Dermoscopic Structure based Skin Cancer Segmentation and Classification using Deep Learning Techniques

By

Sannia Arshad

137-FBAS/PHDCS/F15

Supervisor

Dr. Tehmina Amjad

Assistant Professor

IIUI Islamabad

Co-Supervisor

Dr. Ayyaz Hussain

Professor

QAU Islamabad

Department of Computer Science

Faculty of Computing and Information Technology

International Islamic University, Islamabad

2024

Computer software
Computer programs
Skin cancer

INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD
FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

Date: 9 Aug, 2024

Final Approval

It is certified that we have read this thesis, entitled "Dermoscopic Structure based Skin Cancer Segmentation and Classification using Deep Learning Techniques" submitted by Ms Sannia Arshad, Registration No. 137-FBAS/PHDCS/F15. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for the award of the degree of PhD in Computer Science.

Committee

External Examiner:
Dr. Abdul Basit,
Associate Professor,
Department of Computer Science,
CUST University, Islamabad

External Examiner:
Dr. Hassan Mujtaba,
Professor,
FAST University, Islamabad

Internal Examiner:
Dr. Muhammad Nadeem,
Chairperson, Assistant Professor,
Department of Computer Science,
FCIT, IIUI

Supervisor:
Dr. Tehmina Amjad,
Ex-Assistant Professor,
Department of Computer Science,
FCIT, IIUI

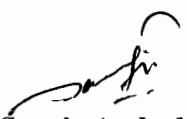
Co-Supervisor:
Dr. Ayyaz Hussain,
Professor,
Department of Computer Science,
QAU, Islamabad

Dissertation

This dissertation is submitted to
Department of Computer Science
Faculty of Computing and Information Technology
International Islamic University, Islamabad
As a Partial Fulfilment of the Requirement for the
Degree of Doctor of Philosophy (Computer
Science)

Declaration

I hereby declare that this thesis **"Dermoscopic Structure based Skin Cancer Segmentation and Classification using Deep Learning Techniques"** neither as a whole nor as a part has been copied out from any source. It is further declared that I have done this research with the accompanied report entirely based on my personal efforts, under the proficient guidance of my supervisors Dr. Tehmina Amjad and Dr. Ayyaz Hussain. I also declare that the work presented in this report has not been submitted in support of any other application or degree or qualification in any other University or Institute.



Sannia Arshad
(137-FBAS/PHDCS/F15)

Acknowledgement

In the name of Allah, the most Merciful and the most Gracious, the Lord of the worlds, and prayers and peace is upon Muhammad (S.A.W) His servant and messenger.

Alhamdulillah, all praise to Allah for giving me strength and the patience to complete my dissertation finally, after all difficulties and challenges. Words are inadequate to describe my gratitude and appreciation to Him in the whole process of completing this research.

This dissertation would not have been completed without the guidance, help, support, and efforts of many people.

I would like to express my deep gratitude to my supervisor Dr. Tehmina Amjad for his constant help and guidance throughout my PhD research work. I would like to extend my gratitude to my Co-supervisor Dr. Ayyaz Hussain for her help and advice regarding the process of writing this dissertation.

My deepest gratitude goes to my parents for their support and encouragement. They have been such a source of strength and encouragement throughout my PhD journey. Today whatever I have achieved in my life, all the credit goes to my parents.

I owe special thanks to my all friends for the support, encouragement, patience and for being there for me when I needed it.

Last but not least, my special thanks from the depth of my heart goes to my beloved husband Dr. Zain Ul Abideen, who was always there for me, and supported me throughout my thesis work. Without him I was not able to complete my work. He has always promoted me in every field of my life and for academic excellence.

Sannia Arshad

(137-FBAS/PHDCS/F15)

Dedicated to

My parents and my husband

Abstract

Skin is considered as one of the prime structures in human body. It helps to defend our body from the harmful rays of sun. From the last few years, there is a great rise in skin infections around the world. This is due to the change in global warming and human behaviour towards environment. Among these infections and diseases, skin cancer is one of the most spreading diseases; however, it can be cured by detecting it earlier. There are various kinds of skin cancer for example, melanoma, Basal Cell Carcinoma (BCC), Seborrheic Keratosis (SK), and more. From the last few decades, a lot of research has been made in the development and improvement of the automated systems for the recognition and segmentation of the lesions of skin. There are further various types and sub-types of skin cancer such as melanocytic, malignant, and non-malignant that make a hierarchy in a sequence. Introducing this hierarchy in our algorithm can help to develop an explainable system that can assist a practitioner in a much better way. There are also different dermoscopic structures or patterns present in all these types of skin cancer lesions including pigment network, globules, and negative network etc. These structures can help to identify and differentiate among various skin cancer types and their diagnosis. In this thesis, we have proposed a hierarchical model by introducing some variations in the existing GoogleNet model, named H-GoogleNet model. A three levels hierarchy of skin lesions is embedded within a single model such as in H-GoogleNet so that it will predict and generate a hierarchy of skin lesion type and its sub-type till three levels. Level 1 shows the origin, level 2 depicts degree of malignancy, and level 3 represents the differential or fine class. Auxiliary classifiers and inception blocks are added in the GoogleNet model to make it work for taxonomic level classification. Various optimizers and loss functions are used in the proposed model to get more accurate results. Along hierarchical classification, we also proposed a ResNet50-UNet model for the structural segmentation of skin lesions within a single image. To handle the class imbalance problem, present in skin lesion classification datasets, we have applied data augmentation techniques. For data imbalance, present in structural segmentation dataset of skin lesion, for five attributes, we have proposed a hybrid loss function so that we can handle the imbalance segmentation data of skin lesion. International Skin Imaging Collaboration (ISIC) 2018 classification dataset of skin cancer is used to evaluate our

proposed hierarchical classification model. ISIC 2018 segmentation dataset is used for the evaluation of proposed structural segmentation model. The results show that the proposed H-GoogleNet model and attributes segmentation model generates a good hierarchical classification of skin lesion and attribute level segmentation respectively. The proposed models are also compared with the existing literature. The proposed H-GoogleNet and ResNet50-UNet models can help the dermatologists to get an explainable sequence of skin lesion origin, degree of malignancy, and fine class, and distinguish among various lesion structures present in skin lesions.

Research Contribution:

Sannia Arshad, Tehmina Amjad, Ayyaz Hussain, Imran Qureshi, and Qaisar Abbas. 2023. "Dermo-Seg: ResNet-UNet Architecture and Hybrid Loss Function for Detection of Differential Patterns to Diagnose Pigmented Skin Lesions" *Diagnostics* 13, no. 18: 2924. 2023 https://doi.org/10.3390/diagnostics13182924

Table of Contents

List of Figures

List of Tables

List of Abbreviations

WHO	World Health Organization
UV	Ultraviolet
DNA	Deoxyribose Nucleic Acid
ACS	American Cancer Society
US	United States
BCC	Basal Cell Carcinoma
SCC	Squamous Cell Carcinoma
SKMCH	Shaukat Khanum Memorial Cancer Hospital
CASH	Color, Architecture, Symmetry and Homogeneity
ABCD	Asymmetry, Border, Color, Diameter
CAD	Computer Aided Diagnosis
CT	Computed Tomography
PET	Positron Emission Tomography
MRI	Magnetic Resonance Imaging
KLR	Kernel Logistic Regression
RGB	Red, Green, Blue
ANN	Artificial Neural Network
SVM	Support Vector Machine
KNN	K Nearest Neighbors
DANAOS	Diagnosis and Neural Analysis of Skin cancer
FFA	Firefly Algorithm
FCM	Fuzzy C Mean
LBP	Local Binary Pattern
CNN	Convolutional Neural Network
H-VGG-19	Hierarchical Visual Geometry Group - 19
H-GoogleNet	Hierarchical Google Network
YOLO	You Only Look Once
GPU	Graphical Processing Unit
DBN	Deep Belief Network
VGG	Visual Geometry Group
ISBI	International Symposium on Biomedical Imaging
ISIC	International Skin Imaging Collaboration
HD-CNN	Hierarchical Deep Convolutional Neural Network
H-CNN	Hierarchical Convolutional Neural Network
LSTM	Long Short-Term Memory

ReLU	Rectified Linear Unit
FC	Fully Connected
RAM	Random Access Memory
ELU	Exponential Linear Unit
CONV	Convolutional
GANS	Generative Adversarial Network
WCE	Weighted Cross Entropy
ADAM	Adaptive Momentum
SGD	Stochastic Gradient Descent
CAM	Class Activation Map
NV	Nevus
AKIEC	Actinic Keratosis
BKL	Benign Keratosis
DF	Dermatofibroma
VASC	Vascular
BACC	Balanced Accuracy
ROC-AUC	Receiver Operating Curve – Area Under the Curve
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
ResNet	Residual Network
IoU	Intersection over Union
BN	Batch Normalization
MP	Maximum Pooling
BN-ACT	Batch Normalization Activation
FTL	Focal Loss Tversky
TI	Tversky Index
NA	Not Available

Chapter 1

Introduction

Chapter 1

1 Introduction

We all know that cancer, no matter what type of it and in which body part, is a disease of cells that are basic building blocks of the human body. These cells make up tissues and organs of our body. Naturally, the human body keeps on making new cells to help our body to grow, replace worn-out or dead cells, and heal injuries or any other infectious and non-infectious disease. All this happens in an orderly way, so that the new one cell replaces the older one, and this process lasts throughout our life. However, sometimes, this normal growth of cells can turn into abnormal growth. This may happen due to some external effect of our surroundings or sometime due to internal effects of our body. No matter what the cause, this abnormal growth of cells develops a specific cancer on that part of the body. These abnormal cells have a great potential to spread and may invade the nearby tissues, destroying normal cells and developing a cancer resultantly. This cancer can be in any organ of the human body such as, brain, lung, blood, breast, kidney and many more. Like all these body organs, cancer can also develop in the skin, the largest organ of our body, causing skin cancer.

In our routine life, when we talk about the human body, we usually circle around our significant body parts, like brain, lungs, kidney, liver, heart, eyes etc. and rarely do we do this with our skin and its diseases. Like any other organ in our body, skin also suffers from diseases including acne, infections, allergy, eczema and many more. Among all these diseases, the most dangerous and fatal disease is skin cancer. It is the one causing ultimate deaths worldwide and this death rate is increasing yearly with the passage of time [1]. It is one of the widespread cancers in USA and Europe. World Health Organization (WHO) assess that the cancer cases will get double in the next few decades [1]. From these cancer assessments, one of the most alarming is skin cancer.

1.1 Who is at Risks? / Risk Factors

Anyone can develop skin cancer, but the risk is higher in people with lack of skin pigmentation, fair and freckled skin if it burns easily, very short or intense exposure to UV radiations, playing too much outdoor games, family history, lots of irregular or uneven moles on their body and many other causes. People who mostly work outdoors are at higher risk of skin cancer due to harmful

sunrays and damaging emissions of some industrial materials. UV radiation can also be produced from artificial sources, such as solariums (sunbeds/tanned beds). People using solariums have great chances of skin cancer. These solariums are mostly used in European countries. It is one of the greatest causes of intense skin cancer spread.

Skin cancer mostly develops on skin bare to the sun including, lips, ear, neck, hand, and face. It appears rarely on parts of skin not revealed to the sunlight (see Figure 1.1). Skin cancer develops when mutations in the skin cell's DNA take place causing abnormal growth of cells.

Figure 1.1: Examples of Skin Cancer

1.2 Reduce Risk Factors

Skin cancer can be lowered by ensuring these precautions:

1) Limit exposure to sunlight.
2) Examine skin for irregular variations to locate skin cancer at earliest.
3) Wearing clothes that cover your arms and legs can be helpful to prevent skin cancer.
4) Applying a suitable sunscreen can also help to prevent skin cancer.
5) Regular self-examination of your skin could lower the chances of skin cancer.
6) People using sun lamps and internal tanning beds are at great risk. They should cut off these artificial things from their life.

1.3 Worldwide Surveys of Skin Cancer

Skin cancer is among the most spreading cancers today. According to the recent facts and figures of American Cancer Society (ACS) in an annual published report 2020 [2], around 100,350 new cases and 6,850 deaths due to Melanoma are reported in the United States (US) till 2020. This report also gives the probability (%) of developing invasive skin cancer during selected age intervals by gender, in the US, 2014-2016. Its complete detail can be seen in [2]. It is assessed that around 9,500 people are diagnosed with skin cancer every day in the U.S. [3]. The more deep detail can be seen in [3][4][5].

1.4 Skin Cancer in Pakistan

Skin cancer is also reported in Pakistan every year. There are some survey reports [6 – 8] giving facts and figures of melanoma and few other skin cancer types such as BCC, SCC etc. in Pakistan. According to Shaukat Khanum Memorial Cancer Hospital (SKMCH) annual registry report of 2018 [7], skin cancer was among the top ten cancers viewed in all age groups. A total of 267 cases of skin cancer were reported only in this hospital. Out of 267, total 163 cases of male, and 104 cases of female were reported. In all 267 cases, Basal Cell Carcinoma (BCC), Nodular BCC was the most frequent skin cancer reported. After that, Squamous Cell Carcinoma (SCC), Malignant Melanoma were on second and third number respectively. The deeper fact sheet for skin cancer cases in SKMCH Pakistan, in year 2018, can be seen in [7]. There is also a collective registry report of SKMCH from 1994 till 2018 [6] for all types of cancers. This report tells that 2,440 cases of Melanoma and other malignant neoplasm of skin were reported from year 1994-2018. Out of 2,440 cases, 960 female cases and 1,480 were male cases. According to the statistics of WHO of 2019, about skin cancer in Pakistan [8], 428 new cases, 313 death cases and 922 5-year prevalence cases were reported. The deeper details of all these reports can be read in [6 – 8].

1.5 Skin Cancer Diagnosis

Skin is the most important part of the human body. It performs very important functions to help protect our body from various infections, germs, harmful rays of sun, to regulate our body temperature etc. It has three most important layers: *Epidermis, dermis* and *subcutis* [9]. The upper layer is epidermis. It has a basal layer containing melanocytes in it. Melanocytes release a pigment, named melanin, giving color to the skin such as brown, tan etc. The deeper detail of skin structure can be seen in Chapter 2. The abnormal production of melanin pigment causes the development of

melanocytic malignant melanoma. This abnormal production of melanin may be due to the harmful UV sun rays. This malignant melanoma is the extreme form of skin cancer causing ultimate killing if not identified at earlier stages. There are also some other kinds of skin cancer including BCC and SCC that are malignant. There are some types of lesions that are benign, but they can turn to malignant if not treated seriously such as melanocytic nevus, dermatofibroma, vascular, seborrheic keratosis, actinic keratosis etc. Some skin cancer lesions look similar in appearance that even the experts may not be able to differentiate among them. Sometimes, it is not possible to differentiate between various types of skin cancer lesions due to the similarity of their structures with one another. Due to these reasons, efforts have been made in the last few decades to enhance the clinical analysis of skin cancer lesions. All these efforts include some imaging techniques-dermoscopy [10], and clinical diagnosis procedures including Pattern Analysis [11], ABCD rule [12][13][14], 7-Point Checklist [15][16], Menzies method [17], CASH algorithm [18] etc. to help diagnose the skin cancer earlier. However, there are also some computer aided methods [19][20][21][22][23][24][25] that are in use by dermatologists. These systems are clinically oriented Computer Aided Diagnostic (CAD) systems. These systems are fully automated. The deeper detail of these methods is discussed in Chapter 3.

1.6 Research Challenges & Motivation

Precise identification of skin cancer type includes various challenges. There is a great need for self-explainable diagnostic system that could provide explainable outputs to the physicians such as it should not only provide the fine classification of skin lesion to dermatologists but also talk about the root class or origin of a specific skin lesion. There is one more challenge in which one type of skin lesion may mimic with the other one resulting in confusion for the dermatologist to take final decision. The appearance of these lesions may be similar and sometimes they also have the same structure in them. These similarities make it difficult for dermatologists to diagnose the actual type of lesion and it may misdiagnose the skin lesion type. This misclassification may lead to the wrong treatment of skin lesion and may cause the loss of human life.

The structures or attributes present in skin lesion may overlap with other structures in lesion. This overlapping cause misdiagnosis of skin lesion. There is also a great need for structural segmentation of skin lesions. From these structures, very useful information can be found. These structures may include pigment networks, cobblestones, milia-like cysts, and many others.

There is also a great challenge of class imbalance problem in the available skin cancer datasets. These datasets do not have the same number of instances in each class of skin lesion.

There is one more challenge of noise present in skin lesion images. These noises may include clinical noise, acquisition noise, and hair noise.

1.7 Challenges in Skin Lesion Classification

There are various challenges in publicly available skin lesion datasets including class imbalance, variations in skin lesions, and structural similarities. In this section, we have discussed these challenges in detail.

1.7.1 Imbalanced Dataset

Most of the available skin lesion datasets [26] [27] have serious imbalance in the number of samples among various skin lesion classes. For example, one class contains large number of samples, and the other one contains limited number of samples. The training of a CNN model with such imbalanced dataset may results in poor performance on minority classes such as it will show biasness toward majority class. This imbalance number of samples, in various skin cancer datasets are shown in Table 1.1.

Table 1.1: Number of samples and class imbalance in public skin cancer datasets.

Classes	Datasets	
	ISIC 2018	ISIC 2019
Melanoma	1113	4522
Nevus	6705	12875
Basal Cell Carcinoma	514	3323
Actinic Keratosis	327	867
Benign Keratosis	1099	2624
Dermatofibroma	115	239
Vascular Lesion	142	253
Squamous Cell Carcinoma	NA	628
Total	**10015**	25,331

1.7.2 Variations in skin lesions

There are various challenges in the visual characteristics of skin lesion images, shown in Table 1.2, given as follows:

1.7.2.1 Various Shapes and Sizes

In skin lesion datasets, large variation in the shape and size of skin lesions is present in skin lesion images that increase their complexity. These variations make the skin cancer classification a difficult task.

1.7.2.2 Irregular Boundaries

A lot of skin lesion images have irregular boundaries that makes them difficult to segment and classify for segmentation and classification techniques respectively.

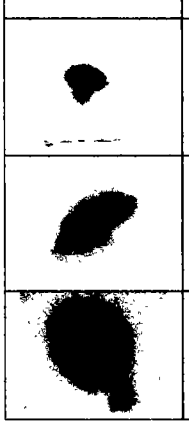
1.7.2.3 Noise and Artifacts

Various kinds of noise are introduced in datasets while acquiring images with acquisition devices. The classification and segmentation processes are affected by these noise and artifacts. This noise includes color illuminations, light reflection, blood vessels, hair artifacts, and bubbles.

1.7.2.4 Low Contrast

The low contrast present between lesion area and the background normal skin that cause the difficulty in accurate segmentation of skin lesion.

Table 1.2: Challenges in skin lesion dataset

Various Shapes	Various Sizes	Irregular Boundaries	Noise and Artifacts	Low Contrast
				

1.8 Problem Statement

1. There is lack of explainability in the existing CNN models to get the hierarchy of skin lesions which makes it difficult for dermatologists to make accurate decisions about skin lesion type and its sub-types.
2. In addition, various structures in skin lesion may appear in more than one of its types creating hurdles for dermatologists to make accurate decision about final skin lesion type.
3. Moreover, the existing imbalance data handling techniques may not perform well while dealing with structural level segmentation of skin lesions.

1.9 Research Questions

Research Question 1: How can we design a self-explainable CNN model by embedding in inherent hierarchical structure of skin lesion for its taxonomic classification?

Research Question 2: How can we perform the structure level segmentation of skin lesions?

Research Question 3: How can we resolve the existing class imbalance problem present in the structural segmentation dataset?

Research Question 4: How can we highlight and interpret the skin lesion regions and localize them, that contribute the most to predict an output?

1.10 Research Aim and Objectives

This thesis aims to develop Deep Learning based algorithms for the hierarchical classification and structural segmentation of skin lesions in dermoscopic images.

The objectives of the thesis are as follows:

- **Hierarchical Classification of Skin Cancer**

To propose an explainable CNN model for the hierarchical classification of skin cancer lesions, by using their inherent hierarchical structure.

- **Structural Segmentation of Skin Lesions**

To propose a CNN model for the structural level segmentation of skin lesion attributes.

- **Handling Class Imbalance in Dataset**

To handle class imbalance problem, present in skin cancer lesion datasets to make the hierarchical classification and structural segmentation more robust.

- **Structure Visualization and Interpretation**

To visualize and interpret the regions of the image contributing most to make prediction at each hierarchical level of classification through gradient class activation maps.

1.11 Research Contributions

Following are the contributions of this research thesis:

1. We have proposed an explainable CNN model for three-level hierarchical classification of skin cancer lesions, by using their inherent hierarchical structure, to make the CAD system self-explainable.
2. We have proposed a CNN model for structural level segmentation and localization of skin lesions attributes.

3. The problem of class imbalance present in ISIC skin cancer dataset has been addressed to make the taxonomic classification and structural segmentation more robust.
4. The proposed system also provides visualization at each hierarchical level of classification, through gradient class activation maps highlighting the image regions participating most in output.

1.12 Dataset

The efficiency of the proposed models for hierarchical classification and structural segmentation of skin cancer lesions is validated on dataset of International Skin Imaging Collaboration (ISIC) archive [31][32]. This dataset contains separate individual sets for training, validation, and testing. It is the standard benchmarked dataset for dermoscopic skin cancer images available on international level and released yearly. ISIC is a collective effort of academic and industry aiming to improve melanoma diagnosis and lowering melanoma death rate by aiding the application of digital skin imaging techniques. The more detail about ISIC archive can be seen in [33]. The deeper detail of this dataset is discussed later in Chapter 4.

1.13 Thesis Organization

Chapter 2 talks about the structure level detail of skin including its layers. The details of types of skin cancer lesions are discussed in this chapter. A complete hierarchy of skin cancer lesion types is given in pictorial form. This chapter also discusses the various structures or attributes that are present in skin lesions. Medical diagnosis of skin lesion by using various imaging techniques is also discussed. Clinical diagnosis of skin lesions including all clinical methods are also described in this chapter.

Chapter 3 provides a summary of previous research on skin cancer segmentation and classification using traditional Machine Learning techniques and Deep Learning based techniques. Previous work is divided into four categories: (1) Computer Aided Diagnostic (CAD) Systems for skin cancer diagnosis, (2) traditional skin cancer segmentation techniques, (3) CNN- based skin segmentation and classification techniques, (4) skin cancer hierarchical classification and structural segmentation techniques. This section provides extensive literature over the last 10 years on representation and classification methods in all categories. Gaps or limitations in previous research are identified and the problem is formulated at the end of this chapter.

Chapter 4 presents a CNN method for the explainable hierarchical classification of skin lesion. In this chapter, we proposed hierarchical classification named as Hierarchical GoogleNet (H-GoogleNet). The proposed architecture is illustrated in detail. Experimental setups and results are also given in detail. The proposed architecture is also compared with existing literature.

Chapter 5 presents a method for structural level segmentation of skin lesion. UNet model is applied with ResNet-50 as backbone architecture along a hybrid loss function. Existing architectures of UNet and ResNet are discussed in detail. The proposed ResNet50-UNet architecture is also discussed in detail. Experimental setup and obtained results are provided at the end of chapter. The proposed model is evaluated on ISIC dataset. A comparison of the proposed model is also given in this chapter.

Chapter 6 concludes the thesis with a discussion of contributions that are made for hierarchical classification of skin lesions, by proposing H-GoogleNet model, and structural segmentation model for skin lesions. Further enhancements are also discussed in this chapter. Future work is also described at the end of the chapter.

Chapter 2

Dermoscopy Image Analysis and Medical Diagnosis of Skin Cancer

Chapter 2

2 Dermoscopy Image Analysis and Medical Diagnosis of Skin cancer

In this chapter a detailed description of skin image analysis and medical diagnosis of skin cancer lesion is given in detail.

2.1 Skin Structure

Before moving toward the more detail of skin cancer and its dermoscopic image analysis, we first throw some light on the structure of the normal skin, in detail, to have its clear understanding.

Skin is the main part of the body. It performs very important functions such as protecting us from microbes (germs causing diseases), helps to normalize our body temperature and allows the sensation to touch, heat and cold. From top to bottom, the skin has three layers; *Epidermis, dermis, subcutis* [9] (see Figure 2.1). *Epidermis* is the uppermost layer of the skin. It is composed of connective tissue. *Epidermis* prevents entry of substances and organisms into the body. It has some more sub-layers in it, named as *basal cell layer, squamous cell layer, stratum granulosum, stratum lucidum* and *stratum corneum* (see Figure 2.2). *The basal cell layer* of *epidermis* has three main types of cells: *keratinocytes, melanocytes, and Langerhans cells. Melanocytes* cells produce a skin pigment known as *melanin. Melanin* gives color to the skin such as brown, dark brown or tan etc. *Melanocytes cells* protect the skin from dangerous UV sun rays by enhancing melanin's production. The second type of cells in Basal Cell layer is *keratinocytes*. It produces *keratin*. The second layer of the skin is *dermis*. It is the supportive layer of the skin. It consists of hair follicles and sweat glands. It also consists of collagens fibers, sensors, receptors, blood vessels and nerve ends. *Dermis* is further parted into two layers: *papillary region* and *reticular region*. *Dermis* and *epidermis* are tightly connected through *basement membrane*. The third layer of the skin is *subcutis*. It is the fat layer immediately below the dermis layer.

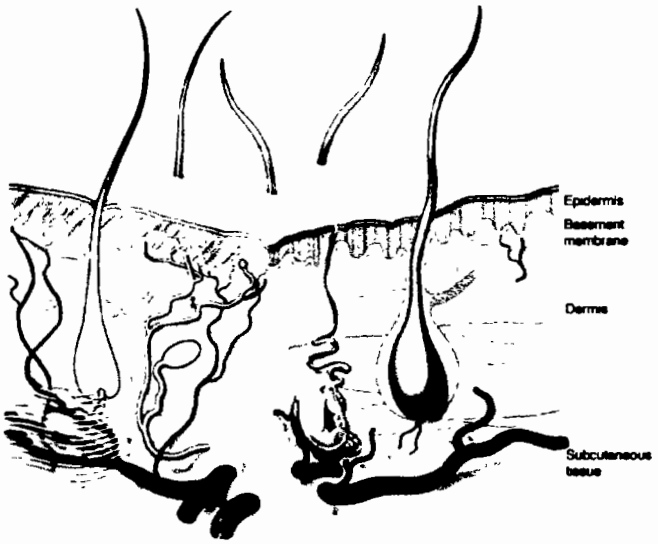


Figure 2.1: Normal skin structure [9]

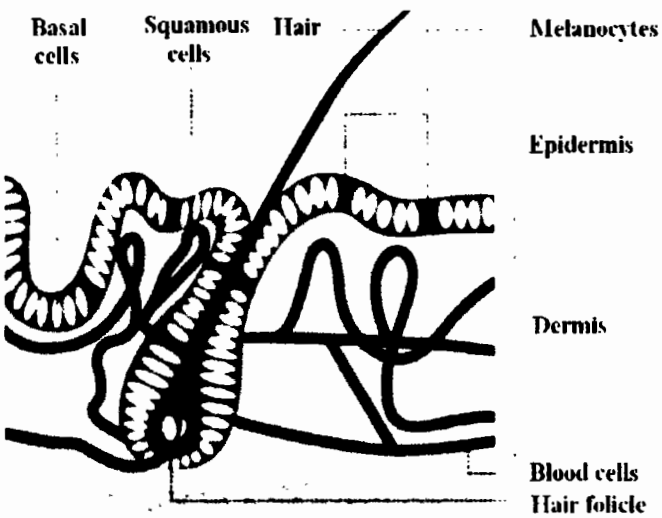


Figure 2.2: Normal skin main components [9]

2.2 Types of Skin Cancer Lesions

There are various types of skin lesions that can be categorized in a hierarchical way [34]. There exists a hierarchical classification structure of skin cancer lesions [9][34] (see Figure 2.3 & Figure 2.4). At the first step, each lesion is divided according to its origin such as according to its type of skin cells. In this step, the lesion is divided into *melanocytic* and *non-melanocytic* lesions. *Melanocytic* lesion like melanoma, develops from melanocytes cells. *Non-melanocytic* lesions develop from various types of skin cells like basal cells or squamous cells. After getting the origin

of lesion, the next step is to classify the category of lesion such as whether it is *benign* or *malignant*. *Melanocytic benign* lesions are non-cancerous in nature and are not harmful. On the other side, *melanocytic malignant* lesions are cancerous, and they spread rapidly, hence are life threatening. If the melanocytic lesion is benign, it is then classified into further types of lesions such as *clark nevus*, *spits nevus* and *blue nevus* as shown in Figure 2.3. If the melanocytic lesion is malignant, then it is classified as *melanoma*. After that, *melanoma* is further divided into its sub-types such as *superficial*, *lentigo*, *acral lentiginous*, *nodular*, *mucosal*, *desmoplastic*, *lentiginous* etc. The difference among all these lesions is made based on the presence or absence of some dermoscopic structures or attributes in the lesion. For example, these structures may be *pigment network*, *nodular*, *white streaks*, *globular*, *dots and globules*, *cobblestone pattern*, *reticular pattern*, *branched streaks*, *structureless areas*, *blotches*, *regression*, *blue-white veil*, *milia-like cysts*, *comedo-like openings*, *fingerprint like structures*, *fissure ridges*, *leaf-like areas*, *blue-grey globules* etc. [35] (see Figure 2.5). All above lesions are divided based on these dermoscopic structures present in them.

After that, the second major category of skin lesion is *non-melanocytic* lesions. *Non-melanocytic* lesions are also segregated into *benign* and *malignant*. *Benign* lesions are further categorized as *vascular lesions*, *dermatofibroma*, *seborrheic keratosis*, *neurofibroma* and *actinic keratosis* as shown in Figure 2.3. *Non-melanocytic malignant* lesions are categorized as *BCC* and *SCC*. *BCC* is the most frequent form of skin cancer. It develops very slowly so it is less dangerous as compared to *malignant melanoma*. There are more sub-types of *BCC* such as *nodular*, *superficial*, *morphoeic*, *basosquamous and pigmented* etc. Similarly, *SCC* can be further sub-divided as *adenoid*, *intraepidermal*, *lymphoepithelial*, *papillary*, *spindle cell* and *verrucous* etc. (see Figure 2.3). There are some example images of various kinds of skin lesions and their structures [35] (see Figure 2.4 & Figure 2.5). All this categorization is based on the structural information of skin lesion. There can be similar structures present in different lesions such as they can be overlapped or have same characteristics. This similarity makes it difficult to differentiate among various types of skin lesions, especially; melanocytic from non-melanocytic lesions [36]. It is very challenging to diagnose skin cancer type due to these similarities. This is because sometimes melanocytic malignant melanoma impersonates the look of other skin lesions i.e. of non-melanocytic lesions [36] creating difficulties for practitioner in accurate diagnosis of origin of lesion. That is why there

is need to model an automated system that can accurately identify the origin of lesion, level of malignancy, till diagnosis of its fine class. This will help the practitioner in accurate decision making and to get rid of confusions that occur due to structure's similarities among various lesions during skin cancer analysis.

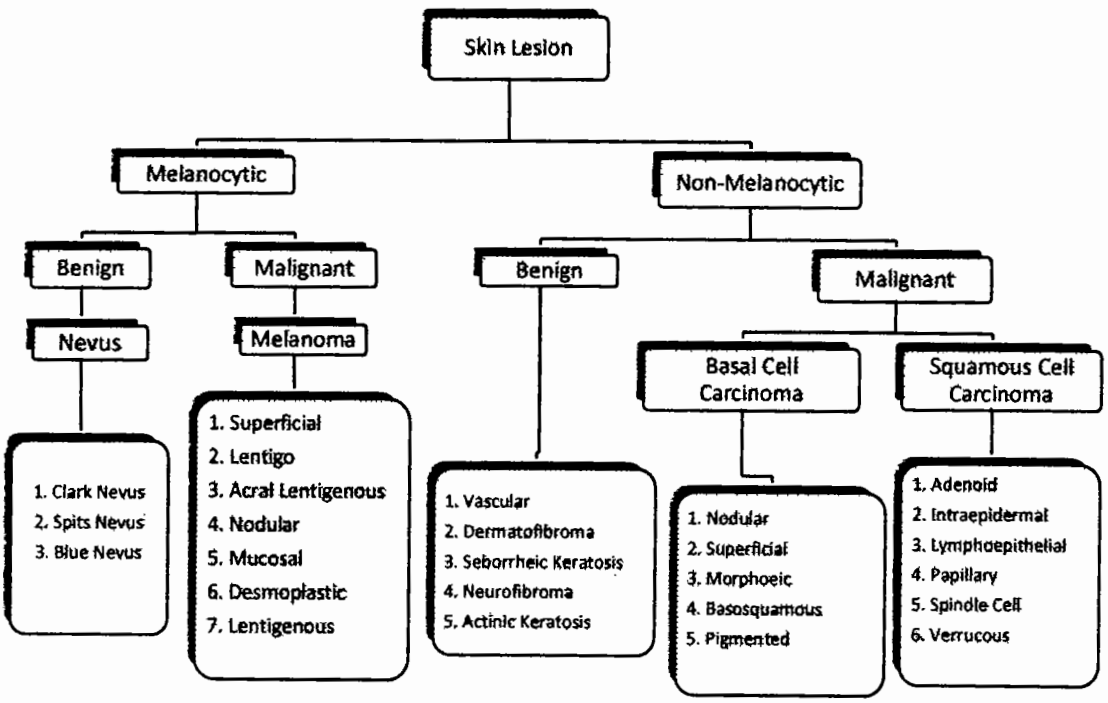


Figure 2.3: Skin lesions hierarchical classification [9]

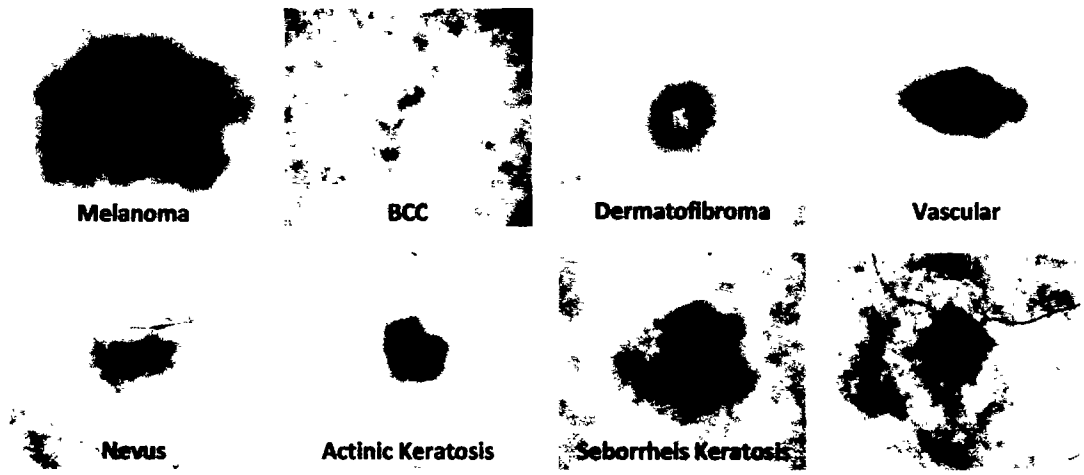


Figure 2.4: Examples of various types of dermoscopic skin lesions.

