# COMPUTER-AIDED DIAGNOSIS FOR EARLY DETECTION OF MELANOMA BASED ON DEEP-LEARNING TECHNIQUES

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Supervisors

Prof. Hamid GholamHosseini

Prof. Roopak Sinha

27-06-2022

By

Ranpreet Kaur

School of Engineering, Computer and Mathematical Sciences

# Abstract

Melanoma is the deadliest form of skin cancer, with a high mortality rate every year, and New Zealand is known to be one of the countries with the highest incidence of this disease. Overexposure to ultraviolet sun rays causes the upper layers of the skin to produce a pigment known as melanin, the primary cause of melanoma. An early diagnosis and prognosis of melanoma can improve survival rates before it becomes dangerous. There is a high demand for computerized automated skin lesion analysis techniques at low cost since trained specialists are in limited supply, and manual diagnosis involves high costs. The increasing use of non-invasive methods in diagnosing malignant melanoma reduces the need for biopsies.

A primary objective of this thesis is the development of computerized detection systems for analyzing lesions and distinguishing melanoma from other types of skin cancer. The thesis focuses mainly on designing novel methods for three major phases of automatic melanoma diagnosis process: a) Pre-processing deals with removal of noise artefacts such as hairlines and improving image's contrast, b) Lesion segmentation to accurately extract lesion region using deep learning, and c) Melanoma classification using deep learning for a fast and accurate detection.

For the first phase, an algorithm Intensity Adjustment-based Hair Removal (IA-HR) employing morphological operators is developed to remove the hairlines, which are a significant problem in skin image samples. Additionally, a Multi-scale Context Aggregation Convolutional Neural Network (MCACNN) is used to enhance the contrast and

resolution of images. A cleaned dataset is then generated using these pre-processing methods. A class imbalance problem is also addressed using data augmentation methods.

To determine lesion borders and extract lesion information, deep learning networks are used. Two network designs were constructed; one was based on encoder-decoder layered patterns (EDNet) and the other on atrous convolutions (DilatedSkinNet). As DilatedSkinNet shows higher average accuracy, thus it is preferred to EDNet for segmentation tasks.

Finally, in the classification stage, we present the design of a multi-layer deep convolutional neural network (DCNN) named as LCNet to distinguish melanoma from benign tumors. The designed classification network is a lightweight network having a smaller number of learnable parameters. Due to the less complex architecture of classification network, it takes less inference time. Moreover, the network can also diagnose diseases efficiently and accurately without pre-processing and segmentation techniques, unlike traditional machine learning classifiers that rely heavily upon these initial steps.

The study examined, however, the effects of applying designed pre-processing and segmentation methods on segmentation and classification performance. The aim of this effort is to further enhance the performance of the classifier by preparing more rich and clean data for training. The classification network is fed with hairlines-free, high-contrast, and segmented images, and its performance is compared to raw images. Our experiments showed that the classification network efficiently processes raw and complex data by offering an accuracy (ACC) of $90.92 \pm 1.0\%$. The accuracy performance of classification model is improved with pre-processed data as $92.47\%$ and with pre-processed+segmented data as $93.40\%$ indicating the classification model's performed higher with pre-processing and segmentation operations to distinguish melanoma vs benign. Additionally, it is observed that noise-free and cleaned data using IA-HR and

MCACNN methods improved the performance of the segmentation approach.

In our study, we found that the proposed pre-processing and segmentation methods could improve the performance of deep learning-based classifiers. Furthermore, there are many areas of melanoma diagnosis process where the proposed approaches can be successfully applied. The denoising method can be used to clean skin samples without causing any discomfort to the patients such as hairlines can be automatically eliminated from images and contrast can be enhanced. Another use of segmentation model to extract lesion region and to perform detailed analysis of it. Additionally, segmentation method can be employed to generate accurate and smooth ground truth labels for new samples that are currently annotated manually by experts. The classification model can be used to classify melanoma in less time such as in 1.3 seconds as predicted by our model. This method may also be used to generate a handcrafted feature extraction process if a machine learning-based classifier is employed to diagnose skin cancer. The DCNN classification model also showed more improvement in diagnosing melanoma when trained on a large, balanced and pre-processed dataset. In its future scope, embedded systems such as FPGA based system-n-chip and other resource-constrained implementations can benefit from the designed classification network.

# Contents

# List of Tables

# List of Figures

# Publications

### *Journal Publications*

R. Kaur, H. GholamHosseini, R. Sinha & M. Lindén, "Automatic Lesion Segmentation Using Dilated Convolutional Deep Neural Networks in Dermoscopic Skin Cancer Images", *BMC Medical Imaging*, vol.22,no.1,pp.1-13,2022. [Online]. Available: https://doi.org/10.1186/s12880-022-00829-y.

R. Kaur, H. GholamHosseini & R. Sinha, "Skin Lesion Segmentation using an Improved Framework of Encoder-Decoder Based Convolutional Neural Network", *International Journal of Imaging Systems and Technology, Wiley Online Library*, 2022. [Online]. Available: https://doi.org/10.1002/ima.22699.

R. Kaur, H. GholamHosseini, R. Sinha, & M. Lindén, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images", *Sensors*, vol.22, no.3, pp.1134, 2022. [Online]. Available: https://doi.org/10.3390/s22031134.

R. Kaur, H. GholamHosseini & R. Sinha, "Hairlines Removal and Low Contrast Enhancement of Melanoma Skin Images Using Convolutional Neural Network with Aggregation of Contextual Information", *Biomedical Signal Processing and Control*, vol.76, pp.103653, 2022.
[Online]. Available: https://doi.org/10.1016/j.bspc.2022.103653.

### *Conferences Publications*

R. Kaur, H. GholamHosseini & R. Sinha, "Deep Convolutional Neural Network for Melanoma Detection using Dermoscopy Images", *In* $42^{nd}$ *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, Canada*, IEEE, pp.1524-1527, 20-24 July, 2020.
[Online]. Available: 10.1109/EMBC44109.2020.9175391.

R. Kaur, H. GholamHosseini & R. Sinha, "Skin Cancer Classification Utilizing Atrous Convolutions in a Fine-Tuned Network", *In* $25^{th}$ *International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPCV'21), Las Vegas, Nevada, USA*, Springer, 26-29 July, 2021.

R. Kaur, H. GholamHosseini & R. Sinha, "Lesion Border Detection of Skin Cancer Images Using Deep Fully Convolutional Neural Network with Customized Weights", *In* $43^{rd}$ *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico*, IEEE, pp. 3035-3038, 1-5 November, 2021. [Online]. Available: 10.1109/EMBC46164.2021.9630512.

R. Kaur, H. GholamHosseini & R. Sinha, "Synthetic Images Generation Using Conditional Generative Adversarial Network for Skin Cancer Classification", *2021 IEEE Region 10 Conference (TENCON), Auckland, New Zealand*, IEEE, pp.381-386, 07-10 December, 2021. [Online]. Available: 10.1109/TENCON54134.2021.9707291.

***Book Chapters***

R. Kaur, H. GholamHosseini & R. Sinha, "Deep Learning in Medical Applications: Lesion Segmentation in Skin Cancer Images Using Modified and Improved Encoder-Decoder Architecture", *International Symposium on Geometry and Vision*, vol. 1386, pp. 39-52, Springer, Cham, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-72073-5-4.

***Research Symposiums***

R. Kaur, H. GholamHosseini & R. Sinha, "How Machine Learning Can Help with Early Detection of Melanoma?", *Queenstown Research Week (QRW), Queenstown, New Zealand*, $30^{th}$ August 2022.

R. Kaur, "From Machine Learning to Deep Learning: Experimental Comparison of Machine Learning and Deep Learning for Skin Cancer Image Segmentation", *AUT Postgraduate Research Symposium, Auckland, New Zealand*, $26^{th}$ November, 2021.

# Acknowledgements

I would like to thank the following people, without whom I would not have been able to complete this research and without whom I would not have made it through my Ph.D. study.

I would like to express my sincere gratitude to my primary supervisor, Professor Hamid GholamHosseini, for providing continuous and valuable support throughout my Ph.D. study. His support, encouragement, and motivation would have made this journey possible. He helped me develop innovative ideas for progressing in my research and writing research papers and thesis reports. For me, having the opportunity to work with him has been an enormous blessing.

I would also like to express my deepest thanks to Professor Roopak Sinha, my secondary supervisor, for his continuous support, motivation, and helpful suggestions throughout this research process. With his guidance, I was able to improve my research activities significantly. I owe him my heartfelt appreciation.

Further, I would like to thank AUT University for providing resources for executing research experiments and offering me financial support through teaching employment, summer research scholarships, and funding research papers.

Last but not least, thanks go to my beloved husband and son, Narinderjeet and Sidakveer, for their love, support, and patience throughout my work, who never left me in my hard times and kept me motivated. Furthermore, I would like to thank my parents, brother, and his family for supporting me emotionally and financially in New Zealand throughout my Ph.D. research. I would also like to thank all my friends and colleagues who helped me through this long journey of study and my life in general.

# Dedication

*I dedicate this study with extreme gratitude to my beloved family, Dr.Hamid GholamHosseini, Dr.Roopak Sinha, and all of my friends and colleagues without whom none of my success would be possible.*

# Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning.

_____

Signature of candidate

Date: 27-06-2022

# List of Acronyms

| | |
|---|---|
| ANN | Artificial Neural Network |
| ACC | Accuracy |
| AK | Actinic Keratosis |
| ADAM | Adaptive Moment Estimation |
| ANOVA | Analysis of Variance |
| Bal ACC | Balanced Accuracy |
| BEN | Benign |
| BCC | Basal Cell Carcinoma |
| BKL | Benign Keratosis |
| BRISQUE | Blind Image Spatial Quality Evaluator |
| BN | Batch Normalization |
| CNN | Convolutional Neural Network |
| CONV | Convolutional |
| CAD | Computer-aided diagnosis |
| CLAHE | Contrast Limited Adaptive Histogram Equalization |
| DCNN | Deep Convolutional Neural Network |
| DF | Dermatofibroma |
| DECONV | Deconvolutional |
| DNN | Deep Neural Network |
| EDNet | Encoder-Decoder Network |
| FC | Fully Connected |
| FCN | Fully Convolutional Network |
| FCNN | Fully Convolutional Neural Network |
| FN | False Negative |
| FP | False Positive |
| FPR | False Positive Rate |
| FA | Firefly Algorithm |
| GMM | Gaussian Mixture Model |
| GPU | Graphics Processing Unit |
| HOG | Histogram of Gradients |
| IOU | Intersection Over Union |
| ISIC | International Skin Imaging Collaboration |
| IA-HR | Intensity Adjustment-based Hair Removal |
| JAC | Jaccard |

| | |
|---|---|
| KNN | K Nearest Neighbour |
| LEM | Lentigo Maligna |
| LCNet | Lesion Classification Network |
| MCACNN | Multi-scale Context Aggregation Convolutional Neural Network |
| MRAB | Multiscale Residual Aggregation Blocks |
| MEL | Melanoma |
| MSE | Mean Square Error |
| NAIT | National Animal Identification and Tracing |
| NV | Nevus |
| NIQE | Natural Image Quality Evaluator |
| PET | Positron Emission Tomography |
| PSO | Particle Swarm Optimization |
| PH | Pedro Hispano |
| PSNR | Peak Signal-to-Noise Ratio |
| PReLU | Piecewise Rectified Linear Unit |
| POOL | Pooling |
| PRE | Precision |
| ReLU | Rectified Linear Unit |
| RGB | Red, Green and Blue |
| RMS(E) | Root Mean Square (Error) |
| ROI | Region of Interest |
| REC | Recall |
| SSIM | Structural Similarity Index |
| SGD(M) | Stochastic Gradient Descent (Momentum) |
| SVM | Support Vector Machine |
| SEK | seborrheic keratosis |
| SCC | Squamous Cell Carcinoma |
| SGO | Social Group Optimization |
| SPE | Specificity |
| SEN | Sensitivity |
| TN | True Negative |
| TP | True Positive |
| UV | Ultra-Violet |
| UNK | Unknown |
| WHO | World Health Organization |

# Chapter 1

# Introduction

Skin cancer is an abnormal growth of cells in the outermost layer of the skin, known as the epidermis. These cancerous cells invade surrounding tissues and form a malignant tumor. The primary cause of skin cancer is direct exposure to UV radiation, and the elderly are more prone to developing cancer than other age groups. There are many risk factors associated with melanoma, including fair skin, overexposure to the sun, sunburn, genetic predisposition, and weakened immune system [2, 3]. Early detection can reduce the mortality rate [4] to a great extent. Because it is highly treatable if diagnosed at an early stage, thus it is significant to aim for a diagnostic system that detects melanoma on time to reduce mortality, morbidity, and cost involved during the clinical diagnosis process.

## 1.1 Study Overview and Motivation

Melanoma is prevailing all over the globe causing 132,000 deaths each year estimated by WHO [5]. The American cancer research institute reported that melanoma is the $5^{th}$ most common cancer form in men and women [6]. There are two major types of skin cancer: melanoma and non-melanoma. The malignant form of skin cancer

is melanoma, a severe disease because of a high level of metastasis, thus lethal. On the other hand, non-melanoma types include various other forms of cancer such as nevus (NV), dermatofibroma (DF), seborrheic keratosis (SEK), Lentigo maligna (LEM), actinic keratosis (AK), basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), where BCC and SCC are the commonly found cancer types. These two forms are rarely metastasized; thus, they are less lethal.

Several factors motivated this research:

- The first reason is the increasing incidence of skin cancer across the globe, making it the cause of highest number of deaths each year [7]. The World Health Organization (WHO) reported that melanoma and non-melanoma skin cancer rank among the top causes of death, accounting for approximately 10 million deaths globally in 2020 [8]. In 2017, the American cancer society stated that in the USA, there were 1,688,780 new cancer cases diagnosed and 600,920 cancer deaths occurred [9]. There were 1,735,350 new cancer cases diagnosed and 609,640 deaths occurred in 2018 [10], 1,762,450 new cancer cases and 606,880 cancer deaths in 2019 [11], and 1,806,590 new cancer cases found and 606,520 cancer deaths happened in 2020 [12]. The next most affected countries are Australia, and New Zealand, where skin cancer cases keep on escalating yearly [13]. It was anticipated in 2020 that there be 16221 cases in Australia and 2500 in New Zealand [14]. Therefore, investing research efforts in generating skin cancer detection techniques is the primary concern for the research community.

- Another reason is the level of difficulty medical experts face in identifying melanoma due to its complex nature [15, 16]. Several factors contribute towards this complexity: (a) irregular border (b) different shapes (c) intensity variations (d) color illuminations (e) evolution over time. Visual examination of different skin lesions can be difficult due to similarities between them that can lead to

incorrect identification. Another major challenge in skin cancer diagnosis is the similarity between the different types of cancer; for example, NV skin cancer type is most commonly misunderstood as melanoma because it shares a similar lesion formation along with color, texture, and diameter. Hence, it is vital to distinguish several cancer forms from each other, which is an arduous task to consider by experts manually.

- Furthermore, skin diagnosis is a very labor-intensive and slow process in the current clinical settings. The trained specialists such as dermatologists commonly diagnose melanoma from dermoscopic lesion images based on ABCDE (Asymmetry, Border, Color, Diameter, and Evolution) metrics. Lesion moles are asymmetrical as half of the mole does not match the other half, thus giving asymmetry. Border features represent irregular, ragged, and blurred edges that are difficult to examine through the naked eye. Color information is also not evenly distributed over the lesion area, having black, grey, pink, and white shades. Another feature is diameter which is essential to track for differentiating types of cancer such as melanoma have diameter above 6mm. Lastly, evolution is another significant feature that stands for E in the ABCDE rule. Despite the importance of all factors in the diagnosis process, these may take time from months to years to become available to the experts. Because lesions take time to evolve by increasing their size and changing color and shape, the patient has to visit the clinic many times for regular checkups. Every time manually keeping a record of lesion information is very labor-intensive. Also, during that time, the effects of skin cancer may become worse, and it is too late to cure it.

- Another factor is that the current clinical diagnosis process is not very accurate. To diagnose skin cancer, dermatologists screen the infected skin area and visually examine skin cells formation using biopsy and histopathology. They also consider

patients' clinical information such as age, location site, diameter, evolved, and bleed. These features help them understand the nature of the lesion; however, it remains a challenging task. To increase the reliability of the diagnosis process, they use dermatoscopes, an imaging technique that magnifies the lesion to have a clear picture of the skin surface and its structures. This procedure can only be used by trained specialists who are in limited supply. Moreover, an expert can achieve up to 65% to 75% accuracy using dermoscopy [17]. There are variations in accuracy depending upon the expert's visual acuity and knowledge [10] because the similarity between different types of skin lesions makes the visual examination process very hard and can lead to wrong identification results.

- Finally, lesion diagnosis is an expensive procedure. It is estimated that the average cost involved in the USA for skin cancer treatment is approximately $44.9 million for existing patients and $932.5 million for newly diagnosed patients from all age groups [18, 19]. As per the New Zealand census, $28.6 million costs are involved in cancer treatment [20], thus making it prohibitively expensive. There is no surprise that melanoma has become a critical public health concern.

The reasons listed above support the need for a computer-aided diagnosis (CAD) system to assist dermatologists in the early detection of melanoma. By monitoring the disease as efficiently as possible, CAD can help increase the survival rate since patients have a high chance of recovering. These automated tools are user-friendly and can be used by non-dermatologists also. The primary application of CAD tools is in clinical sites to assist medical experts in routine checkups by having a quick and economic assessment. This research can revolutionize the medical industry by eliminating or optimizing currently-used manual and time-consuming procedures for diagnosing melanoma. Furthermore, it is the fastest way to monitor infected skin areas and their evolution.

In developing a CAD system, the research community focused on introducing machine learning and image processing techniques for the classification of melanoma. Machine learning offers statistical algorithms for classification tasks that rely on two main steps: training and testing on data. Later in 2016, traditional machine learning algorithms had workflow consisting of pre-processing, feature extraction, and classification. Few studies such as [21, 22] presented the use of hand-crafted features and others also proposed feature calculation based on the 7-point checklist [23, 24] to train computer vision algorithms like support vector machines (SVM) [25, 26], K-nearest neighbour (KNN) [27, 28] and artificial neural networks (ANN) [29] for classifying skin cancer. There were some drawbacks in using these traditional methods; first, these techniques are not efficient in handling complex and large amounts of data; secondly, hand-crafted features do not have generalization capability to distinguish the similar shape of lesion structures [30]. Lastly, the 7-point checklist or ABCDE criteria are suitable only for pigmented lesions, which means they can not classify BCC or SCC non-pigmented lesions. Moreover, a required level of expertise is required while segmenting lesion regions and extracting optimal features. The use of poor segmentation techniques and wrong selection of features can lead towards lower classification results [31].

Recently introduced deep learning approaches offered numerous advantages towards automatic and accurate identification. In particular, a convolutional neural network (CNN) is the preferable choice to apply for the classification task. Therefore, developing new and efficient approaches can lead towards achieving higher accuracy and fast melanoma detection. Further, there are rapid advancements in embedded system applications, which can create portable devices for medical analysis by employing deep learning-based efficient detection techniques.

## 1.2 Research Problems/Questions

RQ1 **Whether to use dermoscopic or clinical images?**

For the accurate development of the CAD system, the first step involves the collection of images. For skin cancer data, there are two choices between types of images to use. The first is dermoscopic images acquired through a special device called dermatoscopes. Other types of skin cancer images are clinical images. The main difference between these two is in screen visualization. Clinical images are captured using high-resolution cameras under appropriate distance and lighting conditions. These images show skin lesions on the skin surface very clearly and naturally, similar to what the dermatologist observes with the naked eye. In contrast, dermoscopic images are sometimes low resolution and have changed color. Figure 1.1 shows two types of images that were captured by a standard camera and dermatoscopes. We used dermoscopic images for this research because large amounts of data are required to train deep learning networks for generalization. Thus, dermoscopic images are available from benchmark data sets, but clinical images are relatively scarce.

RQ2 **Use of Pre-processing Technique: Hair Removal and Contrast Enhancement**

RQ2 (a) *How to identify efficient pre-processing techniques to remove hairlines without losing object information in pictures of skin taken to detect melanoma?*

The main issues that occur during the acquisition process are contamination of various noisy elements such as poor illumination, hairlines, color variations, and ruler marks. Figure 1.2 illustrates some of the

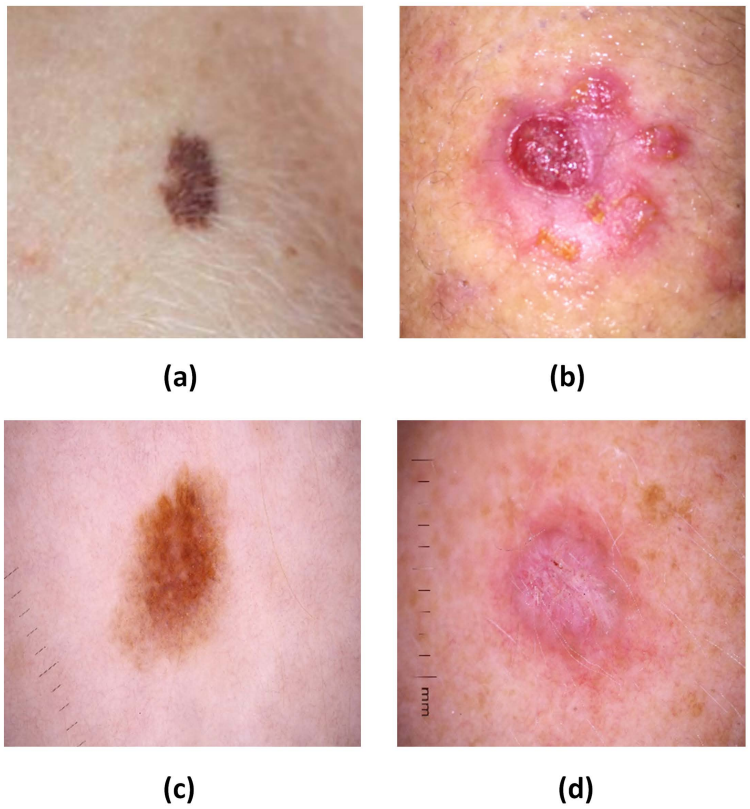



Figure 1.1: Clinical images (top) and dermoscopic images (bottom).

challenging cases that are the outcome of the image acquisition process achieved through dermatoscopes and cameras. One alternative to improve the image acquisition process is high-resolution cameras, magnifying lenses, regulating light sources that can enhance visibility while capturing skin images. However, this technical support contributes only 50% towards achieving accuracy [32]. Alternatively, the development of suitable image pre-processing techniques to eliminate these artefacts to increase the accuracy of the system can be adopted. The given images from benchmark datasets were collected in uncontrolled conditions and unknown environments; thus, image processing techniques are considered for pre-processing these images to remove hairlines and enhance their contrast using deep learning framework.

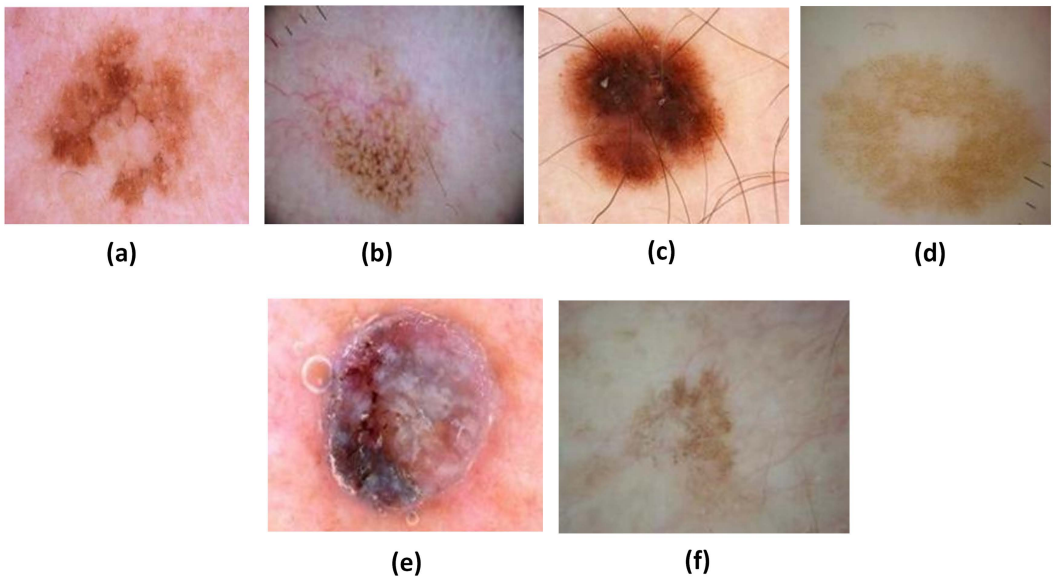



Figure 1.2: Examples of noise artifacts (a) irregular boundaries (b) blood vessels (c) hairlines (d) color illumination (e) bubbles (f) low contrast.

RQ3 **Lesion Segmentation**

RQ3 (a) *To what extent can lesion segmentation algorithms based on machine learning and deep learning techniques be effective in terms of accuracy?*

An automatic segmentation approach is required to assist dermatologists in understanding the nature and pattern of the lesion. The structural features such as its shape, area, and diameter are the main distinguishing point to consider while identifying melanoma. Moreover, this method is significant for automatically generating the ground truth images which were previously annotated manually by dermatologists. Dermatologists commonly diagnose melanoma from dermoscopic lesion images based on ABCDE (Asymmetry, Border, Color, Diameter, and Evolution) characteristics, which is time-consuming if performed manually. Therefore, the lesion segmentation method helps to quickly calculate statistical

values of these metrics to analyze the type of cancer. Implementing such an automatic segmentation technique aims to determine whether segmented and cleaned lesion image contributes more towards achieving higher classification performance. Therefore, for detecting lesion boundaries through software based approach, two main challenges are addressed. First is the selection between traditional machine and deep learning technique to address this task and secondly the development of a novel method.

RQ4 **Melanoma Classification**

RQ4 (a) *What benefits can a novel deep learning-based classifier provide for the early detection of melanoma, especially in terms of classification accuracy, over existing methods?*

The early detection of skin cancer disease can limit its progression to severe levels by providing timely treatment. Design of an efficient classification approach can assist in diagnosing melanoma with more accuracy. Various machine learning and pre-trained deep learning networks are present to address classification tasks. However, their suitability on skin cancer data is a significant concern because these techniques are data-dependent. Therefore, the aim is to propose a novel classification model to achieve an acceptable high classification accuracy for recognizing cancer at an early stage facing all challenging trade-offs between high accuracy rate and optimal design of a network.

RQ5 **Improved Classification**

RQ5 (a) *To what extent does pre-processing of the images, such as the removal of hairlines as in RQ2 and segmentation as in RQ3, improve or affect*

*the performance of the classifier developed in response to RQ4?*

The standard computer-aided systems have pre-processing, segmentation, feature extraction, and classification phases as necessary steps to do classification. The advantage of deep learning approaches is that they do not require these phases explicitly. The deep learning approaches efficiently calculate features and train the network to perform the classification task. However, this research also investigates the impact of using methods discovered in RQ2 and RQ3 on classification performance. To achieve this, a classifier is trained with pre-processed and raw images differently, and their performance is compared based on statistical metrics.

## 1.3 Contributions and Research Outcomes

### 1.3.1 Data preparation

For the given research work, the publicly available dataset consisted of dermoscopic skin cancer images of different types of cancer is used. The dataset have lesion samples of various type of cancer collected from more than 2000 patients. The availability of data was hampered by two major issues. In the first place, the datasets are highly imbalanced, with different distributions of samples in each class, and this imbalance affects the performance of the classifier. Therefore, a data oversampling technique is implemented to address this problem by generating synthetic images and balancing data distribution. Secondly, there were many duplicate samples present, resulting in overfitting and bias in the network. A duplicate removal method was applied to solve this problem.

### 1.3.2 Pre-processing algorithm

To handle noisy images, specifically hairy images, the intensity-based illumination correction algorithm is designed for the removal of thin and light hairlines that obstruct the accurate extraction of the lesion object. The developed algorithm is a novel method based on morphological operations to remove hairlines. The contrast and resolution of images are degraded after hair removal, which is not good for classification as color and texture are important when diagnosing lesions. Thus, the images were enhanced using a context aggregation-based deep learning network.

### 1.3.3 Lesion border detection

Extracting region of interest (ROI), i.e., lesion from images, is a challenging computer vision problem due to the varying nature of skin cancer images. Two types of methodologies were designed known as lesion segmentation techniques; the first is a hybrid design of soft computing and machine learning, and the second is based on deep learning. Further, as part of deep learning frameworks, two novel designs were developed to perform lesion segmentation and compare their performance. The primary aim is to achieve accurate regions with high performance. These methodologies were designed from scratch and compared to select the best one for cascading it in the classification phase.

### 1.3.4 Classification

After detecting ROI, the next step of CAD is a design of a classifier. A novel DCNN classifier is designed focusing on the unique design of network's layers, selection of suitable hyper-parameters, and use of training algorithm for meeting high accuracy and low execution time. Additionally, the performance of a classifier is analyzed using raw and pre-processed images.

## 1.4 Thesis Structure

The thesis is organized into seven main chapters following the introduction chapter:

- **Chapter 2:** presents background information about various computer vision and image processing techniques proposed related to skin cancer and melanoma. Specifically, it provides a literature review of several CAD systems including, pre-processing, segmentation, and classification task.

- **Chapter 3:** provides details of the proposed system design including the datasets used for each phase of the system such as pre-processing, segmentation and classification. It also discussed the data augmentation methods applied to the data during experimental evaluation.

- **Chapter 4:** deals with the details of the pre-processing method introduced as the initial step of the classification system. It also provides a systematic description of the technique proposed to eliminate hairlines and improve contrast from noisy images.

- **Chapter 5:** explains the design of two CNNs proposed for performing lesion segmentation, including results and discussion. It also describes the results of a few pre-trained networks used for comparison.

- **Chapter 6:** deals with the method proposed for melanoma classification. It explains the detailed design of the deep learning network used to classify melanoma among other skin cancer types. An explanation of other existing networks that were experimented on the adopted dataset is given for comparing the system's performance based on evaluation metrics. Additionally, we investigated the use of adding pre-processing and segmentation steps before classification to analyze their impact on the performance of a classifier. This chapter compares the results

using pre-processed images (using hair removal algorithm and segmentation) and raw images (without pre-processing steps).

- **Chapter 7:** includes overall conclusion and discussion of achieved results including future scope.

# Chapter 2

# Background and Literature Review

## 2.1 Introduction

This chapter aims to provide an introduction to melanoma and review the computer vision approaches subject to image pre-processing, lesion segmentation, and melanoma classification among various skin cancer types. Firstly, a brief introduction about the problem domain is explained, i.e., structure of the skin, nature of different kinds of cancer, key points to distinguish them from each other, and currently used clinical procedures. Secondly, the methodology used for performing literature studies on each phase of the CAD system is explained. Finally, traditional machine learning and recent deep learning techniques applied to hair removal, segmentation, and the development of cancer detection systems are discussed. The advantages and disadvantages of techniques proposed in the literature for each phase of CAD system are explained followed by the literature gaps found by our research analysis. Through our research work, the efforts were made to fill those gaps by proposing novel and efficient techniques for each phase of melanoma CAD system targeting higher performance and low inference time.

## 2.2 Background Information on Melanoma

### 2.2.1 Melanoma's structure and nature

There are three major layers of skin on the human body (Figure 2.1) [1, 33], epidermis, dermis, and subcutaneous fatty tissue. The epidermis layer consists of two types of cells, i.e., squamous and melanocytes cells. The proliferation of melanocytes and the overgrowth of fibrous tissues in the deeper layers of the skin are the main causes of skin cancer. The cancerous cells develop in the outermost epidermis skin layer; thus, it is visible. The melanocytes produce melanin in the body, responsible for producing skin color. Sometimes these cells cluster to form moles that may develop to malignant later and become melanoma.

Skin cancer is more likely to occur if the skin is fair, excessive sun exposure, sunburned, have a weak immune system, and have a family history of the disease [34, 35]. Skin cancer may occur in many different shapes, sizes, and colors, and in many cases, it is indistinguishable from a normal skin mole. Therefore, it isn't easy to recognize it with the naked eye.

### 2.2.2 Global prevalence

The incidence of melanoma has been increasing across the globe, and the disease accounts for 15.0% of all deaths in New Zealand, according to statistics calculated in 2018 [36]. According to a recent study conducted by the cancer society, melanoma is the $19^{th}$ most commonly found cancer problem worldwide, with New Zealand having the highest rate [34]. Table 2.1 describes the recent data about melanoma cases reported in countries such as America, New Zealand and Australia in 2018, 2019 and 2020 [34, 37, 38, 39].

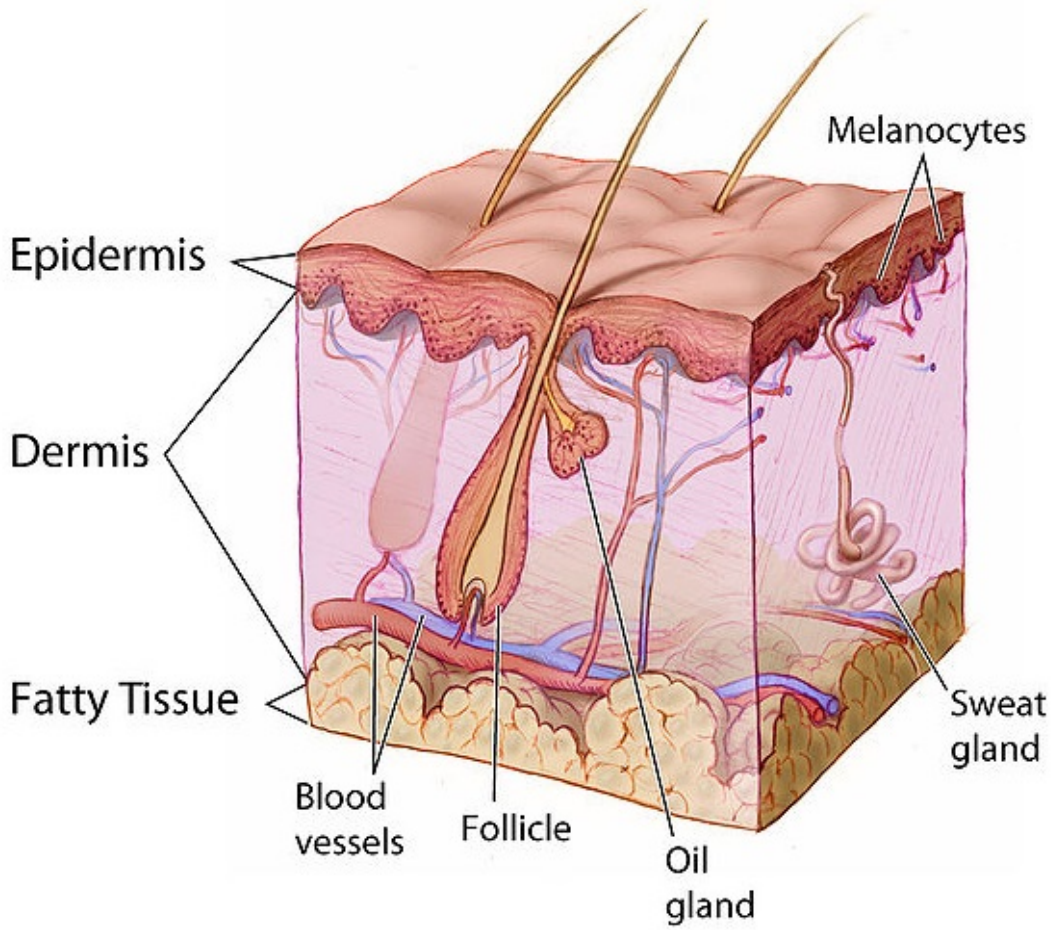



Figure 2.1: Skin structure [1].

Table 2.1: Number of melanoma cases reported in America, Australia, and New Zealand in 2018, 2019, 2020 and 2021.

| **Country** | **No. of cases 2018** | **No. of cases 2019** | **No. of cases 2020** | **No. of cases 2021** |
|---|---|---|---|---|
| USA | 609,640 | 606,880 | 606,520 | 608,570 |
| Australia | 34,200 | 145,000 | 16,221 | 16,878 |
| New Zealand | 2,473 | 2,727 | 2,500 | — |

### 2.2.3 Types of cancer

There are seven different types of skin cancer [40] as given in Figure 2.2. The first type is melanoma which is considered malignant whereas, other cancer types fall into

a benign category. Among the different types of skin cancer, melanomas are more aggressive and less common. It can spread quickly to other parts of the body, making it the most dangerous. It appears in brown with dark color speckles. The diameter of melanoma's lesion is 6mm and contains irregular borders such as ragged edges, uneven shapes, and blurry. The second category is benign further having several variations. For example, melanocytic nevus (NV) contains nevus cells that occur due to the proliferation of melanocytes cells in the body. These can occur on any part of the body, and its shape differs depending upon the body site. They appear pink or brown, and their diameter varies from millimeters to centimeters. Then basal cell carcinomas (BCC), also known as non-melanoma, are a common type of skin cancer. The younger people are more prone to it, and the primary emergence sites are the upper trunk and shoulders. It is difficult to identify due to its irregular shape. Another category is keratosis cancers such as actinic (AK), squamous cell carcinoma (SCC), and benign keratosis (BKL), which occurs due to spending long hours in the sun having a weak immune system. The clinical properties of this type involve a flat or thickened appearance and brown or white color. These types of cancer can be differentiated from each other based on the size of the lesion, where AK tends to be smaller than SCC and BK. The final form of cancer identified in this work is dermatofibroma (DF), formed due to the overgrowth of fibrous tissues in the deeper layer of the skin. It is not as harmful as others, and its appearance is similar to other lesion shapes. The primary goal of this research is to classify melanoma from others.

## 2.3 Clinical Procedures

To determine whether the suspicious skin lesion is melanoma or not, the doctors take a sample of tissue from the skin. The procedure is known as biopsy, specifically called a punch biopsy, in which whole lesion tissue is removed from the skin with a circular

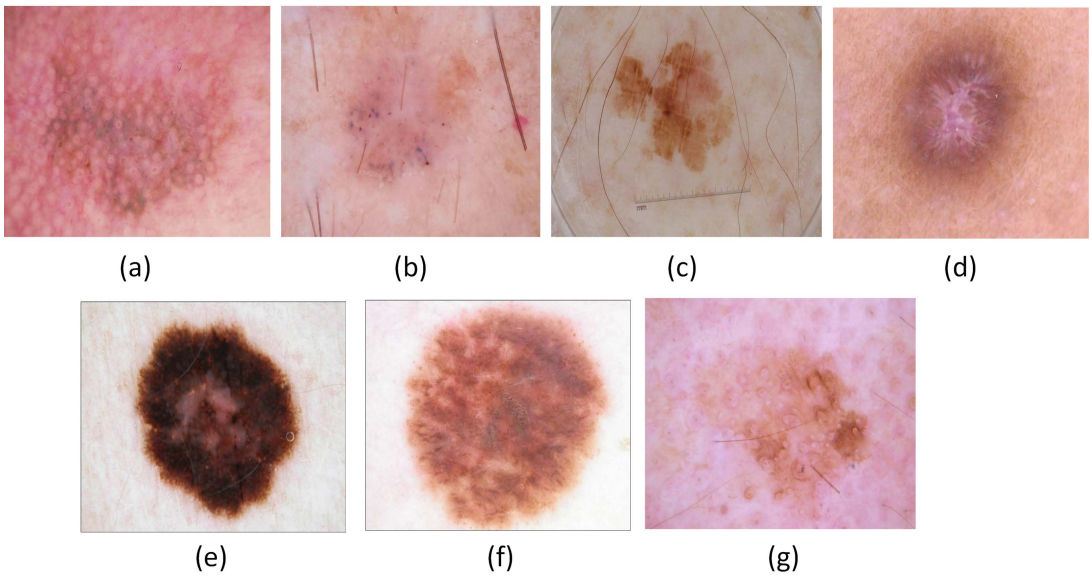



Figure 2.2: Different skin cancer types (a) AK (b) BCC (c) BKL (d) DFF (e) Melanoma (f) NV (g) SCC.

blade. To assign a stage to melanoma, a few further investigations are done, like finding the thickness of the mole and whether it spread to lymph nodes. A CT scan, X-ray, and positron emission tomography (PET) scan may be needed in some cases to perform the advanced search. However, this procedure is excruciating and uncomfortable for the patient. Currently, capturing photographs of infected skin areas through cameras or dermatoscopes is an effective tool in clinical practices and research environments to analyze skin lesion patterns. According to the review conducted in [41, 42] researchers concluded that dermatoscopes are very efficient tools in diagnosing cancerous and non-cancerous skin lesions. It works like a magnifying lens that has the capability of viewing skin surface ten times larger [43]. The light source helps examiners view skin clearly in the presence of light variations. Dermatoscopes are considered accurate tools as compared to the visual examination performed by trained specialists [41]; however, few factors are affecting the accuracy of dermatoscopes. Those factors include unclean skin, dark skin complexion, and using different dermatoscopes. Thus, automatic

computer-aided diagnosis is a more robust and accurate method.

### 2.3.1 Clinical images

Clinical images are close-up images captured from the skin surface using high-grade camera devices from a suitable distance. The lesion context appears large because the image is focused only on the region, excluding surrounding anatomic structures. These images represent the region of interest similar to what is observed by an expert with a naked eye. Clinical images are also referred to as macroscopic images and help keep track of lesion changes over time. The most common sites for photography are faces, ears, palms, toes, and fingers. The advantage of capturing images through customer-graded cameras is that it causes less anxiety and pain to the patient than other screening tools.

### 2.3.2 Dermoscopy images

Dermoscopic images are an output of a dermoscopy procedure in which skin surface is examined in detail. The device used to perform dermoscopy is dermatoscopes which are hand-held devices to view the outer layer of skin in detail which is difficult to observe by the naked eye. This is a non-invasive method for enhancing a doctor's view of the skin by using a magnifying lens and a powerful lighting system. A dermoscope is a lightweight, battery-powered device useful in taking still photography or video. The captured pigmented lesion is evaluated using dermoscopy in terms of color and structure.

### 2.3.3 Clinical metrics to diagnose melanoma

In the early stages of melanoma, it isn't elementary to detect it. The initial examination can be done by checking for symptoms. In clinical settings, dermatologists perform a

visual examination of a skin lesion based on ABCDE criteria or 7-point checklist [44]. The ABCDE examination describes five characteristics and are displayed in Figure 2.3.

- **Asymmetry:** Cancerous mole is asymmetrical and looks different from each side, whereas non-cancerous cell seems to be symmetrical.
- **Border:** Cancer has irregular, ragged, and notched boundaries as compared to non-malignant, which have smooth borders.
- **Color:** Melanoma contains different shades of color such as black, brown, pink tan, or sometimes blue or white pigmentation.
- **Diameter:** The size of the mole keeps on changing with time; thus, a diameter above 6mm might be considered melanoma.
- **Evolution:** The continuous change in the appearance of the lesion over weeks is a sign of melanoma.

The complex lesion patterns and variations in human interpretations introduce inaccuracies, thus not accurately analysed.

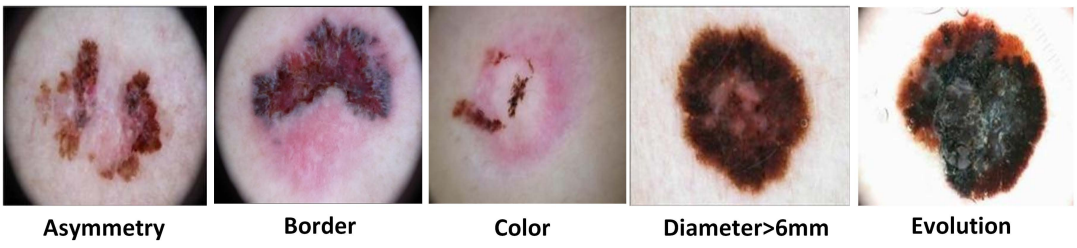



Figure 2.3: ABCDE diagnostic metrics of skin cancer.

## 2.4 Systematic Literature Review

The primary aim of performing literature analysis is to identify the gap and study the existing approaches to skin cancer detection systems. The literature review is conducted

on the following key areas of skin cancer detection system by considering relevant papers:

- pre-processing of images related to hair removal, contrast adjustment, and unwanted artefacts.
- identifying accurate region of interest that involves segmentation of lesion.
- melanoma diagnosis that results in the classification of skin lesion.

The search questions designed to refine the literature review process:

1. What are the image processing-based hair removal methods from skin cancer images?
2. How is the segmentation process performed to extract lesions using machine learning and deep learning approaches?
3. How are deep learning approaches used to detect skin cancer?

### 2.4.1 Search strategy

A total of nine databases were referenced including PubMed, Scopus, IEEE Xplore, Springer, ACM digital library, Elsevier, and Scholar. For the extraction of relevant information, a web search is conducted based on the following parameters:

- Search keywords based on hair removal, lesion segmentation, and cancer classification.
- Words related to search keywords.
- Search by combining one or more keywords.

### 2.4.2 Search keywords

Table 2.2 lists the keywords used to search for information related to each phase of the CAD system.

