Revolutionizing Skin Lesion Classification using Novel Deep Learning Models

Thesis submitted in partial fulfilment of the requirements for the degree of

Master of Technology
in
Computer Technology

Submitted by

Amartya Ray
Regn. No. - 160077 of 2021 - 2022
Exam Roll No. - M6TCT24005
Class Roll No. - 002110504006
Session - 2021 - 2024

Under the supervision of

Prof. Ram Sarkar

Department of Computer Science and Engineering
Jadavpur University
188, Raja S.C. Mallick Rd,
Kolkata - 700032, West Bengal, India

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF RECOMMENDATION

This is to certify that the work embodied in this thesis entitled **"Revolutionizing Skin Lesion Classification using Novel Deep Learning Models"** has been satisfactorily completed by **Amartya Ray** (Registration Number 160077 of 2021 - 2022; Class Roll No. 002110504006; Examination Roll No. M6TCT24005). It is a bonafide piece of work carried out under my supervision and guidance at Jadavpur University, Kolkata, for partial fulfilment of the requirements for the awarding of the **Master of Technology** in **Computer Technology** degree of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, during the academic year 2021 - 2024.

Prof. Ram Sarkar,
Department of Computer Science and Engineering,
Jadavpur University.
(Supervisor)

Forwarded By:

Prof. Nandini Mukhopadhyay,
Head, Department of Computer Science and Engineering,
Jadavpur University.

Prof. Dipak Laha,
Dean, Faculty of Engineering and Technology,
Jadavpur University.

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled **"Revolutionizing Skin Lesion Classification using Novel Deep Learning Models"** is a bonafide record of work carried out by **Amartya Ray** (Registration Number 160077 of 2021 - 2022; Class Roll No. 002110504006; Examination Roll No. M6TCT24005) in partial fulfilment of the requirements for the award of the degree of **Master of Technology** in **Computer Technology** in the Department of Computer Science and Engineering, during the period of September 2021 to May 2024. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose of which it has been submitted.

_____

Signature of Examiner 1
Date:

_____

Signature of Examiner 2
Date:

Department of Computer Science and Engineering
Faculty of Engineering and Technology
Jadavpur University, Kolkata - 700032

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that the thesis entitled **"Revolutionizing Skin Lesion Classification using Novel Deep Learning Models"** contains literature survey and original research work by the undersigned candidate, as a part of his degree of **Master of Technology** in **Computer Technology, Jadavpur University**. All the information have been obtained and presented in accordance with academic rules and ethical conduct.
I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Amartya Ray

Examination Roll No.: M6TCT24005

Registration No.: 160077 of 2021 - 2022

Thesis Title: Revolutionizing Skin Lesion Classification using Novel Deep Learning Models

Signature of the Candidate:

. . . . . .

Dedicated to my unwavering pillars of support,
my family and friends,
whose boundless love and selflessness have propelled every phase of my journey.

. . . . . .

Acknowledgements

I am profoundly grateful and indebted to my supervisor, **Dr. Ram Sarkar**, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India. His constant support, supervision, and expertise were instrumental in shaping the direction and quality of this thesis work. I am sincerely thankful for his steadfast dedication throughout this endeavor.

I express my sincere gratitude to the Head of the Department, **Dr. Nandini Mukhopadhyay**, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for granting me permission to conduct my research at this esteemed institution.

I must convey my sincere appreciation to **Prof. Mahantapas Kundu**, Coordinator of the **Centre for Microprocessor Applications for Training, Education and Research (CMATER) Lab**, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for allowing me to carry out this study in the renowned CMATER lab and providing valuable suggestions.

I extend my special thanks to my teachers, **Prof. Subhadip Basu**, **Prof. Debotosh Bhattacharjee**, **Prof. Nibaran Das**, and **Dr. Mita Nasipuri (former Professor)**, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, for imparting their invaluable knowledge to me.

I am grateful to my colleagues and labmates, Mr. Sujan Sarkar, Mr. Soumyajit Gayen, Mr. Akash Halder, Mr. Diptarka Mandal, Mr. Debasmit Roy, Mr. Anurup Naskar, Mr. Ankit Das, and friends, Mr. Arunava Roy, Mr. Arunava Banerjee, and Mr. Anik Biswas for their unwavering support, validation, and understanding throughout this journey. Their companionship and camaraderie have been invaluable.

I must also extend heartfelt thanks to my family members for their encouragement and support in preparing this thesis.

Amartya Ray

Examination Roll No.: M6TCT24005

Registration No.: 160077 of 2021 - 2022

Department of Computer Science and Engineering,

Jadavpur University.

Abstract

Skin cancer is a significant global health concern, and timely, accurate diagnosis is crucial for improving patient outcomes. Recent advancements in deep learning have revolutionized automated skin cancer classification, offering powerful tools for dermatology. This thesis presents three novel approaches to enhance skin lesion classification accuracy using deep learning techniques.

The first method utilizes a deep feature fusion network (DFF-Net), enhanced by squeeze-and-excitation attention, to tackle the classification task. Images are initially cropped using a contour-based approach, and features are then extracted from these cropped images using Xception and BiT-M models. These features are fused and processed through a squeeze-and-excitation attention module, followed by pooling and fully connected layers for final classification. A combined loss function is introduced to further boost performance.

The second method introduces a novel edge-guided dual-branched network (EDB-Net) for classification. One branch processes edge maps of dermoscopic images, while the other focuses on the original images. Both branches use a DenseNet-based backbone with dense and transition blocks to extract features. The second branch incorporates an additional attention module to highlight salient regions in the images, capturing more relevant features. A novel strategy modulates attention maps using previously extracted edge features, creating edge-guided features that refine the representations. The final stage aggregates both attended and edge-guided features, which are then flattened and fed into dense layers for classification.

The third method employs a Dempster-Shafer theory based classifier combination network (DSCC-Net) that combines the outputs of three attention-based convolutional neural networks. The base learners — ResNet50, InceptionResNetV2, and DenseNet201 — are each enhanced with soft attention modules to improve focus on relevant features. These base learners generate confidence scores which are then combined using the popular Dempster-Shafer theory.

The three proposed models are evaluated on three challenging skin cancer datasets: HAM10000, ISIC 2016, and ISIC 2017. These approaches exhibit superior performance compared to many recent methods in the literature, achieving state-of-the-art accuracy in skin lesion classification and demonstrating significant advancements in this domain.

Keywords: Skin Cancer, Deep Learning, Medical Image, Feature Fusion, Attention Module, Edge Maps, Classifier Combination, Dempster-Shafer Theory

Contents

Abstract i

Table of Contents ii

List of Figures iv

List of Tables vii

List of Abbreviations ix

1 Introduction 1
- 1.1 Objectives . . . 2
- 1.2 Scope of the Thesis . . . 3
- 1.3 Organization of the Thesis . . . 4

2 Literature Survey 5

3 Methods and Materials 13
- 3.1 DFF-Net . . . 13
  - 3.1.1 Image Preprocessing and Augmentation . . . 14
  - 3.1.2 Xception . . . 15
  - 3.1.3 BiT-M . . . 15
  - 3.1.4 Squeeze-and-Excitation Attention Module: . . . 16
  - 3.1.5 Combined Loss Function . . . 16
- 3.2 EDB-Net . . . 17
  - 3.2.1 Novel Method for Edge Map Generation . . . 18
    - 3.2.1.1 Preprocessing . . . 18
    - 3.2.1.2 K-means Clustering . . . 18
    - 3.2.1.3 Adaptive Histogram Equalization . . . 18
    - 3.2.1.4 Edge Detection . . . 19
  - 3.2.2 Proposed Dual-Branched CNN . . . 20
    - 3.2.2.1 Branch 1 - Edge Feature Extraction . . . 21
    - 3.2.2.2 Branch 2 - Dermoscopic Feature Extraction . . . 22
    - 3.2.2.3 Modulation of Attention Maps . . . 22
    - 3.2.2.4 Integration of Edge-Enhanced Attention . . . 23
  - 3.2.3 Loss Function . . . 24
- 3.3 DSCC-Net . . . 24
  - 3.3.1 Data Augmentation . . . 24

3.3.2 Soft Attention Module . . . 25
3.3.3 ResNet50 . . . 26
3.3.4 InceptionResNetV2 . . . 26
3.3.5 DenseNet201 . . . 27
3.3.6 Dempster-Shafer Theory based Classifier Combination . . . 27
3.3.7 Loss Function . . . 27

4 Results and Discussion **29**
4.1 Datasets . . . 29
4.1.1 HAM10000 . . . 29
4.1.2 ISIC Archive . . . 31
4.2 System Configuration . . . 31
4.3 Evaluation Metrics . . . 32
4.3.1 Accuracy . . . 32
4.3.2 Precision . . . 32
4.3.3 Sensitivity . . . 33
4.3.4 Specificity . . . 33
4.3.5 F1 Score . . . 33
4.3.6 Receiver Operating Characteristic Curve . . . 33
4.3.7 Area Under the Curve . . . 34
4.4 DFF-Net . . . 34
4.4.1 Training Setup . . . 34
4.4.2 Experimental Results . . . 34
4.5 EDB-Net . . . 36
4.5.1 Training Setup . . . 36
4.5.2 Experimental Results . . . 36
4.6 DSCC-Net . . . 37
4.6.1 Training Setup . . . 38
4.6.2 Experimental Results . . . 38
4.7 Model Interpretability using Grad-CAM . . . 40
4.8 Comparison with Recent Past Methods . . . 41

5 Conclusions **47**
5.1 Limitations . . . 48
5.2 Scope for Future Work . . . 49

List of Publications **51**

Bibliography **53**

List of Figures

1.1 Types of skin cancer 2

3.1 Architecture of the proposed DFF-Net model 14
3.2 Architecture of the proposed EDB-Net model 21
3.3 Architecture of the CBAM block 22
3.4 Architecture of the proposed DSCC-Net model 25
3.5 Overview of the SA unit 26

4.1 Sample images of skin lesions in HAM10000 and ISIC 2017 datasets 30
4.2 Confusion matrix of DFF-Net on (a) HAM1000 dataset; (b) ISIC 2017 dataset 35
4.3 Confusion matrix of EDB-Net on HAM10000 dataset 38
4.4 Confusion matrix of (a) ResNet50; (b) InceptionResNetV2; (c) DenseNet201; (d) DSCC-Net on HAM10000 dataset 39
4.5 Grad-CAM results on (a) HAM10000 dataset using DSCC-Net; (b) HAM10000 dataset using EDB-Net; (c) ISIC 2017 dataset using DFF-Net 41
4.6 (a) Training curve; (b) Loss curve for HAM10000 dataset 44
4.7 (a) Training curve; (b) Loss curve for ISIC 2016 dataset 44
4.8 (a) Training curve; (b) Loss curve for ISIC 2017 dataset 44
4.9 ROC curves for (a) HAM1000 dataset; (b) ISIC 2016 dataset; (c) ISIC 2017 dataset; (d) Accuracy comparison across proposed models and datasets 45

List of Tables

4.1 Datasets for skin cancer classification 30
4.2 Ablation results for selecting the best model on ISIC 2017 dataset . 35
4.3 Classification results achieved by DFF-Net across three datasets . . 35
4.4 Ablation study on HAM10000 dataset for selecting the best backbone network 36
4.5 Ablation results for selecting the best model on HAM1000 dataset . 37
4.6 Classification results achieved by EDB-Net across three datasets . . 37
4.7 Hyper-parameters for base learners utilized by DSCC-Net 38
4.8 Performance comparison among base learners and DSCC-Net on HAM10000 dataset 40
4.9 Classification results achieved by DSCC-Net across three datasets . 40
4.10 Performance comparison of the proposed models with some recent methods on HAM10000 dataset 42
4.11 Performance comparison of the proposed models with some recent methods on ISIC 2016 dataset 43
4.12 Performance comparison of the proposed models with some recent methods on ISIC 2017 dataset 43

List of Abbreviations

A	
AI	**A**rtificial **I**ntelligence
AUC	**A**rea **U**nder the **C**urve
B	
BGR	**B**lue-**G**reen-**R**ed
BiT-M	**Bi**g **T**ransfer-**M**edium
BPA	**B**asic **P**robability **A**ssignment
C	
CBAM	**C**onvolutional **B**lock **A**ttention **M**odule
CLAHE	**C**ontrast **L**imited **A**daptive **H**istogram **E**qualization
CNN	**C**onvolutional **N**eural **N**etwork
D	
DABN	**D**eep **A**ttention **B**ranch **N**etwork
DFF-Net	**D**eep **F**eature **F**usion **Net**work
DL	**D**eep **L**earning
DS	**D**empster **S**hafer
DSCC-Net	**D**empster-**S**hafer theory-based **C**lassifier **C**ombination **Net**work
E	
EDB-Net	**E**dge-guided **D**ual-**B**ranched **Net**work
F	
FCRN	**F**ully **C**onvolutional **R**esidual **N**etwork
FN	**F**alse **N**egative

FP **F**alse **P**ositive

G

GAN **G**enerative **A**dversarial **N**etwork

H

HAM **H**uman **A**gainst **M**achine

HED **H**olistically nested **E**dge **D**etector

HSV **H**ue-**S**aturation-**V**alue

I

ISBI **I**nternational **S**ymposium on **B**iomedical **I**maging

ISIC **I**nternational **S**kin **I**maging **C**ollaboration

L

LAB **L**ightness-**A**-**B**

LBPH **L**ocal **B**inary **P**attern **H**istogram

LGBM **L**ight **G**radient **B**oosting **M**achine

M

ML **M**achine **L**earning

P

PCA **P**rincipal **C**omponent **A**nalysis

R

ReLU **Re**ctified **L**inear **U**nit

RF **R**andom **F**orest

ROC **R**eceiver **O**perating **C**haracteristic

ROI **R**egion **O**f **I**nterest

S

SA **S**oft **A**ttention

SAM **S**oft **A**ttention **M**odule

SEA **S**queeze-and-**E**xcitation **A**ttention

SELU	**S**caled **E**xponential **L**inear **U**nit
SGD	**S**tochastic **G**radient **D**escent
SoTA	**S**tate-**o**f-**T**he-**A**rt
SVM	**S**upport **V**ector **M**achine
T	
TL	**T**ransfer **L**earning
TN	**T**rue **N**egative
TP	**T**rue **P**ositive
U	
UV	**U**ltra **V**iolet
V	
ViT	**Vi**sion **T**ransformer
W	
WCSS	**W**ithin-**C**luster **S**um of **S**quares

CHAPTER 1

Introduction

Over the past few years, DL techniques have significantly transformed medical image analysis, especially in skin cancer classification. The prevalence of skin cancer has witnessed a steady increase over time, largely attributed to increased exposure to detrimental UV radiation from the sun [1]. Frequent and intense sun exposure can lead to sunburns, elevating the likelihood of developing skin cancer. Additional significant causes include the presence of abnormal moles in the body and a family history of skin cancer [2]. Skin cancer is broadly categorized into two main types: melanoma and non-melanoma [3]. Figure 1.1 illustrates the categorization of skin cancer into different types. While basal cell carcinoma and squamous cell carcinoma constitute the most prevalent malignant forms, they are not as harmful as melanoma [4]. Melanoma poses a higher risk of metastasis and contributes significantly to mortality related to skin cancer [5]. Early diagnosis is crucial for the effective cure of melanoma, as it spreads rapidly and can lead to a distressing and fatal outcome for the affected individual [6]. Improving the accuracy and robustness of skin cancer classification holds the potential to elevate precise diagnosis, clinical decision-making, treatment planning, and ultimately, patient outcomes in the field of dermatology. DL, with its exceptional ability to automatically learn hierarchical features from raw data, has emerged as a powerful tool to tackle the challenges in skin cancer classification. This thesis presents a deep dive into DL methodologies for classifying dermoscopic images of skin lesions, introducing three innovative approaches to enhance classification accuracy and robustness. By building on previous research insights [7] and leveraging advancements in DL architectures and methodologies, this study aims to advance the current capabilities in skin cancer classification.