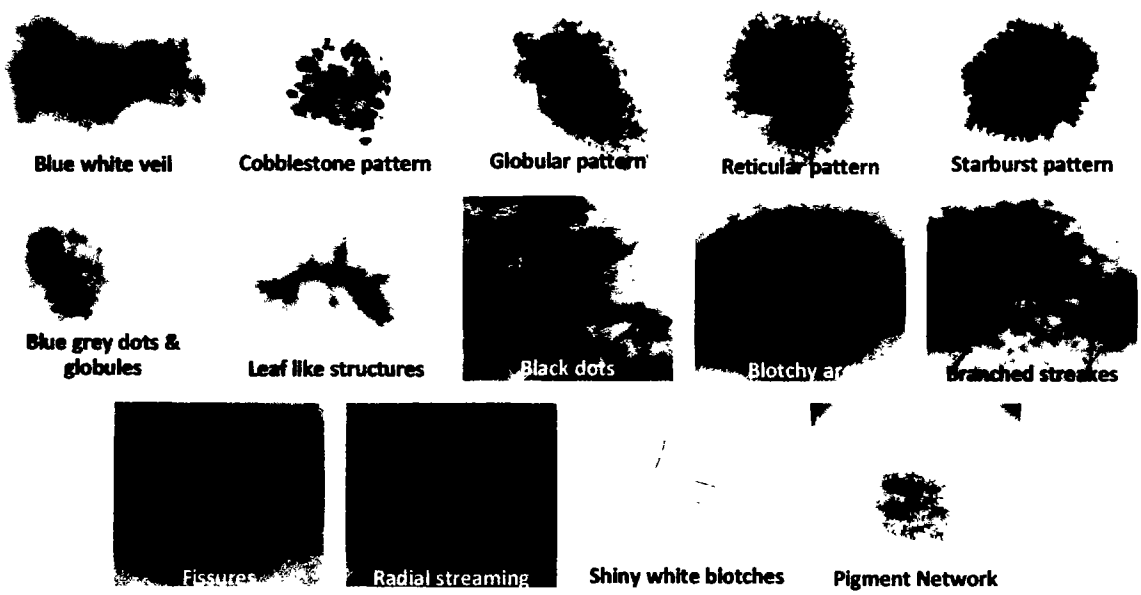


Figure 2.5: Examples of various dermoscopic structures present in different skin cancer types [35]

To deal with the above discussed issues, dermoscopy lets the visualization of attributes and patterns that are not noticeable by naked eye, enabling dermatologists to make more accurate assessments. Analyzing these structures plays a crucial role in the accurate classification of skin cancer types.

Here we will discuss the key dermoscopic structures commonly observed in skin cancer lesions and their significance in clinical practice:

1) **Pigment Network**: The pigment network is a fundamental dermoscopic structure observed in skin cancer lesions [35]. It refers to the regular or irregular distribution of pigmented lines and grids within the lesion. The evaluation of the pigment network can provide insights into the organization and pattern of pigmented cells, which aids in distinguishing benign lesions from malignant ones. Irregular or disrupted pigment network patterns are often indicative of malignancy. It is a key feature in the diagnosis of melanoma. The analysis of the pigment network involves evaluating its distribution, thickness, regularity, and overall arrangement. Deviations from a regular, symmetrical pattern may indicate malignancy.
2) **Globules:** Globules are round or ovoid structures observed within the skin lesion [35]. They can vary in size, color, and distribution. The analysis of globules helps in differentiating between benign and malignant lesions. In benign lesions, globules are usually regular, symmetric, and evenly distributed. Malignant lesions, on the other hand, may exhibit irregular, asymmetrical, or clustered globules, which can be indicative of malignancy. Globules are round or ovoid structures observed within the lesion. They can vary in size, color, and distribution. Globules can be pigmented or non-pigmented and may provide valuable information for differentiating melanoma from benign lesions. The analysis of globules includes assessing their size, shape, color uniformity, and arrangement.
3) **Streaks and Lines**: Streaks and lines refer to linear structures observed within the lesion [35]. These structures can provide valuable diagnostic information. In benign lesions, streaks and lines are often uniform, thin, and evenly distributed. In contrast, malignant lesions may exhibit thick, irregular, and asymmetric streaks or lines, which can indicate malignancy. Additionally, the presence of a radial streaking pattern, extending outward from the centre of the lesion, can be indicative of invasive melanoma. Streaks are linear or irregular pigmented structures observed within the lesion. They can be thin or thick and may have various colors, including brown, black, or even blue gray. The presence of streaks can be associated with melanoma, particularly if they are asymmetric, irregular, or demonstrate abrupt changes in color or thickness.

4) **Negative Network**: The negative network, also known as the hypopigmented or depigmented network [35], appears as lighter or colorless lines or areas within the lesion. It represents the absence of pigmentation in certain areas. The presence of a negative network can be suggestive of certain types of melanoma or other non-melanocytic skin cancers.
5) **Milia-like Cysts:** Milia-like cysts are small, white, and yellowish cystic structures [35] observed within the skin lesion. These structures resemble milia, which are tiny epidermal cysts commonly seen in healthy skin. In the context of skin cancer, the presence of milia-like cysts can be indicative of specific subtypes or serve as a diagnostic clue for differentiating benign and malignant lesions.

2.3 Medical Diagnosis of Skin Lesion

In this section, we discuss the medical diagnosis of skin lesions by using skin lesion imaging techniques, clinical diagnosis and by using automated procedures.

2.3.1 Skin Lesion Imaging Technique - Dermoscopy

Numerous image acquisition methods have been developed for the deep examination of skin lesions. The most common technique used for lesion analysis and its regular follow up is clinical imaging. These clinical images are attained by using digital cameras. Though, this common method does not extract images with good resolution, resulting no deeper inspection of skin lesion such as its structural level details. The deeper inspection of lesion allows to measure its growth [37][38]. Other imaging techniques such as Computerize Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) are also used to analyze skin lesions, but these approaches are not useful to diagnose the lesion at earlier (*in-situ*) stage. It is necessary to diagnose any type of skin lesion at an *in-situ* stage. To achieve this goal, dermoscopy [10] or microscopy is used to extract the deeper detail of skin lesion. Macroscopy is another device to capture deeper details of skin lesions.

Dermoscopy [10] is a non-invasive imaging technique to capture the deep structure level detail of skin lesion by going deeper into epidermis and dermis. It is also known as dermatoscopy. It was first introduced in 1987. Dermoscopy can clearly visualize the morphological structures of skin lesion that are not clearly observable by naked eye. To capture an image of skin lesion, first gel or oil is applied on skin specific area to analyze the pigmented structures of lesion by using dermatoscope, a clinical device to capture skin lesion image (see Figure 2.6). The gel or oil is

applied to minimize the reflection of dermatoscope light and get the transparent epidermis and dermis. By doing this, the underlying structures of skin lesion become more visible. These visible structures are further utilized for the first step diagnosis (i.e., discriminate between melanocytic and non-melanocytic lesions) and the second step diagnosis as well, such as benign and malignant. Dermatoscope is used by the expert dermatologists and trained practitioners to examine the skin lesion structures and patterns in detail. There are few examples of digitally acquired images of skin lesions by dermoscopy (see Figure 2.7).

Figure 2.6: Examples of dermatoscope.

Figure 2.7: Examples of dermoscopic images of skin lesions.

2.4 Clinical Diagnosis of Skin Cancer

From last few decades, a lot of effort has been made in the clinical diagnosis of skin cancer such as, melanoma, BCC, SCC by using some clinical manual techniques. These major techniques include, pattern analysis technique [11], ABCD *(Asymmetry, Border, Color, Diameter)* rule of

dermoscopy for self-screening [12][13][14], 7-Point Checklist [12] [15][16], Menzies method [17], CASH algorithm [18] etc. These methods are used by trained general practitioners for the accurate diagnosis of melanoma, BCC, and other types of skin cancer, and to differentiate among melanocytic and non-melanocytic lesion. These clinical techniques have given an acceptable accuracy rate in skin cancer diagnosis but in expert hands, as every dermatologist is not enough trained in use of these manual methods. The lack of trained practitioners to use manual clinical methods is the main reason to meet the need of some automated systems such as CAD systems to achieve better diagnosis and accurate second opinion. There is also a need to reduce the dependency on practitioners or to train them enough so that they can diagnose the disease easily without putting much effort.

Following are some major clinical diagnosis techniques that are used by dermatologists in routine:

2.4.1 Pattern Analysis

In dermoscopy, the lesion's broad category or origin is identified first to determine whether the lesion is melanocytic or non-melanocytic. Pattern analysis[11] method is preferably used by expert dermatologists to diagnose melanocytic lesion and to differentiate among benign melanocytic lesions and malignant melanoma by analyzing their morphological features within the lesion. This method supposes that there are various patterns, also called global features, present in all types of skin lesion. A definite pattern is described by one or more dermoscopic attributes (see Figure 2.5 in Section 2.2), covering some parts or the whole lesion. These morphological features and dermoscopic structures may include some reticular patterns, globular patterns, cobblestone patterns, starburst patterns, homogenous patterns, parallel pattern and other than that colors, regularity or irregularity of the lesion boundary, pigmentation intensity and the surface of the lesion [35]. The detail of these specific dermoscopic structures [11][34][35] is as follows:

1) Reticular Pattern:

This reticular pattern is described by pigment network present in various parts of the lesion. This pattern is most commonly present in melanocytic lesions developed from melanocytes.

2) Parallel Pattern:

It is defined by pigmented lines. These lines are organized parallelly.

3) Homogenous Pattern:

These patterns appear as diffuse pigmentation that are typically brown, blue-grey, or grey black.

4) Starburst Pattern:

These patterns are described by pigmented streaks that are present in lesions in radial arrangement. These streaks are localized at the border of the lesion.

5) Globular Pattern

These patterns are present in the form of several oval structures. These oval structures are known as dots and globules. They have different colors and sizes.

6) Cobblestone Pattern

These are alike globular patterns except in this, the globules are strictly combined.

Pattern analysis method evaluates the dermoscopic criteria qualitatively and exhibit the maximum diagnostic accuracy as compared to other clinical methods such as ABCD rule [12][13], 7-Point Checklist [16], Menzies [17] and CASH algorithm [18] when used by experienced dermatologists. By using pattern analysis method, both melanocytic and non-melanocytic (first step diagnosis) skin lesions can be recognized and diagnosed. The second step decision such as benign or malignant also depends on various dermoscopic structures along its shapes that form a pattern. Pattern analysis led to a precise decision achieved by dermatologists. Though, the valuation of the local dermoscopic attributes and its pattern description is specific and lacks reproducibility. To handle these issues, more constrained procedures have been developed. This is achieved using Table 2.1, describing melanocytic algorithm [39].

Table 2.1: Melanocytic algorithm [39]

Step	Dermoscopic Criteria	Type of Lesion
I	Streaks Pigment Network Homogenous Blue Pigmentation Brown to black dots	Melanocytic
II	Milia-like cysts Comedo-like openings	Seborrheic Keratosis
III	Arborizing vessels Leaf-like areas Irregular grey-blue globules	Basal Cell Carcinoma
IV	Red lacunas Red bluish to red-black homogenous areas	Vascular
V	Central white patch surrounded by pigment network	Dermatofibroma
VI	None of the above criteria	Melanocytic

2.4.2 ABCD Rule

ABCD rule of dermoscopy [12][13][14] is an algorithm that evaluate four various types of features such as *Asymmetry (A), Border (B), Color (C) and Dermoscopic structure (D)* over lesion (see Figure 2.8). This method can classify the lesion into benign and malignant. It is a semi-quantitative scoring system. There are some specific criteria based on which these ABCD features are scored quantitatively. These criteria are given in Table 2.1. Scores are weighted depending on the values given in Table 2.2. Total dermoscopy score (TDS) is calculated using equation (2.1).

$$TDS = A_{score} \times 1.5 + B_{score} \times 0.4 + C_{score} \times 0.6 + D_{score} \times 0.7 \quad (2.1)$$

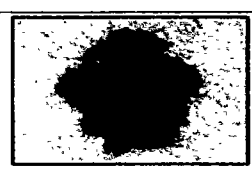
	Asymmetry	Border	Color	Diameter
Benign	 Symmetrical	 Even Edges	 One Shade	 Less than 6mm
Malignant	 Asymmetrical	 Uneven Edges	 More Shades	 More than 6mm

Figure 2.8: ABCD rule for malignant melanoma and benign nevi [12][13]

After that, the final diagnosis is based on the TDS value.

Table 2.2: ABCD rule of dermoscopy [12]

Criterion	Description	Score	Weight
Asymmetry	Axis is drawn to measures contour, colors, and structures.	0 - 2	1.5
Border	Sudden or random ending of pigment pattern at the border.	0 - 8	0.4
Color	Presence of 6 colors 1 – 6 (white, red, light brown, dark brown, blue grey and black)	1 - 6	0.6
Dermoscopic structure	Presence of network, streaks, dots, and globules.	1 - 5	0.7

The detail of Table 2.2 is given as follows:

1) Asymmetry:

To calculate asymmetry, the lesion is segmented by two perpendicular axes. The asymmetry score is computed depending on the shape, color sharing, and lesion structure on either side of axis.

2) Border:

The score of borders is calculated by examining the intensity of lesion border. For this, the lesion is parted into eight segments. In each segment, an abrupt cut-off of pigment pattern at the boundary of the lesion gets a score 1, while a steady, unclear cut-off gets a score 0. The maximum border score is 8 and the minimum score is 0.

3) Color:

A broader color range of melanocytic lesions can be easily detected with the help of a dermatoscope. In this, a total of six discriminant colors namely: black, blue-grey, light-brown, dark brown, white, and red are counted. The maximum score is 6 if all colors are present and minimum score is 0.

4) Dermoscopic Structures:

Various dermoscopic structures namely, pigment network, dots, streaks, globules, cobblestone are present in skin lesions. The maximum and minimum score for D is 5 and 1 respectively.

2.4.3 3-Point Checklist

3-Point Checklist [40] is a manual clinical method used to distinguish malignant lesions (melanoma or BCC, SCC) from benign pigmented lesion. This method was developed in 2004. This method relies upon three specific criteria, namely, *Asymmetry in structure, Atypical pigment network and blue-white structures*. A 3-point checklist is used for early detection of skin lesions. This method allows non-experts not to miss the detection of melanoma. These three criteria are as follows:

1) Asymmetry in Structure

Asymmetry of color and structure in one or two perpendicular axes.

2) Atypical Network

In this, irregular pigment network with thick lines is present.

3) Blue-White Structure

In this, blue or white color is present on the lesion. A combination of blue-white veil and regression structures may also be there.

The major purpose of the 3-Point Checklist method is to identify whether the examined lesion should go for biopsy or not. In this method, the exceptional attributes of lesion are not assessed.

2.4.4 7-Point Checklist

7-Point Checklist [12][15][16] is another manual technique that is used by dermatologists for skin cancer diagnosis. It is used for the dermoscopic difference between benign melanocytic lesions and malignant melanoma. In this, seven different criteria are used for the diagnosis of skin lesion type. These criteria are basically dermoscopic structures, separated into two classes namely: major and minor criteria as shown in Table 2.3.

Table 2.3: 7-Point Checklist [16]

	Local Features	Score
	Major Criteria	
1	Atypical Vascular Pattern	2
2	Blue-White Veil	2
3	Atypical Pigment Network	2
	Minor Criteria	
1	Irregular dots or globules	1
2	Irregular streaks	1
3	Regression structures	1
4	Irregular pigmentation	1

The major criteria contain atypical vascular pattern, blue-white veil, and atypical pigment network. On the other side, minor criteria include irregular dots/globules, irregular streaks, regression structures and irregular pigmentation. If anyone criterion is present in skin lesion, it will get a score 1 or 2 depending on the criteria, as shown in Table 2.3. After getting separate scores, they are summed up to compute the entire score for the lesson. If the total score is greater than 3 the lesion is diagnosed as malignant melanoma, otherwise benign.

2.4.5 Menzies Method

Menzies method [17] is another clinical technique used by dermatologists to discriminate the dermoscopic features of malignant melanoma and benign melanocytic lesion. In this, two groups of features are defined: one group for positive features and the second group for negative features. *Negative features* are present in benign lesions and *positive features* are present in malignant melanoma. *Negative group of features* include single color and symmetrical pattern, such as of

color and structure, about any axis over the center of the lesion. On the other hand, group of positive features include nine features: blue-white veil, multiple brown dots, pseudopods, radial streaming, scar-like depigmentation, peripheral black dots/globules, several colors (5 to 6), various blue/grey dots, and broad-end network. For melanocytic lesion to be identified as malignant melanoma, at least one positive feature should be present in the lesion and no negative feature should be there. In this method, 5 to 6 multiple colors include grey, red, tan, black, dark brown and blue.

2.4.6 CASH Method

CASH [18] is one more clinical method for skin lesion diagnosis. This method is appropriate for less skilled practitioners. It is a scoring system just like ABCD rule. CASH stands for ***Color, Architecture, Symmetry and Homogeneity***. It is particularly used for the difference between benign melanocytic lesion and malignant melanoma. In this method, the colors include light brown, dark brown, black, red, white, and blue.

Chapter 3

Literature Review

Chapter 3

3 Literature Review

In this chapter, a thorough study of previous approaches for skin cancer segmentation and classification is provided. Since a large amount of literature about skin cancer segmentation and classification is available, this chapter will focus on the research that is closely linked to the work presented in this thesis. This chapter provides a review of the existing methods such as Machine Learning and Deep Learning based skin cancer segmentation and classification techniques. Section 3.1 covers existing Computer Aided Diagnostic (CAD) systems. Traditional skin cancer segmentation techniques are reviewed in Section 3.2. In Section 3.3, CNN based skin cancer segmentation and classification techniques are reviewed. Finally, existing techniques for hierarchical based skin cancer classification and structural based skin lesion segmentation are explored in Section 3.4. After an extensive review, research gaps are presented in Section 3.5 and problems are formulated in Section 3.6.

Literature review is divided into four categories as follows:

3.1 Computer Aided Diagnostic (CAD) systems for skin cancer

3.2 Traditional skin cancer segmentation techniques.

3.3 CNN-based skin cancer segmentation and classification techniques.

3.4 Skin Cancer hierarchical classification and structural segmentation techniques.

3.1 Computer Aided Diagnosis (CAD) Systems for Skin Cancer

It is a great challenge to diagnose the skin cancer lesion at an earlier stage, using dermoscopy images. Even using clinical methods, discussed in Chapter 2, such as ABCD rule [12][13], Pattern Analysis [11], 3-Point Checklist [39], 7-Point Checklist [16], Menzies method [17] and CASH algorithm [18], it is still not easy to discriminate melanocytic lesions from non-melanocytic lesions and to further differentiate among their sub-types. This is because some lesions appear alike one another and may also have overlapped structures in them. This challenge may cause an increase in the number of unwanted or useless histological inspections. The lack of clinical experience of some practitioners, when working with dermoscopic images, may also be the reason for incorrect diagnosis of skin lesion. Moreover, many dermatologists may conflict in their analysis of skin

lesions. This is because the valuation of the various dermoscopic standards is based on the understanding of dermatologists.

All these factors encouraged the need of computer-aided systems for the accurate diagnosis and assessment of skin cancer lesions. Computer Aided Diagnostic (CAD) systems are one of the examples of that kind of system. CAD systems have various strengths and advantages that can be very cooperative for experts and dermatologists. One of these advantages is that the diagnosis of skin lesion is not dependent on the person who is using it. Another advantage is that this CAD system can be applied for the follow-up of a particular lesion. The CAD system can be operated by both skilled and non-skilled dermatologists, such as it works as a second judgment tool for dermatologists.

In literature, there exist various CAD systems [19 - 25], [41] that have been proposed for early detection and classification of skin lesions. These automated systems perform pre-defined phases involving skin image acquisition, image pre-processing, skin lesion segmentation, feature extraction, feature selection and finally the classification steps successfully. Following is given a review of some of these CAD systems in detail for skin lesions. One major limitation of these systems is that they can diagnose and differentiate only melanocytic benign and malignant melanoma but not among any other type of lesion.

One of these kind of systems is designed by J. Lopez-Labraca *et al* [19]. They proposed a CAD system founded on dermoscopic structures of melanoma to diagnose it. Skin lesion is segmented into dermoscopic structures by applying the combination of Latent Topic Modelling (LTM) and Kernel Logistic Regression (KLR). Soft segmentation maps are used to segment the skin lesion instead of using the hard segmentation. They performed category-based segmentation using supervised and unsupervised methods. They also have extracted structure-specific features along with generic features. In this, for each individual structural pattern, an individual classifier is constructed. To construct these individual classifiers or experts, they proposed a wide set of clinical features that are used by dermatologists for skin lesion diagnosis. They have also designed one more approach to choose those specific features that are more related to describe each of the dermoscopic structures. At the end, decisions of all classifiers are fused together by means of Bayesian method to provide the doctor useful information. However, this method only considered

the structures of melanoma and did not consider the dermoscopic structures of any other type of skin lesions.

Another CAD tool for the classification of melanocytic lesions into benign and malignant is presented by S. Pathan [20]. She has developed a hair detection method that detects and removes both light and dark hair to pre-process the image. The hair is removed by extracting the blue channel of the RGB image. For boundary segmentation, fuzzy C-Mean clustering is used to localize the lesions by choosing the number of clusters as 2. After boundary segmentation of foreground image, an optimal feature set including color, shape and texture based on ABCD rule was developed to extract color and shape features. For classification, Artificial Neural Network (ANN) is used. Experiments were accomplished using PH2 dataset of skin lesions. The accuracy of 82% was achieved by the proposed system along with sensitivity 85.71% and specificity 81.25%. This method also only focused on the classification of melanocytic lesion by considering only benign nevus and malignant melanoma.

S. Jain *et al* [21] presented a CAD tool to analyse the size, texture and shape features for segmentation and classification stage of skin lesion. This method was also only proposed for melanoma detection by applying image processing techniques. The input image was first pre-processed for contrast enhancement. To segment the image lesion, edge detection is used after applying Otsu's automated threshold. The segmented image is further used to extract features including geometrical features (area, irregularity index, circularity index and greatest diameter) and ABCD features too. After that, these features are passed to the classifier that classifies the skin lesion into benign or malignant by comparing them with the pre-defined threshold.

One more similar method for automated segmentation and classification of skin lesions was designed in [22]. Skin lesion images were enhanced by applying image enhancement techniques. To filter the noise, present in skin lesion images, Gaussian smoothing filter was used of size 3x3. After that, morphological operations such as erosion and dilation are applied for further image enhancement. Region growing method was used for image segmentation, by determining initial seed points. To characterize the lesion part, color histogram and texture features are used by the authors. Two classifiers namely, Support Vector Machine (SVM) and k-Nearest Neighbor *k-NN* for the classification purpose are used individually and their ensemble too. 46.71% F-measure

score is obtained using SVM classifier and 34% F-measure score with k-NN classifier is obtained. F-measure score of 61% was achieved by fusing SVM and k-NN together.

The enhancement of skin lesion image using sigmoidal function and then the segmentation of melanoma and nevus is performed by combining morphological operations and thresholding [23]. The system involves three main steps namely, hair detection along with inpainting, contrast enhancement and lesion area segmentation. The pixels containing hair information are enhanced by Gabor wavelet-based directional filters. After that, these pixels are in-painted by proposed neighbourhood-based region filling algorithm. Experiments were performed on the European database of dermoscopic images. This system is also focused only on the boundary segmentation of melanocytic melanoma and nevus lesion. There is no structural level detail segmentation of skin lesion is performed. Their focus is on the hair enhancement and their removal and finally the lesion boundary segmentation from the background.

Along with the classification of melanocytic lesions there is also intense need for non-melanocytic lesions classification [24] while designing a CAD system. A computer aided system was designed by R. Suganya [24] to classify melanocytic lesion from non-melanocytic lesion along with second step decision to differentiate among benign nevus and melanoma (melanocytic) and malignant BCC and benign SK (non-melanocytic). First, hair artifacts are removed by applying median filter. For lesion boundary segmentation, K-mean clustering was used. After color and texture features extraction, the two-step classification of lesion is performed by using SVM classifier. Experiments are performed on Dermweb dataset of skin lesion. This method achieved 96.8%, 89.3% and 95.4% of accuracy, specificity, and sensitivity respectively. However, this method lacks how structural level detail of each type of lesion was extracted. The summary of all the above discussed CAD systems is given in Table 3.1.

3.1.1 Commercial CAD System Devices

There are also some commercial CAD system devices [42 - 46] have also been developed and practically installed in hospitals and dermatology centers to provide accurate second opinion to the experts or dermatologists.

3.1.1.1 SolarScan

One of these CAD system devices is *SolarScan* [42] (see Figure 3.1), that is developed by Polar-technics Ltd in Australia in 2004. This device can detect warning signs of skin cancer at an early

stage. Unnecessary surgical procedures are reduced using this device. This device is in use in medical centers in Australia, US, and Europe from several years. The developers of this device claim that this device can determine melanoma even before it reveals its features. There is also an image analysis software that is installed along with *SolarScan* device. First, *SolarScan* captures the image of a patient's skin spot. The image analysis software removes artifacts from skin lesion images such as hair, bubbles, gel etc. and then analyzes the features of lesion such as shape, color, and texture etc. After that, SolarScan compares these features with the images of melanoma and non-melanoma in database and then finally return the results.

3.1.1.2 DermoGenius-Ultra

DermoGenius-Ultra [43] (see Figure 3.1), was developed by LINOS Photonics Inc. Germany, as shown in Figure. It is an image capturing device with a 5 mega-pixel attached camera. This device is easily moveable from one place to another. It has two image acquisition modes: live mode and standard mode. Live mode is employed for immediate mole inspection without predefining a localization typically done with an optical dermoscope. In standard mode, the lesion localization is defined first and then the pictures are taken one after the other. The device stores all captured images of one patient in a database.

3.1.1.3 DBDermo-MIPS

DBDermo-MIPS [44] (see Figure 3.1), was developed at the university of Siena in Italy. It is a digital dermoscopy analyzer. This device does clinical examination of lesion in three major steps: In step 1, the clinician captures the skin lesion area by using camera attached with *DBDermo-MIPS*. A series of images is taken of one patient and then stored in a database. This device is also used to acquire, frame, and analyze the digital images of skin lesions. It calculates various features related to the geometry, color distribution and internal pattern of the lesion.

3.1.1.4 Diagnosis and Neural Analysis of Skin cancer (DANAOS)

Diagnosis and Neural Analysis Of Skin cancer (DANAOS) expert system [45] (see Figure 3.1), developed by Center of Neuroinformatic. This device performs image processing steps, feature extraction and then pattern classification. This system can assist in the analysis of melanoma, non-melanoma, vascular changes in lesion and before and afterwards documentation of skin treatment. The device was developed for the computer-aided diagnosis of pigmented skin lesions using neural networks.

3.1.1.5 Melafind

Melafind [46] (see Figure 3.1), was developed by Electro-Optical Sciences Inc., Irvington etc. The device is only used for the diagnosis of melanoma. It is not used for all akin lesions. It uses an imaging system to capture images of skin lesions. It simply uses comparison of the captured image with the images stored in database. It purely results in yes or no. It intends to provide additional information on melanocytic lesions such as melanoma.

Though, there is still a great need to adopt more accurate computer-based algorithms in routine diagnostic process.

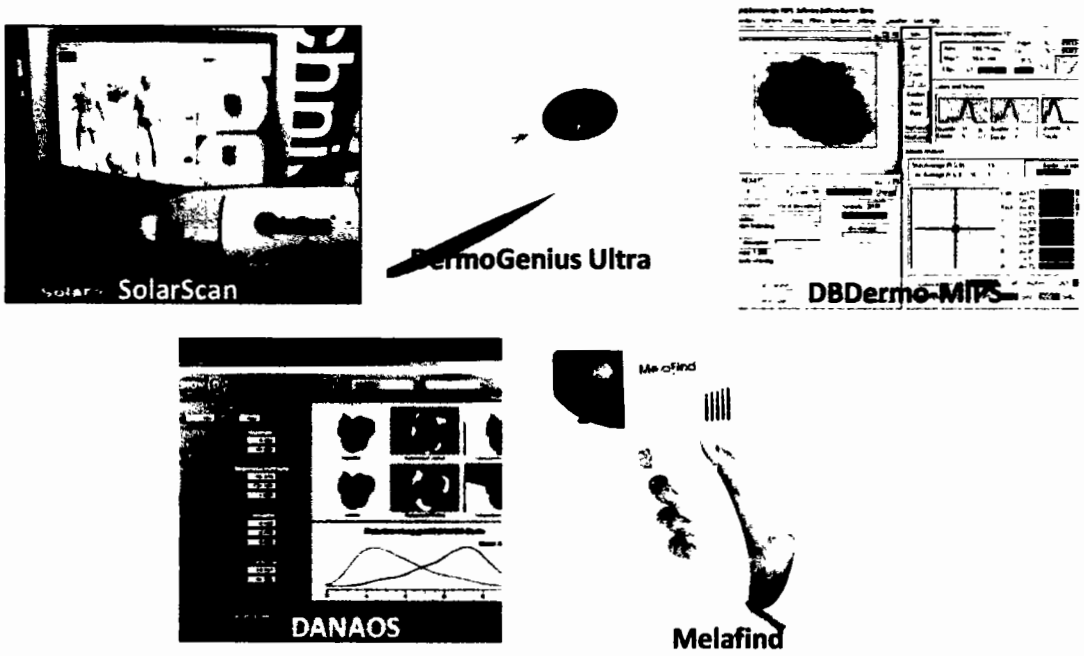


Figure 3.1: CAD system devices.

All these CAD systems have been developed to assist the practitioners or dermatologists for the detection and diagnosis of skin cancer including melanoma, SCC, BCC etc. There has been a great boost in interest in computer-aided diagnosis of skin cancer and its types. The aim of this increase is to eliminate bias and ambiguity from the diagnostic procedure and offer a trustworthy second judgement to dermatologists.

Table 3.1: Summary of CAD systems

Sr No.	Reference	Technique / Tool proposed	Dataset used	Implementation tool	Evaluation performed	Research findings
1	[19]	1. Dermoscopic structure-based CAD system for melanoma detection. 2. Individual classifier for each individual structural pattern.	EDRA Interactive Atlas of Dermoscopy [47]	MATLAB	AUC	System provides additional information to dermatologists
2	[20]	Hair detection and removal, boundary segmentation using fuzzy C-Mean clustering, ANN based classification.	PH2	MATLAB	Accuracy, sensitivity, specificity,	FCM clustering gave better results for lesion's localization. Weighting exponent gave better segmentation.
3	[21]	CAD tool to analyse the size, texture and shape features for segmentation and classification stage of skin lesion	PH2	MATLAB	Masks	Tool is useful for rural areas where experts may not be available.
4	[22]	Developed method for Preprocessing, segmentation, and classification	Own private data	MATLAB	Measure of Overlap, Measure of under segmentation, Dice similarity, Precision, Recall, F-Measure, Accuracy	System performed well for segmentation and classification.
5	[23]	Hair removal and boundary segmentation of melanoma and nevus	European database	MATLAB	TDR, FPR, ER	System handles the problem of unwanted artifacts. Boundary segmentation performed well.
6	[24]	Classification of melanocytic and non-melanocytic and further classification of benign and malignant.	Dermweb dataset	MATLAB	Sensitivity, Specificity, Accuracy	The system distinguishes well among various skin lesions using SVM classifier.

3.2 Traditional Skin Cancer Segmentation Techniques

The most useful and informative skin lesion features have their great impact in the classification of skin lesions. These features are obtained by the accurate segmentation of skin lesion from its surrounding tissue. Various approaches have been proposed in literature [48 - 55] for skin lesion segmentation. These methods are based on traditional Machine Learning techniques. All these

methods are limited to only boundary segmentation of skin lesions but not their structural segmentation.

