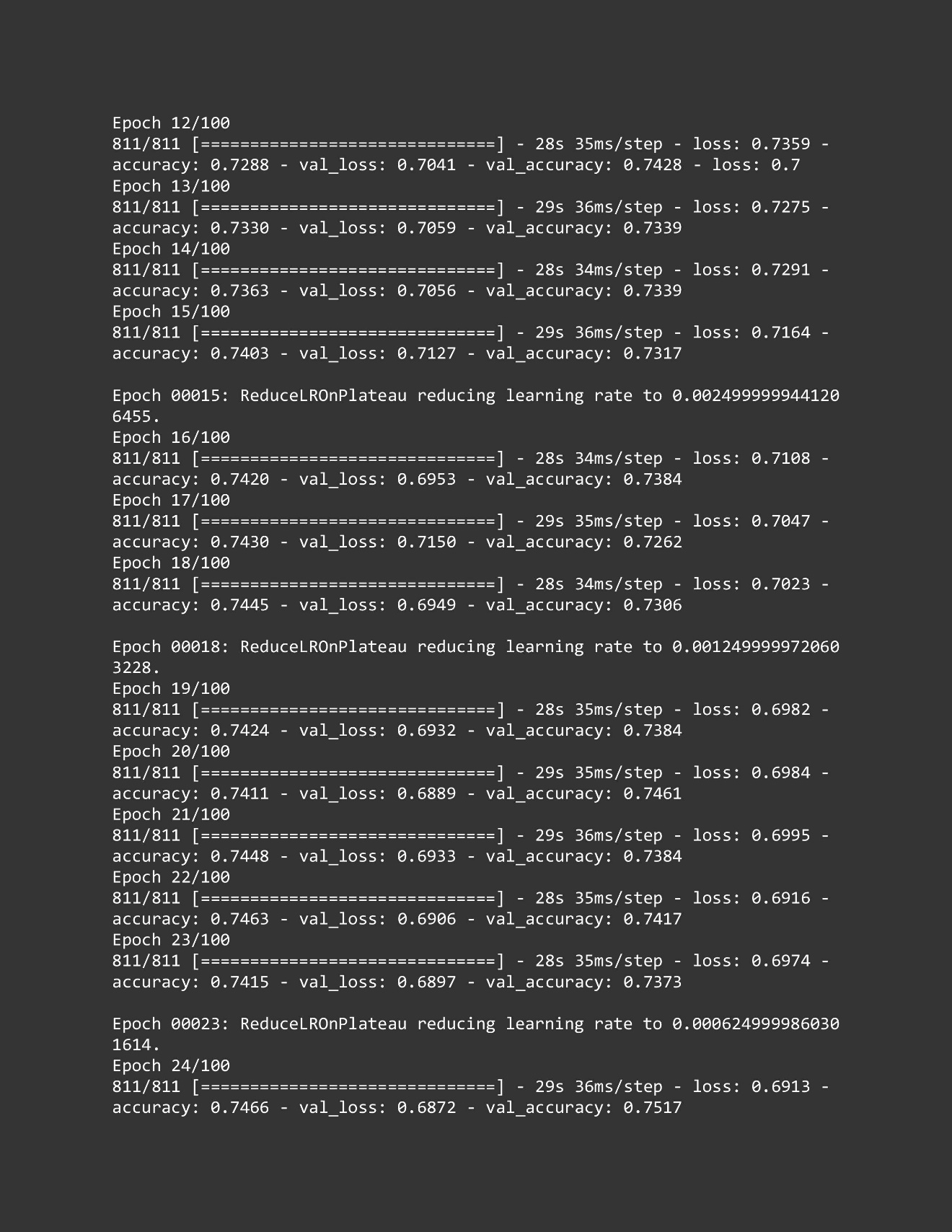
**CONCLUSION**

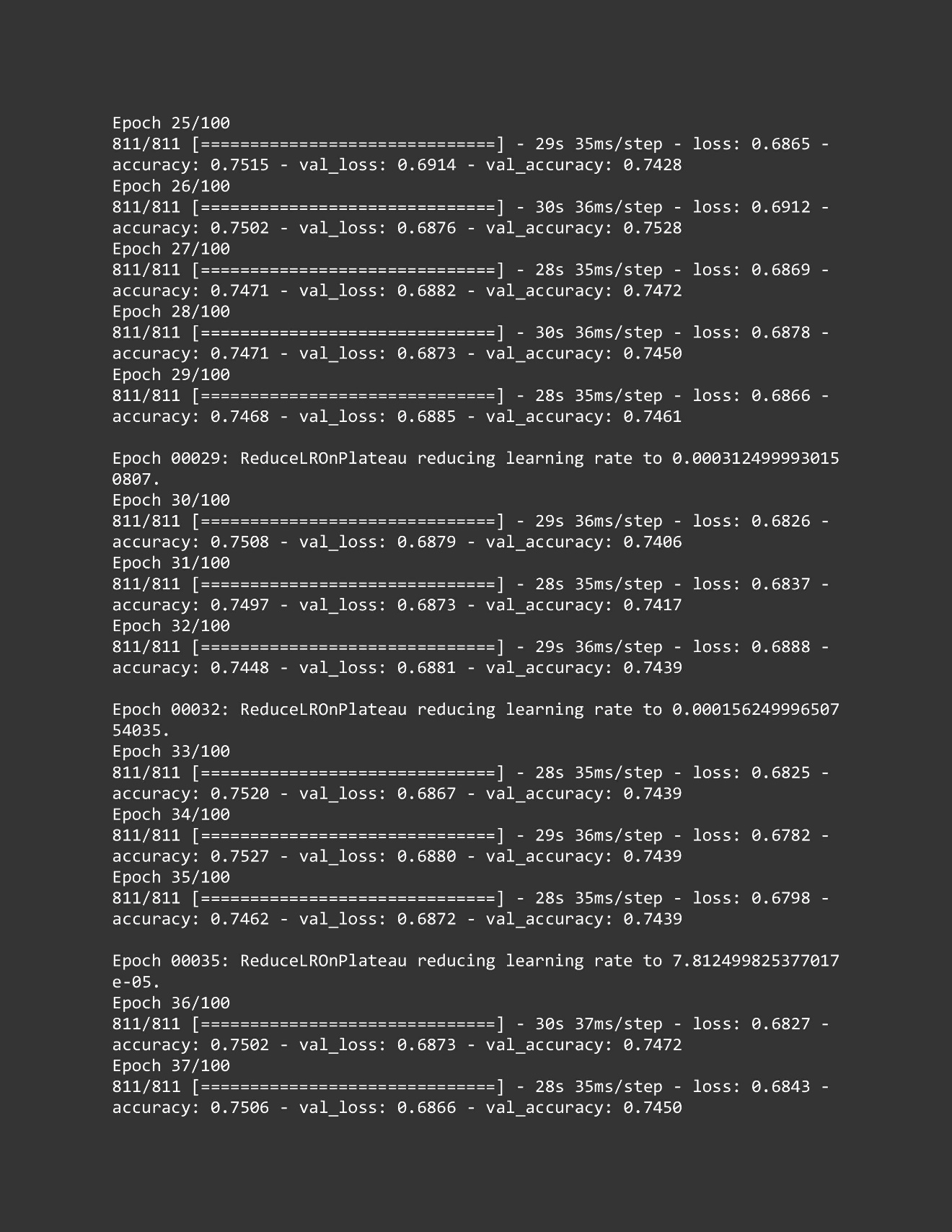
In this paper, the proponents proposed a model that segments a dermoscopic image by isolating and comparing results from both unaltered and U-Net segmented. The results showed poor performance since VGG16 and ResNet50 models did not converge well, this may be caused by the imbalanced dataset. The images had to be resized in order to be compatible with the U-Net model. Callback methods during the training process, resampling the dataset, and data augmentation were used to decrease overfitting. These did effectively reduce overfitting, however, there is still a vast difference on the number of images between the 7 classes. This was the most impactful modification to the model, adding class weights to each of the models. This increased accuracy by 12.7% and reduced loss by 14.8% on the VGG16 model.

That said, the focus of the study was whether segmentation yielded positive affects to the classification process. While the results of the segmentation model were adequate for the purposes of the study, there is plenty of room for improvement for accurately segmenting non-melanoma classes. Despite this, the results did show improvements in the accuracy predictions on the better segmented images. Further research may be able to maintain the model proposed in this research of segmentation followed by classification and produce more favorable results by internalizing the obstacles encountered in this study.

# APPENDICES

# VGG16 Training Results





# 

# 

# 

# 

# ResNet50 Training Results

# 

# 

# 

# 

# 

# 

# 

# 

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

# REFERENCES

[1] Key Statistics for Melanoma Skin Cancer. American Cancer Society. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. Accessed: 2019-04-19.