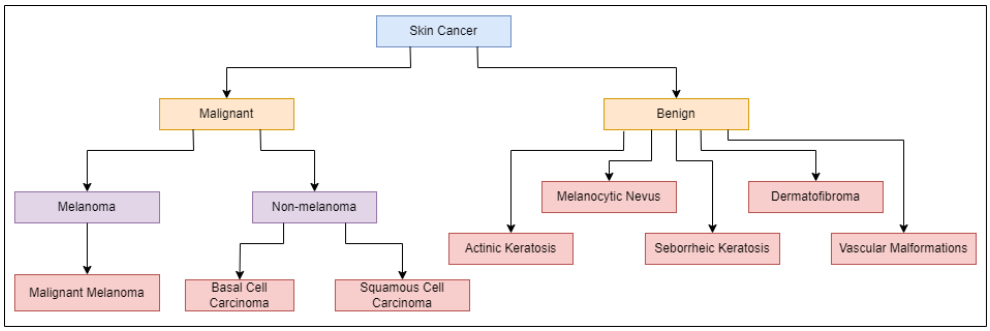


Figure 1.1: Types of skin cancer

1.1 Objectives

The economic impact of skin cancer on healthcare systems is significant, given the expenses related to diagnosis, treatment, long-term care, and extensive medical training. Skin cancer diagnosis primarily involves three methods: self-examination, dermoscopy, and skin biopsy. Accurate self-examination of skin cancer is highly challenging due to the striking resemblance between malignant and benign skin conditions. Dermoscopy, a widely accepted imaging technique among dermatologists, involves magnifying the surface of a skin lesion for enhanced visibility during examination. However, its practice is limited to trained physicians, relying heavily on their visual skills and experience [8]. Skin biopsy, another method used by doctors, entails the removal of a sample from a suspected skin lesion for examination to determine its cancerous nature [9]. Despite being a reliable diagnostic tool, this procedure is characterized by pain and time consumption. These challenges highlight the necessity for novel and innovative AI approaches in visualizing and diagnosing skin cancer. Skin cancer detection has traditionally relied on extracting handcrafted features followed by classification, which often fails to capture the subtle characteristics of skin lesions and tends to yield undesirable results, particularly in the presence of artifacts and noise in medical images. In contrast, DL-based approaches offer a promising alternative by leveraging large-scale annotated datasets and advanced neural network architectures to automatically learn discriminative features directly from the images. The motivations behind this thesis work involve developing novel DL-based approaches to combat challenges like inter-class similarity and intra-class variability of lesions for the accurate multi-class classification of skin cancer.

1.2 Scope of the Thesis

In this study, three novel DL approaches specifically designed for skin cancer classification have been proposed. These approaches feature innovative methodologies and architectures that address key challenges in differentiating various types of skin lesions. The methods focus on effectively managing the diverse appearances of lesions, reducing noise and artifacts from dermoscopic images, and improving both classification accuracy and generalization capabilities. A thorough evaluation of the three methods has been conducted across three challenging and commonly used real-world datasets for assessing proposed techniques in skin cancer classification, namely HAM10000, ISIC 2016, and ISIC 2017.

- In the first approach, a deep feature fusion network (DFF-Net) is proposed. Initially, an efficient preprocessing technique based on contour detection is leveraged to crop the images, focusing on the lesion areas. Features are then extracted from the cropped images using Xception [10] and Google's BiT-M model [11], and fused using an element-wise multiplication technique along the channel dimension. Subsequently, a customized squeeze and excitation [12] based attention module is employed to emphasize important features after feature fusion. Additionally, the network is optimized using a combination of categorical cross entropy loss [13] and triplet loss [14] to reduce intra-class variability and increase inter-class dissimilarity.

- In the second approach, a novel edge-guided dual-branched network (EDB-Net) is introduced. One branch is dedicated to processing the edge maps generated from the dermoscopic images, while the other branch focuses solely on the original dermoscopic images. These edge maps aid in detecting irregularities in shape and symmetry, which are key indicators of potential malignancy. Additionally, these edge maps offer an enhanced focus on the ROIs. Both branches employ a DenseNet-based backbone [15] to extract features from their respective inputs. In the first branch, a combination of dense and transition blocks effectively captures edge features from the input edge maps. In the second branch, DenseNet201 has been utilized as the backbone, supplemented with an attention module. This module enhances the network's capability to focus on salient regions within the dermoscopic images, thereby capturing relevant features vital for classification. Furthermore, a novel strategy has been elucidated to modulate the attention maps using the previously extracted edge features. By using enhanced attention maps, edge-guided features are generated, leading to refined feature representations that harness complementary information from both sources. In

the final stage of the network architecture, both the attended features and the edge-guided features are aggregated to preserve vital edge information along with high-level semantic information.

- In the third approach, a novel network (DSCC-Net) is proposed that combines the outcomes of three attention-based CNNs using DS theory [16]. This classifier combination approach utilizes insights from multiple models which are discriminative in nature, thereby reducing overfitting and greatly improving the model's ability to classify unseen data. DS theory is particularly beneficial in handling multiple classifiers that may produce conflicting predictions. Initially, three base learners — ResNet50 [17], InceptionResNetV2 [18], and DenseNet201 — are trained, each aided with an attention module. These base learners generate confidence scores which are then combined using DS theory. Finally, the class with the highest belief value is selected to make the final decision.

1.3 Organization of the Thesis

The sections of this thesis are organized as outlined below:

- Chapter 1 presents a brief introduction to skin cancer and outlines the objectives and scope of this thesis.
- In chapter 2, a comprehensive literature survey is presented, examining existing research, methodologies, challenges, and breakthroughs in the field of skin cancer classification.
- Subsequently, chapter 3 details the innovative methodologies proposed for accurate skin cancer classification.
- In chapter 4, the various datasets used to assess the proposed methods are explored and the exhaustive results from experiments conducted on these challenging datasets are presented. This chapter also includes quantitative evaluations and critically discusses the implications of these findings.
- Finally, chapter 5 summarizes the key conclusions drawn from the study and offers insights into future research directions. This concluding chapter synthesizes the main findings, highlights overarching themes, and emphasizes the significance of the study's contributions to the field of skin cancer classification.

CHAPTER 2

Literature Survey

The emergence of DL has yielded rapid growth in the fields of pattern learning, image classification and recognition. DL models are trained on input data and not programmed explicitly. Following the training phase, these models serve as domain experts in their respective fields. Deep neural networks play a pivotal role in the classification of skin cancer. In recent years, researchers have explored novel approaches using CNNs, GANs, ViTs and segmentation-guided classification techniques to delineate ROIs from the dermoscopic images and extract meaningful features. Additionally, the introduction of hybrid and multimodal techniques has led to improved performance across skin cancer datasets. In spite of challenges related to robust model generalization, interpretability, and explainability, recent advancements highlight the potential of DL-based methods to reform skin cancer diagnosis, treatment planning, and prognosis assessment. Several recent publications employing various DL approaches for skin cancer classification are discussed in this section.

CNNs excel at analyzing dermoscopic images, identifying irregularities, borders, textures, and other visual cues crucial for distinguishing between different types of skin lesions. Mridha et al. [19] introduced a customized CNN architecture comprising two blocks for the feature extraction phase. In block 1, there were 2 convolutional layers with a kernel size of 3, accompanied by a pooling layer with a stride of 1, and a dropout layer. Block 2 included two convolutional layers with a kernel size of 3, a pooling layer with a stride of 2, and a dropout layer. The output from block 2 was flattened and subsequently passed through a final dense layer. Rastegar et al. [20] proposed a deep CNN with 69 layers, aiming to extract detailed and discriminative features from skin lesion images. The network consisted of multiple convolutional layers with different filter sizes, 3×3, 5×5, 7×7, and depths. The network also contained residual layers, Inception modules

and pooling layers accompanied by a final fully connected classification layer.

TL plays a pivotal role in skin cancer classification by leveraging knowledge learned from pre-trained models on large, diverse datasets to improve the performance of models trained on smaller skin cancer datasets. This approach is particularly beneficial due to the limited availability of skin cancer-related data. A pre-trained deep CNN architecture, VGG16 [21] with the last 3 fine-tuned layers and 5 convolutional blocks was proposed by Kalouche et al. [22]. For fine-tuning, they employed an SGD optimizer [23] with a low learning rate. Emara et al. [24] employed the InceptionV4 backbone [18] and introduced modifications by incorporating feature reuse through a residual connection. This connection played a crucial role in merging features extracted from earlier layers with those from high-level layers, contributing to an enhancement in the classification performance.

Le et al. [25] utilized the ResNet50 backbone with additional modifications for the classification of 7 types of skin cancer. Their modifications included the use of global average pooling instead of simple average pooling and the introduction of a dropout layer between the last 2 fully connected layers. Furthermore, they used a combination of weighted loss and focal loss to optimize their model. Sagar et al. [26] employed several pre-trained models for the binary classification of melanoma. They performed experiments using InceptionV3 [27], Inception-ResNetV2, ResNet50, MobileNet [28], and DenseNet169, out of which ResNet50 emerged with superior performance. Shen et al. [29] leveraged a low cost and high performance data augmentation strategy along with TL for automatic skin cancer screening in rural communities. Their network was built on EfficientNetB7 architecture [30].

Incorporating attention mechanisms within CNNs for skin cancer classification enhances models' ability to concentrate on crucial features within the skin lesions by assigning weights to the feature maps according to their relevance to the lesions. This integration also helps suppress image artifacts, like portions of uninfected skin, hair, and veins, contributing to more accurate and precise diagnostic outcomes. Ding et al. [31] proposed the DABN model, incorporating attention branches to enhance traditional deep CNNs. In the training stage, the attention branch was crafted to acquire the class activation maps, subsequently serving as attention maps directing the network's focus to discriminative regions in skin lesions. The DABN model incorporated 2 attention branches into the baseline architecture, which consisted of 4 dense blocks, 3 transition layers, and a classification layer. Datta et al. [32] proposed a skin cancer classification model using InceptionResNetV2 as backbone, aided with a soft attention unit. Here, the soft attention unit consists of two phases, a bilinear attention layer and a step to compute the weighted feature maps.

In the context of skin cancer, where lesion appearance can vary widely, ensemble techniques enable the system to recognize a broader range of features associated with different types of skin lesions. Harangi et al.'s study [33] involved using various deep CNN architectures like AlexNet [34], VGGNet, and GoogLeNet [35]. The final prediction was determined through a weighted majority vote, with each CNN's vote being weighted by its confidence as indicated by the softmax output. Ichim et al. [36] examined two ensemble models. The first model consisted of three neural networks, MobileNet, DenseNet121, and DenseNet169, with an ensemble of individual decisions determined by the weights associated with each individual network. The second model incorporated two networks, MobileNet and DenseNet169, and followed a horizontal voting approach, where the ensembling was determined by the voting from the best models associated with the considered number of epochs. The second ensemble strategy was observed to deliver superior results compared to the first.

Hasan et al. [37] introduced a hybrid CNN model comprising three distinct feature extractor modules, which were integrated to enhance the depth of feature maps for skin lesions. The fused feature maps underwent classification using separate fully connected layers, and their predictions were then ensembled to determine the lesion class. In the first level of ensembling, feature fusion was conducted through both channel averaging and channel concatenation. Ultimately, the output probability was determined by averaging the outputs of the fully connected layers, referred to as second level ensembling.

GANs have the potential to generate synthetic images for underrepresented skin lesion classes, thereby alleviating class imbalances and augmenting datasets related to skin cancer. Rashid et al. [38] introduced a skin lesion classification system based on GANs. In their approach, they performed data augmentation on a training set of images by incorporating synthetic skin lesion images generated using a GAN. The generator module in their system employed a deconvolutional network, while the discriminator module used a CNN as the classifier. The CNN was trained to classify skin lesions into seven different categories. Cheng et al. [39] introduced a GAN architecture featuring multiple convolutional layers and upsampling in the generator module. Their discriminator module consisted of a CNN and a gradient penalty function, aimed at improving image quality and preventing artifacts.

The role of ViTs in skin cancer classification involves leveraging their ability to learn representations from images broken down into patches and capture intricate patterns and features that distinguish among different types of skin lesions. Moreover, ViTs provide attention maps, highlighting areas, where the model focuses its attention, contributing to interpretability. Aladhadh et al. [40] designed

a two-tier framework to classify skin cancer. In the first stage, they applied various data augmentation techniques to tackle class imbalance. In the second stage, they employed a medical ViT, where the lesion images of size 72×72 were fed as input and each image was split into nine patches. Their transformer comprised three layers: an embedding layer, an encoder layer and a classifier layer. In the embedding layer, the transformer processed each patch as an individual token and then mapped it to a specific dimension with a learnable linear projection. The encoder layer contained self-attention and concatenation layers. The classifier layer predicted the final classification decision.

A four-block approach was proposed by Yang et al. [41]. In the first block, seven different classes of cancer were balanced using various data augmentation methods. The image restructuring block formed the second block, which was responsible for splitting a 2D input image into a sequence of patches of the same size. These patches underwent flattening into tokens with consistent dimensions, followed by positional embedding to retain spatial information. The resulting output served as the input for the subsequent transformer encoder block. The third block comprised the transformer encoder block, featuring N repeated layers. Each layer contained a multi-head self-attention layer and a fully connected feed-forward network. The features derived from this encoder were employed for cancer classification in the final classification block. This block included a flatten layer, two normalization layers, a dense layer, and a softmax layer.

Segmentation-guided classification techniques help isolate the lesion from the surrounding skin and other artifacts in the image. This reduction in noise and background interference leads to a cleaner input for the classification model, potentially improving its performance. These techniques enhance the extraction of features specifically from the identified ROI. Yu et al. [42] introduced a two-stage framework for melanoma detection. The first stage involved lesion segmentation, where an FCRN with 16 residual blocks was employed to accurately delineate the skin lesion from the surrounding healthy skin. This ensured that the subsequent classification focused specifically on the relevant region. In the second stage, a ResNet architecture was utilized to classify the segmented lesion as either melanoma or non-melanoma.

Hasan et al. [43] proposed the Dermo-DOCTOR system, utilizing end-to-end dual encoders for both segmentation and classification tasks. The model incorporated two distinct encoders, each specialized in extracting different features from input images. Encoder 1 focused on global features, capturing the overall structure and shape of the lesion, while encoder 2 concentrated on local features, extracting fine-grained details within the lesion. These encoders were seamlessly integrated into a single, end-to-end trainable architecture, enabling simultaneous

detection and recognition. The features extracted from both encoders were fused and directed into two separate branches: the detection branch, responsible for localizing the precise boundaries of the lesion within the image, and the recognition branch, which classified the lesion into different categories.

The combination of DL and ML techniques (hybrid approaches) holds significant importance for image classification tasks. DL methods, especially pre-trained CNN models, excel in feature extraction from images. When coupled with traditional ML algorithms like SVM [44] or RF [45], these hybrid models can leverage the strengths of both approaches, potentially leading to improved classification accuracy. 8 pre-trained CNN architectures, VGG16, VGG19, ResNet50, ResNet101, InceptionV3, DenseNet121, MobileNet, and Xception were used to extract deep features from dermoscopic lesion images by Gajera et al. [46]. They used SVM as a classifier to train the extracted features from all the deep CNNs, out of which DenseNet121 yielded the highest accuracy for melanoma detection. Keerthana et al. [47] utilized a hybrid CNN architecture of DenseNet201 and MobileNet to capture both low-level features like textures and edges and high-level features like lesion patterns and shapes. Then, they applied PCA to reduce the dimensionality of the extracted features and improve computational efficiency. Finally, they leveraged an SVM classifier trained on the set of features with reduced dimensionality.