One of these segmentation methods is proposed by S. Garg and B. Jindal [48]. The method is composed of pre-processing and segmentation steps. The proposed method used threshold and morphological operations to reduce artifacts. A hybrid of K-Mean and firefly algorithm (FFA) is used to perform segmentation of skin lesion. During segmentation, K-Mean identify the exact lesion region. After that, FFA is applied to optimize the clustering and to get high accuracy. The experiments are performed on ISIC 2017 [27] and PH2 dataset. The model achieved the accuracy rate of 99.1% on ISIC dataset and 98.9% on PH2 dataset.

N. Durgarao et al. [49] proposed a technique for skin lesion segmentation based on three steps including: segmentation, feature extraction, and classification. For segmentation they have used fuzzy C-Mean (FCM) clustering and features are extracted by a hybrid of local vector pattern (LVP) and local binary pattern (LBP) that are used as an input to fuzzy classifier. Their major contribution is the optimization of member function in fuzzy classifier by introducing DOROA algorithm.

A combination of CNN model, named as YOLO, and grabcut algorithm [50] was proposad by H. M Unver et al. for skin leison segmentation. The proposed algorithm works in four steps: removal of hairs on the lesion, location of lesion detection, lesion segmentation, and post processing of lesion. The model was assessed on PH2 dataset and ISIC 2017 [27]. The model performed well on ISIC dataset.

Another model [51] for skin lesion segmentation proposed by N. Fulgencio et al. was proposed based on novel adaption of superpixels techniques. It gave suitable results on ISIC 2017 dataset. They have also introduced a modified image registration approach to measure the progress of features depicting the skin lesion, taking two images of the same lesion to capture different stages of the same lesion.

A threefold method [52] for the detection and classification of skin lesion using segmentation and feature selection methods was proposed by Talha *et al.* They used three color spaces to separate foreground image from background. They introduced a weighting criteria to choose the best solution based on extensive feature analysis, using associated labels, and central distance. Further,

an improved form of feature extraction and dimensionality reduction was introduced. The method performs well and give better results as compared to existing techniques.

R. Kaur et al. introduced fifteen thresholding methods for the segmentation of Basal Cell Carcinoma (BCC) [53] along with two error metrics: relative XOR error and lesion capture ratio. Lesion segmentation is also performed by using geodesic active contour technique [54] to get accurate contours of lesion by initializing them automatically. This strategy overcome the sticking produced by the noise such as hair. The techniques perform well in the presence of various artifacts and also in the presence of variation in structure, weak boundary strength, color, and structure. The borders are recreated by following bordor smoothing, calculation of spectral difference modification in otsu threshold and inlet removal.

Chun-yan yu *et al* [55] proposed an active contour model that is based on level set method. To get image contour, it uses region information. They applied distance regularization so that the level set deviation from signed distance function can be penalized. They have used synthetic and real image data for the experiment purpose. The results shows the usefulness of the proposed model. The summary of these skin lesion traditional segmentation techniques is given in Table 3.2.

Table 3.2: Traditionsl skin cancer segmentation techniques

Sr. No.	Reference	Method	Dataset	Results
1	[48]	hybrid of K-Mean and firefly algorithm (FFA)	ISIC 2017, PH2	Accuracy = 99.1% on ISIC dataset and 98.9% on PH2 dataset.
2	[49]	Fuzzy C-Mean for segmentation. Features extracted by LVP and LBP. For classification, used DOROA algorithm.	PH2	Acc = 0 87 SE = 0.43, SP = 0.99, PR = 0.99, F1-Score = 0.60
3	[50]	YOLO and Grabcut	ISIC 2017, PH2	PH2: Acc = 94.4, IoU = 90 ISIC 2017: Acc = 96, IoU = 86
4	[51]	Image registration approach	ISIC 2017	IoU = 0 76, Acc = 0.95
5	[52]	Threefold method: introduced a weighting criteria to choose the best solution based on extended feature analysis	PH2, ISIC 2016	PH2: Acc = 92.54, ISIC 2016: Acc = 91.53
6	[53]	Fifteen thresholding methods	Clinical dataset	Error rates calculated

Sr. No.	Reference	Method	Dataset	Results
7	[54]	Used geodesic active contour	Clinical dataset	Median XOR border error = 6.7%
8	[55]	Active contour model based on level set method.	Synthetic dataset	IoU = 94.45%

3.3 CNN based Skin Cancer Segmentation and Classification Techniques

Skin cancer classification involves Machine Learning as well as Deep Learning techniques [29][47][55 - 69]. Deep Learning has become much more popular in the last few years with the evolution of increasing computational power (GPU) and the availability of large image datasets. They have made their own attractive place and almost replaced the traditional Machine Learning techniques for the classification tasks. Now, Deep Learning techniques have numerous applications in the medical imaging especially in skin lesions or skin cancer classification tasks. In this section, we review the application of standard CNN architectures for the classification of skin lesion types such as how these pre-trained models have been utilized to improve the classification of skin lesion types. In typical CAD systems, classification is the last step for skin lesion diagnosis. To fulfill the task of skin lesion classification, several deep learning methods have been proposed in literature [29][47] [55 - 69]. In the following, we have given a brief review of the existing deep learning methods for skin lesions classification. The summary of these techniques is given in Table 3.3.

One of the earliest work that used deep learning method for skin lesion classification was given by Masood *et al.* [56]. They first applied the threshold-based algorithm to detect the skin lesions. After that, the features were extracted by applying some machine learning techniques. Deep Belief Networks (DBNs) and Support Vector Machine (SVM) are combined in the model to classify the skin lesions using extracted features. A hybrid CNN ensemble method [57] was developed by Mahbod et al. in literature. In this, for the classification of skin lesions, inter-architecture and intra-architecture networks fusions were combined. Networks of different architectures were fine-tuned multiple times by using their different settings. After that, from various sets of fine-tuned networks the results were combined. By doing so, their proposed method got better results on the ISIC 2017 dataset for classification task. The advantage of this fusion method was that it does not need general pre-processing, lesion area segmentation or any kind of additional training data. Fabio

Perez et al [58] evaluated 9 CNN standard architectures for the classification of melanoma. All these nine architectures were applied on five set of splits created on ISIC 2017 dataset. From this work it was found that, ensemble of more than one or two CNN models could perform well as compared to the individual one for the classification of melanoma. Four distinct CNN architectures such as VGG 16, VGG 19, Inception V3 and MobileNet were applied with transfer learning using dilated convolution by Ratul *et al* in [59] to classify malignant skin lesions. From these four architectures, Inception V3 gave higher classification accuracy along with precision, recall and F-1 score. They used HAM10000 skin cancer dataset for training, validation, and testing of the architectures. Codella N *et al* [60] proposed a method for melanoma classification in dermoscopy images by a combination of sparse coding, deep learning, and SVM. This method reduces the need of labelled data. They have used ISIC dataset to evaluate the performance of proposed method. The CNN was trained on natural photographs to get lesions feature descriptors, and then combined with the representation of sparse coding. An ensemble of CNN models such as ResNext, SENet and EfficientNet was used by Gessert *et al* [61] by applying search strategy. They combine meta data for example age, gender etc. using dense neural network. After that, their features are fused together with the CNN. At the end, all these models were combined utilizing some ensembling approach to get optimal models' subset. Due to the limited availability of skin cancer datasets, transfer learning approach adopted in various existing works for skin lesions classification task. One of that kind of works was proposed by Liao [62]. They employed three pretrained networks including VGG 16, VGG 19 and GoogleNet using transfer learning approach. These network models trained on dermnet dataset. There is one more approach proposed by the same author Liao *et al* in [63]. They employed pretrained AlexNet model using transfer learning approach for the lesion targeted and disease targeted classification. For lesion targeted classification, they utilized multi-label image classifier. Similarly, for the disease targeted classification, multi-class image classifier was trained. Codella, N *et al* [64] combined deep residual networks, CNN, U-Net architecture along with sparse coding and hand coded feature representation for melanoma classification. A pretrained AlexNet architecture using transfer learning approach was also employed by Hosny *et al* [65]. The authors have applied data augmentation on the dataset using transformations, fine-tuned the architectural weights, and replaced classification layer with soft-max layer. To assess the performance of AlexNet model, the authors have utilized three datasets

including DermQuest, ISIC, and MED-NODE. Similarly, M.A. Kassem *et al.* [30] also applied a pretrained GoogleNet (Inception V1) by using transfer learning. ISIC 2019 dataset was used for the model evaluation. Their proposed method can classify eight different classes of skin lesions. For evaluation of the proposed model, sensitivity, accuracy, specificity, and precision are used as performance measures. GoogleNet and AlexNet architectures using transfer learning have also been used by various other authors [66]. Alqudah *et al* in [66] employed AlexNet and GoogleNet architectures along with transfer learning and (Adaptive Momentum Learning Rate) ADAM as an optimizer, for the skin lesions images classification. Their proposed model was tested on ISIC dataset having three classes such as seborrheic keratosis, melanoma and benign. A pretrained Inception V3 (GoogleNet) model was applied using transfer learning approach by Zhang et al [41] for dermoscopic image classification of skin lesions into four classes. For model evaluation, they used a private dataset collected from hospital. Their results showed that the model performed very well on skin lesions diagnosis. Similarly, Inception V3 model was also used by De Vries T et al [67] by fine-tuning it for skin lesions classification utilizing two distinct scales of resolution of input images. These two scales are coarse scale and finer scale. In this, the coarse scale represents the general shape and context of the skin lesion. On the other side at the finer scale, textural detail and low-level properties of lesions are exposed that played their important role to distinguish between skin lesion classes. To evaluate model performance, ISIC 2017 dataset was used. Akram et al [68] introduced a deep learning framework to integrate deep features information in order to get discriminant feature vector to differentiate among skin lesion classes. Entropy controlled neighbourhood component analysis was used to choose discriminant features and dimensionality reduction. The authors have used Inception-ResNet V2, DenseNet-201, and Inception V3 models and their selected layers as classifiers. The proposed technique was assessed using four different skin cancer datasets including ISBI 2017, ISIC MSK, ISIC UDA and PH2 datasets. There are also some more applications of deep learning techniques exist in literature. One of these approach is based on VGG-16 architecture for melanoma classification [69]. In this, the model was trained in three diverse ways: First, the model was trained from scratch. Second, the transfer learning approach was applied to get features from VGG-16 model. At the last, transfer learning and fine-tuning both are applied on the network. The proposed model performed well on skin cancer classification task.

Table 3.3: Summary of CNN-based skin cancer classification techniques

Sr. No.	1	2	3	4
Reference	[56]	[57]	[58]	[59]
Authors	Masood *et al*	Mahbod *et al*	Perez *et al.*	Ratul *et al*
Year	2015	2019	2019	2019
Dataset	Clinical database of dermoscopic of the images came from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital.	ISIC 2017	ISIC 2017	ISIC 2018
No. of Classes		03	03	07
Image Type	Dermoscopic	Dermoscopic	Dermoscopic	Dermoscopic
No. of Images	290	Train images = 2000 Test images 600	Train images = 2000 Test images 600	Train images = 8011 Test images = 1002 Validation images = 1002
Technique Developed	Semi-supervised learning model for automated recognition of melanoma. Deep Belief Network is developed using the combination of labelled and unlabelled instances. SVM is used for classification.	Ensemble of CNNs of different architectures. The technique consists of various sets of CNNs of different architectures to denote levels of different feature abstraction.	Created an ensemble of 9 CNN architectures on five set of splits of dataset for skin lesion classification.	They have used four CNN models named, VGG 16, VGG 19, MobileNet, and Inception V3 with transfer learning to implement dilated convolution.
Performance	Accuracy = 89%	Accuracy = 87.7%		Top 1 accuracy: VGG 16 = 87.42% VGG 19 = 85.02% MobileNet = 88.22% Inception V3 = 89.81%

7	6	5	Sr. No.
[62]	[61]	[60]	**Reference**
Liao *et al*	Gessert *et al*	Codella *et al*	**Authors**
2016	2019	2015	**Year**
DermNet, OLE	ISIC 2019	ISIC 2016	**Dataset**
	08	02	**No. of Classes**
	Dermoscopic	Dermoscopic	**Image Type**
Train images = 23,000	Train images = 25,000	Total images = 2624	**No. of Images**
They employed three pretrained networks including VGG 16, VGG 19 and GoogleNet using transfer learning approach.	An ensemble of CNN models such as ResNext, SENet and EfficientNet was used by applying search strategy. They combined metadata using dense neural network and then fused their features together All these models combined by some ensembling approach to get optimal model subset.	The method proposed by combining deep learning, sparse coding and SVM algorithms	**Technique Developed**
Dermnet: Top 1 accuracy = 73.1% Top 5 accuracy = 91%	Balanced Accuracy = 74.2%	Accuracy for first task = 93.1% Accuracy for second task =73.9%	**Performance**

11	10	9	8	Sr. No.
[30]	[65]	[64]	[63]	**Reference**
Kassem *et al*	Hosney *et al*	Codella *et al*	Liao *et al*	**Authors**
2020	2019	2017	2016	**Year**
ISIC 2019	MED-NODE, Derm (IS & Quest) and ISIC 2017	ISIC 2016	Data collected from various sources including: AtlasDerm, Danderm, Derma, DermIS,DermNet, DermQuest	**Dataset**
08	MED-NODE & Derm = 02 ISIC 2017 = 03	02		**No. of Classes**
Dermoscopic	Dermoscopic	Dermoscopic		**Image Type**
Train images = 25,000		Total images = 2624	Total images = 75,665	**No. of Images**
They applied a pretrained GoogleNet by using transfer learning.	Data augmentation and transfer learning applied using AlexNet model.	They combined deep residual networks, CNN, U-Net architecture along with sparse coding and hand coded feature representation for melanoma classification	They applied pretrained AlexNet model using transfer learning approach for the lesion targeted and disease targeted classification. For lesion targeted classification, they utilized multi-label image classifier.	**Technique Developed**
Accuracy =94.92% Sensitivity =79.8% Specificity =97% Precision =80.36%	Accuracy = MED-NODE = 96.86% Derm (IS & Quest) = 97.7% ISIC 2017 = 95.51%	Accuracy = 76% Specificity = 62% Sensitivity = 82%	Top 1 Accuracy = 27.6% Top 5 Accuracy = 57.9%	**Performance**

Sr. No.	12	13	14	15
Reference	[66]	[67]	[68]	[69]
Authors	Alqudah *et al*	De Vries *et al*	Akram *et al*	Romero *et al*
Year	2019	2017	2020	2017
Dataset	ISIC 2017	ISIC 2017	ISIC 2017, PH2, ISIC MSK, ISIC UDA	ISIC 2016
No. of Classes	03	03	ISIC 2017 = 03 PH2 = 02	02
Image Type	Dermoscopic	Dermoscopic	Dermoscopic	Dermoscopic
No. of Images	Train images = 2000 Test images 600	Train images = 2000 Test images 600	PH2 Total images = 200	Total images = 2624
Technique Developed	AlexNet and GoogleNet architectures along with transfer learning and ADAM as an optimizer.	Inception V3 model was used by fine-tuning it for skin lesions classification utilizing two distinct scales of resolution of input images. These two scales are coarse scale and finer scale.	They introduced a deep learning framework to integrate deep features information to get discrimunant feature vector to differentiate among skin lesion classes.	A pre-trained model VGG-16 is used for skin lesion classification.
Performance	Accuracy = 92.2%	Accuracy = 90.3%	Accuracy ISIC 2017= 95 9% ISIC UDA= 97.1% ISIC MSK= 99 2% PH2= 98 8%	Sensitivity= 78.66%

In literature, there have been proposed various deep learning-based skin lesion segmentation methods. A method proposed in [70] optimize the initial contour using genetic algorithm without

the edge detection for detection of the skin lesion boundary. To perform segmentation in dermoscopic samples, a fully convolutional encoder-decoder model was optimized using exponential neighbourhood grey wolf optimization algorithm [71]. The researcher proposed a novel CNN-based architecture [72] using end-to-end atrous spatial pyramid pooling for segmentation of the lesion.

One more system is proposed to segments skin lesions using Retina-DeepLab, graph-based techniques, and Mask R-CNN [73]. A dense encoder-decoder-based framework is proposed in which the combination of ResNet and DenseNet is utilized for segmentation improvement. Further, ASPP is employed to get multiscale contextual information and skip connections are used to recover the information [74]. An automated lesion segmentation using an adaptive dual attention component with three characteristics is proposed in [75]. Its first property is two global context modelling schemes that are integrated with ADAM. The second property is to support the multi-scale fusion for good segmentation. The third useful property is to use spatial information weighted technique to lower redundancies [75].

One more effective approach [76] based on enhanced fully convolutional network (iFCN) segments the skin lesion without pre-processing or post-processing the skin lesion image. It contributes to the identification of the centre location of the lesion and clear the details on the edges by eliminating the unwanted effects [76]. Another technique is proposed that automatically segments the skin lesion and introduces a novel segmentation approach named FC-DPN made with the combination of a fully convolutional and dual path network [77]. In literature, an attentive border-aware system is proposed [78] for segmentation of multi-scale lesions through adversarial schooling. It consist of various sections, including ResNet34 as the encoding, and decoder, skip connection based on Scale Att-ASPP and PPM at the peak of the last convolutional layer in the encoding path [78]. The Mask R-CNN-based technique was proposed in [79] for skin lesion segmentation. This technique consists of two parts, creating the bounding boxes of candidate object with RPN and Fast R-CNN classifier and a branch of binary mask prediction [80]. The authors in [81] proposed a novel method of lesion segmentation by proposing the fusion of YOLOv3 and the GrabCut algorithm [81]. Another lesion segmentation approach such as encoder-decoder model is proposed based on Pyramid Scene Parsing Network. It makes use of pyramid pooling blocks, and a skip connection which can look for lost spatial details and accumulate the

global context [82]. A combination of DeepLabv3+ and Mask R-CNN was proposed to enhance the performance of segmentation model to locate the skin lesion accurately [83]. In literature, the authors in [84] also described the relevancy of deep learning models by employing a pre-trained VGG16 model as an encoder and combined it with DeeplabV3, Seg- Net decoder, and TernausNet to perform the segmentation of skin lesion [84].

A deep learning technique to improve the major task of skin lesion segmentation is applied by using 46 layered U-net model to get a good lesion segmentation performance [85]. An approach based on dense deconvolutional framework was proposed to deal with the challenges of different sizes and the appearance of skin lesions [86]. In this, the deconvolutional layers are utilized to unchanged the dimension of input/output. The chained residual pooling extracts the contextual information and then fuses multi-level features. To improve the prediction masks and reduce the auxiliary loss, a hierarchical supervision is added [86].

In Table 3.4, a summary of the segmentation models, as discussed above, is provided. It consists of models, datasets, and the result of the experiment based on accuracy and Jaccard.

3.4 Skin Cancer Hierarchical Classification and Structural Segmentation Techniques

The method for hierarchical classification using deep learning technique is proposed in the Hierarchical Deep CNN (HD-CNN) [87]. In this, Deep Neural Network (DNN) is trained as an n-way classifiers, by assuming that classes have flat relations to each other. Although, few of the classes are much confusing as compared to other classes. For example, it is very easy to differentiate apple from bus as compared to apple from orange. It first uses CNN to distinguish easily distinguishable classes. After that, it separated the fine classes.

Table 3.4: CNN based skin cancer segmentation techniques.

Reference	Year	Methods	Datasets	Results
[70]	2021	GA, Initial contour optimization, and Dull razor,	ISIC 2016	0.83 JACC
[71]	2022	EN-GWO, FCEDN	ISIC 2016 17	0. 96 JACC 0.87 JACC
[72]	2021	Downsampling, augmentation, Atrous dilation CNN	ISIC 2016 17 18	0 9 JACC 0 82 JACC 0.89 JACC
[73]	2021	R-CNN, Retina-DeepLab, graph-related approach	ISBI 2017 PH2 DermQuest	0 94 JACC 0.90 ACC 0 99 ACC
[74]	2021	Data augmentation, encoder-decoder framework	ISIC 2018	0.97 ACC

Reference	Year	Methods	Datasets	Results
[75]	2020	Bicubic interpolation, Adaptive dual attention Module, data augmentation	ISBI 2017 ISIC 2018	0.96 ACC 0.95 ACC
[76]	2020	IFCN	PH2 ISBI 2017	0 97 ACC 0.95 ACC
[77]	2020	Augmentation, FC-DPN	ISBI 2017 PH2	0 80 JACC 0.84 JACC
[78]	2020	augmentation, PPM, GAN, ResNet34, Scale-Att-ASPP	ISBI 2016 ISBI 2017 PH2	0.96 ACC 0.93 ACC 0.11 DV
[80]	2020	R-CNN	ISIC	0.91 recall
[81]	2019	morphological operations, YOLO, Dull razor, GrabCut	PH2 ISBI 2017	0.93ACC 0.93 ACC
[82]	2019	encoder-decoder model with multi-resolution skip connections	ISIC 2018	0.84 JACC
[83]	2019	DeeplabV3+, and Mask R-CNN, Morphological operations.	ISIC 2017 PH2	0.79 JACC 0.84 JACC
[84]	2019	Data augmentation, DeeplabV3, SegNet, VGG16 encoder	ISIC 2018	0.88 JACC
[85]	2019	Linear filter, U-Net 46 layered, U-Net 32 layered	ISIC 2018	0.93 JACC
[86]	2018	Dense CNN	ISBI 2016 ISBI 2017	0.96 ACC 0.94 ACC

Component-wise pretraining is performed in it. After that, global fine-tuning is performed along multinomial logistic loss regularized by coarse category consistency. The extended version [88] of HD-CNN creates three different HD-CNN models. Different combinations of layers are used in these three HD-CNN to compare their performance. This model has the hierarchical structure in CNN achieving desirable results. Firstly, the model pays attention to classifying the coarse categories and then focuses on categorizing the fine classes falling in the same coarse categories. The training complexity of the model was not increased to implement this model. Two training steps are involved in it including pre-training of coarse and fine categories and then fine-tuning of the model. This model has some limitations such as it cannot perform multiple hierarchy levels. This limitation leads to the development of Branch-CNN (B-CNN) [89]. Its concept is based on the fact that the initial layers in CNN learn the low-level features of an image and last (higher) layers learn high-level features. This B-CNN architecture produces various predictions such as from origin level to fine level, which correspond to the hierarchical structure of the class. In this, the authors have proposed Branch Training Strategy (BTS) for B-CNN training. This strategy is helpful to reduce the loss by adjusting parameters at the output layers. By following the proposed

structure of the model and training strategy, it will first learn at coarse level and then to fine level features along layer's blocks. In B-CNN, the errors are bounded to sub-category, either of first level, second level, or the fine level classes. The method hierarchical CNN (H-CNN) [90] is proposed in literature for the fashion image data classification. The H-CNN structure is applied on VGG-16 and VGG-19 both. This model produces hierarchical level detail of input image by creating the model in a way so that it can produce labels sequentially from low level to higher level.

There are two methods [28][29] exist in literature for explainable classification of skin lesions. These methods solve the problem for two levels of hierarchical classification. They have used encoders such as DenseNet-161, VGG-16, and ResNet-50 separately with decoder LSTM to perform two level classification. They have not dealt with the problem of class imbalance present in ISIC dataset. They have used ISIC 2017 and ISIC 2018 dataset having two classes and seven classes in them respectively. Further, they have not shown the results in the form of evaluation measures such as sensitivity, precision, specificity, and F1-score for the hierarchical level-1 and level-2 classification [28][29]. The results are reported only for fine-level classification. For hierarchy, they have reported the results by using gradient class activation maps (grad-CAMs). One more method is introduced by Benyahia *et al.* [91] for hierarchical classification of skin lesion using CNN models. In this method, they have used five independent CNN models arranged in a hierarchical way. Each CNN performs individually to decide a specific type of skin lesion. Experiments are performed on ISIC 2019 dataset to assess the function of their suggested algorithm.

It is to honestly state that there does not exist any further method in literature for hierarchical classification of skin cancer lesions. The above discussed techniques [87 - 90] have not been used for skin cancer datasets. There is a great need to do explainable classification of skin lesions to make the system self-explainable. The summary of these techniques is given in Table 3.5.

Table 3.5: Summary of existing hierarchical classification techniques

Sr. No.	Reference	Methods	Datasets	Results
1	[87]	HD-CNN	CIFAR100, ImageNet	Top-1 error: 2.65%, 3.1%,
2	[88]	Extended version of HD-CNN	CIFAR100, ImageNet	Top-1 error: 1.1%

Sr. No.	Reference	Methods	Datasets	Results
3	[89]	Branch (B-CNN)	MNIST, CIFAR10, CIFAR 100	MNIST: 99.4% CIFAR10: 84.4% CIFAR100: 64.42%
4	[90]	Hierarchical (HCNN) using VGG-16, VGG-19	MNIST image dataset	VGG-16: 93.5 VGG19: 93.3
5	[28]	DenseNet151, ResNet-inception, Attention module, LSTM	ISIC 2017, ISIC 2018	2017 SE=71.9, SP=86.5, AUC=87 9 2018: SE=78, SP=97, AUC=80.2
6	[29]	VGG16, DenseNet161, ResNet50, Attention module, LSTM	ISIC 2018	SE=69.0, SP=96.8, AUC=95.2
7	[91]	Hierarchy of five independent CNN models (using DenseNet201, ResNet50)	ISIC 2019	ResNet50: Acc = 77.6 SE = 76.6, SP = 95.8, PR = 66.3 DenseNet201, Acc = 79.4, SE=76.3, SP=95.8, 69.1

As came under discussion in Section 3.2 and 3.3 boundary level segmentation of skin lesions has been explored extensively, nevertheless the literature for the structural segmentation of skin lesions is comparatively limited. Nguyen D et al [92] suggested TATL a novel transfer learning approach for skin detecting traits and through extensive experiments conducted on the ISIC 2018 and ISIC 2017 datasets, after which the effectiveness of TATL was validated. It is noteworthy that their suggested model overtook the prevailing methods while applying merely $1/30^{th}$ of parameters compared to the ISIC 2018 competition champion. The TATL approach came up with remarkable improvements in diagnosing methods of skin lesion, already pre-trained on ImageNet making it specifically efficient for structures with limited training samples. TATL (Task Agnostic Transfer Learning) is stimulated by the dermatologists' behavior in the skincare perspectives, employing an attribute-agnostic segmenter for detection of skin lesion regions. The gained intelligence is then transferred to the classifiers qualified for specific attribute detection for enhancement of the diagnostic process. The approach of attribute-agnostic segmenter of TATL by exclusively concentrating on the identification of skin attribute regions allows it to collect abundant data from all structures, facilitating data transfer among various features and mitigating the issue of limited training data existing for rare attributes.

The other approach suggested by Kadir M. et al [93] for structural segmentation of skin lesions that involves fine-tuning CNN provides user feedback on two grounds simultaneously: the classification of lesion and visual explanation which accompanies the classification. In their research they introduced a new CNN architecture that integrates the Grad-CAM technique which provides visual description of the models' results during the process. In their approach they examined the way of response of CNNs to this dual feed mechanism and found that this method brought about improvement both in interpretability and accuracy of the model, hence providing better understanding of the decision-making process of the network.

By using feedback of the user, the authors came up with the finding that fine-tuning their model based on both classification and visual description heads to ameliorate visual explanations without compromising accuracy of classification. Such an advancement has the potential to enhance the user confidence in CNN-based lesion classifiers, as it enables the user to better understand and interpretation of the model's decision through the provided visual explanations.

In Task 2 of the ISIC 2018 Challenge, Nguyen D et al [94], introduced another innovative approach for automated prediction of dermoscopic structures. This approach brings in use the Attention U-Net model and multi-scale images as input. For enhancement of the model' performance, the authors have implemented a transfer learning strategy by undertaking the adjustment of the weights of a pre-trained network utilized for segmentation to train the deep neural network for attributes' extraction. Their suggested algorithm resulted in better performance than that of LeHealth [95] and NMN [101].

In the series of developments of attribute level segmentation algorithms, Jahanifar M et al [96] proposed a segmentation model which by using convolutional neutral networks (CNNs) integrated transfer learning to segment lesions and their structures. The model dealt with an architecture of encoder-decoder supporting different pre-trained models in the encoding process, hence created projected maps by blending wide-scale information in decoding process via pyramid pooling approach. To resolve the issue of limited training data and to increase the model' generalization, a complete package of innovative domain-centered augmentation techniques was introduced. These algorithms identify the real differences surfaced in dermoscopy images. The mentioned algorithms enjoyed top of the list position on the leaderboard for ISIC 2018 attribute detection task. Another significant deep learning technique named as the superpixel attention network (SANet) was

introduced [97] in this domain in which the input images are classified into small sections, then reordered by applying the random shuffle mechanism (RSM) and finally, the SANet is employed to collect discrete features and recreate the input images. The given model is composed of two different units called as superpixel average pooling module and superpixel attention module in such a way that the first one employed to restructure the issue of superpixel classification as a problem of superpixel segmentation.

The first one is employed to reframe the issue of superpixel classification to make it as superpixel segmentation problem. To get discriminative superpixel regions and feature channels, the superpixel attention module (SAM) is used. The authors developed global balancing loss as a loss function to resolve the issue of class imbalance in ISIC 2018 Task 2 dataset. their method shows better performance on the said dataset.

Additionally, a new technique [98] was suggested by utilizing collection of deep learning models for segmentation of skin lesion attributes. The researchers [99] developed a complete solution for detection of dermoscopic attributes by using the ISIC competition dataset. They discovered the ability to enhance performance by integrating the descriptions/labels of different modalities to be segmentation masks. To do so, a suggestion was presented regarding the leveraging of the task of segmentation to be conducive and the sharing of information between the two tasks. The task was accomplished by training a Y-Net neutral network algorithm which proved helpful in facilitating wight sharing between segmentation and classification. Moreover, many techniques were employed for combination of the training of Y-Net applying labels of segmentation and classification.

In Kawahara J et al. [100] approach, the classification of clinical dermoscopic structures within the super-pixels is performed by dealing it as a segmentation problem. They applied an extensive convolutional neural network especially framed to detect dermoscopic structures in dermoscopy images. The design of their neural network was integrated maps of interceded features derived from different intermediate layers of the network. Fixing the problem of the imbalanced labels, they introduce a negative multilabel Dice-F1 score as a loss function, made for estimating and minimizing the score for each label throughout the minibatch.

Koohbanani N et al. [101] suggested an architecture to embody transfer learning into skin cancer lesions' segmentation and their attributes by applying CNNs. This structure draws insight from the

well-known and worldly recognized UNet architecture. Different pre-trained networks are leveraged in the encoding process, while pyramid pooling process is employed in decoding path to collect multi-scale knowledge and project the prediction map. To resolve the issue of inadequate availability of training data and boost the generalization of their model, various innovative augmentation techniques were applied while training the network. To deal with the challenges that appeared in the training phase the authors framed a loss function.

Nunnari F et al. [102] studied to examine the relationship between the graphic characteristics and regions documented by CNNs cultured for classification. By conducting tryouts of two varying neural network architectures with changing depth and resolution of the final convolutional layer, they measured the efficacy of thresholded Grad-CAM saliency maps in identifying graphical features related to skin cancer. Gonzalez Ivan [103] incorporated the skills of dermatologists into the commonly used structure of CNNs. Their technique focuses on framing multiple networks that includes recognition of lesion area, and segmentation of lesions into attribute structures. Moreover, the author has established novel blocks of CNNs to seamlessly link this data into the administering of diagnosis.

Another multi-task U-Net model was proposed in [104] for auto-detection of attributes of melanoma lesion. One of the two tasks it is designed for is classification for determination of the existence of lesion attributes, and the other is segmentation for identification and demarcation of attributes within the images. Using this multi-task model, the Jaccard index score gained was 0.43 on the dataset of the ISIC 2018 task 2, and the method got 5^{th} position in last leaderboard of ISIC 2018. Another approach proposed by Labraca J et al. [105] form improvement of diagnostic procedure for doctors. The approach supports soft segmentation of based on dermoscopic structure to raise a combination of classifiers specific to different dermoscopic structures in such a way that each classifier is focused on differentiating between benign lesions and melanomas grounded on a specific dermoscopic structure. Then by applying Bayesian method the results of all the individual classifiers are combined which not only extends the ultimate diagnosis but also provides supplementary knowledge to the doctor. The experts of skin lesions' specific attributes give their opinions and talk about their doubts. The model combined these opinions of experts, and their doubts related to diagnosis to further improve the diagnosis process. The summary of all these discussed techniques is elaborated in Table 3.6.

To summarize the series of developments in the structure segmentation of skin lesions in the light of the literature studied, it can be concluded that the available literature does not meet satisfaction by virtue of many unresolved challenges and issues. These challenges include insufficient annotated data, imbalanced datasets and presence of complicated structures exhibiting diverse appearances within each lesion.

To tackle the given challenges, several previous techniques have relied upon standard deep learning techniques and transfer learning algorithms which showed efficacy in alleviating data shortage problems. Still there are some explorable techniques like one-shot learning and zero-shot learning, which are not extensively accessed for structure segmentation in skin lesions.