Table 2.2: Search keywords.

| Primary keyword | Combined set of words |
|---|---|
| Skin cancer | skin cancer types, skin diseases, skin treatment, cancer diseases, skin cancer types, skin cancer treatment, automatic cancer diagnosis |
| Melanoma | melanoma + skin + cancer, melanoma cases, melanoma causes, melanoma symptoms, melanoma treatment |
| Classification | melanoma diagnosis + deep learning, convolutional neural networks, melanoma classification + machine learning |
| Segmentation | melanoma + lesion extraction, segmentation techniques |
| Pre-processing | hair removal + skin cancer images, image processing, noise removal, image enhancement |
| Deep learning | melanoma classification + deep learning, lesion segmentation + deep learning, image enhancement + deep learning |

### 2.4.3 Selection criteria

The research papers considered for review are selected based on specific parameters such as refereed journal papers, the language of articles preferred English only, the year of publication considered between 2011 and 2021, and the paper's topic is focused within the domain. The final study includes a selection of the most relevant research papers based on the three stages of the CAD system, namely pre-processing, segmentation, and classification. Figure 2.4 shows that initially 71,800 records were retrieved from various databases related to skin cancer. Some records were not related to the keywords and research questions. Also, some papers were repeated across multiple databases. Thus, 43,900 duplicate records were removed, and 27,900 records remain to be studied. Next, 19,060 records were filtered that were primarily focused on pre-processing, such as hair-lines removal and enhancement, lesion segmentation, and melanoma classification. In addition, 18,162 papers published in conferences, book chapters, and symposiums were removed, leaving 253 papers for study. Final analysis of 148 records was conducted, consisting of 36 papers on pre-processing, 68 papers on segmentation, and 44 papers on classification of melanoma. The implication of the appropriate technique is ultimately

application dependent; thus, the literature review focused on finding the gap to develop a highly accurate system compared to other state-of-the-art methods.

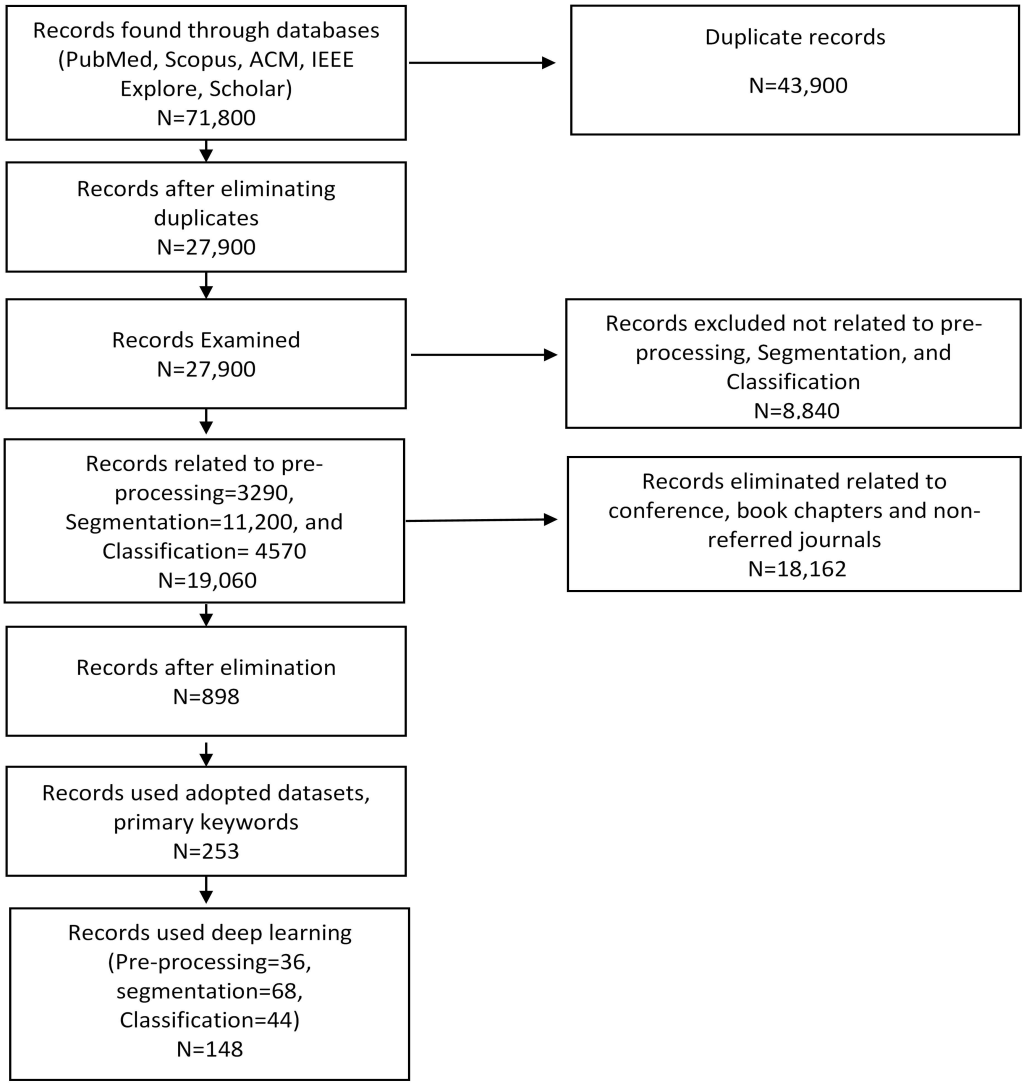



Figure 2.4: Selection and screening process.

## 2.5 Literature Review on Melanoma CAD System

The review is categorized according to different phases of the proposed CAD system as shown in Figure 2.5 which is mainly concerned with three tasks:

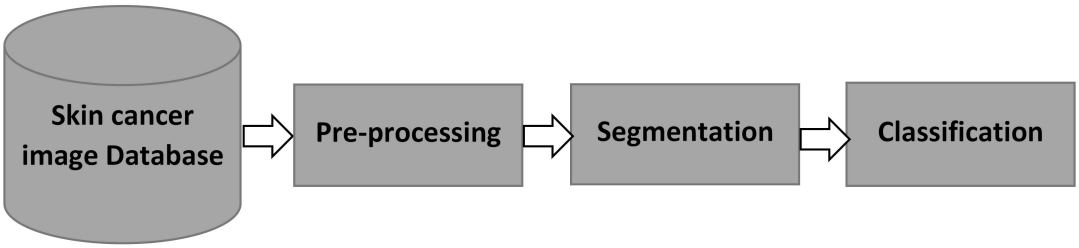



Figure 2.5: CAD system for melanoma skin cancer.

### 2.5.1 Hairlines removal and contrast enhancement

The pre-processing phase of any classification system facilitates removing the noise artifacts, unwanted pixels and enhancing the quality of images. The melanoma images extensively require pre-processing because skin images are poor in illumination, contain hairs, color variations, and noise. Figure 1.2 illustrated some examples of these cases. During acquisition, transmission, and mathematical computation, noise greatly affects image quality. There are a number of reasons why lesion images are prone to high noise variations, including low lighting conditions, limited resources, and limited ability to capture lesion images using clinical instruments like dermatoscopes. An accurate diagnosis of a disease requires that precise information be collected from the lesions. Pre-processing the lesion area is challenging due to the heterogeneous appearance of the lesion. Additionally, hairlines introduce noise in the form of edges and poor colour contrast, resulting in weak feature extraction. Although hairlines may pose limited difficulties for human inspection, they may pose problems for automated image analysis steps such as segmentation and classification. There is an increasing need to enhance images by applying high-quality pre-processing procedures. In designing real-time CAD tools, automatic pre-processing methods are helpful since they preserve lesion information at low computational costs and can effectively eliminate irrelevant details.

The inefficiency of hair removal methods leads to over-segmentation and poor

classification features [45], which results in a poor feature set for classification. Researchers used two phases of image processing to eliminate hairlines; (a) detection and elimination of hair pixels using any image enhancement methods, and (b) finding the color and texture of overlapping skin areas and repairing them using inpainting techniques such as bi-linear interpolation, histogram equalizations, and fast marching. It has previously been observed that researchers tried to detect and improve skin cancer image quality by image enhancement techniques to achieve accurate lesions. These image enhancement methods are divided into four major categories; *Histogram equalization* spreads the most frequent intensities over the image and increases the image contrast from low to high. One of the most commonly used histogram-based methods is contrast limited adaptive histogram equalization (CLAHE) [46, 47] used for hair removal from melanoma images. *Morphological filters* deal with the form and structure of the image and are quite effective in removing noise from medical images. It is a collection of non-linear operations to extract shape-related image components such as boundaries, skeletons, and convex hull [48, 49, 50]. The most common morphological operations are area closing, opening, and erosion. *Gaussian filters* perform subtraction operation on images to reduce noise and preserve edge information [51, 52]. Lastly, *edge detection techniques* also proved to be effective in detecting hairlines, edges, and ruler marks; however, these can not be used alone. Some morphological operations and image restoration methods are used to achieve the objectives [53, 54]. The use of these enhancement techniques motivated many other studies for the hair removal process in melanoma images [55, 56]. For example, an initial attempt was made in 1997 by Tim Lee [57] to propose an approach, a dull razor to occlude hairlines from melanoma images. In their work, the morphological closing operation is employed as a hair detector that determines dark hairlines and smoothes out all the black hair. Then, the bi-linear interpolation operation is used as an inpainting method to replace corresponding pixels with non-hair pixel values. In 2009, Xie et al. [58] proposed a

top-hat operator to detect hairlines and segment them based on a statistic threshold value. Then, occluded hair information in the images is repaired using the PDE-based inpainting method.

A similar type of study is conducted in [59], given a new algorithm for removing hairlines from images by employing a bottom-hat operator as a detector. It is applied on the Y-channel image colour space to find thick hairlines followed by binarization operation. After that, histogram function and morphological closing operations are used as an inpainting method to repair images by replacing hair pixels with the highest intensity value. Another work by Abbas et al. [52] designed three steps; (a) derivative of gaussian to detect hairs, (b) enhancement using morphological operators, and (c) repairing is done using fast marching inpainting technique. In a review study [55], a comparative analysis was conducted for several hair removal methods based on three classes: linear interpolation, inpainting by partial differential equations and exemplar-based methods. There is an example of linear interpolation hair removal algorithms in [57, 60]. Few papers have attempted to implement the concept of applying inpainting via partial differential equations outlined in [55, 61]. Moreover, little attention is paid to exemplar-based inpainting methods [62, 63]. The dermoscopic skin cancer images also suffer from other issues such as blurring, low contrast and resolution, including hairlines, which need to be addressed to get cleaned images. For instance, Ren S. et al. [64] presented a multiscale resolution-based residual network to generate images with precise outline details. Another work given by Zhang J. et al. [45] proposed a high-resolution multiscale network (HRNet) to produce high-resolution features of the images. Their network employs multiple branches of parallel convolutional blocks, which are fully connected to make the flow of information smooth between different branches.

The primary concern of employing hair removal techniques is to boost the performance of the CAD system and using it in clinical care to clean noisy images. The

Table 2.3: Summary of studies for hair removal process on melanoma images.

| Method | References | Advantages | Disadvantages |
|---|---|---|---|
| Histogram equalization | [65] | Straight forward method and increases the contrast of an image. | It increases the contrast of noise in the background area. |
| Morphological operations | [48, 52, 66] | More suitable to reduce the high level of noise. Preserve critical information. | The selection of the correct structuring element is a quite difficult process. |
| Gaussian filters | [51] | Fast and effective in reducing noise. | Time-consuming and eliminates important information from images. |
| CLAHE | [47] | Gives excellent results for increasing image contrast. | Slow, and needs a high number of operations per pixel. Not effective for color images. |
| Edge operators | [53, 54] | Simple and easy to implement. | Not suitable for very noisy images. |

benefits and limitations of image enhancement methods are listed in Table 2.3. In much of the published literature, the focus is on removing thick and dark hairs while ignoring the detection of light and thin hairlines. It is quite challenging to detect light and thin hairlines in low contrast images. Another challenging task is to preserve pixel information such as edges, color, texture, lesion border, and contrast after removing hair artifacts which is not addressed in the literature studies. Hence, in this research study, eliminating hairlines and preserving its contextual information is a major concern.

### 2.5.2 Lesion segmentation

In image analysis, segmentation is a crucial step for distinguishing object descriptions from background images [56]. Segmenting images is not standardized because the areas of interest and description of objects vary from application to application. As a result, the technique to be used depends on the nature of the problem. Melanoma lesions are

difficult to segment because of their significant variations in color, size, shape, location, and texture. As shown in Figure 2.6, the approaches applied in lesion segmentation are categorized into six categories.

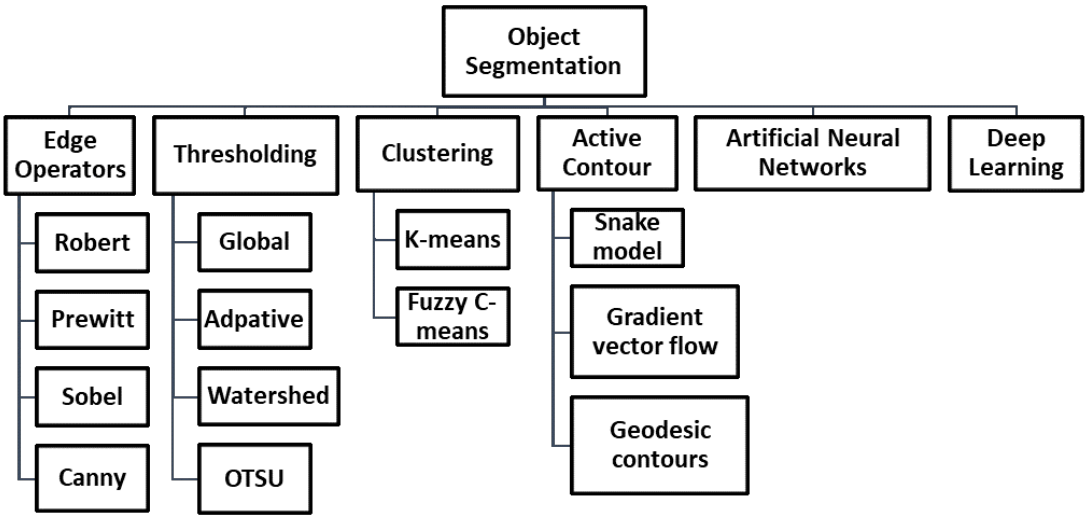



Figure 2.6: Types of segmentation approaches.

- **Edge operators**-based methods utilize edge operators to collect information about edges of the object [67, 68]. Some studies employed edge detection methods on skin cancer images to find discontinuities based on points, lines, and edges. Out of the all edge operators canny is the most effective in detecting fine and sharp edge boundaries as utilized in these studies [69, 70].
- **Thresholding** methods are used to differentiate the lesion from the surrounding tissue by determining the object's information based on one or more threshold values [71, 72, 73, 74]. The authors in [75] applied an adaptive method on skin data to calculate the optimum value of the threshold for separating the lesion area from its background. Another example of thresholding techniques are Otsu [76], and watershed [77] implemented for analyzing the presence of a lesion.
- **Clustering** is an unsupervised learning approach and aims at dividing similar kinds of pixels values into groups based on some optimization function. These

methods form homogeneous clusters by extracting region information by grouping similar pixels into one cluster and others into different clusters [78, 79]. Among the various clustering methods in the literature, K-Means and fuzzy C-means are the most commonly used for image segmentation. We have cited those which were applied for skin lesion segmentation [80, 46].

- **Active contour** methods were widely used for skin lesion segmentation [54, 81, 82] to smooth the image shape and to form the closed contour. Contour is a collection of points that are interpolated as linear, splines, and polynomial curves. The work presented in [54] proposed a geodesic active contour method for lesion segmentation, which determines the accurate boundaries of the lesion despite noise artefacts, color variations, and poor edges.
- **Artificial intelligence** methods include the use of classifiers such as ANNs, decision trees, SVM [83, 84]. ANN has been widely applied to lesion segmentation and given better results as compared to other baseline methods [85, 86]. ANN faces some the challenges, such as local optima problems, high training time, optimized network structure, suitable weights, number of layers, and unable to find low-cost functions. Thus, many studies made attempts to improve their performance using soft-computing algorithms to find an optimal solution such as evolutionary [87], social group optimization (SGO) algorithm [88], particle swarm optimization (PSO) [83], and artificial bee colony algorithm. [89].
- **Deep learning** techniques have been frequently applied in the field of image segmentation due to their remarkable performance and offered higher accuracy. Deep neural networks label each pixel with a class of object or non-object region and are effective in semantic segmentation applications. It was attempted in 2013 to segment X-ray images using the CNN approach [90]. In order to reduce training time, the researchers applied patch-wise classification on raw images to identify bone tissues. In 2015, Melinscak et al. [91] proposed a similar approach

for segmenting blood vessels from fundus images. They achieved an average accuracy of 94% using a deep convolutional neural network with max-pooling layers instead of subsampling or downsampling layers.

A fully convolutional neural (FCN) network was proposed by Long et al. [92] in 2015 as the first end-to-end pixel-wise segmentation technique. To generate segmentation maps, they utilized AlexNet, VGGNet, and GoogleNet classification models to replace convolutional layers with fully connected ones. This approach was faster than the patch classification approach. Hong et al. [93] proposed deconvolution networks (DeconvNet) as an extension of FCN networks. The class-oriented feature maps created by this architecture reduce the training and inferencing time by decoupling classification and segmentation tasks. Pooling layers was the main problem with FCN, as it increased the field of view and discarded essential information. In contrast, semantic segmentation requires precise knowledge of class maps and must preserve 'where' information.

In 2017, Badrinarayanan et al. presented a network architecture based on encoder-decoders called SegNet [94], which is similar to VGG16. With this network, the encoder reduces spatial dimensionality with pooled layers, and the decoder gradually regains object information and spatial details. In the decoder section, it applies the pooling indices computed by the max-pooling layers of the encoder section. A similar network, U-Net [95], has been successfully applied to segment medical images using encoder-decoder architecture. In 2018, Chen et al. [96] introduced deep convolutional nets called DeepLab networks to address the segmentation task. These networks are based on dilated/atrous convolutions that explicitly control resolution to preserve feature map information computed in the deep convolutional layers. It differs from SegNet or U-Net in the ability to expand the field of view of filters to have a greater context of information without increasing the number of parameters. The details of these networks can be found

in review papers [97, 98].

- **Deep learning in lesion segmentation** The studies that specifically focused on the segmentation of lesions from skin cancer images applying deep learning methods include work of Al-Masni et al. [99] that presented an extension of FCN called fully resolution CNN (FrCN) for lesion segmentation to extracts image features without using any pre-processing technique. VGG16 layers inspired the network by removing its subsampling layers to preserve its full resolution of individual pixels. Their network outperformed FCN, VGG16, SegNet. Another work given in [100] extended VGG16 architecture to perform lesion segmentation by eliminating its fully connected layers and resizing the final feature map so that it matches with the output size. Lei Bi et al. [101] proposed an FCN based ResNet model that learns the visual features of the lesion corresponding to each class based on their probability. This network is significantly different from other networks as it segments images according to their category and learns more deep features. The work proposed by Yu. et al. [30] presented a very deep residual network having 50 layers to calculate rich and more significant features for accurate recognition. Residual learning was applied to prevent overfitting and degradation problems of the convolutional neural networks. A similar type of lesion segmentation work can be seen in the studies [99, 102, 103] that have given different approaches for lesion segmentation either inspired from existing networks or extended them using transfer learning.

  There are other variations considered in using deep learning techniques such as layer structure of network architecture was changed, applied a transfer learning approach [104, 105, 106] and used fine-tuned training parameters [107, 102, 108, 30] aiming to reach an optimal solution that reduces computation time, power and increase network's accuracy. The benefits and drawbacks of given segmentation methods are summarized in Table 2.4. Out of these, deep learning proved a

powerful method capable of extracting features from the raw data and training the network to predict the output. The use of deep neural networks is rapidly increasing on the subject of image segmentation and classification.

Table 2.4: Comparison of image segmentation methods.

| **Method** | **References** | **Advantages** | **Disadvantages** |
|---|---|---|---|
| Edge detection | [68, 70] | -Detect finer edges and sharp edge boundaries as compared to other operators. | -Sensitive to noise artifacts.<br>-Not suitable for image set containing too many edges. |
| Thresholding | [75, 77] | -Do not require any initial information about images. | -Wrong selection of threshold results into under or over-segmentation. |
| Clustering | [80, 79, 109] | -Suitable for homogeneous and large data sets. | -Wrong selection of initial centroid value affects the performance. |
| Active contour | [54, 82, 110] | -Enhances contour outline, energies can be defined manually. | -Slow computation performance in case of large size images.<br>-Not able to segment the nearest objects efficiently. |
| Artificial intelligence | [85] | -Better handling of information and able to solve complex problems.<br>-Low error rate and uses learned knowledge to solve problems. | -High cost involved.<br>-High training time required for large and complex datasets. |
| Deep learning | [111] | -Capable of self-organization.<br>-Accept raw images and do not need preprocessing and feature extraction.<br>-Perform optimized learning. | -Layer architecture is highly complex.<br>-Extremely expensive and needs high hardware like GPU machines.<br>-It requires a very large amount of data to perform better. |

Recent developments have heightened the need for a melanoma detection system, and considerable literature has grown up around the theme of lesion segmentation. It has been analyzed that the most popular network's choices for semantic

segmentation are U-Net, FCN, and SegNet, containing sampling and upsampling layers to maintain the spatial resolution of the output. The major problem in these networks is that they suffer from a low spatial resolution output due to the repetitive use of max-pooling and striding at consecutive layers. Moreover, the optimized, comprehensive, and customized frameworks proposed in the literature for lesion segmentation tasks still exhibit difficulties, such as low performance or high execution time due to complex architectures. These networks have not given satisfactory performance on the adopted dataset. Hence, there is probably room for further improvement in terms of accuracy by designing a new architecture. In the research work, we have designed CNN networks based on the concept of sampling-upsampling layers and atrous convolutions to segment lesions maintaining high resolution effectively. The networks were designed from scratch by organizing different layers of the network, fine-tuning hyperparameters, and using a suitable loss function.

### 2.5.3 Classification

This section deals with explaining state-of-the-art approaches given based on the classification phase of the diagnostic system. A computer-aided system dates back to 1998, and several classification methods were proposed in the literature to recognize melanoma and its types. Image-based classification systems are developed by using traditional machine learning techniques such as K nearest neighbour (KNN), support vector machines (SVM), artificial neural networks (ANN) [27, 112]. In contrast, the choice of employing classification techniques has shifted to deep learning approaches [113] from recent years. Deep learning approaches have shown remarkable performance in solving classification problems [114, 100]. In deep learning, the CNN is an efficient method to extract global data consisting of information about edges, curves, shapes to

generate a complex feature set to perform classification. The CNN network comprises many layers such as convolutional, normalization, activation, pooling layers, and fully connected layers. An early adopter Krizhevsky et al. [115] using deep convolutional neural networks to classify 1.2 million images containing 1000 different classes in the ImageNet, the first-ever study to do so, published a paper in 2012. Many such types of CNN networks were proposed for classification tasks, for example, VGG16 & 18, Inception, GoogleNet, ResNet, and MobileNet but not limited to. These networks were trained on different image datasets, and their application to any other problem domain is entirely data-dependent. A recent survey on deep learning architectures used for melanoma classification can be found in [116, 117] that discussed various types of deep learning architectures, challenges, and trade-offs between them.

The widely adopted approach is to use pre-trained networks as transfer learning to classify skin cancer types. With transfer learning, the need for annotation is reduced by transferring the knowledge of deep learning models. The models are fine-tuned and applied to solve the problem of a new task. Khalid et al. [118] proposed a transfer learning approach by replacing the softMax layer of the original architecture with AlexNet to classify three types of skin cancer based on the $PH^2$ dataset. Additionally, they used data augmentation and fine-tuning parameters to make the network well suited for the problem. In 2019, Jason et al. [119] proposed a fusion method by utilizing the ResNet50 network's learning knowledge and handcrafted features from the clinical information module. The prediction score of each category from both types of information modules was combined, and the final prediction score was identified using logistic regression. More work-related to the transfer learning approach for skin cancer detection can be seen in [120, 121, 122, 123]. Kassani et al. [124] performed a comparative study between different types of deep convolutional neural networks such as Xception, AlexNet, VGGNet16 & 19, and ResNet50 architecture for skin cancer classification. It was observed that augmentation and pre-processing steps positively

affect accuracy ratio, and out of given architectures, ResNet50 has better performance. The problem in utilizing pre-trained networks for new classification tasks using transfer learning is a mismatch between learned features on natural images and medical images. Because those networks were trained on ImageNet public dataset, it is an ineffective practice to apply learned knowledge from different data to medical images.

Some attempts were made to increase the model's accuracy by combining deep neural networks and supervised machine learning classifiers. In 2018, Ulzii et al. [125], proposed a technique that employed a pre-trained AlexNet and the SVM classifier to classify 3753 skin lesion images. They explained the advantages of AlexNet in terms of a few training parameters and validity. The work of Mahbod et al. in 2019 [126] presented three deep learning networks AlexNet, VGG16, and ResNet18, as feature generators. They trained the SVM classifier with each feature set of three neural networks and fused their results to make final predictions to detect three categories of skin cancer. They concluded in their study that merging multiple features from various networks leads to a better classification rate. Kassani et al. [124] performed a comparative analysis between different types of deep convolutional neural networks such as Xception, AlexNet, VGGNet16/19, and ResNet architecture for medical applications. It was concluded that out of given architectures, ResNet had given a better performance. Another study proposed by Gessert et al. [127] given an ensemble model comprising three networks such as EfficientNets, SENet, and ResNet for addressing classification tasks of the ISIC 2019 challenge. They applied the loss balancing technique to overcome the class imbalance problem and added unknown metadata as an additional branch to the network to increase the classifier's performance. Another work given by Andre et al. [128] implemented eight pre-trained models to develop an ensemble model for the prediction of eight types of cancer categories. They replaced classifiers to adjust the model according to the dataset keeping all classification layers and obtaining higher accuracy. The method proposed by To T.D et al. [129] classified different types of

cancer from clinical and macroscopic images. They applied the divide and conquered technique to many small binary and 3-ary problems for easy processing. To overcome the class imbalance problem, they applied an upsampling approach, and two networks named, EfficientNet and InceptionResNet were trained using focal binary loss function.

Many contributions were made by developing new CNN architectures from scratch, focusing on skin cancer data. An optimal network design requires extensive experimental study to analyze the impact of layers, parameters, and loss functions to obtain higher performance. A new architecture of CNN trained ISIC dataset was proposed by Imran et al. [130] for the classification of eight types of skin cancer with an accuracy of 89.58%. The network was structured into three blocks consisting of 68 convolutional layers and other layers such as batch normalization and ReLU layer. Each block has a varying number of parameters to extract complex lesion information. Another work presented a new design of a network known as DermaKNet [131] that employed ResNet50 layers as a base model. The network consists of modulation blocks, polar pooling blocks, and asymmetry blocks interconnected to pass lesion features to the last fully connected layer. The network outperformed other state-of-the-art studies with an AUC score of 91.7. Similarly, many new ideas were proposed to motivate the design of deep neural networks, which can be seen in detail in these studies [132, 133, 134, 135].

## 2.6 Existing Literature Gaps

Recent developments have heightened the need for a melanoma detection system, and considerable literature has grown up around the theme of removing noise artifacts, lesion segmentation, and classification. In the context of the pre-processing phase, investigators given several methodologies based on illumination correction [65, 46], Gaussian filters [51, 52], and morphological filtering [48, 66]. However, the major difficulty recorded in the pre-processing phase is still required to preserve color resolution,

Table 2.5: Comparison of image segmentation methods.

| **Method** | **References** | **Advantages** | **Disadvantages** |
|---|---|---|---|
| Machine learning | [27] | - Simple to implement, handle noisy data and performs well in case of large input data.<br>- Ability to work with incomplete information and able to perform parallel processing. | - Selection of initial seed points.<br>- Suitable for binary classification.<br>- Suffers from local minima and overfitting.<br>- Requires high processing time for large and complex networks. |
| Deep learning | [116, 136] | - Automatically deduce features and optimally tune them according to an expected outcome.<br>- Perform parallel computation.<br>- Flexible architecture. | - Require a very large amount of data to process.<br>- High power computational hardware required. |

border, and texture information while removing noise artifacts. These characteristics are significant to differentiate melanoma from a normal mole. Moreover, there are only a few recent studies that have developed techniques for repairing hairlines in skin images, according to the literature analysis. Also, those studies mostly focused on the removal of hairlines and ignored the effects on lesion patterns. The elimination of hair pixels can result in unwanted blurring, discolored images, and poor textures in melanoma images. Existing studies have proposed ways to occlude thick and dark hairlines, which are easy to detect due to a large difference in intensity between hair and lesion pixels. However, it is much more difficult to detect thin and light hairlines. Furthermore, the literature studies do not utilize an extensive dataset. Thus, the proposed research study aims for a robust method of detecting hairlines while simultaneously preserving pixel information after hair removal process. The development of an effective pre-processing model can be helpful in cleaning hairs from infected skin lesions without causing discomfort to patients.

Other significant contributions are found in segmenting lesion objects. Several

attempts have been made to extract accurate lesion areas employing edge detectors, thresholding, clustering, active contours, artificial neural networks, and deep learning techniques. Out of the given methodologies, artificial intelligence and deep learning is the most appropriate choice to extract lesion object. Here, a significant concern is required to address the gap by proposing a neural network able to extract lesion information within less execution time and with high accuracy. A considerable amount of literature has been published on the melanoma classification phase too. Various literature studies have implemented several standard classifiers such as machine learning, neural networks, feature extractors, and deep learning. However, to achieve generalization, less complex architecture, and low inference time is still an open challenge in this area. Thus, the design of a lightweight CNN offering high performance needs to be considered. Furthermore, to the best of our knowledge, there are very few studies that have conducted experiments on a complex dataset published by ISIC [137] in 2020, consisting of a lot of duplicate samples, which could lead to over-fitting. Also, the dataset contained samples of dark skin and aged individuals with complex lesion patterns.

The advantages and disadvantages of various segmentation and classification techniques used for melanoma problem so far are analyzed in Table 2.4 and 2.5. It is observed that researchers have proposed various optimized, extended, and customized frameworks in the literature, still exhibit some difficulties such as the need for efficient hair removal techniques that do not lose important information, segmentation techniques to extract correct lesion area, and need of accurate classification models.

In the discussed state-of-the-art methods, there is a compromise between achieving high accuracy in less computational time. As the size of the dataset increases, execution time also increases. Thus, we aim to propose a novel design of a classification model that classifies a large dataset in less computational time with limited GPU resources. Moreover, there are limited studies that analyzed the impact of pre-processed and

segmented images on classification performance. To make the classification model more generalized, it is trained for large set of dermoscopic images. As a final point, it is imperative to mention that to reach a fully automated system, large datasets and high-power computational hardware are required to achieve better accuracy.

## 2.7 Conclusion

This chapter has presented an overview of the melanoma structure, its types, and current clinical approaches used to diagnosis it. In addition, it also highlighted the advantage of employing an automatic computerized system for early and fast detection. The CAD system is categorized into three phases: pre-processing, segmentation and classification. The type of techniques proposed in the literature for each stage of the CAD system is thoroughly discussed. We reviewed recent works related to these steps, and a summary of them is illustrated in tables including their advantages and disadvantages. A comprehensive analysis of literature review is provided in the form of research gaps to have clear insights about the research question designed in this study. The next chapter discusses implementation details of the system design and various datasets used for the experimental evaluation of proposed methods.

# Chapter 3

# Overall Design Flow and the Selected Datasets

In this chapter, the overall design of the proposed system is discussed for a better understanding of the methods designed towards melanoma diagnosis. It is followed by a discussion of various datasets used in this research for different phases of system and basic operations for data cleaning. The literature study conducted in the previous chapter has revealed a high demand for robust, accurate, and cost-efficient automatic melanoma detection systems to assist clinical experts in the initial assessment. As New Zealand has a high rate of melanoma, this study can help in the early diagnosis of it. The focus of our research is the design of the melanoma CAD system, where each phase is implemented by proposing novel techniques to obtain the best performance as compared to state-of-the-art methods. The approaches were implemented and executed on high computing powered GPU hardware machines such as GeForce RTX 2060 and 2080 with computation capacity '7.5' to achieve better performance in low computation time.

## 3.1 System Overview

The overall design of the system composed of four phases, starting with basic image operations, hair removal, segmentation and classification as given in Figure 3.1. The first phase was to collect skin cancer datasets from various web sources and resize them into suitable dimensions. Then the images were pre-processed to eliminate any noisy elements, specifically hairlines, because these can degrade the performance of the classifier. After pre-processing, lesion information was extracted using segmentation techniques. At the final stage, a classifier was developed to detect melanoma from benign cancer. The classifier was trained and evaluated in two stages by feeding it with two different types of datasets such as with raw images and pre-processed images to analyze the impact of the pre-processing and segmentation on the classifier's performance. The primary aim was to gain high classification accuracy for recognizing melanoma at an early stage facing all challenging trade-offs between high accuracy rate and design of optimal network to process a large amount of data, including noise.

Different types of quantitative metrics are used to evaluate the segmentation and classification approaches. For example, accuracy, precision, recall, specificity, F1-score, and inference time are used to test classification approach. Additionally Jaccard score, and dice coefficients are used to evaluate the segmentation approach. System modelling and simulation is an iterative process that helps in improving performance based on these measures. The high performance will bridge the research gap raised after the literature review and enhance the system's design to achieve the expected results. The main focus of this chapter is to explain the datasets, discussing challenges during data preparation such as sorting of data and removal of duplicates, and basic image processing operations. In the next chapters, the phases of CAD system such as pre-processing, segmentation and classification will be explained in detail.

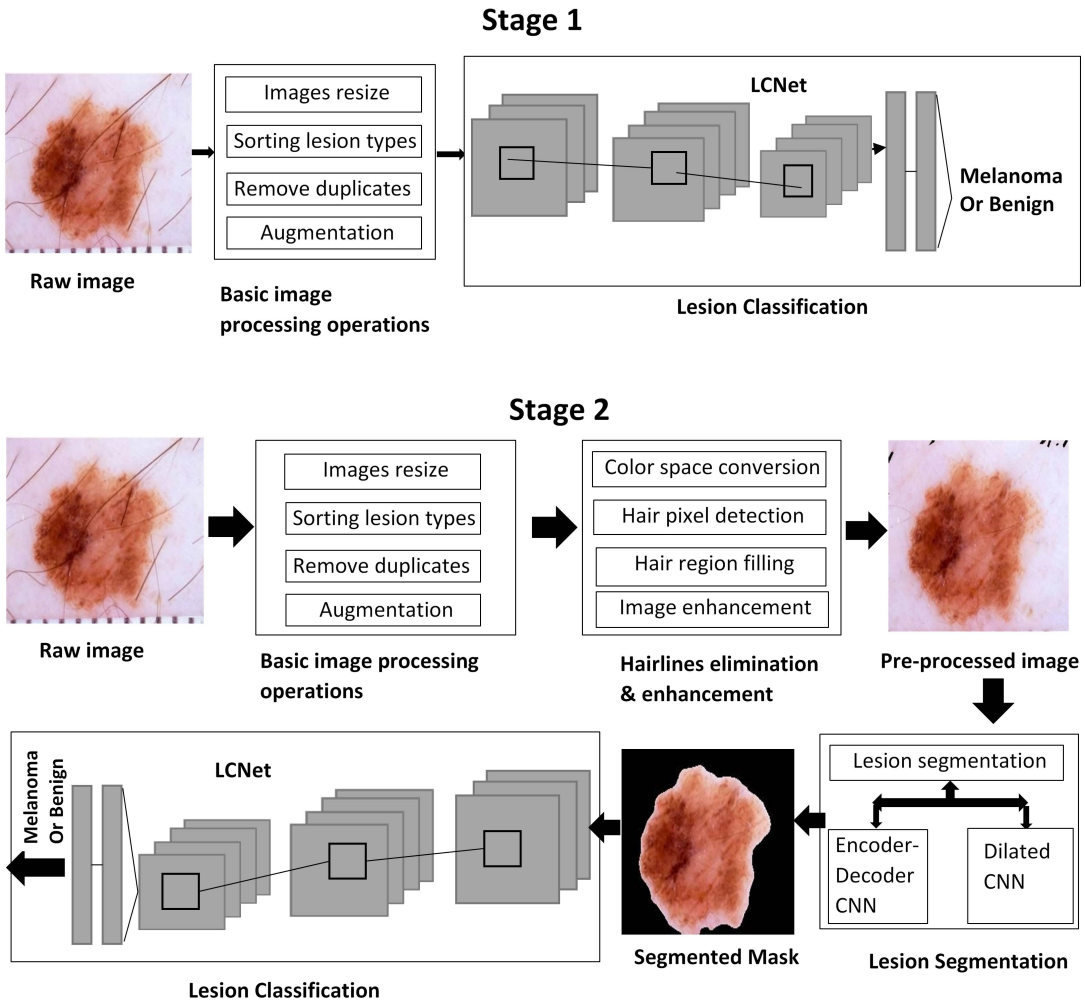



Figure 3.1: Design flow of melanoma detection system.

## 3.2 Datasets

A number of well-known datasets from open-access dermatology repositories were used in this study, including the ISIC dermoscopic archive [16, 138], HAM10000 [139], and $PH^2$ [140]. There are five variations of ISIC dataset from 2016 to 2020, some of them are the same samples, some of them are new samples. This research focuses on dermoscopic images that are generated by dermatoscopes. For the segmentation and classification phases, different datasets were designed. In order to validate the performance of the segmentation method, the ground truth images were considered

as reference images. There were various types of skin lesion used for assessing the performance of the classification method. While preparing data for different phases of the system, some significant steps were performed on it. This detail is provided in Table 3.1 and 3.2 for each individual dataset.

- **ISIC 2016:** The 'Skin lesion analysis towards melanoma detection' utilizes the ISIC repository of dermatologist-annotated skin lesions. The dataset was categorized into 900 training and 379 test images. Additionally, it consists of ground truth binary mask annonated by human experts for the consideration of comparisons.

- **ISIC 2017:** This dataset consists of 2,000 skin lesions for training purpose, 150 validation, and 366 test images. The ground truth samples are also provided to compare the segmentation outputs.

- **ISIC 2018:** This challenge provides separate 2,594 training images to engage for segmentation and classification task. A validation set of 100 and blind test set of 1,000 samples was given to perform testing of the method.

- **ISIC 2019:** It contains 25,331 dermoscopic training images distributed among eight different types of lesion which were broadly classified into benign and malignant. The test set have 8,239 lesions which need to be classified by the method.

- **ISIC 2020:** The dataset have 33,126 image samples which were divided into training, validation and test sets while the training and testing of the proposed methods. It consists of many duplicate lesion samples.

- **PH$^2$:** It is obtained from the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal. This dataset contained a total of 200 images along with their ground truths.

- **HAM10000:** The dataset named HAM10000 ('Human Against Machine with 10000 training images') provided for automated diagnosis of pigmented skin lesions contains 9,900 dermoscopic lesion samples.

Table 3.1: Distribution of lesion samples in different datasets used for segmentation task.

| **Dataset** | **Training** | **Training ground truth** | **Validation data** | **Validation ground truth** | **Test** | **Test ground truth** |
|---|---|---|---|---|---|---|
| ISIC 2016 | 900 | 900 | — | — | 379 | 379 |
| ISIC 2017 | 2000 | 2000 | 150 | 150 | 366 | 366 |
| ISIC 2018 | 2594 | 2594 | 100 | 100 | 1000 | — |
| $PH^2$ | 200 | 200 | — | — | — | — |

Table 3.2: Distribution of lesion samples in different datasets used for classification task.

| **Datasets** | **Training** | **Validation** | **Test** | **Total** |
|---|---|---|---|---|
| ISIC | 9017 | 1288 | 2577 | 12882 |
| HAM10000 | 6930 | 990 | 1980 | 9900 |

## 3.3 Elimination of Duplicate Samples and Sorting of Data

Originally, the data were not categorized according to cancer types, so they were sorted into different classes by writing a script code in MATLAB. The image read function uses the ID from the metadata file to move that image into the corresponding category. Another major issue with the data was the presence of duplicate samples. There were the same skin samples but with a different name and similar lesion structures, making it difficult to distinguish them by naked eye. The method first converts the image to grayscale and then calculates its histogram. After that, the mean of each image is determined, and the correlation index of each image is compared with the other. When

the correlation between two images is greater than a threshold of 0.99, the images are considered identical, and one copy is discarded. The steps to eliminate duplicate samples in a dataset is explained in Algorithm 1.

**Algorithm 1** Duplicate image elimination method

**INPUT:** Load set of images
**for** i:1 to length(image-set) **do**
  Read $I_i$(x,y)
  Convert $I_i$(x,y) RGB image into grayscale, $I_i^{'}$(x,y)
  Calculate histogram of image, $I_i^{'}$(x,y) as h1= imhist($I_i^{'}$(x,y))
  Calculate mean of image, $I_i^{'}$(x,y) as h1$'$=mean(vec2mat(h1))
  **for** j:1 to i **do**
    Read $I_j$(x,y) // Read all images till the count of i and compare with $I_j$(x,y)
    Convert $I_j$(x,y) RGB image into grayscale, $I_j^{'}$(x,y)
    Calculate histogram of image, $I_j^{'}$(x,y) as h2= imhist($I_j^{'}$(x,y))
    Calculate mean of image, $I_j^{'}$(x,y) as h2$'$=mean(vec2mat(h2))
    **if** correlation-index(h1$'$, h2$'$)$\geq$ 0.99 **then**
      Image $I_i^{'}$(x,y) and $I_j^{'}$(x,y) are same, thus one copy of image is saved.
    **else**
      Both images are saved.
    **end if**
  **end for**
**end for**
**OUTPUT:** Return image dataset after removing duplicates.

## 3.4 Resizing

Available data was 8-bit RGB images with different size ranges from $540 \times 722$ to $4499 \times 6748$ dimensions. These images cannot be used directly for the network's training due to their large size; thus, images were resized into three dimensions such as $300 \times 300 \times 3$ for pre-processing, $192 \times 256 \times 3$ for segmentation, and for classification $128 \times 128 \times 3$ according to the network's design and reduce computation speed. These different image sizes for each phase were finalized after tuning the best performance on the validation sets. Images were resized using the nearest neighbour interpolation

technique [141]. This method determined the color and intensity of unknown pixel positions using available information of surrounding pixels. The principle idea used in this method is to have an original image as a reference image based on which a new rescaled image is constructed. The resulting image is smaller in size without the loss of its resolution. When the size of an image decreases, the empty spaces in the original image are removed and reconstructed with a smaller size. Consider, an original image have height and width as $w_1, h_1$ and the resized image have dimensions as $w_2, h_2$. For finding the ratio in vertical (y) and horizontal (x) direction to perform rescaling is given as:

$$\mathrm{x} = \frac{\mathrm{w}_1}{\mathrm{w}_2}, \quad \mathrm{y} = \frac{\mathrm{h}_1}{\mathrm{h}_2}$$
$$where \;\; w_2, h_2 \neq 0 \tag{3.1}$$

Once the ratio is determined, the original image is rescaled along three channels (RGB) to a smaller size by interpolating pixels values, maintaining the horizontal to the vertical ratio. For example, in Figure 3.2 the original image of size $1024 \times 768 \times 3$ decreased to $192 \times 256 \times 3$ dimension. In this research work, the images are resized to different dimensions using this approach to reduce the computational time and to fine-tune the networks which are developed for different stages of melanoma detection system.

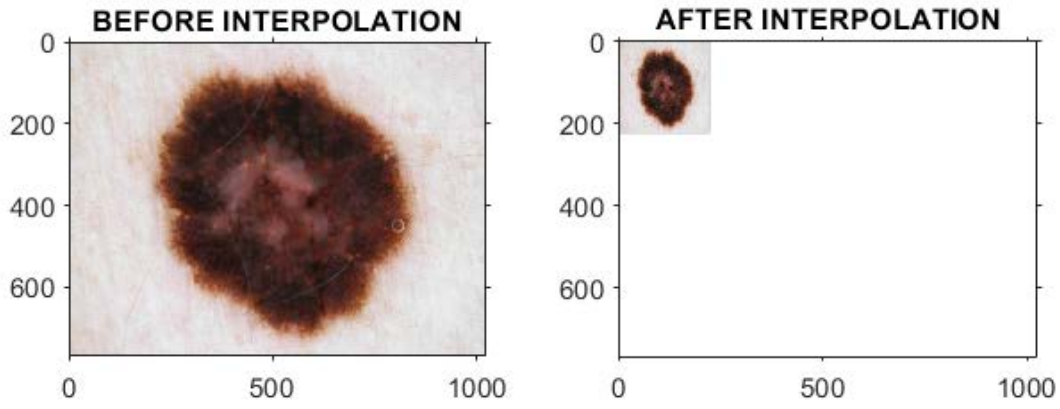



Figure 3.2: Image resizing using nearest neighbor interpolation method.

## 3.5 Data Augmentation

Data augmentation is a popular technique to increase training data manifolds to mitigate the lack of data. The artificial generation of images resembling original images is challenging due to unstable information in the skin lesions, such as irregular borders, diameter, shape, color, and texture. The ISIC datasets are highly skewed among various types of skin cancer. The data imbalance problem causes the network to get biased towards classes with many samples compared to those with low samples. Moreover, the visual similarity between different lesion types is high. Due to the high intraclass variations, it is challenging to differentiate skin lesions. Data augmentation is one such method applied over data to balance the data of minority classes. In order to balance the data distribution among skin cancer classes and avoid overfitting, the augmented images should look completely different at every epoch and be completely unique. The most common techniques used to increase the data is transformations [142, 95]. It involves translations, rotation, scaling and reflection of images to produce different images from original. Another approach is random oversampling that improves a classifier's performance by adding more image samples in the minority class.

### 3.5.1 Transformations

This method was applied during the network training to increase the number of overall samples in the dataset. The amount of data is increased in the training set by applying rotation, reflection, and scaling operations to generate new samples. The transformation operations such as reflection of $[-20^0, 20^0]$, rotation ranges from $[0, 360^0]$ and scaling in X & Y direction by $[-5, 5]$ factors was applied. The effect of transformations is displayed in the following figures, where Figure 3.3a displays original images, and Figure 3.3b shows augmented images.

(a) Original images

(b) Synthetic images

Figure 3.3: Augmentation data.