Tembhurne et al. [48] proposed a multi-branch approach by combining ML and DL techniques for skin lesion classification. In the DL branch, they employed a pre-trained VGG16 network to extract high-level features and perform image classification. In the ML branch, they leveraged the contourlet transform technique [49] and LBPH [50] to extract texture and color features from the image. These features were then concatenated, subjected to dimensionality reduction via PCA, and fed into two ML models — logistic regression and linear SVM. The final classification was determined by combining the outcomes from both branches through a voting mechanism, resulting in the categorization of images as malignant or benign.

Multimodal techniques in skin cancer classification combine varied data streams, including images, clinical data, and pathology reports, providing a comprehensive and robust diagnostic solution. These approaches enhance generalization and improve diagnostic confidence by capturing multifaceted patterns. Integrating complementary information enhances interpretability, fostering a nuanced comprehension of skin lesions. Yap et al. [51] introduced a multimodal fusion model which combined information from three modalities: macroscopic image, dermoscopic image, and patient metadata. The macroscopic images were analyzed by a CNN for spatial features and the dermoscopic images were analyzed by a different CNN for finer details. Patient metadata was encoded with a separate network

using an embedding layer. The extracted features were fused and fed into a final neural network classifier to categorize the lesion images into malignant and benign.

Omeroglu et al. [52] introduced a multi-branch structure for multi-label skin lesion classification. They used two branches in their feature extraction phase, a dermoscopy branch and a clinical branch. A modified Xception architecture was used to extract visual features in the dermoscopy branch, whereas, they processed clinical data into numerical representations in the clinical branch. They also employed a soft attention module to analyze feature maps from both branches and assign weights based on their relevance to specific lesions. Subsequently, they designed a hyperbranched fusion block to combine weighted feature maps from different scales within each branch and across branches, creating a richer and more comprehensive representation of the lesion. Finally, they used a multi-label classification layer to compute output probabilities for each possible skin lesion label.

SM et al. [53] employed EfficientNetB6 as the backbone of their model to extract features related to melanoma and non-melanoma lesions. They also designed a simple neural network to train the contextual information given in the ISIC 2020 dataset [54]. The extracted features from both networks were concatenated and trained using an LGBM classifier [55]. In addition to this, they also utilized the Ranger optimizer [56] to improve the model's generalizability and overall performance.

The numerous research methodologies and articles, discussed in this literature survey, have been successful in addressing many challenges regarding skin cancer classification. Despite promising outcomes, the current methodologies fall short in delivering consistently strong performance within real-world settings. Some of the unresolved issues and open challenges that necessitate further attention and exploration are outlined below:

1. **Inter-Class Similarity:** A crucial challenge lies in distinguishing between different skin lesions due to their subtle variations and similarities. Both benign and malignant lesions often share visual characteristics, such as color, texture, shape, and patterns, blurring the distinction between harmless moles and potentially cancerous growths.

2. **Intra-Class Variability:** Another challenge stems from the variability within the same class of skin lesions. Even among malignant lesions, there exists considerable diversity in appearance. For example, melanomas can display a broad spectrum of colors, shapes, and patterns, complicating the establishment of clear boundaries between various types of lesions.

3. **Small Lesion Detection:** Detecting small and subtle skin lesions, espe-

cially in high-resolution images, is an uphill challenge. Generic DL models may fail to identify tiny lesions or produce false negatives due to their size and the presence of various imaging artifacts like portions of uninfected skin, hair, and veins.

4. **Limited Data and Imbalanced Datasets:** DL models require a large amount of labelled data for effective training. However, obtaining a diverse and comprehensive dataset for skin cancer images, with different types, stages, and demographics, can be challenging. Additionally, imbalances in class distribution can lead to biased models that perform poorly on underrepresented classes, thereby hindering the accurate classification of less common types of skin cancer, potentially leading to misdiagnosis or overlooking of critical cases.

5. **Limited Annotated Data:** Annotating and labelling skin lesion images by experts is time-consuming, expensive, and prone to human error. Developing efficient and reliable methods for annotating data is crucial to facilitate model training.

6. **Robustness to Variations:** Skin cancer images can exhibit variations due to inconsistent lighting conditions, background noise, camera and image quality. Models need to be robust to such variations to ensure accurate and consistent performance.

7. **Generalization to Diverse Population:** Models developed on data from specific demographics may lack generalizability across diverse skin colors, ethnicities, ages, or geographical regions. This is particularly important in skin cancer classification where the model should be able to perform accurately on various skin types.

8. **Interpretability and Explainability:** Skin cancer classification requires accurate and interpretable predictions to aid medical professionals in making informed decisions. DL models often lack transparency, making it challenging to explain their decisions to clinicians and patients. Interpretability is extremely crucial for trust and acceptance, especially in medical applications. Clinicians need to understand why a model arrived at a particular diagnosis to trust and validate its recommendations.

CHAPTER 3

Methods and Materials

This section provides a detailed discussion of the three novel DL methods proposed in this thesis. It outlines the architectural designs, training procedures, loss functions, and optimization strategies tailored for each method to address specific challenges in skin cancer classification. This section aims to provide a comprehensive overview of the underlying intuition behind the development of these methods by elaborating on their respective architectural designs. It evaluates their effectiveness and robustness in accurately identifying cancerous regions. Through this discussion, the thesis advances SoTA techniques in skin cancer classification, promoting better diagnostic outcomes and treatment strategies.

3.1 DFF-Net

In this section, a deep feature fusion network (DFF-Net) is proposed. This research presents a novel approach to enhance multi-class classification on challenging image datasets, specifically addressing the difficulties posed by the HAM1000, ISIC 2016, and ISIC 2017 datasets. It introduces a method which combines the features of two DL models, leveraging the feature fusion strategy. The main contributions of this work involves applying an efficient preprocessing technique based on contour detection and cropping the images to focus on the lesion areas. Following this, the method integrates Xception and Google Brain's BiT-M models as base learners, harnessing their respective capabilities to generate distinct feature maps obtained from the two models. These feature maps are then fused through an element-wise multiplication technique along the channel dimension, enabling the extraction of important and versatile features. A customized SEA module is employed to emphasize the most crucial features within the fused feature maps. Notably, to introduce enhanced non-linearity, the SELU activation function is incorporated alongside the ReLU activation function in the fully connected layers.

The findings highlight the effectiveness of this fusion-based technique and the integrated SEA module, offering significant implications for advancing the accuracy of multi-class skin cancer classification models, particularly when dealing with complex skin cancer datasets. In addition to this, a combination of categorical cross-entropy loss and triplet loss is introduced to evaluate the model's performance by quantifying the difference between the predicted output and the actual target values. Figure 3.1 illustrates the architecture of DFF-Net.

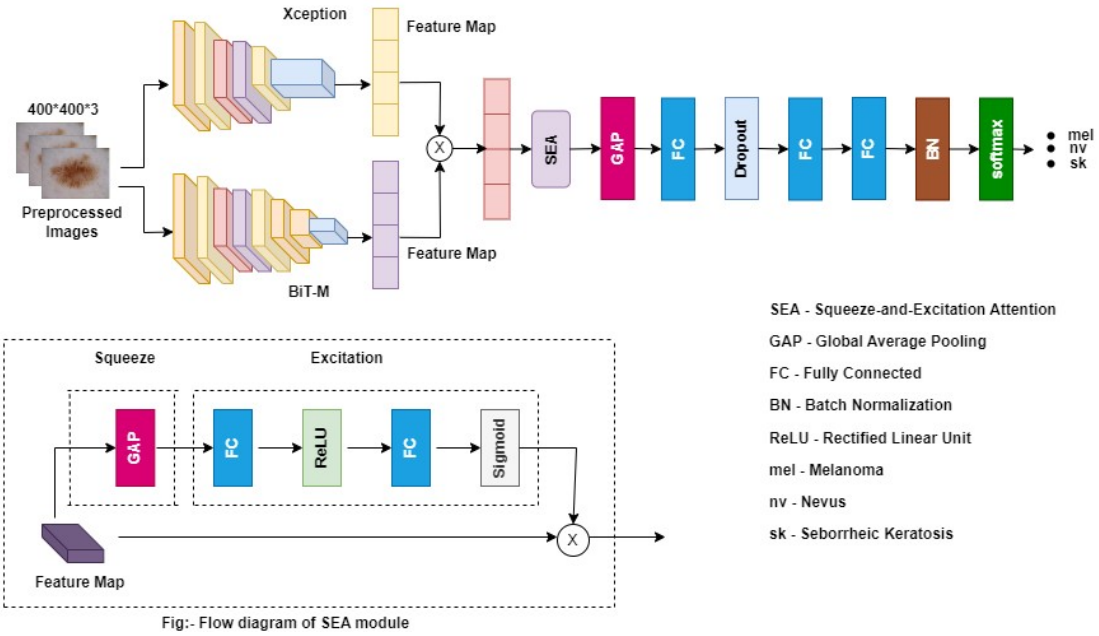


Figure 3.1: Architecture of the proposed DFF-Net model

3.1.1 Image Preprocessing and Augmentation

Classifying skin lesion images poses a significant challenge due to the necessity of focusing solely on a restricted area of relevance within the image. Meanwhile, various peripheral elements, such as sections of unaffected skin, hair, veins, and several other artefacts are deemed irrelevant to the classification task. Therefore, to increase the effectiveness of the proposed model, it becomes crucial to preprocess the images and convert them into a more appropriate format.

Here, the input images are resized to 400×400 dimensions using an interpolation technique called inter-linear exact interpolation. This resizing serves two main purposes: to reduce feature loss from the original high-resolution images and to alleviate computational load during network training while preserving most of the image content and features. The resized images are also cropped by extracting the extreme points of the contour surrounding the infected area and removing the rest. To identify the extreme points of the contour that encompass the infected

region, various image processing steps are employed. Initially, the image is converted to grayscale and subjected to a slight Gaussian blur filter for smoothing. Following this, the resulting image is thresholded, and a series of erosion and dilation operations are applied to eliminate unwanted noise. Once noise reduction is complete, the largest contours within the thresholded image are identified. Finally, the four extreme points — top, bottom, left, and right — are determined from these contours. These extreme points are subsequently used to crop the original image, effectively removing any extraneous regions that might introduce noise during the training process.

Since skin cancer datasets contain a relatively smaller number of images for training, several data augmentation techniques have been used to generate additional training examples from the existing datasets, effectively expanding the datasets' diversity, in order to develop a more robust model. Several image manipulation methods such as rotation, flipping, zooming, fading, shearing, and alterations in color space, have been employed to augment the dataset.

3.1.2 Xception

Xception is a deep CNN that draws inspiration from the Inception architecture, where Inception modules have been replaced with depth-wise separable convolutions. These depth-wise separable convolutions significantly reduce the number of parameters compared to traditional convolutional layers, leading to more efficient training and reduced computational burden. In the proposed DFF-Net, the final classification layer from the original Xception model is excluded. Images with 400×400 dimensions are used as input, and the resulting feature maps of size 13×13 are utilized for the novel feature fusion method.

3.1.3 BiT-M

BiT models, a Google Brain innovation, are a family of large scale pretrained neural networks that specialize in various computer vision tasks, primarily focusing on image classification. These models are notable for their effectiveness in TL. BiT models have been pretrained on diverse datasets to excel in numerous downstream tasks. Their architecture relies on a simplified ResNetV2 base, featuring deep residual blocks with skip connections. They have shown improved performance over individual TL models since their pre-trained weights serve as a strong initialization point for fine-tuning specific datasets. In this work, the BiT-M model, pre-trained on ImageNet-21k, which contains 14M images, is implemented. The images of size 400×400 serve as input to BiT-M. The output

is then reshaped to match the expected shape of the output feature maps from Xception before performing feature fusion.

3.1.4 Squeeze-and-Excitation Attention Module:

Based on Hu's work [12], a SEA module is employed to focus on relevant features after fusing the feature maps obtained from the two base learners, aiding the model to learn discriminative features that are crucial for making accurate predictions. The squeeze operation involves reducing the spatial dimensions of the feature maps by computing the average value of each feature map across its spatial dimensions, resulting in a single value per channel. The excitation operation consists of a small neural network comprising a couple of fully connected layers followed by activation functions (ReLU) and gating mechanisms (sigmoid). Within the excitation phase, the network learns channel-specific significance or weights, which are then applied to the feature maps to lay more emphasis on relevant features and suppress less relevant ones. In this model, after performing feature fusion, the feature maps with dimensions 13×13 form the input to the global average pooling layer within the squeeze phase. The features with reduced dimensions are then passed through the neural network in the excitation phase. Finally, an element-wise multiplication of these maps with the feature maps obtained after fusion is performed. The workflow of the SEA module is depicted in Figure 3.1.

3.1.5 Combined Loss Function

Since multi-class classification is being performed here, categorical cross entropy loss is employed to optimize the model.

$$L_{CCE} = -\sum_{i=1}^{C} t_i log(f(s)_i) \tag{3.1}$$

In Equation 3.1, $f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^{C} e^{s_j}}$. C represents the number of classes, and t_i signifies the ground truth label. s_i denotes the score generated by the CNN for each class i within the range of C and $f(s)_i$ represents the outcome after applying softmax to these scores.

Furthermore, triplet loss is implemented to address the problem of inter-class similarity. Triplet loss involves having sets of three data points: an anchor, a positive (similar class to that of the anchor class) example, and a negative (different class to that of the anchor class) example. This loss function incentivizes the model to minimize the distance between the anchor and the positive example,

while simultaneously maximizing the distance between the anchor and the negative example by a specified margin. Here, in Equation 3.2, $f(x)$ takes x as input and x^a, x^p and x^n denote anchor, positive and negative images respectively. α is the bias which acts as a threshold value.

$$L_{triplet} = \sum_{i=1}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right] \tag{3.2}$$

Equation 3.3 combines Equation 3.1 and Equation 3.2, and refers to the combined loss function used to measure the performance of the proposed model. In Equation 3.3, w serves as the scalar weight for determining the relative contribution of the two different types of loss functions. For all experiments, the value of w is assigned as 0.5.

$$L_{combined} = w \times L_{CCE} + (1 - w) \times L_{triplet} \tag{3.3}$$

3.2 EDB-Net

In dermatology, skin cancer classification relies heavily on evaluating asymmetry, border irregularity, color variation, and lesion diameter — commonly known as the ABCD rules. Dermatologists utilize these criteria to assess skin lesions for signs of malignancy. Edge information retrieved from edge maps of dermoscopic images has the potential to capture aspects of the "A" (asymmetry) and "B" (border irregularity) components of the ABCD rules by delineating lesion boundaries and contours. These maps aid in detecting irregularities in shape and symmetry, key indicators of potential malignancy. This motivates the development of an edge-guided neural network, prioritizing the recognition of intricate edge patterns to improve classification accuracy and diagnostic efficacy in dermatology.

In this work, a novel edge-guided dual-branched network (EDB-Net) is introduced. This architecture comprises two branches: one branch is dedicated to processing the edge maps generated from the dermoscopic images, while the other branch focuses solely on the original dermoscopic images. In both branches, a DenseNet-based backbone is employed, consisting of dense and transition blocks, to extract features from their respective inputs. The dense blocks promote feature reuse and facilitate the propagation of gradient information across the network. Through the iterative application of convolutional layers, combined with batch normalization and ReLU activation within dense blocks, the network adeptly captures features at varied levels of abstraction. Meanwhile, the transition blocks complement the dense blocks by facilitating feature extraction at multiple resolutions. By downsampling the feature maps, these blocks enable the extraction of

coarse-grained features while preserving fine details present in the images.

3.2.1 Novel Method for Edge Map Generation

In this section, a detailed explanation of the novel methodology developed for creating edge maps from dermoscopic images is provided. Each step of the algorithm mentioned in 1 for generating edge maps is elaborated here, presenting a comprehensive overview of the method.

3.2.1.1 Preprocessing

Median blur is applied to mitigate noise from the dermoscopic image in the BGR color space. This operation involves computing the median value for each pixel within a 5×5 square neighborhood centred around its position. For a pixel at position (x, y), the operation can be represented mathematically as $Output(x, y) = Median(Neighborhood(x, y))$. This process enhances the overall quality of the image for subsequent analysis.