Table 3.6: Summary of attribute level segmentation techniques for skin lesions

Sr. No.	Reference	Methods	Datasets	Results
1	[92]	Task Agnostic Transfer Learning (TATL) using U-Shape with b0-EfficientNet L-Shape with b0-EfficientNet	ISIC 2018 Task 2	**Mean IoU (U-Shape with b0-EfficientNet)** Pigment network: 0.565 Globules: 0.373 Milia: 0.157 Negative Network: 0.268 Streaks: 0.243 **Mean IoU (L-Shape with b0-EfficientNet)** Pigment network: 0.562 Globules: 0.356 Milia: 0.168 Negative Network: 0 292 Streaks: 0.252
2	[93]	User feedback based CNN model	ISIC 2018 Task 2	IoU = 0.13
3	[94]	Attention-UNet model using transfer learning	ISIC 2018 Task 2	**Mean IoU** Pigment network: 0.535 Globules: 0.312 Milia: 0.162 Negative Network: 0 187 Streaks: 0.197
4	[96]	Transfer learning based UNet model with multi-scale convolution (MSC) block, pyramid pooling paradigm	ISIC 2018 Task 2	**Mean IoU** Pigment network: 0.563 Globules: 0.341 Milia: 0.171 Negative Network: 0 228 Streaks: 0.156
5	[97]	Super pixel attention network including (superpixel average pooling and super pixel attention module)	ISIC 2018 Task 2	**Mean IoU** Pigment network: 0 576 Globules: 0.346 Milia: 0 251 Negative Network: 0 286 Streaks: 0.248
6	[98]	Ensemble of deep learning models	ISIC 2018 Task 2	Not reported

Sr. No.	Reference	Methods	Datasets	Results
7	[99]	Leveraged the task of segmentation and enabled sharing of knowledge between two tasks using Y-Net	ISIC 2018 Task 2	F1-Score: 0.502
8	[101]	UNet with pyramid pooling paradigm	ISIC 2018 Task 2	**Mean IoU** Pigment network: 0.544 Globules: 0.252 Milia: 0.165 Negative Network: 0.285 Streaks: 0.123
9	[102]	CNNs and Grad-CAMs	ISIC 2018 Task 2	IoU: 0.143
10	[103]	Multiple CNNs with novel blocks	ISIC 2017	Not reported
11	[104]	Multi-task UNet model	ISIC 2018 Task 2	Mean IoU: 0 433

3.5 Existing Research Limitations

In this chapter, the existing techniques for skin cancer segmentation and classification are discussed in detail. Existing methods have shown promising results in various skin cancer segmentation and classification tasks. However, there are several open issues and problems that need to be solved.

In existing literature, there is less attention on inherent hierarchical structure of skin lesion [47][55 - 68]. There is a lack of self-explainable CNN models in existing literature. Much more focus is given on binary classification and fine level classification. Skin lesions are divided into various types. These divisions are based on their origin, degree of malignancy and final diagnosis type.

Existing research has paid less attention to structural or attribute level segmentation of skin lesion[70 - 86]. Mostly work is done on the segmentation of boundaries of skin lesion but there is lack of work on the identification or localization of structures or features present in skin lesions.

Existing methods for skin lesion traditional segmentation and classification have shown promising results in the presence of class imbalance problems, exist in skin lesion dataset. These techniques handled class imbalance issues very well. However, these class imbalance handling techniques are giving unpleasant results when working in the scenario of structural segmentation of skin lesions.

The focus of this research is to propose self-explainable CNN based diagnostic model that can embed inherent hierarchical structure of skin lesion in it to self-explain its taxonomy. Besides, we have also focused on interpreting the structural segmentation of skin lesion in the presence of imbalanced dataset.

3.6 Problem Formulation

1) **Hierarchical Classification:** Let $D(X,Y)$ be a labeled skin image dataset where $X = \{x_1, x_2, x_3 \ldots\ldots x_n\}$ is a dataset containing n samples, and $Y = \{y_1, y_2, y_3, \ldots\ldots y_m\}$ contains label information of corresponding sample in X. A sample image x_i is represented by $[H \times W \times Ch]$ where H is height, W is width, and Ch is 3-channel RGB. From each class $j, where\ j = 0,1,2, \ldots\ldots 6$, a specific ratio is selected as train set, validation set, and test set. Let $X_{train}, X_{valid}, and\ X_{test}$ represents training, validation, and test sets. Let $HL_{L1}, HL_{L2}, and\ HL_{L3}$ represents the hierarchical labels of skin lesion classes at *level-1, level-2, and level-3* respectively, where $HL_{L1} = \{melanocytic, nonmelanocytic\}, and$ $HL_{L2} = \{benign, malignant, nbenign, nmalignant\}$
$HL_{L3} = \{nevus, melanoma, vascular, dermatofibroma, benign\ keratosis,$ $actinic\ keratosis, basal\ cell, squamous\ call\}$. Let the hierarchical labels mapping is as
$HL_{L3} \rightarrow HL_{L2} = \{0{:}\,0, 1{:}\,1,2{:}\,2,3{:}\,2,4{:}\,2,5{:}\,2,6{:}\,3,7{:}\,3\}$
$HL_{L2} \rightarrow HL_{L1} = \{0{:}\,0,1{:}\,0,2{:}\,1,3{:}\,1\}$. After that, the objective is to find the hierarchical labels or accurate taxonomy of the lesion such as
$melanocytic \rightarrow benign \rightarrow nevus$, and
$nonmelanocytic \rightarrow nmalignant \rightarrow squamous\ cell$, where $\rightarrow$ represents the relationship between the origin of the lesion, degree of malignancy, until a differential diagnosis is reached. The problem is broken into the following parts:
Let the function for feature extraction of individual class be denoted by Ψ and defined as $\left[\{O_i^j\}_{i=1}^m\right]_{j=1}^C = \Psi\left[x_i^{j\,m}{}_{i=1}\right]_{j=1}^C$. Here, Ψ represent the feature extraction model i.e., H-GoogleNet. For the recognition of origin, degree, and fine-level class, a classification function $g(.)$ is applied on extracted features as $g{:}\left[\{O_i^j\}_{i=1}^m\right]_{j=1}^C \rightarrow [\{l_i^j\}_{i=1}^m]$. It returns a discrete value $\{l_i^j\}_{i=1}^m$ consist of individual class labels of origin, degree, and fine-level class predicted by $g(.)$. A sequence of predicted taxonomic class labels is generated to see the predicted explainability of the model.

2) **Attribute Segmentation:** Let $D(X,Y)$ be a labeled skin image segmentation dataset where $X = \{x_1, x_2, x_3 \ldots\ldots x_n\}$ is a dataset containing n samples, and $Y = \{y_1, y_2, y_3, y_4, y_5\}$ contains label information of corresponding sample in X. Each image in dataset has five labeled masks

against it for the presence or absence of five attributes (represented by masks) in an image. A sample image x_i is represented by $[H \times W \times Ch]$ where H is height, W is width, and Ch is 3-channel RGB. Each attribute is taken as a class. So, there are five labeled classes i.e., pigment, negative, globules, streaks, and milia in dataset. Each class has several samples in it. The label of each class is given by mask. Each mask has the foreground as white pixels and background as black pixels. The segmentation will be performed on this mask of the image.

Chapter 4

Explainable H-GoogleNet Model for Skin Cancer Hierarchical Classification

Chapter 4

4 Explainable H-GoogleNet Model for Skin Cancer Hierarchical Classification

In this chapter a detailed description of proposed hierarchical deep learning-based model is provided.

4.1 Introduction

Classification of skin cancer lesions has always been a challenging and difficult step due to the presence of various dermoscopic structures as well as various categories of skin lesions. Particularly, the hierarchical levels of skin cancer types are challenging for clinical practitioners when they need to identify types and sub-types of the lesions. Numerous skin cancer classical classification techniques have been proposed in existing works. Literature in Chapter 3 revealed that traditional classification techniques cannot handle the challenges of hierarchical levels classification of skin lesions.

In this chapter, an explainable hierarchical H-GoogleNet model has been proposed for hierarchical level classification of skin lesions (see Figure 4.4). Three hierarchical levels of skin cancer types are classified using this model. Three more inception blocks are added before level-1 classification. Further, two modified auxiliary classifiers are added at two levels to get two outputs from a single model at level 1 and level 2. The model is trained using training data from ISIC 2018 dataset. After that, the experiments are performed on test dataset and the results of proposed model are also compared with the existing methodologies.

4.2 Motivation

This chapter presents the hierarchical H-GoogleNet classification model for classification of skin lesion at three levels. It is proposed to represent the multi levels of classification by adding multiple layers at two different levels to get three outputs from a single model. This proposed model, for three levels skin lesion classification, is different from explainable skin lesion diagnosis using taxonomies [29]. They used LSTM along with various CNN models (VGG-16, DenseNet-161, ResNet-50). CNN based models are used for feature extraction and LSTM used for sequential diagnosis of skin lesion using those features. This method is limited to two levels of classification. Further, there are no results shown for two-level classification using evaluation measures such as sensitivity, specificity, precision, AUC-ROC. The results are shown only for the fine (last) level classification that do not give support for hierarchical level

classification. Moreover, they have used multiple models for this purpose. On the other side, the strategy to introduce multiple levels of classification within a single model has been proved to show better results for hierarchical classification [90]. This thesis proposed to use GoogleNet model to embed three levels of classification in it. The three inception blocks and more layers are added to perform hierarchical level classification. The proposed model is not affected by any kind of noise present in the image.

4.3 The Architecture of CNN

CNN is a basic model in deep learning. In this, every layer is interconnected to produce feedback. Its architecture has several layers to extract both local and global information from every image. A traditional CNN architecture has various layers including convolution, activation, pooling, fully connected, and softmax layer. Batch normalization and dropout layers are also there to resolve the issue of overfitting and generalization. The CNN architecture takes an input image in the input layer (first layer). This layer takes input as $H \times W \times ch$ where ch represents the number of channels. The convolutional layer is the second layer in which the features are extracted from the given input image by using the operation of convolution using dot product. This layer is calculated as given in equation (4.1).

$$O_{x,y} = \sum_{i=0}^{r} \sum_{j=0}^{r} w_{i,j} z_{i,j} \tag{4.1}$$

Where $O_{x,y}$ represents the output of the convolutional layer and $w_{i,j}$ represents the weights and $z_{i,j}$ denotes the input pixel. This layer outputs many positive and negative values of pixels. To fix the handling of negative values, an activation layer is added to the network. This layer may use any activation functions including *sigmoid, ReLU,* and *tanh.* After convolution layer, there is a pooling layer to reduce the spatial dimension of the image. The size of the filter and its stride are the two key parameters in a pooling layer. Mathematically, the output of pooling layer is calculated as given in equation (4.2).

$$I_{output\_size} = \frac{\left(I_{input\_size} - K_{kernel_{size}} + 2P_{padding}\right)}{S_{stride}} + 1 \tag{4.2}$$

Where I_{output_size} is the size (width and height) of the output image, $K_{kernel_{size}}$ is the size of the filter, and S_{stride} is the number of strides. In the fully connected FC layer, the processed features are transformed to a 1-dimensional array. This layer extracts features for classification.

4.4 Techniques used in Proposed Model

The following techniques are used in the proposed H-GoogleNet model.

4.4.1 GoogleNet Architecture

A simple way of producing powerful deep learning models is to add more layers and neurons to make it deep. The deeper the model the more powerful the architecture, but it has its own problems. In VGG-19 and various other CNN models, deeper layers demand more resources and are prone to overfitting in the case of hierarchical level classification. The architecture of GoogleNet [106] is different in a way that it is composed of 'sparsely connected neural network' instead of 'fully connected neural network'. The sparsely connected neural network architecture is different in a way it works efficiently in limited computational resources as the network width and depth increases. The architecture of GoogleNet is described in Table 4.1. The architecture of inception block in GoogleNet model is shown in Figure 4.1.

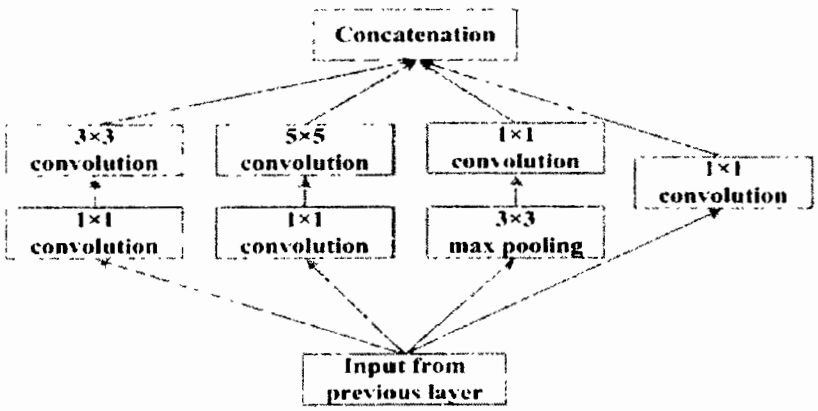


Figure 4.1: Architecture of inception block in GoogleNet model [106]

Table 4.1: Details of layers in GoogleNet architecture [106]

type	patch size / stride	output size	depth	#1x1	#3x3 reduce	#3x3	#5x5 reduce	#5x5	pool proj	params	ops
convolution	7 x 7/2	112 x 112 x 64	1							2.7K	34M
max pool	3 x 3/2	56 x 56 x 64	0								
convolution	3 x 3/1	56 x 56 x 192	2		64	192				112K	360M
max pool	3 x 3/2	28 x 28 x 192	0								
inception (3a)		28 x 28 x 256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28 x 28 x 480	2	128	128	192	32	96	64	380K	304M
max pool	3x3/2	14 x 14 x 480	0								
inception (4a)		14 x 14 x 512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14 x 14 x 512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14 x 14 x 512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14 x 14 x 528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14 x 14 x 832	2	256	160	320	32	128	128	840K	170M
max pool	3x3/2	7 x 7 x 832	0								
inception (5a)		7 x 7 x 832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7 x 7 x 1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7x7/1	1 x 1 x 1024	0								
dropout (40%)		1 x 1 x 1024	0								
linear		1 x 1 x 1000	1							1000K	1M
softmax		1 x 1 x 1000	0								

The input shape given to the GoogleNet is (224, 224). In Table 4.1, the patch size is the rectangular sweeping window that is used by convolutional layers and pooling layers. Stride refers to the size by which pixels shift over input image. In this, the output size is the output shape that is generated after the input is passed from a layer. Depth of the network defines the levels in the architectural component. 1x1, 3x3 and 5x5 are the convolutional filters utilized in inception module of architecture. On the other side, '3x3 reduce' and '5x5 reduce' represent the filters of 1x1 size used in network before convolutional layers. In Table 4.1, 'Pool proj' represents the number of filters of size 1x1 that are used after the pooling layers inside the inception module of network. 'Params' are the number of weights that are used in architecture components. 'Ops' represents the number of mathematical operations conducted in a component of architecture. Googlenet architecture is designed to decrease the input image size but to retain the 'spatial' information.

The first layer in the network uses patch size of 7x7. The key purpose of this layer is to reduce the image size by retaining spatial information. The max pooling layer is added after one convolutional layer. At the second convolutional layer, the image size is reduced by the factor of four. After passing from the second maxpooling layer, the input size is reduced by the factor of eight. Several feature maps are produced by this time. The second convolutional layer leverages the depth of 2 and has a 1x1 convolutional block. This 1x1 convolutional block reduces the dimensions. This dimensionality reduction reduces the computational load as the number of operations of layers are reduced. There are nine inceptions modules that are introduced in GoogleNet architecture as shown in Figure 4.2. The two maxpooling layers are also added in between the inception module. The purpose of adding maxpooling layers is to reduce the input size. This also contributes to reducing the computational load. An inception module constitutes the following components:

Input layer — 1x1 Conv — 3x3 Conv — 5x5 Conv — 1Maxpool — concatenation layer

The 1x1 convolutional layer reduces the dimensions of input. It calculates the element-wise product of all values in the image. This layer does not learn spatial features of the input image, but it learns the channel pattern. So, this layer fulfils our two objectives: first, it reduces the computational load and second, it learns pattern across depth or channel. The 3x3 and 5x5 filter sizes of the convolutional layer learn 'spatial' patterns across depth, height, and width. The concatenation layer concatenates the feature maps with outputs and produces the single output of inception module. After the inception modules, average pooling layers and final dense layers are added. A dropout of 40% is also added to avoid overfitting.

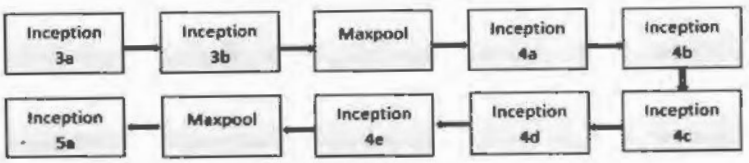


Figure 4.2: Inception modules introduced in GoogleNet architecture [106]

Another major contribution GoogleNet architecture made is to include auxiliary classifiers. Auxiliary classifiers are added to overcome vanishing gradient problem. This problem arises when the network is extensive, there are chances that the gradient value produced is very small and thus the weights are not updated. It means that the network is not learning during training.

4.4.2 Data Augmentation

In the ISIC 2018 datasets, the fewest examples are from the vascular class, and these are 100 in total. The training of a deep learning model with 100 examples is not an appropriate approach. In our proposed work, data augmentation technique is applied to up-sample the low class and down-sample the high class. We have applied reflection, translation, rotate to augment our dataset, as shown in Figure 4.3, to do oversampling of minority class and under-sampling of majority class. By doing so, we can achieve an equal number of instances for each class to get the generalized performance of proposed model. The proposed model generates three level outputs, so the size of model becomes huge. Due to resource limitations, we are not able to explore the complete ISIC 2018 dataset. The ISIC 2018 dataset contains 10,015 samples and seven classes. We have only sampled 7000 examples from each due to resource limitations in hand and multiple hierarchical levels.

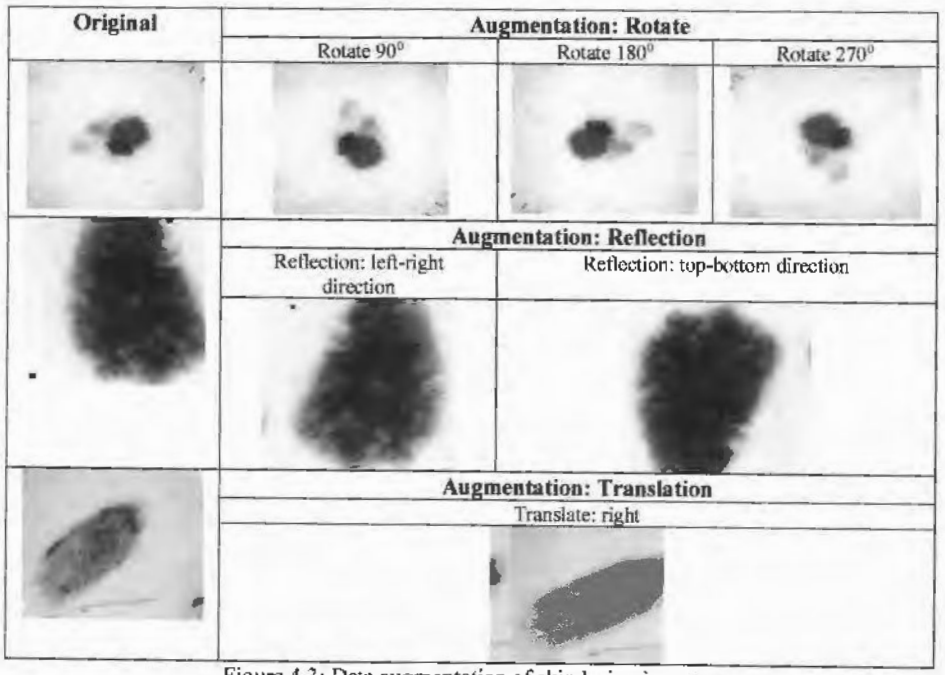


Figure 4.3: Data augmentation of skin lesion images.

4.4.3 Gradient Class Activation Maps

In machine learning as well as deep learning based models, the reliability is questioned because of the interpretability issues. Machine learning models are considered as black box so, nobody can tell why a model has reached to such a problem. This explainability issue get more serious in medical imaging. Physicians or practitioners are clueless over the model decisions. Due to this, it is significant to enhance the reliability of the deep learning models to gain the confidence of the practitioners in finding the accurate results with details along with visualization.

We cannot tell how a decision is made by a specific model, but we are able to answer "why" a certain decision is made by that model. For this purpose, visualization methods have been introduced that plot values computed by the model on the image to localize a certain region. This highlighted region explains the reason of making a certain decision. To enhance the model reliability, we have used the gradient class activation maps in our proposed work to increase the explainability of the model. It highlights a particular region to explain the reason of a particular decision. This is very helpful in terms of reliability in a way that a physician now will know that why the model has predicted a certain class.

4.5 Proposed Model Overview

In this section we have given an overview of the proposed H-GoogleNet architecture for three level hierarchical classification of skin lesion. The proposed model uses GoogleNet model which is modified to work for the hierarchical level classification purpose. The proposed architecture uses auxiliary classifiers to get the output at multiple levels.

The proposed model is shown in Figure 4.4. The complete detail of newly added layers is shown in Annex-I. In our proposed architecture, we have added multiple layers to the standard architecture of GoogleNet. Our proposed model takes $Image_{input} = Height_{224} \times Width_{224} \times Channel_3$ in the input layer. There are no changes made to the standard inception block of original Googlenet. The details of the inception module are given in section 4.4. After input layer, there is a convolutional layer, Maxpooling layer, 2 more Conv2D layers, maxpooling layer, inception 3a block, inception 3b block, maxpooling layer, and inception 4a block. All these layers are the same as in the standard GoogleNet architecture.

After inception 4a block, we have added three more inception blocks (inception 4a1, inception 4a2, inception 4a3) in our proposed model as shown in Figure 4.4. All these three newly added inception blocks have input layer, number of filters = 192 of 1x1 conv layer in the first path, number of filters=96 of 1x1 conv layer in the second path, number of filters=208 of 3x3 conv

layer in the second path, number of filters= 16 corresponding to 1x1 conv layer in the third path, number of filters= 48 corresponding to 5x5 conv layer in the third path, number of filters= 64 corresponding to 1x1 conv layer in the fourth path. These three newly added inception blocks are added to get binary classification label at hierarchical level 1. After adding three inception blocks, an Average pooling AVGPOOL layer is added of pool size 5x5 and stride=3. A convolutional Conv2D layer is added after Avgpool layer. This Conv2D layer has filters=128 with kernel size = 1x1. ReLU activation function is used in this Conv2D layer. This newly added Conv2D layer takes input from Avgpool layer. Further, the output of this Conv2D layer is flattened by adding Flatten layer. The flattened output is then passed to the newly added Dense layer. The Dense layer takes 1024 features that are flattened. ReLU function is used as an activation function. After that Dropout layer is added which dropped 50% of the neurons. The benefit of using a dropout layer is that it prevents the model from overfitting. By doing so, some of neurons (filters) and their connections are disabled so preventing model to rely too much on a single neuron. The computation cost of the model also gets reduced by adding this dropout layer. One more Dense layer is added which uses sigmoid function as a classifier to do binary classification at level 1. The output of inception 4a is passed to inception 4b which already exists in original Googlenet model. After inception blocks 4c and 4d, an Avgpool layer, Conv2D layer, flatten layer, Dense layer, Dropout layer, and Dense layer are added. The parameters of these newly added layers are the same as of previously added Avgpool, Conv2D, flatten, Dense, and Dropout layer. The only difference is in the Dense layer in which softmax activation is applied as classifier to perform multi-classification at level 2 of hierarchical classification.

The good point of any CNN based architecture is that the features extracted at various layers give different types of information. This information is useful and can be used to do classification. Less deep features give less information and more deep features give extra detail of information. The three more inception blocks are added to extract the deep detailed features for level 1 binary classification and level 2 multiclassification. All the layers from the start to the end are responsible for making the classification decisions at all three levels. The first classification level uses sigmoid activation function and binary crossentropy loss function. The second and third level uses softmax classifier and categorical crossentropy loss function. For every image, the model predicts its class at three levels such as major class, sub-class, and fine-grained class, by producing a hierarchy. The H-GoogleNet model was executed on Kaggle's GPU. The model is too heavy resulting in crash of RAM on Kaggle platform. To decrease the

model size, during execution, maxpooling layers are introduced at multiple levels. The selection and sequence of these extra added layers was decided after performing several trials to get the best performance results. The level one and level two outputs are generated from modified auxiliary classifiers.

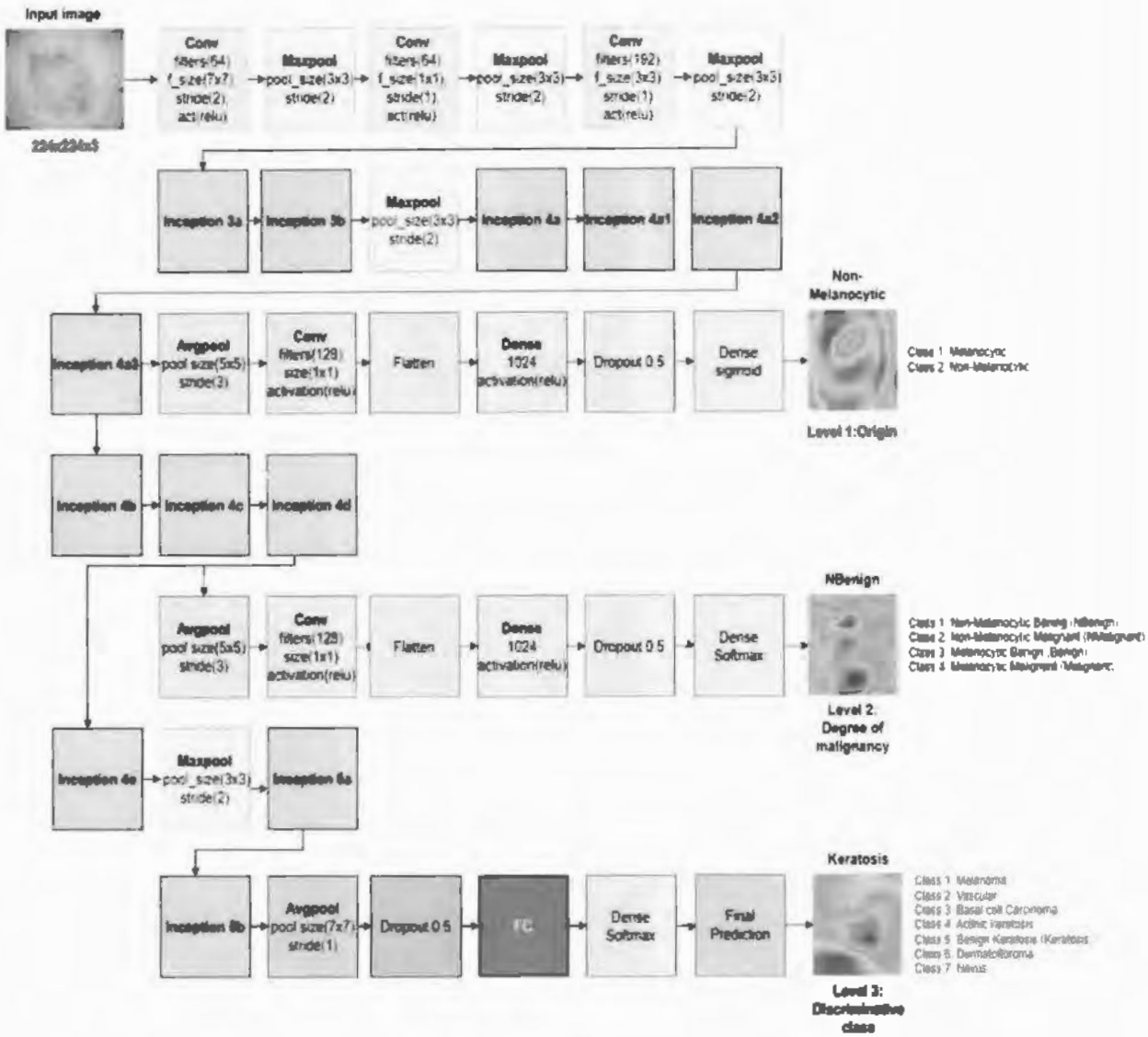


Figure 4.4: Proposed Hierarchical GoogleNet Model

4.6 Proposed H-GoogleNet Model for Skin Lesions Hierarchical Classification

In this section we have discussed the steps involved in the proposed H-GoogleNet model for skin lesions hierarchical classification. The model consists of the following steps:

4.6.1 Data Preparation

Let $D = (X, Y)$ be an augmented labeled dataset where $X = \{x_1, x_2, x_3, \ldots, x_n,\}$ is a dataset containing n samples and $Y = \{y_1, y_2, y_3, \ldots, y_n,\}$ have the labels detail of related samples in X. An instance x_i in a dataset D is denoted by a t-dimensional feature vector $x_i = \{f_1, f_2, f_3, \ldots, f_t,\}$. Let x_i^j represents the samples belonging to class j $where\ j = 0,1,2, \ldots 6$. From each class j, 80% of samples are selected for the training dataset. The remaining 20% instances from each class are taken as test dataset. Let X_{train} signifies the training dataset and X_{test} represents the test set. Similarly, Y_{train} represents the labels of training data and Y_{test} represents the labels of test data.

Mapping Class Labels Hierarchy

The process of mapping the hierarchy of class labels, at each level, is comprised of following steps:

1) Let $indices = \{\}$ represents the list of classes with indices as

$$indices = \{AKIEC: 0,\ \ BCC: 1,\ BKL: 2,\ \ DF: 3,\ MEL: 4,\ \ NV: 5,\ \ VASC: 6\} \quad (4.3)$$

where 0,1,…,6 represents the labels for mentioned classes.

2) Let the classes at $level\ 1$, $level\ 2$, and $level\ 3$ represented as

$$Classes_{level\ 1} = \{Melanocytic, Non\_Melanocytic\} \quad (4.4)$$

$$Classes_{level\ 2} = \{Malignant, Benign, NMalignant, NBenign\} \quad (4.5)$$

$$Classes_{level\ 3} = \{AKIEC, BCC, BKL, DF, MEL, NV, VASC\} \quad (4.6)$$

3) After defining the classes, the mapping for hierarchy is formulated as

$$level_3 \rightarrow level_2 = \{0: 3, 1: 2, 2: 3, 3: 3, 4: 0, 5: 1, 6: 3\} \quad (4.7)$$

where '0' in $\{0: 3\}$ represents $AKIEC$ from $Classes_{level\ 3}$, and '3' represents corresponding $NBenign$ from $Classes_{level\ 2}$. The remaining mapping is mapped in the same way.

4) Similarly, the mapping of hierarchy from $level\ 2$ to $level\ 1$ is formulated as

$$level_2 \rightarrow level_1 = \{0: 0, 1: 0, 2: 1, 3: 1\} \quad (4.8)$$

5) After that, the mapped labels for $level_2$ are extracted separately, by using formulated mapping, for all training and test sets as

$$level2^{y_i}_{train\_labels} = T(y_i) \qquad \forall\ y_i \in Y_{train}\ \ and\ i = 1,2, \ldots, size_{train} \quad (4.9)$$

where $T(.)$ represents the transformation function, transforming / mapping the $label^{y_i}_{level\_3}$ to the corresponding $Class^{label}_{level\_2}$, defined as:

$$T(y_i) = level_3 \rightarrow level_2[y_i] \quad (4.10)$$

and $level2_{train\_labels}$ is the obtained mapped labels $\forall\ y_i$ in Y_{train}.

6) Similarly, for $level_1$

$$level1^{y_i}_{train\_labels} = T\left(level2^{y_i}_{train\_labels}\right) \quad \forall\ y_i \in Y^{level2}_{train} \quad and\ i = 1,2,\dots,size_{train} \tag{4.11}$$

$$where \quad T\left(level2^{y_i}_{train\_labels}\right) = level_2 \rightarrow level_1[y_i] \tag{4.12}$$

7) Similarly, the mapped labels of X_{test} for $level_2$ and $level_1$ are extracted and the following test set labels are obtained as

$$level2^{y_i}_{test\_labels}$$

$$level1^{y_i}_{test\_labels}$$

4.6.2 H-GoogleNet Architecture

Once we have prepared the X_{train}, X_{test} and their corresponding labels for all three levels Y_{train}, $level2^{y_i}_{train\_labels}$, $level1^{y_i}_{train\_labels}$, and Y_{test}, $level2^{y_i}_{test\_labels}$, $level1^{y_i}_{test\_labels}$, and the defined model architecture, we use them to train our proposed H-GoogleNet model and then test it respectively. The proposed H-GoogleNet model consists of the following steps:

1) The proposed model takes an input image $x_i(h \times w \times ch)$ in its input layer.
2) Input feature maps of the input layer are convolved with the convolutional layer $CONV\_1$ as

$$O_{x,y} = \sum_{i=0}^{r=h} \sum_{j=0}^{r=w} x_{i,j} w_{i,j} + b \tag{4.13}$$

where $x_{i,j}$ is the pixel location of input image, and $w_{i,j}$ is the weight value of the filter or kernel at (i,j). The $\#\ of\ filters = 64$ and $filter\ size = 7 \times 7$ with $stride = 2$. This layer is followed by an activation function to add non-linearity. ReLU activation function is used in several layers of this architecture.

$$f(\alpha) = \begin{cases} 0 & for\ \alpha < 0 \\ \alpha & \alpha \geq 0 \end{cases} \tag{4.14}$$

$$\max(0, x) \tag{4.15}$$

3) Maxpooling layer is further added after $CONV$ layer to lessen the size of feature maps. The size of the maxpooling window is $maxpool\ size = 3 \times 3$ with $stride = 2$. The mathematical form of maxpooling is given in equation 4.2.