### 3.5.2 Random oversampling

Oversampling randomly duplicates examples from the minority class and adds them to the training set. Image samples are selected from the training set randomly with replacement. The minority class examples can be selected to form a more balanced dataset from the original and then returned or "replaced" in the original dataset, allowing them to be selected again. It is an effective method for the datasets where data is highly skewed among different cancer types. For example, if any majority cancer class has 1000 samples and the minority class has 500 samples, this technique oversamples the minority class to contain 1000 samples. This procedure is applied to balance the classes for the classification stage in which number of samples are determined for each class and more randomized data is added to the class having less number of samples. It is computationally expensive because it involves generating synthetic samples; therefore, the random oversampling is performed only for minority classes. The method presented modified image samples to the learner model on every epoch to prevent overfitting and bias problem. This technique leads towards improvement in the performance of the classification model.

## 3.6 Conclusion

A large amount of balanced data is needed for deep learning to work well. Consequently, we applied augmentation techniques to enlarge the training dataset and overcome the underfitting and overfitting problems. The data in the minority classes was increased by generating synthetic samples. Data augmentation was applied during the training process using transformation and random oversampling. This method presents a distinct batch of image data to the learner model over each epoch. The augmentation methods proved efficient in achieving better performance than the original datasets. Various experiments were performed using the discussed augmentation methods and achieved satisfactory results. The next chapter discusses a novel method for the proposed system's pre-processing stage, specifically the hairlines removal process and contrast improvement, including their experimental results.

# Chapter 4

# Pre-processing

## 4.1 Introduction

In this chapter, an introduction to noise problem is provided that immensely affects the quality of skin images. It is followed by the mathematical and implementation details of two pre-processing methods proposed to deal with hairlines problem and image enhancement. Finally, the experimental results illustrate importance of the proposed method which is published in the following journal articles as given:

- R. Kaur, H. GholamHosseini & R. Sinha, "Skin Lesion Segmentation using an Improved Framework of Encoder-Decoder Based Convolutional Neural Network", *International Journal of Imaging Systems and Technology, Wiley Online Library*, 2022.

- R. Kaur, H. GholamHosseini & R. Sinha, "Hairlines Removal and Low Contrast Enhancement of Melanoma Skin Images Using Convolutional Neural Network with Aggregation of Contextual Information", *Biomedical Signal Processing and Control*, vol.76, pp.103653, 2022.

A common problem in medical images is the presence of noise artifacts. Noise occurs due to various external and environmental factors. Medical images are significantly corrupted by noise during acquisition, transmission, and mathematical computation. The visual quality of digital images could be degraded by noise artifacts which reduces the precision and accuracy of image examination and interpretation operation. The low lighting conditions, and limited resources to capture medical images through clinical instruments lead to high noise variations. The performance of any image analysis system intensively depends upon the image quality. It is of paramount importance to perform disease diagnosis by extracting accurate and precise information from medical images. Thus, it is crucial to apply denoising techniques to cancer images to improve quality since it removes a barrier for feature extraction, segmentation, and classification. This research work primarily aims to address two issues related to pre-processing stage. Firstly, to remove light and thin hairlines from the skin cancer images and secondly, to regain the contrast and improve the image resolution after the hair occlusion process. To respond these issues, efficient methods were developed that preserves the lesion patterns while keeping its computational cost low so that it can be incorporated into the design of a real-time melanoma CAD system. Two approaches have been designed to achieve this task: hairlines removal and resolution enhancement. The hair removal technique deals with the adjustment of intensity values. It repairs light and thin hairlines so that they do not interfere with the lesion texture and structures. After removing hairs, the images are passed to the next stage of the system, where the degraded quality of the images restored using the contextual information preserving CNN model. The straightforward layout of pre-processing steps is represented in Fig. 4.1.

The primary research contributions made in this phase are listed as follows:

- Design of a hair removal algorithm using morphological operators to eliminate hairlines.

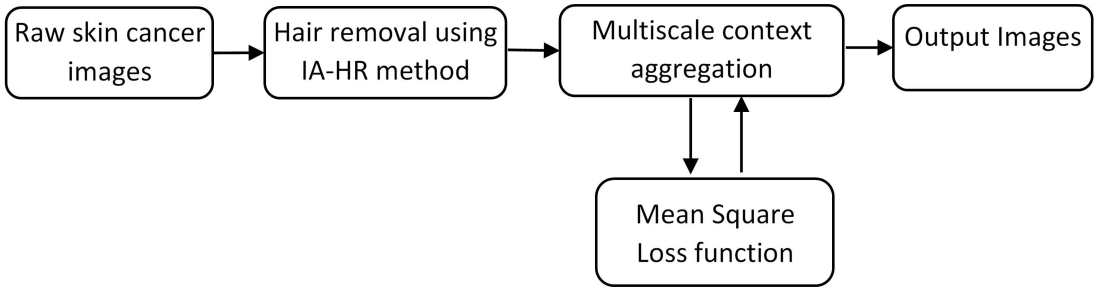



Figure 4.1: Systematic model for pre-processing process.

- Implementation of CNN model to acquire rich context and a broad scale of information from the lesion images.
- Estimation of contextual information, multi-level color features, and contrast to apply on the processed images for regaining visual equivalence.
- Performance comparison between proposed and other popular denoising methods.

## 4.2 Intensity Adjustment Hair Removal (IA-HR) Technique

Skin cancer images contain hairlines, as shown in Figure 4.2 affecting the system's performance with a high gradient. Typically, hair detection algorithms assume that hair pixels have colour slightly darker than lesion or background areas. There have been a few algorithms proposed to identify hairlines throughout the literature, and most of them can effectively detect thick hairlines, and only a few of them can find light and thin lines. In addition to this, the primary challenge is to preserve lesion information, border contours, and shape after the hair removal process. Therefore, a hair removal method is presented to address these issues based on morphological, thresholding, and pixel interpolation operations. This approach is named as Intensity Adjustment Hair Removal (IA-HR) algorithm as it deals with the adjustment of intensity values. It

repairs thin and light hairlines that interfere with the lesion texture and structure. The proposed approach is divided into four systematic steps, as shown in Figure 4.3; (a) color space transformation, (b) hairlines detection using morphological closing and subtraction operation, (c) filtration of non-hairy pixels with the use of binarization process and morphological dilation operation used to repair broken boundaries, and finally (d) texture restoration performed using interpolation technique.

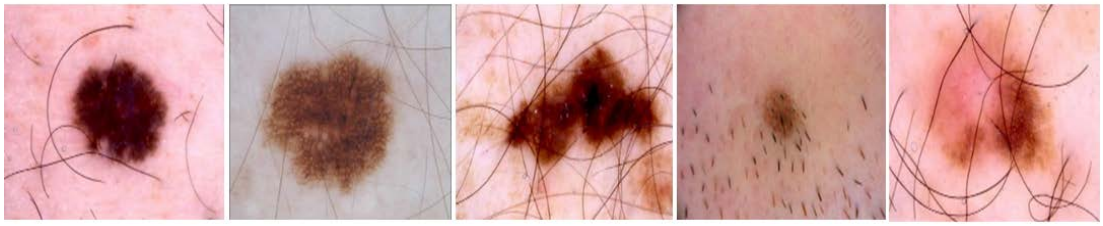


Figure 4.2: Examples of hairy images from five patients.

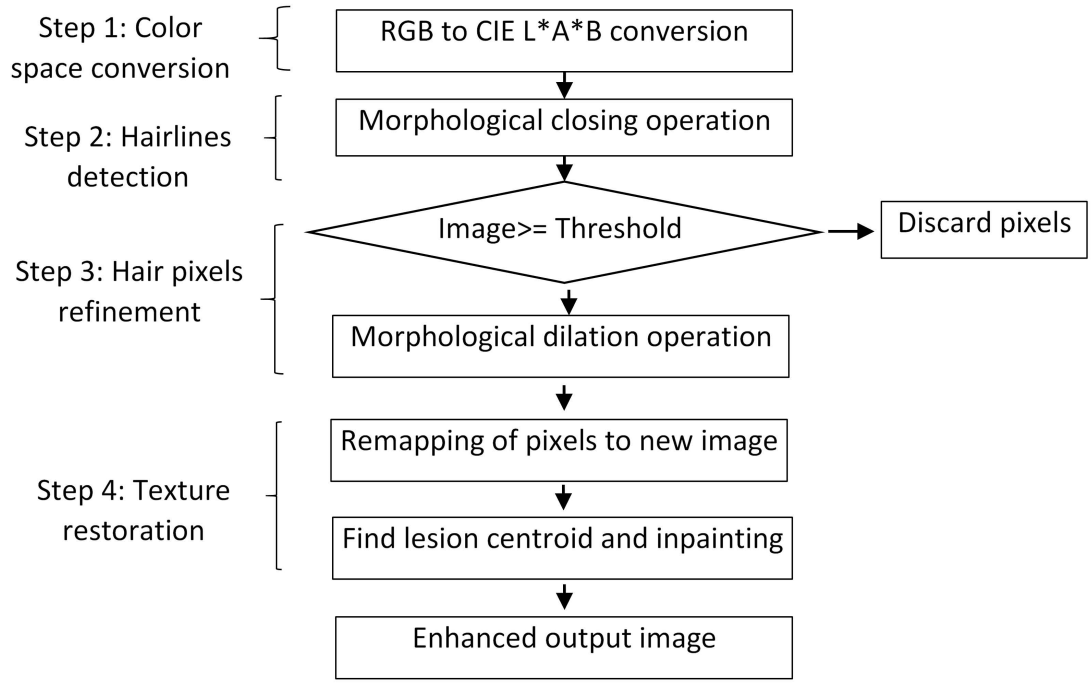



Figure 4.3: Systematic flow diagram of the proposed hair removal method.

Morphological image processing considers the form and structure of the image and is quite effective in removing noise elements in medical images. This technique

extracts shape-related image components such as boundary lines, skeletons, and convex hulls. Using morphological filtering, basic operations such as erosion, dilation, opening, and closing compare each pixel with its neighbours over a template called structuring element (SE). The neighbourhood pixels are processed either in a four-connected or eight-connected manner, the former defines pixel positions horizontally and vertically, and the latter considers diagonal positions. The SE is positioned at all possible locations in an image and compared with the adjacent pixels to perform logical operations such as AND, OR, NOT, and XOR.

### 4.2.1 Color space transformations

The first step of this method is to convert the $RGB$ colour space to a perceptually uniform $CIE\ L*a*b$ with formula given below for processing on a common platform. This colour space closely approximates human visual perception, and thus, the use of this colour space produces a more accurate result. Moreover, morphological operators are easy to perform on $CIE\ L*a*b$ color space as compared to the $RGB$.

$$L^* = 116\ f(\frac{Y}{Y_n}) - 16 \quad if\ \frac{Y}{Y_n}) > 0.008856 \tag{4.1}$$

$$a^* = 500\ [(\frac{X}{X_n}) - (\frac{Y}{Y_n})] \tag{4.2}$$

$$b^* = 200\ [(\frac{Y}{Y_n}) - (\frac{Z}{Z_n})] \tag{4.3}$$

The terms used in above equation $L^*$ indicates lightness, $a^*$, and $b^*$ are chromaticity coordinates. These color space coordinates are calculated using formula in which $X$, $Y$, $Z$ describe the colour perceived for image pixels and $X_n$, $Y_n$, $Z_n$ describe a specified black or white stimulus in an image.

### 4.2.2 Hairlines detection

To roughly detect the presence of hairs, first, an image closing operation is applied using (4.4) to close the irrelevant portions of an image. It also smoothes the object contours from the outside, eliminates small holes, and fill gaps in the boundaries.

$$\mathrm{Close_{img}} = (\mathrm{I(x, y)} \oplus \mathrm{SE}) \ominus \mathrm{SE} \tag{4.4}$$

The dilation operation ($\oplus$) is performed between the original image $I$ and the structuring element $SE$ followed by erosion operation ($\ominus$). Structuring is a critical part of morphological dilation and erosion operations because it permits further analysis of the input image. A SE is a matrix that identifies the pixel in the image being processed and defines the neighbourhood used in the processing of each pixel. Basically, SE is a matrix or a template that is used to traverse over an image. The SE is positioned at every possible location in the image, and it is compared with the pixels connected to it. A SE is considered Fit when all its pixels cover the pixels of the object. Whenever at least one of the pixels in the structuring element covers a pixel in the object it is called a Hit. A miss occurs when no pixels in the structuring element cover the pixels of the object. A SE may be any shape, typically its size and shape are selected to process the input image. In 4.4 close operation, 8-pixel connectivity was applied using $SE$ of shape 'disk' and size '3' that moves in all directions to achieve better results, as shown in Fig. 4.4. The selection of an appropriate value of $SE$ helps to deal with an image's structure closely.

The subtraction between $Close_{img}$ and an original image $I$ as given in (4.5) is performed to detect changes between two images and for leveling uneven parts of an image. $Diff_{img}$ contains the difference of pixels between the two images.

$$\mathrm{Diff_{img}} = \mathrm{Close_{img}} - \mathrm{I} \tag{4.5}$$


Figure 4.4: Structuring element movement over an image.

### 4.2.3 Refinement of hair pixels

It is a vital task to refine hair pixels by separating foreground pixels from the background. It can be achieved through the binarization process by selecting an optimal value of a threshold (T=25). The comparison is performed to filter the pixels of an image using (4.6). This comparison sets background pixels to '0' and highlights foreground pixels, including hair pixels values to '1'. It is checked that if any pixel value in $Diff_{img}$ is greater than the threshold, then those pixels are saved in $Dilation_{img}$, otherwise rejected.

$$\text{Dilation}_{\text{img}} = \text{Diff}_{\text{img}} > \text{T} \tag{4.6}$$

Further, dilation operation is applied to the resultant image using (4.7) for bridging the gaps and repairing the broken boundaries. Weak areas of hair occluded lines are filled through this operation to connect line pixels in all eight directions. This operation enlarges brighter regions and shrinks darker areas of an image.

$$\text{Dilation}_{\text{img}} = \text{Dilation}_{\text{img}} \oplus \text{SE} \tag{4.7}$$

### 4.2.4 Restoration of image

In this step, image values are transported to a new image in a coherence direction utilizing the remapping of structure. Therefore, the interpolation of pixels is carried out to construct an image called $New_{img}$. The size of this image is the same as that of the original image $I$. Non-hair pixels are copied to $New_{img}$ depending upon (4.8). This condition iterates through the whole image and copies those pixels of an original image to $New_{img}$ for which $Dilation_{img}$ is false, i.e., background pixels. Otherwise, it copies all pixels which belong to the lesion area based on the threshold. If any portion contains hairs, then it is replaced with the intensity of non-hair pixels. The remapping process is initiated using a seed point calculated using (4.9).

$$\text{New}_{\text{img}} = \begin{cases} \text{I}, \text{ if Dilation}_{\text{img}} == 0, & \text{pixels based on threshold} \end{cases} \tag{4.8}$$

$$\text{centroid} = (\text{X} - \text{C})^2 + (\text{Y} - \text{C})^2 < \text{C}^2 \tag{4.9}$$

$X$ and $Y$ are the dimensions of an image and the value of constant $C$ is fine-tuned equal to '180'. In this manner, the algorithm detects thick and thin hairlines in dermoscopic images and paint them with the lesion color. The image reconstructed with high contrast and converted back to its original RGB color space. The most challenging task is the appropriate value of $SE$, threshold (T), and finding the lesion's centroid (C) which is selected after conducting several experiments. In the end, an image is converted back to original colour space.

## 4.3 Multiscale Context Aggregation CNN (MCACNN)

A deep learning-inspired MCACNN [143] is proposed that approximates various image operators such as tone mapping, the original image's style, and the dehazing effect in

medical images. Skin images were intensively affected by hair elimination operation in the previous step; therefore, MCACNN is efficiently employed to improve the image contrast. The MCACNN first be trained on a set of images for processing high-frequency details to achieve greater accuracy. Upon training the network, degraded images can be processed directly without need for conventional processing. The network aggregates multiscale information of images and preserves local and global information at the deeper level of the network. It efficiently extracts information under the different receptive fields using atrous convolutions in the feature extraction layer. The design of the network is inspired by a basic CNN. It incorporates residual connections forming a deep network to preserve the image's information and enhance the non-linear mapping function at different levels.

The MCACNN model minimizes noise while preserving edge information degraded during the hair elimination. The architecture of the MCACNN model is displayed in Figure 4.5. The network works on image patches of size $256 \times 256$ extracted from original medical image data. Patch size is extracted based on the network's receptive field that captures all high-level features from the spatial region of an image. The network consists of multiscale residual aggregation blocks (MRAB) with different layers. Each layer has a respective function to calculate local and global features to form improved and high-resolution images.

### 4.3.1 MCACNN layers detail

The MCACNN have eight context modules named MRBL blocks composed of multiple layers, including the input, 2D convolutional, batch normalization, leaky activation, and regression layer. Two more customized layers, adaptive normalization $\mu$ and $\lambda$, are designed to adjust and strengthen the normalization and identity branches. The context modules boost performance by aggregating multi-scale contextual information.

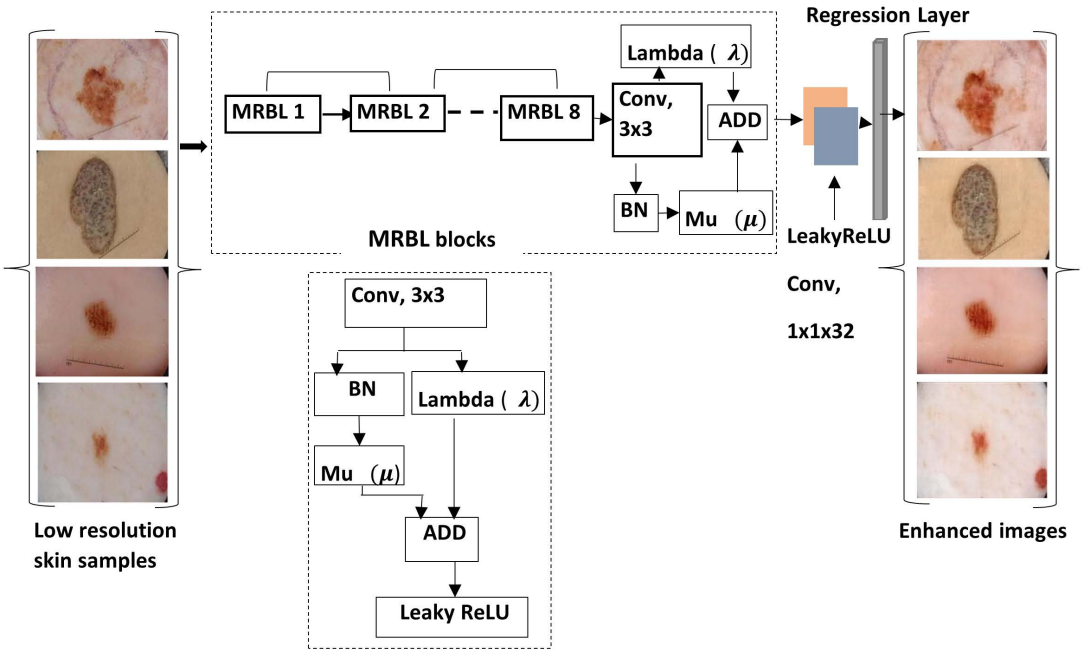



Figure 4.5: The proposed MCACNN model's design.

It receives feature maps as input and generates them to pass on to the next MRBL module. The description of MRBL how it works is given as follows:

The network receives an image patch $I(x, y)$ of size $256 \times 256 \times 3$, defining the spatial region to view all high-level features of an image. In the MRBL block, a convolutional operation applies 32 convolutions of size $3 \times 3$ over an image patch with a different dilation rate. Dilation rate for each convolution layer $Conv_i, \ldots, Conv_N$, where $N = 8$ is set as $2, 4, 6, 8, 16, 32, 64$ and $128$. It increases with the successive use of the convolution layer in the network. The network systematically uses dilation convolutions to aggregate the contextual information without systematically degrading the image's resolution. The dilation rate embedded in each convolution layer helps to receive a wider view of the feature map. Moreover, convolution operations extract high-level to low-level features by applying different dilation rates and filters. In a convolution operation, padding is also set to zero to maintain the same size of feature maps as input images. Padding inserts extra bits to the input image to adjust the size

imbalance between an input image and a convolutional filter.

Further, adaptive normalization $\mu$ layers adjust the strength of batch normalization (BN) branches, and adaptive normalization $\lambda$ adjust the strength of identity branches. These layers help strengthen the batch normalization and identity normalization by backpropagating the derivative of the loss function. The output of both $\mu$ and $\lambda$ scaling layers is combined and passed to an activation function, leaky ReLU. This activation function filters information computed by the previous layer that propagates in a forward direction through the network. It performs an element-wise operation on input and accepts input if it is positive. Otherwise, it changes to a positive value by multiplying it with positive scalar '$a$' if negative.

$$\text{leakyReLU} = \begin{cases} a * x, & x < 0, \\ x, & x \geq 0 \end{cases} \tag{4.10}$$

The value of scalar '$a$' is set as '$0.2$' for negative values. Each MRBL block works in a similar fashion employing this sequence of layers. The last block of the context aggregation module employs a convolutional layer with filters of size $3 \times 3$ without dilation rate. It is cascaded by batch normalization, and adaptive normalization $\mu$, and $\lambda$ layers. The output from both layers combined using additional layers and passed to the leaky layer. Finally, a convolutional operation uses 32 filters of size $1 \times 1$ followed by a regression layer to reconstruct the output $K(x, y)$. The regression layers utilize mean square error (MSE) function as loss function to determine the learnability of the aggregation operation as:

$$\text{Loss} = \sum_{\text{i=1}}^{\text{R}} \frac{(\text{I}_\text{i} - \text{K}_\text{i})^2}{\text{R}} \tag{4.11}$$

In this loss function, the mean square error between the target image $I$ and the network's predicted image $K$ is computed for $R$ responses.

### 4.3.2 Network training and hyperparameter settings

To generalize the learning capability of the MCACNN effectively, it is trained on skin cancer image samples. Adaptive moment estimation (ADAM) [144] optimizer algorithm was used to train the network for $20$ epochs (fine-tuned) with a mini-batch size of $16$. The network training parameters, weights, and bias are updated using a momentum gradient descent algorithm to minimize loss values. It calculates an adaptive learning rate for every network parameter instead of stochastic gradient descent, which updates all weights at once. In the original gradient descent algorithm, updating the network in the direction of the negative value of the loss function is given as follows:

$$\theta_{i+1} = \theta_i - \alpha \nabla L(\theta_i) \tag{4.12}$$

The learning rate, $\alpha > 0$ is taken as '$0.01$', $\theta$ is the parameter vector, and $L(\theta)$ is the loss function for iterations $i$. The stochastic gradient method evaluates the gradient of the loss function, $\nabla L(\theta)$ across all training sets. On the other hand, the ADAM optimiser adds a 'momentum' term to the parameter gradients and their squared values, as well as an element-wise average of them. The above equation can be rewritten as:

$$m_i = \beta_1 m_i - 1 + (1 - \beta_1) \nabla L(\theta_i) \tag{4.13}$$

$$v_i = \beta_2 v_i - 1 + (1 - \beta_2) [\nabla L(\theta_i)]^2 \tag{4.14}$$

The optimizer used moving average to update the network as:

$$\theta_{i+1} = \frac{\theta_i - \alpha m_i}{v_i + \epsilon} \tag{4.15}$$

The gradient decay factors $\beta_1$ and $\beta_2$ were set to '$0.95$', respectively and squared gradient decay factors set as '$0.99$' while the learning rate $\alpha$ was set at '$0.03$'.

## 4.4 Experimental Results

In this section, the quantitative results of the MCACNN model together with the IA-HR method are presented on the skin cancer dataset. First, detail of the dataset, evaluation metrics, and results of the IA-HR method are discussed. Then the outcome of the MCACNN model is explained. The proposed system is compared against other denoising techniques to analyze its effectiveness.

### 4.4.1 Datasets

The benchmark skin cancer dataset ISIC 2019 [139, 16, 138] is used in the experimental evaluation of this method. It contains 25,331 dermoscopic training and 8239 test images collected from different patients as illustrated in Table 4.1. Several diagnostic categories of lesions are present in the ISIC 2019 training dataset such as Actinic keratoses (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC). The original images are RGB of size $600 \times 450$, resized to $300 \times 300$ using the bilinear interpolation method. This technique effectively rescues the image's size maintaining aspect ratio so that information does not get lost during the reconstruction process. The network training is performed using training images and then applied on 8239 test images to estimate the contextual information. The MCACNN model learns contextual representation of each type of skin lesion sample from training set so that it can effectively estimate it for each lesion sample present in the test set.

Table 4.1: Distribution of different types of lesion samples in a training set.

| **Dataset** | **AK** | **BCC** | **BKL** | **DF** | **MEL** | **NV** | **VASC** | **Total** |
|---|---|---|---|---|---|---|---|---|
| Training | 867 | 3323 | 2624 | 239 | 4522 | 12875 | 881 | 25331 |
| Test | —- | —- | —- | —- | —- | —- | —- | 8239 |

### 4.4.2 Image quality evaluation metrics

The performance of the proposed methods is measured based on five image quality metrics, peak signal-to-noise ratio (PSNR), mean square error (MSE), structural similarity (SSIM) index [145], blind image spatial quality evaluator (BRISQUE) [146], and a natural image quality evaluator (NIQE) [147]. These metrics were commonly used in the literature [148] to assess the enhancement quality of the images. A PSNR is the ratio of the maximum power of the image to the peak noise error that affects the fidelity of the image. Its value should be high to ensure better image quality. In this case, R is set to 255.

$$\mathrm{PSNR} = 10\log_{10}\frac{\mathrm{R}^2}{\mathrm{MSE}} \tag{4.16}$$

The SSIM metric quantifies the quality loss associated with the hair removal during an image's processing. It measures three properties of an image, such as structure $s(x, y)$ comprises of a pattern of pixel intensities, luminance $l(x, y)$, and contrast $c(x, y)$ where x,y are pixels of an image $I(x, y)$ in x-axis and y-axis. This metric closely perceives structure as humans do, thus, it is significant to measure for lesion images because structural information is very important to determine lesion type.

$$\mathrm{SSIM(x,y)} = [\mathrm{l(x,y)}^{\alpha}, \mathrm{c(x,y)}^{\beta}, \mathrm{s(x,y)}^{\gamma}], \quad \alpha, \beta, \gamma = 1 \tag{4.17}$$

Next is the BRISQUE is a good image quality evaluator to measure different types of distortions that are present in the different types of lesions. Finally, a NIQE also measures arbitrary distortions which can be perceived subjectively.

### 4.4.3 Results of IA-HR method

The visual quality of digital images is harshly degraded by noise artifacts, reducing the precision and accuracy of image examination and interpretation operation. The

low lighting conditions, limited resources to capture medical images through clinical instruments lead to high noise variations in lesion images. The first significant issue handled is the removal of hairlines using the IA-HR method. It applies morphological operators to detect hairlines, and a new image is constructed through a remapping procedure while replacing hair pixels with neighboring pixels' intensities. Figure 4.6 displays the stepwise output of the IA-HR method that accepts raw images (Figure 4.6a) containing hairlines and produces cleaned images.

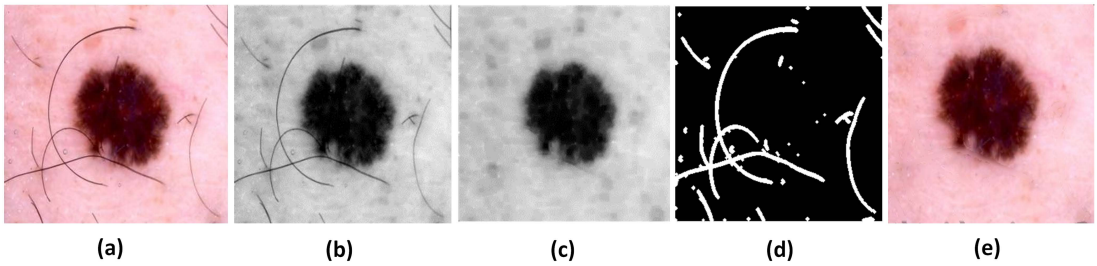



Figure 4.6: Results of IA-HR method: (a) Original images (b) CIE L*a*b color space (c) Dilation, erosion and closing operation (d) Hair mask (e) Output image.

This algorithm firstly converts RGB color space to $CIE\ L*a*b$ (Figure 4.6b), then applied dilation and erosion morphological operations on an input image to close irrelevant portions of the lesion and smooth contours (Figure 4.6c). After that, subtraction is performed between the original and output images from the previous step to find varying intensity levels. Next, a threshold function with value, T= 25 (fine-tuned) is applied to differentiate the lesion from the background that generates a binary mask highlighting hairlines (Figure 4.6d). The final step is to fill detected hair pixels with the intensity of neighboring pixels either belonging to lesion or background skin area (Figure 4.6e). Additionally, in Figure 4.7, the output of more sample images using the IA-HR method is shown, which signifies that the proposed noise removal method can eliminate thin and light hairlines from lesion images.

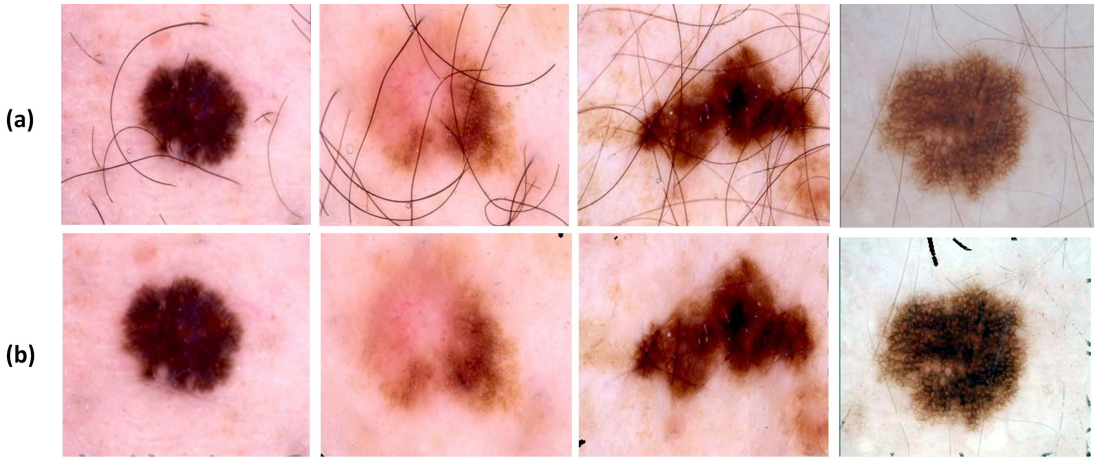
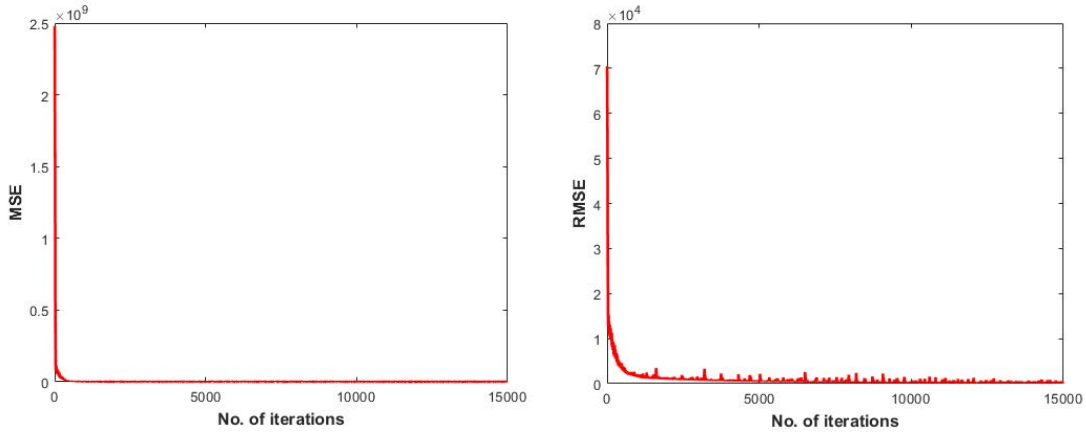



Figure 4.7: Some output examples using the IA-HR method.

### 4.4.4 Results of MCACNN model

The MCACNN model is trained on the output images produced by the IA-HR method. The model learns by predicting how well it fits the training data through a loss function. Through the loss function, the prediction error is reduced. The graphs in Fig. 4.8 display the MSE and RMSE loss computed by the MCACNN model during the training process. Standard deviation of residuals is known as Root Mean Square Error (RMSE). A residual refers to how far the data points are from the regression line, and a RMSE refers to how widely spread out the residuals are. As a result, it tells how concentrated the data is around the line of best fit. Over the number of iterations, the loss decreases continuously resulting into less error.

Table 4.2 illustrates the statistical calculation and comparison of image quality metrics for each lesion type using CLAHE, bilateral filtering, and MCACNN. The value of PSNR and SSIM should be high, BRIS and NIQE should be low for better image quality. It is seen in Table 4.2 that MCACNN obtain a higher value of PSNR on BCC, BKL, DF, MEL, NV, and VASC as 45.08, 45.10, 42.49,44.35, 40.33, 40.46, and 44.94, respectively. In addition to this, BRIS and NIQE are lower in the case of MCACNN as compared to CLAHE and the bilateral filtering approach.




(a) MSE Loss (b) RMSE Loss

Figure 4.8: Loss prediction by MCACNN during training process.

Table 4.2: Statistical performance analysis on each sample of lesion type.

| **Lesion Type** | **CLAHE** | | | | **Bilateral Filtering** | | | | **MCACNN** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **PSNR** | **BRIS** | **SSIM** | **NIQE** | **PSNR** | **BRIS** | **SSIM** | **NIQE** | **PSNR** | **BRIS** | **SSIM** | **NIQE** |
| AK | 16.34 | 31.24 | 0.800 | 5.50 | **45.08** | 55.21 | **0.995** | 5.95 | 33.59 | 29.20 | 0.946 | 5.47 |
| BCC | 21.21 | 20.46 | 0.742 | 3.75 | 42.01 | 45.91 | 0.995 | 5.01 | **45.10** | **17.50** | **0.999** | **3.67** |
| BKL | 18.28 | 20.49 | 0.910 | **3.78** | 41.79 | 34.27 | 0.980 | 4.89 | **42.49** | **11.98** | **0.999** | 3.83 |
| DF | 19.86 | 25.71 | 0.942 | 4.22 | 23.42 | 48.06 | **0.991** | 5.25 | **44.35** | **23.22** | 0.972 | **3.99** |
| MEL | 23.52 | 23.40 | 0.750 | 4.58 | 35.77 | 52.57 | 0.989 | 5.61 | **40.33** | **21.40** | **0.998** | **4.44** |
| NV | 16.30 | 19.14 | 0.744 | **3.94** | 29.45 | 38.56 | 0.996 | 4.64 | **40.46** | **15.96** | **0.998** | 4.02 |
| VASC | 16.97 | 43.74 | 0.892 | 4.45 | 22.10 | 60.01 | 0.970 | 5.54 | **44.94** | **40.84** | **0.987** | **4.37** |

The perceptual quality analysis of various lesion types is shown in Figure 4.9. Contextual information from original images is estimated to improve the contrast of processed images lost during hair removal. As compared to other contrast enhancement techniques, such a contrast limited adaptive histogram (CLAHE) [149] and bilateral filtering [150], the output images produced by MCACNN have higher contrast and resolution than the original images and IA-HR processed images. It is seen in the results that the CLAHE contrast enhancement method makes the lesion image brighter but changes color which is not good because color is a very important component in distinguishing lesion types. Similarly, bilateral filtering is not an ideal choice for skin cancer data because it only smoothes the images and does not produce sharp and high-resolution images.

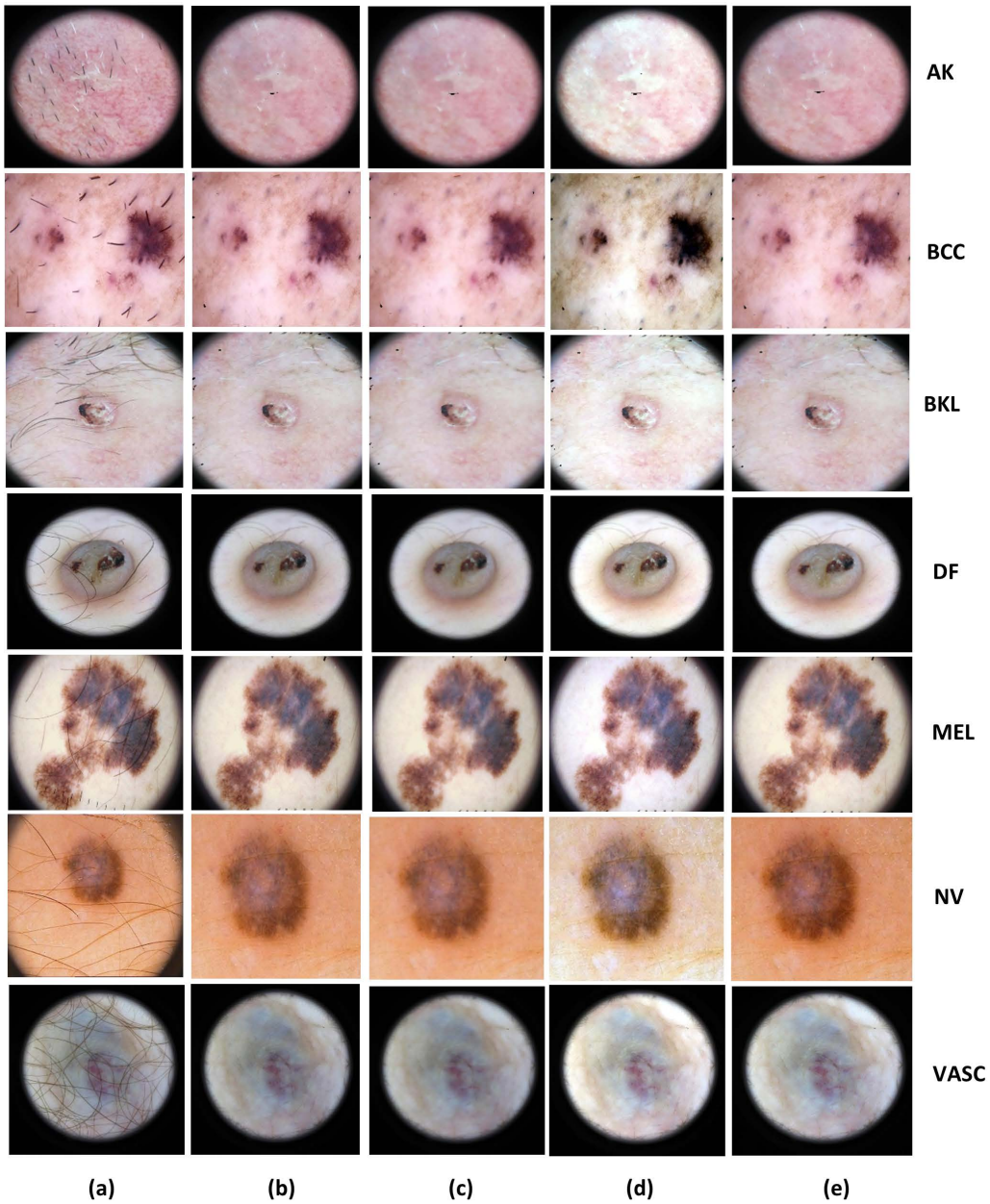



Figure 4.9: Perceptual quality analysis on seven types of lesions using: (a) Original (b) IA-HR (c) Bilateral filtering (d) CLAHE (e) MCACNN methods.

The MCACNN outperformed CLAHE and bilateral method by producing high-quality images and achieving a good score of quality metrics. These methods are applied to the output images produced by the IA-HR method to improve color and resolution which was lost during the hair occlusion process. Another advantage of using MCACNN

is that it is trained for images to estimate the color and contextual information of each lesion type and then applied on the test images for their improvement which saves time in processing each image unlike CLAHE and bilateral.

In addition to this, the average values of four metrics on the test set are as given in Table 4.3 that illustrates the MCACNN model outperformed other methods. The SSIM ranges from 0 to 1, and 1 means perfect match between the processed image and original image's contrast, structure, and resolution. In our case, the MCACNN model achieved average SSIM value '0.9976' which is better than bilateral and CLAHE method. Moreover, the average value of PSNR is higher and BRIS and NIQE are lower in the case of MCACNN model as compared to CLAHE and the bilateral filtering approach.

Table 4.3: Statistical performance analysis on test set.

| **Method** | **PSNR** | **BRIS** | **SSIM** | **NIQE** |
|---|---|---|---|---|
| CLAHE [149] | 22.16 | 34.43 | 0.6862 | 4.18 |
| Bilateral filtering [150] | 25.22 | 47.40 | 0.9529 | 5.18 |
| MCACNN (proposed) | **44.37** | **34.15** | **0.9976** | **4.01** |

The ANOVA test is performed between MCACNN, CLAHE and bilateral method based on different image quality metrics in Figure 4.10. In these graphs, the centre red mark indicates median of all values for the respective parameters on the test set and notches represents the minimum and maximum range of parameter's value varies for each image. The MCACNN model achieved high median value for SSIM and low for SSIM and BRISQUE as compared to other two methods. It indicates that MCACNN model is effective in gaining contrast, sharpness, tone reproduction and resolution of processed images by IA-HR method.

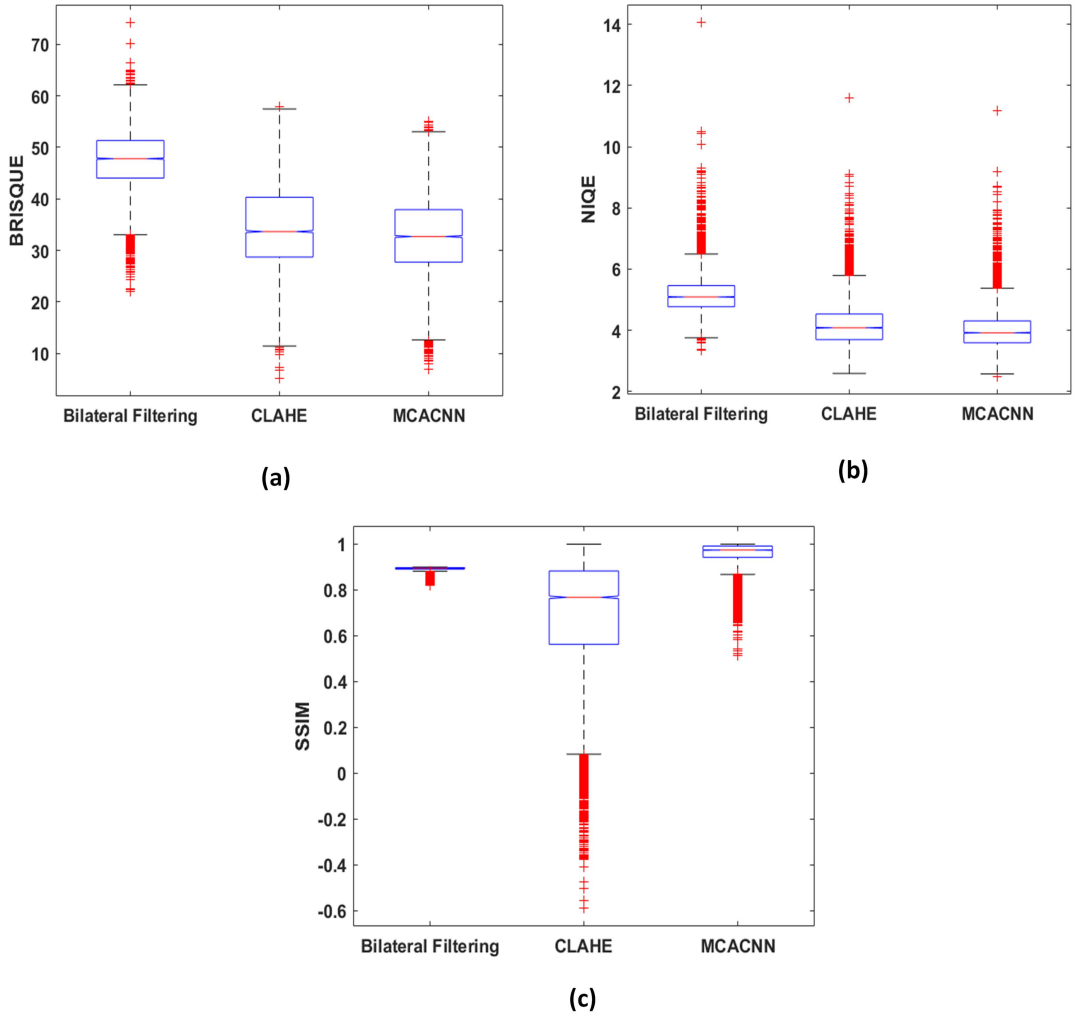



Figure 4.10: Performance analysis using ANOVA test based on (a) BRISQUE (b) NIQE and (c) SSIM parameters on test dataset.

## 4.5 Comparison with Other Methods

To prove the robustness of the proposed method, a comparison is conducted with other state-of-the-art methods. There are very limited latest studies that worked on the concept of hair removal from skin cancer images like [151, 152, 153]. These studies considered a small set of images in the test set, ignoring very dense hairy images, in contrast, a large dataset with very complex images is adopted in research study. A comparison between the proposed method and a few more current state-of-the-art methods are illustrated in

Table 4.4. It is analyzed from the results that the proposed approach obtained a higher SSIM and PSNR comparatively on different datasets adopted by each study.

Table 4.4: Comparison with few other state-of-the-art studies.

| **Method** | **Year** | **Mean SSIM** | **Mean PSNR** |
|---|---|---|---|
| T. Lee (Dull Razor) [57] | 1997 | 0.890 | 29.15 |
| Q. Abbas [55] | 2011 | 0.867 | 26.08 |
| A. Huang [154] | 2013 | 0.851 | 27.00 |
| M.T.B Toosi [53] | 2013 | 0.864 | 26.57 |
| F. Xie [58] | 2015 | 0.921 | 33.09 |
| P. Bibloni [155] | 2017 | 0.885 | 29.78 |
| L. Talavera [153] | 2020 | 0.926 | 35.13 |
| M. Attia [151] | 2020 | 0.930 | — |
| W. Li [152] | 2021 | 0.599 | — |
| Our Method | 2022 | **0.997** | **44.73** |

Note: Best performance shown in bold.