3.2.1.2 K-means Clustering

K-means clustering, an unsupervised learning algorithm, partitions data points into K clusters. The objective function of this algorithm, denoted in Equation 3.4, measures the WCSS.

$$\text{WCSS} = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{3.4}$$

In Equation 3.4, K represents the number of clusters, C_i denotes the i^{th} cluster and x signifies a data point in C_i. μ_i and $\| \cdot \|$ denote the centroid of C_i and Euclidean distance respectively. The algorithm iteratively updates cluster assignments and centroids to minimize this objective function. In the methodology, K-means clustering is applied to the preprocessed image with K=8 clusters. The cluster centres are computed, and each pixel in the image is assigned to its nearest cluster centre. Ultimately, the image is reconstructed using these cluster centres, yielding a segmented image characterized by pixels with similar attributes within each cluster.

3.2.1.3 Adaptive Histogram Equalization

The K-means segmented image undergoes a conversion from BGR color space to HSV color space. This transformation facilitates the independent processing of color and intensity information. Subsequently, CLAHE is applied exclusively to the value channel, representing image intensity or brightness. The hue and

saturation channels remain unaltered to preserve the color information. CLAHE restricts contrast enhancement to localized regions by dividing the image into small tiles and applying histogram equalization individually to each tile. The threshold parameter for contrast limiting is specified as 3, with a tile grid size of 8×8 for CLAHE initialization. Then, the individual hue, saturation, and value channels are merged together to produce the enhanced image in the LAB color space. Finally, the image is converted from LAB to BGR color space, thereby restoring the image to its original color representation while incorporating enhanced contrast in the value channel, due to the CLAHE operation.

3.2.1.4 Edge Detection

The BGR image is processed in a way suitable for input to a HED model for predicting edge maps. This processing typically involves scaling pixel values and performing mean subtraction. Following this, the resultant image is utilized as input for the pre-trained HED model. Unlike conventional edge detectors, which rely on low-level features such as gradients or filters, HED employs DL to acquire rich hierarchical representations of edges from the input image. The fundamental concept underlying HED involves conducting edge detection across multiple scales, followed by fusing the multi-scale information to generate a holistic edge map.

HED consists of a stem network, side output layers, and a weighted fusion layer. A pre-trained CNN model like VGGNet serves as the backbone for hierarchical feature extraction from the input image. The feature maps from the stem network are fed into the side output layers across different stages. Each side output layer produces an edge map at a specific scale, capturing edges at varying levels of granularity. These edge maps from all the side output layers are then fused together using the weighted fusion layer. The weighted fusion layer assigns learned weights to the edge maps from each side output layer and combines them, thereby allowing selective emphasis or de-emphasis based on the significance for edge detection.

In this work, HED is utilized solely for generating edge maps, without altering the model's parameters or performing any training or optimization. The pre-trained weights are directly employed to predict edge maps for the input image. During the forward pass of the input image through the HED model, the image traverses through the layers of the network, and edge features are computed at each layer. Following this process, the HED model generates a set of edge maps representing detected edges in the input image. These edge maps encompass edges identified at diverse scales and levels of detail, capturing both fine and coarse edges within the image. For the input image, edge map predictions from the side

output layers ($\hat{Y}_{\text{side}}^{(1)}, \ldots, \hat{Y}_{\text{side}}^{(M)}$) and the weighted-fusion layer ($\hat{Y}_{\text{wfuse}}$) are obtained. Here, M denotes the number of side-output layers. The final consolidated edge map ($\hat{Y}_{\text{HED}}$) is achieved by aggregating these generated edge maps denoted in Equation 3.5.

$$\hat{Y}_{\text{HED}} = \text{Average}(\hat{Y}_{\text{wfuse}}, \hat{Y}_{\text{side}}^{(1)}, \ldots, \hat{Y}_{\text{side}}^{(M)}) \tag{3.5}$$

The resulting edge map undergoes rescaling to ensure that edge intensities fall within the appropriate range for visualization, typically between 0 and 255.

Algorithm 1 Algorithm for Edge Map Generation

Input: Dermoscopic image
Output: Edge map of the input image

1. Apply median blur with a 5×5 kernel to the dermoscopic image.
2. Perform K-means clustering with $K = 8$ on the image.
3. Convert the clustered image from BGR to HSV color space.
4. Split the HSV image into three channels: hue (H), saturation (S), and value (V).
5. Apply CLAHE to the value (V) channel.
6. Merge the modified value (V) channel with the original hue (H) and saturation (S) channels to obtain the enhanced LAB image.
7. Convert the enhanced LAB image back to BGR color space.
8. Construct a processed image by performing pixel scaling and mean subtraction on the BGR image.
9. Set the resultant image as the input to the pre-trained HED model.
10. Perform a forward pass of the input image through the network.
11. Combine the edge map predictions from the side output layers and the weighted-fusion layer using Equation 3.5
12. Return the final consolidated edge map $\hat{Y}_{\text{HED}}$.

3.2.2 Proposed Dual-Branched CNN

This section deals with the architectural details of the proposed dual-branched CNN. Within the first branch, a combination of dense and transition blocks is leveraged to effectively capture edge features from the input edge maps. In the second branch, DenseNet201 is utilized as the backbone, supplemented with an

attention module. This module enhances the network's capability to focus on salient regions within the dermoscopic images, thereby capturing relevant features vital for classification. Furthermore, a novel strategy is proposed to modulate the attention maps using the previously extracted edge features. By generating edge-guided features using the enhanced attention maps, the aim is to refine the feature representations, harnessing complementary information from both sources. In the final stage of the network architecture, both the attended features and the edge-guided features are aggregated. Subsequently, these features are flattened and fed into the final dense layers for classification. This innovative approach facilitates the generation of more discriminative and contextually relevant features for classification. Throughout this section, a detailed exploration of the dual-branched network architecture is provided, elucidating the rationale behind each architectural choice and discussing the integration of edge features and attention mechanisms for improved classification performance. Figure 3.2 illustrates the architecture of EDB-Net.

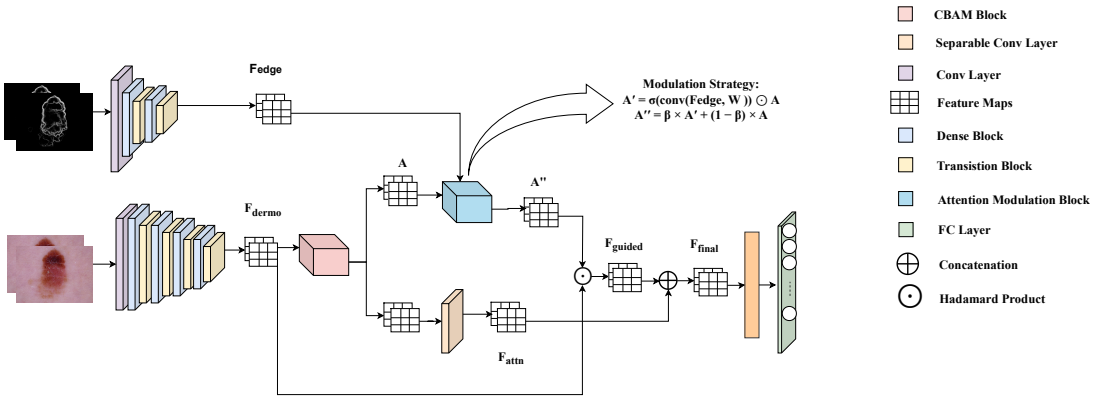


Figure 3.2: Architecture of the proposed EDB-Net model

3.2.2.1 Branch 1 - Edge Feature Extraction

Here, a set of dense and transition blocks is used for extracting the edge features, F_{edge}, from the input edge maps. The architectural sequence commences with an initial convolution operation to extract basic features from the input images. This convolutional layer is followed by subsequent batch normalization and ReLU activation to normalize and introduce non-linearity into the feature maps. A sequence of dense blocks and transition blocks is then iteratively applied to capture increasingly complex edge features. Within this iterative process, two dense-transition iterations are specified, each comprising a dense block with 4 layers and a growth rate of 16, followed by a transition block. The decision to limit the number of dense-transition iterations underscores the aim to prioritise the extraction of edge features while minimizing the likelihood of the network acquiring

extraneous information. This strategic approach aligns with the research goal of developing a specialized model tailored specifically for the edge feature extraction task.

3.2.2.2 Branch 2 - Dermoscopic Feature Extraction

Here, the DenseNet201 CNN is used to extract features, F_{dermo}, from the dermoscopic images. The DenseNet201 model is customized by excluding the last 28 layers and integrating CBAM to the dense block within the network architecture, where the image feature map size is 7×7. The CBAM block involves the computation of both channel and spatial attention mechanisms. Figure 3.3 depicts the architecture of this block. Channel attention is computed by globally pooling the input tensor, F_{dermo}, along the channel dimension, extracting both average and maximum channel-wise activations. These activations are then fused through dense layers to generate the channel attention map. Spatial attention is computed by averaging and max-pooling activations along spatial dimensions, followed by convolutional operations to derive the spatial attention map. The resultant channel and spatial attention maps are then fused through element-wise multiplication, yielding a unified attention map. This attention map is further refined via additional convolutional operations to produce a final attention map, A. This refined attention map, A, is subsequently applied to the input features, F_{dermo}, through element-wise multiplication. Following this operation, a SeparableConv2D layer is employed, resulting in enhanced feature representations, F_{attn}. These refined representations encapsulate both channel-wise and spatially relevant information, facilitating improved performance in the subsequent classification task.

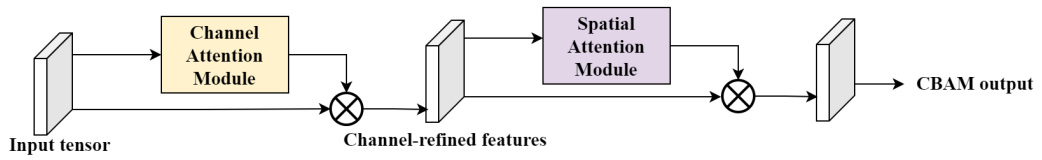


Figure 3.3: Architecture of the CBAM block

3.2.2.3 Modulation of Attention Maps

Here, the attention map A is refined through the integration of edge features, F_{edge}. These features are instrumental in modulating the attention map, extracted using the CBAM block, enhancing the model's ability to focus on ROIs defined by prominent edges within the input images. This modulation process is governed by a gating mechanism equipped with learnable convolutional weights W, which

adaptively capture contextually relevant edge patterns. The gating mechanism is formulated in Equation 3.6.

$$A' = \sigma(\text{conv}(F_{\text{edge}}, W)) \odot A \tag{3.6}$$

Here, in Equation 3.6, σ represents the sigmoid activation function, that normalizes the convolution output, ensuring that the values lie within the range $[0, 1]$. This normalization is crucial as it conditions the modulation to be conducive for multiplicative scaling. The operation $\odot$ represents the Hadamard product. This operation allows the edge features to selectively enhance or suppress features within the attention map A, based on the edge information. This process aligns the model's focus with the most salient regions marked by significant edge features

Further, to introduce flexibility and maintain a balance between the original and edge-enhanced attention maps, a hybrid modulation strategy is applied to generate a refined attention map A'', defined in Equation 3.7.

$$A'' = \beta \times A' + (1 - \beta) \times A \tag{3.7}$$

Here, in Equation 3.7, β is a trainable parameter that determines the extent to which the edge-modulated attention map A' influences the final attention representation A''. Such a hybrid approach not only enriches the attention mechanism with precise, edge-based contextual information but also preserves the integrity of the initial attention cues. This strategy ensures that the network remains robust to variations in edge relevance across different contexts.

3.2.2.4 Integration of Edge-Enhanced Attention

This stage is pivotal for incorporating the edge-modulated attention map A'' into the feature maps to enhance the model's feature representations. Initially, the Hadamard product is employed to fuse the features extracted during the dermoscopic feature extraction phase, F_{dermo}, with the refined attention map A''. This operation, denoted in Equation 3.8, produces the edge-guided feature maps, F_{guided}, which are enhanced by the attentional focus driven by edge-specific information.

$$F_{\text{guided}} = F_{\text{dermo}} \odot A'' \tag{3.8}$$

Subsequently, to ensure a comprehensive feature representation, these edge-guided features F_{guided} are concatenated with F_{attn}, the features initially generated using the CBAM block. This concatenation, shown in Equation 3.9, helps preserve the integrity of both the original attention-focused features and the newly integrated

edge-enhanced features, culminating in an enriched feature set F_{final}.

$$F_{\text{final}} = F_{\text{guided}} \oplus F_{\text{attn}} \tag{3.9}$$

Finally, the enriched feature set F_{final} is flattened and fed into the final dense layers to perform the classification task.

3.2.3 Loss Function

In all the experiments, categorical cross entropy loss has been applied to optimize the networks.

$$Loss = -\sum_{i=1}^{C} t_i log(f(s)_i) \tag{3.10}$$

Here, in Equation 3.10, $f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^{C} e^{s_j}}$, C denotes the number of classes, and t_i is the ground truth label. s_i is the score generated by the CNN for each class i in C and $f(s)_i$ denotes the result after softmax gets applied to these scores.

3.3 DSCC-Net

This work proposes a novel network (DSCC-Net) that combines the outcomes of three attention-based CNNs using DS theory. This classifier combination approach utilizes insights from multiple discriminative models, thereby reducing overfitting and significantly improving the model's ability to classify unseen data. Additionally, a combination scheme can mitigate individual models' errors and produce more reliable predictions. DS theory, a mathematical evidence theory useful for reasoning under uncertainty and combining evidence from multiple sources, is particularly beneficial in handling multiple classifiers that may produce conflicting predictions. By using belief functions to represent conflicts in individual classifier outputs, this approach can lead to better decision-making. These factors motivated the development of the proposed method. For this method, three base learners — ResNet50, InceptionResNetV2, and DenseNet201 — each aided by an attention module, are trained. These base learners produce confidence scores, which are then combined using DS theory, as elaborated in section 3.3.6. Final decisions are made by selecting the class with the highest belief value. The architecture of DSCC-Net is presented in Figure 3.4.

3.3.1 Data Augmentation

Skin cancer datasets are highly imbalanced since the distribution of the different classes has an uneven structure. Therefore, to tackle the problem of over-

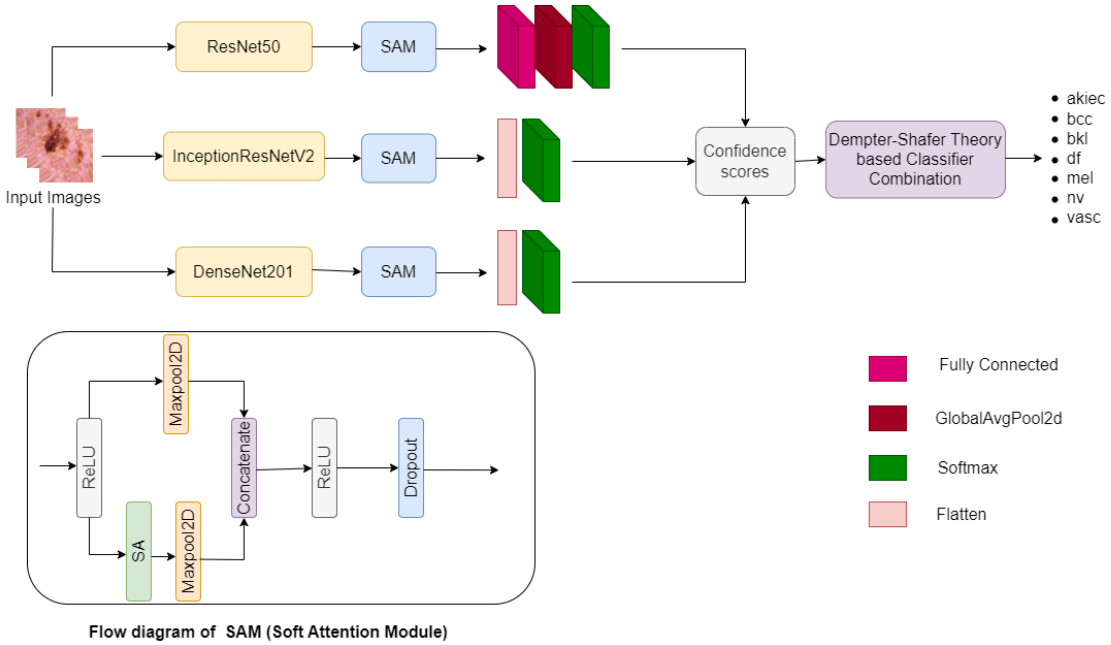


Figure 3.4: Architecture of the proposed DSCC-Net model

fitting, and to reduce the probability of misclassification, data augmentation has been used as an oversampling technique to balance the imbalanced datasets. Data augmentation operations used in this work comprises various geometric image transformation techniques like rotation, zooming, flipping, scaling, color space transformations, cropping, fading, and shearing.