[2] Dermoscopy for the Family Physician. Encyclopedia of Mental Disorders. https://www.aafp.org/afp/2013/1001/p441.html. Accessed: 2019-04-19.

[3] Philippines Economic Update April 2017: Http://www.workldbank.org/en/news/feature/2017/05/04/philippines-economic-update-april-2017. Accessed: 2019 – 04 – 29.

[4] Mishra, N., Celebi, M. 2016. An Overview of Melanoma Detection in Dermoscopy Images Using Image Processing and Machine Learning. https://arxiv.org/pdf/1601.07843.pdf. Accessed: 2019-04-20.

[5] Sun Protection Tips for Tropical Getaways. https://www.skincancer.org/prevention/tropical. Accessed: 2019-04-29.

[6] Kawahara, J., Hamarneh, G. 2019. Fully Convolutional Neural Networks to Detect Clinical Dermoscopic Features. IEEE Journal of Biomedical and Health Informatics, Vol. 23, No. 2, March 2019

[7] N.C.F. Codella 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),” in IEEE Int. Symp. Biomed. Image., 2018, pp. 168–172.

[8] Codella, N., Nguyen, Q., Pankanti, S., Gutman, D., Helba, B. 2016. Deep learning ensembles for melanoma recognition in dermoscopy images. https://arxiv.org/pdf/1610.04662.pdf. Accessed: 2019-05-14.

[9] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, J. Smith. “Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images”. Proceedings of the 6th International Workshop on Machine Learning in Medical Imaging (MLMI), LNCS 9352, pp. 118-126. 2015.

[10] H. Liao, “A Deep Learning Approach to Universal Skin Disease Classification”. CSC 400 Graduate Project Report. Available: https://www.cs.rochester.edu/u/hliao6/projects/other/skin\_project\_report.pdf

[11] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. In Journal of Biomedical Informatics, 34(1), 2001, 28-36., 2001.

[12] N. Codella, Q.B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern, and J.R. Smith. Deep learning ensembles for melanoma recognition in dermoscopy images. In arXiv preprint arXiv:1610.04662, 2016.

[13] C. Barata, M. Ruela, M. Francisco, T. Mendon¸ca, and J.S. Marques. Two systems for the detection of melanomas in dermoscopy images using texture and color features. In IEEE Systems Journal, 8(3), 2014.

[14] M. Walter. Is this the end? machine learning and 2 other threats to radiology’s future. [online]. http://www.radiologybusiness.com/topics/technology-management/ end-machine-learning-and-2-other-threats-radiology%E2%80%99s-future, 2016.

[15] S. Jha. Will computers replace radiologists? [online]. http://www.medscape.com/ viewarticle/863127, 2016.

[16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431-3440), 2015.

[17] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1520-1528), 2015.

[18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haner. Gradient-based learning applied to document recognition. In Proceedings of the IEEE, 86(11), pp.2278-2324., 1998.

[19] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A.C. Berg. ImageNet large scale visual recognition challenge. In International Journal of Computer Vision, 115(3), 2015.

[20] S.J. Pan and Q. Yang. A survey on transfer learning. In IEEE Transactions on knowledge and data engineering, 22(10), pp.1345-1359., 2010.

[21] CS231n Course. Transfer learning. [online]. http://cs231n.github.io/ transfer-learning/, 2016.

[22] X. Li, B. Aldridge, L. Ballerini, R. Fisher, and J. Rees. Depth data improves skin lesion segmentation. In International Conference on Medical Image Computing and Computer- Assisted Intervention, Sep 2009.

[23] S. Han, M. Kim, W. Lim, G. Park, I. Park, S. Chang. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm.

Journal of Investigative Dermatology, Volume 138, Issue 7, 1529 – 1538

[24] 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*.   Springer, 2013, pp. 63–86.

[25] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman, and N. Petkov, “Med-node: a computer-assisted melanoma diagnosis system using non-dermoscopic images,” *Expert systems with applications*, vol. 42, no. 19, pp. 6578–6585, 2015.

[26] S. Kornblith, J. Shlens, and Q. V. Le, “Do better imagenet models transfer better?” *arXiv preprint arXiv:1805.08974*, 2018.

[27] Wurm EM, Soyer HP. "Scanning for melanoma". Australian Prescriber (33): 150–55. 2010.

1. https://gist.github.com/baraldilorenzo/07d7802847aaad0a35d3 [↑](#footnote-ref-1)
2. https://www.kaggle.com/keras/resnet50 [↑](#footnote-ref-2)