Another comparison is conducted on a different dataset named H13Sim-data [156]. The given studies in Table 4.5 are tested on this same dataset. First, to fairly illustrate the proposed method's performance, the IA-HR method was applied to eliminate hairs from the 13 benchmark images, as displayed in Figure 4.11 and 4.12. Then the MCACNN model was used to increase the resolution and smooth the images. The results in Table 4.5 show that the MCACNN approach achieved high SSIM and PSNR compared to other methods, which indicates high-quality images whereas, the MSE of Xie. et al. [58] is better than our method.

Table 4.5: Comparison with other state-of-the-art methods on H13Sim-data.

| **Method** | **Mean MSE** | **Mean SSIM** | **Mean PSNR** |
|---|---|---|---|
| Xie [58] | **16.792** | 0.975 | 36.968 |
| Abbas [55] | 133.351 | 0.906 | 27.619 |
| Lee (Dull Razor) [57] | 51.118 | 0.956 | 32.486 |
| Huang [154] | 44.494 | 0.951 | 32.594 |
| Toossi [53] | 150.676 | 0.891 | 26.943 |
| Hair Removal [156] | 61.334 | 0.917 | 30.879 |
| Giuliana [157] | 59.626 | 0.838 | 31.384 |
| Our Method | 87.262 | **0.980** | **37.864** |

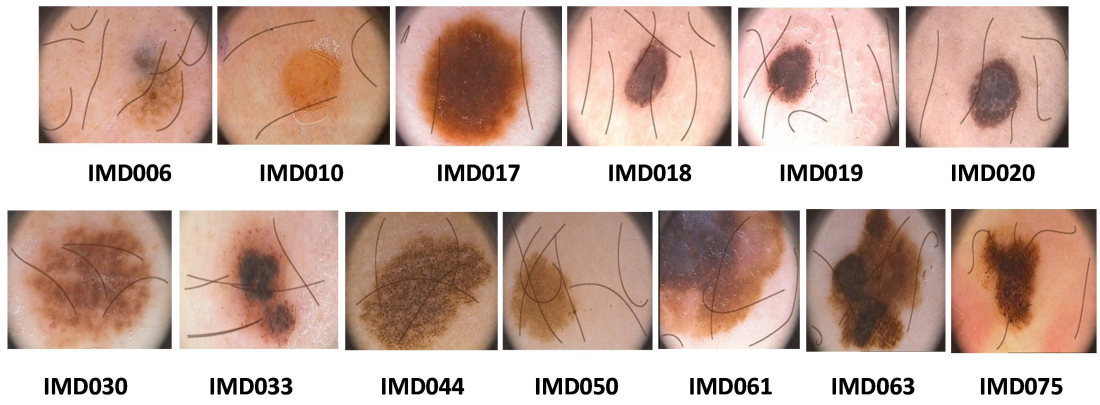



Figure 4.11: Original image samples of H13-Sim dataset.

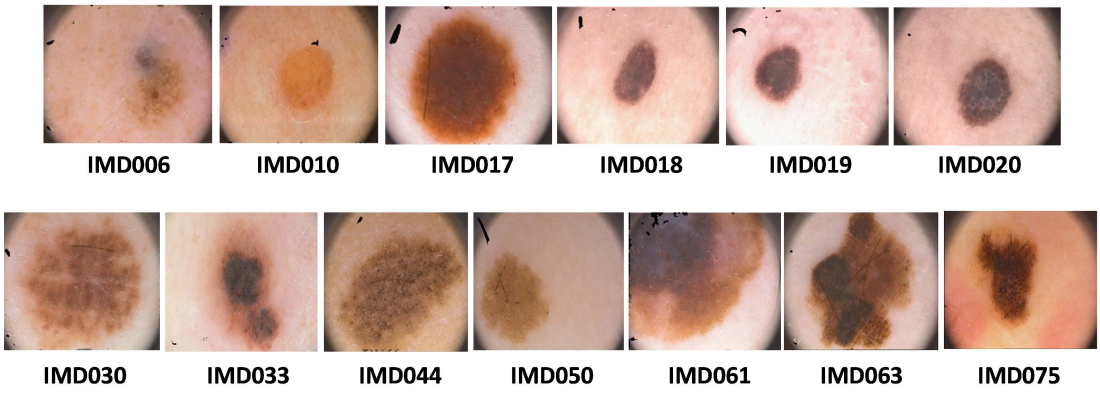



Figure 4.12: Output images using the proposed method on H13-Sim dataset.

## 4.6 Conclusion

Skin lesion images are highly noisy and suffer from poor color contrast, low illumination, and resolution. The primary objective of pre-processing stage was to remove hairlines artifacts from lesion images and improve their contrast after the hair occlusion process. Thus, the proposed methods organized into two steps; for the hairline's elimination process, a novel algorithm called IA-HR was designed to employ morphological operators to detect hair pixels and fill those regions using an average intensity interpolation scheme. At the second stage, the MCACNN was utilized to enhance the color contrast of processed images by extracting rich contextual information and covering a broader scale of data. The design of both methods is discussed, and results

were evaluated based on a few image quality metrics such as PSNR, BRISQUE, NIQE, and SSIM. It is concluded that the IA-HR method is efficient to remove light and thin hairlines from the lesion images and the context aggregation network was effective in producing a high contrast and smooth images. The next chapter discusses the lesion segmentation phase of the proposed system's design to extract lesion information.

# Chapter 5

# Lesion Segmentation

## 5.1 Introduction

The lesion segmentation task is discussed in this chapter to address the research question, RQ3. The clinical tools used by medical experts in inspecting infected skin areas are biopsy, histopathological analysis, and dermoscopy. These manual analysis methods are very subjective, challenging, error-prone, and time-consuming. Thus, the automatic segmentation process allows a quantitative study of the clinical and structural parameters. In CAD, segmentation is the most critical step for the detailed analysis of lesion structure. Due to the heterogeneous appearance of the lesion area having different color, size, shape, location, and texture, object analysis can be quite difficult. The lesion characteristics should be analyzed for clear distinction of melanoma since benign moles are often confused with melanoma moles. A significant challenge in segmentation is localizing the precise boundary of lesions. As a result, techniques were presented for automatically segmenting infected areas from skin photographs to help clinicians better understand lesion patterns.

The choice between methods based on machine and deep learning has always been challenging for performing lesion segmentation. The goal is to design an efficient

and high-performing technique for segmenting lesion regions. The best approach is investigated by implementing different techniques. Initially, the experiments were done by implementing a machine learning-based optimised K-means using the firefly algorithm (FA) to extract lesion objects. Then few existing semantic segmentation networks based on deep learning were modified and fine-tuned on lesion data. These methods are discussed in more detail in Appendix A. It is found in the literature and from the experimental results of the K-means + FA algorithm that deep learning networks proved to be more efficient than supervised machine learning techniques. There are few challenges reported in the literature while using deep learning such as optimal design of network architecture, enough data for training, noise variations, and complexity of network. These issues are successfully handled in the proposed work by designing a new CNN frameworks which are less complex, efficient and process noise free skin cancer data. Two different novel architectures based on CNNs were developed to achieve accurate segmentation. The comparison of these networks was performed to determine the most effective segmentation method for lesion data. This chapter first discusses method 1 followed by method 2, the results of both networks are presented and compared. The methods discussed in this chapter are published in different journal articles as given below:

- R. Kaur, H. GholamHosseini, R. Sinha & M. Lindén, "Automatic Lesion Segmentation Using Dilated Convolutional Deep Neural Networks in Dermoscopic Skin Cancer Images", *BMC Medical Imaging*, vol.22,no.1,pp.1-13,2022. [Online]. Available: https://doi.org/10.1186/s12880-022-00829-y.

- R. Kaur, H. GholamHosseini & R. Sinha, "Skin Lesion Segmentation using an Improved Framework of Encoder-Decoder Based Convolutional Neural Network", *International Journal of Imaging Systems and Technology, Wiley Online Library*, 2022. [Online]. Available: https://doi.org/10.1002/ima.22699.

## 5.2 Method 1: Encoder-Decoder CNN (EDNet)

Convolutional neural networks are capable of producing high-quality object segmentation. The major challenge in segmentation networks is the reconstruction of images from feature maps. Thus, the segmentation task is more complicated than classification as it involves converting a vector to an image. In EDNet, DarkNet is used as a backbone unit to create a lightweight network with fewer learning parameters and is trained end-to-end for lesion segmentation. This network is comprised of two modules, contraction and expansion, as shown in Figure 5.1. In the contraction layers, input images are compressed into latent spaces, and the expansion layers reconstruct the output images from this latent space. It is designed to deal with the diverse characteristics of lesions.

### 5.2.1 Contraction module

The contraction or encoder module has a depth of four blocks consisting of multiple convolutional, batch normalization, PReLU and pooling layers in each block. The network accepts RGB lesion images of size $192 \times 256 \times 3$. A two-dimensional image represents pixels located in positions $[x, y]$ by a two-dimensional data of size $M \times N$:

$$[\mathrm{x}, \mathrm{y}] = [\mathrm{x}_1\mathrm{y}_1, \mathrm{x}_2\mathrm{y}_2, \ldots\ldots\ldots \mathrm{x}_\mathrm{M}\mathrm{y}_\mathrm{N}] \tag{5.1}$$

The 2-D feature extraction layer produces a feature vector by performing convolutions on image data. Two parameters, padding and stride, are manually set as '1' and '2' are applied to each convolutional layer which slides 256 kernels of size $3 \times 3$ over an input image. Border pixels are handled differently based on the padding parameter. It is sometimes necessary to add padding bits to image borders when the filter does not fit perfectly over an input image while convolving over it. The stride parameter controls the kernel's step size while it slides. For the stride of '2', filter moves two pixels at a

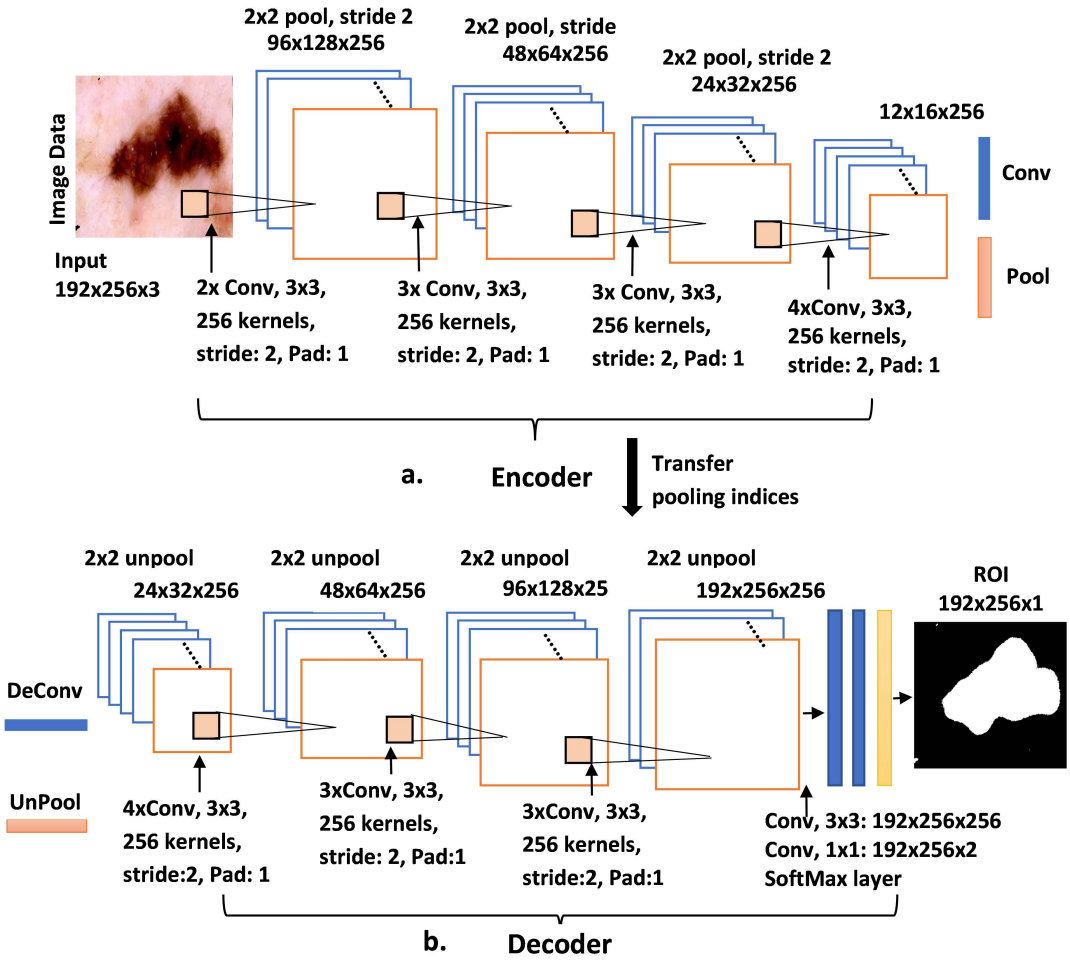



Figure 5.1: Architecture of EDNet Model.

time over an image. The following equation is used to perform a convolution function involving multiple kernels, padding, and stride:

$$\mathrm{Conv[x,y]} = \sum_{\mathrm{i=1}}^{\mathrm{s}} \sum_{\mathrm{j=1}}^{\mathrm{s}} (\mathrm{I_{x-i,y-j}} \times \mathrm{K_{i,j}},\ \mathrm{n_f}) \tag{5.2}$$

In the output feature map $Conv[x, y]$, pixels with positions $[x, y]$ are represented in the spatial domain, the input image is $I$, and the kernel size is $s$. Feature map $Conv[x, y]$ with multiple channels $n_f$ is produced from input image $I_{x,y}$ with kernel $K_{i,j}$.

A batch normalization layer regulates the learning process by standardizing each

mini-batches input to avoid overfitting problems. The mini-batch size is a variation of gradient descent that divides the training dataset into small batches to calculate the error of the network and update its coefficients. Samples are processed in batches before updating the network. For stabilizing the learning process and reducing epochs, it is usually used before the activation function PReLU. It is a piecewise linear function applied as the activation function in the network. By using the formula outlined below, it passes the output element onto the next layer directly if it is positive; otherwise, it multiplies all negative values by a fixed scalar $a$ and gives it to the next layer:

$$\mathrm{PReLU} = \begin{cases} \mathrm{x} \times \mathrm{a}, & \mathrm{x} < 0 \\ \mathrm{x}, & \mathrm{x} \geq 0 \end{cases} \tag{5.3}$$

Following the convolutional operation, the feature map becomes larger. As a result, the pooling operation is used to reduce the number of learning parameters as well as computations. Using the following formula, it calculates the maximum for each region in the feature map and divides it into pooling regions:

$$\mathrm{FM}[\mathrm{x}, \mathrm{y}] = \mathrm{f_p} \times \mathrm{C_{x,y}} \tag{5.4}$$

The output feature map is shown by the $FM[x, y]$ variable, $f_p$ is the pooling function, and the input feature map is shown by the $C_{x,y}$ variable. Feature maps $FM[x, y]$ present lesion features identified from the layers of the contractions. Segmentation maps have a low resolution, so spatial information was recovered by using the expansion module.

### 5.2.2 Expansion module

When successive pooling layers are applied in the encoder section, the down-sampling layers cause spatial information loss and may affect the localization of lesion boundaries.

In addition, the loss of the significant features leaves the network unable to distinguish between lesions and background regions. In order to develop rich feature patterns with various levels of complexity, the contraction module contains layers with high connectivity between them, whereas the expansion module comprises deconvolutional layers and unpooling operations to improve resolution and regain spatial information. The unpooling layers use the same pooling indices as those in the encoder section to generate a more sparse and enlarged activation map. In unpooling, activations are repositioned in original positions to capture image-oriented features. In this way, they contribute to the reconstruction of image structures. The output from different deconvolution layers is combined to create a flat feature vector which is converted into a probability map using the softmax function. Finally, pixel-wise binary classification determines whether a pixel is a background or a lesion. The layer descriptions for the contraction and expansion section can be found in Table 5.1.

Table 5.1: Detail of layers of EDNet Model.

| **Contraction section** | | | | **Expansion section** | | | |
|---|---|---|---|---|---|---|---|
| **Block** | **Layer** | **Kernel size, maps** | **Output size** | **Block** | **Layer** | **Kernel size, maps** | **Output size** |
| INPUT LAYER | Input | 192x256x3 | 192x256x3 | | | | |
| BLOCK 1 | Conv1_1 | 3x3, 256 | 192x256x256 | BLOCK 4 | UnPool4 | 2x2, 256 | 24x32x256 |
| | Conv1_2 | 3x3, 256 | 192x256x256 | | Deconv4_5 | 3x3, 256 | 24x32x256 |
| | Pool1 | 2x2, 256 | 96x128x256 | | Deconv4_4 | 3x3, 256 | 24x32x256 |
| | | | | | Deconv4_3 | 3x3, 256 | 24x32x256 |
| | | | | | Deconv4_2 | 3x3, 256 | 24x32x256 |
| | | | | | Deconv4_1 | 3x3, 256 | 24x32x256 |
| BLOCK 2 | Conv2_1 | 3x3, 256 | 96x128x256 | BLOCK 3 | UnPool3 | 2x2, 256 | 48x64x256 |
| | Conv2_2 | 3x3, 256 | 96x128x256 | | Deconv3_3 | 3x3, 256 | 48x64x256 |
| | Conv2_3 | 3x3, 256 | 96x128x256 | | Deconv3_2 | 3x3, 256 | 48x64x256 |
| | Pool | 2x2, 256 | 48x64x256 | | Deconv3_1 | 3x3, 256 | 48x64x256 |
| BLOCK 3 | Conv3_1 | 3x3, 256 | 48x64x256 | BLOCK 2 | Unpool2 | 2x2, 256 | 96x128x256 |
| | Conv3_2 | 3x3, 256 | 48x64x256 | | Deconv2_3 | 3x3, 256 | 96x128x256 |
| | Conv3_3 | 3x3,256 | 48x64x256 | | Deconv2_2 | 3x3, 256 | 96x128x256 |
| | Pool3 | 2x2, 256 | 24x32x256 | | Deconv2_1 | 3x3, 256 | 96x128x256 |
| BLOCK4 | Conv4_1 | 3x3,256 | 24x32x256 | BLOCK 1 | Unpool1 | 2x2, 256 | 192x256x256 |
| | Conv4_2 | 3x3,256 | 24x32x256 | | Conv1_1 | 3x3, 256 | 192x256x256 |
| | Conv4_3 | 3x3,256 | 24x32x256 | | Conv1_2 | 1x1, 256 | 192x256x2 |
| | Conv4_4 | 3x3,256 | 24x32x256 | | Softmax | ——— | 192x256x2 |
| | Conv4_5 | 3x3,256 | 24x32x256 | | Pixel_classi | Tversky Loss | 192x256x2 |
| | Pool4 | 2x2,256 | 12x16x256 | | | | |

### 5.2.3 Loss function

In order to calculate the error between the predicted and expected output, a Tversky loss function [158] ('$TL_c$') is implemented. Based on the segmented image $P$ and the ground truth image $T$, this loss function determines the overlapping ratio between the two images. The loss between $P$ and $T$ can be expressed as follows:

$$\mathrm{Loss} = \sum_{\mathrm{c=1}}^{\mathrm{c}} 1 - \mathrm{TL_c} \tag{5.5}$$

$$\mathrm{TL_c} = \frac{\sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{P_{cm}T_{cm}}}{\sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{P_{cm}T_{cm}} + \alpha \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{P_{cm}T_{c`m}} + \beta \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{P_{c`m}T_{cm}}} \tag{5.6}$$

Here, $c$ belongs to the number of classes i.e., c=2 (lesion or background) and $c$' defines the non lesion. In the formula, $P_{cm}$ is the probability of pixel $n$ be a lesion and $P_{c`m}$ is the probability of pixel to be non-lesion in the predicted image $P$. Similarly, $T_{cm}$ defined to '1' if pixels belongs to lesion and '0' if associated with background region i.e., non-lesion and vice-versa for the $T_{c`m}$. $N$ is the number of elements across the two dimensions of $P$. The weighing factors $\alpha$ and $\beta$ (which are set to 0.7 and 0.3, respectively) determine each class's contribution to the loss achieved by a false positive or false negative. For the imbalanced data, this higher value of $\beta$ produced good performance. In this case, the loss function addressed the class imbalance effectively when the background region was larger than the lesion. Because it controls loss values to better scale segmentation outcomes and produce fine edges, it has outperformed other loss functions such as cross-entropy and dice coefficient. This loss function is employed in EDNet for cancer images since the lesion occupies less area than the background.

### 5.2.4 Hyperparameters setting

Deep learning offers better results on large datasets, so the network is trained with augmented data to increase its performance. Rotation and scaling are used to increase

the amount of data with data augmentation. By using the early stopping function, the number of epochs is finalized to '30'. If the accuracy performance of a network isn't improving, it helps to monitor it. During the stochastic gradient descent (SGD) training algorithm, the network's weights and bias values are updated to minimise the loss function by making small progresses over each iteration in a direction of the negative gradient given as:

$$\theta_{l+1} = \theta_l - \alpha \bigtriangledown \mathrm{E}(\theta_l) \tag{5.7}$$

Here, $\alpha$ represents the initial learning rate, $l$ indicates the number of iterations, $\theta$ represents the parameter vector, and $E(\theta)$ represents the Tversky loss function. The SGD evaluates the gradient and updates the network parameters using a subset of the training data defined by a mini-batch size taken as '16' at each iteration. The complete pass of the training algorithm over the whole training data using mini-batches is defined as one epoch. The training algorithm is stochastic because it updates the parameters from the entire dataset using mini-batch size.

The issue with the standard SGD is that it oscillates along the path of steepest descent achieving optimum. This is because, in SGD the exact derivative is not calculated, it is always an estimated value over a mini-batches. It means the training of networks do not go in optimal direction always, because derivatives are noisy. Thus, addition of momentum works best by computing weighed averages to reduce oscillations as given:

$$\theta_{l+1} = \theta_l - \alpha \bigtriangledown \mathrm{E}(\theta_l) + \gamma \mathrm{E}(\theta_l - \theta_{l-1}) \tag{5.8}$$

In training parameters, $\gamma$ finds the gradient value of the last step's momentum value set as a step in the training process. The momentum $\gamma$ is added to reduce the oscillations along the steepest descent path towards optimum. Other hyperparameters, such as convolutional kernel size, number of kernels at each layer, and training parameters (see

Table 5.2), were finalised after several experimental trials.

Table 5.2: Optimized hyperparameters for EDNet model.

| **Parameters** | **Values** |
|---|---|
| Input Image size | $192 \times 256 \times 3$ |
| Batch size | 16 |
| Learning parameter, $\alpha$ | $1 \times 10^{-3}$ |
| Epochs | 30 |
| Momentum, $\gamma$ | 0.9 |
| Loss function, $E(\theta)$ | Tversky loss function |
| Optimizer | SGDM |

## 5.3 Method 2: DilatedSkinNet Deep Neural Network (DDNN)

The heterogeneous appearance of the lesion area in terms of color, size, shape, location, and texture makes the segmentation task very challenging. Therefore, there is a need for an efficient segmentation approach to assist dermatologists in understanding the nature and pattern of the lesion area. Thus, based on the concept of dilated convolutions, which was orginally proposed by Chen et al. [159, 96], a second method has been developed for segmentation tasks, called DilatedSkinNet. A wide field of view was achieved by using atrous rates in the DNN architecture. Research on the development of dilated DNN networks continued for different applications, and a high execution time was one of the significant challenges required to overcome. As shown in Figure 5.2, we designed a trainable DNN architecture consisting of 16 feature extraction layers with different dilation factors. The structure of DilatedSkinNet is interpreted in two primary steps: feature extraction and pixel classification i.e., segmentation. Using multiple scaling rates, the network used atrous convolutions to enlarge the filter's view. The pooling operation used in other semantic networks such as FCN, VGG16, and encoder-decoder

reduces the dimensional size of output feature vectors and the loss of information. In contrast, dilated convolutions expand the receptive field's view to obtain in-depth information without using pooling operation and are suitable for maintaining the spatial resolution of the segmented image.

The major contributions in the design of this method are listed as: (a) a new design of the CNN network organized into five building blocks for extracting low-level features to high-level features to allow auto segmentation on the whole image rather than processing a patch or subset, (b) incorporating an atrous dilation rate in each convolution feature layer to capture lesion and image at different scales, (c) a careful placement of many leakyReLU activation functions in each block instead of standard ReLU because the former has a small slope for negative values that makes the network learns faster and is more balanced. This section discusses the working of network including description of its various layers such as feature extraction, atrous convolutions, normalization, activation function, and segmentation.

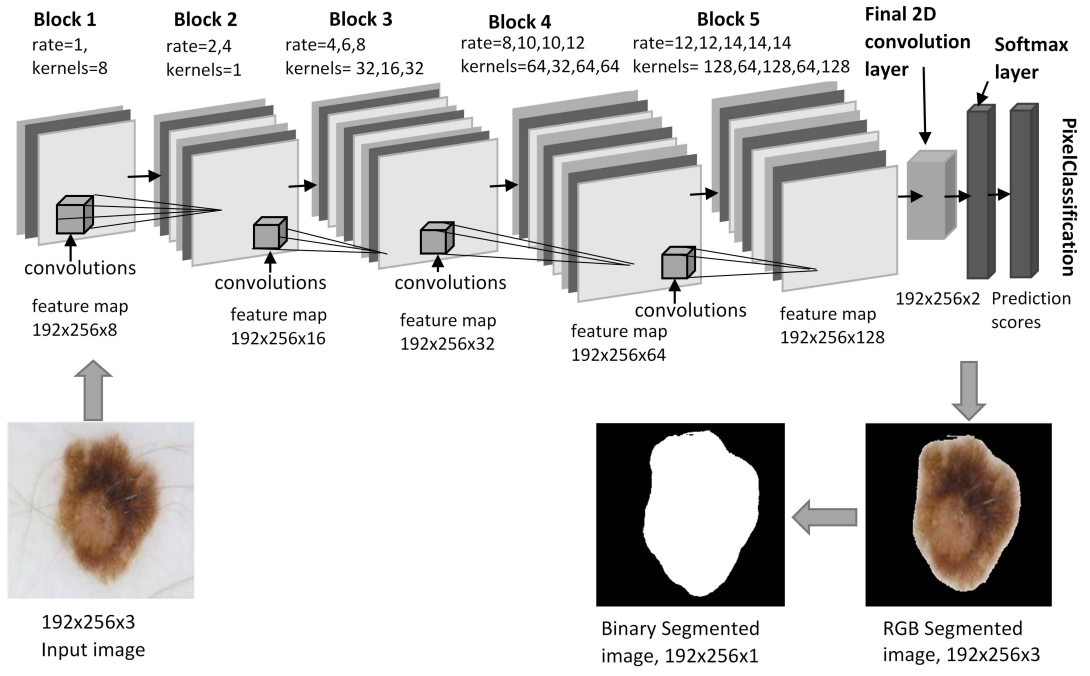



Figure 5.2: An overall structure of the DilatedSkinNet.

### 5.3.1 Feature extraction

The initial image is divided into multiple sub-blocks based on a specific size, and those sub-blocks compute features of the input image. The convolutional layer in the network is responsible for performing the feature calculation process. The feature calculation is a process where a small matrix called kernel $K_{i,j}$ or filter slides over an image $I_{x,y}$ and returns a feature map $C(I,K)_{x,y}$ by transforming image pixel's values as:

$$\mathrm{C(I,K)_{x,y}} = \sum_{\mathrm{i=1}}^{\mathrm{m_h}} \sum_{\mathrm{j=1}}^{\mathrm{m_w}} \sum_{\mathrm{k=1}}^{\mathrm{m_c}} \mathrm{I_{x+i-1,y+j-1,k}} * \mathrm{K_{i,j,k}} \tag{5.9}$$

A height, a width, and the number of channels in an input image $I_{x,y}$ are given by $m_h$, $m_w$, and $m_c$ respectively. The number of channels of a kernel $K_{i,j}$ needs to be the same as the channels of an input image. Stride and padding are two of the other parameters that need to be set artificially in the convolution layer. Because of the dilation rate set in each convolutional layer, these values are zero. The output feature maps are generated by applying multiple filters to each convolutional layer (see Table 5.3). The advantage of using multiple filters is that they can calculate many different features independently for each image, resulting in a rich feature map that subsequent layers can use. Two kernel sizes $3 \times 3$ and $1 \times 1$ were used in the convolutional layers.

### 5.3.2 Atrous or Dilation convolutions

Atrous convolutions with different dilation rates were used at each convolution layer to extract a rich feature map. The filter can utilize a larger field of view (the space of the input vector that a layer can see) so that it can assimilate a larger context. Therefore, it provides an efficient methodology to determine the best trade-off between correct localization and context absorption without increasing the computational parameters. With convolutional layers, features are extracted by sliding a filter over an image. We

used atrous convolutions instead of general convolutions, which are decisive for extracting more contextual information. The standard convolutional operation is described in (5.10) with the dilation rate '1'.

$$C[i] = \sum_{s=1}^{S} I[i + s.r] * K[s], \quad r = 1 \tag{5.10}$$

whereas, (5.11) describes the atrous convolution operation when the dilation rate is set more than '1'.

$$C[i] = \sum_{s=1}^{S} I[i + s.r] * K[s], \quad r \geq 2 \tag{5.11}$$

Rather than using the same dilation rate at every layer, different atrous rate is employed in each convolutional layers to avoid the 'gridding effect' that occurs due to the same dilation rate. At each successive 2-dimensional convolutional layer, the dilation factor increased by '2'. The Figure 5.3 represents the impact of using dilation rates '1', '2' and '4' on a filter's view over an input image.

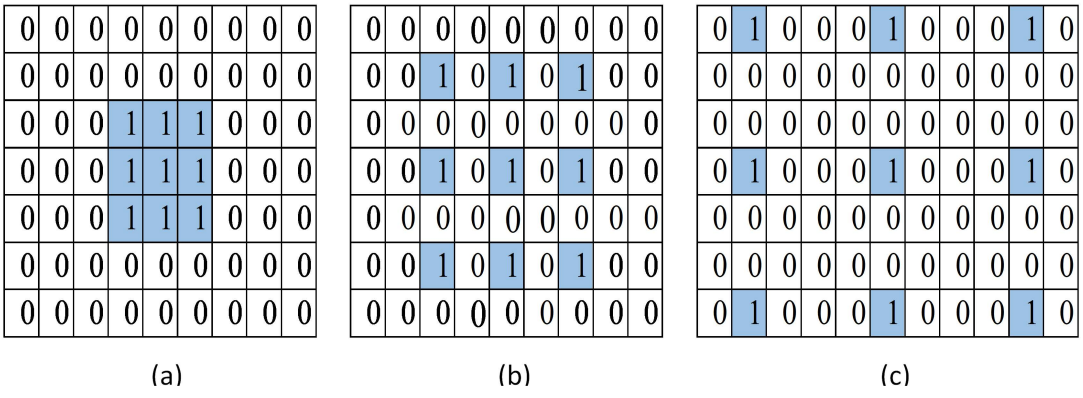



Figure 5.3: Standard convolution (a) rate=1, atrous convolution (b) rate=2 (c) rate=4

### 5.3.3 Data normalization

The batch normalisation is used to speed up the process of network learning using internal normalisation values. The term 'batch' refers to the group or set of data that

is processed at a time, whereas a 'mini-batch' means splitting the training dataset into small batches for the purpose of calculating network errors and updating network coefficients. Each input is normalized according to the mini-batch size in this layer. The batch normalisation layer performs a series of operations to normalise the data. In the first step, the standardisation process converts the input data batch so that its mean is $1$ and its standard deviation is $0$. The mean $(\mu)$ is computed using (5.12) by summing up all inputs $(x_i)$ of the batch (B) and divided by the total number of inputs (n). The resultant vector contains each input sample's mean value. The variance $(\sigma_B^2)$ is obtained by squaring the standard deviation of the input as given in (5.13). It is computed by taking the square of each input sample $(x_i)$ in the current batch (B) subtracted from the mean $(\mu_B)$.

$$\text{mean}, \mu_{\mathrm{B}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{x_i} \tag{5.12}$$

$$\text{variance}, \sigma_{\mathrm{B}}^2 = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{x_i} - \mu_{\mathrm{B}})^2 \tag{5.13}$$

In the following operation (5.14), the mean value of the current batch is subtracted from each input instance and divided by the square root of addition between variance $(\sigma_B^2)$ and smoothing term $\epsilon$. This term is set as '$0.00005$' to avoid division by a zero number in the calculations.

$$\hat{\mathrm{x_i}} = \frac{\mathrm{x_i} - \mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2 + \epsilon}} \tag{5.14}$$

Finally, the output of the batch normalization process is obtained by re-scaling $\gamma$ and offsetting $\beta$ of the input values using (5.15). These two parameters are learned during the training of the CNN network and should be optimal to ensure accurate normalization.

$$\mathrm{BN}_{\gamma\beta}(\mathrm{x_i}) \quad \text{or} \quad \mathrm{y_i} = \gamma \hat{\mathrm{x_i}} + \beta \tag{5.15}$$

To stabilise the learning process and reduce epochs, batch normalisation is usually applied between the convolutional layer and the leakyReLU layer. The batch normalisation occurs after each convolutional layer in order to reduce the internal covariate shift, which significantly improves the learning efficiency of the network.

### 5.3.4 Activation function

The activation function used in this network is leakyReLU that multiplies all negative values by a fixed scalar $s$ and applies the threshold function to each element in the input. If the function outputs a positive value, then it passes that value directly to the next layer; otherwise, it outputs the multiplied value by $s$ in the following formula:

$$\text{leakyReLU} = \begin{cases} s \times x, \ x < 0, \\ \\ x, \ \ x \geq 0 \end{cases} \tag{5.16}$$

In the DilatedSkinNet, the scalar '$s$' value is set as '$0.1$'. Activation function transforms the weighted sum of inputs from the node into the node's activation value.

### 5.3.5 Segmentation

The feature maps were flattened by using a $1 \times 1$ convolutional layer rather than fully connected layers at the end of the network. A SoftMax layer was then used to classify each image pixel into two defined classes, i.e. lesion and background. By applying SoftMax, the weighted sum feature values are normalized to probability scores between 0 and 1. As a result of this layer, each pixel was mapped to defined classes according to its probability. Based on the probability score generated by the SoftMax layer, the pixel classification layer created a categorical label (background or lesion) for each pixel. A loss function is also used to calculate the network's prediction error rate using this layer.

A network is trained for several iterations in order to minimize the loss. The structural detail of each layer used in the network with the layer's name, kernel size, number of filters, dilation rate and the size of output feature map is explained in Table 5.3.

Table 5.3: Architecture details of the DilatedSkinNet.

| **Block** | **Layer** | **Kernel size, Feature Maps** | **Atrous Dilation Rate** | **Output size** |
|---|---|---|---|---|
| Input Layer | Input | 192x256x3 | ——— | 192x256x3 |
| Block 1 | Conv1 | 3x3, 8 | 1 | 192x256x8 |
| Block 2 | Conv_21 | 3x3, 16 | 2 | 192x256x16 |
| | Conv_22 | 3x3, 16 | 4 | 192x256x16 |
| Block 3 | Conv_31 | 3x3, 32 | 4 | 192x256x32 |
| | Conv_32 | 1x1, 16 | 6 | 192x256x16 |
| | Conv_33 | 3x3, 32 | 8 | 192x256x32 |
| Block 4 | Conv_41 | 3x3, 64 | 8 | 192x256x64 |
| | Conv_42 | 1x1, 32 | 10 | 192x256x32 |
| | Conv_43 | 3x3, 64 | 10 | 192x256x64 |
| | Conv_44 | 3x3, 64 | 12 | 192x256x64 |
| Block 5 | Conv_51 | 3x3, 128 | 12 | 192x256x128 |
| | Conv_52 | 1x1, 64 | 12 | 192x256x64 |
| | Conv_53 | 3x3, 128 | 14 | 192x256x128 |
| | Conv_54 | 1x1, 64 | 14 | 192x256x64 |
| | Conv_55 | 3x3, 128 | 14 | 192x256x128 |
| | FinalConv | 1x1, 2 | — | 192x256x2 |
| | Softmax | — | — | 192x256x2 |
| OutPutMap | Pixel Classification | Cross entropy | loss function | 192x256x2 |

### 5.3.6 Model training

Loss function measures the difference between the prediction $P$ and the target score $T$ so that the network can be trained and optimized appropriately. The weighted cross-entropy loss [160] function was employed to measure the network's error as:

$$\mathrm{Loss} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{K}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{w_i T_{n_i} log(P_{n_i})} \tag{5.17}$$

A vector of weights determined by the network for each class is $w$, where $N$ is the number of observations, $K$ is the number of classes. By making small changes in the direction of optimization, stochastic gradient descent algorithm reduces the loss value by adjusting the weights and biases of the network.

$$\theta_{i+1} = \theta_i - \alpha \bigtriangledown L(\theta_i) \tag{5.18}$$

In this equation, $i$ is the number of iterations, $\alpha > 0$ is the learning parameter (set as $1 \times 10^{-3}$), $\theta$ is a parameter vector, and $\bigtriangledown L(\theta_i)$ is the gradient of the loss function. The algorithm evaluates the gradient at each iteration and updates parameters over a mini-batch set. The larger weight values can cause a network to be stuck into the local minima; thus, momentum term is added in the gradient descent algorithm to reduce the oscillations as follows:

$$\theta_{i+1} = \theta_i - \alpha \bigtriangledown L(\theta_i) + \gamma(\theta_i - \theta_{i+1}) \tag{5.19}$$

The following hyperparameters values as shown in Table 5.4 were set for network's training.

Table 5.4: Optimised hyperparameters for DilatedSkinNet.

| **Parameter** | **Values** |
|---|---|
| Input image size | $192 \times 256 \times 3$ |
| Batch size | 16 |
| Learning parameter, $\alpha$ | 0.01 |
| L2 regularization | 0.005 |
| Momentum, $\gamma$ | 0.9 |
| Epochs | 30 |
| Loss function $E(\theta)$ | weighted cross entropy |
| Optimiser | SGDM |

## 5.4 Results

The section presents qualitative and quantitative information derived from both methods of segmentation. The dataset and evaluation metrics are described first, followed by the experimental outcome of method 1 (EDNet) and method 2 (DilatedSkinNet). Then the performance of proposed networks is compared with traditional base networks such as SegNet, U-Net, and FCN. The final step is to compare the performances of each model with other state-of-the-art methods.

### 5.4.1 Datasets

These methods were evaluated on three benchmark datasets containing dermoscopic images of different skin cancer types like BCC, MEL, SCC, and NV including their ground truths that were used for validation, training, and testing purposes. The datasets were obtained from open-access dermatology repositories, ISIC archive [16, 139, 161, 162]. The ISIC 2016 dataset includes 900 training images and 379 testing images, the ISIC 2017 dataset contains 2100 training images, 150 validation images, and 366 testing images, and the ISIC 2018 dataset contains 2594 training images and 100 validation images. There were no external validation data available for the ISIC 2016 challenge. As a result, the ISIC 2016 training set was divided into training (630 images) and validation sets (270 images) in a 7:3 ratio. Similarly, the ISIC 2018 challenge did not provide any ground truth for its test data. Hence, 30% percent of ISIC 2018 training data was used for testing. The details of data available based on three years divided into different sets is given in Table 5.5.

The proposed approaches target dermoscopic images, which were produced by dermatoscopes. Available images were 8-bit with size ranges from $540 \times 722$ to $4499 \times 6748$ pixels. These images cannot be used directly for the network's training due to their large size as it increases computation cost. Thus, images were downsized to

Table 5.5: Details of the ISIC challenge data.

| Dataset name | Training | Validation | Test |
|---|---|---|---|
| ISIC 2016 | 630 | 270 | 379 |
| ISIC 2017 | 2000 | 150 | 366 |
| ISIC 2018 | 1816 | 100 | 780 |
| Total | 4446 | 520 | 1525 |

$192 \times 256 \times 3$ dimensions using the nearest-neighbour interpolation technique. The principle idea used in this resizing method was to have an original image as a reference image. A new rescaled image is constructed, resulting in a smaller size maintaining aspect ratio and resolution as of the original image. Data augmentation with rotation at an angle between $[15^o, 20^o]$, scaling and translation with factors $[-6, 5]$ was applied to enlarge the training dataset and to overcome the problem of underfitting or overfitting that occurs in neural networks owing to the insufficient availability of data.

### 5.4.2 Evaluation metrics

A set of performance metrics was used to evaluate the effectiveness of the segmentation technique. The first metric is accurate classification (ACC), which measures the percentage of pixels correctly classified out of all pixels. Based on the Dice coefficient (DICE), we can estimate the matching score between segmented output and ground truth. The Intersection of Lesion (IoU), also known as the Jaccard Index (JAC), defines intersection of the output lesion with a ground truth image. The worst to best score of ACC, DICE, and JAC ranges from 0 to 1, where their value close to '1' indicates best and '0' represents worst performance. The higher the value of these parameters, the better the segmentation.

$$\mathrm{ACC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{5.20}$$

$$\text{DICE} = \frac{2* \mid S_{G(x_i)} \cap S_{P(x_i)} \mid}{\mid S_{G(x_i)} \mid + \mid S_{P(x_i)} \mid} \quad \text{OR} \quad \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{5.21}$$

$$\text{IoU or JAC} = \frac{\mid S_{G(x_i)} \cap S_{P(x_i)} \mid}{\mid S_{G(x_i)} \cup S_{P(x_i)} \mid} \quad \text{OR} \quad \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{5.22}$$

The $S_{G(x_i)}$, $S_{P(x_i)}$ represents segmented results in ground truth and predicted $x_i$ image samples, respectively. Further, all these performance metrics relating to elements of the confusion matrix are computed from the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). When lesion pixels are classified correctly, they are classified as TP, otherwise they are classified as FN. On the other hand, if background pixels, which are non-lesion pixels, are correctly segmented as non-lesion pixels are deemed TN, otherwise FP.

### 5.4.3 Results of method 1: EDNet

The EDNet model is evaluated across three datasets (ISIC 2016-2018) containing separate training and testing images (see Table 5.5). Corresponding ground truth images were used as a reference to assess the model's performance. The first row in Figure 5.4(a) displays original dermoscopic images with their corresponding ground truths as shown in the second row of Figure 5.4(b). In the last row, Figure 5.4(c) demonstrates the segmented lesions generated by EDNet achieving a similar ROI corresponding to their ground truths. In Figure 5.5, the boxes shows that the EDNet achieved more than 90% ACC and 80% IoU index represented by median value on three sets. Furthermore, Figure 5.6 provides an additional illustration of performance of the EDNet by computing confusion matrix for the ISIC 2016-2018 test sets. The count of pixels correctly classified into two classes, i.e. lesion and background, is determined by the sum of diagonal positions (TP and TN). It indicates that there are fewer false positives in three datasets than true positives.

While designing EDNet, several key points have been considered to obtain the

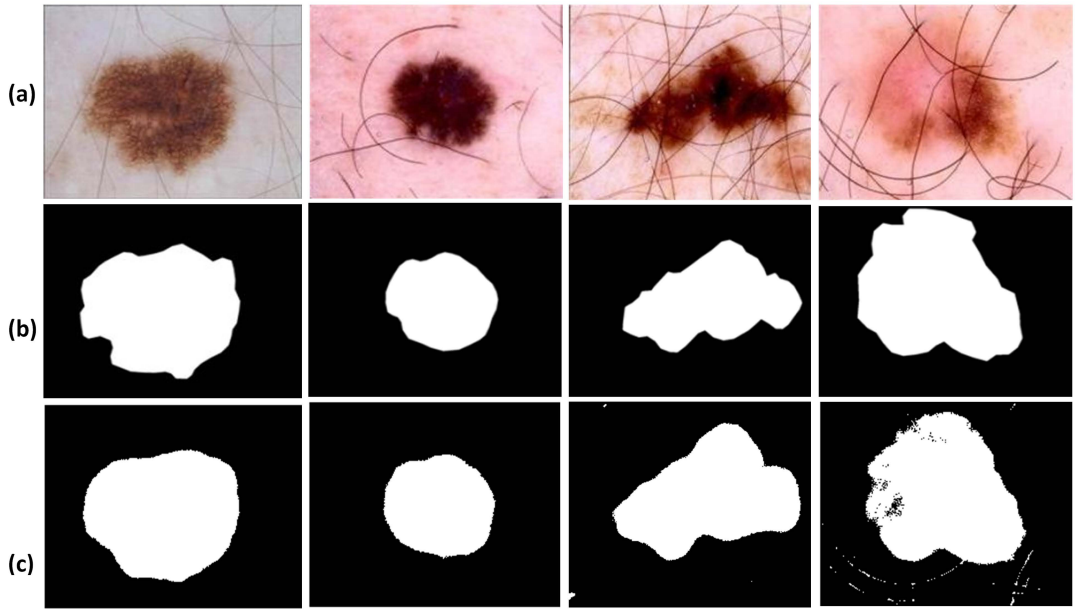



Figure 5.4: Qualitative results of the EDNet model: (a) original (b) ground truth (c) segmented images.