4) Two more $CONV$ layers $CONV\_2$ and $CONV\_3$ are added after maxpool layer. The $CONV\_2$ layer has $\#\ of\ filters = 64$ and $filter\ size = 1 \times 1$ with $stride = 1$. $CONV\_3$ layer has $\#\ of\ filters = 192$ and $kernel\ size = 3 \times 3$ with $stride = 1$.
5) Similarly, further maxpooling layer, inception blocks 3a and 3b, maxpooling layer, and more inception blocks are added. The formulation of inception block is given in Figure 4.1.
6) Three more inception blocks 4a1, 4a2, and 4a3 are added in our H-GoogleNet model before adding modified auxiliary classifier for level-1 classification as shown in Figure 4.4.
7) A modified auxiliary classifier is added after inception block 4a3. This auxiliary classifier has average pooling layer as

$$f_k^{(a)} = \frac{1}{X_k}\sum_{x \in X_k} x \tag{4.16}$$

where $f_k^{(a)}$ is the output average pooled map, X_k is the size of the input feature map, and x is the individual feature value in the feature map. After $Avgpool$ layer in modified auxiliary classifier, $CONV2D$ layer is added having $\#\ of\ filters = 128$ and $kernel\ size = 1 \times 1$ with ReLU activation. After that the output feature maps are flattened and a dense fully connected (FC) layer is added as

$$y(x_i) = f(\sum_{i=1}^{n} w_i s_i) \tag{4.17}$$

where $f(.)$ is the activation function ReLU. w_i is the weight multiplied with input s_i. $y(x_i)$ is the obtained feature vector. After that, dropout layer is added as

$$q = 1 - p \tag{4.18}$$

Where p is the dropout rate which is $p = 0.5$ in our proposed model. The dropout layer is added to reduce the computations of the model and to get rid of overfitting. After getting reduced feature vector, the classification at level-1 is performed by applying sigmoid activation function at this layer as

$$S(x) = \frac{1}{1+e^{-x}} \tag{4.19}$$

Up till now, the classification at level-1 is performed and a class label is predicted for binary classification using this modified auxiliary classifier.

8) After that, binary cross-entropy loss function is applied to calculate the loss at classification level-1 as

$$CrossEntropy_{binary} = -\frac{1}{M}\sum_{i=1}^{M} l_i log\hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i) \tag{4.20}$$

9) If a large loss is obtained, then optimization is applied to update the wights of the model as

$$w_{new} = w_{old} - \eta \frac{\partial L}{\partial w_{old}} \tag{4.21}$$

10) Repeat steps 2-9 for all samples in X_{train} for *level* 1 binary classification until the optimized weights are obtained.

11) Along above, in parallel, after inception 4a3 block, inception 4b, 4c, and 4d are there as in the original GoogleNet model.

12) After inception block 4d, another modified auxiliary classifier is added, the same as given in step 7 above, for level-2 multi-class classification.

13) For level-2 multi-class classification, softmax activation function is used as

$$softmax(y_j) = \frac{\exp(y_j)}{\sum_{k=1}^{m} \exp(y_k)} \qquad for\ j = 0,1,2, \ldots\ldots, n \tag{4.22}$$

14) For this level-2 multi-class classification, categorical cross-entropy loss function is applied as

$$CrossEntropy_{categorical} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{4.23}$$

Weights are optimized as discussed in step 9.

15) Repeat steps 2-7 and 11-14 for all samples in X_{train} for *level* 2 multiclass classification and until the optimized weights are obtained.

16) Along above, in parallel, after inception 4d, there are inception 4e, maxpool, inception 5a and 5b blocks. The layers detail of these blocks is given in Table 4.1.

17) Finally, the level-3 multi-class classification is carried out to determine the fine-level class.

18) Loss function and optimizer are applied to reduce the loss and update the weights respectively.

19) Repeat steps 2-6, 11, and 16-18 for all samples in X_{train} for *level* 3 multiclass classification and until the optimized weights are obtained.

4.6.3 Learning Rate

Learning rate is one of the important parameters in any CNN based model. In the training of a deep learning based model, initially the learning rate is very significant. The too large learning rate led to poor results. On the other side, too small learning rate make it difficult for the model to converge during training process. In the proposed H-GoogleNet model, a learning rate schedular is used to schedule the learning rate for the model after various number of epochs. Initially the learning rate is set to 0.0001. After that, we have kept it on changing after 42 epochs and then 52 epochs.

4.6.4 Optimizer

In any deep learning model, an optimizer is used to alter weights during CNN based model training during backpropagation. After getting the loss value from the loss function, the weights are modified again using an optimizer function. We do this to optimize the performance of CNN model to get better convergence effect. From this we can observe that, the selection of a good optimizer is of great concern. The training of the model will be difficult to converge if a bad optimizer is selected. In the proposed H-GoogleNet model, after performing multiple experiments, we have selected Adaptive Moment Estimation (ADAM) based on its performance on our ISIC datasets. The ADAM optimizer is the second-order moment function. It is an adaptive learning rate method such as for various parameters, it calculates individual learning rates. The equation for ADAM optimizer is as

$$w_{t+1} = w_t - \alpha m_t \tag{4.24}$$

where,

$$m_t = \beta m_{t-1} + (1-\beta)\left[\frac{\delta L}{\delta w_t}\right] \tag{4.25}$$

In this, m_t is the aggregate of gradients at time t (current) and initially m_t= 0. m_{t-1} represent the aggregate of gradients at time *t-1* (previous). w_t is the weight at time t. δL is the derivative of loss function. δw_t is the derivative of weight at time t. β is the moving average parameter and its constant default value 0.9 is used in the proposed work.

4.6.5 Procedure Steps

In this section, we have discussed about the stages of the proposed procedure.

i. Load Data

At first place, the ISIC dataset of skin cancer is loaded. The dataset is separated into 80% training dataset and 20% test dataset.

ii. Customized Layers of GoogleNet

In the second place, the customized layers of CNN are used to introduced variations in GoogleNet model to include the three-level hierarchy in the model.

iii. Activation Function

Rectified Linear Unit (ReLU) and sigmoid are used as an activation function in the H-GoogleNet model.

iv. Loss Functions

Various loss functions are used in the proposed model. They work well in the presence of imbalance data.

v. Gradient Class Activation Map

The gradient class activation maps are used for the visualization of locations in a lesion that participate more in the identification of the class of skin lesion.

vi. Optimizer

ADAM is used as an optimizer.

vii. Network Evaluation using Test Data

In this step, the classification is performed by using test dataset to assess the performance of the model. Softmax classifier is used at the third level classification of skin lesions.

4.7 Experiments and Results

In this section, the experiments are performed to assess the performance of the proposed H-GoogleNet model on ISIC 2018 dataset. The detailed description is given in the succeeding sections.

4.7.1 Dataset

To assess deep learning models for automated segmentation and classification of skin cancer lesions, various standard databases are openly accessible. The basic intent of these openly available databases is to assess the strength of automated skin cancer screening as well as to compare the obtained results of proposed models with existing techniques. One of these skin imaging databases is International Skin Imaging Collaboration (ISIC) datasets [26][27]. The ISIC [26][27] is a platform for skin imaging, specifically for skin cancer diagnosis. It provides a centralized platform for data sharing and analysis. The ISIC has published various datasets from year 2016 to 2020 named ISIC 2016, ISIC 2017, ISIC 2018, ISIC 2019, and ISIC 2020. These databases include separate datasets for segmentation and classification tasks. The ISIC 2016 dataset contains 900 instances for classification task and two classes (malignant and benign). ISIC 2017 has 4000 instances for three classes including melanoma, nevus, and seborrheic keratosis. On the other side, ISIC 2018 dataset has 10,015 images for classification task and has seven classes. ISIC 2019 dataset has 25,331 images with eight classes. In our proposed work we have used 2018 dataset of seven classes. The complete details of the ISIC 2018 datasets used for the training and evaluation of the proposed Network is described in Table 4.2.

Table 4.2: A brief overview of experimental ISIC datasets [26][27]

Class No	Class Name	Number of Instances (2018)
1	MEL	1113
2	NV	6705
3	BCC	514
4	AKIEC	327
5	BKL	1099
6	DF	115
7	VASC	142
	Total	**10015**

The classification task of these datasets is available in the form of train set, validation set, and the test set. The ground truth data is only available for training data but not for validation and test set. Due to this reason, we have splitted our ISIC 2018 training data into further training data, validation data, and test data along with their ground truths. All these instances belong to seven classes named as: Actinic keratosis (AKIEC), melanoma (MEL), benign keratosis (BKL), melanocytic nevus (NV), dermatofibroma (DF), vascular lesion (VASC), and basal cell carcinoma (BCC). The total data is divided into 80% train and 20% test data from each eight classes. All available ISIC dataset versions are highly imbalanced. In our used ISIC 2018 dataset, all seven classes are highly imbalanced as shown in Table 4.2. Vascular and dermatofibroma classes have almost ignorable number of samples that could lead the model toward biasness. Actinic keratosis and basal cell carcinoma also have a smaller number of samples. Nevus is the majority class as it has many samples as compared to the other classes in the dataset. The reason why this dataset is highly imbalanced is that there are a smaller number of patients with BCC, AKIEC, DF, and VASC in Europe. Due to this many samples are not available. On the other side, there is a large growth of patients having nevus, and melanoma. To prepare this dataset for hierarchical classification, a mapping of hierarchy is formulated as given in section 3.6. We have applied data augmentation techniques to get the balanced dataset. By doing so, we have equalized the number of samples in each class, such as 1000 in each class as shown in Table 4.3, so the model may not show biasness toward any class, and it will show generalized performance.

Table 4.3: Balanced dataset after applying data augmentation on ISIC 2018 dataset.

Class No	Class Name	Number of Instances
1	MEL	1000
2	NV	1000
3	BCC	1000
4	AKIEC	1000
5	BKL	1000
6	DF	1000

Class No	Class Name	Number of Instances
7	VASC	1000
	Total	**7000**

4.7.2 Performance Evaluation Metrics

The performance evaluation measures accuracy, sensitivity, specificity, precision, F-score, and AUC-ROC are selected for validation of the designed model on ISIC 2018 dataset. These measures are calculated by using the following equations.

1) Accuracy

Accuracy is used to measure how correctly a classification test recognizes a condition. It is the fraction of accurate results to the entire number of instances.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \tag{4.26}$$

2) Sensitivity / Recall

Sensitivity measures the capability of the developed method to properly detect the patients with disease who do have the condition. It is the fraction of those whose test result is positive for the disease, among those who have the illness.

$$Sensitivity(Recall) = \frac{TP}{TP+FN} \tag{4.27}$$

3) Precision

Precision (positive predicted value) is proportion of related examples among the retrieved ones.

$$Precision = \frac{TP}{TP+FP} \tag{4.28}$$

4) F1-Score

It summarizes predictive performance of a model by integrating precision and recall. It is a harmonic mean of precision and recall.

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4.29}$$

5) ROC-AUC Score

The ROC curve depicts the performance of a model at various classification threshold. ROC-AUC is a graph between true positive rate (TPR) (recall / sensitivity) and false positive rate.

4.7.3 Experimental Setup

Experiments of proposed H-GoogleNet model are performed on ISIC 2018 dataset. The proposed model is executed on Kaggle and Colab with 12GB available RAM of GPU. Experiments are performed by using the ADAM optimizer with binary cross entropy and multi-class crossentropy loss functions. The images of training set and test set are resized to [224 x 224 x 3] resolution to be acceptable for H-GoogleNet architecture. The original GoogleNet

model is pre-trained on ImageNet dataset (ILSVRC) having 1000 classes. In this work, we have not used pre-trained weights of ImageNet dataset by using transfer learning. The reason for not using pre-trained weights using transfer learning is that when we add more layers in the original architecture of GoogleNet, the learned weights do not remain the same. Their sequence got disturbed, and they do not remain available for processing. That's why we need to train the model from scratch and learn new weights. Due to this, all new weights are learned from scratch in the model. The weights are initialized randomly using a random function. After that, we set training options of the H-GoogleNet architecture. After several iterations and running the experiment several times, we selected the mini-batch size=32 and maximum epochs = 60. For ADAM, the initial learning rate is set to 0.0001, as it gives good results on this learning rate. The parameters setting of H-GoogleNet model is given in Table 4.4. From this table, it can be observed that the learning rates are set after reaching a certain number of epochs such as after reaching epoch 20, the learning rate is set to 0.0002, and 0.00005 after reaching epoch 40 for ADAM optimizer. All these optimized learning rates are obtained after performing several experiments several times for various fine-tuned parameters and hyperparameters of the model. The function of Loss Weight Modifier is introduced to modify the weights if a certain number of epochs are reached before reaching stopping criteria such as if epoch reaches 15, 25, and 35 then modify the alpha, beta, and gamma values accordingly. The dataset is shuffled in every epoch during training so that every instance had a chance of being part of training data from validation set but the $batch\ size = 32$ in every epoch remains the same. The proposed model is trained by using an early stopping criterion such as $patience = 10,\ and\ verbose = 1$. In our dataset, there are two classes at level 1, four classes at level 2, and seven classes at level 3 such as level 1 has binary classification and level 2, and level 3 have multi-class classification.

Table 4.4: Experimental parameters setting for H-GoogleNet Model training.

Optimizer	Initial Learn Rate	LR after epoch 20	LR after epoch 40	No. of Epochs	Batch size	Early Stopping Criteria	
ADAM	0.001	0.0002	0.00005	60	32	patience	verbose
						10	1

4.7.4 Experimental Results

In this section, we have discussed the experimental results of our proposed H-GoogleNet model on ISIC 2018 in detail. The results are given in tabular form along with various plots in the following sections.

Results and Evaluation with ADAM Optimizer

In this section, the experiments are performed using ADAM optimizer while using binary cross entropy and multi-class cross entropy as loss functions. Average experimental results of accuracy, precision, recall, F1-score, and ROC-AUC score for three hierarchical levels are shown in Table 4.5. Optimizer plays its role during backpropagation to optimize the model training and reduce the value of loss function to improve the model performance.

From Table 4.5 we observed that H-GoogleNet is giving better performance with ADAM. The termination epoch is set by introducing early stopping criteria. The training is stopped at epochs = 51 for ADAM optimizer which shows that the model is not showing any improvement after epochs = 51. The accuracy, precision, recall, F1-score, and ROC-AUC score show acceptable results. From Table 4.5 we can observe that, at level 2 (L_2) the performance of model is going down as compared to level 1 and level 3. This is because the complexity of classes increases at level 3 due to the similarity among classes as they all are benign and malignant. The model faces difficulties in classification due to these huge similarities.

Further, the evaluation measures at individual classes of each level are also shown in Table 4.6, 4.7, & 4.8. From these individual results, it is observed that our proposed model performed very well for each individual class at all three levels. The confusion matrices results are shown in Figure 4.5 (a, b, & c). The training and validation losses curves of the H-GoogleNet model with ADAM optimizer are shown in Figure 4.6 for all three levels. In this, the number of epochs is given on x-axis and loss values on y-axis.

From Figure 4.6, we can observe that training loss of the model is continuously decreasing even after epoch 50 for level 3 classification using ADAM optimizer. This is also the same for level 1 and level 2 classification. On the other side, validation loss of the model for level 2 and level 3 is fluctuating continuously such as the validation loss is kept on increasing and decreasing till the last epoch.

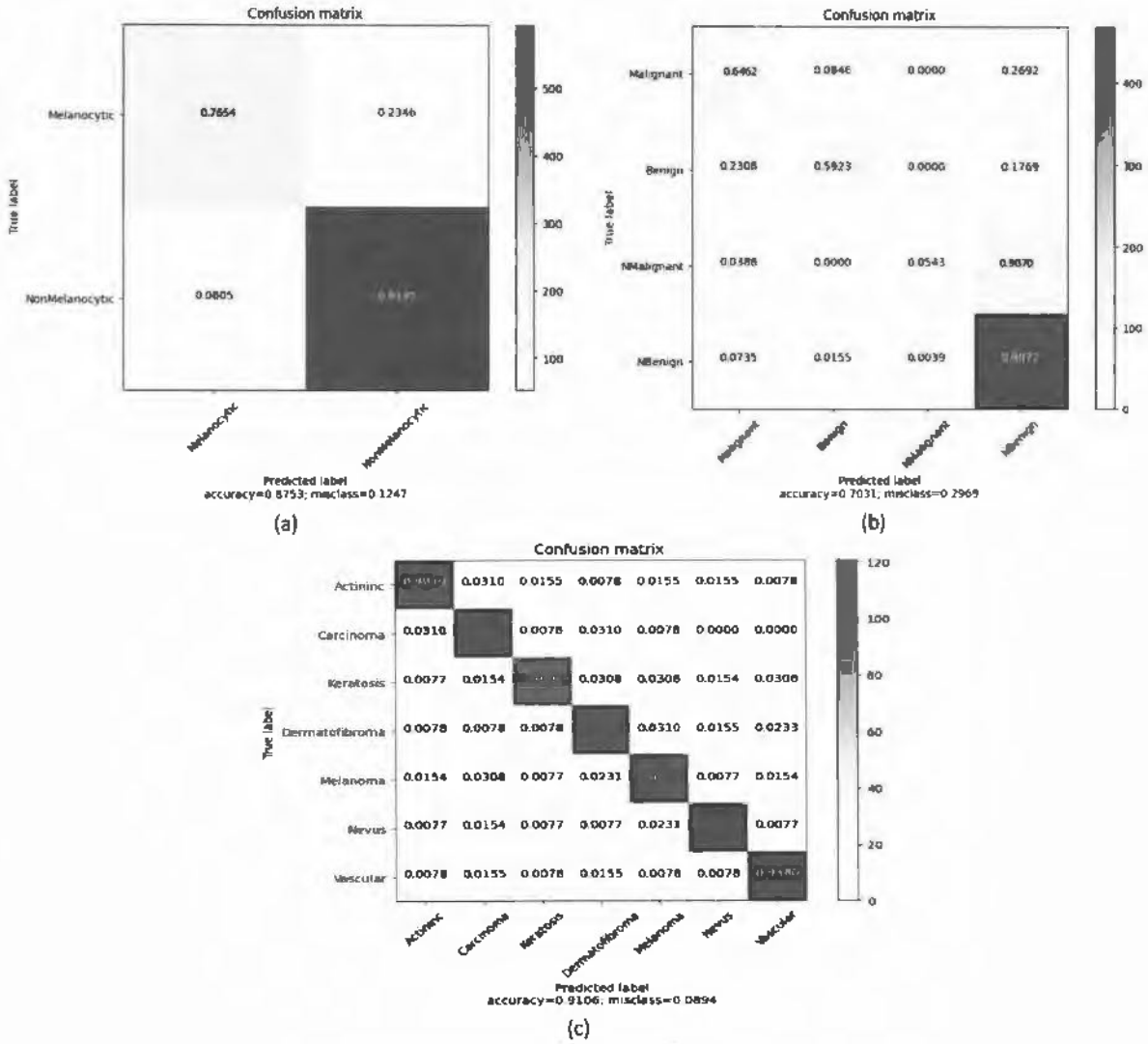


Figure 4.5: Confusion matrices produced at (a) level-1 (b) level-2 (c) level-3 of classification.

Table 4.5: Average evaluation measures of proposed model for classification of test dataset at three levels.

Optimizer	ADAM		
Initial Learn Rate	0.001		
Termination epoch #	51		
	L_1	**L_2**	**L_3**
Accuracy	0.875	0.703	0.911
Precision	0.85	0.73	0.91
Recall	0.84	0.55	0.91
Specificity	0.84	0.86	0.98
F1-Score	0.85	0.54	0.90
ROC-AUC Score	0.80	0.70	0.965

It is observed that training and validation loss at level 1 and level 2 classification is less as compared to level 3 classification. It shows that features extracted till level 1 and level 2 give better performance. Furthermore, features extracted at level-1 are less complex as compared to level-2. Training and validation loss at level 1 (binary level) is very less as compared to remaining two levels. Further, ROC-AUC plots of H-GoogleNet for all three levels are

depicted in Figure 4.7 (a, b, c) with ADAM optimizer. These ROC-AUC plots give a clear picture of the performance of the model. The predicted probabilities of the model are shown in Figure 4.8 (a, b, c) for all three-level classifications. The results shown here are produced randomly. The Gradient Class Activation Maps (Grad-CAMs) of predicted hierarchy of skin lesion types, with focus on locations participating more in the prediction of skin lesion types, are shown in Figure 4.9. From these activation maps, the visualization of spatial location of actual lesion's structure become clearer. From these activation maps, we become more able to get knowledge about which area of the lesion is participating more at any level of classification of skin lesion. To elaborate on it, in Figure 4.9, the regions highlighted in red are the ones participating more in the classification of lesion at three levels. It is also observed from Figure 4.9 that the model may also focus on that area of skin that does not have the lesion. This is because the model faces some structural complexities in the skin lesion image, and it may confuse the foreground with the background in some images. The structure present at the boundary of skin lesion may resemble the background skin structure creating complexities for the model.

Table 4.6: Evaluation measures of individual classes at class level 1 of test set

Class	Precision	Recall	Specificity	F1-Score	AUC-ROC Score
Melanocytic	0.79	0.77	0.92	0.78	0.84
Non-Melanocytic	0.91	0.92	0.77	0.91	0.75
Average	0.85	0.84	0.84	0.85	0.80

Table 4.7: Evaluation measures of individual classes at class level 2 of test set

Class	Precision	Recall	Specificity	F1-Score	AUC-ROC Score
Malignant	0.62	0.55	0.91	0.58	0.75
Benign	0.80	0.59	0.98	0.68	0.78
NMalignant	0.78	0.05	1.0	0.10	0.53
NBenign	0.73	0.91	0.55	0.81	0.73
Average	0.73	0.55	0.86	0.54	0.70

Table 4.8: Evaluation measures of individual classes at class level 3 of test set

Class	Precision	Recall	Specificity	F1-Score	AUC-ROC Score
Actinic	0.92	0.91	0.99	0.91	0.94
BCC	0.89	0.92	0.98	0.90	0.95
Keratosis	0.94	0.87	0.99	0.90	0.93
Dermatofibroma	0.89	0.91	0.98	0.90	0.94
Melanoma	0.89	0.90	0.98	0.89	0.94
Nevus	0.94	0.93	0.99	0.93	0.96
Vascular	0.92	0.94	0.99	0.93	0.96
Average	0.91	0.91	0.98	0.91	0.95

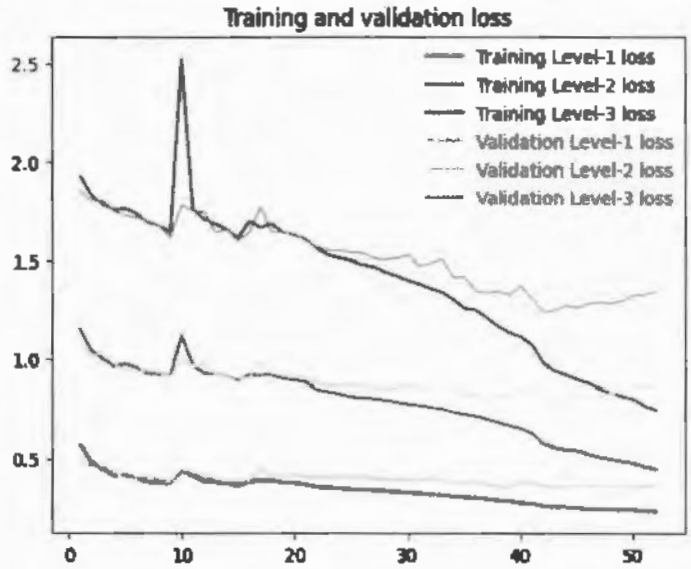
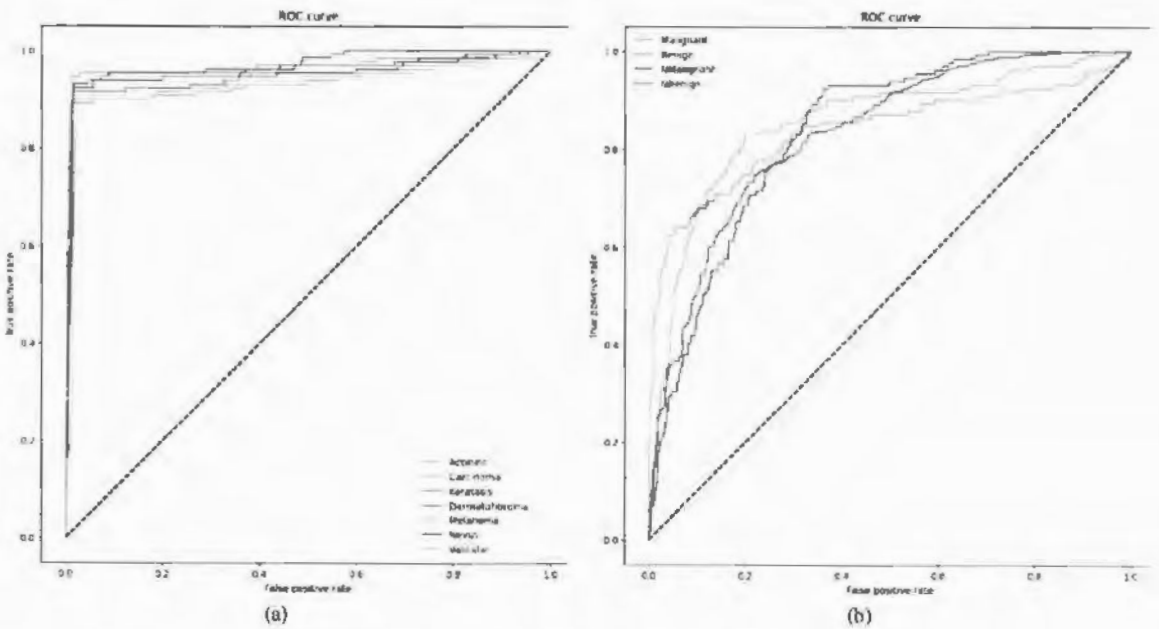


Figure 4.6: Training and validation loss of proposed H-GoogleNet model.



(a) (b)

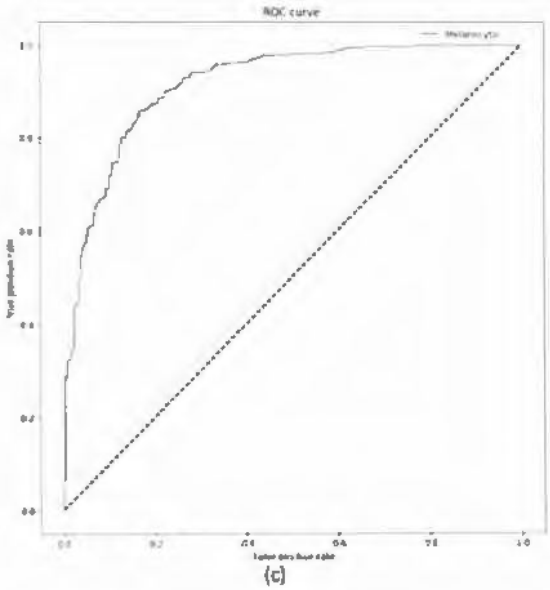


Figure 4.7: ROC-AUC curves of proposed H-GoogleNet model with ADAM optimizer.

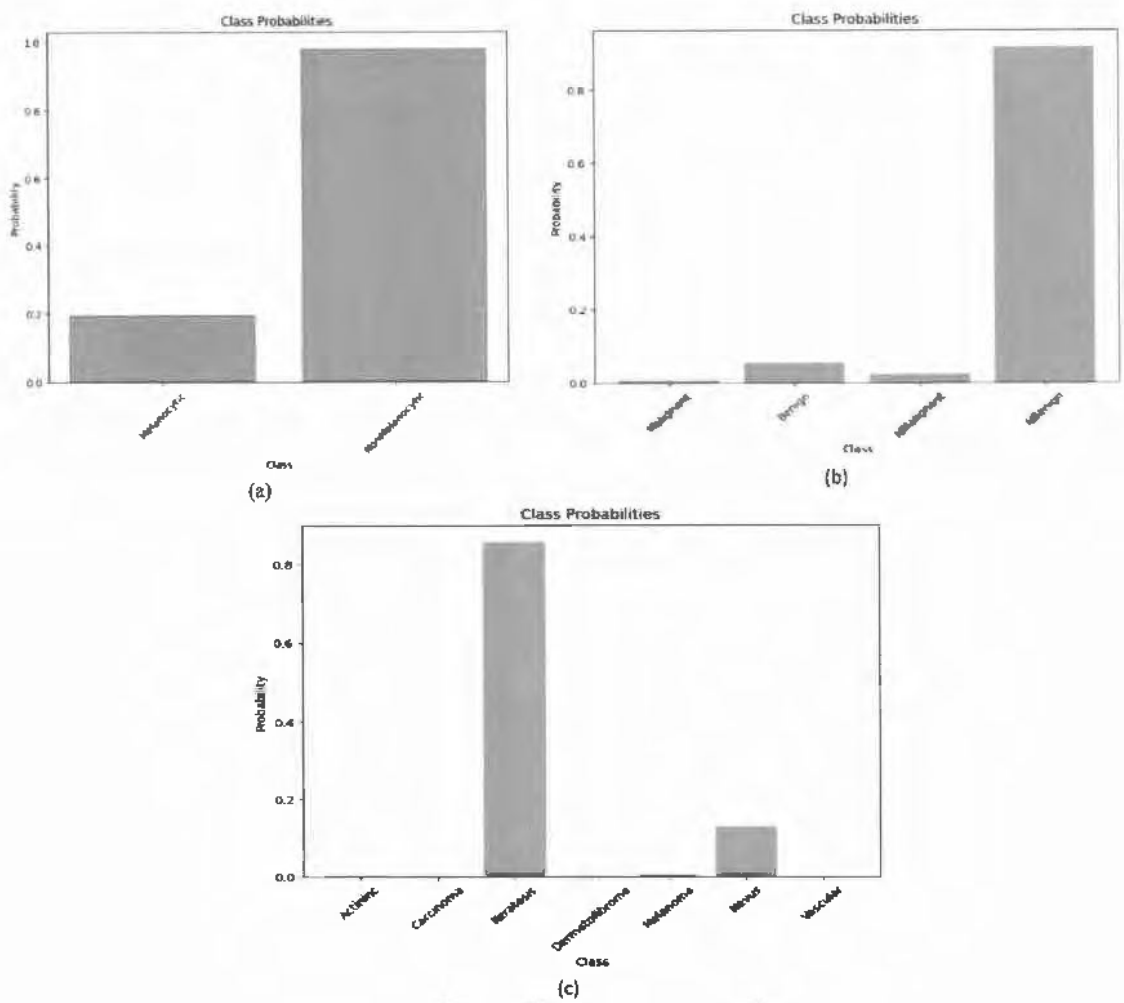


Figure 4.8: Predicted probabilities of skin cancer lesion at (a) hierarchical level-1 (b) hierarchical level-2 (c) hierarchical level-3 using proposed model.

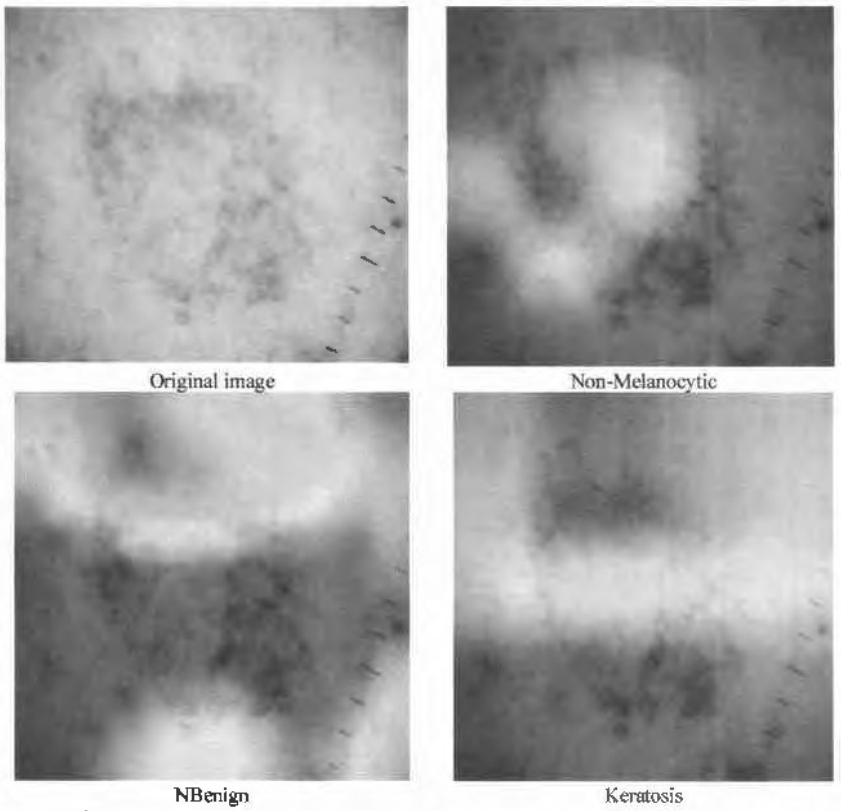


Figure 4.9: Grad-CAMs of predicted hierarchy of skin lesion types with focus on locations participating more in the prediction of skin lesion types, with proposed H-GoogleNet model.

4.8 Comparison and Discussion

In this section, we have shown the comparison of our proposed model with the competitor and given the discussion.