3.3.2 Soft Attention Module

Inspired by Datta's work [32], SA has been employed in this method for skin cancer classification. The classification of skin lesion images poses significant challenges, as only a limited portion of the image is relevant, while various peripheral artifacts like portions of uninfected skin, hair, and veins are considered irrelevant. SA is a technique based on weighted sum, where it disregards irrelevant areas by assigning lower weights to their corresponding feature maps, while assigning higher weights to sensitive areas' feature maps. Therefore, SA is utilized to place more emphasis on relevant features of the skin lesion images. This ensures that when CNNs extract image features, the SAM captures only the appropriate areas.

$$f_{sa} = \gamma t((\sum_{k=1}^{K} softmax(W_k * t))) \tag{3.11}$$

Here, in Equation 3.11, feature tensor forms the input to a 3D convolution layer with weights $W_k \in \mathbb{R}^{h \times w \times d \times K}$, where K is the number of 3D weights. Subsequently, a softmax function is used to normalize each of the K attention maps.

These maps are then combined into a composite attention map which behaves as a weighting function α. This α value is multiplied with t, and then further scaled by a learnable scalar γ. Finally, the scaled attentive features f_{sa} are combined with the input tensor t. Figure 3.5 depicts the overview of this SA unit.

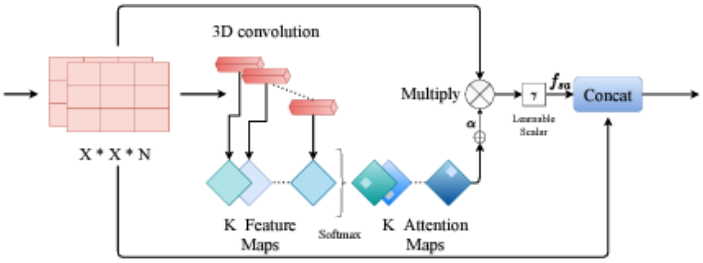


Figure 3.5: Overview of the SA unit

In this work, the SA layer is passed through a maxpool layer with 2×2 pool size. It is then combined with the standard maxpool layer of the corresponding CNN model used. This concatenated layer is then passed through a ReLU function and followed by a dropout layer to regularize the output of the SA layer. The workflow of SAM is depicted in Figure 3.4.

3.3.3 ResNet50

ResNet50 is a deep CNN that reformulates layers as learning residual functions, emphasizing hierarchical feature learning. In this approach, the last three layers of the original ResNet50 model are excluded. The SAM is then attached after the 5th convolutional block, where the feature map size is 7×7. The output from the SAM is flattened and passed through a final output layer containing seven neurons, followed by a softmax activation unit. This process generates the confidence scores.

3.3.4 InceptionResNetV2

InceptionResNetV2 is a deep CNN that combines the Inception architecture with residual connections. The Inception modules extract multi-scale features, while the residual functions capture hierarchical features. In this approach, the last 28 layers of the original InceptionResNetV2 model are excluded. The SAM is then added to the InceptionResNet block of the model, where the feature map size is 8×8. The output from the SAM is flattened, and after passing through the final dense layer with seven neurons and a softmax activation unit, the confidence scores are obtained.

3.3.5 DenseNet201

DenseNet201 is developed based on the idea that CNNs with relatively shorter connections between layers close to the input and output layers are more accurate and efficient to train. In this approach, the last 28 layers of the original DenseNet201 model are excluded. The SAM is then attached to the dense block, where the feature map size is 7×7. The output from the SAM is flattened and passed through a final dense layer consisting of seven neurons, followed by a softmax activation unit, which provides the confidence scores.

3.3.6 Dempster-Shafer Theory based Classifier Combination

DS theory is a mathematical theory, which considers all possible outcomes of a problem, and is widely used to express uncertainty. In the proposed approach, each base learner produces different outcomes for each test image. For every subset of hypotheses, these base learners specify a BPA to denote the degree of confidence of that image belonging to that particular subset. The BPAs from all base learners are combined using Dempster's rule of combination given in Equation 3.12. In Equation 3.12, m_1 and m_2 are the BPAs of subsets B and C and $m_{1,2}$ is the combined BPA, where A is the intersection of subsets B and C. After the combination, final decisions are made by selecting the hypothesis with the highest belief value. Algorithm 2 gives the details of this approach for a single skin lesion image.

$$m_{1,2}(A) = \begin{cases} 0, & \text{if } A = \Phi \\ \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \Phi} m_1(B) m_2(C)}, & \text{if } A \neq \Phi \end{cases} \tag{3.12}$$

3.3.7 Loss Function

In all the experiments, categorical cross entropy loss has been applied to optimize the networks.

$$Loss = -\sum_{i=1}^{C} t_i log(f(s)_i) \tag{3.13}$$

Here, in Equation 3.13, $f(s)_i = \frac{e^{s_i}}{\sum_{j=1}^{C} e^{s_j}}$, C denotes the number of classes, and t_i is the ground truth label. s_i is the score generated by the CNN for each class i in C and $f(s)_i$ denotes the result after softmax gets applied to these scores.

Algorithm 2 Dempster-Shafer Theory based Classifier Combination Scheme

$Pred_i \leftarrow Confidence\ scores\ for\ all\ classes\ given\ by\ the\ ith\ classifier$ ▷ $i \in [1...\mathbb{N}]$
$Labels \leftarrow Set\ of\ all\ output\ classes$ ▷ $Labels = \{1, 2, ..., C\}$, where C is no. of classes
function DS_BASED_CLASSIFIER($Pred_i, Labels$)
  for $i \leftarrow 1$ to N **do** ▷ Create power set for each classifier by using Labels
    $P_i \leftarrow CreatePowerset(Labels)$
    for $e \leftarrow 1$ to $length(P_i)$ **do** ▷ Assign elements of power set
      $P_i[e].value \leftarrow AssignMinValue(Pred_i, P_i[e])$ ▷ Assign minimum confidence score from labels belonging to the subset $P_i[e]$
    end for
  end for
  for $i \leftarrow 1$ to N **do** ▷ Convert power sets into BPAs
    $sum \leftarrow \sum_{e=1}^{length(P_i)} P_i[e].value$
    $P_i.value \leftarrow P_i.value/sum$
  end for
  $Combined_mass \leftarrow DSCombinationRule(P_1, P_2, ..., P_N)$ ▷ Applying Equation 3.12
  $Predicted_subset \leftarrow Max_subset(Combined_mass)$ ▷ $Predicted_subset$ stores the subset with maximum mass value
  if $length(Predicted_subset) == 1$ **then** ▷ Label prediction for singleton subset
    return $Output_label \leftarrow Predicted_subset$
  end if
  if $length(Predicted_subset) > 1$ **then** ▷ Label prediction if subset contains multiple labels
    for $i \leftarrow 1$ to N **do**
      $SUM[i] \leftarrow \sum_{j \in Predicted_subset} Pred_i[j]$
    end for
    $MC \leftarrow Max_classifier(SUM)$ ▷ Store classifier id with maximum total confidence score according to $Predicted_subset$
    $max \leftarrow -1$
    for $k \leftarrow 1$ to $length(Predicted_subset)$ **do**
      if $Pred_{MC}[Predicted_subset[k]] > max$ **then**
        $Output_label \leftarrow Predicted_subset[k]$
      end if
    end for
  end if
  return $Output_label$ ▷ Final output
end function
output_label ←DS_BASED_CLASSIFIER($Pred_i, Labels$)

CHAPTER 4

Results and Discussion

The results presented in this section highlight the efficiency of the proposed models in the field of skin cancer classification. Section 4.1 details the different datasets considered, section 4.2 highlights the system configuration that is used for performing all the experiments, and section 4.3 elaborates on the evaluation metrics that are used. Section 4.4, section 4.5, and section 4.6 show the results and highlights the analysis that is drawn from the three methods proposed in this thesis. Finally, section 4.7 presents the interpretability of the proposed models using Grad-CAM [57], and section 4.8 presents the comparison of the proposed methods with several recent existing methods for skin cancer classification.

4.1 Datasets

Numerous computer vision-based systems aimed at diagnosing skin cancer have been proposed. However, effectively evaluating their diagnostic accuracy and validating predicted results necessitates a trustworthy compilation of dermoscopic images. Hence, the importance of having a standardized and reliable collection of dermoscopic images becomes very essential. This section delves into three commonly used real-world datasets — HAM10000, ISIC 2016, and ISIC 2017 — that are used for assessing the proposed techniques in skin cancer classification. Table 4.1 gives the details of these datasets in brief and Figure 4.1 presents sample images of skin lesions taken from HAM10000 and ISIC 2017 datasets.

4.1.1 HAM10000

The Human Against Machine dataset [58], also known as HAM10000, is the latest publicly available skin lesions dataset comprising dermoscopic images sized at 450×600 pixels. It took two decades to assemble this dataset. These images primarily originated from Cliff Rosendahl's skin cancer practice in Queensland,

Table 4.1: Datasets for skin cancer classification

Name	**Year**	**# Classes**	**# Images**	**Official Split**
HAM10000	2018	7	10015	No
ISIC 2016	2016	2	1279	Yes (Train: 900, Test: 379)
ISIC 2017	2017	3	2750	Yes (Train: 2000, Valid: 150, Test: 600)

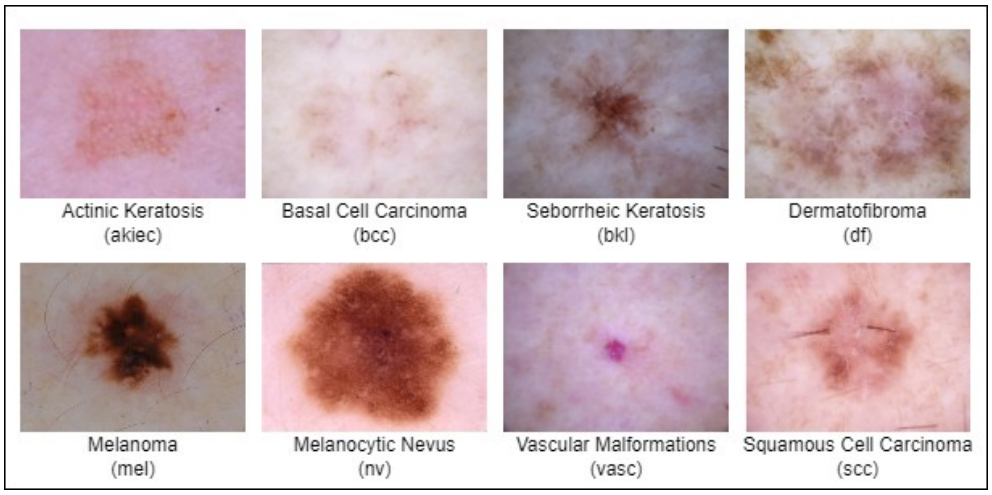


Figure 4.1: Sample images of skin lesions in HAM10000 and ISIC 2017 datasets

Australia, and the Dermatology Department of the Medical University of Vienna, Austria. Collected from diverse populations, various acquisition and cleaning methods were employed, along with the development of semi-automatic workflows, to address diversity issues. The dataset contains 10,015 images categorized into seven skin disease groups: 327 actinic keratosis and intraepithelial carcinoma images, 514 basal cell carcinoma images, 1099 benign keratosis images, 115 dermatofibroma images, 1113 melanoma images, 6705 melanocytic nevus images, and 142 vascular malformation images. Since there is no official split for this dataset, an 80-20 split was employed for all experiments, consistent with the approach used by other SoTA methods.

4.1.2 ISIC Archive

The ISIC archive [59] is a collection of numerous skin cancer datasets. These datasets were introduced by ISIC at the ISBI as individual challenges in order to improve melanoma diagnosis. For evaluating the proposed methods in this thesis, two datasets from this archive have been selected.

The ISIC 2016 challenge [60] contains dermoscopic images segregated into training and testing subsets. The training set consists of 900 images, while the testing subset comprises 379 images. These images are categorized into two classes: malignant melanomas and benign nevi, with around 30.30% representing melanoma lesions and the rest belonging to benign nevi.

The ISIC 2017 challenge [61] comprises 2,750 images, belonging to three classes: melanoma, nevus and seborrheic keratosis. The dataset comprises 2,000 training images, 150 validation images, and 600 testing images spanning all three classes. Notably, images in this dataset are captured by various imaging devices, leading to non-uniform resolutions. These images exhibit a range of resolutions, from smaller sizes like 300×200 pixels to larger resolutions such as 6000×4000 pixels. This diversity in image resolutions mirrors real-world scenarios, introducing complexity when analyzing images sourced from different devices.

4.2 System Configuration

A Jupyter Notebook with a 12 GB NVIDIA Tesla T4 GPU in Google's collaborative environment is used to run the whole set of experiments. Tensorflow, Keras, OpenCV, Matplotlib, scikit-learn, NumPy, and pandas are the main open-source modules that have been used for implementing the suggested methods in Python.

4.3 Evaluation Metrics

Evaluation metrics play a critical role in gauging the effectiveness and utility of a model. These metrics are instrumental in determining a model's ability to accurately predict the appropriate type or label for an input in classification tasks. In this section, the commonly employed evaluation metrics utilized for skin cancer classification tasks are explored. Before delving into the discussion on these metrics, it's essential to familiarize ourselves with the subsequent terms:

- **True Positive:** This refers to the number of instances that are actually positive and are correctly predicted as positive by the model. In a medical context, it would be when the model correctly identifies a patient as having a particular condition or disease.
- **False Positive:** This refers to the number of instances that are actually negative but are incorrectly classified as positive by the model. In a medical scenario, it would mean the model falsely indicates a patient as having a condition when they don't.
- **False Negative:** This refers to the number of instances that are actually positive but are incorrectly predicted as negative by the model. In a medical context, it would mean the model fails to identify a patient who actually has a condition, falsely indicating them as healthy.
- **True Negative:** This refers to the number of instances that are actually negative and are correctly predicted as negative by the model. In a medical setting, it would mean the model correctly identifies a healthy individual as not having a particular condition.