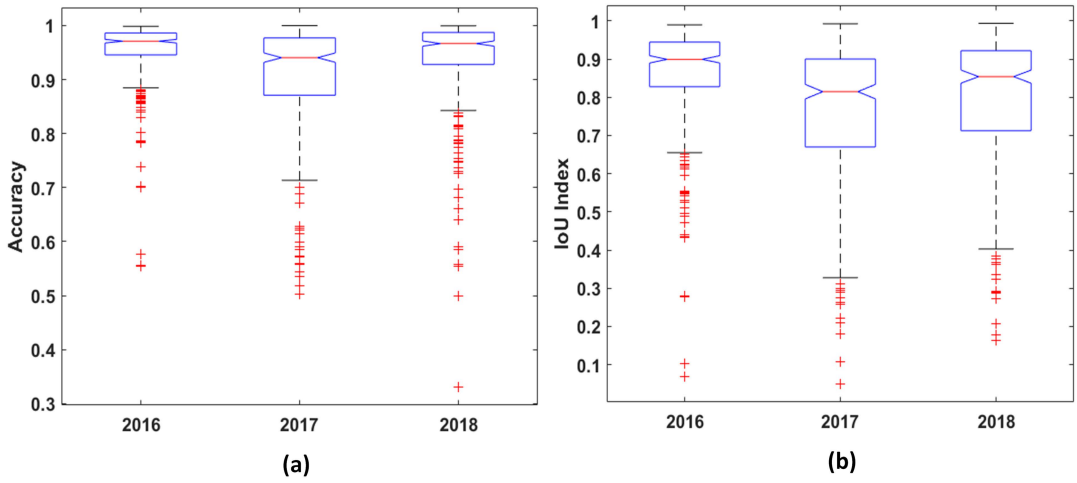



Figure 5.5: ANOVA graphs of EDNet for ACC and IoU parameters on ISIC 2016-2018 on test sets.

highest performance. After several experiments, the hyperparameters were finalised, and performance was affected by the choice of hyperparameters as described in Table 5.2 which were fine-tuned for EDNet. Thus, to show the impact of selecting different parameter values on the network's performance, the experiments were conducted on the

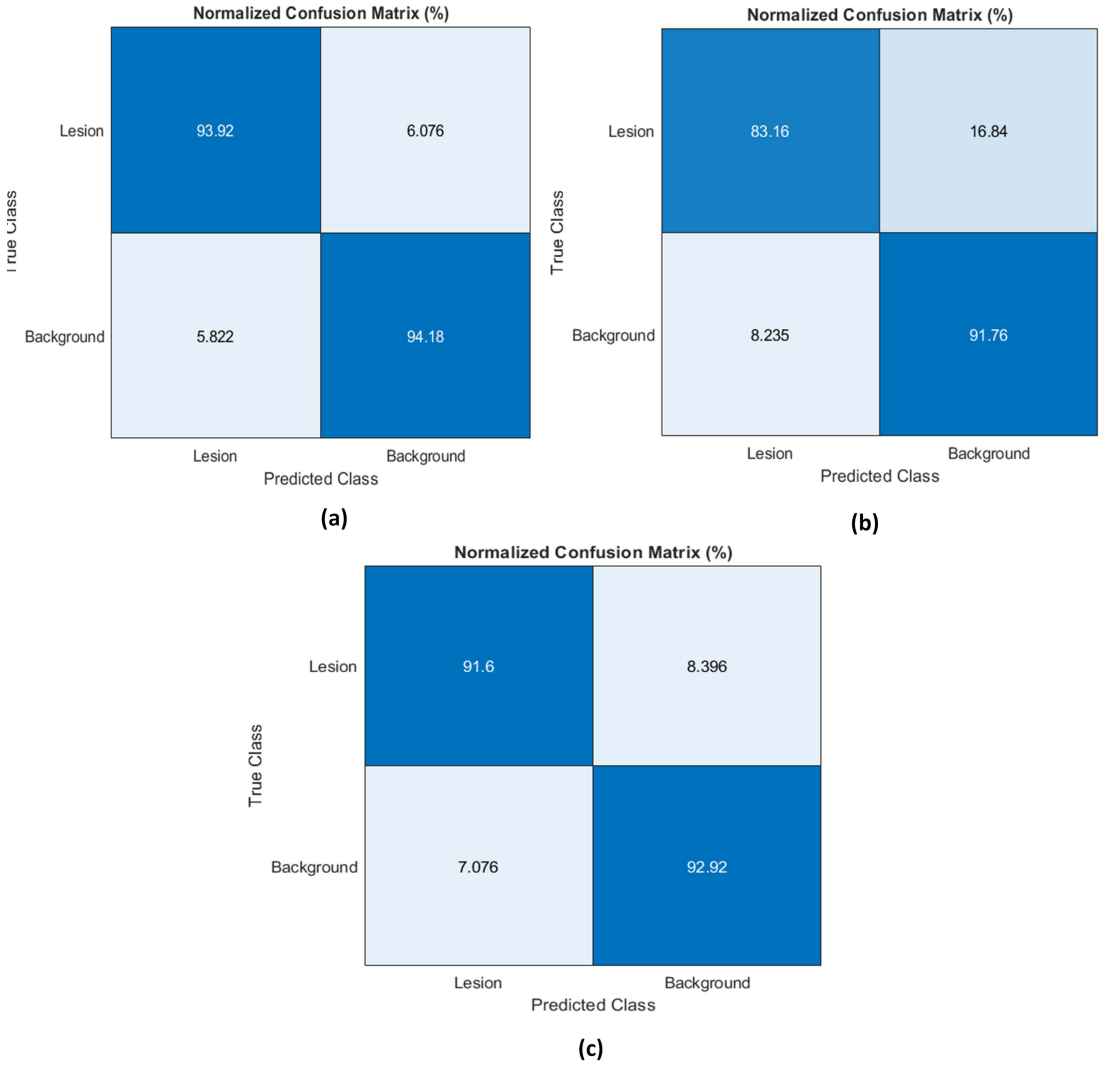



Figure 5.6: Confusion Matrix of EDNet for two classes, Lesion and Background on (a) ISIC 2016 (b) ISIC 2017 (c) ISIC 2018 test sets.

ISIC 2017 dataset that contains 2000 training, 150 validation, and 366 test images. The following parameters were selected with different values to observe the variations in the performance on test set.

- Network depth in terms of blocks of layers such as 2, 3, 4, and 5.
- Number of kernels such as 64,128, 256, and 512.
- Optimization algorithms such as SGDM, ADAM, and RMS.

- Loss functions like Cross entropy, DICE, Tversky loss.

As given in Table 5.6, 5.7, 5.8, 5.9 and Figure 5.7, it is analyzed that the network using 256 number of kernels, a depth of 4, SGDM optimization algorithm and Tversky loss function presented a higher performance on ISIC 2017 test set as compared to other choices.

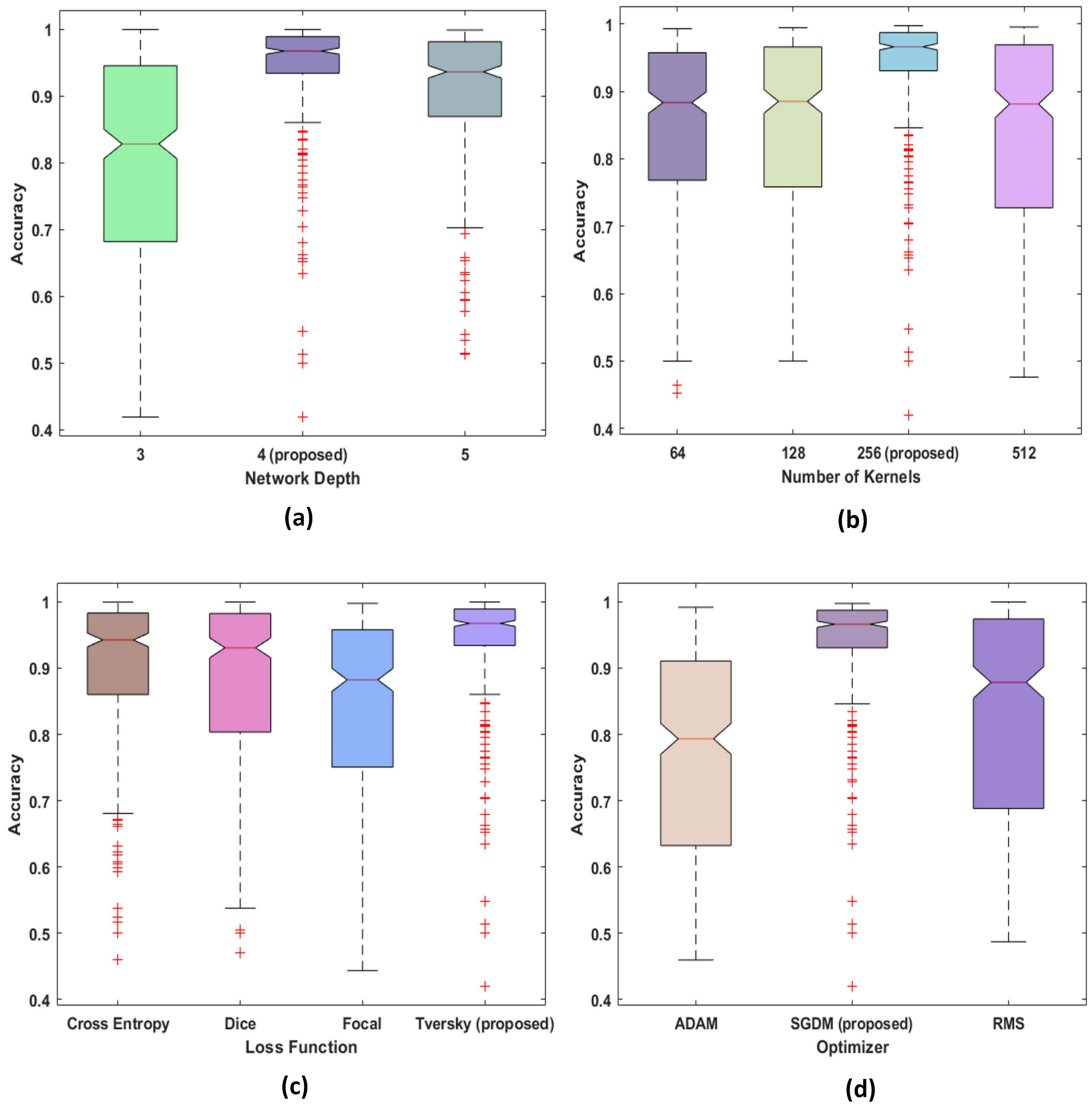



Figure 5.7: Accuracy analysis of EDNet on ISIC 2017 test set using different network's parameters: (a) Network's depth (b) Kernel size (c) Loss function (d) Optimization algorithm.

Table 5.6: Performance of EDNet on ISIC 2017 test set with different network depths.

| **Network's Depth in Blocks** | **ACC** | **IoU** | **DICE** |
|---|---|---|---|
| 3 | 0.776 | 0.665 | 0.566 |
| **4 (proposed)** | **0.874** | **0.770** | **0.869** |
| 5 | 0.876 | 0.791 | 0.595 |

Table 5.7: Performance of EDNet on ISIC 2017 test set with different number of kernels.

| **Number of Kernels** | **ACC** | **IoU** | **DICE** |
|---|---|---|---|
| 64 | 0.807 | 0.737 | 0.594 |
| 128 | 0.810 | 0.741 | 0.612 |
| **256 (proposed)** | **0.874** | **0.770** | **0.869** |
| 512 | 0.781 | 0.709 | 0.605 |

Note: The network depth is fixed as 4.

Table 5.8: Performance of EDNet on ISIC 2017 test set with different optimization algorithms.

| **Optimization Algorithm** | **ACC** | **IoU** | **DICE** |
|---|---|---|---|
| ADAM | 0.718 | 0.633 | 0.633 |
| RMSE | 0.763 | 0.687 | 0.600 |
| **SGDM (proposed)** | **0.874** | **0.770** | **0.869** |

Note: The fixed parameters are as 256 number of kernels, and 4 depth.

Table 5.9: Performance of EDNet on ISIC 2017 test set with different loss functions.

| **Loss Function** | **ACC** | **IoU** | **DICE** |
|---|---|---|---|
| Cross entropy | 0.872 | **0.786** | 0.622 |
| Dice loss | 0.837 | 0.762 | 0.634 |
| Focal loss | 0.811 | 0.719 | 0.513 |
| **Tversky loss (proposed)** | **0.874** | 0.770 | **0.869** |

Note: The fixed parameters are as 256 number of kernels, 4 depth, and SGDM optimization algorithm.

### Comparison with baseline models

The effectiveness of EDNet is demonstrated in Table 5.10 by comparing its performance with other existing popular networks such as SegNet, U-Net, and FCN. A fair comparison was conducted by fine-tuning these networks under the same parameters settings and datasets. The networks were trained on limited hardware available resources as GeForce RTX 2060. EDNet outperformed SegNet, U-Net, and FCN with higher values

on three test datasets. The primary advantage of the EDNet is that it involves fewer learnable parameters, making the network lightweight and fast compared to the given approaches. The ANOVA test based on the IoU index has also been used to analyze the performance difference between the EDNet and other baseline networks. The boxes represent the statistical differences between the means of IoU for different approaches. As seen in Figure 5.8, the EDNet achieved a higher median than others, indicating that it outperformed other given approaches.

Table 5.10: Performance comparison between EDNet and other existing networks on ISIC 2016-2018 test sets.

| **Methods** | **ISIC 2016** | | | **ISIC 2017** | | | **ISIC 2018** | | | **Parameters** |
|---|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **IoU** | **DICE** | **ACC** | **IoU** | **DICE** | **ACC** | **IoU** | **DICE** | |
| EDNet | **0.941** | **0.887** | **0.940** | **0.874** | **0.770** | **0.869** | **0.923** | **0.855** | **0.922** | **14 M** |
| FCN [92] | 0.918 | 0.817 | 0.851 | 0.810 | 0.614 | 0.693 | 0.870 | 0.672 | 0.779 | 423 M |
| SegNet [94] | 0.908 | 0.813 | 0.907 | 0.822 | 0.679 | 0.818 | 0.880 | 0.730 | 0.879 | 29 M |
| U-Net [95] | 0.854 | 0.798 | 0.832 | 0.764 | 0.687 | 0.696 | 0.842 | 0.793 | 0.815 | 31 M |

Note: The higher performance is marked in bold.

### Comparison with other related studies

The EDNet model was compared with the winning teams of the ISIC challenge and other recent studies based on ISIC datasets. The performance on ISIC 2016, 2017 and 2018 test sets having 379, 366, and 778 samples, respectively was compared based on the average IoU index. It can be seen in Tables 5.11, 5.12 and 5.13 that the EDNet model outperformed the given studies on the ISIC 2016, 2017, and 2018 datasets with an average IoU of 88.7% and 77.0%, and 85.5% respectively which is significantly higher than the other related studies.

The EDNet model outperformed other recent studies and winners of the ISIC challenge for the three respective years (2016, 2017 & 2018) with a higher IoU index. The network's framework is dense and complex to compute feature set for classifying image's pixels into the background or lesion objects. The results showed the capability of the network in segmenting skin lesions with higher accuracy and IoU index compared

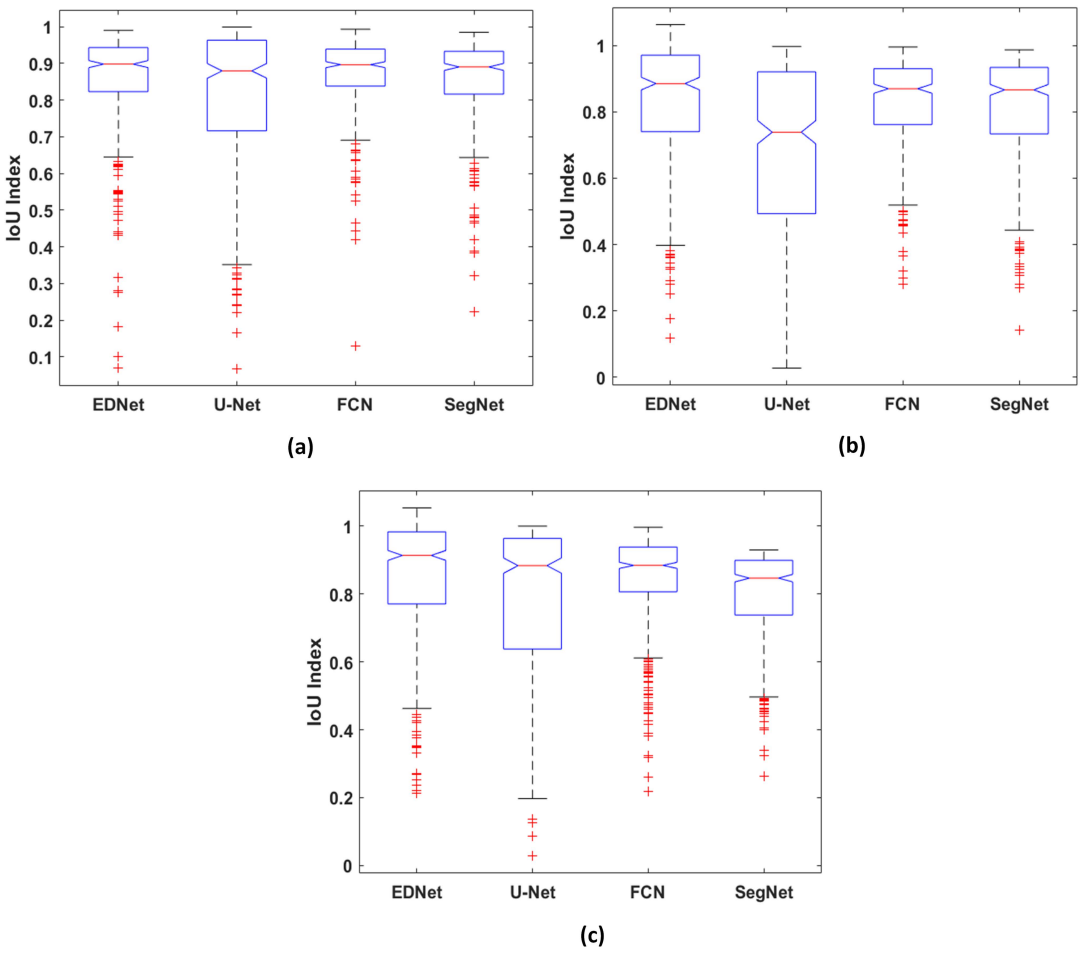



Figure 5.8: Comparison of the EDNet with other networks using ANOVA test based on IoU index for: (a) ISIC 2016 (b) ISIC 2017 (c) ISIC 2018 test sets.

Table 5.11: Performance comparison of the EDNet with other state-of-the-art methods on ISIC 2016 test set.

| **References** | **Model** | **IoU** |
|---|---|---|
| **Our method** | EDNet | **0.887** |
| M.A.Al-Masni [15] | FrCN | 0.789 |
| L.Yu [30] | CUMED | 0.829 |
| U.Sanchez [163] | ExB | 0.852 |
| M.M.Rahman [164] | Adaptive thresholding and Region growing | 0.822 |
| L.Huang [165] | OSO–FCNs | 0.811 |
| F.Xie [166] | CNN with HRFB | 0.858 |
| M.K.Hassan [167] | DSNet | 0.859 |

Table 5.12: Performance comparison of the EDNet with other state-of-the-art methods on ISIC 2017 test set.

| References | Model | IoU |
|---|---|---|
| **Our method** | EDNet | **0.770** |
| Y.Yuan [168] | Convolutional-Deconvolutional neural network | 0.765 |
| M.Berseth [169] | U-Net | 0.762 |
| L.Bi [170] | MResNet | 0.760 |
| A.Manegola [171] | RECOD Titans (U-Net) | 0.754 |
| K.Zafar [172] | Res+U-Net | 0.772 |
| B.Lei [173] | U-Net-SCDC | 0.771 |
| K.Abhishek [174] | Illumination-based Transformations | 0.752 |

Table 5.13: Performance comparison of the EDNet with other state-of-the-art methods on ISIC 2018 test set.

| References | Model | IoU |
|---|---|---|
| **Our method** | **EDNet** | **0.855** |
| B.Lei [173] | Generative adversarial network | 0.824 |
| C.Qian [175] | MaskRCNN2 + Segmentation | 0.802 |
| H.Du [176] | ensemble_with_CRF_v3 | 0.799 |
| Y.Ji [177] | Deep convolutional neural network | 0.799 |
| Y.Xue [178] | Adversarial Learning | 0.798 |
| R.Ali [179] | Deep convolutional neural network and SVM | 0.735 |

to other state-of-the-art approaches. It is concluded that network designed in encoder-decoder fashion with a careful decision of layer ordering, hyperparameters, and cleaned datasets exhibit higher performance.

### 5.4.4 Results of method 2: DilatedSkinNet

DilatedSkinNet was also trained on the same datasets, ISIC 2016-2018 training samples, and evaluated against unseen test sets. The concept of dilation rate was incorporated into the design of the network. The impact of using atrous convolutions and leakyReLU activation function on the system's performance is analyzed by training the network both with and without them. We demonstrate in Table 5.14 how the DilatedSkinNet with atrous convolutions and leakyReLU achieves higher ACC, IoU, and DICE indexes with

low training time than the network without them. Using DilatedSkinNet, the inference time was shown to be shorter than a network without atrous and leakyReLU functions. The visual outputs predicted by the proposed model for a few samples are shown in Figure 5.9 that closely resembles the expected ground truths.

Table 5.14: Performance comparison of the DilatedSkinNet with its modified architectures on ISIC 2016-2018 test sets.

| **Methods** | **ISIC 2016** | | | | **ISIC 2017** | | | | **ISIC 2018** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** |
| DilatedSkinNet | **0.950** | **0.904** | **0.949** | **10s** | **0.888** | **0.793** | **0.884** | **10s** | **0.942** | **0.891** | **0.942** | **14s** |
| Without atrous convolutions | 0.853 | 0.735 | 0.513 | 12s | 0.758 | 0.556 | 0.460 | 16s | 0.825 | 0.635 | 0.510 | 21s |
| Without leakyReLU | 0.935 | 0.855 | 0.622 | 11s | 0.874 | 0.784 | 0.578 | 14s | 0.863 | 0.736 | 0.556 | 19s |

Note: Time is in seconds on test sets. The higher values are marked in bold.

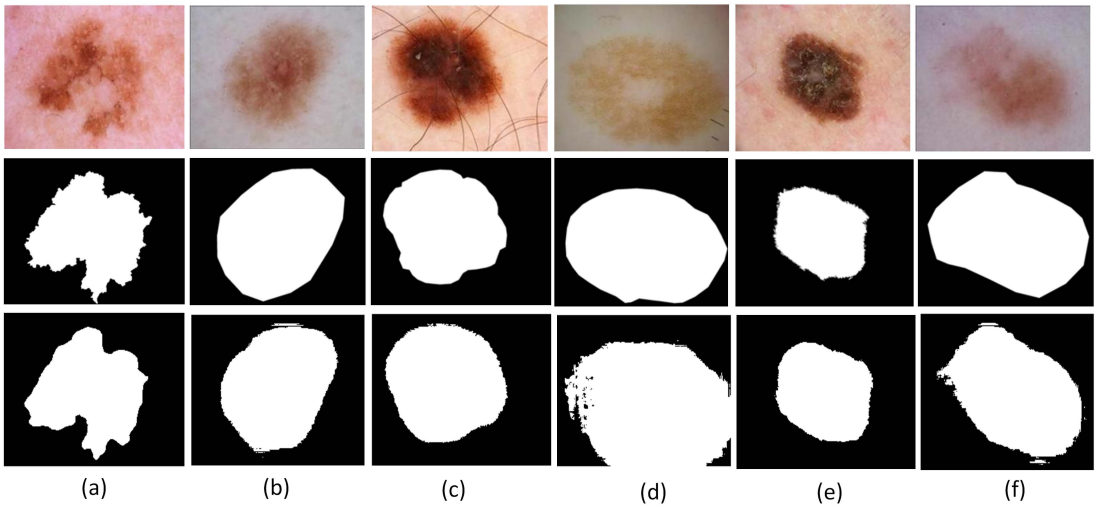



Figure 5.9: Segmentation result of DilatedSkinNet; first row shows original images, second row are ground truth and the last row are segmented images on exemplary lesion samples: (a) irregular boundaries (b) blood vessels (c) hairlines (d) color illumination (e) bubbles (f) low contrast.

### Comparison with baseline models

The DilatedSkinNet segmentation framework performed better than the existing frameworks. In Table 5.15, the results of the given model are demonstrated in comparison to

the existing semantic segmentation networks named U-Net, SegNet, and DeepLabv3+. For a fair comparison, these existing networks were also trained on the same datasets and fine-tuned by replacing their pixel classification block with new layers for segmenting an image into two classes; lesion or background. These baseline networks were trained end-to-end with the same hyperparameter configuration which is used for DilatedSkinNet segmentation model.

Table 5.15: Performance comparison of the DilatedSkinNet with U-Net, SegNet, and DeepLabv3+ on ISIC 2016-2018 test sets.

| **Methods** | **ISIC 2016** | | | | **ISIC 2017** | | | | **ISIC 2018** | | | | **Test time** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** | |
| SegNet [94] | 0.908 | 0.813 | 0.907 | 16s | 0.822 | 0.679 | 0.818 | 16s | 0.880 | 0.730 | 0.879 | 33s | 0.5s |
| U-Net [95] | 0.854 | 0.798 | 0.832 | 11s | 0.764 | 0.687 | 0.696 | 10s | 0.842 | 0.793 | 0.815 | 24s | 0.3s |
| DeepLabv3+ [96] | **0.952** | 0.892 | **0.952** | 16s | 0.878 | 0.730 | 0.881 | 19s | 0.939 | 0.888 | 0.941 | 20s | 0.7s |
| DilatedSkinNet | **0.950** | **0.904** | **0.949** | **10s** | **0.888** | **0.793** | **0.884** | **10s** | **0.942** | **0.891** | **0.942** | **14s** | **0.1s** |

Note: Time is in seconds on complete test sets. Test time is on single image in seconds (s). The best values are marked in bold.

The performance of these networks was recorded on the individual test sets of three years. DilatedSkinNet achieved an average ACC 95.0%, IoU 90.4%, and DICE 94.9% on ISIC 2016, ACC 88.8%, IoU 79.3%, DICE 88.4% on ISIC 2017, and ACC 94.2%, JAC 89.1%, DICE 94.2% on ISIC 2018 dataset. In contrast to this, the IoU index computed by the U-Net was 79.8%, 68.7%, and 79.3%, SegNet computed IoU of 81.3%, 67.9%, 73.0%, and DeepLabv3+ gained IoU of 89.2%, 73.0%, and 88.8% on ISIC 2016, 2017 and 2018 datasets, respectively which is lower than DilatedSkinNet. The results show that the proposed network generalized well on the test sets compared to the state-of-the-art semantic segmentation networks. Moreover, the primary advantage of DilatedSkinNet is that it yields high performance in less inference time comparatively, as given in Table 5.15. The inference time of DilatedSkinNet on ISIC 2016-2018 sets is 10 seconds, 10 seconds, and 14 seconds respectively. The plots in Figure 5.10 and 5.11 shows higher accuracy achieved by the DilatedSkinNet on ISIC 2016-2018 test and validation sets, respectively, in comparison to the SegNet, U-Net and DeepLabv3+ networks.

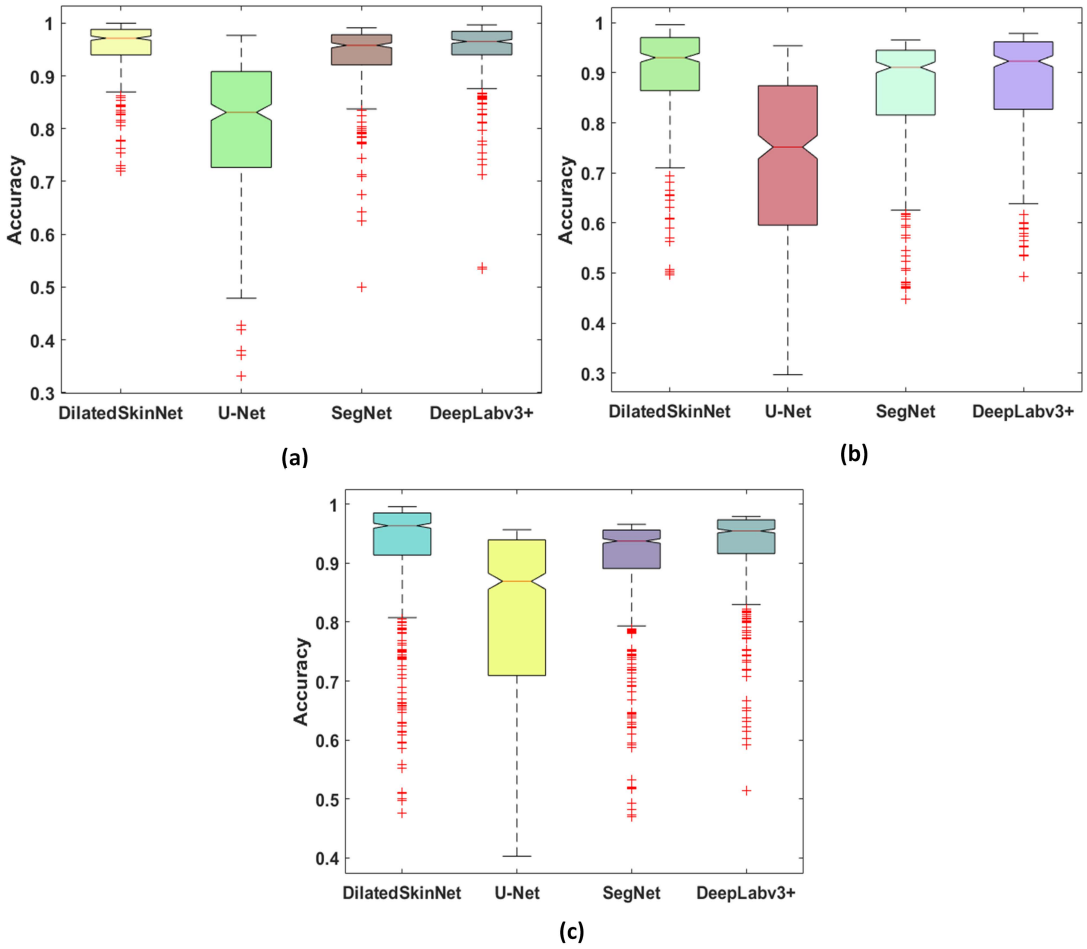



Figure 5.10: Comparison of the DilatedSkinNet with other networks using ANOVA test based on accuracy for: (a) ISIC 2016 (b) ISIC 2017 (c) ISIC 2018 test sets.

The box plots in Figure 5.12 show that DilatedSkinNet was efficient in extracting lesion information with a high IoU score as compared to the other models. The experimental results proved that DilatedSkinNet is a less complex network and efficient for calculating the expected information rather than designing a heavy network such as DeepLabv3+.

### Comparison with Other Related Studies

To prove the robustness of the network, we compared our model with the top winners of the ISIC challenge 2016-2018 as given in Table 5.16. The vital parameter used by the

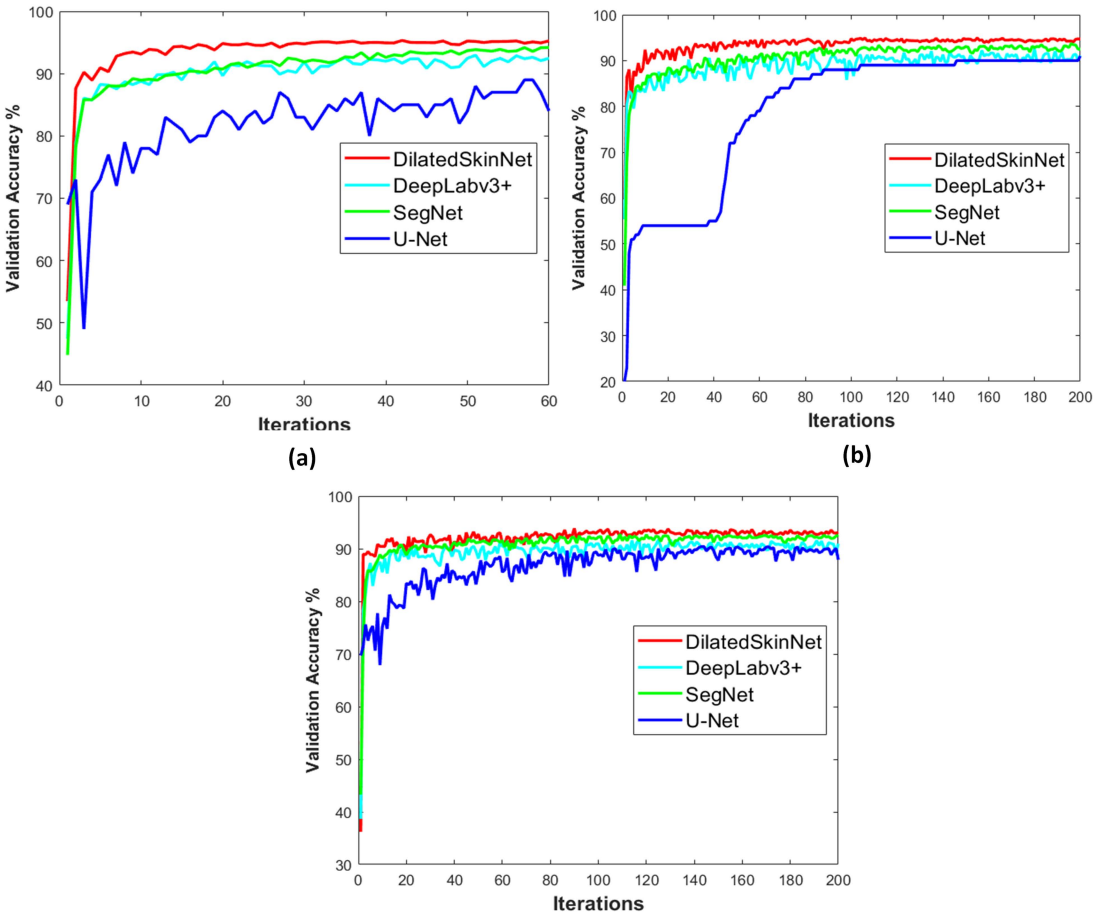



Figure 5.11: Comparison of DilatedSkinNet based on accuracy over iterations on the ISIC (a) 2016 (b) 2017 (c) 2018 validation sets.

ISIC challenge to announce the winner was the IoU index for which DilatedSkinNet achieved the highest value as 90.4% on ISIC 2016, 79.3% on ISIC 2017, and 89.1% on ISIC 2018 dataset. The comparison of DilatedSkinNet with the recent studies published in the years 2019-2021 is also given in the last two rows. It is presented that DilatedSkinNet outperformed all given studies by offering higher IoU scores.

## 5.5 Conclusion

Automatic segmentation of skin lesions is crucial for the melanoma detection system. In this chapter, we discussed two novel segmentation methods specifically designed

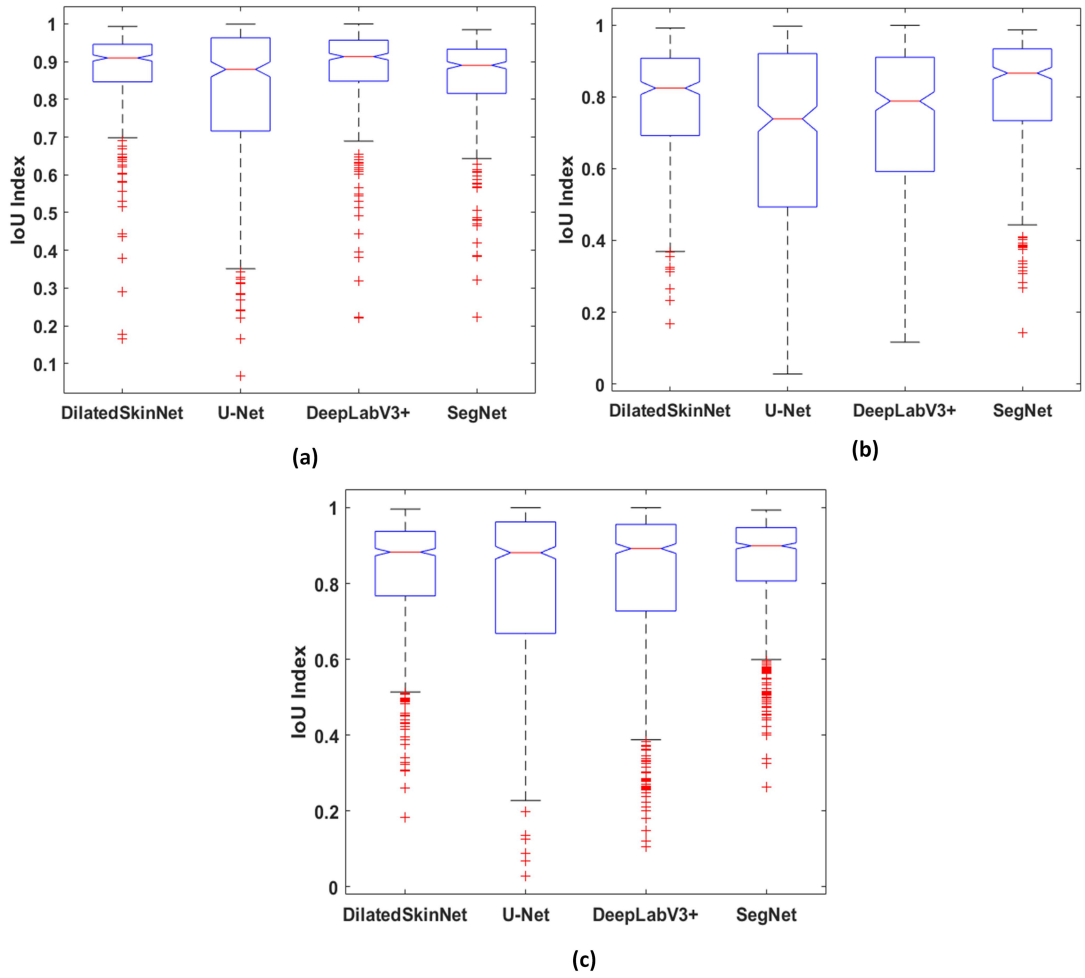



Figure 5.12: ANOVA graphs of the DilatedSkinNet with other networks based on Jaccard index/IoU on: (a) ISIC 2016 (b) ISIC 2017 (c) ISIC 2018 test sets.

Table 5.16: Performance comparison of the DilatedSkinNet with other state-of-the-art methods on ISIC 2016-2018 test sets.

| **ISIC 2016** | | **ISIC 2017** | | **ISIC 2018** | |
|---|---|---|---|---|---|
| **References** | **IoU** | **References** | **IoU** | **References** | **IoU** |
| DilatedSkinNet | **0.904** | DilatedSkinNet | **0.793** | DilatedSkinNet | **0.891** |
| L. Yu [30] | 0.829 | Y.Yuan [168] | 0.765 | C. Qian [175] | 0.802 |
| U. Sanchez [163] | 0.843 | M. Berseth [169] | 0.762 | H. Du [176] | 0.799 |
| M. Rahman [164] | 0.822 | L. Bi [170] | 0.760 | Y. Ji [177] | 0.799 |
| L. Huang [165] | 0.811 | Menegola [171] | 0.754 | Y. Xue [178] | 0.798 |
| F. Xie [166] | 0.858 | K. Zafar [172] | 0.772 | R. Ali [179] | 0.735 |
| MK. Hassan [167] | 0.859 | MP. Pour [180] | 0.782 | B. Lei [173] | 0.824 |

Note: The higher values are marked in bold.

to extract lesions from dermoscopic skin cancer images. The performance of these networks was improved using a large dataset and optimal hyperparameters. The EDNet model was designed based on encoder-decoder architecture, and the DilatedSkinNet employed atrous convolutions. The EDNet framework consisted of sampling and up-sampling layers utilizing the max-pooling indices of the encoder into the decoder section. These indexes helped to maintain the original spatial resolution of the output map. The DilatedSkinNet model used atrous dilations to expand the receptive field of the input vector without using pooling layers. Dilation rate allowed each convolution output to contain a wide range of information without extra computations and without losing the image's resolution. This network also achieved higher performance by minimizing the cross-entropy loss across mini-batches and extracting relevant features from the different dermoscopic skin cancer images for generating segmented maps. Both networks successfully segmented the majority of challenging cases such as irregular boundaries, gel bubbles, low contrast, and color illumination. If the performance of both networks is compared, it is concluded that the EDNet achieved higher performance compared to the DilatedSkinNet. On ISIC 2016-2018 test sets, EDNet's gained average accuracies as 94.1%, 87.4%, and 92.3%, whereas the DilatedSkinNet achieved average accuracies of 95.0%, 88.8%, and 94.2% on the same test sets, which is marginally higher than EDNet's. Therefore, DilatedSkinNet would be used as a segmentation step in evaluating the classification model's performance using segmented data to distinguish melanoma and non-melanoma. The next chapter will discuss the details of the classification model designed to address research question RQ4.

# Chapter 6

# Melanoma Classification

## 6.1 Introduction

The diagnosis process of melanoma through automatic computer vision techniques is challenging. However, the use of deep learning as a machine vision tool can solve some of these challenges. Using deep convolutional neural networks (DCNNs) to classify melanoma and benign skin cancers, an automated melanoma classifier is presented in this chapter. DCNN was carefully crafted by arranging many layers in an orderly fashion to extract low-level and high-level details in the skin images. Following the preprocessing and segmentation stages, the classification stage was designed to distinguish melanoma from benign skin cancer. In the following subsections, we discuss the details of the network's architecture followed by experimental results to illustrate how well the designed DCNN can identify melanoma cancer. The complete network design and results discussed in this chapter are submitted in the following journal:

- R. Kaur, H. GholamHosseini, R. Sinha, & M. Lindén, "Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images", *Sensors*, vol.22, no.3, pp.1134, 2022.
  [Online]. Available: https://doi.org/10.3390/s22031134.

Every year, millions of people die from skin cancer, a widespread and dangerous disease. The control over mortality rate due to cancer is challenging; however, early diagnosis can increase the survival rate. An expert dermatologist performs a series of steps, starting with a visual inspection of a skin lesion by the naked eye, then dermoscopy which is a magnifying lens to view lesion pattern in detail, and finally, the biopsy is performed [181]. These conventional methods are time-consuming, expensive, and laborious. Achieving accurate diagnosis is entirely subjective depending upon the expert's skillset resulting in variations in their predictions. For the diagnosis of a skin lesion at the earliest stage, comprehensive research solutions are required to identify a skin lesion considering its data complexity and presenting its diagnosis at an early stage. According to literature analysis, the traditional machine learning classifiers such as SVM [27], DT [182], and ANN [183] have been utilized by many studies. However, these machine learning methods pose disadvantages in computing highly complex skin data with low contrast, noise, and hairy images. In contrast, these approaches efficiently work only on noise-free and good-quality image data. To classify skin cancer, a rich set of features must be considered, including the lesion's color, texture, and structure. The poor feature information causes a reduction in the performance because the skin lesions differ significantly between classes and within classes [25]. Traditional methods that rely on supervised machine learning are inadequate because each lesion has different pattern. The application of deep learning techniques to skin classification is, on the other hand, efficient at allowing dermatologists to diagnose skin lesions with a high degree of accuracy even when they are complex.

Figure 6.1 shows the three primary types of skin cancer: Basal, Squamous, and Melanocytes [184]. Abnormal cell growth in the body's skin called basal cell carcinoma grows very slowly and does not spread to other parts. This form of cancer often recurs, so eliminating it from the body is essential. Squamous cells penetrate deeper into the skin than Basal cells and are more prone to spreading elsewhere on the body. In

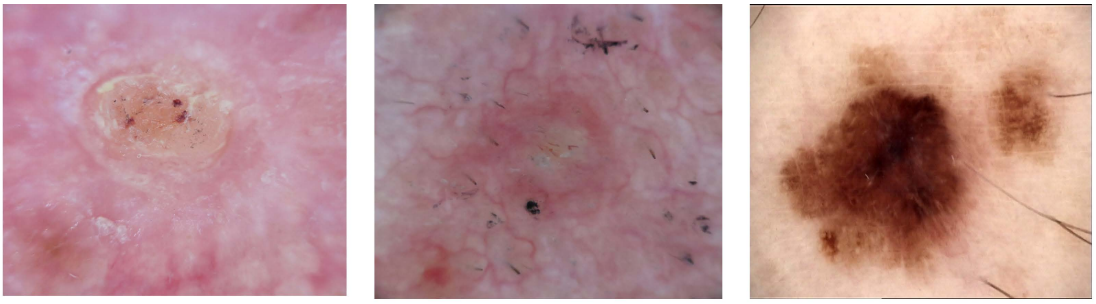



Figure 6.1: Primary lesion types: (a) Basal (b) Squamous (c) Melanoma.

the last type, Melanocytes, melanin is produced when the skin is exposed to sunlight, which gives it its dark brown or tan appearance. Melanin in these cells shields the skin from sunlight, but if it accumulates in the body, it can cause cancerous moles, a condition is known as melanoma. Basal and squamous cancers are considered benign, whereas Melanocytes are considered malignant and can be life-threatening due to their tendency to cause extensive damage to surrounding tissues. Despite the fact that deep learning approaches are highly effective for handling complex data, skin dataset remains a challenging for classification task due to the following reasons:

(a) There is an imbalance in the classes of skin lesions in given datasets.

(b) Several artefacts are present in lesions, including hairlines, gel bubbles, ruler marks, and low contrast.

(c) As a result of high intraclass differences and interclass similarities, it is difficult to distinguish lesion types.

(d) presence of dulpicate lesions causes bias and overfitting.

Designing a melanoma classification network using deep learning techniques came with a few challenges, including:

- Finding a high prediction rate while dealing with class imbalance problems.

- Designing a network architecture that is less complex and lighter.
- Obtaining a low inference time.

Transfer learning fails to produce efficient results on skin cancer classification problems using existing deep neural models because they are trained on different datasets such as ImageNet. Hence, proposed DCNN model was designed, implemented and evaluated to classify melanoma using skin cancer data.