4.8.1 Comparison of H-GoogleNet with Existing CNN Based Explainable Model

In this section, we have given a comparison between our proposed H-GoogleNet model and existing explainable model [29]. This comparison is performed for the taxonomic classification and explanation of skin lesion by locating the regions participating the more in the diagnosis of accurate skin lesion type. The average results of the comparison are shown in Table 4.9. The evaluation results for individual classes at each level are shown in Table 4.10. For comparison, we have shown results using specificity, sensitivity, and AUC values, as our competitor has used only these measures to evaluate their model. From Table 4.9, it can be observed that our proposed model depicts better average results as compared to existing model in terms of specificity, sensitivity, and AUC. Our model gives 98.0% average specificity for level 3 classification. The existing model [29] shows 96.8% specificity for fine level classification which is much less than our obtained specificity. Similarly, the proposed model shows

acceptable average specificity results for level 1 and level 2 classification such as 84% and 86% respectively. The competitor has not shown the evaluation results at level 1 and level 2 in their work, so we are not able to compare the gained results for level 1 and level 2 classification. Further, the proposed model shows 91% average sensitivity for level 3 classification which is much better than existing model [29] giving only 69% average sensitivity. For level 1 and level 2 classification, our obtained average sensitivity is 84% and 55% respectively. The obtained AUC score of our proposed model is 95% which is a little bit less than the AUC score of existing models [29]. The existing model is showing 95.2% AUC score which is 0.2% more than our obtained AUC score. As far as the performance of the proposed model for all individual seven classes at fine level is concerned, the results are shown in Table 4.10. From this table it can be observed that our proposed model shows better sensitivity results for all classes at level 3 (fine level) as compared to competitor [29]. The proposed model shows sensitivity scores as 90%, 93%, 92%, 91%, 87%, 91%, and 94% for melanoma, nevus, BCC, actinic, keratosis, dermatofibroma, and vascular respectively. The specificity of the proposed model is fluctuating or showing less scores for few classes at fine level as compared to existing model [29]. The individual results for specificity at level 3 are shown in Table 4.10. Our proposed model shows better specificity scores for melanoma, nevus, and keratosis as 98%, 99%, and 99% respectively. For our proposed model , the specificity score is going down for BCC, actinic, dermatofibroma, and vascular as compared to competitor [29] but the overall average specificity score of our proposed model is better than the competitor [29]. This increase in average is because our model is showing good specificity score for melanoma such as 98% that is 8% more than the competitor's specificity score for melanoma and it is a huge difference. Similarly, the same is the case for nevus with a difference of 6% in specificity score of both models. The AUC score of our proposed model for melanoma is 94% at level 3 classification that is much better than the competitor [29]. Overall, the performance of the proposed model is better than the competitor when we analyze the obtained results. The results for sequence of hierarchy generated by the proposed model with predicted probabilities is given in section 4.6.4.

Table 4.9: Comparison of proposed model with existing CNN model for three level hierarchical classification.

	Attention LSTM [29]			**Proposed H-GoogleNet**		
	L_1	**L_2**	**L_3**	**L_1**	**L_2**	**L_3**
Avg. Specificity (SP)	NA	NA	96.8	84.0	86.0	**98.0**
Avg. Recall / SE	NA	NA	69.0	84.0	55.0	**91.0**
Avg. ROC-AUC	NA	NA	**95.2**	80.0	70.0	95.0

Table 4.10: Comparison of proposed model with existing CNN model for three level hierarchical classification (individual classes at each level).

	Attention-LSTM [29]			Proposed H-GoogleNet		
	SE	SP	AUC	SE	SP	AUC
Melanoma	67.8	90.1	86.1	**90.0**	**98.0**	**94.0**
Nevus	82.0	93.7	**96.1**	**93.0**	**99.0**	96.0
BCC	74.2	**98.4**	**98.5**	**92.0**	98.0	95.0
Actinic	60.5	**99.5**	**95.5**	**91.0**	99.0	94.0
Keratosis	72.8	96.6	**94.4**	**87.0**	**0.99**	93.0
Dermatofibroma	68.2	**99.8**	**97.2**	**91.0**	98.0	94.0
Vascular	57.1	**99.7**	**98.5**	**94.0**	99.0	96.0

4.8.2 Discussion

The above given experiments and their comparison show that the proposed H-GoogleNet performed better than existing work [29].

From the above given experiments, it can be observed that, H-GoogleNet model gives good results with ADAM optimizer. In proposed H-GoogleNet, the auxiliary classifiers are used to classify the first two levels of classification. From experiments, it is observed that classification level 3 accuracy is better than level 1 and level 2 accuracy. This model stopped after completing 55 training epochs. The model ran for the longest time on 12 GB RAM of GPU on Kaggle. Moreover, H-GoogleNet is light weight on computational resources.

For visualization of structures, present in skin lesions, Grad-CAMs are plotted to comprehend the specific decision of the proposed model. From experiments, it is observed that the proposed H-GoogleNet model localizes different regions for all three individual levels of same skin lesion, to take accurate decision. It implies that the model looks at different regions to predict the sequence of hierarchy.

Chapter 5

Skin Cancer's Structure Segmentation using ResNet50-UNet Model with Hybrid Loss Function

Chapter 5

5 Skin Cancer's Structure Segmentation using ResNet50-UNet Model with Hybrid Loss Function

5.1 Introduction

The most common use of CNN based models is for the feature extraction and classification tasks, to get the output of an image as a single class label. In the previous chapter, we proposed a hierarchical based CNN model named H-GoogleNet for explainable classification of skin lesion, where till the last layer we get a hierarchy of classes produced by the model. Along with the above, however, for many visual tasks of biomedical image processing, there is a great need that the output should include localization. We should be able to find structures and spatial location (with name) within an image that are participating more in the identification and classification of skin lesion type and its structure.

In this chapter, we have addressed this problem and proposed a solution to deal with the attribute or structure level segmentation of skin lesion. We have applied segmentation model named UNet [107] with a backbone model such as ResNet-50 [108], for the attribute or structure segmentation and localization of skin lesion attribute inside an image. By doing this, we will be more able to get the presence of structure. By knowing the presence of numerous structures such as pigment network, streaks, milia, globules etc. in a lesion, we would be more able to get knowledge about the lesion type. Resultantly, the explainability of the model will increase. This structural level information will help the practitioner to know about the accurate class of skin lesion.

In this work, we have worked with some limitations due to the absence of skin lesion type information and its structure type of information, present in them, at one place. The ISIC dataset does not have these types of information at one place, such as in one dataset.

5.2 U-Net Architecture

In this section, we have discussed the U-Net based model exist in literature. UNet architecture [107] is based on fully convolutional networks. The intent of the U-Net is to extract both the context features as well as the localization. U-Net [107] consist of two parts: encoder and decoder. Encoder and decoder are the first and second half of the model respectively. The encoder is also called the contracting path and decoder as expansive path. The encoder part of

the U-Net model is a pre-trained network which uses any CNN based architecture as a backbone, such as ResNet and VGG etc. In U-Net architecture, at expansion side, two 3x3 unpadded convolutional blocks are applied that are followed by rectified linear unit (ReLU), maxpooling operation of 2x2 and a stride of 2 for downsampling. This downsampling is used to encode the input image into feature maps at various levels. The number of features is doubled at each downsampling step.

After that, decoder, also called expansive path, is the second part of the UNet architecture. In this part, discriminative features learnt by encoder are projected onto the pixel space (higher resolution) to get segmentation. Each step of the decoder comprises of upsampling of the feature map. This upsampling is further followed by the 2x2 convolution, also known as up-convolution. This upsampling is half the number of feature channels. There is a concatenation with the cropped feature map from the concatenating path, and 3x3 convolutions, followed by ReLU. This network has 23 convolutional layers. UNet model was introduced in 2014 for semantic segmentation of biomedical images. Semantic segmentation is basically classification per pixel of an image. This concept is employed in medical imaging to localize a certain part of an image usually where the abnormality exists. In semantic segmentation unlike classification, the input and output hold the same shape. The architecture of UNet is shown in Figure 5.1.

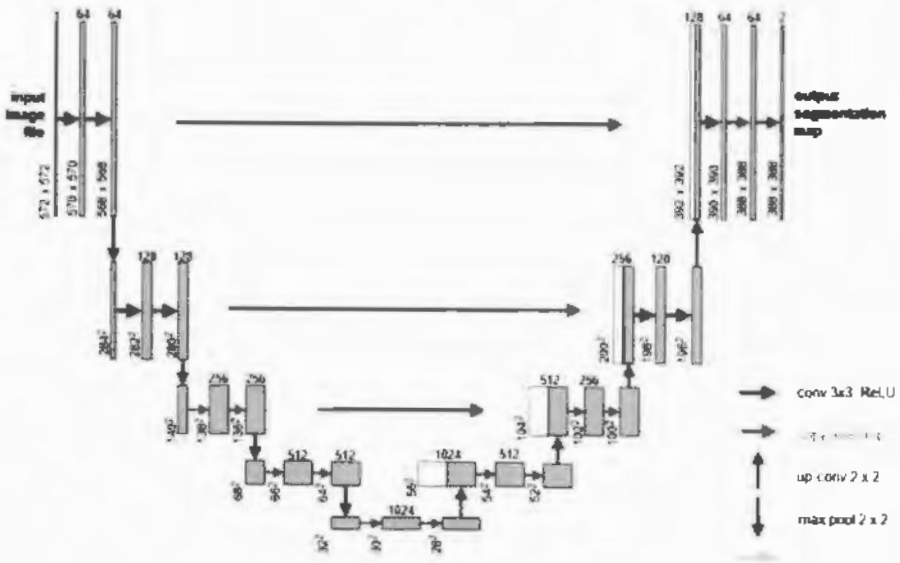


Figure 5.1: UNET Architecture[107]

The final layer in the architecture is 1x1 convolution which is utilized to map feature vector of 64 components to given number of classes. In the proposed technique, the encoder layers of UNet were replaced with ResNet architecture (used as backbone).

5.3 ResNet Architecture

ResNet [108] was first introduced in solving the degradation or vanishing gradient problem of the deeper networks. After converging, the deeper networks start to degrade. This degradation is not because of overfitting. Instead, the training accuracy of the model is degraded. The ResNet architecture solved this degradation problem. In the architecture, the stacked layers are explicitly mapped to residual mapping. In preceding architectures, the stacked layers were mapped directly to the required primary mapping. In the mathematical notation, let $\mathcal{H}(x)$ be the desired underlying mapping. In ResNet architecture, another mapping $G(x) := \mathcal{H}(x) - x$ is defined, and stacked nonlinear layers are mapped to this mapping. The original mapping is re-expressed in the form $G(x) + x$. The hypothesis is that optimization of residual mapping is easier than optimization of unreferenced mapping. The $G(x) + x$ formulation can be executed in the feed forward network with 'shortcut connections'. The basic building block of residual learning is shown in Figure 5.2.

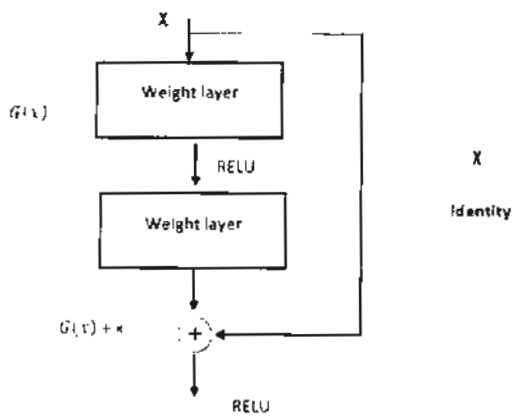


Figure 5.2: Building block of residual learning [108]

In the ResNet architecture, shortcut connections enable "identity mappings". Further, the outputs of these shortcut connections are combined with the outputs of the stacked layers. Notably, these identity mappings do not lead to extra computational overhead. The models like ResNet in Deep residual networks, reveal amended outcomes as the model goes deeper and deeper, as compared to the models that face deprivation in their results as the depth increases. Various versions of ResNet are introduced in literature based on different layer counts such as 18-layer, 34-layer, 50-layer, 101-layer, and 152-layer as shown in Table 5.1. In our proposed model, for the structure segmentation task, ResNet-50 architecture is used as backbone in U-Net model.

Table 5.1: ResNet model versions [108]

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112x112	7 x 7, 64, stride 2				
conv2_x	56 x 56	3 x 3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
Conv3_x	28 x 28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 32$
Conv4_x	14 x 14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
Conv5_x	7 x 7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1 x 1	Average pool, 100-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

5.4 Structure Segmentation Model

In this section, we have detailed our proposed hybrid model for structural segmentation of skin lesion. The proposed model is named ResNet50-UNet.

5.4.1 Model Overview

An overview of our proposed ResNet50-UNet model is given in detail in this section. Due to the presence of similar or sometime overlapped structures, in skin lesions, their segmentation at semantic level becomes an extremely challenging task. To address this challenge, we proposed a novel attribute's level segmentation model using hybrid loss function. This proposed model is used to perform pixel level classification or semantic segmentation of structures present in skin lesion's images. An overview of the proposed model is given in Figure 5.3. The proposed model comprises two parts including an encoder and decoder like the original UNet [94]. In this, the section of encoder is swapped by ResNet-50 model [95]. The ResNet-50 is used as a backbone in the proposed UNet model. This encoder-decoder is shown in Figure 5.4 and Figure 5.6. The ResNet50 encoder part down-sample the input image of skin lesion to extract the high level details of akin lesion attributes. Till the last layer, the ResNet50 encoder extracts the feature maps having very high level information but with low resolution. This high level information is then upsampled in the decoder section of the UNet model. During up-sampling, these low level feature maps are also concatenated with high level feature maps (in encoder), via skip connections. By doing so, the decoder will be more able to up-sample

the low level feature maps by combining them with extra information from skip connections. At the end, there is a convolutional layer that convolves the final up-sampled feature maps and apply activation function to predict the class of pixels. The encoder (ResNet50) uses the learned weights of the model by transfer learning. The weights of the decoder are initialized and then updated during backpropagation. There are five attributes in skin lesion segmentation task, so this proposed model is learned individually for each attribute. In total, five ResNet50-UNet models are there.

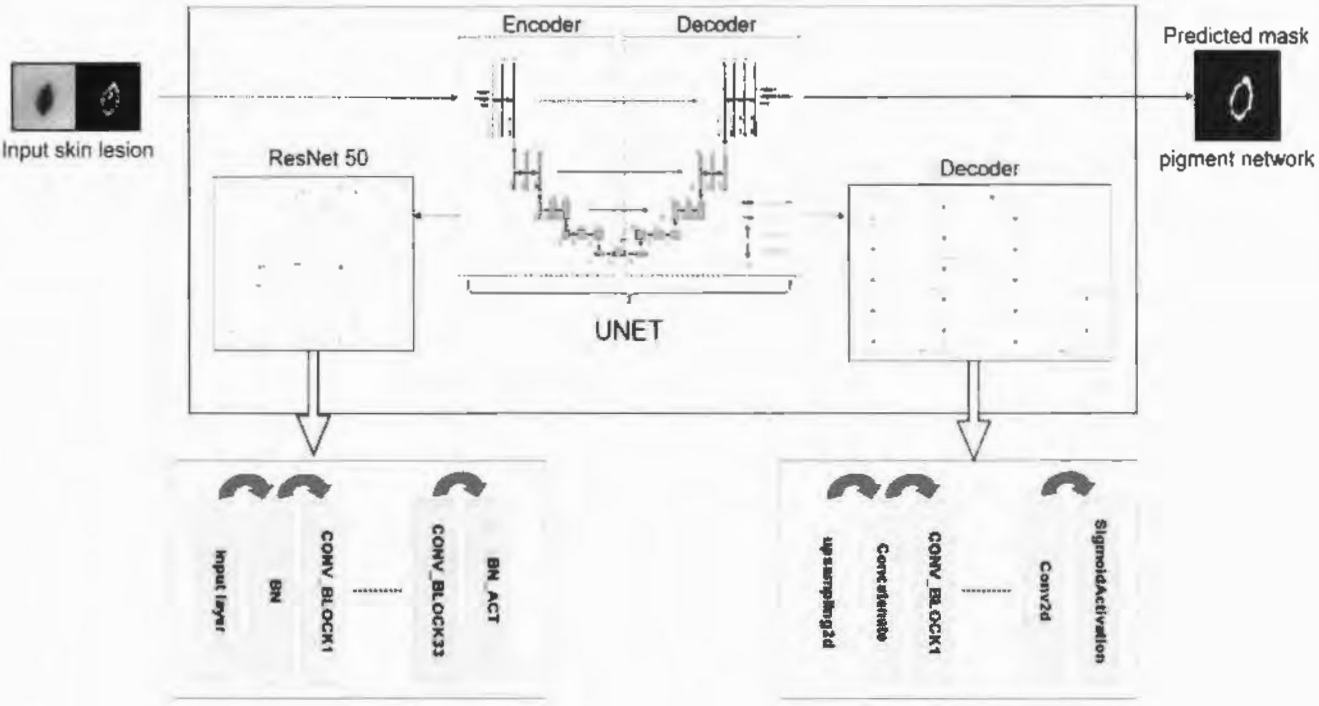


Figure 5.3: Attribute segmentation model overview

5.4.2 ResNet50-UNet Model

In our proposed work, we swapped the encoder block of UNet with ResNet50 model to make use of residual mapping. The architecture of the encoder is shown in Figure 5.4. In this, the BN represents Batch Normalization, MP indicates Maxpooling, while BN_ACT is Batch Normalization and ReLU activation. Furthermore, the CONV_BLOCK represents the convolution layer, batch normalization and activation layer as shown in Figure 5.5.

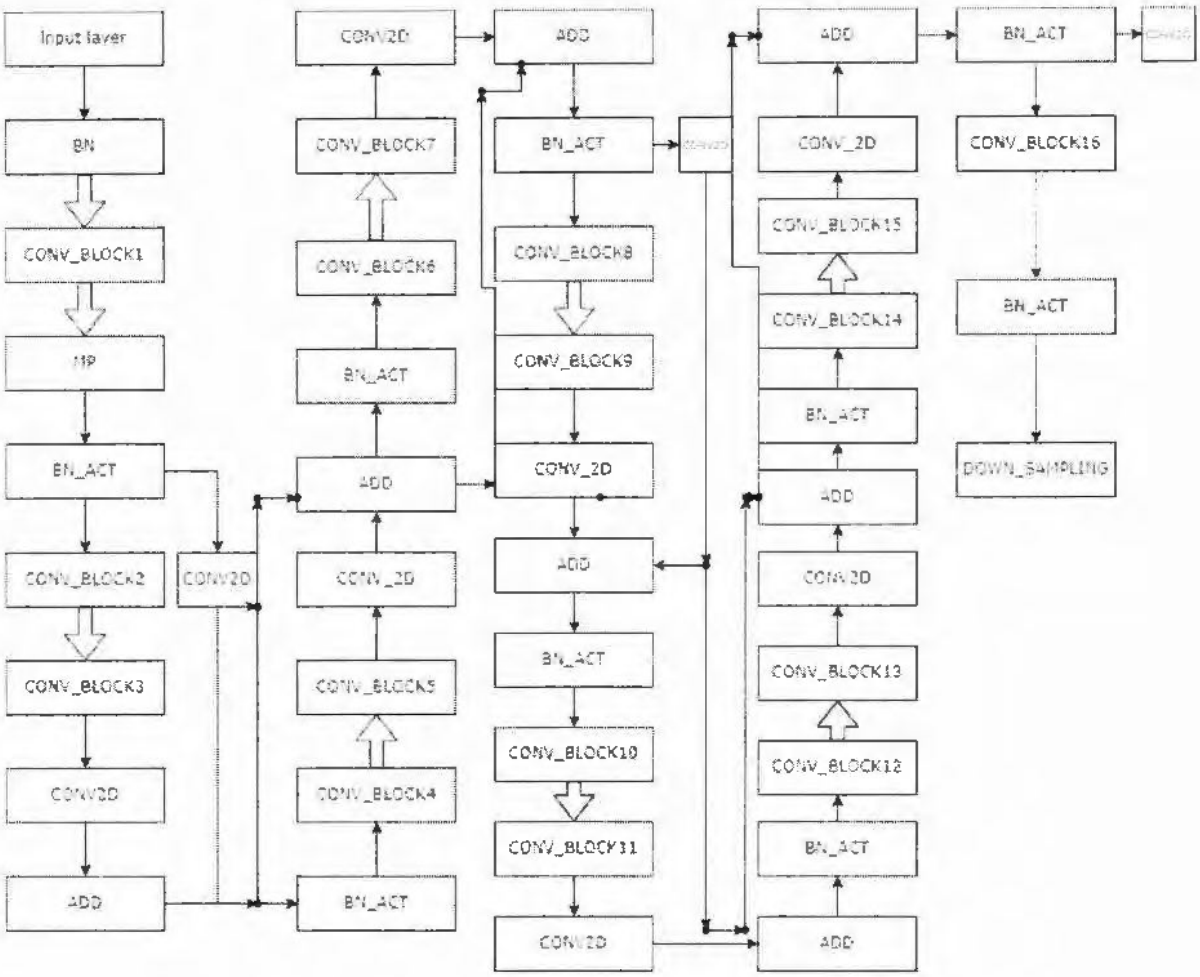


Figure 5.4: The encoder architecture (ResNet-50 backbone) in ResNet50-UNet

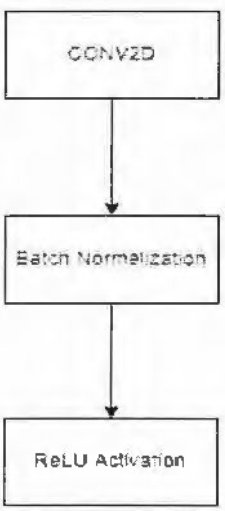


Figure 5.5: CONV_BLOCK

In Figure 5.4, the zero-padding layer is represented by the double arrow. Figure 5.4 shows half part of the ResNet50 architecture. Altogether, 33 CONV_BLOCKS are there. Unfortunately, due to space limitation they are restricted to CONV_BLOCK 16 over here. The remaining CONV_BLOCKS follow the same arrangement. In this, zero padding is used to connect two CONV_BLOCKS. A part of addition comes after every two CONV_BLOCKS. For this, a layer for addition is included to the existing layer as shown in Figure 5.4. Furthermore, after activation layer there are two CONV2D layers at various places. To concatenate part of the upsampling layers these activation layers are connected. A final batch normalization layer and activation layer is there after CONV_BLOCK 33. Finally, the upsampling section starts after that. An entire architecture of the upsampling portion is given in Figure 5.6. A concatenate layer and two CONV_BLOCKS are there after upsampling the layer. From the encoder part, the activation functions are concatenating to the upsampling layer of decoder part through all the four concatenation layers. The attribute segmentation is performed in the decoder part of the model while upsampling the feature maps.

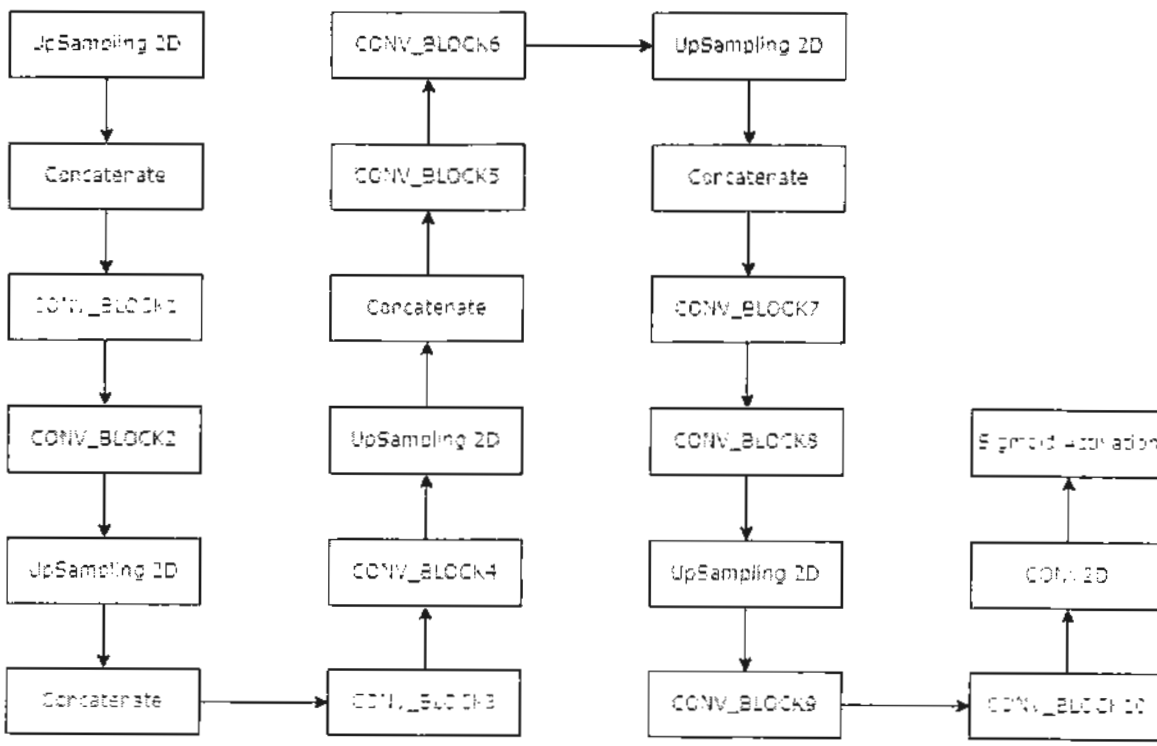


Figure 5.6: The decoder block in ResNet50-UNet architecture.

5.4.3 Hybrid Loss Function

Even though the data imbalance issue is resolved by downsampling the maximum class, the ratio of the positive class is still very low. Segmentation is applied as per pixel classification. This task is binary segmentation such as the positive and negative class. The negative class represents black color and positive class represents white color. The negative class is still dominating over the positive class. It is because the attributes occupy a very less area, for example, here is attribute image of milia-like cyst. The sample mask of the presence of attribute, like milia, in skin lesion image is shown in Figure 5.7.

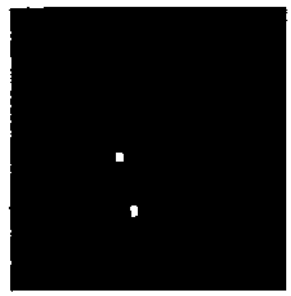


Figure 5.7: Mask of the presence of milia like cyst in skin lesion image.

The positive class is occupying very little space as shown in Figure 5.7. Most of the space is captured by the negative class in this image. This is the image with positive class in it. One more issue is the presence of blank images in the given dataset. Resultantly, after downsampling the images to positive class, the data imbalance issue remains there. To resolve this issue, specific loss functions are used to handle imbalance. For the given problem in hand, we have used hybrid loss function, a combination of focal Tversky loss and IoU segmentation loss as given in equation (5.1).

$$Hybrid\ loss = focal\ Tversky\ loss + iou\ segmentation \tag{5.1}$$

5.4.3.1 Focal Tversky Loss

The focal Tversky loss was proposed [4] to handle the issue of class imbalance. The mathematical formulation of the focal Tversky loss is given in equation (5.2)

$$FTL_c = \sum_c (1 - TI_c)^{1/\gamma} \tag{5.2}$$

where the value of γ is within the range [1,3] and TI is Tversky index. TI is the generalization of the dice loss function. By using this, the balance between False positive and false negative can be made flexible. The mathematical representation of TI is given in equation (5.3).

$$TI_c = \frac{\sum_{i=1}^{N} p_{ic}\, g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{i\bar{c}} g_{ic} + \beta \sum_{i=1}^{N} p_{ic} g_{i\bar{c}} + \epsilon} \tag{5.3}$$

Here, p_{ic} is the probabolity that pixel i belongs to lesion class c, and $p_{i\bar{c}}$ is the probability that pixel i belongs to non lesion calss $\bar{c}$. Same is case for g_{ic} and $g_{i\bar{c}}$. α *and* β are hyperparameters that should be tuned according to the ratio of class.

In Focal tversky loss, if a particular pixel is misclassified and the value of TI is greater, the loss remains unchanged. On the other side, if the TI is small and the pixel is misclassified by model, the value of focal tversky will decrease as a result. If the value of γ is greater then 1, it shows that the less accurate predictions are more focused by FTL. The work presented in [4] experimented numerous values of γ and concluded that 4/3 peformed best. In our proposed work, we have used the same concluded value.

5.4.3.2 IOU Segmentation Loss

To evaluate the performance of segmentation models, the IoU score is used as an evaluation measure. It calculates the similarity of actual pixels with predicted ones. Mathematically, IoU can be defined as given in equation (5.4).

$$IoU = \frac{TP}{FP+TP+FN} \tag{5.4}$$

where, TP indicates true positive, FP is false positive, and FN as false negative. Although the IoU score is a count-based metric, the output of the model is typically the computed probabilities of the pixels about their belonging to various classes. Therefore, the IoU count is given that approximated over the probabilities. The IoU count is defined in equation (5.5).

$$IoU = \frac{I(X)}{U(X)} \tag{5.5}$$

Where $I(X)$ can be approximated as given in equation (5.6)

$$I(X) = \sum_{v \epsilon V} X_v * Y_v \tag{5.6}$$

X represents the output of the model for a set of pixels V of the images. Y represents the actual pixels and $Y \epsilon \{0,1\}^V$. The mathematical form of approximating $U(X)$ is given in equation (5.7).

$$U(X) = \sum_{v \epsilon V} (X_v + Y_v - X_v * Y_v) \tag{5.7}$$

The IoU loss is as given in equation (5.8).

$$L_{IoU} = 1 - IoU = 1 - \frac{I(X)}{U(X)} \tag{5.8}$$

The objective function is given in equation (5.9).

$$\arg\min_{w} L_{IoU} = 1 - IoU \tag{5.9}$$

This equation (5.9) is solved by using stochastic gradient descent. The final output is computed as given in equation (5.10).

$$\frac{\partial L_{IoU}}{\partial X_v} = \begin{cases} \frac{-1}{U(X)} & if\ Y_v = 1 \\ \frac{I(X)}{U(X)^2} & otherwise \end{cases} \tag{5.10}$$

After calculating the gradients, the derivatives can be computed by applying the chain rule. These two different loss functions are combined to achieve a hybrid loss function. Both loss functions are good at dealing with class imbalance, so their hybrid is applied in our proposed work.

5.5 Experiments and Results

In this section, the experiments are performed to assess the performance of our proposed structural segmentation model. These experiments were conducted on ISIC 2018 segmentation dataset.

5.5.1 Dataset

To perform our experiments, we have used ISIC 2018 dataset available for the attribute segmentation task. The available dataset is highly imbalanced. There is a total of 2594 images. For each image, five labels are there representing the names of attributes or structures given in ISIC 2018 segmentation dataset. These five attributes are not equally present in each image. Table 5.2 shows the distribution of attributes in lesion image. From this table, it can be seen that there are 3098 images in total. This is because, in one image there could be more than one structures present. The ground truth masks for these images are available separately for each attribute. Due to this, the number of images increases against each individual attributes.

Table 5.2: Distribution of attributes in ISIC 2018 skin lesion image dataset of segmentation

Attributes	Number of Images Present	Percentage of images
Streaks	100	3.2%
Pigment Network	1523	49.2%
Globules	603	19.5%
Negative Network	190	6.1%
Milia-like cysts	682	22.0%
Total	3098	100%

To achieve our proposed objective, each structure is segmented separately in our proposed model. Resultantly, five segmentation models are trained for five structures in skin lesion images independently. Moreover, as the data is highly imbalanced, the down-sampling technique is used to resolve the data imbalance problem along with hybrid loss function.

5.5.2 Evaluation Measures

Mean-IoU score is used as evaluation measure as follows:

$$Mean_{IoU} = \frac{1}{N}\sum_{i=1}^{N} IoU \tag{5.12}$$

5.5.3 Experimental Setup

To evaluate the performance of our proposed model, all experiments are conducted on Kaggle and Colab platforms using Jupyter Notebook. Firstly, the ISIC dataset was prepared to rescale the input images to a resolution of 512×512. This rescaling is performed to make images compatible with the model during training. After that, the whole dataset was split into training, validation, and test sets with a ratio of 80:20. In the UNet architecture, we have used ResNet-50 model as a backbone. While performing experiments, we set the $batch\ size = 8$ due to the limitations of available RAM on Kaggle, as exceeding this value caused memory crash. The parameters' settings for the proposed model are given in Table 5.3. To train our proposed model, we executed $\#\ of\ epochs = 60$ by setting early stopping criteria. The ADAM optimizer is utilized along $learning\ rate = 0.001$. The loss function that we have employed in our proposed model is given in equation (5.1). To assess the performance of our proposed model, we used the Mean Jaccard index or IoU as an evaluation measure. The weights in the encoder part, in our proposed model, are learned from the Imagenet dataset as they are the best learned weights showing excellent performance. The parameters and hyperparameters of our proposed model are fine-tuned to optimize its performance and to achieve the best possible scores. During fine-tuning of the proposed model, various specific layers were intentionally removed or added to ResNet-50 encoder, depending on the specific requirements of our proposed model.

Table 5.3: Proposed model training parameters settings

Model Training Parameters	Values
Image Resolution	512 x 512 x 3
Batch size	8
No. of epochs	60
Early Stopping at	
Globules	36
Pigment Network	49
Negative Network	59

Model Training Parameters	Values
Streaks	50
Milia	30
Learning Rate	0.001
Optimizer	Adam
Loss function	Hybrid

5.5.4 Experimental Results and Discussion

To assess the performance of our proposed model, we have used the mean Intersection over Union (IoU) as an evaluation measure. To get mean IoU, we calculated the intersection over union for each class and then took the average of these computed IoU values. The obtained results of the mean IoU scores are shown in Table 5.4.

Table 5.4: Mean IoU test data results of structural segmentation of skin lesion using proposed model.