4.3.1 Accuracy

Accuracy [62] computes the proportion of accurately predicted samples, encompassing both true positives and true negatives ($TP + TN$), relative to the total sample size. It assesses the overall correctness of the model's predictions. Equation 4.1 denotes the formula for accuracy.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{4.1}$$

4.3.2 Precision

Precision [62] quantifies the proportion of accurately predicted positive observations (TP) relative to all predicted positive instances ($TP + FP$). It measures

the accuracy of positive predictions, indicating the relevance of selected items within the predictions. Equation 4.2 denotes the formula for precision.

$$Pre = \frac{TP}{TP + FP} \tag{4.2}$$

4.3.3 Sensitivity

Sensitivity [62], also known as recall, measures the ratio of correctly predicted positive observations (TP) concerning all actual positive instances ($TP + FN$). This metric signifies the model's capability to recognize all relevant instances, assessing its completeness in identification. Equation 4.3 denotes the formula for sensitivity.

$$Sen = \frac{TP}{TP + FN} \tag{4.3}$$

4.3.4 Specificity

Specificity [62] computes the proportion of accurately predicted negative observations (TN) in relation to all actual negative instances ($TN + FP$). This metric evaluates the model's capability to accurately recognize negative instances. Equation 4.4 denotes the formula for specificity.

$$Spe = \frac{TN}{TN + FP} \tag{4.4}$$

4.3.5 F1 Score

F1 score [62] represents the harmonic mean of precision and recall, amalgamating both metrics into a unified measure. By striking a balance between precision and recall, it offers a singular score that comprehensively considers both false positives and false negatives. Equation 4.5 denotes the formula for F1 score.

$$F1\ score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4.5}$$

4.3.6 Receiver Operating Characteristic Curve

The ROC curve [63] graphically depicts how well a classification model performs across a range of thresholds. By plotting the true positive rate ($sensitivity$) against the false positive rate ($1-specificity$) across various thresholds, each point on the curve reflects a $sensitivity/specificity$ pair linked to a specific threshold setting. A superior model showcases an ROC curve closer to the top-left corner, signalling increased sensitivity and specificity across different threshold values.

4.3.7 Area Under the Curve

The AUC [63] metric measures the total area enclosed by the ROC curve, summarizing a classification model's performance across all feasible thresholds. A higher AUC value, closer to 1, signifies the model's enhanced ability to differentiate between classes. An AUC of 0.5 indicates random chance, while an AUC of 1 represents a perfect classifier.

4.4 DFF-Net

This section outlines the training configuration and empirical findings derived from the training and testing phases of DFF-Net proposed in this study.

4.4.1 Training Setup

To capture and extract unique patterns relevant to the skin cancer domain, the network has been retrained on the training set by fine-tuning all layers for 60 epochs with an early stopping patience of 30. For all training-related tasks, a learning rate of 0.01 and the Adam optimizer [64] with an epsilon value of 0.1 are used.

4.4.2 Experimental Results

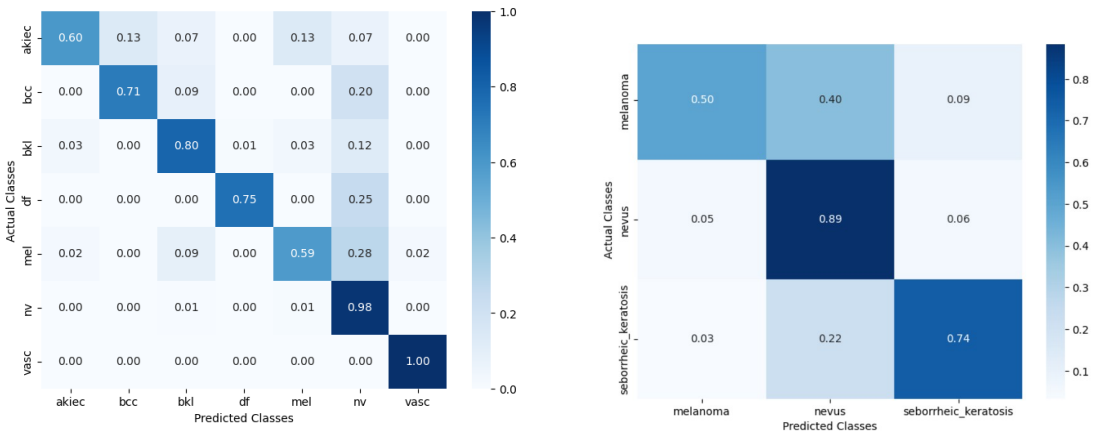
To select the primary base learners for this method, several models were trained including AlexNet, VGG16, ResNet50, Xception, and BiT-M, out of which Xception and BiT-M emerged with superior performance. These findings have prompted an exploration of additional architectural configurations using Xception and BiT-M as backbone networks. The different configurations are as follows:
(i) Preprocessing + Xception + SEA module
(ii) Preprocessing + BiT-M + SEA module
(iii) Preprocessing + fusion of Xception and BiT-M + SEA module + L_{CCE}
(iv) Preprocessing + fusion of Xception and BiT-M + SEA module + $L_{combined}$ (DFF-Net)

From Table 4.2, it can be seen that the introduction of the contour-crop based preprocessing and the SEA module provides a necessary boost to the overall performance by removing the unimportant artefacts of the images and only focusing on the lesion areas. A further improvement in accuracy and F1-score is achieved by implementing the proposed feature fusion technique using the features extracted from Xception and BiT-M models. Figure 4.2a and Figure 4.2b depict the confusion matrix of DFF-Net on HAM10000 and ISIC 2017 datasets respectively.

Table 4.3 reports the classification results of this method across all three datasets considered for this thesis.

Table 4.2: Ablation results for selecting the best model on ISIC 2017 dataset

Model	Accuracy	Precision	Recall	F1-score
Xception	0.748	0.752	0.745	0.748
BiT-M	0.738	0.732	0.736	0.734
(i)	0.770	0.773	0.761	0.767
(ii)	0.767	0.764	0.762	0.763
(iii)	0.780	0.782	0.775	0.778
(iv)	**0.795**	**0.791**	**0.795**	**0.792**



(a) (b)

Figure 4.2: Confusion matrix of DFF-Net on (a) HAM1000 dataset; (b) ISIC 2017 dataset

Table 4.3: Classification results achieved by DFF-Net across three datasets

Dataset	Accuracy	Precision	Recall	F1-Score
HAM10000	0.922	0.923	0.923	0.922
ISIC 2016	0.856	0.854	0.856	0.855
ISIC 2017	0.795	0.791	0.795	0.792

Accurately classifying skin cancer includes considerable complexity, and skin cancer datasets pose specific challenges of their own. To address this, a novel approach is introduced using deep feature fusion with Xception and BiT-M models as backbones, along with a customized SEA module. This fusion combines multi-level and multi-scale features extracted from different layers of the base learners, which leads to improved representations and enhanced discriminative power. The ablation study ensures that Xception and BiT-M capture different aspects of visual

information. Feature fusion leverages the complementary nature of these models to provide feature maps rich with diverse information. The main advantage of this approach is that by utilizing CNNs with diverse architecture as feature extractors, deep feature fusion can be applied to a wide array of classification problems.

4.5 EDB-Net

This section presents the training configuration and empirical results obtained from the training and testing phases of the EDB-Net model proposed in this study.

4.5.1 Training Setup

To capture and extract unique patterns relevant to the skin cancer domain, the network has been retrained on the training set by fine-tuning all layers for 60 epochs with an early stopping patience of 30. For all training-related tasks, a learning rate of 0.01 and the Adam optimizer with an epsilon value of 0.1 are used.

4.5.2 Experimental Results

Notably, data augmentation was not utilized, as the objective for this work was to perform multi-class classification, capable of handling significant dataset imbalances. To select the primary backbone network, several models were trained on the HAM10000 dataset, including VGG16, ResNet34, ResNet50, MobileNetV2 [65], InceptionResNetV2, DenseNet121, and DenseNet201, out of which DenseNet201 emerged with the best performance. The results presented in Table 4.4 justifies this finding.

Table 4.4: Ablation study on HAM10000 dataset for selecting the best backbone network

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
VGG16	0.869	0.862	0.869	0.865
ResNet34	0.862	0.858	0.862	0.860
ResNet50	0.872	0.877	0.868	0.874
MobileNetV2	0.858	0.852	0.858	0.849
InceptionResNetV2	0.892	0.888	0.892	0.890
DenseNet121	0.877	0.871	0.868	0.870
DenseNet201	**0.894**	**0.900**	**0.897**	**0.900**

These findings have prompted an exploration of additional architectural configurations using DenseNet201 as the backbone on the HAM10000 dataset. The

different configurations are as follows:
(i) DenseNet201 + CBAM
(ii) DenseNet201 + SAM
(iii) DenseNet201 + CBAM + edge guidance (EDB-Net)
(iv) DenseNet201 + SAM + edge guidance

Table 4.5: Ablation results for selecting the best model on HAM1000 dataset

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
DenseNet201	0.894	0.900	0.897	0.900
(i)	0.913	0.910	0.913	0.911
(ii)	0.908	0.910	0.908	0.909
(iii)	**0.927**	**0.924**	**0.927**	**0.926**
(iv)	0.920	0.919	0.920	0.919

As shown in Table 4.5, integrating edge information enhances accuracy by at least 1.2% compared to using only an attention mechanism. During testing, two types of attention mechanisms were utilized: CBAM and SAM. The proposed EDB-Net demonstrates an improvement over the baseline network supported solely by an attention module. Table 4.6 presents the classification results of this method across all three datasets considered in this thesis. Figure 4.3 depicts the confusion matrix of EDB-Net on the HAM10000 dataset. Despite the absence of data augmentation, EDB-Net demonstrates a high level of accuracy in identifying the different classes. This highlights the efficacy of the proposed method in effectively distinguishing between the various categories within the dataset, showcasing its robustness and potential for accurate skin cancer classification.

Table 4.6: Classification results achieved by EDB-Net across three datasets

Dataset	**Accuracy**	**Precision**	**Recall**	**F1-Score**
HAM10000	0.927	0.924	0.927	0.926
ISIC 2016	0.848	0.850	0.848	0.848
ISIC 2017	0.786	0.788	0.786	0.787

4.6 DSCC-Net

This section outlines the training setup and empirical findings derived from the training and testing phases of DSCC-Net proposed in this study.

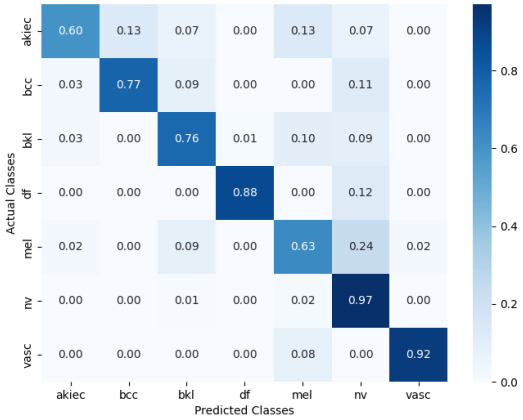


Figure 4.3: Confusion matrix of EDB-Net on HAM10000 dataset

4.6.1 Training Setup

Each of the attention-aided CNNs utilized by DSCC-Net is retrained on the train set for 100 epochs by fine-tuning all their layers. For training all base learners, the Adam optimizer has been used with an epsilon value of 0.1. The remaining hyper-parameters are mentioned in Table 4.7.

Table 4.7: Hyper-parameters for base learners utilized by DSCC-Net

Model	**Image Size**	**Learning Rate**	**Early Stop**
ResNet50	224×224	0.0001	40
InceptionResNetV2	299×299	0.01	30
DenseNet201	224×224	0.01	30

4.6.2 Experimental Results

Table 4.8 depicts the performance of the three individual attention-aided CNN models and DSCC-Net in terms of accuracy, precision, recall, and F1-score on the HAM10000 dataset. DSCC-Net achieves an overall accuracy of 0.932, which is an improvement over the best performing base learner. Table 4.9 reports the classification results of this method across all three datasets considered for this thesis.

Classifying skin cancer accurately can be very complex and heavily imbalanced datasets like HAM10000 comes with its own set of challenges. It is evident from Figure 4.4a, Figure 4.4b and Figure 4.4c that individual soft attention-based CNN models extract distinct features from the images and are diverse in nature while classifying the seven different classes. In an attempt to leverage the best out of these models, these networks are combined using DS theory. This diversity allows DSCC-Net to learn differently from the base learners. From Figure 4.4d, it can be

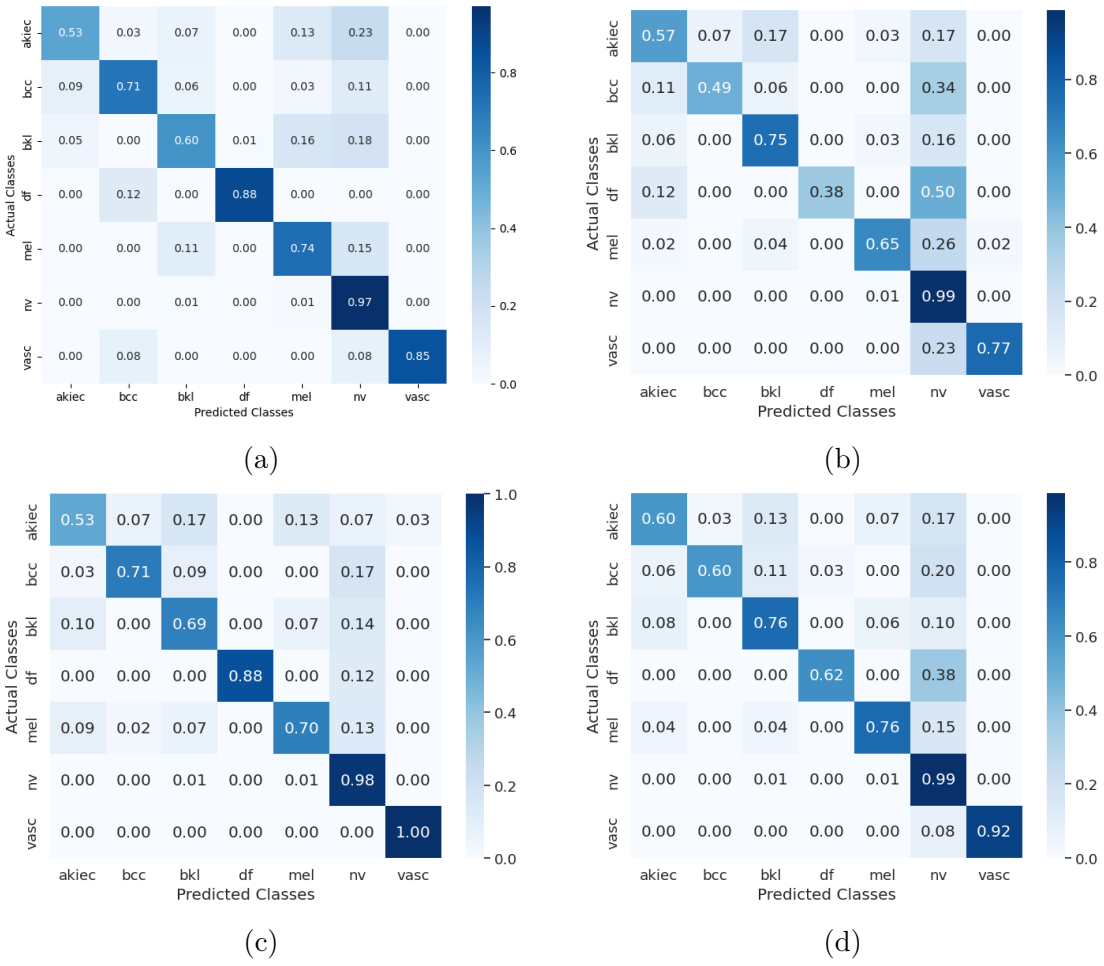


Figure 4.4: Confusion matrix of (a) ResNet50; (b) InceptionResNetV2; (c) DenseNet201; (d) DSCC-Net on HAM10000 dataset

Table 4.8: Performance comparison among base learners and DSCC-Net on HAM10000 dataset

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
ResNet50	0.910	0.913	0.910	0.910
InceptionResNetV2	0.919	0.917	0.919	0.910
DenseNet201	0.922	0.923	0.923	0.922
DSCC-Net	**0.932**	**0.931**	**0.932**	**0.932**

Table 4.9: Classification results achieved by DSCC-Net across three datasets

Dataset	**Accuracy**	**Precision**	**Recall**	**F1-Score**
HAM10000	0.932	0.931	0.932	0.932
ISIC 2016	0.868	0.862	0.866	0.864
ISIC 2017	0.792	0.791	0.792	0.791

concluded that the suggested approach is capable of performing the classification task with greater consistency as it is able to identify all seven classes with reasonably high accuracy (high sensitivity). In all the above experiments, the best performing individual model on HAM10000 in terms of accuracy is DenseNet201 which could identify melanoma (the most deadly type) with an accuracy of only 70%. However, DSCC-Net achieves a 76% accuracy for the melanoma class, which is a significant improvement. The main advantage of this combination procedure is that it can be implemented without focusing much on intricate details about feature extraction and classification. The nodes in the last layer of the individual soft attentive CNN models contain a final score for each of the seven output classes and only these scores need to be considered while applying the DS theory based combination technique. Therefore, this approach can be extended to any kind of classifier combination problem, not just limited to the skin cancer domain. Also, the proposed model does not require any further training or tuning since it only takes the confidence scores into account for combination. Therefore, it does not involve any extra cost or effort to achieve an optimal solution.