## 6.2 Lesion Classification Network (LCNet)

A deep learning network named lesion classification network (LCNet) is based on DCNN models and explicitly designed to classify skin lesions, a task that becomes increasingly difficult for clinicians. The LCNet architecture consists of 11 blocks, as shown in Figure 6.2. This network is constructed by repeating blocks 4 & 5, 7 & 8, and 10 & 11 from top to bottom according to rates of 2, 4, and 2, respectively. The network's first layer accepts an image whose dimensions are $128 \times 128$ and has three channels i.e., RGB. It then follows a convolutional operation which slides $8$ kernels of size $3 \times 3$ over an image with a stride of $2$. Feature calculation is the primary purpose of this layer; when a kernel is overlapped over an image, the pixel values are transformed as follows:

$$Conv[x, y] = \sum_{i=0}^{s} \sum_{j=0}^{s} (I_{x-i,y-j} * K_{i,j},\ n_f) \tag{6.1}$$

During convolution, $Conv[x, y]$ represent the output feature map having transformed pixels in the spatial domain, $s$ represent the size of kernel $K$ which is multiplied with the pixels of input image $I$ for multiple channels $n_f$. This layer output a feature map and passes to the next layer, i.e., max-pooling, which transforms the regions of the feature map by taking the maximum value of the regions. Feature maps are reduced

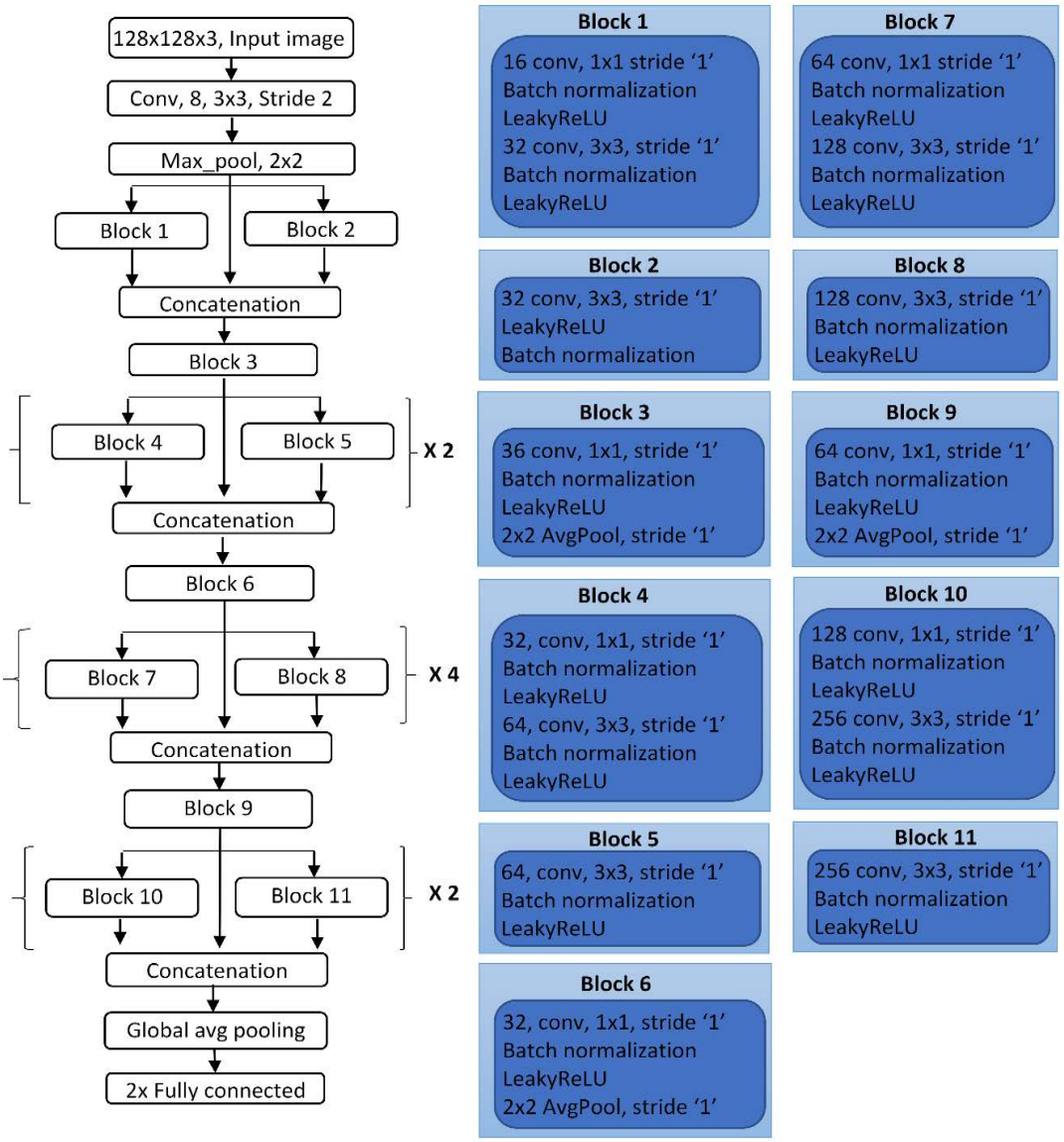



Figure 6.2: The design of the LCNet model.

in size through pooling operation. Each block has three main layers: convolutional, batch normalization (BN), and leakyReLU. A batch normalization process normalizes the input feature maps from previous layers. It prevents overfitting by regulating the learning process of the network. The activation function used is leakyReLU which has the advantage of having a slight slope for negative values rather than zero slopes, as is the case with the standard ReLU. By multiplying negative values by a scalar value

$s$ = 0.3, the LeakyReLU function transforms negative values into positive values:

$$leakyReLU = \begin{cases} x \times s, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{6.2}$$

A convolutional layer, a BN layer, and a leakyReLU layer are used twice in Block 1. The first convolutional layer contains 16 kernels with a size of $1 \times 1$, while the second layer contains 32 kernels with a size of $3 \times 3$ with a stride of 1. Generally, a stride represents the number of steps it takes for the filter map to slide across an image. On the other hand, block 2 contains three layers as a convolutional layer with 32 filters having the size of $3 \times 3$, followed by leakyReLU and BN layers. A combined feature set is then obtained by concatenating the feature sets generated by blocks 1 and 2 and the pooling indices derived from max-pooling, which is then passed on to block 3. Blocks 4, 7, and 10 follow a similar pattern to block 1, except the number of kernels increases from 32, 64, 64, 128, 128, and 256. In the successive convolutional layers, the number of filters has increased to 64, 128, 256 of size $3 \times 3$ in blocks 5, 8, and 11. Finally, the number of filters in blocks 3, 6, and 9 are 36, 32, and 64 of size $1 \times 1$ followed by a layer of average pooling. This block used an average pooling operation instead of max-pooling to calculate the average for each patch of the feature map that overlaps the filter window. This layer downsamples to an average value in the window with a filter size of $2 \times 2$.

A deep network of repeating blocks allows for the extraction of lesion information, including edges, colors, and complex patterns, in the form of a feature map. Using a global average pooling method and a fully connected layer, a single feature vector for each class category is generated at the network's end. The softmax function calculates a confidence score to interpret the probabilities of falling into a certain class. Compared to other state-of-the-art networks, the proposed network generates fewer learnable parameters, making it less complex and lightweight. To illustrate, the number of

learnable parameters used in the studies [185, 186, 187, 188] were respectively 134.2M, 19M, 4.15M and 138M. LCNet, on the other hand, achieved high performance by optimizing parameters and kernels as 3.3M and 3.1K, respectively.

Computational costs and memory limitations are common problems with deep neural networks. In order to resolve this issue, the original images were resized to smaller dimensions using bi-linear interpolation. The rescaling operation preserves contextual information about lesions that would otherwise be lost for skin classification. In addition, augmentation operations and random oversampling are used to deal with the skewed distribution of lesion samples. To balance the distribution, more instances are created for underrepresented classes. The presence of noise artifacts and the high similarity ratio between classes and differences within classes make classification difficult. Consequently, the proposed network with 31 convolutional layers can efficiently extract low to high-level information. Backpropagation is used to optimize the network weights by reducing loss based on gradient values. To achieve this, a stochastic gradient optimizer (SGDM) [189] updates the weights of the network as well as its biases to make small changes in the direction of optimization.

$$\theta_{i+1} = \theta_i - \alpha \bigtriangledown L(\theta_i) \tag{6.3}$$

In this case, $i$ represents the number of iterations, $\alpha > 0$ represents the learning parameter ('0.001'), $\theta$ describes a parameter vector, and $\bigtriangledown L(\theta_i)$ represents the gradient of the loss function. The algorithm evaluates the gradient at each iteration and updates parameters over a mini-batch set. Due to large weight values, a network may become stuck in a local minima. The momentum term $\gamma$ is therefore added to the gradient descent algorithm to reduce oscillations in the following manner:

$$\theta_{i+1} = \theta_i - \alpha \bigtriangledown L(\theta_i) + \gamma(\theta_i + \theta_{i+1}) \tag{6.4}$$

To measure error between prediction score P and target T, the LCNet uses a cross-entropy loss function [190]. The weighted cross-entropy loss function calculates the error as follows:

$$Loss = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{N} w_i T_{i_j} log(P_{i_j}) \tag{6.5}$$

The number of observations is indicated as $N$, $K$ is the number of classes, and $w$ is the weight determined by the network. Table 6.1 described the hyperparameters utilized by the LCNet for training.

Table 6.1: Hyperparameter used in the training of the LCNet.

| Learning algorithm | Learning rate | Mini-batch size | Epochs | Activation function | Data augmentation | Momentum | Regularization |
|---|---|---|---|---|---|---|---|
| SGDM | 0.001 | 32 | 100 | LeakyReLU | Random oversampling, rotation, translation, and scaling | 0.99 | 0.0005 |

## 6.3 Datasets

The dermoscopic datasets used for the training and evaluation of the LCNet model were obtained from the ISIC 2020 challenge [191] and HAM10000 [139]. The original ISIC 2020 dataset consists of 33,126 lesion samples collected from 2056 patients. There are 16 samples per patient in this dataset, which are recorded over time to observe changes. Several of the lesions are very similar and come from the same patient, thus creating duplicates [137]. In this case, a network may exhibit biases and suffer from overfitting. Therefore, data normalization was performed to eliminate duplicates and prevent the network from experiencing these issues. Using the duplicate removal method, the dataset left 12,882 skin samples made up of 7093 BEN samples and 5789 MEL samples. Furthermore, the present dataset contains challenging samples, such as lesion samples with dark skin and samples of older people. Another dataset used was HAM10000 in

which there were 9900 total samples divided into three classes BEN with 6705 image samples, 1113 MEL samples, and 2082 were unknown samples referred to as UNK. The classes in these datasets were highly unbalanced, which negatively affected the model's accuracy. For example, MEL samples were fewer than BEN samples in the ISIC and HAM10000 datasets. The two datasets, ISIC and HAM10000 are divided into three subsets, training, validation, and test sets (see Table 6.2, & 6.3). The training set was created from 70% of the total samples, the validation set from 10%, and the test set from the remaining 20%. To provide enough training to the network, the training set contains a larger proportion of samples. Validation data is used to hypertune the parameters of the network based on its performance. A final evaluation is conducted by using the test data.

### 6.3.1 Data normalization

It was found that some skin samples had similar lesion structures but different names, resulting in duplicate data. As a result of data normalization, operations like updating and deleting anomalies are eliminated. The algorithm 1 explained in chapter 3 is applied to remove duplicate samples. The primary image $I_i(x, y)$ and images in comparison set are initially converted to grayscale as $I'_i(x, y)$ and $I'_j(x, y)$, and its histogram is calculated as $h_i$ and $h_j$.

$$h_i = hist(I'_i(x, y)) \quad h_j = hist(I'_j(x, y)) \tag{6.6}$$

We then calculate the average of each histogram using the average function:

$$h'_i = mean(h_i) \quad h'_j = mean(h_j) \tag{6.7}$$

The correlation index is calculated using 6.8 between primary image and every other

image in the comparison set. An image is deemed identical if its correlation is greater than equal to 0.99 then one copy of the image is discarded. Following the removal of redundant data, the final distribution of samples used for experiments is shown in Tables 6.2 and 6.3.

$$Correlation = \frac{\sum_x \sum_y (I(x,y) - h'_i)(I'(x,y) - h'_j)}{\sqrt{\left(\sum_x \sum_y (I(x,y) - h'_i)^2\right)\left(\sum_x \sum_y (I'(x,y) - h'_j)^2\right)}} \tag{6.8}$$

Table 6.2: Distribution of ISIC data among training, validation, and test sets.

| **Classes** | **Training samples 70%** | **Augmented Training samples** | **Validation samples 10%** | **Test samples 20%** | **Total samples 100%** |
|---|---|---|---|---|---|
| MEL | 4052 | 4965 | 579 | 1158 | 5789 |
| BEN | 4965 | 4965 | 709 | 1419 | 7093 |
| Total | 9017 | 9930 | 1288 | 2577 | 12882 |

Table 6.3: The HAM10000 data distribution among training, validation, and test sets.

| **Classes** | **Training samples 70%** | **Augmented Training samples** | **Validation samples 10%** | **Test samples 20%** | **Total samples 100%** |
|---|---|---|---|---|---|
| MEL | 779 | 4694 | 111 | 223 | 1113 |
| BEN | 4694 | 4694 | 670 | 1341 | 6705 |
| UNK | 1457 | 4694 | 209 | 416 | 2082 |
| Total | 6930 | 14082 | 990 | 1980 | 9900 |

### 6.3.2 Basic pre-processing operations

For the preparation of image samples, standard pre-processing operations were applied. Within each category, there are varying resolution dimensions ranging from $576 \times 768$ to $1024 \times 1024$ with three color channels RGB. Using bilinear interpolation, the images were resized to $128 \times 128$ dimensions, preserving their aspect ratio while minimizing computational costs. In the beginning, no noise removal method was used to remove hairlines, gel bubbles, ruler marks, or ink stains because the LCNet model's design is capable of efficiently processing raw images that are contaminated with noise artifacts. A discussion of experimental results follows after data has been pre-processed and

segmented by means of pre-processing and segmentation methods, and LCNet is used to perform a classification task.

### 6.3.3 Data augmentation

The adopted datasets were highly skewed among a wide variety of skin cancer types, as explained above. In data imbalance, the network is biased in favor of classes that have large samples over those that do not. For example, the MEL samples in ISIC and HAM10000 sets are smaller than the BEN samples. For two datasets, Table 6.2 and 6.3 illustrate the distribution of data samples within different classes. Some classes were extended, such as the 4052 MEL samples in the ISIC and the 779 MEL samples in the HAM10000, by using the random oversampling method to generate more artificial samples. Data augmentation techniques were applied more to the underrepresented classes than oversampled classes to address data undersampling, skewness, and image sample scarcity. Moreover, online data augmentation was applied during the network training using three common operations: rotation from $-300$ to $+300$, scaling with factors $0.8$ in X-direction, and $1.0$ in Y-direction translation of $-5$ and $+5$. These operations were only applied on the training sets, whereas validation and test sets were not augmented, and their original data distribution was used during the validation and testing process. Figure 6.3 shows the augmented samples for the classes MEL and BEN.

## 6.4 Performance Metrics

Various classification performance metrics such as precision (PRE), recall (REC), accuracy (ACC), specificity (SPE), F1-Score, and balanced accuracy (Bal ACC) [192, 193] and learnable parameters are considered to evaluate the model. The mathematical

Figure 6.3: Augmented data samples using translation, rotation, and scaling.

formulas used to calculate the values of these metrics are given as:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{6.9}$$

$$\text{PRE} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6.10}$$

$$\text{REC} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6.11}$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{6.12}$$

$$\text{F1} - \text{Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{6.13}$$

$$\text{Bal ACC} = \frac{\text{SPE} + \text{SEN}}{2} \tag{6.14}$$

TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives in a confusion matrix. TP represents the number of lesion samples correctly classified as melanoma, TN represents the number of lesion samples correctly classified as benign, FP represents the ratio of samples incorrectly classified as melanoma, and

FN represents the images determined to be benign when they are melanoma. An ACC is defined as the fraction of correctly identified samples and the number of predictions based on these parameters. PRE and REC are significant metrics to evaluate the model's performance as PRE measures all positive predicted rates. In contrast, REC calculates the actual positive ratio out of all positively identified samples. The model's ability to identify the TN of each class is measured by a metric called SPE. Lastly, F1-Score calculates the harmonic mean of PRE and REC by considering FP and FN. Its value close to 1 indicates the perfect PRE and REC.

## 6.5 Results of LCNet

Several experiments for skin lesion classification were conducted on different dermoscopic lesion images to evaluate the performance of the LCNet. It was tested on two different large sets, the ISIC for two classes MEL and BEN and HAM10000 for three classes MEL, BEN, and UNK. In HAM10000, the third class UNK stands for unknown skin cancer type. Several state-of-the-art methods depend highly on the ROI-specific feature calculation process to achieve a high classification rate. In contrast, the LCNet does not require extensive feature calculation or post-processing operations. It is trained end-to-end on dermoscopic images to distinguish melanoma and other lesion types. Hyperparameters (see Table 6.1) are finalized after monitoring the network's highest performance on the validation data by conducting several experiments. The network training was done on the hardware configuration of 'GeForce RTX 2080 SUPER' with a computation capacity of 7.5. Moreover, the inference time on the ISIC test set with 2577 images was 76 seconds, HAM10000 with 1980 test images was 54.4 seconds.

The scarcity of lesion samples in different classes prevented by using data augmentation and oversampling methods to resolve the bias issue. In data extension, firstly, the training of the LCNet was done on the augmented training set using the fine-tuned

hyperparameters. The impact of using data oversampling on the network's performance is shown in Tables 6.4 and 6.5 that explains there is an increase in the values of metrics on both test datasets.

Table 6.4: Impact of data oversampling on the performance of LCNet on ISIC test set.

| **Approach** | **ACC** | **PRE** | **REC** | **SPE** | **F-Score** | **Bal ACC** |
|---|---|---|---|---|---|---|
| No oversampling | 90.12 | 89.01 | 92.21 | 89.11 | 91.19 | 90.66 |
| Oversampling | **90.92** | **91.12** | **92.29** | **89.29** | **91.70** | **90.79** |

Table 6.5: Impact of data oversampling on the performance of LCNet on HAM10000 test set.

| **Approach** | **ACC** | **PRE** | **REC** | **SPE** | **F-Score** | **Bal ACC** |
|---|---|---|---|---|---|---|
| No oversampling | 87.04 | 72.25 | 65.45 | 85.82 | 67.70 | 75.64 |
| Oversampling | **93.74** | **84.56** | **89.40** | **94.87** | **86.74** | **92.14** |

Training progress was monitored on the validation set of the ISIC and HAM10000 datasets. The validation set contains a different proportion of lesion samples from all classes, and hyperparameters were tuned on the validation set to improve the performance. Thus, the final values were selected based on the best output offered by the network on the validation set having the lowest loss and high accuracy. Finally, the trained model with fine-tuned parameters was used to evaluate the test set unseen by the network. The graphical view in Figure 6.4 shows behaviour of the LCNet on the ISIC and HAM10000 validation sets by plotting their performance between loss and accuracy over the number of epochs. It displays the network's accuracy progressively increasing towards higher values over the number of iterations per epoch. The higher performance is noticed on the ISIC due to the large size of the dataset.

Further, in Figure 6.5a, it is given that the LCNet model achieved TP as 1293, TN as 1050, FN as 108, and FP as 126 on the ISIC tests dataset. Similarly, it is illustrated from Figure 6.5b that LCNet predicted 191, 1687, 70, and 32 samples as TP, TN, FP, and FN for MEL class, 1223, 596, 43, 118 samples as TP, TN, FP, FN for BEN class and 380,

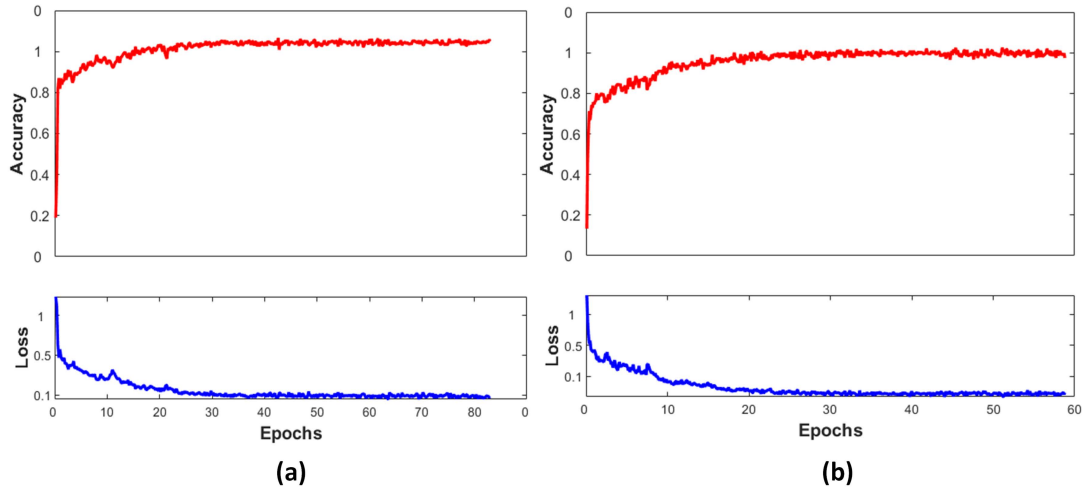



Figure 6.4: Classification accuracy and loss curves of the LCNet with the number of epochs on the validation set (a) MEL vs BEN lesion classes on ISIC (b) MEL vs BEN vs UNK lesion classes on HAM10000 datasets.

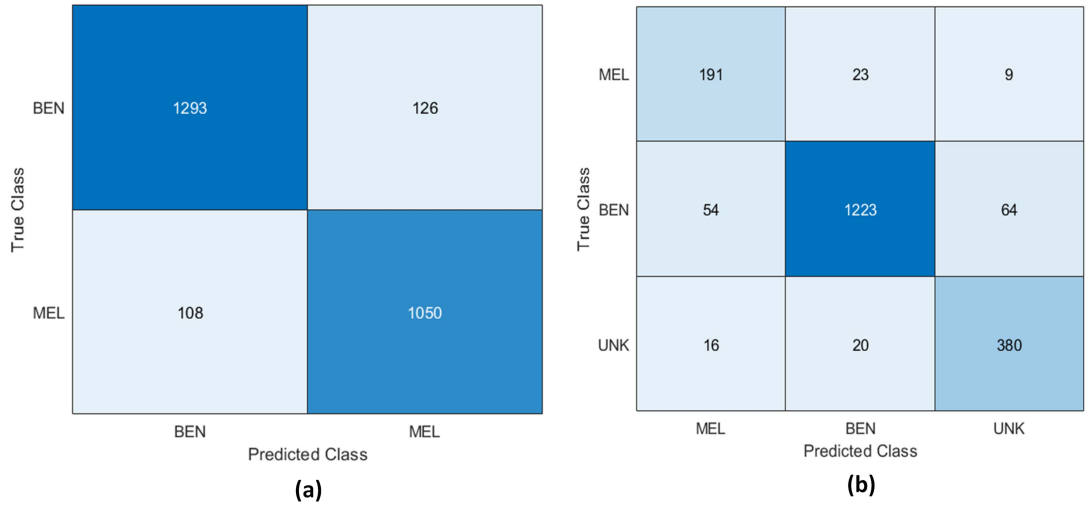



Figure 6.5: Confusion matrix of LCNet on (a) ISIC (b) HAM10000 test sets.

1491, 155 and 36 samples as TP, TN, FP, and FN for UNK class on the HAM10000 test data. Overall the false prediction rate of LCNet for each class is comparatively low, resulting in high performance of the model. Similarly, Figure 6.6 demonstrates the true positive vs. false positive curves [194] illustrating the trade-off between sensitivity and specificity achieved by the model with the area under the curve (AUC) for MEL class as '0.9642' and '0.9060' on ISIC and HAM10000 test set, respectively.

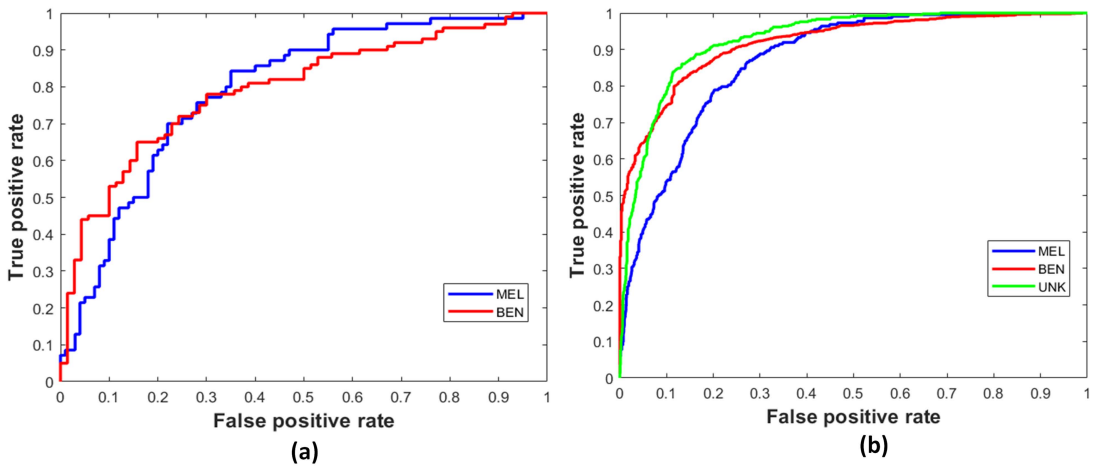



Figure 6.6: Sensitivity vs specificity curves for each class on (a) ISIC (b) HAM10000 test set.

### 6.5.1 Comparison with other related studies

The LCNet surpassed the other state-of-the-art approaches for skin lesion classification as given in Table 6.6. It shows the comparison between the methods with the best results highlighted in bold based on metrics such as ACC, PRE, REC, SPE, F1-Score, and learnable parameters. Only the PRE and REC of [186] on the HAM10000 dataset is higher than the LCNet model, whereas the ACC, SPE, and F-score of the LCNet is the highest and a smaller learnable parameters among all given studies on HAM10000 data. Furthermore, the LCNet model showed higher values for all metrics than other studies for the ISIC dataset. The related work published between 2020 and 2021 that used the same datasets with slight variations in sample size were used for the comparison. The results are also compared using Area under the curve (AUC) in Table 6.7. In the ISIC 2020 challenge, AUC was the primary metric to determine the winner. There were many duplicates in the dataset, and to the best of our knowledge, no study has tried to eliminate them. According to our findings, the LCNet model has the highest AUC score and the number of learnable parameters is the least, making it a lightweight and simpler network.

Table 6.6: Performance comparison of LCNet with other state-of-the-art methods.

| Studies | Dataset | No. of images | ACC% | PRE% | REC% | SPE% | F1-Score% | Parameters (millions) |
|---|---|---|---|---|---|---|---|---|
| H.W. Huang [186] | HAM10000 | 10015 | 85.50 | **91.10** | **96.00** | 80.10 | — | 19 M |
| M.A. Khan [187] | HAM10000 | 10000 | 86.50 | 87.01 | 85.57 | — | 86.28 | 4.15 M |
| M.A. Khan [195] | HAM10000 | 10000 | 87.02 | — | 86.98 | — | — | — |
| A. Hekler [183] | HAM10000 | 11444 | 82.95 | — | 89.0 | 81.5 | — | >23 M |
| Srivastava et al. [188] | HAM10000 | 10015 | 83.69 | 84.00 | 84.00 | — | 82.00 | 138 M |
| A. Ameri [196] | HAM10000 | 3400 | 84.00 | — | 81.00 | 88.00 | — | 62.3 M |
| F.Afza [197] | HAM10000 | 10015 | 85.80 | 86.28 | 86.00 | — | 86.14 | — |
| **LCNet Model** | HAM10000 | 9900 | **93.74** | 84.56 | 89.40 | **94.87** | **86.74** | **3.32 M** |
| S. Nasiri [198] | ISIC | 1346 | 75.00 | 77.00 | 73.00 | 78.00 | 75.00 | — |
| M.R.Hasan [185] | ISIC | 6594 | 93.18 | — | — | — | — | 134.2 M |
| T.J Brinker[199] | ISIC | 12378 | — | 74.10 | 87.50 | 60.00 | — | >23 M |
| Kwasigroch, A. [200] | ISIC | 13600 | 77.00 | — | — | — | — | 7.18 M |
| **LCNet Model** | ISIC | 12882 | **90.92** | **91.12** | **92.29** | **89.29** | **91.70** | **3.32 M** |

Table 6.7: Performance comparison of LCNet with other related studies based on AUC score.

| Studies | Dataset | Dulicate Removed | Model | AUC |
|---|---|---|---|---|
| S. Karki [201] | 2020 | NO | Ensemble Nets | 94.11 |
| Q. Ha [202] | 2020 | NO | Ensemble Efficient-Net | 94.90 |
| M. O'Brien [203] | 2020 | NO | DNN | 59.1 |
| Proposed | 2020 | Yes | LCNet | **96.42** |

### 6.5.2 Comparison with baseline CNN models

In Table 6.8, the performance of baseline CNN models such as ResNet18, Inceptionv3, and AlexNet was displayed. These popular networks were fine-tuned on the adopted datasets, and then the comparison was performed to evaluate the proposed network's performance. It is clear from Table 6.8 that the proposed model outperformed given networks by achieving the highest values for metrics ACC, PRE, and REC on the HAM10000 and ISIC test sets. LCNet obtained highest values for all performance metrics on both datasets as compared to ResNet18, Inceptionv3, and AlexNet. Additionally, the number of learnable parameters are extremely low in case of LCNet.

In Figure 6.7 and 6.8, the confusion matrices illustrating TP, TN, FP and FN score achieved by the ResNet18, Inceptionv3, AlexNet and LCNet model on the HAM10000 and ISIC test set are displayed, respectively. The LCNet model scored the highest TP

Table 6.8: A comparison between proposed LCNet with baseline CNN models on HAM10000 and ISIC datasets.

| Approach | HAM10000 | | | | | ISIC | | | | | Parameters | Test time |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | PRE | REC | SPE | F1-Score | ACC | PRE | REC | SPE | F1-Score | | |
| ResNet18 | 81.62 | 57.58 | 48.89 | 76.27 | 49.46 | 85.32 | 84.86 | 88.54 | 81.54 | 86.66 | 11M | 171s |
| Inceptionv3 | 82.26 | 62.17 | 51.73 | 78.55 | 50.49 | 90.50 | 92.82 | 90.52 | 90.48 | 91.66 | 24M | 260s |
| AlexNet | 78.48 | 54.6 | 33.33 | 66.67 | 26.92 | 49.17 | 13.11 | 70.72 | 46.72 | 22.12 | 62.3M | 166s |
| **LCNet** | **93.74** | **84.56** | **89.40** | **94.87** | **86.74** | **90.92** | **91.12** | **92.29** | **89.29** | **91.70** | **3.32M** | **76s** |

Note: Execution time is on test sets in seconds (s).

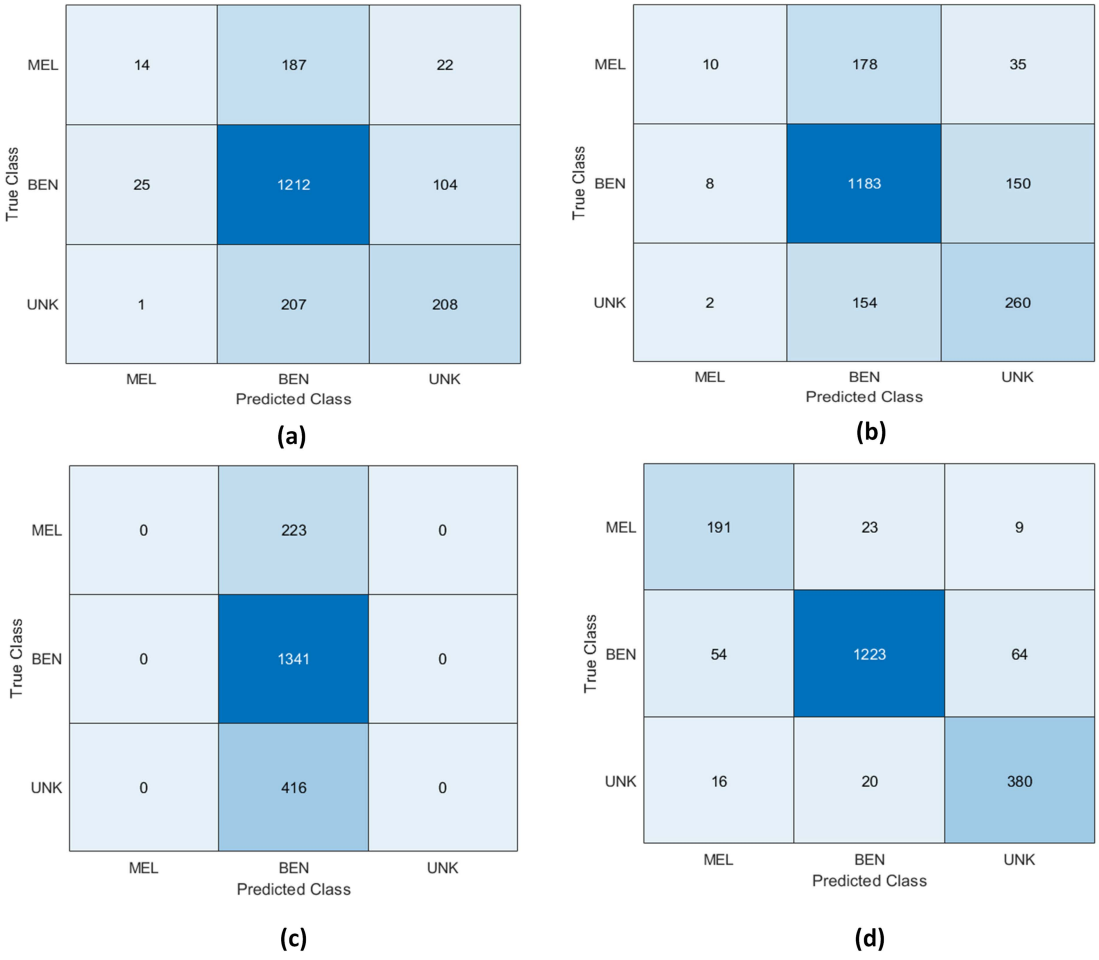



Figure 6.7: Confusion matrix on HAM10000 test data using (a) ResNet18 (b) InceptionV3 (c) AlexNet (d) LCNet.

and TN on both datasets compared to other baseline models. The metrics gave a better insight into the correctly classified and misclassified samples based on the parameters of the confusion matrix (i.e., TP, TN, FP, and FN).

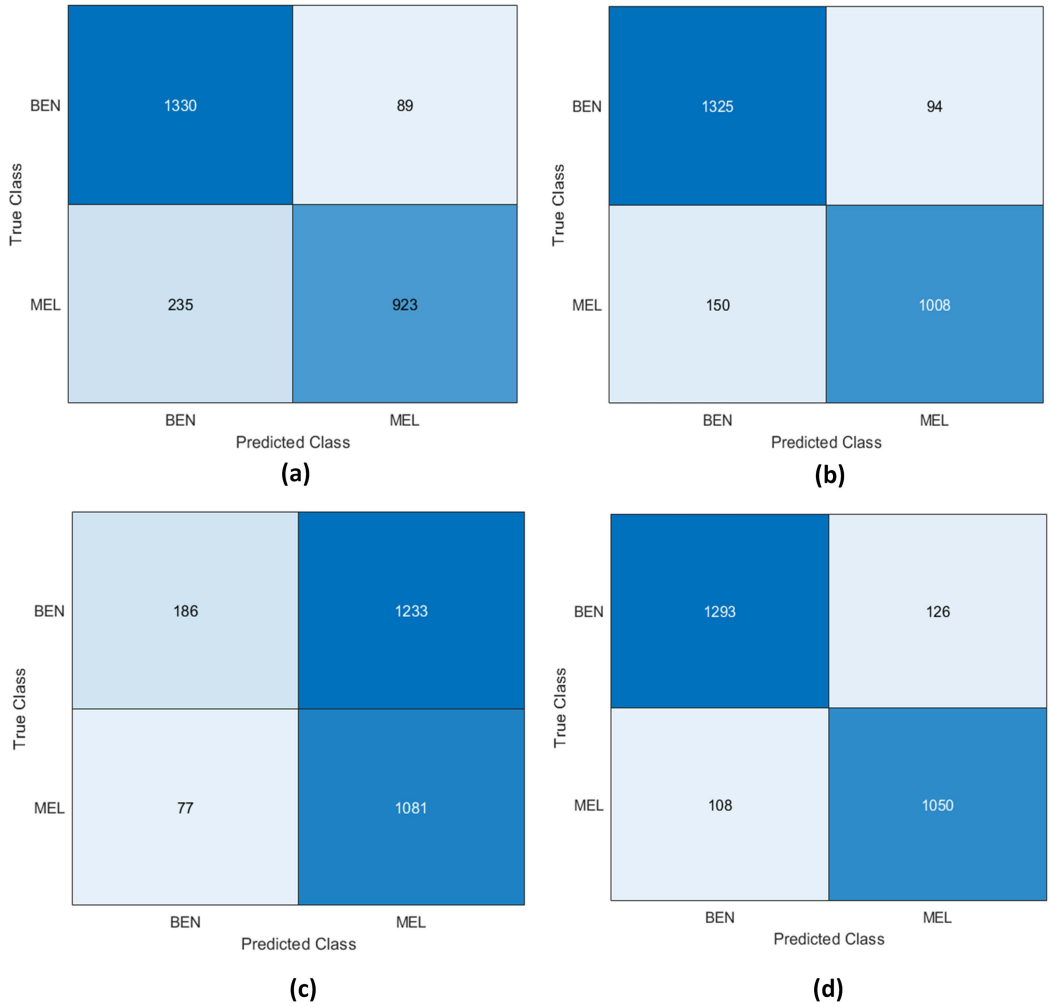



Figure 6.8: Confusion matrix on ISIC test data using (a) ResNet18 (b) InceptionV3 (c) AlexNet (d) LCNet.

The experimental outcomes proved that the LCNet model performs better in classifying melanoma with a balanced and large dataset. The LCNet model obtained ACC, PRE, and REC of 93.74%, 84.56%, and 89.40%, respectively, for the three-class classification of MEL vs BEN vs UNK on the HAM10000 dataset. At the same time, the values for these metrics on the ISIC test set for two-class classification, i.e., MEL vs BEN, were 90.92%, 91.12%, and 92.29%, consequently. It is observed that the proposed LCNet model gained a higher ACC, PRE, and REC on both datasets, which is higher among other popular models comparatively. The validation performance in terms of

accuracy over the iterations is also plotted as shown in Figure 6.9. It demonstrates that LCNet gained the highest accuracy rate on validation data as compared to ResNet18, Inceptionv3, and AlexNet.

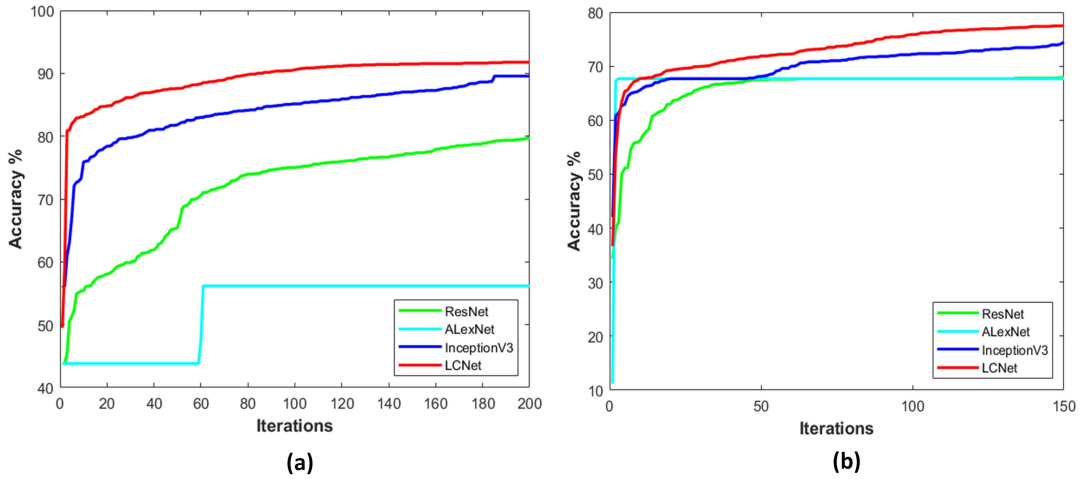



Figure 6.9: Comparison of validation accuracy of models over iterations on (a) ISIC (b) HAM10000.

## 6.6 Impact of pre-processed data

The performance of segmentation (DilatedSkinNet) and classification (LCNet) models using pre-processed data is examined in this section. Pre-processed data is data that has been cleaned by using a hair removal algorithm (IA-HR), and their contrast improved using MCACNN. The statistical results of a segmentation model followed by a classification model based on performance metrics are presented here.

### 6.6.1 Results of DilatedSkinNet on pre-processed data

In Chapter 5, a DilatedSkinNet segmentation model was developed to segment lesions for the ISIC 2016 to 2018 datasets. The model was evaluated using both raw and pre-processed data (cleaned using the IA+HR and MCACNN models) and results are

presented here. According to Table 6.9, cleaned data improved model performance slightly by a difference of 1.38% ACC, 6.6% IoU and 1.45% DICE score on ISIC 2016, 0.6% DICE on ISIC 2017 and 1.22% ACC, 2.11% IoU and 1.19% DICE score on ISIC 2018 data. However, the ACC and IoU is higher on ISIC 2017 raw data as compared to pre-processed data. Figure 6.10 displays ANOVA test on ISIC 2016-2018 test set indicating higher median IoU score presented with red centre line in the boxes achieved by DilatedSkinNet on pre-processed data as compared to original data. Thus, hair removal can be utilized in order to eliminate hairlines from dermoscopic images prior to the accurate extraction of lesions.

Table 6.9: Performance analysis of DilatedSkinNet on the pre-processed data.

| **Data** | **ISIC 2016** | | | | **ISIC 2017** | | | | **ISIC 2018** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** | **ACC** | **IoU** | **DICE** | **Time** |
| Raw | 93.63 | 87.86 | 93.54 | 10.8 | **93.63** | **93.54** | 87.86 | 10 | 93.06 | 87.02 | 93.06 | 14 |
| Pre-processed | **95.01** | **90.46** | **94.99** | **10.2** | 88.88 | 79.30 | **88.46** | 10 | **94.28** | **89.13** | **94.25** | **14** |

Note: Time is in seconds on test sets. The higher values are marked in bold.