Class / Attribute	Mean IOU
Pigment Network	0.67
Negative Network	0.58
Milia like cysts	0.53
Globules	0.66
Streaks	0.53

By considering the obtained mean IOU scores for various attributes or classes, the performance of structure segmentation model can be summarized as follows:

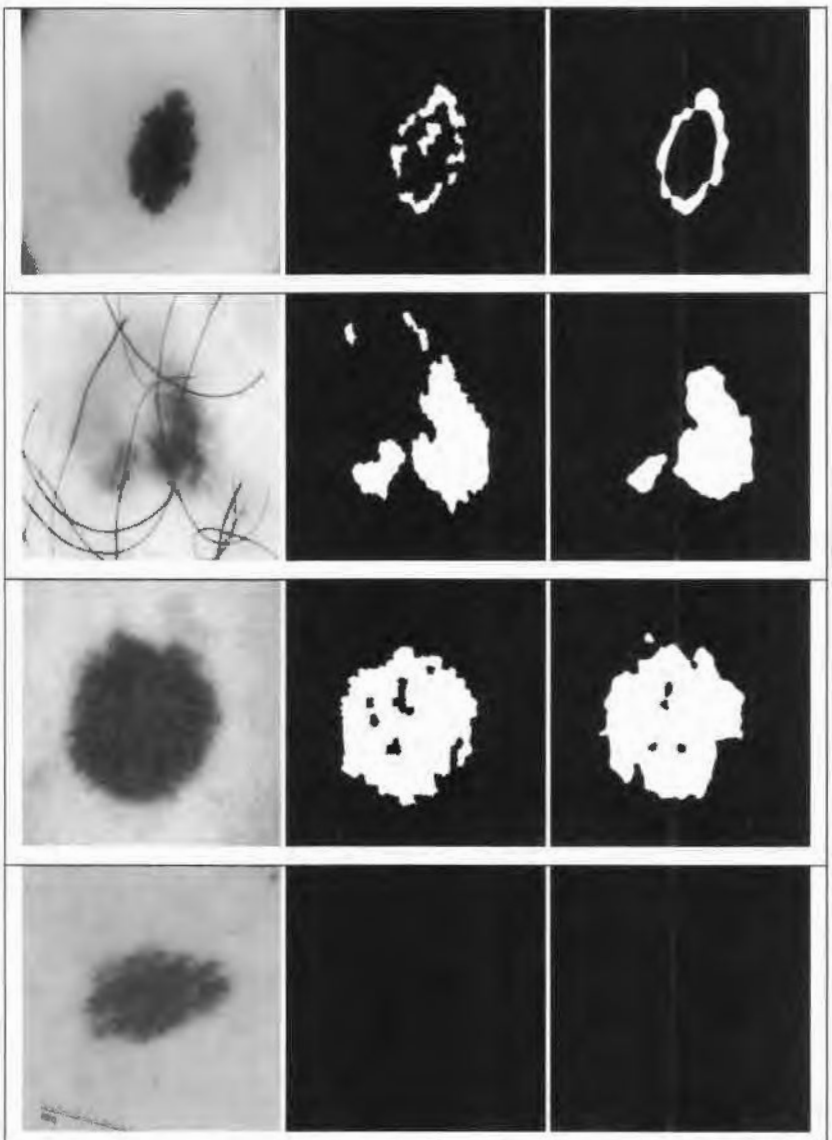
One of the attributes named pigment network is successfully segmented by our proposed model by achieving the mean IoU score 0.67. The obtained result shows that the proposed model has successfully identified the patterns of pigment network present in the skin lesion and its boundaries as well. The negative network attribute is segmented with a mean IOU score of 0.58. Even though the acquired score is lower than the IoU score of pigment network, the model shows better performance to segment the presence of network patterns in the lesions. Another attribute named milia-like cysts segmented well by achieving mean IOU score 0.53. The proposed model showed an average score to identify and segment that pattern accurately in the skin lesion. While evaluating the performance of proposed segmentation model, we observed that it faced various challenges while capturing the boundaries of structures and the patterns itself accurately that are associated with them. Furthermore, the proposed ResNet-UNet model obtained a mean IOU score of 0.66 while segmenting the attribute of globules in the lesions. Globule's structure plays its vital role to determine various types of skin lesion. Fortunately, the proposed model has shown better performance for the identification and segmentation of

this pattern within the skin lesion. The streaks attribute also achieved a mean IOU score of 0.53 showing the good model's performance to segment streak like structures.

The experimental results of proposed Res-UNet model for structural segmentation of attributes such as pigment network, negative network, milia like cyst, globules, and streaks, present in skin lesion image are shown in Figures 5.8, 5.9, 5.10, 5.11, 5.12 . The results show the original attribute masks and predicted mask by the proposed model. These results show that the proposed framework has performed well, and it gives acceptable results for attribute level segmentation. Presence of pigment network alone or with milia-like cyst in one lesion leads toward the presence of malignant lesion [109]. On the other side, absence of pigment and the presence of milia in one lesion leads toward benign lesion [109]. Pigment Network is a primary attribute which plays an important role in the diagnosis of malignant melanoma and other pigmented skin lesions. Identification and segmentation of this structure accurately, by a diagnostic system, ean assist a dermatologist to distinguish benign lesions from potentially malignant ones.

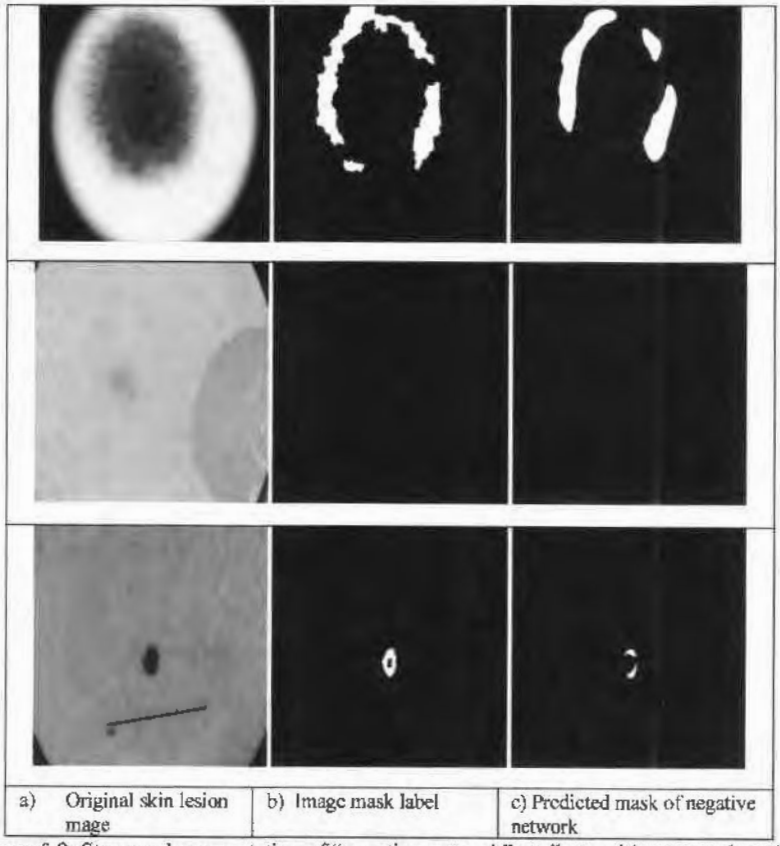
Negative networks are very challenging and scarce to detect. These are thin hair-like structures having low contrast against the texture of skin lesion [110]. They are normally light as compared to the main color of lesion. Due to this, they may be easily skipped as white colored hairs in the image. This negative network attribute may be present in nevus as well as in melanoma type of skin lesion. It is very difficult to separate nevus from melanoma but the existence of negative network always gave indication of melanoma [110]. The negative network attribute is frequently observed in melanoma skin lesion type. Our proposed model performed well to detect this complex negative network attribute even in the presence of these hurdles that are discussed above. Our model gives IoU = 0.58 for negative network attribute segmentation. In Figure 5.9, the predicted masks of negative network are given which leads towards the presence of melanoma skin lesion type.

Similarly, it is also very difficult to distinguish between basal cell carcinoma and melanoma if globules attribute is present in both. Streaks attribute may appear at the exterior edge of a skin lesion [110]. Its presence can possibly reveal different skin lesions, forcing stronger association towards melanoma. However, it is important to observe that the presence of streaks attribute alone does not provide a sufficient diagnosis for melanoma or any other specific skin lesion. To achieve a more accurate diagnosis, it is necessary to consider the presence of other attributes in conjunction with streaks.



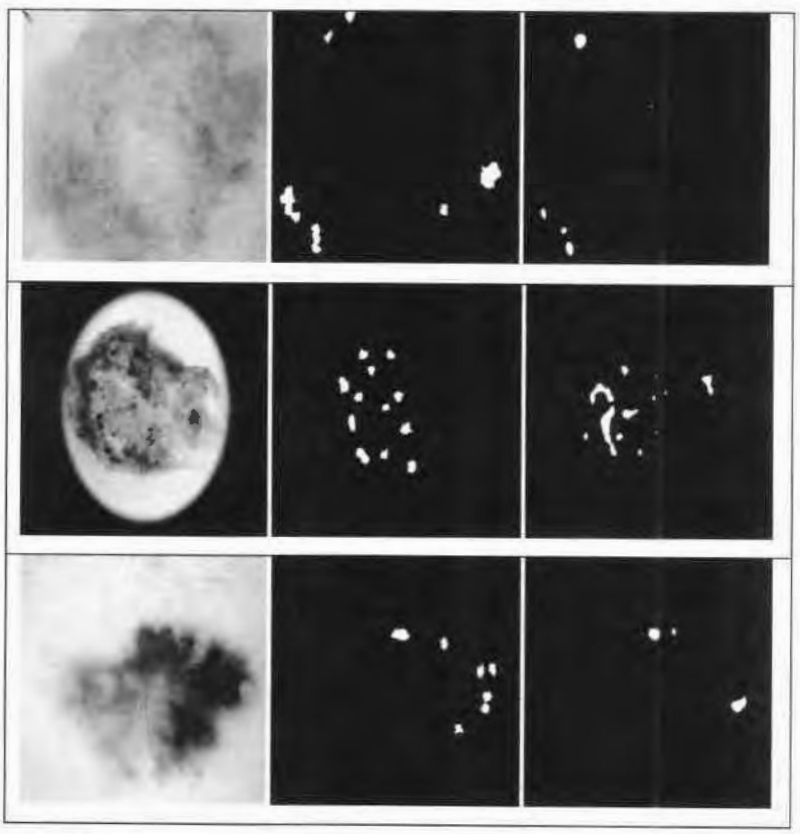
a) Original skin lesion image	b) Image mask label	c) Predicted mask of pigment network

Figure 5.8: Structural segmentation of "pigment network" attribute with proposed model.



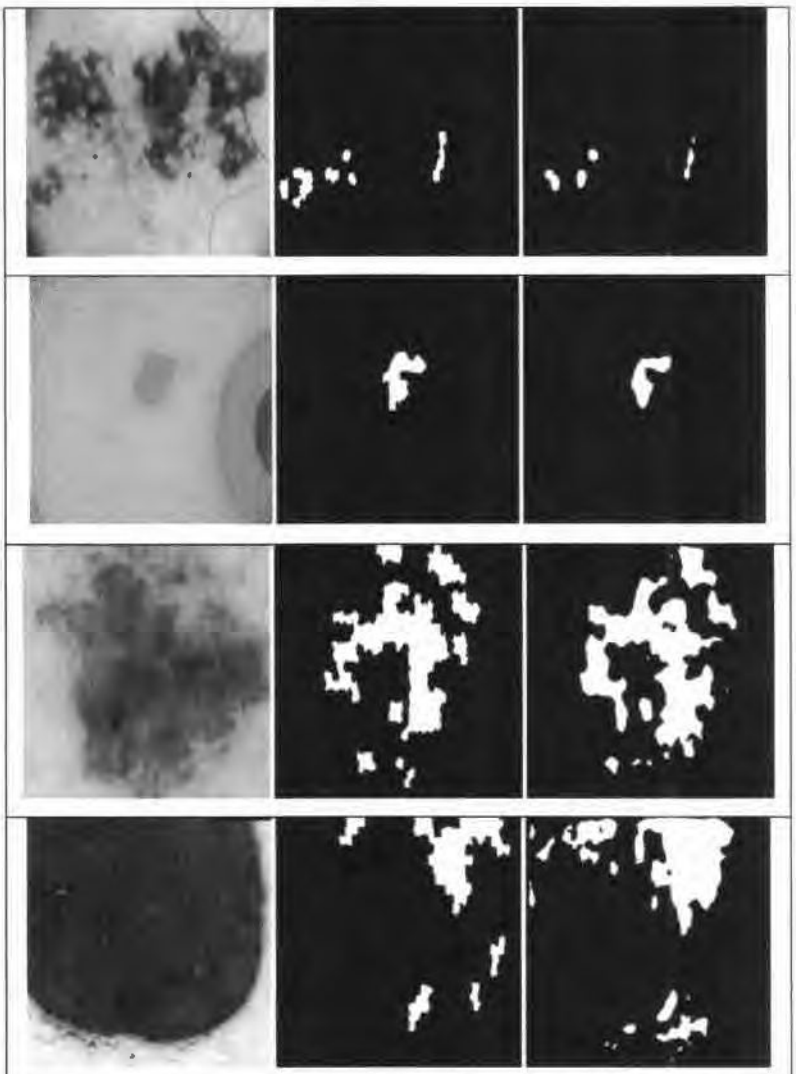
a) Original skin lesion mage	b) Image mask label	c) Predicted mask of negative network

Figure 5.9: Structural segmentation of "negative network" attribute with proposed model.



a) Original skin lesion mage	b) Image mask label	c) Predicted mask of milia like cyst

Figure 5.10: Structural segmentation of "milia like cyst" attribute with proposed model.



a) Original skin lesion mage	b) Image mask label	c) Predicted mask of globules attribute

Figure 5.11: Structural segmentation of "globules" attribute with proposed model.

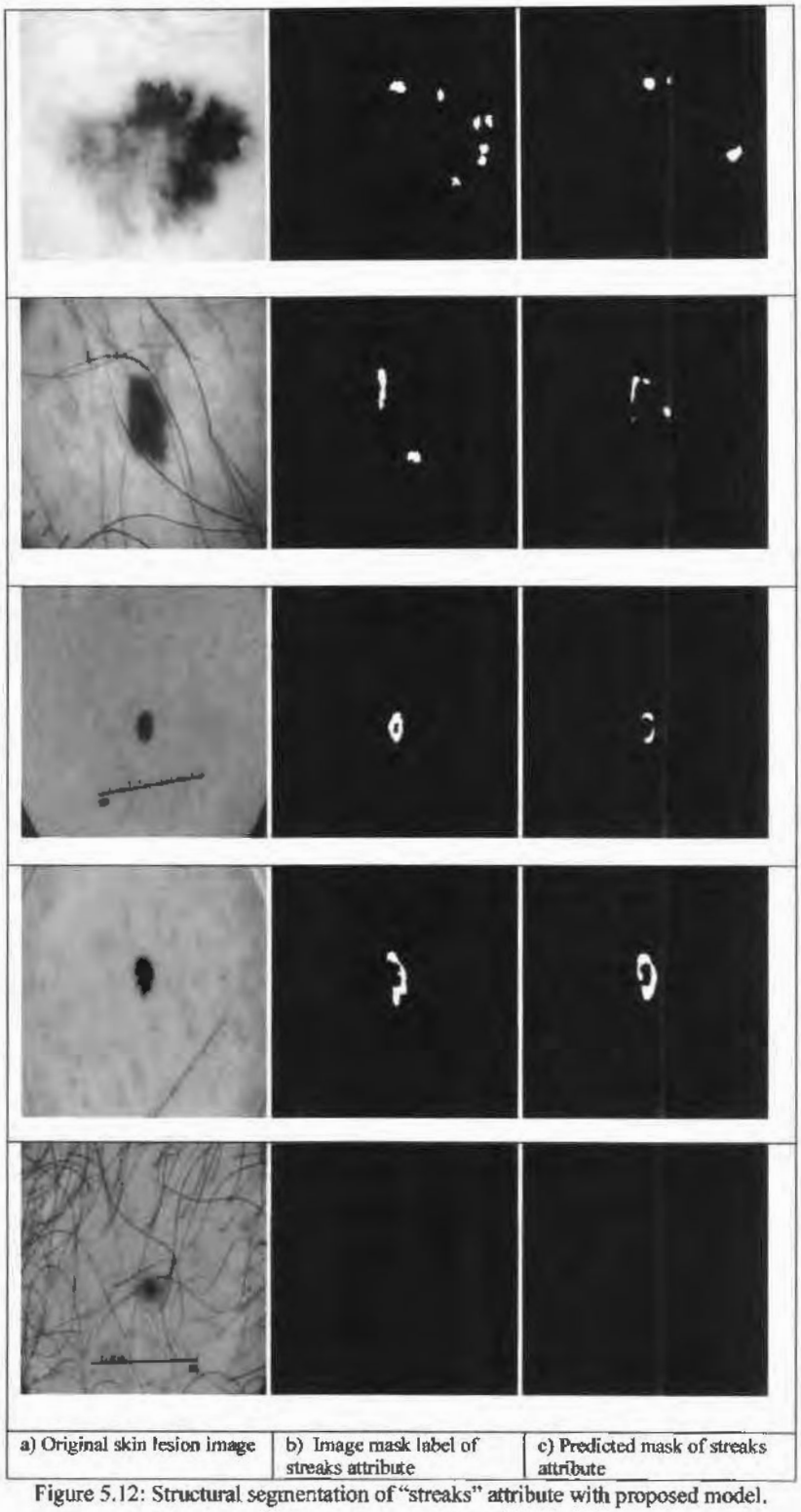


Figure 5.12: Structural segmentation of "streaks" attribute with proposed model.

The accurate identification and segmentation of dermoscopic structures by a segmentation model has a direct impact on the diagnostic accuracy of skin lesion type. The mean IOU values acquired for various structures give an assessment of how good the model has aligned with the ground truth segmentations. Greater mean IOU values indicate a better similarity between

predicted and true segmentations, representing improved diagnostic accuracy. It is particularly important in differentiating between benign and malignant lesions, as well as distinguishing between various other types of skin cancer. The training and validation loss curves for individual five learned models for each attribute (five attributes) are shown in Figure 5.13. The computational efficiency of our proposed model is shown in Table 5.5.

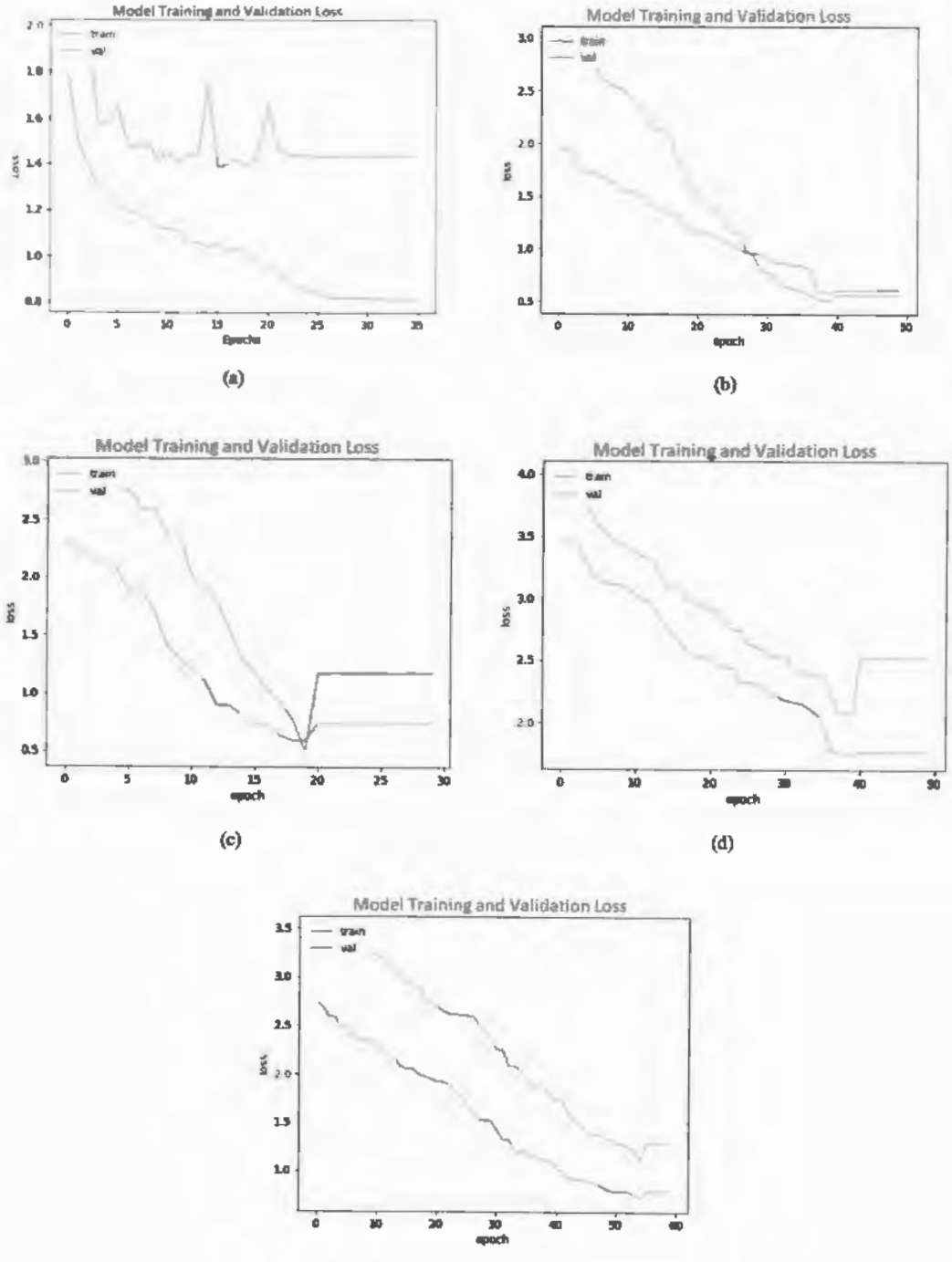


Figure 5.13: Training and validation loss curves of (a) Globules (b) pigment network (c) milia like cysts (d) Streaks (e) Negative attribute.

Table 5.5: Computational efficiency of proposed model

Factors	Values
Execution Time	02 hrs 40 minutes
Total No. of parameters	32,561,114
Trainable parameters	9,058,644
Non-trainable parameters	23,502,470
Space required	12 GB RAM

5.6 Comparison of Proposed model with Existing Techniques

A detailed comparison of the obtained results of our proposed ResNet50-UNet model with the existing state-of-the-art techniques is given in this section. We have also compared our proposed model with the winner of ISIC 2018 [101][95]. The performance of various backbone architectures, used with U-Net, is compared for skin lesion attribute segmentation with our proposed ResNet50-UNet by using our hybrid loss function. The results of comparison of IoU (Jaccard Index) are shown in Table 5.6. Additionally, Table 5.6 shows a comparison of our proposed structural segmentation model for skin lesions with existing state-of-the-art approaches. We analyzed the performance of various backbone architectures used with U-Net for skin lesion attribute segmentation. The Mean IoU is used as an evaluation measure, and the results are shown in Table 5.4. Our proposed model shows vigorous performance as compared to existing state-of-the-art techniques for attribute level segmentation of skin lesions.

The ResNet50-UNet model shows better performance as compared to the approach presented in [96] for the segmentation of negative network attribute. They applied various pre-processing techniques such as contrast reduction, hair occlusion, and sharpness improvement to enhance the performance of their segmentation. It is important to note that, our proposed model scored Jaccard similarity or IoU of 0.58 against negative network attribute segmentation, while [96] reported IoU values of 0.149, 0.189, 0.213, and 0.228 with base networks ResNet151, ResNetv2, DenseNet169, and their proposed ensemble, respectively. Unlike [96], we did not apply any dataset augmentation or pre-processing techniques, yet our proposed method yielded better results in comparison.

Moreover, our method presented significantly better results for attribute segmentation compared to the findings in [92]. In [92], they employed b0-EfficientNet as the backbone in conjunction with UNet and LinkNet architectures. Our proposed model also performed well when compared to the attention UNet model presented in [94]. Similarly, our proposed model exhibited competitive performance against the second-ranked winner of the ISIC 2018

challenge [97] and NMN's method [96], which come forward as the winner of the ISIC 2018 challenge.

Table 5.6: Comparison of proposed model with state-of-the-art methods measuring IOU score.

Sr. No.	Method	Pigment Network	Globules	Milia-Like Cysts	Negative Network	Streaks
1	ResNet-151 [96]	0.527	0.304	0.144	0.149	0.125
2	ResNet-v2 [98]	0.539	0.310	0.159	0.189	0.121
3	DenseNet-169 [98]	0.538	0.324	0.158	0.213	0.134
4	b0-EfficientNet [92]	0.554	0.324	0.157	0.213	0.139
5	U-Eff (TATL) [99]	0.565	0.373	0.157	0.268	0.243
6	L-Eff(TATL) [99]	0.562	0.356	0.168	0.292	0.252
7	Ensemble [98]	0.563	0.341	0.171	0.228	0.156
8	Attention UNet [94]	0.535	0.312	0.162	0.187	0.197
9	LeHealth method (Second ranked ISIC 2018 challenge) [97]	0.482	0.239	0.132	0.225	0.145
10	NMN's method [96]	0.544	0.252	0.165	0.285	0.123
11	SANet [97]	0.576	0.346	0.251	0.286	0.248
12	**Proposed ResNet-50 with U-Net**	**0.67**	**0.66**	**0.53**	**0.58**	**0.53**

5.7 Discussion

In this chapter, we have discussed the issue of dermoscopic segmentation of skin lesions at attribute level. A novel ResNet50-UNet framework is presented that uses ResNet-50 architecture as backbone in UNet model. ADAM optimizer is used to update the weights of the network. To handle the class imbalance problem, present in ISIC 2018 task 2 attribute segmentation dataset, we have proposed a hybrid of two loss functions named focal Tversky loss and IoU loss functions. By using this proposed loss function, the data imbalance problem is handled to some extent.

Experiments are conducted to compare the performance of proposed attribute segmentation model with existing state-of-the-art approaches such as U-Eff (TATL), L-Eff (TATL), attention UNet, LeHealth, NMN, and SANet. It has been shown that the proposed attribute segmentation model gives better segmentation results as compared to the competitors, in the presence of all five attributes of skin lesion. In discussion, we have shown that the analysis of dermoscopic structures present in skin cancer lesions provides valuable information for the diagnosis and management of skin cancer. By carefully examining the pigment network, negative network, milia-like cysts, globules, and streaks, dermatologists can make informed decisions and

differentiate between benign and malignant lesions. The analysis of our obtained results for dermoscopic structures in skin cancer lesions from a medical perspective highlights the importance of accurate identification and segmentation of these structures for diagnostic accuracy and patient care. The obtained results provide insights into the model's performance for different structures, indicating areas of strength and potential improvement.

Chapter 6

Conclusion and Future Work

Chapter 6

6 Conclusion and Future Work

6.1 Conclusion

This thesis has set to propose CNN based classification models for the hierarchical classification of skin lesions and a UNet based segmentation model for the dermoscopic structure segmentation of skin lesion. Mainly, the thesis is geared towards solving the challenges in classification of skin lesions at taxonomic levels to embed the explainability in a CNN model. Further, the thesis is concerned to deal with the challenge of complex attribute level skin lesion segmentation that are present as structures in skin lesion. There is a great challenge to segment the skin lesion attribute as they have similar and overlap structures among them and there is a low contrast present in some of them. These challenges lead towards creating hurdles for the practitioners during the diagnosis process of skin lesion and their attributes.

Comprehensive literature survey (Chapter 3) concluded that although many approaches have been proposed for traditional skin lesion fine level classification and for the boundary level segmentation of skin lesion including a few approaches for attribute level segmentation of skin lesions. There is still a great need to introduce the explainability in the deep learning based models for skin lesion classification and structural level segmentation. This thesis is a small step toward this effort.

6.2 Contribution

1. The problem of three-level hierarchical classification of skin cancer lesions, by localizing and visualizing their structural level features is resolved to make the CAD system self-explainable. The proposed method achieved 91% sensitivity, specificity score 98% and AUC score 95%. Our experimental results proved that the proposed method performs well as compared to competitor.
2. We have proposed a model for the structural level segmentation and localization of skin lesion's features. The proposed model works well in the presence of imbalance dataset. This attribute segmentation model has shown promising results. Promising mean IoU scores are obtained for all five attributes segmentation.

3. The problem of class imbalance present in ISIC skin cancer dataset has been addressed to make the hierarchical classification and structural segmentation more robust. To resolve the class imbalance issue present in ISIC Task 2 attribute segmentation dataset, we proposed a hybrid loss function that performs well for imbalance dataset.
4. The proposed system also provides visualization at each hierarchical level of classification, through gradient class activation maps, which highlight the image regions that contribute the most to predict output and that are interpretable.

6.3 Limitations

Our proposed work has certain limitations.

1. The available standard ISIC 2018 has separate datasets for classification and segmentation tasks. It does not provide comprehensive and combined information about both the skin lesion type and its specific structural / attributes level detail within one dataset.
2. Due to the above reason, we are not able to know about the specific names or labels of the attributes present in the skin lesion after determining its hierarchy.
3. We have produced GradCAMs visualizations, but these visualizations do not talk about label of that specific attribute identified in that region. It only captures the region, and we visualize it through GradCAMs and getting only the knowledge about participated region in decision making.
4. We have not determined the skin lesion type after performing attribute level segmentation as we do not have the class labels in hand about that specific skin lesion type.
5. Due to resource limitations in hand, we are not able to train our proposed models on large dataset. The performance of the proposed model could be enhanced by increasing the hardware resources as we will be more able to load large datasets and process it.

6.4 Future Enhancements

This study is focused on three levels hierarchical classification of skin lesions by predicting a sequence of hierarchy and complex attribute level segmentation of skin lesions. Although acceptable results are obtained in all proposed modules , however, there are still future directions that need to be considered.

1. We are intending to work more on embedding further hierarchical levels in deep learning based models to make them self-explainable.
2. In future, we are planning to embed further deep hierarchy of skin lesions in other various deep learning based models.

3. There is need to work on more attributes of skin lesions and to handle the overlapping of these structures present in lesions that leads towards creating confusions during the decision of skin lesion types.
4. There is still needed to further improve the explainability of the CNN models as it can be seen from our proposed models obtained results that it is not performing well on some images especially in structural segmentation of skin lesions.

7 References

[1] J. Ferlay *et al.*, "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012," *Eur. J. Cancer*, vol. 49, no. 6, pp. 1374–1403, 2013, doi: 10.1016/j.ejca.2012.12.027.

[2] NIH National Cancer Institute, "Cancer Facts & Figures 2020," *CA. Cancer J. Clin.*, pp. 1–76, 2020.

[3] C. B. Rogers HW, Weinstock MA, Feldman SR, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population. JAMA Dermatol.," 2015.

[4] "American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society; 2019.," 2019.

[5] J. A. Siegel RL, Miller KD, "Cancer statistics, 2019. CA Cancer J Clin. 2019," 2019.

[6] F. B. Shahid Mahmood, Raqib Faraz, Aneel Yousaf, Ain ul Quader, Hina Asif, Adna Atif, Lubna Nadeem, Natasha Perveen, Roma Tanveer, Azam Hussain, "Collective Cancer Registry Report From December 1994 Till December 2018, Of The Shaukat Khanum Memorial Cancer Hospital & Research Center, Pakistan," 2018.

[7] F. B. Aneel Yousaf, Shahid Mahmood, Raqib Faraz, Ain ul Quader, Hina Asif, Adna Atif, Lubna Nadeem, Natasha Parveen, Roma Tanveer, Azam Hussain, "Annual Cancer Registry Report-2018, Of The Shaukat Khanum Memorial Cancer Hospital & Research Center, Pakistan," 2018.

[8] P. F. S. of Cancer, "International Agency for Research on Cancer World Health Organization (WHO)," 2018.

[9] "https://dermnetnz.org/topics/the-structure-of-normal-skin/." [Online]. Available: https://dermnetnz.org/topics/the-structure-of-normal-skin/. [Accessed: Oct 2022]

[10] H. Zalaudek, Iris and Argenziano, Giuseppe and Di Stefani, Alessandro and Ferrara, Gerardo and Marghoob, Ashfaq A and Hofmann-Wellenhof, Rainer and Soyer, H Peter and Braun, Ralph and Kerl, "Dermoscopy in general dermatology," *Dermatology*, vol. 212, no. 1, pp. 7–18, 2006.

[11] K. W. H. Pehamberger, A. Steiner, "In vivo epiluminescence microscopy of pigmented skin lesions. I: Pattern analysis of pigmented skin lesions.," *J. Am. Acad. Dermatol.*, vol. 17, pp. 571–583, 1987.

[12] S. M. Strayer and P. Reynolds, "Diagnosing skin malignancy: Assessment of predictive

clinical criteria and risk factors," *J. Fam. Pract.*, vol. 52, no. 3, pp. 210–218, 2003.

[13] P. B. W. Stolz, A. Riemann, A. B. Cognetta, L. Pillet, W. Abmayr, D. Holzel and and O. B.-F. F. Nachbar, M. Landthaler, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma.," *Eur. J. Dermatology,* vol. 4, pp. 521–527, 1994.

[14] A. B. W. Stolz, O. Braun-Falco, P. Bilek, M. Landthaler, W. H. C. Burgdorf and Cognetta., "colour Atlas of Dermatoscopy," *Blackwell, Berlin, 2 Ed.*, 2002.

[15] G. Fabbrocini, "Computer Vision Techniques for the Diagnosis of Skin Cancer," no. September, 2014, doi: 10.1007/978-3-642-39608-3.

[16] M. D. G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis.," *Arch. Dermatology,* vol. 134, pp. 1563–1570, 1998.

[17] W. H. M. S. W. Menzies, C. Ingvar, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma.," *Melanoma Res.*, vol. 6, no. 1, pp. 55–62, 1996.

[18] A. K. J. Henning, S. Dusza, S.Wang, A. Marghoob, H. Rabinovitz, D. Polsky, "The CASH (colour, architecture, symmetry, and homogeneity) algorithm for der- moscopy.," *J. Am. Acad. Dermatol.*, vol. 56, no. 1, pp. 45–52, 2007.

[19] J. López-Labraca, M. Á. Fernández-Torres, I. González-Díaz, F. Díaz-de-María, and Á. Pizarro, "Enriched dermoscopic-structure-based cad system for melanoma diagnosis," *Multimed. Tools Appl.*, 2017, doi: 10.1007/s11042-017-4879-3.