4.7 Model Interpretability using Grad-CAM

This section employs Grad-CAM, a visualization technique, to provide insightful visual explanations of model predictions. Grad-CAM generates gradient-weighted class activation maps, highlighting the regions in input images that significantly influence the model's decision-making process. These visualizations offer valuable insights into how neural networks arrive at their predictions, enhancing

transparency and interpretability in AI systems.

Figures 4.5a, 4.5b, and 4.5c depict the Grad-CAM results for the three proposed models across a few complicated instances from two datasets. The Grad-CAM visualization reveals the model's focus on skin lesions, offering insights into its decision-making. The original image displays typical lesion characteristics like uneven boundaries and varied pigmentation. Grad-CAM overlays a heatmap, ranging from blue to red and yellow, indicating region importance for the model. Red areas denote maximal activation, often highlighting central, darker lesion regions with irregular borders or color aberrations. Yellow and orange areas show lesser activation but still contribute to the decision process. Green regions represent moderate activation, providing supplementary context, while blue areas indicate minimal activation, typically found in background regions or featureless skin parts. Notably, Figure 4.5b showcases EDB-Net's capability to delineate even intricate lesion borders, affirming the efficacy of the edge-guided dual-branched CNN architecture.

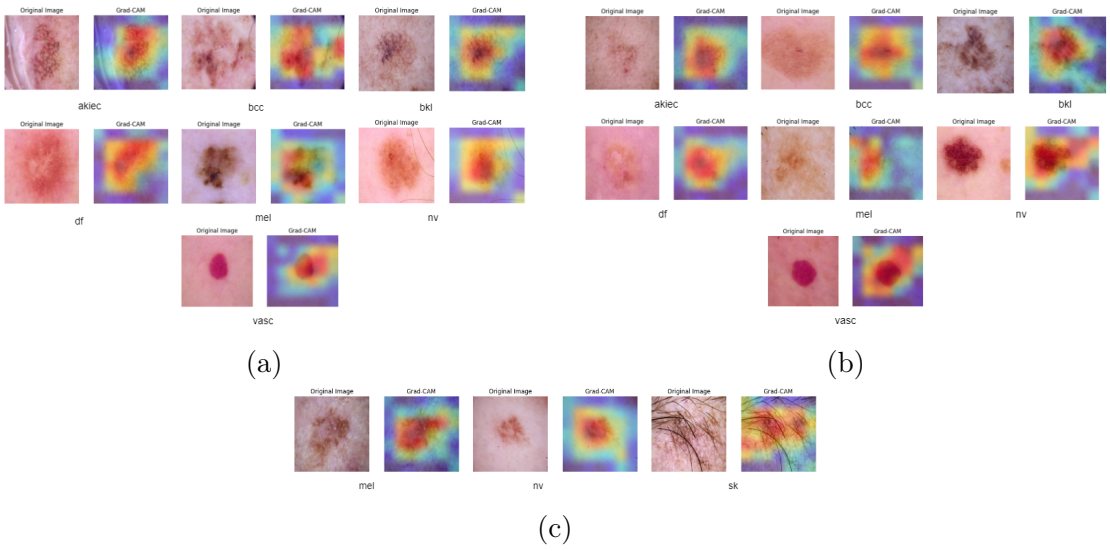


Figure 4.5: Grad-CAM results on (a) HAM10000 dataset using DSCC-Net; (b) HAM10000 dataset using EDB-Net; (c) ISIC 2017 dataset using DFF-Net

4.8 Comparison with Recent Past Methods

In this section, the three proposed models are compared with several existing methods for skin cancer classification across three datasets. As shown in Table 4.10, DSCC-Net outperforms Deepskin [66], one of the leading existing methods, in terms of both F1-score and accuracy on the HAM10000 dataset. Additonally, Table 4.11, highlights that while the model proposed by Gajera et al. [67], achieves the highest accuracy, its relatively low sensitivity score is a notable

flaw. This poses a significant limitation, as misclassifying an individual with cancer as not having the condition carries a higher risk than the opposite scenario. In contrast, DSCC-Net offers a balanced performance in terms of both accuracy and sensitivity. Moreover, Table 4.12 indicates that DFF-Net demonstrates superior efficiency compared to Dermo-DOCTOR [43], one of the best existing methods, where Hasan et al. used dual encoders to first segment the region of interest before conducting skin lesion classification.

The training and loss curves for the best performing method for each dataset are shown in Figure 4.6, Figure 4.7, and Figure 4.8, respectively. The corresponding ROC curves for each dataset are illustrated in Figure 4.9. Bar chart 4.9d illustrates a comparative analysis of the accuracy of the three proposed models — DFF-Net, EDB-Net, and DSCC-Net — across three datasets: HAM10000, ISIC 2016, and ISIC 2017.

Table 4.10: Performance comparison of the proposed models with some recent methods on HAM10000 dataset

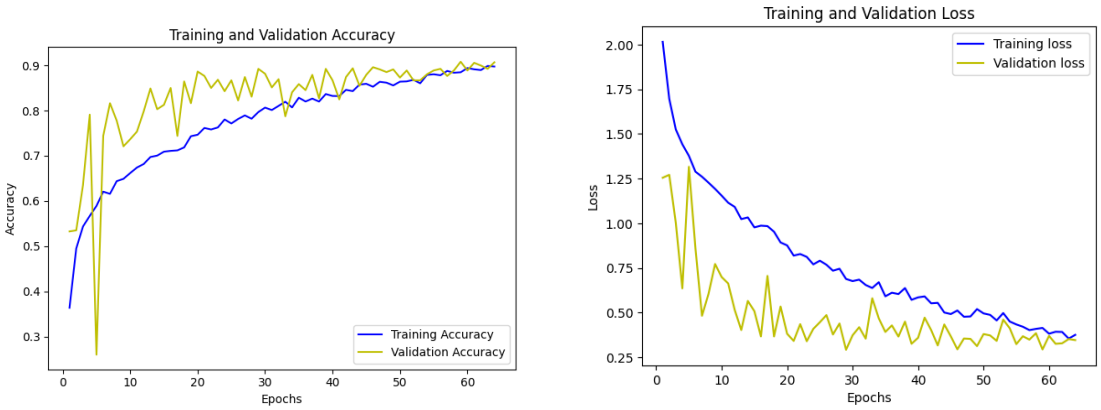
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Aldwgeri et al. [68], 2019	0.801	-	-	-
Gessert et al. [69], 2020	0.926	-	-	-
Yao et al. [70], 2021	0.864	-	-	-
Iqbal et al. [71], 2021	0.888	0.905	0.888	0.891
Datta et al. [32], 2021	0.916	0.915	0.916	0.914
Shen et al. [29], 2022	0.853	-	-	-
Sai Charan et al. [72], 2022	0.886	-	-	-
Mridha et al. [19], 2023	0.820	-	-	-
Gururaj et al. [66], 2023	0.912	-	-	0.917
Roy et al. [73], 2023	-	0.723	0.706	0.702
DFF-Net	0.922	0.923	0.923	0.922
EDB-Net	0.927	0.924	0.927	0.926
DSCC-Net	**0.932**	**0.931**	**0.932**	**0.932**

Table 4.11: Performance comparison of the proposed models with some recent methods on ISIC 2016 dataset

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Yu et al. [42], 2016	0.855	-	0.507	-
Saba et al. [74], 2019	0.786	-	0.667	-
Kalwa et al. [75], 2019	0.813	-	0.573	-
Gajera et al. [67], 2022	**0.871**	-	0.480	-
Gajera et al. [46], 2023	0.804	-	0.560	-
Sahoo et al. [76], 2024	0.781	-	0.780	-
DFF-Net	0.856	0.854	0.856	0.855
EDB-Net	0.848	0.850	0.848	0.848
DSCC-Net	0.868	**0.862**	**0.866**	**0.864**

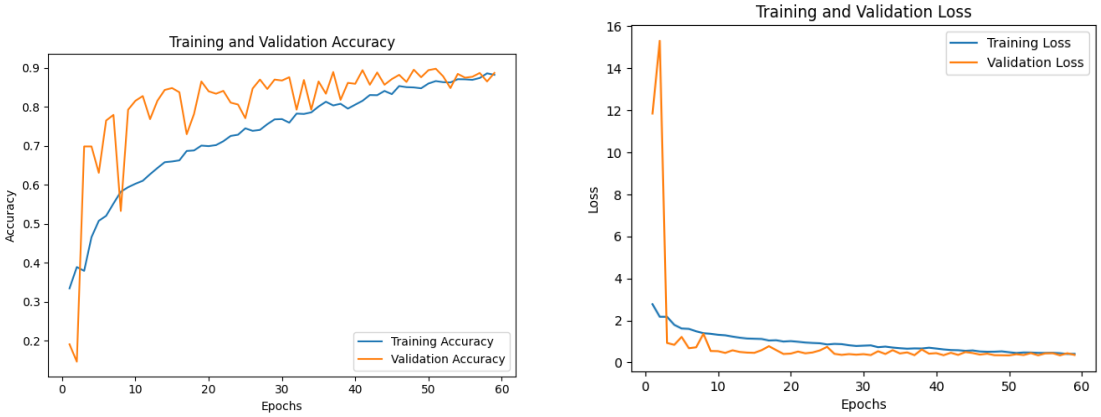
Table 4.12: Performance comparison of the proposed models with some recent methods on ISIC 2017 dataset

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Serte et al. [77], 2019	-	-	0.750	-
Zhang et al. [78], 2019	-	-	0.770	-
Al-Masni et al. [79], 2020	-	-	0.760	-
Yu et al. [80], 2020	-	0.720	-	-
Hasan et al. [43], 2021	0.780	0.790	0.780	-
DFF-Net	**0.795**	**0.791**	**0.795**	**0.792**
EDB-Net	0.786	0.788	0.786	0.787
DSCC-Net	0.792	0.791	0.792	0.791



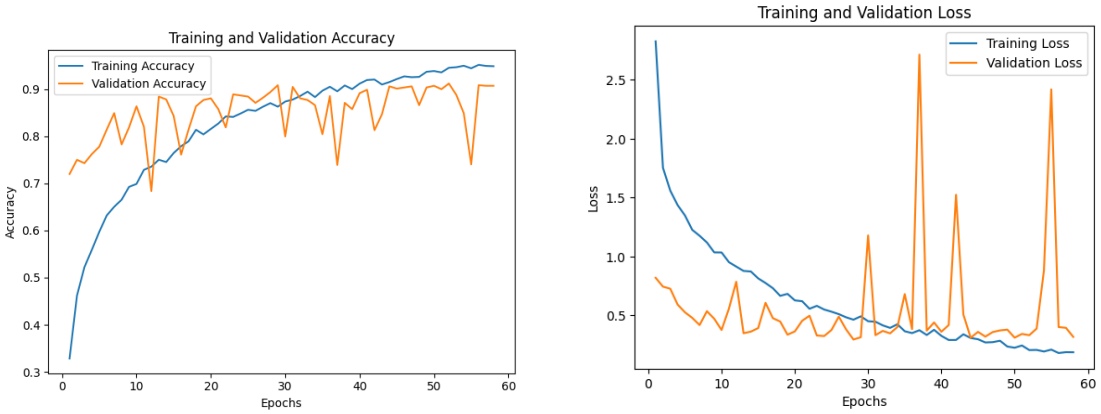
(a) (b)

Figure 4.6: (a) Training curve; (b) Loss curve for HAM10000 dataset



(a) (b)

Figure 4.7: (a) Training curve; (b) Loss curve for ISIC 2016 dataset



(a) (b)

Figure 4.8: (a) Training curve; (b) Loss curve for ISIC 2017 dataset

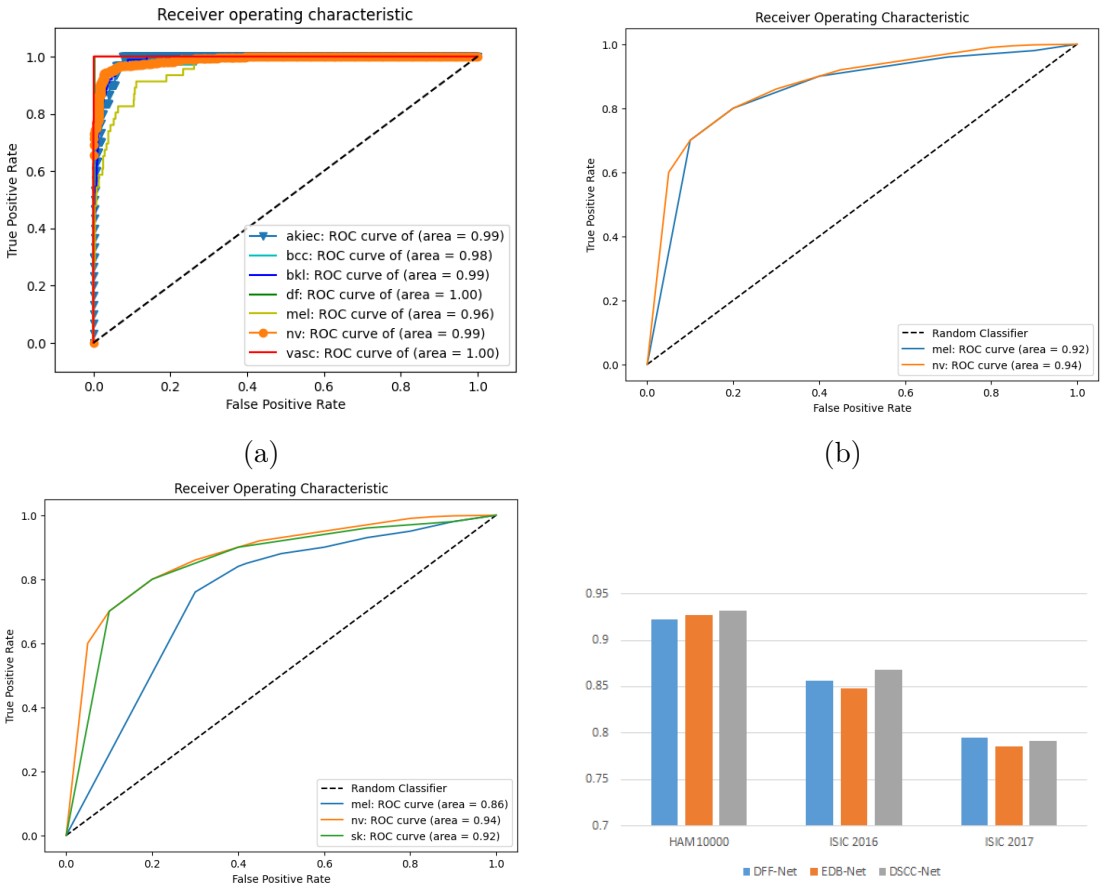


Figure 4.9: ROC curves for (a) HAM1000 dataset; (b) ISIC 2016 dataset; (c) ISIC 2017 dataset; (d) Accuracy comparison across proposed models and datasets

CHAPTER 5

Conclusions

This thesis introduces novel DL approaches specifically designed for skin cancer classification across various datasets. Through comprehensive experimentation and validation, these methodologies demonstrate exceptional effectiveness in accurately identifying and classifying different types of skin cancer, thus making a significant contribution to the field of medical imaging.