### 6.6.2 Results of LCNet on pre-processed data

In this section, the performance of the LCNet is evaluated on the pre-processed and segmented images to respond to the RQ5 related to observing the impact of pre-processing and segmentation on classification accuracy of LCNet. A large dataset ISIC (see Table 6.2) was adopted to conduct the experiments. First, the IA-HR algorithm (explained in Chapter 4) was applied to this dataset to remove hairlines, and their contrast was increased using MCACNN (refer to Chapter 4), then the segmentation method, DilatedSkinNet (refer to Chapter 5), was applied to extract accurate lesion ROI. Finally, the LCNet was trained, validated, and evaluated with training, validation, and test samples of the ISIC dataset as defined in Table 6.2. The same hyperparameters, as explained in the previous sections in Table 6.1 were used while training the LCNet on pre-processed data. The training progress of LCNet on raw data and pre-processed

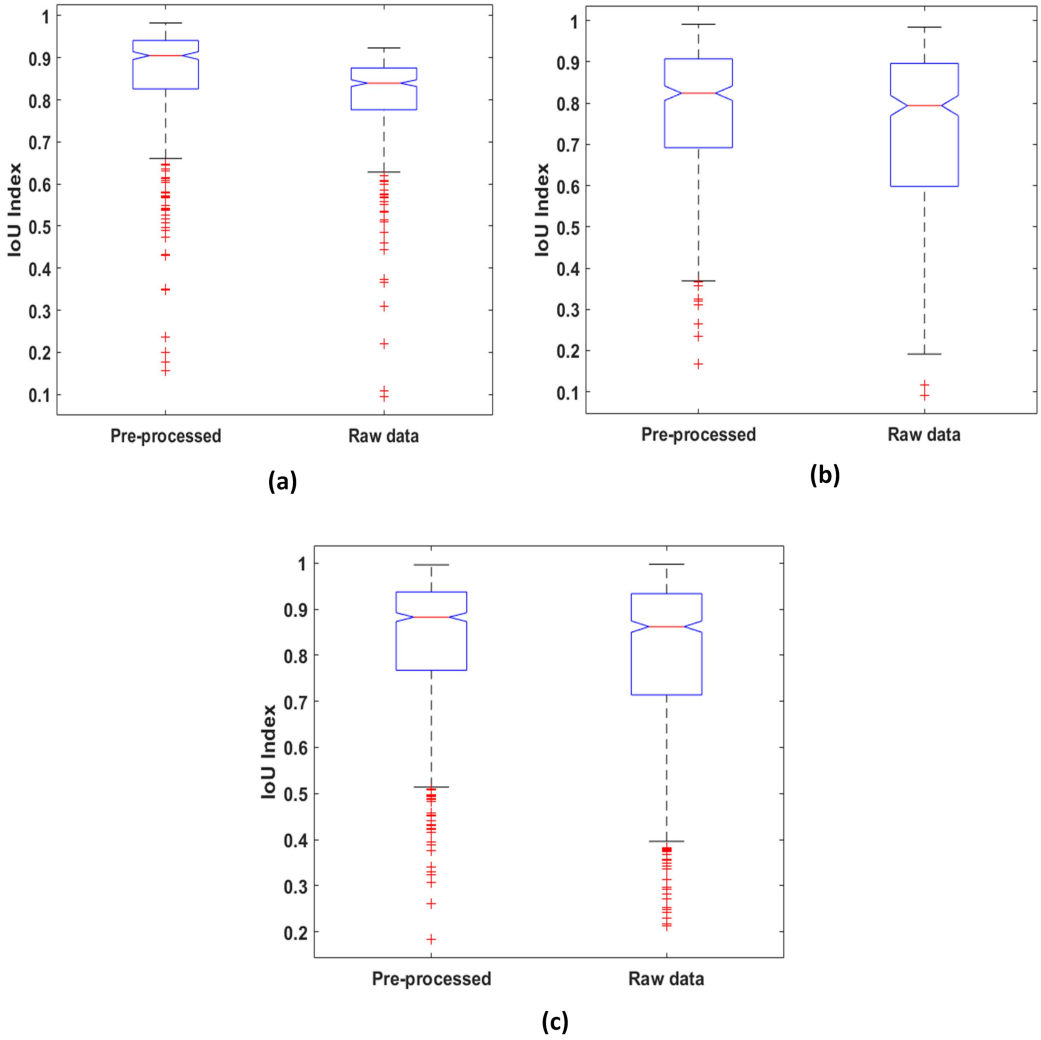
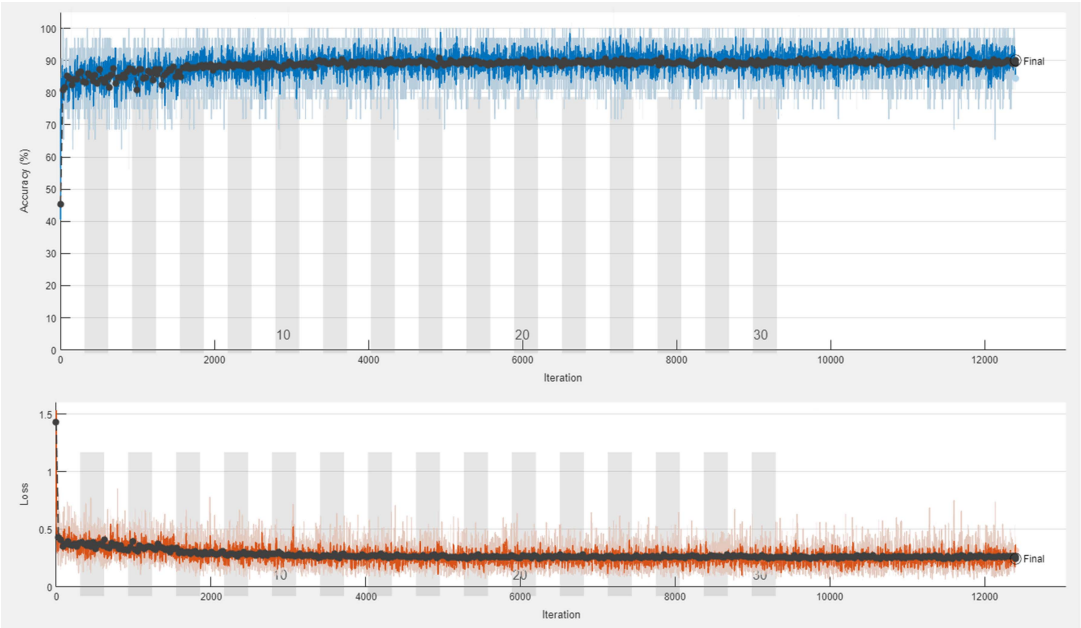
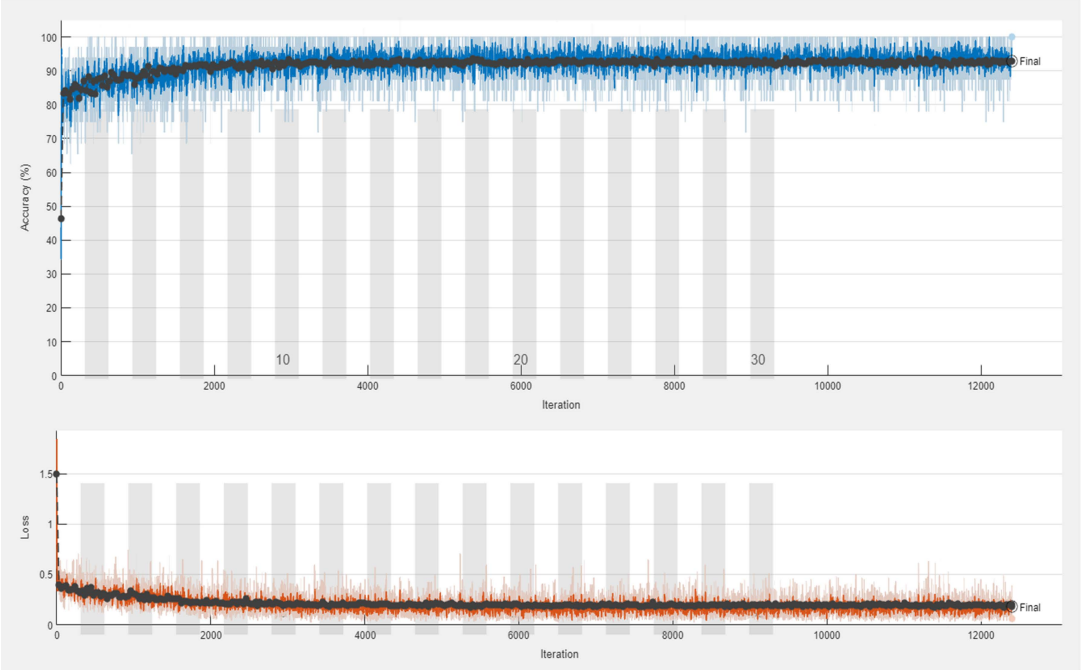



Figure 6.10: ANOVA test on ISIC 2016-2018 test data based on IoU index using DilatedSkinNet.

data is displayed in Figure 6.11. The validation accuracy and loss plot is displayed in Figure 6.12 indicating the highest accuracy and lowest loss values achieved by the LCNet on the pre-processed+segmented images as compared to raw data. LCNet was assessed based on ACC, PRE, REC, SPE and F-Score metrics calculated on a test set for analyzing the performance. Therefore, in Figure 6.13 it is shown that LCNet predicts the highest number of TP and TN on pre-processed+segmented data, whereas




**(a) Training-progress on Raw data**




**(b) Training-progress on pre-processed data**

Figure 6.11: Training progress of LCNet on (a) Raw (b) Pre-processed data.

it is comparatively low on raw data, which is noisy and have poor contrast. The overall performance of the LCNet has greatly improved by applying pre-processing and

segmentation methods.

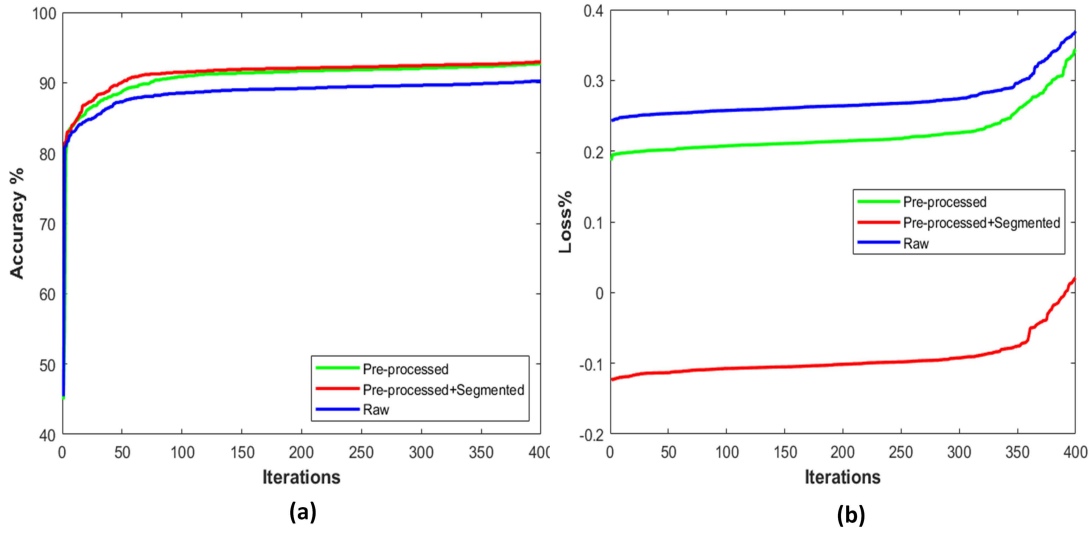



Figure 6.12: Performance of LCNet on different types of validation data: (a) accuracy (b) loss.

Table 6.10: Performance comparison of LCNet on the variations of datasets.

| **Data** | **ACC** | **PRE** | **REC** | **SPE** | **F-Score** | **Execution time (s)** | **Per image test time (s)** |
|---|---|---|---|---|---|---|---|
| Raw | 90.92 | 92.11 | 92.29 | 89.29 | 91.70 | 76 | 1.8 |
| Pre-processed | 92.47 | 93.80 | 92.62 | 92.28 | 93.21 | 19 | 1.1 |
| Pre-processed + Segmented | **93.40** | **93.45** | **94.51** | **92.08** | **93.98** | 17 | 1.3 |

Note: Execution time is on complete test set. The best values are marked in bold.

From the statistical results presented in Table 6.10, it is observed that the performance of LCNet is significantly improved using external pre-processing operations, such as the removal of hairlines, increasing contrast and lesion segmentation. The network efficiently analyzes very complex lesion patterns from enhanced images by computing rich features at multiple layers. On raw data, the model gained an ACC of 90.92% but increased to 92.47% and 93.40% using pre-processed, and pre-processed+segmented data respectively. In addition, the inference time on raw images is high as 76 seconds, while processed and segmented images only require 19 and 17 seconds correspondingly for the same test set of 2577 samples. The reason for low inference time is that networks

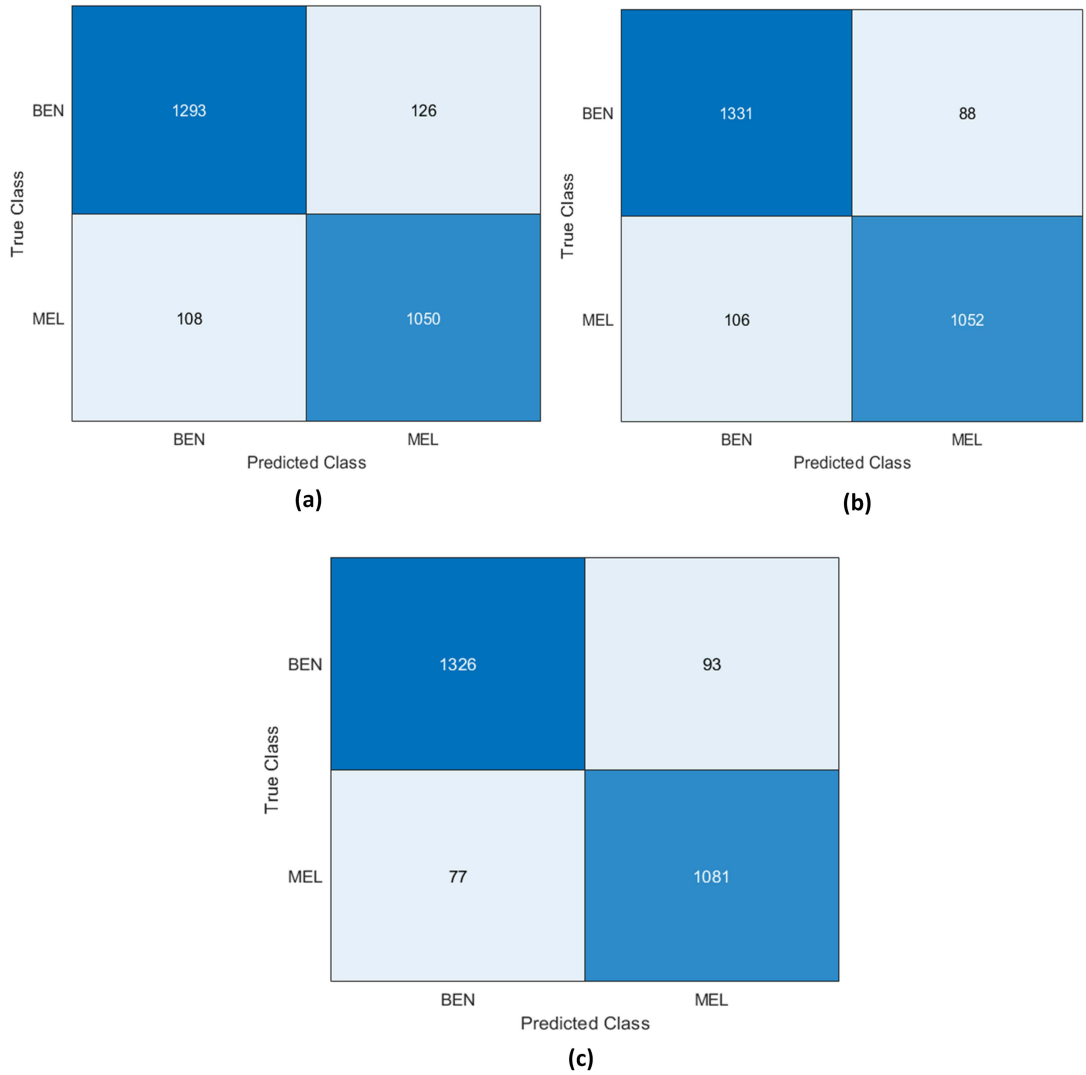



Figure 6.13: Confusion matrix on variations of data (a) Raw (b) Pre-processed (c) Pre-processed + Segmented.

take less time in computing lesion features from an image having ROI only rather than processing the whole original image. The test time on single image is quite low using LCNet for melanoma classification as 1.8 seconds on raw, 1.1 seconds on pre-processed and 1.3 seconds on pre-processed+segmented image. Finally, in comparison to the existing CNN models and state-of-the-art studies, LCNet showed improved accuracy and reduced architectural and computational complexity. In addition, an evaluation of the LCNet on a large and enhanced dermoscopic dataset revealed that it performed well

despite irrelevant visual indicators and imaging artifacts.

## 6.7 Conclusion

Skin cancer is a global health problem, and the development of an automatic melanoma detection system plays a significant role in its early diagnosis. The LCNet model, inspired by the deep convolutional neural network for skin cancer classification, was trained in an end-to-end manner on dermoscopic skin cancer images. Two different datasets, ISIC and HAM10000, were incorporated to perform the experiments. It was challenging to establish an automatic framework to classify melanoma lesions due to high inter-similarities and intra-class variations. With the design of a few pre-processing steps such as image resizing, oversampling, and augmentation, an accurate model, was designed for melanoma classification. The experimental results showed that the proposed model achieved higher performance than the selected studies and pre-trained classification models. The LCNet model is reliable in predicting the correct lesion category with a high true positive rate, thus strongly satisfying the use of AI in solving medical problems as a diagnostic tool. The performance of LCNet is also evaluated on pre-processed datasets, including denoised and segmented data using the IA-HR method and DilatedSkinNet model. LCNet's performance was significantly improved by the pre-processed data, reflecting the significance of employing pre-processing and segmentation models on raw data to improve its visual representation. The following chapter presents the concluding remarks and suggests directions for future work.

# Chapter 7

# Discussion and Conclusions

A major focus of this thesis was the development of a computer-aided melanoma diagnosis system based on dermoscopic skin cancer images using deep learning techniques. Furthermore, a variety of pre-processing techniques have been developed to obliterate hairlines with image processing operations and segment lesion locations using convolutional neural networks. A brief introduction is presented in this chapter, along with the significance of conducting this research and response to the research questions. Throughout chapter 2, we discuss the background details and state-of-the-art methods related to the pre-processing, segmentation, and classification stages. The overall system design was explained in chapter 3, explaining various steps that were conducted in this research. In addition, the detail of the benchmark datasets used for the experimental study were explained. As presented in chapter 4, we implemented a hair elimination method that effectively removed hairlines from lesion images and improved their color contrast using MCACNN deep learning. Our approaches outperformed other related studies when the results were compared based on image quality metrics.

Chapter 5 presented a deep learning-based segmentation network for separating lesions from background regions. Two deep neural networks with different architectures were investigated in this phase, one employs encoder-decoder and the other relies on

dilated convolutions. An experimental study led to the identification of the best network for detecting lesions with high accuracy and Jaccard score. In addition to this, we implemented a machine learning based optimised K-means method to extract the lesion region. The most interesting finding was that machine learning based methods were not able to segment complex skin lesion patterns and were highly dependent upon noise-free images having high resolution. Deep learning, on the other hand, does not require any external pre-processing operations. Comparing the two novel deep learning-based segmentation networks, we achieved high performance with DilatedSkinNet over EDNet model. In chapter 6, we discussed a novel framework for identifying melanoma versus benign cancer using LCNet novel framework. This chapter describes our contributions to the field and explores future possibilities.

## 7.1 Contributions and Outcomes

The contributions of this research are as follows:

- **Development of efficient hair removal technique**

  In original skin cancer dataset, images have a lot of hairlines and poor resolution, which adversely affect subsequent phases of the CAD system. In particular, hairlines make it difficult to identify precise lesion regions during segmentation, especially when using machine learning methods. Human inspection of hairlines in images is not likely to cause problems, but automatic image analysis may be affected. The explanation of hair removal method (IA-HR) using morphological operations was presented in chapter 4. The novel method we developed has been designed especially for skin lesion images that detect hairlines pixels and paint them with the average intensity of neighbouring pixels. The purpose of implementing this method was to generate images that are clean and high in resolution. A further investigation was conducted to determine whether the

segmentation and classification network would perform better when presented with clean images.

- **Machine vs Deep learning for segmentation**

  To examine the most effective approach for segmentation tasks, we explored machine learning and deep learning options. We investigated several approaches at this point, including optimized K-means, encoder-decoder-based CNNs, and dilated convolutions. The use of deep learning for automatic lesion segmentation was demonstrated to be effective. All of these techniques were trained on large datasets, the network's performance was monitored on validation data and tested on unseen data. In the first method, traditional K-means was improved by hybridizing an optimization algorithm called FA by incorporating an average intensity function. We observed that this method was very dependent upon noise-free, high contrast data, so we used the IA-HR method to produce clean and non-hairy images as input. In contrast to deep learning, this method was not able to achieve high performance. Furthermore, it lacks the ability to learn, so each image in the dataset must be applied separately.

- **Effective lesion segmentation approach**

  For segmentation using deep learning, two convolutional neural network designs were explored. The first step consisted of designing an EDNet framework that added novelty by providing its unique design of encoder and decoder modules. The advantage of this network was that it produced a high-resolution segmentation map and did not lose spatial information. In addition to this, it has given a high performance on raw images which were noisy. The second segmentation model, DilatedSkinNet incorporates atrous convolutions in the layers of a deep neural network. The dilation rate helps to obtain a wider spatial view of the feature map. Our study compared these three approaches and concluded that DilatedSkinNet

based CNNs performed better than optimized K-means and EDNet. The EDNet achieved a IoU score of 88.7%, 77.0%, 85.5% on the ISIC 2016, 2017, and 2018 test set. Optimized K-means ranked lower with an average IoU score of 49.8%, 48.17%, and 47.80% on ISIC 2016, 2017 and 2018 test set respectively. In comparison, DilatedSkinNet gained 90.4%, 79.8%, 89.1% of IoU index on the ISIC 2016, 2017, and 2018 test sets. There were two main objectives for implementing this phase. Firstly, the segmentation network can be used to generate ground truth images that have been manually annotated by experts in the current practice and is very time consuming. Secondly, this phase's output can be used to improve the diagnostic system's performance during the classification phase.

- **Development of accurate classification approach**

  The novel development of a deep learning classifier (LCNet) that can distinguish between melanoma and non-melanoma lesions was discussed in chapter 6. Our investigation of several different pre-trained networks has included ResNet18, AlexNet, and Inceptionv3 to classify melanoma. Experimental results indicated superior performance of the LCNet network over existing networks. LCNet is a novel layered design developed specifically for melanoma detection using dermoscopic images of skin cancer. Training of the network was done with the augmented dataset. A separate validation dataset was used for monitoring and tuning hyperparameters, and an unknown dataset was used for evaluating them. We carefully built the network architecture after a series of trials, adding specific layers in a particular fashion, deciding the number of kernels at each layer and the depth of the network. To prove the robustness of the network, we evaluated it on two large datasets, ISIC and HAM10000 and achieved an average accuracy of $90.92 \pm 1.0\%$ for two-class (MEL vs BEN) and $93.74 \pm 1.0\%$ for three-class

(MEL vs BEN vs UNK) respectively.

- **Analysis of using pre-processed data**

  Finally, we evaluated the performance of the DilatedSkinNet using pre-processed data, and the performance of the LCNet classifier using raw, pre-processed, and pre-processed+segmented data. According to the experimental findings, DilatedSkinNet performed slightly better with pre-processed data. Moreover, the LCNet network has proven effective in classifying pre-processed images with high accuracy $93.40\%$ as compared to noisy, complicated, and hairy original images. Apart from that, it does not require any special feature extraction process as LCNet automatically computes rich feature set through its deep layers to distinguish melanoma. In general, the pre-processing and segmentation algorithms proposed in this study can be used as a pre-processing step for increasing performance of deep learning-based classifier and can also be used in machine learning-based traditional classifiers since those classifiers rely heavily on high-contrast and cleaned datasets.

## 7.2 Concluding Remarks

This research has achieved its objective by developing a method for classifying melanoma and benign skin cancers using the deep learning technique. The designed classifier has been proven successful in performing an initial assessment of malignant skin cancer from the dermoscopic picture of the lesion without the need for manual pain-inducing procedures. Moreover, the proposed deep learning-based classification technique does not require the hand-crafted feature extraction step computation. The network was trained end-to-end to identify test samples with less inference time and high accuracy. The deep learning-based classification model is more effective, fast, and accurate than

traditional machine learning-based supervised classifiers. While designing an efficient classification model was challenging, the main challenge was building a lightweight network that could be trained with a large amount of data to make it more robust. To the best of our knowledge, the proposed classification network has novel architecture and is less complex than several studies in the literature.

The research has also proposed novel noise removal methods such as removing hairlines that can obscure critical information when analyzing skin cancers. The hairlines are usually shaved or removed from the infected skin area through excruciating procedures that cause discomfort and stress to patients. This software-based method can replace invasive procedures for removing hairlines from skin images without causing any pain to the patient.

Furthermore, the novel segmentation method was designed to separate the lesion objects from their background. An accurate lesion mask is generated using a deep learning model despite complexities present in the images, such as ruler marks, gel bubbles, and hairs. For accurate localization of the cancerous region of a skin lesion, timely detection and delineation of lesion boundaries are fundamental. To assess the structural characteristics of lesions, experts manually annotate ground truth masks to understand the primary assessment criteria, which include size, diameter, and asymmetry. Therefore, the proposed segmentation model can be successfully used to generate a ground truth mask for skin samples with clear boundaries and whose structure can be easily assessed.

This study's potential outcome and benefits include a robust classification model to diagnose malignant skin cancer in minimum time. Hence, it enables practical usability for real-time analysis, such as through mobile and embedded platforms, and a safer and potentially more convenient approach for monitoring health. An accurate and robust classification model developed through deep learning eliminates the need to train the model on every new patient or perform calibrations to maintain accuracy.

Consequently, this methodology is more likely to be utilized in healthcare facilities. As a result of the CNN framework's flexibility to handle highly complex lesion samples, the proposed model can be efficiently applied to hand-held device applications. Hand-held devices may be able to use it due to their ability to learn features without calibration or supervision. The expensive process of feature extraction does not need to be utilized either.

## 7.3 Future Work

Each step of the proposed method has demonstrated high potential towards skin cancer analysis and diagnosis based on experimental results. There are several opportunities to pursue further research in this area. The following are a few of them:

- Implementation of LCNet on the hardware platform to develop portable mobile device for easy and fast diagnosis assessment.
- Employing high-computing devices such as GPUs with more power for training and testing a large amount of data to increase the performance of the LCNet.
- Investigating hand-crafted features and combining them with deep neural features can be experimented for better performance.
- Use of clinical images for the evaluation of LCNet and comparison of its performance with the proposed system.
- Utilization of classification network for resource-constrained implementation, such as embedded systems.

# References


[1] National Cancer Institute, "Skin anatomy," https://visualsonline.cancer.gov/details.cfm?imageid=4604, 2010, Accessed on: 24 April 2020.

[2] American Cancer Society, "About melanoma skin cancer," http://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma, Accessed on: 23 March 2019.

[3] F. Bogo, F. Peruch, A. Fortina, E. Peserico, M. Celebi, T. Mendonca, and J. Marques, "Where's the lesion? variability in human and automated segmentation of dermoscopy images of melanocytic skin lesions," pp. 67–96, 2015.

[4] C. Curiel-Lewandrowski, S. C. Chen, S. M. Swetter, M. P. W. G.-P. S. L. Sub-Committee *et al.*, "Screening and prevention measures for melanoma: Is there a survival advantage?" *Current Oncology Reports*, vol. 14, no. 5, pp. 458–467, 2012.

[5] W. H. Organization, "Radiation: Ultraviolet (UV) radiation and skin cancer," https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer, 2017, Accessed on: 21 October 2019.

[6] American Cancer Society's (ACS), "Cancer facts and figures 2021," https://www.cancer.net/cancer-types/melanoma/statistics, Accessed on: 10 January 2021.

[7] P. Fontanillas, B. Alipanahi, N. A. Furlotte, M. Johnson, C. H. Wilson, S. J. Pitts, R. Gentleman, and A. Auton, "Disease risk scores for skin cancers," *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021.

[8] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: Cancer today. Lyon: International agency for research on cancer; 2018," 2020.

[9] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 3, pp. 177–193, 2017.

[10] K. D. Miller, A. Goding Sauer, A. P. Ortiz, S. A. Fedewa, P. S. Pinheiro, G. Tortolero-Luna, D. Martinez-Tyson, A. Jemal, and R. L. Siegel, "Cancer

statistics for hispanics/latinos, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 425–445, 2018.

[11] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[12] R. L. Siegel, K. D. Miller, A. Goding Sauer, S. A. Fedewa, L. F. Butterly, J. C. Anderson, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 3, pp. 145–164, 2020.

[13] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[14] The Skin Cancer Foundation, "Skin cancer facts and statistics," https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/, Accessed on: 12 January 2021.

[15] M. A. Al-Masni, D.-H. Kim, and T.-S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Computer Methods and Programs in Biomedicine*, vol. 190, pp. 105 351–105 363, 2020.

[16] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler *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 *2018 IEEE 15$^{th}$ International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 168–172.

[17] N. Nami, E. Giannini, M. Burroni, M. Fimiani, and P. Rubegni, "Teledermatology: state-of-the-art and future perspectives," *Expert Review of Dermatology*, vol. 7, no. 1, pp. 1–3, 2012.

[18] E. S. Ruiz, F. C. Morgan, C. M. Zigler, R. J. Besaw, and C. D. Schmults, "Analysis of national skin cancer expenditures in the United States medicare population, 2013," *Journal of the American Academy of Dermatology*, vol. 80, no. 1, pp. 275–278, 2019.

[19] G. P. Guy Jr, S. R. Machlin, D. U. Ekwueme, and K. R. Yabroff, "Prevalence and costs of skin cancer treatment in the US, 2002- 2006 and 2007- 2011," *American Journal of Preventive Medicine*, vol. 48, no. 2, pp. 183–187, 2015.

[20] G. P. Guy Jr, D. U. Ekwueme, F. K. Tangka, and L. C. Richardson, "Melanoma treatment costs: a systematic review of the literature, 1990–2011," *American Journal of Preventive Medicine*, vol. 43, no. 5, pp. 537–545, 2012.

[21] S. E. Umbaugh, R. H. Moss, W. V. Stoecker, and G. A. Hance, "Automatic color segmentation algorithms-with application to skin tumor feature identification," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, no. 3, pp. 75–82, 1993.

[22] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 9, pp. 837–845, 1994.

[23] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.

[24] A. Masood and A. Ali Al-Jumaily, "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms," *International Journal of Biomedical Imaging*, vol. 2013, pp. 1–22, 2013.

[25] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2015, pp. 118–126.

[26] J. Scharcanski and M. E. Celebi, *Computer vision techniques for the diagnosis of skin cancer*. Springer, 2013.

[27] A. Murugan, S. A. H. Nair, and K. S. Kumar, "Detection of skin cancer using svm, random forest and knn classifiers," *Journal of Medical Systems*, vol. 43, no. 8, pp. 1–9, 2019.

[28] A. Victor and M. Ghalib, "Automatic detection and classification of skin cancer," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 3, pp. 444–451, 2017.

[29] I. A. Ozkan and M. Koklu, "Skin lesion classification using machine learning algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 4, pp. 285–289, 2017.

[30] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2016.

[31] M. A. Marchetti, N. C. Codella, S. W. Dusza, D. A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M. E. Celebi, J. L. DeFazio *et al.*, "Results of the 2016 international skin imaging collaboration international symposium on biomedical

imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images," *Journal of the American Academy of Dermatology*, vol. 78, no. 2, pp. 270–277, 2018.

[32] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk *et al.*, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[33] J. McGrath, R. Eady, and F. Pope, "Anatomy and organization of human skin," *Rook's Textbook of Dermatology*, vol. 1, pp. 3–2, 2004.

[34] Massey University of New Zealand, "Environmental health indicators new zealand," https://www.ehinz.ac.nz/indicators/uv-exposure/melanoma/, Accessed on: 10 July 2020.

[35] P. Autier, D. for Epimel, and J.-F. Eortc Melanoma Cooperative Group, "Influence of sun exposures during childhood and during adulthood on melanoma risk," *International Journal of Cancer*, vol. 77, no. 4, pp. 533–537, 1998.

[36] Derma Plus, "Skin cancer-index 2018," https://derma.plus/en/skin-cancer-index-2018/, Accessed on: 10 Mar 2019.

[37] American Cancer Society, "Cancer facts and figures," https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/, Accessed on: 16 July 2021.

[38] Australian Government Cancer Australia, "Melanoma of the skin," https://www.canceraustralia.gov.au/cancer-types/melanoma/statistics, Accessed on: 5 January 2022.

[39] Australian Institute of Health and Welfare, "Cancer in australia 2019," https://www.aihw.gov.au/reports/cancer/cancer-in-australia-2019/summary, Accessed on: 6 January 2021.

[40] D. Schadendorf, A. C. van Akkooi, C. Berking, K. G. Griewank, R. Gutzmer, A. Hauschild, A. Stang, A. Roesch, and S. Ugurel, "Melanoma," *The Lancet*, vol. 392, no. 10151, pp. 971–984, 2018.

[41] J. Dinnes, J. J. Deeks, N. Chuchu, L. F. di Ruffano, R. N. Matin, D. R. Thomson, K. Y. Wong, R. B. Aldridge, R. Abbott, M. Fawzy *et al.*, "Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults," *Cochrane Database of Systematic Reviews*, no. 12, 2018.

[42] J. Kato, K. Horimoto, S. Sato, T. Minowa, and H. Uhara, "Dermoscopy of melanoma and non-melanoma skin cancers," *Frontiers in Medicine*, vol. 6, pp. 180–187, 2019.

[43] G. Kamińska-Winciorek and W. Placek, "The most common mistakes on dermatoscopy of melanocytic lesions," *Advances in Dermatology and Allergology*, vol. 32, no. 1, p. 33, 2015.

[44] G. Di Leo, A. Paolillo, P. Sommella, and G. Fabbrocini, "Automatic diagnosis of melanoma: A software system based on the 7-point check-list," in *2010 43$^{rd}$ Hawaii international conference on system sciences*. IEEE, 2010, pp. 1–10.

[45] Z. Zhang, W. V. Stoecker, and R. H. Moss, "Border detection on digitized skin tumor images," *IEEE transactions on Medical Imaging*, vol. 19, no. 11, pp. 1128–1143, 2000.

[46] S. M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation of dermoscopic images using GrabCut and K-means algorithms," *IET Computer Vision*, vol. 12, no. 8, pp. 1088–1095, 2018.

[47] J. L. Garcia-Arroyo and B. Garcia-Zapirain, "Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding," *Computer Methods and Programs in Biomedicine*, vol. 168, pp. 11–19, 2019.

[48] J. A. Salido and C. Ruiz Jr, "Hair artifact removal and skin lesion segmentation of dermoscopy images," *Asian Journal of Pharmaceutical and Clinical Research*, vol. 11, no. 3, 2018.

[49] S. Joseph and J. R. Panicker, "Skin lesion analysis system for melanoma detection with an effective hair segmentation method," in *2016 International Conference on Information Science (ICIS)*. IEEE, 2016, pp. 91–96.

[50] I. S. Zaqout, "An efficient block-based algorithm for hair removal in dermoscopic images," *Computer Optics*, vol. 41, no. 4, 2017.

[51] S. Khalid, U. Jamil, K. Saleem, M. U. Akram, W. Manzoor, W. Ahmed, and A. Sohail, "Segmentation of skin lesion using cohen - daubechies - feauveau biorthogonal wavelet," *SpringerPlus*, vol. 5, no. 1, pp. 1–17, 2016.

[52] Q. Abbas, I. F. Garcia, M. Emre Celebi, and W. Ahmad, "A feature-preserving hair removal algorithm for dermoscopy images," *Skin Research and Technology*, vol. 19, no. 1, pp. e27–e36, 2013.

[53] M. T. B. Toossi, H. R. Pourreza, H. Zare, M.-H. Sigari, P. Layegh, and A. Azimi, "An effective hair removal algorithm for dermoscopy images," *Skin Research and Technology*, vol. 19, no. 3, pp. 230–235, 2013.

[54] R. Kasmi, K. Mokrani, R. Rader, J. Cole, and W. Stoecker, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," *Skin Research and Technology*, vol. 22, no. 2, pp. 208–222, 2016.

[55] 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.

[56] M. Sridevi and C. Mala, "A survey on monochrome image segmentation methods," *Procedia Technology*, vol. 6, pp. 548–555, 2012.

[57] 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.

[58] F. Xie, Y. Li, R. Meng, and Z. Jiang, "No-reference hair occlusion assessment for dermoscopy images based on distribution feature," *Computers in Biology and Medicine*, vol. 59, pp. 106–115, 2015.

[59] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE transactions on medical imaging*, vol. 20, no. 3, pp. 233–239, 2001.

[60] 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 *Medical Imaging 2010: Image Processing*, vol. 7623. SPIE, 2010, pp. 1436–1443.

[61] D. H. Chung and G. Sapiro, "Segmenting skin lesions with partial-differential-equations-based image processing algorithms," *IEEE transactions on Medical Imaging*, vol. 19, no. 7, pp. 763–767, 2000.

[62] Q. Abbas, I. Fondón, and M. Rashid, "Unsupervised skin lesions border detection via two-dimensional image analysis," *Computer methods and programs in biomedicine*, vol. 104, no. 3, pp. e1–e15, 2011.

[63] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[64] S. Ren, D. K. Jain, K. Guo, T. Xu, and T. Chi, "Towards efficient medical lesion image super-resolution based on deep residual networks," *Signal processing: Image communication*, vol. 75, pp. 1–10, 2019.

[65] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," *Dermoscopy Image Analysis*, vol. 10, pp. 97–129, 2015.

[66] I. Lee, X. Du, and B. Anthony, "Hair segmentation using adaptive threshold from edge and branch length measures," *Computers in Biology and Medicine*, vol. 89, pp. 314–324, 2017.

[67] C. A. Z. Barcelos and V. Pires, "An automatic based nonlinear diffusion equations scheme for skin lesion segmentation," *Applied Mathematics and Computation*, vol. 215, no. 1, pp. 251–261, 2009.

[68] H. de Souza Ganzeli, J. G. Bottesini, L. de Oliveira Paz, and M. F. S. Ribeiro, "SKAN: Skin scanner-system for skin cancer detection using adaptive techniques," *IEEE Latin America Transactions*, vol. 9, no. 2, pp. 206–212, 2011.

[69] P. Sabouri, H. GholamHosseini, and J. Collins, "Border detection of skin lesions on a single system on chip," in *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*. Springer, 2014, pp. 465–471.

[70] J. Yasmin and M. Sathik, "An improved iterative segmentation algorithm using canny edge detector for skin lesion border detection," *International Arab Journal of Information Technology (IAJIT)*, vol. 12, no. 4, pp. 325–332, 2015.

[71] K.-A. Norton, H. Iyatomi, M. E. Celebi, S. Ishizaki, M. Sawada, R. Suzaki, K. Kobayashi, M. Tanaka, and K. Ogawa, "Three-phase general border detection method for dermoscopy images using non-uniform illumination correction," *Skin Research and Technology*, vol. 18, no. 3, pp. 290–300, 2012.

[72] Q. Abbas, I. F. Garcia, M. Emre 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. e490–e497, 2013.

[73] 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.

[74] P. G. Cavalcanti and J. Scharcanski, "A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images," *Computer Methods and Programs in Biomedicine*, vol. 112, no. 3, pp. 684–693, 2013.

[75] A. Gupta, A. Issac, M. K. Dutta, and H.-H. Hsu, "Adaptive thresholding for skin lesion segmentation using statistical parameters," in *2017 31$^{st}$ International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2017, pp. 616–620.

[76] S. Jain, N. Pise *et al.*, "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, pp. 735–740, 2015.

[77] H. Wang, X. Chen, R. H. Moss, R. J. Stanley, W. V. Stoecker, M. E. Celebi, T. M. Szalapski, J. M. Malters, J. M. Grichnik, A. A. Marghoob *et al.*, "Watershed segmentation of dermoscopy images using a watershed technique," *Skin Research and Technology*, vol. 16, no. 3, pp. 378–384, 2010.

[78] H. Castillejos, V. Ponomaryov, L. Nino-de Rivera, and V. Golikov, "Wavelet transform fuzzy algorithms for dermoscopic image segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2012, 2012.

[79] S. Sookpotharom, "Border detection of skin lesion images based on fuzzy C-means thresholding," in *2009 Third International Conference on Genetic and Evolutionary Computing*. IEEE, 2009, pp. 777–780.

[80] H. Zhou, G. Schaefer, A. H. Sadka, and M. E. Celebi, "Anisotropic mean shift based fuzzy C-means segmentation of dermoscopy images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 26–34, 2009.

[81] F. F. X. Vasconcelos, A. G. Medeiros, S. A. Peixoto, and P. P. Reboucas Filho, "Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour," *Cognitive Systems Research*, vol. 55, pp. 44–59, 2019.

[82] M. Bayraktar, S. Kockara, T. Halic, M. Mete, H. K. Wong, and K. Iqbal, "Local edge-enhanced active contour for accurate skin lesion border detection," *BMC Bioinformatics*, vol. 20, no. 2, pp. 87–97, 2019.

[83] T. Y. Tan, L. Zhang, C. P. Lim, B. Fielding, Y. Yu, and E. Anderson, "Evolving ensemble models for image segmentation using enhanced particle swarm optimization," *IEEE access*, vol. 7, pp. 34 004–34 019, 2019.

[84] R. B. Oliveira, E. Mercedes Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: A review," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 127–141, 2016.

[85] J. Jiang, P. Trundle, and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 617–631, 2010.

[86] M. J. Moghaddam and H. Soltanian-Zadeh, "Medical image segmentation using artificial neural networks," *Artificial Neural Networks-Methodological Advances and Biomedical Applications*, pp. 121–138, 2011.

[87] X. Yuan, N. Situ, and G. Zouridakis, "Automatic segmentation of skin lesion images using evolution strategies," *Biomedical Signal Processing and Control*, vol. 3, no. 3, pp. 220–228, 2008.

[88] N. Dey, V. Rajinikanth, A. S. Ashour, and J. M. R. Tavares, "Social group optimization supported segmentation and evaluation of skin melanoma images," *Symmetry*, vol. 10, no. 2, pp. 51–72, 2018.

[89] M. Aljanabi, Y. E. Özok, J. Rahebi, and A. S. Abdullah, "Skin lesion segmentation method for dermoscopy images using artificial bee colony algorithm," *Symmetry*, vol. 10, no. 8, pp. 347–365, 2018.

[90] C. Cernazanu-Glavan and S. Holban, "Segmentation of bone structure in X-ray images using convolutional neural network," *Adv. Electr. Comput. Eng*, vol. 13, no. 1, pp. 87–94, 2013.

[91] M. Melinscak, P. Prentasic, and S. Loncaric, "Retinal vessel segmentation using deep neural networks." in *VISAPP (1)*, 2015, pp. 577–582.

[92] 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*, 2015, pp. 3431–3440.

[93] S. Hong, H. Noh, and B. Han, "Decoupled deep neural network for semi-supervised semantic segmentation," *arXiv preprint arXiv:1506.04924*, 2015.

[94] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[95] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[96] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[97] E. Okur and M. Turkan, "A survey on automated melanoma detection," *Engineering Applications of Artificial Intelligence*, vol. 73, pp. 50–67, 2018.

[98] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[99] M. A. Al-Masni, M. A. Al-Antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221–231, 2018.

[100] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *2016 IEEE 13$^{th}$ International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1397–1400.

[101] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognition*, vol. 85, pp. 78–89, 2019.

[102] S. Vesal, N. Ravikumar, and A. Maier, "SkinNet: A deep learning framework for skin lesion segmentation," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2018, pp. 1–3.

[103] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.

[104] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, "A novel multi-task deep learning model for skin lesion segmentation and classification," *arXiv preprint arXiv:1703.01025*, 2017.

[105] J. Kawahara and G. Hamarneh, "Fully convolutional neural networks to detect clinical dermoscopic features," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 578–585, 2018.

[106] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, and M. T. Mahmood, "Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering," *International Journal of Medical Informatics*, vol. 124, pp. 37–48, 2019.

[107] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *2016 23$^{rd}$ International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 337–342.

[108] X. Zhang, "Melanoma segmentation based on deep learning," *Computer Assisted Surgery*, vol. 22, no. sup1, pp. 267–277, 2017.

[109] Y. Guo, A. S. Ashour, and F. Smarandache, "A novel skin lesion detection approach using neutrosophic clustering and adaptive region growing in dermoscopy images," *Symmetry*, vol. 10, no. 4, pp. 119–138, 2018.

[110] F. Riaz, S. Naeem, R. Nawaz, and M. Coimbra, "Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 489–500, 2018.

[111] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation," *Informatics in Medicine Unlocked*, vol. 18, pp. 100 297–100 309, 2020.

[112] M. Vidya and M. V. Karki, "Skin cancer detection using machine learning techniques," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. IEEE, 2020, pp. 1–5.

[113] A. Rezvantalab, H. Safigholi, and S. Karimijeshni, "Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms," *arXiv preprint arXiv:1810.10348*, 2018.

[114] J. Premaladha and K. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of Medical Systems*, vol. 40, no. 4, pp. 1–12, 2016.

[115] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[116] M. Dildar, S. Akram, M. Irfan, H. U. Khan, M. Ramzan, A. R. Mahmood, S. A. Alsaiari, A. H. M. Saeed, M. O. Alraddadi, and M. H. Mahnashi, "Skin cancer detection: A review using deep learning techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, pp. 5479–5501, 2021.

[117] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. Von Kalle, "Skin cancer classification using convolutional neural networks: systematic review," *Journal of Medical Internet Research*, vol. 20, no. 10, pp. e11 936–11 944, 2018.

[118] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *International Biomedical Engineering Conference (CIBEC)*. IEEE, 2018, pp. 90–93.

[119] J. R. Hagerty, R. J. Stanley, H. A. Almubarak, N. Lama, R. Kasmi, P. Guo, R. J. Drugge, H. S. Rabinovitz, M. Oliviero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1385–1391, 2019.

[120] R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber, M. Bakhtyar, I. Mehmood, O.-Y. Song, and M. Maqsood, "Region-of-interest based transfer learning assisted framework for skin cancer detection," *IEEE Access*, vol. 8, pp. 147 858–147 871, 2020.

[121] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, pp. 100 036–44, 2021.

[122] L. Alzubaidi, M. Al-Amidie, A. Al-Asadi, A. J. Humaidi, O. Al-Shamma, M. A. Fadhel, J. Zhang, J. Santamaría, and Y. Duan, "Novel transfer learning approach for medical imaging with limited labeled data," *Cancers*, vol. 13, no. 7, pp. 1590–1612, 2021.

[123] D. N. Le, H. X. Le, L. T. Ngo, and H. T. Ngo, "Transfer learning with class-weighted and focal loss function for automatic skin cancer classification," *arXiv preprint arXiv:2009.05977*, 2020.

[124] S. H. Kassani and P. H. Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, pp. 76–83, 2019.

[125] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, 2018.

[126] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1229–1233.

[127] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data," *MethodsX*, vol. 7, pp. 100 864–100 872, 2020.

[128] A. G. Pacheco, A.-R. Ali, and T. Trappenberg, "Skin cancer detection based on deep learning and entropy to detect outlier samples," *arXiv preprint arXiv:1909.04525*, 2019.

[129] T. D. Tô, D. T. Lan, T. T. H. Nguyen, T. T. N. Nguyen, H.-P. Nguyen, L. Phuong, and T. Z. Nguyen, "Ensembled skin cancer classification (ISIC 2019 challenge submission)," Ph.D. dissertation, ISIC2019, 2019.

[130] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, pp. 101 843–101 853, 2021.

[131] I. Gonzalez-Diaz, "Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 547–559, 2018.

[132] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2092–2103, 2019.

[133] M. A. A. Milton, "Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018: Skin lesion analysis towards melanoma detection challenge," *arXiv preprint arXiv:1901.10802*, 2019.

[134] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, pp. 556–572, 2018.

[135] Z. Lai and H. Deng, "Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[136] J. Höhn, A. Hekler, E. Krieghoff-Henning, J. N. Kather, J. S. Utikal, F. Meier, F. F. Gellrich, A. Hauschild, L. French, J. G. Schlager *et al.*, "Integrating patient data into skin cancer classification using convolutional neural networks: Systematic review," *Journal of Medical Internet Research*, vol. 23, no. 7, pp. e20 708–20 722, 2021.

[137] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, E. Chousakos, N. Codella, M. Combalia, S. Dusza, P. Guitera, D. Gutman *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific Data*, vol. 8, no. 1, pp. 1–8, 2021.

[138] M. Combalia, N. C. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig *et al.*, "BCN20000: Dermoscopic lesions in the wild," *arXiv preprint arXiv:1908.02288*, 2019.

[139] 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.

[140] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "$PH^2$-A dermoscopic image database for research and benchmarking," in *2013 $35^{th}$ Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 5437–5440.