[20] K. G. Pathan, Sameena and Siddalingaswamy, PC and Lakshmi, Lasya and Prabhu, "Classification of Benign and Malignant Melanocytic Lesions : A CAD Tool," no. September, 2017.

[21] S. Jain, V. Jagtap, and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 736–741, 2015, doi: 10.1016/j.procs.2015.04.209.

[22] R. Sumithra, M. Suhil, and D. S. Guru, "c," *Procedia Comput. Sci.*, vol. 45, no. C, pp. 76–85, 2015, doi: 10.1016/j.procs.2015.03.090.

[23] U. Jamil, M. U. Akram, S. Khalid, S. Abbas, and K. Saleem, "Computer Based Melanocytic and Nevus Image Enhancement and Segmentation," *Biomed Res. Int.*, vol. 2016, 2016, doi: 10.1155/2016/2082589.

[24] R. Suganya, "An automated computer aided diagnosis of skin lesions detection and classification for dermoscopy images," *2016 Int. Conf. Recent Trends Inf. Technol. ICRTIT 2016*, 2016, doi: 10.1109/ICRTIT.2016.7569538.

[25] J. Garnavi, Rahil and Aldeen, Mohammad and Bailey, "Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1239–1252, 2012.

[26] N. C. F. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," *Proc. - Int. Symp. Biomed. Imaging*, vol. 2018-April, pp. 168–172, 2018, doi: 10.1109/ISBI.2018.8363547.

[27] N. Codella *et al.*, "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)," pp. 1–12, 2019.

[28] M. Barata, Catarina and Marques, Jorge S and Emre Celebi, "Deep Attention Model for the Hierarchical Diagnosis of Skin Lesions," 2019.

[29] C. Barata, M. E. Celebi, and J. S. Marques, "Explainable skin lesion diagnosis using taxonomies," *Pattern Recognit.*, vol. 110, p. 107413, 2021, doi: 10.1016/j.patcog.2020.107413.

[30] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020, doi: 10.1109/ACCESS.2020.3003890.

[31] H. Tschandl, P., Rosendahl, C. & Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.," *Sci. Data 5, 180161*, vol. doi:10.103, 2018.

[32] J. M. Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, "'BCN20000: Dermoscopic Lesions in the Wild,'" *arXiv:1908.02288*, 2019.

[33] "https://www.isic-archive.com/." [Online]. Available: https://www.isic-archive.com. [Accessed: January 2022]

[34] "https://dermnetnz.org/cme/lesions/." [Accessed: Nov 2022]

[35] DVD-ROM, "Dermoscopy: A practical guide.," *Am. Acad. Dermatology*.

[36] M. Scott M. Strayer, MD, MPH, Peter Reynolds, "Diagnosing skin malignancy:

Assessment of predictive clinical criteria and risk factors," *J. Fam. Pract.*, vol. 52, no. 3, pp. 210–218, 2003.

[37] A. M. Loane, H. Gore, R. Corbett, K. Steele, C. Mathews, S. Bloomer, D. Eedy, R. Telford and R. Wootton, "Effect of camera performance on diagnostic accuracy: preliminary results from the northern ireland arms of the uk multicentre teledermatology trial," *J. Telemed. Telecare,* vol. 3, no. 2, pp. 83–88, 1997.

[38] D. M. S. Singh, J. Stevenson, "An evaluation of polaroid photographic imaging for cutaneous-lesion referrals to an outpatient clinic: a pilot study," *Br. J. Plast. Surg.*, vol. 54, no. 2, pp. 140–143, 2001.

[39] V. H.- G. Argenziano, H. P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, M. Delfino, A. Ferrari and and I. H. W. Wellenhog, D. Massi, G. Mazzocchetti, M. Scalvenzi, "Interactive Atlas of Dermoscopy.," *EDRA Med. Publ. New Media*, 2000.

[40] A. Soyer, H Peter and Argenziano, Giuseppe and Zalaudek, Iris and Corona, Rosamaria and Sera, Francesco and Talamini, Renato and Barbato, Filomena and Baroni, Adone and Cicale, Lorenza and Di Stefani, "Three-point checklist of dermoscopy," *Dermatology*, vol. 208, no. 1, pp. 27–31, 2004.

[41] M. N. Bajwa *et al.*, "Computer-aided diagnosis of skin diseases using deep neural networks," *Appl. Sci.*, vol. 10, no. 7, pp. 1–13, 2020, doi: 10.3390/app10072488.

[42] V. N. Skladnev *et al.*, "Diagnostic Feature Extraction in Dermatological Examination," vol. 10/478078, 2004.

[43] J. B. M. Jamora, B. Wainwright, S. Meehan, "Improved identification of potentially dangerous pigmented skin lesions by computerised image analysis," *Arch. of Dermatology*, vol. 139, pp. 195–198, 2003.

[44] P. B. P. Rubegni, M. Burroni, G. Cevenini, R. Perotti, G. Dell'Eva and and L. A. M. Fimiani, "Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study," *J. Invest. Dermatol.*, vol. 119, pp. 471–474, 2002.

[45] A. R. K. Hoffmann, T. Gambichler, "Diagnostic and neural analysis of skin cancer. A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy," *Br. J. Dermatol.*, vol. 149, pp. 801–809, 2003.

[46] M. Elbaum, "Computer-aided melanoma diagnosis," *Dermatol. Clin.*, vol. 20, pp. 735–747, 2002.

[47] G. VD Argenziano G, Soyer HP, "Interactive Atlas of Dermoscopy," 2002.

[48] S. Garg and B. Jindal, "Skin lesion segmentation using k-mean and optimized fire fly algorithm," *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 7397–7410, 2021, doi: 10.1007/s11042-020-10064-8.

[49] N. Durgarao and G. Sudhavani, "Diagnosing skin cancer via C-means segmentation with enhanced fuzzy optimization," *IET Image Process.*, vol. 15, no. 10, pp. 2266–2280, 2021, doi: 10.1049/ipr2.12194.

[50] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," *Diagnostics*, vol. 9, no. 3, 2019, doi: 10.3390/diagnostics9030072.

[51] F. Navarro, M. Escudero-Viñolo, and J. Bescós, "Accurate Segmentation and Registration of Skin Lesion Images to Evaluate Lesion Change," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 2, pp. 501–508, 2019, doi: 10.1109/JBHI.2018.2825251.

[52] T. Akram, M. A. Khan, M. Sharif, and M. Yasmin, "Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features," *J. Ambient Intell. Humaniz. Comput.*, vol. 0, no. 0, p. 0, 2018, doi: 10.1007/s12652-018-1051-5.

[53] R. Kaur *et al.*, "Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images," *Ski. Res. Technol.*, pp. 416–428, 2016, doi: 10.1111/srt.12352.

[54] R. Kasmi, K. Mokrani, R. K. Rader, J. G. Cole, and W. V Stoecker, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," *Ski. Res. Technol.*, vol. 22, no. 2, pp. 208–222, 2016, doi: 10.1111/srt.12252.

[55] C. Y. Yu, W. S. Zhang, Y. Y. Yu, and Y. Li, "A novel active contour model for image segmentation using distance regularization term," *Comput. Math. with Appl.*, vol. 65, no. 11, pp. 1746–1759, 2013, doi: 10.1016/j.camwa.2013.03.021.

[56] A. Masood, A. Al-Jumaily, and K. Anam, "Self-supervised learning model for skin cancer diagnosis," *Int. IEEE/EMBS Conf. Neural Eng. NER*, vol. 2015-July, pp. 1012–1015, 2015, doi: 10.1109/NER.2015.7146798.

[57] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Comput. Med. Imaging Graph.*, vol. 71, pp. 19–29, 2019, doi: 10.1016/j.compmedimag.2018.10.007.

[58] F. Perez, S. Avila, and E. Valle, "Solo or Ensemble? Choosing a CNN Architecture for Melanoma Classification," 2019.

[59] M. Aminur Rab Ratul, M. Hamed Mozaffari, W. S. Lee, and E. Parimbelli, "Skin lesions

classification using deep learning based on dilated convolution," *bioRxiv*, 2019, doi: 10.1101/860700.

[60] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9352, no. October, pp. 118–126, 2015, doi: 10.1007/978-3-319-24888-2_15.

[61] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," *arXiv*, pp. 1–10, 2019.

[62] H. Liao, "A deep learning approach to universal skin disease classification.," *Univ. Rochester Dep. Comput. Sci. CSC*, 2016.

[63] H. Liao, Y. Li, and J. Luo, "Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks," *arXiv*, pp. 355–360, 2018.

[64] N. C. F. Codella *et al.*, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM J. Res. Dev.*, vol. 61, no. 4, pp. 1–28, 2017, doi: 10.1147/JRD.2017.2708299.

[65] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," *PLoS One*, vol. 14, no. 5, pp. 1–17, 2019, doi: 10.1371/journal.pone.0217293.

[66] A. M. Alqudah, H. Alquran, and I. A. Qasmieh, "Segmented and non-segmented skin lesions classification using transfer learning and adaptive moment learning rate technique using pretrained convolutional neural network," *J. Biomimetics, Biomater. Biomed. Eng.*, vol. 42, pp. 67–78, 2019, doi: 10.4028/www.scientific.net/JBBBE.42.67.

[67] T. De Vries and D. Ramachandram, "Skin lesion classification using deep multi-scale convolutional neural networks," *arXiv*, 2017.

[68] T. Akram *et al.*, "A multilevel features selection framework for skin lesion classification," *Human-centric Comput. Inf. Sci.*, vol. 10, no. 1, 2020, doi: 10.1186/s13673-020-00216-y.

[69] A. Romero Lopez, X. Giro-I-Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," *Proc. 13th IASTED Int. Conf. Biomed. Eng. BioMed 2017*, pp. 49–54, 2017, doi: 10.2316/P.2017.852-053.

[70] N. Ashour, A.S.; Nagieb, R.M.; El-Khobby, H.A.; Elnaby, M.M.A.; Dey, "Genetic algorithm-based initial contour optimization for skin lesion border detection.," *Multimed. Tools Appl.*, vol. 80, 2583–2, pp. 2583–2597, 2021.

[71] R. Mohakud, R.; Dash, "Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN.," *J. King Saud Univ.-Comput. Inf. Sci*, vol. 34, pp. 9889–9904, 2022.

[72] M. Kaur, R.; Gholam, H.; Sinha, R.; Lindén, "Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images.," *BMC Med. Imaging*, vol. 22, pp. 1–13, 2021.

[73] M. Bagheri, F.; Tarokh, M.J.; Ziaratban, "Skin lesion segmentation from dermoscopic images by using Mask R-CNN, Retina-Deeplab, and graph-based methods.," *Biomed. Signal Process. Control*, vol. 67, p. 102533, 2021.

[74] L. Qamar, S.; Ahmad, P.; Shen, "Dense Encoder-Decoder–Based Architecture for Skin Lesion Segmentation," *Cogn. Comput*, vol. 13, pp. 583–594, 2021.

[75] J. Wu, H.; Pan, J.; Li, Z.; Wen, Z.; Qin, "Automated Skin Lesion Segmentation Via an Adaptive Dual Attention Module," *IEEE Trans. Med. Imaging*, vol. 40, pp. 357–370, 2020.

[76] U. Öztürk, ,S.; Özkaya, "Skin Lesion Segmentation with Improved Convolutional Neural Network.," *J. Digit. Imaging*, vol. 33, pp. 958–970, 2020.

[77] J. Shan, P.; Wang, Y.; Fu, C.; Song, W.; Chen, "Automatic skin lesion segmentation based on FC-DPN," *Comput. Biol. Med*, vol. 123, p. 103762., 2020.

[78] G. Wei, Z.; Shi, F.; Song, H.; Ji, W.; Han, "Attentive boundary aware network for multi-scale skin lesion segmentation with adversarial training.," *Multimed. Tools Appl*, vol. 79, pp. 27115–27136, 2020.

[79] M. Khan, M.A.; Akram, T.; Zhang, Y.-D.; Sharif, "Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework.," *Pattern Recognit. Lett*, vol. 143, pp. 58–66, 2021.

[80] H. Huang, C.; Yu, A.;Wang, Y.; He, "Skin Lesion Segmentation Based on Mask R-CNN," *Proc. 2020 Int. Conf. Virtual Real. Vis. (ICVRV), Recife, Brazil,*.

[81] E. Ünver, H.M.; Ayan, "Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm," *Diagnostics*, vol. 9, 72., 2019.

[82] M. A. Shahin, A.H.; Amer, K.; Elattar, "Deep Convolutional Encoder-Decoders with Aggregated Multi-Resolution Skip Connections for Skin Lesion Segmentation.," *Proc.*

2019 IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019), Venice, Italy, pp. 451–454.

[83] M. H. Goyal, M.; Oakley, A.; Bansal, P.; Dancey, D.; Yap, "Skin Lesion Segmentation in Dermoscopic Images with Ensemble Deep Learning Methods.," *IEEE Access*, vol. 8, pp. 4171–4181.

[84] S. Lameski, J.; Jovanov, A.; Zdravevski, E.; Lameski, P.; Gievska, "Skin lesion segmentation with deep learning," in *In Proceedings of the IEEE EUROCON 2019-18th International Conference on Smart Technologies, Novi Sad, Serbia,* pp. 1–5.

[85] S. Hasan, S.N.; Gezer, M.; Azeez, R.A.; Gulsecen, "Skin Lesion Segmentation by using Deep Learning Techniques.," in *In Proceedings of the 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey,* pp. 1–4.

[86] B. Li, H.; He, X.; Zhou, F.; Yu, Z.; Ni, D.; Chen, S.;Wang, T.; Lei, "Dense Deconvolutional Network for Skin Lesion Segmentation.," *IEEE J. Biomed. Heal. Inform.*, vol. 23, pp. 527–537.

[87] R. Yan, Z. , Jagadeesh, V. , DeCoste, D. , Di, W. , & Piramuthu, "HD-CNN: Hierarchical deep convolutional neural network for image classification.," *Proc. Int. Conf. Comput. Vis. , 2 .*, 2015.

[88] Yan, Z. , Zhang, H. , Piramuthu, R. , Jagadeesh, V. , DeCoste, D. , Di, W., "HD-CNN: Hierarchical deep convolutional neural networks for large scale visual recognition.," *Proc. IEEE Int. Conf. Comput. Vi- sion (pp. 2740–2748) .*, 2015.

[89] M. Zhu, X., & Bain, "B-CNN: Branch convolutional neural network for hierarchical classification," *arXiv: 1709.09890 (Preprint).*, 2017.

[90] Y. Seo and K. shik Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert Syst. Appl.*, vol. 116, pp. 328–339, 2019, doi: 10.1016/j.eswa.2018.09.022.

[91] S. Benyahia, B. Meftah, and O. Lézoray, "Hierarchical Approach for the Classification of Multi-class Skin Lesions Based on Deep Convolutional Neural Networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13364 LNCS, pp. 139–149, 2022, doi: 10.1007/978-3-031-09282-4_12.

[92] D. M. H. Nguyen *et al.*, "TATL : Task agnostic transfer learning for skin attributes detection," *Med. Image Anal.*, vol. 78, p. 102359, 2022, doi: 10.1016/j.media.2022.102359.

[93] M. A. Kadir, F. Nunnari, and D. Sonntag, "Fine-tuning of explainable CNNs for skin

lesion classification based on dermatologists' feedback towards increasing trust," pp. 1–13, 2023.

[94] D. M. H. Nguyen, A. Ezema, F. Nunnari, and D. Sonntag, "A visually explainable learning system for skin lesion detection using multiscale input with attention u-net," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12325 LNAI, pp. 313–319, 2020, doi: 10.1007/978-3-030-58285-2_28.

[95] Y. . Zou, J., Ma, X., Zhong, C., Zhang, "Dermoscopic image analysis for ISIC challenge 2018," *arXiv Prepr. arXiv1807.08948*, 2018.

[96] M. Jahanifar, N. Z. Tajeddin, N. A. Koohbanani, A. Gooya, and N. Rajpoot, "Segmentation of Skin Lesions and their Attributes Using Multi-Scale Convolutional Neural Networks and Domain Specific Augmentations," pp. 1–18, 2018.

[97] P. Id, "SANet : Superpixel Attention Network for Skin Lesion Attributes Detection."

[98] A. Bissoto, F. Perez, V. Ribeiro, M. Fornaciali, S. Avila, and E. Valle, "Deep-Learning Ensembles for Skin-Lesion Segmentation, Analysis, Classification: RECOD Titans at ISIC Challenge 2018," pp. 1–5, 2018.

[99] I. Saitov, T. Polevaya, and A. Filchenkov, "Dermoscopic attributes classification using deep learning and multi-Task learning," *Procedia Comput. Sci.*, vol. 178, no. 2019, pp. 328–336, 2020, doi: 10.1016/j.procs.2020.11.034.

[100] J. Kawahara and G. Hamarneh, "Fully convolutional neural networks to detect clinical dermoscopic features," *IEEE J. Biomed. Heal. Informatics*, 2018.

[101] N. Koohbanani, N.A., Jahanifar, M., Tajeddin, N.Z., Gooya, A., Rajpoot, "Leveraging transfer learning for segmenting lesions and their attributes in dermoscopy images," *arXiv Prepr. arXiv*, vol. 1809.10243, 2018.

[102] D. Nunnari, F., Kadir, M.A., Sonntag, "On the overlap between grad-cam saliency maps and explainable visual features in skin cancer images," *Int. Cross-Domain Conf. Mach. Learn. Knowl. Extr. Springer*, pp. 241-253.

[103] I. González-Díaz, "DermaKNet: Incorporating the Knowledge of Dermatologists to Convolutional Neural Networks for Skin Lesion Diagnosis," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 2, pp. 547–559, 2019, doi: 10.1109/JBHI.2018.2806962.

[104] R. R. E. Z. Chen, X. Dong, J. Wu, H. Jiang, X. Li, "Lesion attributes segmentation for melanoma detection with deep learning," *bioRxiv Prepr. bioRxiv*, vol. 381855, 2018.

[105] J. ;M. A. F.-T. G.-D. D.-M. and A. P. Lopez-Labraca, "Enriched dermoscopicstructure-

based cad system for melanoma diagnosis," *Multimed. Tools Appl.*, vol. 77, no. 10, pp. 12171–12202, 2018.

[106] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.

[107] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[108] K. He, "Deep Residual Learning for Image Recognition."

[109] K. K. dermoscopedia R. Braun, "Dermoscopic Structures," *[Online; accessed 21-January-2019]*., 2019.

[110] N. G. Marghoob, K. Liopyris, and N. Jaimes, "Dermoscopy: A review of the structures that facilitate melanoma detection," *J. Am. Osteopath. Assoc.*, vol. 119, no. 6, pp. 380–390, 2019, doi: 10.7556/jaoa.2019.067.

8 Annexures

Annex-I: Detail of the layers in proposed H-GoogleNet model

Model: "GoogLeNet"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 224, 224, 3)]	0	[]
conv2d (Conv2D)	(None, 109, 109, 64)	9472	['input_1[0][0]']
max_pooling2d (MaxPooling2D)	(None, 54, 54, 64)	0	['conv2d[0][0]']
conv2d_1 (Conv2D)	(None, 54, 54, 64)	4160	['max_pooling2d[0][0]']
conv2d_2 (Conv2D)	(None, 54, 54, 192)	110784	['conv2d_1[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, 26, 26, 192)	0	['conv2d_2[0][0]']
conv2d_4 (Conv2D)	(None, 26, 26, 96)	18528	['max_pooling2d_1[0][0]']
conv2d_6 (Conv2D)	(None, 26, 26, 16)	3088	['max_pooling2d_1[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, 26, 26, 192)	0	['max_pooling2d_1[0][0]']
conv2d_3 (Conv2D)	(None, 26, 26, 64)	12352	['max_pooling2d_1[0][0]']
conv2d_5 (Conv2D)	(None, 26, 26, 128)	110720	['conv2d_4[0][0]']
conv2d_7 (Conv2D)	(None, 26, 26, 32)	12832	['conv2d_6[0][0]']

```
conv2d_8 (Conv2D)              (None, 26, 26, 32)   6176       ['max_pooling2d_2[0][0]']

concatenate (Concatenate)      (None, 26, 26, 256)  0          ['conv2d_3[0][0]',
                                                                'conv2d_5[0][0]',
                                                                'conv2d_7[0][0]',
                                                                'conv2d_8[0][0]']

conv2d_10 (Conv2D)             (None, 26, 26, 128)  32896      ['concatenate[0][0]']

conv2d_12 (Conv2D)             (None, 26, 26, 32)   8224       ['concatenate[0][0]']

max_pooling2d_3 (MaxPooling2D) (None, 26, 25, 256)  0          ['concatenate[0][0]']

conv2d_9 (Conv2D)              (None, 26, 26, 128)  32896      ['concatenate[0][0]']

conv2d_11 (Conv2D)             (None, 26, 26, 192)  221376     ['conv2d_10[0][0]']

conv2d_13 (Conv2D)             (None, 26, 26, 96)   76896      ['conv2d_12[0][0]']

conv2d_14 (Conv2D)             (None, 26, 26, 64)   16448      ['max_pooling2d_3[0][0]']

concatenate_1 (Concatenate)    (None, 26, 26, 480)  0          ['conv2d_9[0][0]',
                                                                'conv2d_11[0][0]',
                                                                'conv2d_13[0][0]',
                                                                'conv2d_14[0][0]']

max_pooling2d_4 (MaxPooling2D) (None, 12, 12, 480)  0          ['concatenate_1[0][0]']

conv2d_16 (Conv2D)             (None, 12, 12, 96)   46176      ['max_pooling2d_4[0][0]']

conv2d_18 (Conv2D)             (None, 12, 12, 16)   7696       ['max_pooling2d_4[0][0]']

max_pooling2d_5 (MaxPooling2D) (None, 12, 12, 480)  0          ['max_pooling2d_4[0][0]']

conv2d_15 (Conv2D)             (None, 12, 12, 192)  92352      ['max_pooling2d_4[0][0]']

conv2d_17 (Conv2D)             (None, 12, 12, 208)  179920     ['conv2d_16[0][0]']

conv2d_19 (Conv2D)             (None, 12, 12, 48)   19248      ['conv2d_18[0][0]']

conv2d_20 (Conv2D)             (None, 12, 12, 64)   30784      ['max_pooling2d_5[0][0]']

concatenate_2 (Concatenate)    (None, 12, 12, 512)  0          ['conv2d_15[0][0]',
                                                                'conv2d_17[0][0]',
                                                                'conv2d_19[0][0]',
                                                                'conv2d_20[0][0]']

conv2d_22 (Conv2D)             (None, 12, 12, 96)   49248      ['concatenate_2[0][0]']

conv2d_24 (Conv2D)             (None, 12, 12, 16)   8208       ['concatenate_2[0][0]']

max_pooling2d_6 (MaxPooling2D) (None, 12, 12, 512)  0          ['concatenate_2[0][0]']

conv2d_21 (Conv2D)             (None, 12, 12, 192)  98496      ['concatenate_2[0][0]']

conv2d_23 (Conv2D)             (None, 12, 12, 208)  179920     ['conv2d_22[0][0]']

conv2d_25 (Conv2D)             (None, 12, 12, 48)   19248      ['conv2d_24[0][0]']
```

```
conv2d_26 (Conv2D)              (None, 12, 12, 64)   32832      ['max_pooling2d_6[0][0]']

concatenate_3 (Concatenate)     (None, 12, 12, 512)  0          ['conv2d_21[0][0]',
                                                                 'conv2d_23[0][0]',
                                                                 'conv2d_25[0][0]',
                                                                 'conv2d_26[0][0]']

conv2d_28 (Conv2D)              (None, 12, 12, 96)   49248      ['concatenate_3[0][0]']

conv2d_30 (Conv2D)              (None, 12, 12, 16)   8208       ['concatenate_3[0][0]']

max_pooling2d_7 (MaxPooling2D)  (None, 12, 12, 512)  0          ['concatenate_3[0][0]']

conv2d_27 (Conv2D)              (None, 12, 12, 192)  98496      ['concatenate_3[0][0]']

conv2d_29 (Conv2D)              (None, 12, 12, 208)  179920     ['conv2d_28[0][0]']

conv2d_31 (Conv2D)              (None, 12, 12, 48)   19248      ['conv2d_30[0][0]']

conv2d_32 (Conv2D)              (None, 12, 12, 64)   32832      ['max_pooling2d_7[0][0]']

concatenate_4 (Concatenate)     (None, 12, 12, 512)  0          ['conv2d_27[0][0]',
                                                                 'conv2d_29[0][0]',
                                                                 'conv2d_31[0][0]',
                                                                 'conv2d_32[0][0]']

conv2d_34 (Conv2D)              (None, 12, 12, 96)   49248      ['concatenate_4[0][0]']

conv2d_36 (Conv2D)              (None, 12, 12, 16)   8208       ['concatenate_4[0][0]']

max_pooling2d_8 (MaxPooling2D)  (None, 12, 12, 512)  0          ['concatenate_4[0][0]']

conv2d_33 (Conv2D)              (None, 12, 12, 192)  98496      ['concatenate_4[0][0]']

conv2d_35 (Conv2D)              (None, 12, 12, 208)  179920     ['conv2d_34[0][0]']

conv2d_37 (Conv2D)              (None, 12, 12, 48)   19248      ['conv2d_36[0][0]']

conv2d_38 (Conv2D)              (None, 12, 12, 64)   32832      ['max_pooling2d_8[0][0]']

first_level_layer (Concatenate  (None, 12, 12, 512)  0          ['conv2d_33[0][0]',
)                                                                'conv2d_35[0][0]',
                                                                 'conv2d_37[0][0]',
                                                                 'conv2d_38[0][0]']

conv2d_41 (Conv2D)              (None, 12, 12, 112)  57456      ['first_level_layer[0][0]']

conv2d_43 (Conv2D)              (None, 12, 12, 24)   12312      ['first_level_layer[0][0]']

max_pooling2d_9 (MaxPooling2D)  (None, 12, 12, 512)  0          ['first_level_layer[0][0]']

conv2d_40 (Conv2D)              (None, 12, 12, 160)  82080      ['first_level_layer[0][0]']

conv2d_42 (Conv2D)              (None, 12, 12, 224)  226016     ['conv2d_41[0][0]']

conv2d_44 (Conv2D)              (None, 12, 12, 64)   38464      ['conv2d_43[0][0]']

conv2d_45 (Conv2D)              (None, 12, 12, 64)   32832      ['max_pooling2d_9[0][0]']
```

```
concatenate_5 (Concatenate)    (None, 12, 12, 512)  0           ['conv2d_40[0][0]',
                                                                 'conv2d_42[0][0]',
                                                                 'conv2d_44[0][0]',
                                                                 'conv2d_45[0][0]']

conv2d_47 (Conv2D)             (None, 12, 12, 128)  65664       ['concatenate_5[0][0]']

conv2d_49 (Conv2D)             (None, 12, 12, 24)   12312       ['concatenate_5[0][0]']

max_pooling2d_10 (MaxPooling2D  (None, 12, 12, 512)  0          ['concatenate_5[0][0]']
)

conv2d_46 (Conv2D)             (None, 12, 12, 128)  65664       ['concatenate_5[0][0]']

conv2d_48 (Conv2D)             (None, 12, 12, 256)  295168      ['conv2d_47[0][0]']

conv2d_50 (Conv2D)             (None, 12, 12, 64)   38464       ['conv2d_49[0][0]']

conv2d_51 (Conv2D)             (None, 12, 12, 64)   32832       ['max_pooling2d_10[0][0]']

concatenate_6 (Concatenate)    (None, 12, 12, 512)  0           ['conv2d_46[0][0]',
                                                                 'conv2d_48[0][0]',
                                                                 'conv2d_50[0][0]',
                                                                 'conv2d_51[0][0]']

conv2d_53 (Conv2D)             (None, 12, 12, 144)  73872       ['concatenate_6[0][0]']

conv2d_55 (Conv2D)             (None, 12, 12, 32)   16416       ['concatenate_6[0][0]']

max_pooling2d_11 (MaxPooling2D  (None, 12, 12, 512)  0          ['concatenate_6[0][0]']
)

conv2d_52 (Conv2D)             (None, 12, 12, 112)  57456       ['concatenate_6[0][0]']

conv2d_54 (Conv2D)             (None, 12, 12, 288)  373536      ['conv2d_53[0][0]']

conv2d_56 (Conv2D)             (None, 12, 12, 64)   51264       ['conv2d_55[0][0]']

conv2d_57 (Conv2D)             (None, 12, 12, 64)   32832       ['max_pooling2d_11[0][0]']

second_level_layer (Concatenat  (None, 12, 12, 528)  0          ['conv2d_52[0][0]',
e)                                                               'conv2d_54[0][0]',
                                                                 'conv2d_56[0][0]',
                                                                 'conv2d_57[0][0]']

conv2d_60 (Conv2D)             (None, 12, 12, 160)  84640       ['second_level_layer[0][0]']

conv2d_62 (Conv2D)             (None, 12, 12, 32)   16928       ['second_level_layer[0][0]']

max_pooling2d_12 (MaxPooling2D  (None, 12, 12, 528)  0          ['second_level_layer[0][0]']
)

conv2d_59 (Conv2D)             (None, 12, 12, 256)  135424      ['second_level_layer[0][0]']

conv2d_61 (Conv2D)             (None, 12, 12, 320)  461120      ['conv2d_60[0][0]']

conv2d_63 (Conv2D)             (None, 12, 12, 128)  102528      ['conv2d_62[0][0]']
```

```
conv2d_64 (Conv2D)              (None, 12, 12, 128)  67712       ['max_pooling2d_12[0][0]']

concatenate_7 (Concatenate)     (None, 12, 12, 832)  0           ['conv2d_59[0][0]',
                                                                  'conv2d_61[0][0]',
                                                                  'conv2d_63[0][0]',
                                                                  'conv2d_64[0][0]']

max_pooling2d_13 (MaxPooling2D  (None, 5, 5, 832)    0           ['concatenate_7[0][0]']
)

conv2d_66 (Conv2D)              (None, 5, 5, 160)    133280      ['max_pooling2d_13[0][0]']

conv2d_68 (Conv2D)              (None, 5, 5, 32)     26656       ['max_pooling2d_13[0][0]']

max_pooling2d_14 (MaxPooling2D  (None, 5, 5, 832)    0           ['max_pooling2d_13[0][0]']
)

conv2d_65 (Conv2D)              (None, 5, 5, 256)    213248      ['max_pooling2d_13[0][0]']

conv2d_67 (Conv2D)              (None, 5, 5, 320)    461120      ['conv2d_66[0][0]']

conv2d_69 (Conv2D)              (None, 5, 5, 128)    102528      ['conv2d_68[0][0]']

conv2d_70 (Conv2D)              (None, 5, 5, 128)    106624      ['max_pooling2d_14[0][0]']

concatenate_8 (Concatenate)     (None, 5, 5, 832)    0           ['conv2d_65[0][0]',
                                                                  'conv2d_67[0][0]',
                                                                  'conv2d_69[0][0]',
                                                                  'conv2d_70[0][0]']

average_pooling2d (AveragePool  (None, 3, 3, 512)    0           ['first_level_layer[0][0]']
ing2D)

average_pooling2d_1 (AveragePo  (None, 3, 3, 528)    0           ['second_level_layer[0][0]']
oling2D)

conv2d_72 (Conv2D)              (None, 5, 5, 192)    159936      ['concatenate_8[0][0]']

conv2d_74 (Conv2D)              (None, 5, 5, 48)     39984       ['concatenate_8[0][0]']

max_pooling2d_15 (MaxPooling2D  (None, 5, 5, 832)    0           ['concatenate_8[0][0]']
)

conv2d_39 (Conv2D)              (None, 3, 3, 128)    65664       ['average_pooling2d[0][0]']

conv2d_58 (Conv2D)              (None, 3, 3, 128)    67712       ['average_pooling2d_1[0][0]']

conv2d_71 (Conv2D)              (None, 5, 5, 384)    319872      ['concatenate_8[0][0]']

conv2d_73 (Conv2D)              (None, 5, 5, 384)    663936      ['conv2d_72[0][0]']

conv2d_75 (Conv2D)              (None, 5, 5, 128)    153728      ['conv2d_74[0][0]']

conv2d_76 (Conv2D)              (None, 5, 5, 128)    106624      ['max_pooling2d_15[0][0]']

flatten (Flatten)               (None, 1152)         0           ['conv2d_39[0][0]']

flatten_1 (Flatten)             (None, 1152)         0           ['conv2d_58[0][0]']
```

```
third_level_layer (Concatenate  (None, 5, 5, 1024)  0           ['conv2d_71[0][0]',
 )                                                               'conv2d_73[0][0]',
                                                                 'conv2d_75[0][0]',
                                                                 'conv2d_76[0][0]']

dense (Dense)                  (None, 1024)         1180672     ['flatten[0][0]']

dense_1 (Dense)                (None, 1024)         1180672     ['flatten_1[0][0]']

GAPL (GlobalAveragePooling2D)  (None, 1024)         0           ['third_level_layer[0][0]']

dropout (Dropout)              (None, 1024)         0           ['dense[0][0]']

dropout_1 (Dropout)            (None, 1024)         0           ['dense_1[0][0]']

dropout_2 (Dropout)            (None, 1024)         0           ['GAPL[0][0]']

level1 (Dense)                 (None, 1)            1025        ['dropout[0][0]']

level2 (Dense)                 (None, 4)            4100        ['dropout_1[0][0]']

output (Dense)                 (None, 7)            7175        ['dropout_2[0][0]']

==================================================================================================
Total params: 9,644,428
Trainable params: 9,644,428
Non-trainable params: 0
```