By tackling the challenges posed by inter-class similarity, intra-class variability of skin lesions, and dataset imbalances, these novel methods lay the groundwork for improved classification performance. The results obtained surpass the SoTA for three challenging skin cancer datasets, HAM10000, ISIC 2016, and ISIC 2017, underscoring the superiority of the proposed methods. The integration of these DL techniques has the potential to revolutionize medical image analysis, equipping clinicians with invaluable tools to combat skin cancer and improve healthcare outcomes on a global scale. As the field progresses, bridging the gap between AI-driven systems and clinical practice will be pivotal, ultimately aiming to improve diagnostic results, effective treatment planning, and overall patient outcomes in dermatology.

The first method, DFF-Net, revolves around employing a fusion of deep features aided by squeeze-and-excitation attention to address the classification task. Initially, all images are cropped using a contour-based approach. Subsequently, features are extracted from these cropped images using the Xception and BiT-M models. These fused features are then subjected to a squeeze-and-excitation attention module, followed by additional pooling and fully connected layers to perform the final classification. Furthermore, a combined loss function is introduced to enhance performance. On the challenging HAM10000, ISIC 2016, and ISIC 2017 datasets, this model achieves accuracies of 0.922, 0.856, and 0.795, respectively.

The second method, EDB-Net, employs a novel edge-guided dual-branched

CNN to perform the classification task. One branch processes edge maps from dermoscopic images, while the other focuses on the original dermoscopic images. Both branches utilize a DenseNet-based backbone, which includes dense and transition blocks to extract features from their respective inputs. The second branch also incorporates an additional attention module to emphasize salient regions within the dermoscopic images, thereby capturing more relevant features. Furthermore, a novel strategy is introduced to modulate the attention maps using previously extracted edge features, creating edge-guided features that refine the feature representations. In the final stage, both attended and edge-guided features are aggregated, flattened, and fed into dense layers for classification. This approach generates more discriminative and contextually relevant features, significantly enhancing classification performance. This model achieves accuracies of 0.927, 0.848, and 0.786 on HAM10000, ISIC 2016, and ISIC 2017 datasets, respectively. Notably, this method does not utilize data augmentation, yet it performs with great efficiency, underscoring its capabilities.

The third method, DSCC-Net, employs a classifier combination approach instead of relying on a single classifier. This approach initially utilizes three base learners — ResNet50, InceptionResNetV2, and DenseNet201 — each enhanced with a soft attention module. The outcomes of these classifiers are then combined using the popular DS theory. The method has been evaluated on three datasets. On the class-imbalanced HAM10000 dataset, this model achieves an impressive accuracy of 0.932. On the ISIC 2016 and 2017 datasets, it achieves accuracies of 0.868 and 0.792, respectively.

5.1 Limitations

While the proposed methods offer innovative approaches to skin cancer classification, it is important to acknowledge some disadvantages to maintain an unbiased perspective.

- **DFF-Net:** The model encounters difficulty in distinguishing melanoma images, often misclassifying many of them as nevus due to the minimal inter-class disparity between these two categories. Additionally, combining features from different CNN models can sometimes introduce redundant information or conflicting features, potentially hindering the model's overall performance. Moreover, the inherent complexity of handling multiple CNN models increases the computational burden and may lead to overfitting if not managed properly.

- **EDB-Net:** Although this model presents a novel approach, its efficiency

heavily relies on the quality of the generated edge maps, which significantly impacts the overall performance of the system. Additionally, finding optimal hyperparameters for the attention module and edge feature modulation strategy poses challenges, necessitating extensive experimentation. There are also potential limitations in generalizing this method to other domains or tasks, as it is specifically tailored for dermoscopic image classification.

- **DSCC-Net:** The model faces challenges in classifying categories with fewer samples. As depicted in Figure 4.4d, the model demonstrates difficulty in accurately classifying images of the akiec and bcc classes. Additionally, the DS theory based combination is computationally intensive; increasing the number of classifiers leads to a corresponding rise in hypotheses and their associated BPAs. Moreover, the method remains static throughout its application. If the feature representation is biased towards any particular class, the method's fixed configuration may struggle to adapt to this bias effectively.

5.2 Scope for Future Work

In the medical field, the application of DL models to real-life scenarios demands perfection, leaving no room for error. Hence, it is imperative to continue refining and improving the performance of these models. This section outlines several potential avenues for future research aimed at enhancing the accuracy and efficiency of skin cancer classification models.

- **Improving Minority Class Accuracy:** Future investigations may delve into specific loss functions tailored to address class imbalances. Such functions could ensure that less represented classes receive sufficient attention during training, potentially ameliorating their classification performance.

- **Synthetic Data Generation:** In addition to conventional data augmentation techniques, the exploration of GANs for synthetic data generation holds promise. By augmenting the training dataset with diverse and realistic samples, this approach may bolster model generalization and robustness.

- **Focus on Region of Interest:** Methodologies spotlighting image ROIs could be investigated. By honing in on relevant features, such techniques may facilitate more precise and accurate classification.

- **Feature Selection:** Post feature fusion, the application of feature selection methods holds potential for identifying and leveraging the most pertinent

features. This endeavor can streamline model performance by eliminating extraneous information.

- **Resource Efficient Models:** Given the computational demands associated with ensemble approaches, the adoption of lightweight CNN models merits exploration. These models can optimize resource utilization without compromising performance, thereby enhancing system efficiency and scalability.

By addressing these future research directions, the aim is to develop skin cancer classification models that not only achieve higher accuracy and efficiency but also demonstrate greater robustness and reliability in clinical applications. Continuous improvement in these areas will be crucial for advancing the field of medical image analysis and enhancing patient care on a global scale.

List of Publications

Accepted for Publication

1. S. Sarkar*, **A. Ray***, D. Kaplun, R. Sarkar, "A Combination of Soft Attention aided CNN Models using Dempster-Shafer Theory for Skin Cancer Classification", International Conference on Current Problems of Applied Mathematics and Computer Systems, CPAMCS-2023, Lecture Notes in Networks and Systems, Springer Nature.

2. D. Mandal*, **A. Ray***, S. Sarkar*, O. Ojjela, R. Sarkar, "A Robust Deep Feature Fusion Model for Skin Cancer Classification", International Conference on Advanced Computing and Applications, ICACA-2024, Lecture Notes in Networks and Systems, Springer Nature.

Communicated (under review)

1. **A. Ray***, S. Sarkar*, R. Sarkar, "Decoding Skin Cancer Classification: Perspectives, Insights, and Advances through Researchers' Lens", Archives of Computational Methods in Engineering, Springer Nature.

Drafting Stage

1. **A. Ray**, S. Gayen, R. Sarkar, "EDB-Net: A Novel Edge-guided Dual-Branched Network for Skin Cancer Classification"

Bibliography

[1] C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1096–1109, 2018.

[2] B. K. Armstrong and A. Kricker, "Skin cancer," *Dermatologic Clinics*, vol. 13, no. 3, pp. 583–594, 1995.

[3] M. Elgamal, "Automatic skin cancer images classification," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 3, 2013.

[4] V. Madan, J. T. Lear, and R.-M. Szeimies, "Non-melanoma skin cancer," *The lancet*, vol. 375, no. 9715, pp. 673–685, 2010.

[5] M. A. Linares, A. Zakaria, and P. Nizran, "Skin cancer," *Primary care: Clinics in office practice*, vol. 42, no. 4, pp. 645–659, 2015.

[6] M. Q. Khan, A. Hussain, S. U. Rehman, *et al.*, "Classification of melanoma and nevus in digital images for diagnosis of skin cancer," *IEEE Access*, vol. 7, pp. 90 132–90 144, 2019.

[7] M. Dildar, S. Akram, M. Irfan, *et al.*, "Skin cancer detection: A review using deep learning techniques," *International journal of environmental research and public health*, vol. 18, no. 10, p. 5479, 2021.

[8] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: a cancer journal for clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[9] R. Skaggs and B. Coldiron, "Skin biopsy and skin cancer treatment use in the medicare population, 1993 to 2016," *Journal of the American Academy of Dermatology*, vol. 84, no. 1, pp. 53–59, 2021.

[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[11] A. Kolesnikov, L. Beyer, X. Zhai, *et al.*, "Big transfer (bit): General visual representation learning," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, Springer, 2020, pp. 491–507.

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[13] E. Gordon-Rodriguez, G. Loaiza-Ganem, G. Pleiss, and J. P. Cunningham, "Uses and abuses of the cross-entropy loss: Case studies in modern deep learning," 2020.

[14] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3*, Springer, 2015, pp. 84–92.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[16] G. Shafer, "Dempster's rule of combination," *International Journal of Approximate Reasoning*, vol. 79, pp. 26–40, 2016.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[19] K. Mridha, M. M. Uddin, J. Shin, S. Khadka, and M. Mridha, "An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system," *IEEE Access*, 2023.

[20] H. Rastegar and D. Giveki, "Designing a new deep convolutional neural network for skin lesion recognition," *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18 907–18 923, 2023.

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

[22] S. Kalouche, A. Ng, and J. Duchi, "Vision-based classification of skin cancer using deep learning," *2015, conducted on Stanfords Machine Learning course (CS 229) taught*, 2016.

[23] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[24] T. Emara, H. M. Afify, F. H. Ismail, and A. E. Hassanien, "A modified inception-v4 for imbalanced skin cancer classification dataset," in *2019 14th International Conference on Computer Engineering and Systems (ICCES)*, IEEE, 2019, pp. 28–33.

[25] 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.

[26] A. Sagar and J. Dheeba, "Convolutional neural networks for classifying melanoma images," *bioRxiv*, pp. 2020–05, 2020.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[28] A. G. Howard, M. Zhu, B. Chen, *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[29] S. Shen, M. Xu, F. Zhang, *et al.*, "A low-cost high-performance data augmentation for deep learning-based skin lesion classification," *BME Frontiers*, 2022.

[30] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.

[31] S. Ding, Z. Wu, Y. Zheng, *et al.*, "Deep attention branch networks for skin lesion classification," *Computer methods and programs in biomedicine*, vol. 212, p. 106 447, 2021.

[32] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*, Springer, 2021, pp. 13–23.

[33] B. Harangi, A. Baran, and A. Hajdu, "Classification of skin lesions using an ensemble of deep neural networks," in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, 2018, pp. 2575–2578.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[35] C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[36] L. Ichim, R. I. Mitrică, and D. Popescu, "Detection of melanomas using ensembles of deep convolutional neural networks," in *2023 13th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, IEEE, 2023, pp. 1–6.

[37] M. K. Hasan, M. T. E. Elahi, M. A. Alam, M. T. Jawad, and R. Martí, "Dermoexpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation," *Informatics in Medicine Unlocked*, vol. 28, p. 100 819, 2022.

[38] H. Rashid, M. A. Tanveer, and H. A. Khan, "Skin lesion classification using gan based data augmentation," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 916–919.

[39] A. Cheng, J. Frank, and I. Gallegos, "Conditional gan for dermatology image generation,"

[40] S. Aladhadh, M. Alsanea, M. Aloraini, T. Khan, S. Habib, and M. Islam, "An effective skin cancer classification mechanism via medical vision transformer," *Sensors*, vol. 22, no. 11, p. 4008, 2022.

[41] G. Yang, S. Luo, and P. Greer, "A novel vision transformer model for skin cancer classification," *Neural Processing Letters*, pp. 1–17, 2023.

[42] 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.

[43] M. K. Hasan, S. Roy, C. Mondal, *et al.*, "Dermo-doctor: A framework for concurrent skin lesion detection and recognition using a deep convolutional neural network with end-to-end dual encoders," *Biomedical Signal Processing and Control*, vol. 68, p. 102 661, 2021.

[44] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[45] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.

[46] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "A comprehensive analysis of dermoscopy images for melanoma detection via deep cnn features," *Biomedical Signal Processing and Control*, vol. 79, p. 104 186, 2023.

[47] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with svm classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, p. 100 069, 2023.

[48] J. V. Tembhurne, N. Hebbar, H. Y. Patil, and T. Diwan, "Skin cancer detection using ensemble of machine learning and deep learning techniques," *Multimedia Tools and Applications*, pp. 1–24, 2023.

[49] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[50] M. Pietikäinen, "Local binary patterns," *Scholarpedia*, vol. 5, no. 3, p. 9775, 2010.

[51] J. Yap, W. Yolland, and P. Tschandl, "Multimodal skin lesion classification using deep learning," *Experimental dermatology*, vol. 27, no. 11, pp. 1261–1267, 2018.

[52] A. N. Omeroglu, H. M. Mohammed, E. A. Oral, and S. Aydin, "A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105 897, 2023.

[53] J. SM, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 15 763–15 778, 2023.

[54] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific data*, vol. 8, no. 1, p. 34, 2021.

[55] J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, and W. Zeng, "Light gradient boosting machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data," *Agricultural water management*, vol. 225, p. 105 758, 2019.

[56] L. Wright and N. Demeure, "Ranger21: A synergistic deep learning optimizer," *arXiv preprint arXiv:2106.13731*, 2021.

[57] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[58] 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.

[59] *The international skin imaging collaboration archive*, `https://www.isic-archive.com/`.

[60] D. Gutman, N. C. Codella, E. Celebi, *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1605.01397*, 2016.

[61] N. C. Codella, D. Gutman, M. E. Celebi, *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 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, 2018, pp. 168–172.

[62] S. A. Hicks, I. Strümke, V. Thambawita, *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Scientific reports*, vol. 12, no. 1, p. 5979, 2022.

[63] R. Kumar and A. Indrayan, "Receiver operating characteristic (roc) curve for medical researchers," *Indian pediatrics*, vol. 48, pp. 277–287, 2011.

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

[65] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[66] H. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya, and F. Flammini, "Deepskin: A deep learning approach for skin cancer classification," *IEEE Access*, 2023.

[67] H. K. Gajera, D. R. Nayak, and M. A. Zaveri, "Fusion of local and global feature representation with sparse autoencoder for improved melanoma classification," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2022, pp. 5051–5054.

[68] A. Aldwgeri and N. F. Abubacker, "Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images," in *Advances in Visual Informatics: 6th International Visual Informatics Conference, IVIC 2019, Bangi, Malaysia, November 19–21, 2019, Proceedings 6*, Springer, 2019, pp. 214–226.

[69] 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, p. 100 864, 2020.

[70] P. Yao, S. Shen, M. Xu, *et al.*, "Single model deep learning on imbalanced small datasets for skin lesion classification," *IEEE transactions on medical imaging*, vol. 41, no. 5, pp. 1242–1254, 2021.

[71] 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, p. 101 843, 2021.

[72] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to classify skin lesions using dermoscopic images," *arXiv preprint arXiv:2008.09418*, 2020.

[73] D. Roy, R. Pramanik, and R. Sarkar, "Margin-aware adaptive-weighted-loss for deep learning based imbalanced data classification," *IEEE Transactions on Artificial Intelligence*, 2023.

[74] T. Saba, M. A. Khan, A. Rehman, and S. L. Marie-Sainte, "Region extraction and classification of skin cancer: A heterogeneous framework of deep cnn features fusion and reduction," *Journal of medical systems*, vol. 43, no. 9, p. 289, 2019.

[75] U. Kalwa, C. Legner, T. Kong, and S. Pandey, "Skin cancer diagnostics with an all-inclusive smartphone application," *Symmetry*, vol. 11, no. 6, p. 790, 2019.

[76] S. R. Sahoo, R. Dash, and R. K. Mohapatra, "Fusion of deep and wavelet feature representation for improved melanoma classification," *Multimedia Tools and Applications*, pp. 1–27, 2024.

[77] S. Serte and H. Demirel, "Gabor wavelet-based deep learning for skin lesion classification," *Computers in biology and medicine*, vol. 113, p. 103 423, 2019.

[78] 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.

[79] 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, p. 105 351, 2020.

[80] Z. Yu, F. Jiang, F. Zhou, *et al.*, "Convolutional descriptors aggregation via cross-net for skin lesion recognition," *Applied Soft Computing*, vol. 92, p. 106 281, 2020.