[141] N. Jiang and L. Wang, "Quantum image scaling using nearest neighbor interpolation," *Quantum Information Processing*, vol. 14, no. 5, pp. 1559–1571, 2015.

[142] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[143] Q. Chen, J. Xu, and V. Koltun, "Fast image processing with fully-convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2497–2506.

[144] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[145] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[146] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[147] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[148] Y. Fu and S. Wang, "A no reference image quality assessment metric based on visual perception," *Algorithms*, vol. 9, no. 4, pp. 87–108, 2016.

[149] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics Gems*, pp. 474–485, 1994.

[150] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 839–846.

[151] M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, and A. Yazdabadi, "Realistic hair simulator for skin lesion images: A novel benchemarking tool," *Artificial Intelligence in Medicine*, vol. 108, pp. 101 933–101 945, 2020.

[152] W. Li, A. N. J. Raj, T. Tjahjadi, and Z. Zhuang, "Digital hair removal by deep learning for skin lesion segmentation," *Pattern Recognition*, vol. 117, pp. 107 994–108 009, 2021.

[153] L. Talavera-Martinez, P. Bibiloni, and M. Gonzalez-Hidalgo, "Hair segmentation and removal in dermoscopic images using deep learning," *IEEE Access*, vol. 9, pp. 2694–2704, 2020.

[154] A. Huang, S.-Y. Kwan, W.-Y. Chang, M.-Y. Liu, M.-H. Chi, and G.-S. Chen, "A robust hair segmentation and removal approach for clinical images of skin lesions," in *2013 35$^{th}$ Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 3315–3318.

[155] P. Bibiloni, M. González-Hidalgo, and S. Massanet, "Skin hair removal in dermoscopic images using soft color morphology," in *Conference on artificial intelligence in medicine in Europe*. Springer, 2017, pp. 322–326.

[156] H. Mirzaalian, T. K. Lee, and G. Hamarneh, "Hair enhancement in dermoscopic images using dual-channel quaternion tubularness filters and mrf-based multilabel optimization," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5486–5496, 2014.

[157] G. Ramella, "Hair removal combining saliency, shape and color," *Applied Sciences*, vol. 11, no. 1, pp. 447–475, 2021.

[158] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2017, pp. 379–387.

[159] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[160] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.

[161] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "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)," 2016.

[162] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019.

[163] U. Sanchez, "ISBI 2016 challenge results," https://challenge.kitware.com/submission/56fe2b60cad3a55ecee8cf74, 2016, Accessed on: 16 April 2019.

[164] M. M. Rahman and N. Alpaslan, "Automated melanoma recognition in dermoscopic images based on extreme learning machine (elm)," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. SPIE, 2017, pp. 278–285.

[165] L. Huang, Y.-g. Zhao, and T.-j. Yang, "Skin lesion segmentation using object scale-oriented fully convolutional neural networks," *Signal, Image and Video Processing*, vol. 13, no. 3, pp. 431–438, 2019.

[166] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 186, pp. 105 241–105 251, 2020.

[167] M. K. Hasan, M. T. E. Elahi, M. A. Alam, M. T. Jawad, and R. Martí, "Dermo-Expert: Skin lesion classification using a hybrid convolutional neural network

through segmentation, transfer learning, and augmentation," *Informatics in Medicine Unlocked*, pp. 100 819–100 836, 2022.

[168] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," *arXiv preprint arXiv:1703.05165*, 2017.

[169] M. Berseth, "Isic 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:1703.00523*, 2017.

[170] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," *arXiv preprint arXiv:1703.04197*, 2017.

[171] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. Avila, and E. Valle, "Recod titans at isic challenge 2017," *arXiv preprint arXiv:1703.04819*, 2017.

[172] K. Zafar, S. O. Gilani, A. Waris, A. Ahmed, M. Jamil, M. N. Khan, and A. Sohail Kashif, "Skin lesion segmentation from dermoscopic images using convolutional neural network," *Sensors*, vol. 20, no. 6, pp. 1601–1615, 2020.

[173] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang, and S. Wang, "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Medical Image Analysis*, vol. 64, pp. 101 716–101 728, 2020.

[174] K. Abhishek, G. Hamarneh, and M. S. Drew, "Illumination-based transformations improve skin lesion segmentation in dermoscopic images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 728–729.

[175] C. Qian, T. Liu, H. Jiang, Z. Wang, P. Wang, M. Guan, and B. Sun, "A detection and segmentation architecture for skin lesion segmentation on dermoscopy images," *arXiv preprint arXiv:1809.03917*, 2018.

[176] Hao Du and Jeon Young Seok and Dianwen Ng Ngiam and Kee Yuan and Mengling Feng, "Team holidayburned at ISIC challenge 2018," https://challenge.isic-archive.com/leaderboards/2018, 2018, Accessed on: 18 May 2019.

[177] Y. Ji, X. Li, G. Zhang, D. Lin, and H. Chen, "Automatic skin lesion segmentation by feature aggregation convolutional neural network," 2018.

[178] Yuan Xue and Lijun Gong and Wei Peng and Xiaolei Huang and Yefeng Zheng, "Automatic skin lesion analysis with deep networks," https://challenge.isic-archive.com/leaderboards/2018, 2018, Accessed: 19 June 2019.

[179] R. Ali, R. C. Hardie, M. S. De Silva, and T. M. Kebede, "Skin lesion segmentation and classification for ISIC 2018 by combining deep CNN and handcrafted features," *arXiv preprint arXiv:1908.05730*, 2019.

[180] M. P. Pour and H. Seker, "Transform domain representation-driven convolutional neural networks for skin lesion segmentation," *Expert Systems with Applications*, vol. 144, pp. 113 129–113 141, 2020.

[181] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[182] C. Dhivyaa, K. Sangeetha, M. Balamurugan, S. Amaran, T. Vetriselvi, and P. Johnpaul, "Skin lesion classification using decision trees and random forest algorithms," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2020.

[183] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal, R. C. Maron, C. Berking, S. Haferkamp, J. Klode, D. Schadendorf *et al.*, "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer*, vol. 120, pp. 114–121, 2019.

[184] American cancer society, "Skin cancer," https://www.cancer.org/cancer/skin-cancer.html/, 2019, Accessed on: 16 May 2019.

[185] M. R. Hasan, M. I. Fatemi, M. Monirujjaman Khan, M. Kaur, and A. Zaguia, "Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[186] H.-W. Huang, B. W.-Y. Hsu, C.-H. Lee, and V. S. Tseng, "Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers," *The Journal of Dermatology*, vol. 48, no. 3, pp. 310–316, 2021.

[187] M. A. Khan, Y.-D. Zhang, M. Sharif, and T. Akram, "Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification," *Computers and Electrical Engineering*, vol. 90, pp. 106 956–106 973, 2021.

[188] V. K. Shrivastava, K. Parvathi *et al.*, "Multiclass skin lesion classification using image augmentation technique and transfer learning models," *International Journal of Intelligent Unmanned Systems*, 2021.

[189] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[190] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2019.

[191] Sloan Kettering Cancer Center, "The international skin imaging collaboration," https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main/, 2019, Accessed on: 13 Feb 2019.

[192] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[193] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2020.

[194] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[195] M. A. Khan, K. Muhammad, M. Sharif, T. Akram, and S. Kadry, "Intelligent fusion-assisted skin lesion localization and classification for smart healthcare," *Neural Computing and Applications*, pp. 1–16, 2021.

[196] A. Ameri, "A deep learning approach to skin cancer detection in dermoscopy images," *Journal of Biomedical Physics & Engineering*, vol. 10, no. 6, pp. 801–807, 2020.

[197] F. Afza, M. Sharif, M. Mittal, M. A. Khan, and D. J. Hemanth, "A hierarchical three-step superpixels and deep learning framework for skin lesion classification," *Methods*, vol. 202, pp. 88–102, 2022.

[198] S. Nasiri, J. Helsper, M. Jung, and M. Fathi, "DePicT melanoma Deep-CLASS: A deep convolutional neural networks approach to classify skin lesion images," *BMC Bioinformatics*, vol. 21, no. 2, pp. 1–13, 2020.

[199] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, T. Holland-Letz *et al.*, "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *European Journal of Cancer*, vol. 113, pp. 47–54, 2019.

[200] A. Kwasigroch, M. Grochowski, and A. Mikołajczyk, "Neural architecture search for skin lesion classification," *IEEE Access*, vol. 8, pp. 9061–9071, 2020.

[201] S. Karki, P. Kulkarni, and A. Stranieri, "Melanoma classification using efficientnets and ensemble of models with different input resolution," in *2021 Australasian Computer Science Week Multiconference*, 2021, pp. 1–5.

[202] Q. Ha, B. Liu, and F. Liu, "Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge," *arXiv preprint arXiv:2010.05351*, 2020.

[203] M. O'Brien, J. Bukowski, G. Hager, A. Pezeshk, and M. Unberath, "Evaluating neural network robustness for melanoma classification using mutual information," in *Medical Imaging 2022: Image Processing*, vol. 12032. SPIE, 2022, pp. 173–177.

[204] X.-S. Yang, "Firefly algorithm, levy flights and global optimization," in *Research and Development in Intelligent Systems XXVI*. Springer, 2010, pp. 209–218.

[205] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[206] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

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

# Appendix A

In this chapter, we describe the other approaches used to address the lesion segmentation task within our initial experiments. Firstly, optimization of the K-means method and results are discussed followed by FCNN and SegNet models. These methods are already existing ones that have been modified a little for use on the skin cancer dataset. Based on these methods, the following papers have been published:

- R. Kaur, "From Machine Learning to Deep Learning: Experimental Comparison of Machine Learning and Deep Learning for Skin Cancer Image Segmentation", *AUT Postgraduate Research Symposium, Auckland, New Zealand*, 26 November, 2021.

- R. Kaur, H. GholamHosseini & R. Sinha, "Lesion Border Detection of Skin Cancer Images Using Deep Fully Convolutional Neural Network with Customized Weights", *In* $43^{rd}$ *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico*, IEEE, pp. 3035-3038, 1-5 November, 2021. [Online]. Available: 10.1109/EMBC46164.2021.9630512.

- R. Kaur, H. GholamHosseini & R. Sinha, "Deep Learning in Medical Applications: Lesion Segmentation in Skin Cancer Images Using Modified and Improved Encoder-Decoder Architecture", *International Symposium on Geometry and Vision*, vol. 1386, pp. 39-52, Springer, Cham, 2021. [Online]. Available:

https://doi.org/10.1007/978-3-030-72073-5-4.

## A.1 Optimized K-means for Segmentation

Our experimentation on segmenting lesions from skin cancer images began with machine learning techniques. An example of such a method is the K-means method, which was optimized with the help of a firefly algorithm (FA). K-means clustering approach is based on the principle of identifying the number of clusters and every pixel point assigned to the nearest cluster by calculating Euclidean distance. The clustering is also a kind of an optimization method in which the minimum sum of Euclidean distance between the centroid of clusters and the objects is determined. The K-means algorithm first initialized with $k$ = 2, the number of clusters having centers $c_1, c_2 \ldots\ldots, c_k$, created randomly. For every pixel position in an image, it assigns each point into the cluster having a minimum distance to centroid value, where distance calculated as follows:

$$E = \sum_{i=1}^{K} \sum_{j=1}^{N} |X_i^j - C_j|^2 \tag{A.1}$$

The number of clusters are represented by $K$ for $N$ classes, $X_i^j$ is the position vector of the object $i$ and $C_j$ its centroid for the cluster $j$. The K-means divides the data set into two clusters by using Euclidean distance as the similarity index. This distance should be minimum for pixel positions inside one same cluster as compared to other pixels in different clusters. The K-means performs an iterative process until the centroids do not change. In this study, we observe that the K-means algorithm yielded poor results as a result of the incorrect selection of the initial centroid value, thus FA is combined with K-means to improve its performance. The FA [204] optimization was applied to the K-means segmented outputs to replace any pixels inside the lesion region that were incorrectly classified by the K-means. The K-means segmentation results

were improved by incorporating an average function as the fitness function in the FA algorithm. This function replaces the intensity value in the neighborhood of a pixel with a mean value calculated from the intensity values in the neighborhood.

$$f_{th} = \frac{\sum_{i=1}^{pl} Data(i)}{Feature_{len}}, \quad f_{fit} = \begin{cases} 1 & if\ pixel's\ intensity\ is\ less\ than\ f_{th} \\ 0 & otherwise \end{cases} \tag{A.2}$$

The number of iterations $i$ performed over the segmented lesion image given by K-means until the population limit is reached, denoted by $pl$. The threshold value is calculated by taking the ratio of an array of lesion pixels, $Data(i)$ and features length, $Feature_{len}$. This $f_{th}$ gives average intensity value of neighboring pixels in the lesion region. Thus, fitness function, $f_{fit}$ replaces pixel '1' if its intensity value is less than a threshold computed by fitness function and '0' otherwise. In this manner, FA performs an iteration process for finding those pixels which are incorrectly identified by the K-means and replaces them with the average intensity value of the neighboring pixels.

### A.1.1 Results of optimized K-means

To conduct the experiments using K-means+FA method, three well-known benchmark datasets ISIC 2016, ISIC 2017 and ISIC 2018 containing dermoscopic images were used. The dataset ISIC 2016 contains 379 testing samples, ISIC 2017 contains 2000 testing images, and ISIC 2018 consists of 778 test samples. The evaluation parameters computed were ACC, DICE, JAC or IoU, SEN, and SPE to analyze the performance. We kept the same image size of $192 \times 256$ in this case as we did in Chapter 5 when discussing proposed segmentation methods. The input skin cancer images first pre-processed using IA-HR method (See chapter 4) to eliminate hairlines, then optimized K-means was applied to extract lesion region. The outcome of this method on ISIC 2016 and ISIC 2017 dataset is given in Figure A.1 and Table A.1. From the results, it

is analysed that K-means is not efficient in separating lesion from its background. It produced partially segmented lesions which is not good for recognition task. Moreover, the IoU score and other performance metrics are quite low as compared to deep learning networks developed in Chapter 5. This method also consumes high execution time compared to the DilatedSkinNet. Hence, it is clear that deep learning networks performs better than machine learning based supervised method for skin lesion segmentation task. This machine learning-based approach is not efficient to process complex lesion patterns.

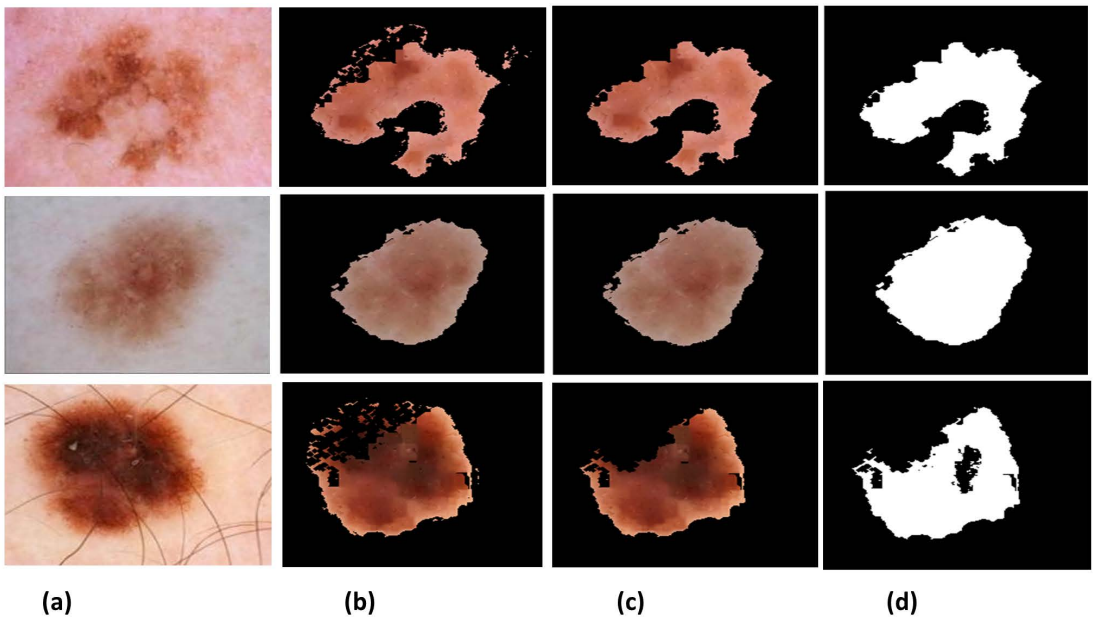



Figure A.1: Segmentation results of optimized K-means: (a) Original images (b) K-means output (c) Optimized K-means with FA output (d) Final segmented image.

Table A.1: Evaluation of optimized K-means + FA on the ISIC 2016, 2017 and 2018 test sets.

| **Dataset** | **No. of Images** | **ACC** | **DICE** | **IoU** | **SEN** | **SPE** | **Time (sec)** |
|---|---|---|---|---|---|---|---|
| ISIC 2016 | 379 | 85.14 | 58.81 | 49.83 | 55.34 | 94.41 | 169.82 s |
| ISIC 2017 | 366 | 84.61 | 57.07 | 48.17 | 53.93 | 94.20 | 178.94 s |
| ISIC 2018 | 778 | 84.79 | 55.72 | 47.80 | 55.61 | 92.93 | 1593.7s |

## A.2 Other Deep Learning-based Segmentation Methods

In deep learning, two existing frameworks (FCNN) [92] and SegNet [94] were modified and implemented to extract lesion region. These two frameworks have shown remarkable performance on the semantic segmentation task in the literature studies [205]. In order to fine-tune these networks on the skin cancer datasets, a few modifications were done in their layers and parameters.

### A.2.1 Fully convolutional neural network (FCNN)

As shown in Figure A.2, a FCNN segmentation model was constructed using five blocks with multiple layers in each block. The network consists of 16 feature extraction or convolutional layers along normalization, pooling and activation function. It also have transposed layers, crop layers, and double convolution layers of size $1 \times 1$ at the end of the network. It is a multilayer convolutional network to analyze input data by overlapping many kernels on an input image. Several parameters, including ratio of kernels, their size, stride, and padding, are artificially set in the convolutional layer. The numbers of kernels were increasing at double rate in convolutional layer of each block with uniform size $3 \times 3$. The padding parameter specifies how the input image will be handled when it does not fit properly, and the stride parameter specifies how much the kernel will move over an image. To calculate the weights, a function proposed by He. et. al. [206] is used in the convolutional layer.

$$\sigma^2 = \frac{2}{(1+s^2)m} \quad where \quad m = kernelsize(i) * kernelsize(j) * numchannels \tag{A.3}$$

The average value of the normalized distribution was set to zero, the weights are

initialized, and the variance is calculated with the formula above. The leakyReLU layer was scaled by an amount $s$ set as '0.3' after fine-tuning, and $m$ was obtained by multiplying depth and width of a kernel. The leakyReLU activation function is then used to improve the network's performance without creating an overfitting problem or adding any additional computational cost. If '$x$' is positive, leakyReLU passes the sample to the next layer, while if it is negative, it multiplies the sample by any scalar value. Between convolution and leakyReLU, a layer of batch normalization is applied to normalize data based on the size of each mini-batch. A mini-batch size is a variant of gradient descent that divides the training dataset into small batches, which is used to calculate network error and update network coefficients. Batch size refers to how many samples are processed before a network is updated.

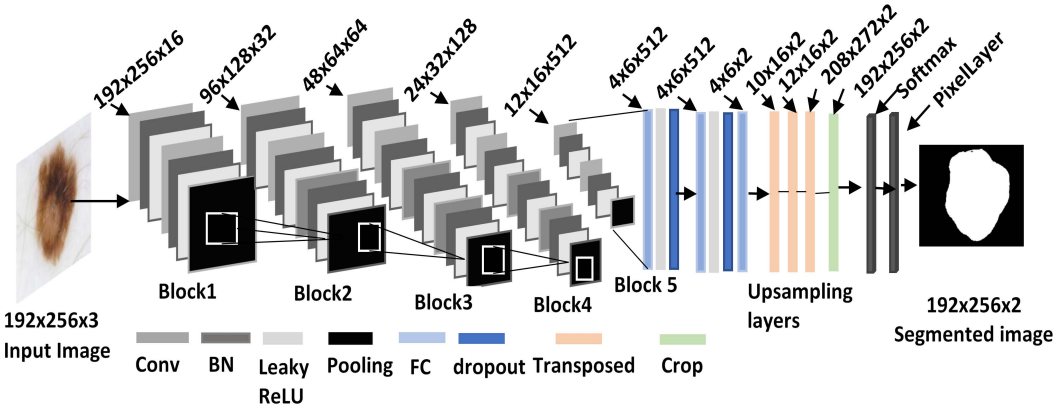



Figure A.2: The design of FCNN.

As the network is designed to obtain deep features, spatial resolution is lost due to successive pooling layers. These layers reduce the size of the feature map as the depth of the network increases. The feature map is transformed back to its original dimensions by inserting two up-sampling layers at the end. The concept of dense feature maps is achieved by using two convolutional layers of size $1 \times 1$ instead of a fully connected layer at the end of the network. To interpret pixels' probabilities, the Softmax function [160] is then used to convert feature map coefficients between 0 and

1. Furthermore, the cross-entropy loss function used to determine whether there was a difference between the predicted P samples and the target T samples.

### A.2.2 Results of FCNN

We compared the performance of the FCNN against the ISIC 2016, ISIC 2017, and ISIC 2018 test datasets. The network performance was evaluated objectively based on ACC, IoU, and DICE metrics listed in Table A.2. The segmentation results using FCNN model also presented in Figure A.3 displaying segmented lesion corresponding to original samples. It is observed that FCNN model was able to extract lesion areas accurately to some extent, however, the overall performance is lower than DilatedSkinNet.

Table A.2: Lesion segmentation performance of FCNN on different datasets.

| **Dataset** | **Number of images** | **ACC** | **IoU** | **DICE** |
|---|---|---|---|---|
| ISIC 2016 | 379 | 0.9188 | 0.8174 | 0.8512 |
| ISIC 2017 | 366 | 0.8105 | 0.6148 | 0.6934 |
| ISIC 2018 | 778 | 0.8700 | 0.6721 | 0.7790 |

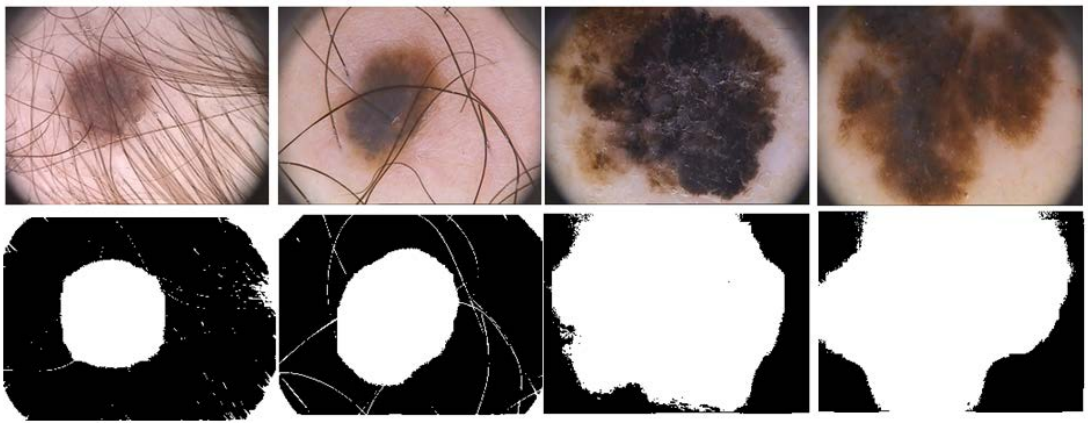


Figure A.3: Segmentation output of FCNN model (top row) original (bottom row) segmented lesions.

### A.2.3 Modified SegNet

Adapted from VGG19 [207] deep neural network, whose layers are being modified and optimized for designing deep encoder-decoder architecture for lesion segmentation. Figure A.4 illustrates the CNN architecture that was developed for skin cancer segmentation. Originally, encoder-decoder networks were used to segment road scenes using semantics. We designed the proposed network by changing the layer configuration, network depth, and hyperparameters. It is composed of two halves - the encoder and decoder sections. The encoder section has four layers that include multiple convolutional, batch normalization, leaky ReLU, and max-pooling layers. On the other hand, the lower half was composed of the decoder, which had deconvolution layers, batch normalization, leaky ReLU layers, and max-unpooling layers according to the encoder layers, followed by a cross-entropy loss function-based segmentation layer. This network performs upsampling in the decoder section using the pooling indices calculated by the encoder section's max-pooling layers. As a result of removing fully connected layers in previous networks, the encoder section was made smaller, reducing the number of parameters that could be learned.

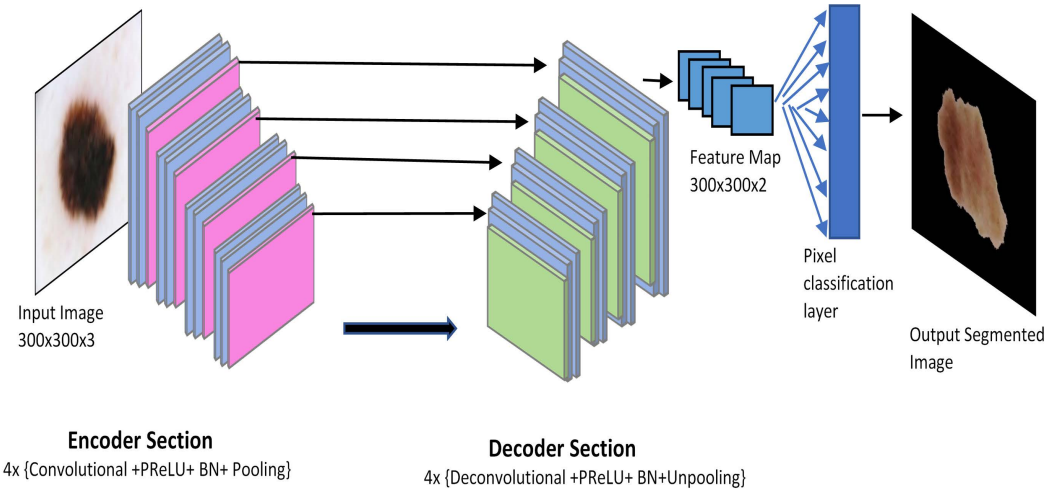



Figure A.4: The proposed architecture of the modified SegNet.

The encoder section made use of kernels to perform a convolutional operation to generate feature maps. By applying element-wise leaky rectified linear units (PReLU), the negative values were set to any fixed positive value and then normalized via batch normalization layer. Using a $2 \times 2$ window size, a stride of 2, and padding of 0, a max-pooling operation was implemented. In order to reduce the size of feature maps while preserving significant characteristics, max pooling was used. Pooling operation can be applied using the equation as:

$$FM[x, y] = f_p \times I_{x,y} \tag{A.4}$$

$FM[x, y]$ shows the output feature map, $f_p$ is the pooling function, $I_{(x,y)}$ is the input feature map from the previous layer. With maximum pooling, a maximum value is calculated from the portion overlapped by the kernel. The spatial resolution of the output feature map is reduced by the repeated use of max-pooling and striding at successive layers. Convolutional networks that use a 'transposed convolutional layer' consume more memory and time than conventional neural networks. For more robust classification, several convolutional layers were successively added to achieve high translation invariance in the original VGG19 network. Nonetheless, there was a loss in spatial resolution that rendered the segmentation unsuitable. Thus, boundary information must be preserved after subsampling layers. The spatial resolution was regained using upsampling that utilized maximum-pooling indices, i.e. the location of maximum features was stored and used in the decoder section. The probabilities corresponding to each pixel in the feature map were generated by a softmax and pixel classification layer.

### A.2.4 Results of Modified SegNet

The evaluation of modified SegNet model is performed on the same datasets ISIC 2016-2018 having similar distribution of samples in each test set. Presented in Table A.3 are the results produced by this model when applied to the ISIC 2016, ISIC 2017, and ISIC 2018 test datasets. This model achieved satisfactory values for ACC and DICE metrics, however, the IoU index is very low on all three datasets. The extracted lesion regions are shown in Figure A.5 corresponding to each original image. It can be seen that output lesion produced by SegNet are very rough, partially segmented, and have inaccurate borders as compared to ground truths.

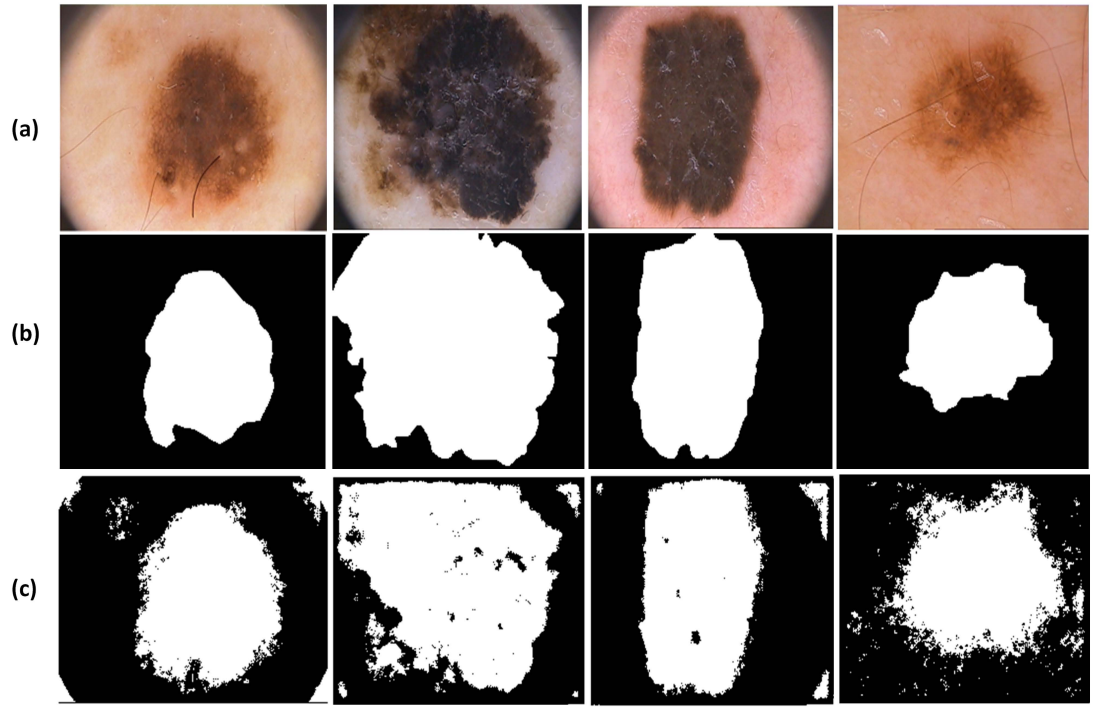



Figure A.5: Segmentation output of SegNet model (a) original (b) ground truth (c) segmented lesions.

Table A.3: Lesion segmentation performance of modified SegNet on different test sets.

| Dataset | Number of images | ACC | IoU | DICE |
|---|---|---|---|---|
| ISIC 2016 | 379 | 0.9082 | 0.8134 | 0.9071 |
| ISIC 2017 | 366 | 0.8219 | 0.6790 | 0.8180 |
| ISIC 2018 | 778 | 0.8802 | 0.7308 | 0.8794 |

## A.3 Comparison of Implemented Segmentation Models

Initially, machine learning-based optimized K-means using FA algorithm was implemented for segmentation tasks as discussed above. It is observed from the results that this method was not very efficient in producing higher performance in terms of ACC, IoU and DICE on the skin cancer datasets. Moreover, this method required highly cleaned images and hairlines to be completely occluded, whereas deep learning approaches can process raw images, and is capable of detecting lesion accurately and efficiently. Therefore, to explore the advantages of deep learning techniques for lesion segmentation, few existing segmentation frameworks such as FCNN [92] and SegNet [94] were implemented as initial experimental analysis. These methods were evaluated on the ISIC 2016, 2017 and 2018 datasets and their performance is compared with the proposed novel methods, EDNet and DilatedSkinNet (see Chapter 5).

Table A.4: Performance comparison between segmentation networks implemented on ISIC 2016-2018 datatsets.

| Methods | ISIC 2016 | | | ISIC 2017 | | | ISIC 2018 | | | Per image test time |
|---|---|---|---|---|---|---|---|---|---|---|
| | ACC | IoU | DICE | ACC | IoU | DICE | ACC | IoU | DICE | |
| K-means+FA | 0.851 | 0.498 | 0.588 | 0.846 | 0.481 | 0.570 | 84.79 | 47.80 | 55.72 | 3.3s |
| FCNN | 0.918 | 0.817 | 0.851 | 0.810 | 0.614 | 0.693 | 0.870 | 0.672 | 0.779 | 0.4 |
| SegNet | 0.908 | 0.813 | 0.907 | 0.821 | 0.679 | 0.818 | 0.880 | 0.730 | 0.879 | 0.6s |
| EDNet | 0.941 | 0.887 | 0.940 | 0.874 | 0.770 | 0.869 | 0.923 | 0.855 | 0.922 | 0.3s |
| DilatedskinNet | **0.950** | **0.904** | **0.949** | **0.888** | **0.793** | **0.884** | **0.942** | **0.891** | **0.942** | **0.2s** |

Note: Test time is on single image. The best values are marked in bold.

Table A.4 shows that DilatedSkinNet outperformed other semantic segmentation models implemented in this study. It achieved an average pixel accuracy of 95.0%,

88.8%, and 94.2% on the ISIC 2016-2018 datasets respectively. The effectiveness of DilatedSkinNet model in segmenting accurate and smooth lesion is illustrated in Figure A.6. It shows that DilatedSkinNet produced the most accurate and smooth lesion among the models, while optimized-K-means produced a partially segmented lesion, FCNN resulted in a rough border, SegNet produced poor segmentation, and EDNet produced inaccurate and rough borders.

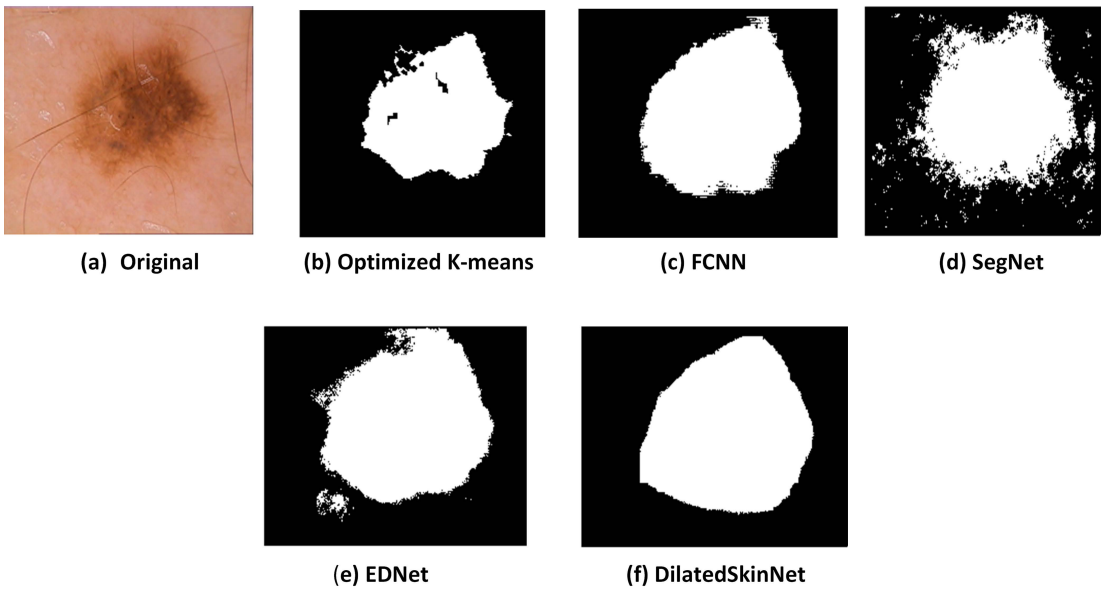



Figure A.6: Comparison of segmentation techniques on Lesion sample.

# Appendix B

## B.1 Other Classification Networks

In our initial attempt to develop a classification network for melanoma detection, we developed a small multi-layer CNN model. The work based on this network is published in the following article:

- R. Kaur, H. GholamHosseini & R. Sinha, "Deep Convolutional Neural Network for Melanoma Detection using Dermoscopy Images", *In* $42^{nd}$ *Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, Canada*, IEEE, pp.1524-1527, 20-24 July, 2020. [Online]. Available: 10.1109/EMBC44109.2020.9175391.

Figure B.1 shows an architecture of CNN designed from scratch having six blocks composed of multiple layers. The network performs melanoma classification accepting input image vector and producing probability score for each sample. Each block contained five main layers and performed different operations, such as convolution, batch normalization, ReLU, cross normalization and max pooling.

It is a multi-layer CNN that process image's information by using a set of convolutional kernels or filters. In each convolutional layer, multiple filters of fixed size $3 \times 3$ were employed that slides over an image to compute different set of features.

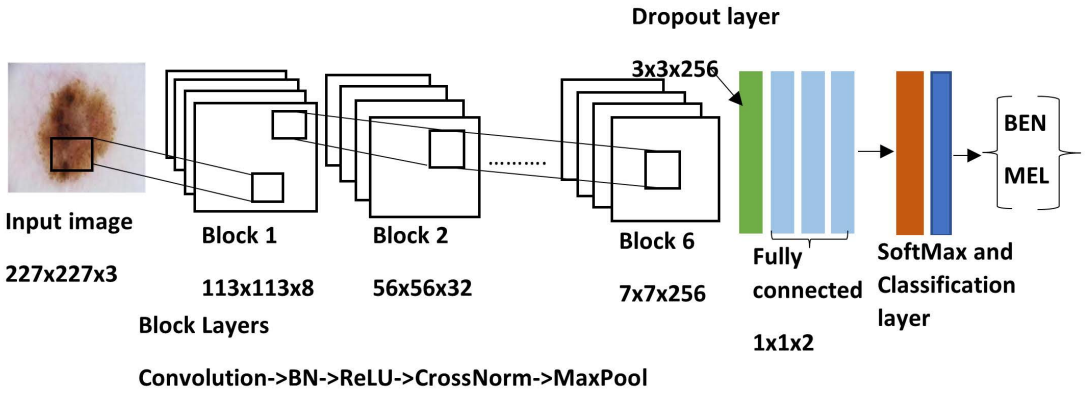



Figure B.1: Classification CNN network architecture.

Various filter sizes such as $11 \times 11$, $5 \times 5$, and $3 \times 3$ were tested during the design of the network, however, filter size $3 \times 3$ was found to be more effective. Each layer inputs the subsequent layer in the network to allow the flow of information.

The specification of the layers used in each block of the the network as given in Table B.1 are described as follows:

- **Input layer (input):** The input images of size $227 \times 227 \times 3$ given to the network through input layer in which the first two parameters denote the height and width of an image and third value stands for number channels in an image. In this case, it is '3' representing RGB image with Red, Green and Blue channels.

- **Convolutional layers ($Conv_1 - Conv_6$):** In the network, there are total of six convolutional layers that have different number of kernels as 8, 16, 32, 64, 128, and 256 increasing at double rate in each block. The size of each kernel used is $3 \times 3$, padding '0' and stride of '1'. This layer is responsible for extracting features from the input image $I_{x,y}$ by sliding multiple filters $K_{x,y}$. It computes feature maps $FM$ using the following formula and pass those maps to the next layers in the blocks.

$$FM[x, y] = I_{x,y} \times K_{x,y} \tag{B.1}$$

Table B.1: Detail of layers of melanoma classification CNN.

| **Block** | **Layer** | **Kernel size, maps** | **Output size** |
|---|---|---|---|
| INPUT LAYER | Input | $227 \times 227 \times 3$ | $227 \times 227 \times 3$ |
| BLOCK 1 | Conv_1 | $3 \times 3$, 8 | $227 \times 227 \times 8$ |
| | BN_1 | —– | $227 \times 227 \times 8$ |
| | Crossnorm_1 | —– | $227 \times 227 \times 8$ |
| | MaxPool_1 | $2 \times 2$,8 | $113 \times 113 \times 8$ |
| BLOCK 2 | Conv_2 | $3 \times 3$, 16 | $113 \times 113 \times 16$ |
| | BN_2 | —– | $113 \times 113 \times 16$ |
| | Crossnorm_2 | —– | $113 \times 113 \times 16$ |
| | MaxPool_2 | $2 \times 2$,8 | $56 \times 56 \times 16$ |
| BLOCK 3 | Conv_3 | $3 \times 3$, 32 | $56 \times 56 \times 32$ |
| | BN_3 | —– | $56 \times 56 \times 32$ |
| | Crossnorm_3 | —– | $56 \times 56 \times 32$ |
| | MaxPool_3 | $2 \times 2$,8 | $28 \times 28 \times 32$ |
| BLOCK4 | Conv_4 | $3 \times 3$, 64 | $28 \times 28 \times 64$ |
| | BN_4 | —– | $28 \times 28 \times 64$ |
| | Crossnorm_4 | —– | $28 \times 28 \times 64$ |
| | MaxPool_4 | $2 \times 2$ | $14 \times 14 \times 64$ |
| BLOCK5 | Conv_5 | $3 \times 3$, 128 | $14 \times 14 \times 128$ |
| | BN_5 | —– | $14 \times 14 \times 128$ |
| | Crossnorm_5 | —– | $14 \times 14 \times 128$ |
| | MaxPool_5 | $2 \times 2$ | $7 \times 7 \times 128$ |
| BLOCK6 | Conv_6 | $3 \times 3$, 128 | $7 \times 7 \times 256$ |
| | BN_6 | —– | $7 \times 7 \times 256$ |
| | Crossnorm_6 | —– | $7 \times 7 \times 256$ |
| | MaxPool_6 | $2 \times 2$ | $3 \times 3 \times 256$ |
| Final layers | Dropout layer | —— | $3 \times 3 \times 256$ |
| | $3 \times FC$ | —— | $1 \times 1 \times 2$ |
| | Softmax | —— | $1 \times 1 \times 2$ |
| | Classification layer | —— | —— |

- **Batch normalization** ($BN\_1 - BN\_6$)**:** This layer in the network normalizes each input based on the mini-batch size. It is positioned between the convolutional and ReLU layers to reduce the number of epochs, and stabilize the learning process.

- **ReLU activation**($ReLU\_1 – ReLU\_6$)**:** By using threshold function, $ReLU(x)$ = $max(0, x)$, it sets all negative values to '0' for each element in the input.

- **Cross-normalization**($Crossnorm\_1 - Crossnorm\_6$)**:** The layer normalizes local responses channel-by-channel. In this layer, each element is replaced with

a normalized value obtained from neighboring channels. Using the following formula, each element of the input vector is normalized:

$$X' = \frac{X}{\left(C + \frac{\alpha * SS}{windowsize}\right)^{\beta'}} \tag{B.2}$$

$C$, $\alpha$ and $\beta$ are the hyperparameters in the normalization process, and $SS$ is the sum of the square of the elements in the normalization window. In our network, the values of the parameters chosen as $windowsize$ = 5, $\alpha$=0.001, $\beta$=0.75, and $C$=2.

- **Max-pooling ($maxpool\_1 - maxpool\_5$):** Downsampling is performed using this layer by dividing the input feature map into several regions and then computing the maximum of each region to form a new feature map. Pool size in max-pooling layers is $2 \times 2$, stride is '2', and padding is '0'.

$$FM[x, y] = f_p \times I_{x,y} \tag{B.3}$$

- **Dropout:** A 50% dropout rate is used in this layer to prevent a network from overfitting by setting the neurons within hidden layers to '0' after each training phase.

- **Fully connected ($FC1 - FC3$):** By combining the weights and bias values, the input is multiplied by the fully connected layer. Based on the backpropagation network, the weights are used to predict final probabilities for each label, such as 'malignant' or 'benign'. The network consists of three fully interconnected layers.

- **Softmax and loss function):** It is the output activation unit that is used after fully connected layers. To calculate the accuracy of the classification network, a

loss function is applied by subtracting the predicted and actual outputs. As a loss function, the sum of squares function was used in place of the commonly used cross-entropy loss function.

The ISIC skin cancer dataset with a total of 12,882 images as given in Table B.2 consisting of benign and malignant melanoma samples was used for experimental study. The distribution of samples into training, validation and test set was done randomly in the ratio of 70%, 10%, and 20% respectively. A lot of time and patience was spent to tune the hyperparameters of CNN for the given dataset. It was quite challenging to figure out which parameters worked best together. Different kernel sizes, filter values, number of layers, topologies, and training options were used to design the network for gaining the best performance.

Table B.2: Distribution of benign and melanoma samples in the dataset.

| **Type of samples** | **Training 70 %** | **Validation 10 %** | **Test 20 %** |
|---|---|---|---|
| Melanoma | 4052 | 579 | 1158 |
| Benign | 4965 | 709 | 1419 |
| Total | 9017 | 1288 | 2577 |

Table B.3: Performance of Classification network.

| **Method** | **ACC %** | **PRE %** | **REC %** | **SPE %** | **F-score %** |
|---|---|---|---|---|---|
| CNN | 79.69 | 83.12 | 79.48 | 79.48 | 81.26 |

Table B.4: Performance of Classification network.

| **Method** | **ACC %** | **PRE %** | **REC %** | **SPE %** | **F-score %** |
|---|---|---|---|---|---|
| CNN | 79.69 | 83.12 | 79.48 | 79.48 | 81.26 |
| LCNet | **90.92** | **91.12** | **92.29** | **89.29** | **91.70** |

The proposed classification network achieved an average ACC of 79.69%, SEN of 79.48%, and PRE of 83.12%, as shown in Table B.3. These results have demonstrated that a small CNN is efficient in performing binary classification and identifying

melanoma vs benign; however, its accuracy is still under the expected target i.e., above 90%. Therefore, LCNet (refer to Chapter 6) was designed to achieve high performance. Table B.4 compares performance of this small CNN with our final classification network, LCNet. We observed that LCNet with a deeper layered design is more efficient in achieving higher accuracy on the same dataset.